

GIScience & Remote Sensing



ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgrs20

Regionalized classification of stand tree species in mountainous forests by fusing advanced classifiers and ecological niche model

Panfei Fang, Guanglong Ou, Ruonan Li, Leiguang Wang, Weiheng Xu, Qinling Dai & Xin Huang

To cite this article: Panfei Fang, Guanglong Ou, Ruonan Li, Leiguang Wang, Weiheng Xu, Qinling Dai & Xin Huang (2023) Regionalized classification of stand tree species in mountainous forests by fusing advanced classifiers and ecological niche model, GIScience & Remote Sensing, 60:1, 2211881, DOI: <u>10.1080/15481603.2023.2211881</u>

To link to this article: <u>https://doi.org/10.1080/15481603.2023.2211881</u>

| 9 | © 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. | + | View supplementary material 🕼 |
|-----------|---|---|---------------------------------------|
| | Published online: 12 May 2023. | | Submit your article to this journal 🛛 |
| hl | Article views: 160 | Q | View related articles 🕑 |
| CrossMark | View Crossmark data 🖸 | | |

OPEN ACCESS OPEN ACCESS

Regionalized classification of stand tree species in mountainous forests by fusing advanced classifiers and ecological niche model

Panfei Fang^a, Guanglong Ou^b, Ruonan Li^c, Leiguang Wang^d, Weiheng Xu^d, Qinling Dai^e and Xin Huang^f

^aFaculty of Forestry, Southwest Forestry University, Kunming, China; ^bKey Laboratory of State Forestry Administration on Biodiversity Conservation in Southwest China Southwest Forestry University, Kunming, China; ^cFaculty of Geography, Yunnan Normal University, Kunming, China; ^dInstitute of Big Data and Artificial Intelligence, Southwest Forestry University, Kunming, China; ^eArt and Design College, Southwest Forestry University, Kunming, Yunnan, China; ^fSchool of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China

ABSTRACT

Though many new remote sensing technologies have been introduced to analyze forests, regionalscale species-level mapping products are still rare, especially in large mountainous areas. Tree species abundance, low spectral separability among species and huge computing demand are hindrances for obtaining an accurate stand tree species map. This study addressed these problems by synergizing regionalization, multiple feature fusion, and model fusion and proposed a new machine learning workflow. The whole area, i.e. Yunnan province in China (approximately 390,000 km²), was firstly divided into 8 distinct floristic regions according to the distributions and phylogenetic relationships of native tree species. Thereafter, with Google Earth Engine (GEE) platform, multiple data sets, including Sentinel-2 imagery, SRTM DEM, and WorldClim bioclimatic, were collected to construct a high-dimensional feature pool for each region. Thirdly, the maximum entropy model (MaxEnt), generally used for predicting ecological niche, and three classifiers, i.e. Random Forest (RF), Support Vector Machines (SVM), and Extreme Gradient Boosting (XGBoost), were used to pre-classify environmental and remote sensing data, respectively. After that, two types of decision fusion strategies, parallel and serial ensembles, were applied to fuse preclassification probability maps for each sub-region. Finally, the spatial distribution of 19 forest stand species over the whole Yunnan Province was obtained by mosaicking the best classification results from 8 sub-regions. Our method achieves an overall accuracy of 72.18% on the entire validation dataset. The decision fusion models significantly improve the classification accuracy, with the eight partitioned best fusion models improving the accuracy by 7.33%-25.39% on average compared to base classifiers. This study demonstrates that the spatial partitioning strategy and the decision fusion integrating a proper machine learning algorithm and ecological niche model can significantly improve the classification accuracy of forest stand species in montane forests

ARTICLE HISTORY

Received 20 December 2022 Accepted 4 May 2023

Tavlor & Francis

Taylor & Francis Group

KEYWORDS

Tree species classification; Vegetation regionalization; Maximum entropy model; Machine learning; Decision fusion

1. Introduction

Information on the distribution of forest stand species is essential for the development of forest management policies and the sustainable use of forest resources (Fassnacht et al. 2014; Guo et al. 2020) as well as the assessment of biodiversity, habitat quality, carbon cycling, and biomass (Kollert et al. 2021; Ørka et al. 2013; Waser et al. 2011).

The spatial distribution of tree species is traditionally obtained through labor-intensive field surveys. Thus, inventory data are expensive and time-consuming to create, especially in a large mountainous region. Remote sensing brings opportunities for costefficiently and time-efficiently obtaining vegetationrelated information (Grabska, Frantz, and Ostapowicz 2020). However, unlike other land cover types, the extremely low separability among tree species in spectral signature makes mapping tree species is still a big challenge. In this context, various data and advanced processing techniques are often needed, to meet the corresponding requirements with the desired accuracy and details of forest stand tree species mapping.

Multimodal remote sensing sensors provide key information for describing tree species. High spatial resolution multispectral images have rich spectral and textural information, and hyperspectral images with nano-scale spectral resolution are rich in spectral information (Wan et al. 2021); LiDAR data provide

CONTACT Leiguang Wang 🐼 wlgbain@126.com

B Supplemental data for this article can be accessed online at https://doi.org/10.1080/15481603.2023.2211881.

^{© 2023} The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

geometric information related to tree vertical structure, such as tree height, canopy diameter, and leaf area (Feng et al. 2020; Ke, Quackenbush, and Im 2010; Michaowska and Rapiński 2021). Current tree species classification research focuses on algorithm-driven or data-driven methods (Fassnacht et al. 2016). Machine learning algorithms have potent capabilities to map non-normally distributed feature into classes (Bhatt et al. 2022). Typical models including Random Forest (RF), Support Vector Machines (SVM), and Extreme Gradient Boosting (XGBoost) have been widely applied to classification. To date, with the increasing availability of earth observation data, Deep Learning (DL) models have drawn more attention. DL models can learn discriminant features from data through multi-layer neural networks, accurately capture nonlinear relationship between data and classes, and thus achieving end-to-end image classification (LeCun, Bengio, and Hinton 2015).

These data and classification models have been widely used to map tree species with high accuracies. For instance, very-high-resolution (VHR) Worldview-2 images and RF were used to identify the three most abundant Pine species in Galicia, with a classification accuracy of 91% (Alonso, Picos, and Armesto 2021). Feng et al. collaborated aerial hyperspectral with LiDAR data and used RF to identify 10 tree species in a forest farm with an OA of 96.10% (Feng et al. 2020). Zhang et al. applied 3DCNN to tree species classification, achieving a high accuracy of 93.14% while improving computational efficiency (Zhang, Zhao, and Zhang 2020). Unfortunately, these studies cannot be extended to large regions or globally due to high data collection costs.

Free-accessible images acquired by low- and medium-resolution satellites, like MODIS, Landsat, and Sentinel serials, are often used for mapping forest ecosystem over a large area. Many forest-related products, such as GLAD Forest (Hansen et al. 2013), FROM_GLC (Gong et al. 2013), and iMap World 1.0 (Liu et al. 2021), have been produced. However, these products treat forests as a single class or divide forests into several forest types. These researches focus on analyzing forest distribution and dynamic changes but cannot provide the spatial distribution of tree species. How to obtain spatial distribution information of large-scale tree species economically, efficiently, and accurately is still an important and urgent problem. Currently, Sentinel-2 (S2) satellite constellation may be the best choice for tree species classification in large regions. Several studies have demonstrated the high potential of unique red-edge band and high-density time series of S2 images for vegetation mapping in large regions (Grabska, Frantz, and Ostapowicz 2020; Welle et al. 2022). However, new challenges arise when classifying a geographically heterogeneous region such as the Yunnan Province.

The Yunnan province is in southwest China (Figure 1), which covers an area about the size of Germany. Mapping such a large area firstly implies processing massive amounts of data, and with its cloudy and rainy climate, more images need to be processed to synthesize high-quality images. Secondly, the horizontally and vertically natural conditions are highly variable, and the increased geospatial heterogeneity of tree species' spectra and phenology in a large area challenge the mapping process (Shirazinejad, Zoej, and Latifi 2022). Thirdly, Yunnan is the central part of the Indo-Burma biodiversity hotspot (Myers et al. 2000), the composition of tree species in this area is highly complex, and the local distribution of species is also highly variable. These factors mentioned above increase the difficulty of tree species mapping.

When studies focus on developing and optimizing the classification process of tree species in large areas, one aspect that deserves significant attention is the influence of environment on the distribution of tree species. Environmental factors, such as rainfall, temperature, geomorphology, and soil, have essential effects on the distribution of tree species (Yu et al. 2020). However, the spatial resolution of current environmental data is mostly lower than 1 km. For example, the TerraClimate dataset provides 4 km of temperature data (Abatzoglou et al. 2018), and the WorldClim dataset provides 1 km of bioclimatic variables (Hijmans et al. 2005). Owing to the resolution gap with remotely sensed images, environmental data are rarely incorporated in classification processes based on remotely sensed data and machine learning. Environmental data are more often used for species distribution modeling, using sample points of species associated with environmental variables to map the potential distribution areas of target species. For example, Zhou et al. used the MaxEnt model to predict the potential distribution of Cunninghamia



Figure 1. (a) Yunnan, a global biodiversity hotspot region, located in southwestern China. (b) The elevation map of the Yunnan. (c) Eight distinct floristic sub-regions of Yunnan as basic unit for classification (Li et al. 2015).

lanceolata in China (Zhou et al. 2021). Although machine learning classification and ecological niche model prediction have been extensively studied, the combination of the two has rarely been explored. Researchers are therefore missing the opportunity to synergize the two domains to obtain more accurate tree species distributions (Mouta et al. 2021). Therefore, a rising question is how to effectively integrate ecological niche models and machine learning to do classification over such a large area.

The development of remote sensing cloud computing technology and the emergence of platforms have brought new solutions to these problems. Google Earth Engine (GEE) is a cloud computing platform, whose powerful computing capabilities have changed the traditional paradigm of remote sensing data processing and analysis (Tamiminia et al. 2020). Operations such as cloud masking, image composition within time windows and time series analysis are easily implemented on GEE. Meanwhile, the GEE platform also is convenient for users to employ multiple machine learning classifiers and ecological niche models to form a processing chain. In addition, spatial partitioning has been proved to improve classification accuracy for large areas with high environmental heterogeneity (Costa et al. 2022). It is more precise and reasonable to introduce eco-geographical regionalization strategy into classification process (Moraes et al. 2021).

The blooming development of multimodal sensors and technical revolution of processing platforms and methods open the door for classifying tree species over a large area. However, analyzing the data at the desired accuracy raises new challenges. In this context, we aim to build a classification chain for accurately mapping forest stand species in the whole Yunnan province by exploring floristic regionalization, possible featurelevel and decision-level fusion strategies, and combinations. The objectives of this study were (i) to evaluate the effectiveness of using the floristic regionalization for forest stand tree species classification, (ii) to explore the significant features for separating tree species, and (iii) to develop an effective method for forest stand species classification by integrating ecological niche model and machine learning.

2. Materials

2.1. Study area

Yunnan is the most southwestern province of China (Figure 1(a)) and situated in a mountainous area. It has significant environmental heterogeneity, with diverse topography, landforms, and climate. As shown in Figure 1(b), the topography of Yunnan Province exhibits a high northwest and low southeast, with a stepped slope in the terrain, ranging in elevation from the highest mountain top 6740 m in the northwest to the lowest valley bottom 76.4 m in the southeast (Yang et al. 2004). The territory contains a variety of landform types including high mountains, hills, intermountain basins, river valleys, and karsts. Yunnan has the characteristics of low latitude, monsoon, and mountain plain climates and crosses seven climate zones: northern tropical, southern subtropical, middle subtropical, northern subtropical, southern temperate, middle temperate, and highland climate.

As shown in Figure 1(c), Yunnan is bordered by the southeastern edge of the Tibetan Plateau in the northwestern part of the country and by the Southeast Asian countries in the western and southern parts, a huge meeting place of floras and formations (Li and D 1986). Its northwestern part is introduced by the Himalayan vegetation zone and the ancient southern continental flora, while the eastern part has numerous central and southern Chinese

flora, and the northern part has the Emei, Qinling, and northwestern flora (Li and Pei 1991). The flora of southern Yunnan part is more closely related to the Indo-Malaysian flora, while the flora of southeastern Yunnan is more closely associated with the East Asian flora (Hua and Heil 2017). Geological activity and diverse topography and climate have not only forged Yunnan's unique phytogeographic divergence and vegetation geography but also made it a wellknown global biodiversity hotspot (Liu et al. 2021).

2.2. Multi-source data

This study used four types of data, Sentinel-2 images, WorldClim data, Digital Elevation Model (DEM) data and Forest Management Inventory data (FMI), to perform tree species classification.

2.2.1. Sentinel-2 data

High-quality images with low cloud ratio in Yunnan are often difficult to obtain due to persistent cloud cover. We counted the number of cloud-free images in three periods. As shown in Figure 2, there were no valid observations for some pixels throughout 2016. In contrast, the qualified observations exceed 10 over the most tiles from 7/2015 to 12/2017 and exceed 50 from 2016 to 2020. Therefore, to obtain high-quality images and extract dense time-series related features, we used 31,440 top-of-atmosphere (TOA) images from 72 MGRS tiles between 7/2015 to 12/2017 were used to extract spectral and texture features, and the images from 2016 to 2020 were used to construct the time series



Figure 2. Valid observations of Sentinel-2 images in different periods. (a) 2016.1 to 2016.12; (b) 2015.7 to 2017.12; (c) 2016.1 to 2020.12.

data. The temporal mismatch between the extracted features and FMI data year may affect the reliability of the classification. For this problem, the details were discussed in section 5.

2.2.2. Bioclimatic and topographic data

The distribution of tree species on the earth's surface is neither random nor uniform, but geographically specific and characterized by specific environmental and climatic factors (Zeb et al. 2021). Topographic and climatic heterogeneity has an important impact on tree species distribution. Therefore, bioclimatic and topographic-related factors are considered to be introduced for classification.

Bioclimatic data with a spatial resolution of 1 km were obtained from the WorldClim website (https://www. worldclim.org/data/worldclim21). These data consist of 19-dimensional bioclimatic variables with biological significance, such as mean annual temperature, annual precipitation, annual temperature difference, and precipitation during the rainy and dry (Hijmans et al. 2005). Moreover, topographic data were obtained from the Platform Space Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM).

2.2.3. Forest Management Inventory data

The Forest Management Inventory (FMI) data 2016 in China was employed as the reference data for our study. FMI aims to determine the attributes including location, natural condition, tree species composition, and distribution of each forest stand. FMI data is officially produced by local forestry and grassland administrations, under the conduction of the national administration of China. As the data has been calibrated with forest sample plots to control the error components of inventory, the data can provide a reliable reference for the different tree species.

Yunnan has about 18,000 plant species (Liu and Peng 2016), and it is impossible to identify all of them by using remote sensing imagery. We counted each forest stand's distribution area in each sub-region based on FMI and selected the top 9 species in each sub-region, which resulted in 19 species in total (Table 1).

3. Methodology

This paper proposes a classification framework combining machine learning and ecological niche models

| | I. Observeu | i ciass | anu | ιιs | acion | / . |
|--|-------------|---------|-----|-----|-------|---------------|
|--|-------------|---------|-----|-----|-------|---------------|

| | Target legend | Abbreviation |
|---------|---------------------------------|--------------|
| Level 1 | Non-forest land | / |
| | Forest land | / |
| Level 2 | Pinus yunnanensis | PY |
| | Cunninghamia lanceolata | CL |
| | Quercus L. | QL |
| | Other broadleaves | OB |
| | Eucalyptus L.Herit | EL |
| | Pinus armandii Franch. | PAF |
| | Alnus cremastogyne Burk. | ACB |
| | Betula L. | BL |
| | Cupressus funebris Endl. | CFE |
| | PopulusL. | PL |
| | Pinus kesiya var. langbianensis | PKV |
| | Abies fabri | AF |
| | Pinus densata Mast. | PDM |
| | Sassafras tsumu | ST |
| | Hevea brasiliensis | HB |
| | Acacia dealbata Link. | ADL |
| | Juglans regia | JR |
| | Picea asperata Mast. | PAM |
| | Larix gmelinii | LG |

to determine the spatial distribution of the major forest stand species in Yunnan Province. Figure 3 illustrates the whole workflow, including seven steps: (1) Spatial division; (2) Data pre-processing; (3) Feature extraction; (4) Feature reduction; (5) Classification with component classifiers; (6) Classifier fusion ensemble; and (7) Accuracy evaluation and mapping.

3.1. Floristic regionalization

We divided Yunnan Province into eight sub-regions based on the Yunnan flora system (Li et al. 2015). This system is produced by combining the distribution and phylogenetic relationships of seed plants of the genus 1983 in Yunnan. As shown in Figure 1(c), the eight sub-regions were used as the basic units for independent species mapping. Feature screening, model training, and classification were performed independently for each subregion. The goal is to make the classification and post-classification analysis applicable to local conditions and tree species, avoiding the usage of consistent feature sets for classification across the study area.

3.2. Data pre-processing

3.2.1. Sentinel-2 and topographic variables

We conducted the following steps on the acquired Sentinel-2 images to obtain cloud-free, seamless image composite. Firstly, pixels obscured and



Figure 3. Technique workflow of research.

covered by clouds were removed based on the adjusted cloud score algorithm (Oreopoulos, Wilson, and Várnai 2011). Secondly, post-cloud masking image data from 7/2015 to 12/2017 were composed by median reducer in GEE. The composite image was used to extract spectral, vegetation index, and texture features. Thirdly, we constructed the NDVI and Red-Edge Position Index (REP) time series at 5-day period, using Sentinel-2 images from 1/2016 to 12/2020. We adopted the median of all high-quality observations available within the 5-day time window to represent the observation value. When no good quality observations were available within the 5-day period, we used linear interpolation to fill the data gaps. Although these methods address the issue of missing pixel values, there is still a significant amount of noise in the data product, which hinders further analysis and utilization of the data. To mitigate noise, we applied a Savitzky-Golay (SG) filter to smooth the NDVI and REP time series, using a moving window of size 9 and a polynomial filter order of 2.

3.2.2. *Reference data.* The quality and quantity of the training samples are critical for training models. FMI data were filtered to obtain accurate and representative samples. To obtain samples for forest and non-forest classification, we conducted the following three steps. Firstly, the plots of forest (pure forest, shrub economic forest, mixed forest, other shrub woodlands, other special shrubs, economic forest, open woodland, bamboo forest) and non-forest (construction land, watershed,

cropland grassland, and other non-forest lands) were selected from FMI data. Secondly, the selected candidates were further purified according to the NDVI standard deviation calculated from pixels in each plot candidate. Candidates with high standard deviation value were removed, and centroids of the rest plot were selected as sample points. Finally, the rest samples were further purified through visually inspecting the overlaying of sample points and Sentinel-2 median composite image.

Similar steps were also employed for producing tree species samples. Firstly, plots, where the first dominant species accounts for more than 65% of the total stock volume, were seemed as pure and preserved. Secondly, the top 9 tree species in terms of coverage in each sub-region were counted from FMI, respectively, and corresponding plots were selected from the pure plots of each sub-region. Thirdly, with each plot as a basic unit, the standard deviations of Blue, NIR, SWIR, NDVI, and Greenness bands and the area were calculated, and plots with low standard deviations and large areas were preserved. Fourthly, centroids of the preserved subcompartments were selected as candidate sample points and finally purified by visually inspecting the overlaying of sample points and image composite.

The samples of each category were divided into a training sample set (75%) and a validation sample set (25%), and Figure 4 shows the numbers of training and validation data.

3.3. Feature pools and feature reduction

3.3.1. Mining features from multiple data

Tree species differ in canopy shape, phenology, and habitat. To fully understand and accurately distinguish forest from non-forest and identify species, based on comprehensively reviewing the literature, this paper extracts various features, including spectral bands, texture features, vegetation indices, and other environmental features.

Ten bands (Blue, Green, Red, RedEdge1, RedEdge2, RedEdge3, NIR, Narrow NIR, SWIR1, SWIR2) were selected from the S2 image compositing data. Based on the 10 bands, the Greenness and Wetness components of Tasseled Cap Transform were also calculated as well as 16 vegetation indices, including Triangular Vegetation Index (TVI), MERIS Terrestrial Chlorophyll Index (MTCI), Normalized Burn Ratio (NBR), Normalized Difference Built-up Index (NDBI), Inverted Red-Edge Chlorophyll Index (IRECI), Modified Chlorophyll Absorption Ratio Index (MCARI), Land Surface Water Index (LSWI), Normalized Difference 750/705 Chl NDI (Chl NDI), Normalized Difference Red-Edge Index (NDRE), Red-Edge Position Index (REP), Normalized Difference Vegetation Index (NDVI), Normalized Difference Senescent Vegetation Index (NDSVI), Normalized Difference Tillage Index (NDTI), Normalized Difference Water Index (NDWI), and Red Edge Chlorophyll Index (Clrededge). The formulas for vegetation indices are shown in Table 2.



Figure 4. The number of samples for each observed class.

| Indices | Formulation | Reference |
|-----------|--|---------------------------------------|
| TVI | $\textbf{0.5} \times (\textbf{120} \times (\rho_{RedEdge1} - \rho_{Green}) - (\textbf{200} \times (\rho_{Red} - \rho_{Green})))$ | (Main et al. 2011) |
| MTCI | $(\rho_{\text{RedEdge2}} - \rho_{\text{RedEdge1}}) / (\rho_{\text{RedEdge1}} - \rho_{\text{NIR}})$ | (Dash and Curran 2004) |
| NBR | $(\rho_{\text{NIR}} - \rho_{\text{SWIR2}})/(\rho_{\text{NIR}} + \rho_{\text{SWIR2}})$ | (Long et al. 2019) |
| IRECI | $(ho_{RedEdge3} - ho_{Red1}) / (ho_{RedEdge1} / ho_{RedEdge2})$ | (Rozenstein et al. 2019) |
| MCARI2 | $((\rho_{RedEdge2} - \rho_{RedEdge1}) - 0.2 \times (\rho_{RedEdge2} - \rho_{Green})) \times (\rho_{RedEdge2} / \rho_{RedEdge1})$ | (Wu et al. 2008) |
| LSWI | $(\rho_{\text{NIR}} - \rho_{\text{SWIR1}})/(\rho_{\text{NIR}} + \rho_{\text{SWIR1}})$ | (Chandrasekar et al. 2010) |
| NDRE | $(ho_{NIR}- ho_{RedEdge1})/(ho_{NIR}+ ho_{RedEdge1})$ | (Ahamed et al. 2011) |
| Chl NDI | $(\rho_{RedEdge2} - \rho_{RedEdge1})/(\rho_{RedEdge2} + \rho_{RedEdge1})$ | (Richardson, Duigan, and Berlyn 2002) |
| REP | $705 + 35 \times (0.5 \times (\rho_{Red} + \rho_{RedEdge3}) - \rho_{RedEdge1}) / (\rho_{RedEdge2} - \rho_{RedEdge1})$ | (Schlerf, Atzberger, and Hill 2005) |
| NDVI | $(\rho_{NIR} - \rho_{Red})/(\rho_{NIR} + \rho_{Red})$ | (Carlson and Ripley 1997) |
| NDSVI | $(ho_{SWIR1} - ho_{Red})/(ho_{SWIR1} + ho_{Red})$ | (Zheng et al. 2019) |
| NDTI | $(ho_{SWIR1} - ho_{RedEdge2})/(ho_{SWIR1} + ho_{RedEdge2})$ | (Zhong, Hu, and Zhou 2019) |
| NDSI | $(ho_{SWIR1} - ho_{SWIR2})/(ho_{SWIR1} + ho_{SWIR2})$ | (Richardson, Duigan, and Berlyn 2002) |
| NDBI | $(ho_{SWIR1} - ho_{NIR})/(ho_{SWIR1} + ho_{NIR})$ | (Zhang, Odeh, and Han 2009) |
| NDWI | $(oldsymbol{ ho}_{Green} - oldsymbol{ ho}_{NIR})/(oldsymbol{ ho}_{Green} + oldsymbol{ ho}_{NIR})$ | (Gao 1996) |
| Clrededge | $ ho_{ m NIR}/ ho_{ m RedEdge1}-1$ | (Duan et al. 2021) |

Table 2. The vegetation indices and corresponding formulas used in this study.

In addition, we computed six texture features (Sum Average, Correlation, Variance, Dissimilarity, Contrast, and Cluster shade) for each of the 10 spectral bands and the NDVI, so there are 66 texture features in total. Five-day interval time series data indices were computed from both NDVI and REP indices, resulting in a feature set with 146 bands (73×2). We also obtained 19 dimensions of bioclimatic variables from the WorldClim data and 3 topographic factors, namely Elevation, Slope, and Aspect, from the SRTM data using the GEE built-in functions.

Based on the basic features mentioned above, two feature pools were constructed. As shown in Table 3, the first feature pool consists of 37-dimensional features, for forest and non-forest classification. The second feature pool consists of 257-dimensional features (Table 4) for forest stand species classification. Among them, 22dimensional environmental features, including 19dimensional bioclimatic features and 3-dimensional topographic features, were used as input features for MaxEnt; 237-dimensional feature sets, named the *ALL Feature Set*, were selected as inputs to three machine learning models, SVM, RF, and XGBoost.

3.3.2. Feature selection

The number and quality of features largely determine the algorithm's efficiency. Selecting features with optimal separability is crucial for computational efficiency and classification performance. Thus, we optimized features before classification.

For the MaxEnt model, the factors were evaluated and selected by the following three steps. (1) Spearman

| Table 3. The feature candidates for the forest mask generation | Table 3 | . The | feature | candidates | for the | forest | mask | generation |
|--|---------|-------|---------|------------|---------|--------|------|------------|
|--|---------|-------|---------|------------|---------|--------|------|------------|

| Туре | Proxies | Number of features |
|--------------------------------|---|-----------------------|
| Spectral Bands | Blue, Green, Red, RedEdge1, RedEdge2, RedEdge3, NIR, Narrow NIR, SWIR1, SWIR2 | 10 |
| Vegetation Index | TVI, LSWI, NDWI, NDBI, NBR, CIrededge, NDSI, NDTI, ChI NDI, REP, NDRE, MCARI2, MTCI, IRECI, NDVI, NDSVI | 16 |
| Texture | Sum Average, Correlation, Variance, Dissimilarity, Contrast, Cluster shade (Calculated for NDVI) | 6=1×6 |
| Tasseled Cap Transformation | Wetness, Greenness | 2 |
| Topographic | Elevation, Aspect, Slope | 3 |

| Туре | Proxies | Number of features |
|--------------------------------|---|-----------------------|
| Spectral Bands | Blue, Green, Red, RedEdge1, RedEdge2, RedEdge3, NIR, Narrow NIR, SWIR1, SWIR2 | 10 |
| Vegetation Index | TVI, LSWI, NDWI, NDBI, NBR, CRE, NDSI, NDTI, ĆHI NDI, REP, NDRE, MCARI2, MTCI, IRECI, NDVI, NDSVI | 16 |
| Texture | Sum Average, Correlation, Variance, Dissimilarity, Contrast, Cluster shade (Calculated for Blue, Green, Red, RedEdge1, RedEdge2, RedEdge3, NIR, Narrow NIR, SWIR1, SWIR2) | 60=10×6 |
| Tasseled Cap Transformation | Wetness, Greenness | 2 |
| Topographic | Elevation, Aspect, Slope | 3 |
| Time Series | NDVI, REP | 146=73×2 |
| WorldClim | Bio01, Bio02, Bio03, Bio04, Bio05, Bio06, Bio07, Bio08, Bio09, Bio10, Bio11, Bio12, Bio13, Bio14, Bio15, Bio16, Bio17, Bio18, Bio19 | 19 |

correlation analysis was first performed on the variables. (2) Pre-training was performed using all environmental variables, and the importance of each dimensional variable was evaluated using Jackknife test. (3) The variable most significant in the Jackknife test was selected among the highly correlated variables (absolute value of correlation ≥ 0.8) (i.e. the model generated from all variables except this variable and the model generated using only this variable had the smallest difference in regularized training gain). Highly correlated and redundant variables were removed, and each environmental factor that dominates the prediction was finally identified. The selected features consist of an optimal feature set for the MaxEnt predicting model and called Feature Set 1 in the following section.

For SVM, RF, and XGBoost classifier, the relevance hierarchical clustering method (You et al. 2021) was used to select the best input features. This feature selection process is performed in four steps. The normalization method is firstly used to map all features to a distribution with mean value of 0 and standard deviation value of 1. Pre-training is secondly performed to determine the number of feature dimensions when the accuracy reaches saturation point. The feature importance of the features is evaluated by the Gini importance or mean reduction impurity calculated by the RF classifier and ranked by feature importance score. Spearman correlation analysis is performed on the most important features. The most important features are clustered into several clusters by the maximum depth threshold, and finally the features with the highest feature importance score in each cluster are retained. In this study, the maximum depth threshold is set to 0.5 to preserve the important features with high separability and low correlation among the selected features. The optimal features are selected for stand tree species classification, referred to as Feature set 2 in the following sections, and the set of best features for classification and non-forest classification, hereinafter referred to as Feature set 3.

It is worth noting that, as the above-mentioned above selection process independently operated in each floristic region, *Feature Set 1, Feature Set 2*, and the *Feature Set 3* may be different from region to region.

3.4. Classification with component classifiers

MaxEnt derives the constraints on species distribution, then seeks the possible distribution with maximum entropy. When the entropy achieves maximum, the species occurrence probability distribution is closest to the true distribution of species (Phillips, Anderson, and Schapire 2006). The prediction results of MaxEnt are reflected by the probability of species occurrence. And the larger the value, the more suitable environmental conditions for species survival. Due to their superior and stable classification performance, three machine learning classifiers, including RF, SVM, and XGBoost, are the most popular and widely used in various applications (Abdi 2020; Grabska, Frantz, and Ostapowicz 2020).

The four component models are trained and validated using the same samples, and empirically adjustment method was used to obtain the hyperparameters for each classifier. The MaxEnt model classification was implemented on local computer using the maxent software (version 3.4.1), while the three machine learning classifiers were implemented on the GEE platform.

3.5. Classification by multi-classifier fusion

After the classification results of the four component classifiers are obtained, we construct two paradigms to fuse MaxEnt and the three machine learning classifiers parallelly (Figure 5(a)) and serially (Figure 5(b)), respectively. The integration schematic of the two paradigms is shown in Figure 5.

The first paradigm is called the parallel ensemble model. Classification is performed by four component classifiers parallelly and followed by decision fusion based on a weighted voting strategy. The final fusion objective function is defined as (1).

$$Y^{*} = \underset{Y}{\operatorname{argmax}} \left(\sum_{i \in S} \sum_{j=1}^{4} \log \left(w_{j} l_{i, y_{i}}^{\operatorname{Md}_{j}} \right) \right)$$
(1)

where S is the set of pixel locations, $I_{i,y_i}^{Md_j}$ denotes a probability predicted by the model *j*, that a pixel i belongs to a given class y_i , and w_j is a weight measuring confidence of the model *j*. Intuitively, the setting of w_j is related to the training performance of models. Therefore, weights for each component classifier are calculated based on the training accuracies with $w_i = \frac{Acc(C_i)}{\sum_{j=1}^{4} Acc(C_j)}$ and $\sum_{j} w_j = 1$. The second integration paradigm is a serial mode,

The second intégration paradigm is a serial mode, where the output of the MaxEnt classifier, in the form of class-specific possibilities, is stacked together with Feature set 2 and fed into the three machine learning classifiers, respectively. In this process, the possibility of each class is seemed as a feature.



Figure 5. Schematic diagram of two decision fusion strategies.

4. Results

We implemented the proposed classification scheme in Yunnan province. First, we divided Yunnan Province into eight sub-regions. Then, the best classification features were selected for each subzone from the candidate features. Finally, two levels of classification, forest cover and tree species, were executed in each sub-region. In this section, experimental setting, the mapping results of forest cover and species are described. (Grabska et al. 2019; Li et al. 2022). Therefore, the experiments focus on analyzing the performance of species-level classifications. As shown in Table 5, totally 11 feature-classifier combinations are compared to analyze the importance of features and the effectiveness of the proposed fusing strategies. Moreover, visual inspection and quantitative metrics, including confusion matrix and derived Overall Accuracy (OA) and Kappa, are employed for assessing classification performance.

4.1. Experimental setting and accuracy assessment

Many studies proved that forest-type classification could often be achieved with high accuracy

Feature Set 2

Feature Set 2

Feature Set 2

Ensemble

4.2. Forests cover mapping

XGBoost

XGBoost

Weighted voting

SVM RF

SVM

RF

Figure 6 shows the number of features selected and the classification accuracy for each sub-region for both levels of forest cover and tree species

Abbreviation

M XALL SALL RALL

XOPT SOPT

ROPT

MXSR

MX

MS

MR

| Туре | Feature setting | Classifier/fusion rule |
|-----------|-----------------|------------------------|
| Component | Feature Set 1 | MaxEnt |
| | All Feature Set | XGBoost |
| | All Feature Set | SVM |
| | All Feature Set | RF |

The probability outputs of M, XOPT, SOPT, ROPT

The probability output of M, SOPT and Feature Set 2

The probability output of M, ROPT and Feature Set 2

The probability output of M and Feature Set 2

Table 5. Feature-classifier combinations explored in this study.



Figure 6. The numbers of optimal features and overall accuracies in the 8 sub-regions. Darker red tones mean higher accuracy for species classification.

classification. Depending on the spatial location, the number of features used for forest classification ranged from 8 to 11 dimensions, and the number of features for tree species classification ranged from 26 to 41 dimensions. The OA ranged from 93.91% to 98.06% at the forest level, and from 70.12% to 75.24% at the tree species level.

The forest and non-forest classification was firstly performed. Since the classification error in this level will transfer to the species-level classification, the classification accuracy is expected as high as possible. In this study, the OA is 96.88%, and the Kappa coefficient is 0.96, according to 27,296 reference samples. The classification results are shown in Figure 7. The classification results are highly consistent with the forest distribution in the field. Therefore, this result can be used for producing forest mask.

4.3. Result at species level

In each sub-region, the accuracy of the 11 classification schemes was evaluated based on the validation samples. The OAs of each partition are shown in Table 6 and Figure 6.

The accuracies of the seven component classifiers do not exceed 67.40% in all eight partitions, and these accuracies may be insufficient to support further analysis and decision-making. In contrast, the integrated model



Figure 7. Forest/non-forest classification result of Yunnan province.

significantly improves the classification accuracy. Among the integration models, the average accuracy of MXSR over eight regions is 69.51%, and the average accuracy of the three serial integration models, MS, MX, and MR, is higher than 71.2%. Overall, the classification performance of the integrated classifier is much higher than that of the component classifier. Among the ensemble models, three serial integrated ensembles are superior to the parallel integrated model.

By mosaicking the MR classification results of the eight sub-regions, we obtained the spatial distribution map of 19 target forest species in Yunnan Province at 10 m resolution (Figure 8). The inset details the spatial distribution of tree species in the selected sites. The accuracy was evaluated by 53,934 validation sample points, and the overall accuracy was 72.18% with

a kappa coefficient of 0.69. The heat map of confusion matrix is given in Figure 9, from which the four classes of *Larix gmelinii*, *Hevea brasiliensis*, *Abies fabri*, and *Juglans regia* achieved high accuracies, while *Alnus cremastogyne Burk.*, *Quercus L.*, and *Sassafras tsumu* were classified with low accuracies.

5. Discussion

We have divided the Yunnan region into eight subregions, with features extracted from S2, SRTM, and WorldClim data, and used four component classifiers to map the spatial distribution of forest stand species. This section discusses the impacts of floristic regionalization, multiple data and the component classifiers and integrated models on the classification performance.

| Table | 6. | Accuracy | asses | sment | for | different | clas | sification | scenarios | in | each | sub- |
|------------|----------|--------------|------------|-------------|--------|-----------|-------------|------------|-----------|----|------|-------|
| region. Tl | he highe | st accuracie | s over all | sub-regions | are in | bold. | | | | | | |
| | | I | II | | | IV | V | VI | VII | | VIII | Mean |
| Classifier | | | | | | Overa | ll accuracy | y (%) | | | | |
| М | 5 | 53.44 | 49.11 | 49.58 | | 42.59 | 42.59 | 41.33 | 51.68 | 4 | 8.68 | 47.38 |
| XALL | 6 | 57.76 | 65.70 | 67.92 | | 67.35 | 68.39 | 65.17 | 67.20 | 6 | 4.28 | 66.72 |
| SALL | 6 | 58.95 | 66.61 | 69.42 | | 65.39 | 68.31 | 65.37 | 68.74 | 6 | 6.39 | 67.40 |
| RALL | 6 | 58.57 | 66.57 | 68.25 | | 65.39 | 66.79 | 65.21 | 66.56 | 6 | 4.80 | 66.52 |
| XOPT | 6 | 57.30 | 65.07 | 66.35 | | 64.14 | 66.08 | 63.20 | 65.39 | 5 | 9.63 | 64.65 |
| SOPT | 6 | 57.58 | 66.63 | 65.30 | | 66.23 | 66.56 | 63.60 | 67.74 | 6 | 0.35 | 65.50 |
| ROPT | 6 | 57.83 | 65.53 | 66.92 | | 64.87 | 66.41 | 64.30 | 66.14 | 6 | 0.35 | 65.29 |
| MXSR | 7 | /1.89 | 72.22 | 71.55 | | 66.46 | 70.99 | 67.70 | 71.07 | 6 | 4.19 | 69.51 |
| MX | 7 | /3.76 | 71.15 | 72.88 | | 69.49 | 71.37 | 68.79 | 74.02 | 6 | 8.43 | 71.24 |
| MS | 7 | 2.93 | 71.65 | 72.31 | | 70.16 | 70.79 | 68.13 | 75.16 | 7 | 0.47 | 71.45 |
| MR | 7 | 4.44 | 72.21 | 74.12 | | 70.34 | 72.52 | 70.12 | 75.24 | 7 | 1.98 | 72.62 |



Figure 8. The spatial details of the forest stand species map in Yunnan Province in 2016. Site a (103.279°E, 26.851°N); Site b (103.626°E, 25.910°N); Site c (103.031°E, 25.434°N); Site d (104.568°E, 23.026°N); Site e (101.347°E, 21.946°N); Site f (99.173°E, 25.318°N); Site g (99.688°E, 27.743°N).

5.1. The necessity of floristic regionalization

Yunnan Province contains a variety of landform types including high mountains, hills, intermountain basins, river valleys, and karsts, as well as seven climate types and seven vegetation types. The corresponding physical differences in climate and biota between the North and South are equivalent to the differences from Hainan Island, China to Changchun, Northeast China (Yang et al. 2004). Spatial zoning is a practical solution for large areas with different growth conditions and complex tree species composition.

In summary, the floristic regionalization can bring four obvious strengths. First, zoning according to vegetation zones creates zonation with uniform ecological and spectral characteristics. Compared to the whole area, a sub-region has a higher consistency of geological history, geomorphology, climate, and vegetation composition, reducing spatial and temporal variability in the tree species' spectral and phenological response (Cano et al. 2017). Second, each partition has a different species composition, and feature selection was independently performed for each partition to avoid classification with a set of feature combinations over a very large area. Third, the partitioning increases the proportion of the number of samples and alleviates the spatial and quantitative imbalance of samples to some extent. Fourth, by classifying the eight regions sequentially over the GEE platform, the computational burden and demand on storage are also reduced.

5.2. Variable importance assessment

5.2.1. The importance of Sentinel-2 data in stand species classification

The mean spectral reflectance of the 19 tree species extracted from July 2015 to December 2017 median composite images are shown below in Figure 10.

| PY85052626481341754232643529141266870011100342CL29040813642645101157100616444293660000QL7082194179460207318266628392145612146671QL313177150476207318266628392145013758992QL3131771501763087427430500106000075299PAF53481509159255193157122670170001449ACE2241302092429921341030180010011010101ACE2241302092429921341030141300111111111111111111111111111 <th>R PAM</th> <th>ADL JF</th> <th>8 A</th> <th>HB</th> <th>ST</th> <th>PDM</th> <th>AB</th> <th>PKV</th> <th>PL</th> <th>CFE</th> <th>BL</th> <th>ACB</th> <th>PAF</th> <th>EL</th> <th>OB</th> <th>QL</th> <th>CL</th> <th>PY</th> <th></th> | R PAM | ADL JF | 8 A | HB | ST | PDM | AB | PKV | PL | CFE | BL | ACB | PAF | EL | OB | QL | CL | PY | |
|---|-------|--------|-----|-----|----|------|------|------|-----|-----|-----|------|------|------|------|------|------|------|-----|
| CL2904088136482451011571006164444293.6.0.0.0.QL708219417946020731826662839214561214671OB31950642647591541873019920369554327558892CL31317715017630874274305000106000072929PAF53481509159255193157122670170004444ACB224130209242299213310301800201810ACB701079712626556898001800230101010CFE10627423651242104103200002300010CFE106274236512421010114000000000000< | 34 25 | 10 34 | 1 | 1 | 0 | 87 | 6 | 12 | 14 | 29 | 35 | 264 | 423 | 175 | 134 | 648 | 262 | 8595 | PY |
| QL70821941794602073182666283921456121466714OB3195004264759154187301992036950432755889925EL3131771501763087427430506010660000075889926PAF53481509159255193157122677011700000444444ACB22413020924229921331030180002000180GE10617797126265589800180002100180GE107971262656898001800022300180GE107971262656898001800000000000000000000000000000000000< | 0 0 | 0 0 | | 6 | 3 | 29 | 4 | 44 | 16 | 6 | 100 | 157 | 101 | 45 | 482 | 136 | 4088 | 290 | CL |
| OB319506426475915418730199203695432755895EL31317715017630874274305001060000752917PAF53481509159255193157122670170004441ACB2241302092422992133410301800022300181BL76107971262655689800460002021000CFE1062742365124210410322000021000PK323465617702012014100 <t< td=""><td>7 11</td><td>6 7</td><td></td><td>14</td><td>12</td><td>56</td><td>14</td><td>92</td><td>3</td><td>8</td><td>62</td><td>266</td><td>318</td><td>207</td><td>460</td><td>4179</td><td>219</td><td>708</td><td>QL</td></t<> | 7 11 | 6 7 | | 14 | 12 | 56 | 14 | 92 | 3 | 8 | 62 | 266 | 318 | 207 | 460 | 4179 | 219 | 708 | QL |
| EL 313 177 150 176 3087 42 74 30 50 0 106 0 0 0 0 75 29 1 PAF 534 81 509 159 25 5193 157 12 26 7 0 1 70 0 0 4 </td <td>9 52</td> <td>8 9</td> <td></td> <td>55</td> <td>27</td> <td>43</td> <td>5</td> <td>69</td> <td>3</td> <td>20</td> <td>99</td> <td>301</td> <td>187</td> <td>154</td> <td>4759</td> <td>426</td> <td>506</td> <td>319</td> <td>OB</td> | 9 52 | 8 9 | | 55 | 27 | 43 | 5 | 69 | 3 | 20 | 99 | 301 | 187 | 154 | 4759 | 426 | 506 | 319 | OB |
| PAF 534 81 509 159 25 5193 157 12 26 7 0 1 70 0 0 4 4 4 ACB 224 130 209 242 29 92 1334 10 3 0 18 0 0 2 0 0 18 10 BL 76 107 97 126 26 5 56 898 0 0 46 0 0 2 3 0 18 0 CFE 106 27 4 23 65 124 21 0 410 3 2 0 0 0 0 21 0 0 0 10 11 0 0 10 11 10 10 11 10 10 11 10 | 29 0 | 75 29 | 7 | 0 | 0 | 0 | 0 | 106 | 0 | 50 | 30 | 74 | 42 | 3087 | 176 | 150 | 177 | 313 | EL |
| ACB 224 130 209 242 29 92 1334 10 3 0 18 0 0 2 0 0 18 18 BL 76 107 97 126 26 5 56 898 0 0 46 0 0 2 3 0 0 0 16 17 12 2 12 10 10 3 2 0 0 12 3 0 0 1 </td <td>4 6</td> <td>4 4</td> <td></td> <td>0</td> <td>0</td> <td>70</td> <td>1</td> <td>0</td> <td>7</td> <td>26</td> <td>12</td> <td>157</td> <td>5193</td> <td>25</td> <td>159</td> <td>509</td> <td>81</td> <td>534</td> <td>PAF</td> | 4 6 | 4 4 | | 0 | 0 | 70 | 1 | 0 | 7 | 26 | 12 | 157 | 5193 | 25 | 159 | 509 | 81 | 534 | PAF |
| BL 76 107 97 126 26 5 56 898 0 0 46 0 0 2 3 0 0 0 CF 106 27 4 23 65 124 21 0 410 3 2 0 0 0 0 21 0 0 14 0 0 0 0 0 10 10 < | 18 2 | 0 18 | | 0 | 2 | 0 | 0 | 18 | 0 | 3 | 10 | 1334 | 92 | 29 | 242 | 209 | 130 | 224 | ACB |
| CFE 106 27 4 23 65 124 21 0 410 3 2 0 0 0 0 21 0 1 0 1 0 1 0 1 0 1 0 < | 0 0 | 0 0 | | 3 | 2 | 0 | 0 | 46 | 0 | 0 | 898 | 56 | 5 | 26 | 126 | 97 | 107 | 76 | BL- |
| PL 41 13 38 2 14 57 37 28 7 141 0 0 0 0 0 11 0 11 0 0 0 0 0 11 0 0 0 0 0 11 0 0 0 0 0 0 0 0 11 0 | 0 0 | 21 0 | 2 | 0 | 0 | 0 | 0 | 2 | 3 | 410 | 0 | 21 | 124 | 65 | 23 | 4 | 27 | 106 | CFE |
| PKV 32 34 65 61 77 0 20 12 0 0 1192 0 0 0 66 0 0 0 0 0 0 1192 0 0 10 10 00 | 0 0 | 1 0 | | 0 | 0 | 0 | 0 | 0 | 141 | 7 | 28 | 37 | 57 | 14 | 2 | 38 | 13 | 41 | PL |
| AB 3 5 11 3 0 0 0 0 0 1112 43 0 </td <td>0 0</td> <td>0 0</td> <td></td> <td>6</td> <td>0</td> <td>0</td> <td>0</td> <td>1192</td> <td>0</td> <td>0</td> <td>12</td> <td>20</td> <td>0</td> <td>77</td> <td>61</td> <td>65</td> <td>34</td> <td>32</td> <td>PKV</td> | 0 0 | 0 0 | | 6 | 0 | 0 | 0 | 1192 | 0 | 0 | 12 | 20 | 0 | 77 | 61 | 65 | 34 | 32 | PKV |
| PDM 91 99 15 24 0 97 59 6 0 0 0 25 1617 0 < | 0 68 | 0 0 | | 0 | 0 | 43 | 1112 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 11 | 5 | 3 | AB |
| ST 0 21 10 82 0 3 0 17 0 0 0 0 91 0 0 0 0 HB 2 0 2 56 1 0 0 0 0 4 0 0 0 746 0 0 0 | 0 33 | 0 0 | | 0 | 0 | 1617 | 25 | 0 | 0 | 0 | 6 | 59 | 97 | 0 | 24 | 15 | 99 | 91 | PDM |
| HB 2 0 2 56 1 0 0 0 0 0 4 0 0 0 746 0 0 | 0 0 | 0 0 | | 0 | 91 | 0 | 0 | 0 | 0 | 0 | 17 | 0 | 3 | 0 | 82 | 10 | 21 | 0 | ST |
| | 0 0 | 0 0 | 5 | 746 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 56 | 2 | 0 | 2 | HB |
| ADL 27 0 4 1 161 7 11 0 11 0 0 0 0 0 442 9 | 9 0 | 442 9 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 | 11 | 7 | 161 | 1 | 4 | 0 | 27 | ADL |
| JR 7 0 17 7 56 4 11 0 0 0 0 0 0 0 0 0 7 481 | 81 0 | 7 48 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 4 | 56 | 7 | 17 | 0 | 7 | JR |
| PAM 5 0 6 28 0 2 1 0 0 0 117 114 0 0 0 0 4 | 0 450 | 0 0 | | 0 | 0 | 114 | 117 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 28 | 6 | 0 | 5 | PAM |
| LG 0 0 2 0 0 0 0 0 0 0 0 0 37 13 0 0 0 0 2 | 0 20 | 0 0 | | 0 | 0 | 13 | 37 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | LG |

Figure 9. Confusion matrix heatmap of tree species classification.

Although S2 can provide more detailed spectral and spatial information than Landsat images, it is still insufficient to distinguish between tree species. This conclusion is also confirmed by the OA shown in Table 6, where the classification accuracy using the features from S2 failed to reach 70%.

The reflectance differences among the species in the Red-Edge to SWIR1 bands are higher than in other spectral bands, and in this spectral band, the reflectance of broadleaf trees was higher than that of conifers, with higher values for *Betula L*. and *Hevea brasiliensis*, while *Picea asperata Mast.*, *Abies fabri*, and *Larix gmelinii* had lower reflectance, and in this spectral range *Sassafras tsumu* and *Larix gmelinii* in reflectance differed significantly from the other species, indicating

the usefulness of the three bands in distinguishing tree species. However, as expected, the reflectance differences between coniferous and broadleaf trees are much smaller in other bands, and there is a large overlap between the different tree species.

To evaluate the contribution of texture and time-series features to classification, based on three feature combinations, we assessed the separability between tree species. Three feature combinations include (a) raw bands and vegetation indices, (b) spectral features and texture features, and (c) spectral features and time series features.

The Jeffries-Matusita (JM) (Ma et al. 2021) distance between tree species is calculated and shown in Figure 11. Higher JM values indicate greater



Figure 10. The mean spectral reflectance of 19 tree species, calculated from the Sentinel-2 bands.

separability of the two classes. In this study, the JM distances are classified into four classes: strongly separable (1.9–2.0), better separable (1.8–1.9), weakly separable (1.7–1.8), and poorly separable (<1.7). The JM distance between two different categories is expected to be greater than 1.8 to obtain a satisfactory classification. Several species do not co-exist in the same subregion, so the JM distances were not calculated.

In Scheme a, the separability of most tree species pairs is poor due to the high similarity in spectral reflectance of the tree species. In Scheme b, separability between species improved significantly with the inclusion of textural features, but sassafras was still difficult to discriminate from *Sassafras tsumu*, *Pinus yunnanensis*, and *Alnus cremastogyne Burk.*. In Scheme c, the JM distances between species all reached over 1.9, suggesting that the time series feature can increase the separability of species.

5.2.2. The importance of environmental data in stand species classification

Regional arboreal species composition can be largely explained in terms of a long history of biogeographic and evolutionary events. Remote sensing-based tree species classification usually concentrates on data and methods, often ignoring the rich context that environmental data can provide.

In our study, the most important variable for stand species classification is Elevation, ranked highest in evaluation quantitative importance scores in all eight subregions. By overlaying the classification results with the elevation maps, the elevation percentage map of each tree species was obtained (Figure 12), which also reflected the differences in elevation distribution of each tree species. For example, *Larixgmelinii* and *Abies fabri* are distributed above 3200 m, while *Heveabrasiliensis* and *Sassafras tsumu* occur at lower elevations below 1500 m. Numerous studies confirm that topographic features facilitate tree species classification, especially for species that follow an atural height gradient (Grabska, Frantz, and Ostapowicz 2020).

Each species has its own preferred temperature range, acceptable temperature extremes, and rainfall. A large amount of literature indicates that climate conditions are the key factor determining the large-scale distribution patterns of organisms (Datta, Schweiger, and Kühn 2020). Our study uses environmental factors and the MaxEnt model to assign the probability of tree species occurrence for each geographical location. As shown in Table 7, adding these probability layers in two forms to the process based on remote sensing image classification can improve classification accuracy. The constraint of environmental factors on species distribution increases classification performance. Figure 9 shows that the classification accuracy of the categories with higher environmental constraints is also higher, such as Hevea brasiliensis, Abies fabri, and Larix gmelinii. Other species may be excluded due to lower environmental tolerance in the growing areas of these species, making these species have higher homogeneity. However, in our study, the range of some tree species was over-predicted. The climate variables used indicate that these tree species should have a wider geographical range than in reality, which may be because the climate



Figure 11. JM distances between species with (a) spectral bands and vegetation indices shown in Table 4, (b) spectral features and texture features, and (c) spectral features and time series features.

factors we used are not the main limiting factors for the geographical range of tree species (Wiens and Graham 2005). In addition, the characteristics of the environmental factors used have a resolution of 1 km, and the low resolution ignores the micro-scale processes of local effects of tree species. Many tree species require specific small-scale habitat attributes (Sinclair, White, and Newell 2010). We believe that future research needs to consider more ecological and geographical factors, such as the CHELSA dataset, which contains explicit indices for many ecological and physiological processes. Our test results in local areas show that this data can provide more accurate distribution ranges. Of course, to generate high-precision spatiotemporal microclimate, it may be necessary to combine global climate datasets with in situ microclimate measurements, long-term meteorological station data, and high-resolution remote sensing data (Bobrowski, Weidinger, and Schickhoff 2021).

5.3. Species mapping performance by fusion multiple models

To evaluate the mapping performance of the MR serial model, we compare the classification performance of the M, ROPT, and MR models in this section. The MR method achieves the best accuracy in the vast majority of sub-regions (7 out of 8). Table 7 visually demonstrates the advantages of the MR integrated model. The accuracy difference between MR and M for the eight sub-regions is from 21.00% to 31.06% with an average difference of 25.39%. The difference in accuracy with

| LG- | 0% | 0% | 0% | 0% | 0% | 0% | 0.11% | 30.1% | 69.78% | 75% |
|-------|----------|--------|-------------|-----------|------------|----------|--------|-----------|--------|---------|
| AF - | 0% | 0% | 0% | 0% | 0% | 0% | 8.16% | 49.82% | 42% | 68% |
| PAM - | 0% | 0% | 0% | 0% | 0% | 4.69% | 41.03% | 52.49% | 1.73% | 00 /0 |
| PDM - | 11.59% | 7.8% | 6.09% | 4.64% | 3.12% | 2.35% | 41.5% | 22.88% | 0% | 60% |
| PF - | 1.16% | 5.7% | 4.76% | 10.69% | 31.8% | 33.2% | 11.7% | 0.98% | 0% | |
| JR - | 0% | 0.92% | 10.69% | 28.41% | 41.06% | 18.08% | 0.71% | 0.13% | 0% | 53% |
| ADL - | 0% | 0% | 4.13% | 41.12% | 38.24% | 16.39% | 0% | 0% | 0% | |
| PL - | 0.43% | 2.74% | 8.67% | 46.84% | 38.33% | 2.94% | 0% | 0% | 0% | 45% |
| ACB - | 4.87% | 7.98% | 16.53% | 26.01% | 27.18% | 15.25% | 2.16% | 0% | 0% | |
| QL - | 3.58% | 9.35% | 13.86% | 15.95% | 19.41% | 18.3% | 10.84% | 7.2% | 1.51% | 38% |
| PY- | 0.76% | 9.25% | 16.77% | 23.44% | 28.83% | 14.9% | 5.94% | 0.12% | 0% | |
| EL - | 4.2% | 10.35% | 32.43% | 33.93% | 12.89% | 3.99% | 1.74% | 0.42% | 0% | 30% |
| CFE - | 1.13% | 12.36% | 28.17% | 24.26% | 31.24% | 2.5% | 0.18% | 0.14% | 0% | |
| BL - | 0.27% | 12.43% | 43.75% | 23.55% | 8.62% | 8.25% | 3.06% | 0% | 0% | 23% |
| OB - | 17.57% | 27.11% | 22.21% | 14.47% | 9.66% | 4.43% | 2.98% | 1.57% | 0% | |
| CL - | 8.12% | 30.58% | 28.83% | 15.58% | 7.84% | 6.05% | 2.93% | 0% | 0% | 15% |
| PKV- | 8.76% | 45.34% | 39.52% | 6.17% | 0.21% | 0% | 0% | 0% | 0% | |
| ST - | 39.19% | 50.1% | 7.52% | 3.11% | 0% | 0% | 0% | 0% | 0% | 8% |
| HB - | 75.05% | 24.91% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 09/ |
| | ent | .511 | -710 | -2112 | -011 | - 679 | .011 | -110 | .011 | 0 /0 |
| 76 | 9602 068 | 134.5- | 161 15 677- | 1983-083- | 23031 2095 | 2720-726 | 32495 | 3891 291- | 6740 | |
| | 90 | 12 | 10. | 190 | 250 | 212 | 34 | 30/ | | |

Figure 12. Distribution map of forest stand species in the altitude range.

 Table 7. Comparison of classification accuracies of different methods in each sub-region.

| Region | М | ROPT | MR | MR-M | MR-ROPT |
|--------|--------|--------|--------|--------|---------|
| 1 | 53.44% | 67.83% | 74.44% | 21.00% | 6.61% |
| II | 49.11% | 65.53% | 72.20% | 23.09% | 6.67% |
| III | 49.58% | 66.92% | 74.12% | 24.54% | 7.20% |
| IV | 42.59% | 64.87% | 70.34% | 27.75% | 5.47% |
| V | 41.46% | 66.41% | 72.52% | 31.06% | 6.11% |
| VI | 41.33% | 64.30% | 70.12% | 28.79% | 5.47% |
| VII | 51.68% | 66.14% | 75.24% | 23.56% | 9.10% |
| VIII | 48.68% | 60.35% | 71.98% | 23.30% | 11.63% |
| Mean | 47.23% | 65.29% | 72.62% | 25.39% | 7.33% |

ROPT is 11.63% at the maximum and 5.47% at the minimum. Although the M classifier obtained the lowest OA in each partition, adding the classification results of the M classifier as features to ROPT resulted in a better classifier than the component classifier MR. The great advantage in classification accuracy proves the effectiveness and potential of the MR-integrated model for mapping tree species in forest stands. In addition, the classification accuracy of the MR serial integration method was higher than 70% in 8 partitions with different ecological conditions, which proved that the method can be used for tree stand species mapping under diverse ecological conditions with good stability.

Figure 13 shows the S2 false color images (the combination of bands 3, 4, and 8), the FMI data, and the classification results of the three methods, from four selected regions for visual evaluation. We can find that the classification results of different classifiers differed greatly. The M classifier with low resolution environmental data for classification predicted poorly that the geographically expressed area of each tree species was larger than its actual distribution area, which is consistent with the conclusions reached in previous studies (Chandra et al. 2021; Pshegusov et al. 2022). The RF model, which uses higher resolution remote sensing data for classification, provides finer details than the M model.



Figure 13. The spatial details of classification results in forest stand species. (a) Original image, (b) the Forest Management Inventory data, (c) M classification, (d) ROPT classification, and (e) MR classification.

However, there are still a lot of misclassifications and noise due to the low separability of tree species. The MR integrated by both can alleviate the problems of both, the integrated classifier MR achieves the best classification result, and its classification results are more consistent with the FMI data. MR by taking advantage of different data and classifiers. Environmental data provide a geographical representation of the ecological niche of tree species, and remote sensing images provide macroscopic and fine scale spectral, textural, and phenological differences of tree species. Orchestrating the interplay of various data and theories with modeling has been identified as a promising approach to obtaining species information (Maréchaux et al. 2021). Studies on species distribution and invasions (Engler et al. 2013; Kattenborn et al. 2019) also confirm that using integrated datasets and combining ecological niche models and machine learning algorithms can benefit improve classification accuracy.

In addition, we collected typical tree species classification cases in recent years and compared the results of the cases with ours. A brief overview is shown in Figure 14, and more detailed information can be found in the literature to read the original content (Bjerreskov, Nord-Larsen, and Fensholt 2021; Boschetti et al. 2007; Engler et al. 2013; Fang et al. 2020; Grabska, Frantz, and Ostapowicz 2020; Grabska et al. 2019; Hościło and Lewandowska 2019; Immitzer, Atzberger, and Koukal 2012; Ke, Quackenbush, and Im 2010; Li, Hu, and Noland 2013; Ma et al. 2021; Plakman et al. 2020; Shirazinejad, Zoej, and Latifi 2022; Sun et al. 2019; Wang and Ren 2021; Wessel, Brandmeier, and Tiede 2018).

Because these cases tested different area sizes, classification methods and tree species, it is not reasonable to directly compare precision. However, we can still compare individual study cases in terms of the following aspects. The species number, reported in listed study cases ranged from 3 to 19 species, the area tested ranged from 5.23 km² to 43,000 km², the resolution of the data used ranged from a maximum of 0.1 m to a minimum of 16 m, and the overall accuracies ranged from 61.3% to 97%. Compared with the listed cases, the number of tree species (19) and the size of the study area (394100 km²) in our study have a definite advantage. The overall accuracy of our study was 72.18%, which is low compared to some study cases. However, we did not use ultra-high resolution or additional input data (e.g. LiDAR). This accuracy is acceptable given the scope and class numbers of the study. This precision is significantly better than the results reported for the 19 tree species study (61.3%) (Fang et al. 2020) and comparable to



Figure 14. Brief review on research cases in tree species classification.

the reported for the 18 tree species cases (73.25%) (Sun et al. 2019), but the area of the two cases is 8.6 km² and 177 km², respectively, which is much smaller than our study. In contrast, the proposed method can achieve accurate and stable mapping of tree species distribution in large areas under different ecological conditions.

5.4. Limitations and future work

The variety of multiple data in terms of spatial resolution, acquisition date and times may affect the classification performance and the spatial distribution patterns of the categories.

In the GEE platform, based on the geographical reference, multiple features extracted from datasets, including S2, SRTM DEM, and WorldClim bioclimatic, can be easily resampled to the same pixel size and stacked together. However, the up-sampling process of low-resolution data sets cannot provide extra information than the original data set. The resolution is usually coarse, and they often do not really describe the spatial distribution of tree species but rather to predict habitat suitability conditions at a given site (Engler et al. 2013). This is also the main reason that the probabilistic output of MaxEnt model was seen as an extra feature set, rather than directly assigned as a component classifier like RF or SVM. As we can see through Figure 13, when the low spatial resolution probability map output from the MaxEnt model is integrated with the machine learning model, it significantly changes the spatial pattern of the output map, and species with fewer samples may be smoothed out by the dominant category. While this improves classification accuracy overall, we believe that the effect of this integration is not always positive, and that the introduction of the low spatial resolution probability maps output by the MaxEnt model may reduce the classification accuracy of certain spatial locations and specific categories, and reduce the distribution area of small sample categories.

Expected for spatial mismatch, the temporal mismatch among multiple data is another factor that may increase uncertainties of classification. To avoid missing some key tree species phenology, we decided to construct 5-day intensive time series data. However, in most areas of the study area, data from both S2A/B satellites still do not allow for 5-day repeat observations, and until March 2017, a single satellite required 10 days, which combined with the cloudy and rainy climate, resulting in fewer pixels available for image acquisition. To construct High-Spatiotemporal-Resolution time series data, we used S2 data for 5 years from 2016 to 2020, which were synthesized according to the time of image acquisition in the year (DOY, Day of Year) to construct a 5-day dense NDVI and REP time series. In this study, we used sample data from the 2016 FMI. This operation led to the mismatch in acquisition time between the sample data and the imagery used. This mismatch may introduce uncertainties.

In this study, we predicted the distribution of 19 tree species in Yunnan by integrating machine learning classifiers and ecological niche models and further experiments and validations can be conducted in other types of forests in the future. We used only four models in this experiment, and more models can be introduced in the future to increase the diversity and complementarity of models. In addition, with growing data availability, more eco-physiological and remote sensing data can be integrated into the models. This process will benefit species classification by a better understanding of the species physiological limits and selecting features more sensitive to species range limits.

6. Conclusions

Detailed and accurate mapping of forest stand species has an important practical need. To achieve effective mapping of the mountain vegetation, it is necessary to fully consider the ecological characteristics of tree species and the displayed remote sensing characteristics and to use multiple data and classification methods for integrated discrimination. In this study, the spatial distribution of 19 stand tree species in Yunnan Province was mapped based on the GEE platform, in collaboration with remote sensing image machine learning algorithm classification and ecological niche modeling, with an overall accuracy of 72.18% assessed through 53,934 validation sample points. Higher classification accuracy was obtained by integrating decision fusion models from both domains compared to supervised classification or ecological niche modeling of environmental data. The study shows that integrating

machine learning algorithms and ecological niche models is effective in regions with high environmental heterogeneity. Spectral, spatial, and temporal extracted from remote sensing data and various environmental variables contribute to tree species classification.

Acknowledgments

We thank anonymous reviewers and editors for their valubale comments and suggestions, which greatly improved the quality of our maniucrpt.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The work was supported by the National Natural Science Foundation of China [32160369]; Major Scientific and Technological Projects of Yunnan Province [202202AD080010]; National Natural Science Foundation of China [31860182]; National Natural Science Foundation of China [32060320]; National Natural Science Foundation of China [41961053].

References

- Abatzoglou, J. T., S. Z. Dobrowski, S. A. Parks, and K. C. Hegewisch. 2018. "TerraClimate, a High-Resolution Global Dataset of Monthly Climate and Climatic Water Balance from 1958–2015." *Scientific Data* 5 (1): 1–12. doi:10. 1038/sdata.2017.191.
- Abdi, A. M. 2020. "Land Cover and Land Use Classification Performance of Machine Learning Algorithms in a Boreal Landscape Using Sentinel-2 Data." *GlScience & Remote Sensing* 57 (1): 1–20. doi:10.1080/15481603.2019.1650447.
- Ahamed, T., L. Tian, Y. Zhang, and K. C. Ting. 2011. "A Review of Remote Sensing Methods for Biomass Feedstock Production." *Biomass & bioenergy* 35 (7): 2455–2469. doi:10. 1016/j.biombioe.2011.02.028.
- Alonso, L., J. Picos, and J. Armesto. 2021. "Forest Cover Mapping and Pinus Species Classification Using Very High-Resolution Satellite Images and Random Forest." *ISPRS Annals of the Photogrammetry, Remote Sensing & Spatial Information Sciences* 3: 203–210. doi:10.5194/isprsannals-V-3-2021-203-2021.
- Bhatt, P., A. Maclean, Y. Dickinson, and C. Kumar. 2022. "Fine-Scale Mapping of Natural Ecological Communities Using Machine Learning Approaches." *Remote Sensing* 14 (3): 563. doi:10.3390/rs14030563.

- Bjerreskov, K. S., T. Nord-Larsen, and R. Fensholt. 2021.
 "Classification of Nemoral Forests with Fusion of Multi-Temporal Sentinel-1 and 2 Data." *Remote Sensing* 13 (5): 950. doi:10.3390/rs13050950.
- Bobrowski, M., J. Weidinger, and U. Schickhoff. 2021. "Is New Always Better? Frontiers in Global Climate Datasets for Modeling Treeline Species in the Himalayas." *Atmosphere* 12: 543. doi:10.3390/atmos12050543.
- Boschetti, M., L. Boschetti, S. Oliveri, L. Casati, and I. Canova. 2007. "Tree Species Mapping with Airborne Hyper-spectral MIVIS Data: The Ticino Park Study Case." *International Journal of Remote Sensing* 28 (6): 1251–1261. doi:10.1080/ 01431160600928542.
- Cano, E., J. -P. Denux, M. Bisquert, L. Hubert-Moy, and V. Chéret. 2017. "Improved Forest-Cover Mapping Based on MODIS Time Series and Landscape Stratification." *International Journal of Remote Sensing* 38 (7): 1865–1888. doi:10.1080/ 01431161.2017.1280635.
- Carlson, T. N., and D. A. Ripley. 1997. "On the Relation Between NDVI, Fractional Vegetation Cover, and Leaf Area Index." *Remote Sensing of Environment* 62 (3): 241–252. doi:10. 1016/S0034-4257(97)00104-1.
- Chandrasekar, K., M. Sesha Sai, P. Roy, and R. S. Dwevedi. 2010. "Land Surface Water Index (LSWI) Response to Rainfall and NDVI Using the MODIS Vegetation Index Product." International Journal of Remote Sensing 31 (15): 3987–4005. doi:10.1080/01431160802575653.
- Chandra, N., G. Singh, S. Lingwal, J. S. Jalal, M. S. Bisht, V. Pal, M. P. S. Bisht, B. Rawat, and L. M. Tiwari. 2021. "Ecological Niche Modeling and Status of Threatened Alpine Medicinal Plant Dactylorhiza Hatagirea D. Don in Western Himalaya." Journal of Sustainable Forestry 41 (10): 1–17. doi:10.1080/ 10549811.2021.1923530.
- Costa, H., P. Benevides, F. D. Moreira, D. Moraes, and M. Caetano. 2022. "Spatially Stratified and Multi-Stage Approach for National Land Cover Mapping Based on Sentinel-2 Data and Expert Knowledge." *Remote Sensing* 14 (8): 1865. doi:10.3390/rs14081865.
- Dash, J., and P. Curran. 2004. "The MERIS Terrestrial Chlorophyll Index." International Journal of Remote Sensing 25 (23): 5403–5413. doi:10.1080/0143116042000274015.
- Datta, A., O. Schweiger, and I. Kühn. 2020. "Origin of Climatic Data Can Determine the Transferability of Species Distribution Models." *NeoBiota* 59: 61–76. doi:10.3897/neo biota.59.36299.
- Duan, B., S. Fang, Y. Gong, Y. Peng, X. Wu, and R. Zhu. 2021. "Remote Estimation of Grain Yield Based on UAV Data in Different Rice Cultivars Under Contrasting Climatic Zone." *Field Crops Research* 267: 108148. doi:10.1016/j.fcr.2021. 108148.
- Engler, R., L. T. Waser, N. E. Zimmermann, M. Schaub, S. Berdos, C. Ginzler, and A. Psomas. 2013. "Combining Ensemble Modeling and Remote Sensing for Mapping Individual Tree Species at High Spatial Resolution." *Forest Ecology and Management* 310: 64–73. doi:10.1016/j.foreco.2013.07.059.
- Fang, F., B. E. McNeil, T. A. Warner, A. E. Maxwell, G. A. Dahle, E. Eutsler, and J. Li. 2020. "Discriminating Tree Species at

Different Taxonomic Levels Using Multi-Temporal WorldView-3 Imagery in Washington DC, USA." *Remote Sensing of Environment* 246: 111811. doi:10.1016/j.rse.2020. 111811.

- Fang, P., X. Zhang, P. Wei, Y. Wang, H. Zhang, F. Liu, and J. Zhao. 2020. "The Classification Performance and Mechanism of Machine Learning Algorithms in Winter Wheat Mapping Using Sentinel-2 10 M Resolution Imagery." Applied Sciences 10 (15): 5075. doi:10.3390/app10155075.
- Fassnacht, F. E., H. Latifi, K. Stereńczak, A. Modzelewska, M. Lefsky, L. T. Waser, C. Straub, and A. Ghosh. 2016. "Review of Studies on Tree Species Classification from Remotely Sensed Data." *Remote Sensing of Environment* 186: 64–87. doi:10.1016/j.rse.2016.08.013.
- Fassnacht, F. E., C. Neumann, M. Forster, H. Buddenbaum, A. Ghosh, A. Clasen, P. K. Joshi, and B. Koch. 2014.
 "Comparison of Feature Reduction Algorithms for Classifying Tree Species with Hyperspectral Data on Three Central European Test Sites." *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 7 (6): 2547–2561. doi:10.1109/jstars.2014.2329390.
- Feng, B., C. Zheng, W. Zhang, L. Wang, and C. Yue. 2020. "Analyzing the Role of Spatial Features When Cooperating Hyperspectral and LiDar Data for the Tree Species Classification in a Subtropical Plantation Forest Area." *Journal of Applied Remote Sensing* 14 (02): 022213. doi:10. 1117/1.JRS.14.022213.
- Gao, B. -C. 1996. "NDWI—A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space." *Remote Sensing of Environment* 58 (3): 257–266. doi:10.1016/ S0034-4257(96)00067-3.
- Gong, P., J. Wang, L. Yu, Y. Zhao, Y. Zhao, L. Liang, Z. Niu, et al. 2013. "Finer Resolution Observation and Monitoring of Global Land Cover: First Mapping Results with Landsat TM and ETM+ Data." International Journal of Remote Sensing 34 (7): 2607–2654. doi:10.1080/01431161.2012.748992.
- Grabska, E., D. Frantz, and K. Ostapowicz. 2020. "Evaluation of Machine Learning Algorithms for Forest Stand Species Mapping Using Sentinel-2 Imagery and Environmental Data in the Polish Carpathians." *Remote Sensing of Environment* 251: 112103. doi:10.1016/j.rse.2020.112103.
- Grabska, E., P. Hostert, D. Pflugmacher, and K. Ostapowicz. 2019. "Forest Stand Species Mapping Using the Sentinel-2 Time Series." *Remote Sensing* 11 (10): 1197. doi:10.3390/rs11101197.
- Guo, Y., Z. Li, E. Chen, X. Zhang, L. Zhao, E. Xu, Y. Hou, and R. Sun. 2020. "An End-To-End Deep Fusion Model for Mapping Forests at Tree Species Levels with High Spatial Resolution Satellite Imagery." *Remote Sensing* 12 (20): 3324. doi:10.3390/rs12203324.
- Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, et al. 2013. "High-Resolution Global Maps of 21st-Century Forest Cover Change." *Science* 342 (6160): 850–853. doi:10.1126/science. 1244693.
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. "Very High Resolution Interpolated Climate Surfaces

for Global Land Areas." International Journal of Climatology: A Journal of the Royal Meteorological Society 25 (15): 1965–1978. doi:10.1002/joc.1276.

- Hościło, A., and A. Lewandowska. 2019. "Mapping Forest Type and Tree Species on a Regional Scale Using Multi-Temporal Sentinel-2 Data." *Remote Sensing* 11 (8): 929. doi:10.3390/ rs11080929.
- Hua, Z., and M. Heil. 2017. "Biogeographical Divergence of the Flora of Yunnan, Southwestern China Initiated by the Uplift of Himalaya and Extrusion of Indochina Block." *PLos One* 7 (9): 7. doi:10.1371/journal.pone.0045601.
- Immitzer, M., C. Atzberger, and T. Koukal. 2012. "Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data." *Remote Sensing* 4 (9): 2661–2693. doi:10.3390/rs4092661.
- Kattenborn, T., J. Lopatin, M. Förster, A. C. Braun, and F. E. Fassnacht. 2019. "UAV Data as Alternative to Field Sampling to Map Woody Invasive Species Based on Combined Sentinel-1 and Sentinel-2 Data." *Remote Sensing* of Environment 227: 61–73. doi:10.1016/j.rse.2019.03.025.
- Ke, Y., L. J. Quackenbush, and J. Im. 2010. "Synergistic Use of QuickBird Multispectral Imagery and LIDAR Data for Object-Based Forest Species Classification." *Remote Sensing* of Environment 114 (6): 1141–1154. doi:10.1016/j.rse.2010. 01.002.
- Kollert, A., M. Bremer, M. Löw, and M. Rutzinger. 2021. "Exploring the Potential of Land Surface Phenology and Seasonal Cloud Free Composites of One Year of Sentinel-2 Imagery for Tree Species Mapping in a Mountainous Region." International Journal of Applied Earth Observation and Geoinformation 94: 102208. doi:10.1016/j.jag.2020. 102208.
- LeCun, Y., Y. Bengio, and G. Hinton. 2015. "Deep Learning." Nature 521 (7553): 436–444. doi:10.1038/nature14539.
- Li, X., and W. D. 1986. "The Plant Geography of Yunnan Province, Southwest China." *I Historical Biogeography: Journal of Biogeography* 13 (5): 367–397. doi:10.2307/ 2844964.
- Li, R., P. Fang, W. Xu, L. Wang, G. Ou, W. Zhang, and X. Huang. 2022. "Classifying Forest Types Over a Mountainous Area in Southwest China with Landsat Data Composites and Multiple Environmental Factors." *Forests* 13 (1): 135. doi:10. 3390/f13010135.
- Li, J., B. Hu, and T. L. Noland. 2013. "Classification of Tree Species Based on Structural Features Derived from High Density LiDar Data." Agricultural and Forest Meteorology 171-172: 104–114. doi:10.1016/j.agrformet.2012.11.012.
- Li, R., N. J. Kraft, J. Yang, and Y. Wang. 2015. "A Phylogenetically Informed Delineation of Floristic Regions Within a Biodiversity Hotspot in Yunnan, China." *Scientific Reports* 5 (1): 1–7. doi:10.1016/S0034-4257(96)00178-2.
- Li, W., and S. Pei. 1991. "A Study of Flora of Yunnan." *Guihaia* 11: 293–303.
- Liu, H., P. Gong, J. Wang, X. Wang, G. Ning, and B. Xu. 2021. "Production of Global Daily Seamless Data Cubes and Quantification of Global Land Cover Change from 1985 to

2020-iMap World 1.0." *Remote Sensing of Environment* 258: 112364. doi:10.1016/j.rse.2021.112364.

- Liu, F., J. Hu, F. Yang, and X. Li. 2021. "Heterogeneity-Diversity Relationships in Natural Areas of Yunnan, China." *Chinese Geographical Science* 31 (3): 506–521. doi:10.1007/s11769-021-1207-7.
- Liu, Z., and H. Peng. 2016. "Notes on the Key Role of Stenochoric Endemic Plants in the Floristic Regionalization of Yunnan." *Plant Diversity* 38 (6): 289–294. doi:10.1016/ S0034-4257(96)00067-3.
- Long, T., Z. Zhang, G. He, W. Jiao, C. Tang, B. Wu, X. Zhang, G. Wang, and R. Yin. 2019. "30 M Resolution Global Annual Burned Area Mapping Based on Landsat Images and Google Earth Engine." *Remote Sensing* 11 (5): 489. doi:10.3390/rs11050489.
- Main, R., M. A. Cho, R. Mathieu, M. M. O'Kennedy, A. Ramoelo, and S. Koch. 2011. "An Investigation into Robust Spectral Indices for Leaf Chlorophyll Estimation." *Isprs Journal of Photogrammetry & Remote Sensing* 66 (6): 751–761. doi:10. 1016/j.isprsjprs.2011.08.001.
- Ma, M., J. Liu, M. Liu, J. Zeng, and Y. Li. 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. doi:10.3390/f12121736.
- Maréchaux, I., F. Langerwisch, A. Huth, H. Bugmann, X. Morin, C. P. O. Reyer, R. Seidl, et al. 2021. "Tackling Unresolved Questions in Forest Ecology: The Past and Future Role of Simulation Models." *Ecology & Evolution* 11 (9): 3746–3770. doi:10.1002/ece3.7391.
- Michaowska, M., and J. Rapiński. 2021. "A Review of Tree Species Classification Based on Airborne LiDar Data and Applied Classifiers." *Remote Sensing* 13 (3): 353. doi:10. 3390/rs13030353.
- Moraes, D., P. Benevides, H. Costa, F. D. Moreira, and M Caetano. 2021. "Assessment of the Introduction of Spatial Stratification and Manual Training in Automatic Supervised Image Classification." In *Earth Resources and Environmental Remote Sensing/GIS Applications XII*, LNEE, 291–298. Singapore: SPIEdoi:10.1007/978-981-16-0289-4_39.
- Mouta, N., R. Silva, S. Pais, J. M. Alonso, J. F. Gonçalves, J. Honrado, and J. R. Vicente. 2021. "The Best of Two Worlds'—Combining Classifier Fusion and Ecological Models to Map and Explain Landscape Invasion by an Alien Shrub." *Remote Sensing* 13 (16): 3287. doi:10.3390/rs13163287.
- Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. "Biodiversity Hotspots for Conservation Priorities." *Nature* 403 (6772): 853–858. doi:10.1038/35002501.
- Oreopoulos, L., M. J. Wilson, and T. Várnai. 2011. "Implementation on Landsat Data of a Simple Cloud-Mask Algorithm Developed for MODIS Land Bands." *IEEE Geoscience & Remote Sensing Letters* 8 (4): 597–601. doi:10. 1109/LGRS.2010.2095409.
- Ørka, H. O., M. Dalponte, T. Gobakken, E. Næsset, and L. T. Ene. 2013. "Characterizing Forest Species Composition Using Multiple Remote Sensing Data Sources and Inventory Approaches." Scandinavian Journal of Forest Research 28 (7): 677–688. doi:10.1080/02827581.2013.793386.

- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. "Maximum Entropy Modeling of Species Geographic Distributions." *Ecological Modelling* 190 (3–4): 231–259. doi:10.1016/j.ecolmodel.2005.03.026.
- Plakman, V., T. Janssen, N. Brouwer, and S. Veraverbeke. 2020. "Mapping Species at an Individual-Tree Scale in a Temperate Forest, Using Sentinel-2 Images, Airborne Laser Scanning Data, and Random Forest Classification." *Remote Sensing* 12 (22): 3710. doi:10.3390/rs12223710.
- Pshegusov, R., F. Tembotova, V. Chadaeva, Y. Sablirova, M. Mollaeva, and A. Akhomgotov. 2022. "Ecological Niche Modeling of the Main Forest-Forming Species in the Caucasus." *Forest Ecosystems* 9: 100019. doi:10.1016/j.fecs. 2022.100019.
- Richardson, A. D., S. P. Duigan, and G. P. Berlyn. 2002. "An Evaluation of Noninvasive Methods to Estimate Foliar Chlorophyll Content." *The New Phytologist* 153 (1): 185–194. doi:10.3390/rs12203324.
- Rozenstein, O., N. Haymann, G. Kaplan, and J. Tanny. 2019.
 "Validation of the Cotton Crop Coefficient Estimation Model Based on Sentinel-2 Imagery and Eddy Covariance Measurements." *Agricultural Water Management* 223: 105715. doi:10.1016/j.agwat.2019.105715.
- Schlerf, M., C. Atzberger, and J. Hill. 2005. "Remote Sensing of Forest Biophysical Variables Using HyMap Imaging Spectrometer Data." *Remote Sensing of Environment* 95 (2): 177–194. doi:10.1016/j.rse.2004.12.016.
- Shirazinejad, G., M. Zoej, and H. Latifi. 2022. "Applying Multidate Sentinel-2 Data for Forest-Type Classification in Complex Broadleaf Forest Stands." Forestry: An International Journal of Forest Research 95 (3): 363–379. doi:10.1093/for estry/cpac001.
- Sinclair, S. J., M. D. White, and G. R. Newell. 2010. "How Useful are Species Distribution Models for Managing Biodiversity Under Future Climates?" *Ecology and Society* 15. doi:10.5751/ES-03089-150108.
- Sun, Y., J. Huang, Z. Ao, D. Lao, and Q. Xin. 2019. "Deep Learning Approaches for the Mapping of Tree Species Diversity in a Tropical Wetland Using Airborne LiDar and High-Spatial-Resolution Remote Sensing Images." *Forests* 10 (11): 1047. doi:10.3390/f10111047.
- Tamiminia, H., B. Salehi, M. Mahdianpari, L. Quackenbush, S. Adeli, and B. Brisco. 2020. "Google Earth Engine for Geo-Big Data Applications: A Meta-Analysis and Systematic Review." *Isprs Journal of Photogrammetry & Remote Sensing* 164: 152–170. doi:10.1016/j.isprsjprs.2020.04.001.
- Wang, X., and H. Ren. 2021. "DBMF: A Novel Method for Tree Species Fusion Classification Based on Multi-Source Images." *Forests* 13 (1): 33. doi:10.3390/f13010033.
- Wan, H., Y. Tang, L. Jing, H. Li, F. Qiu, and W. Wu. 2021. "Tree Species Classification of Forest Stands Using Multisource Remote Sensing Data." *Remote Sensing* 13 (1): 144. doi:10. 3390/rs13010144.
- Waser, L. T., C. Ginzler, M. Kuechler, E. Baltsavias, and L. Hurni. 2011. "Semi-Automatic Classification of Tree Species in Different Forest Ecosystems by Spectral and Geometric Variables Derived from Airborne Digital Sensor (ADS40)

24 🔶 P. FANG ET AL.

and RC30 Data." *Remote Sensing of Environment* 115 (1): 76–85. doi:10.1016/j.rse.2010.08.006.

- Welle, T., L. Aschenbrenner, K. Kuonath, S. Kirmaier, and J. Franke. 2022. "Mapping Dominant Tree Species of German Forests." *Remote Sensing* 14 (14): 3330. doi:10.3390/rs14143330.
- Wessel, M., M. Brandmeier, and D. Tiede. 2018. "Evaluation of Different Machine Learning Algorithms for Scalable Classification of Tree Types and Tree Species Based on Sentinel-2 Data." *Remote Sensing* 10 (9): 1419. doi:10.3390/ rs10091419.
- Wiens, J. J., and C. H. Graham. 2005. "Niche Conservatism: Integrating Evolution, Ecology, and Conservation Biology." *Annual Review of Ecology, Evolution, and Systematics* 36 (1): 519–539. doi:10.1146/annurev.ecolsys.36.102803.095431.
- Wu, C., Z. Niu, Q. Tang, and W. Huang. 2008. "Estimating Chlorophyll Content from Hyperspectral Vegetation Indices: Modeling and Validation." *Agricultural and Forest Meteorology* 148 (8–9): 1230–1241. doi:10.1016/j.agrformet.2008.03.005.
- Yang, Y., K. Tian, J. Hao, S. Pei, and Y. Yang. 2004. "Biodiversity and Biodiversity Conservation in Yunnan, China." *Biodiversity* & Conservation 13 (4): 813–826. doi:10.1023/B:BIOC. 0000011728.46362.3c.
- You, N., J. Dong, J. Huang, G. Du, G. Zhang, Y. He, T. Yang, Y. Di, and X. Xiao. 2021. "The 10-M Crop Type Maps in Northeast China During 2017–2019." *Scientific Data* 8 (1): 1–11. doi:10. 1038/s41597-021-00827-9.
- Yu, X., D. Lu, X. Jiang, G. Li, Y. Chen, D. Li, and E. Chen. 2020. "Examining the Roles of Spectral, Spatial, and Topographic Features in Improving Land-Cover and Forest Classifications in a Subtropical Region." *Remote Sensing* 12 (18): 2907. doi:10.3390/rs12182907.

- Zeb, S. A., S. M. Khan, and Z. Ahmad. 2021. "Phytogeographic Elements and Vegetation Along the River Panjkora -Classification and Ordination Studies from the Hindu Kush Mountains Range." *The Botanical Review* 87 (4): 1–25. doi:10. 1007/s12229-021-09247-1.
- Zhang, Y., I. O. Odeh, and C. Han. 2009. "Bi-Temporal Characterization of Land Surface Temperature in Relation to Impervious Surface Area, NDVI and NDBI, Using a Sub-Pixel Image Analysis." *International Journal of Applied Earth Observation and Geoinformation* 11 (4): 256–264. doi:10.1016/j.jag.2009.03.001.
- Zhang, B., L. Zhao, and X. Zhang. 2020. "Three-Dimensional Convolutional Neural Network Model for Tree Species Classification Using Airborne Hyperspectral Images." *Remote Sensing of Environment* 247: 111938. doi:10.1016/j. rse.2020.111938.
- Zheng, G., A. Bao, X. Li, L. Jiang, C. Chang, T. Chen, and Z. Gao. 2019. "The Potential of Multispectral Vegetation Indices Feature Space for Quantitatively Estimating the Photosynthetic, Non-Photosynthetic Vegetation and Bare Soil Fractions in Northern China." *Photogrammetric Engineering & Remote Sensing* 85 (1): 65–76. doi:10.14358/ PERS.85.1.65.
- Zhong, L., L. Hu, and H. Zhou. 2019. "Deep Learning Based Multi-Temporal Crop Classification." *Remote Sensing of Environment* 221: 430–443. doi:10.1016/j.rse.2018.11.032.
- Zhou, Y., Z. Zhang, B. Zhu, X. Cheng, L. Yang, M. Gao, and R. Kong. 2021. "MaxEnt Modeling Based on CMIP6 Models to Project Potential Suitable Zones for Cunninghamia Lanceolata in China." *Forests* 12 (6): 752. doi:10.3390/ f12060752.