

Article

Mapping Forest Canopy Height via Self-Attention Multisource Feature Fusion and a Blending-Based Heterogeneous Ensemble Model

Jing Tian ^{1,2,*} , Pinghao Zhang ¹ , Pinliang Dong ³ , Wei Shan ² , Ying Guo ² , Dan Li ¹, Qiang Wang ¹ and Xiaodan Mei ¹

¹ Heilongjiang Institute of Technology, College of Surveying and Mapping Engineering, Harbin 150050, China; zhangpinghao@hljit.edu.cn (P.Z.); lidan@hljit.edu.cn (D.L.); wangqiang1@hljit.edu.cn (Q.W.); meixiaodan@hljit.edu.cn (X.M.)

² Consulting and Design Institute, Northeast Forestry University, Harbin 150040, China; shanwei@nefu.edu.cn (W.S.); ying-guo@nefu.edu.cn (Y.G.)

³ Department of Geography and the Environment, University of North Texas, Denton, TX 76203, USA; pinliang.dong@unt.edu

* Correspondence: tianjing@hljit.edu.cn

Highlights

What are the main findings?

- The introduction of the self-attention (SA) mechanism significantly enhances the ability to mine potential complementary information between multisource features and further strengthens the expression of high-dimensional features.
- Three heterogeneous models as base-learners, namely the deep neural network (DNN), extreme gradient boosting (XGBoost), and residual network (ResNet), were adopted, combined with random forest as the meta-learner, to construct an SA-Blending heterogeneous ensemble framework, effectively improving the prediction performance of the model.

What are the implications of the main findings?

- The introduction of a self-attention mechanism significantly improved the performance of each base model, validating the effectiveness of complementary correlation information between cross-modal data in feature fusion.
- The SA-Blending heterogeneous ensemble model effectively compensates for the limitations of single models in feature learning and nonlinear representation by integrating the structural advantages of different base models, demonstrating better generalization and adaptability and offering a promising solution for accurate forest canopy height estimation.

Abstract

The accuracy of forest canopy height estimation is crucial for forest resource management and ecosystem carbon sequestration. However, existing approaches often face limitations in effectively integrating multisource remote sensing data, feature representation, and model learning strategies. To enhance the prediction performance of the model in complex terrain and multisource data environments, this study comprehensively used ICESat-2/ATLAS photon point clouds, Sentinel-2/MSI multispectral imagery, and SRTM-DEM to construct a remote sensing-driven multisource feature system, which eliminated redundant interference using permutation feature importance analysis. Additionally, a self-attention (SA) mechanism was introduced to strengthen high-dimensional feature representation.



Received: 16 December 2025

Revised: 2 February 2026

Accepted: 15 February 2026

Published: 18 February 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Three heterogeneous models, incorporating deep neural network (DNN), extreme gradient boosting (XGBoost), and residual network (ResNet), were independently applied for forest canopy height estimation and were further used as base learners, with a random forest as the meta-learner, and an SA-Blending heterogeneous ensemble model that combines a blending technique with an SA mechanism was proposed to enhance the accuracy of forest canopy height estimation. To evaluate the SA optimization strategy and the role of multisource fusion, this study used the original features, SA-optimized features, and multisource fusion features (i.e., the concatenation and fusion of original features and self-attention mechanism features) as inputs to comprehensively compare the performance of each single model and the integrated model. The results show that: (1) The self-attention mechanism significantly improves the prediction performance of heterogeneous models. Compared with original features inputs, the R^2 of DNN (SA-Only) and XGBoost (SA-Only) increased to 0.706 and 0.708, respectively, and the RMSE decreased to 1.691 m and 1.613 m. Although the R^2 for ResNet (SA-Only) decreased slightly to 0.699 and the RMSE increased to 1.712 m, the overall impact was not significant. (2) Under the condition of multisource fusion feature input, DNN+SA, XGBoost+SA, and ResNet+SA all demonstrated higher fitting accuracy and stability, verifying the enhancing effect of the SA mechanism on the association expression of multisource information. (3) The SA-Blending model achieved the best overall performance, with R^2 of 0.766 and RMSE of 1.510 m. It outperformed individual models and the SA-optimized model in terms of overall accuracy, stability, and robustness. The results can provide technical support for high-precision forest canopy height mapping and are of great significance for ecological monitoring applications.

Keywords: forest canopy height; self-attention mechanism; blend ensemble learning; multisource features; multi-source remote sensing

1. Introduction

Forest canopy height is a key structural parameter for indicating forest carbon storage, evaluating forest ecological services, and formulating effective forest management strategies to address global climate change [1]. High-accuracy, spatially continuous forest canopy height mapping based on multisource remote sensing provides a fundamental basis for forest carbon stock estimation at regional and global scales, and serves as an essential input for biomass modeling, carbon cycle analysis, and studies of ecosystem processes [2,3]. In addition, reliable forest canopy height information enables the long-term monitoring of forest structural dynamics, facilitates the assessment of forest disturbance and recovery processes, and supports data-driven decision making in sustainable forest management [4,5]. Thus, high-precision canopy height products not only advance scientific understanding of forest carbon interactions but also offer important practical value for forest resource monitoring, ecosystem conservation, and climate change mitigation strategies.

The application of remote sensing technology in forest surveying has greatly improved the efficiency of forest management and minimized ground data collection efforts [6,7]. Estimating forest canopy height using satellite data primarily involves synthetic aperture radar (SAR) and light detection and ranging (LiDAR) data [8]. For SAR, forest canopy height estimation mainly relies on interferometric SAR (InSAR) technology [9], which is a multifaceted process influenced by various factors, including the time intervals and distance between two data acquisitions, the canopy-penetrating capabilities of the operating wavelength, and field conditions. Furthermore, while microwave radar functions effectively in all weather conditions, including rain and cloudy, its signal is influenced by topographical relief and

is susceptible to saturation. These restrictions hinder the capacity of SAR interferometry to produce precise measurements of forest canopy height [10–12]. As for airborne LiDAR, also known as airborne laser scanning (ALS), it can accurately estimate tree height but is unfeasible for large-scale applications due to the high cost of the data [13]. In contrast, spaceborne LiDAR benefits from the satellite platform's high orbit and expansive perspective, making it a feasible solution for collecting broad-scale forest canopy height data. One of the most recent spaceborne LiDAR systems is the advanced topographic laser altimeter system (ATLAS) on board the Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2), which was launched on 15 September 2018. For altimetry observations, ICESat-2/ATLAS used a photon-counting technology system with a footprint diameter of about 14 m, allowing it to produce a variety of geophysical products over various surface types [14]. The ATL08 is one of the products that provides canopy height percentiles along 100 m segments using returned photons classified as ground, noise, canopy, or top of canopy, but in cases of high forest coverage, the existing ATL08 data products are difficult to apply directly to the extraction of forest canopy height [15]. Since ICESat-2 data are obtained along transects at discrete intervals, coupling spatially continuous optical data, such as Sentinel-2 imagery, with ATLAS-derived vegetation attributes is an effective method for estimating spatially continuous forest canopy heights [16,17]. Compared with other optical data, Sentinel-2 data provide moderate spatial resolution, and the red-edge bands of their multispectral imager (MSI) exhibit heightened sensitivity to vegetation growth, thereby providing more precise vegetation growth information [17].

Thus far, several studies have demonstrated that cooperative modeling based on LiDAR data and optical remote sensing can effectively improve the accuracy of forest canopy height inversion at regional or large scales [18,19]. Commonly used methods include data-driven and physical models [20]. Approaches can be classified as either parameter-reliant or parameter-independent. Parametric methods are primarily represented by traditional statistical regression, whereas nonparametric methods encompass a range of machine learning techniques, including random forests, support vector machines, and extreme gradient boosting [3]. Although those methods have been developed and perform well at estimating forest canopy height, each algorithm has its own scope of application, and none performs well in all situations. Thus, to improve prediction accuracy, compared with selecting a single model to estimate forest canopy height, ensemble learning is a good alternative method that combines the advantages of multiple base learners, each a functionally independent classifier or regressor. Even when an individual base learner yields an inaccurate prediction, the remaining models can compensate for this error and enhance the overall estimation performance [3,21]. Blending, as a representative of the heterogeneous ensemble techniques, can effectively reduce model noise by combining multiple models to enhance generalization. Compared with a single model, ensemble learning usually achieves higher fitting accuracy [22].

However, the existing studies on estimating forest canopy height still have significant deficiencies in feature construction and model design. The majority of recent research uses either a single model for fitting or straightforward feature concatenation. This approach cannot fully capture the complex, highly nonlinear relationships among multisource features [23], such as photon point clouds, spectral information, terrain factors, and forest age. Due to the lack of effective modeling of the complementarity and interaction mechanisms between cross-modal features, traditional methods often fail to comprehensively reflect the true physical properties of forest canopy structure, failing to fully utilize the overall information contained in multiple sources of data, thereby limiting further improvement in the accuracy of forest canopy height estimation.

This study proposes an SA-blending heterogeneous ensemble framework for high-precision forest canopy height estimation, which effectively combines the strengths of self-attention mechanism and blending ensemble algorithm. Multisource remote sensing data, including ICESat-2/ATLAS photon point clouds, Sentinel-2/MSI multispectral imagery, and DEM data from the Shuttle Radar Topography Mission (SRTM) were integrated to construct a multisource feature system. A permutation-based feature importance analysis was employed to quantitatively evaluate and filter candidate features, thereby eliminating redundant information and optimizing the model's input structure. Subsequently, a self-attention (SA) mechanism was introduced to deeply explore the latent correlations and complementarities among multisource features, enhancing the discriminative power and high-dimensional representational capacity of the feature set. For model development, three heterogeneous learning algorithms—Deep Neural Network (DNN), Extreme Gradient Boosting (XGBoost), and Residual Neural Network (ResNet)—were independently used to estimate forest canopy height. These three models were then integrated as base learners, while Random Forest (RF) was adopted as the meta-learner to construct the SA-Blending heterogeneous ensemble framework. To systematically evaluate the effectiveness of the feature optimization strategy and the SA-Blending model, three types of feature inputs—original features, SA-optimized features, and cross-modal fusion features (combining original and SA-optimized features)—were separately fed into the three single models and the SA-Blending model, for comparative accuracy assessment. The objective is to achieve high-precision forest canopy height mapping within the study area and to provide methodological support for forest parameter inversion under complex stand structural conditions.

2. Materials and Methods

2.1. Study Area

The study area is the Mohe Forestry Bureau (121°11'22" to 123°16'10"E, 52°16'58" to 53°32'46"N), situated in the northwestern region of Mohe City within the Greater Khingan Mountains, Heilongjiang Province, China. The area has a cool-temperate continental monsoon climate, with annual precipitation of 350–500 mm, predominantly in July and August. The landscape mostly has low mountains and undulating hills, displaying intricate geomorphological characteristics with a general elevation gradient that diminishes from south to north. The region possesses extensive forest resources, featuring a forest coverage rate of 92.4%, and is characterized by primary natural forests typical of cold-temperate coniferous ecosystems. The predominant coniferous species are *Larix gmelinii* and *Pinus sylvestris* var. *mongolica*, whereas the principal broadleaved species are *Betula platyphylla* and *Populus davidiana*. The understory vegetation includes temperate deciduous shrubs, cultivated crops, marshy grasslands, and swamp meadow communities [24]. Figure 1 depicts the geographical location of the study region.

2.2. Data Acquisition and Processing

2.2.1. ICESat-2/ATLAS Data and Processing

ICESat-2 provides 23 products describing the surface characteristics (ATL00–ATL24, without ATL05 and ATL18) [25,26]. The ATL08 is the land and vegetation data product derived from ATL03 photon data following noise filtering and photon classification procedures. It provides terrain and canopy structural metrics computed along-track within 100 m segments, including estimates such as relative canopy height and relative terrain elevation [14,25]. The level-3a ATL08 data from 2019, in Hierarchical Data Format Version 6 (HDF-6), were used in this study to ensure the chronological validity of the verified data and the coverage of the study area. The data were freely downloaded from the National

Snow and Ice Data Center (<https://nsidc.org/data/ICESat-2/> accessed on 10 July 2024), which are delivered in the WGS84 ellipsoidal and vertical datum [27,28].

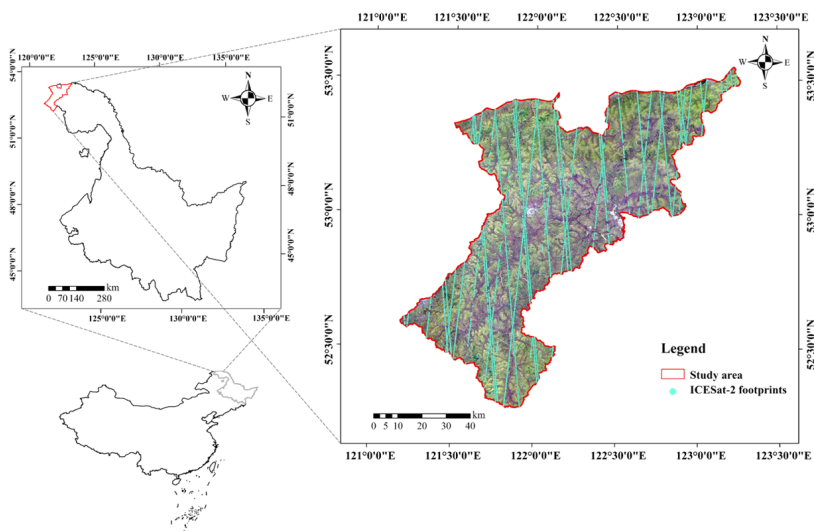


Figure 1. Map of the study area and distribution of ICESat-2 data. The true-color image (right) consists of the Red, Green, and Blue bands of Landsat-8 OLI. The blue dots represent the footprint in the study area. The red boundary (Mohe district) in the upper-left corner is the study area in Heilongjiang, and the bottom-left image is the scaled map of China.

To acquire more precise canopy information, the ATL08 data were filtered to exclude low-quality or evident errors data, primarily by removing cloud-affected photons using the cloud flag parameter (`cloud_flag_atm` > 0) and excluding high uncertainty ICESat-2 photons based on the canopy height uncertainty parameter (`h_canopy_uncertainty` = 3.4028235×10^{38}), and retaining only nighttime observations (`night_flag` = 1) to mitigate solar radiation background noise and enhance signal reliability. Subsequently, using field-measured tree heights in the study area as a benchmark, photons with elevations ranging from 2 to 60 m were retained to mitigate the impact of outliers on subsequent modeling. In summary, a total of 28,369 valid ICESat-2/ATL08 footprint data points were acquired, and five relative height (RH) index parameters, namely “`h_max_canopy`,” “`h_canopy`,” “`h_median_canopy`,” “`h_min_canopy`,” and “`h_mean_canopy`”, were extracted (Table 1).

Table 1. ICESat-2/ATLAS related indicator parameters information.

Standard Name	Description
<code>h_max_canopy</code>	RH100, maximum of individual absolute canopy heights within segment.
<code>h_canopy</code>	RH98, 98% height of all the individual canopy relative heights for the segment above the estimated terrain surface.
<code>h_median_canopy</code>	RH50, the median of individual relative canopy heights within segment.
<code>h_min_canopy</code>	RHmin, the minimum of relative individual canopy heights within segment.
<code>h_mean_canopy</code>	RHmean, mean of individual relative canopy heights within segment.

2.2.2. Sentinel-2 Image Preprocessing and Spectral Variable Calculation

While ICESat-2 cannot provide wall-to-wall coverage, combining it with continuous remote sensing data, such as those from Sentinel-2, proved to be an effective method for mapping forest canopy heights [3,29]. Sentinel-2 was launched in June 2015, onboard multispectral instrument (MSI), which detects a broad electromagnetic spectrum (443 nm to 2202 nm) in 13 bands with a swath width of 290 km and a spatial resolution of 10 m (four visible and near-infrared bands), 20 m (six red-edge and shortwave infrared band),

and 60 m (three atmospheric correction bands) [30]. Among them, the three red-edge bands are widely used to monitor vegetation due to their sensitivity to plant growth [31].

In this study, the Sentinel-2A image (path/row:123/23) in the vegetation growing season from 22 September 2019 was acquired from the European Copernicus Data Centre (<https://browser.dataspace.copernicus.eu/> accessed on 18 January 2024) as a Level-2A product with the UTM/WGS84 projection and less than 5% cloud cover. The Sentinel-2A image was atmospherically corrected by the Sen2Cor module, and the pixels were transformed to account for surface reflectance. To enhance the consistency of multisource remote sensing data, the 20 m and 60 m spatial resolution bands of Level-2A data products were resampled to 10 m using bilinear interpolation in SNAP 13.0.0 software, and saved as GeoTIFF files. Subsequently, the study area was extracted using a vector boundary as a mask in Esri's ArcGIS 10.8 software. Given the heterogeneity of canopy structures and the diverse tree species within the study area, ten vegetation indices, which represent the physiological, biochemical, and structural characteristics of the vegetation canopy, and eight spectral reflectance bands were extracted as spectral features to establish correlations between ICESat-2 and the mapping (Table 2).

Table 2. List of spectral features from Sentinel-2 used in this study.

Type (Source)	Indices	Reference
Normalized difference vegetation index (NDVI)	$(\text{NIR} - \text{Red})/(\text{NIR} + \text{Red})$	[32]
Enhanced vegetation index (EVI)	$2.5 \times (\text{NIR} - \text{Red})/(\text{NIR} + 6 \times \text{Red} - 7.5 \times \text{Blue} + 1)$	[33]
Green normalized difference vegetation index (GNDVI)	$(\text{NIR} - \text{Green})/(\text{NIR} + \text{Green})$	[34]
Normalized difference red edge (NDRE)	$(\text{NIR} - \text{RedEdge1})/(\text{NIR} + \text{RedEdge1})$	[35]
Difference vegetation index (DVI)	$\text{NIR} - \text{Red}$	[36]
Modified soil adjusted vegetation index (MSAVI)	$(2 \times \text{NIR} + 1 - \sqrt{(2 \times \text{NIR} + 1)^2 - 8 \times (\text{NIR} - \text{Red})})/2$	[37]
Red-edge normalized difference vegetation (NDVI _{re})	$(\text{RedEdge1} - \text{Red})/(\text{RedEdge1} + \text{Red})$	[38]
Ratio vegetation index (RVI)	NIR/Red	[39]
Normalized difference moisture index (NDMI)	$(\text{NIR} - \text{SWIR1})/(\text{NIR} + \text{SWIR1})$	[40]
Global environment monitoring index (GEMI)	$\left(\frac{2 \times (\text{NIR}^2 - \text{Red}^2) + 1.5 \times \text{NIR} + 0.5 \times \text{Red}}{\text{NIR} + \text{Red} + 0.5} \right) \times \left(1 - 0.25 \times \left(\frac{2 \times (\text{NIR}^2 - \text{Red}^2) + 1.5 \times \text{NIR} + 0.5 \times \text{Red}}{\text{NIR} + \text{Red} + 0.5} \right) \right) - \frac{\text{Red} - 0.125}{1 - \text{Red}}$	[41]
Spectral reflectance	Band2 (Blue), Band3 (Green), Band4 (Red), Band5 (Red edge1). Band 6 (Red edge2), Band 7 (Red edge3), Band 8 (NIR) Band 8a (Narrow NIR)	

Note: red, green, blue, NIR, SWIR1, and RedEdge1 correspond to the reflectivity of bands B4, B3, B2, B8, B11, and B5, respectively.

2.2.3. STRM DEM

Terrestrial factors, such as elevation, slope, and aspect, are important biogeographical parameters that partially control vegetation distribution and growth [42]. The Shuttle Radar Topography Mission (SRTM) obtained elevation data on a near-global scale to generate the most complete high-resolution digital topographic database of Earth. This data will be processed to produce a rectified, terrain-corrected mosaic of approximately 80% of Earth's land surface topography (between 60°N and 56°S) at 30 m resolution [43]. We derived the elevation, slope, and aspect from the SRTM 1 Arc-Second Global elevation data at a resolution of 1 arc-second and then resampled it to a 10 m resolution through the cubic convolution interpolation method (Cubic Convolution) in ArcGIS Pro 3.0 software, as the auxiliary data for the ICESat-2/ATLAS and Sentinel-2A training model [44], as shown in Figure 2.

2.2.4. Field Data Collection and Preprocessing

This study utilized the Forest Resources Second Class Survey from the Mohe Forestry Bureau, which is carried out every 10 years, as its fundamental ground truth data for obtaining reference data on forest canopy height and forest age. The data collection was conducted from July to October 2019, coinciding with the growing season of forest

vegetation, and established a database of forest resources. The sub-compartment is the basic unit of forest resource planning and design investigation, statistics, and management, which contains information such as the area of all kinds of forest land, the information on tree species, the characteristic features of the dominant forest type, including average tree height, diameter at breast height, canopy density, and forest age, etc. The average tree height of each sub-compartment was recorded as the average height of 3–5 dominant trees, and individual tree heights were measured in situ using a laser rangefinder. Forest age was primarily determined by the dominant tree species in the main stand layer, while for planted forests, stand age was calculated as the sum of the planting year interval and the seedling age at the time of planting, whereas for natural forests, stand age was defined as the average age of the dominant tree species. Age classes were subsequently assigned according to the stand origin and mean age and were recorded using standardized categorical codes. In this study, the average tree height and forest age were extracted from a total of 36,786 sub-compartment of the Forest Resources Second Class Survey data and converted into vector layer data in the UTM/WGS84 projection, which served as field measurements to validate the generated forest canopy height map. The spatial distribution of the vector sub-compartment data is shown in Figure 3.

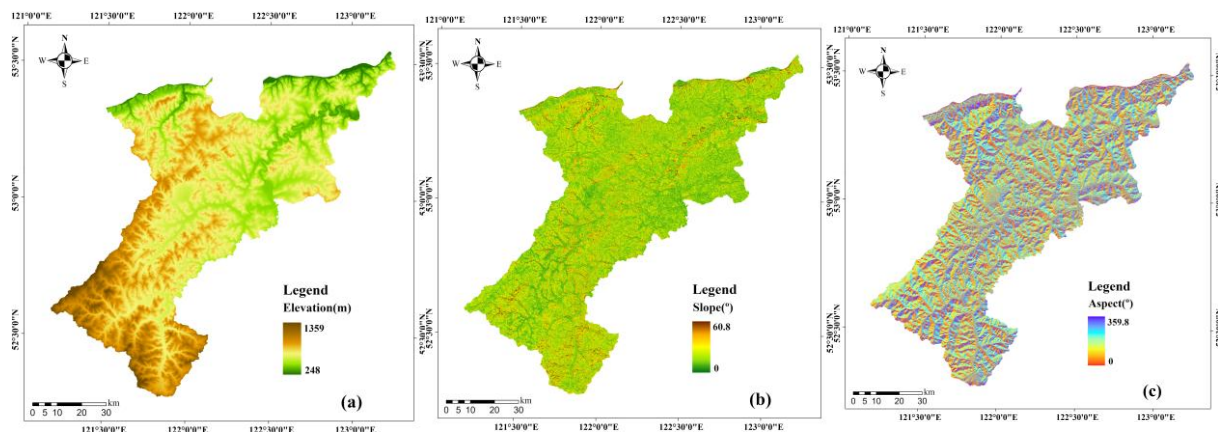


Figure 2. SRTM data. Here, (a), (b), and (c) represent the elevation, slope, and aspect data extracted by SRTM, respectively.

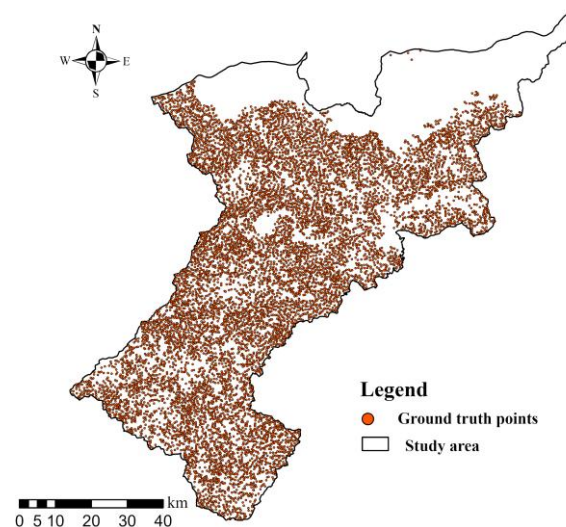


Figure 3. Ground truth points. A red dot represents a sub-compartment that is a business unit with the same attributes planned within the scope of the forest compartment.

2.2.5. Forest and Non-Forest

To consider the effects of non-forest areas, we used a 30 m resolution forest-type classification map of the study area derived from the Landsat 8-OLI imagery in 2018 [24]. Despite a one-year difference between the acquisition of Sentinel-2A images and the forest-type data, no significant deforestation or disturbances transpired in the research area during this timeframe, and alterations in forest cover type and non-forest land type were negligible. Consequently, this study deems the effect of this temporal delay on the research outcomes to be insignificant. ArcGIS Pro software was employed in data processing to differentiate between forest and non-forest areas; all classifications, excluding coniferous forest, broadleaved forest, and mixed forest, were consistently consolidated into non-forest areas. A reclassification tool generated a mask file for non-forest land, while bilinear interpolation resampled the 30 m resolution to 10 m, supplying essential data for later forest canopy height mapping to mitigate the impact of non-forest areas.

2.2.6. Sample Dataset Preprocessing

Multisource remote sensing features for the study area were derived through resampling, vegetation index computation, and spectral reflectance feature selection. Upon obtaining the ICESat-2/ATLAS photon point cloud data, all multisource datasets were spatially overlaid and matched, then partitioned into 10 m × 10 m grid cells using a fish net layout method. Based on the geolocation information of ATL08 photons, multisource feature data corresponding to each grid cell were extracted. After removing empty and invalid records, a total of 9667 valid point-scale samples were obtained. This sample dataset comprises five ICESat-2/ATLAS relative height metrics, eight original spectral reflectance bands, ten vegetation indices, forest age, and three topographic factors, resulting in a set of 27 multisource features along with field-measured canopy height. These datasets served as training and validation samples for forest canopy height modeling and accuracy assessment.

2.3. Method

The workflow of the method adopted in this paper is shown in Figure 4.

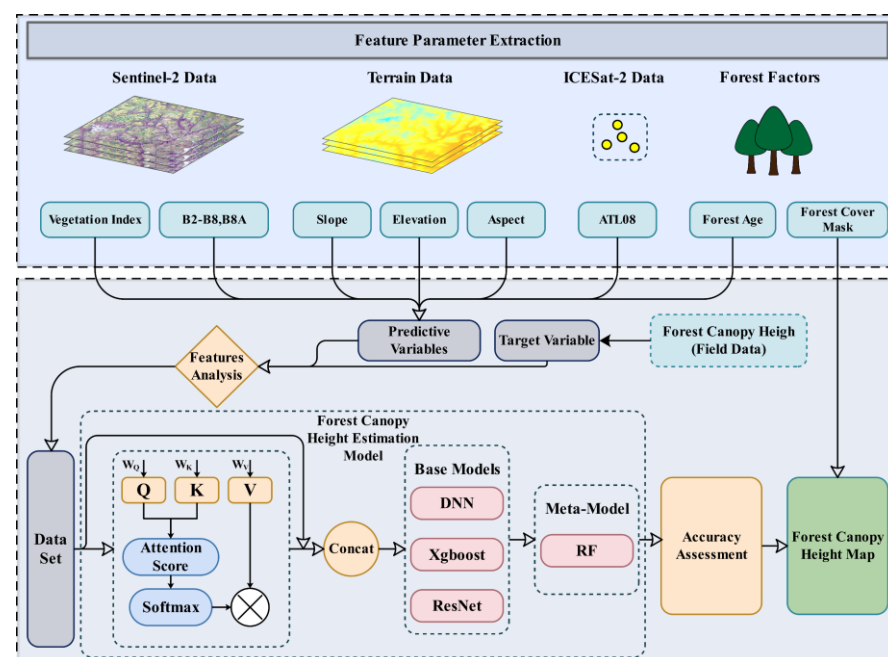


Figure 4. Workflow for estimating forest canopy height.

2.3.1. Feature Selection of Predictor Variables

This study utilized a permutation-based feature importance method to optimize the feature-set inputs for forest canopy height estimation and improve model adaptability in complex geographic environments. This model-checking technique evaluates each variable's contribution to the predicted outcome by randomly permuting feature values and analyzing the resulting changes in model performance. A higher importance score indicates that the feature is more essential to the model's predictive performance. Utilizing the Scikit-learn library in Python3.10, an important threshold of 10% was set to quantitatively assess 27 candidate variables. Features exhibiting a contribution degree within the lowest 10%, deemed low-relevance for forest canopy height or redundant between features, were subsequently eliminated to decrease the number of features and dimensions, thereby improving the comprehension of the relationship between features and eigenvalues. Ultimately, the most significant characteristics for the estimation task were preserved to enhance the discriminative capacity of the feature set and improve the model's generalization performance [45].

2.3.2. Multisource Feature Fusion with Self-Attention Mechanism

Self-attention is a deep learning mechanism that captures internal dependencies within a sequence through dynamically assigned weights, with its primary objective being the modeling of complex relational structures among sequence elements [46]. The structure of the self-attention mechanism is shown in Figure 5, which primarily consists of three vectors: query (Q), key (K), and value (V). The three vectors are obtained by multiplying the input time series by three matrices.

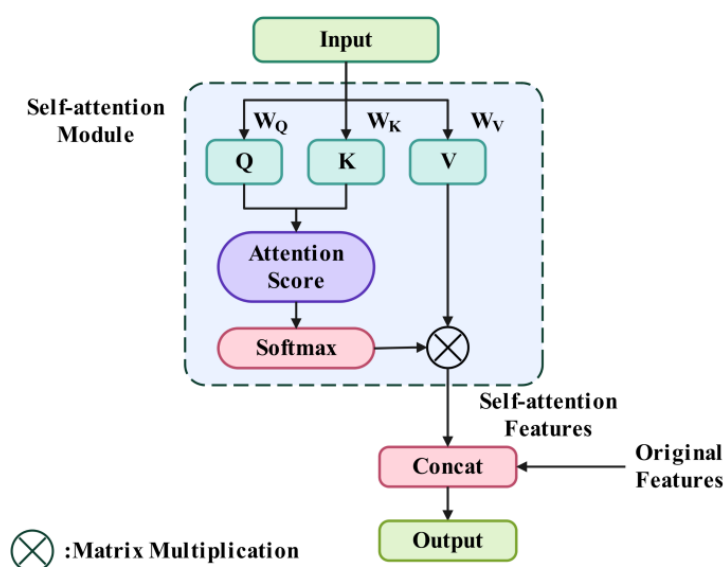


Figure 5. Self-attention mechanism module. The Self-attention module's output features are concatenated with the original features to further integrate local dependencies with the original semantic information, thereby achieving cross-modal data fusion. The fused features are then used as the input to the SA-Blending heterogeneous model.

When applied to input feature images, the self-attention mechanism quantifies the similarity or importance among feature elements, enabling the adaptive refinement and enhancement of feature representations, thereby substantially improving the model's capability in feature extraction and structural modeling [47]. Typically, self-attention is implemented using scaled dot-product attention, in which attention weights are computed exclusively within the sequence to characterize intrinsic relationships among its elements.

The aggregated global weighted features are then fused with the original input feature map via channel-wise addition to produce the final module output. The computation is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

where QK^T denotes the dot-product matrix between the query (Q) and key (K), representing the attention intensity of each query over all keys; $\frac{1}{\sqrt{d_k}}$ is a scaling factor used to stabilize gradient propagation and prevent numerical instability caused by excessively large dot-product values, and softmax is the normalization function.

To deeply excavate the complex interdependencies among features, the multisource feature set—screened based on permutation-based feature importance ranking—was jointly fed into the feature space of the self-attention (SA) mechanism. This design enables the model to autonomously identify and enhance the representations of critical features while effectively suppressing the influence of redundant information. Through dynamic weight allocation and semantic relationship modeling, the SA module integrates complementary and heterogeneous information across multiple dimensions, including spectral and index features, environmental drivers, and horizontal forest structure. This unified representation facilitates cross-modal data fusion, which subsequently serves as the input feature set for the proposed SA-Blending heterogeneous model.

2.3.3. SA-Blending Heterogeneous Ensemble Model of Multisource Fusion

Deep neural network

A fully connected deep neural network (DNN) model architecture was designed that consists of one input layer, eleven hidden layers, and one output layer. The input layer receives preprocessed feature vectors, with the number of nodes being consistent with the feature dimension of the training samples. The fully connected hidden layers consist of two layers of 512 neurons, two layers of 256 neurons, five layers of 128 neurons, and two layers of 64 neurons. The hidden layers introduce a nonlinear transformation through the Rectified Linear Unit (ReLU) activation function and apply Dropout (dropout rate 0.1) after each layer to prevent overfitting. During training, the Adam optimizer is used to accelerate convergence, and the loss function is the mean squared error (MSE), aiming to minimize the sum of the squared errors between the predicted and the true values. This network systematically abstracts features layer by layer and compresses dimensions, ultimately mapping the information to a single output node for regression-based prediction of forest canopy height [48,49].

Extreme gradient boosting

Extreme Gradient Boosting (XGBoost) is an efficient ensemble algorithm optimized based on Gradient Boosting Decision Trees (GBDTs) [50]. In this study, to enhance the model's generalization capability, XGBoost incorporates L1/L2 regularization and pruning strategies, along with row-wise random sampling, effectively controlling model complexity and preventing overfitting. In terms of computational optimization, XGBoost uses second-order Taylor expansion to refine the loss function, thereby increasing the precision of gradient updates [51].

Residual network

The ResNet network employs a hybrid architecture integrating hierarchical feature extraction with deep residual learning. The architecture designed for this study comprises an input layer, a one-dimensional convolutional layer, three residual modules, a fully connected layer, and an output layer. The initial feature extraction stage applies a 7×1 one-

dimensional convolutional kernel (Conv1D) to perform spatial feature mapping, followed by batch normalization (Batch Normalization, BN), ReLU activation, and max pooling operations, which collectively mitigate data noise and achieve preliminary dimensionality reduction [52]. In the deep feature learning module, a hierarchical configuration of three residual blocks is implemented with a stacking pattern of [2]. Relative to ResNet18, the proposed model moderately reduces both network depth and channel width, preserving representational capacity while enhancing computational efficiency. Each residual block incorporates two successive convolutional, normalization, and activation operations, and utilizes skip connections to enable identity mapping, thereby facilitating gradient propagation and strengthening feature representation. At the regression stage, adaptive average pooling is applied to standardize the feature tensor into a fixed dimension. Subsequently, a four-layer fully connected network with progressively decreasing neuron counts (512, 256, 128, and 64) performs nonlinear transformations, extending beyond the single-layer design of conventional ResNet to ensure comprehensive utilization of deep features during gradual dimension reduction. Model optimization is conducted using the Adam optimizer in conjunction with the mean squared error (MSE) loss function, balancing prediction accuracy and convergence speed. The final output consists of a single regression node for predicting forest canopy height.

Random forest

Random Forest (RF) is a nonparametric statistical estimation method based on decision tree ensembles [53,54]. In the blending ensemble model architecture, RF is used as a meta-learner to integrate predictions from heterogeneous base models (DNN, XGBoost, and ResNet). By leveraging the robustness of RF in handling nonlinear dependencies among features, it achieves the efficient integration of multiple model prediction information, thereby obtaining more stable and more accurate forest canopy height estimation results.

Blending algorithm

On the basis of obtaining multisource fusion features based on the self-attention mechanism, to address the uncertainty in the estimation results of a single model, the blending ensemble learning algorithm is introduced to construct the SA-Blending heterogeneous ensemble model for estimating forest canopy height.

Blending is an ensemble learning method that focuses on model fusion, where the core idea is to integrate the strengths of multiple base models through a hierarchical collaborative mechanism. By exploring the multisource feature space from different perspectives, while avoiding data leakage, it enhances overall prediction performance and improves predictive accuracy. This study employs a two-layer blending ensemble learning framework, comprising base learners and meta-learners. The base learner consists of three heterogeneous models, namely DNN, XGBoost, and ResNet, designed to extract complementary information from multisource features. The meta-learner uses a random forest as the fusion and inference module for the ensemble results. During model training, each base learner independently learns the nonlinear mapping relationship between input features and target variables on the training set and integrates its output predictions as the input features for the meta-model. Subsequently, the meta-learner is trained a second time using the synthesized features to capture the collaborative and differential information between base models, thereby achieving a higher level of generalized inference. The overall process structure of the blending ensemble algorithm is shown in Figure 6.

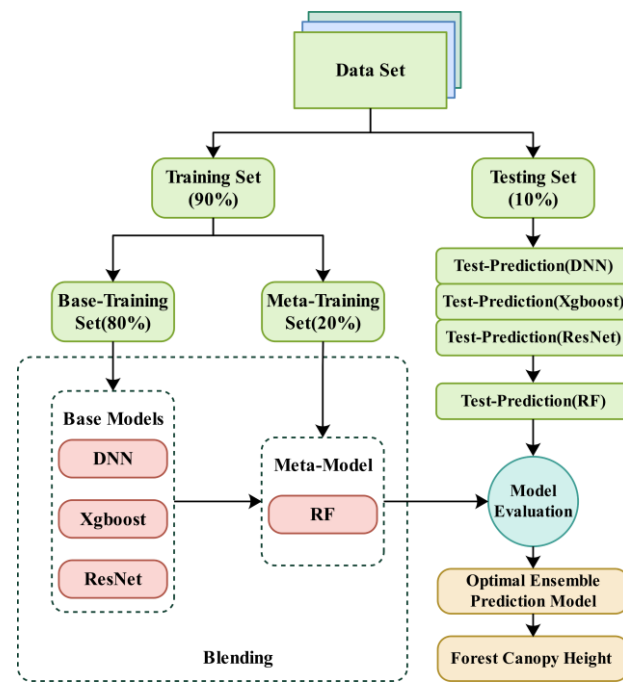


Figure 6. Blending ensemble learning algorithm framework diagram.

2.3.4. Model Performance Evaluation

Since the SA-Blending heterogeneous model is a regression problem, three evaluation metrics to assess the performance of each model were selected, namely the coefficient of determination (R^2), the root mean square error (RMSE), and the mean bias. Furthermore, the best-performing model was used to map forest canopy height in the study area. The calculation formula of the evaluation parameters is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (3)$$

$$Bias = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \quad (4)$$

where y_i represents the measured tree height, i.e., the true value; $\hat{y}_i$ denotes the tree height predicted based on the model; $\bar{y}$ is the average of all observed values; n is the number of validation samples. The model with higher R^2 value and lower RMSE and bias is preferred for forest canopy height estimation.

3. Results

3.1. Permutation Feature Importance Results

Figure 7 displays the feature selection outcomes from the permutation-based feature significance method. With an importance value of 0.9887, feature importance analysis revealed that forest age was the most important factor influencing the prediction of forest canopy height. While the significance values of the remaining characteristics varied from 0.2204 to 0.5539, elevation (0.6557) and Normalized Difference Moisture Index (NDMI, 0.6291) also showed significant influence. RVI, GEMI, and MSAVI, on the other hand, had importance scores of 0.220, 0.125, and 0.111, respectively. These values were all in the bottom 10%, indicating that they contributed little to the model's prediction and were thus

excluded. Twenty-four feature variables, including vegetation indices (NDVI, EVI, GNDVI, NDVIre, NDRE, NDMI, DVI), photon point cloud features (RH100, RH98, RH50, RHmean, RHmin), topographic factors (elevation, slope, aspect), multispectral reflectance bands (B2–B8, B8a), and forest age, were ultimately kept as model inputs after the aforementioned screening. Finally, a feature system that accounts for both representation capability and modeling efficiency was constructed and used to precisely estimate forest canopy height in the subsequent stage.

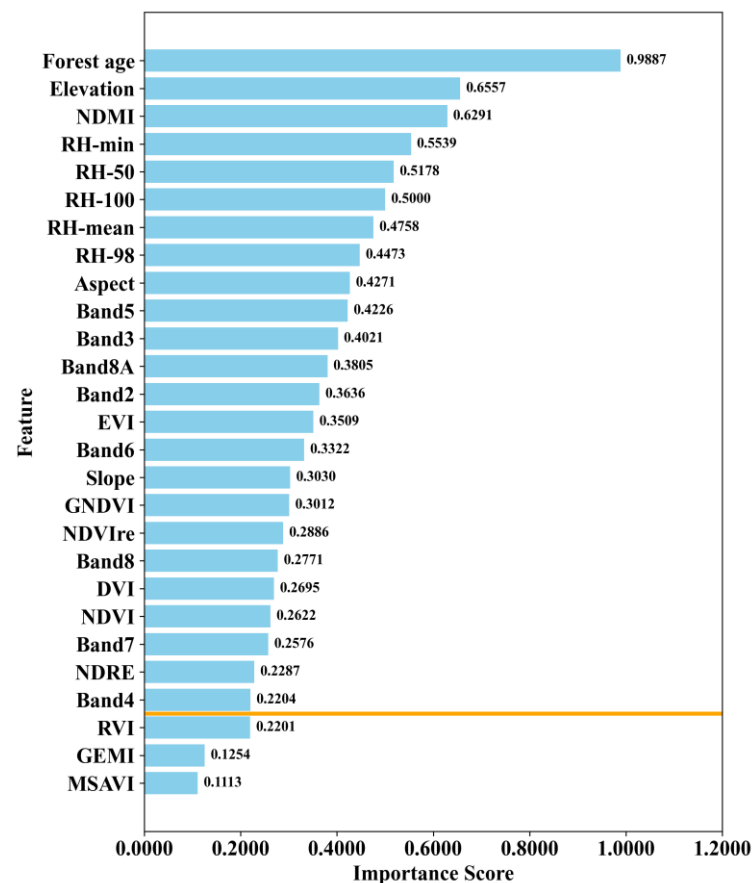


Figure 7. Results of feature importance ranking using permutation-based feature significance method. The orange line represents the dividing line for the last 10% of the characteristics.

3.2. Model Testing and Training Results

To improve the model's training efficacy on multisource features, using the StandardScaler 1.7.0 tool from the scikit-learn library, each dimension of the features is transformed into a standard normal distribution with a mean of 0 and a standard deviation of 1, so that the feature dimensions of the dataset have the same distribution scale to reduce the impact of the difference in numerical scales on model training and improve the stability of model training. Finally, a dataset of 9667 samples is generated, with each sample consisting of 24 independent variables (after feature elimination) and one dependent variable (measured tree height value).

Concurrently, to meet the sample partitioning requirements of the SA-Blending heterogeneous ensemble learning framework, the aforementioned original dataset was first randomly split into a training set and test set at a 9:1 ratio, with the test set independent of the training procedure and used solely for the final assessment of model generalization performance. On this basis, the training set was further stratified and divided into a base-model training subset (base-training set) and a meta-model training subset (meta-training set) at an 8:2 ratio. The base-training set was used to train all base learners, whereas the

predictor variables in the meta-training subset were derived from the outputs of each base learner applied to this subset; the corresponding ground-measured tree heights were retained as the response variable. On the independent test set, the trained base learners were applied to generate predictions for all test samples, yielding three sets of prediction results, each consistent with the number of test set samples. These prediction results were then concatenated along the feature dimension to construct the input feature matrix for the meta-learner, which was subsequently fed into the Random Forest meta-learner to produce the final fused predictions.

The model in this study was constructed and trained in Python. The deep learning model was built using the PyTorch (version 2.5.1) open source framework and integrated with NVIDIA CUDA 12.6 to support GPU-accelerated computing. To achieve the best model performance, this study used the optimal hyperparameters determined via grid search. The specific ranges of the hyperparameters are listed in Table 3. The key hyperparameter settings for DNN, XGBoost, ResNet, and RF listed in Table 4 are the optimal configurations for system optimization.

Table 3. The specific hyperparameter range.

Model	Hyperparameter	Search Range
DNN	batch_size	[200, 300]
	n_estimators	[200, 400]
XGBoost	learning_rate	[0.01, 0.1]
	min_child_weight	[2, 10]
	max_depth	[10, 30]
	Gamma	[5, 10]
	subsample	[0.1, 1]
	reg_alpha	[0.1, 0.3]
	reg_lambda	[0.1, 0.3]
ResNet	batch_size	[300, 500]
RF	n_estimators	[200, 600]
	min_samples_split	[4, 15]
	min_samples_leaf	[2, 8]
	max_features	{“sqrt”, “log2”}
	max_depth	[6, 50]

Table 4. Details of the four models’ hyperparameter setups.

Model	DNN	XGBoost	ResNet	RF
Optimal hyperparameters	epochs = 400 batch_size = 270 loss = ‘mse’ optimizer = Adam()	n_estimators = 200 gamma = 6 min_child_weight = 5 max_depth = 20 learning_rate = 0.01 subsample = 0.8 reg_alpha = 0.2 reg_lambda = 0.2	epochs = 50 batch_size = 400 loss = ‘mse’ optimizer = Adam()	n_estimators = 200 min_samples_split = 10 min_samples_leaf = 4 max_features = sqrt max_depth = 7

Note: n_estimators indicates the number of trees; reg_alpha and reg_lambda respectively represent the L1 and L2 regularization coefficients.

3.3. Comparative Analysis of Model Accuracy

During model training, each base learner is configured with three feature input schemes to systematically assess the effects of different feature combinations on model performance: (1) training exclusively with the original features to evaluate the model’s fundamental performance; (2) utilizing solely the features generated from the self-attention

mechanism (SA-Only) to evaluate the independent contribution of this mechanism in feature representation; and (3) employing a concatenated feature set that includes both the original features and the high-order features generated by the self-attention mechanism (+SA), representing a co-optimization of the original features and the attention mechanism (+SA). An SA-Blending heterogeneous ensemble framework was constructed by integrating three heterogeneous base learners (DNN, ResNet, and XGBoost) with a Random Forest meta-learner, as per the third feature configuration. In total, ten model configurations were systematically tested and compared to provide a comprehensive evaluation of predictive performance.

Figure 8a–c depicts the loss trajectories of the DNN, ResNet, and XGBoost models throughout the training process, according to different feature combinations. The loss values of all models decreased monotonically with increasing iterations, signifying that each model converged stably. Nonetheless, significant disparities emerge across models with different feature combinations in terms of convergence rate and curve smoothness, indicating differences in optimization efficiency and training stability. In Figure 8a, the convergence rate of the original DNN architecture is comparatively sluggish (green curve), whereas the loss curve of DNN (SA-only) exhibits considerable fluctuations during both the initial and final phases of training (orange curve), suggesting that dependence solely on self-attention mechanism features, devoid of the original feature constraints, weakens the model's training stability. In contrast, the DNN+SA structure exhibits a more rapid decrease in loss and smaller fluctuations in the loss curve during the first 100 epochs (blue curve), indicating that the self-attention module effectively guides gradient optimization and improves training performance. Figure 8b illustrates that the ResNet+SA architecture (blue curve) achieves a stable convergence phase after roughly 20 epochs. In comparison to the original ResNet model (green curve) and the ResNet (SA-only) structure (orange curve), its loss diminishes more rapidly and exhibits a smoother trajectory, indicating that the synergistic optimization of original and self-attention mechanism features markedly enhances the model's fitting capacity and bolsters the stability of gradient propagation during deep network training. In Figure 8c, the three structures XGBoost, XGBoost (SA-only), and XGBoost+SA all maintained low RMSE values throughout the training process. Among them, xgboost+SA achieved the fastest descent rate and the lowest RMSE value during iterations, demonstrating the model's good robustness and convergence stability. Figure 8d depicts the distribution of SHAP feature importance for the Blending framework utilizing RF as a meta-model about the prediction results of each base learner. From a comprehensive contribution perspective, ResNet's prediction results demonstrate the widest range of SHAP values, indicating its significant impact on the final model output. The SHAP value distributions of XGBoost and DNN show greater concentration, suggesting their contributions are more uniform, though with slightly diminished relative weights. The meta-model allocates greater weights to ResNet in the fusion process, while maintaining XGBoost and DNN as auxiliary information sources, illustrating the synergistic benefits of model fusion.

The analysis indicates that incorporating the self-attention (SA) mechanism significantly enhances the performance of all base models, with the most remarkable results observed when combined with original features, thereby validating the effectiveness of the potential complementarity and correlation between cross-modal information in feature fusion. The heterogeneous Blending model, guided by the self-attention mechanism, can effectively integrate the structural characteristics and representational advantages of various base models, resulting in significant information complementarity and performance improvements, while overcoming the limitations of singular modeling methods and enhancing the effectiveness of multisource forest canopy height prediction models.

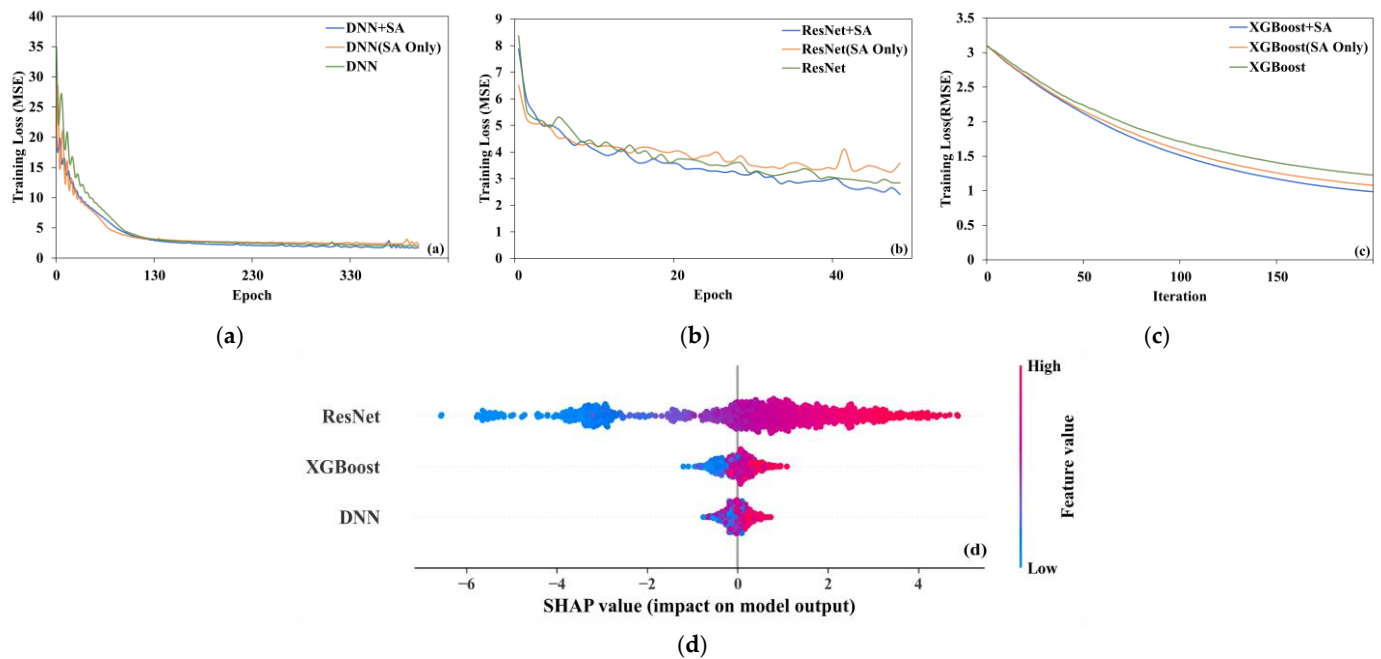


Figure 8. Model training diagram. (a), (b) and (c) respectively represent the loss change curves of the DNN, ResNet, and XGBoost models during the training stage. (d) is the SHAP feature contribution map of the Blending meta-model. The horizontal axis represents SHAP values, indicating the intensity and direction of the feature's influence on the model output. A positive SHAP value indicates a positive contribution of the feature to the prediction result, while a negative value indicates a negative impact. The vertical axis represents the names of different base models. The color of the points indicates the size of the prediction result in the test set, with blue representing low values and red representing high values.

3.4. Comparison of Model Performance Results

To comprehensively assess the forecasting accuracy of the ten models on the test dataset, we conducted a systematic analysis of their performance metrics, as detailed in Table 5. The DNN model attained an R^2 of 0.703, an RMSE of 1.702 m, and a Bias of 0.006, indicating that its predictions showed no significant systematic deviation. Upon integrating self-attention mechanism features, the DNN (SA-Only) model exhibited a slight enhancement in R^2 to 0.706; however, its bias significantly escalated to 0.301, indicating that dependence exclusively on SA features may lead to considerable systematic overestimation, presumably due to inadequate alignment between the SA representations and the model's intrinsic feature-learning architecture. Conversely, the DNN+SA configuration, which combines original features with attention-enhanced representations, yielded a significant performance improvement: R^2 rose from 0.703 to 0.727, RMSE declined from 1.702 m to 1.631 m, and Bias remained relatively low at 0.076. The results indicate that the self-attention mechanism significantly enhances the model's capacity to capture long-range dependencies in the feature space, therefore enhancing predictive accuracy and overall training stability.

The overall performance of ResNet parallels that of DNN, with an R^2 of 0.704. Nonetheless, its bias is negative at -0.133 , signifying a degree of systematic underestimating. Upon the incorporation of self-attention mechanism features, the R^2 of ResNet (SA-Only) diminishes to 0.699, while the bias shifts positively to 0.102, indicating that exclusive reliance on self-attention mechanism features fails to adequately align with the residual structure and may jeopardize the stability of local convolutional features. The ResNet+SA model demonstrates a notable enhancement in performance, with R^2 rising to 0.722, RMSE declining to 1.645 m, and bias markedly improved to 0.015. This signifies that the combined effect of

self-attention mechanism and residual architecture results in enhanced and more stable predictive ability.

Table 5. Test set coefficient of determination (R^2), the root mean square error (RMSE), and the mean Bias for forest canopy height. The best performance values are in bold.

Model	R^2	RMSE	Bias	Figure 9
DNN	0.703	1.702	0.006	Figure 9a
DNN (SA-Only)	0.706	1.691	0.301	Figure 9b
DNN+SA	0.727	1.631	0.076	Figure 9c
ResNet	0.704	1.698	−0.133	Figure 9d
ResNet (SA-Only)	0.699	1.712	0.102	Figure 9e
ResNet+SA	0.722	1.645	0.015	Figure 9f
XGBoost	0.693	1.729	0.046	Figure 9g
XGBoost (SA-Only)	0.708	1.686	0.071	Figure 9h
XGBoost+SA	0.733	1.613	0.059	Figure 9i
SA-Blending	0.766	1.510	0.067	Figure 9j

Compared with DNN and ResNet, the XGBoost model performs inferiorly, with an R^2 of 0.693 and an RMSE of 1.729 m, albeit with a minimal Bias of 0.046. The implementation of a self-attention mechanism markedly enhances performance, elevating R^2 to 0.708, which signifies a substantial improvement in the feature weighting capacity of the tree model. This pertains to the absence of explicit tools for representing feature interactions in tree models. The XGBoost+SA fusion model attains superior performance, evidenced by a notable rise in R^2 to 0.733 and a large reduction in RMSE to 1.613 m, outperforming all individual models. The Bias is maintained within a tolerable range (0.059), indicating that SA-guided enhanced feature representation can effectively offset the structural restrictions of tree models.

The SA-Blending heterogeneous integration model exhibited superior performance following the amalgamation of the strengths of each separate model. The performance metrics attained their optimal level in this study ($R^2 = 0.766$, RMSE = 1.510, Bias = 0.067). In comparison to the optimal single model XGBoost+SA, the R^2 of this model improved by 0.033, while the RMSE decreased by 0.103 m. This clearly illustrates that the incorporation of cross-modal features and self-attention mechanisms inside the Blending framework may efficiently extract and amalgamate complimentary information from several models, thereby substantially improving overall predictive performance. This outcome confirms the benefits and applicability of SA-Blending in the multisource features forest canopy height inversion problem.

Figure 9 shows the consistency scatter distribution between the predicted and observed values of forest canopy height by the 10 models. Overall, all three base models (XGBoost, DNN, and ResNet) exhibited varying degrees of systematic bias, either overestimating or underestimating. After introducing the self-attention mechanism, the scatter distribution of each model converged significantly toward the 1:1 baseline, and a marked improvement in the fit between predictions and observations, demonstrating that the self-attention mechanism effectively strengthens the model's selective focus on key features, thereby enhancing its ability to represent complex spatial structures and multisource features.

Compared with single-model structures, the SA-Blending heterogeneous ensemble model further reduced the prediction bias, with its scatter distribution evenly distributed on both sides of the 1:1 line, exhibiting optimal fit consistency. This result fully validates the superiority of the proposed SA-Blending heterogeneous ensemble framework in forest

canopy height estimation, demonstrating its ability to achieve higher prediction accuracy and robustness by fusing complementary information from multiple models.

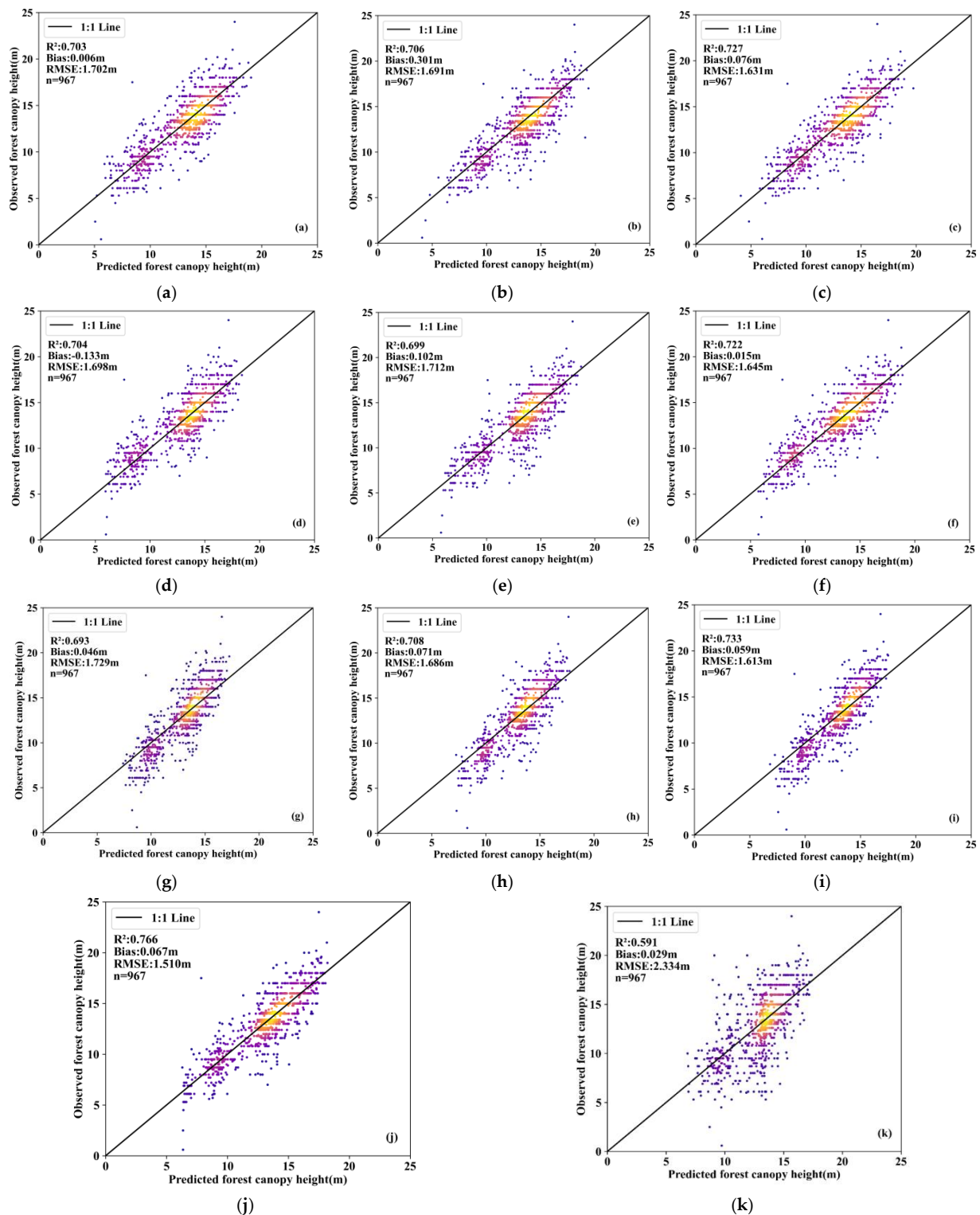


Figure 9. Scatter plot of predicted forest canopy height against the observed values. The brighter the color of the scatter points, the denser the distribution of forest canopy height. (a) DNN, (b) DNN (SA-Only), (c) DNN+SA (d) ResNet, (e) ResNet (SA-Only), (f) ResNet+SA, (g) XGBoost, (h) XGBoost (SA-Only), (i) XGBoost +SA, (j) SA-Blending, (k) SA-Blending (removed forest age feature).

Further analysis using the absolute error box plot (Figure 10) and permutation test results (Table 6) reveals that, compared with baseline models (DNN, ResNet, and XGBoost), the proposed SA-Blending heterogeneous ensemble model exhibits significantly lower median absolute error (red line) and mean absolute error (green rhombus), and displays the narrowest interquartile range, indicating that its predictions are closer to the measured tree height and demonstrate stronger robustness under complex environmental conditions. Regarding significance testing, we conducted pairwise comparisons of prediction errors between different models. All test results satisfied $p < 0.05$, demonstrating that the SA-Blending model's advantage in accuracy improvement is statistically significant and not due to random fluctuations, thus further validating the reliability and superiority of this method in forest canopy height estimation.

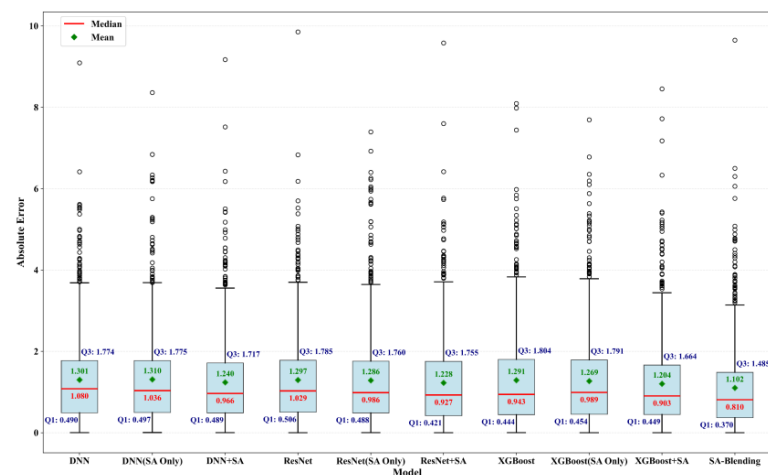


Figure 10. Absolute error box plot. The small circles represent the error samples that are greater than the edge lines of the box plot.

Table 6. The result of the permutation test.

Model Comparison	Sample Size	p-Value	Significant ($p < 0.05$)
SA-Blending vs. DNN	967	0.000	✓
SA-Blending vs. DNN (SA-Only)		0.000	✓
SA-Blending vs. DNN+SA		0.000	✓
SA-Blending vs. ResNet		0.000	✓
SA-Blending vs. ResNet (SA-Only)		0.000	✓
SA-Blending vs. ResNet+SA		0.000	✓
SA-Blending vs. XGBoost		0.000	✓
SA-Blending vs. XGBoost (SA-Only)		0.000	✓
SA-Blending vs. XGBoost+SA		0.000	✓
DNN+SA vs. DNN		0.005	✓
DNN+SA vs. DNN (SA-Only)		0.003	✓
ResNet+SA vs. ResNet		0.006	✓
ResNet+SA vs. ResNet (SA-Only)		0.020	✓
XGBoost+SA vs. XGBoost		0.000	✓
XGBoost+SA vs. XGBoost (SA-Only)		0.000	✓

Note: The ✓ represents the test result satisfied $p < 0.05$

3.5. Ablation Experiment

To further verify the significance of the forest age feature, this study designed an ablation experiment. Specifically, the forest age variable was completely removed from the input features, and the SA-Blending heterogeneous ensemble model was retrained and evaluated. Except for whether the forest age feature was included in the model, other parameters and validation metrics remained consistent with those described above. The evaluation result is shown in Figure 9k. After removing the forest age feature, the R^2 of the SA-Blending heterogeneous ensemble model was 0.591, and the bias was 0.029, representing

decreases of 0.175 and 0.029 respectively, while the RMSE was 2.334, an increase of 0.824. Although the bias decreased slightly, reflecting a change in the overall system bias. Overall, the accuracy of the SA-Blending heterogeneous ensemble model significantly decreased after removing the forest age feature, validating the importance of age characteristics in this study.

3.6. Mapping Wall-to-Wall Map of Forest Canopy Height

Based on the accuracy comparison results above, we selected the SA-blending heterogeneous integration model's point scale results to generate a continuous forest canopy height map. Continuous spatial feature information and canopy height extrapolation algorithms provide the possibility of mapping forest canopy height in wall-to-wall areas.

After obtaining the point-scale forest canopy height data, multisource features were aligned to a consistent spatial extent for regional extrapolation. Given that the five relative height metrics of the ICESat-2 photon point cloud (h_{max_canopy} , h_{canopy} , h_{median_canopy} , h_{min_canopy} , h_{mean_canopy}) are discrete spatial data, this study applied the empirical Bayesian Kriging (EBK) interpolation method to extend these features, generating continuous surface data aligned with the 10 m spatial range of 19 feature variables for raster stacking, creating a multivariable feature raster dataset for regional extrapolation. During the extrapolation process, to ensure consistency in the numerical scales of the feature variables, this study used the standardization parameters saved during the training phase to perform scale-consistency processing on the multivariable feature raster dataset. In the regional extrapolation phase, a pixel-by-pixel scanning approach was employed, in which the 24 raster feature data corresponding to each pixel were sequentially read in a "left-to-right, top-to-bottom" order. Based on the SA-Blending heterogeneous ensemble model weight file saved during the training phase, a complex nonlinear mapping relationship between the multisource data and forest canopy height was fitted, ultimately producing the canopy height value for each pixel, which was then written to the resulting raster image. To minimize the impact of non-vegetated areas on the forest canopy height estimation, non-forest land information was removed using the non-forest land mask vector data provided in Section 2.2.5. Finally, a 10 m resolution forest canopy height distribution map for the study area was obtained, as shown in Figure 11.

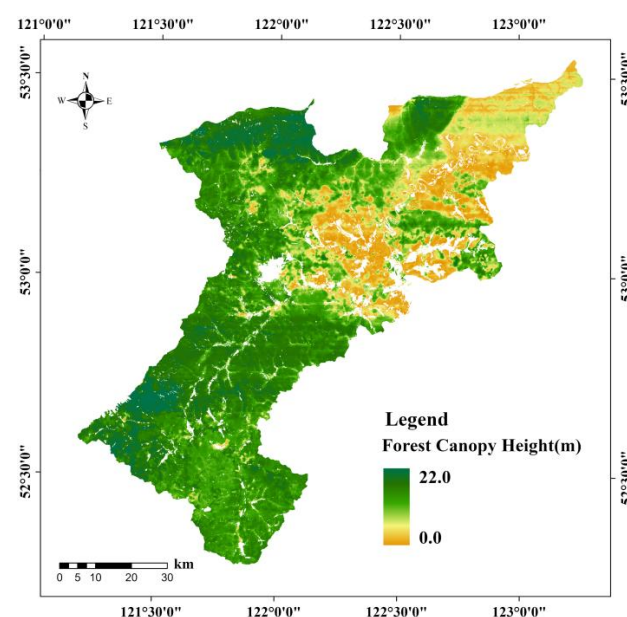


Figure 11. Spatial distribution of forest canopy height predictions in the study area, obtained using SA-Blending. The white areas represent non-forest areas.

4. Discussion

4.1. Forest Canopy Height Estimation Algorithms

In recent decades, many studies have exploited complementary information from multiple remotely sensed data to improve forest canopy height estimation. However, the in-depth mining of feature information is still insufficient. The diversity of different models and algorithms in terms of structural advantages, feature expression capabilities, and generalization performance needs to be further systematically explored.

Existing studies have employed self-attention mechanisms or stacking-based ensemble learning algorithms to estimate the forest canopy height, respectively and have achieved relatively satisfactory results. For example, Xiao et al. used ARFCNet, which synergizes convolutional, self-attention, and upsampling mechanisms to map forest canopy height [55]. Jiang et al. used a stacking algorithm to predict forest canopy height and achieved the best prediction accuracy [3]. Nonetheless, the majority of the pertinent work concentrates on refining a specific methodological approach related to feature augmentation or model ensembling. No research has integrated the self-attention mechanism with the Blending heterogeneous ensemble framework to estimate forest canopy height. The self-attention mechanism offers substantial advantages in multisource feature representation and noise suppression grid. Subsequently, these were combined with the representation, while Blending effectively integrates the structural characteristics of multiple base learners and mitigates error accumulation among them. Consequently, combining these two approaches has the potential to yield complementary advantages, which can not only enhance feature expression capabilities but also improve the model's robustness and generalization proficiency. In this study, we propose, for the first time, an SA-Blending heterogeneous integration model that combines the deep feature extraction of the self-attention mechanism with the model fusion technique of Blending, aimed at enhancing the accuracy and stability of forest canopy height mapping.

In terms of feature extraction, this study focused on exploring the changes in the model's prediction performance after introducing the self-attention mechanism. The results in Figure 9 show that after concatenating and fusing the original features with the high-order features generated by the self-attention mechanism, the forest canopy height prediction accuracy of the three base models (XGBoost, DNN, and ResNet) has significantly improved. Compared to relying solely on the original input features, the self-attention mechanism can explicitly capture long-range dependencies across multisource data and dynamically optimize the weighted combination of multiple feature sources (photon features, spectral information, terrain variables, and forest age), effectively enhancing feature expression and task relevance. The self-attention mechanism not only strengthens the model's ability to identify key structural information but also reduces redundant and noisy features in the original multisource data by lowering the weights, thereby improving the overall prediction performance. This result further indicates that, in multisource remote sensing scenarios, introducing self-attention-driven high-order feature fusion is an effective way to improve the accuracy of forest canopy height estimation.

The efficacy of the proposed SA-Blending heterogeneous ensemble model is further supported by the fact that, when given the same multisource feature input (i.e., the concatenation and fusion of the original features and the self-attention mechanism features), the Blending heterogeneous ensemble framework outperformed the single models XGBoost, DNN, and ResNet in estimating forest canopy height. Through the synergistic effect of meta-learners, Blending can fully integrate the complementary advantages of various base learners, achieving more stable predictive capabilities globally than single-model learning mechanisms that rely on a single structure or local features. Additionally, the higher-order features generated by the self-attention mechanism provide richer cross-modal associa-

tion information for the ensemble process, enabling the ensemble model to maintain high generalization performance even under complex forest structure conditions.

The bias produced by the SA-Blending heterogeneous ensemble model, however, is marginally greater than that of the three basic models'. The reason may be due to the Blending framework's tendency to improve prediction stability and reduce overall error variance during optimization, leading to a slight systematic divergence in some local high-value regions. Another possibility is that the high-dimensional fusion features generated by the self-attention mechanism amplify sensitivity to specific forest canopy height segments in a portion of the feature space, leading the meta-learner to slightly deviate in its estimates of those segments during the weighting process. Nevertheless, this deviation did not affect the model's overall accuracy and remained within a reasonable range. Overall, the slight increase in bias can be addressed by introducing more refined bias-correction strategies in the subsequent research framework or by further optimizing the normalization and selection methods of self-attention features, to maintain high accuracy while reducing systematic bias.

4.2. Key Drivers for Estimating Forest Canopy Height

Forest age is an essential biological indicator of the forest growth stage and structural changes, and it directly influences the potential differences in forest canopy height. In their work, Luo et al. highlighted the influence of forest age on changes in forest structure; however, they used forest age as the foundation for building a forest canopy height extrapolation model rather than as a predictive variable to estimate forest canopy height [10]. Thus, it is still worthwhile to investigate the usefulness of forest age as a feature variable in estimating forest canopy height. This work considered forest age a crucial variable within the multisource feature system, fusing it collaboratively with other features via a self-attention mechanism, which served as input for the SA-blending heterogeneous integration model. This strategy significantly improved feature expression capabilities and the depth of cross-modal correlation information extraction, resulting in the precise estimation of forest canopy height. The feature importance ranking results in Figure 7 further confirm that forest age is a key factor in estimating forest canopy height. The contribution of forest age to the model's predictive performance is significantly higher than that of other environmental factors, such as slope and spectral features. Unlike traditional spectral features, forest age can directly reflect the intrinsic changes in forest growth from a biological perspective, thereby influencing the spatial distribution and structural characteristics of the forest canopy height. Consequently, incorporating forest age into the prediction model not only helps to enhance the prediction accuracy of the model but also provides a novel approach for remote sensing data-driven forest resource monitoring. Future studies may investigate the synergistic impacts of forest age alongside other environmental variables to enhance the precision and application of forest canopy height estimation.

Additionally, the results in Figure 7 show that among the topographical factors, the importance of elevation is not only higher in terms of feature significance than that of slope and aspect, but also significantly surpasses that of spectral features suggesting that elevation has a potential dominant effect on forest canopy height estimation and its geographical distribution. This conclusion is consistent with the findings of Xiao et al. [55]. To further reveal the intrinsic relationship between elevation and forest canopy height, this study conducts a spatial overlay analysis of elevation and forest canopy height estimated map, and presents the results alongside three-dimensional terrain visualization, as shown in Figure 12. It is evident from the visual interpretation results in Figure 11 that the spatial heterogeneity of forest canopy height shows a highly consistent spatial coupling pattern with topographic undulation. Low-canopy areas are primarily concentrated in lower

terrain and relatively gentle topography, while high-canopy areas are usually distributed in mountainous terrain units at higher altitudes. The underlying ecological mechanism lies in the fact that high-altitude regions often have lower levels of human disturbance, stronger soil moisture retention capacity, and more favorable water and heat resource allocation, thereby providing more stable and suitable habitat conditions for tree growth. Conversely, low-altitude regions are more vulnerable to the combined effects of human activity, natural disturbances, and land-use changes, which can degrade forest structure and prevent canopy height growth. The results of feature importance ranking and spatial visualization analysis jointly verified the crucial role of elevation in the spatial distribution and estimation of forest canopy height. This finding not only deepened our understanding of the driving mechanism of spatial heterogeneity in forest canopy height but also provided key theoretical support for constructing forest canopy height inversion models that consider the effect of terrain gradients. At the same time, this finding has significant practical significance for optimizing monitoring and management strategies for mountain forest resources, especially in complex terrain areas where high-precision canopy height mapping is conducted.

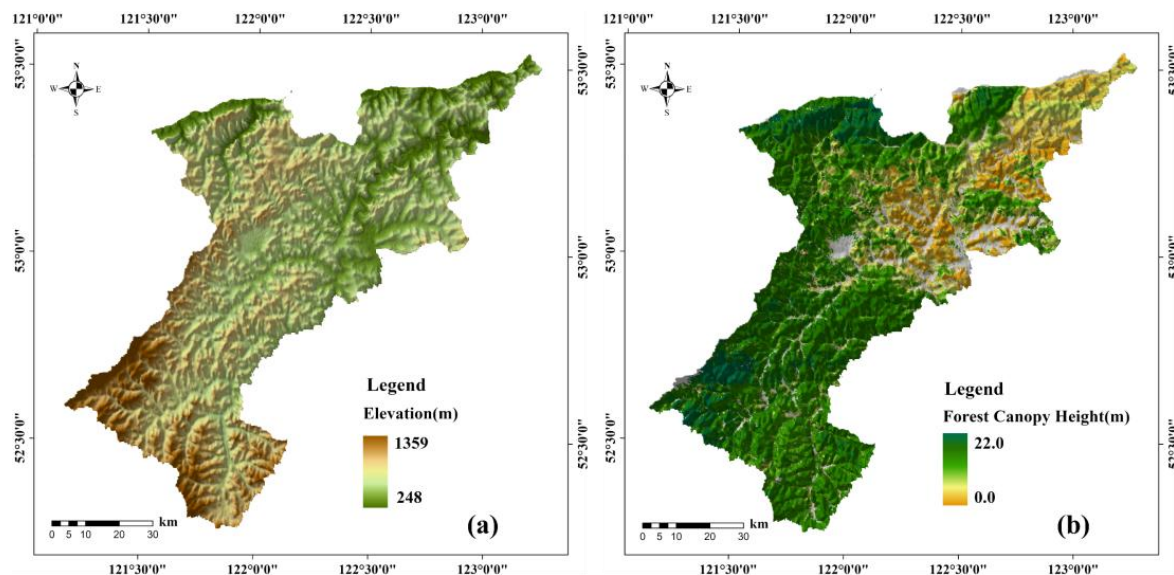


Figure 12. Map of three-dimensional visualization. (a) The elevation distribution map of the study area; and (b) Three-dimensional map of forest canopy height and elevation overlay.

4.3. Limitations and Perspectives

Although the multi-model fusion framework proposed in this study has significantly improved the accuracy of estimating forest canopy height, there are still issues of prediction Bias, especially for samples with extremely high or low canopy heights. Since such samples account for a relatively small proportion in the training dataset (as shown in Figure 13), each model is prone to being affected by the imbalance in sample distribution when learning the distribution and structural patterns of relevant features, thereby weakening the ability of the self-attention mechanism to adjust the cross-modal feature weights under extreme conditions effectively and adaptively. Additionally, factors such as shadow occlusion in complex terrain, the limited penetration of photon point clouds in high-vegetation-density areas, and differences in spatial resolution across sensors may further interfere with the model's representation of local canopy structure, leading to prediction errors [56,57]. In subsequent studies, methods such as the increasing sample size and adopting multiscale or hierarchical feature modeling strategies can be used to improve the model's fit and reduce prediction bias.

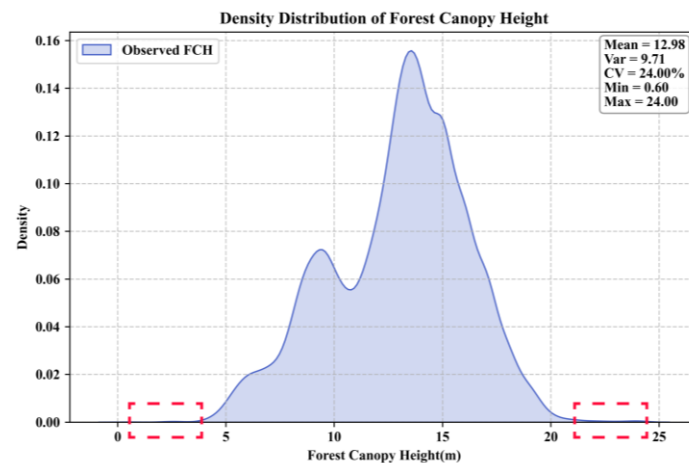


Figure 13. The distribution of forest canopy height sample data. The mean, coefficient of variation, maximum value, and minimum value are all shown in the upper-right corner of the figure. The red dashed box marks the initial and final areas with low sample density.

5. Conclusions

In this study, we explored the effect of fusing multisource spaceborne LiDAR data and optical imagery to improve the accuracy of regional-scale forest canopy height estimates. A multisource remote sensing feature system was built using ICESat-2/ATLAS photon point clouds, Sentinel-2/MSI multispectral imagery, and SRTM-DEM topographic data, after eliminating redundant information interference using a permutation feature importance analysis method, as the model's input variables. Additionally, we used DNN, XGBoost and ResNet as base learners and RF as the meta-learner, proposed an SA-Blending heterogeneous ensemble model that combines a blending technique with a self-attention mechanism to enhance the accuracy of forest canopy height estimation and spatial mapping capabilities. The following are the primary conclusions: (1) The self-attention mechanism effectively enhances the representation ability of high-dimensional features by adaptively focusing on the potential complementarity among cross-modal features (photon point clouds, spectral features, forest age, and terrain factors). Following the implementation of the self-attention mechanism, the DNN(SA-Only) and XGBoost(SA-Only) models' prediction performance improved considerably compared to using the original feature set. The corresponding R^2 values rose to 0.706 and 0.708, and the RMSE values decreased to 1.691 m and 1.613 m, respectively. Although the R^2 for ResNet(SA-Only) slightly reduced to 0.699 and the RMSE increased to 1.712 m, the overall impact was limited. (2) Under identical conditions of multisource feature input (i.e., the concatenation and fusion of original features and those derived from the self-attention mechanism), DNN+SA, ResNet+SA and XGBoost+SA all showed improved fitting accuracy and stability, with R^2 increasing to 0.727, 0.722, and 0.733, while RMSE decreased to 1.631 m, 1.645 m and 1.613 m, respectively. The results validated the efficacy of cross-modal feature fusion via the self-attention mechanism for estimating forest canopy height, indicating that the self-attention mechanism substantially enhances the association expression capability of multisource features, consequently improving model performance. (3) By using multisource features as the input (i.e., the concatenation and fusion of the original features and the self-attention mechanism features), the SA-Blending heterogeneous ensemble model achieved the best prediction performance with an R^2 of 0.766, and an RMSE of 1.510 m, which significantly outperformed all single models and the self-attention mechanism optimized model structure. This result indicates that the SA-Blending heterogeneous ensemble model can effectively integrate the structural advantages of different models, overcome the limitations of a single modeling strategy, minimize error accumulation, and has stronger robustness and generalization ability. It

is particularly suitable for forest canopy height estimation under complex terrain and multisource data conditions.

In summary, the results of this study demonstrated that the proposed SA-Blending heterogeneous ensemble model using multisource features can effectively estimate and map the spatial distribution of forest canopy height, providing a theoretical basis and technical support for regional and even global forest resource monitoring and forest management. The optimal model parameters and cross-modal features used in this study have regional applicability. When applied to other regions, corresponding adjustments should be made based on the specific situation.

Author Contributions: Conceptualization, J.T. and P.D.; methodology, J.T., P.Z. and Q.W.; formal analysis, P.Z. and J.T.; validation, D.L., Y.G. and Q.W.; investigation, P.Z., W.S. and X.M.; resources, J.T.; data curation, J.T. and P.Z.; writing—original draft preparation, J.T., P.Z. and P.D.; writing—review and editing, J.T., P.D. and P.Z.; visualization, J.T., P.D. and P.Z.; supervision, J.T., P.D., P.Z., D.L., W.S. and Y.G.; project administration, X.M.; funding acquisition, J.T., D.L. and Q.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Basic Scientific Research Operating Expenses of Heilongjiang Provincial Universities, grant number 2022GJ02, Heilongjiang Provincial Natural Science Foundation of China, grant number PL2024D018, National Key Research and Development Program of China, grant number 2023YFD2200804 and Longjiang Project Young Goose Innovation Team Support Program, grant number 2025CYLJ01.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: The authors extend their sincere appreciation to the foresters at Mohe Forestry Bureau for their invaluable support in data collection and for generously sharing their insights into the local forests.

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

References

1. Rahman, M.F.; Onoda, Y.; Kitajima, K. Forest canopy height variation in relation to topography and forest types in central Japan with LiDAR. *For. Ecol. Manag.* **2022**, *503*, 119792. [\[CrossRef\]](#)
2. Li, Y.; Lu, D.; Lu, Y.; Li, G. Examining the Impact of Topography and Vegetation on Existing Forest Canopy Height Products from ICESat-2 ATLAS/GEDI Data. *Remote Sens.* **2024**, *16*, 3650. [\[CrossRef\]](#)
3. Jiang, F.; Zhao, F.; Ma, K.; Li, D.; Sun, H. Mapping the forest canopy height in Northern China by synergizing ICESat-2 with Sentinel-2 using a stacking algorithm. *Remote Sens.* **2021**, *13*, 1535. [\[CrossRef\]](#)
4. Liu, C.; Gong, W.; Shi, S.; Wang, T.; Xu, T.; Shi, Z.; Niu, J. Deep learning-driven forest canopy height mapping in boreal regions through multi-source remote sensing fusion: Integrating Sentinel-1/2, PALSAR, and ICESat-2/LVIS data. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *143*, 104766. [\[CrossRef\]](#)
5. Solberg, S.; Hansen, E.H.; Gobakken, T.; Naesset, E.; Zahabu, E. Biomass and InSAR height relationship in a dense tropical forest. *Remote Sens. Environ.* **2017**, *192*, 166–175. [\[CrossRef\]](#)
6. Pourshamsi, M.; Garcia, M.; Laval, M.; Balzter, H. A Machine-Learning Approach to PolInSAR and LiDAR Data Fusion for Improved Tropical Forest Canopy Height Estimation Using NASA AfriSAR Campaign Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3453–3463. [\[CrossRef\]](#)
7. Li, Y.; Li, C.; Li, M.; Liu, Z. Influence of variable selection and forest type on forest aboveground biomass estimation using machine learning algorithms. *Forests* **2019**, *10*, 1073. [\[CrossRef\]](#)
8. Tamiminia, H.; Salehi, B.; Mahdianpari, M.; Goulden, T. State-wide forest canopy height and aboveground biomass map for New York with 10 m resolution, integrating GEDI, Sentinel-1, and Sentinel-2 data. *Ecol. Inform.* **2024**, *79*, 102404. [\[CrossRef\]](#)
9. Xu, K.; Zhao, L.; Chen, E.; Li, K.; Liu, D.; Li, T.; Li, Z.; Fan, Y. Forest Height Estimation Approach Combining P-Band and X-Band Interferometric SAR Data. *Remote Sens.* **2022**, *14*, 3070. [\[CrossRef\]](#)
10. Luo, Y.; Qi, S.; Liao, K.; Zhang, S.; Hu, B.; Tian, Y. Mapping the forest height by fusion of ICESat-2 and multi-source remote sensing imagery and topographic information: A case study in Jiangxi province, China. *Forests* **2023**, *14*, 454. [\[CrossRef\]](#)

11. Wang, M.; Sun, R.; Xiao, Z. Estimation of forest canopy height and aboveground biomass from spaceborne LiDAR and Landsat imageries in Maryland. *Remote Sens.* **2018**, *10*, 344. [CrossRef]
12. Ghosh, S.M.; Behera, M.D.; Kumar, S.; Das, P.; Prakash, A.J.; Bhaskaran, P.K.; Roy, P.S.; Barik, S.K.; Jeganathan, C.; Srivastava, P.K. Predicting the forest canopy height from LiDAR and multi-sensor data using machine learning over India. *Remote Sens.* **2022**, *14*, 5968. [CrossRef]
13. Potapov, P.; Li, X.; Hernandez-Serna, A.; Tyukavina, A.; Hansen, M.C.; Kommareddy, A.; Pickens, A.; Turubanova, S.; Tang, H.; Silva, C.E. Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sens. Environ.* **2021**, *253*, 112165. [CrossRef]
14. Tiwari, K.; Narine, L.L. A comparison of machine learning and geostatistical approaches for mapping forest canopy height over the southeastern US using ICESat-2. *Remote Sens.* **2022**, *14*, 5651. [CrossRef]
15. Dong, J.; Ni, W.; Zhang, Z.; Sun, G. Performance of ICESat-2 ATL08 product on the estimation of forest height by referencing to small footprint LiDAR data. *Natl. Remote Sens. Bull.* **2021**, *25*, 1294–1307. [CrossRef]
16. Yuanyuan, W.; Guicai, L.; Jianhua, D.; Zhaodi, G.; Shihao, T.; Cheng, W.; Qingni, H.; Ronggao, L.; Jing, M.C. A combined GLAS and MODIS estimation of the global distribution of mean forest canopy height. *Remote Sens. Environ.* **2016**, *174*, 24–43. [CrossRef]
17. Li, W.; Niu, Z.; Shang, R.; Qin, Y.; Wang, L.; Chen, H. High-resolution mapping of forest canopy height using machine learning by coupling ICESat-2 LiDAR with Sentinel-1, Sentinel-2 and Landsat-8 data. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *92*, 102163. [CrossRef]
18. Lin, X.; Xu, M.; Cao, C.; Dang, Y.; Bashir, B.; Xie, B.; Huang, Z. Estimates of forest canopy height using a combination of ICESat-2/ATLAS data and stereo-photogrammetry. *Remote Sens.* **2020**, *12*, 3649. [CrossRef]
19. Wang, J.; Shen, X.; Cao, L. upscaling forest canopy height estimation using waveform-calibrated GEDI spaceborne LiDAR and Sentinel-2 data. *Remote Sens.* **2024**, *16*, 2138. [CrossRef]
20. Coops, N.C.; Tompalski, P.; Goodbody, T.R.; Queinnec, M.; Luther, J.E.; Bolton, D.K.; White, J.C.; Wulder, M.A.; van Lier, O.R.; Hermosilla, T. Modelling lidar-derived estimates of forest attributes over space and time: A review of approaches and future trends. *Remote Sens. Environ.* **2021**, *260*, 112477. [CrossRef]
21. Zhang, B.; Zhang, L.; Yan, M.; Zuo, J.; Dong, Y.; Chen, B. High-resolution mapping of forest parameters in tropical rainforests through AutoML integration of GEDI with Sentinel-1/2, Landsat 8 and ALOS-2 data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 9084–9118. [CrossRef]
22. Wen, L.; Hughes, M. Coastal wetland mapping using ensemble learning algorithms: A comparative study of bagging, boosting and stacking techniques. *Remote Sens.* **2020**, *12*, 1683. [CrossRef]
23. Zhang, H.; Li, L.; Liu, D. Survey of Multimodal Data Fusion Research. *J. Front. Comput. Sci. Technol.* **2024**, *18*, 2501–2520. [CrossRef]
24. Li, B.; Ren, H.-e.; Dong, P.; Tian, J. Comparison of convolutional neural network and support vector machine for identification of forest types and burned areas. *J. Appl. Remote Sens.* **2024**, *18*, 014531. [CrossRef]
25. Rai, N.; Ma, Q.; Poudel, K.P.; Himes, A.; Meng, Q. Evaluating the uncertainties in forest canopy height measurements using ICESat-2 data. *J. Remote Sens.* **2024**, *4*, 0160. [CrossRef]
26. Kong, D.; Pang, Y. ICESat-2 data denoising and forest canopy height estimation using Machine Learning. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *135*, 104263. [CrossRef]
27. Ice, Cloud, and Land Elevation Satellite (ICESat-2) Project Algorithm Theoretical Basis Document (ATBD) for Land Vegetation Along-Track Products (ATL08), version 7. pdf. Available online: <https://nsidc.org/data/atl08/versions/7#anchor-documentation>. (accessed on 10 July 2024).
28. Neuenschwander, A.; Pitts, K. The ATL08 land and vegetation product for the ICESat-2 Mission. *Remote Sens. Environ.* **2019**, *221*, 247–259. [CrossRef]
29. Neuenschwander, A.L.; Magruder, L.A. The potential impact of vertical sampling uncertainty on ICESat-2/ATLAS terrain and canopy height retrievals for multiple ecosystems. *Remote Sens.* **2016**, *8*, 1039. [CrossRef]
30. Ettehadi Osgouei, P.; Kaya, S.; Sertel, E.; Alganci, U. Separating built-up areas from bare land in mediterranean cities using Sentinel-2A imagery. *Remote Sens.* **2019**, *11*, 345. [CrossRef]
31. Liang, H.; Bie, Q.; Shi, Y.; Deng, X.; Li, X. Estimation of canopy height is conducted by integrating multi-source remote sensing data from ICESat-2 and GEDI. *Remote Sens. Technol. Appl.* **2025**, *40*, 202–214. Available online: <http://www.rsta.ac.cn/EN/10.11873/j.issn.1004-0323.2025.1.0202>. (accessed on 16 December 2025).
32. Farbo, A.; Sarvia, F.; De Petris, S.; Basile, V.; Borgogno-Mondino, E. Forecasting corn NDVI through AI-based approaches using sentinel 2 image time series. *ISPRS J. Photogramm. Remote Sens.* **2024**, *211*, 244–261. [CrossRef]
33. Zhou, J.; Zhou, Z.; Zhao, Q.; Han, Z.; Wang, P.; Xu, J.; Dian, Y. Evaluation of different algorithms for estimating the growing stock volume of Pinus massoniana plantations using spectral and spatial information from a SPOT6 image. *Forests* **2020**, *11*, 540. [CrossRef]

34. Genç, Ç.Ö.; Altunel, A.O. Monitoring the operational changes in surface reflectances after logging, based on popular indices over Sentinel-2, Landsat-8, and ASTER imageries. *Environ. Monit. Assess.* **2025**, *197*, 120. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Voitik, A.; Kravchenko, V.; Pushka, O.; Kutkovetska, T.; Shchur, T.; Kocira, S. Comparison of NDVI, NDRE, MSAVI and NDSI indices for early diagnosis of crop problems. *Agric. Eng.* **2023**, *27*, 47–57. [\[CrossRef\]](#)
36. Adamu, B.; Ibrahim, S.a.; Rasul, A.; Whanda, S.J.; Headboy, P.; Muhammed, I.; Maiha, I.A. Evaluating the accuracy of spectral indices from Sentinel-2 data for estimating forest biomass in urban areas of the tropical savanna. *Remote Sens. Appl. Soc. Environ.* **2021**, *22*, 100484. [\[CrossRef\]](#)
37. Alikhanova, S.; Tarantino, C.; Bull, J.W. Tracking Vegetation Dynamics in Drylands with MSAVI: Insights from the South Aral Sea. *Earth Syst. Environ.* **2025**, *9*, 1–13. [\[CrossRef\]](#)
38. Ma, T.; Hu, Y.; Wang, J.; Beckline, M.; Pang, D.; Chen, L.; Ni, X.; Li, X. A Novel Vegetation Index Approach Using Sentinel-2 Data and Random Forest Algorithm for Estimating Forest Stock Volume in the Helan Mountains, Ningxia, China. *Remote Sens.* **2023**, *15*, 1853. [\[CrossRef\]](#)
39. Liu, J.; Fan, J.; Yang, C.; Xu, F.; Zhang, X. Novel vegetation indices for estimating photosynthetic and non-photosynthetic fractional vegetation cover from Sentinel data. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *109*, 102793. [\[CrossRef\]](#)
40. Strashok, O.; Ziemiańska, M.; Strashok, V. Evaluation and Correlation of Sentinel-2 NDVI and NDMI in Kyiv (2017–2021). *J. Ecol. Eng.* **2022**, *23*, 212–218. [\[CrossRef\]](#)
41. Guo, Z.; Kurban, A.; Ablekim, A.; Wu, S.; Van de Voorde, T.; Azadi, H.; Maeyer, P.D.; Dufatanye Umwali, E. Estimation of photosynthetic and non-photosynthetic vegetation coverage in the lower reaches of Tarim river based on sentinel-2a data. *Remote Sens.* **2021**, *13*, 1458. [\[CrossRef\]](#)
42. Gao, S.; Zhu, J.; Fu, H. A rapid and easy way for national forest heights retrieval in China using ICESat-2/ATL08 in 2019. *Forests* **2023**, *14*, 1270. [\[CrossRef\]](#)
43. Liu, K.; Song, C.; Zhao, S.; Wang, J.; Chen, T.; Zhan, P.; Fan, C.; Zhu, J. Mapping inundated bathymetry for estimating lake water storage changes from SRTM DEM: A global investigation. *Remote Sens. Environ.* **2024**, *301*, 113960. [\[CrossRef\]](#)
44. USGS EROS Archive–Digital Elevation–Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global. Available online: <https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1>. (accessed on 16 February 2024).
45. Chen, C.; Liang, J.; Sun, W.; Yang, G.; Meng, X. An automatically recursive feature elimination method based on threshold decision in random forest classification. *Geo-Spat. Inf. Sci.* **2025**, *28*, 1494–1519. [\[CrossRef\]](#)
46. Qu, Y.; Baghbaderani, R.K.; Qi, H.; Kwan, C. Unsupervised pansharpening based on self-attention mechanism. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 3192–3208. [\[CrossRef\]](#)
47. Gao, X.; Zhang, Z.; Mu, T.; Zhang, X.; Cui, C.; Wang, M. Self-attention driven adversarial similarity learning network. *Pattern Recognit.* **2020**, *105*, 107331. [\[CrossRef\]](#)
48. Reyad, M.; Sarhan, A.M.; Arafa, M. A modified Adam algorithm for deep neural network optimization. *Neural Comput. Appl.* **2023**, *35*, 17095–17112. [\[CrossRef\]](#)
49. Montavon, G.; Samek, W.; Müller, K.-R. Methods for interpreting and understanding deep neural networks. *Digit. Signal Process.* **2018**, *73*, 1–15. [\[CrossRef\]](#)
50. Chen, T. *XGBoost: A Scalable Tree Boosting System*; Cornell University: Ithaca, NY, USA, 2016. [\[CrossRef\]](#)
51. Luo, M.; Wang, Y.; Xie, Y.; Zhou, L.; Qiao, J.; Qiu, S.; Sun, Y. Combination of feature selection and catboost for prediction: The first application to the estimation of aboveground biomass. *Forests* **2021**, *12*, 216. [\[CrossRef\]](#)
52. Gao, M.; Qi, D.; Mu, H.; Chen, J. A Transfer Residual Neural Network Based on ResNet-34 for Detection of Wood Knot Defects. *Forests* **2021**, *12*, 212. [\[CrossRef\]](#)
53. Salman, H.A.; Kalakech, A.; Steiti, A. Random forest algorithm overview. *Babylon. J. Mach. Learn.* **2024**, *2024*, 69–79. [\[CrossRef\]](#)
54. Zhu, W.; Li, Y.; Luan, K.; Qiu, Z.; He, N.; Zhu, X.; Zou, Z. Forest Canopy Height Retrieval and Analysis Using Random Forest Model with Multi-Source Remote Sensing Integration. *Sustainability* **2024**, *16*, 1735. [\[CrossRef\]](#)
55. Xiao, K.; Zhao, X.; Ding, Y.; Huang, C.; Lin, J.; Mai, Y.; Sun, Y.; Xin, Q. Ultra-high Spatial Resolution Mapping of Urban Forest Canopy Height with Multimodal Remote Sensing Data and Deep Learning Method. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 9865–9882. [\[CrossRef\]](#)
56. Wang, S.; Liu, C.; Li, W.; Jia, S.; Yue, H. Hybrid model for estimating forest canopy heights using fused multimodal spaceborne LiDAR data and optical imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *122*, 103431. [\[CrossRef\]](#)
57. Neuenschwander, A.; Guenther, E.; White, J.C.; Duncanson, L.; Montesano, P. Validation of ICESat-2 terrain and canopy heights in boreal forests. *Remote Sens. Environ.* **2020**, *251*, 112110. [\[CrossRef\]](#)

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.