Exploitation of Time Series Sentinel-2 Data and Different Machine Learning Algorithms for Detailed Tree Species Classification

Yanbiao Xi, Chunying Ren, Qingjiu Tian, Yongxing Ren, Xinyu Dong, and Zhichao Zhang

***Abstract*—The classification of tree species through remote sensing data is of great significance to monitoring forest disturbances, biodiversity assessment, and carbon estimation. The dense time series and a wide swath of Sentinel-2 data provided the opportunity to map tree species accurately and in a timely manner over a large area. Many current studies have applied machine learning (ML) algorithms combined with Sentinel-2 images to classify tree species, but it is still unclear which algorithm is more effective in the automotive extraction of tree species. In this study, five machine learning algorithms were compared to identify the composition of tree species with multi-temporal Sentinel-2 images in the JianShe forest farm, Northeast China. Three major types of deep neural networks (Conv1D, AlexNet and LSTM) were tested to classify Sentinel-2 time series, which represent three disparate but effective strategies to apply sequential data. The other two models are Support Vector Machine (SVM) and Random Forest (RF), which are renowned for extensive adoption and high performance for various remote sensing applications. The results show that the overall accuracy of neural network models is better than that of SVM and RF. The Conv1D model had the highest classification accuracy (84.19%), followed by the LSTM model (81.52%), and the AlexNet model (76.02%). For non-neural network models, RF's classification accuracy (79.04%) is higher than that of SVM (72.79%), but lower than that of Conv1D and LSTM. Therefore, the deep neural networks combined with multi-temporal Sentinel-2 images can efficiently improve the accuracy of tree species classification.**

***Index Terms*—Deep learning; sequential pattern; tree species classification; Sentinel-2 image**

I. Introduction

THE number and distribution of tree species are related to ecosystem parameters such as biodiversity and habitat quality and are important indicators of forest ecological value [1], [2]. Additionally, the composition of tree species is the fundamental content of forest resource surveys and monitoring [3], which is the basis of forest ecological planning and forest policy making [4]. Therefore, quick and accurate tree species mapping is highly relevant for many ecological and forestry applications.

Compared with traditional field surveys, remote sensing can obtain tree species information without destroying the forest structure, particularly within large and inaccessible areas [5]. Remote sensing images with different spatial and temporal resolutions, such as MODIS, Landsat, SPOT, and Quickbird, have been widely used in tree species classification [6], [7], [8]. Limited by high data costs, weather condition, area coverage, and acquisition time, most remote sensing data, such as very high-resolution (VHR), Unmanned Aerial Vehicle (UAV), and LiDAR images, have difficulty making detailed mappings of large areas of tree species. With the launch of Sentinel-2A/2B satellites from the European Space Agency (ESA), high quality images with high spatial, spectral and temporal resolution have become freely available [9], [10]. Detailed spectral information with bands in the visible, Red-Edge, Near-Infrared (NIR), and Shortwave-Infrared (SWIR) wavelengths has high potential to discriminate between different tree species. In addition, multi-temporal data can capture phenological differences between tree species [9], [11]. A number of recent studies explored the potential of multi-temporal Sentinel-2 data for tree species classification. Immitzer *et al*. [9] analyzed 12 tree species classes in Central Europe with all possible combinations of multi-temporal Sentinel-2 scenes throughout year. They achieved overall accuracies of up to 85.7%. Persson *et al*. [12] classified five forest classes in Sweden with an overall accuracy of 88% using Sentinel-2 images from April, May, July, and October. Four tree species classes in Ebersberger forest and Freisinger forest were separated with accuracies of over 88% by Wessel *et al*. [13]. Lim *et al*. [14] achieved an overall accuracy of 90% in the classification of five tree species in Goseong-gun, South Korea with multi-temporal Sentinel-2 images. These studies clearly demonstrate that the combination of abundant spectral information and time series Sentinel images provide an extraordinary advantage in distinguishing tree species.

In terms of tree species classification, one of the key issues is automatic and precise classification methods. Currently, commonly used methods are traditional machine learning

This work was supported by the National Key Research and Development project of China (No.2016YFC0500300, 2017YFD0600903), and the National Natural Science Foundation of China (42001349, 41771370). *(Corresponding author: Chunying Ren.)*

Yanbiao Xi, Qingjiu Tian, Xinyu Dong, and Zhichao Zhang are with the International Institute for Earth System Science, Nanjing University, Nanjing 210023, China (e-mail: xiyb@smail.nju.edu.cn; tianqj@nju.edu.cn; DG1827003@smail.nju.edu.cn; zhangzc@smail.nju.edu.cn).

Chunying Ren is with the Key Laboratory of Wetland Ecology and Environment, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China (e-mail: renchy@iga.ac.cn).

Yongxing Ren is with the College of Earth Science, Jilin University, Changchun 130100, China (e-mail: ryx20@mails.jlu.edu.cn).

methods such as Support Vector Machine (SVM) [13], Random Forest (RF) [9], [14], Extreme Gradient Boosting (XGB) [11], and K-nearest neighbor (KNN) [15]. Lim *et al.* [16] applied RF and SVM with Hyperion and Sentinel-2 data to classify tree species and achieved overall accuracy of 0.99 and 0.97, respectively. Grabska *et al.* [11] developed a workflow to obtain tree species maps with machine learning algorithms (RF, SVM, and XGB) using multi-temporal Sentinel-2 images and environmental data at the regional scale. Wu and Zhang [17] evaluated the capability of KNN and SVM classifiers to identify combined airborne hyperspectral data of tree species with simultaneously acquired LiDAR data in southern China. The most important process in these methods is to extract features from remote sensing images, which are bands or vegetation indices highly relevant to tree species, such as SWIR bands, NDVI, and texture features, and then to input these features into the model as a mixed variable set for classification [14]. Although the mixed variable set which is beneficial to the classification of tree species is selected from a large number of bands, they all weaken the information superiority of the original data to a certain extent, which affects the potential mining of information [18]. Recently, deep learning, as a subset of machine learning, has been applied to tree species classification [18],[20]. One of its most important advantages is the ability to input high-dimensional data and extract effective deep features from the original image in an end-to-end manner [18], [21]. However, most deep learning algorithms only apply to single scene images and require very high spatial resolution (VHR) images, especially for tree species scales [18], [22], [23]. Currently, extracting the comprehensive information from multi-temporal images for automatic and rapid tree species classification is still a challenge.

In this study, we attempted to apply a variety of neural network models (deep convolutional neural network and recurrent neural network, DCNN and RNN) to multi-temporal Sentinel-2 images to make a comparative analysis with traditional machine learning algorithms. The specific objectives of this research were to: (1) explore potential of multi-temporal Sentinel-2 images for tree classification with different machine learning algorithms; (2) compare the performance of deep neural networks with traditional machine learning algorithms for tree species classification. The method of this paper can provide an important reference in tree species classification research.

II. Materials

A. Study Area

The study area is located at the southeast region of Jilin Province, northeast China, near the Changbai Mountains, and ranges from 42°50′34.4″N ~ 42°59′51.1″N and 127°45′8.8″E ~ 127°54′14.5″E(Fig. 1). The study area covers an area of nearly 86.78km2 with an average elevation of 660m. It belongs to the temperate continental monsoon climate with an annual average temperature of 5.78 ℃, and is characterized by low temperature, heavy precipitation, strong wind, and a short growing period [24]. The average annual rainfall is between 700 and 1400 mm, and the average relative humidity is 70%. The forest vegetation in the study area is mainly natural forest, of which coniferous forests are mainly Korean pine (*Pinus tabuliformis*), Dragon spruce (*Picea asperata Mast*), and Dahurian larch (*Larix gmelinii*), and broadleaf forests are mainly Amur linden (*Tilia amurensis*), White birch (*Betula platyphylla*), Manchurian walnut (*Juglans mandshurica*), Aspen (*Populus tremula*), Manchurian ash (*Fraxinus mandshurica*) [25].

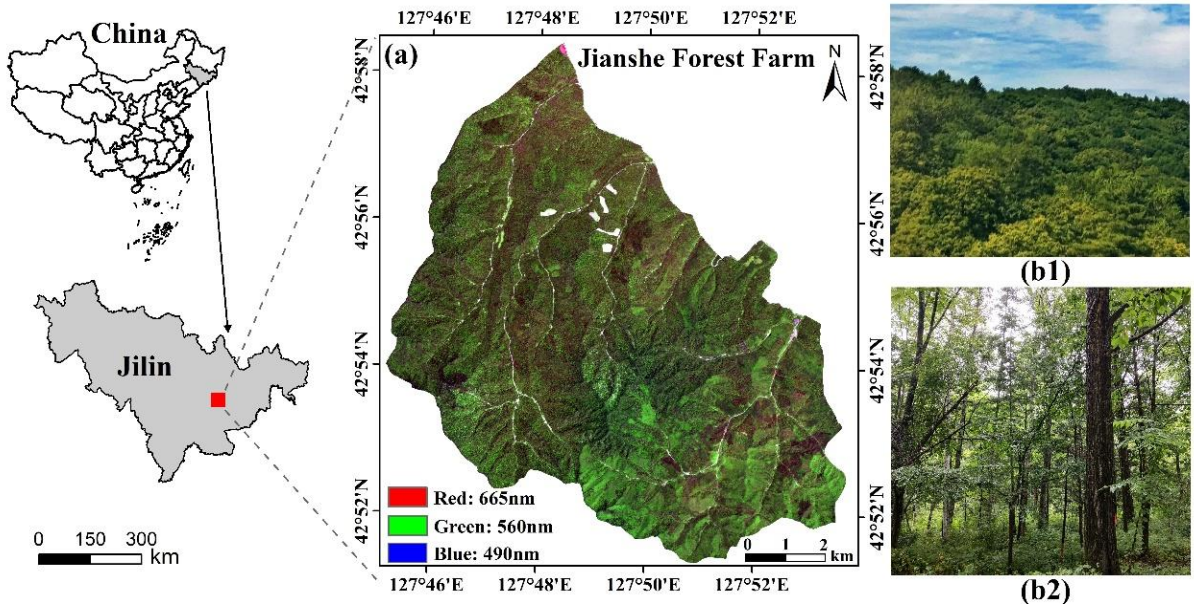

Fig. 1 Study area: JianShe State Owned Forest Farm, Jilin Province, China. (a) The location of the study area. (b1), (b2) Ground level images of the study site.

B. Satellite data

According the Copernicus programme of the European Space Agency, multi-temporal Sentinel-2 images were acquired from the Copernicus Open Access (COA) Hub of the European Space Agency (ESA, 2015) (www.scihub.copernicus.eu/). Twenty-four Sentinel-2 tiles of Level 1C or 2A were downloaded from the Copernicus Open Access (COA) Hub,

corresponding to all available acquisitions between January 1 and December31, 2018 (Table 1). The Sentinel-2 images are 100 km x100km ortho-images with UTM projection and WGS-84 ellipsoid. In this study, all Sentinel-2 images are located in UTM zone 52 N. For cloud-contaminated images, images of the same period in adjacent years (2017 and 2019) were used as an alternative. Only 10 spectral bands were used in this study, including 10-m spatial resolution bands (blue, green, red, and NIR bands) and 20-m spatial resolution bands (Red edge, NIR narrow, and SWIR bands). The Sentinel images were atmospherically corrected using the Sen2Cor plug-in provided by the European Space Agency (ESA) [26]. The 20-m spatial resolution bands were resampled to 10m using the bi-cubic convolution interpolation method on the Sentinel Application Platform (SNAP).

TABLE I
SPECIFICATIONS OF SENTINEL-2 IMAGES FOR THE JIANSHE FOREST FARM THAT WERE USED IN THIS STUDY

Data	Acquisition Time	Cloud Coverage (%)	Product Level
06 January 2018	02:30:59	0.03	S2 L1C
13 January 2018	02:20:39	6.11	S2 L1C
02 February 2018	02:19:09	37.52	S2 L1C
20 February 2018	02:27:31	19.48	S2 L1C
09 March 2018	02:16:01	0.0	S2 L1C
19 March 2018	02:16:01	27.34	S2 L1C
16 April 2018	02:25:49	0.0	S2 L1C
03 May 2019	02:16:11	1.53	S2 L2A
21 May 2018	02:25:51	0.0	S2 L1C
23 May 2018	02:16:49	0.72	S2 L1C
02 June 2018	02:16:39	0.0	S2 L1C
22 June 2018	02:15:59	0.0	S2 L1C
10 July 2019	02:25:59	6.67	S2 L2A
25 July 2018	02:25:49	21.33	S2 L1C
11 August 2018	02:15:59	3.37	S2 L1C
29 August 2017	02:25:49	1.47	S2 L1C
18 September 2019	02:25:51	6.29	S2 L2A
20 September 2018	02:15:59	0.0	S2 L1C
03 October 2018	02:25:49	0.28	S2 L1C
13 October 2018	02:26:29	0.0	S2 L1C
02 November 2018	02:28:39	0.0	S2 L1C
12 November 2018	02:29:29	0.0	S2 L1C
04 December 2019	02:20:41	6.33	S2 L2A
12 December 2018	02:31:09	3.11	S2 L1C

In addition, vegetation indices (VIs) were added to eliminate the influence of some environmental factors, and could highlight the growth characteristics of tree species. The Normalized Difference Vegetation Index (NDVI) is widely applied as a sensitive indicator that can be used to monitor density and intensity of green vegetation growth and phenological variations in time-series analyses [27].

$$NDVI = \frac{\rho_{NIR}-\rho_R}{\rho_{NIR}+\rho_R} \quad (1)$$

Where ρ_{NIR} and ρ_R correspond to near-infrared and red spectral bands of Sentinel 2, respectively.

Moreover, the atmosphere has different attenuation coefficients for different bands, resulting in higher red bands and lower near-infrared bands than true reflectivity. The Enhanced Vegetation Index (EVI) can reduce background value and atmospheric influence [28].

$$EVI = 2.5\left[\frac{\rho_{NIR}-\rho_R}{\rho_{NIR}+C_1\rho_R+\mathrm{L}}\right] \quad (2)$$

Where ρ_{NIR} and ρ_R correspond to near-infrared and red spectral bands of Sentinel 2, respectively. C_1 and L are coefficients to correct for atmospheric condition (i.e., aerosol resistance). For the standard Sentinel-2 EVI product, L=1 and C1=2.4, respectively.

C. Field botanical surveys

Field surveys were conducted from June to July 2018. Based on the principles of spatial distribution uniformity and road accessibility, a total of 286 plots (50 m×50 m) were obtained in the study area. The survey factors include longitude and latitude of the sample plot recorded by Global Positioning System (GPS), aspect, slope, tree height, tree species, and basal area (tree diameters over 10 cm).

Based on the basal area factor (BAF) [29], any tree species with a basal area frequency above 50% in each plot was selected as the dominant tree species (Fig. 2). According to the field surveys, eight dominant tree species classes were identified, comprising Amur linden (*Tilia amurensis*), Korean pine (*Pinus tabuliformis*), Dahurian larch (*Larix gmelinii*), Aspen (*Populus tremula*), Manchurian ash (*Fraxinus mandshurica*), Manchurian walnut (*Juglans mandshurica*), White birch (*Betula platyphylla*), and Dragon spruce (*Picea asperata Mast*).

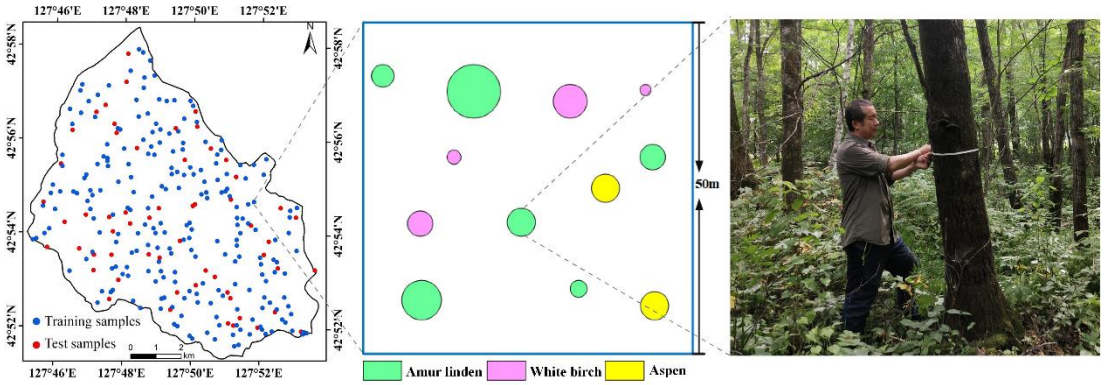

Fig. 2 The diagram of dominant tree species

D. Dataset Partition

According to the boundary of these sample plots, we extracted pixels with different tree species from time series Sentinel-2 images as a dataset. Thus, each pixel is time series data containing Sentinel-2 information. The pixels of each tree

species obtained from each plot are shown in Table 2. The sample data of the whole study area were split into two datasets: training and validation samples. Based on the ten-fold cross-validation technique, we randomly selected 70% of the dataset as training samples and the remaining 30% as held-out validation samples. In addition, 20% of the training samples were further split for model optimization during training, the final classification results were evaluated using the validation samples.

TABLE II
SATELLITE DATA WITH THE DATA OF ACQUISITION, SPATIAL RESOLUTION AND SOURCE

Category Code	Description	Number of Plots	Number of Pixels
AL	Amur linden	31	781
KP	Korean pine	28	690
DL	Dahurian larch	36	896
AS	Aspen	26	657
WB	White birch	57	1413
MW	Manchurian walnut	42	1044
MA	Manchurian ash	41	1019
DS	Dragon spruce	25	625

III. METHODS

In this study, we adopted three major types of deep neural networks (Conv1D, AlexNet and LSTM) were tested to classify Sentinel-2 time series, which represent three disparate but effective strategies to apply sequential data. In addition, SVM and RF as the representative non-neural-network classifier to compare with three neural network models, since these classifiers is renowned for high performance and is often established as the baseline model in classification tasks. The main steps of the data processing workflow in this study are presented in Fig. 3.

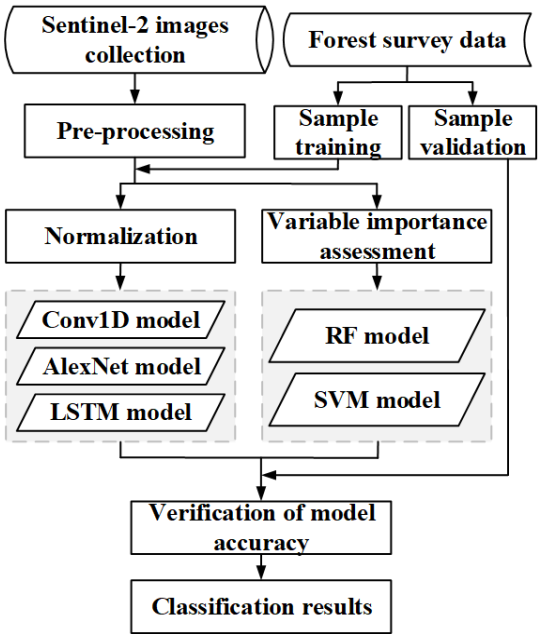

Fig. 3 Framework of the tree species classification based on five machine learning algorithms

A. Neural Network Classifiers

Deep learning in neural networks can extract more abstract feature through multi-layer neural networks to better describe the complex structure of high-dimensional data [22]. Three major deep neural networks were applied to classify tree species with multi-temporal Sentinel-2 images in this study. The first is one-dimensional convolutional neural network (Conv1D) [20], 30], which is a branch of convolutional neural network (CNN). The second is the AlexNet model, which is two-dimensional convolutional neural networks, and was first proposed by Krizhevsky *et al.* [31]. The third is long short-term memory (LSTM) in the recurrent neural network (RNN) family [32]. For high-dimensional data containing multiple bands and multiple time phase features, these models input the data as a whole and autonomously learn features from the data.

The Conv1D includes input, convolutional, pooling, flatten, and fully connected layers (Fig. 4). The time series images are used as the input of the input layer. Then Conv1D applies the convolution kernel to capture the temporal pattern or shape of the input series, gradually extracting deep features by multiple convolutional layers [33]. Each convolution layer applied rectified linear unit (ReLU) as the activation function, which can significantly prevent overfitting and accelerate the training process [34]. The pooling layers operate on the feature maps to aggregate the information within the given neighborhood window with a max or average pooling operation. Last, the fully connected layers produce the predictive probabilities of all the required classification results in the input data. A Conv1D layer has a pattern or shape template in each of its channels and matches the patterns with the input through convolution, so it has good performance for continuous sequence data such as multi-temporal data [35], sea surface oil spill [36], and vibration signal [37].

AlexNet, a 2-D convolutional neural network, is a powerful model capable of achieving high accuracies with image classification (Fig. 5). Similar to the structure of Conv1D model, AlexNet is composed of 5 convolutional layers and 3 fully connected layers. The first layer, the second layer and fifth layer followed by overlapping max pooling layer, and the third, fourth and fifth convolutional layers are connected directly. Then, the network is composed of a few fully connected layers with dropout and threshold, and Softmax classifier is managed to order the images into different classes. Dropout is a commonly used algorithm for training neural networks [31]. It randomly drops units (along with their connections) from the neural network to solve the overfitting problem. For both of these architectures, the stochastic gradient descent (SGD) algorithm is used to optimize the whole cost function during the back-propagation optimization procedure.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3098817, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

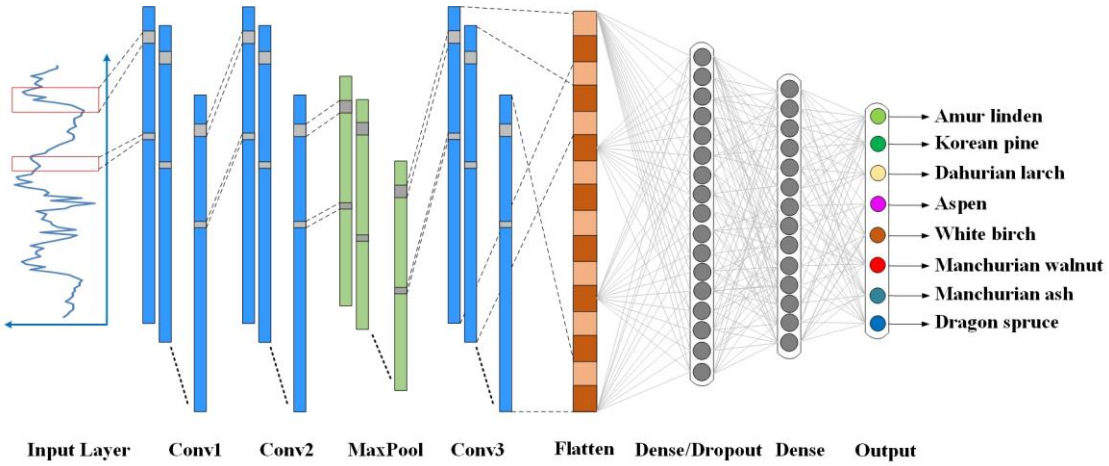

Fig.4. One dimensional convolutional neural network

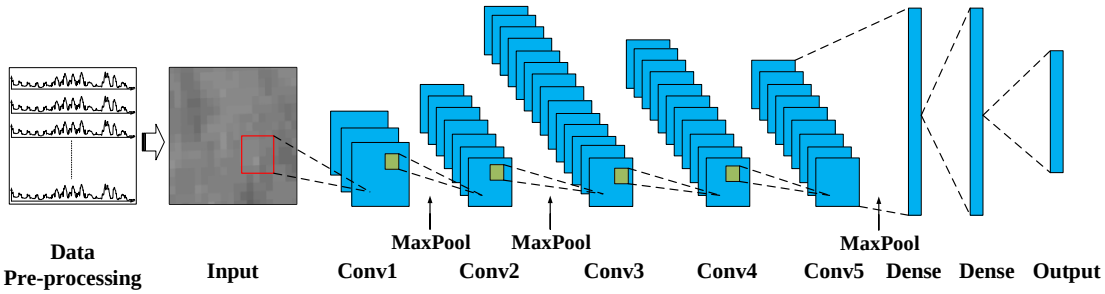

Fig. 5. The architecture of AlexNet model

For both of the above classifiers, the input data format is different. Firstly, each pixel in the image is extracted as a set of time series data by the field samples. Then, the pixel values of each band are arranged horizontally into a one-dimensional vector as input to the Conv1D model, while these values are arranged vertically as a two-dimensional array and converted into grayscale images as input to AlexNet.

Recurrent neural networks (RNN) are another class of artificial neural networks (ANNs) that are often used to process sequential data [32]. As an improved version of RNN, LSTM can retain information for much longer periods of time due to its recurrent structure and gating mechanisms and is regarded as potentially significant for applications related to time series [38]. As shown in Fig. 6(a), LSTM has a chain-like structure. The basic building block of LSTM is composed of a cell (the memory part of the LSTM unit) and three kinds of gates to regulate a cell state, including an input gate (σ_1), an output gate (σ_2), and a forget gate (σ_3) (Fig. 6b). The sigmoid layer gives a number (0-1) as output, this number describes how much information to let through. If the output is closer to 1, the memory cell information is transferred to the hidden state for use. Otherwise, the memory cell information is held by itself. The internal structure of LSTM is shown in Fig. 6(b), any one of the black lines represents the transmission of the vector. The circle with the plus and the cross sign is the cell state and acts as the key to LSTM. The connected lines represent the connections of the vectors, while the separated lines indicate that the content is copied and distributed to different locations. In addition, an LSTM unit is designed to record values over different time intervals, long or short [39]. LSTM can deal with the gradient attenuation problem in recurrent neural networks, better capturing the dependence of time series data with different lengths [40].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3098817, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

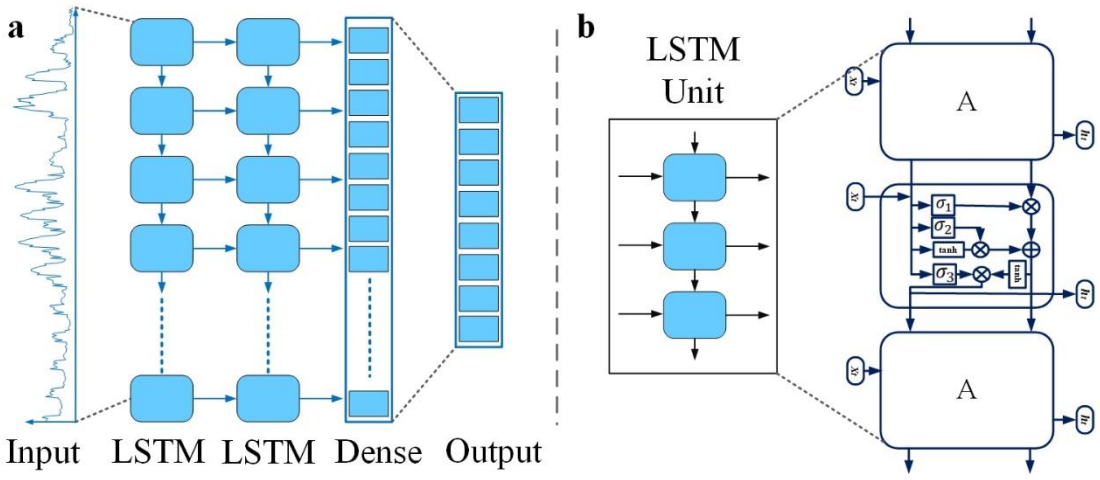

Fig. 6 (a) The architecture of LSTM model, (b) each LSTM unit

B. *Support Vector Machine and Random Forest*

Support Vector Machine (SVM) and Random Forest (RF) were selected as representatives of feature extraction classifiers [1]. These classifiers are renowned for extensive adoption and reliable performance for various remote sensing applications such as paddy rice classification [41], soil moisture monitoring [42], forest aboveground biomass estimation [43], and tree crown delineation [44].

The RF is a classifier that includes a large number of decision tree classifiers, each tree is trained with 70% randomly selected training samples to solve the same problem. In addition, each tree is created to the maximum possible size without pruning. Then, the final classification is conducted based on a majority vote of the trees in the forest. The RF classifier is a non-parametric method and has strong robustness and noise insensitivity to overfitting. At present, RF has been successfully applied to the classification of tree species [16], [45], [46].

The SVM classifier was suggested by Vapnik [47], as it finds the best decision hyper-plane by separating positive samples and negative samples, which can manage classification problems in multi-dimensional data, and has been widely used for tree species classification [48]. This hyper-plane maximizes the distance between different training samples. The SVM is trained through samples to create a reliable model that correctly classifies positive and negative samples [35], [49].

In this study, the application and data processing of RF and SVM models are shown in Fig. 7. Firstly, the multi-temporal Sentinel-2 images and sample data were preprocessed, then the variable importance assessment were executed from S2 and sample data, and the optimal combination of feature variables was selected. Variable importance assessment is a vital step when high-dimensional datasets are used [16]. Using only the most important variables can result in higher classification accuracies than using all available features. High-dimensional datasets with a large number of predictors may result in model overfitting [35]. Furthermore, removing irrelevant or redundant variables and creating a sparse subset for classification helps to simplify statistical problems and shortens data processing times [43]. Finally, hyper-parameters of the RF and SVM models are tuned, and the classification results are assessed.

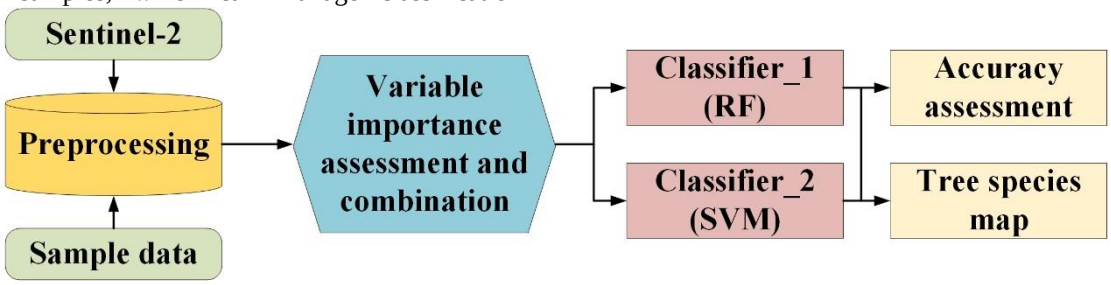

Fig. 7 Scheme of forest tree species mapping with different methods of Sentinel-2 time series.

C. *Accuracy Assessment*

To show the classification performance of each classifier clearly, the confusion or error matrices, kappa coefficient and overall accuracy were calculated in this study. The kappa accuracy evaluates how well the classification performs compared to randomly assigning values. The overall accuracy is a ratio of correctly classified pixels to all verification pixels expressed as a percentage. In addition, we also used the macro-average of F1 score as an indicator of classification capability. For each class, F1 score is the harmonic mean of producer's accuracy (PA or being regarded as recall) and user's accuracy (UA or being regarded as precision). F1-score for each class is shown as follows:

$$F1_c = \frac{2}{\frac{1}{Recall}+\frac{1}{Precision}} = \frac{2}{\frac{1}{A_{prod}}+\frac{1}{A_{user}}} \quad (3)$$

where $F1_c$ is the F1-core of a single class, A_{prod} is the

producer's accuracy of the class, and A_{user} is the user's accuracy of the class. The user's accuracy (UA) is calculated as the ratio of the total number of correctly classified samples for a particular class to the total number of reference samples. The producer's accuracy (PA) refers to the ratio of accurately classified reference samples of a particular class to the total number of the reference samples for that class.

IV. Results

A. Hyper-parameters in Neural Network Models

In machine learning, the selection of hyper-parameter is important for model optimization and accuracy improvement [19]. Compared with other parameters obtained through training, the values of the hyper-parameter are used to control the learning process [18], [52]. Hyper-parameters of CNN include the number of hidden layers, kernel size, and dropout rate. Because of the variability of the specialized architectures, there is no universal formula set for the optimal combination of hyperparameters [30].

For Conv1D model, the model complexity is directly related to the fitting result. One way to deal with underfitting and overfitting is to adjust the complexity of the model for the data set. Therefore, a relatively simple model with only one convolutional layer was set at the beginning, new models were generated by adding a new layer, or replacing a part of the network with a more complex component [30]. In this way, the tested model grew in size and complexity until classification accuracy did not improve further. As shown in Fig. 8, the training time also increases with the increasing number of layers, but the classification accuracy is not improved. As a result, the neural network with 5 convolutional layers was applied for classification. Specifically, the optimized Conv1D-based model mainly includes five Conv1D layers with a width (kernel size) of 3 and an inception module. The output of the inception module is a concatenation of a width 3 convolutional layer. The first layer, the third layer and fifth layer followed by overlapping max pooling layer, and the second and fourth layers are connected directly.

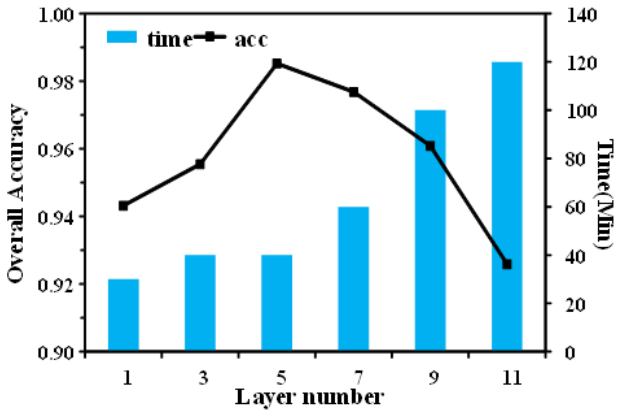

Fig. 8 Impact of layer number on classification accuracy and efficiency for Conv1D

The AlexNet model structure in this study will not be changed as it is a typical two-dimensional convolutional neural network, however, other hyper-parameters will be adjusted. Fig. 9 shows the influence of different convolution kernels on classification accuracy. Based on the accuracy of the last five epochs, we selected the convolution kernel of 5 as the optimal parameter for classification. In addition, the number of iterations is 120, and 30% was selected for the dropout rate in this study. The stochastic gradient descent (SGD) algorithm is used to optimize the whole cost function during the back-propagation optimization procedure.

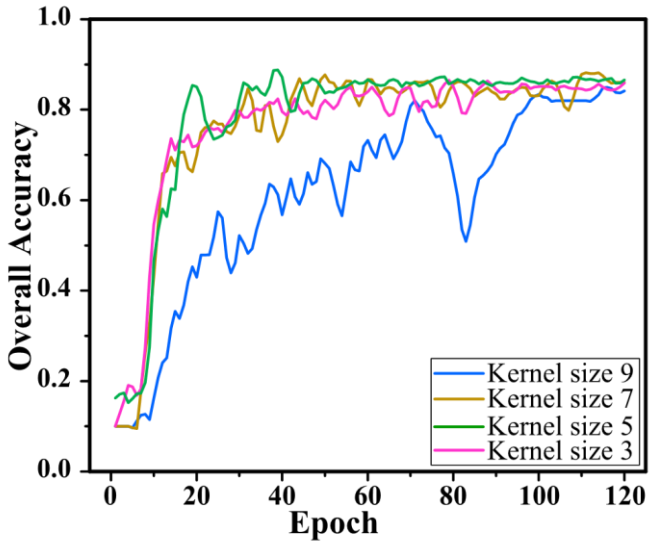

Fig. 9. Impact of kernel size on accuracy for AlexNet

The LSTM-based model was optimized in a similar way to get minimum reconstruction error results. After testing, a two-layer LSTM model with an adaptive moment estimation (Adam) optimizer was built in this study. Arslan and Sekertekin [38] and Santosh *et al.* [39] also indicated that when the amount of sample data is not particularly large, two layers were shown to be sufficient to detect more complex features. Parameters of Adam were fixed as: β1=0.9, β2=0.999, and ε=1e-07, with a learning rate decay of 0.001. LSTM network was separately applied to 24 Sentinel-2 images in order to obtain time-series classification results at each pixel. In addition, all other parameters were optimized empirically based on standard practice in deep network modelling. For example, selecting the optimal dropout rate and pooling layers to prevent overfitting; the learning rate and batch size were tested to learn the deep features.

These algorithms are programmed based on the Tensorflow-Keras deep learning framework which was implemented in Python 3.6. The training process of Conv1D-based networks took about 1 to 5h to obtain stable validation accuracy, and the time of the selected architecture was about 2.5h. The training time of the AlexNet-based networks was relatively long, about 5 to 10h, and the selected one took 8h to finish training. The training process of LSTM-based networks took about 2 to 6h to obtain stable accuracy, and the time of the selected architecture was about 4h.

B. Feature importance and hyper-parameters in RF and SVM Models

In order to select only the most important features for classification for RF and SVM classifier, evaluation was

performed by using mean decrease in accuracy (MDA). MDA, as the most popular measure for feature evaluation, is considered more straightforward, reliable, and easier to understand [53]. These statistics were calculated in R software using the random forest package, the results are shown in Fig. 10.

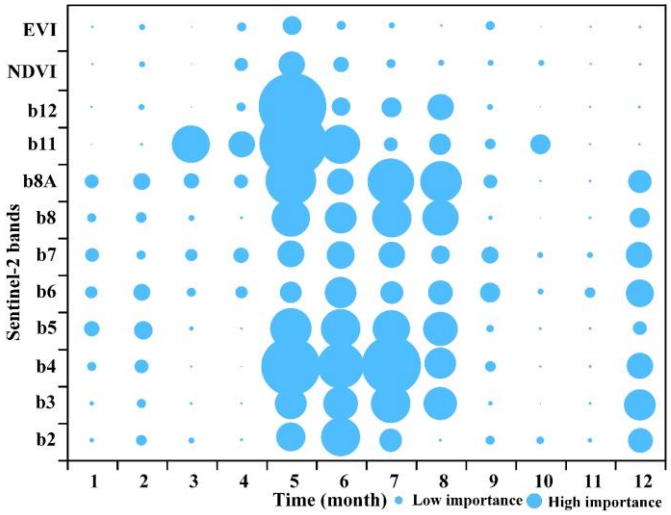

Fig. 10 Aggregated feature importance derived from MDA. Notes that a larger dot size indicates a higher importance

For the spectral features of Sentinel-2, the most important bands are the Red band (B4) and the SWIR bands (B8A, B12 and B11), followed by the Red-Edge band 1 (B5). For the time phase features of Sentinel-2, images from May, June and July are more useful compared to those from late summer and autumn, as shown in Fig. 10. Moreover, while images from March and December show importance to some extent, they are not as good as those from the growing season. In general, amongst the Sentinel-2 bands, B4, B11, and B12 in May, B2, B4, and B11 in June, and B4 and B8A in July are more conducive to tree species separation. Therefore, based on feature contribution rate, the first 20 features input into the RF and SVM classifier for training and classification. The first 20 features are shown in Table 3.

TABLE III
MOST IMPORTANT VARIABLES SELECTED FOR TREE SPECIES CLASSIFICATION BY MDA

Rank	Band	Data
1	B11 (1610nm)	03 May 2018
2	B12 (2190nm)	03 May 2018
3	B4 (665nm)	10 July 2019
4	B11 (1610nm)	23 May 2018
5	B4 (665nm)	03 May 2018
6	B8A (865nm)	03 May 2018
7	B4 (665nm)	23 May 2018
8	B4 (665nm)	25 July 2018
9	B8A (865nm)	10 July 2019
10	B4 (665nm)	02 June 2019
11	B11 (1610nm)	02 June 2018
12	B8A (865nm)	09 August 2018
13	B2 (490nm)	02 June 2018
14	B8A (865nm)	09 August 2018
15	B5 (705nm)	03 May 2018
16	B5 (705nm)	23 May 2018
17	B11 (1610nm)	19 March 2018
18	B7 (783nm)	10 July 2019
19	B5 (705nm)	10 July 2019
20	B2 (490nm)	22 June 2018

In addition, the RF and SVM models are also optimized with hyperparameters to obtain the best classification accuracy, the hyperparameters were configured by the classification accuracy of the validation set. Details of hyperparameter optimization for RF and SVM are shown in Table 4. Among them, the "rbf" refers to the radial basis function.

TABLE IV
HYPER-PARAMETER VALUES OF RF AND SVM CLASSIFIERS.

Classifier	Hyper-parameter	Selected values
RF	n_estimators	600
	max_features	auto
	bootstrap	True
	max_depth	None
	min_samples_split	2
	min_samples_leaf	1
SVM	C	120
	gamma	auto
	kernel	rbf

C. Classification of Tree Species and Accuracy Evaluation

Confusion matrices, Macro F1-score, and kappa coefficient of the test set were calculated to evaluate the performance of the five models. Table 5 shows the accuracy comparison of different classification methods. The results show that the overall accuracy of neural network models is better than that of traditional machine learning model. Among the neural network models, the Conv1D model had the highest classification accuracy (84.27%), followed by the LSTM model (81.20%), while the AlexNet model had the lowest classification accuracy (75.88%). For non-neural network models, RF's classification accuracy is higher than that of SVM, with accuracies of 79.14% and 72.22%, respectively. Table 5 also shows the Macro F1 scores of different tree species. Dahurian Larch, Aspen, and White Birch scores were all higher in these models. In contrast, the scores of Korean Pine and Manchurian walnuts were low in all models. The F1 scores of other tree species, such as Amur Linden, Manchurian ash and Dragon Spruce, were significantly different in these models.

TABLE V
COMPARISONS OF CLASSIFICATION ACCURACY AMONG DIFFERENT METHODS (MEAN VALUE ±STANDARD DEVIATION AFTER 10 ITERATIONS)

Metric		Method				
		Conv1D	AlexNet	LSTM	RF	SVM
Overall accuracy (%)		84.27(±*0.010*)	75.88(±*0.010*)	81.20(±*0.010*)	79.14(±*0.015*)	72.22(±*0.012*)
Kappa		0.82(±*0.012*)	0.72(±*0.013*)	0.78(±*0.012*)	0.76(±*0.020*)	0.68(±*0.014*)
	Amur linden	85.66	70.90	81.03	77.43	68.05

Macro F1 score (%)	Korean pine	75.59	57.91	70.78	67.74	57.40
	Dahurian larch	84.81	80.53	82.53	82.91	79.16
	Aspen	91.06	80.08	88.17	83.28	80.22
	White birch	86.82	83.31	86.16	85.86	78.20
	Manchurian walnut	76.07	58.82	66.55	63.61	62.27
	Manchurian ash	77.50	78.91	78.33	79.39	59.32
	Dragon spruce	79.93	76.93	71.83	73.49	70.89

The confusion matrices for all models are demonstrated in Fig. 11, which provide an overview of the classification errors between the tree species. The confusion matrix showed that the producer's accuracy of Amur linden, Dahurian larch, Aspen, and White birch (over 75%) is higher in all models. The producer's accuracy of Korean pine and Manchurian walnut is lower in all models, while Manchurian ash and Dragon spruce show greater precision difference in these models. For example, the producer's accuracy of Manchurian ash based on LSTM and RF models is 81% and 82%, respectively, while SVM's accuracy is 50%. The main error of classification results involved Manchurian walnut, Manchurian ash, and Korean pine, which were also classified with low accuracies (59% in the LSTM, 50% in the SVM and 56% in the AlexNet). It is important to note that the classification accuracy of Conv1D model has a higher accuracy than other models. In the Conv1D model, Dahurian larch and Aspen had the highest classification accuracy, reaching as high as 89%, while Manchurian walnut had a lower classification accuracy (72%).

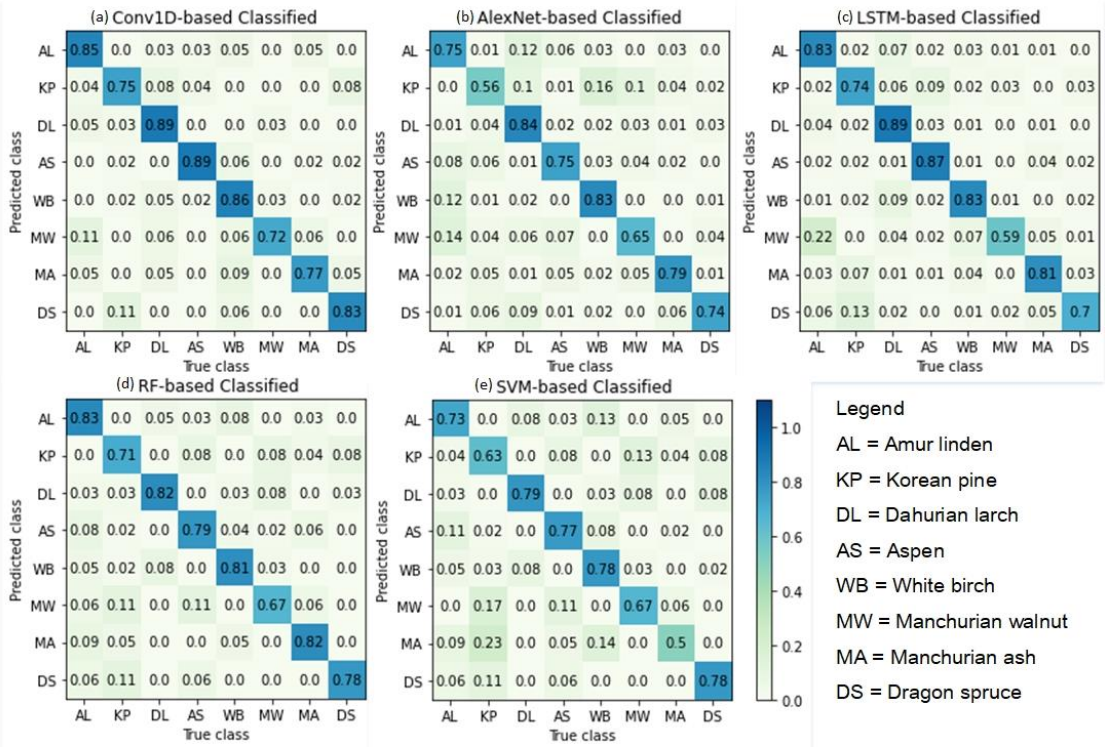

Fig. 11 Confusion matrices show the accuracy of tree species classification using Sentinel-2 images. The diagonal cells show the classification accuracy of each tree species, while the off-diagonal cells show the percentage of classification error between them. Refers (a) to the Conv1D-based classification, (b) to the AlexNet-based classification, (c) to the LSTM-based classification, (d) to the RF-based classification, and (e) to the SVM-based classification.

The classification results of all models are shown in Fig. 12. The proportion of White birch in the study area was the highest (18km^2), about 20%, followed by Manchurian walnut (13km^2) and Manchurian ash (13km^2), about 15%. The area ratio of other tree species ranges from 5% to 12%. These models can effectively distinguish different tree species, but the classification results differed in terms of specific distributions due to the different algorithms of each model. Visual assessment comparing these classification results consistently found that the Conv1D classifier resulted in a more homogeneous and generalized classification of species than other classifiers. However, the spatial distribution of different tree species showed significant differences in all models. Korean pine, for example, is more dispersed in the AlexNet and LSTM models, and more clustered in other models. In addition, for some species with a small proportion (Amur linden), some classifiers (AlexNet and RF) have obvious over-mapping. Assessed against field observations, a comparative analysis of these classification maps shows that, coniferous forests (Dahurian larch and Dragon spruce) are classified with high accuracy and distinct boundaries. For broadleaf forests, Aspen and White birch are widely distributed throughout the entire area, which leads to more misclassification of other species. In terms of tree species distribution, Amur linden is mainly

distributed in the west and southeast of the study area along the ridge. Aspen covers the least area, and the distribution is fragmentary. Dahurian larch is mainly distributed in the northwest and southeast of the study area, while a small amount is distributed in the western parts. Korean pine and Dragon spruce are mainly distributed in the middle of the study area, while Manchurian walnut is distributed in the central and southern parts.

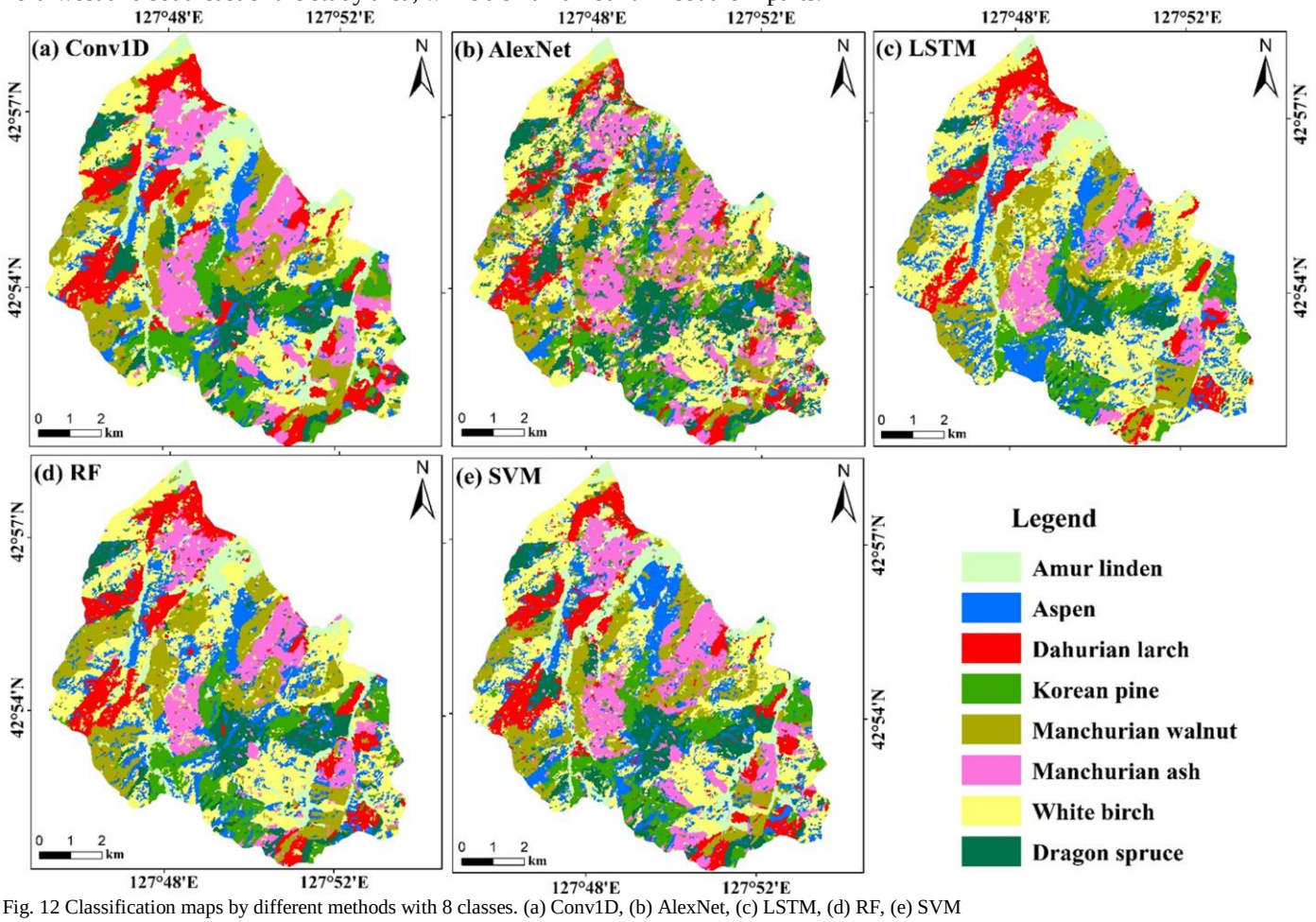

Fig. 12 Classification maps by different methods with 8 classes. (a) Conv1D, (b) AlexNet, (c) LSTM, (d) RF, (e) SVM

V. Discussion

A. Great potential of multi-temporal Sentinel-2 images for tree species classification

Based on Sentinel-2 time series, the utility of five machine learning algorithms were evaluated for mapping tree species; the overall accuracy of all models is higher than 70%, especially deep neural network models (over 80%). Therefore, these results indicate that multi-temporal Sentinel-2 images have great potential for tree species classification, and is closely related to the high spatial, temporal, and spectral resolution.

Previous studies in mixed forests have shown that Landsat images with 30m spatial resolution were not sufficient to accurately classify tree species due to species heterogeneity [46], [54]. Soleimannejad *et al.* [46] examined the potential of medium, semi-high and high spatial resolution data (Landsat-8 OLI sensor, Sentinel-2 MSI, and IRS-Pan sharpened-P5/6) for trees species classification in the Hyrcanian forest of Iran. They obtained more accurate results by using higher spatial resolution data. Higher spatial resolution is necessary for fine discrimination of tree species. Very high resolution (VHR) satellite data (WorldView-2, SPOT, IKONOS, and QuickBird) were successfully used in mapping fine tree species distribution, even individual tree crowns [55]-[57]. However, the use of VHR images is often limited by high costs and limited spatial coverage. Fassnacht *et al.* [1] reviewed many literatures related to the classification of tree species and noted that the best classification accuracy could be obtained when the pixel size was 0.3m or 8m. Freely available Sentinel-2 images have a spatial resolution of 10 m (close to 8 m), which provides an opportunity to achieve large scales and detailed mapping of tree species. Grabska *et al.* [11] achieved a large forest stand species mapping in the Polish Carpathian Mountains (approx. 20,000 km2) with different machine learning algorithms and multi-temporal Sentinel-2 images. Similarly, in an area in southern Poland (over 3800km2), eight tree species were separated with accuracies of over 85% by Hościło and Lewandowska [45]. Therefore, the emergence of Sentinel-2 images (10m spatial resolution) might be a better option for tree species identification, especially at a large spatial extent.

The rich spectral information and unique band setting in Sentinel-2 images play an important role in tree species classification. Previous studies by Mngadi *et al.* [58], Immitzer

et al. [9], Wessel et al. [13], and Persson et al. [12] have revealed the importance of Red-Edge, Near-Infrared (NIR), and Shortwave-Infrared (SWIR) bands for increasing vegetation sensitivity and spectral response. In this study, the importance of these bands was also confirmed by using the MDA algorithm (Fig. 8). The Red-Edge bands are related to vegetation leaf properties, such as photosynthetic pigments, biomass and structural carbohydrates [58], while NIR and SWIR bands are dominated by water content, lignin, starch, and nitrogen [1, 9]. Differences in chemical and physical properties between tree species result in distinctive spectral response, which is the main driver to discriminate species. It is worth noting that in some studies, vegetation indices were equally important in tree species classification [59], [60]. However, vegetation indices (NDVI and EVI) showed a lower sensitivity in this study, which may be related to the higher canopy density.

The high revisit interval of Sentinel-2 provides a great advantage in monitoring vegetation phenology. Phenological variations are more easily captured with the use of multi-temporal imagery, which is helpful for accurate mapping of tree species. As shown in Fig. 8, the Sentinel-2 images in May, June, and July were more sensitive to the phenological variations of tree species in this study. The spectral values of these months represented the temporal dynamic differences in canopy structure and biochemical characteristics among different tree species [45, 61]. Similarly, Immitzer *et al.* [9] concluded that the images in April, May, and June were more beneficial for tree species classification. Persson *et al.* [12] found that late spring and early summer images were optimal for discriminating of species. The same result was obtained by evaluating the importance of time phase. However, Hościło and Lewandowska [45] found that mid-autumn imagery performed the best for tree species classification, which may be connected with the influence of the background signal and should be examined in future studies.

B. The optimal neural network classifier for time series Sentinel-2 images

In this study, three neural networks with different structures were applied to classify tree species. For Conv1D model, we connect each band horizontally into a one-dimensional array according to the timing sequence and extract the features by applying convolution kernel of different sizes. When the convolution kernel is working, it will scan the input features regularly, perform matrix operations on the input features, and add bias values in the receptive field. Then, the convolutional layer applies the learned parameters to the next layer and extracts deeper features. In the AlexNet model, each band is vertically connected into a two-dimensional array by the timing sequence order and converted into a grayscale image. According to these grayscale images, convolution kernels with different sizes are used for feature extraction and classification. The data input of LSTM model is the same as that of Conv1D, which is widely used in dealing with time series problems because it overcomes the problem of gradient disappearance and gradient explosion existing in recurrent neural networks.

Although all three methods can classify time sequential patterns, there are significant differences in the ability to identify trees with similar spectrums. The mean spectral values of different tree species within all years as well as their standard deviations are shown in Fig. 13. The spectral values of Amur Linden and Manchurian Walnut showed no significant difference in distance. However, the classification accuracy of Amur Linden for the three models (Conv1D, AlexNet, and LSTM) was 85%, 75%, and 82%, respectively, and that of Manchurian Walnut was 72%, 65%, and 59%, respectively. Something similar happened with White birch and Aspen. Conv1D achieved higher classification accuracy than AlexNet and LSTM. This is similar to Conv1D having a higher overall accuracy than LSTM in a study focused on crop classification with multi-temporal Landsat imagery in California, USA [30].

Therefore, for sequential data, the application of one-dimensional convolution kernel can better recognize and extract insensitive feature information, which is helpful in the separation of similar tree species. Compared with 2D convolution (AlexNet), the hierarchical feature generation process of the Conv1D-based model provides a flexible way to formulate and identify complex sequential patterns in Sentinel-2 data. Lin *et al.* [37] gave the same explanation when applying Conv1D to predict surface roughness. Similar results were reported by Zhong *et al.* [30] and Yildirim *et al.* [33]. As a non-convolutional neural network, LSTM can select and record longer time series and important nodes, but as shown in Figure 13, the spectral range of some tree species (White birch, Manchurian ash and Korean pine) is quite broad, which may decrease accuracy because error nodes are recorded when processing long sequences.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3098817, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

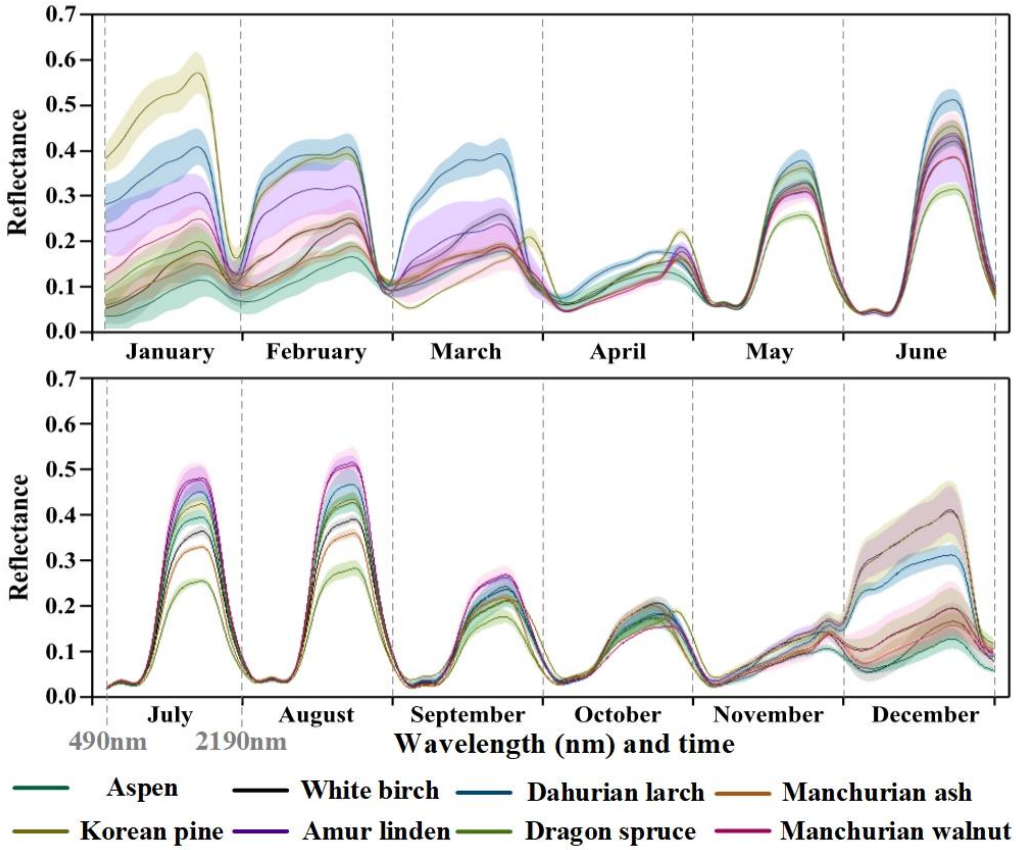

Fig. 13 Spectral reflectance graph of forest species during the year

C. Comparison between deep neural networks and traditional machine learning algorithms

In this study, 288 Sentinel-2 bands of different phases were used for training and classification in the neural network model. For RF and SVM classifier, if all bands were involved for classification, calculation would be difficult due to high dimension and data overload. Therefore, feature extraction is of great significance for the processing of time series of Sentinel-2 data. Fig. 14 shows the outcomes of all methods for our dataset. Raw images are shown on the left, the next five columns on the right contain the prediction results produced by each method. It can be seen that with a spatial resolution of 10m, an accurate boundary of tree species cannot be obtained, especially in mixed coniferous and broadleaf forests, which presents great uncertainty for the classification of tree species. Therefore, the final chosen method might be different image to image since the decision has to be based on the tradeoff between "image type vs cost effectiveness".

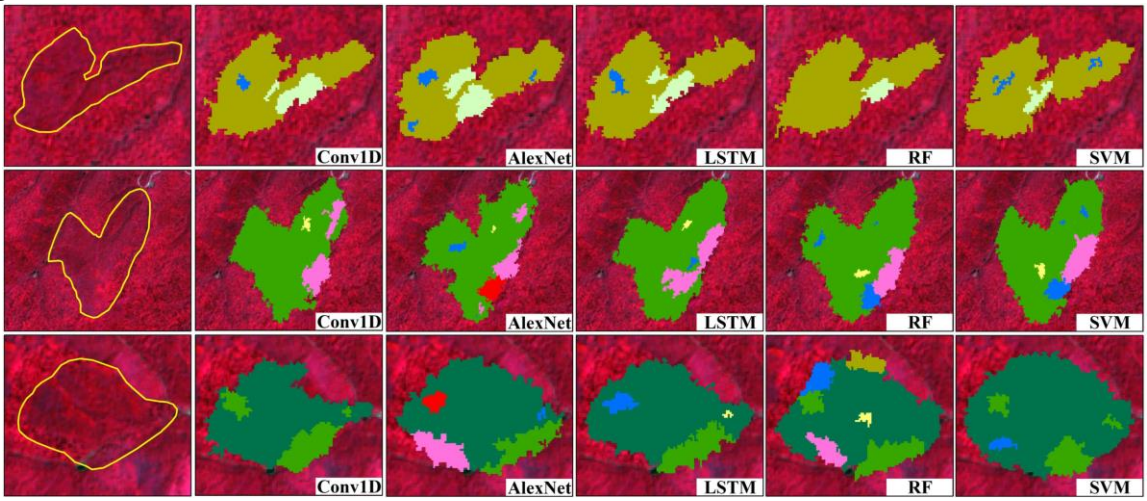

Fig. 14 Visual comparisons of the prediction results from these methods on the Sentinel-2 data set. Notes that the legend for these figures is the same as that in Fig. 12

As noted in the introduction, most studies using multi-temporal Sentinel-2 imagery only used traditional algorithms

(RF and SVM). The RF classifier, as an ensemble method, consists of a combination of tree classifiers, which can reduce overfitting effects and has become the most popular algorithm in remote sensing classification tasks. Similarly, SVMs, as a high-performance algorithm, were developed to use different kernel functions such as the radial basis function to solve non-linear problems. In this study, we found that the RF classifier (79.04%) performed slightly better than the SVM classifier (72.79%). The good performance of RF classifiers was confirmed in other classification studies [16], [55], [61]. However, some studies have also shown contrary results [11], [13]. For large and complex mixed-variable datasets, there are several reasons for performance differences of machine learning (ML) algorithms. The first is the unbalanced training samples (e.g., quality, number, and characteristics of training samples per class). The tree species with fewer samples, such as Korean pine, are often less trained or over-trained by the effects of ML algorithms, which leads to lower accuracies (RF=69.39%; SVM=57.69%) [11], [14]. The second is the patchy nature of the forest structure. As discussed by Wessel *et al.* [13], Sentinel-2 images can achieve higher classification results for forests with less complex structures and background signal effects. Furthermore, the setting of hyperparameters also has an impact on classification results related to the image used. Extensive testing is necessary in order to obtain optimal parameter values, but to some extent, this also brings uncertainty in accuracy.

Generally, the Conv1D classifier provided the highest accuracies, followed by LSTM and RF algorithms. Limited by spatial resolution, deep learning algorithms are rarely applied to medium spatial resolution images, yet medium resolution images of time series provide a new opportunity for the application of deep learning. Our results show that the deep learning algorithm has higher classification accuracy than the traditional algorithm for mixed datasets with high dimensions. This is related to the difference of ML algorithms itself. For instance, when Conv1D carries out the convolution procedure on the local area of the input data, convolution kernel detects specific features on the input feature map to fulfill weight-sharing on the same input feature map, which can effectively reduce the complexity of the network and the number of training parameters [30], [35]. Although deep neural networks provide higher classification accuracy, there are some problems that must be considered, namely, deep learning algorithms requires sufficient training samples, more time, and a large number of hyperparameter setting. For the same large dataset, the traditional machine learning algorithm expends less time. Thus, we need to cautiously select remote sensing images and methods considering the extent, tree species composition complexity and required accuracy.

VI. Conclusions

To explore which ML algorithms can effectively extract various information of images to achieve higher accuracy, three deep neural networks (Conv1D, AlexNet and LSTM) and two traditional ML algorithms (RF and SVM) were used for comparative analysis. The classification results show that neural network models are better than traditional ML algorithms. Among the neural network models, the Conv1D model had the highest classification accuracy (84.19%), followed by the LSTM model (81.52%) and the AlexNet model (76.02%). For non-neural network models, RF's classification accuracy (79.04%) is higher than that of SVM (72.79%), but lower than that of Conv1D and LSTM. For large and complex mixed-variable datasets, deep neural networks can extract deep features and capture certain temporal patterns through multi-layer neural networks to improve classification accuracy. However, unbalanced training samples, the patchy nature of forest structures, and the setting of hyperparameters also affect the accuracy of ML algorithm. Thus, we should choose the appropriate classification algorithm after prudently considering the differences in area, forest structure and required accuracy. This study shows that, among numerous alternative classification methods, deep neural network models combined with multi-temporal Sentinel-2 images provide a new option for tree species mapping in future studies.

Acknowledgment

We thank the National Earth System Science Data Center for providing geographic information data.

References

[1] F. E. Fassnacht et al., "Review of studies on tree species classification from remotely sensed data," Remote Sens. Environ., vol. 186, pp. 64-87, Aug. 2016.

[2] M. A. Cho et al., "Mapping tree species composition in South African savannas using an integrated airborne spectral and LiDAR system," Remote Sens. Environ., vol. 125, pp. 214-226, Aug. 2012.

[3] A. Hovi, L. Korhonen, J. Vauhkonen, and I. Korpela, "LiDAR waveform features for tree species classification and their sensitivity to tree- and acquisition related parameters," Remote Sens. Environ., vol. 173, pp. 224-237, Sep. 2016.

[4] J. Heinzel and B. Koch, "Exploring full-waveform LiDAR parameters for tree species classification," Int. J. Appl. Earth Obs. Geoinf., vol. 13, no. 1, pp. 152-160, Sep. 2011.

[5] M. Dalponte, L. Bruzzone, and D. Gianelle, "Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data," Remote Sens. Environ., vol. 123, pp. 258-270, Apr. 2012.

[6] S. R. Kim et al., "Forest cover classification by optimal segmentation of high resolution satellite imagery," Sensors (Basel), vol. 11, no. 2, pp. 1943-58, Feb. 2011.

[7] Y. Shi et al., "Tree species classification using plant functional traits from LiDAR and hyperspectral data," Int. J. Appl. Earth Obs. Geoinf., vol. 73, pp. 207-219, Jun. 2018

[8] L. Li, N. Li, D. Lu, and Y. Chen, "Mapping Moso bamboo forest and its on-year and off-year distribution in a subtropical region using time-series Sentinel-2 and Landsat 8 data," Remote Sens. Environ., vol. 231, Jun. 2019.

[9] M. Immitzer, M. Neuwirth, S. Böck, H. Brenner, F. Vuolo, and C. Atzberger, "Optimal Input Features for Tree Species Classification in Central Europe Based on Multi-Temporal Sentinel-2 Data," Remote Sens., vol. 11, no. 22, Nov. 2019.

[10] J. Reiche, E. Hamunyela, J. Verbesselt, D. Hoekman, and M. Herold, "Improving near-real time deforestation monitoring in tropical dry forests by combining dense Sentinel-1 time series with Landsat and ALOS-2 PALSAR-2," Remote Sens. Environ., vol. 204, pp. 147-161, Oct. 2018.

[11] E. Grabska, D. Frantz, and K. Ostapowicz, "Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the Polish Carpathians," Remote Sens. Environ., vol. 251, Sep. 2020.

[12] M. Persson, E. Lindberg, and H. Reese, "Tree Species Classification with Multi-Temporal Sentinel-2 Data," Remote Sens., vol. 10, no. 11, Nov. 2018.

[13] M. Wessel, M. Brandmeier, and D. Tiede, "Evaluation of Different Machine Learning Algorithms for Scalable Classification of Tree Types and Tree Species Based on Sentinel-2 Data," Remote Sens., vol. 10, no. 9, 2018.

[14] J. Lim, K.-M. Kim, E.-H. Kim, and R. Jin, "Machine Learning for Tree Species Classification Using Sentinel-2 Spectral Information, Crown Texture, and Environmental Variables," Remote Sens., vol. 12, no. 12, Jun. 2020.

[15] M. Amani, B. Salehi, S. Mahdavi, J. E. Granger, B. Brisco, and A. Hanson, "Wetland Classification Using Multi-Source and Multi-Temporal Optical Remote Sensing Data in Newfoundland and Labrador, Canada," Can. J. Remote Sens., vol. 43, no. 4, pp. 360-373, Apr. 2017.

[16] J. Lim, K.-M. Kim, and R. Jin, "Tree Species Classification Using Hyperion and Sentinel-2 Data with Machine Learning in South Korea and China," ISPRS Int. J. Geo-Inf., vol. 8, no. 3, Mar. 2019.

[17] Y. Wu and X. Zhang, "Object-Based Tree Species Classification Using Airborne Hyperspectral Images and LiDAR Data," Forests, vol. 11, no. 1, Dec. 2019.

[18] B. Zhang, L. Zhao, and X. Zhang, "Three-dimensional convolutional neural network model for tree species classification using airborne hyperspectral images," Remote Sens. Environ., vol. 247, Jun. 2020.

[19] H. Hamraz, N. B. Jacobs, M. A. Contreras, and C. H. Clark, "Deep learning for conifer/deciduous classification of airborne LiDAR 3D point clouds representing individual trees," ISPRS J. Photogramm. Remote Sens., vol. 158, pp. 219-230, Oct. 2019.

[20] Y. Xi et al., "Mapping Tree Species Composition Using OHS-1 Hyperspectral Data and Deep Learning Algorithms in Changbai Mountains, Northeast China," Forests, vol. 10, no. 9, Sep. 2019.

[21] X. Han, Y. Zhong, L. Cao, and L. Zhang, "Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification," Remote Sens., vol. 9, no. 8, Aug. 2017.

[22] Q. Yuan et al., "Deep learning in environmental remote sensing: Achievements and challenges," Remote Sens. Environ., vol. 241, Feb. 2020.

[23] M. P. Ferreira et al., "Individual tree detection and species classification of Amazonian palms using UAV images and deep learning," For. Ecol. Manage., vol. 475, Jul. 2020.

[24] L. Chen, C. Ren, B. Zhang, Z. Wang, and Y. Xi, "Estimation of Forest Above-Ground Biomass by Geographically Weighted Regression and Machine Learning with Sentinel Imagery," Forests, vol. 9, no. 10, 2018.

[25] C. Ren, B. Zhang, Z. Wang, L. Li, and M. Jia, "Mapping Forest Cover in Northeast China from Chinese HJ-1 Satellite Data Using an Object-Based Algorithm," Sensors (Basel), vol. 18, no. 12, Sep. 2018.

[26] J. A. Benediktsson et al., "Sen2Cor for Sentinel-2," presented at the Image and Signal Processing for Remote Sensing XXIII, 2017.

[27] G. M. Gandhi, S. Parthiban, N. Thummalu, and A. Christy, "Ndvi: Vegetation Change Detection Using Remote Sensing and Gis – A Case Study of Vellore District," Procedia Comput. Sci., vol. 57, pp. 1199-1210, 2015.

[28] B. Matsushita, W. Yang, J. Chen, Y. Onda, and G. Qiu, "Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to Topographic Effects: A Case Study in High-density Cypress Forest," Sensors (Basel), vol. 7, no. 11, pp. 2636-2651, Nov. 2007.

[29] L. C. Wensel, J. Levitan, and K. Barber, "Selection of Basal Area Factor in Point Sampling," Journal of Forestry, vol. 78, no. 2, pp. 83-84, 1980.

[30] L. Zhong, L. Hu, and H. Zhou, "Deep learning based multi-temporal crop classification," Remote Sens. Environ., vol. 221, pp. 430-443, Nov. 2019.

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds.: Curran Associates, Inc., pp. 1097—1105, 2012.

[32] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," Physica D, vol. 404, Jan. 2020.

[33] O. Yildirim, U. B. Baloglu, and U. R. Acharya, "A Deep Learning Model for Automated Sleep Stages Classification Using PSG Signals," Int J Environ Res Public Health, vol. 16, no. 4, Feb. 2019.

[34] K. Eckle and J. Schmidt-Hieber, "A comparison of deep networks with ReLU activation function and linear spline-type methods," Neural Netw, vol. 110, pp. 232-242, Feb. 2019.

[35] D. Guidici and M. Clark, "One-Dimensional Convolutional Neural Network Land-Cover Classification of Multi-Seasonal Hyperspectral Imagery in the San Francisco Bay Area, California," Remote Sens., vol. 9, no. 6, Jun. 2017.

[36] B. Liu, Y. Li, G. Li, and A. Liu, "A Spectral Feature Based Convolutional Neural Network for Classification of Sea Surface Oil Spill," ISPRS Int. J. Geo-Inf., vol. 8, no. 4, Mar. 2019.

[37] W.-J. Lin, S.-H. Lo, H.-T. Young, and C.-L. Hung, "Evaluation of Deep Learning Neural Networks for Surface Roughness Prediction Using Vibration Signal Analysis," Appl. Sci., vol. 9, no. 7, Apr. 2019.

[38] N. Arslan and A. Sekertekin, "Application of Long Short-Term Memory neural network model for the reconstruction of MODIS Land Surface Temperature images," J. Atmos. Sol. Terr. Phys., vol. 194, Aug. 2019.

[39] T. Santosh, D. Ramesh, and D. Reddy, "LSTM based prediction of malaria abundances using big data," Comput Biol Med, vol. 124, p. 103859, Sep 2020.

[40] C. Xiao, N. Chen, C. Hu, K. Wang, J. Gong, and Z. Chen, "Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach," Remote Sens. Environ., vol. 233, Aug. 2019.

[41] Q. Yin, M. Liu, J. Cheng, Y. Ke, and X. Chen, "Mapping Paddy Rice Planting Area in Northeastern China Using Spatiotemporal Data Fusion and Phenology-Based Method," Remote Sensing, vol. 11, no. 14, Jul. 2019.

[42] C. Zhang, D. R. Mishra, and S. C. Pennings, "Mapping salt marsh soil properties using imaging spectroscopy," ISPRS J. Photogramm. Remote Sens., vol. 148, pp. 221-234, Jan. 2019.

[43] L. Chen, Y. Wang, C. Ren, B. Zhang, and Z. Wang, "Optimal Combination of Predictors and Algorithms for Forest Above-Ground Biomass Mapping from Sentinel and SRTM Data," Remote Sens., vol. 11, no. 4, Feb. 2019.

[44] R. Blomley, A. Hovi, M. Weinmann, S. Hinz, I. Korpela, and B. Jutzi, "Tree species classification using within crown localization of waveform LiDAR attributes," ISPRS J. Photogramm. Remote Sens., vol. 133, pp. 142-156, Aug. 2017.

[45] A. Hościło and A. Lewandowska, "Mapping Forest Type and Tree Species on a Regional Scale Using Multi-Temporal Sentinel-2 Data," Remote Sens., vol. 11, no. 8, Apr. 2019.

[46] L. Soleimannejad, S. Ullah, R. Abedi, M. Dees, and B. Koch, "Evaluating the potential of sentinel-2, landsat-8, and irs satellite images in tree species classification of hyrcanian forest of iran using random forest," J. Sustainable For., vol. 38, no. 7, pp. 615-628, Feb. 2019.

[47] V. N. Vapnik, "An Overview of Statistical Learning Theory," IEEE Trans. Neural Netw. Learn. Syst., vol. 10, pp. 988-999, 1999.

[48] Z. Chunhui, G. Bing, Z. Lejun, and W. Xiaoqing, "Classification of Hyperspectral Imagery based on spectral gradient, SVM and spatial random forest," Infrared Phys. Technol., vol. 95, pp. 61-69, Oct. 2018.

[49] D. McInerney, P. Kempeneers, M. Marron, and R. E. McRoberts, "Analysis of broadleaf encroachment in coniferous forest plantations using multi-temporal satellite imagery," Int. J. Appl. Earth Obs. Geoinf., vol. 78, pp. 130-137, Dec. 2019.

[50] W. Tang, J. Hu, H. Zhang, P. Wu, and H. He, "Kappa coefficient: a popular measure of rater agreement," Shanghai Arch Psychiatry, vol. 27, no. 1, pp. 62-7, Feb. 2015.

[51] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation," in AI 2006: Advances in Artificial Intelligence (Lecture Notes in Computer Science, pp. 1015-1021, 2006.

[52] S. Nezami, E. Khoramshahi, O. Nevalainen, I. Pölönen, and E. Honkavaara, "Tree Species Classification of Drone Hyperspectral and RGB Imagery with Deep Learning Convolutional Neural Networks," Remote Sens., vol. 12, no. 7, Mar. 2020.

[53] E. Grabska, P. Hostert, D. Pflugmacher, and K. Ostapowicz, "Forest Stand Species Mapping Using the Sentinel-2 Time Series," Remote Sens., vol. 11, no. 10, May. 2019.

[54] M. L. Clark, "Comparison of multi-seasonal Landsat 8, Sentinel-2 and hyperspectral images for mapping forest alliances in Northern California," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 159, pp. 26-40, Nov. 2020.

[55] M. Immitzer, C. Atzberger, and T. Koukal, "Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data," Remote Sensing, vol. 4, no. 9, pp. 2661-2693, Sep. 2012.

[56] Y. Ke, L. J. Quackenbush, and J. Im, "Synergistic use of QuickBird multispectral imagery and LIDAR data for object-based forest species classification," Remote Sens. Environ., vol. 114, no. 6, pp. 1141-1154, Jan. 2010.

[57] D. Lobo Torres et al., "Applying Fully Convolutional Architectures for Semantic Segmentation of a Single Tree Species in Urban Environment on

High Resolution UAV Optical Imagery," Sensors (Basel), vol. 20, no. 2, Jan. 2020.

[58] M. Mngadi, J. Odindi, K. Peerbhay, and O. Mutanga, "Examining the effectiveness of Sentinel-1 and 2 imagery for commercial forest species mapping," Geocarto Int., pp. 1-12, Feb. 2019.

[59] G. Vaglio Laurin et al., "Discrimination of tropical forest types, dominant species, and mapping of functional guilds by hyperspectral and simulated multispectral Sentinel-2 data," Remote Sens. Environ., vol. 176, pp. 163-176, Feb. 2016.

[60] T.-B. Ottosen, G. Petch, M. Hanson, and C. A. Skjøth, "Tree cover mapping based on Sentinel-2 images demonstrate high thematic accuracy in Europe," INT J APPL EARTH OBS., vol. 84, Aug. 2020.

[61] Y. Meng et al., "Tree Species Distribution Change Study in Mount Tai Based on Landsat Remote Sensing Image Data," Forests, vol. 11, no. 2, Jan. 2020.

Yanbiao Xi received the M.S. degree in Cartography and geography information system from University of Chinese Academy of Sciences, Beijing, China. He is currently pursuing the Ph.D. degree in geography at Nanjing University, Nanjing, China.
His research interest includes forest mapping and monitoring, machine learning algorithm, and land cover classification.

Chunying Ren received the Ph.D. degree in Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, China, in 2008.
She is currently a Professor of Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences. Her research interests include forest remote sensing, remote sensing data intelligent processing, and Google Earth Engine. Dr.Ren serves as a Regular Reviewer for the multiple journals such as Remote Sensing of Environment, Remote Sensing, Forest Ecology and Management, and so on.

Qingjiu Tian received the Ph.D. degree in Nanjing University, Nanjing, China, in 2003.
He is currently a Professor of International Institute for Earth System Science, Nanjing University. He has authored over 200 peer-reviewed journal papers in the Remote Sensing of Environment, ISPRS Journal of Photogrammetry and Remote Sensing, and so on. Dr. Tian serves as an Associate Editor-in-Chief for journal of remote sensing (Chinese). He serves as a Regular Reviewer for the multiple journals such as Remote Sensing, ISPRS Journal of Photogrammetry and Remote Sensing, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and forests.

Yongxing Ren received the M.S. degree in Jilin University, Changchun, China. He is currently pursuing the Ph.D. degree in geography at Jilin University, Changchun, China.
His research interest includes soil contamination, machine learning algorithm, and environment remote sensing.

Xinyu Dong received the B.S and M.S. degrees from the Fujian Agriculture and Forestry University, Fuzhou, China, in 2015 and 2018, respectively.
He is currently pursuing the Ph.D. degree with Nanjing University, Nanjing, China. His research interests include remote sensing image processing and object detection on medium or high-resolution remote sensing images

Zhichao Zhang received the B.S and M.S. degrees from the Nanjing Agricultural University, Nanjing, China, in 2012 and 2016, respectively.
He is currently pursuing the Ph.D. degree with Nanjing University, Nanjing, China. His research interests include remote sensing target detection and recognition.