Evaluating ICESat-2 and GEDI with Integrated Landsat-8 and PALSAR-2 for Mapping Tropical Forest Canopy Height

Aobo Liu [1], Yating Chen [1,*] and Xiao Cheng [2,3]

1 College of Geography and Environment, Shandong Normal University, Jinan 250014, China; lab2016bnu@foxmail.com
2 School of Geospatial Engineering and Science, Sun Yat-sen University, and Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China; chengxiao9@mail.sysu.edu.cn
3 Key Laboratory of Comprehensive Observation of Polar Environment, Sun Yat-sen University, Ministry of Education, Zhuhai 519082, China
* Correspondence: chenyt2016bnu@gmail.com

Abstract: Mapping forest canopy height is critical for climate modeling and forest management, and tropical forests present unique challenges for remote sensing due to their dense vegetation and complex structure. The advent of ICESat-2 and GEDI, two advanced lidar datasets, offers new opportunities for improving canopy height estimation. In this study, we used footprint-level canopy height products from ICESat-2 and GEDI, combined with features extracted from Landsat-8, PALSAR-2, and FABDEM products. The AutoGluon stacking ensemble learning algorithm was employed to construct inversion models, generating 30 m resolution continuous canopy height maps for the tropical forests of Puerto Rico. Accuracy validation was performed using the high-resolution G-LiHT airborne lidar products. Results show that tropical forest canopy height inversion remains challenging, with all models yielding relative root mean square errors (rRMSE) exceeding 0.30. The stacking ensemble model outperformed all base learners, and the GEDI-based map had slightly higher accuracy than the ICESat-2-based map, with RMSE values of 4.81 and 4.99 m, respectively. Both models showed systematic biases, but the GEDI-based model exhibited less underestimation for taller canopies, making it more suitable for biomass estimation. The proposed approach can be applied to other forest ecosystems, enabling fine-resolution canopy height mapping and enhancing forest conservation efforts.

Keywords: canopy height; tropical forest; ICESat-2; GEDI; stacking ensemble learning

1. Introduction

Forests play a crucial role in global ecosystems, providing essential services such as carbon sequestration, soil conservation, and water regulation [1–3]. Forests and other terrestrial ecosystems absorb approximately 3.5 ± 0.9 billion metric tons of carbon annually, accounting for about 30% of global carbon emissions from fossil fuels [3]. As climate change intensifies, forests' role as carbon sinks has become increasingly important in the global carbon cycle [4]. Large-scale, continuous forest monitoring is essential to enhance our understanding of the global carbon cycle and identify critical carbon sinks [5]. This has been recognized as a key element in international agreements such as the Paris Agreement [6] and the 2030 Agenda for Sustainable Development [7], which emphasize the need for dynamic monitoring and quantitative assessment of forest structure parameters.

Canopy height, defined as the distance between the canopy top and the ground surface, is a key indicator for estimating forest biomass, carbon storage, and productivity [8]. Accurate canopy height mapping is critical for estimating forest biomass and carbon sequestration potential, which are essential for understanding climate change impacts [9]. Remote sensing techniques, including optical, microwave, and lidar-based approaches, have been widely applied to map forest canopy height [10–18]. However, optical and

shorter-wavelength microwave remote sensing methods often face limitations in dense forests, where signal saturation makes it difficult to accurately estimate canopy height and biomass [19,20].

Lidar (Light Detection and Ranging) technology has emerged as a unique tool for canopy height and biomass mapping, as it can penetrate dense canopies and directly capture accurate forest structure information [21,22]. Unlike optical and SAR sensors, lidar's nature allows for precise vertical forest structure measurements, overcoming signal saturation issues in dense forest environments [23]. While ground-based and airborne lidar systems provide high-resolution forest monitoring data, their coverage is limited, particularly in remote or underdeveloped regions. In 2018, NASA launched two spaceborne lidar missions—ICESat-2 (The Ice, Cloud, and Land Elevation Satellite-2) [24] and GEDI (Global Ecosystem Dynamics Investigation) [25]—which offer global coverage and enable large-scale, high-resolution forest structure monitoring, addressing the coverage limitations of ground-based and airborne lidar systems.

The GEDI mission, mounted on the International Space Station, provides over 10 billion waveform measurements of forest vertical structure at a 25 m footprint resolution [25]. GEDI's optimized lidar system offers higher spatial density and vertical sensitivity, making it well-suited for capturing forest canopy height and aboveground biomass [26]. In contrast, ICESat-2, primarily designed for monitoring polar ice sheets, employs photon-counting lidar technology characterized by high repetition frequency, high peak power, narrow pulse width, and high sensitivity [27]. However, its sensor settings are less optimized for forest monitoring compared to GEDI [28]. Despite their different design purposes, ICESat-2 and GEDI provide complementary datasets for canopy height mapping [29]. GEDI focuses on tropical and temperate regions between 51.6°N and 51.6°S, while ICESat-2 extends coverage up to 88°N and 88°S, offering critical data for high-latitude forests beyond GEDI's reach [30]. By integrating these lidar datasets with optical and SAR imagery, it is possible to produce high-resolution, spatially continuous canopy height maps on a global scale [31,32].

Several studies have explored the use of ICESat-2 and GEDI data for global and national forest canopy height mapping [30–33]. Notably, Potapov et al. [31] combined GEDI-derived canopy height data with multi-temporal Landsat-8 indices to produce a 30 m resolution global canopy height map. Similarly, Lang et al. [32] employed a probabilistic deep learning model, utilizing GEDI and Sentinel-2 data, to generate a global canopy height map at a finer 10 m resolution. Liu et al. [33] integrated GEDI and ICESat-2 canopy height estimates with Sentinel-2 spectral features, producing a 30 m resolution canopy height map specifically for China. Sothe et al. [30] demonstrated the value of combining GEDI and ICESat-2 data with PALSAR and Sentinel imagery for spatially continuous mapping of forest canopy height in Canada. These global and national-scale studies highlight the increasing potential of spaceborne lidar missions to support high-resolution forest monitoring.

To bridge the insights from previous studies to our specific focus, we target tropical forests, where the intricate vertical layering and dense canopy make accurate remote sensing inversion challenging [34–36]. First, we provide a comparative assessment of ICESat-2 and GEDI data for canopy height inversion in tropical forests, an area that has not been extensively explored. Second, by integrating these lidar datasets with multispectral and SAR imagery, we create a more comprehensive feature set to enhance the accuracy of canopy height predictions. This integration helps mitigate issues such as signal saturation and the complexity of vertical forest structures [15]. Lastly, we employ advanced machine learning techniques, specifically leveraging stacking ensemble models within the Auto-Gluon framework [37], to optimize predictions and provide a robust assessment of ICESat-2 and GEDI for tropical forest monitoring.

In this study, we aim to generate high-precision, spatially continuous canopy height maps at a 30 m resolution for tropical forests in Puerto Rico. Our approach leverages the strengths of ICESat-2 and GEDI data while introducing innovations in feature extraction and machine learning techniques to improve canopy height predictions. The findings from

this study have important implications for improving canopy height inversion models and advancing large-scale forest monitoring efforts, particularly in tropical regions with complex forest structures.

2. Materials and Methods

2.1. Study Area

The study area, Puerto Rico (Figure 1), is an island territory of the United States located in the Caribbean Sea, spanning from 65.6°W to 67.3°W longitude and 17.9°N to 18.5°N latitude. The main island of Puerto Rico stretches approximately 180 km from east to west and 65 km from north to south, covering an area of about 9.8×10^3 km^2. It is the third-largest island under U.S. jurisdiction, with a population of around 3.2 million. The island's topography is predominantly mountainous, with 60% of its land area characterized by rugged terrain and complex understory vegetation. Expansive coastal plains are found in both the northern and southern parts of the island.
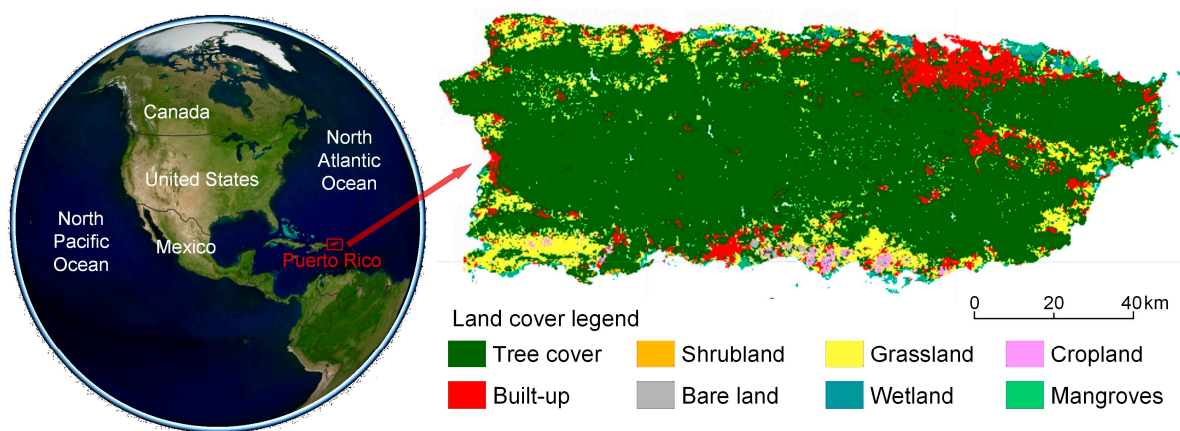

Figure 1. Location and land cover types of the study area.

The central mountain range, known as the Cordillera Central, runs east–west across the island, with the highest peak, Cerro de Punta, reaching an elevation of 1338 m. Puerto Rico's tropical forest ecological zone experiences a tropical maritime monsoon climate, with an average annual temperature of 26 °C and minimal seasonal variation. Year-round temperatures generally range between 20 °C and 30 °C, and the island receives an average annual precipitation of about 1400 mm [38]. The primary land cover types in Puerto Rico include forests, shrublands, grasslands, built-up areas, water bodies, and wetlands [39]. Common tree species found in the island's tropical forests include Prestoea acuminata (palm), Manilkara bidentata (balata), Cocos nucifera (coconut), and Piones mangroves [40]. The study area was selected because of its diverse and structurally complex forests, which present both challenges and representativeness, and because publicly available airborne lidar data are available for independent accuracy validation.

2.2. Data

Figure 2 presents a flowchart outlining the data and processing methods employed in this study. This study utilized a combination of remote sensing datasets to construct canopy height inversion models for tropical forests. The primary datasets included ICESat-2 and GEDI lidar products, which were used to obtain footprint-level forest canopy height measurements. High-resolution airborne lidar data from the G-LiHT (Goddard's Lidar, Hyperspectral and Thermal Imager) [41] were employed for independent accuracy validation. Additionally, Landsat-8 multispectral imagery was used to extract spectral and textural features, while PALSAR-2 SAR imagery was used to derive polarization features. Topographic features were obtained from the FABDEM (Forest And Buildings removed Copernicus Digital Elevation Model) terrain product [42]. To ensure consistency and comparability in the analysis, all datasets were carefully processed for spatial resolution and temporal

alignment. This preprocessing step was essential to guarantee that the data sources were appropriately matched in terms of spatial and temporal scales, allowing for accurate canopy height mapping across the study area.
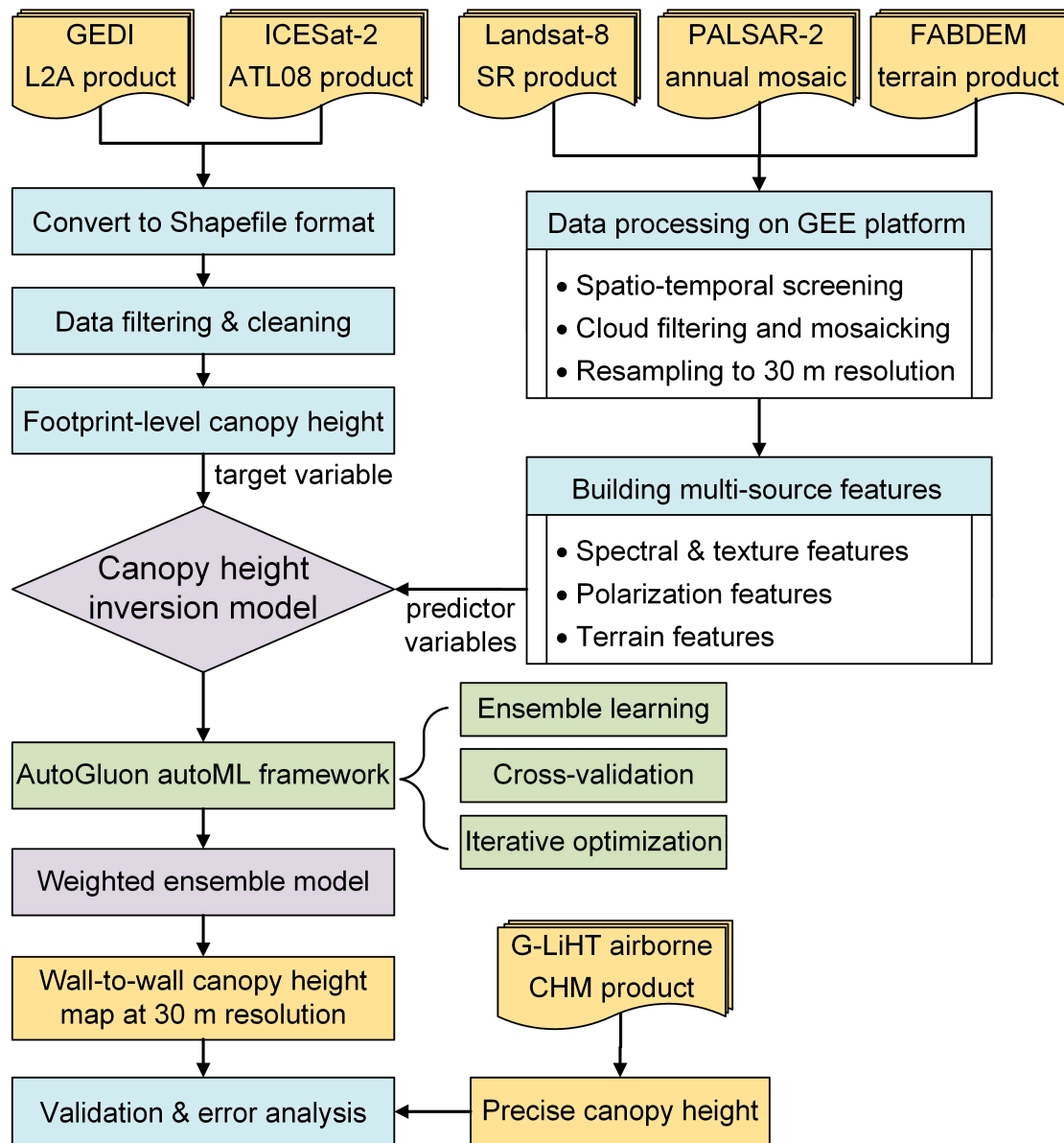
Figure 2. Flowchart illustrating the process of generating spatially continuous canopy height maps using publicly accessible data products and the Google Earth Engine cloud platform.

2.2.1. GEDI L2A Product

GEDI provides four levels of scientific data products, with the L2A product offering footprint-level height metrics (footprint diameter of 25 m), including ground elevation and relative height (RH) percentiles. GEDI uses full-waveform lidar technology, capturing vertical structure information by fitting Gaussian distributions to waveform peaks and calculating normalized cumulative waveform energy [25]. This process allows for the identification of ground, canopy top, and relative heights within each footprint (Figure 3). In this study, we used the latest version of the GEDI L2A product (Version 2; https://lpdaac.usgs.gov/products/gedi02_av002/; accessed on 10 October 2024), released by the U.S. Geological Survey (USGS), and processed the data using USGS-provided data prep scripts, which automate spatial queries, batch downloads, spatial clipping, and parameter filtering.
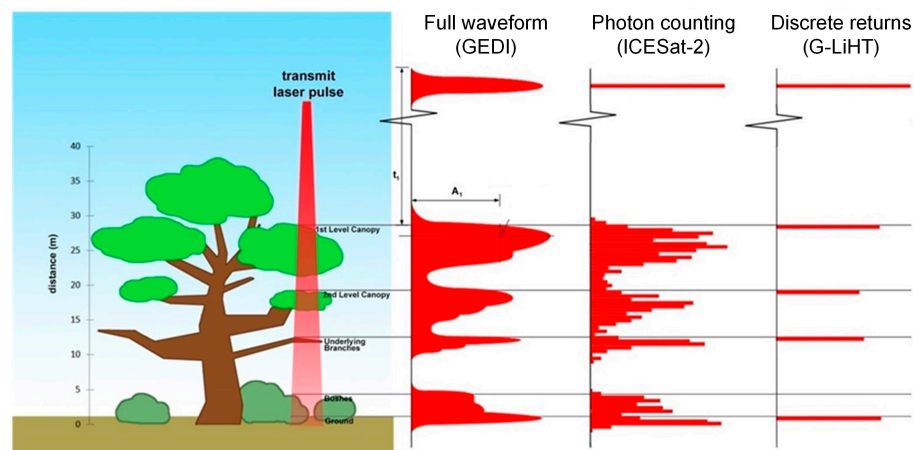
Figure 3. Comparison of three lidar detection technologies. Adapted from [29].

The parameters of the GEDI L2A data product used are listed in Table 1. Canopy height percentiles (RH 1-100) are provided at 1% intervals. To mitigate the impact of noise, we used the RH98 parameter to represent canopy top height, maintaining consistency with the ICESat-2 ATL08 product. To ensure accurate footprint-level canopy height data, we applied established filtering criteria [15,29]. First, invalid waveform data were removed based on the quality_flag parameter (data with quality_flag = 1 are retained). Next, footprints acquired during daytime and those with weaker coverage beams were filtered using the delta_time and beam_flag parameters (BEAM0101, BEAM0110, BEAM1000, and BEAM1011 beams acquired at night are retained). We utilized 68 GEDI L2A files collected between May and September from 2019 to 2021, yielding approximately 170,000 valid footprint-level canopy height estimates after filtering and cleaning.

Table 1. Parameters of the GEDI L2A product used in this study.

Variable Name	Description
lon_lowestmode	Longitude of the footprint center
lat_lowestmode	Latitude of the footprint center
elev_lowestmode	Ground elevation value derived from waveform inversion
RH 1-100	Canopy height percentiles derived from waveform inversion
quality_flag	Flag used for quality assessment
delta_time	Used to determine whether the data were acquired during day/night
beam_flag	Used to identify whether the beam is a coverage or power beam
Sensitivity	Signal-to-noise ratio measure related to canopy cover

2.2.2. ICESat-2 ATL08 Product

ICESat-2 provides 21 standardized data products across four levels. Among the Level 3 products, the ATL08 land and vegetation height product contains terrain and canopy height measurements along the satellite's track [43]. The ICESat-2 ATLAS instrument uses photon counting technology, featuring micropulse, multi-beam, and high-frequency capabilities, firing laser pulses at a rate of 10 kHz to acquire photon cloud data. Each ICESat-2 laser footprint measures approximately 12 m in diameter [28]. In regions with vegetation cover, each ATLAS laser pulse typically returns an average of 0 to 4 photons [44]. As a result, the ATL08 product is segmented along the track at intervals of 20 or 100 m to ensure a sufficient number of photons are available for reliable height estimates [43]. The ATL08 product records a large number of photon returns, forming a distribution histogram similar to a full-waveform return by recording individual photon energy levels (Figure 3).

In this study, we used the latest version of the ICESat-2 ATL08 product (Version 6; https://nsidc.org/data/atl08/versions/6/; accessed on 10 October 2024) provided by the National Snow and Ice Data Center (NSIDC), and processed the data using the NSIDC data

access and service API, which supports customized spatiotemporal queries and variable filtering. The ATL08 product contains hundreds of parameters, and the key parameters used in this study are listed in Table 2. Due to the sensitivity of the photon counting sensor to solar noise, we used the 98th percentile of canopy height (RH98) to represent the canopy top height [23]. The ATL08 product provides terrain and canopy height estimates for 12 by 20 m geosegments, each containing at least 50 photon returns (n_seg_ph $\geq$ 50). We applied several filtering criteria to retain accurate canopy height estimates from ICESat-2 data. We selected strong beam data acquired at night using the night_flag and ground_track_flag (referred to as /gtx) parameters, excluded data with a signal-to-noise ratio (SNR) below 1, and removed potentially erroneous canopy height estimates caused by cloud cover (layer_flag) or snow cover (segment_snowcover). To ensure phenological consistency, we retrieved ATL08 data acquired between May and September from 2019 to 2021, resulting in 53 files containing approximately 60,000 valid footprint-level canopy height estimates.

Table 2. Parameters of the ICESat-2 ATL08 product used in this study.

Parameter	Description
longitude	Longitude of the segment center
latitude	Latitude of the segment center
h_te_mean	Mean elevation of ground photons within the segment
h_canopy	98th percentile canopy height within the segment
n_seg_ph	Number of ground and canopy photons detected in the segment
SNR	Ratio of signal photons to noise photons
ground_track_flag	Identifies beam strength based on orbital information
night_flag	Indicates whether data were acquired during day or night
layer_flag	Indicates whether data are affected by cloud cover
segment_snowcover	Indicates whether data are affected by snow cover

2.2.3. G-LiHT Airborne Lidar Product

In this study, the G-LiHT airborne lidar (https://glihtdata.gsfc.nasa.gov/; accessed on 10 October 2024) canopy height model (CHM) was used for independent accuracy validation. G-LiHT is an airborne system that integrates scanning lidar and high-resolution cameras to map the composition, structure, and function of terrestrial ecosystems [41]. G-LiHT uses the Riegl VQ 480i dual-scanning lidar with real-time waveform processing technology. It operates at a laser wavelength of 1550 nm (near-infrared) and a sampling frequency of 50 to 300 kHz. The footprint diameter is 0.3 m, with a sampling density of 12 points/m^2. The system's calibrated horizontal and vertical accuracies are both 5 cm, and the CHM product has a spatial resolution of 1 m [44].

The G-LiHT airborne lidar system provides discrete return data with a small footprint, with each laser pulse recording 1 to 5 return signals (Figure 3). Compared to satellite-based lidar systems, airborne lidar systems like G-LiHT offer higher sampling densities and smaller footprint sizes, resulting in more accurate terrain and canopy height measurements [45]. The G-LiHT lidar data for Puerto Rico were collected in 2017, and the spatial coverage is shown in Figure 4.
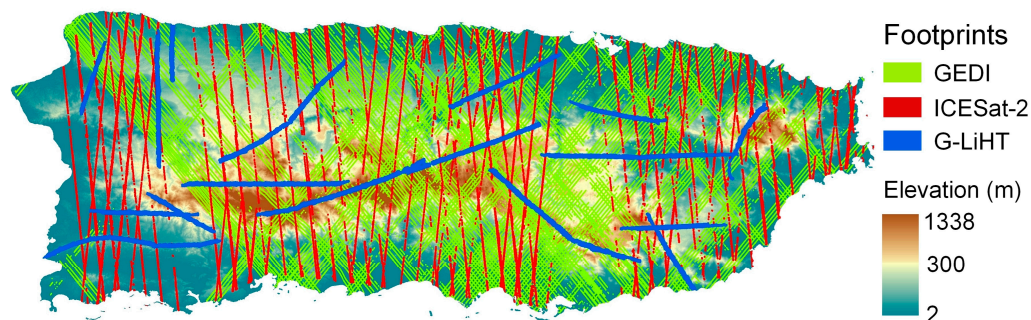

Figure 4. Spatial coverage of GEDI, ICESat-2, and G-LiHT data.

2.2.4. Landsat-8 Multispectral Imagery

Since GEDI and ICESat-2 provide spatially discrete and sparse samples, combining these lidar datasets with spatially continuous optical or SAR satellite imagery is essential for generating wall-to-wall canopy height maps (Figure 2). In this study, we chose 30 m resolution Landsat-8 imagery instead of 10 m Sentinel-2 imagery because the footprint size of GEDI and ICESat-2 better matches Landsat's pixel resolution. The Landsat program, a joint Earth observation initiative by the USGS and NASA, has provided continuous global observations since 1972, offering one of the most widely used sources of medium-to-high spatial resolution multispectral satellite data. Landsat-8 is equipped with two sensors—the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS)—capturing data across 11 spectral bands, including visible, near-infrared, shortwave infrared, and thermal infrared [46]. Here, we used the surface reflectance (SR) data product, available on the Google Earth Engine (GEE) platform [47,48], which has undergone geometric correction, radiometric calibration, and atmospheric correction. The Landsat-8 imagery was temporally aligned with the GEDI and ICESat-2 data, covering the period from May to September between 2019 and 2021. A median composite method was applied to generate a synthetic image for the analysis.

2.2.5. PALSAR-2 SAR Imagery

SAR satellites offer the advantage of all-weather, day-and-night imaging, with the added ability to penetrate vegetation canopies to some extent, making them widely used for forest height and biomass estimation [49]. On May 2014, the Japan Aerospace Exploration Agency launched the Advanced Land Observing Satellite-2 (ALOS-2), equipped with the L-band PALSAR-2 sensor [50]. PALSAR-2 operates at a wavelength of approximately 23.6 cm and a frequency of 1.2 GHz. Compared to Sentinel-1's C-band, PALSAR-2's L-band offers greater penetration, making it better suited for complex canopy structures in tropical forest ecosystems [35]. In this study, we used the PALSAR-2 annual mosaic product provided by the GEE platform, with a spatial resolution of 25 m. This product is created by mosaicking PALSAR-2 strip images, which undergo orthorectification and slope correction. Additionally, destriping techniques are applied to balance intensity differences between adjacent strips caused by variations in surface moisture conditions [51]. We utilized the HH and HV polarization backscatter coefficients from the PALSAR-2 annual mosaic product and converted the original 16-bit digital number (DN) values into decibels (dB) to improve data distribution (refer to Equation (1)), which facilitates visualization and data analysis [52].

$$\gamma_0 = 10 \log_{10}\left(DN^2\right) - 83.0 \text{ dB} \tag{1}$$

2.2.6. FABDEM Terrain Product

Here, we utilized the FABDEM terrain product [42] as auxiliary data to provide terrain features, including elevation and slope, which are instrumental in enhancing the accuracy of forest canopy height estimations. FABDEM is derived from the Copernicus GLO-30 DEM product by eliminating height biases caused by buildings and trees [53]. This dataset is available globally with a spatial resolution of 1 arc second (~30 m). By removing forest and building cover, the FABDEM provides more accurate terrain information beneath the forest canopy, thereby supporting more precise inversion of forest canopy height [53].

2.3. Method

2.3.1. Feature Extraction

After preprocessing the Landsat-8, PALSAR-2, and FABDEM datasets, we extracted spectral, textural, polarization, and terrain features to construct the canopy height inversion model (Figure 2). The spectral features included six bands from Landsat-8 (B2-B7), as well as the Normalized Difference Vegetation Index (NDVI), Tasseled Cap Greenness (TCG; [54]), and the Kernel NDVI (kNDVI; [55]), calculated using Formulae (2) to (4).

$$NDVI = (NIR - R)/(NIR + R) \tag{2}$$

$$TCG = -0.16 \cdot B - 0.28 \cdot G - 0.49 \cdot R + 0.79 \cdot NIR - 0.0002 \cdot SWIR1 - 0.14 \cdot SWIR2 \tag{3}$$

$$kNDVI = \tanh(((NIR - R)/2\sigma)^2) = \tanh(NDVI^2) \tag{4}$$

In these formulae, B, G, R, NIR, SWIR1, and SWIR2 represent the blue, green, red, near-infrared, shortwave infrared 1, and shortwave infrared 2 bands, respectively. kNDVI utilizes kernel methods from machine learning to linearize NDVI, mitigating the non-linearity and saturation effects commonly seen in traditional vegetation indices, which enhances its performance in densely vegetated areas [56]. Additionally, texture features were calculated using a gray-level co-occurrence matrix [57] with a 3×3 kernel based on kNDVI, including metrics such as entropy, energy, contrast, and correlation. Polarization features were extracted from the HH and HV backscatter coefficients of PALSAR-2. Finally, terrain features such as elevation, slope, aspect, and hillshade were derived from the FABDEM product.

2.3.2. Model Construction

This study utilized the AutoGluon automated machine learning framework [37] to construct canopy height inversion models, enabling the extrapolation of canopy heights across the entire study area from GEDI and ICESat-2 footprint-level lidar data. Machine learning models are widely used for canopy height inversion, aiming to establish relationships between forest canopy height and spatially continuous satellite imagery features. In this study, footprint-level canopy height estimates from the GEDI L2A and ICESat-2 ATL08 products were used as target variables, while spectral, textural, polarization, and terrain features extracted from Landsat-8, PALSAR-2, and FABDEM served as predictor variables.

AutoGluon supports the automation of key tasks such as model construction, hyperparameter optimization, and model evaluation, demonstrating excellent performance in classification and regression tasks across various data types, including images, tables, and text [58]. It incorporates a range of algorithms, such as k-nearest neighbors [59], random forests [60], neural networks [61], extremely randomized trees [62], and three gradient-boosting decision tree algorithms (XGBoost [63], LightGBM [64], and CatBoost [65]). These models are commonly used machine learning methods for regression tasks and provide a strong baseline. Additionally, one of AutoGluon's key features is the stacking ensemble model, which combines predictions from multiple base learners to create a meta-model that improves overall accuracy. In a two-layer stacking framework (Figure 5), the first layer consists of base models like random forests and neural networks, while the second layer uses the predictions from these base models as inputs to generate the final prediction. This stacking approach enhances predictive performance compared to individual models by combining the strengths of multiple algorithms. Each model captures different patterns and relationships within the data, and by aggregating their predictions, the stacking model reduces individual model biases and variance. This leads to more robust and accurate predictions, as the ensemble can balance the weaknesses of each single model while leveraging their strengths [58,66].

In this study, seven base models and a stacking ensemble model were constructed. All model parameters were initialized with default values. For the k-nearest neighbors algorithm, weights were set based on inverse distance. For the random forest and extremely randomized tree algorithms, the number of trees was set to 300, with the number of randomly selected features being the square root of the total number of features. Feature selection was based on entropy or Gini index. For the neural network, the number of epochs was set to 500, the activation function was ReLU, the batch size was set to 512, and the base and target learning rates were set to 0.0003 and 1.0, respectively. For XGBoost, LightGBM, and CatBoost, the number of iterations was set to 10,000. The boosting type for LightGBM was set to traditional gradient boosting, and the learning rate for CatBoost was set to 0.1.
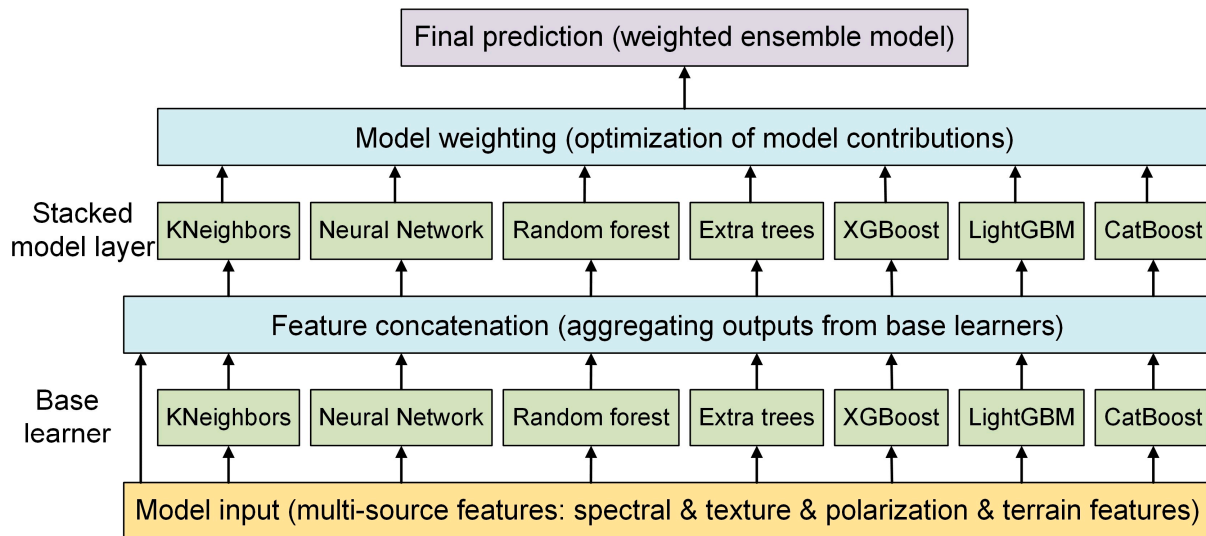
Figure 5. Diagram of the AutoGluon stacking ensemble learning model.

During model training, parameters such as the training strategy, K-fold cross-bagging strategy, training time limit, and early stopping criteria were set to automatically split the input data into training and validation sets and optimize hyperparameters within a predefined search space. Specifically, the training strategy was set to "best_quality"; the num_bag_folds parameter was set to five, meaning five-fold cross-validation was performed, with 80% of the data used for training and 20% for independent validation; the training time limit (time_limit) was set to 3600 s; and the auto_stack parameter was set to True to enable multi-layer stacking, with the number of stacking layers adjusted according to model training speed.

After model training, feature importance evaluation [67,68] was applied to iteratively refine the input features. To assess multicollinearity and reduce redundancy among predictor variables, we used the Variance Inflation Factor (VIF) to check for collinearity [69]. In each iteration, features with low importance or high multicollinearity were removed. The iterative optimization process continued, utilizing cross-validation to fine-tune model performance until the most accurate and stable model was achieved.

2.3.3. Accuracy Validation

We validated the accuracy of the generated tropical forest canopy height maps for Puerto Rico using the G-LiHT airborne lidar product. The G-LiHT product has a spatial resolution of 1 m, so it was resampled to 30 m to match the spatial resolution of the canopy height maps being evaluated. After resampling, a total of 38,235 validation points were obtained across the study area. Five accuracy metrics were used to assess the product's accuracy: (1) mean bias, (2) mean absolute error (MAE), (3) correlation coefficient (R), (4) root mean square error (RMSE), and (5) relative RMSE (rRMSE).

3. Results

3.1. Model Performance Comparison

To evaluate the potential of the ICESat-2 ATL08 and GEDI L2A products for mapping tropical forest canopy height, we integrated their footprint-level canopy height estimates with spectral, textural, polarization, and terrain features derived from Landsat-8, PALSAR-2, and FABDEM data. Canopy height inversion models were constructed, and mapping was performed at a spatial resolution of 30 m to match the resolution of each data product. Figures 6 and 7 show the accuracy evaluation results for seven base models and the stacking ensemble model, based on the ICESat-2 ATL08 and GEDI L2A products, respectively. It is important to note that these seven base models are also the Level 2 models within AutoGluon, constructed through feature concatenation (Figure 5) to enhance predictive

performance. Each subplot illustrates the relationship between predicted and observed canopy heights for a specific model, along with key accuracy metrics.
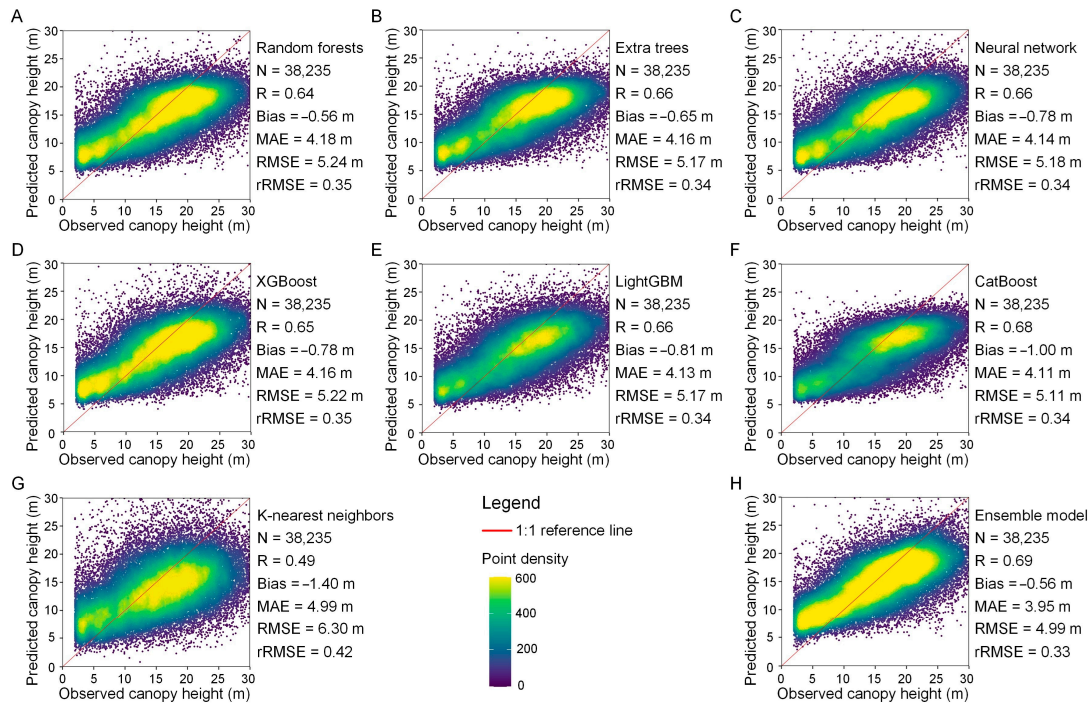

Figure 6. Accuracy comparison of seven base models and the stacking ensemble model for canopy height inversion using ICESat-2 ATL08 data. (**A**) random forests, (**B**) extremely randomized trees, (**C**) neural networks, (**D**) XGBoost, (**E**) LightGBM, (**F**) CatBoost, (**G**) k-nearest neighbors, and (**H**) the stacking ensemble model.
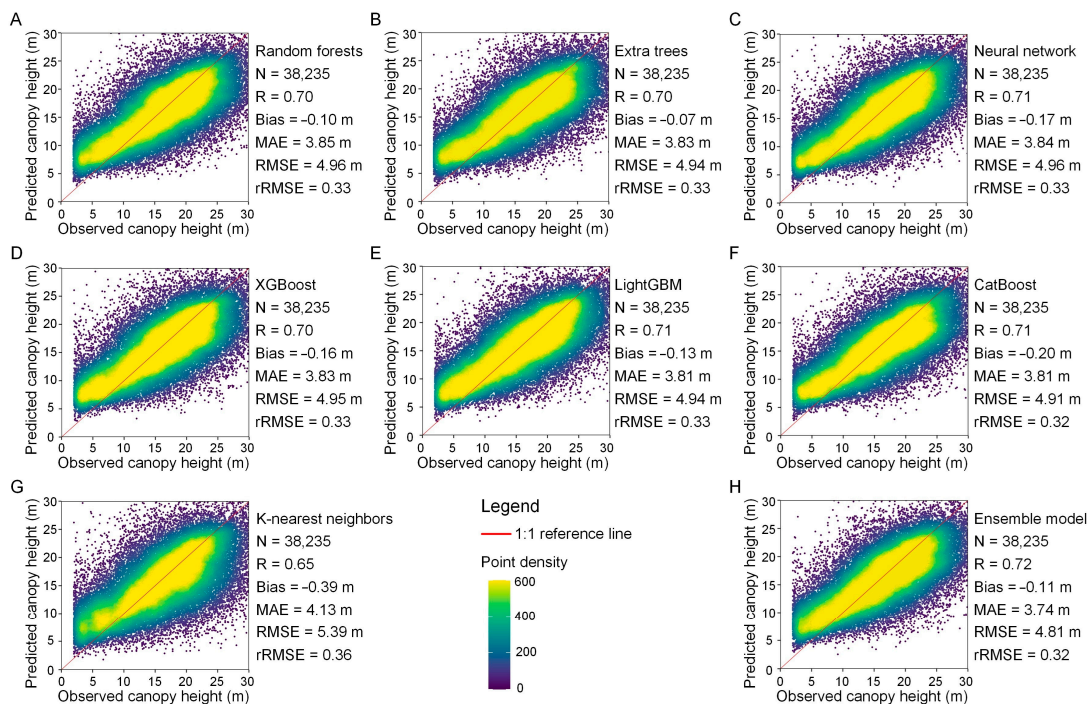

Figure 7. Accuracy comparison of seven base models and the stacking ensemble model for canopy height inversion using GEDI L2A data. (**A**) random forests, (**B**) extremely randomized trees, (**C**) neural networks, (**D**) XGBoost, (**E**) LightGBM, (**F**) CatBoost, (**G**) k-nearest neighbors, and (**H**) the stacking ensemble model.

The ensemble model consistently reduces bias and error compared to the base models for both datasets. For ICESat-2, the ensemble reduced the RMSE to 4.99 m, improving on the 5.11 m achieved by the best-performing base model. Similarly, for GEDI, the ensemble achieved an RMSE of 4.81 m, outperforming the best base model by 0.1 m. The bias for both ensemble models is also significantly lower than that of most base models, indicating that the stacking approach helps mitigate the systematic overestimation and underestimation observed in individual models.

Certain base models demonstrate consistent strengths across both datasets, with CatBoost and LightGBM standing out as top performers, achieving low RMSE values and high correlation. Their strong performance is due to the use of gradient-boosting decision tree (GBDT) algorithms, which are well-suited for handling non-linear relationships and complex feature interactions. In contrast, the k-nearest neighbors model consistently underperforms, with the highest RMSE and lowest correlation, highlighting its limitations in modeling the complex variability of canopy height in both datasets. Additionally, models like Random Forest and Extra Trees performed moderately well, maintaining balanced error metrics but still showing limitations in predicting the extremes of canopy height. These models tended to predict the middle range of canopy heights relatively well but encountered challenges at the higher end, where predicted values diverged from the observed data.

When comparing individual base models, the GEDI-based models generally exhibit slightly better performance across most metrics. For instance, the correlation (R) values for all base models are higher in the GEDI dataset, with CatBoost, LightGBM, and Neural Networks reaching 0.71. In contrast, for ICESat-2, the highest correlation among base models was 0.68 (CatBoost). This suggests that GEDI, with its higher vertical sensitivity and denser sampling, provides a more accurate input for modeling canopy height. Another important observation is that the bias in the GEDI-based models is generally lower than in the ICESat-2 models. For example, the bias for CatBoost in the GEDI model is −0.20 m, while for ICESat-2 it is −1.00 m. This indicates that ICESat-2 data may be more prone to underestimation, particularly for taller canopy heights, a limitation less pronounced in the GEDI data.

In summary, while each base model has strengths and weaknesses, the stacking ensemble model clearly outperforms them by combining their predictive capabilities. The result is a more accurate and reliable canopy height model that minimizes bias and errors, making it better suited for large-scale canopy height mapping.

3.2. Comparison of Canopy Height Models Based on GEDI and ICESat-2

Figure 8 presents the performance of the stacking ensemble models for ICESat-2 and GEDI data. The GEDI-based model performed slightly better, with a correlation of 0.72 and an RMSE of 4.81 m, while the ICESat2-based model had a correlation of 0.69 and an RMSE of 4.99 m. The rRMSE values were very close, at 0.32 for GEDI and 0.33 for ICESat-2, indicating comparable overall accuracy.

The residual distribution plots (Figure 8C,D) further highlight the difference in performance between the two models. The GEDI-based model shows a more symmetric residual distribution centered around zero, with a mean residual of −0.11 m, suggesting less systematic bias. In contrast, the ICESat-2-based model has a slightly larger mean residual of −0.56 m, indicating a tendency to underpredict canopy heights, particularly for taller canopies. Additionally, the standard deviation of residuals for the GEDI-based model (4.81 m) is slightly lower than that of the ICESat-2 model (4.96 m), reflecting a more consistent prediction across the dataset.

Figure 9A,B displays the top ten most important variables in the stacking models based on ICESat-2 and GEDI data, respectively. Several key features are shared between the two models, with kNDVI standing out as one of the most important. The prominence of kNDVI highlights its advantages over traditional NDVI in handling dense vegetation. kNDVI helps mitigate the nonlinearity and saturation issues typically found in tropical

forests, leading to more reliable canopy height predictions [56]. Other spectral bands, such as Red, NIR, and SWIR1, further enhance the models' ability to capture vegetation health and structural information.
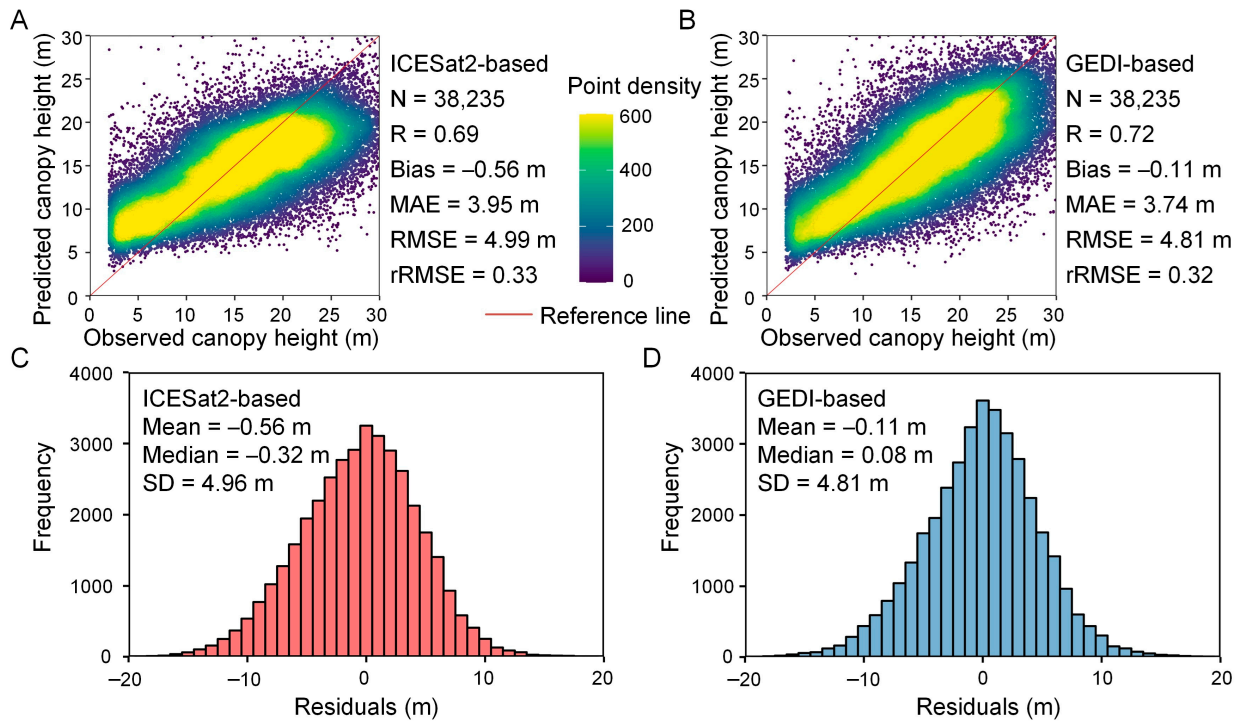

Figure 8. Performance comparison of stacking ensemble models based on (**A**) ICESat-2 and (**B**) GEDI data, showing predicted vs. observed canopy height, and the corresponding residual distributions for (**C**) ICESat-2 and (**D**) GEDI.
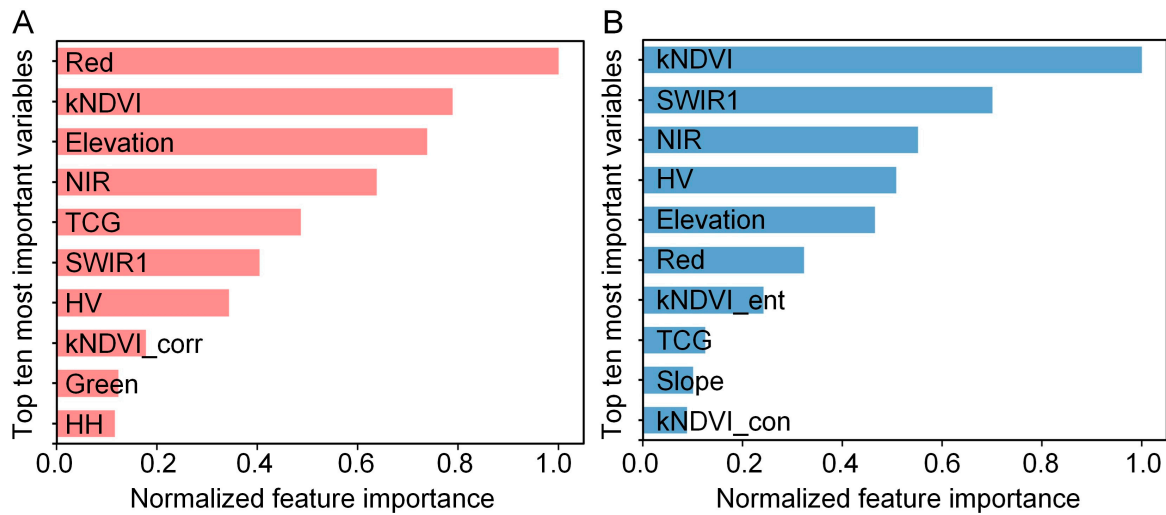

Figure 9. Top ten most important variables for canopy height prediction in stacking ensemble models based on (**A**) ICESat-2 and (**B**) GEDI data.

In addition to spectral indices, the HV polarization from PALSAR-2 is another critical feature in both models. Its better performance compared to HH polarization can be attributed to HV's ability to capture volume scattering, which is particularly useful in tropical forests with complex canopy structures. This makes HV especially valuable for predicting canopy height in such environments. Terrain features, particularly elevation, are highly ranked, likely due to the association between altitude and forest types in Puerto Rico. This highlights the importance of topography in predicting canopy height, as elevation

directly influences species distribution and forest structure. Additionally, texture features derived from kNDVI, such as entropy (kNDVI_ent), contrast (kNDVI_con), and correlation (kNDVI_corr), provide valuable insights into canopy structure by capturing spatial variability and heterogeneity. These texture metrics are useful in dense environments, where complex canopy structures make accurate predictions more challenging [35].

3.3. Canopy Height Product Mapping

Figure 10 presents three visualizations comparing terrain elevation and canopy height predictions for Puerto Rico. The terrain elevation (Figure 10A) highlights the central mountain range, which significantly influences the distribution of forest types across the island. Canopy height maps derived from ICESat-2 (Figure 10B) and GEDI (Figure 10C) data show a general alignment with topography, as taller canopies tend to occur in mountainous areas. However, the GEDI-based map provides generally higher canopy height estimates compared to the ICESat-2-based map. These differences in canopy height predictions could have implications for biomass estimation and forest structure analysis.
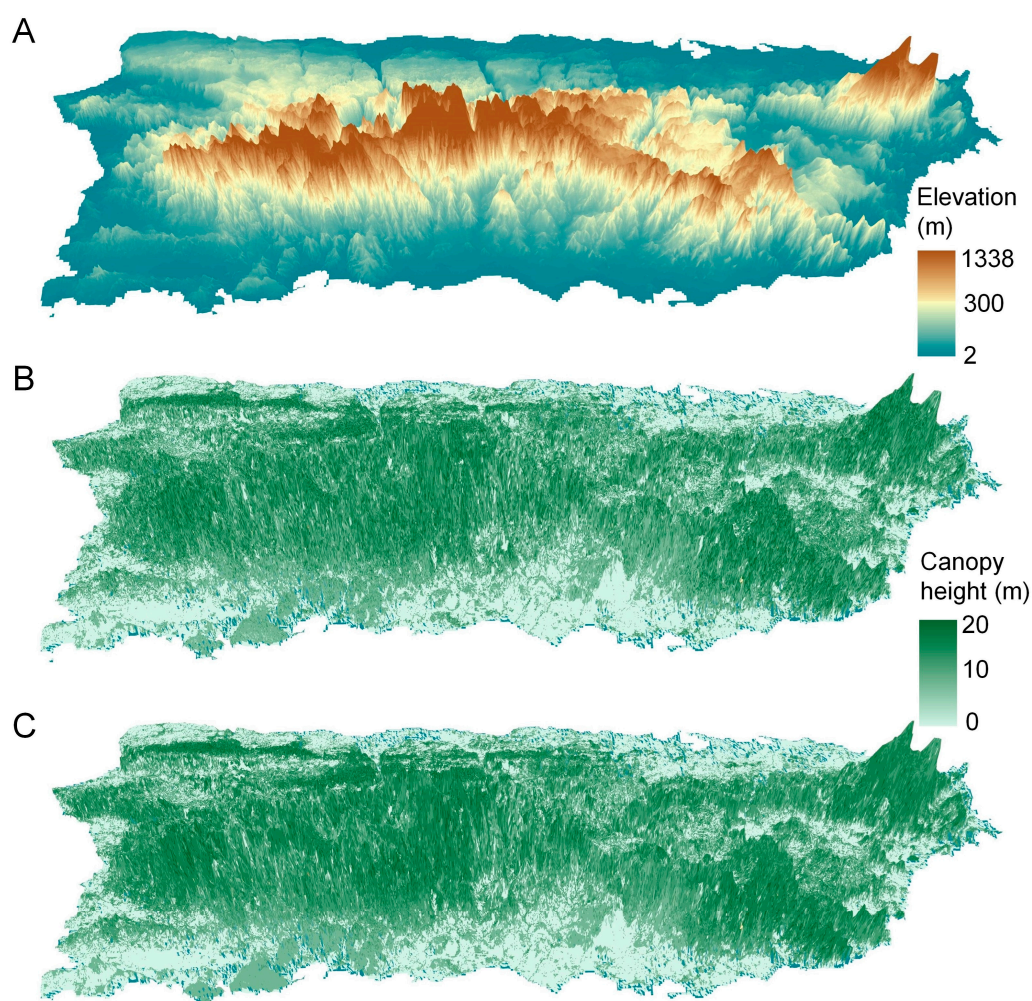

Figure 10. Three-dimensional terrain and canopy height maps. (**A**) Terrain elevation map of Puerto Rico; (**B**) canopy height map based on ICESat-2 data; (**C**) canopy height map based on GEDI data.

Figure 11 reveals key patterns in the residual distributions of canopy height predictions. As canopy height increases, the residuals exhibit a systematic shift from positive values (overestimation) to negative values (underestimation). This trend is evident in both models, but it is more pronounced in the ICESat-2-based model. For canopy heights between 2 and 4 m, both models tend to overestimate, with residuals averaging around 4 to 7 m. This overestimation in shorter canopy heights may be attributed to uncertainties

in distinguishing low vegetation or terrain features from the canopy itself. As the canopy height increases, particularly in the 16–20 m range, both models exhibit a noticeable shift towards negative residuals, indicating an increasing underestimation of canopy heights. This underestimation becomes even more pronounced for canopy heights over 20 m, with residuals reaching as low as −8 m for the tallest trees. Notably, the GEDI-based model performs better in this range, showing less severe underestimation compared to the ICESat-2-based model. This can be attributed to GEDI's denser sampling and greater sensitivity to vertical forest structure, allowing for more accurate capture of taller canopy heights [29]. This is particularly important for forest biomass estimation, where taller trees play a critical role in contributing to total biomass. Additionally, the relatively large sample size for canopy heights above 20 m further underscores the significance of accurately predicting tall canopies.
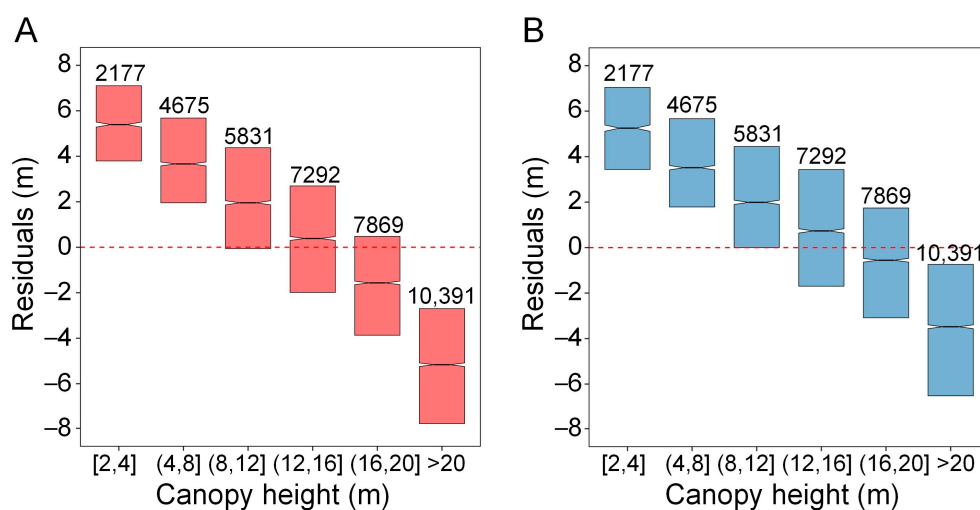

Figure 11. Residual distribution of predicted canopy heights based on (**A**) ICESat-2 and (**B**) GEDI data. The population size of each bin is indicated next to the corresponding box plot.

The residual distribution indicates a systematic bias: shorter tree canopies tend to be overestimated, while taller canopies are generally underestimated. This pattern may be attributed to the uneven distribution of forest height samples or inherent biases in the ICESat-2 ATL08 and GEDI L2A products. As shown in the forest height sample distribution (Figure 12), the canopy height estimates from ICESat-2 and GEDI footprint-level data are more concentrated in the lower canopy compared to the G-LiHT product, with particularly sparse sampling for taller canopies above 20 m. Additionally, although the limited number of overlapping footprints between GEDI, ICESat-2, and G-LiHT (see Figure 4) prevents statistical analysis for this study area, existing accuracy assessment results have highlighted that ICESat-2 tends to overestimate low canopy heights and underestimate high canopy heights, while the geolocated GEDI product performs better [29]. The overestimation of shorter trees may stem from difficulties in differentiating between the canopy and understory layers, or from how lower-canopy returns are processed. In contrast, the underestimation of taller trees, particularly in the ICESat-2 model, could be attributed to the sparse sampling of both lidar systems. The highest points of the canopy may not always generate return signals or photon counts, resulting in an overall lower estimate of canopy height for the tallest trees [23].
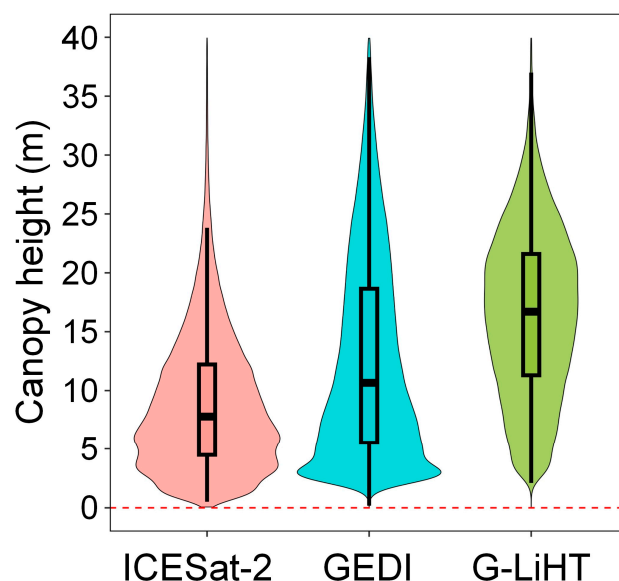
Figure 12. Comparison of canopy height distributions estimated from G-LiHT, ICESat-2, and GEDI footprint-level data.

4. Discussion

Our study highlights the practical utility of ICESat-2 and GEDI spaceborne lidar products in accurately estimating tropical forest canopy height, especially when combined with multispectral, SAR, and terrain data to produce spatially continuous 30 m resolution canopy height maps. The integration of Landsat-8, PALSAR-2, and FABDEM resulted in a more comprehensive feature set, surpassing conventional approaches that predominantly rely on spectral data alone. This research also underscores the importance of incorporating kNDVI and its texture features into canopy height inversion models to enhance prediction accuracy. While traditional vegetation indices like NDVI often face saturation challenges in dense tropical forests, kNDVI provides more reliable information on vegetation structure. The addition of texture metrics provides further insight by capturing subtle variations in canopy density and spatial complexity, which are favorable for understanding forest structure.

Previous studies [9,16–18] have emphasized the strong performance of machine learning methods, such as Random Forest, in estimating forest canopy height. Our study extends these insights by employing a more advanced ensemble learning approach. The stacking ensemble models developed using the AutoGluon automated machine learning framework consistently outperformed individual base models in terms of inversion accuracy. Our canopy height model for Puerto Rico shows slightly higher overall accuracy compared to previous studies on tropical forests in South America and Africa [35], temperate forests in China [33], and boreal forests in Canada [30], all of which reported RMSE values around 5 m. This indicates that our approach has contributed to enhancing the accuracy of forest canopy height inversion. Accurate canopy height estimation can greatly enhance the precision of carbon stock assessments, providing critical information for climate change mitigation strategies. Additionally, the ability to monitor changes in forest structure over time supports biodiversity conservation efforts by identifying areas of habitat degradation and facilitating targeted protection measures.

Our residual analysis provides valuable insights for improving canopy height mapping using ICESat-2 and GEDI data. The sparse sampling of taller canopies in the ICESat-2 and GEDI datasets may introduce bias in estimating forest structure at greater heights (Figure 12). Both datasets tend to underestimate taller canopies, and given the disproportionate contribution of tall trees to forest biomass and carbon storage, accurate height estimates are critical for forest monitoring and conservation efforts. Future research should prioritize addressing the systematic biases in canopy height predictions and explore the

integration of complementary datasets to enhance overall accuracy [70]. A promising direction is to combine GEDI and ICESat-2 data with higher-resolution airborne lidar products for localized validation and improved model precision. Additionally, while our machine learning models performed well overall, potential overfitting or underfitting in certain height ranges remains a concern. Future research could focus on integrating more diverse datasets to complement ICESat-2 and GEDI data. The upcoming long-wavelength SAR missions, such as the P-band BIOMASS mission [71], offer the potential to penetrate dense forests more effectively, providing enhanced indicators for tropical forest canopy height inversion.

In conclusion, this study underscores the importance of integrating multiple data sources for accurate monitoring of forest structure. The integration of ICESat-2 and GEDI data with enhanced remote sensing features and advanced machine learning techniques allows for high-precision, large-scale canopy height mapping. By advancing both the methodology and data used in canopy height modeling, this research lays a foundation for future studies aimed at further improving accuracy.

Author Contributions: Conceptualization, A.L. and Y.C.; methodology, A.L.; software, A.L.; validation, A.L. and Y.C.; formal analysis, A.L., Y.C. and X.C.; data curation, A.L.; writing—original draft preparation, A.L.; writing—review and editing, A.L., Y.C. and X.C.; visualization, A.L.; supervision, Y.C.; project administration, Y.C.; funding acquisition, A.L., Y.C. and X.C. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data can be sourced from the following providers: The Landsat-8 Surface Reflectance data, PALSAR-2 annual mosaic product, and the FABDEM Terrain Product are available through Google Earth Engine (https://developers.google.com/earth-engine/datasets/; accessed on 10 October 2024). The G-LiHT Airborne Lidar Product (accessed on 10 October 2024) is accessible at https://glihtdata.gsfc.nasa.gov/. The GEDI L2A Product (accessed on 10 October 2024) can be obtained from https://lpdaac.usgs.gov/products/gedi02_av002/, and the ICESat-2 ATL08 Product (accessed on 10 October 2024) is available at https://nsidc.org/data/atl08/versions/6/.

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

References

1. Wani, A.M.; Sahoo, G. Forest Ecosystem Services and Biodiversity. In *Spatial Modeling in Forest Resources Management: Rural Livelihood and Sustainable Development*; Shit, P.K., Pourghasemi, H.R., Das, P., Bhunia, G.S., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 529–552, ISBN 978-3-030-56542-8.
2. Bastin, J.-F.; Finegold, Y.; Garcia, C.; Mollicone, D.; Rezende, M.; Routh, D.; Zohner, C.M.; Crowther, T.W. The Global Tree Restoration Potential. *Science* **2019**, *365*, 76–79. [CrossRef] [PubMed]
3. Friedlingstein, P.; O'Sullivan, M.; Jones, M.W.; Andrew, R.M.; Gregor, L.; Hauck, J.; Le Quéré, C.; Luijkx, I.T.; Olsen, A.; Peters, G.P.; et al. Global Carbon Budget 2022. *Earth Syst. Sci. Data* **2022**, *14*, 4811–4900. [CrossRef]
4. Pugh, T.A.M.; Lindeskog, M.; Smith, B.; Poulter, B.; Arneth, A.; Haverd, V.; Calle, L. Role of Forest Regrowth in Global Carbon Sink Dynamics. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 4382–4387. [CrossRef]
5. Harris, N.L.; Gibbs, D.A.; Baccini, A.; Birdsey, R.A.; de Bruin, S.; Farina, M.; Fatoyinbo, L.; Hansen, M.C.; Herold, M.; Houghton, R.A.; et al. Global Maps of Twenty-First Century Forest Carbon Fluxes. *Nat. Clim. Chang.* **2021**, *11*, 234–240. [CrossRef]
6. Atmadja, S.S.; Duchelle, A.E.; Sy, V.D.; Selviana, V.; Komalasari, M.; Sills, E.O.; Angelsen, A. How Do REDD+ Projects Contribute to the Goals of the Paris Agreement? *Environ. Res. Lett.* **2022**, *17*, 044038. [CrossRef]
7. Colglazier, W. Sustainable Development Agenda: 2030. *Science* **2015**, *349*, 1048–1050. [CrossRef]

8.   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]

9.   Nandy, S.; Srinet, R.; Padalia, H. Mapping Forest Height and Aboveground Biomass by Integrating ICESat-2, Sentinel-1 and Sentinel-2 Data Using Random Forest Algorithm in Northwest Himalayan Foothills of India. *Geophys. Res. Lett.* **2021**, *48*, e2021GL093799. [CrossRef]

10. Lei, Y.; Treuhaft, R.; Gonçalves, F. Automated Estimation of Forest Height and Underlying Topography over a Brazilian Tropical Forest with Single-Baseline Single-Polarization TanDEM-X SAR Interferometry. *Remote Sens. Environ.* **2021**, *252*, 112132. [CrossRef]

11. Potapov, P.; Tyukavina, A.; Turubanova, S.; Talero, Y.; Hernandez-Serna, A.; Hansen, M.C.; Saah, D.; Tenneson, K.; Poortinga, A.; Aekakkararungroj, A.; et al. Annual Continuous Fields of Woody Vegetation Structure in the Lower Mekong Region from 2000–2017 Landsat Time-Series. *Remote Sens. Environ.* **2019**, *232*, 111278. [CrossRef]

12. Meddens, A.J.H.; Vierling, L.A.; Eitel, J.U.H.; Jennewein, J.S.; White, J.C.; Wulder, M.A. Developing 5 m Resolution Canopy Height and Digital Terrain Models from WorldView and ArcticDEM Data. *Remote Sens. Environ.* **2018**, *218*, 174–188. [CrossRef]

13. Nie, S.; Wang, C.; Xi, X.; Luo, S.; Zhu, X.; Li, G.; Liu, H.; Tian, J.; Zhang, S. Assessing the Impacts of Various Factors on Treetop Detection Using LiDAR-Derived Canopy Height Models. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 10099–10115. [CrossRef]

14. Benson, M.L.; Pierce, L.; Bergen, K.; Sarabandi, K. Model-Based Estimation of Forest Canopy Height and Biomass in the Canadian Boreal Forest Using Radar, LiDAR, and Optical Remote Sensing. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 4635–4653. [CrossRef]

15. 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]

16. Rishmawi, K.; Huang, C.; Zhan, X. Monitoring Key Forest Structure Attributes across the Conterminous United States by Integrating GEDI LiDAR Measurements and VIIRS Data. *Remote Sens.* **2021**, *13*, 442. [CrossRef]

17. Tommaso, S.D.; Wang, S.; Lobell, D.B. Combining GEDI and Sentinel-2 for Wall-to-Wall Mapping of Tall and Short Crops. *Environ. Res. Lett.* **2021**, *16*, 125002. [CrossRef]

18. Zhu, X.; Wang, C.; Nie, S.; Pan, F.; Xi, X.; Hu, Z. Mapping Forest Height Using Photon-Counting LiDAR Data and Landsat 8 OLI Data: A Case Study in Virginia and North Carolina, USA. *Ecol. Indic.* **2020**, *114*, 106287. [CrossRef]

19. Healey, S.P.; Yang, Z.; Gorelick, N.; Ilyushchenko, S. Highly Local Model Calibration with a New GEDI LiDAR Asset on Google Earth Engine Reduces Landsat Forest Height Signal Saturation. *Remote Sens.* **2020**, *12*, 2840. [CrossRef]

20. Joshi, N.; Mitchard, E.T.A.; Brolly, M.; Schumacher, J.; Fernández-Landa, A.; Johannsen, V.K.; Marchamalo, M.; Fensholt, R. Understanding 'Saturation' of Radar Signals over Forests. *Sci. Rep.* **2017**, *7*, 3505. [CrossRef]

21. Coops, N.C.; Tompalski, P.; Goodbody, T.R.H.; 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]

22. Bolton, D.K.; Tompalski, P.; Coops, N.C.; White, J.C.; Wulder, M.A.; Hermosilla, T.; Queinnec, M.; Luther, J.E.; van Lier, O.R.; Fournier, R.A.; et al. Optimizing Landsat Time Series Length for Regional Mapping of Lidar-Derived Forest Structure. *Remote Sens. Environ.* **2020**, *239*, 111645. [CrossRef]

23. Neuenschwander, A.L.; Magruder, L.A. Canopy and Terrain Height Retrievals with ICESat-2: A First Look. *Remote Sens.* **2019**, *11*, 1721. [CrossRef]

24. Markus, T.; Neumann, T.; Martino, A.; Abdalati, W.; Brunt, K.; Csatho, B.; Farrell, S.; Fricker, H.; Gardner, A.; Harding, D.; et al. The Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2): Science Requirements, Concept, and Implementation. *Remote Sens. Environ.* **2017**, *190*, 260–273. [CrossRef]

25. 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]

26. Schneider, F.D.; Ferraz, A.; Hancock, S.; Duncanson, L.I.; Dubayah, R.O.; Pavlick, R.P.; Schimel, D.S. Towards Mapping the Diversity of Canopy Structure from Space with GEDI. *Environ. Res. Lett.* **2020**, *15*, 115006. [CrossRef]

27. Magruder, L.A.; Brunt, K.M.; Alonzo, M. Early ICESat-2 on-Orbit Geolocation Validation Using Ground-Based Corner Cube Retro-Reflectors. *Remote Sens.* **2020**, *12*, 3653. [CrossRef]

28. Magruder, L.A.; Brunt, K.M. Performance Analysis of Airborne Photon- Counting Lidar Data in Preparation for the ICESat-2 Mission. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2911–2918. [CrossRef]

29. Liu, A.; Cheng, X.; Chen, Z. Performance Evaluation of GEDI and ICESat-2 Laser Altimeter Data for Terrain and Canopy Height Retrievals. *Remote Sens. Environ.* **2021**, *264*, 112571. [CrossRef]

30. 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]

31. Potapov, P.; 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]

32. Lang, N.; Kalischek, N.; Armston, J.; Schindler, K.; Dubayah, R.; Wegner, J.D. Global Canopy Height Regression and Uncertainty Estimation from GEDI LIDAR Waveforms with Deep Ensembles. *Remote Sens. Environ.* **2022**, *268*, 112760. [CrossRef]

33. Liu, X.; Su, Y.; Hu, T.; Yang, Q.; Liu, B.; Deng, Y.; Tang, H.; Tang, Z.; Fang, J.; Guo, Q. Neural Network Guided Interpolation for Mapping Canopy Height of China's Forests by Integrating GEDI and ICESat-2 Data. *Remote Sens. Environ.* **2022**, *269*, 112844. [CrossRef]

34. Roy, D.P.; Kashongwe, H.B.; Armston, J. The Impact of Geolocation Uncertainty on GEDI Tropical Forest Canopy Height Estimation and Change Monitoring. *Sci. Remote Sens.* **2021**, *4*, 100024. [CrossRef]

35. Ngo, Y.-N.; Ho Tong Minh, D.; Baghdadi, N.; Fayad, I. Tropical Forest Top Height by GEDI: From Sparse Coverage to Continuous Data. *Remote Sens.* **2023**, *15*, 975. [CrossRef]

36. Lahssini, K.; Baghdadi, N.; le Maire, G.; Fayad, I. Influence of GEDI Acquisition and Processing Parameters on Canopy Height Estimates over Tropical Forests. *Remote Sens.* **2022**, *14*, 6264. [CrossRef]

37. Erickson, N.; Mueller, J.; Shirkov, A.; Zhang, H.; Larroy, P.; Li, M.; Smola, A. AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data. *arXiv* **2020**, arXiv:2003.06505. Available online: https://arxiv.org/abs/2003.06505v1 (accessed on 7 September 2024).

38. Daly, C.; Helmer, E.H.; Quiñones, M. Mapping the Climate of Puerto Rico, Vieques and Culebra. *Int. J. Climatol.* **2003**, *23*, 1359–1381. [CrossRef]

39. Zanaga, D.; Van De Kerchove, R.; Daems, D.; De Keersmaecker, W.; Brockmann, C.; Kirches, G.; Wevers, J.; Cartus, O.; Santoro, M.; Fritz, S.; et al. ESA WorldCover 10 m 2021 V200. 2022. Available online: https://zenodo.org/records/7254221 (accessed on 10 October 2024).

40. Lugo, A.E.; Helmer, E. Emerging Forests on Abandoned Land: Puerto Rico's New Forests. *For. Ecol. Manag.* **2004**, *190*, 145–161. [CrossRef]

41. Cook, B.D.; Corp, L.A.; Nelson, R.F.; Middleton, E.M.; Morton, D.C.; McCorkel, J.T.; Masek, J.G.; Ranson, K.J.; Ly, V.; Montesano, P.M. NASA Goddard's LiDAR, Hyperspectral and Thermal (G-LiHT) Airborne Imager. *Remote Sens.* **2013**, *5*, 4045–4066. [CrossRef]

42. Hawker, L.; Uhe, P.; Paulo, L.; Sosa, J.; Savage, J.; Sampson, C.; Neal, J. A 30 m Global Map of Elevation with Forests and Buildings Removed. *Environ. Res. Lett.* **2022**, *17*, 024016. [CrossRef]

43. Neuenschwander, A.; Pitts, K. The ATL08 Land and Vegetation Product for the ICESat-2 Mission. *Remote Sens. Environ.* **2019**, *221*, 247–259. [CrossRef]

44. 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]

45. Sumnall, M.J.; Hill, R.A.; Hinsley, S.A. Comparison of Small-Footprint Discrete Return and Full Waveform Airborne Lidar Data for Estimating Multiple Forest Variables. *Remote Sens. Environ.* **2016**, *173*, 214–223. [CrossRef]

46. Chen, Y.; Liu, A.; Cheng, X. Landsat-Based Monitoring of Landscape Dynamics in Arctic Permafrost Region. *J. Remote Sens.* **2022**, *2022*, 9765087. [CrossRef]

47. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]

48. Liu, A.; Wu, Q.; Cheng, X. Using the Google Earth Engine to Estimate a 10 m Resolution Monthly Inventory of Soil Fugitive Dust Emissions in Beijing, China. *Sci. Total Environ.* **2020**, *735*, 139174. [CrossRef] [PubMed]

49. Sinha, S.; Jeganathan, C.; Sharma, L.K.; Nathawat, M.S. A Review of Radar Remote Sensing for Biomass Estimation. *Int. J. Environ. Sci. Technol.* **2015**, *12*, 1779–1792. [CrossRef]

50. Kankaku, Y.; Suzuki, S.; Osawa, Y. ALOS-2 Mission and Development Status. In Proceedings of the 2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS, Melbourne, VIC, Australia, 21–26 July 2013; pp. 2396–2399.

51. Shimada, M.; Itoh, T.; Motooka, T. Regenerated ALOS-2/PALSAR-2 Global Mosaics 2016 and 2014/2015 for Forest Observations. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 2454–2457.

52. Koyama, C.N.; Watanabe, M.; Hayashi, M.; Ogawa, T.; Shimada, M. Mapping the Spatial-Temporal Variability of Tropical Forests by ALOS-2 L-Band SAR Big Data Analysis. *Remote Sens. Environ.* **2019**, *233*, 111372. [CrossRef]

53. Marsh, C.B.; Harder, P.; Pomeroy, J.W. Validation of FABDEM, a Global Bare-Earth Elevation Model, against UAV-Lidar Derived Elevation in a Complex Forested Mountain Catchment. *Environ. Res. Commun.* **2023**, *5*, 031009. [CrossRef]

54. Baig, M.H.A.; Zhang, L.; Shuai, T.; Tong, Q. Derivation of a Tasselled Cap Transformation Based on Landsat 8 At-Satellite Reflectance. *Remote Sens. Lett.* **2014**, *5*, 423–431. [CrossRef]

55. Camps-Valls, G.; Campos-Taberner, M.; Moreno-Martínez, Á.; Walther, S.; Duveiller, G.; Cescatti, A.; Mahecha, M.D.; Muñoz-Marí, J.; García-Haro, F.J.; Guanter, L.; et al. A Unified Vegetation Index for Quantifying the Terrestrial Biosphere. *Sci. Adv.* **2021**, *7*, eabc7447. [CrossRef] [PubMed]

56. Wang, Q.; Moreno-Martínez, Á.; Muñoz-Marí, J.; Campos-Taberner, M.; Camps-Valls, G. Estimation of Vegetation Traits with Kernel NDVI. *ISPRS J. Photogramm. Remote Sens.* **2023**, *195*, 408–417. [CrossRef]

57. Roberti de Siqueira, F.; Robson Schwartz, W.; Pedrini, H. Multi-Scale Gray Level Co-Occurrence Matrices for Texture Description. *Neurocomputing* **2013**, *120*, 336–345. [CrossRef]

58. Mueller, J.; Shi, X.; Smola, A. Faster, Simpler, More Accurate: Practical Automated Machine Learning with Tabular, Text, and Image Data. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event, CA USA, 6–10 July 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 3509–3510.

59. A Branch and Bound Algorithm for Computing K-Nearest Neighbors. Available online: https://ieeexplore.ieee.org/abstract/document/1672890 (accessed on 7 September 2024).
60. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
61. Abdi, H.; Valentin, D.; Edelman, B. *Neural Networks*; SAGE: Newcastle upon Tyne, UK, 1999; ISBN 978-0-7619-1440-2.
62. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely Randomized Trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]
63. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794.
64. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Curran Associates, Inc.: Brooklyn, NY, USA, 2017; Volume 30.
65. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased Boosting with Categorical Features. In Proceedings of the Advances in Neural Information Processing Systems, Montréal, QC, Canada, 3–8 December 2018; Curran Associates, Inc.: Brooklyn, NY, USA, 2018; Volume 31.
66. Sim, S.; Im, J.; Jung, S.; Han, D. Improving Short-Term Prediction of Ocean Fog Using Numerical Weather Forecasts and Geostationary Satellite-Derived Ocean Fog Data Based on AutoML. *Remote Sens.* **2024**, *16*, 2348. [CrossRef]
67. Chen, Y.; Cheng, X.; Liu, A.; Chen, Q.; Wang, C. Tracking Lake Drainage Events and Drained Lake Basin Vegetation Dynamics across the Arctic. *Nat. Commun.* **2023**, *14*, 7359. [CrossRef]
68. Liu, A.; Chen, Y.; Cheng, X. Monitoring Thermokarst Lake Drainage Dynamics in Northeast Siberian Coastal Tundra. *Remote Sens.* **2023**, *15*, 4396. [CrossRef]
69. Liu, A.; Chen, Y.; Cheng, X. Effects of Thermokarst Lake Drainage on Localized Vegetation Greening in the Yamal–Gydan Tundra Ecoregion. *Remote Sens.* **2023**, *15*, 4561. [CrossRef]
70. Zhu, X.; Nie, S.; Wang, C.; Xi, X.; Lao, J.; Li, D. Consistency Analysis of Forest Height Retrievals between GEDI and ICESat-2. *Remote Sens. Environ.* **2022**, *281*, 113244. [CrossRef]
71. Quegan, S.; Le Toan, T.; Chave, J.; Dall, J.; Exbrayat, J.-F.; Minh, D.H.T.; Lomas, M.; D'Alessandro, M.M.; Paillou, P.; Papathanassiou, K.; et al. The European Space Agency BIOMASS Mission: Measuring Forest above-Ground Biomass from Space. *Remote Sens. Environ.* **2019**, *227*, 44–60. [CrossRef]