

Article

Interpretable Multi-Sensor Fusion of Optical and SAR Data for GEDI-Based Canopy Height Mapping in Southeastern North Carolina

Chao Wang ^{1,2,*} , Conghe Song ¹ , Todd A. Schroeder ³, Curtis E. Woodcock ⁴ , Tamlin M. Pavelsky ², Qianqian Han ⁵ and Fangfang Yao ⁶

¹ Department of Geography and Environment, University of North Carolina, Chapel Hill, NC 27599, USA; csong@email.unc.edu

² Department of Earth, Marine and Environmental Sciences, University of North Carolina, Chapel Hill, NC 27599, USA; pavelsky@email.unc.edu

³ USDA Forest Service, Southern Research Station, Forest Inventory and Analysis, Knoxville, TN 37919, USA; todd.schroeder@usda.gov

⁴ Department of Earth and Environment, Boston University, Boston, MA 02215, USA; curtis@bu.edu

⁵ Faculty of Geo-Information Science and Earth Observation, University of Twente, 7522 NH Enschede, The Netherlands; q.han@utwente.nl

⁶ Environmental Institute, University of Virginia, Charlottesville, VA 22904, USA; fangfang.yao@colorado.edu

* Correspondence: chao.wang@unc.edu

Abstract: Accurately monitoring forest canopy height is crucial for sustainable forest management, particularly in southeastern North Carolina, USA, where dense forests and limited accessibility pose substantial challenges. This study presents an explainable machine learning framework that integrates sparse GEDI LiDAR samples with multi-sensor remote sensing data to improve both the accuracy and interpretability of forest canopy height estimation. This framework incorporates multitemporal optical observations from Sentinel-2; C-band backscatter and InSAR coherence from Sentinel-1; quad-polarization L-Band backscatter and polarimetric decompositions from the Uninhabited Aerial Vehicle Synthetic Aperture Radar (UAVSAR); texture features from the National Agriculture Imagery Program (NAIP) aerial photography; and topographic data derived from an airborne LiDAR-based digital elevation model. We evaluated four machine learning algorithms, K-nearest neighbors (KNN), random forest (RF), support vector machine (SVM), and eXtreme gradient boosting (XGB), and found consistent accuracy across all models. Our evaluation highlights our method's robustness, evidenced by closely matched R^2 and RMSE values across models: KNN (R^2 of 0.496, RMSE of 5.13 m), RF (R^2 of 0.510, RMSE of 5.06 m), SVM (R^2 of 0.544, RMSE of 4.88 m), and XGB (R^2 of 0.548, RMSE of 4.85 m). The integration of comprehensive feature sets, as opposed to subsets, yielded better results, underscoring the value of using multisource remotely sensed data. Crucially, SHapley Additive exPlanations (SHAP) revealed the multi-seasonal red-edge spectral bands of Sentinel-2 as dominant predictors across models, while volume scattering from UAVSAR emerged as a key driver in tree-based algorithms. This study underscores the complementary nature of multi-sensor data and highlights the interpretability of our models. By offering spatially continuous, high-quality canopy height estimates, this cost-effective, data-driven approach advances large-scale forest management and environmental monitoring, paving the way for improved decision-making and conservation strategies.

Keywords: forest canopy height; LiDAR; synthetic aperture radar; red edge; machine learning



Academic Editors: Xiaoning Song, Ronghai Hu, Fangcheng Zhou and Wenchao Han

Received: 17 March 2025

Revised: 19 April 2025

Accepted: 23 April 2025

Published: 25 April 2025

Citation: Wang, C.; Song, C.; Schroeder, T.A.; Woodcock, C.E.; Pavelsky, T.M.; Han, Q.; Yao, F. Interpretable Multi-Sensor Fusion of Optical and SAR Data for GEDI-Based Canopy Height Mapping in Southeastern North Carolina. *Remote Sens.* **2025**, *17*, 1536. <https://doi.org/10.3390/rs17091536>

Copyright: © 2025 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/>).

1. Introduction

Forests, covering roughly 31% of the Earth's land area [1], are fundamental to biodiversity conservation, carbon sequestration, and climate regulation [2]. They play a crucial role in maintaining energy balance, regulating the water cycle, and controlling atmospheric CO₂ concentrations. Additionally, forests provide critical habitats for diverse species [3], thereby offering ecosystem services vital for human well-being, including air and water purification, soil stabilization, and resources provision [4]. However, forests face increasing threats from both natural disturbances and human activities, leading to widespread deforestation, degradation, and fragmentation [5,6]. These disturbances undermine the structure and functions of forests and even have profound implications for regional climates, including altering precipitation patterns and temperatures [7], which in turn jeopardize environmental stability.

Monitoring forest structural characteristics, particularly forest canopy height, is essential for accessing forest ecosystem resilience and vulnerability [8]. Forest canopy height, often represented as a Canopy Height Model (CHM), serves as a vital indicator of ecosystem productivity, carbon stocks, and biodiversity [9]. Accurate CHM estimation provides key insights that can support forest management, conservation efforts, and restoration initiatives. Recent advancements in remote sensing (RS) technologies have significantly improved our ability to monitor CHM [10]. Airborne Light Detection and Ranging (LiDAR) has been recognized for its accuracy in height measurements but is limited by high costs and restricted spatial and temporal coverage. In contrast, spaceborne missions such as the Global Ecosystem Dynamics Investigation (GEDI) and the Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2) provide extensive data on global forest structure [11]. Specifically, GEDI offers an unprecedented capability for monitoring vertical vegetation structures globally through full-waveform LiDAR data [12]. However, these spaceborne LiDAR missions are by design sampling-based, leading to spatial discontinuities in collected CHM data [13].

To achieve spatially continuous CHM mapping across forest landscapes, it is essential to integrate GEDI height measurements with other spatially continuous, remotely-sensed datasets. Recent studies highlight the potential of combining spatially sparse LiDAR data with continuous observations from optical sensors (e.g., Landsat and Sentinel-2) or Synthetic Aperture Radar (SAR) sensors (e.g., PALSAR-2 and Sentinel-1), which allows for the prediction of spatially continuous CHM [14,15]. While these platforms were not initially designed for direct vegetation structure monitoring, their observations can indirectly inform canopy height estimation due to correlations between the reflective and radiative properties captured by optical sensors, structural information derived from SAR data, and forest canopy characteristics [16–18]. For example, Potapov [19] proved the feasibility of estimating forest CHM using spectral reflectance, vegetation indices, and tasseled cap greenness derived from Landsat data. Similarly, Ehlers [16] found that texture metrics derived from remotely sensed imagery could be correlated with the structure of the forest canopy, such as crown diameter and canopy height, indicating the value of integrating different metrics derived from remotely sensed imagery.

Unlike optical sensors, SAR sensors, with their cloud-penetrating capabilities, emerge as a valuable alternative for consistent acquisition of Earth observation data, especially in cloud-prone or dense forest regions [17,20]. Integrating GEDI data with both optical and SAR sensors addresses challenges such as spatial gaps, signal saturation, and cloud cover, which arise when relying solely on optical data [19]. By leveraging the complementary strengths of optical and SAR sensors, we can gain a more complete picture of the canopy structures. As demonstrated by studies such as Li et al.'s [21], including backscatter data from Sentinel-1, along with red-edge bands from Sentinel-2 improved the accuracy of forest CHM predictions. Building upon the advantages of SAR data, Polarimetric SAR (PolSAR)

provides even more detailed insights into the physical scattering mechanisms of forest canopies. PolSAR data capture information about the polarization state of electromagnetic waves, which is sensitive to vegetation structure and moisture content [22]. This makes PolSAR an ideal component for integration with LiDAR to improve CHM estimation. For instance, Pourshamsi et al. [23] effectively merged PolSAR parameters with LiDAR samples to accurately estimate forest height. Their approach mirrored patterns observed in estimations based on Polarimetric Interferometric SAR (PolInSAR), demonstrating the efficacy of combining PolSAR and LiDAR data to improve forest CHM predictions [24]. These advancements highlight the growing consensus that leveraging multisource Earth observation data, enhanced by machine learning techniques, is key to advancing forest height estimation.

Machine learning (ML) techniques have advanced significantly in improving CHM predictions, particularly with the shift from traditional ML models to more advanced deep learning approaches. For instance, Lang et al. [25] used deep learning models to estimate CHM from Sentinel-2 data. Tolan et al. [26] advanced this further by applying self-supervised vision transformers to Maxar imagery for high-resolution CHM mapping. These developments highlight the growing role of ML in enhancing canopy height predictions. However, despite these advances, significant limitations remain. Many of these models still rely on single-source optical datasets such as Sentinel-2 and Maxar imagery, limiting the potential for leveraging more comprehensive insights of multisource remote sensing datasets. Furthermore, while cutting-edge methodologies such as Convolutional Neural Networks and vision transformers have demonstrated superior performance [18,25,26], they often function as “black boxes”, offering limited transparency regarding the contributions of individual data sources and features. In addition, incorporating multisource remote sensing data into these complex models can result in high computational costs, which pose practical challenges.

Our study addresses challenges in forest CHM estimation by focusing on two key areas: integrating diverse remote sensing data sources and enhancing ML model interpretability. We evaluate four widely used regression algorithms—K-nearest neighbors (KNN), random forest (RF), support vector regression (SVM), and extreme gradient boosting (XGB)—because of their proven accuracy across diverse data modalities [18,27–29]. Model interpretability is achieved using SHAP (SHapley Additive exPlanations), a cooperative game-theory framework that attributes each feature’s marginal contribution to a prediction [30]. We set out two main objectives: (1) evaluate the performance of ML models for forest CHM estimation using integrated airborne and spaceborne remote sensing data; and (2) enhance model interpretability and practical applicability by ranking feature importance, optimizing data integration, and assessing spatial distribution patterns of estimated forest CHM. The remainder of this paper is organized as follows: Section 2 introduces the study area, describes the multisource remote sensing dataset, and outlines the preprocessing methods used in this study. In Section 3, we present the model performance and SHAP-derived feature contributions of four ML algorithms (i.e., KNN, RF, SVM, and XGB) for estimating forest canopy height. Section 4 discusses the interpretations of data-driven insights, highlighting limitations and future challenges. Finally, Section 5 summarizes the key findings and contributions of this work.

2. Methods

2.1. Study Area

Our study site is located in southeastern North Carolina (NC), USA, covering an area of about 10,000 km² (Figure 1). Situated within the Southern Coastal Plain, this region is characterized by a rich diversity of forested ecosystems, including coastal plain forests,

bottomland hardwood forests, and pine forests predominantly consisting of Loblolly pine (*Pinus taeda*). These ecosystems provide vital habitats for various wildlife species and are crucial for maintaining the region's ecological balance and economic prosperity, with timberland constituting over 70% of the area in several counties. Given the recent history of extreme events triggered by tropical storms, notably Hurricane Florence in 2018, this area presents a unique opportunity to study the resilience of forest ecosystems in the face of natural disasters. Specifically, our research focuses on leveraging ML algorithms and multisource remote sensing data to enhance forest canopy height estimation. This approach aims to provide a more accurate assessment of forest structure and dynamics, crucial for understanding the impact of environmental stressors such as hurricanes on forest health and growth. Through accurate canopy height measurements across the southern part of North Carolina, we seek to help provide a more accurate understanding of how forest ecosystems respond to and recover from such events, offering insights that could guide future conservation and management efforts.

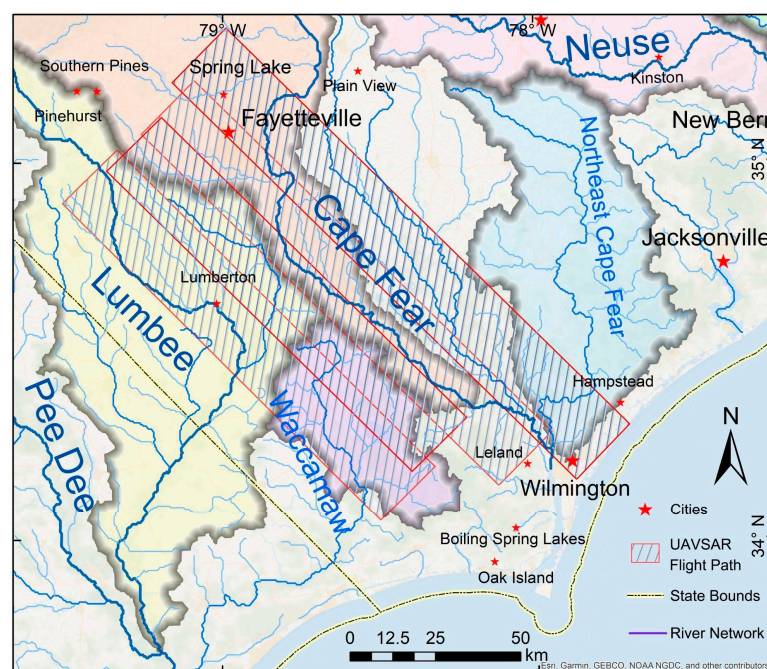


Figure 1. Study area location map showing 4 UAVSAR scenes (i.e., red polygons) collected in southeastern North Carolina on 18 September 2018.

2.2. Datasets

2.2.1. GEDI LiDAR Data and ALS Data

GEDI is equipped with full-waveform LiDAR sensors that capture a 4.2 km-wide swath along its orbit. It employs three lasers generating eight beams spaced about 600 m apart, each with a footprint diameter of roughly 25 m (Table 1) [13]. Version 2 of GEDI's L2A Elevation and Height Metrics products offers detailed canopy height data, with an improved geolocation accuracy of about 10 m [31]. This dataset includes latitude, longitude, elevation, and surface energy metrics from RH0 up to RH100 (i.e., the relative height at which all the energy is returned), with RH98 used in our study to represent forest canopy height, following Lang et al. [25].

Table 1. Summary of multi-sensor data sources, acquisition dates, spatial resolution, and its related processing techniques used for this study.

Source	Data Type	Variables Description	Acquisition Date	Spatial Resolution	Resampling Method	Processing Method
Sentinel-1	Spaceborne Radar	-C-band backscatter VV/VH	2016–2017	10 m	Bilinear interpolation	As per Bauer-Marschallinger et al. [32]
		-Radar InSAR Coherence	1 December 2019–30 November 2020	90 m	Bilinear interpolation	As per Kelldorfer et al. [33]
Sentinel-2	Spaceborne Optical	-Spectral bands (visible, NIR, red edge, SWIR) -Vegetation Index	2018–2019	10 m (visible, NIR); 20 m (red edge, SWIR)	Bilinear interpolation	Median composite of four seasonal cloud-free spectral bands and vegetation index
NAIP	Airborne Optical	-Four bands in the visible and NIR, -First two principal components -Derived texture variables	November 2018	0.3 m	Not applicable	PCA and textural calculations
UAVSAR	Airborne Radar	-L-band full polarimetric backscatter -Polarimetric decompositions variables	September 2018	5 m	Bilinear interpolation	As per Wang et al. [34]
PALSAR-2	Spaceborne Radar	-L-band HH/HV backscatter	2019	25 m	Bilinear interpolation	As per Shimada et al. [35]
ALS	Airborne Laser Scanning	-Elevation/Topographic index	2014	1 m	Not applicable	Mean aggregation over 25 m by 25 m grid in GEE
		-Canopy height model	2014	1 m	Not applicable	98th percentile over 25 m by 25 m grid in GEE
GEDI	Spaceborne LiDAR	-Canopy height model	2019–2020	25 m	Not applicable	Median composite of growing season data (2019–2020)

To minimize measurement variability and geolocation errors, we filtered GEDI data following the GEDI Level 2 User Guide [31]. Focusing on leaf-on conditions during the 2019–2020 growing season (June–August), we selected over 200,000 GEDI shots for robust analysis of forest canopy heights (Table 1). To ensure GEDI data accuracy, we compared it with high-resolution airborne laser scanning (ALS) data from Phase 2 and 3 airborne LiDAR data of the NCFMP (North Carolina Floodplain Mapping Program) [36,37]. We calculated the 98th percentile of ALS canopy heights within a 25 m by 25 m grid. Although the ALS dataset was collected in 2014 and the GEDI data in 2019–2020, this temporal mismatch does not undermine the analysis. Our goal was not to reassess GEDI’s absolute accuracy, but to evaluate whether systematic bias differs between full-power and coverage beams. Given the large and balanced sample sizes across beam types and their concentration in forested areas, any local forest changes over time are assumed to introduce random rather than systematic error, making the ALS data suitable for this comparison (Tables S1 and S2). We also compared GEDI’s full-power and coverage beams to ensure that discrepancies stem from inherent beam characteristics rather than temporal changes in forest structure observed by ALS and GEDI (Figures S1 and S2).

2.2.2. SAR Dataset from Sentinel-1, PALSAR-2 and UAVSAR

Sentinel-1, a key component of the Copernicus Programme’s Sentinel missions and managed by the European Space Agency (ESA), is equipped with a C-band SAR instrument. This advanced technology enables the acquisition of detailed SAR images of the Earth’s surface, even in challenging weather and low-light conditions. For our study, we used the normalized Sentinel-1 Global Backscatter products, which contain dual polarizations (VV and VH backscatter, Table 1), as processed by Bauer-Marschallinger et al. [32]. The data aggregation of Sentinel-1 radar scenes covers the period from 2016 to 2017. This was followed by the normalization of backscatter to a reference incidence angle (38°) using a linear regression model. The process also includes quality checks, noise removal, and harmonization of data from multiple orbits to reduce artifacts and ensure consistency in the mosaic. Additionally, we incorporated the seasonal Sentinel-1 InSAR coherence

dataset preprocessed by Kelldorfer et al. [33] as part of our predictor features. This dataset includes median seasonal coherences calculated from 6-, 12-, 18-, 24-, 36-, and 48-day repeat-pass intervals at VV polarization.

ALOS-2/PALSAR-2, operated by Japan Aerospace Exploration Agency (JAXA), provides L-band SAR backscatter data, including HH (co-polarized) and HV (cross-polarized) polarization. For our study, we used the “Global PALSAR-2/PALSAR Yearly Mosaic (v2)” product, which compiles seamless global backscatter data (gamma-naught) at a 25 m resolution, acquired from PALSAR-2 SAR strip images (Table 1). The dataset selected for this analysis corresponds to the year 2019, ensuring relevance and timeliness of the data used.

Uninhabited Aerial Vehicle Synthetic Aperture Radar (UAVSAR), a fully polarimetric airborne L-band radar instrument, is equipped on the NASA Gulfstream-III aircraft operated by the Armstrong Flight Research Center. UAVSAR operates within a side incidence angle range from about 22° to 67° , covering a 22 km swath in the cross-track direction and extending up to 300 km in the direction of flight. The spatial resolution of UAVSAR is relatively high at about 5 m [38]. In this study, we used multi-looked and ground range projected UAVSAR imagery products and focused on four overlapping flight paths (i.e., 13,510, 13,511, 31,509, and 31,510; Figure 1) acquired on 18 September 2018 (Table 1). These flight lines were selected because they cover southeastern NC, which corresponds to the area of the Cape Fear, Lumbee, and Waccamaw Rivers and their tributaries. We addressed the issue of the decreasing illumination gradient across the span from the near range to the far range (Figure 2A), due to the side-looking geometry of SAR, as described by Ulaby et al. [39]. Following the UAVSAR processing methodology detailed by Wang et al. [34], we conducted radiometric normalization on each polarization of backscatter data, as well as on the polarimetric decomposition parameters derived from UAVSAR. These adjustments effectively corrected the illumination issue and produced seamless UAVSAR mosaic images, as illustrated in Figure 2B.

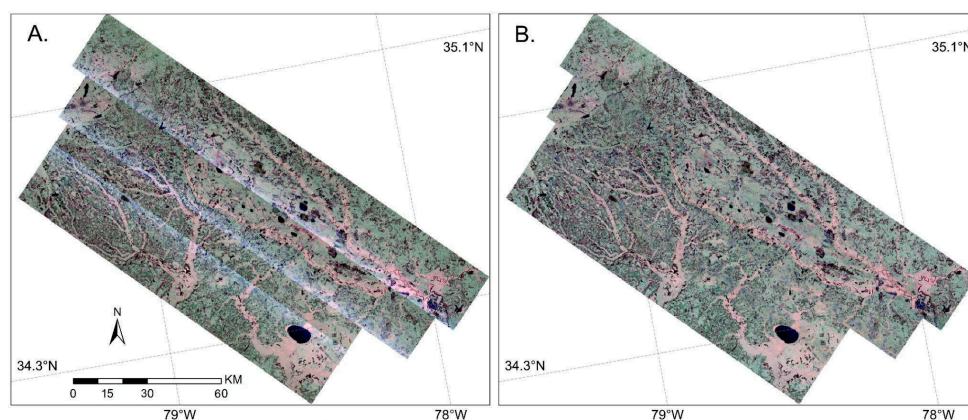


Figure 2. UAVSAR Pauli decomposition before (A) and after (B) correction, with the RGB color model used to represent different scattering mechanisms: red for single bounce, green for double bounce, and blue for volume scattering.

The polarimetric decomposition parameters include three derived from Pauli decomposition (PD), three derived from Freeman–Durden decomposition (FDD), and four derived from Cloude–Pottier decomposition (CPD) methods [40,41].

2.2.3. Optical Dataset from NAIP and Sentinel 2

The NAIP (National Agriculture Imagery Program) is a program run by the U.S. Department of Agriculture (USDA) that acquires high-resolution 4-band visible–infrared aerial imagery of the United States from a low altitude. The NAIP imagery is typically collected at a spatial resolution of 1 m or better and serves multiple purposes, such as

mapping and monitoring agricultural land, forests, and other natural resources. We first conducted Principal Component Analysis (PCA) on NAIP aerial imagery from Fall 2018, extracting the first two principal components (Table 1). Following the methodology detailed by Ehlers et al. [16], we derived texture information related to crown diameter by calculating the ratio of variances at 2 m to variances at 3 m within each 24 m grid for every principal component, then resampled the results to 25 m [42,43].

Sentinel-2 is a satellite mission operated by ESA as part of its Copernicus Programme. It collects multispectral optical imagery of the Earth's surface from space and has wider spatial coverage, with a spatial resolution of up to 10 m. Additionally, the multispectral instrument onboard Sentinel-2 captures spectral information across multiple bands, facilitating the identification of spectral bands closely associated with canopy height variations. To reduce temporal mismatch, we processed the Sentinel-2 data collected from 2018 to 2019 and acquired seasonal composited spectral bands and vegetation indices via Google Earth Engine (GEE) (Table 1). Specifically, we divided cloud-free Level-2A Sentinel-2 surface reflectance images into four seasonal composites: spring (1 March–31 May), summer (1 June–31 August), fall (1 September–30 November), and winter (1 December–28 February). For each season, we generated median composites across all spectral bands to mitigate the effects of cloud contamination, atmospheric variability, and sensor noise, thereby ensuring stable and representative reflectance values throughout the year.

2.2.4. Multisource Remotely Sensed Data Harmonization

To ensure spatial alignment across all data sources, each dataset was reprojected to a common coordinate reference system: EPSG: 32617 (WGS 84/UTM Zone 17N). All layers were then resampled to a 25 m spatial resolution, corresponding to the native footprint diameter of GEDI observations, to enable pixel-level correspondence and integration within GEE (Table 1). Specifically, continuous variables, such as Sentinel-2 surface reflectance, Sentinel-1 backscatter (σ^0), and UAVSAR polarimetric products, were resampled using bilinear interpolation. This method minimizes aliasing and preserves radiometric integrity across spatial scales. For lower-resolution datasets (e.g., 90 m), such as Sentinel-1 InSAR coherence products, we also applied bilinear interpolation to downscale them to 25 m. While this process does not introduce additional spatial detail, it maintains a continuous surface suitable for ML inputs and prevents sharp artificial edges that may arise with nearest-neighbor resampling. We did not choose bicubic interpolation, which, while producing smoother results, is more computationally intensive and risks over-smoothing fine-scale features. In contrast, very high-resolution datasets, including NAIP 1 m aerial imagery and ALS CHM, were aggregated to 25 m $\times$ 25 m blocks using different strategies suited to their data characteristics. For instance, ALS CHM data were calculated as the 98th percentile of 1 m CHM values within each grid cell measuring 25 m by 25 m in GEE. This percentile-based aggregation preserved the vertical structure relevant to canopy height while aligning spatially with GEDI footprints.

2.3. Model Building and Feature Group Analysis

All datasets were re-projected to the same coordinate reference system EPSG: 32617 (WGS 84/UTM Zone 17N) and resampled to a common 25 m spatial resolution to align with the footprint resolution of GEDI. Bilinear interpolation was applied for continuous data, such as spectral bands and SAR backscatter coefficients, to maintain data integrity during resampling. For higher-resolution datasets, such as NAIP aerial imagery and the LiDAR-derived DEM, block aggregation was used to ensure spatial alignment with the other datasets.

We used canopy height data derived from GEDI, integrating the predictor variables outlined in Section 2.2 to create our training and test datasets. After removing NaN (Not-a-Number) values, 165,580 GEDI footprints remained for model development, and 38,198 additional footprints were held out as the independent validation dataset. For the development of our ML models, the data footprints were randomly divided (Table 2): 80% were allocated for training, while the remaining 20% were reserved for testing the model's performance.

Table 2. Overview of the datasets used for training and testing the ML models (independent validation samples were filtered based on 30 m spatial distance with training samples).

ID	Set Name	Data Percentage	Sample Count	Usage Description
1	Training	$80\% \times 80\% = 64\%$	132,464	Used for training the ML models
2	Testing	$80\% \times 20\% = 16\%$	33,116	Used for evaluating the ML models on unseen data without considering spatial autocorrelation
3	Independent validation	20%	38,198	Used for validating the models on unseen data with spatial independence and for calculating performance metrics

We trained our canopy-height models using four widely used regression algorithms—KNN, RF, SVM, and XGB. The KNN, RF, and SVM models were implemented using scikit-learn v1.3.2, whereas XGB model used the xgboost v2.0.3 library. Hyper-parameters were optimized through an exhaustive grid search with five-fold cross-validation, simultaneously maximizing validation R^2 and minimizing RMSE.

To evaluate the contribution of diverse remotely sensed data to canopy height estimation, we trained a series of XGB models on (i) individual feature groups and (ii) their pairwise and cumulative combinations. Performance was assessed on the independent test and validation sets with RMSE and R^2 for each group. Additionally, we conducted a comparative analysis between our studies and existing global canopy height products from prior studies by Potapov et al. [19], Lang et al. [25], and Tolan et al. [26], aiming to evaluate the accuracy and reliability of our models.

2.4. Rationalizing Predictions

Although it is a challenging task to understand the complexities of ‘black-box’ ML models, there are several techniques available that offer insight into the significance of different features. For instance, the SHAP values method, based on cooperative game theory, can be used effectively to quantify each feature's contribution to the model's predictions [30]. Specifically, features with wider SHAP value ranges are considered higher influential features, as their variability significantly affects model outputs. In our study, we adopted the Shapley values approach via the Python package “shap” (v 0.45.0) to investigate how attributes derived from multisource remotely sensed data contribute to canopy height estimation.

While SHAP values are invaluable for deciphering feature importance, they do not directly account for the potential inaccuracies that individual features may introduce into the model. To address this limitation, we derived two additional metrics from the SHAP values calculated using the XGB model: prediction contribution and error contribution. Prediction contribution measures the positive impact of features on model accuracy, while error contribution quantifies the extent to which features contribute to errors in predictions. We assessed these metrics throughout both the training and validation phases of the XGB model, enhancing our ability to understand and improve how each feature influences the model's performance on unseen datasets.

3. Results

3.1. Model Results

Figure 3 presents validation scatter plots of four ML forest canopy height estimates: KNN, RF, SVM, and XGB. Each algorithm used an all-inclusive feature set, designated as Group 14 in Table 3. The results revealed a consistent level of accuracy across the models, indicated by R^2 values of 0.496, 0.510, 0.544, and 0.548 for KNN, RF, SVM, and XGB, respectively. While the performance of all four models was comparable, the SVM and XGB algorithms demonstrated a marginally higher accuracy, followed by RF with moderate accuracy, and KNN lagged slightly behind. The accuracy of each model is also supported by their RMSE values: 5.13 m for KNN, 5.06 for RF, 4.88 for SVM, and 4.85 for XGB. Additionally, the scatter plots in Figure 3 revealed a consistent pattern of estimation errors across all models, i.e., overestimate shorter forest canopies and underestimate taller ones. The performance metrics of various remotely sensed feature groups on XGB CHM model accuracy (as reflected by RMSE and R^2 values).

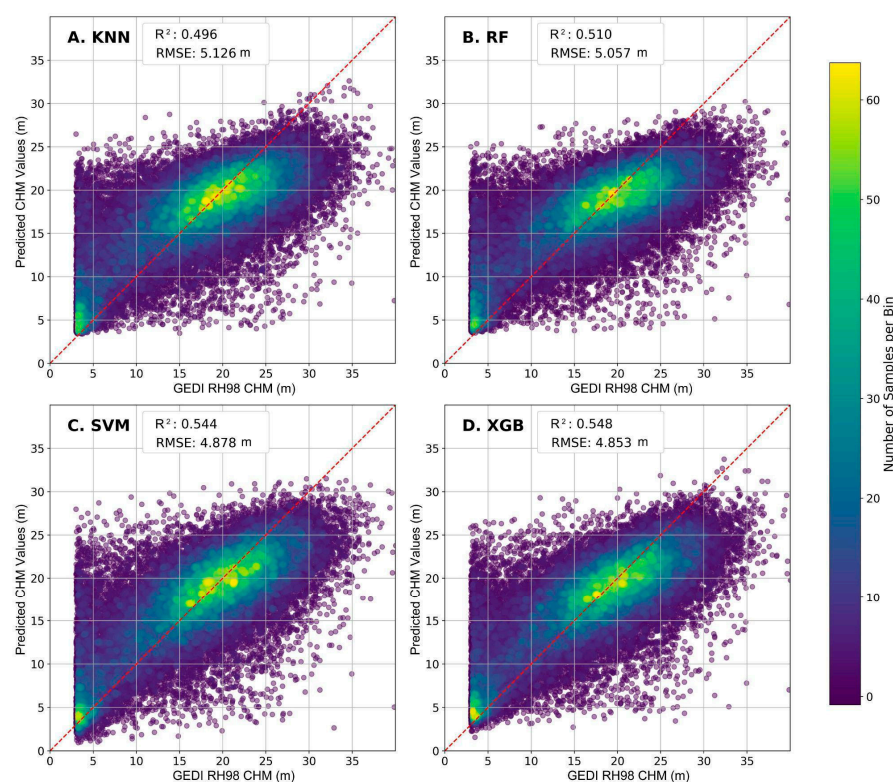


Figure 3. The scatter plots of predicted vs. GEDI-derived RH98 CHM by four ML algorithms: (A) for KNN, (B) for RF, (C) for SVM, (D) for XGB (the red dashed line marks the 1:1 relationship; each plot is labeled with the corresponding RMSE and R^2).

Table 3. The performance metrics of various remotely sensed feature groups on XGB CHM model accuracy (as reflected by RMSE and R^2 values).

Group No.	Feature Group Composition	RMSE	R^2
1	Optical features from Sentinel-2	5.086	0.504
2	Optical features from Sentinel-2 with Terrain	5.030	0.515
3	Optical features from NAIP	6.793	0.115
4	Combined Optical features from Sentinel-2 and NAIP	5.061	0.509
5	SAR features from Sentinel-1 backscatter	6.953	0.073
6	SAR features from Sentinel-1 coherence	6.652	0.151
7	Combined SAR features from Sentinel-1 coherence and backscatter	6.479	0.195

Table 3. Cont.

Group No.	Feature Group Composition	RMSE	R ²
8	SAR features from PALSAR-2	6.718	0.135
9	SAR features from UAVSAR	6.407	0.213
10	Combined SAR features from Sentinel-1, PALSAR-2, and UAVSAR	5.949	0.321
11	Combined Optical (Sentinel-2), SAR (Sentinel-1), and Terrain	4.946	0.531
12	Combined Optical (Sentinel-2), SAR (UAVSAR), and Terrain	4.954	0.529
13	Terrain features	7.196	0.007
14	All Inclusive (Sentinel-2, NAIP, Sentinel-1 backscatter and coherence, PALSAR-2 backscatter, UAVSAR, and Terrain features)	4.853	0.548

3.2. Impact of Diverse Remotely Sensed Feature Groups on Estimating Canopy Heights

This section presents the performance of various remotely sensed feature groups in estimating canopy height using the XGB algorithm. Results from optical (Sentinel-2 and NAIP), SAR (Sentinel-1, PALSAR-2, and UAVSAR), and terrain data, as well as their combinations, are summarized in Table 3, highlighting their contribution to model accuracy.

3.2.1. Optical Data Performance

Using only Sentinel-2 optical data (Group 1, Table 3) effectively estimated forest canopy height, achieving an RMSE of 5.086 m and an R² of 0.504, as shown in Table 3. Including terrain variables with Sentinel-2 data (Group 2) slightly enhanced the model's accuracy, reducing the RMSE to 5.030 m and increasing the R² to 0.515. This improvement underscores the added value of integrating terrain information to boost optical data performance. Conversely, models exclusively based on terrain variables (Group 13) exhibited the weakest performance, with an RMSE of 7.196 m and an R² of merely 0.007. Combining NAIP with Sentinel-2 data also slightly improved model accuracy compared to Sentinel-2 alone, by lowering the RMSE to 5.061 m and raising the R² to 0.509. However, relying only on NAIP imagery (Group 3) yielded a higher RMSE of 6.793 m and a notably lower R² of 0.115, indicating its limited effectiveness for canopy height estimation when used alone.

3.2.2. SAR Data Performance

The effectiveness of using only SAR data to estimate canopy height was found to be limited. The use of Sentinel-1 backscatter data alone (Group 5) resulted in an RMSE of 6.953 m and an R² of 0.073, the lowest among the SAR features (Table 3). This indicates its limited capability for estimating canopy height independently in our study area. In comparison, L-band SAR backscatter data from PALSAR-2 (Group 8, RMSE: 6.718 m, R²: 0.135) and UAVSAR (Group 9, RMSE: 6.407 m, R²: 0.213) each demonstrated better performance over the C-band SAR backscatter data. Using Sentinel-1 coherence data alone (Group 6) yielded better performance compared to backscatter alone, with an RMSE of 6.652 m and an R² of 0.151. This suggests that coherence features offer an enhanced value in modeling efforts. The integration of Sentinel-1 backscatters and coherence data (Group 7) further enhanced model performance, reducing the RMSE to 6.479 m and increasing the R² to 0.195. This improvement highlights the advantage of integrating both backscatter and coherence SAR features. The integration of Sentinel-1 with PALSAR-2 and UAVSAR SAR data (Group 10) notably enhanced the model's accuracy, achieving an RMSE of 5.949 m and an R² of 0.321.

3.2.3. Integrated Data Performance

Our analysis highlighted the beneficial synergy of integrating optical and SAR datasets with terrain variables. The fusion of Sentinel-2 optical data, Sentinel-1 SAR data, and terrain variables (Group 11) yielded better performance against any evaluated individual

feature groups, with an RMSE of 4.946 m and an R^2 of 0.531. This setup was marginally outperformed by the integration of S2 optical data, UAVSAR SAR data, and terrain variables (Group 12, RMSE: 4.954 m, R^2 : 0.529), which demonstrated comparable accuracy. Notably, the most comprehensive model (Group 14), incorporating all optical, SAR, and terrain features, emerged as the top performer among all evaluated feature groups, achieving an RMSE of 4.853 m and an R^2 of 0.548.

3.3. Rationalizing Predictions by Ranking Variable Importance for CHM Models

In our study, the SHAP summary plots (Figure 4) provided a visual exploration of how individual features influence the predictions of each ML model. The features along the y-axis were sorted in descending order based on their impact magnitude, with each feature's SHAP value calculated for every sample. The color-coded points on the plot illustrate that both the high and the low values of features can drive the model's output, with the color indicating whether the feature value of a specific sample is relatively high or low compared to the rest of the samples. Through this analysis, we observed insightful similarities and contrast patterns of feature importance across the KNN, RF, SVM, and XGB models.

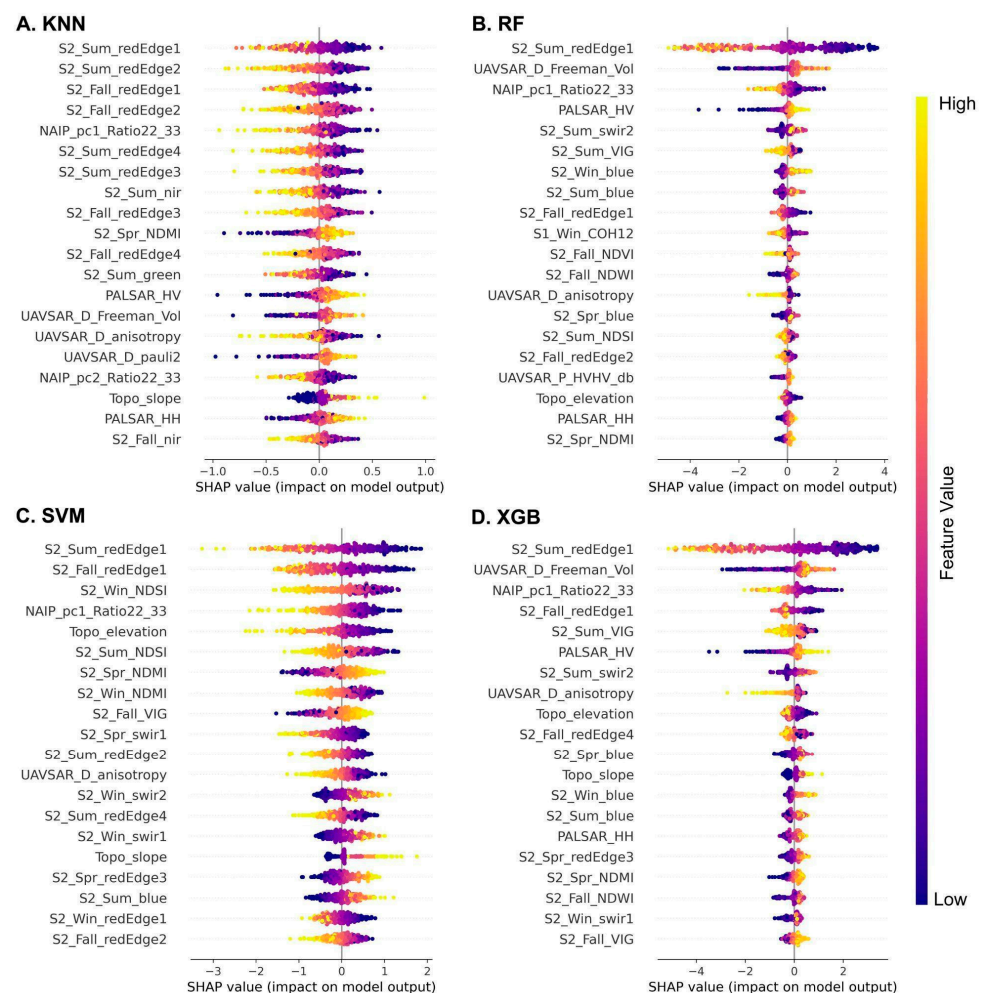


Figure 4. Feature importance in ML models assessed with SHAP values. Each subplot ((A) for KNN, (B) for RF, (C) for SVM, (D) for XGB) showcases the influence of individual features on the model's predictions, with each dot on each plot representing a specific sample. The y-axis features were organized in descending order based on the mean absolute SHAP values.

Notably, the Sentinel-2 red-edge band 1 composite in summer emerges as a consistent influencer across all models, although it varies. Similarly, the texture derived from the first principal component of NAIP stood out as a consistently important feature, securing a position within the top five for its impact in the RF, SVM, and XGB models. This highlights its robust contribution to model outputs. The FDD-derived volume scattering from UAVSAR data shows a considerable wide range of SHAP values in RF and XGB models. This spread reflects its variable, yet substantial effects. It indicates the generalizable predictive power of these models in estimating canopy height. Conversely, some features display a distinct model-specific relevance. For instance, the PALSAR-2 HV polarization band significantly impacts the RF and XGB models, but is less influential in others, whereas topographic elevation is emphasized in the SVM model. This illustrates how different models prioritize and interpret features uniquely.

Diving deeper into the behavior of individual features, such as the CPD-derived anisotropy and FDD-derived volume scattering features from UAVSAR data, we observed interesting influences across the models. In the KNN model, the CPD-derived anisotropy feature plays a substantial but not overwhelming role, with a moderate range of SHAP values. In contrast, the FDD-derived volume scattering feature in the RF and SVM models displays a broader spread of values, indicating a more impactful role in determining the model's output. In the SVM model, the FDD-derived volume scattering feature shows a more consistent but limited influence, while the XGB model demonstrates a higher sensitivity to both features, with a wider range of SHAP values. These differences highlight how each model uniquely processes and interprets features, especially when dealing with complex features such as CPD-derived anisotropy and FDD-derived volume scattering. In addition, the rankings of the Sentinel-1 winter-season coherence feature, based on a 12-day repeat cycle, indicated an intermediate importance in the RF model and even lesser impact in the other models, underlining the contrasting ways in which temporal SAR coherence data contributed to predictive modeling.

To move beyond conventional perspective of feature importance rankings, we examined both how much each feature contributes to the model's predictions and how much it contributes to the residual error. Figure 5 visualizes this dual relationship, prediction contribution versus error contribution, for both training and validation datasets. From this analysis, we observed that most features with higher Prediction Contributions typically have lower Error Contributions, reflecting their robust contribution on model accuracy. The general patterns observed in this analysis align well with our initial findings from the SHAP summary plots. For instance, the feature of Sentinel-2 red-edge band 1 composite in summer acts as the most important factor in both training and validation datasets. In addition, the zoomed-in panels (Figure 5B,D) illuminated the detailed impacts of various features, showcasing their subtle yet crucial roles in the model's predictive performance and generalization ability. For instance, FDD-derived volume scattering from UAVSAR data ranks as the second most important feature in terms of both Prediction and Error Contributions in the training dataset. Although its error contribution remains consistent, its prediction contribution drops to third in the validation dataset. Similarly, the PALSAR-2 HV polarization band, initially the third most important feature for error contribution and fifth for prediction contribution in the training dataset. However, in the validation dataset, the error contribution drops to ninth, and its prediction contribution falls to twelfth.

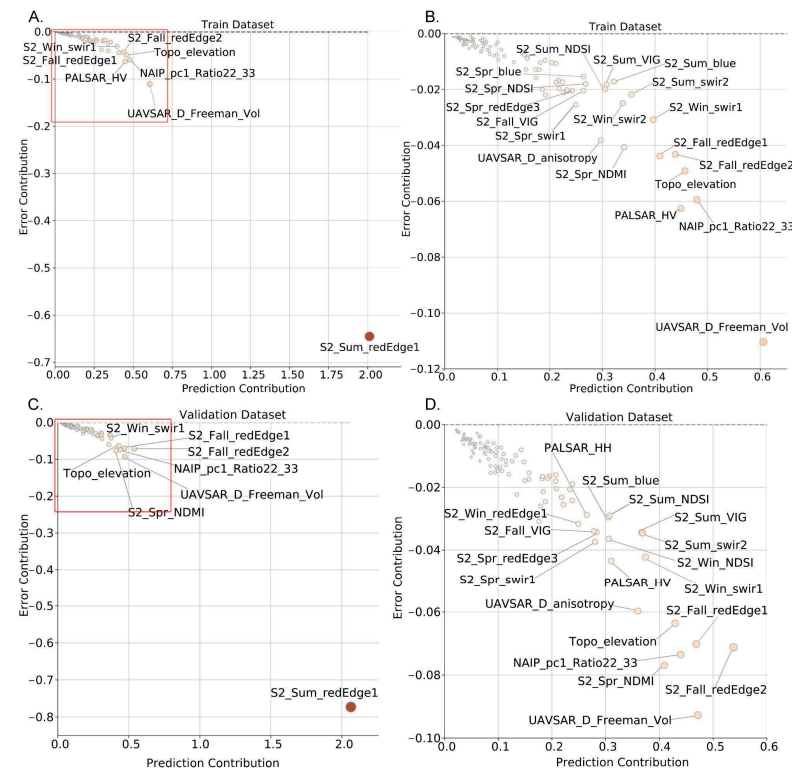


Figure 5. Prediction contribution versus error contribution on the training (A,B) and validation (C,D) datasets from the XGB model. The left panels display an overall view, whereas the right panels offer a detailed view of the region outlined by the red box in the left panels, allowing for a closer examination of features with less pronounced but contributes to the model's predictive accuracy. The size and color of the dots correspond to their prediction contribution.

4. Discussion

4.1. Spatial Distribution of CHM and Comparison with Existing Products

The spatial distribution of forest canopy height estimations, derived from four distinct ML algorithms, demonstrated remarkable consistency (Figure S3). This agreement among the outcomes from the four ML algorithms underscores the robustness of our developed models. The canopy height maps generated in this study accurately capture and illustrate the variations in canopy height across the study area, offering a detailed visual representation of the forest's upper canopy layer (Figure 6). Notably, areas depicted in yellow in Figure 6, indicating taller canopy heights, are predominantly found clustering around river floodplains, forming critical ecological corridors that enhance conservation efforts [44]. This spatial pattern aligns with ecological expectations that riverine environments, due to enhanced water availability and fertile soil conditions, typically support taller vegetation [45]. For instance, aligning with existing canopy height maps, our spatial distribution maps clearly identified elevated canopy heights within the Waccamaw River watershed (as shown at the center bottom of Figure 6), confirming the spatial distribution's accuracy.

To facilitate a comparative analysis of canopy height estimations, we created a detailed scatter plot that compares measured canopy heights by GEDI with estimated canopy heights from both our study and existing products for the validation samples, as illustrated in Figure 7. We used the GEDI RH98 CHM data from the validation dataset as a reference and compared it against three recently developed global-scale canopy-top height maps from Potapov et al. [19], Lang et al. [25], and Tolan et al. [26]. Our findings were promising: the canopy height estimation from XGB closely aligned with GEDI RH98 CHM measurements, achieving an impressive highest R^2 of 0.55 and the lowest RMSE of 4.86 m. In comparison, the global product developed by Potapov et al. [19] showed weak alignment

with the reference CHM from GEDI RH98 in our study area. Potapov's model, potentially constrained by its reliance on the RF regression model which struggles to capture high canopy heights, achieved a lower R^2 of -0.056 (Figure 7A). Lang et al. [25]'s model showed moderate performance with an R^2 of 0.062 . Tolan's model performed better, with an R^2 of 0.24 , indicating better agreement with GEDI canopy height measurements and suggesting its usefulness in global assessments that require some localized accuracy. Notably, the global canopy height maps produced by Potapov et al. [19] and Tolan et al. [26] were based on the GEDI RH95 CHM data. In contrast, our study, along with that of Lang et al. [25], employed the GEDI RH98 CHM to represent the top canopy.

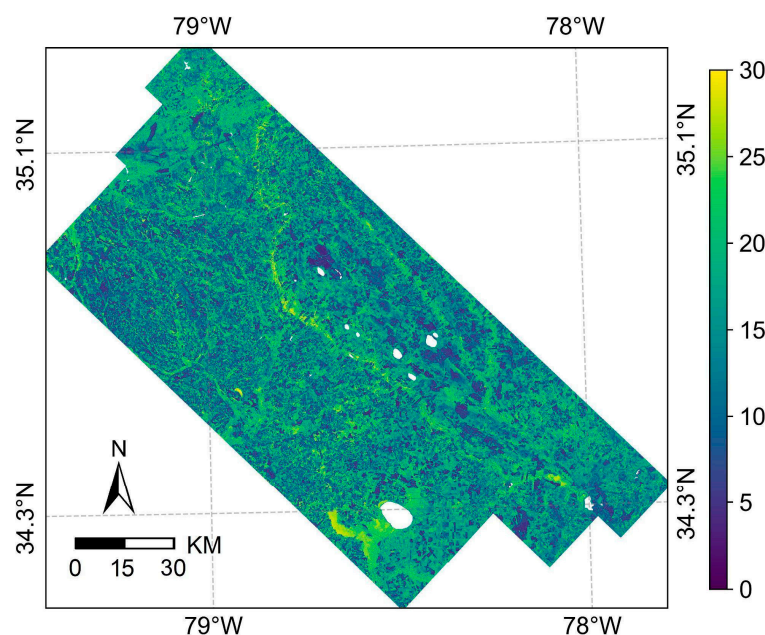


Figure 6. The spatial distribution of canopy heights predicted by XGB.

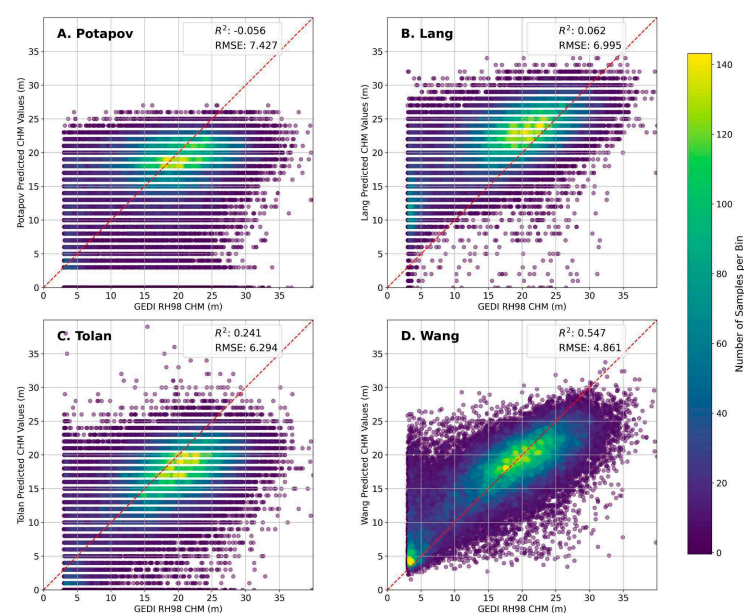


Figure 7. Scatter plots comparing canopy height estimates from various models with canopy heights measured by GEDI. Each subplot corresponds to different model: (A) Potapov et al. [19], (B) Lang et al. [25], (C) Tolan et al. [26], and (D) our XGB model (Wang). These plots allow for direct visual evaluation of model performance, which is quantitatively supported by the calculated R^2 and RMSE values for each panel.

4.2. Advantage of GEDI for CHM Mapping over Southeastern NC

Though GEDI posts are spatially discontinuous, it provides densely sampled LiDAR observations on a global scale, with 25 m diameter footprints. During its two-year mission, GEDI remarkably provided approximately 10 billion cloud-free observations [13]. This high-density sampling, enabled by GEDI's advanced power and coverage beams, makes it an invaluable tool for assessing vegetation canopy heights. In addition, the availability of a consistent GEDI dataset from 2019 onwards mitigates temporal alignment challenges, thus enhancing the integration process with multisource remotely sensed data, as noted by Tamiminia et al. [15]. This consistency is crucial for large-scale studies. It also simplifies the integration of GEDI data with datasets from recently launched satellite missions, offering a more cohesive approach to multisource remotely sensed data analysis.

The quality of GEDI data products is influenced by many factors, making data quality filtering a necessary step [15,46]. However, the choice of filtering criteria is challenging. Overly rigorous quality control measures may substantially reduce the sample diversity of GEDI data, thereby leading to incomplete exploitation of its capabilities. For instance, many studies have raised concerns about the lower quality of coverage beam data compared to power beam data and simply choose to disregard the coverage data [19,47]. Despite these concerns, coverage beam data may still prove useful in certain areas. Specifically, we found a statistically significant difference in canopy height accuracy between full-power and coverage beams, but no practically meaningful difference (Tables S1 and S2, Figure S2), suggesting that both beam types can be effectively used in canopy height modeling within our study area. This result challenges the practice of restricting analyses to higher sensitivity power beam data, which effectively excludes nearly half of the available GEDI observations [48]. Such practices lead to the exclusion of nearly half of the valuable GEDI data that could otherwise enrich CHM mapping efforts. Recently, some strategies were developed to fully use GEDI data. For instance, Wang et al. [49] proposed incorporating all high-quality GEDI observations with existing canopy height maps to maximize the utility of GEDI data, aiming for enhanced accuracy in comprehensive CHM mapping. In light of our findings and literature [46], we recommend an adaptation strategy by pre-evaluating both power and coverage beams to ensure the maximization of GEDI data usage over large-scale studies.

4.3. Multisource Remotely Sensed in Estimating Canopy Height

Our study has demonstrated the advantages of integrating GEDI data with multisource remotely sensed data, including optical and SAR technologies such as NAIP, Sentinel-2, Sentinel-1, PALSAR-2, and UAVSAR, for mapping CHM. This robust approach enabled the production of spatially continuous CHM maps. The generated CHM maps show improved accuracy over existing global canopy height datasets in the forested regions of southeastern North Carolina (Figure 7). While the integration of a full suite of multiple data sources achieved the highest accuracy (RMSE: 4.853, R^2 : 0.548) in the canopy height maps, the increase was modest, with the R^2 value showing a slight improvement of 0.04 compared to models that used Sentinel-2 data alone. Despite this relatively slight increase, the resulting canopy height map leverages the unique strengths of each sensor, demonstrating better overall performance than maps created from single-source data, as evidenced by the lowest RMSE and an improved R^2 value among the models evaluated (Table 3).

A significant contribution of this study lies in our use of explainable ML methods, specifically SHAP analysis, to interpret the feature importance of diverse remotely sensed data in canopy height estimation. By applying SHAP to our XGB model, we were able to clearly identify the most significant contributions features, including the summer composite of Sentinel-2's red-edge bands and UAVSAR's volume scattering (Figure 4). These features

play a crucial role in improving the accuracy of CHM estimation, as the red-edge bands help distinguish vegetation types and stages of growth, while UAVSAR's volume scattering provides key insights into forest structure, including canopy complexity and biomass. This emphasis on explainable ML not only enhanced model transparency but also helped us uncover previously unexplored relationships between diverse data sources and canopy height. For instance, while the importance of Sentinel-2 red-edge bands and seasonal spectral infrared bands from Landsat has been recognized in prior studies [21,50,51], our analysis is, to the best of our knowledge, the first to highlight the significance of seasonal red-edge variation in improving CHM models. Similarly, the detailed structural information captured by UAVSAR's polarimetric SAR data adds a new dimension to canopy height estimation, going beyond what traditional SAR data alone can offer. Additionally, the inclusion of NAIP data in our analysis provides ultra-high-resolution crown diameter related texture features with that complement freely available Sentinel-2 satellite datasets. These findings fill a gap in literature by linking explainable model outputs to specific remote sensing features, offering actionable insights into data fusion strategies for future studies.

Integrating optical data from Sentinel-2 and terrain variables with SAR data from Sentinel-1 (Group 11 in Table 3) or UAVSAR (Group 12 in Table 3) has yielded only marginal improvements in forest canopy height prediction compared to using Sentinel-2 data and terrain variables alone (Group 2 in Table 3). Although the combination benefits from the canopy penetration capabilities of L-band SAR data from UAVSAR, which can provide a comprehensive view of forest structures, the enhancements in accuracy remain slight. This observation aligns with findings reported by Sothe et al. [11], who noted that including L-band SAR PALSAR-2 data could improve models, but not dramatically. Given these outcomes, while the integration of SAR and optical data can slightly enhance CHM accuracy, the modest benefits should be carefully weighed against the simplicity and lower implementation costs of relying solely on Sentinel-2 data. However, explainable ML techniques such as SHAP allow us to better understand why certain datasets, such as UAVSAR and NAIP, contribute more to specific forest conditions, guiding more efficient data fusion efforts in the future. Users in regions such as Eastern North Carolina should assess whether the advantages of a multisource approach justify the additional complexity and costs.

Beyond the dual polarization SAR data provided by Sentinel 1 or PALSAR-2, UAVSAR's polarimetric SAR (PolSAR) data introduces an additional layer of depth to vegetation SAR backscatter analysis [23]. UAVSAR offers unparalleled insights into forest structure interactions with electromagnetic waves, marking a major advancement in our modeling efforts [52]. By differentiating between various scattering mechanisms within the canopy, such as surface, double-bounce, and volume scattering, it provides unique perspectives on forest structure [53]. Each mechanism reveals relevant information of the specific forest's physical attributes, from the ground interface to the complex interactions within the tree trunks and canopies [54]. Our analysis on Section 3.2.2 demonstrates the advantage of integrating multiple SAR data sources to refine canopy height estimation models. For instance, the incorporation of UAVSAR data further enriches the accuracy and reliability of CHM mapping by offering detailed information about the structural complexity of forests. Our use of UAVSAR sets a foundation for future studies, as the forthcoming NISAR mission will offer quad-pol L-band SAR images over some targets, allowing our framework to be directly adapted and expanded for broader canopy height estimation applications.

Our analysis of multisource remotely sensed data revealed a noteworthy role for coherence data from the winter season Sentinel-1 InSAR in the RF model. Although the Sentinel-1 winter-season coherence feature, based on a 12-day repeat cycle, demonstrated only slight importance, its impact, while modest, illuminates the complex interactions

between temporal SAR coherence data and predictive modeling, particularly within deciduous forests. The seasonal coherence variations are likely linked to phenological changes, suggesting that even subtle SAR data characteristics can inform our understanding of canopy dynamics. This finding underscores the value of continued exploration of multitemporal SAR observations, especially in anticipation of future missions such as the NISAR, which, with its longer wavelength, is expected to offer deeper insights into forest canopy dynamics. These insights could aid in significantly improving forest structure estimation. For instance, understanding how temporal coherence changes in SAR observations can enhance the predictive accuracy of canopy height models by accurately capturing the seasonal phenological patterns of different tree species.

4.4. The Performance of Different ML Models in Canopy Height Estimation

The choice and configuration of ML models are crucial for the accuracy of estimations. In this study, we evaluated the performance of our ML models: KNN, RF, SVM, and XGB, in estimating CHM using multisource remotely sensed data. While all models performed relatively well, slight differences in accuracy were observed (Figure 3). Despite its slightly lower accuracy, the inefficiency of KNN during the prediction phase in high-dimensional space makes it the least competitive option [29]. SVM, adapted for regression through SVR (support vector regression), demonstrated significant effectiveness in modeling complex relationships between remotely sensed features and CHM, even surpassing the performance of RF. This superior performance can be largely attributed to its specialized kernel design, which effectively captures complex patterns in the high-dimensional data, as noted by [28]. However, the absence of parallel processing support for SVM within the open-source “sklearn” package significantly slows its computational efficiency. Despite the intermediate accuracy level of the RF model, the reliability and computational efficiency of the RF model should not be discounted [55]. Due to its robustness and ease of use, the RF model offers a balanced approach to tackling the multiple challenges posed by diverse remotely sensed data. XGB emerged as the best performance model due to its ability to handle complex, nonlinear relationships inherent in multisource remotely sensed data, a scenario often encountered in dense forested areas with substantial canopy height variability. This can be attributed to XGB’s gradient boosting framework. Unlike RF, which builds trees in parallel, XGB benefits from a sequential tree-building algorithm that corrects the residuals of previous ones. This iterative correction of residuals enables XGB to adapt more accurately to complex data patterns, such as those encountered in canopy height estimation. The inclusion of regularization terms in XGB helps prevent overfitting, a significant advantage when dealing with the diverse multisource remotely sensed data used in this study.

In addition to evaluating ML model performance, we adopted explainable ML approaches by applying SHAP analysis to assess the contribution of individual variables addressed. Our study contributes to the field by demonstrating the value of explainable ML in multisource remote sensing applications for forest canopy height estimation. By integrating a variety of data sources and emphasizing model interpretability, we provide actionable insights that can inform forest management practices and environmental monitoring efforts. Our findings bridge the gap between high-performance models and the need for transparency, offering a balanced perspective on both innovation and practical application.

4.5. Limitation, Uncertainty, and Future Directions

Generating spatially continuous forest canopy height maps from multisource remotely sensed datasets involves several challenges and uncertainties. The accuracy of our height estimations relies heavily on the quality of GEDI data and the effectiveness of ML models in linking GEDI measurements to ancillary remote sensing variables. Dubayah et al. [13]

and subsequent studies, such as those by Adam et al. [56] and Potapov et al. [19], have pointed out specific limitations of GEDI. One key issue is that GEDI's laser pulses, while capable of penetrating the forest canopy, often fail to reach the ground, and thus tend to underestimate the height of taller trees. Adam et al. [56] found underestimation of deciduous forests that have closed canopies with fewer gaps for the laser impulse to penetrate. Additionally, due to its coarse footprint and horizontal geolocation issue, GEDI may not accurately capture canopy height variability in heterogeneous landscapes. In our analysis, we used Version 2 of GEDI data, which incorporates significant geolocation error corrections compared to Version 1. According to Tang et al. [57], these corrections have reduced geolocation errors to within 10 m, which is well below the 25 m resolution used in this study. While this improvement greatly minimizes positional uncertainties, some residual errors may still cause small-scale discrepancies between full-power and side beams (Tables S1 and S2, Figures S1 and S2). However, these errors are unlikely to substantially affect our overall conclusions. While not necessary for this study, the GEDI simulator described by Hancock et al. [58] could serve as a valuable tool for enhancing the calibration and validation of future LiDAR-based CHM estimations.

Accurately linking GEDI height measurements with ancillary remotely sensed variables is crucial for robust ML models in canopy height estimation [49]. SAR data, with their unique sensitivity to vegetation's dielectric and structural properties, provide insights beyond what optical data alone can offer [18,23]. However, environmental factors such as vegetation water content, canopy geometry, and soil moisture can influence SAR accuracy [11,59], introducing variability in CHM estimates and posing significant challenges. With optical observations, the spectral reflectance captured in each pixel often represents a mixture of both overstory and understory vegetation, complicating the direct correlation with canopy height [18]. This is particularly problematic in regenerating forests where high spectral reflectance from understory, combined with overstory thinning and shading, leads to an overestimation of canopy height. Such scenarios are particularly common in the Southern States, where frequent forestry operations, such as thinning and logging, along with natural regeneration processes, result in diverse canopy heights [60]. These forests, with varying stand ages, tend to be aggregated into single pixels, thereby introducing estimation errors.

Given these complex and often interacting sources of uncertainty, gaining insight into how input features affect model predictions is essential. Feature interactions play a crucial role in model performance, and SHAP analysis offers a robust framework for understanding how features interact and how their combined effects influence predictive outcomes. By quantifying both individual and pairwise feature contributions, SHAP can reveal key interactions that drive predictions, inform feature engineering, and guide model optimization. This aspect will be explored in future work to enhance our understanding of the model's behavior.

5. Conclusions

This study advances forest canopy height estimation by integrating multisource remote sensing data with explainable ML algorithms, focusing on the complex forest ecosystems of southeastern North Carolina, USA. By leveraging a diverse array of airborne and spaceborne remote sensed data, including GEDI canopy height data, multitemporal optical observations from Sentinel-2; C-band backscatter and InSAR coherence from Sentinel-1; quad-polarization L-band backscatter and polarimetric decompositions from UAVSAR; high-resolution aerial photography from NAIP; and terrain variables from an airborne LiDAR-based digital elevation model, we developed a robust data-driven approach that enhances estimation accuracy and provides deeper insights into the contributions of individual data sources.

Our use of explainable ML, specifically SHAP analysis, provided valuable insights into the importance of various predictors, addressing the “black box” issue often associated with advanced methodologies such as deep learning. Notably, we identified the summer composite of Sentinel-2’s red-edge bands and the volume scattering component derived from UAVSAR as the most significant contributors to our optimal model. Additionally, the inclusion of high-resolution NAIP data demonstrated their value in capturing fine-scale canopy variations that are not detectable by existing satellite sources. While the integration of diverse data sources improved model accuracy, the enhancements were modest compared to the models that used only Sentinel-2 optical data. This finding suggested that the benefits of incorporating additional datasets should be carefully balanced against factors such as data availability, processing complexity, and costs.

We recommend that future studies explore the necessity and impact of different data sources, particularly with upcoming satellite missions such as NISAR, which may offer alternative options and further enhance the adaptability of our methodology. By advancing towards more efficient, interpretable, and cost-effective models, the derived dataset can better support sustainable forest management and contribute to the preservation of vital forest ecosystems.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs17091536/s1>, Figure S1. Scatter plots of the forest canopy height between GEDI L2A and ALS data. A and C depict scatterplots for full power laser beams, whereas B and D show data from coverage laser beams. In (A) and (B), the ALS data were resampled using the default setting of mean to align with the GEDI dataset, while ALS data were specifically extracted from the 98th percentile height in (C) and (D); Figure S2. Boxplots illustrating the differences in canopy height between ALS-RH98 and GEDI-RH98 at the GEDI footprints, comparing results from full power laser beams with those from coverage laser beams; Table S1. Error metrics for canopy-height differences (GEDI–ALS) by beam type (Negative bias values indicate GEDI underestimation relative to ALS. RMSE = root mean squared error; MAE = mean absolute error); Table S2. Between-beam significance tests for systematic bias (GEDI–ALS canopy height) (Welch’s *t*-test and the Mann–Whitney U test evaluate equality of means and distributions, respectively; Cohen’s *d* provides the standardized effect size; All *p* values are two tailed).

Author Contributions: Conceptualization, C.W., C.S., T.A.S. and C.E.W.; methodology, C.W., C.S., T.A.S., C.E.W., T.M.P., Q.H. and F.Y.; software, C.W., C.S. and Q.H.; validation, C.W., C.S., Q.H. and F.Y.; formal analysis, C.W., C.S., T.A.S. and F.Y.; investigation, C.W., C.S., C.E.W., T.M.P. and F.Y.; resources, C.S. and T.A.S.; data curation, C.W. and C.S.; writing—original draft, C.W.; review and editing, C.W., C.S., T.A.S., C.E.W., T.M.P., Q.H. and F.Y.; visualization, C.W., C.S., T.A.S., T.M.P., Q.H. and F.Y.; supervision, C.S. and C.E.W.; project administration, C.S. and T.A.S.; funding acquisition, C.S. and T.A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by a collaborative research project between the U.S. Forest Service and the University of North Carolina at Chapel Hill with a grant number of 22-CR-11330145-046.

Data Availability Statement: The UAVSAR data utilized in this research are available from the NASA Jet Propulsion Laboratory and can be accessed at <https://uavsar.jpl.nasa.gov/> (accessed on 22 April 2025). The NAIP imagery used in this study is available from the USDA’s Geospatial Data Gateway, which can be accessed at <https://gdg.sc.egov.usda.gov/> (accessed on 22 April 2025). GEDI data are provided by NASA Goddard Space Flight Center, accessible through <https://gedi.umd.edu/data/> (accessed on 22 April 2025). The Sentinel-1 and Sentinel-2 imagery used in this study is available via the Copernicus Open Access Hub, which can be reached at <https://browser.dataspace.copernicus.eu/> (accessed on 22 April 2025). Additionally, the NAIP, GEDI, Sentinel-1, and Sentinel-2 datasets can be accessed via the Google Earth Engine at <https://earthengine.google.com/> (accessed on 22 April 2025). The forest canopy height dataset generated from this study is available at <https://doi.org/10.5281/zenodo.11176727> (accessed on 22 April 2025).

Acknowledgments: We gratefully acknowledge the members of the UNC’s Remote Sensing and Ecological Modeling Group and the Global Hydrology Lab for their constructive feedback on the figures. Portions of this manuscript were prepared by a U.S. Government employee on official time, it is therefore considered in the public domain and not subject to copyright.

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

References

1. FAO. *Global Forest Resources Assessment 2020: Main Report*; FAO: Rome, Italy, 2020.
2. De Frenne, P.; Zellweger, F.; Rodríguez-Sánchez, F.; Scheffers, B.R.; Hylander, K.; Luoto, M.; Vellend, M.; Verheyen, K.; Lenoir, J. Global buffering of temperatures under forest canopies. *Nat. Ecol. Evol.* **2019**, *3*, 744–749. [[CrossRef](#)] [[PubMed](#)]
3. Cazzolla Gatti, R.; Di Paola, A.; Bombelli, A.; Noce, S.; Valentini, R. Exploring the relationship between canopy height and terrestrial plant diversity. *Plant Ecol.* **2017**, *218*, 899–908. [[CrossRef](#)]
4. Srivathsa, A.; Vasudev, D.; Nair, T.; Chakrabarti, S.; Chanchani, P.; DeFries, R.; Deomurari, A.; Dutta, S.; Ghose, D.; Goswami, V.R.; et al. Prioritizing India’s landscapes for biodiversity, ecosystem services and human well-being. *Nat. Sustain.* **2023**, *6*, 568–577. [[CrossRef](#)]
5. Schwartz, N.B.; Budsock, A.M.; Uriarte, M. Fragmentation, forest structure, and topography modulate impacts of drought in a tropical forest landscape. *Ecology* **2019**, *100*, e02677. [[CrossRef](#)]
6. Adrah, E.; Wan Mohd Jaafar, W.S.; Omar, H.; Bajaj, S.; Leite, R.V.; Mazlan, S.M.; Silva, C.A.; Chel Gee Ooi, M.; Mohd Said, M.N.; Abdul Maulud, K.N.; et al. Analyzing Canopy Height Patterns and Environmental Landscape Drivers in Tropical Forests Using NASA’s GEDI Spaceborne LiDAR. *Remote Sens.* **2022**, *14*, 3172. [[CrossRef](#)]
7. Jucker, T.; Hardwick, S.R.; Both, S.; Elias, D.M.O.; Ewers, R.M.; Milodowski, D.T.; Swinfield, T.; Coomes, D.A. Canopy structure and topography jointly constrain the microclimate of human-modified tropical landscapes. *Glob. Change Biol.* **2018**, *24*, 5243–5258. [[CrossRef](#)]
8. Zhang, Z.; Li, X.; Liu, H. Biophysical feedback of forest canopy height on land surface temperature over contiguous United States. *Environ. Res. Lett.* **2022**, *17*, 034002. [[CrossRef](#)]
9. Torresani, M.; Rocchini, D.; Alberti, A.; Moudry, V.; Heym, M.; Thouverai, E.; Kacic, P.; Tomelleri, E. LiDAR GEDI derived tree canopy height heterogeneity reveals patterns of biodiversity in forest ecosystems. *Ecol. Inform.* **2023**, *76*, 102082. [[CrossRef](#)]
10. Coops, N.C.; Tompalski, P.; Goodbody, T.R.H.; Queinnec, M.; Luther, J.E.; Bolton, D.K.; 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](#)]
11. Sothe, C.; Gonsamo, A.; Lourenço, R.B.; Kurz, W.A.; Snider, J. Spatially Continuous Mapping of Forest Canopy Height in Canada by Combining GEDI and ICESat-2 with PALSAR and Sentinel. *Remote Sens.* **2022**, *14*, 5158. [[CrossRef](#)]
12. Hakkenberg, C.R.; Tang, H.; Burns, P.; Goetz, S.J. Canopy structure from space using GEDI lidar. *Front. Ecol. Environ.* **2023**, *21*, 55–56. [[CrossRef](#)]
13. Dubayah, R.; Blair, J.B.; Goetz, S.; Fatoyinbo, L.; Hansen, M.; Healey, S.; Hofton, M.; Hurtt, G.; Kellner, J.; Luthcke, S.; et al. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth’s forests and topography. *Sci. Remote Sens.* **2020**, *1*, 100002. [[CrossRef](#)]
14. Xi, Z.; Xu, H.; Xing, Y.; Gong, W.; Chen, G.; Yang, S. Forest canopy height mapping by synergizing ICESat-2, Sentinel-1, Sentinel-2 and topographic information based on machine learning methods. *Remote Sens.* **2022**, *14*, 364. [[CrossRef](#)]
15. 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](#)]
16. Ehlers, D.; Wang, C.; Coulston, J.; Zhang, Y.; Pavelsky, T.; Frankenberg, E.; Woodcock, C.; Song, C. Mapping Forest Aboveground Biomass Using Multisource Remotely Sensed Data. *Remote Sens.* **2022**, *14*, 1115. [[CrossRef](#)]
17. Ge, S.; Su, W.; Gu, H.; Rauste, Y.; Praks, J.; Antropov, O. Improved LSTM Model for Boreal Forest Height Mapping Using Sentinel-1 Time Series. *Remote Sens.* **2022**, *14*, 5560. [[CrossRef](#)]
18. Silveira, E.M.O.; Radeloff, V.C.; Martinuzzi, S.; Martinez Pastur, G.J.; Bono, J.; Politi, N.; Lizarraga, L.; Rivera, L.O.; Ciuffoli, L.; Rosas, Y.M.; et al. Nationwide native forest structure maps for Argentina based on forest inventory data, SAR Sentinel-1 and vegetation metrics from Sentinel-2 imagery. *Remote Sens. Environ.* **2023**, *285*, 113391. [[CrossRef](#)]
19. Potapov, P.V.; Li, X.; Hernandez-Serna, A.; Tyukavina, A.; Hansen, M.C.; Kommareddy, A.; Pickens, A.; Turubanova, S.; Tang, H.; Silva, C.E.; et al. Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sens. Environ.* **2021**, *253*, 112165. [[CrossRef](#)]
20. Torres de Almeida, C.; Gerente, J.; Rodrigo dos Prazeres Campos, J.; Caruso Gomes Junior, F.; Providelo, L.A.; Marchiori, G.; Chen, X. Canopy Height Mapping by Sentinel 1 and 2 Satellite Images, Airborne LiDAR Data, and Machine Learning. *Remote Sens.* **2022**, *14*, 4112. [[CrossRef](#)]

21. 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]
22. Lee, J.-S.; Pottier, E. *Polarimetric Radar Imaging: From Basics to Applications*; CRC Press: Boca Raton, FL, USA, 2009.
23. Pourshamsi, M.; Xia, J.; Yokoya, N.; Garcia, M.; Laval, M.; Pottier, E.; Balzter, H. Tropical forest canopy height estimation from combined polarimetric SAR and LiDAR using machine-learning. *ISPRS J. Photogramm. Remote Sens.* **2021**, *172*, 79–94. [CrossRef]
24. 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]
25. Lang, N.; Jetz, W.; Schindler, K.; Wegner, J.D. A high-resolution canopy height model of the Earth. *Nat. Ecol. Evol.* **2023**, *7*, 1778–1789. [CrossRef] [PubMed]
26. Tolan, J.; Yang, H.-I.; Nosarzewski, B.; Couairon, G.; Vo, H.V.; Brandt, J.; Couprie, C. Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar. *Remote Sens. Environ.* **2024**, *300*, 113888. [CrossRef]
27. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y.; Cho, H.; Chen, K.; Mitchell, R.; Cano, I.; Zhou, T. *Xgboost: Extreme Gradient Boosting*, R Package Version 0.4-2. 2015; 1–4.
28. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [CrossRef]
29. Zhang, L.; Zhang, X.; Shao, Z.; Jiang, W.; Gao, H. Integrating Sentinel-1 and 2 with LiDAR data to estimate aboveground biomass of subtropical forests in northeast Guangdong, China. *Int. J. Digit. Earth* **2023**, *16*, 158–182. [CrossRef]
30. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4768–4777.
31. Dubayah, R.; Beck, J.; Wirt, B.; Armston, J.; Hofton, M.; Luthcke, S.; Tang, H. GLOBAL Ecosystem Dynamics Investigation (GEDI) Level 2 User Guide. Available online: https://lpdaac.usgs.gov/documents/986/GEDI02_UserGuide_V2.pdf (accessed on 1 September 2023).
32. Bauer-Marschallinger, B.; Cao, S.; Navacchi, C.; Freeman, V.; Reuß, F.; Geudtner, D.; Rommen, B.; Vega, F.C.; Snoeij, P.; Attema, E.; et al. The normalised Sentinel-1 Global Backscatter Model, mapping Earth’s land surface with C-band microwaves. *Sci. Data* **2021**, *8*, 277. [CrossRef]
33. Kellndorfer, J.; Cartus, O.; Laval, M.; Magnard, C.; Milillo, P.; Oveisgharan, S.; Osmanoglu, B.; Rosen, P.A.; Wegmüller, U. Global seasonal Sentinel-1 interferometric coherence and backscatter data set. *Sci. Data* **2022**, *9*, 73. [CrossRef]
34. Wang, C.; Pavelsky, T.M.; Yao, F.; Yang, X.; Zhang, S.; Chapman, B.; Song, C.; Sebastian, A.; Frizzelle, B.; Frankenberg, E.; et al. Flood Extent Mapping During Hurricane Florence with Repeat-Pass L-Band UAVSAR Images. *Water Resour. Res.* **2022**, *58*, e2021WR030606. [CrossRef]
35. Shimada, M.; Itoh, T.; Motooka, T.; Watanabe, M.; Shiraishi, T.; Thapa, R.; Lucas, R. New global forest/non-forest maps from ALOS PALSAR data (2007–2010). *Remote Sens. Environ.* **2014**, *155*, 13–31. [CrossRef]
36. Partners, O. North Carolina Statewide Lidar DEM 2014 Phase 2 from 2010-06-15 to 2010-08-15. NOAA National Centers for Environmental Information. Available online: <https://www.fisheries.noaa.gov/inport/item/49412> (accessed on 5 January 2025).
37. Partners, O. North Carolina Statewide Lidar DEM 2015 Phase 3 from 2010-06-15 to 2010-08-15. NOAA National Centers for Environmental Information. Available online: <https://www.fisheries.noaa.gov/inport/item/49413> (accessed on 5 January 2025).
38. Rosen, P.A.; Hensley, S.; Wheeler, K.; Sadowy, G.; Miller, T.; Shaffer, S.; Zebker, H. UAVSAR: New NASA Airborne SAR System for Research. *IEEE Aerosp. Electron. Syst. Mag.* **2007**, *22*, 21–28. [CrossRef]
39. Ulaby, F.T.; Moore, R.K.; Fung, A.K. *Microwave Remote Sensing: Active and Passive. Volume 3-From Theory to Applications*; Artech House: Norwood, MA, USA, 1986.
40. Cloude, S.R.; Pottier, E. An entropy based classification scheme for land applications of polarimetric SAR. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 68–78. [CrossRef]
41. Freeman, A.; Durden, S.L. A three-component scattering model for polarimetric SAR data. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 963–973. [CrossRef]
42. Song, C.; Woodcock, C.E. Estimating tree crown size from multiresolution remotely sensed imagery. *Photogramm. Eng. Remote Sens.* **2003**, *69*, 1263–1270. [CrossRef]
43. Song, N.; Zhang, F. The changing process and mechanism of the farming-grazing transitional land use pattern in Ordos. *Acta Geogr. Sin.* **2007**, *62*, 1299–1308.
44. Pander, J.; Mueller, M.; Geist, J. Habitat diversity and connectivity govern the conservation value of restored aquatic floodplain habitats. *Biol. Conserv.* **2018**, *217*, 1–10. [CrossRef]
45. Klein, T.; Randin, C.; Körner, C. Water availability predicts forest canopy height at the global scale. *Ecol. Lett.* **2015**, *18*, 1311–1320. [CrossRef]

46. Zhu, X.; Nie, S.; Zhu, Y.; Chen, Y.; Yang, B.; Li, W. Evaluation and Comparison of ICESat-2 and GEDI Data for Terrain and Canopy Height Retrievals in Short-Stature Vegetation. *Remote Sens.* **2023**, *15*, 4969. [\[CrossRef\]](#)
47. Fayad, I.; Baghdadi, N.; Lahssini, K. An Assessment of the GEDI Lasers' Capabilities in Detecting Canopy Tops and Their Penetration in a Densely Vegetated, Tropical Area. *Remote Sens.* **2022**, *14*, 2969. [\[CrossRef\]](#)
48. Rishmawi, K.; Huang, C.; Schleeweis, K.; Zhan, X. Integration of VIIRS Observations with GEDI-Lidar Measurements to Monitor Forest Structure Dynamics from 2013 to 2020 across the Conterminous United States. *Remote Sens.* **2022**, *14*, 2320. [\[CrossRef\]](#)
49. Wang, C.; Elmore, A.J.; Numata, I.; Cochrane, M.A.; Lei, S.; Hakkenberg, C.R.; Li, Y.; Zhao, Y.; Tian, Y. A Framework for Improving Wall-to-Wall Canopy Height Mapping by Integrating GEDI LiDAR. *Remote Sens.* **2022**, *14*, 3618. [\[CrossRef\]](#)
50. Perez, G.G.; Bourscheidt, V.; Lopes, L.E.; Takata, J.T.; Ferreira, P.A.; Boscolo, D. Use of Sentinel 2 imagery to estimate vegetation height in fragments of Atlantic Forest. *Ecol. Inform.* **2022**, *69*, 101680. [\[CrossRef\]](#)
51. Arévalo, P.; Baccini, A.; Woodcock, C.E.; Olofsson, P.; Walker, W.S. Continuous mapping of aboveground biomass using Landsat time series. *Remote Sens. Environ.* **2023**, *288*, 113483. [\[CrossRef\]](#)
52. Fatoyinbo, T.; Armston, J.; Simard, M.; Saatchi, S.; Denbina, M.; Laval, M.; Hofton, M.; Tang, H.; Marselis, S.; Pinto, N.; et al. The NASA AfriSAR campaign: Airborne SAR and lidar measurements of tropical forest structure and biomass in support of current and future space missions. *Remote Sens. Environ.* **2021**, *264*, 112533. [\[CrossRef\]](#)
53. Wang, C.; Pavelsky, T.M.; Kyzivat, E.D.; Garcia-Tigreros, F.; Podest, E.; Yao, F.; Yang, X.; Zhang, S.; Song, C.; Langhorst, T.; et al. Quantification of wetland vegetation communities features with airborne AVIRIS-NG, UAVSAR, and UAV LiDAR data in Peace-Athabasca Delta. *Remote Sens. Environ.* **2023**, *294*, 113646. [\[CrossRef\]](#)
54. Zhang, Z.; Ni, W.; Sun, G.; Huang, W.; Ranson, K.J.; Cook, B.D.; Guo, Z. Biomass retrieval from L-band polarimetric UAVSAR backscatter and PRISM stereo imagery. *Remote Sens. Environ.* **2017**, *194*, 331–346. [\[CrossRef\]](#)
55. Han, Q.; Zeng, Y.; Zhang, L.; Cira, C.-I.; Prikaziuk, E.; Duan, T.; Wang, C.; Szabó, B.; Manfreda, S.; Zhuang, R.; et al. Ensemble of optimised machine learning algorithms for predicting surface soil moisture content at a global scale. *Geosci. Model Dev.* **2023**, *16*, 5825–5845. [\[CrossRef\]](#)
56. Adam, M.; Urbazaev, M.; Dubois, C.; Schmullius, C. Accuracy Assessment of GEDI Terrain Elevation and Canopy Height Estimates in European Temperate Forests: Influence of Environmental and Acquisition Parameters. *Remote Sens.* **2020**, *12*, 3948. [\[CrossRef\]](#)
57. Tang, H.; Stoker, J.; Luthcke, S.; Armston, J.; Lee, K.; Blair, B.; Hofton, M. Evaluating and mitigating the impact of systematic geolocation error on canopy height measurement performance of GEDI. *Remote Sens. Environ.* **2023**, *291*, 113571. [\[CrossRef\]](#)
58. Hancock, S.; Armston, J.; Hofton, M.; Sun, X.; Tang, H.; Duncanson, L.I.; Kellner, J.R.; Dubayah, R. The GEDI Simulator: A Large-Footprint Waveform Lidar Simulator for Calibration and Validation of Spaceborne Missions. *Earth Space Sci.* **2019**, *6*, 294–310. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Choi, C.; Cazcarra-Bes, V.; Guliaev, R.; Pardini, M.; Papathanassiou, K.P.; Qi, W.; Armston, J.; Dubayah, R.O. Large Scale Forest Height Mapping by Combining TanDEM-X and GEDI Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 2374–2385. [\[CrossRef\]](#)
60. Schroeder, T.A.; Obata, S.; Papeş, M.; Branoff, B. Evaluating Statewide NAIP Photogrammetric Point Clouds for Operational Improvement of National Forest Inventory Estimates in Mixed Hardwood Forests of the Southeastern U.S. *Remote Sens.* **2022**, *14*, 4386. [\[CrossRef\]](#)

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.