



# Article Comparing Algorithms for Estimation of Aboveground Biomass in *Pinus yunnanensis*

Tianbao Huang <sup>1,2,3</sup>, Guanglong Ou <sup>3</sup>, Hui Xu <sup>3</sup>, Xiaoli Zhang <sup>3</sup>, Yong Wu <sup>3</sup>, Zihao Liu <sup>3</sup>, Fuyan Zou <sup>1,2</sup>, Chen Zhang <sup>1,2</sup> and Can Xu <sup>1,2,\*</sup>

- <sup>1</sup> Kunming General Survey of Natural Resources, China Geological Survey, Kunming 650111, China; huangtianbao@swfu.edu.cn (T.H.); zoufuyan@mail.cgs.gov.cn (F.Z.); zhangchen002@mail.cgs.gov.cn (C.Z.)
- <sup>2</sup> Technology Innovation Center for Natural Ecosystem Carbon Sink, Ministry of Natural Resources, Kunming 650111, China
- <sup>3</sup> Key Laboratory of Southwest Mountain Forest Resources Conservation and Utilization, Ministry of Education, Southwest Forestry University, Kunming 650233, China; olg2007621@swfu.edu.cn (G.O.); swfc213@126.com (H.X.); karojan@swfu.edu.cn (X.Z.); yongwu@swfu.edu.cn (Y.W.); liuzihao@swfu.edu.cn (Z.L.)
- \* Correspondence: xucan@mail.cgs.gov.cn

Abstract: Comparing algorithms are crucial for enhancing the accuracy of remote sensing estimations of forest biomass in regions with high heterogeneity. Herein, Sentinel 2A, Sentinel 1A, Landsat 8 OLI, and Digital Elevation Model (DEM) were selected as data sources. A total of 12 algorithms, including 7 types of learners, were utilized for estimating the aboveground biomass (AGB) of Pinus yunnanensis forest. The results showed that: (1) The optimal algorithm (Extreme Gradient Boosting, XGBoost) was selected as the meta-model (referred to as XGBoost-stacking) of the stacking ensemble algorithm, which integrated 11 other algorithms. The  $\mathbb{R}^2$  value was improved by 0.12 up to 0.61, and RMSE was decreased by 4.53 Mg/ha down to 39.34 Mg/ha compared to the XGBoost. All algorithms consistently showed severe underestimation of AGB in the Pinus yunnanensis forest of Yunnan Province when AGB exceeded 100 Mg/ha. (2) XGBoost-Stacking, XGBoost, BRNN (Bayesian Regularized Neural Network), RF (Random Forest), and QRF (Quantile Random Forest) have good sensitivity to forest AGB. QRNN (Quantile Regression Neural Network), GP (Gaussian Process), and EN (Elastic Network) have more outlier data and their robustness was poor. SVM-RBF (Radial Basis Function Kernel Support Vector Machine), k-NN (K Nearest Neighbors), and SGB (Stochastic Gradient Boosting) algorithms have good robustness, but their sensitivity was poor, and QRF algorithms and BRNN algorithm can estimate low values with higher accuracy. In conclusion, the XGBoost-stacking, XGBoost, and BRNN algorithms have shown promising application prospects in remote sensing estimation of forest biomass. This study could provide a reference for selecting the suitable algorithm for forest AGB estimation.

**Keywords:** forest aboveground biomass; machine learning algorithm; remote sensing; *Pinus yunnanensis* forest; stacking ensemble; Yunnan Province; China

## 1. Introduction

As a crucial quantitative and qualitative indicator of forest ecosystems, forest biomass holds great importance in swiftly acquiring information on the quantity of biomass within forests [1,2]. Nevertheless, traditional remote sensing methods for estimating forest biomass suffer from disadvantages, such as low efficiency, high cost, and ecological damage. As a result, there is a growing interest in biomass estimation based on remote sensing methods to overcome the shortage above. Meanwhile, to enhance the accuracy of remote sensing estimation of forest aboveground biomass (AGB), researchers have conducted extensive studies utilizing diverse image data sources and algorithms [3,4].



Citation: Huang, T.; Ou, G.; Xu, H.; Zhang, X.; Wu, Y.; Liu, Z.; Zou, F.; Zhang, C.; Xu, C. Comparing Algorithms for Estimation of Aboveground Biomass in *Pinus yunnanensis*. *Forests* **2023**, *14*, 1742. https://doi.org/10.3390/f14091742

Academic Editors: Svein Solberg and Chunying Ren

Received: 10 July 2023 Revised: 18 August 2023 Accepted: 25 August 2023 Published: 28 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). In AGB estimation of remote sensing, the data sources typically include lidar data, multispectral data, synthetic aperture radar (SAR), hyperspectral data, and others. While lidar data is not affected by data saturation effects and exhibits high estimation accuracy, its widespread application is limited due to its high cost [4]. For free data source images, Landsat 8 OLI and Sentinel 2A multispectral images have been widely used in forest AGB remote sensing estimation due to their advantages, such as wide coverage, free acquisition, high spatial-temporal resolution, mature technology, sensitivity to SWIR band, and rededge bands, etc., which help to monitor vegetation leaf characteristics [5–9]. In addition, the free open-access Sentinel 1A SAR, a long-wave active sensor with the advantages of day and night operation, no rain and cloud interference, senses forest geometry better than passive optical sensors and provides valuable data for mapping forest AGB [10]. Numerous studies have also shown that collaborative estimation of forest AGB using multi-source remote sensing imagery could improve estimation accuracy, especially in regions with high heterogeneity [11,12].

AGB estimation of remote sensing had many uncertainties caused by remote sensing data sources, prediction models, forest physical environment, mixed pixel, sample biomass calculation error, sampling error, image time mismatch, and other influencing factors, which limited the accuracy of remote sensing estimation of AGB [13,14]. Among these, the model plays a crucial role in the AGB estimation of remote sensing, and the selection and performance of the model directly affect the accuracy and reliability of the AGB estimate [4,15], so it was important to select a suitable algorithm for the AGB model to improve the accuracy of remote sensing estimation of AGB. Machine learning algorithms have been widely used in the estimation of forest AGB by Remote Sensing because they can capture complex non-linear relationships between variables in multiple data sources and have high estimation accuracy [16,17]. Simultaneously, the seven types of learners algorithms, such as Bagging Learners, Boosting Learners, neural networks, linear-based learners, kernel-based learners, K-nearest neighbor learners, and stacking ensemble algorithms, have been gradually used for the estimation of forest AGB by Remote Sensing.

Review of the application of seven types of learners algorithms mentioned above that are used in AGB estimation of remote sensing, the commonly used kNN algorithm has been widely applied due to its simplicity and good estimation performance [18,19]. The Random Forest (RF) algorithm has been generally recognized as an excellent choice for bagging learners due to its robustness and high accuracy, establishing it as the most commonly employed algorithm [20]. However, it is important to note that the utilization of Quantile Random Forest (QRF), an enhanced variant of the Random Forest algorithm, has relatively few applications in estimating forest AGB by remote sensing [21,22]. Extreme gradient boosting (XGBoost), an advanced tree boosting system and an enhancement of gradient boosting based on Boosting Learners, has demonstrated outstanding performance in estimating AGB from remote sensing data [23–25]. Furthermore, Guneralp et al. (2014) [26] have shown that Stochastic Gradient Boosting (SGB) outperforms other algorithms, including MARS (Multivariate Adaptive Regression Splines) and Cubist (Cubist), in the forest AGB estimation of remote sensing within the context of Boosting Learners. The Quasi-Recurrent Neural Network (QRNN) algorithm, as a part of Neural Networks Learners, has started to gain attention in the field of estimation of forest AGB by Remote Sensing. Li et al. (2023) [27] have demonstrated the capability of QRNN to effectively enhance forest AGB estimation using remote sensing data. Furthermore, Bayesian Regularized Neural Networks (BRNN) have addressed the challenges of overfitting and robustness that are commonly associated with artificial neural networks. Although BRNN has attracted considerable interest in other domains, its application in remote sensing for forest AGB estimation remains relatively unexplored [28,29]. Among linear-based learners, Alvarez-Mendoza et al. (2022) [30] have demonstrated the favorable estimation performance of Bayesian Ridge Regression (BRR) in estimating grassland forest AGB through remote sensing. Moreover, the Elastic Network Program (EN) represents a regularized version of linear regression that combines the characteristics of Tikhonov Regularization (ridge) regression and Least Absolute Shrinkage

and Selection Operator (LASSO) regression [31]. By incorporating the properties of both algorithms, it produces estimates that can be interpreted as Bayesian posterior modes under a prior distribution implied by the elastic network form. Despite the potential benefits, there have been limited studies employing remote sensing for AGB using the EN [31]. Kernel-based learners referred to a family of algorithms that utilized kernel functions to project low-dimensional data into a higher-dimensional space, enabling linear separability. These algorithms were capable of handling nonlinear problems while still relying on linear algebra [32]. Ghosh et al. (2023) [33] demonstrated that the Gaussian Process algorithm, as a kernel-based learner, enhanced the accuracy of the estimation of forest AGB by remote sensing. Additionally, Support Vector Machine (SVM), a popular machine learning technique, was found to have widespread application in the estimation of forest AGB using remote sensing resources [4]. SVM employed a kernel function, such as the radial basis function (RBF), to process nonlinear data by mapping it into a higher-dimensional feature space. The SVM with RBF Kernel algorithm (SVM-RBF) excelled at modeling nonlinear relationships between input and output variables and exhibited robustness against noise and outliers [34,35].

Moreover, integration algorithms have been shown to exhibit higher estimation accuracy compared to single algorithms, such as the Stacking ensemble algorithm, as one of the classic integration algorithms, combines the strengths of various models and has been increasingly applied in the estimation of AGB using remote sensing resources [36,37]. In Stacking ensemble models, the diversification of the base models plays a vital role in enhancing the integrated model [38]. Additionally, selecting models with high generalization capability in the second layer as meta-models can effectively address and rectify any bias present in the first-layer base learners towards the training data. The data generated in the first layer for secondary prediction can further enhance the performance of the first layer. Therefore, the model selection in the second layer holds significant importance status [39]. By comparing multiple algorithms, selecting the optimal algorithm for remote sensing estimation of AGB became an important pathway to improve the accuracy of AGB estimation of remote sensing [16,38,40].

Although most of the different algorithms have been used in AGB estimation by remote sensing, there are still incomplete comparisons of algorithms for different learners, and some algorithms have not been investigated in forest AGB estimation, especially in highly heterogeneous landscapes. Yunnan Province is located in a longitudinal ridge and valley area with complex geological conditions and a special geographical location, resulting in high forest heterogeneity [41–43]. Accurately estimating forest AGB using remote sensing in such areas is undoubtedly a challenge [17]. For this reason, this study selected Sentinel 2A, Sentinel 1A, Landsat 8 OLI, and Digital Elevation Model (DEM) as data sources and selected 12 algorithms that pairs of bagging learners, boosting learners, neural networks, linear-based learners, kernel-based learners, KNN and stacking ensemble learners to explore the remote sensing estimation of *Pinus yunnanensis* forests in Yunnan Province.

The aims of this study were: Comparing the performance of 12 algorithms from 7 types of learners on AGB estimation of *Pinus yunnanensis* forests in highly heterogeneous landscapes.

#### 2. Study Area and Materials

#### 2.1. Study Area

Yunnan Province is located between 97°31′–106°11′ E and 21°8′–29°15′ N, on the Yunnan-Guizhou Plateau, predominantly mountainous and highland, with a total area of approximately 394,000 square kilometers in southwestern China, bordering the southeastern edge of the Tibetan Plateau [41,44]. The terrain slopes from northwest to southeast, with an altitude of 74–6457 m. Yunnan has a highland tropical monsoon climate with average summer and winter temperatures of 19–22 °C and 6–8 °C, respectively. *Pinus yunnanensis* is an endemic species of southwestern China. It generally grows in the plateau mountains and medium-high valleys at an altitude of 250–3500 m and was concentrated at an altitude of 1600–2900 m. The main dominant tree species in Yunnan Province are *Pinus Yunnanensis*,

*Pinus kesiya, Pinus armindii,* oaks, *Alnus nepalensis* and other tree species. The *Pinus yunnanensis* was the forest type with the largest distribution area in Yunnan Province, and its horizontal distribution extended to 28°23′33″ N in the north, 23°01′20″ N in the south, 97°46′39″ E in the east and 105°54′05″ E in the west [45]. *Pinus yunnanensis* not only plays an important role in the ecological benefits of soil and water conservation in plateau areas but also brings high economic and social benefits [45]. Figure 1 shows the location of the study area.



**Figure 1.** Location of the study area: (**a**) the location of Yunnan province in China, (**b**) sample plots of *Pinus yunnanensis* in Yunnan province, (**c**) the Landsat 8 OLI image.

## 2.2. AGB Calculation of Pinus yunnanensis Sample Plots

The ground data were derived from the 2021 survey of 210 *Pinus yunnanensis* forest plots in the Continuous Forest Inventory (CFI) of Yunnan Province, and the distribution of sample plots is shown in Figure 1. The basic information, such as the dominant species, the diameter at breast height (*DBH*) of individual trees, tree height, average height and the coordinate, and plot coordinates, was recorded by terrestrial RTK. The survey accuracy met the requirements of the Technical Regulations for Continuous Inventory of Forest Resources. Calculation of the aboveground biomass of individual *Pinus yunnanensis* trees is based on Liu et al. (2015), and the R<sup>2</sup> was 0.99 in equation [46], the equation was:

$$M = 0.048 \times DBH^{1.9276} \times H^{0.9638} \tag{1}$$

where DBH (cm) is the average diameter at breast height (1.3 m), H (m) is the average tree height, and M is the aboveground biomass of a single standing tree (kg).

To obtain the AGB of the sample plot, the unit was converted into the value per hectare using Equation (2). The final *AGB* statistical data of the *Pinus yunnanensis* forests are shown in

$$AGB = \frac{n \times M}{25.8 \times 25.8} \times \frac{10,000}{1000}$$
(2)

where *M* was the aboveground biomass of a single standing tree (kg), *n* was the number of trees in the sample plot, and *AGB* was the AGB of the sample plot (Mg/ha). Seventy percent modeling and 30 percent model evaluation were adopted. The sample basic information is shown in Figure 2. The 147 plots were used for model construction, and 63 plots were used for model evaluation, and there was little difference between the model sample and the test sample (p = 0.94 for Wilcoxon test).



Figure 2. Basic overview of the samples.

2.3. Remote Sensing Data Acquisition and Variable Extraction

The DEM data was from the Geospatial Data Cloud (http://www.gscloud.cn/ accessed on 9 July 2023) at a spatial resolution of  $30 \times 30$  m (obtained by space-borne sensors). Sentinel 1A, Sentinel 2A, and Landsat 8 OLI were downloaded from Google Earth Engine (https://code.earthengine.google.com/ accessed on 9 July 2023) to match the survey data. Sentinel 2A and Landsat 8 OLI data were surface reflectance products that selected less than 3% cloud shadow and 5% cloud to synthesize from median values of the Yunnan area in 2021 January-December. The Landsat 8 OLI was from "LANDSAT/LC08/C01/T1\_SR", and Sentinel 2A was from "COPERNICUS/S2\_SR" in Google Earth Engines. Sentinel 1A was from "COPERNICUS/S1\_GRD" in Google Earth Engines. The Sentinel-1 mission provides data from a dual-polarization C-band Synthetic Aperture Radar (SAR) instrument at 5.405 GHz (C band). This collection includes the S1 Ground Range Detected (GRD) scenes, processed using the Sentinel-1 Toolbox to generate a calibrated, ortho-corrected product. Besides, the image synthesis time was on 20 January 2023 and resampled by  $30 \times 30$  m. Subsequently, a 30 m resolution DEM was used for the terrain correction of Sentinel 2A, Landsat 8 OLI, and Sentinel 1A. The vegetation indices, single band, and texture features of Sentinel 2A and Landsat 8 OLI were calculated in ENVI 5.3 [47,48]. Landsat 8 OLI included 7 spectral bands, 17 vegetation indices, and 168 texture variables (3  $\times$  3, 5  $\times$  5, 7  $\times$  7 were from the gray-level co-occurrence matrix GLCM). Sentinel 2A included 12 spectral bands, 18 vegetation indices, and 168 texture variables ( $3 \times 3, 5 \times 5, 7 \times 7$  grey-scale co-occurrence matrix feature GLCM). The spectral variables are shown in Table 1.

Table 1. Spectral variables.

Image Source	Index	Abbreviation	Formula
Sentinel 1A	vertical transmit-vertical channel	VV	-
	vertical transmit-horizontal channel	VH	-

Table 1. Cont.

Image Source	Index	Abbreviation	Formula
	B2-Blue, B3-Green, B4-Ged, B5-Gegetation red edge, B6-Vegetation red edge, B7-Vegetation red edge, B8-NIR, B9-Water vapour, B10-SWIR-Cirrus, B11-SWIR,	B2, B3, B4, B5, B6, B7, B8, B9, B10	-
	ratio vegetation index	RVI	B8/B4
	difference vegetation index	DVI	B8 - B4
	weighted difference vegetation index	WDVI	$B8 - 0.5 \times B4$
	infrared vegetation index	IPVI	B8/(B8 + B4)
	perpendicular vegetation index	PVI	$\sin(45) \times B8 - \cos(45) \times B8$
	normalized difference vegetation index	NDVI	(B8 - B4)/(B8 + B4)
	NDVI with band4 and band5	NDVI45	(B5 - B4)/(B5 + B4)
	NDVI of green band	GNDVI	(B7 – B3)/(B7 + B3)
	inverted red edge chlorophyll index	IRECI	(B7 - B4)/(B5/B6)
Sentinel 2A	soil adjusted vegetation index	SAVI	1.5  imes (B8 - B4) / 8  imes (B8 + B4 + 0.5)
	transformed soil-adjusted vegetation index	TSAVI	$0.5 \times (B8 - 0.5 \times B4 - 0.5)/(0.5 \times B8 + B4 - 0.15)$
	modified soil-adjusted vegetation index	MSAVI	$(2 - \text{NDVI} \times \text{WDVI}) \times$ $(B8 - B4)/8 \times (B8 + B4 -$ $1 - \text{NDVI} \times \text{WDVI})$
	sentinel-2 red edge position index	S2REP	$705 + 35 \times [(B4 + B7) / 2 - B5] \times (B6 - B5)$
	red edge infection point index	REIP	$700 + 40 \times [(B4 + B7) / 2 - B5] / (B6 - B5)$
	atmospherically resistant vegetation index	ARVI	$B8 - (2 \times B4 - B2)/$ $B8 + (2 \times B4 - B2)$
	pigment-specific simple ratio chlorophyll index	PSSRa	B7/B4
	meris terrestrial chlorophyll index	MTCI	(B6 - B5)/(B5 - B4)
	modified chlorophyll absorption ratio index	MCARI	$[(B5 - B4) - 0.2 \times (B5 - B3)] \times (B5 - B4)$
Landsat 8 OLI	band1—coastal aerosol, band2—blue (BLU), band3—green (GRN), band4—red (RED), band5—near-infrared (NIR), band6—shortwave infrared 1 (SWIR1), and band7—shortwave infrared 2 (SWIR2).	B1, B2, B3, B4, B5, B6, B7	-
	normalized difference vegetation index	NDVI	(B5 - B4)/(B5 + B4)
	NDVI with band3 and band4	ND43	(B4 - B3)/(B4 + B3)
	NDVI with band6 and band7	ND67	(B6 – B7)/(B6 + B7)
	NDVI with band3 and band5 with band6	ND563	((B5 + B6) - B3)/(B5 + B6 + B3)

Image Source	Index	Abbreviation	Formula
	difference vegetation index	DVI	B5 – B4
	soil-adjusted vegetation index	SAVI	$((1+0.5) imes( ext{B5}- ext{B4}))/(0.5+ ext{B5}+ ext{B4})$
	ratio vegetation index	RVI	B4/B3
	brightness vegetation Index	В	$\begin{array}{c} 0.2909 \times \text{B2} + 0.2493 \times \text{B3} + \\ 0.4806 \times \text{B4} + 0.5568 \times \text{B5} + \\ 0.4438 \times \text{B6} + 0.1706 \times \text{B7} \end{array}$
	greenness vegetation Index	G	$\begin{array}{c} -0.2728 \times B2 - 0.2174 \times B3 \\ -0.5508 \times B4 + 0.7221 \times B5 \\ +0.0733 \times B6 - 0.1648 \times B7 \end{array}$
	temperature vegetation index	W	$\begin{array}{c} 0.1446 \times B2 + 0.1761 \times B3 + \\ 0.3322 \times B4 + 0.3396 \times B5 - \\ 0.6210 \times B6 - 0.4186 \times B7 \end{array}$
	atmospherically resistant vegetation index	ARVI	$(B5 - (2 \times B4 - B2))/(B5 + (2 \times B4 - B2))$
	mid-infrared temperature vegetation index	MV17	(B5 – B7)/(B5 + B7)
	modified soil adjusted vegetation index	MSAVI	$\begin{array}{c}(2\times B5 + 0.25 - ((2\times B5 + \\ 0.25)^2 - 8\times (B5 - B4))^{0.5})/2\end{array}$
	multiband Linear combination of band2 with band3 and band4	VIS234	B2 + B3 + B4
	multiband Linear combination	ALBEDO	B2 + B3 + B4 + B5 + B6 + B7
	Simple Ratio Index	SR	B5/B4
	improved vegetation index	SAV12	$\frac{\text{B5} + 0.5 - ((\text{ B5} + 0.5)^2 - 2 \times (\text{B5} - \text{B4}))^{0.5}}{2 \times (\text{B5} - \text{B4})^{0.5}}$
	optimized Simple Ratio vegetation Index	MSR	$(B5/B4 - 1)/(B5/B4 + 1)^{0.5}$
	karst terrain factor 1	KT1	$\begin{array}{c} 0.304 \times B2 + 0.279 \times B3 + \\ 0.474 \times B4 + 0.559 \times B5 + \\ 0.508 \times B6 + 0.186 \times B7 \end{array}$
	principal component 1—factor A	PC1-A	$\begin{array}{c} 0.054 \times B2 + 0.130 \times B3 + \\ 0.143 \times B4 + 0.595 \times B5 + \\ 0.709 \times B6 + 0.321 \times B7 \end{array}$
	principal component 1—factor B	PC1-B	$            0.140 \times B2 + 0.242 \times B3 + \\             0.313 \times B4 + 0.262 \times B5 + \\             0.739 \times B6 + 0.457 \times B7                                 $
	principal component 1—factor P	PC1-P	$\begin{array}{c} 0.056 \times B2 + 0.079 \times B3 + \\ 0.127 \times B4 - 0.845 \times B5 - \\ 0.490 \times B6 - 0.143 \times B7 \end{array}$

Table 1. Cont.

# 3. Research Method

In this study, Sentinel 1A, Sentinel 2A, Landsat 8 OLI, and DEM were utilized as data sources. The research was conducted based on 210 FCI *Pinus yunnanensis* forest plots in Yunnan Province. A total of 12 algorithms from 7 types, including Bagging learners, Boosting learners, neural network, linear-based learners, kernel-based learners, kNN, and Stacking ensemble, were constructed for the study. Among these algorithms, the Stacking ensemble algorithm integrates 11 algorithms, namely Bagging learner, Boosting learner,

neural network, linear learner, kernel function learner, and kNN, from 6 types of learners. The Stacking ensemble then selected the optimal algorithm from these 6 types of learners as its meta-model. The research workflow is illustrated in Figure 3.



Figure 3. Research technical route.

#### 3.1. Variable Selection

Linear stepwise regression (LSR) is a commonly employed variable selection method in remote sensing-based estimation of forest AGB [20,48]. It aims to identify useful variables from a pool of redundant data. LSR introduces characteristic variables into the model and conducts significance tests one by one to identify statistically significant variables that fall within a specified range (p < 0.05). These selected variables form the final combination in LSR [48]. To ensure the accuracy of the estimation model, collinearity between the selected trait variables is assessed for each variable combination. Collinearity refers to the presence of strong correlations between predictor variables, which can introduce bias in the model estimation. In this context, the variance inflation factor (VIF) is used as a measure of collinearity between the trait variables. A VIF threshold of 10 is commonly applied to detect and address collinearity issues [38,49]. Variables with VIF values exceeding this threshold are considered to have high collinearity and may be excluded from the model to mitigate bias.

## 3.2. Model Construction

Grid search is a common tuning method that optimizes model performance by searching for the best combination of hyperparameters in a predefined hyperparameter grid [50]. This research adopts 5 cross-validation search algorithms to find the optimal combination of parameters, parameters CARET used to wrap the default value (https://topepo.github.io/caret/ accessed on 9 July 2023). The hyperparameters of the algorithm are shown in Table 2.

Algorithm	R Packages	Hyperparameters Tuned
Random Forest	Random Forest	mtry (Randomly Selected Predictors)
Quantile Random Forest	Quantreg Forest	mtry (Randomly Selected Predictors)
Gaussian Process	kernlab	none
Stochastic Gradient Boosting	gbm, plyr	n.trees (Boosting Iterations), interaction.depth (Max Tree Depth), shrinkage (Shrinkage), n.minobsinnode (Min. Terminal Node Size)
Support Vector Machines with Radial Basis Function Kernel	kernlab	Sigma (Sigma), C (Cost)
Bayesian Regularized Neural Networks	brnn	neurons
Quantile Regression Neural Network	qrnn	n.hidden (Hidden Units), penalty (Weight Decay), bag (Bagged Models)
Bayesian Ridge Regression	monomvn	None
Gaussian Process	kernlab	None
Elasticnet	elasticnet	fraction (Fraction of Full Solution), lambda (Weight Decay)
K-nearest neighbor	none	k (Neighbors)
Extreme gradient boosting	xgboost, plyr	nrounds, max_depth, eta, gamma, subsample, colsample_bytree, rate_drop, skip_drop, min_child_weight
Stacking ensemble	caretEnsemble, mlbench, caret	-

Table 2. Algorithms for tuning hyperparameters.

notes: "-", the hyperparameters are determined by the meta-model of Stacking ensemble algorithm.

# 3.3. Model Evaluation

Using the sample independence test to calculate its coefficient of determination ( $R^2$ ) and root mean square error (RMSE) metrics for model evaluation.

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} (y_{i} - \hat{y}_{i})^{2}}{\sum_{i=1}^{n} (y_{i} - \overline{y}_{i})^{2}}$$
(3)

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}}$$
(4)

where *n* is the number of sample observations, and  $y_i$  is the actual value;  $\hat{y}_i$  is the estimated value, and  $\overline{y}_i$  is the mean of the observed sample.

# 4. Results

#### 4.1. Variable Selection

A total of 224 feature variables (35 vegetation indices, 19 single bands, 168 texture features, VV, VH) were selected using the LSR method. Figure 4 shows the selection results, and it can be seen that there was no strong multi-collinearity between variables. Eight variables were selected to participate in the model construction, including three vegetation indices (S2PSSRa, S2NDVI45, S2REP), four texture features (L8\_b7\_EN7, L8\_b5\_CR7, L8\_b6\_SM5, L8\_b3\_CR5) and one terrain factor (Slop). The results showed S2PSSRa had the highest correlation with forest AGB. The figure also showed that the biomass samples were skewed to the right.



**Figure 4.** Correlation of variables (the blue diagonal line is the data distribution frequency map, the lower left corner is the scatterplot relationship between the two variables, and the upper right corner is the Pearson correlation coefficient between the two variables. Note: "\*" stands for  $p \le 0.05$ , "\*\*" stands for  $p \le 0.01$ , "\*\*" stands for  $p \le 0.001$ ).

## 4.2. Model Evaluation

Figures 5 and 6 were the fitting diagrams and AGB maps of the models, respectively. Figure 5 shows that the goodness of fit at a single algorithm level was XGBoost > BRNN > EN > GP > RF > QRF > BR > QRNN > SGB > SVM-RBF > kNN. In addition, it can be seen from the AGB maps that the distribution of high and low AGB values was almost consistent across all algorithms. XGBoost, BRNN, RF, and QRF had good sensitivity and the range of AGB estimates was reasonable with good robustness, among which QRNN, GP, and EN had more outliers and poor robustness. The SVM-RBF, k-NN, and SGB also had good robustness, but their sensitivity was less than that of XGBoot, BRNN, RF, and QRF algorithms, and high-value underestimation was evident. For this purpose, the XGBoot algorithm was chosen as a meta-model of the Stacking ensemble to integrate 11 algorithms. It can be seen that the stacking integrated algorithm has the highest estimation accuracy  $(R^2 = 0.61, RMSE = 39.34 Mg/ha)$ . Based on the XGBoost-stacking algorithm integrated with 11 algorithms, R<sup>2</sup> increased by 0.12 and RMSE decreased by 4.53 Mg/ha compared to the single optimal algorithm, XGBoost algorithm. Good sensitivity and robustness were also reflected in the AGB maps of the XGBoost-stacking algorithm. However, it can be seen from Figure 5 that although the integrated stacking algorithm mitigated high-value underestimation to some extent, all algorithms showed high-value underestimation in forest AGB at around 100 Mg/ha, especially the SVM-RBF algorithm. However, for low

values, the BRNN algorithm was more practical and could estimate low AGB values. In conclusion, XGBoost, BRNN, and XGBoost-stacking algorithms have a good application prospect in AGB estimation of remote sensing, and high-value underestimation was still an important factor affecting the accuracy of AGB estimation.



Figure 5. AGB fitting scatter diagram of 12 algorithms.



Figure 6. AGB maps of 12 algorithms in study areas.

# 5. Discussion

Theoretically, Sentinel 1A data have a strong ability to penetrate forest stands and may well reflect the vertical structure of forest stands, which correlates well with AGB [4]. However, in this study, the backscattering coefficient of Sentinel 1A was not selected by LSR variables to participate in the model construction. This may be because the correlation between the backscatter coefficient of SAR and the AGB of the forest is easily affected by complex terrain. High-precision terrain correction for SAR images can improve image quality, and the accuracy of the DEM greatly affects image quality [51–54]. For example, in Vatandaşlar's research, 1-m resolution data from the Shuttle Radar Topography Mission (SRTM) was used to perform terrain correction on SAR data, and a good estimation effect was obtained in mountainous landscapes [55]. However, Pinus yunnanensis in Yunnan Province generally grows in plateau mountains and mid-altitude valleys at an elevation of 250–3500 m, and the terrain is relatively complex. In this study, 30 m DEM was used to perform terrain correction on Sentinel 1A data. The DEM data are rough, which may be the reason for the correlation between Sentinel 1A and AGB. Besides, this study employed the LSR variable selection method to identify variables with high linear correlation to forest AGB. To control collinearity between variables and mitigate estimation instability, a VIF threshold of <10 was applied. Additionally, variable selection methods, such as LASSO [56] and Boruta [17] may help prevent the exclusion of SAR data with low linear correlation coefficients from the analysis. If the variable selection process is avoided, directly combining SAR and optical variables to build a model can improve the accuracy of AGB estimation. Moreover, existing studies have demonstrated that integrating texture measurements from SAR images with forest auxiliary information can further enhance the AGB estimation in mountainous forest remote sensing [57,58], which should be improved in future research. Furthermore, from the perspective of the importance of the variables, the correlation of the vegetation index was higher than that of the texture features, and the forest structure of coniferous forests was simpler than that of deciduous forests, reflecting the higher correlation of the vegetation index in forests with simpler forest structure [4,13].

The XGBoost algorithm demonstrated excellent fitting and robustness in this study, primarily due to its inclusion of regularization techniques and pruning strategies. These components play a crucial role in controlling the complexity of the model and mitigating the risk of overfitting [59]. XGBoost has integrated the prediction results of all the basic learners. Furthermore, during the learning and storage process, XGBoost has utilized various methods to address the challenge of missing values encountered at different nodes. Additionally, XGBoost has provided support for custom loss functions and incorporated regular terms into the objective function to simplify the learning model and improve the overall learning effectiveness. As a result, the XGBoost-based algorithm has proven to be effective for estimating forest AGB. BRNN were more robust than QRNN because they can control the number of effective parameters for training through a Bayesian criterion and are insensitive to the architecture of the network [60]. It has been observed that the algorithm employing the regularization techniques had promising prospects for the estimation of forest AGB through remote sensing. Particularly in the areas characterized by high forest heterogeneity, the algorithm exhibited better robustness and superior fitting performance. In addition, the Stacking integrated algorithm had the highest estimation accuracy in this study, the  $R^2$  was 0.61, and the RMSE was 39.34 Mg/ha. The estimation performance of the Stacking algorithm was largely dependent on the performance of the meta-model. This research compared the performance of 11 machine learning algorithms and selected the optimal algorithm of XGBoost as the meta-model for the Stacking ensemble. The XGBoost-stacking algorithm not only combined the excellent performance of the XGBoost algorithm and Stacking algorithm but also integrated the excellent performance of six kinds of learning algorithms of the basic model. Therefore, the XGBoost-stacking algorithm could significantly improve the remote sensing estimation performance of forest AGB by improving the generalization ability of the model [36]. Although the accuracy was improved, the increase was not large. Besides, if we can screen from the model level, eliminate redundant model variables, and select truly useful model variables for Stacking ensemble, the estimation accuracy may be better improved. Furthermore, only the fusion strategy of stacking algorithms was considered in this study. The fusion strategies of other algorithms, such as blending ensemble learning [61] and averaging algorithms, can be explored in future research. The saturation effect was a common phenomenon in AGB estimation based on optical remote sensing [4]. The saturation phenomenon was serious according to the scatter distribution of predicted and observed values, and there was a serious underestimation for higher AGB values.

Compared with similar studies, it can be seen that the AGB of Pinus yunnanensis was underestimated when AGB was greater than 100 Mg/ha, which led to a lower accuracy of estimation in this study. The saturation effect threshold of the data in this study was lower than the AGB saturation of pine forests in Zhejiang Province of 159 Mg/ha, as reported by Zhao et al. (2016) [62]. This may be due to the AGB samples showing a right-skewed distribution that most of the values clustered in the low-value and fewer AGB samples with a high value, indicating a certain degree of forest heterogeneity, which was also an important reason for the serious underestimation of the higher values (Figure 4). At the same time, the structure and habitat of the forest were more complex in Yunnan Province compare to Zhejiang Province. Compared to Tang et al. (2022) and Chen et al. (2022) [41,44] remote sensing assessment of forest AGB in Yunnan Province, the remote sensing estimation accuracy of AGB in this study was still lower. If the hierarchical estimation of *Pinus yunnanensis* forest could be implemented according to the characteristics of topography and phenology, the data saturation phenomenon could be reduced, and the estimation accuracy would be improved. Due to the limitation of the sample size, stratified estimation was not possible in this study, but it could be supplemented in future studies. In addition, LiDAR and high-resolution optical remote sensing data can provide the vertical distribution information of vegetation and richer spectral characteristics, and introducing them into the remote sensing estimation of Pinus yunnanensis AGB may overcome and reduce the data saturation effect and improve the estimation accuracy [63,64]. Some studies showed that temperature factors had a significant influence on coniferous forests in Yunnan Province. Thus, adding environmental factors, such as temperature, would reduce the phenomenon of underestimation and overestimation [17,65,66]. In addition, the combination of geostatistical methods could be used for the next step study, as it can also reduce the spatial heterogeneity of forest images, as well as the data saturation effect, which may further improve the estimation accuracy [67].

### 6. Conclusions

Research shows that among the 12 algorithms, the fitting performance rank was XGBoost-Stacking > XGBoost > BRNN > EN > GP > RF > QRF > BR > QRNN > SGB > SVM-RBF > kNN. The stacking ensemble, with XGBoost as the meta-model, achieved the highest estimation accuracy, with an  $R^2$  value of 0.61 and an RMSE of 39.34 Mg/ha. When compared to the single optimal XGBoost algorithm, the stacking ensemble showed an improvement of 0.12 in  $R^2$  and a reduction of 4.53 Mg/ha in RMSE. The Stacking, XGBoost, BRNN, RF, and QRF models had a good sensitivity to AGB, which obtained a reasonable AGB estimation range and good robustness. On the contrary, the QRNN, GP, and EN models had more outlier data and poor robustness. SVM-RBF, k-NN, and SGB algorithms also had good robustness, but their sensitivity was worse than that of XGBoot, BRNN, RF, and QRF algorithms, and many of larger values were underestimated. All algorithms underestimated the values when the forest AGB > 100 Mg/ha, especially the SVM-RBF algorithm. However, for lower values, the BRNN algorithm was more practical and could estimate lower AGB with more accuracy. In conclusion, XGBoost, BRNN, and XGBoot-Stacking had a good application prospect in AGB estimation of remote sensing, and high-value underestimation was still an important factor affecting the accuracy of AGB estimation.

In the optical remote sensing-based estimation of *Pinus yunnanensis* forest AGB in highly heterogeneous areas of Yunnan Province, the saturation effect was still an important factor affecting the accuracy. XGBoot-Stacking could improve the estimation accuracy and selecting an appropriate algorithm to participate in the AGB remote sensing estimation is the key step to reducing the estimation errors. This study could provide a reference for selecting suitable algorithms and data sources in AGB estimation.

**Author Contributions:** Conceptualization: T.H., X.Z., Y.W., G.O. and C.X.; data curation: T.H. and Z.L.; formal analysis: T.H. and C.X.; funding acquisition: C.X.; investigation: F.Z. and C.Z.; methodology: T.H. and G.O.; project administration: C.X., C.Z. and F.Z.; resources: C.X., F.Z. and C.Z.; software: T.H. and Z.L.; supervision: C.X.; validation: T.H. and G.O.; visualization: T.H.; writing original draft: T.H.; writing review and editing: T.H., Y.W., X.Z. and H.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the comprehensive survey of carbon sinks in typical areas of China by Kunming Natural Resources Survey Center of China Geological Survey (DD20220877) and the Expert Workstation of Yunnan Province of China grant number 2018IC100).

Data Availability Statement: No new data is created.

Conflicts of Interest: The authors declare no conflict of interest.

#### References

- 1. Qian, C.; Qiang, H.; Wang, F.; Li, M. Estimation of Forest Aboveground Biomass in Karst Areas Using Multi-Source Remote Sensing Data and the K-DBN Algorithm. *Remote Sens.* **2021**, *13*, 5030. [CrossRef]
- 2. Bonan, G.B. Forests and climate change: Forcings, feedbacks, and the climate benefits of forests. *Science* **2008**, *320*, 1444–1449. [CrossRef]
- 3. McRoberts, R.; Tomppo, E. Remote sensing support for national forest inventories. *Remote Sens. Environ.* **2007**, *110*, 412–419. [CrossRef]
- Lu, D.; Chen, Q.; Wang, G.; Liu, L.; Li, G.; Moran, E. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. *Int. J. Digit. Earth* 2014, 9, 63–105. [CrossRef]
- 5. Avitabile, V.; Baccini, A.; Friedl, M.A.; Schmullius, C. Capabilities and limitations of Landsat and land cover data for aboveground woody biomass estimation of Uganda. *Remote Sens. Environ.* **2012**, *117*, 366–380. [CrossRef]
- Su, H.; Liu, H.; Heyman, W.D. Automated Derivation of Bathymetric Information from Multi-Spectral Satellite Imagery Using a Non-Linear Inversion Model. *Mar. Geod.* 2008, 31, 281–298. [CrossRef]
- Welle, T.; Aschenbrenner, L.; Kuonath, K.; Kirmaier, S.; Franke, J. Mapping Dominant Tree Species of German Forests. *Remote Sens.* 2022, 14, 3330. [CrossRef]
- 8. Caughlin, T.T.; Barber, C.; Asner, G.P.; Glenn, N.F.; Bohlman, S.A.; Wilson, C.H. Monitoring tropical forest succession at landscape scales despite uncertainty in Landsat time series. *Ecol. Appl.* **2021**, *31*, e02208. [CrossRef]
- 9. Cooper, S.; Okujeni, A.; Pflugmacher, D.; van der Linden, S.; Hostert, P. Combining simulated hyperspectral EnMAP and Landsat time series for forest aboveground biomass mapping. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *98*, 102307. [CrossRef]
- 10. Huang, X.; Ziniti, B.; Torbick, N.; Ducey, M. Assessment of Forest above Ground Biomass Estimation Using Multi-Temporal C-band Sentinel-1 and Polarimetric L-band PALSAR-2 Data. *Remote Sens.* **2018**, *10*, 1424. [CrossRef]
- 11. Wang, X.; Liu, C.; Lv, G.; Xu, J.; Cui, G. Integrating Multi-Source Remote Sensing to Assess Forest Aboveground Biomass in the Khingan Mountains of North-Eastern China Using Machine-Learning Algorithms. *Remote Sens.* **2022**, *14*, 1039. [CrossRef]
- 12. Zhang, Y.; Wang, N.; Wang, Y.; Li, M. A new strategy for improving the accuracy of forest aboveground biomass estimates in an alpine region based on multi-source remote sensing. *GIScience Remote Sens.* **2023**, *60*, 2163574. [CrossRef]
- 13. Lu, D.; Chen, Q.; Wang, G.; Moran, E.; Batistella, M.; Zhang, M.; Vaglio Laurin, G.; Saah, D. Aboveground forest biomass estimation with Landsat and LiDAR data and uncertainty analysis of the estimates. *Int. J. For. Res.* **2012**, 2012, 1–16. [CrossRef]
- Araza, A.; de Bruin, S.; Herold, M.; Quegan, S.; Labriere, N.; Rodriguez-Veiga, P.; Avitabile, V.; Santoro, M.; Mitchard, E.T.A.; Ryan, C.M.; et al. A comprehensive framework for assessing the accuracy and uncertainty of global above-ground biomass maps. *Remote Sens. Environ.* 2022, 272, 112917. [CrossRef]
- 15. Shettles, M.; Temesgen, H.; Gray, A.N.; Hilker, T. Comparison of uncertainty in per unit area estimates of aboveground biomass for two selected model sets. *For. Ecol. Manag.* **2015**, *354*, 18–25. [CrossRef]
- 16. Tang, J.; Liu, Y.; Li, L.; Liu, Y.; Wu, Y.; Xu, H.; Ou, G. Enhancing Aboveground Biomass Estimation for Three Pinus Forests in Yunnan, SW China, Using Landsat 8. *Remote Sens.* **2022**, *14*, 4589. [CrossRef]
- 17. Huang, T.; Ou, G.; Wu, Y.; Zhang, X.; Liu, Z.; Xu, H.; Xu, X.; Wang, Z.; Xu, C. Estimating the Aboveground Biomass of Various Forest Types with High Heterogeneity at the Provincial Scale Based on Multi-Source Data. *Remote Sens.* **2023**, *15*, 3550. [CrossRef]
- 18. Gao, Y.; Lu, D.; Li, G.; Wang, G.; Chen, Q.; Liu, L.; Li, D. Comparative Analysis of Modeling Algorithms for Forest Aboveground Biomass Estimation in a Subtropical Region. *Remote Sens.* **2018**, *10*, 627. [CrossRef]

- 19. Ronoud, G.; Fatehi, P.; Darvishsefat, A.A.; Tomppo, E.; Praks, J.; Schaepman, M.E. Multi-Sensor Aboveground Biomass Estimation in the Broadleaved Hyrcanian Forest of Iran. *Can. J. Remote Sens.* **2021**, *47*, 818–834. [CrossRef]
- Jiang, F.; Kutia, M.; Ma, K.; Chen, S.; Long, J.; Sun, H. Estimating the aboveground biomass of coniferous forest in Northeast China using spectral variables, land surface temperature and soil moisture. *Sci. Total Environ.* 2021, 785, 147335. [CrossRef] [PubMed]
- Prasad, A.M.; Iverson, L.R.; Liaw, A. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems* 2006, 9, 181–199. [CrossRef]
- Francke, T.; López-Tarazón, J.A.; Schröder, B. Estimation of suspended sediment concentration and yield using linear models, random forests and quantile regression forests. *Hydrol. Process.* 2008, 22, 4892–4904. [CrossRef]
- 23. Li, Y.; Li, M.; Wang, Y. Forest Aboveground Biomass Estimation and Response to Climate Change Based on Remote Sensing Data. *Sustainability* **2022**, *14*, 14222. [CrossRef]
- 24. Geng, L.; Che, T.; Ma, M.; Tan, J.; Wang, H. Corn Biomass Estimation by Integrating Remote Sensing and Long-Term Observation Data based on Machine Learning Techniques. *Remote Sens.* **2021**, *13*, 2352. [CrossRef]
- Tian, Y.; Zhang, Q.; Huang, H.; Huang, Y.; Tao, J.; Zhou, G.; Zhang, Y.; Yang, Y.; Lin, J. Aboveground biomass of typical invasive mangroves and its distribution patterns using UAV-LiDAR data in a subtropical estuary: Maoling River estuary, Guangxi, China. *Ecol. Indic.* 2022, 136, 108694. [CrossRef]
- Güneralp, İ.; Filippi, A.M.; Randall, J. Estimation of floodplain aboveground biomass using multispectral remote sensing and nonparametric modeling. *Int. J. Appl. Earth Obs. Geoinf.* 2014, 33, 119–126. [CrossRef]
- Li, L.; Zhou, B.; Liu, Y.; Wu, Y.; Tang, J.; Xu, W.; Wang, L.; Ou, G. Reduction in Uncertainty in Forest Aboveground Biomass Estimation Using Sentinel-2 Images: A Case Study of Pinus densata Forests in Shangri-La City, China. *Remote Sens.* 2023, 15, 559. [CrossRef]
- Choubin, B.; Hosseini, F.S.; Fried, Z.; Mosavi, A. Application of Bayesian Regularized Neural Networks for Groundwater Level Modeling. In Proceedings of the 2020 IEEE 3rd International Conference and Workshop in Óbuda on Electrical and Power Engineering (CANDO-EPE), Budapest, Hungary, 18–19 November 2020; pp. 000209–000212.
- 29. Garg, D.; Mishra, A. Bayesian regularized neural network decision tree ensemble model for genomic data classification. *Appl. Artif. Intell.* **2018**, *32*, 463–476. [CrossRef]
- Alvarez-Mendoza, C.I.; Guzman, D.; Casas, J.; Bastidas, M.; Polanco, J.; Valencia-Ortiz, M.; Montenegro, F.; Arango, J.; Ishitani, M.; Selvaraj, M.G. Predictive Modeling of Above-Ground Biomass in Brachiaria Pastures from Satellite and UAV Imagery Using Machine Learning Approaches. *Remote Sens.* 2022, 14, 5870. [CrossRef]
- 31. Hans, C. Elastic Net Regression Modeling with the Orthant Normal Prior. J. Am. Stat. Assoc. 2011, 106, 1383–1393. [CrossRef]
- Halme, E.; Pellikka, P.; Mõttus, M. Utility of hyperspectral compared to multispectral remote sensing data in estimating forest biomass and structure variables in Finnish boreal forest. Int. J. Appl. Earth Obs. Geoinf. 2019, 83, 101942. [CrossRef]
- Ghosh, S.S.; Khati, U.; Kumar, S.; Bhattacharya, A.; Lavalle, M. Gaussian process regression-based forest above ground biomass retrieval from simulated L-band NISAR data. Int. J. Appl. Earth Obs. Geoinf. 2023, 118, 103252. [CrossRef]
- 34. Wu, N.; Liu, G.; Wuyun, D.; Yi, B.; Du, W.; Han, G. Spatial-Temporal Characteristics and Driving Forces of Aboveground Biomass in Desert Steppes of Inner Mongolia, China in the Past 20 Years. *Remote Sens.* **2023**, *15*, 3097. [CrossRef]
- Breunig, F.M.; Galvão, L.S.; Dalagnol, R.; Dauve, C.E.; Parraga, A.; Santi, A.L.; Della Flora, D.P.; Chen, S. Delineation of management zones in agricultural fields using cover–crop biomass estimates from PlanetScope data. *Int. J. Appl. Earth Obs. Geoinf.* 2020, *85*, 102004. [CrossRef]
- 36. Zhang, Y.; Ma, J.; Liang, S.; Li, X.; Liu, J. A stacking ensemble algorithm for improving the biases of forest aboveground biomass estimations from multiple remotely sensed datasets. *GIScience Remote Sens.* **2022**, *59*, 234–249. [CrossRef]
- 37. Li, X.; Zhang, M.; Long, J.; Lin, H. A Novel Method for Estimating Spatial Distribution of Forest Above-Ground Biomass Based on Multispectral Fusion Data and Ensemble Learning Algorithm. *Remote Sens.* **2021**, *13*, 3910. [CrossRef]
- Liu, K.; Wang, J.; Zeng, W.; Song, J. Comparison and Evaluation of Three Methods for Estimating Forest above Ground Biomass Using TM and GLAS Data. *Remote Sens.* 2017, 9, 341. [CrossRef]
- 39. Hou, H.; Chen, X.; Li, M.; Zhu, L.; Huang, Y.; Yu, J. Prediction of user outage under typhoon disaster based on multi-algorithm Stacking integration. *Int. J. Electr. Power Energy Syst.* 2021, 131, 107123. [CrossRef]
- López-Serrano, P.M.; López-Sánchez, C.A.; Álvarez-González, J.G.; García-Gutiérrez, J. A Comparison of Machine Learning Techniques Applied to Landsat-5 TM Spectral Data for Biomass Estimation. *Can. J. Remote Sens.* 2016, 42, 690–705. [CrossRef]
- Tang, C.Q.; Han, P.-B.; Li, S.; Shen, L.-Q.; Huang, D.-S.; Li, Y.-F.; Peng, M.-C.; Wang, C.-Y.; Li, X.-S.; Li, W.; et al. Species richness, forest types and regeneration of Schima in the subtropical forest ecosystem of Yunnan, southwestern China. *For. Ecosyst.* 2020, 7, 1–19. [CrossRef]
- 42. Sun, H.; Wang, J.; Xiong, J.; Bian, J.; Jin, H.; Cheng, W.; Li, A.; García Mozo, H. Vegetation Change and Its Response to Climate Change in Yunnan Province, China. *Adv. Meteorol.* **2021**, 2021, 1–20. [CrossRef]
- Tamme, R.; Hiiesalu, I.; Laanisto, L.; Szava-Kovats, R.; Pärtel, M. Environmental heterogeneity, species diversity and co-existence at different spatial scales. J. Veg. Sci. 2010, 35, 843–848. [CrossRef]
- Chen, H.; Qin, Z.; Zhai, D.-L.; Ou, G.; Li, X.; Zhao, G.; Fan, J.; Zhao, C.; Xu, H. Mapping Forest Aboveground Biomass with MODIS and Fengyun-3C VIRR Imageries in Yunnan Province, Southwest China Using Linear Regression, K-Nearest Neighbor and Random Forest. *Remote Sens.* 2022, 14, 5456. [CrossRef]

- 45. Deng, X.; Huang, B.; Wen, Q.; Hua, C.; Tao, J. A research on the distribution of *Pinus yunnanensis* forest in Yunnan Province. *J. Yunnan Univ.* **2013**, *35*, 843–848. [CrossRef]
- 46. Liu, L. Model regression analysis of Pinus yunnanensis biomass in northwest Yunnan. Shandong For. Sci. Technol. 2015, 5–9, 34.
- 47. Zhengqi, G.; Xiaoli, Z.; Yueting, W. Ability evaluation of coniferous forest aboveground biomass inversion using Sentinel-2A multiple characteristic variables. *J. Beijing For. Univ.* 2020, 42, 27–38. [CrossRef]
- 48. Jiang, F.; Kutia, M.; Sarkissian, A.J.; Lin, H.; Long, J.; Sun, H.; Wang, G. Estimating the Growing Stem Volume of Coniferous Plantations Based on Random Forest Using an Optimized Variable Selection Method. *Sensors* 2020, 20, 7248. [CrossRef] [PubMed]
- 49. O'brien, R.M. A Caution Regarding Rules of Thumb for Variance Inflation Factors. Qual. Quant. 2007, 41, 673–690. [CrossRef]
- Shekar, B.H.; Dagnew, G. Grid Search-Based Hyperparameter Tuning and Classification of Microarray Cancer Data. In Proceedings of the 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), Gangtok, India, 25–28 February 2019; pp. 1–8.
- Sinha, S.; Jeganathan, C.; Sharma, L.K.; Nathawat, M.S. A review of radar remote sensing for biomass estimation. *Int. J. Environ. Sci. Technol.* 2015, 12, 1779–1792. [CrossRef]
- Le Toan, T.; Quegan, S.; Davidson, M.W.J.; Balzter, H.; Paillou, P.; Papathanassiou, K.; Plummer, S.; Rocca, F.; Saatchi, S.; Shugart, H.; et al. The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sens. Environ.* 2011, 115, 2850–2860. [CrossRef]
- 53. Khudinyan, M. The Use of Remotely Sensed Data for Forest Biomass Monitoring: A Case of Forest Sites in North-Eastern Armenia. Ph.D. Thesis, NOVA Information Management School (NIMS), Lisbon, Portugal, 2019.
- Shimada, M. Ortho-Rectification and Slope Correction of SAR Data Using DEM and Its Accuracy Evaluation. *IEEE J. Sel. Top.* Appl. Earth Obs. Remote Sens. 2010, 3, 657–671. [CrossRef]
- 55. Vatandaşlar, C.; Abdikan, S. Carbon stock estimation by dual-polarized synthetic aperture radar (SAR) and forest inventory data in a Mediterranean forest landscape. *J. For. Res.* **2021**, *33*, 827–838. [CrossRef]
- 56. Naik, P.; Dalponte, M.; Bruzzone, L. Prediction of Forest Aboveground Biomass Using Multitemporal Multispectral Remote Sensing Data. *Remote Sens.* 2021, *13*, 1282. [CrossRef]
- 57. Nuthammachot, N.; Askar, A.; Stratoulias, D.; Wicaksono, P. Combined use of Sentinel-1 and Sentinel-2 data for improving above-ground biomass estimation. *Geocarto Int.* 2020, 37, 366–376. [CrossRef]
- 58. Forkuor, G.; Benewinde Zoungrana, J.-B.; Dimobe, K.; Ouattara, B.; Vadrevu, K.P.; Tondoh, J.E. Above-ground biomass mapping in West African dryland forest using Sentinel-1 and 2 datasets—A case study. *Remote Sens. Environ.* 2020, 236, 111496. [CrossRef]
- Chen, T.; Guestrin, C. XGBoost. In Proceedings of the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
- 60. Burden, F.; Winkler, D. Bayesian Regularization of Neural Networks. Methods Mol. Biol. 2008, 458, 25–44. [PubMed]
- 61. Wu, T.; Zhang, W.; Jiao, X.; Guo, W.; Alhaj Hamoud, Y. Evaluation of stacking and blending ensemble learning methods for estimating daily reference evapotranspiration. *Comput. Electron. Agric.* **2021**, *184*, 106039. [CrossRef]
- 62. Zhao, P.; Lu, D.; Wang, G.; Wu, C.; Huang, Y.; Yu, S. Examining Spectral Reflectance Saturation in Landsat Imagery and Corresponding Solutions to Improve Forest Aboveground Biomass Estimation. *Remote Sens.* **2016**, *8*, 469. [CrossRef]
- 63. Luo, S.; Wang, C.; Xi, X.; Pan, F.; Peng, D.; Zou, J.; Nie, S.; Qin, H. Fusion of airborne LiDAR data and hyperspectral imagery for aboveground and belowground forest biomass estimation. *Ecol. Indic.* **2017**, *73*, 378–387. [CrossRef]
- 64. Xiaoyi, W.; Huabing, H.; Peng, G.; Caixia, L.; Congcong, L.; Wenyu, L. Forest Canopy Height Extraction in Rugged Areas with ICESat/GLAS Data. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4650–4657. [CrossRef]
- 65. Ni, J. Impacts of climate change on Chinese ecosystems: Key vulnerable regions and potential thresholds. *Reg. Environ. Change* **2010**, *11*, 49–64. [CrossRef]
- Dakhil, M.A.; Xiong, Q.; Farahat, E.A.; Zhang, L.; Pan, K.; Pandey, B.; Olatunji, O.A.; Tariq, A.; Wu, X.; Zhang, A.; et al. Past and future climatic indicators for distribution patterns and conservation planning of temperate coniferous forests in southwestern China. *Ecol. Indic.* 2019, 107, 105559. [CrossRef]
- Sun, S.; Cao, Z.; Zhu, H.; Zhao, J. A Survey of Optimization Methods from a Machine Learning Perspective. *IEEE Trans. Cybern.* 2019, 50, 3668–3681. [CrossRef] [PubMed]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.