RESEARCH ARTICLE

Forest resource value assessment and prediction model based on machine learning

Shijie Zhou*, Jingliang Li

School of Culture and Communication, Loudi Vocational and Technical College, Loudi, Hunan, China.

Received: August 15, 2024; **accepted:** December 21, 2024.

With the acceleration of global urbanization, the importance of urban green space and landscape design has become increasingly prominent. In the process of urbanization, the destruction of natural ecosystems and the reduction of biodiversity have become issues that cannot be ignored. This study proposed a landscape design method that combined ecological principles and technical applications to enhance the ecological value and social benefits of landscapes. The ecological factor data were processed using a customized transformer model, and the impact of design options on ecological balance was evaluated by an ecological adaptability module. The results showed that the proposed method demonstrated significant advantages in species diversity index, species richness index, and ecosystem service score, while maintaining a high visual aesthetic score, which was superior to traditional design methods. Specifically, the values of mean square error (MSE) of the proposed model on the training set, validation set, and test set were 0.032, 0.045, and 0.050. The coefficients of determination (R^2) were 0.92, 0.89, and 0.87. The values of the mean absolute error (MAE) were 0.12, 0.15, and 0.16, respectively. The feature importance analysis showed that the average SHapley Additive exPlanations (SHAP) value of the normalized difference vegetation index (NDVI) was 0.125, and the SHAP value of the tree species composition was 0.110, indicating that these features had an important impact on the model prediction results. These results not only verified the effectiveness of the method, but also provided new ideas and tools for sustainable development in the field of garden landscape design. By integrating ecological principles and advanced technologies, the ecological functionality of garden landscapes had been significantly improved, while the aesthetic appeal of the design had been improved, providing practical methods and tools for achieving ecological sustainability goals.

Keywords: forest resources; valuation; predictive modeling; machine learning; environmental economics.

***Corresponding author:** Shijie Zhou, School of Culture and Communication, Loudi Vocational and Technical College, Loudi 417000, Hunan, China. Email: zhoushijie126@hotmail.com.

Introduction

With the acceleration of global urbanization, the importance of urban green space and landscape design has become increasingly prominent. In the process of urbanization, the destruction of natural ecosystems and the reduction of biodiversity have become issues that cannot be ignored. At the same time, the public's growing demand for a high-quality living environment has promoted the development of landscape design to a higher level. Landscape design is not only an art of beautifying the environment, but also an important means to restore and enhance ecosystem services such as air purification, water quality improvement, and biodiversity protection. However, the existing design methods are still insufficient in the application of ecological principles, especially in complex ecosystems. In addition, traditional methods are difficult to process large-scale, high-dimensional ecological data, which limits their effective application in actual projects [1, 2].

In recent years, the field of landscape design has made significant progress in both theory and practice. The application of ecological principles has gradually become the core guiding ideology of design, emphasizing the simulation of natural ecosystems through reasonable plant configuration, terrain shaping, *etc*., to promote ecological diversity, water cycle management, and microclimate regulation. On the other hand, the introduction of new technologies has brought new possibilities to landscape design. In particular, the development of deep learning technology has provided a powerful tool for processing complex spatiotemporal series data. For example, the Transformer model has achieved great success in the field of natural language processing due to its efficient parallel computing capabilities and powerful sequence modeling capabilities. It has also been gradually applied to ecological research, helping researchers to better understand and simulate the dynamic changes of ecosystems [3]. Although certain achievements have been made in the field of garden landscape design, there are still several problems that need to be solved. First, the existing design methods are still insufficient in the application of ecological principles, especially in complex ecosystems. Second, traditional methods are difficult to process large-scale, high-dimensional ecological data, and more advanced technical means are needed to support them. Third, how to ensure ecological functions while considering visual aesthetics and achieving multi-objective optimization is a major challenge in current design. Finally, the complexity and black box characteristics of deep learning models make it difficult for designers to understand the decision-making process of the model, affecting the practicality and credibility of the model [4].

This study proposed a garden landscape design method that combined ecological principles and technical applications to enhance the ecological value and social benefits of garden landscapes. The specific objectives included analyzing the limitations of existing garden landscape design methods and their impact on the ecological environment, identifying the shortcomings of current design methods, especially those factors that had a negative impact on ecosystem services [5], exploring the application potential of the transformer model in processing garden landscape design data, and studying how to use the model to process complex sequence data such as vegetation type distribution, terrain changes, and hydrological cycles to provide more accurate design decision support. A set of garden landscape design guidelines based on ecological principles and technical tools was also proposed through this study to provide designers with practical tools and methods to help them better achieve the goal of ecological sustainability in future projects [6, 7]. The study evaluated the effect of the proposed method by comparing it with traditional design methods to verify its advantages [8]. The results of this research would improve the ecological functionality of garden landscapes by integrating ecological principles and advanced technologies. The multi-objective optimization strategy ensured that the design scheme had good visual aesthetics while maintaining high ecological benefits. Further, the application of the transformer model in garden landscape design provided new technical means for processing complex ecological data. This proposed design guidelines provided practical tools and methods for garden designers, which would help to achieve the goal of ecological sustainability and improve the urban ecological quality and the life quality of residents.

Materials and methods

Modeling of value assessment and forecasting problems

The objective of forest resource valuation and forecasting is to estimate the value of forest resources at the current point in time and to predict trends in their future value. The value here covers both the economic value of forest resources and their ecological value [9]. The economic value refers to the economic benefits that forest resources can generate through timber sales, non-timber product development, *etc.*, while the ecological value involves the carbon sink function of forests, water nourishment, and biodiversity conservation. To achieve this goal, the forest resource value assessment and prediction problem were modeled as a multivariate prediction task. Specifically, it was assumed that the value of forest resources $V(t)$ was a function of time t and a set of characteristics x as shown in equation 1.

$$V(t) = f(X, t) \tag{1}$$

where x included, but was not limited to, the remote sensing features like normalized difference vegetation index (NDVI), leaf area index (LAI), vegetation cover, *etc.*; the ground survey features including tree species composition, tree age, tree height, tree diameter, *etc.*; the meteorological characteristics such as average temperature, precipitation, humidity, *etc.*; and other relevant features including soil type, elevation, slope, *etc*. Therefore, the goal was to construct a function f such that $V(t) = f(X, t)$ could accurately predict the value of forest resources at different points in time [10]. To construct such a function f, a supervised learning approach was used. Given a set of training data $\{(X_i, V_i)\}_{i=1}^{n}$, where x_i was the feature vector of the i th observation, and V_i was the corresponding forest resource value, a machine learning algorithm could be used to learn a function f that made the predicted value $\hat{V}_i = f(X_i)$ as close as possible to the true value V_i [11].

Innovative solution methods

Forest resource valuation is a multifactorial task with remotely sensed image data providing information from a macroscopic perspective, while ground survey data and meteorological data provide information in more detail and under specific conditions. Combining these data can provide a more comprehensive picture of the state of forest resources and thus improve the accuracy of prediction. In forest resource assessment, convolutional neural network (CNN) was utilized to extract key features in remote sensing images and thus predict the value of forest resources. Meanwhile, other types of input data like ground survey data and meteorological data should be considered. For this reason, a CNN architecture that fused multimodal data was designed (Figure 1) [12].

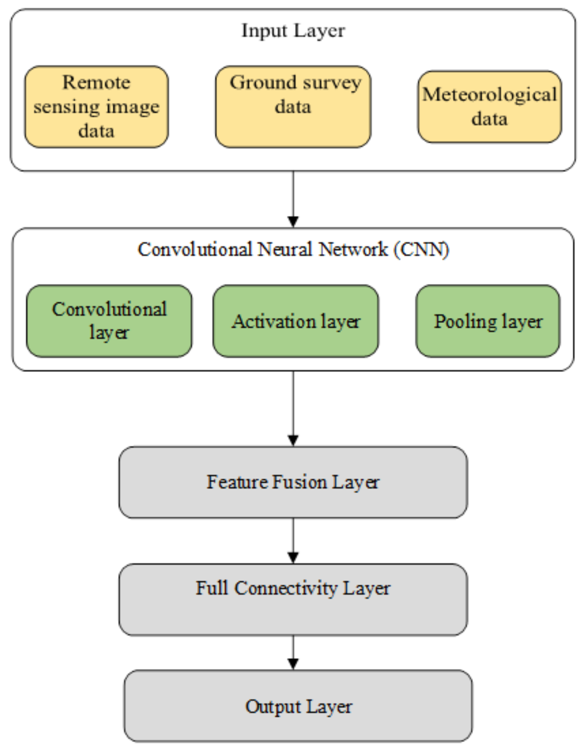

Figure 1. Modeling framework.

The input layer received the pre-processed remote sensing image data, ground survey data, and meteorological data. For image data, assuming the image size was $H \times W$ and the number of channels was C, the input shape was (H, W, C). The ground survey data assumed that the number of features was F_g and the input shape was (F_g). The meteorological data assumed that the number of features was F_m, and the input shape was (F_m). CNN was good at extracting features from image data, especially when dealing with data with spatial structure. With the convolutional layer, the features related to the value of forest resources such as the health of vegetation, tree species distribution, and so on could be automatically learned. The convolutional layer contained multiple convolutional kernels, each of which had a size of $k \times k$ and a step size of s padding of p. The number of output channels was N [13]. The convolution operation was represented by equation 2.

$$Y_{ij} = \sum_{m=0}^{k-1}\sum_{n=0}^{k-1} X_{(i+m)(j+n)} \cdot w_{mn} + b \tag{2}$$

where Y_{ij} was an element in the output feature map. $X_{(i+m)(j+n)}$ was an element in the input feature map. w_{mn} was the weight in the convolution kernel. b was the bias term. In the activation Layer, the $ReLU$ function was usually used to increase the nonlinear expressiveness of the model as shown below.

$$Y'_{ij} = \max(0, Y_{ij}) \tag{3}$$

The pooling layer reduced the spatial dimensions of the data, which reduced the computational complexity and helped the model to extract more representative features, which was especially important for processing high-resolution remote sensing images by preventing overfitting and improving the generalization ability of the model. Commonly used pooling methods are maximum pooling and average pooling. The maximum pooling was expressed as equation 4.

$$P_{ij} = \max\nolimits_{m,n} Y'_{(i+m)(j+n)} \tag{4}$$

The feature fusion layer fused features of remote sensing image, ground survey, and meteorology. Assuming that the remote sensing image feature vector was F_r, the ground survey feature vector was F_g, and the meteorological feature vector was F_{m}, the fused feature vector F could be expressed as equation 5.

$$F = [F_r; F_g; F_m] \tag{5}$$

where ";" denoted the splicing of feature vectors. The fully connected layer spread the fused feature vectors and mapped them to the final output through the fully connected layer. The output dimension was O. The output was calculated as equation 6.

$$Z = WX + b \tag{6}$$

where X was the spread feature vector. W was the weight matrix. b was the bias vector. Z was the output vector. In the output layer, according to the task requirements, the loss function for regression task or classification task could be chosen. For forest resource value prediction, the mean square error (MSE) was usually used as the loss function and was calculated below.

$$\mathrm{Loss} = \frac{1}{N}\sum_{i=1}^{N}(Z_i - V_i)^2 \tag{7}$$

where Z_i was the forest resource value predicted by the model. V_i was the true forest resource value. N was the sample size.

Interpretive enhancements

When constructing forest resource value assessment and prediction models, many types of features including remote sensing features, ground survey data, and meteorological features were usually used. These features had different degrees of influence on predicting forest resource values. To understand how these features affected the model prediction, SHapley Additive exPlanations (SHAP) values were used to quantify the importance of each feature. The framework structure of SHAP was shown in Figure 2. Assuming a prediction model f with an input feature vector $X = (x_1, x_2, \dots, x_n)$, the contribution of each feature x_i to the model prediction $f(X)$ could be expressed by the SHAP value ϕ_i. For a given subset of features S, the marginal contribution could be defined as follows [14].

$$\Delta_f(S \cup \{i\}) - \Delta_f(S) \tag{8}$$

where $\Delta_f(S)$ was the marginal contribution of the feature set S to the predicted value $f(X)$. In other words, $\Delta_f(S \cup \{i\}) - \Delta_f(S)$ was the change in the prediction when the feature x_i was added to the se t S. The calculation of SHAP value involved a weighted average of the marginal contributions of all possible subsets of features S and its mathematical form could be expressed as below.

$$\phi_i(f) = \sum_{S \subseteq N,\ \{i\}} \frac{|S|!(n-|S|-1)!}{n!}[\Delta_f(S \cup \{i\}) - \Delta_f(S)] \tag{9}$$

where N was the set of all features. n was the number of features. S was the subset of features. $|S|$ was the number of features in S. In forest resource valuation and prediction models, SHAP values were used to assess the importance of features and understand how they affected the predictions of the model. Meteorological characteristics such as mean temperature and precipitation reflected the environmental conditions under which the forest grew.

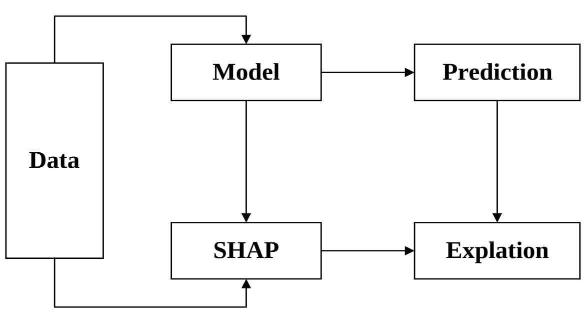

Figure 2. SHAP algorithm framework.

To calculate SHAP values for these features, the set of all features N was first defined, followed by the calculation of marginal contribution of all possible feature subsets S for each feature x_i as specified in equation 10.

$$\Delta_f(S \cup \{i\}) - \Delta_f(S) \tag{10}$$

Weights were then calculated based on the size of the feature subset. All marginal contributions were weighted and averaged to obtain a SHAP value of $\phi_i(f)$ for each feature x_i.

Suppose a forest resource valuation and prediction model were used to assess the impact of the remotely sensed feature NDVI on the model predictions, SHAP value of NDVI could be calculated by defining the feature set of three features including the NDVI, the tree species composition (Tree_Species), and the Average_Temperature, which could be expressed as follows.

$$N = \{NDVI,\ Tree\_Species,\ Average\_Temperature\} \tag{11}$$

The marginal contribution was then calculated by considering all possible subsets of features S, excluding NDVI, as below.

$$\Delta_f(S \cup \{NDVI\}) - \Delta_f(S) \tag{12}$$

For each feature subset S, weights were used to calculate the marginal contribution of NDVI. For the feature subset $S = \{Tree\_Species\}$, the weights were $\frac{1!(3-1-1)!}{3!} = \frac{1}{6}$. The SHAP value for NDVI was ultimately calculated as the weighted sum of the marginal contributions of all feature subsets S as expressed in equation 13.

$$\phi_{NDVI}(f) = \sum_{S \subseteq N,\ \{NDVI\}} \frac{|S|!(3-|S|-1)!}{3!}[\Delta_f(S \cup \{NDVI\}) - \Delta_f(S)] \tag{13}$$

The high SHAP value for NDVI indicated that the feature had a greater positive or negative impact on the prediction of forest resource values. Similarly, SHAP values for other features such as species composition and mean temperature could be calculated and compared for the relative importance. To understand these results more intuitively, SHAP value visualization tools such as SHAP value distribution or dependency charts were employed, which could be helpful to quickly identify which characteristics were the most critical to forest resource values and how they changed over time [15].

Data sets

A total of 40 GB high-resolution multispectral remote sensing image data were obtained from the Sentinel-2 satellite database (https://scihub.copernicus. eu/), while a total of 10 GB ground survey data including the vegetation type distribution and tree species composition data were provided locally by the Beijing Municipal Bureau of Landscape and Forestry (http://yllhj.beijing.gov.cn/). About 20 GB of the meteorological data were obtained from the National Meteorological Bureau, China Meteorological Administration (http://www. cma.gov.cn/). About 30 GB of ecological factor data and terrain data were obtained from the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences (IGSNRR) (Beijing, China) (http://www.igsnrr.cas.cn/). In general, a total of approximately 100 GB of data covering a variety of data types were involved in this research to fully support the research on garden landscape design and ecological balance optimization. The original data were cleaned, integrated, feature engineered, and normalized to ensure the quality of model inputs before being divided into training, validation, and testing sets with the ratio of 70%, 15%, and 15% to guarantee the efficiency of model training and the fairness of testing [16].

Experimental design

A series of delicate data preprocessing steps including removal of anomalies and missing data, integration of data across sources, creation of new features such as normalized difference vegetation index (NDVI), and normalization of the data were implemented to construct a reliable and consistent dataset. The knowledge of domain experts and preliminary data analysis results were combined to carefully select key features for this proposed methodology, while considering a variety of advanced machine learning algorithms to find the best model. Through a rigorous training, validation, and testing process, the performance of the model was comprehensively evaluated using core performance metrics such as mean square error (MSE), coefficient of determination (R^2), and mean absolute error (MAE) to provide a comprehensive measure of the model's ability to generalize and the accuracy of its predictions. Further, the key factors affecting the prediction results of the models were successfully identified through in-depth feature importance analysis and performance comparison with support vector machines (SVM) model, random forest model, linear regression model, and multilayer perceptron model (MLP) for the most suitable models for solving the problem of forest resource valuation.

Results and discussion

Evaluation of proposed model performance

The best model performance indicators were obtained through iterative training and evaluation on the training set, validation set, and test set. The mean square error (MSE), coefficient of determination (R^2), and mean absolute error (MAE) of the model were recorded to quantify the prediction accuracy and generalization ability of the model. Specifically, the mean square errors of the model were 0.032, 0.045, and 0.050 for training set, validation set, and test set, respectively, indicating that the model had a certain generalization ability. The coefficients of determination were 0.92, 0.89, and 0.87 for training, validation, and test sets, respectively, indicating that the model could well explain the changes in the value of forest resources. The mean absolute errors were 0.12, 0.15, and 0.16 for training, validation, and test sets, respectively, further proving the prediction accuracy of the model.

Feature importance analysis

Table 1. Characteristic importance analysis.

Feature name	Mean SHAP value	Rank
Normalized Difference Vegetation Index (NDVI)	0.125	1
Tree species composition	0.110	2
average temperature	0.098	3
measured quantity of rain	0.082	4
age of trees	0.075	5
Leaf Area Index (LAI)	0.068	6
Soil type	0.060	7
tree diameter	0.055	8
humidity level	0.050	9
altitude	0.045	10

Table 2. Comparison of model predictions.

Point of time	Real value	Predicted value	Absolute error
January 1, 2021	1,200	1,180	20
April 1, 2021	1,300	1,290	10
July 1, 2021	1,400	1,380	20
October 1, 2021	1,500	1,490	10
January 1, 2022	1,600	1,580	20
April 1, 2022	1,700	1,690	10
July 1, 2022	1,800	1,790	10
October 1, 2022	1,900	1,890	10
January 1, 2023	2,000	1,990	10
April 1, 2023	2,100	2,090	10

The feature importance analysis was based on the SHAP value calculated after model training. The average SHAP value of each feature was calculated and ranked to reveal the relative impact of different features on the model prediction results. The results showed that the average SHAP value of the normalized difference vegetation index (NDVI) was 0.125, which was ranked the top one feature compared to the other features (Table 1), indicating that NDVI had the greatest impact on the model prediction results. SHAP value of tree species composition was 0.110, which was also a key feature. These data revealed which features played a decisive role in predicting the value of forest resources.

Comparison of model predictions

The comparison between model predictions and actual values showed that the prediction accuracy of the proposed model at different time

Table 3. Distribution of SHAP values.

Features	Minimum SHAP	Maximum SHAP	Mean SHAP
Normalized difference vegetation index (NDVI)	-0.2	0.25	0.125
Tree species composition	-0.15	0.22	0.110
average temperature	-0.12	0.21	0.098
measured quantity of rain	-0.1	0.2	0.082
age of trees	-0.09	0.18	0.075
Leaf Area Index (LAI)	-0.08	0.17	0.068
Soil type	-0.07	0.16	0.060
tree diameter	-0.06	0.15	0.055
humidity level	-0.05	0.14	0.050
altitude	-0.04	0.13	0.045

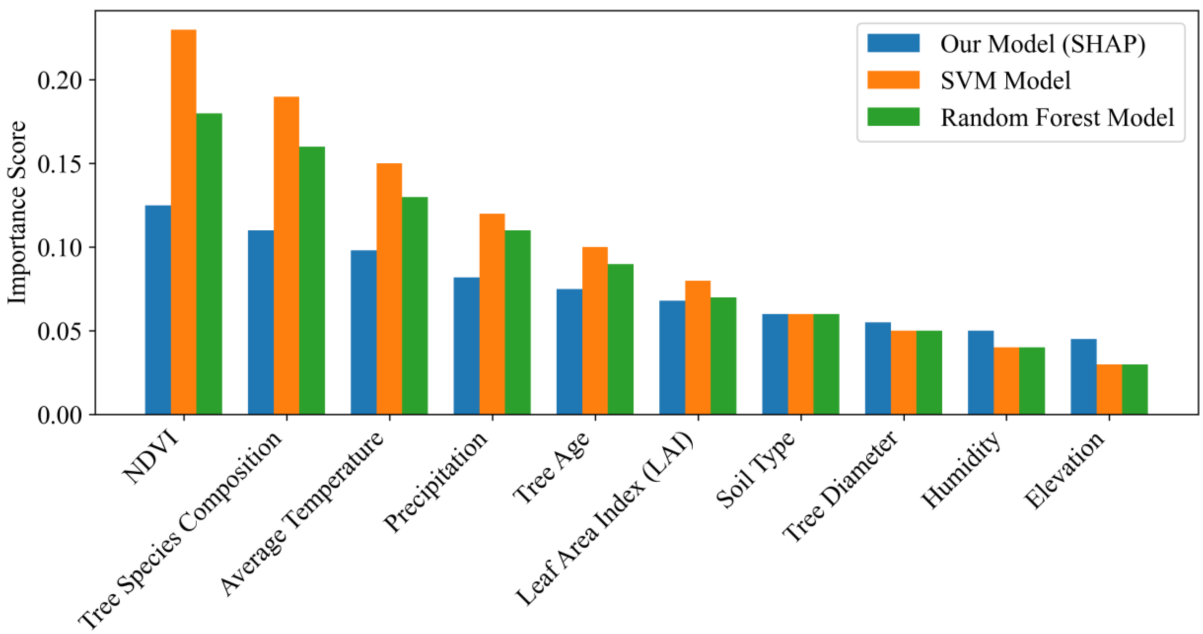

Figure 3. Comparison of feature importance across different models.

points was very close to the actual situation with predicted values slightly less than real values. The results demonstrated that, among 10 different time points, 7 out of 10 points showed absolute errors of only 10, while 3 out of 10 points demonstrated absolute errors of 20 (Table 2). The results indicated that the proposed model could be applied in real practices.

Distribution of SHAP values and comparison of feature importance

The distribution of SHAP values provided the minimum, maximum, and average values of each feature. The results demonstrated that SHAP values of NDVI ranged from -0.2 to 0.25 with an average of 0.125, which was the highest average SHAP value among the other features and

Table 4. Comparison of model performance.

Model name	Mean Square Error (MSE)	Coefficient of determination (R^2)	Mean Absolute Error (MAE)
Proposed SHAP model	0.050	0.87	0.16
SVM model	0.065	0.83	0.18
Random forest model	0.055	0.85	0.17
Linear regression model	0.070	0.81	0.19
Multilayer Perceptron Model (MLP)	0.052	0.86	0.165

showed the direction and degree of influence of NDVI on the prediction results under different scenarios (Table 3). The features' importances were determined through feature importance analysis by different models. The results showed that all three models demonstrated similar trends of feature importance with NDVI, tree species composition, average temperature, precipitation, and tree age as the top 5 important features (Figure 3).

Model performance comparison

The results of model performance comparison demonstrated that the advantages of proposed model overed other models through specific performance indicator values including MSE and R^2. The MSE of proposed model on the test set was 0.050, which was lower than 0.065 of the SVM model, indicating that the proposed model had a smaller prediction error. The R^2 value was also higher than other models, further proving the superiority of proposed model in predicting the value of forest resources (Table 4). Compared with SVM model, the proposed model of this study performed well in performance indicators of MSE, R^2, and MAE, which indicated that the method of fusing multimodal data and using SHAP values to enhance interpretability was effective and helped to improve the predictive performance and interpretability of the model. The results not only verified the effectiveness of the proposed method, but also provided valuable references for future research.

References

1. Astola H, Hame T, Sirro L, Molinier M, Kilpi J. 2019. Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region. Remote Sens Environ. 223:257-273.
2. Vauhkonen J. 2018. Predicting the provisioning potential of forest ecosystem services using airborne laser scanning data and forest resource maps. For Ecosyst. 5:19.
3. Kamal M, Zhang BL, Cao JF, Zhang X, Chang J. 2022. Comparative study of artificial neural network and random forest model for susceptibility assessment of landslides induced by earthquake in the western Sichuan Plateau, China. Sustainability. 14(21):14.
4. Latif QS, Cannon JB, Chabot EJ, Sparks RA. 2022. Simulated treatment effects on bird communities inform landscape-scale dry conifer forest management. Ecol Appl. 32(4):19.
5. Liu Y, Zhao QS, Hu CC, Luo NX. 2023. Prediction of storm surge water level based on machine learning methods. Atmosphere. 14(10):20.
6. Hua WP, Ye HM, Rau JY, Qiu T. 2020. Dynamic prediction of uneven-aged natural forest for yield of Pinus taiwanensis using joint modelling. Environ Monit Assess. 192(4):13.
7. Dobre M, Srivastava A, Lew R, Deval C, Brooks ES, Elliot WJ, *et al*. 2022. WEPPcloud: An online watershed-scale hydrologic modeling tool. Part II. Model performance assessment and applications to forest management and wildfires. J Hydrol. 610:18.
8. Müller MM, Vilà-Vilardell L, Vacik H. 2020. Towards an integrated forest fire danger assessment system for the European Alps. Ecol Inform. 60:10.
9. Lalitha M, Dharumarajan S, Suputhra A, Kalaiselvi B, Hegde R, Reddy RS, *et al*. 2021. Spatial prediction of soil depth using environmental covariates by quantile regression forest model. Environ Monit Assess. 193(10):10.
10. Huang SS, Dong QJ, Zhang X, Deng WS. 2021. Catchment natural driving factors and prediction of baseflow index for Continental United States based on Random Forest technique. Stoch Environ Res Risk Assess. 35(12):2567-2581.
11. Garrido AD, Gobakken T, Hauglin M, Naesset E, Bollandsås OM. 2023. Accuracy assessment of the nationwide forest attribute map of Norway constructed by using airborne laser scanning

data and field data from the national forest inventory. Scand J For Res. 38(1-2):9-22.
12. Shi BW, Wan C, Hu WW, Xu CC, Liu D, Li MX, *et al*. 2023. Analysis and prediction of the changes in groundwater resources under heavy precipitation and ecological water replenishment. J Water Clim Chang. 14(6):1762-1778.
13. Qi B, Yu M, Li YY. 2024. Multi-scenario prediction of land-use changes and ecosystem service values in the Lhasa River Basin based on the FLUS-Markov model. Land. 13(5):24.
14. Karna YK, Penman TD, Aponte C, Hinko-Najera N, Bennett LT. 2020. Persistent changes in the horizontal and vertical canopy structure of fire-tolerant forests after severe fire as quantified using multi-temporal airborne lidar data. For Ecol Manag. 472:12.
15. Zhou P, Zhang HJ, Huang B, Ji YL, Peng SL, Zhou T. 2022. Are productivity and biodiversity adequate predictors for rapid assessment of forest ecosystem services values? Ecosyst Serv. 57:13.
16. Gigovic L, Pourghasemi HR, Drobnjak S, Bai SB. 2019. Testing a new ensemble model based on SVM and random forest in forest fire susceptibility assessment and its mapping in Serbia's Tara national park. Forests. 10(5):21.