Forest Tree Species Classification Based on Deep Ensemble Learning by Fusing High-Resolution, Multitemporal, and Hyperspectral Multisource Remote Sensing Data

Dengli Yu, Lilin Tu, Ziqing Wei, Fuyao Zhu, Chengjun Yu, Denghong Wang, Jiayi Li, *Senior Member, IEEE*, and Xin Huang, *Fellow, IEEE*

***Abstract*—Forest tree species classification has great significance for sustainable development of forest resource. Multisource remote sensing data provide abundant temporal, spatial, and spectral information for tree species classification. However, there lacks tree species classification methods, which comprehensively capture and fuse spatio–temporal–spectral information. Therefore, a tree species classification method based on deep ensemble learning of multisource spatio–temporal–spectral remote sensing data is proposed. First, multitemporal, high-resolution, and hyperspectral data are utilized for training temporal, spatial, and spectral deep networks. Furtherly, deep ensemble learning is developed for the fusion of spatio–temporal–spectral network outputs, where weighted fusion is implemented via dynamic weight optimization based on the spatio–temporal–spatial features. Experimental results indicate that the importance of temporal features is higher than that of spatial information, and spectral networks perform best among all network structures. After the spatio–temporal–spectral ensemble learning, the performance of tree species classification is further improved, and the overall accuracy (OA) of the proposed method reaches above 90%. The proposed algorithm realizes precise and fine-scale tree species classification and provides technique support for the monitoring and conservation of forest resource.**

***Index Terms*—Deep learning, ensemble learning, multisource remote sensing data, spatio–temporal–spectral feature, tree species classification.**

I. Introduction

As THE main body of terrestrial ecosystem, forests play a vital role in biodiversity and ecological balance conservation. Tree species composition is the core content of forest resource monitoring and the key indicator of forest biodiversity measurement and has great significance in a wide range of applications [1], [2], [3].

With the rapid development of remote sensing technology, multisource remote sensing data have been widely used in tree species classification. Among the various remote sensing data sources, multispectral and hyperspectral sensors can capture the spectral response in specific wavelength ranges, which provides spectral information for tree species mapping [3]. High-resolution images involve abundant spatial details, such as structures and textures, leading to fine-scale tree species identification [4]. Multitemporal data can obtain temporal features related to phenological differences of tree species, increasing the stability and separability of tree species classification [5].

For tree species classification methods, deep learning algorithms have drawn unprecedented attention by virtue of its capacity of extracting high-level semantic representations. Cao and Zhang [4] proposed a Res-UNet semantic segmentation network combining with conditional random fields (CRFs) postprocessing for tree species classification based on airborne high-resolution images. Huang et al. [6] designed Transformer4SITS network to extract spectral and temporal features of multitemporal Sentinel-2 images. Zhang et al. [7] proposed the 3-D and 1-D convolutional neural network (CNN) to capture spectral–spatial features of tree species using hyperspectral data. In addition, some researches fused multisource remote sensing data to compensate for the deficiency of single-source data. For example, Chen et al. [8] first utilized high-resolution data to delineate tree canopies and then fed multitemporal images into 3DLSTM network to generate tree species classification results. Wang and Ren [9] proposed a dual-branch network for tree species classification, where spectral and spatial information from hyperspectral and multispectral data is extracted and interacted.

The temporal, spectral, and spatial information play important roles in tree species classification. However, few researches simultaneously capture and fuse the spatio–temporal–spectral features of tree species. In view of this background, a tree species classification method based on deep ensemble learning using spatio–temporal–spectral remote sensing data is proposed in this study, and the performance of tree species classification is improved via combining spatio–temporal–spectral information. Specifically, multitemporal Sentinel-2 data, high-resolution Gaofen-2 data, and Orbita Hyperspectral Satellite (OHS) data are utilized to train temporal, spatial, and spectral networks.

Received 23 September 2025; revised 26 October 2025; accepted 15 November 2025. Date of publication 19 November 2025; date of current version 1 December 2025. This work was supported in part by the National Key Research and Development Program of China under Grant 2024YFF1306102 and in part by the National Natural Science Foundation of China under Grant 42471391 and Grant 42271328. *(Corresponding authors: Lilin Tu; Xin Huang.)*

Dengli Yu, Ziqing Wei, Chengjun Yu, and Denghong Wang are with the Management Bureau of Maolan National Nature Reserve, Libo, Guizhou 558400, China.

Lilin Tu, Fuyao Zhu, Jiayi Li, and Xin Huang are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: tulilin0312@163.com; xhuang@whu.edu.cn).

Digital Object Identifier 10.1109/LGRS.2025.3634553

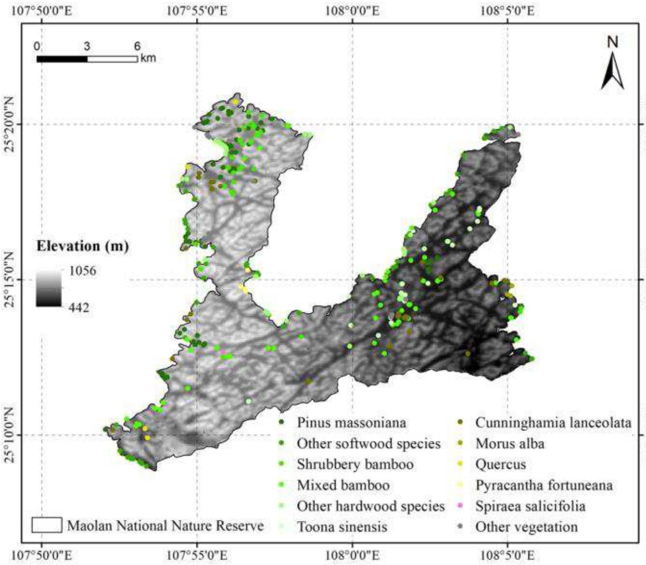


Fig. 1. Overview and label distribution of Maolan National Nature Reserve.

Through the proposed deep ensemble learning method, the outputs of spatio–temporal–spectral networks are fused via dynamic weight optimization mechanism, and tree species classification result after spatio–temporal–spectral fusion is achieved.

II. Study Area and Data

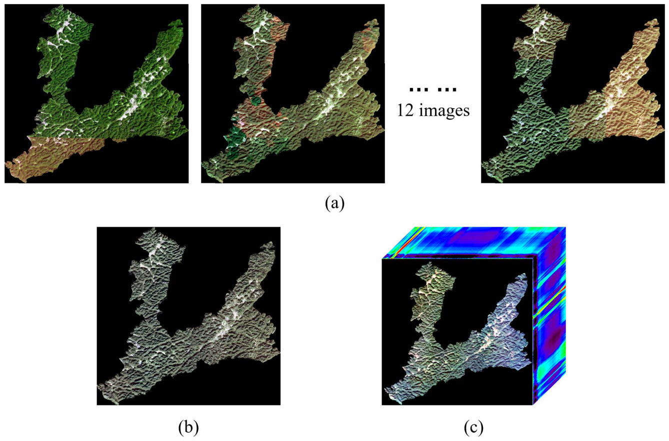
A. Study Area

Maolan National Nature Reserve is selected as the study area, which is located in Libo county of Guizhou province in China, as shown in Fig. 1. The distinctive landforms, mainly consisting of karst peaks, depressions, and basins with the elevation of 442–1056 m, give birth to rich and diverse forest landscapes in this study area. Sample data used in this study derive from the field surveys conducted around 2020, including 12 tree species classes: Pinus massoniana, Other softwood species, Shrubbery bamboo, Mixed bamboo, Other hardwood species, Toona sinensis, Cunninghamia lanceolata, Morus alba, Quercus, Pyracantha fortuneana, Spiraea salicifolia, and Other vegetation. The locations of sample data are highlighted in colored points in Fig. 1 (totally 342 points).

B. Multisource Remote Sensing Data

Multisource data used in this letter include Sentinel-2 (temporal), Gaofen-2 (spatial), and OHS data (spectral), which are visualized in Fig. 2 and introduced as follows.

1) *Sentinel-2 Data:* Sentinel-2 data have 13 spectral bands with the wavelength ranging from 444 to 2202 nm and the spatial resolution ranging from 10 to 60 m. The revisit period is 5 days. We obtain the Sentinel-2 data during 2019 and 2021 and composite 12 seasonal images. Ten of the 13 bands with the resolution of 10 and 20 m are selected and resampled to the resolution of 10 m for network training.
2) *Gaofen-2 Data:* Gaofen-2 satellite carries multispectral and panchromatic sensor with the spatial resolution of 4 and 1 m, respectively. The Gaofen-2 image acquired in November 12, 2020 is utilized. After pan-sharpening, high-resolution image consisting four spectral bands, i.e., RGB and near-infrared (NIR), with the resolution of 1 m is obtained.
3) *OHS Data:* OHS data have 32 spectral bands with the wavelength ranging from 400 to 1000 nm, and the spatial resolution is 10 m. The OHS image acquired by OHS-3B satellite in November 8 of 2020 is used in this study.



(a)

(b)

(c)

Fig. 2. Spatio–temporal–spectral multisource remote sensing data. (a) Sentinel-2 images (temporal). (b) Gaofen-2 image (spatial). (c) OHS hyperspectral image (spectral).

III. Methodology

Aiming at precise tree species classification, in this study, multitemporal, high-resolution, and hyperspectral data are utilized to extract temporal, spatial, and spectral information, respectively. Furtherly, deep ensemble learning algorithm is proposed to fuse the outputs of spatio–temporal–spatial networks. The overall flowchart is illustrated in Fig. 3.

A. Deep Network Structures

Among various deep network structures, CNN, which captures local context using convolution, and transformer, which models long-range dependencies using self-attention, are the network components commonly used for feature extraction. Therefore, the temporal, spatial, and spectral networks based on CNN and transformer are selected for experiments.

1) Temporal Networks: TSViT and ConvLSTM are selected for mining temporal information of tree species. TSViT [10] is a transformer network for processing multitemporal remote sensing images. In TSViT, learnable position encodings for each time step and image patch are introduced, and transformer blocks with temporal-then-spatial self-attention are designed for precise temporal and spatial dependency modeling. In addition, multiple learnable class tokens are utilized to further increase the discriminative ability. ConvLSTM [11] combines the advantages of both CNN and long-short term memory (LSTM). For each ConvLSTM layer in this network, spatial context information is extracted through convolution, and gating mechanism in LSTM is utilized to capture the changes of spatial features over time.

2) Spatial Networks: Two representative networks, SwinT and ConvNeXT, are chosen to extract spatial details of tree species. SwinT [12] is a hierarchical transformer structure,

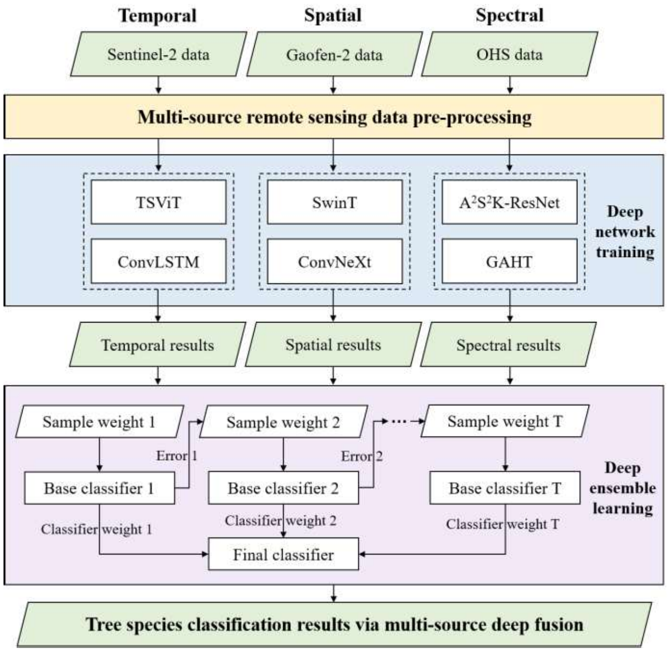


Fig. 3. Flowchart of the proposed deep ensemble learning algorithm. Temporal, spatial, and spectral networks are trained using the corresponding remote sensing data, and the outputs of spatio–temporal–spectral networks are fused by deep ensemble learning based on dynamic weight optimization mechanism to generate final tree species classification results.

which employs window-based and shifted window-based self-attention to extract spatial features both locally and globally. Between each two stages, neighboring image patches are merged to form multiscale representations. ConvNeXT [13] is a CNN designed based on standard ResNet [14] structure with a series of modifications adopted to improve the performance, such as using larger convolutional kernels, designing inverted bottleneck, reducing activations and norms, and so on.

In this study, the encoder–decoder architecture U-Net [15] is utilized to implement tree species classification. Specifically, SwinT and ConvNeXT are incorporated into U-Net as the encoder, and feature maps are gradually upsampled in the decoder to obtain fine-scale classification maps.

3) Hyperspectral Networks: For spectral features extraction, A^{2}S^{2}K-ResNet and GAHT are selected. A^{2}S^{2}K-ResNet [16] is an improved ResNet structure for discriminative spectral–spatial feature extraction. Adaptive spectral–spatial kernel attention is designed to automatically adjust the receptive field, and efficient feature calibration module is introduced to learn cross-channel dependencies. GAHT [17] is a hierarchical transformer network for the extraction of spectral–spatial information. It employs grouped pixel embedding for local spectral feature mining and utilizes transformer blocks to capture long-range dependencies, providing abundant spectral–spatial representations for tree species classification.

B. Spatio–temporal–Spectral Deep Ensemble Learning Algorithm

The main procedure of the proposed deep ensemble learning algorithm is introduced as follows.

TABLE I

HYPERPARAMETER SETTINGS FOR DEEP NETWORK TRAINING

Deep Network	Epochs	Learning Rate	Batch Size	Optimizer
TSViT	150	5e-4	128	Adam
ConvLSTM	400	5e-4	128	Adam
SwinT-UNet	300	1e-4	32	AdamW
ConvNeXt-UNet	300	1e-4	32	AdamW
A^{2}S^{2}K-ResNet	100	5e-4	128	Adam
GAHT	100	5e-4	128	Adam

1) *Step 1 (Multisource Data Preprocessing):* For Sentinel-2 and Gaofen-2 data, images are cropped into the patches of 24 × 24 and 256 × 256 pixels, respectively. For OHS data, patches are generated for each pixel using a window with fixed size.
2) *Step 2 (Deep Network Training):* With the preprocessed data as an input, the temporal, spatial, and spectral networks introduced in Section III-A are trained, and then, classification results of each network are obtained.
3) *Step 3 (Sample Weight Initialization):* The proposed deep ensemble learning method trains several base classifiers using the output features of the spatio–temporal–spectral networks. In this process, each sample is assigned with a weight representing its importance, which is dynamically adjusted afterward. For N training samples, the initial weights are the same for each sample, i.e., $1/N$

$$D_1 = \left(\frac{1}{N}, \frac{1}{N}, \ldots, \frac{1}{N}\right). \tag{1}$$

4) *Step 4 (Iterative Training of Base Classifiers):* For the tth round of the iteration, according to the weight of each sample, a base classifier $H_t(x)$ is trained and error rate e_t is calculated

$$e_t = \sum_{i=1}^{N} D_t(i) \cdot I(H_t(x_i) \neq y_i) \tag{2}$$

where $D_t(i)$ is the sample weight for current iteration and $I(H_t(x_i) \neq y_i)$ is 1 for samples incorrectly classified and 0 for those correctly classified.

According to the error rate e_t, the weight of the base classifier trained in this iteration α_t is calculated

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1-e_t}{e_t}\right). \tag{3}$$

Then, the weights of the training samples are updated, which are used for the next iteration

$$D_{t+1}(i) = \frac{D_t(i) \cdot \exp(-\alpha_t \cdot I(H_t(x_i) = y_i))}{\sum_{j=1}^{N} D_t(j) \cdot \exp\left(-\alpha_t \cdot I\left(H_t\left(x_j\right) = y_j\right)\right)}. \tag{4}$$

The aforementioned process is repeated until the total number of iterations T is reached.
5) *Step 5 (Ensemble of Base Classifiers):* The base classifiers are fused according to classifier weights to obtain the final classifier $H_{\text{final}}(x)$ and tree species classification result

$$H_{\text{final}}(x) = \sum_{t=1}^{T} \alpha_t H_t(x). \tag{5}$$

The core idea of the proposed method is fusing the classification results of multiple networks to generate a strong classifier, where samples incorrectly classified are given more attention in subsequent training, in order to increase the

TABLE II

ACCURACIES OF TREE SPECIES CLASSIFICATION FOR DIFFERENT TYPES OF NETWORKS WITH THE BEST AND SECOND-BEST VALUE BOLDED AND UNDERLINED

Tree species	Temporal networks				Spatial networks				Spectral networks			
	TSViT		ConvLSTM		SwinT		ConvNeXt		A^2S^2K-ResNet		GAHT	
	P	R	P	R	P	R	P	R	P	R	P	R
Pinus massoniana	0.6905	0.7840	0.7369	0.7231	0.6268	0.7454	0.5789	0.5555	**0.7885**	**0.8121**	0.7539	0.8029
Other softwood species	0.6818	0.9438	0.5689	0.8303	0.6363	0.7382	0.6017	0.5673	0.6618	0.8736	**0.7835**	**0.9643**
Shrubbery bamboo	0.8310	0.7664	0.8292	0.7683	0.8830	0.6919	0.6775	0.7152	0.9280	**0.9141**	**0.9438**	0.7545
Mixed bamboo	**0.8781**	0.8433	0.8645	0.8578	0.7717	0.6906	0.5734	0.6378	0.7886	0.7461	0.7356	**0.9629**
Other hardwood species	0.6569	**0.7621**	0.3921	0.4320	0.7804	0.3991	0.2497	0.3600	0.4833	0.3240	**0.9821**	0.3073
Toona sinensis	0.6223	0.6000	0.5075	0.5179	0.7341	0.6487	0.5013	0.5140	**0.7462**	**0.8122**	0.7358	0.7845
Cunninghamia lanceolata	**0.9578**	0.6913	0.7846	0.4435	0.6816	0.5025	0.2582	0.1001	0.5569	0.7845	0.8098	**0.9171**
Morus alba	**1.0000**	**1.0000**	0.0000	0.0000	0.3475	0.9164	0.6647	0.3050	**1.0000**	**1.0000**	**1.0000**	0.8077
Quercus	**1.0000**	0.4761	0.8026	0.2536	0.8333	0.7238	0.2120	0.2904	**1.0000**	0.5977	0.9559	**0.7580**
Pyracantha fortuneana	**0.9928**	**1.0000**	0.6055	0.4783	0.7340	0.4325	0.7439	0.3865	0.9766	**1.0000**	0.9158	0.6960
Spiraea salicifolia	0.8869	**1.0000**	0.0000	0.0000	0.4458	0.8611	0.6059	0.2128	0.9455	0.8062	**1.0000**	0.9070
Other vegetation	0.7566	0.5542	0.3713	0.7301	0.5170	**0.7568**	0.1771	0.1777	**0.7649**	0.5874	0.5429	0.5989
OA	0.7748		0.6951		0.6992		0.5465		0.8041		**0.8044**	
Kappa	0.7265		0.6317		0.6380		0.4482		0.7633		**0.7655**	

TABLE III

ACCURACIES OF TREE SPECIES CLASSIFICATION BASED ON DEEP ENSEMBLE LEARNING OF MULTIPLE NETWORKS AND MULTISOURCE REMOTE SENSING DATA

Tree species	Temporal		Spatial		Spectral		Spatial+Spectral		Spatial+Spectral+Temporal	
	P	R	P	R	P	R	P	R	P	R
Pinus massoniana	0.7503	0.7840	0.6375	0.7566	**0.8669**	0.8514	0.7323	0.8575	0.8398	**0.8952**
Other softwood species	0.7161	0.9011	0.6937	0.7646	0.7745	0.8915	0.8119	0.9118	**0.9054**	**0.9844**
Shrubbery bamboo	0.8672	0.8297	0.8220	0.7343	0.9144	**0.9793**	0.9458	0.7633	**0.9508**	0.9065
Mixed bamboo	**0.9031**	0.8472	0.8155	0.7410	0.7140	**0.9928**	0.8315	0.9394	0.8972	0.9770
Other hardwood species	0.8257	0.4369	0.6852	0.3620	0.9825	0.3128	0.9838	0.4096	**0.9934**	**0.5444**
Toona sinensis	0.6179	0.6718	0.7708	0.7030	0.8896	0.7569	**0.9354**	0.7604	0.9276	**0.8537**
Cunninghamia lanceolata	0.7126	0.7652	0.8046	0.4928	0.5159	0.8066	0.9879	0.9165	**0.9972**	**0.9319**
Morus alba	0.4286	0.8571	0.5885	0.8742	**1.0000**	0.7308	0.9357	0.8933	0.9991	**0.9514**
Quercus	0.9863	0.5988	0.8465	0.7404	**1.0000**	0.5452	0.9912	0.8976	0.9999	**0.9423**
Pyracantha fortuneana	0.9928	**1.0000**	0.8259	0.3832	**1.0000**	0.4160	0.7604	0.8547	0.8312	0.9832
Spiraea salicifolia	**0.9739**	**1.0000**	0.5475	0.8230	0.0000	0.0000	0.6342	0.8828	0.9465	0.9968
Other vegetation	0.7182	**0.8289**	0.5479	0.7432	0.8700	0.4986	0.7297	0.7456	**0.9194**	0.7784
OA	0.8051		0.7212		0.8276		0.8321		**0.9094**	
Kappa	0.7642		0.6618		0.7896		0.7972		**0.8900**	

performance and stability. In this way, the advantages of multisource remote sensing data and networks are combined, and accurate tree species classification map is generated.

IV. RESULTS AND DISCUSSION

A. *Experimental Setting*

The hyperparameter settings for training each deep network are shown in Table I. For deep ensemble learning, classification and regression tree (CART) [18] is taken as the base classifier and the number of iterations T is set to 50.

For experiments, the tree species classification results of temporal, spatial, and spectral networks and data are first compared. Then, tree species classification result using the proposed deep ensemble learning method is analyzed. Samples are divided into training, validation, and testing sets with the ratio of 6:2:2, and the same sample division is adopted for all experiments for fairness comparison. Overall accuracy (OA), kappa coefficient, precision (P), and recall (R) are used to evaluate the performance of tree species classification.

B. *Comparison of Different Types of Networks and Data*

The accuracies of different temporal, spatial, and spectral networks in tree species classification are shown in Table II. For temporal networks, TSViT achieves high precision and recall in most of the tree species classes, while ConvLSTM performs poorly in some tree species. Specifically, Pinus massoniana, Other softwood species, Shrubbery bamboo, Mixed bamboo, and Toona sinensis reach high accuracies for both networks. Morus alba and Spiraea salicifolia are well discriminated in TSViT; however, ConvLSTM fails to recognize them. Spatial networks achieve lower accuracies than temporal networks, implying that spatial details are less important than phenological features for identifying tree species. For overall performance, the accuracy of SwinT is higher than that of ConvNeXT. For specific tree species classes, Pinus massoniana, Other softwood species, Mixed bamboo, Toona sinensis, and Quercus reach highest precision and recall for SwinT, while ConvNeXT obtains highest precision in Morus alba, Pyracantha fortuneana, and Spiraea salicifolia. For spectral networks, the overall accuracies of both A^2S^2K-ResNet and GAHT are higher than temporal and spatial networks, indicating that spectral information is better than temporal and spatial features in tree species classification. Specifically, A^2S^2K-ResNet achieves best precision and recall in Pinus massoniana, Toona sinensis, and Morus alba, while GAHT performs well in identifying Other softwood species, Cunninghamia lanceolate, and Spiraea salicifolia.

C. *Results of Deep Ensemble Learning*

In order to generate high-resolution tree species classification map, tree species classification results of spectral and temporal networks are fused with those of spatial networks using the proposed deep ensemble learning method. It can

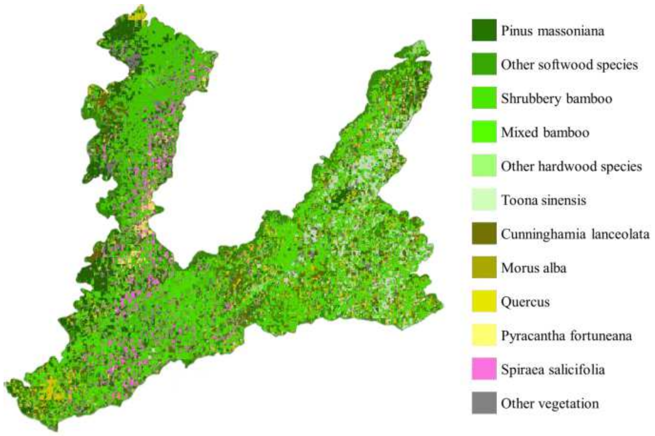


Fig. 4. Tree species classification result of Maolan National Nature Reserve based on deep ensemble learning of spatio–temporal–spectral remote sensing data.

be seen from Table III that the performance of tree species classification is significantly boosted after the deep ensemble learning. On the one hand, compared with the accuracies in Table II, deep ensemble learning of multiple networks achieves better performance than single network for each network type. On the other hand, the overall accuracies for deep ensemble learning of multisource data are higher than the accuracies obtained using single-source data. Specifically, compared with the accuracies of spatial networks, after introducing spectral networks, the increment of precision and recall exceeds 10% for Other softwood species, Cunninghamia lanceolate, and Quercus. When temporal information is further involved, the OA of tree species classification finally reaches 0.9094. As shown in Fig. 4, the fine-scale tree species classification map, which can reflect the actual tree species distribution, is achieved.

V. Conclusion

In this study, a deep ensemble learning method is proposed for effective tree species classification, where temporal, spatial, and spectral information are fused using multisource remote sensing data. First, temporal networks TSViT and ConvLSTM are trained using multitemporal Sentinel-2 data, spatial networks SwinT and ConvNeXT are trained using high-resolution Gaofen-2 data, and spectral networks A^2S^2K-ResNet and GAHT are trained using OHS data. On this basis, the deep ensemble learning method is designed to fuse the outputs of spatio–temporal–spectral networks via dynamic weight optimization mechanism. Experiments are conducted in the study area of Maolan National Nature Reserve. According to the classification results, spectral and temporal information are superior to spatial information in tree species identification. Compared with using single-source data, the accuracy of tree species classification is significantly improved when adopting the proposed spatio–temporal–spectral deep ensemble learning. The proposed method realizes precise and fine-scale tree species classification and provides technique support for forest resource monitoring and conservation. For future work, we plan to introduce more types of remote sensing data source, e.g., LiDAR and SAR, to further increase the accuracy and reliability of tree species classification.

Acknowledgment

The authors would like to thank the editors and anonymous reviewers for providing insightful suggestions to this letter.

References

[1] A. Fichtner, W. Härdtle, Y. Li, H. Bruelheide, M. Kunz, and G. von Oheimb, "From competition to facilitation: How tree species respond to neighbourhood diversity," *Ecol. Lett.*, vol. 20, no. 7, pp. 892–900, Jul. 2017, doi: 10.1111/ele.12786.

[2] M. Jonsson, J. Bengtsson, L. Gamfeldt, J. Moen, and T. Snäll, "Levels of forest ecosystem services depend on specific mixtures of commercial tree species," *Nature Plants*, vol. 5, no. 2, pp. 141–147, Jan. 2019, doi: 10.1038/s41477-018-0346-z.

[3] F. E. Fassnacht et al., "Review of studies on tree species classification from remotely sensed data," *Remote Sens. Environ.*, vol. 186, pp. 64–87, Dec. 2016, doi: 10.1016/j.rse.2016.08.013.

[4] K. Cao and X. Zhang, "An improved Res-UNet model for tree species classification using airborne high-resolution images," *Remote Sens.*, vol. 12, no. 7, p. 1128, Apr. 2020, doi: 10.3390/rs12071128.

[5] R. Vanguri, G. Laneve, and A. Hošciło, "Mapping forest tree species and its biodiversity using EnMAP hyperspectral data along with Sentinel-2 temporal data: An approach of tree species classification and diversity indices," *Ecol. Indicators*, vol. 167, Oct. 2024, Art. no. 112671, doi: 10.1016/j.ecolind.2024.112671.

[6] Z. Huang et al., "A spectral-temporal constrained deep learning method for tree species mapping of plantation forests using time series Sentinel-2 imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 204, pp. 397–420, Oct. 2023, doi: 10.1016/j.isprsjprs.2023.09.009.

[7] B. Zhang, L. Zhao, and X. Zhang, "Three-dimensional convolutional neural network model for tree species classification using airborne hyperspectral images," *Remote Sens. Environ.*, vol. 247, Sep. 2020, Art. no. 111938, doi: 10.1016/j.rse.2020.111938.

[8] C. Chen, L. Jing, H. Li, Y. Tang, F. Chen, and B. Tan, "Using time-series imagery and 3DLSTM model to classify individual tree species," *Int. J. Digit. Earth*, vol. 17, no. 1, Dec. 2024, Art. no. 2308728, doi: 10.1080/17538947.2024.2308728.

[9] X. Wang and H. Ren, "DBMF: A novel method for tree species fusion classification based on multi-source images," *Forests*, vol. 13, no. 1, p. 33, Dec. 2021, doi: 10.3390/f13010033.

[10] M. Tarasiou, E. Chavez, and S. Zafeiriou, "ViTs for SITS: Vision transformers for satellite image time series," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 10418–10428, doi: 10.1109/CVPR52729.2023.01004.

[11] X. Shi, Z. Chen, H. Wang, D. Yeung, W. K. Wong, and W. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 1–9.

[12] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002, doi: 10.1109/ICCV48922.2021.00986.

[13] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11966–11976, doi: 10.1109/CVPR52688.2022.01167.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," 2015, *arXiv:1505.04597*.

[16] S. K. Roy, S. Manna, T. Song, and L. Bruzzone, "Attention-based adaptive spectral–spatial kernel ResNet for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7831–7843, Sep. 2021.

[17] S. Mei, C. Song, M. Ma, and F. Xu, "Hyperspectral image classification using group-aware hierarchical transformer," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5539014, doi: 10.1109/TGRS.2022.3207933.

[18] W.-Y. Loh, "Classification and regression trees," *WIREs Data Mining Knowl. Discovery*, vol. 1, no. 1, pp. 14–23, Jan. 2011, doi: 10.1002/widm.8.