



Linking random forest and auxiliary factors for extracting the major economic forests in the mountainous areas of southwestern Yunnan Province, China

Pei Huang^a, Xiaoqing Zhao^{b,*}, Junwei Pu^a, Zexian Gu^{a,c}, Yan Feng^b, Shijie Zhou^b, Xinyu Shi^b, Yuanyuan Tang^a, Pinliang Dong^d

^a Institute of International Rivers & Eco-Security, Yunnan University, Kunming 650500, China

^b School of Earth Sciences, Yunnan University, Kunming 650500, China

^c Nujiang Forestry and Grassland Administration, Lushui 673100, China

^d Department of Geography and the Environment, University of North Texas, Denton, TX 76203, USA

ARTICLE INFO

Keywords:

Economic forest
Precise extraction of vegetation information
Site conditions
Vegetation indices
Random forest
Mountainous areas of southwestern Yunnan Province

ABSTRACT

Forests are generally extracted from remotely sensed images based on the spectral features, ignoring other important auxiliary information, and the techniques of precise extraction need to be further improved. By using the Sentinel-2 image and auxiliary factors (AFs) including site conditions (SCs) and vegetation indices (VIs), the random forest model with AFs (RF-AFs) was adopted for the extraction of the economic forests in Lancang County, which is a mountainous area with rich biodiversity and is witnessing rapid development of economic forests in Yunnan province of China. The results obtained using the RF-AFs model were compared with those obtained using the random forest model without AFs (RF). The results were as follows: (1) The kappa coefficient for extracting the first-level land use obtained using the RF model was 0.9531. Lancang County is dominated by forests, accounting for 73.76% of the total area. (2) After parameter optimization, the RF-AFs model yielded the highest accuracy in the extraction of the second-level forests, with a kappa coefficient value of 0.9493, which was 14.69% higher than that of the RF model. Thus, the RF-AFs model is more suitable for the precise extraction of economic forests. (3) The evaluation results of the factors' importance of the RF-AFs model showed that the cumulative importance values of SCs such as temperature (TEM), elevation (EL), precipitation (PRE) and VIs such as plant senescence reflectance index (PSRI), enhanced vegetation index (EVI), transformed soil-adjusted vegetation index (TSAVI) was 76.09%, indicating that they were the main factors for the extraction of economic forests. (4) Economic forests are dominated by Simao pines in Lancang County, which are mainly distributed in the central, southwestern and northern regions, accounting for 31.37% of forests area. The proportion of tea plantations, eucalyptus, and rubber trees is 9.05%, 6.71%, and 3.05% of forests area, respectively. The RF-AFs model is conducive for precisely extracting the economic forests and is thus of great significance in studying the ecological and environmental effects of economic forests, performing forestry management, and maintaining regional ecological security.

1. Introduction

As an important component of the ecosystem, vegetation governs the balance of the ecosystem, and acts as an indicator of global change (Yang et al., 2015). Extraction of vegetation information is the basis for studying the coverage status and dynamic changes in the vegetation characteristics. The vegetation distribution, growth, and vigor aid in

environmental monitoring, forestry management, and maintenance of regional ecological security (Savchenko et al., 2020; Wang et al., 2021; Kureel et al., 2022). Efficient and effective extraction of vegetation is the basis for forest research and precise management. Due to the differences in reflectance and emissivity, different vegetation types present different spectral features in the remotely sensed images. Remote sensing technology is one of the most effective methods for extracting vegetation

Abbreviations: SCs, Site conditions; VIs, Vegetation indices; RF, Random forest.

* Corresponding author.

E-mail address: xqzhao@ynu.edu.cn (X. Zhao).

<https://doi.org/10.1016/j.ecolind.2023.110025>

Received 8 September 2022; Received in revised form 31 January 2023; Accepted 12 February 2023

Available online 24 February 2023

1470-160X/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

information (Pal et al., 2018; Fu et al., 2021).

With the development of remote sensing and computer technologies, extensive research activities have been conducted on vegetation information extraction at different scales (Reddy et al., 2015; Du et al., 2018; Geng et al., 2019; Zhao et al., 2019; Li et al., 2020). In terms of data sources, the available satellites for vegetation remote sensing information extraction include MODIS, NOAA, Landsat, SPOT, Sentinel, among others. Sentinel-2 has the features of a double-star network (including A and B satellites), large width, short revisit period, high spatial resolution, free data acquisition, rich waveband information and unique red-edge waveband, thus offering great advantages in vegetation information extraction (Wakulińska and Marcinkowska-Ochtyra, 2020; Pu et al., 2021). In terms of methods for vegetation information extraction, existing methods include visual interpretation (Fritz et al., 1999), supervised and unsupervised classification (Mohammady et al., 2015), expert knowledge (Visser et al., 2018), object-oriented classification (Owers et al., 2016), phenology-based feature extraction (Cao et al., 2016), multi-source remote sensing data fusion (Sankey et al., 2017), machine learning (Liu et al., 2021), and mixed pixel decomposition (Lima et al., 2017). Among them, the machine learning algorithm continuously improves the training and learning ability of the model through repeated learning, and plays an important role in the precise classification of medium and high-resolution vegetation remote sensing (Carranza-García et al., 2019). It further promotes the automatic and intelligent development of the object information extraction. However, different machine learning methods have different accuracy. Studies have found artificial neural network (ANN), support vector machine (SVM) and random forest (RF) generally provide better accuracy compared to other traditional methods (Thanh Noi and Kappas, 2017), while RF has better performance in vegetation information extraction compared to other machine learning techniques (Mao et al., 2020; Talukdar et al., 2020). RF, a machine learning method based on ensemble learning, is advantageous for quickly and efficiently processing large amounts of nonlinear multidimensional data under the combined effect of the randomness of the training sets and the optimal property of node splitting (Melville et al., 2018). The RF model has a strong generalization ability, which can help improve the accuracy and efficiency of the object information extraction (Tian et al., 2016). In recent years, the RF model has been effectively employed for vegetation information extraction (Liu et al., 2018; Niculescu et al., 2020). Most of the aforementioned vegetation information extraction methods are based on the spectral features of the remotely sensed images themselves. However, the phenomenon of different objects having the same spectra and same objects having different spectra is prominent, which leads to great limitation of extracting vegetation information only by the spectral features of remotely sensed images (Le and Ha 2019). Various vegetation indices (VIs) offer a deeper exploration of the spectral features of remote sensing images and are often used as important auxiliary information in vegetation extraction (Sun et al., 2019a; Wu and Zhang 2019; Bai et al., 2021). Based on different band operations, VIs can reduce the interference of vegetation information caused by similar spectral features of remotely sensed images and aid in the precise extraction of vegetation information (Fadaei, 2020; Suchenwirth et al., 2012). Besides, the distribution of different vegetation is also affected by site conditions (SCs) such as topography, climate, soil, and human activities (Waśniewski et al., 2020). Therefore, it is necessary to combine multiple auxiliary information features for vegetation information extraction (Luo et al., 2015). Although some studies have combined vegetation index with elevation factor to extract vegetation information (Barzegar et al., 2015; Liu et al., 2018), the research that combines the VIs with complex SCs such as elevation, slope, climate, and human activities to extract vegetation information is relatively weak. However, vegetation growth is closely related to the above SCs. The role of SCs and VI in the precise extraction of vegetation information is still unclear.

Economic forests are forests which aim to produce wood or other forest products and obtain direct economic benefits (Brukas et al.,

2014). Economic forests have the characteristics of strong photosynthesis, a fast growth rate, and high economic efficiency (Goded et al., 2019). Driven by economic interests, economic forests such as eucalyptus, rubber trees, Simao pines (*Pinus kesiya* var. *Langbianensis*), and tea plantations have witnessed rapid development in China in the past 20 years (Nong et al., 2019; Zhao et al., 2021; Zhao and Gu 2021). However, the large-scale planting of economic forests has led to great changes in the structure and pattern of regional land use. These changes not only threaten the growth space of natural forests but also change the original ecosystem and cause changes in regional ecological security effects (Ahrends et al., 2015; Zhou et al., 2020), especially in Yunnan Province, which is the most biologically diverse province in China (Chu et al., 2019). Therefore, the extraction of economic forests is beneficial for the timely and precise analysis of the spatial distribution pattern of economic forests, and is of great significance for forest management and regional ecological security. However, research on extraction of economic forests is mostly based on direct extraction by using spectral features (Zhao and Xu 2015; Oliveira et al., 2021), with insufficient attention on other important auxiliary information, and the accuracy of information extraction needs to be further improved. Whether considering other AFs of economic forests will help realize a more precise extraction of economic forests is a question worth exploring.

Studies on forest information extraction have rarely focused on the extraction of economic forests in plateaus and mountainous areas with rich biodiversity, large-scale, and intensive planting. Therefore, in this study, by taking Lancang County of southwest Yunnan Province in China (with rich biodiversity and large-scale planting of economic forests) as an example and using the spectral features of field samples in the Sentinel-2A image, training and testing samples of various land use types were selected. Further, AFs including SCs and VIs, were integrated into the RF-AFs model to establish a method for economic forests extraction. The contributions of this study are as follows: (1) Facing the current situation that the research on the precise extraction of economic forests is relatively weak, this study can enrich the theory and method research of precise extraction of economic forests. (2) The RF-AFs model combining VIs and complex SCs is constructed to extract economic forests information, which proves that the simultaneous inclusion of VIs and complex SCs can effectively improve the accuracy of economic forests extraction. (3) The key factors in the precise extraction of economic forests are identified, which can provide factor reference for the precise extraction of economic forests and other forests.

2. Materials and methods

2.1. Study area

Lancang County, with an area of 8807 km², is located in the southwest of Yunnan Province (99°29'~100°35'E, 22°01'~23°16'N), and south of the Hengduan Mountains. It has a high-altitude terrain in the northwest and a low-altitude terrain in the southeast, with large differences in elevation; the highest elevation is 2515 m, and the lowest is 468 m. The vertical climate is obvious, mainly belongs to the subtropical monsoon climate, with abundant rainfall, sufficient sunshine, and distinct dry and rainy seasons. Many rivers belonging to the Lancang River system pass through the study area. The zonal soils include lateritic red earths, humid-thermo ferralitic, red earths, and yellow-brown earths; among these, lateritic red earths are the most widely distributed. Lancang County has complex site conditions and diverse vegetation types. The zonal vegetation comprises subtropical deciduous broadleaf forests and coniferous forests. Large-scale economic forests of Simao pines, eucalyptus, rubber trees and tea plantations have been planted in Lancang County, resulting in extensive changes in the landscape (Zhao et al., 2021). In 2020, the GDP of Lancang County was 12.007 billion yuan, the population was 441,500, and the urbanization rate was 37 %. The geographic location of Lancang County is shown in Fig. 1.

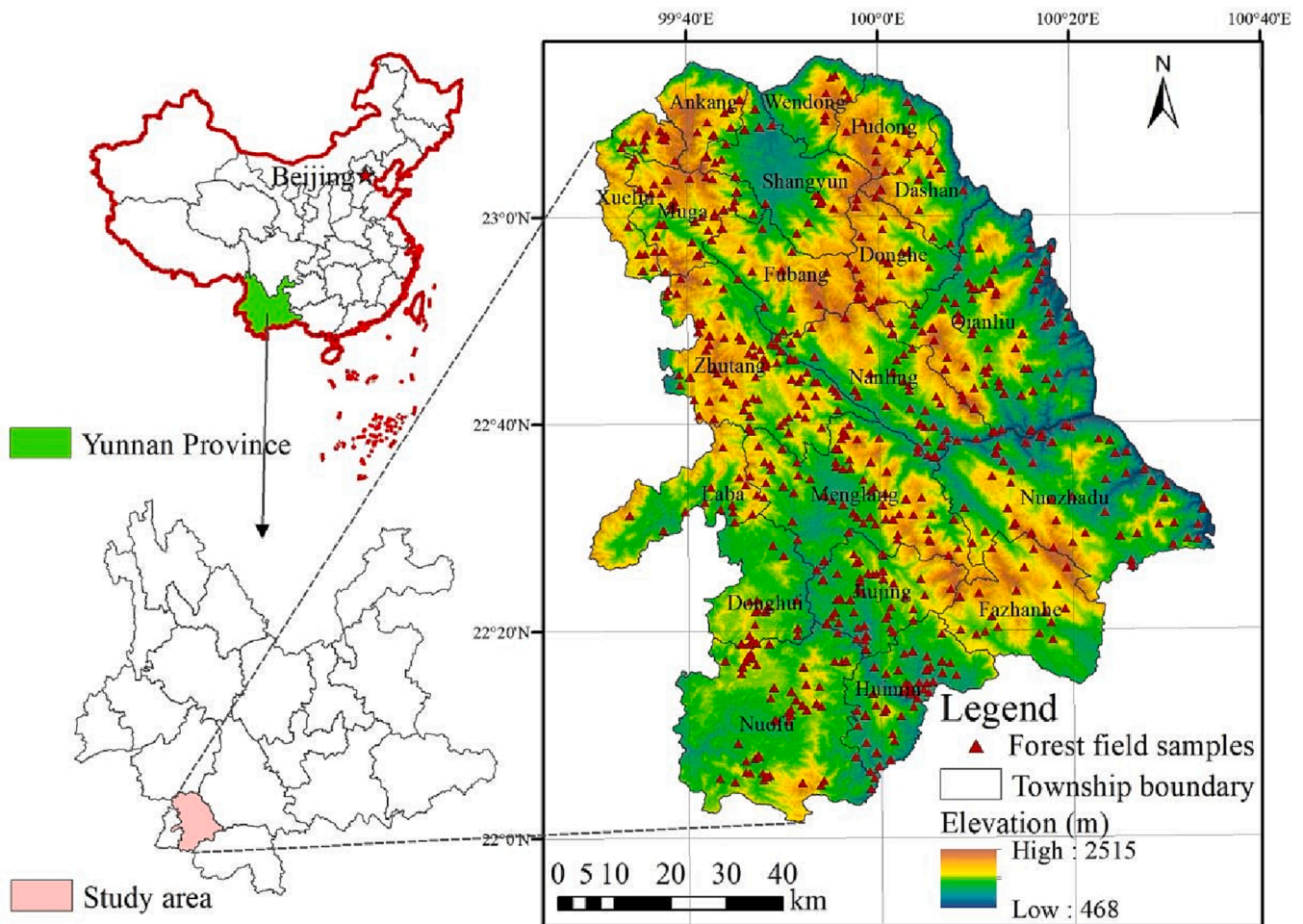


Fig. 1. The location of Lancang County.

2.2. Data source and preprocessing

2.2.1. Satellite image and AFs data

The datasets in this study include a satellite image and AFs (Fig. 2). The Sentinel-2 image was downloaded from the United States Geological Survey (USGS) website (Brede et al., 2020). It was taken on February 4, 2020, with cloud cover below 5%. The Sentinel-2A image, after the process of atmospheric correction and radiometric calibration, was resampled to a pixel size of 10 m × 10 m. AFs included SCs and VIs. The SCs were elevation (EL), slope (SL), aspect (AS), soil type (ST), temperature (TEM), precipitation (PRE) and distance from settlements (DS). The EL, SL and AS data were extracted from digital elevation model (DEM) provided by the Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences (Pu et al., 2021), with a spatial resolution of 30 m × 30 m. The ST data, including red earths, purplish soils, lateritic red earths, yellow earth, yellow-brown earths, paddy soils and humid-thermo ferralitic, was procured from the China soil map based harmonized world soil database (Fischer et al., 2008), with a spatial resolution of 30 m × 30 m. The TEM and PRE were respectively the average annual temperature and annual precipitation of Lancang County in 2020 with a resample resolution of 30 m × 30 m, which were obtained from the National Earth System Science Data Center, National Science & Technology Infrastructure of China (Shi et al., 2022). Referring to the Sentinel-2A image, the construction lands after removing the road and industrial and mining lands in the first-level land use were regarded as the settlements. The buffer zone results of the settlements were regarded as DS data. Combining the infrared and near-infrared bands to create VIs is effective for extracting vegetation information (Fakhri

et al., 2019). With the development of hyperspectral and thermal infrared remote sensing technology, wavebands of satellite images are becoming increasingly abundant (Sun et al., 2019b). Based on previous research (Ren and Feng, 2015; Dong et al., 2019; Solymsi et al., 2019; Wei et al., 2020; Zhu et al., 2021) and considering the actual situation of the Lancang County, the VIs including Ratio Vegetation Index (RVI), Normalized Differential Vegetation Index (NDVI), NDVI of Red-edge Index (NDVire1), Perpendicular Vegetation Index (PVI), Enhanced Vegetation Index (EVI), Soil-Adjusted Vegetation Index (SAVI), Transformed Soil-Adjusted Vegetation Index (TSAVI), Renormalized Difference Vegetation Index (RDVI) and Plant Senescence Reflectance Index (PSRI) are selected and presented in Table 1. The VIs were calculated based on different bands of the Sentinel-2A image. In order to be consistent with the resolution of the image, all AFs data were created or resampled as rasters with a resolution of 10 m × 10 m.

2.2.2. Training and testing samples

The field samples which were obtained via field investigation represented actual land use types. Due to the limited number of field samples, especially for the extraction of vegetation in a large area, it is not enough to use only field samples as the training and testing samples of the model. Large-scale vegetation extraction only by the field samples are quite labor-intensive, and requiring more material resources (eg. survey equipments) and financial resources. Based on the spectral features of the field samples on the satellite image and referenced with the same period of Google Earth high-resolution images online, total samples with the same characteristics as the field samples were selected for the RF model and RF-AFs model by the visual interpretation method

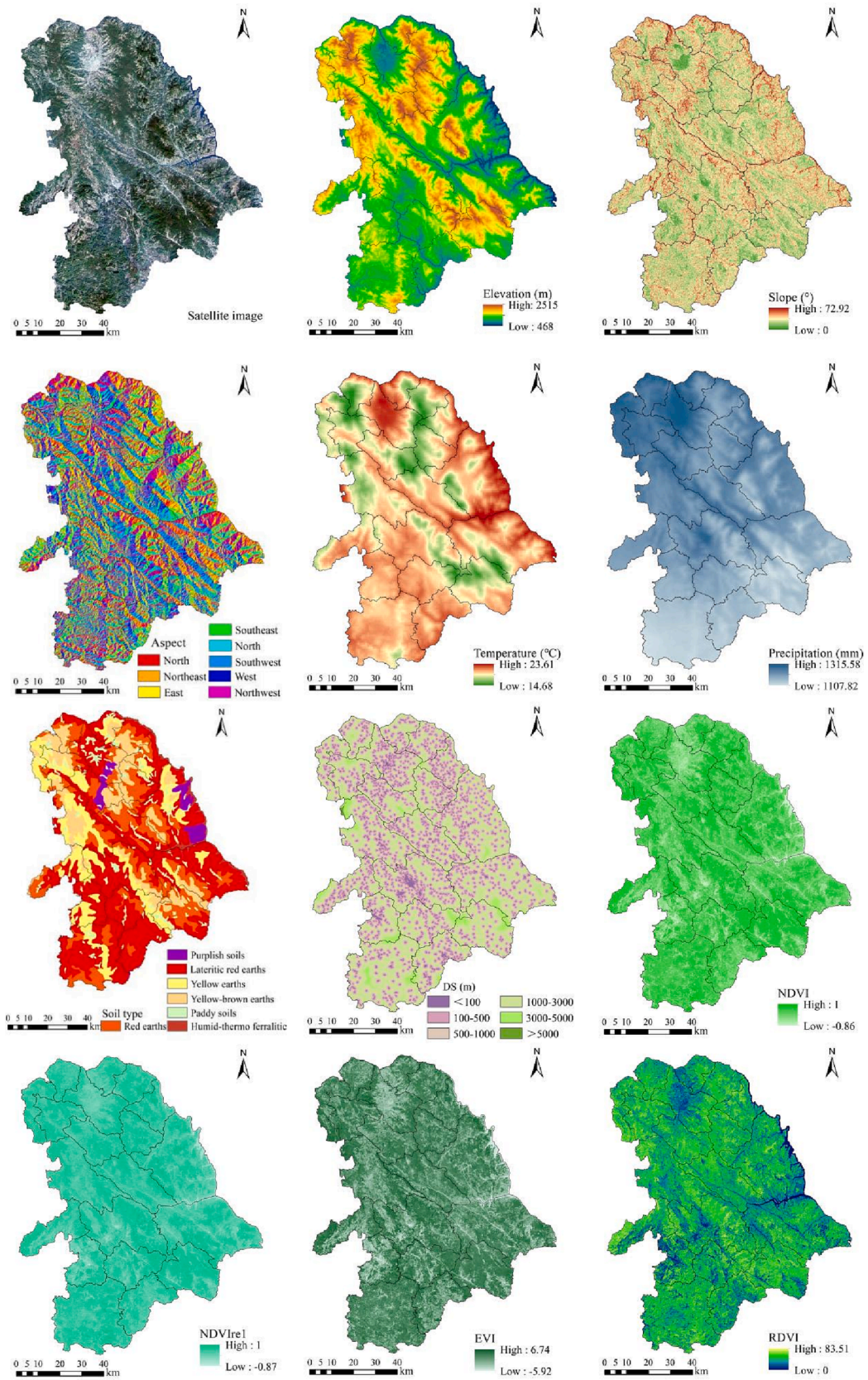


Fig. 2. Satellite image and auxiliary factors.

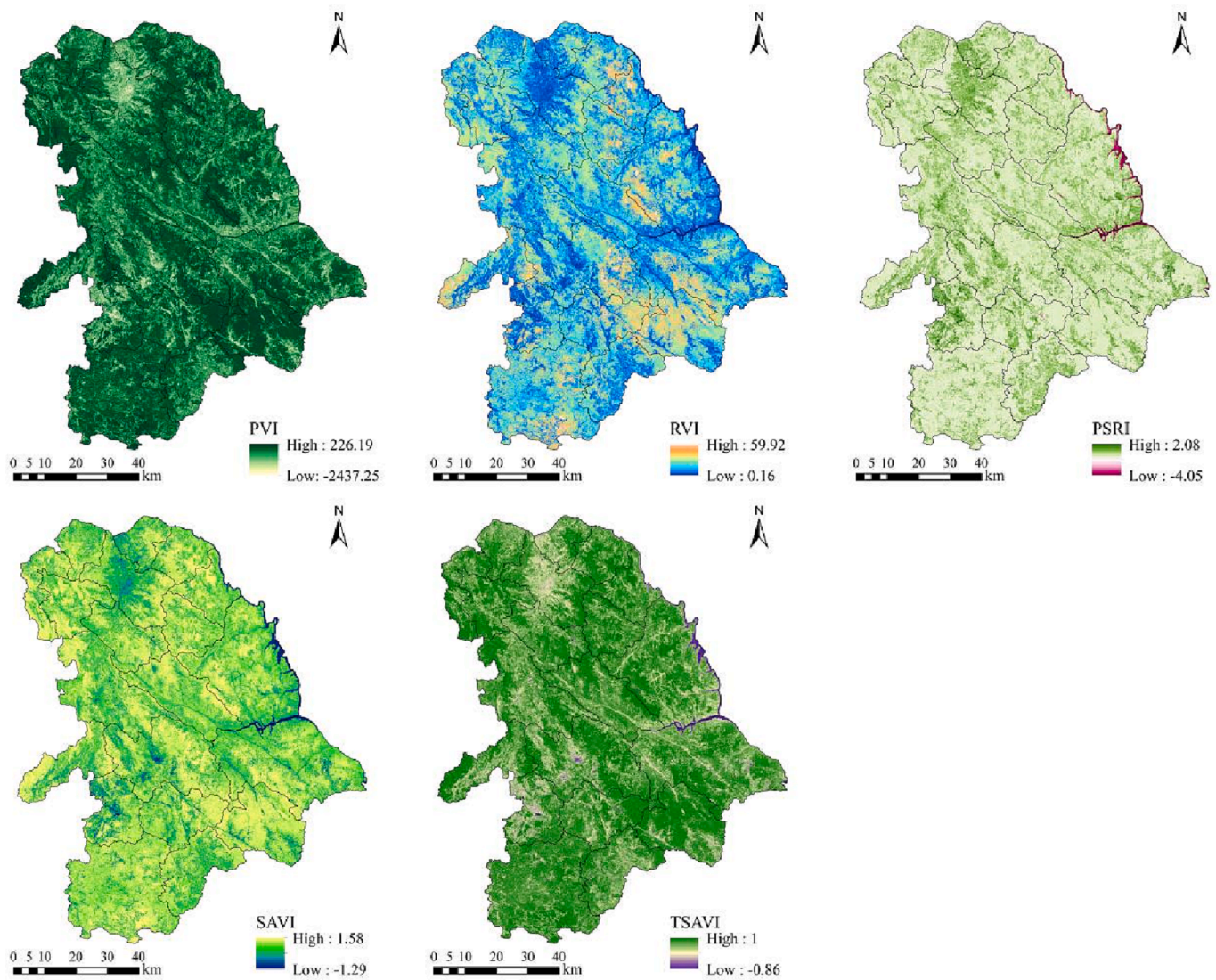


Fig. 2. (continued).

(Thanh Noi and Kappas, 2017). The separability between the different samples of the first-level land use types are above 1.93, while the samples of the second-level forests types are above 1.89, indicating that the samples confusion between different land use type is very low. Training and testing samples were randomly sampled from the total samples by the createDataPartition() function in RStudio with a ratio of 3: 1 (Table 2).

2.3. Methods

2.3.1. Technical flowchart

The technical flowchart is illustrated in Fig. 3. First, the Sentinel-2 image was processed by atmospheric correction and radiometric calibration using the Sen2cor software developed by the European Space Agency (ESA), and the SNAP software was used for band fusion and resampling to acquire a multispectral high-resolution Sentinel-2A image. Second, based on the training samples 1, testing samples 1, and the satellite image data, the RF model in the ENVI platform was used to extract the first-level land use including croplands, forests, water bodies, construction lands, and unutilized lands. The model accuracy was evaluated. Finally, the economic forests were extracted precisely. To verify the advantages of the RF-AFs model in the precise extraction of forests, training samples 2 and testing samples 2 were inputted into the

RF-AFs model and RF model to extract the economic forests of eucalyptus, Simao pines, rubber trees and tea plantations. The accuracy of each model were compared to determine the optimal results. It is difficult for the ENVI platform to integrate multiple AFs into the RF model. RS is a free, open-source, and convenient platform for data statistics and analysis. In RStudio, there are many complete function packages that provide important tools for integrating multiple AFs into the RF model to extract forests information (Grabska et al., 2019). Therefore, the precise extraction of economic forests by using the RF-AFs model was implemented through programming with caret, randomForest package in RStudio of version 1.3.1073.

2.3.2. Random forest (RF) model

The RF model is a machine learning method that combines the bagging method to generate multiple mutually independent training sets and multiple classification and regression trees (CART) for prediction (Huang et al., 2020a). The basic principle (Breiman, 2001) is described as follows:

- 1) k samples are randomly extracted from the original training set D by using the bootstrap sampling method to construct k decision trees models.
- 2) n ($n \leq m$, where m is the total number of features in the sample) features are randomly selected from each sample as the split feature set;

Table 1
Vegetation indices.

Index	Abbreviation	Formula
Ratio Vegetation Index	RVI	$RVI = \rho_8 / \rho_4$
Normalized Differential Vegetation Index	NDVI	$NDVI = \frac{\rho_8 - \rho_4}{\rho_8 + \rho_4}$
NDVI of Red-edge Index	NDVire1	$NDVire1 = \frac{\rho_{8A} - \rho_5}{\rho_{8A} + \rho_5}$
Perpendicular Vegetation Index	PVI	$PVI = \frac{\rho_8 - a\rho_4 - b}{\sqrt{1 + a^2}}$
Enhanced Vegetation Index	EVI	$EVI = \frac{2.5(\rho_8 - \rho_4)}{\rho_8 + 6.0\rho_4 - 7.5\rho_2 + 1}$
Soil-Adjusted Vegetation Index	SAVI	$SAVI = \left(\frac{\rho_8 - \rho_4}{\rho_8 + \rho_4 + L}\right) \times (1 + L)$
Transformed Soil-Adjusted Vegetation Index	TSAVI	$TSAVI = \frac{a(\rho_8 - a\rho_4 - b)}{a\rho_8 + \rho_4 - ab}$
Renormalized Difference Vegetation Index	RDVI	$RDVI = \sqrt{NDVI} \times (\rho_8 - \rho_4)$
Plant Senescence Reflectance Index	PSRI	$PSRI = \frac{\rho_4 - \rho_3}{\rho_6}$

* Note: where $\rho_2, \rho_3, \rho_4, \rho_5, \rho_6$, and ρ_{8A} are the reflectivity corresponding to the blue, green, red, vegetation red edge (band 6), NIR, and vegetation red edge (band 8A) bands in the Sentinel-2A image, reflectivity; a and b are the slope and interception of the soil line, respectively. ($a = 10.849$, $b = 6.604$); and L is the adjustment factor ($L = 0.5$) (Solymosi et al., 2019).

then, the optimal features are determined to grow the nodes. The minimum Gini coefficient is used as the optimal feature splitting criterion.

3) Each tree is grown as much as possible and is not pruned. Random forests are generated by repeating the above steps.

4) Based on the k classification results, each record is voted separately to obtain the final classification result. The importance of each feature in the classification results can be calculated by the mean decreased value of the Gini coefficient at the node split.

The $mtry$ and $ntree$ are two key parameters in the RF model, and the optimal $mtry$ and $ntree$ parameters directly determine the superiority of the extraction results of economic forests (Dobrinić et al., 2021). The $mtry$ is the number of variables randomly sampled as candidates at each split, and the $ntree$ is the number of CART in random forests. These two parameters are determined by the out-of-bag (OOB) error generated during the RF model construction (Nina and John, 2016). The $mtry$ and $ntree$ in the RF-AFs model were calculated using the Random Forest function package in RStudio (Sothe et al., 2020).

2.3.3. Model training and accuracy assessment

The above training and testing samples (Table 2) were mainly used for model training and accuracy evaluation of each model. By using R programming technology, the AFs such as SCs and VIs were linked to the forest types, and the RF-AFs model for precise extracting of economic forests are constructed. Training and testing samples were inputted into the RF-AFs model for model training, parameter optimization, and accuracy evaluation (Fig. 4). For the second-level forests, all the training samples 2 and testing samples 2 in the RF-AFs model included 16 AFs and a unique forest type. Except for the training samples 2 and testing samples 2, the remaining forests samples (including only 16 AFs) were fed into the trained RF-AFs model to identify the specific forest types. The RF model only combined training samples 2 and testing samples 2 (not including AFs) and the satellite image to extract the forest information and evaluate the accuracy directly.

Accuracy assessment is an important step after object classification (Abdi, 2020). The overall accuracy (OA) and kappa coefficient are often used to evaluate the accuracy of the classification results. To compare the effectiveness of different models in the extraction of economic forests, overall accuracy and kappa coefficient were used for the accurate evaluation of the RF model and RF-AFs model.

3. Results

3.1. Extraction of the first-level land use

3.1.1. Accuracy assessment

The training samples 1, testing samples 1, and satellite image data were inputted into the RF model of the ENVI platform for model training and accuracy assessment. The confusion matrix and accuracy of the extraction of the first-level land use is shown in Table 3. The overall accuracy was 0.9645 and the kappa coefficient was 0.9529, which indicates that the extraction accuracy of the first-level land use and the model generalization ability realized using the RF model are relatively high.

3.1.2. Spatial distribution of the first-level land use

The extraction results of the first-level land use (Fig. 5) revealed that Lancang County is dominated by forests, covering an area of 6495.87 km², accounting for 73.76 % of the county. Forests are mainly distributed in the southern regions. Croplands are mainly concentrated in the western and northern regions, while scattered in the eastern and southern regions, covering an area of 2053.65 km² and accounting for 23.32 % of the county. Construction lands are mainly distributed in the central regions, covering an area of 136.30 km² and accounting for 1.54 % of the county. Water bodies are mainly distributed linearly along the eastern county boundary, covering an area of 105.32 km² and accounting for approximately 1.20 % of the county. Grasslands are scattered in the southeast and central regions, covering an area of 9.33 km² and accounting for less than 1 % of the county. The unutilized land area is the least, only 6.53 km².

3.2. Precise extraction of the second-level forests

3.2.1. Optimization of model parameters

Optimization of parameters directly affects the evaluation results of the models. In this study, the training samples of the second-level forests were inputted into the RF model and RF-AFs model for model training to obtain the optimal parameters, respectively. Fig. 6(a) shows that the mean OOB error rate of the RF-AFs model is minimized when $mtry = 13$; the OOB error rate becomes smooth and each forest type can be well distinguished when $ntree \geq 100$ (Fig. 6(b)). Therefore, the $mtry$ and $ntree$ in the RF-AFs model were set as 13 and 100, respectively, to accurately extract the economic forests. The $mtry$ and $ntree$ parameters of the RF model (not including AFs) were set as 3 and 100, respectively, after several rounds of trials in ENVI.

3.2.2. Accuracy assessment


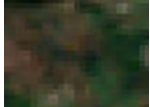



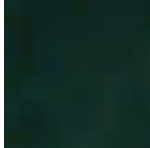





Other remaining forests samples except training samples 2 and testing samples 2 were inputted into the trained RF model and RF-AFs model, respectively, to identify the forest type of each sample. The testing samples and the identified results were used to calculate the accuracy of different models. The confusion matrix, overall accuracy and kappa coefficient of different models were presented in Table 4 and Table 5.

The accuracy assessment results of each model revealed that the overall accuracy and kappa coefficient were above 0.8, indicating that each model has high accuracy. The overall accuracy and kappa coefficient of the RF model were the lowest, 0.8436 and 0.8024, respectively. The RF-AFs model exhibited the highest overall accuracy and the highest kappa coefficient of 0.9600 and 0.9493, which were 11.64 % and 14.69 % higher than those of the RF model, respectively. It further indicates the importance of integrating SCs and VIs for extracting forest information. Overall, the RF-AFs model has an absolute advantage in extracting forest information in this study.

3.2.3. Importance of AFs

The RF-AFs model yields higher accuracy and can identify the

Table 2
Training and testing samples of different levels of geographical objects.

Land use type		Training samples		Testing samples		Total samples		Image characteristics	Description
First-level land use	Croplands	Training samples 1	2250	Testing samples 1	750	3000			Light brown and light green is alternate, but light brown is the dominant color.
	Forests		2250		750	3000		See the Second-level forests.	See the Second-level forests.
	Grasslands		375		125	500			Dark green and dark brown is alternate, but dark green is the dominant color. The texture is rough.
	Water bodies		750		250	1000			The color is dark blue, and the texture is uniform.
	Construction lands		1500		500	2000			Light blue and white is alternate, but light blue is the dominant color. It concentrated in spatial distribution.
	Unutilized lands		150		50	200			White is the dominant color, and the surrounding is dark brown.
Second-level forests	Eucalyptus	Training samples 2	1500	Testing samples 2	500	2000			The color is dark green and the texture is uniform.
	Simao pines		1500		500	2000			The color is light green, but it is darker than rubber trees and other shrubs. The texture is rough.
	Rubber trees		750		250	1000			Light brown and light green is alternate, and its green is the lightest color among all kinds of forests.
	Tea plantations		1125		375	1500			The color is dark purple, and it is concentrated in spatial distribution.
	Other shrubs		375		125	500			Light green and light brown is alternate, but light green is the dominant color.
	Other forests		2250		750	3000			The color is dark green and the texture is rough.

importance of each AFs directly (Fig. 7). The importance value clarifies the influence of each AFs on the precise extraction of economic forests. Fig. 7 shows that the importance of PSRI, TEM, EVI, and EL is high, and all are above 10%. Among them, PSRI exhibited the highest importance value of 25.26%. The importance of TSAVI and PRE was above 5%,

whereas the importance of other indicators was below 5%. Among them, RVI exhibited the lowest importance value of 1.17%. The cumulative importance of the top six factors was 76.09%, indicating that these factors play a decisive role in the forest extraction result. It further proves the importance of SCs such as TEM, EL, and PRE and VIs such as

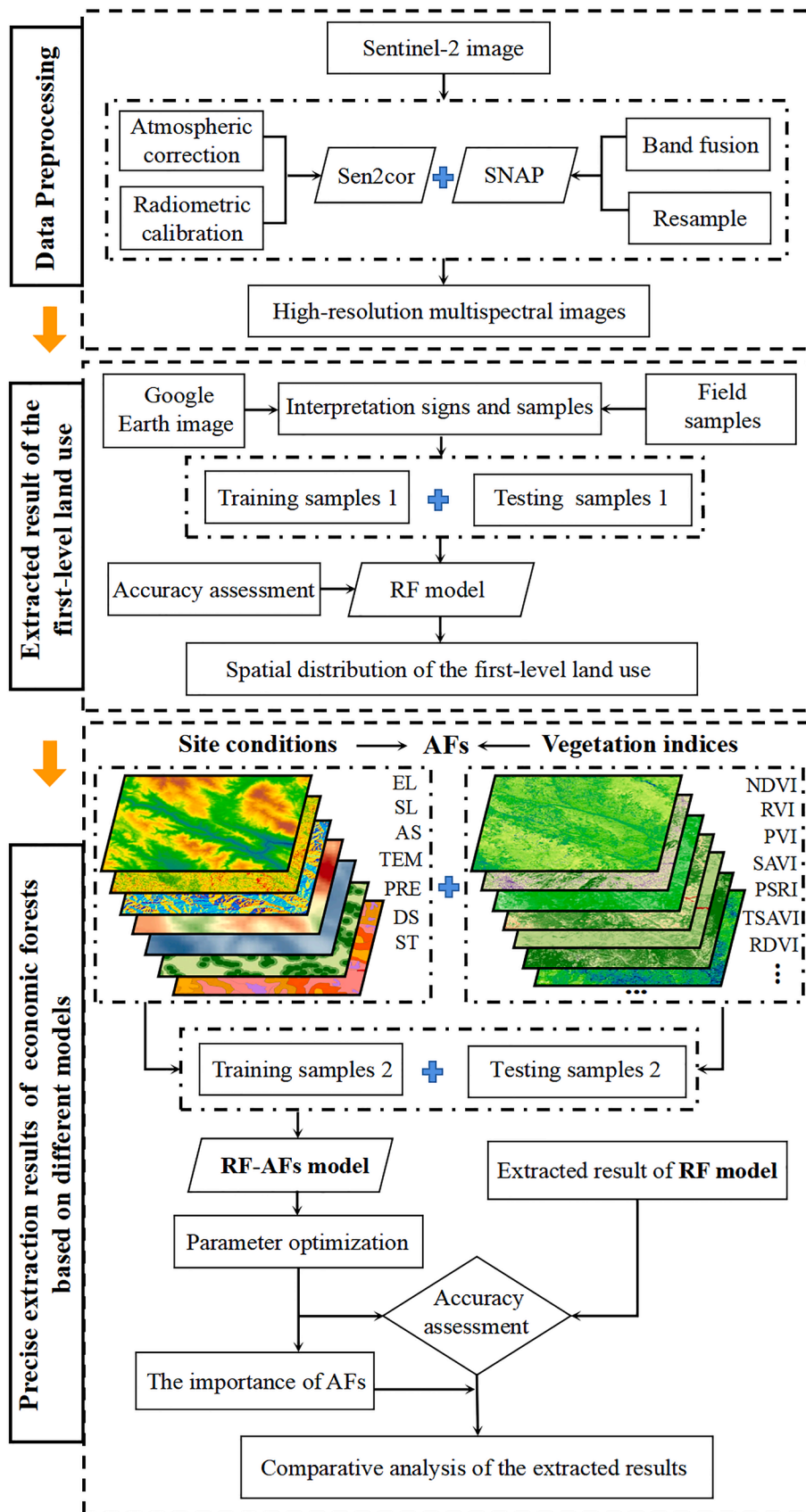


Fig. 3. Technical flowchart.

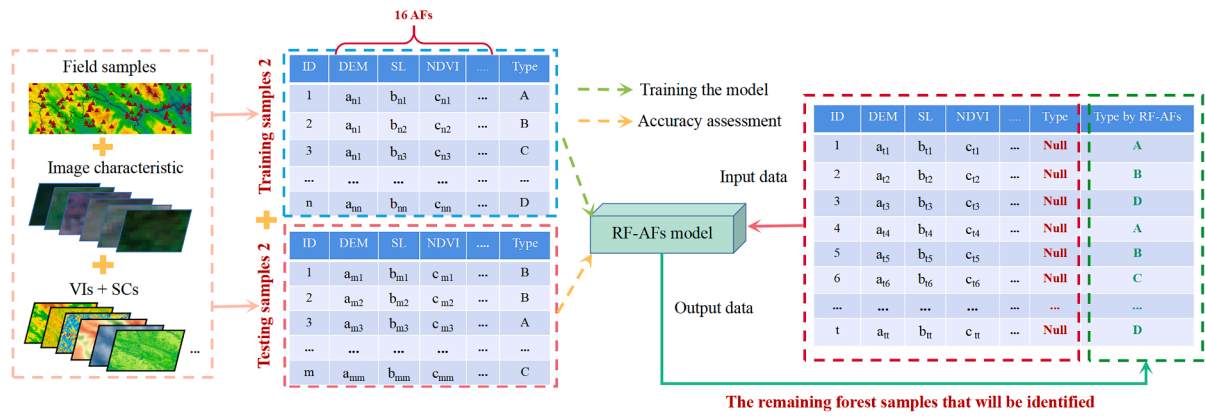


Fig. 4. Schematic of the use of training and testing samples in the RF-AFs model.

Table 3
Confusion matrix and accuracy of the extraction results of the first-level land use.

Samples type Identified type	Croplands	Forests	Grasslands	Water bodies	Construction lands	Unutilized lands
Croplands	720	12	3	0	14	1
Forests	27	710	7	0	6	0
Grasslands	1	6	118	0	0	0
Water bodies	1	0	0	249	0	0
Construction lands	2	2	0	0	496	0
Unutilized lands	1	1	0	0	2	46
OA	0.9645					
Kappa	0.9529					

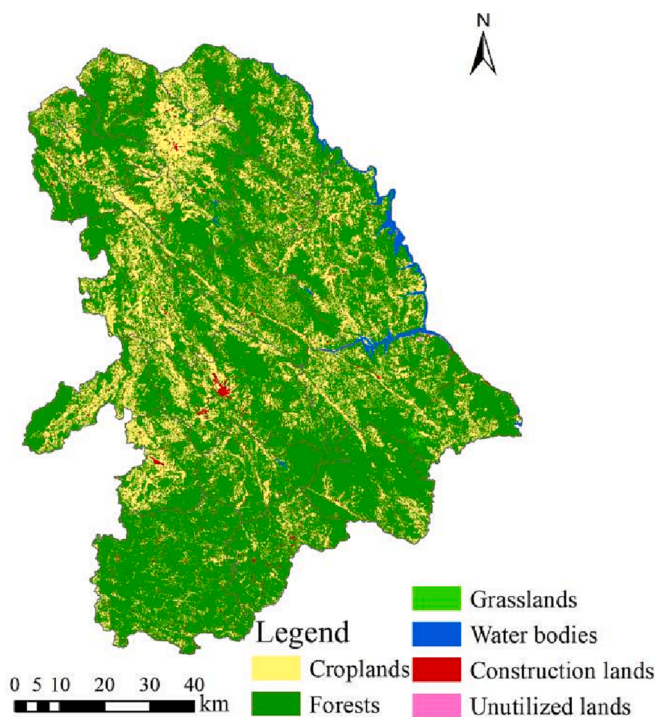


Fig. 5. First-level land use.

PSRI, EVI, and TSAVI in the accurate extraction of economic forests in Lancang County.

3.2.4. Spatial distribution of economic forests

The extraction results of economic forests under different models are

shown in Fig. 8.

The spatial differentiation of each forest type was obvious, and the overall pattern of the classification results of the two models was relatively consistent. The area of economic forests under each model was obtained separately (Table 6). The area of Simao pines, rubber trees, and tea plantations obtained using the different models exhibited obvious differences, whereas that of eucalyptus were relatively similar for two models.

In summary, the RF-AFs model provides the most precise results compared with the RF model. Therefore, the extraction results of the second-level forests obtained using the RF-AFs model were used for further analysis. Simao pines, the main economic forests in Lancang County, are mainly distributed in the central, southwestern, and northern parts of the study area, covering an area of 2037.92 km² and accounts for 31.37 % of forests area. Tea plantations are mainly distributed in the northern and southern regions, covering an area of 587.95 km² and accounting for 9.05 % of forests area. Eucalyptus, as introduced species, are distributed in all townships, covering an area of 435.68 km² and accounting for 6.71 % of forests area. Rubber trees are mainly distributed in the eastern region of the study area, with an obvious spatial distribution pattern following the river valley, covering an area of 197.86 km² and accounting for 3.05 % of forests area. Furthermore, other shrubs are mainly distributed in the central and western regions, covering an area of 185.74 km² and accounting for 2.86 % of forests area. Other forests are mainly distributed in the southeastern, central, and northern regions, covering an area of 3050.72 km² and accounting for 46.96 % of forests area. The above extraction results of economic forests will serve as an important reference for future research on ecological benefits, structural adjustment, and spatial pattern optimization of economic forests in Lancang County.

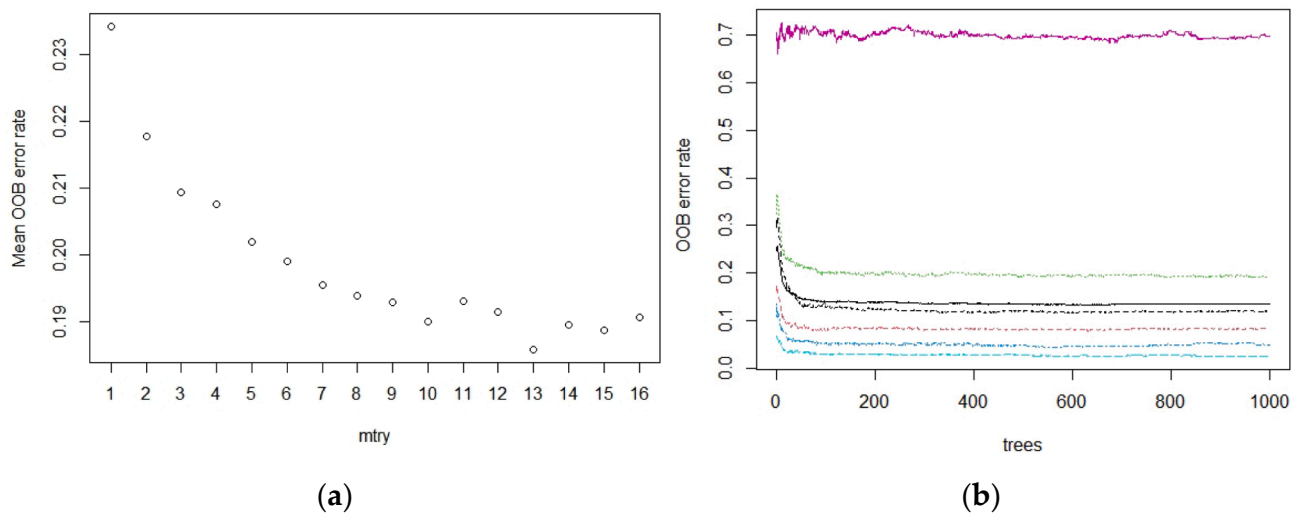


Fig. 6. Determination of optimal parameters for the RF-AFs model: (a) The mean OOB error rate of different *mtry*; (b) The OOB error rate of different *ntree* when *mtry* = 13.

Table 4

Confusion matrix and accuracy of the extraction results of the second-level forests based on the RF model.

Samples type Identified type	Eucalyptus	Simao pines	Rubber trees	Tea plantations	Other shrubs	Other forests
Eucalyptus	466	6	2	0	1	25
Simao pines	4	424	0	4	2	66
Rubber trees	2	1	229	7	3	8
Tea plantations	0	4	2	360	2	7
Other shrubs	9	32	8	3	41	32
Other forests	64	85	7	1	4	589
OA	0.8436					
Kappa	0.8024					

Table 5

Confusion matrix and accuracy of the extraction results of the second-level forests based on the RF-AFs model.

Samples type Identified type	Eucalyptus	Simao pines	Rubber trees	Tea plantations	Other shrubs	Other forests
Eucalyptus	481	4	0	0	0	15
Simao pines	1	485	0	0	0	14
Rubber trees	0	1	245	1	0	3
Tea plantations	2	5	0	356	1	11
Other shrubs	0	7	4	3	83	0
Other forests	0	0	0	0	0	750
OA	0.9600					
Kappa	0.9493					

4. Discussion

4.1. Applicability of the RF-AFs model

Precise extraction of forest information is an important research topic and technical difficulty in the field of forestry remote sensing (Schäfer et al., 2016; Zhang et al., 2019). Existing studies on forest information extraction are mostly based on the spectral features of satellite images, while other important auxiliary information including complex SCs and VIs have not been considered simultaneously, the precise extraction technology of forests needs to be further improved (Liu et al., 2018). In this study, the spectral features of satellite images were used in combination with the visual interpretation of training and testing samples, and a method was developed for precise extraction of economic forests by using an RF-AFs model that integrates relevant AFs, including SCs and VIs. An effective connection between auxiliary information and economic forests types is realized through model

training, so as to realize the mapping and precise extraction of economic forests information from point to surface (Fig. 4). In the RF-AFs model, AFs are linked with the forest category without additional weight calculation and classification standard of AFs in advance. In other words, the RF-AFs model has the function of index importance assessment, which can directly obtain the importance of each evaluation factor. Meanwhile, the RF-AFs model does not de-dimensioning or need hierarchical processing of indicators, and the data can be directly inputted into the model. In addition, the RF-AFs model inherits the advantages of the RF model, and only requires two key parameters for the model optimization, which can be optimized on the open-source platform of RStudio. The RF-AFs model has advantages in the form of the randomness of the training sets and the optimal property of node splitting, resulting in high accuracy (Talukdar et al., 2020). Specifically, the bootstrap method was used to select the training sample sets from the training samples using a random return sampling method. Features for each node splitting in the classification tree are randomly selected,

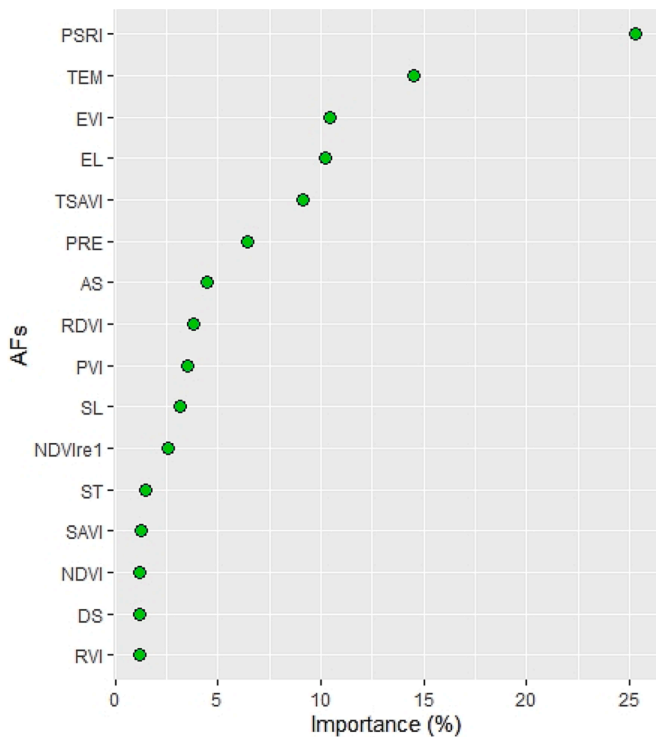


Fig. 7. Importance of AFs.

that is, randomly selecting a part of the sample features on the node and selecting an optimal feature from them as the criterion for the node splitting. This randomness enhances the generalization ability of RF-AFs model, resulting in greater tolerance and flexibility to input features

compared to other machine learning algorithm (Ma et al., 2021). The RF-AFs model is highly suitable for the precise extraction of forest information. Besides, the RF-AFs model is not limited to the extraction of economic forests, but also provides a method reference for the fine extraction of other tree species.

4.2. Discussion of the AFs importance

The RF-AFs model can be used to perform AFs importance analysis, which is convenient for analyzing the degree of influence of each factor on forest classification. The PSRI can maximize the sensitivity of the carotenoid-to-chlorophyll ratio, which is often used in vegetation health monitoring, plant physiological stress detection, crop production, and yield analysis (Guerini Filho et al., 2020). This study further shows that PSRI is also an important factor in the extraction of economic forests as well as other forests. The EVI not only possesses the advantages of NDVI but also overcomes the problems of the saturation of high vegetation coverage areas, incomplete atmospheric correction, and water vapor interference. The EVI can improve the sensitivity of vegetation in biotopes, and has higher sensitivity and superiority for vegetation monitoring (Huang et al., 2020b). Therefore, EVI has high importance in the extraction results of economic forests. The different responses of soil background to spectral radiation affect the accuracy of forest information extraction. The TSAVI takes into account the interaction of electromagnetic radiation, atmosphere, vegetation cover, and soil background noise and minimizes the influence of soil background (Lemenkova, 2021), which contributes to the effective extraction of economic forests. In addition, the precipitation and temperature of the sites are necessary for vegetation growth. Moreover, affected by the EL (Liu et al., 2018), the vegetation growth and their spectral reflection and radiation characteristics differ, which enlarges the variability between different forests. Perhaps these factors could be considered as the AFs in the accurate extraction of the economic forests in other region.

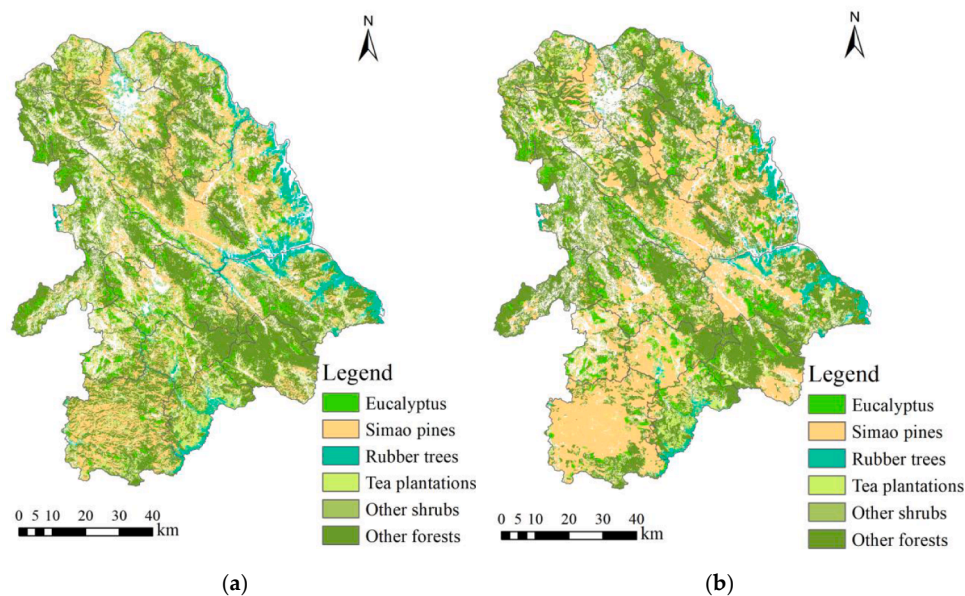


Fig. 8. Spatial distribution of economic forests under different models: (a) RF model; (b) RF-AFs model.

Table 6
Area statistics of the second-level forests in different models (km²).

Model	Eucalyptus	Simao pines	Rubber trees	Tea plantations	Other shrubs	Other forests
RF	445.57	1441.51	345.92	935.79	111.94	3215.14
RF-AFs	435.68	2037.92	197.86	587.95	185.74	3050.72

4.3. Parameters optimization and accuracy assessment

Model parameter optimization is essential for improving model accuracy. In this study, the spectral features and AFs of forests were considered, and training and testing samples were used for model training and model error calculation to determine the optimal parameters of different models. This is the key that the RF-AFs model can achieve high accuracy in the extraction of forests information. The kappa coefficient of the extraction results of forests in the RF-AFs model is 0.9493, which is significantly higher than that of the RF model, confirming the importance of AFs, including SCs and VIs in the precise extraction of forests information.

In the extraction process of the first-level land use, AFs are not considered, and the optimal parameters are not determined because the spectral features of croplands, forests, grasslands, water bodies, construction lands and unutilized lands differ greatly, and these types can be well distinguished without combining multiple AFs. It can be seen from the accuracy results of the first-level land use. However, during the precise extraction of economic forests, different forest types are difficult to distinguish due to their similar spectral features. Therefore, important interpreted features of different economic forests must be considered simultaneously, and other auxiliary features for increasing the variability among different forest types must be explored further to comprehensively improve the extraction accuracy of economic forests.

4.4. Study limitations

This study aimed to develop a method for the precise extraction of economic forests. The major economic forests in Lancang County were selected for empirical demonstration. This method can also be employed for the precise extraction of other forests. However, the textural features of forests are not considered. Considering multiple AFs of forests in the extraction model simultaneously may improve the extraction accuracy of forests. In the future, more AFs such as spectral features, textural features, shape features, and SCs can be integrated simultaneously, and multivariate machine learning algorithms can be used for the precise and intelligent extraction of vegetation information. Finally, there are other hyper parameters in the RF model and RF-AFs model. We only selected the key parameters for optimizing. Considering more hyper parameters for optimizing is useful to obtain a more precise model.

5. Conclusions

Based on the AFs, including SCs and VIs, the RF-AFs model was employed for the extraction of the economic forests in Lancang County of Yunnan Province, China. The main conclusions are as follows:

(1) The kappa coefficient of the extraction results of the first-level land use in Lancang County obtained using the RF model is 0.9529, and Lancang County is dominated by forests. Forests are mainly concentrated in the southern region, accounting for 73.76 % of the territory area. The proportions of croplands, grasslands, water bodies, construction lands, and unutilized lands are 23.32 %, 0.11 %, 1.20 % and 1.54 %, and 0.07 %, respectively.

(2) Economic forests and other forests can be extracted precisely and effectively by integrating AFs, including SCs and VIs. The RF-AFs model, after parameter optimization, achieved good results in the extraction of economic forests. The kappa coefficient of the RF-AFs model is 0.9493, and the extraction accuracy is higher than that of the RF model.

(3) The importance of PSRI, TEM, EVI, EL, TSAVI and PRE are all above 5 %, and the cumulative value of their importance is 76.09 %, which are the main factors affecting the extraction results of economic forests in Lancang County. It may provide factors reference for extracting economic forests and other forests in other regions.

(4) Simao pines constitute the dominant economic forests in Lancang County, and are mainly distributed in the central, southwestern, and northern regions, accounting for 31.37 % of forests area. Tea plantations

are mainly distributed in the northern and southern regions, accounting for 9.05 % of forests area. Eucalyptus are distributed in all townships, accounting for 6.71 % of forests area. Rubber trees are mainly distributed in the eastern river valleys, accounting for 3.05 % of forests area. Furthermore, other shrubs and other forests other than economic forests account for 2.86 % and 46.96 % of forests area, respectively.

CRedit authorship contribution statement

Pei Huang: Conceptualization, Methodology, Software, Investigation, Resources, Formal analysis, Data curation, Visualization, Validation, Writing – original draft, Writing – review & editing, Project administration, Funding acquisition. **Xiaoqing Zhao:** Conceptualization, Methodology, Resources, Data curation, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Junwei Pu:** Software, Investigation, Validation, Writing – original draft. **Zexian Gu:** Software, Investigation, Validation, Writing – original draft. **Yan Feng:** Investigation, Data curation, Visualization, Validation, Writing – review & editing. **Shijie Zhou:** Investigation, Visualization, Validation, Writing – review & editing. **Xinyu Shi:** Software, Investigation, Visualization, Writing – review & editing, Funding acquisition. **Yuanyuan Tang:** Investigation, Visualization, Funding acquisition. **Pinliang Dong:** Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

This work was jointly supported by the National Natural Science Foundation of China (Nos. 42061052, 41361020, 40961031), the Joint Fund of Yunnan Provincial Science and Technology Department and Yunnan University (No. 2018FY001-017), the Construction Project of Graduate Tutor Team in Yunnan Province (No. C176230200), the Postgraduate Innovative Research Project of Yunnan University (Nos. 2020Z46, 2021Z099, 2021Y373). Besides, We want to extend our heartfelt gratitude to all the staff of the Simao Jin Lancang High-Yield Forestry Co., Ltd., Lancang County, Yunnan Province, China for their support, particularly during the field survey.

References

- Abdi, A.M., 2020. Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GISci. Remote Sens.* 57, 1–20. <https://doi.org/10.1080/15481603.2019.1650447>.
- Ahrends, A., Hollingsworth, P.M., Ziegler, A.D., Fox, J.M., Chen, H., Su, Y., Xu, J., 2015. Current trends of rubber plantation expansion may threaten biodiversity and livelihoods. *Glob. Environ. Change.* 34, 48–58. <https://doi.org/10.1016/j.gloenvcha.2015.06.002>.
- Bai, Y., Liang, S., Yuan, W., 2021. Estimating global gross primary production from sun-induced chlorophyll fluorescence data and auxiliary information using machine learning methods. *Remote Sens.* 13, 963. <https://doi.org/10.3390/rs13050963>.
- Barzegar, M., Ebadi, H., Kiani, A., 2015. Comparison of different vegetation indices for very high-resolution images, specific case UltraCam-D imagery. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* 41, 97–104. <https://doi.org/10.5194/isprsarchives-XL-1-W5-97-2015>.
- Brede, B., Verrelst, J., Gastellu-Etchegorry, J.P., Clevers, J.G., Goudzwaard, L., den Ouden, J., Verbesselt, J., Herold, M., 2020. Assessment of workflow feature selection on forest LAI prediction with sentinel-2A MSI, landsat 7 ETM+ and Landsat 8 OLI. *Remote Sens.* 12 (6), 915. <https://doi.org/10.3390/rs12060915>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32. <https://doi.org/10.1023/A:1010933404324>.

- Brukas, V., Mizaras, S., Mizaraitė, D., 2014. Economic forest sustainability: comparison between Lithuania and Sweden. *Forests* 6 (1), 47–64. <https://doi.org/10.3390/f6010047>.
- Cao, P., Zhang, L., Li, S., Zhang, J., 2016. Review on vegetation phenology observation and phenological index extraction. *Adv. Earth Sci.* 31 (4), 365–376.
- Carranza-García, M., García-Gutiérrez, J., Riquelme, J.C., 2019. A framework for evaluating land use and land cover classification using convolutional neural networks. *Remote Sens.* 11, 274. <https://doi.org/10.3390/rs11030274>.
- Chu, S., Ouyang, J., Liao, D., Zhou, Y., Liu, S., Shen, D., Wei, X., Zeng, S., 2019. Effects of enriched planting of native tree species on surface water flow, sediment, and nutrient losses in a Eucalyptus plantation forest in southern China. *Sci. Total Environ.* 675, 224–234. <https://doi.org/10.1016/j.scitotenv.2019.04.214>.
- Dobričić, D., Gašparović, M., Medak, D., 2021. Sentinel-1 and 2 time-series for vegetation mapping using random forest classification: a case study of Northern Croatia. *Remote Sens.* 13 (12), 2321. <https://doi.org/10.3390/rs13122321>.
- Dong, T., Liu, J., Shang, J., Qian, B., Ma, B., Kovacs, J.M., Walters, D., Jiao, X., Geng, X., Shi, Y., 2019. Assessment of red-edge vegetation indices for crop leaf area index estimation. *Remote Sens. Environ.* 222, 133–143. <https://doi.org/10.1016/j.rse.2018.12.032>.
- Du, H., Mao, F., Li, X., Zhou, G., Xu, X., Han, N., Sun, S., Gao, G., Cui, L., Li, Y., Zhu, D., Liu, Y., Chen, L., Fan, W., Li, P., Shi, Y., Zhou, Y., 2018. Mapping global bamboo forest distribution using multisource remote sensing data. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 11, 1458–1471. <https://doi.org/10.1109/JSTARS.2018.2800127>.
- Fadaei, H., 2020. Advanced land observing satellite data to identify ground vegetation in a juniper forest, northeast Iran. *J. For. Res.* 31 (2), 531–539. <https://doi.org/10.1007/s11676-018-0812-5>.
- Fakhri, S.A., Sayadi, S., Latifi, H., Khare, S., 2019. An optimized enhanced vegetation index for sparse tree cover mapping across a mountainous region//2019 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor). IEEE pp 146–151. <https://doi.org/10.1109/MetroAgriFor.2019.8909259>.
- Fischer, G., Nachtergaele, F., Prieler, S., Van Velthuisen, H.T., Verelst, L., Wiberg, D., 2008. Global agro-ecological zones assessment for agriculture (GAEZ 2008). IIASA, Laxenburg, Austria and FAO, Rome, Italy, 10.
- Fritz, R., Frech, I., Koch, B., Ueffing, C., 1999. Sensor fused images for visual interpretation of forest stand borders. *Int. Arch. Photogramm. Remote Sens.* 32, 1–7.
- Fu, B., Xie, S., He, H., Zuo, P., Sun, J., Liu, L., Huang, L., Fan, D., Gao, E., 2021. Synergy of multi-temporal polarimetric SAR and optical image satellite for mapping of marsh vegetation using object-based random forest algorithm. *Ecol. Indic.* 131, 108173. <https://doi.org/10.1016/j.ecolind.2021.108173>.
- Geng, L., Che, T., Wang, X., Wang, H., 2019. Detecting spatiotemporal changes in vegetation with the BFAST model in the Qilian Mountain region during 2000–2017. *Remote Sens.* 11, 103. <https://doi.org/10.3390/rs11020103>.
- Goded, S., Ekroos, J., Domínguez, J., Azcárate, J.G., Guitián, J.A., Smith, H.G., 2019. Effects of eucalyptus plantations on avian and herb species richness and composition in North-West Spain. *Glob. Ecol. Conserv.* 19, e00690. <https://doi.org/10.1016/j.gecco.2019.e00690>.
- Grabska, E., Hostert, P., Pflugmacher, D., Ostapowicz, K., 2019. Forest stand species mapping using the Sentinel-2 time series. *Remote Sens.* 11 (10), 1197. <https://doi.org/10.3390/rs11101197>.
- Guerini Filho, M., Kuplich, T.M., Quadros, F.L.F.D., 2020. Estimating natural grassland biomass by vegetation indices using Sentinel 2 remote sensing data. *Int. J. Remote Sens.* 41, 2861–2876. <https://doi.org/10.1080/01431161.2019.1697004>.
- Huang, Y., Zhang, Z., Huang, X., Hong, C., Wang, M., Zhang, R., Zhang, X., Zeng, J., 2020b. Study on Vegetation Cover Change of Huang Huai Hai Plain Based on MODIS EVI. In *Recent Trends in Intelligent Computing, Communication and Devices*. Springer, Singapore, pp 459–466. https://doi.org/10.1007/978-981-13-9406-5_56.
- Huang, P., Peng, L., Pan, H., 2020a. Linking the random forests model and GIS to assess geo-hazards risk: a case study in Shifang County, China. *IEEE Access* 8, 28033–28042. <https://doi.org/10.1109/ACCESS.2020.2972005>.
- Kureel, N., Sarup, J., Matin, S., Goswami, S., Kureel, K., 2022. Modelling vegetation health and stress using hyperspectral remote sensing data. *Model. Earth Syst. Environ.* 8, 733–748. <https://doi.org/10.1007/s40808-021-01113-8>.
- Le, B.T., Ha, T.T.L., 2019. Hyperspectral image classification based on average spectral-spatial features and improved hierarchical-ELM. *Infrared Phys. Technol.* 102, 103013 <https://doi.org/10.1016/j.infrared.2019.103013>.
- Lemenkova, P., 2021. Distance-based vegetation indices computed by SAGA GIS: a comparison of the perpendicular and transformed soil adjusted approaches for the LANDSAT TM image. *Poljoprivredna tehnika.* 46 (3), 49–60.
- Li, J., Pei, Y., Zhao, S., Xiao, R., Sang, X., Zhang, C., 2020. A review of remote sensing for environmental monitoring in China. *Remote Sens.* 12, 1130–1154. <https://doi.org/10.3390/rs12071130>.
- Lima, M.B.O., Lustosa Jr, I.M., Oliveira, E.M., Ferreira, J.C.B., Soares, K.L., Miguel, E.P., 2017. Artificial neural networks in whole-stand level modeling of Eucalyptus plants. *Afr. J. Agric. Res.* 12, 524–534. <https://doi.org/10.5897/AJAR2016.12068>.
- Liu, M., Fu, B., Xie, S., He, H., Lan, F., Li, Y., Lou, P., Fan, D., 2021. Comparison of multi-source satellite images for classifying marsh vegetation using deeplabv3 plus deep learning algorithm. *Ecol. Indic.* 125, 107562.
- Liu, Y., Gong, W., Hu, X., Gong, J., 2018. Forest type identification with random forest using Sentinel-1A, Sentinel-2A, multi-temporal Landsat-8 and DEM data. *Remote Sens.* 10, 946. <https://doi.org/10.3390/rs10060946>.
- Luo, X., Zhao, W., Wei, S., Fu, Q., 2015. Study on urban remote sensing classification based on improved RBF network and normalized difference indexes. *Int. J. Signal Process. Image Process. Pattern Recogn.* 8, 257–270. <https://doi.org/10.14257/ijisp.2015.8.10.27>.
- Ma, M., Liu, J., Liu, M., Zeng, J., Li, Y., 2021. Tree species classification based on Sentinel-2 imagery and random forest classifier in the eastern regions of the Qilian mountains. *Forests* 12 (12), 1736. <https://doi.org/10.3390/f12121736>.
- Mao, W., Lu, D., Hou, L., Liu, X., Yue, W., 2020. Comparison of machine-learning methods for urban land-use mapping in Hangzhou city, China. *Remote Sens.* 12 (17), 2817. <https://doi.org/10.3390/rs12172817>.
- Melville, B., Lucieer, A., Aryal, J., 2018. Object-based random forest classification of Landsat ETM+ and WorldView-2 satellite imagery for mapping lowland native grassland communities in Tasmania, Australia. *Int. J. Appl. Earth Obs. Geoinf.* 66, 46–55. <https://doi.org/10.1016/j.jag.2017.11.006>.
- Mohammady, M., Moradi, H.R., Zeinivand, H., Temme, A.J.A.M., 2015. A comparison of supervised, unsupervised and synthetic land use classification methods in the north of Iran. *Int. J. Environ. Sci. Technol.* 12, 1515–1526. <https://doi.org/10.1007/s13762-014-0728-3>.
- Niculescu, S., Xia, J., Roberts, D., Billey, A., 2020. Rotation forests and random forest classifiers for monitoring of vegetation in Pays de Brest (France). *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* 43, 727–732. <https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-727-2020>.
- Nina, J., John, M., 2016. *Data Science: Theory, Method and R Language Practice*. China Machine Press, Beijing.
- Nong, M., Leng, Y., Xu, H., Li, C., Ou, G., 2019. Incorporating competition factors in a mixed-effect model with random effects of site quality for individual tree above-ground biomass growth of Pinus kesiya var. langbianensis. *N. Z. J. Forest. Sci.* 49. <https://doi.org/10.33494/nzjfs492019x27x>.
- Oliveira, D., Martins, L., Mora, A., Damásio, C., Caetano, M., Fonseca, J., Ribeiro, R.A., 2021. Data fusion approach for eucalyptus trees identification. *Int. J. Remote Sens.* 42, 4087–4109. <https://doi.org/10.1080/01431161.2021.1883198>.
- Owers, C.J., Rogers, K., Woodroffe, C.D., 2016. Identifying spatial variability and complexity in wetland vegetation using an object-based approach. *Int. J. Remote Sens.* 37, 4296–4316. <https://doi.org/10.1080/01431161.2016.1211349>.
- Pal, S.C., Chakraborty, R., Malik, S., Das, B., 2018. Application of forest canopy density model for forest cover mapping using LISS-IV satellite data: a case study of Sali watershed, West Bengal. *Model. Earth Syst. Environ.* 4, 853–865. <https://doi.org/10.1007/s40808-018-0445-x>.
- Pu, J., Zhao, X., Dong, P., Wang, Q., Yue, Q., 2021. Extracting information on rocky desertification from satellite images: a comparative study. *Remote Sens.* 13, 2497.
- Reddy, C.S., Jha, C.S., Diwakar, P.G., Dadwal, V.K., 2015. Nationwide classification of forest types of India using remote sensing and GIS. *Environ. Monit. Assess.* 187, 1–30. <https://doi.org/10.1007/s10661-015-4990-8>.
- Ren, H., Feng, G., 2015. Are soil-adjusted vegetation indices better than soil-unadjusted vegetation indices for above-ground green biomass estimation in arid and semi-arid grasslands? *Grass Forage Sci.* 70, 611–619. <https://doi.org/10.1111/gfs.12152>.
- Sankey, T., Donager, J., McVay, J., Sankey, J.B., 2017. UAV lidar and hyperspectral fusion for forest monitoring in the southwestern USA. *Remote Sens. Environ.* 195, 30–43. <https://doi.org/10.1016/j.rse.2017.04.007>.
- Savchenko, Y.Y., Goleva, O.G., Korchagina, I.A., Lobanova, Y.S., Borzhikov, T.S., 2020. GIS approaches to creating maps based on vegetation indices for forest management. *IOP Conf. Ser.: Mater. Sci. Eng.* 828 (1), 012021.
- Schäfer, E., Heiskanen, J., Heikinheimo, V., Pellikka, P., 2016. Mapping tree species diversity of a tropical montane forest by unsupervised clustering of airborne imaging spectroscopy data. *Ecol. Indic.* 64, 49–58. <https://doi.org/10.1016/j.ecolind.2015.12.026>.
- Shi, X., Zhao, X., Pu, J., Huang, P., Gu, Z., Chen, Y., 2022. Evolution modes, types, and social-ecological drivers of ecologically critical areas in the Sichuan-Yunnan ecological barrier in the last 15 years. *Int. J. Environ. Res. Public Health* 19 (15), 9206. <https://doi.org/10.3390/ijerph19159206>.
- Solymosi, K.J., Kövér, G., Romvári, R., 2019. The progression of vegetation indices: a short overview. *Acta Agraria Kaposváriensis* 23, 75–90. <https://doi.org/10.31914/AAK.2264>.
- Sothe, C., De Almeida, C.M., Schimalski, M.B., La Rosa, L.E.C., Castro, J.D.B., Feitosa, R. Q., Dalponte, M., Lima, C.L., Liesenborg, V., Miyoshi, G.T., Tommaselli, A.M.G., 2020. Comparative performance of convolutional neural network, weighted and conventional support vector machine and random forest for classifying tree species using hyperspectral and photogrammetric data. *GISci. Remote Sens.* 57 (3), 369–394. <https://doi.org/10.1080/15481603.2020.1712102>.
- Suchenwirth, L., Förster, M., Cierjacks, A., Lang, F., Kleinschmidt, B., 2012. Knowledge-based classification of remote sensing data for the estimation of below- and above-ground organic carbon stocks in riparian forests. *Wetl. Ecol. Manag.* 20 (2), 151–163.
- Sun, S., Li, Z., Tian, X., Gao Z, Wang, C., Gu, C., 2019. Forest Canopy Closure Estimation in Greater Khingan Forest Based on GF-2 Data//IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE: 6640–6643. <https://doi.org/10.1109/IGARSS.2019.8899175>.
- Sun, Y., Qin, Q., Ren, H., Zhang, T., Chen, S., 2019b. Red-edge band vegetation indices for leaf area index estimation from sentinel-2/MSI imagery. *IEEE Trans. Geosci. Remote Sensing.* 58, 826–840. <https://doi.org/10.1109/TGRS.2019.2940826>.
- Talukdar, S., Singha, P., Mahato, S., Pal, S., Liou, Y.A., Rahman, A., 2020. Land-use land-cover classification by machine learning classifiers for satellite observations—A review. *Remote Sens.* 12 (7), 1135. <https://doi.org/10.3390/rs12071135>.
- Thanh Noi, P., Kappas, M., 2017. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors* 18 (1), 18. <https://doi.org/10.3390/s18010018>.
- Tian, S., Zhang, X., Tian, J., Sun, Q., 2016. Random forest classification of wetland landcovers from multi-sensor data in the arid region of Xinjiang, China. *Remote Sens.* 8, 954. <https://doi.org/10.3390/rs8110954>.

- Visser, F., Buis, K., Verschoren, V., Schoelynck, J., 2018. Mapping of submerged aquatic vegetation in rivers from very high-resolution image data, using object-based image analysis combined with expert knowledge. *Hydrobiologia* 812, 157–175. <https://doi.org/10.1007/s10750-016-2928-y>.
- Wakulińska, M., Marcinkowska-Ochtyra, A., 2020. Multi-temporal sentinel-2 data in classification of mountain vegetation. *Remote Sens.* 12, 2696. <https://doi.org/10.3390/rs12172696>.
- Wang, Y., Zhang, X., Guo, Z., 2021. Estimation of tree height and aboveground biomass of coniferous forests in North China using stereo ZY-3, multispectral Sentinel-2, and DEM data. *Ecol. Indic.* 126, 107645 <https://doi.org/10.1016/j.ecolind.2021.107645>.
- Waśniewski, A., Hościlo, A., Zagajewski, B., Moukétou-Tarazewicz, D., 2020. Assessment of Sentinel-2 satellite images and random forest classifier for rainforest mapping in Gabon. *Forests* 11 (9), 941. <https://doi.org/10.3390/f11090941>.
- Wei, M., Qiao, B., Zhao, J., Zuo, X., 2020. The area extraction of winter wheat in mixed planting area based on Sentinel-2 a remote sensing satellite images. *Int. J. Parallel Emergent Distributed Syst.* 35, 297–308. <https://doi.org/10.1080/17445760.2019.1597084>.
- Wu, Y., Zhang, X., 2019. Object-Based tree species classification using airborne hyperspectral images and LiDAR data. *Forests* 11, 32. <https://doi.org/10.3390/f11010032>.
- Yang, Y., Guan, H., Shen, M., Liang, W., Jiang, L., 2015. Changes in autumn vegetation dormancy onset date and the climate controls across temperate ecosystems in China from 1982 to 2010. *Glob. Change Biol.* 21, 652–665. <https://doi.org/10.1111/gcb.12778>.
- Zhao, X., Gu, Z., 2021. Ecological security and optimal allocation of land use in artificial economic forest planting areas in Yunnan Province Science Press, Beijing, pp 3–4.
- Zhang, Z., Li, Z., Tian, X., 2019. Precise identification of forest land types based on high resolution remotely sensed imagery. *J. Zhejiang A&F Univ.* 36, 857–867. In Chinese.
- Zhao, X., Tan, K., Xie, P., Chen, B., Pu, J., 2021. Multiobjective land-use optimization allocation in eucalyptus-introduced regions based on the GMDP-ACO model. *J. Urban Plan. Dev* 147, 05021004. [https://doi.org/10.1061/\(ASCE\)UP.1943-5444.0000664](https://doi.org/10.1061/(ASCE)UP.1943-5444.0000664).
- Zhao, X., Xu, X., 2015. Research on landscape ecological security pattern in a Eucalyptus introduced region based on biodiversity conservation. *Russ. J. Ecol.* 46, 59–70. <https://doi.org/10.1134/S106741361501018X>.
- Zhao, X., Yuan, Y., Song, M., Ding, Y., Lin, F., Liang, D., Zhang, D., 2019. Use of unmanned aerial vehicle imagery and deep learning UNet to extract rice lodging. *Sensors* 19, 3859. <https://doi.org/10.3390/s19183859>.
- Zhou, X., Zhu, H., Wen, Y., Goodale, U.M., Zhu, Y., Yu, S., Li, C., Li, X., 2020. Intensive management and declines in soil nutrients lead to serious exotic plant invasion in Eucalyptus plantations under successive short-rotation regimes. *Land Degrad. Dev* 31, 297–310. <https://doi.org/10.1002/ldr.3449>.
- Zhu, K., Sun, Z., Zhao, F., Yang, T., Tian, Z., Lai, J., Zhu, W., Long, B., 2021. Relating hyperspectral vegetation indices with soil salinity at different depths for the diagnosis of winter wheat salt stress. *Remote Sens.* 13, 250. <https://doi.org/10.3390/rs13020250>.