Improved Forest Signal Detection for Space-Borne Photon-Counting LiDAR Using Automatic Machine Learning

Bo Zhang, Li Zhang, Yong Pang, *Member, IEEE*, Peter North, Min Yan, Hongge Ren, Linlin Ruan, Zhenyu Yang, and Bowei Chen

***Abstract*—NASA's (National Aeronautics and Space Administration) ICESat-2 with a Photon Counting LiDAR (Light Detection And Ranging) Sensor sensitively detects signal photons at high speed with an advanced detection system called the Advanced Topographic Laser Altimeter System (ATLAS). However, the sensor also extracts a large amount of background photon noise coming from the atmosphere, ground, sun, or other radiation. This condition is particularly evident in forest areas. This study proposes an automatic machine learning approach to utilize data for forestry applications to improve data availability compared to NASA's official product.**

Our method uses only a very limited number (10%) of sample points for training, ensuring operational efficiency and training accuracy. We conclude that the integrated learning performance generally outperforms single models, and the mean F1 score of all tests is approximately 0.9. The mean F1 score of the Stacked Ensembles model is 0.957 ahead of the other models. The top three variables used in training models are kNNdist5, kNNdist10, and h. These three variables could explain 51.6% of the components of the models. Over the regions tested, the proposed method could improve the proportion of signals correctly identified by 6.4%, 12.2%, 2.7%, 9.3%, and 1.4% in five datasets. The model performs better in low signal-to-noise (SNR) datasets less than 7.5. The method would be largely unaffected by differences in topography, noise distribution, and SNR. The classifiers could correct misclassified labels in ATL08 products and show good stability in different conditions.

***Index Terms*—Automated machine learning, ICESat-2/ATLAS, photon point cloud filtering, space-borne light detection and ranging (LiDAR).**

Manuscript received 23 March 2023; revised 19 May 2023; accepted 12 June 2023. Date of publication 29 June 2023; date of current version 23 November 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFE0200800, and in part by the National Natural Science Foundation of China under Grant 42001361. *(Corresponding author: Bowei Chen.)*

Bo Zhang, Hongge Ren, and Linlin Ruan are with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, also with the International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: zhangbo203@mails.ucas.ac.cn; renhg@aircas.ac.cn; ruanlinlin20@mails.ucas.ac.cn).

Li Zhang, Min Yan, and Bowei Chen are with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and also with the International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China (e-mail: zhangli@aircas.ac.cn; yanmin@aircas.ac.cn; chenbw@aircas.ac.cn).

Yong Pang is with the Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing 100091, China (e-mail: pangy@ifrit.ac.cn).

Peter North is with the Global Environmental Modelling and Earth Observation (GEMEO), Department of Geography, Swansea University, SA20LS Swansea, U.K. (e-mail: p.r.j.north@swansea.ac.uk).

Zhenyu Yang is with the School of Marine Technology and Geomatics, Jiangsu Ocean University, Lianyungang 222005, China (e-mail: yzyloloo@gmail.com).

Digital Object Identifier 10.1109/JSTARS.2023.3290680

I. Introduction

NASA launched the the Ice, Cloud, and Land Elevation Satellite (ICESat) series satellites to track changes in the terrain of glaciers, sea ice, forests, and other areas [1]. The first generation of ICESat onboard the Geoscience Laser Altimeter System (GLAS) LiDAR system provided a large amount of global ground elevation and 3-D information from 2003 to 2009. The GLAS products were relied upon to accurately estimate key forest parameters such as forest max height, mean height [2], [3], and biomass [4], [5], and showed the great potential for vegetation and ecosystem science of a global space-borne light detection and ranging (LiDAR) system for the first time [6].

As a successor to the ICESat satellite mission, NASA launched the ICESat-2 satellite in September 2018 [7]. The second-generation satellite is equipped with the advanced micropulse multibeam photon-counting laser altimeter (Advanced Terrain Laser Altimeter System (ATLAS)) replacing the full-waveform LiDAR system. The new detection system used a more sensitive single-photon detector to extract higher-accuracy 3-D surface information at 10 kHz high-frequency pulse [8]. ICESat-2 uses three pairs of low-energy 532 nm laser beams (3.2 km separating the pairs, 90 m within pairs). Each beam produces an approximately 17 m diameter footprint at every 0.7 m along-track sampling interval [9], [10], [11].

NASA released the official datasets with photon labels and further extracted forest canopy labels. The ICESat-2 Level-2 A products (ATL03) determine the geodetic location (i.e. the latitude, longitude, and height) of the ground points with the flight times, the observatory position, and attitude. The ATL03 products are used by higher-level (Level-3 A) surface-specific products to determine glacier and ice sheet height, sea ice freeboard, vegetation canopy height, ocean surface topography, and inland water bodies height [12]. The new photon counting system provides individual photon measurements and achieves higher laser repetition rates for improving spatial coverage. However, this operating mode is sensitive to solar background

noise, making noise removal a necessary preprocess before application [13].

Several recent research projects have been conducted on noise removal in photon counting LiDAR. The following algorithms attempted photon classifications by simulated ICESat-2 type data and explored solutions from statistical analysis methods. The Canny edge detection process would provide 97.4% success in identifying signal photons (37.2% false alarms) producing a sufficient evaluation of the surface location in fairly quick processing times [14]. Automatic statistical analysis results have shown it to be highly effective against low signal-to-noise (SNR) datasets such as those resulting from high repetition rate, low pulse energy laser, and single-photon sensitive detectors [15]. A mathematical algorithm is developed using spatial statistics and discrete mathematics concepts. Validation for instrument design shows that ground and canopy elevation and hence, canopy height can be expected to be observable with high accuracy (93.01%–99.57% correctly selected points for a beam with the expected return of 0.93 mean signals per shot (msp), and 72.85%–98.68% for 0.48 msp) [16]. The classic geodesic active contours method demonstrates that this technique can identify the potential signal photons effectively with an error rate of less than 4.2%. The proposed approach is appropriate for the present airborne simulated data with high accuracy for a flat surface with dense canopy [17].

With the development of denoising methods, researchers have tried computer graphics-based methods. An adaptive density model is shown to detect the ground surface and vegetation canopy with better performance for smoother surfaces and lower noise rate conditions in laser altimeter photon-counting data, although the performance evaluation metric F-measure does not vary significantly over a range of noise rates (0.5–5 MHz) [18]. The PSODBSCAN algorithm performs better than the localized statistics algorithm (the mean F value of 0.9759 for the PSODBSCAN, and the mean F value of 0.6978 for the localized statistical algorithm) in the forest region, and does not need manual adjustment of parameters for different test data [19]. Results of noise filtering show that the multilevel filtering process design effectively identifies the background noise and preserves signal photons in the raw data, although validation results obtained over densely vegetated conditions are not impressive [20]. Filtering out the noise photons based on localized statistical analysis by an effective noise removal algorithm was shown to effectively reduce the edge effect and the influence of inconsistent noise photon density [21], [22]. A noise filtering method based on local outlier factor (LOF) extracting the ground and canopy surface shows that methods can detect the potential signal photons effectively from a quite high noise rate environment in relatively rough terrain [23]. LOF modified with ellipse searching area for noise and signal photons shows that the approach has a good performance not only in lower noise rate with relatively flat terrain surface, but also works even for a quite high noise rate environment in relatively rough terrain, and the horizontal ellipse searching area gives the best result compared with the circle or vertical ellipse searching area [24].

Most of the above methods were studies of photon noise removal approaches and attempted to extract canopy and ground signals in forest areas. These existing methods are mostly based on unsupervised methods, whereas supervised approaches have been barely investigated; only one study explored the use of random forest to detect forest signal photons [25]. The simulated data used in this study is relativity ideal compared to ATLAS and need to be tested based on the actual data. The study did not take into account multiple influence factors in different detection environments. The supervised method has the following characteristics.

1) Convenient calculation with a limited number of labels and automatic hyperparameters.
2) The transferability of the model in similar conditions [25].
3) The applicability for different scale datasets [26].
4) The interpretability of results with different category labels or dataset information.

NASA has released some ICESat-2 photon products including forest signal labels, the algorithm employed a DRAGANN method filtering noise, and an iterative filtering method to identify both the ground and top of canopy surfaces [27]. However, due to the complexity of the terrain and atmospheric conditions, some mislabeled photons are prevalent in forest areas on a large scale [28]. We proposed here a supervised classification method for correcting official product labels.

Conventional machine learning techniques require experts to explore the large design space, perform artificial model compression, and make tradeoffs between model size, speed, and accuracy, which are usually suboptimal and laborious. On the other hand, every machine learning system has hyperparameters, which also require extensive testing by experts.

Recently, automatic machine learning (AutoML) techniques have been developed. Automatic hyperparameters are the most fundamental task. The AutoML technique aims to adjust various models at once and allow the models to achieve state-of-the-art performance without any manual intervention ensuring the fairness of model comparisons [29]. Meanwhile, AutoML with the complete evaluation system could sample the model space efficiently improving the quality of model compression. The automatic hyperparameter selection and model compression also improve the efficiency and reproducibility of the research [30]. Therefore, in this article, we choose to conduct experiments in a continuous forest area, extract forest photon features, and denoise photon-counting LiDAR data based on the AutoML method. Compared to the official NASA products, we would like to improve the misclassification of forest photons in forest areas, and further improve the availability of data.

II. Study Area

The study was conducted at Saihanba National Forest Farm in Hebei Province, China (42°19′ – 42°33′ N, 117°07′ – 117°28′ E). Fig. 1 shows a terrain schematic with an elevation range between about 1400–1900 m. The source of the background map is from the Google Earth platform. This figure is also overlaid with the forest farm boundary and the ground trajectory of the six sets of ICESat-2 laser detectors. The Saihanba National Forest Farm is the largest plantation forest in China, located in a mountainous area on the southeastern edge of the Inner Mongolia Plateau.

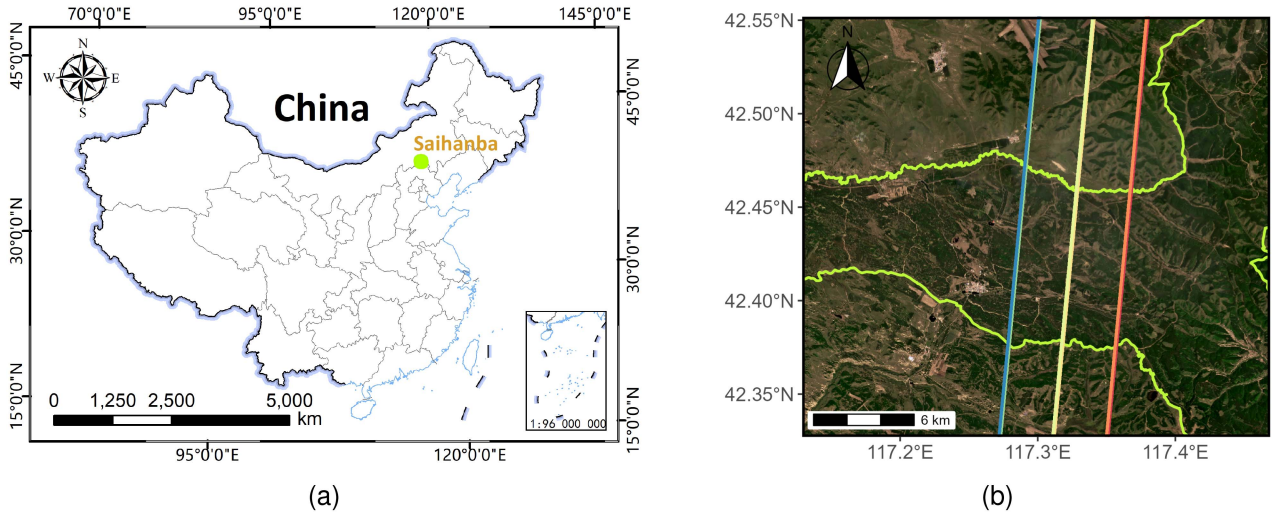


Fig. 1. Location and terrain of the study area. (a) The relative position of the study area on the map of China. (b) The terrain of the study area with the forest farm boundary and the ICESat-2 ground trajectory.

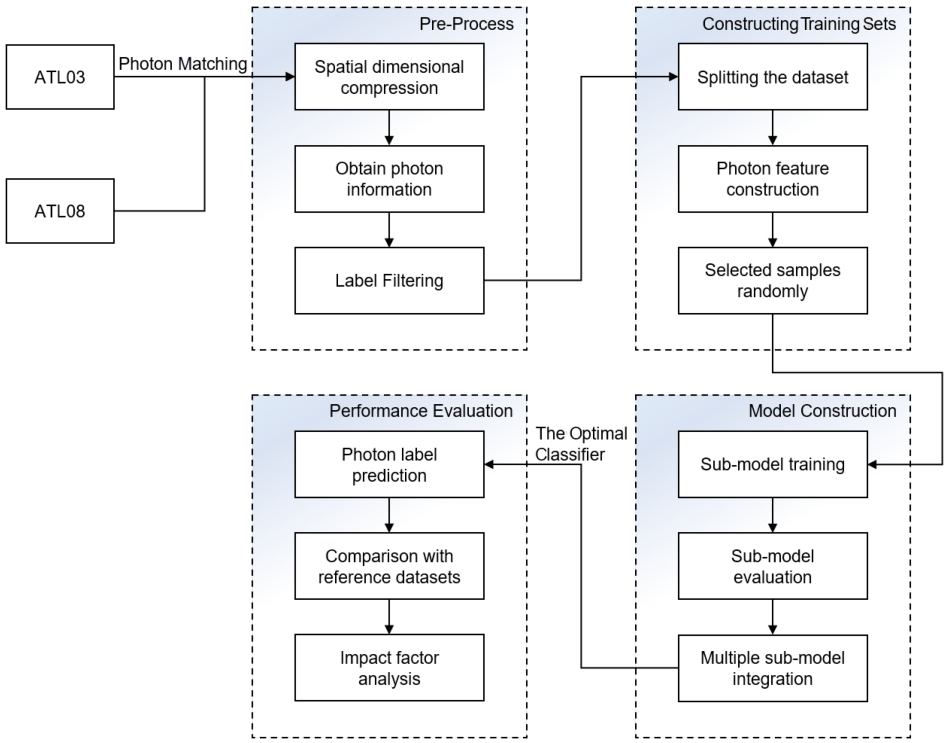


Fig. 2. Flowchart of our methods.

The study area belongs to the typical temperate forest-steppe ecotone with a semiarid and semihumid climate [31]. The study site is characterized by long, cold winters and short springs and autumns with about –1.4 °C annual average temperature. The temperature extremes are between –43.2 °C and 33.4 °C. The annual average rainfall and evaporation are 453.6 and 1388 mm [32]. The dominant tree species include *Pinus sylvestris* var. *mongolica*, *Larix gmelinii*, and *Picea meyeri*, as well as scattered natural secondary deciduous broad-leaved forests of *Betula platyphylla*, and *Ulmus pumila* woodland [33]. The topographic condition and forest environment of Saihanba are representative, while its abundant forest structure and phenological characteristics provide an ideal test site for the observation and model coupling of complex topographic parameters [32].

III. Materials and Methods

The approach for photon classification based on automated machine learning is shown in Fig. 2. First, we clipped the data to roughly match the forest region and paired ATL08 with

TABLE I
ICESAT-2 PRODUCTS USED IN THIS ARTICLE

Datasets	Season	Day/Night	Channel	SNR
20190108154038_01710206_005_01.h5	Winter	Night	Gt1l, Gt2l, Gt3l	4.47, 7.12, 6.72
			Gt1r, Gt2r, Gt3r	9.10, 9.98, 6.76
20201005091904_01710906_005_01.h5	Autumn	Day	Gt1l, Gt2l, Gt3l	2.38, 2.64, 3.22
			Gt1r, Gt2r, Gt3r	0.92, 1.01, 1.09
20191206113615_01711406_005_01.h5	Winter	Night	Gt1l, Gt2l, Gt3l	10.39, 10.37, 11.45
			Gt1r, Gt2r, Gt3r	10.86, 7.96, 8.99
20220102113835_01711406_005_01.h5	Winter	Night	Gt1l, Gt2l, Gt3l	6.02, 5.89, 4.87
			Gt1r, Gt2r, Gt3r	8.23, 8.17, 7.46
20220302203452_10781402_005_01.h5	Spring	Night	Gt1l, Gt2l, Gt3l	7.72, 6.17, 6.74
			Gt1r, Gt2r, Gt3r	9.14, 7.24, 8.97

Note: Only month and day will be presented for subsequent references, such as 0108.

ATL03 photon labels. Then we segmented the dataset by every 10000 m range and randomly selected 10% signal and noise photons in the range. Next, 18 features were constructed that could describe the statistical properties of the photons. We then considered feature relevance for feature reductions and ranked them according to explanatory. The top ten features were used to constructing models. Next, the models were ranked according to performance, and the best-performing model was used to distinguish signal and noise photons. Finally, we evaluated the performance of our method based on AutoML to the official products for different datasets.

A. ICESat-2 Data and Preprocessing

We combined the ICESat-2 ATL03 and ATL08 products to conduct photon classification experiments in the Saihanba Forest Farm. To demonstrate the stability of the method for different terrains, we selected five typical datasets. Table I shows the Season, Day, or Night, and SNR of the datasets which were downloaded from the American National Snow and Ice Data Center. The parameter of "confidence" is provided to classify each photon as being either a likely signal or noise [34]. With high confidence, the photon is more likely a signal. In the experiment, we considered photons with high confidence as signals and the rest as noise. SNR is defined as the ratio of signal photons to the noise photons in every channel. The ICESat-2 Level-2 ATL03 product provides the time, latitude, longitude, and ellipsoid height for each photon [35]. The ICESat-2 Level-3 A ATL08 product estimates terrain height, canopy height, and canopy cover at 100 meters of fixed-length steps along the ground track.

First, due to positioning bias when downloading products from the official web, data cropping was necessary according to the Saihanba Forest Farm shapefile. Then, we converted the ATL08 labels of high and median confidence levels to signal and other photons to noise. Next, the ATL08 photon tags were traced back to add the latitude, longitude, and ellipsoid height information from ATL03. Afterward, the latitude and longitude were downscaled into ICESat-2 satellite ground track distance for facilitating data processing and visualization. Finally, we randomly selected the signal and noise photons from the dataset to construct the training set. Fig. 3 shows an example of the randomly extracted 10% photons in proportion to signal and noise from every 10000 m along-track distance.

B. AutoML Platform and Models Selection

We used the H2O platform in this experiment. H2O is an open-source, in-memory, distributed, fast, and scalable machine learning and predictive analytics platform [36]. It allows users to build machine learning models on big data [37], [38]. The platform was selected for the following features.

1) Includes the most widely utilized algorithms for statistics and machine learning.
2) Models interpretation capabilities to support regression and classification tasks.
3) Automatic feature engineering, model validation, tuning, selection, and deployment.
4) Automatic visualization.
5) Allowing user interaction directly for machine learning operations.

To test the performance of AutoML, six typical machine learning models were selected for experimentation; the selected models are shown in Fig. 4. They are as follows.

1) Distributed Random Forest (DRF) generates a forest of classification or regression trees. Each of these trees is a weak learner built on a subset of rows and columns. Both classification and regression take the average prediction over all of their trees to make a final prediction.
2) Gradient boosting machine (GBM) is a forward learning ensemble method. The guiding heuristic is that good predictive results can be obtained through increasingly refined approximations.
3) Generalized linear models (GLM) estimate regression models for outcomes following exponential distributions. In addition to the Gaussian (i.e., normal) distribution, these include Poisson, binomial, and gamma distributions.
4) Stacked ensembles (SE) methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms.

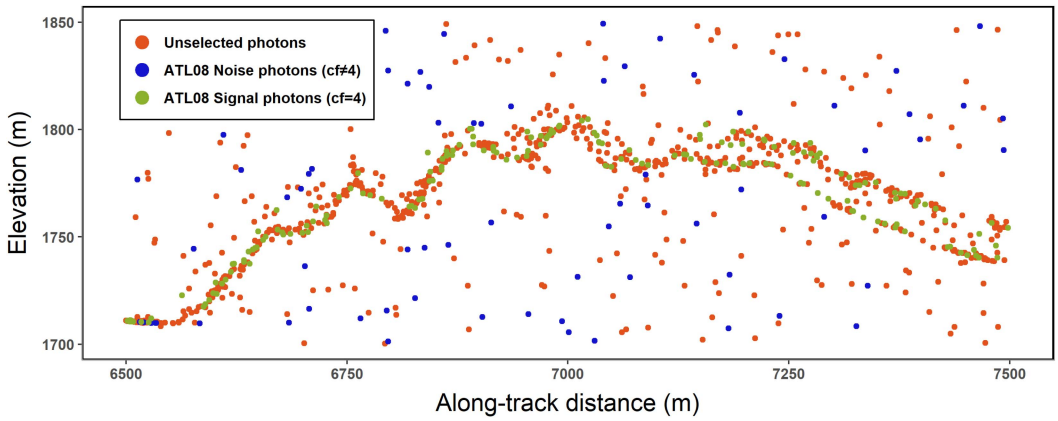


Fig. 3. Result of extracting photon labels randomly for training models. The red, blue, and green labels stand for the selected signal photons with high confidence, selected noise photons, and the rest of the photons.

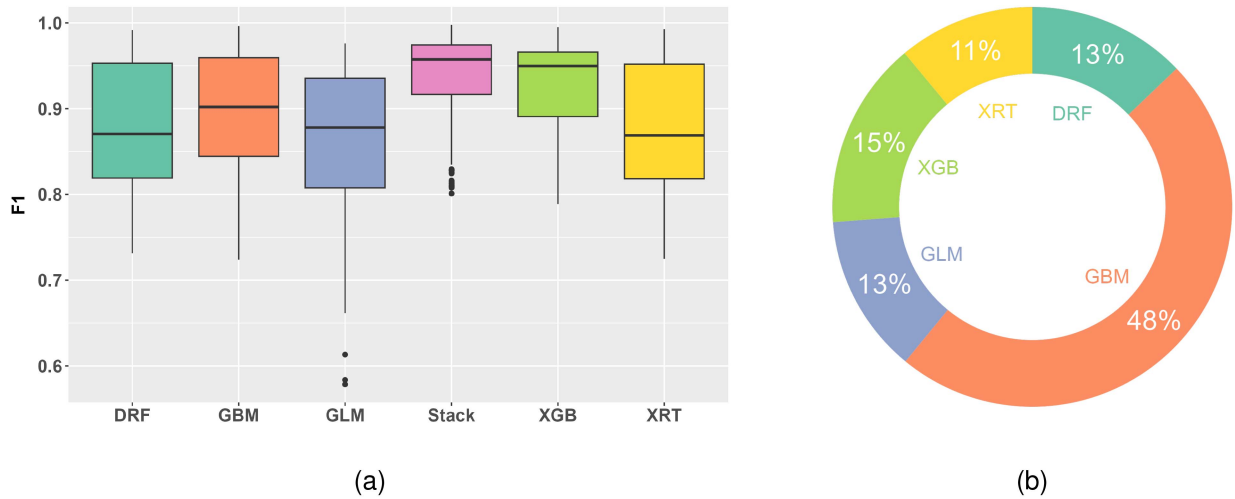


Fig. 4. Performance of models trained by AutoML. (a) The F1 score of models involved in the model-building process of the five datasets. The dark green, orange, blue, pink, green, and yellow boxes stand for the F1 score distribution of the DRF, GBM, GLM, SE, XGB, and XRT models, respectively. (b) The preference distribution of the best performance SE model in submodel combination from the model building process of the five datasets. The numbers represent the frequency percentage of submodels combined with the SE model in the training of the five datasets. The colors correspond to the settings in the above content.

5) eXtreme Gradient Boosting (XGB) refers to the ensemble learning technique of building many models sequentially, with each new model attempting to correct for the deficiencies in the previous model. In tree boosting, each new model that is added to the ensemble is a decision tree.
6) In XRT, randomness goes one step further in the way that splits are computed.

Thresholds are drawn at random for each candidate feature, and the best of these randomly generated thresholds is picked as the splitting rule. We limited the maximum number of models to prevent over-saturation of model training ($max\_models = 20$) and increased performance with the AUC (area under the ROC curve) score by restricting the number of models ($stopping\_metric = auc$).

C. Feature Extraction and Selection

Table II shows the 18 photon features covered in this article, including the height, along-track distance, and kNNdist-N, the difference between the height of a photon and mean, median, percentiles, kurtosis, skewness, standard deviation, variance, minimum height, maximum height, height range, mean absolute deviation, coefficient variation, interquartile range, and canopy relief ratio of all the photons at every 10 m window. N in kNNdist-N means extracting the nearest N points around the sample point. The photon features are defined as (1)

$$F_{m,i} = y_{m,i} - Z_m(x_m, y_m) \tag{1}$$

where $F_{m,i}$ represents the feature calculated for the ith photon in the mth 10 m window, x and y represent the along-track distance and photon height within the window, respectively. The function $Z_m(x_m, y_m)$ is used to calculate the statistical metrics (mean, median, percentiles, kurtosis, skewness, standard deviation, variance, minimum height, maximum height, height range, mean absolute deviation, coefficient variation, interquartile range, and canopy relief ratio) of all photons within the 10 m

TABLE II
DESCRIPTION OF FEATURES FOR TRAINING THE PHOTON-COUNTING LIDAR DATA CLASSIFIER

Num	Feature	Description
1	h	The height of a photon
2	dist	The along-track distance of a photon
3	dist.mean	The difference between the height of a photon and the mean in the surrounding 10 m window
4	dist.median	The difference between the height of a photon and the median in the surrounding 10 m window
5	dist.p(5~95)	The difference between the height of a photon and the (5~95) the percentile in the surrounding 10 m window
6	kNNdist5	The k-nearest neighbour's distance for photons (N = 5)
7	kNNdist10	The k-nearest neighbour's distance for photons (N = 10)
8	dist.sd	The difference between the height of a photon and the standard deviation in the surrounding 10 m window
9	dist.var	The difference between the height of a photon and the variance in the surrounding 10 m window
10	min	The difference between the height of a photon and the minimum height in the surrounding 10 m window
11	max	The difference between the height of a photon and the maximum height in the surrounding 10 m window
12	h.kurtosis	The difference between kurtosis of a photon and the mean value in the surrounding 10 m window
13	h.skewness	The difference between skewness of a photon and the mean value in the surrounding 10 m window
14	range	The difference between the height of a photon and the height range in the surrounding 10 m window
15	mad	The difference between the height of a photon and the mean absolute deviation in the surrounding 10 m window
16	cv	The difference between the height of a photon and the coefficient variation in the surrounding 10 m window
17	iqr	The difference between the height of a photon and the interquartile range in the surrounding 10 m window
18	crr	The difference between the height of a photon and the canopy relief ratio in the surrounding 10 m window

window, defined as the following equation:

$$cv_m = sd_m - \overline{y}_m. \tag{2}$$

Equation (2) describes the calculation of the coefficient variation of photons in the mth 10 m window

$$iqr_m = \theta_{m,75} - \theta_{m,25}. \tag{3}$$

Equation (3) describes the calculation of the interquartile range of photons in the mth 10 m window, where $\theta_{m,p}$ is the pth percentile of the aggregate in the mth 10 m window

$$crr_m = (\overline{y}_m - \min(y_m))/range_m. \tag{4}$$

Equation (4) describes the calculation of the canopy relief ratio of photons in the mth 10 m window, where $range_m$ is the height range of photons in the mth 10 m window.

IV. RESULTS

A. Model Establishment

Here, we developed the AutoML classification model using the training samples. The selected photon samples were sourced from the photon tags in the ATL08 products, and the reference datasets were also validated by visual interpretation. During the selection of photon samples, since the ATL08 datasets included some inaccurate labels, we corrected the labels by careful visual inspection, where necessary. Fig. 4(a) shows the F1 score distribution of the overall classifiers. The F1 score is the harmonic mean of precision and sensitivity, in which the precision is the fraction of true signal photons from all points identified as photons and sensitivity is the fraction of photons considered as signal photons that are correctly identified. F1 score box could be compared in detail mainly by minimal, lower quartile, mean, upper quartile, and maximal values (Table III). The optimal classifiers in the GBM, SE, and XGB methods were largely superior among the overall machine learning models. Although the DRF, GLM, and XRT models were slightly inferior, the models exhibited better stability in classification performance. The SE models combined the advantages of other models into better-performing classifiers. The best-performing classifiers are SE models used for photon predictions in each experiment all along.

TABLE III
MAINLY VALUE OF F1 SCORE BOXES FOR FIG. 4(A)

Model	Minimal	Lower	Mean	Upper	Maximal
DRF	0.731	0.819	0.870	0.953	0.992
GBM	0.724	0.844	0.902	0.959	0.996
GLM	0.578	0.807	0.878	0.936	0.976
SE	0.801	0.916	0.957	0.974	0.998
XGB	0.789	0.891	0.950	0.966	0.995
XRT	0.725	0.817	0.869	0.952	0.993

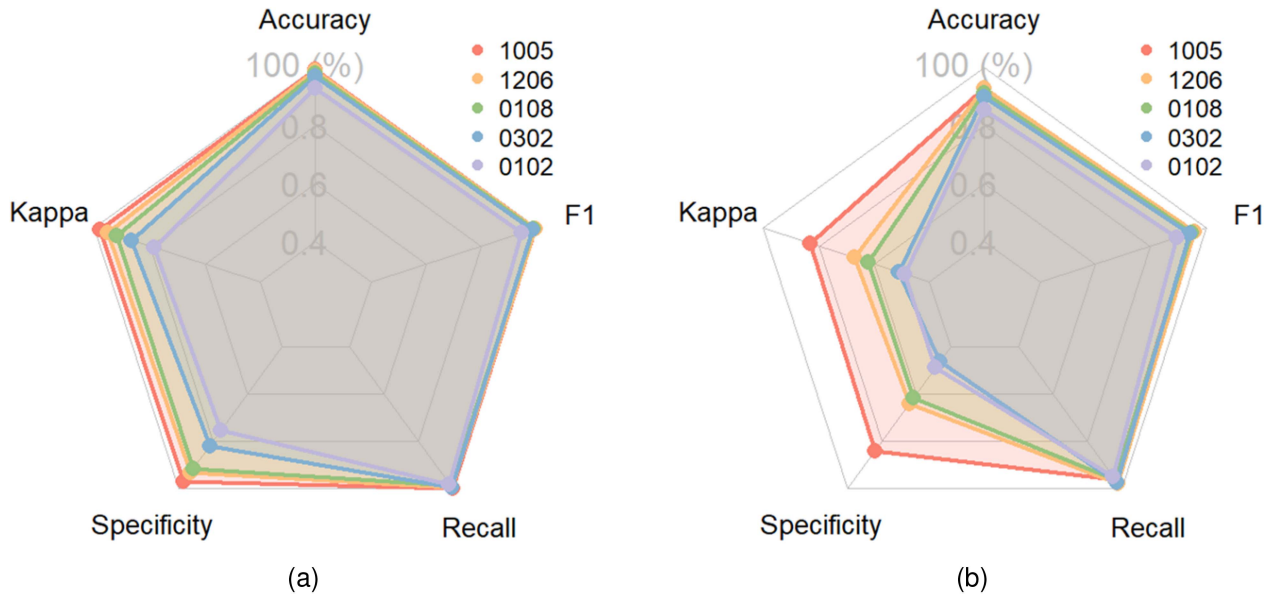


Fig. 5. Statistical indicators of model performance evaluation. (a) The classification performance of the training sets. (b) The classification performance of the test sets. The red, orange, green, blue, and purple frames stand for the indicators of five datasets in Table I.

Fig. 4(b) shows the preferences of the best-performing SE models when integrating submodels in all five datasets. The frequency percentage of DRF, GBM, GLM, XGB, and XRT models integrated into the SE model was 13%, 48%, 13%, 15%, and 11%, respectively. The SE models involved overall other models in the integration of submodels, but the models preferred to aggregate the GBM and XGB models with better performance. The integration strategy of AutoML ensures that the performance of SE models is optimal among the overall models available. Subsequently, the classification experiments were tested by the best-performing SE model.

B. Accuracy Assessments

We performed accuracy assessments of the classifiers and examined the results quantitatively. Photon samples were split according to 70% as the training set and another 30% as the test set. Fig. 5 shows the statistical indicator of performance evaluation: Accuracy, kappa coefficient, specificity, recall, and F1. Only month and day will be presented for subsequent references, such as 0108. The indicators were calculated from the confusion matrix for 0108, 1005, 1206, 0102, and 0302 products. They are defined as follows: Accuracy is the proportion of photons correctly identified as signal and noise in the total photons; kappa coefficient measures the prediction performance of the SE classifier; specificity is the proportion of photons identified as noise photons correctly; recall measures the proportion of signal photons that are classified correctly; and the F1 score. The values of indicators are shown in Table IV. The training results of the five datasets have good performance, while the test results have different degrees of degradation.

C. Variable Importance

Variable importance is calculated by the relative influence of each feature: whether that feature was selected to classify during the model-building process, and how much the squared error improved (decreased). Then, the squared error was normalized to be between 0 to 1. In the experiment, we considered the commonly used statistical features of photon point clouds. Fig. 6 presents the average variable importance of the top ten features used in training models for five datasets. The top three variables are kNNdist5, kNNdist10, and h. These three variables could explain 51.6% of the components of the models. The top ten variables provided 74.3% the explanation of the model components, namely kNNdist5, kNNdist10, h, dist, iqr, h.kurtosis, mad, crr, h.skewness, and dist.sd (Fig. 6). As the number of variables increases, the explanatory ability of the variables decreases gradually. While the number of variables is up to 20 (the top ten variables, range, min, max, cv, dist.p95, dist.mean, dist.p90, dist.p5, dist.median, and dist.p10), the explanatory model ability reaches about 90%. It could be considered that the whole model could be roughly explained. The remaining variables have only

TABLE IV
INDICATOR VALUES OF MODEL PERFORMANCE FOR FIG. 5

Type	Dataset	Accuracy	Kappa	Specificity	Recall	F1
Train	0108	0.979	0.918	0.916	0.989	0.988
	1005	0.992	0.979	0.972	0.999	0.995
	1206	0.989	0.951	0.934	0.998	0.994
	0102	0.930	0.784	0.755	0.981	0.946
	0302	0.969	0.865	0.820	0.995	0.982
Test	0108	0.911	0.622	0.615	0.961	0.948
	1005	0.929	0.829	0.841	0.964	0.943
	1206	0.929	0.672	0.642	0.977	0.960
	0102	0.854	0.490	0.488	0.948	0.893
	0302	0.897	0.509	0.461	0.974	0.942

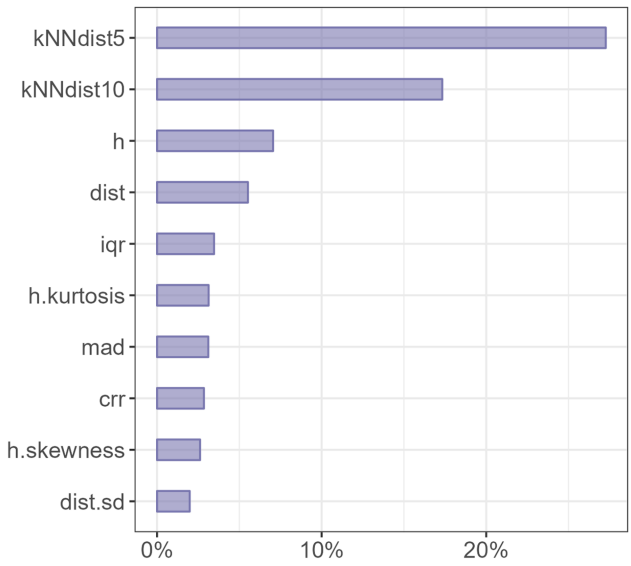


Fig. 6. Variable importance of classification models was calculated by the average of five datasets. Variables have been sorted according to their importance and each variable was defined in Table II. Only the top ten features were shown here.

less than 10% of the explanatory ability, of which most are the percentile variables such as dist.p35. We then selected only the top ten features for the experiment.

In the experiment process, we found that choosing the N value of the kNNdist algorithm would have a great impact on the results (Fig. 6). The kNN method is used to measure the aggregation of photons. Generally, the density of signal photons is much larger than the noise density. While the distance between adjacent signal photons is closer than the distance between adjacent noise photons, which makes the signal photon kNNdist much smaller than the noise kNNdist. Signal photons could be roughly distinguished from noise photons with this feature. We obtained some rules according to the selection of the N value. Fig. 6 also indicates that most of the model components could be explained by the kNNdist feature. This feature is adept at finding the aggregated photons and roughly filtering out the noise from the signal. In general, the aggregation of signal photons is much greater than that of noise photons. However, the condition of noise aggregation is inevitable in the measurement. Thus, when the N value of kNNdist is smaller, the classifier is more sensitive and would get a clearer profile of signal photons. And it would also be more likely to extract the local aggregated noise points. On the contrary, increasing the N value of kNNdist would achieve blurred signal photon contours and have a suppression effect on the aggregated noise in the meantime. Therefore, we make the combination of kNNdist5 and kNNdist10 for improving the comprehensive interpretation of the classifiers.

D. Analysis of Potential Factors on Classification

For analyzing the influence of different factors on the classification results, we considered the following factors in the experiment.

1) The SNR.
2) Seasons.
3) Time.
4) Channels.
5) Terrain.

We statistically estimated the difference between before and after classification regarding the 30 channels of the five datasets. Fig. 7(a) presents the relationship between discrepancies and SNR with different datasets and seasons. Only month and day will be presented for subsequent references, such as 0108. The lower quartile values of 0108, 1005, 1206, 0102, and 0302 are 6.52%, 6.30%, 4.25%, 9.47%, and 6.45%. The upper quartile values are 9.76%, 7.76%, 6.31%, 13.90%, and 8.56%.

Fig. 7(b) presents the relationship between discrepancies and SNR with different channels. The lower quartile values of gt3r, gt2r, gt1r, gt1l, gt2l, and gt3l are 6.30%, 7.51%, 7.24%, 5.99%, 6.52%, and 5.69%, respectively. The upper quartile values are 9.21%, 10.60%, 10.35%, 9.54%, 8.07%, and 6.45%.

To evaluate the stability of the classifier, we consider that the difference between the upper quartile value and the lower quartile value of the box represents the strength of the stability. The lower difference should be, the more stable the classifier. The impact of different potential factors will be developed in the subsequent discussion.

V. Discussion

A. Accuracy Assessment of Different Classifiers

From the result of the proposed methods, it is clear that the training and prediction accuracy could achieved a relatively satisfactory results in generally (Fig. 5). Although the signal and noise photons are not quantitatively balanced, the indicators were still substantial at least. Also, these models are well trained for both signal and noise features. It is not difficult to see from the labelled photons that many signal photons are mistaken as noise in the original dataset, so the ability to distinguish signal photons from noise is important, and the ability of the classifier to learn to distinguish signal photon features needs to be evaluated. Among the commonly used models, the XGB model is a relatively superior model in terms of stability and identification of signal photons. While the SE models are combined based on the commonly used models, the ability to distinguish signal photons is further improved. The classification experiments were tested by the best-performing SE model. The integration strategy of automatic machine learning could extract photonic features more efficiently and ensure the ability of the models to distinguish noise better from different aspects.

Also, we can see that the distribution of photons for the five datasets is shown in Fig. 8. Notably, the good representative indicators of 1005, 1206, and 0108 showed training was satisfactory apart from 0302 and 0102 generally. The difference may be caused by the noise near the ground that is evident in both 0302 and 0102. In addition, less outlier noise and the uneven elevation distribution of noise could also have some negative impact on the classification performance, just like the case in 0108 and 1206. On the other hand, the ability to predict noise photons deteriorated to different degrees in the overall test set.

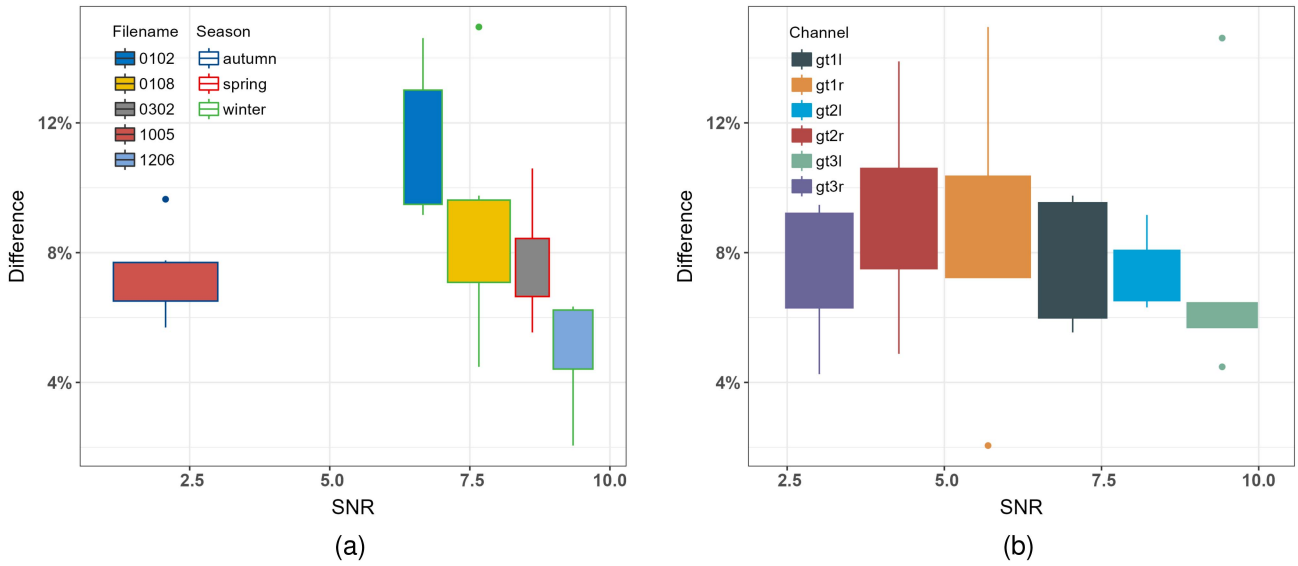


Fig. 7. Influence of different conditions on the classifications. (a) The influence of SNR with different seasons and datasets. The orange, purple, and blue frames stand for the Autumn, Spring, and Winter seasons, and the boxes filled with green, orange, purple, blue, and red colors represent the datasets of 0102, 0108, 0302, 1005, and 1206, respectively. (b) The influence of SNR with different channels. The boxes filled with orange, pink, purple, green, yellow, and blue colors indicate the channels of gt1l, gt1r, gt2l, gt2r, gt3l, and gt3r, respectively.

The signal photons identified as noise in the ATL08 dataset would affect this condition possibly. The condition might also be influenced by the diversity of noise patterns, which were much more difficult to identify than the signal photons, especially when the noise photons were adjacent to the signal photons. The performance of identifying noise photons in the other four datasets was more degraded compared to 1005. Furthermore, from visual interpretation, the lower SNR would have a more positive effect on identifying noise photons. Perhaps the performance of classification would be better for the day datasets.

In summary, the SE model could achieve better results than the traditional single model like XGB, with an improvement of about 0.7%∼9.1%. Also, we found that even with one single model, it could achieve an overall accuracy of 90%∼95% and much improve the correctly labelled photons, both visually and quantitatively. It is worth to mention that our classifiers could distinguish signal photons from noise photons with a very limited number (e.g., 10%) of samples from the AutoML method, which could be beneficial to implement similar transfer learning for a large study area in future works.

B. Discussion of Improvement From Official Products

In order to better evaluate the improvement from our proposed methods, we further investigated the performance under the following different observation conditions.

1) Different SNR from low to high.
2) Flat versus rugged terrains.
3) Signal mistakenly for noise photons.
4) Significant canopy structures.
5) Noise photons with symmetrical versus asymmetrical distribution around signal photons.

To better quantitively assess the impact of these factors, each individual granule fragment is trained separately and the classification results are derived. The datasets chosen for this experiment were with the typical distribution patterns. Fig. 8 shows the distribution of signal and noise photons. The left-hand and right-hand columns represent the ATL08 tags and the classification results predicted by the classifiers, respectively.

Generally, the prediction results not only largely contain the original signal labels, but it was also clear that the classifiers correctly identified signal photons that were misclassified as noise photons in ATL08 datasets. The classifiers played a role both in the presence of signal misidentification at the along-track and elevation orientation. The extensibility of the classifiers achieved a significant improvement in the utilization rate of data. It is worth noting that classifiers were more effective in low SNR than in high SNR datasets, this could be more advantageous in the daytime with a substantial amount of noise.

For the asymmetrical distribution of the dataset, there were no significant under- or misclassifications in the classification results. Furthermore, a large amount of dense noise near the ground had virtually no effect on the accuracy. The classifier showed good performance for signal photons in different terrains. Notably, noise photons bordering the ground were occasionally perceived as signal photons, and we believe that these noise photons were misclassified as the signal would not unduly affect large numbers of photon signals. The automatic identification or even visual interpretation of near-ground noise is difficult and remains a pressing and intractable problem to solve. In addition, the noise near the canopy has a similar distribution to the upper surface of the canopy. Which makes the identification of the upper canopy difficult. In addition, the top of the canopy position is normally not as readily identified as the ground, which still requires verification data for comparison.

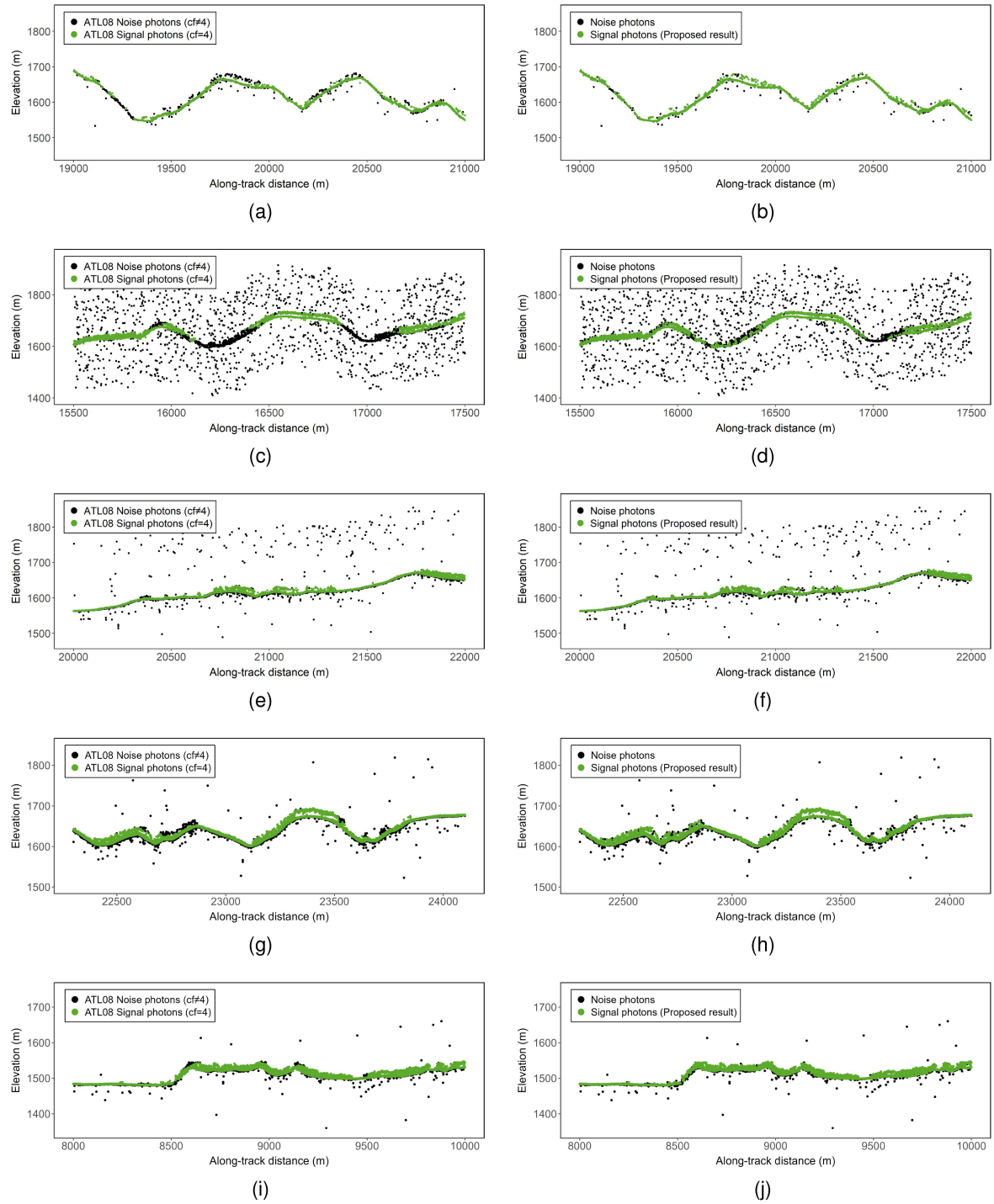


Fig. 8. Classification results and the ATL08 product labels. Each row stands for different datasets in the order of 0108, 1005, 1206, 0102, and 0302, respectively. The left-hand column represents the photon labels of the ATL08 products, and the right-hand column represents the classification results predicted by the classifiers. The green and black labels stand for signal photons and noise photons, respectively. (a) ATL08 labels of 0108. (b) Classification labels of 0108. (c) ATL08 labels of 1005. (d) Classification labels of 1005. (e) ATL08 labels of 1206. (f) Classification labels of 1206. (g) ATL08 labels of 0102. (h) Classification labels of 0102. (i) ATL08 labels of 0302. (j) Classification labels of 0302.

To quantify the ability to identify signal photons, we classified the datasets in Fig. 8 by visual interpretation. And we computed the proportion of signals correctly identified by ATL08 and proposed an approach based on the results of our visual interpretation, and improved the fraction after the proposed algorithm (Table V). Our method improved the proportion of signals correctly identified by 6.4%, 12.2%, 2.7%, 9.3%, and 1.4%, respectively. In particular, the classifiers demonstrated

TABLE V
PROPORTION OF PHOTON SIGNALS IDENTIFIED CORRECTLY

Datasets	ATL08	Proposed	Improved
0108	92.8%	99.2%	6.4%
1005	65.6%	77.8%	12.2%
1206	86.4%	89.1%	2.7%
0102	82.3%	91.6%	9.3%
0302	87.5%	88.9%	1.4%

satisfactory performance in low SNR datasets. In other rough terrain or asymmetrical noise distribution, our method still showed varying degrees of improvement.

In general, the classifiers not only corrected erroneous labels in the ATL08, but also showed good stability for different terrain and SNR. The classification results in forested areas are largely superior to the existing ATL08 products.

C. Assessment of Different Signal-to-Noise Ratio Levels

Based on Table I, the 1005 dataset collected contained more noise. The seasonal characteristics of SNR may not be apparent. The 1005 dataset was recorded by the ATLAS during the daytime, while the other four datasets were acquired after sunset. The different radiant energy from the sun would greatly affect the noise profile, and as expected the noise ratio of 0102, 0108, 0302, and 1206 are shown to be much lower than that of 1005. Compared to the other four datasets, the classifier has a more stable performance in 1005, which might be due to the lower SNR. The model performs better in low SNR datasets less than 7.5. We would expect the model to show better performance than ATL08 products for datasets collected during the day. The solar noise would provide a more balanced noise distribution. In the 1005 dataset, the difference tends to be uniform for the six channels generally. In contrast, in others, the noise in the high SNR datasets is irregularly distributed with a very limited number, which would make the classification more difficult. This may explain the large fluctuation in the classification.

D. Assessment of Different Terrain Conditions

The granules were selected to contain a variety of topographies. For flat terrain, the photon heights are all at the same level and the photon characteristics are similar across bins. For rugged terrain, the height of photons are more scattered, and the number of photons under each ground height is relatively reduced. There are many uncertainties in feature extraction.

Besides the 1005 granule, it can be assumed that all other granules are with similar high SNR. These granules can be compared at the same level. The 0108 and 0102 granules are more rugged compared to 1206 and 0302 granules, and the rugged terrain causes more misclassification labels. Our method corrected more mislabeling in rugged terrain, 6.4% and 9.3%, respectively. For granules in flat terrain, our method corrected only 2.7% and 1.4%. In addition, the rugged terrain reduced the stability of noise removal. The uncertainty was 3.5% and 2% for rugged and flat terrain, respectively.

The differences in the 1206 and 0302 datasets are significantly less than those in the 0108 and 0102 datasets. According to the visual interpretation of Fig. 8, the topographic undulation variation of 0108 and 0102 in the distribution of photon point clouds is complex against 1206 and 0302. The terrain complexity would harm the classification results. At the same time, the point density of the 0108 dataset is less than that of 1206. This makes extracting noise features more difficult and increases the instability of the classification.

E. Assessment of Different Channels

The left channels and right channels of ATLAS acquire signals in different modes. The ATLAS sensor is designed with higher energy in the left channel than that in the right channel. Therefore, the left channels have much higher SNR than the right channels. Fig. 7(b) showed the right channels would usually pick up more noise than the left channels in the same environment.

As it can be seen in Fig. 7(b), data collected from the left channels showed different performance than those from right channels, the difference varies less than that in the right channels. Also, the difference varies less in the left channels. The difference between the left channel and the right channel is approximately 2% and 3.5%, respectively. This seems to contradict the previous statement that a smaller signal-to-noise ratio would provide higher stability of variance changes. This could be caused by the statistical bias that the right channels would extract more photon point clouds than the left channels. As a result, the characteristics of the left channels would be lost in a large number of photon point clouds of the right channels. The previously mentioned points would be satisfied when looking at the characteristics of the right channel only, which supports the analysis in Fig. 7(a). More importantly, the dataset with a low to medium SNR is consistent with the above conclusion. When the SNR increases to a high level, it means that the dataset does not need to be classified again, which is also consistent with our perception. Overall, the right channel has more erroneous labels that need to be corrected.

In short, the terrain complexity could harm classification variation. Second, the signal-to-noise also interferes with the classification difference variation. The model performs better in low SNR datasets less than 7.5. We would expect the model to show better performance than ATL08 products for datasets collected during the day. The higher the SNR, the more complex the terrain will be, which would bring more difficulties to the classification, which is in line with our usual perception. Finally, the right channel will be corrected for more mislabels than the left channel generally.

VI. CONCLUSION

In this study, we proposed a supervised classification method to further improve data availability compared to NASA's official product. We used the AutoML method to develop superior performance models and conducted the ATL08 labels of reclassification experiments in forest regions. Finally, we assessed the

performance of models and analyzed the influence factors of detection results.

The results indicated that the classification results are largely superior to the ATL08 products. The kNNdist feature incorporated in this method can effectively improve the interpretability of the model for the photon denoising process. Over the regions tested, the proposed method could improve the proportion of signals correctly identified by 6.4%, 12.2%, 2.7%, 9.3%, and 1.4% in five datasets. The model performs better in low SNR datasets less than 7.5. The classifiers are not sensitive to the symmetry of photon distribution. Furthermore, the more complex the terrain, the higher the SNR, and the classifications will be more difficult. In general, the classifiers had good stability in aspects of nonuniform distribution, different terrain, and SNR.

Our approach could train models with a few sample points reducing calculation expenses. The method achieved superior detection results than ATL08 products and largely improved data availability.

Acknowledgment

The authors would like to thank the American National Snow and Ice Data Center for providing free ATLAS data to the public. We also appreciate the valuable comments and constructive suggestions from the anonymous reviewers that helped improve the manuscript.

References

[1] National Aeronautics and Space Administration, "ICESat/GLAS overview," Nov. 2021. [Online]. Available: https://nsidc.org/data/icesat

[2] J. Rosette, P. North, and J. Suarez, "Vegetation height estimates for a mixed temperate forest using satellite laser altimetry," *Int. J. Remote Sens.*, vol. 29, no. 5, pp. 1475–1493, Mar. 2008.

[3] M. Simard, N. Pinto, J. B. Fisher, and A. Baccini, "Mapping forest canopy height globally with spaceborne LiDAR," *J. Geophys. Res.*, vol. 116, no. G4, Nov. 2011, Art. no. G04021, doi: 10.1029/2011JG001708.

[4] L. Duncanson et al., "Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California," *Remote Sens. Environ.*, vol. 242, Jun. 2020, Art. no. 111779.

[5] M. Lefsky et al., "Estimates of forest canopy height and aboveground biomass using ICESat," *Geophysical Res. Lett.*, vol. 32, no. 22, Nov. 2005, Art. no. L22S02.

[6] W. Abdalati et al., "The ICESat-2 laser altimetry mission," *Proc. IEEE Proc. IRE*, vol. 98, no. 5, pp. 735–751, May 2010.

[7] National Aeronautics and Space Administration, "ICESat & ICESat-2." [Online]. Available: https://icesat.gsfc.nasa.gov/

[8] T. Markus et al., "The ice, cloud, and land elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation," *Remote Sens. Environ.*, vol. 190, pp. 260–273, Mar. 2017.

[9] L. Magruder, K. Brunt, and M. Alonzo, "Early ICESat-2 on-orbit geolocation validation using ground-based corner cube retro-reflectors," *Remote Sens.*, vol. 12, no. 21, Nov. 2020, Art. no. 3653. [Online]. Available: https://www.mdpi.com/2072-4292/12/21/3653

[10] X. Zhu, S. Nie, C. Wang, and X. Xi, "The performance of ICESat-2's strong and weak beams in estimating ground elevation and forest height," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 6073–6076.

[11] S. B. Luthcke et al., "ICESat2 pointing calibration and geolocation performance," *Earth Space Sci.*, vol. 8, no. 3, Mar. 2021, Art. no. e2020EA001494, doi: 10.1029/2020EA001494.

[12] T. Neumann et al., "The Ice, Cloud, and Land Elevation Satellite-2 mission: A global geolocated photon product derived from the advanced topographic laser altimeter system," *Remote Sens. Environ.*, vol. 233, Nov. 2019, Art. no. 111325.

[13] Z. Xiaoxiao, W. Cheng, X. Xiaohuan, N. Sheng, Y. Xuebo, and L. Dong, "Research progress of ICESat-2/ATLAS data processing and applications," *Infrared Laser Eng.*, vol. 49, no. 11, pp. 1007–2276, 2020.

[14] L. A. Magruder, M. E. Wharton, K. D. Stout, and A. L. Neuenschwander, "Noise filtering techniques for photon-counting ladar data," *Proc. SPIE*, vol. 8379, May 2012, Art. no. 83790Q, doi: 10.1117/12.919139.

[15] K. Horan and J. Kerekes, "An automated statistical analysis approach to noise reduction for photon-counting LiDAR systems," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Melbourne, Australia, 2013, pp. 4336–4339. [Online]. Available: https://ieeexplore.ieee.org/document/6723794/

[16] U. C. Herzfeld et al., "Algorithm for detection of ground and canopy cover in micropulse photon-counting LiDAR altimeter data in preparation for the ICESat-2 mission," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2109–2125, Apr. 2014. [Online]. Available: https://ieeexplore.ieee.org/document/6522499/

[17] B. Chen and Y. Pang, "A denoising approach for detection of canopy and ground from ICESat-2's airborne simulator data in Maryland, USA," *Proc. SPIE*, vol. 9671, Oct. 2015, Art. no. 96711S, doi: 10.1117/12.2202777.

[18] J. Zhang and J. Kerekes, "An adaptive density-based model for extracting surface returns from photon-counting laser altimeter data," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 726–730, Apr. 2015.

[19] J. Huang, Y. Xing, H. You, L. Qin, J. Tian, and J. Ma, "Particle swarm optimization-based noise filtering algorithm for photon cloud data in forest area," *Remote Sens.*, vol. 11, no. 8, Apr. 2019, Art. no. 980.

[20] S. Popescu et al., "Photon counting LiDAR: An adaptive ground and canopy height retrieval algorithm for ICESat-2 data," *Remote Sens. Environ.*, vol. 208, pp. 154–170, Apr. 2018.

[21] S. Nie et al., "Estimating the vegetation canopy height using micro-pulse photon-counting LiDAR data," *Opt. Exp.*, vol. 26, no. 10, pp. A520–A540, May 2018.

[22] L. He, Y. Pang, Z. Zhang, X. Liang, and B. Chen, "ICESat-2 data classification and estimation of terrain height and canopy height," *Int. J. Appl. Earth Observation Geoinformation*, vol. 118, Apr. 2023, Art. no. 103233. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1569843223000559

[23] B. Chen et al., "Potential of forest parameter estimation using metrics from photon counting LiDAR data in howland research forest," *Remote Sens.*, vol. 11, no. 7, Apr. 2019, Art. no. 856.

[24] B. Chen et al., "Ground and top of canopy extraction from photon-counting LiDAR data using local outlier factor with ellipse searching area," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 9, pp. 1447–1451, Sep. 2019.

[25] B. Chen et al., "Forest signal detection for photon counting LiDAR using random forest," *Remote Sens. Lett.*, vol. 11, no. 1, pp. 37–46, Jan. 2020, doi: 10.1080/2150704X.2019.1682708.

[26] A. Jain, P. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000. [Online]. Available: https://ieeexplore.ieee.org/document/824819/

[27] A. Neuenschwander and K. Pitts, "The ATL08 land and vegetation product for the ICESat-2 mission," *Remote Sens. Environ.*, vol. 221, pp. 247–259, Feb. 2019.

[28] M. Brown et al., "Applications for ICESat-2 data from NASA's early adopter program," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 4, pp. 24–37, Dec. 2016.

[29] Y. He, J. Lin, Z. Liu, H. Wang, L.-J. Li, and S. Han, "AMC: AutoML for model compression and acceleration on mobile devices," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 815–832.

[30] M. Feurer and F. Hutter, "Hyperparameter optimization," in *Automated Machine Learning: Methods, Systems, Challenges*, F. Hutter, L. Kotthoff, and J. Vanschoren, Eds. Berlin, Germany: Springer, 2019, pp. 3–33.

[31] Z. Du, W. Wang, W. Zeng, and H. Zeng, "Nitrogen deposition enhances carbon sequestration by plantations in Northern China," *PLoS One*, vol. 9, no. 2, Feb. 2014, Art. no. e87975.

[32] P. Yong, L. Xiaojun, J. Wen, S. Lin, Y. Guangjian, and S. Jiancheng, "The comprehensive airborne remote sensing experiment in Saihanba Forest farm," *J. Remote Sens.*, vol. 25, no. 4, pp. 904–917, 2021.

[33] Y. Xie, J. Zhang, X. Chen, S. Pang, H. Zeng, and Z. Shen, "Accuracy assessment and error analysis for diameter at breast height measurement of trees obtained using a novel backpack LiDAR system," *Forest Ecosystems*, vol. 7, pp. 1–11, May 2020.

[34] A. L. Neuenschwander et al., "ATLAS/ICESat-2 L3A land and vegetation height, Version 5," 2021. [Online]. Available: https://nsidc.org/data/atl08/versions/5

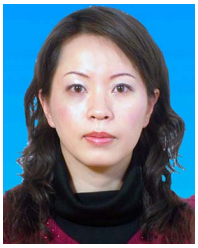
[35] T. A. Neumann et al., "ATLAS/ICESat-2 L2A global geolocated photon data, Version 5," 2021. [Online]. Available: https://nsidc.org/data/atl03/versions/5

[36] H2O.ai, "h2o: R interface for H2O. R package version 3.38.0.2," 2022. [Online]. Available: https://github.com/h2oai/h2o-3

[37] L. Gao, D. Wang, L. Zhuang, X. Sun, M. Huang, and A. Plaza, "BS^3 LNet: A new blind-spot self-supervised learning network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5504218.
[38] D. Wang, L. Zhuang, L. Gao, X. Sun, M. Huang, and A. J. Plaza, "PDB-SNet: Pixel-shuffle down-sampling blind-spot reconstruction network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5511914.

Bo Zhang received the B.E. degree in instrumentation and control engineering from Shandong University, Shandong, China, in 2020. He is currently working toward the Ph.D. degree in cartography and geographic information systemsfrom Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China.

His research interests include forest parameters inversion using space-borne LiDAR and multimodal satellite images fusion, photon point cloud processing, and machine learning algorithms, etc.



Li Zhang received the M.S. degree in geography from South Dakota State University, Brookings, SD, USA, and the Ph.D. degree in cartography and geography information system from Beijing Normal University, Beijing, China.

She is currently a Professor with the Institute of Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. She has been the PI for more than 30 national and provincial sponsored projects. She has authored or coauthored more than 160 scientific papers in the areas of her interest. From 2005 to 2007, she was an Environmental Scientist with United States Geological Survey (USGS), Earth Resources Observation and Science (EROS) Center, South Dakota, USA, working on carbon modeling, ecosystem performance, and vegetation monitoring. Her research interests include remote sensing applications for ecosystem and land-surface dynamics, carbon cycling, and coastal environment monitoring.

Yong Pang (Member, IEEE) received the B.S. degree in forestry from Anhui Agriculture University, Hefei, China, in 1997, the M.Agr. degree in forest management from the Chinese Academy of Forestry, Beijing, China, in 2000, and the Ph.D. degree in cartography and geography information system from the Chinese Academy of Sciences, Beijing, China, in 2006.

From 2006 to 2008, he was a Postdoctoral Researcher with the Department of Forest, Rangeland, and Watershed Stewardship, Colorado State University, Fort Collins, CO, USA. He is currently a Professor with the Research Institute of Forest Resource Information Techniques, Chinese Academy of Forestry. His research interests include surface height and vegetation spatial structure from InSAR and LiDAR, modeling of LiDAR waveforms from forest stands, and development of algorithms for forest parameter retrieval from remote sensing data.

Peter North received the M.A. degree in natural sciences and computing science from the University of Cambridge, Cambridge, U.K., in 1988, and the D.Phil. degree in 3-D computer vision from Sussex University, Sussex, U.K., in 1992.

From 1992 to 2000, he was with the Natural Environment Research Council, Centre for Ecology and Hydrology. He is currently a Professor in physical geography with Swansea University, Swansea, Wales, U.K. His research interests include modeling the interaction of radiation with natural surfaces, where he developed the FLIGHT radiative transfer model, and global retrieval of atmospheric and land surface information from LiDAR, hyperspectral, and multidirectional satellite observations.

Min Yan received the Ph.D. degree in forestry remote sensing with the Chinese Academy of Forestry, Beijing, China, in 2016.

She is currently a Associated Professor with the Chinese Academy of Sciences (CAS), Beijing, China. Her research interests include forest and mangrove carbon cycle and their responses to climate variations.

Dr. Yan is a Member of the International Society for Digital Earth (ISDE).

Hongge Ren received the D.Sc. degree in cartography and geography information system from the Institute of Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, in 2023.

Her research interests include machine learning, classification of satellite remote sensing data, eddy covariance data application, dynamic vegetation model, terrestrial carbon cycle, and water cycle, etc.

Linlin Ruan received the bachelor's degree in geographic information science from Beijing Forestry University, Beijing, China, in 2020. She is currently working toward the Ph.D. degree in cartography and geography information system with the Aerospace Information Research Institute (five-year Ph.D. Program), University of Chinese Academy of Sciences, Beijing, China.

Her research interests include the remote sensing of vegetation ecology, including analyzing the spatiotemporal variations, and drivers of vegetation ecology.

Zhenyu Yang received the M.Sc. degree in master of civil and hydraulic engineering from the School of Marine Technology and Geomatics, Jiangsu Ocean University, Lianyungang, Jiangsu, China, in 2023.

His research interests include social media and big data mining.

Bowei Chen received the B.Sc. degree in forestry from Huazhong Agricultural University and Huazhong University of Science and Technology, Wuhan, China, in 2013 and the Ph.D. degree in forestry remote sensing from the Chinese Academy of Forestry, Beijing, China, in 2019.

He is currently an Assistant Professor with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. His research interests include the use of satellite LiDAR to monitor mangrove productivity and its coastal application, and machine learning algorithms in both satellite images and LiDAR sensors, especially for the ICESat-2 and GEDI missions.