remote sensing

MDPI

Article

Improving National Forest Mapping in Romania Using Machine Learning and Sentinel-2 Multispectral Imagery

Mohamed Islam Keskes, Aya Hamed Mohamed, Stelian Alexandru Borz and Mihai Daniel Niţă *

Department of Forest Engineering, Forest Management Planning and Terrestrial Measurements, Faculty of Silviculture and Forest Engineering, Transilvania University of Brasov, Şirul Beethoven 1, 500123 Brasov, Romania; mohamed.keskes@unitbv.ro (M.I.K.); aya.hamed@unitbv.ro (A.H.M.); stelian.borz@unitbv.ro (S.A.B.)
* Correspondence: mihai.nita@unitbv.ro

Abstract: Forest attributes, such as standing stock, diameter at breast height (DBH), tree height, and basal area, are critical for effective forest management; yet, traditional estimation methods remain labor-intensive and often lack the spatial detail required for contemporary decision-making. This study addresses these challenges by integrating machine learning algorithms with high-resolution remotely sensed data and rigorously collected ground truth measurements to produce accurate, national-scale maps of forest attributes in Romania. To ensure the reliability of the model predictions, extensive field campaigns were conducted across representative Romanian forests. During these campaigns, detailed measurements were recorded for every tree within selected plots. For each tree, DBH was measured directly, and tree heights were obtained either by direct measurement—using hypsometers or clinometers—or, when direct measurements were not feasible, by applying well-established DBH—height allometric relationships that have been calibrated for the local forest types. This comprehensive approach to ground data collection, supplemented by an independent dataset from Brasov County collected using the same protocols, allowed for robust training and validation of the machine learning models. This study evaluates the performance of three machine learning algorithms—Random Forest (RF), Classification and Regression Trees (CART), and the Gradient Boosting Tree Algorithm (GBTA)—in predicting the forest attributes from Sentinel-2 satellite imagery. While Random Forest consistently delivered high R^2 values and low root mean square errors (RMSE) across all attributes, GBTA showed particular strength in predicting standing stock, and CART excelled in basal area estimation but was less reliable for other attributes. A sensitivity analysis across multiple spatial resolutions revealed that the performance of all algorithms varied significantly with changes in resolution, emphasizing the importance of selecting an appropriate scale for accurate forest mapping. By focusing on both the methodological advancements in machine learning applications and the rigorous, detailed empirical forest data collection, this study provides a clear solution to the problem of obtaining reliable, spatially detailed forest attribute maps.

Keywords: forest attributes; spatial resolution; machine learning; remotely sensed data; prediction; performance; decision

Academic Editors: Brenden E. McNeil and Hubert Hasenauer

Received: 12 December 2024
Revised: 8 February 2025
Accepted: 14 February 2025
Published: 19 February 2025

Citation: Keskes, M.I.; Mohamed, A.H.; Borz, S.A.; Niţă, M.D. Improving National Forest Mapping in Romania Using Machine Learning and Sentinel-2 Multispectral Imagery. *Remote Sens.* **2025**, *17*, 715. https://doi.org/10.3390/rs17040715

1. Introduction

Romania's forests are among the most diverse in Europe, hosting a range of ecosystems that significantly contribute to carbon sequestration, biodiversity conservation, and sustainable resource management [1]. The National Forest Inventory (NFI) plays a vital

role in monitoring forest dynamics, shaping policies, and supporting research on forest attributes [2,3]. Advanced methodologies can be developed through NFI data for large-scale forest assessment and management, ensuring informed decision-making and sustainable forestry practices [4]. Romania, after 50 years of forest management planning at a national level, still lacks a unified database to characterize all the forests, with more than 700,000 hectares of forests without a proper inventory [5–7].

In order to sustainably use this resource, effective forest management requires precise and up-to-date information on key attributes [8–10]. Critical parameters, such as standing stock (Vol), basal area (BA), and aboveground biomass (AGB), provide valuable insights into forest structure and productivity [10–12]. Standing stock estimates are essential for sustainable timber harvesting, while BA helps assess stand density and forest health [10]. Additionally, AGB, which can be measured from allometric equations based on diameter at breast height (DBH) and tree height (H) [13], plays a crucial role in evaluating carbon storage potential, influencing decarbonization policies, and supporting climate change mitigation efforts [14]. By integrating NFI data with these key parameters, forest managers can develop more efficient strategies to ensure the long-term health and sustainability of Romania's forests [4,6].

To effectively utilize forest attributes in decision-making, comprehensive and detailed maps of Romania's forest characteristics are essential [4,15,16]. These maps not only capture spatial variability but also highlight high-conservation-value regions and improve resource allocation [17]. National-scale forest mapping allows policymakers and researchers to monitor environmental changes, assess deforestation risks, and develop data-driven conservation strategies [18]. Achieving accurate and up-to-date mapping requires integrating remote sensing, machine learning (ML), and field-based observations [19]. However, Romania's diverse forest landscapes necessitate careful calibration of remote sensing data with regional measurements [17,20]. Variability in data collection methods, sensor limitations, and environmental factors introduce uncertainties, making robust validation frameworks crucial for ensuring the reliability of these forest assessments [20].

Given these challenges, ML models provide a powerful solution for extracting meaningful patterns from complex forest datasets, improving the accuracy and scalability of forest attribute estimation. Selecting the right model is important as different models offer unique advantages and limitations depending on forest structure complexity, data availability, and computational resources. ANN has shown high accuracy in complex, nonlinear forest attribute prediction tasks but requires large datasets and extensive computational resources [21]. Studies demonstrated that ANN outperforms traditional regression models for AGB estimation but suffers from overfitting when data quality is inconsistent [22]. SVM is effective in handling high-dimensional data and has been successfully used for DBH and BA prediction [23]. However, SVM's performance is highly sensitive to kernel selection and parameter tuning, making it less adaptable to varying forest conditions. KNN, while easy to implement and interpret, has limitations in handling large datasets and tends to be affected by noisy inputs, as observed in forest volume prediction L.

Among non-parametric algorithms, RF, GBTA (Gradient Boosting Tree Algorithm), and CART (Classification and Regression Trees) have demonstrated their effectiveness in accurately estimating various forest characteristics such as the aboveground biomass [24–26], forest canopy height [23], and standing stock [27,28]. The RF algorithm minimizes prediction variance in supervised ML by combining multiple decision trees through a process called bagging [29]. Initially, each tree is trained on random subsets of data; during validation, the algorithm averages the predictions to come to a balanced prediction outcome [29,30]. Its effectiveness lies in handling intricate decision boundaries for both categorical and continuous variables, managing noisy and high-dimensional datasets, and

minimizing overfitting [31]. This algorithm incorporates randomization techniques to reduce tree correlation, which enhances the model's accuracy. It excels in handling large datasets with numerous variables and complex interactions between them [32].

The CART algorithm is a versatile binary recursive partitioning method for classification and regression tasks [33]. It constructs a tree structure by iteratively dividing data into increasingly homogeneous subsets based on selected features and threshold values; the process continues until stopping conditions, such as reaching maximum depth or a minimum number of samples per leaf node, are met [34]. The resulting tree structure can be pruned to prevent overfitting, which improves the generalization capabilities [35]. CART accommodates both categorical and continuous variables, handles multi-class tasks, and provides interpretable decision trees for insights into feature–target connections [36].

The GBTA is a powerful ML algorithm that enhances prediction accuracy by building an ensemble of decision trees based on the boosting concept [37]. As the abovementioned methods show, it corrects errors from previous models by iteratively fitting decision trees to residuals and preventing overfitting through regularization [34]. Popular implementations like XGBoost, LightGBM, and GBM extend GBTA's capabilities using different optimization techniques [38]. While GBTA offers high accuracy and adaptability, achieving optimal performance requires meticulous hyperparameter tuning [35].

While some studies [25,39–42] have focused on specific forest attributes or localized areas, a significant gap remains in national-scale forest mapping. ML models perform best in well-defined domains, but localized data can introduce uncertainties in species diversity, topography, and density, limiting generalizability [43]. Moreover, hyperparameter tuning that works well for a limited domain may not hold under broader conditions [44]. Testing ML algorithms at national scales, such as in Romania's diverse ecosystems, provides valuable insights into their adaptability [25,45]. Additionally, model complexity influences predictive capabilities—complex models may capture intricate patterns yet risk overfitting, while simpler models generalize better but might miss fine-scale variations [39,46]. Finally, spatial resolution plays a crucial role, as coarser resolutions may lose detail [47], whereas finer resolutions increase computational demands [45].

Additionally, the choice of spatial resolution affects the performance of ML algorithms at larger scales [48]. Coarser spatial resolutions may lead to loss of detail and less variability in the data, potentially affecting the accuracy of predictions [47]; conversely, finer spatial resolutions can provide more detailed information but may also increase computational complexity and resource requirements [45].

Some studies [49,50] advanced our understanding of mapping forest attributes at a larger scale, such as the national level, but there remains a critical oversight regarding the process of testing on unseen data, which can be provided by independent datasets. This process is essential for assessing the generalizability and robustness of the models [51].

The aim of this study is to model and map key forest characteristics in Romania by integrating ML algorithms with remotely sensed data. By leveraging advanced ML techniques alongside Sentinel-2 satellite imagery, this research seeks to generate comprehensive and detailed maps of forest attributes such as standing stock Vol, BA, DBH, and H at the national level. Given the diversity of Romania's forests, the findings are intended to be representative of temperate forests globally, offering insights that can inform sustainable forest management and policy-making on a broader scale.

To achieve this aim, this study compares the performance of three ML algorithms (RF, GBTA, and CART) in predicting key forest attributes using Sentinel-2 data. This comparison is designed to highlight the strengths and limitations of each algorithm, ultimately guiding the selection of the most appropriate model for accurate forest attribute estimation. Additionally, this research investigates the influence of model complexity on predictive

accuracy, exploring whether increased complexity in the models leads to better performance or if simpler models can deliver comparable results. This aspect of this study will clarify when added complexity is justified in the context of forest mapping. Furthermore, this study assesses the effects of spatial resolution on model performance to identify the optimal resolution needed to capture fine-scale forest attributes accurately. This evaluation aims to provide practical insights into the spatial data requirements necessary for effective, large-scale forest assessments.

2. Materials and Methods

2.1. Study Area and Field Data

Romanian forests exhibit a high degree of variability, encompassing a wide range of forest types, management practices, and ecological features, which makes them representative of broader European forest conditions. The overall climate of Romania is temperate (Figure 1), and the mean elevation is 330 m, but it varies widely from 0 to 2544 m a.s.l. [52]. About 18% of the study area holds a rugged terrain (>15° slope), while the rest is relatively flat (<5°, 52%) or rolling (5–15°, 30%) [53].

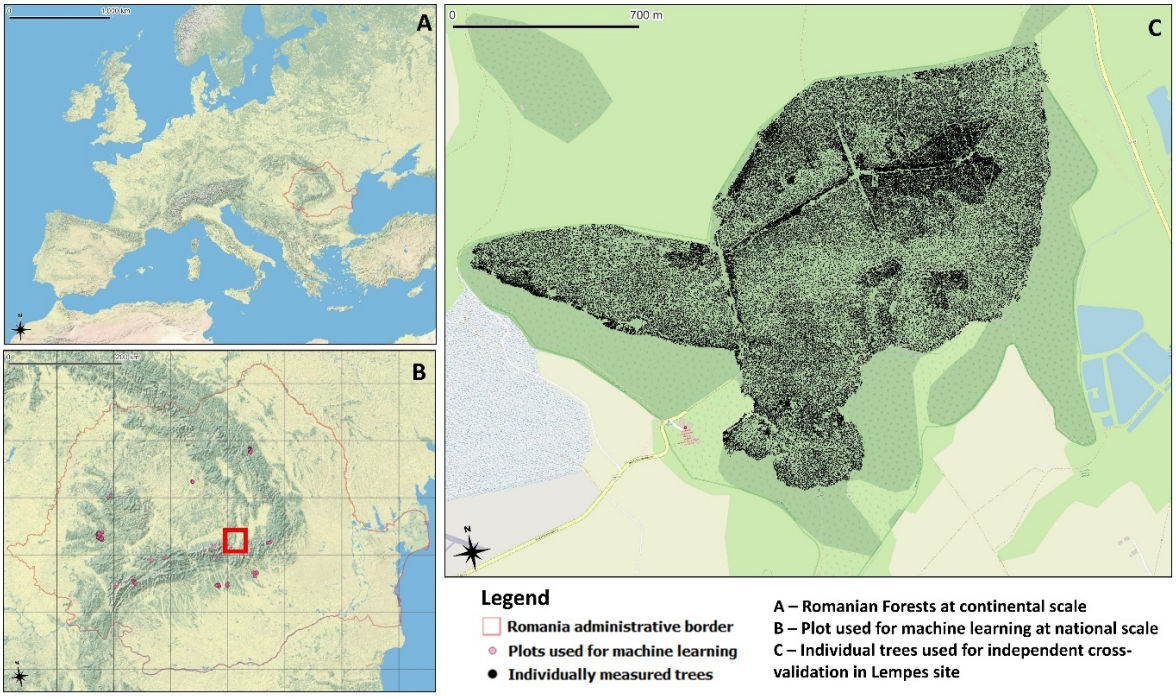

Figure 1. Study area's geographic location: (**A**) Romanian forests at continental scale; (**B**) Plots locations as red dots used at national scale; (**C**) Individual tree measurements used for independent validation in Lempes test area.

Romanian forests encompass a variety of ecosystems, including mixed broadleaved forests, coniferous stands, and temperate woodlands, each defined by unique structural and compositional traits [54]. The country's topography ranges from lowland plains to mountainous regions, creating diverse microclimates and soil conditions that support species such as beech (*Fagus sylvatica*), oak (*Quercus* spp.), and spruce (*Picea abies*) [55]. Forest management practices range from conservation-focused approaches in protected areas to production forestry, employing sustainable techniques like selective logging and continuous cover forestry to balance biodiversity and ecosystem health [56,57].

These diverse forest conditions make Romania an ideal case study for temperate forest ecosystems, with insights applicable across Europe. At the national level, this study relied on ground-based measurements collected from 1325 circular plots (300–500 m^2 each)

distributed across 10 representative forest types (Figure A1 in the Appendix A), sampled from June to September 2022 (Figure 1B). Data included diameter at breast height, tree height, species, and position, gathered via traditional inventory methods and mobile LiDAR [58] (Table A1 in Appendix A). Metrics like BA measured in m^2/ha, forest stand stock volume measured in m^3/ha, average (DBH) measured in cm, and dominant tree height measured in m, were calculated using national equations [59] and extrapolated per hectare based on the type of the plot. The national dataset exhibited significant variability, with BA ranging from 0.01 to 95.19 m^2/ha (mean: 30.55 m^2/ha), DBH averaging 32.19 cm, and Vol at 389.64 m^3/ha, reflecting the structural heterogeneity of Romania's forests. For model training at the national level, 70% of the data was used, leaving 30% for validation to ensure robust evaluations [60].

An independent testing dataset was collected in the Dealul Cetăţii Lempeş area, a temperate forest characterized by diverse habitats and climatic conditions, with a maximum elevation of 704 m. Far from being representative of the national level, this dataset provided a full tree spatial inventory on an area large enough to assess algorithmic predictions (Figure 1C). This dataset consists of individual tree measurements of all the trees on an area of 168 hectares (Table A2 in Appendix A). Forest attribute maps were generated at 10 m, 50 m, and 100 m resolutions by aggregation and extrapolation using QGIS (v3.34.1), enabling detailed analysis of spatial patterns and algorithm performance. The choice of these resolutions was guided by their practical relevance in forestry applications. These resolutions align with the remote sensing literature, where similar scales have been used for biomass estimation and forest structure assessments [61,62].

2.2. Remotely Sensed Data and Feature Extraction

The Google Earth Engine (GEE) platform was used to process Sentinel-2 satellite images via JavaScript [63]. This cloud-based solution efficiently handles large datasets, making it ideal for analyzing extensive forest areas. The study area included all relevant clusters used for training and the test area, ensuring that models were trained using representative remote sensing data.

Sentinel-2 Level-1C products with a 10 m spatial resolution were used, offering high-detail images across multiple spectral bands. Specifically, Bands 2 (Blue), 3 (Green), 4 (Red), 8 (NIR), 11, and 12 (SWIR) were deemed suitable for assessing forest attributes like biomass and canopy cover [64]. The images were acquired from 1 June 2022 to 30 September 2022, precisely matching the field data collection period. This alignment ensures temporal consistency between remote sensing data and ground measurements, minimizing discrepancies due to seasonal variations and enhancing the predictive accuracy of the models.

Essential variables included spectral indices such as NDVI, SAVI, and EVI, as well as slope from SRTM data [65]. These indices were chosen for their effectiveness in quantifying vegetation health and biomass [66,67], and more information about the formula and reason for choosing are detailed in Table A3 in the Appendix A. Cloud masking was performed using the built-in Sentinel-2 cloud probability layer in GEE. This approach effectively removes high-confidence clouds and cloud shadows while preserving vegetation-related spectral information. While alternative methods, such as S2Cloudless, offer advanced cloud detection, they were not applied due to computational constraints and the need for consistency in large-scale processing. Manual inspection confirmed minimal cloud contamination in the selected Sentinel-2 images, ensuring the reliability of the dataset for ML analysis [68].

All spectral bands required for the indices are merged into a single image and reprojected to the specified coordinate system, and water bodies are masked out to improve accuracy. Next, a feature collection is created, which is essential for both training and

validation. In this context, quantitative attributes refer to measurable data extracted from ground-based surveys (e.g., DBH, H) and remotely sensed attributes (e.g., NDVI, slope, aspect). These attributes serve as inputs for ML, where models learn to extract more abstract features that capture complex patterns not directly observable.

The training–validation dataset combines both ground-based measurements and remote sensing-derived attributes. ML models use this data to learn patterns, validate their predictions, and refine their accuracy. Once validated, these models can reliably estimate forest attributes for new data, supporting robust forest analysis.

2.3. Optimizing and Training ML Models for Forest Attribute Mapping in GEE

This study was conducted using GEE, where the entire ML workflow was implemented in JavaScript within the GEE Code Editor. After preparing the dataset by integrating and preprocessing the data, we applied ML algorithms—CART, RF, and GBTA—to model forest attributes. The models were trained and validated within GEE, utilizing the respective built-in classifier functions: RF with ee.Classifier.smileRandomForest(), GBTA with ee.Classifier.smileGradientTreeBoost(), and CART with ee.Classifier.smileCart().

To optimize these models, we followed a two-step tuning approach: initially, we fixed the number of trees at 100 to maintain a balance between computational efficiency and model stability. This helped isolate the effects of other hyperparameters without interference from tree count variations, ensuring a systematic tuning process. Since GEE does not support automated hyperparameter tuning, we performed a manual grid search, selecting hyperparameter ranges based on established guidelines from the GEE API documentation and previous research findings. Once optimal hyperparameters were selected, we varied the number of trees from 100 to 2500 for RF and from 100 to 3500 for GBTA, based on empirical studies [69,70] suggesting that these ranges provide an optimal trade-off between model complexity and computational efficiency, reapplying the tuning process at each step to ensure consistency. The number of trees serves as a key indicator of model complexity, influencing the trade-off between bias and variance [69]. Increasing the number of trees enhances the model's ability to capture patterns in the data, thereby reducing variance and improving predictive performance [69]. However, excessive tree count increases computational cost and may lead to diminishing returns in model performance. By systematically adjusting the number of trees, we aimed to find an optimal balance between performance and computational efficiency [49]. Hyperparameters in Google Earth Engine were manually tuned due to the platform's lack of automated options. For GBTA, we tested learning rates (0.01, 0.05, 0.1, 0.2), max depths (3, 5, 7), and min split losses (0, 0.01, 0.1); for RF, we varied variables per split (auto, sqrt, log2) and min leaf populations (1, 5, 10, 20); for CART, we optimized max depths (5, 10, 15) and min samples split (2, 5, 10). A grid search combined with cross-validation ensured robust model evaluation, balancing computational efficiency and performance [71]. Although methods like Bayesian optimization could improve tuning efficiency, they were not feasible within GEE's API constraints [72].

2.4. Performance Assessment

After model development, we evaluated performance on the validation dataset using R^2, RMSE, rRMSE, and MAE. R^2 indicates goodness-of-fit; RMSE emphasizes larger errors; rRMSE normalizes errors relative to observed values, and MAE provides an outlier-resistant measure. Together, these metrics offer a robust and nuanced assessment of model accuracy. To statistically assess significant differences in performance metrics across various algorithms and complexity levels, we employed non-parametric testing methods using the Scheirer–Ray–Hare rank-based test, an extension of the Kruskal–Wallis H test [73]. Given that goodness-of-fit estimators such as R^2 and RMSE are derived values rather than directly

observed measurements, this approach ensures a more robust statistical validation [74]. We analyzed factors such as algorithm type and model complexity, measured by the number of trees in RF and GBT models, which ranged from 100 to 2500 for RF and 100 to 3500 for GBTA.

Models with the highest R^2 and lowest RMSE, rRMSE, and MAE were considered to have the best performance. This multi-metric approach ensures that models not only fit the data well but also make precise predictions with minimal errors. The Scheirer–Ray–Hare rank-based test results further supported the selection of models with optimal complexity, particularly identifying the number of trees that yielded the most accurate predictions for RF and GBT models.

All analyses were conducted using R version 4.3.2. This evaluation process, combining multiple error metrics and statistical testing, provided a robust framework for selecting the most accurate and reliable models for predicting forest attributes in the study area.

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

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

$$\text{rRMSE} = \frac{RMSE}{\overline{Y}} \times 100\% \quad (3)$$

$$\text{MAE} = \frac{1}{n}\sum_{i=0}^{n}|y_i - \hat{y}_i| \quad (4)$$

2.5. Forest Attribute Mapping and Analysis

The trained ML models were applied to the satellite imagery of the study area to predict forest attributes at a larger scale. This process involved feeding the satellite imagery into the models and obtaining the predicted values for each pixel containing the selected features. For each attribute, maps were developed at the resolutions (10, 50 and 100 m).

To understand how different algorithms predict forest attributes across different resolutions, a detailed analysis was carried out. These maps were compared with the independent dataset that was already prepared for the accuracy assessment (Figure 2). The independent dataset stored as individual tree measurement was assigned geographically on the pixel with the specific resolution of the map (10, 50, or 100 m) and extrapolated as stand characteristics at hectare. Then, R^2, RMSE, MAE, and rRMSE were calculated to quantify the agreement between the predicted values and the validation data for each pixel size. This analysis will determine the optimal resolution that enables the capture of fine-scale forest attributes with the highest level of accuracy.

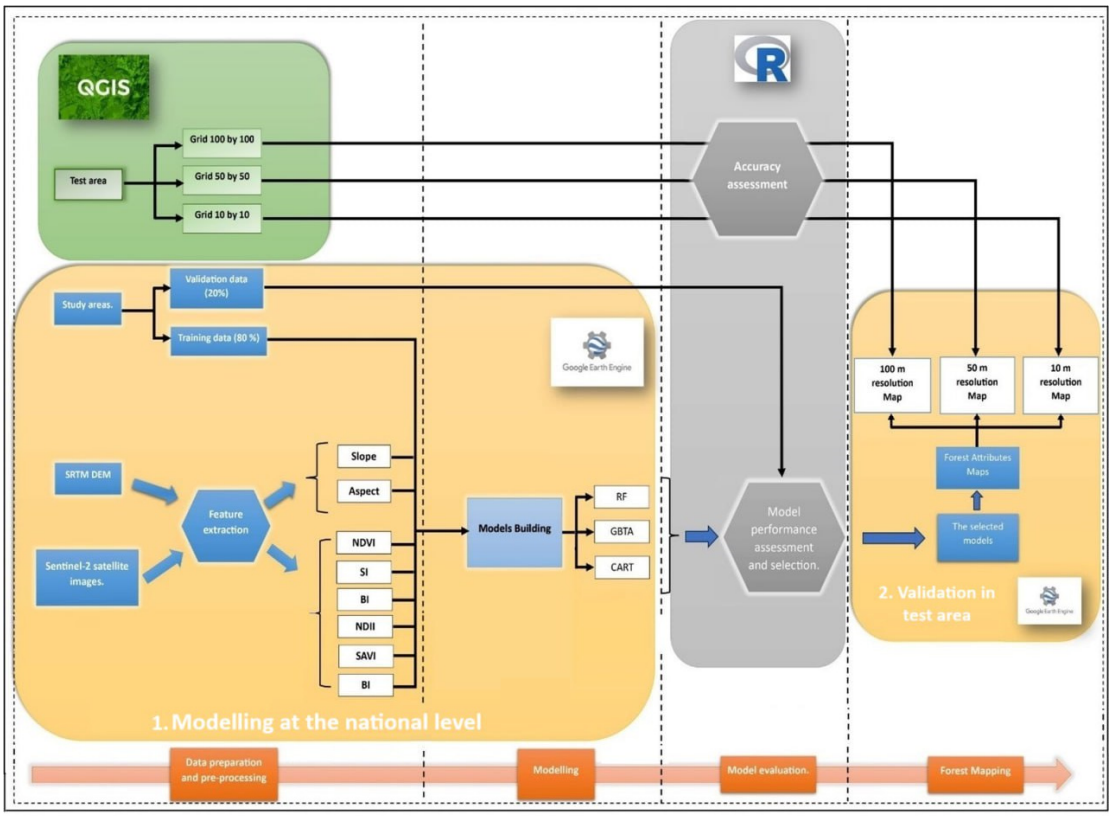

Figure 2. Workflow diagram for data used in research.

2.6. Evaluation of Model-Predicted Volume Accuracy Against NFI Data

To evaluate the accuracy of model-predicted volume estimates, remote sensing-derived raster datasets were processed and compared against NFI data. The NFI data are systematically structured to provide comprehensive forest resource assessments. It consists of a network of systematically distributed sample plots where tree attributes such as DBH, tree height, species, and volume are recorded. It operates on a 5-year cycle with two phases: aerial photo-interpretation using high-resolution orthophotos and field measurements within permanent sample plots [7]. The plots are typically categorized based on predefined DBH classes (<100 mm, 100–199 mm, 200–299 mm, 300–399 mm, 400–499 mm, ≥500 mm), allowing for standardized volume estimation across different forest structures.

The DBH raster was classified into the same predefined DBH classes used in NFI to ensure comparability. The mean volume per hectare () was derived for each DBH class by averaging the total volume within that class, weighted by the number of sample points or pixels within the study area. The computation followed the formula:

$$\overline{V} = \frac{\sum_{i=1}^{n} V_i}{n},$$

where

- $\overline{V}$ is the mean volume per hectare for a given DBH class;
- V_i represents the volume in each sample unit (pixel or plot);
- n is the total number of sample units in the class.

3. Results

3.1. Forest Characteristics Modeling Results at the National Level

The evaluation of multiple ML algorithms for predicting forest stand attributes, summarized in Table 1, highlights significant differences in their performance. Comprehensive

wall-to-wall maps were generated for models with the highest prediction accuracy (Figure 3).

GBTA showed strong predictive capabilities, with R^2 values increasing from 0.66 (100 trees) to 0.92 (2500 trees) for stand average Vol estimation, indicating as significant improved performance with increased model complexity a result further validated by the Scheirer–Ray–Hare rank-based test (see Table A4). Similar trends were observed for other attributes (Figure 4). GBTA predicted higher volumes in regions similar to RF but with notable differences, particularly in the central and northeastern areas (Figure 3).

RF, on the other hand, performed well, explaining over 80% of the variance in the target variables. This regression line performed well despite being slightly less steep than GBTA with 2500 trees, as shown in Figure 5. Consequently, the R^2 value was 94.09%, and the RMSE was approximately 3.56 m^2/ha. In Figure 3, the model predicts higher forest Vol, BA, DBH, and H predominantly in the central and northwestern regions of Romania, with lower values generally toward the borders.

In BA estimations, CART outperformed all other predictors with its high R^2 value of 94.09% and low RMSE of 3.56. Nevertheless, its accuracy did not extend uniformly across all forest attributes, as it failed to consistently outperform RF and GBTA (see Figure 6). Thus, while it excelled in explaining the variance in BA predictions, its performance exhibited variability across different scenarios. Visually, CART presents a more homogeneous distribution with fewer areas of high BA concentration compared to other models.

Table 1. Performance of ML algorithms in the validation datasets attributes and scenarios.

Attribute	Model	N^0 of Trees	RMSE	MAE	rRMSE (%)	R^2
Vol (m^3/ha)	RF	100	107.461	−2.420	0.284	0.810
		1000	107.727	−1.217	0.285	0.810
		3500	107.643	−1.013	0.285	0.810
	GBTA	100	178.085	17.077	0.471	0.660
		1000	89.769	2.105	0.237	0.860
		2500	64.507	1.499	0.171	0.930
	CART	1000	132.663	0.000	0.351	0.630
BA (m^2/ha)	RF	100	6.617	−0.020	0.223	0.830
		1000	6.594	−0.047	0.222	0.830
		3500	6.585	−0.046	0.222	0.830
	GBTA	100	11.520	1.039	0.388	0.710
		1000	5.286	−0.177	0.178	0.889
		2500	3.809	−0.113	0.128	0.940
	CART	1000	3.564	0.000	0.120	0.941
DBH (cm)	RF	100	6.402	−0.005	0.205	0.816
		1000	6.311	−0.029	0.202	0.826
		3500	6.294	−0.034	0.201	0.827
	GBTA	100	10.768	0.923	0.345	0.708
		1000	5.154	−0.061	0.165	0.874
		2500	3.616	3.616	−0.044	0.936
	CART	1000	7.904	0.000	0.253	0.653
H (m)	RF	100	3.207	−0.022	0.135	0.839
		1000	3.198	−0.009	0.134	0.845
		3500	3.192	−0.014	0.134	0.845
	GBTA	100	2.589	−0.290	0.109	0.750
		1000	2.589	−0.290	0.109	0.891
		2500	1.868	−0.199	0.078	0.941
	CART	1000	4.234	0.000	0.178	0.665

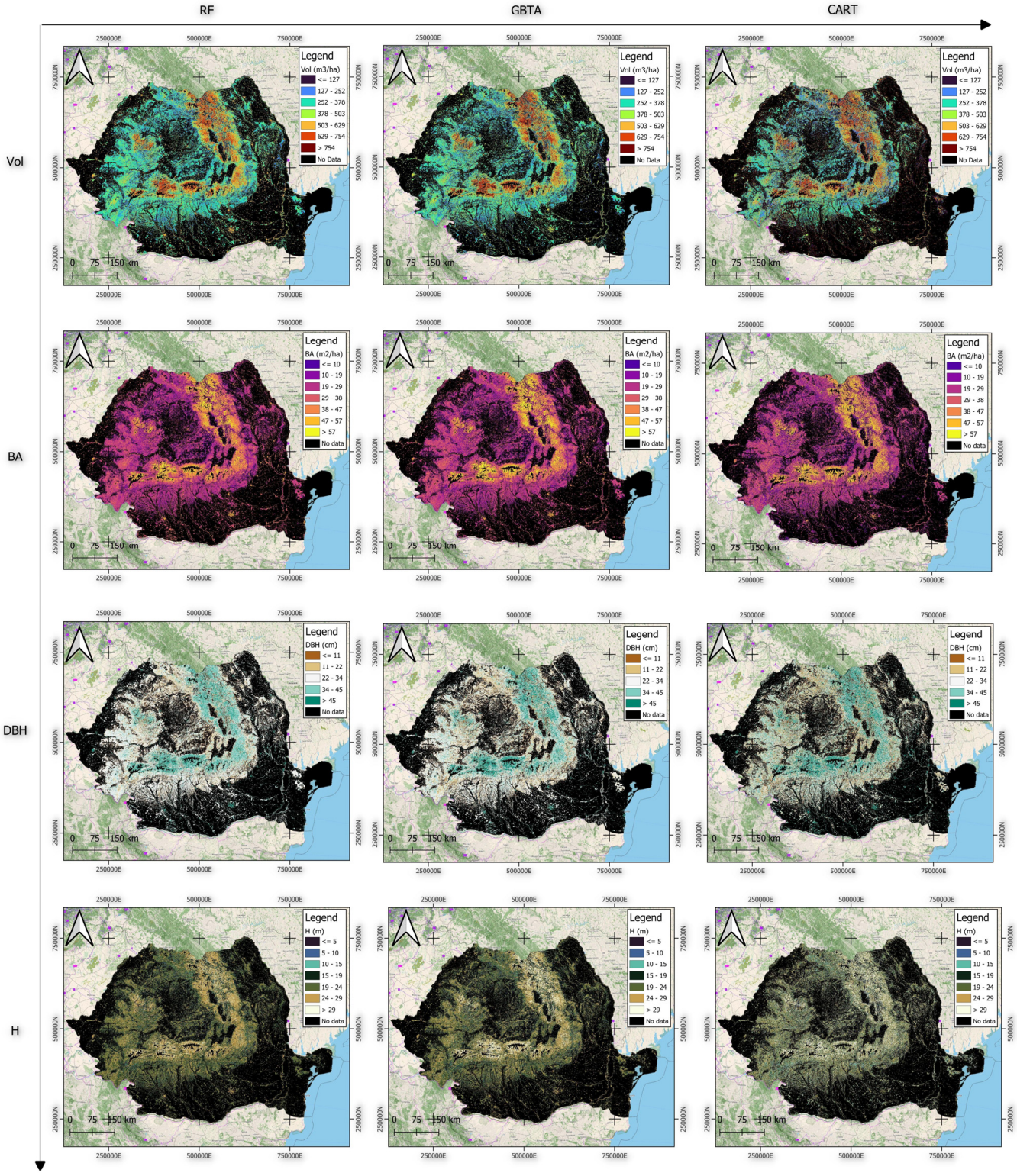

Figure 3. Forest characteristics modeling results at national level.

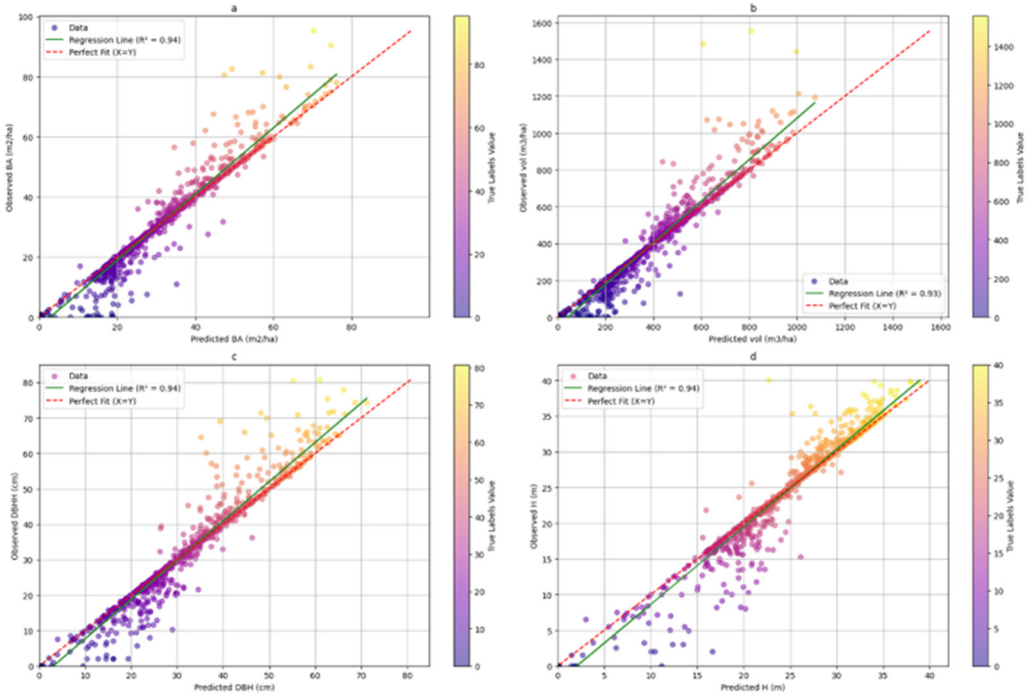

Figure 4. Scatterplots for the GBTA regression model of forest stand attributes: (**a**) BA; (**b**) Vol; (**c**) DBH; (**d**) H.

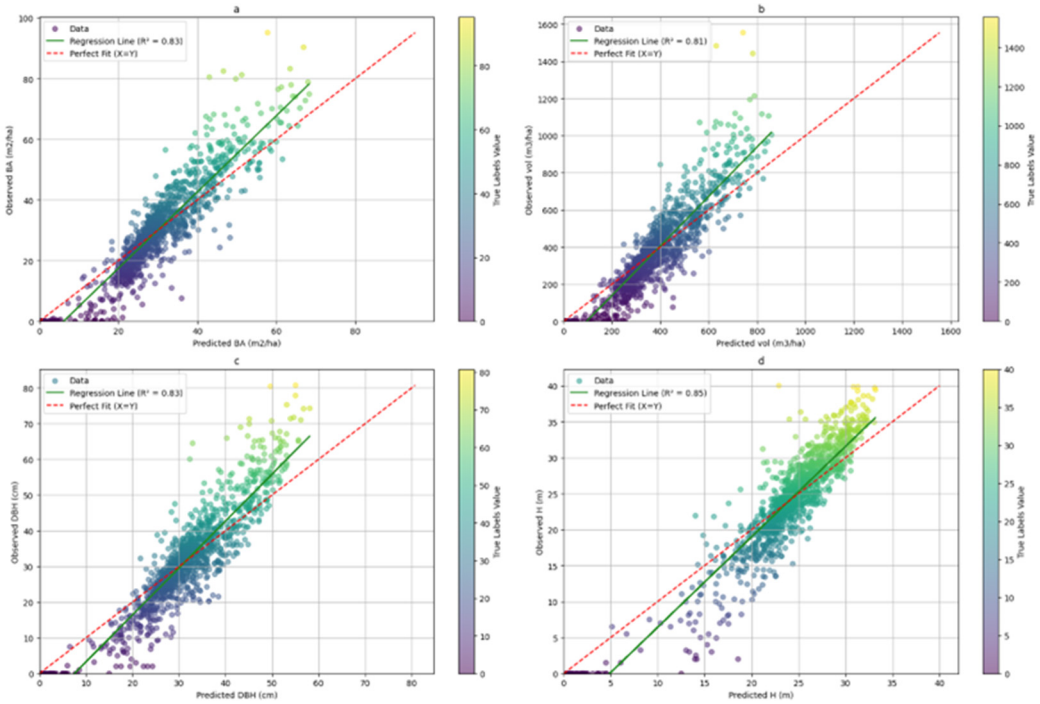

Figure 5. Scatterplots for the RF regression model of forest stand attributes: (**a**) BA; (**b**) Vol; (**c**) DBH; (**d**) H.

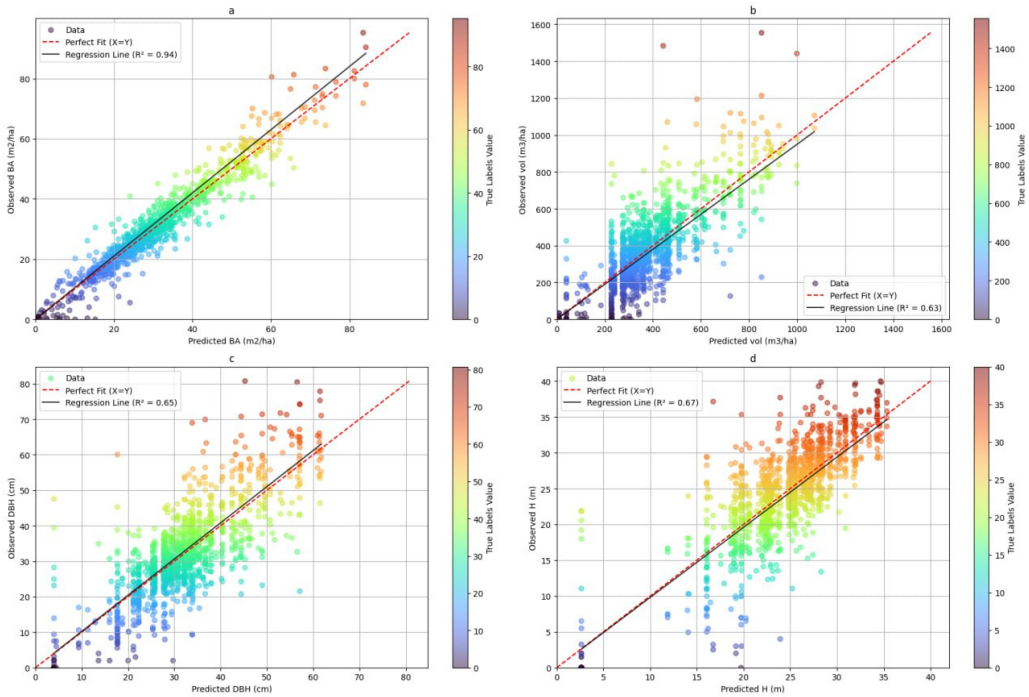

Figure 6. Scatterplots for the CART regression model of forest attributes: (**a**) BA; (**b**) Vol; (**c**) DBH; (**d**) H.

3.2. Model Performance Assessment with Independent Validation Data in the Test Area

The assessment of prediction accuracy and reliability through comparison with independent ground truth data in the Lempes test area indicates that RF and GBTA algorithms with a maximum number of suggested trees demonstrated high performance in predicting various forest attributes during the evaluation with initial validation data. The CART algorithm exhibited reliable outcomes, specifically in predicting BA. Based on these findings, these three models were selected for further analysis. This selection process was crucial to ensure the robustness and generalizability of the chosen models.

3.2.1. Visual Assessment of the Models in Mapping Forest Characteristics

The detailed maps presented in Figures 7–10 illustrate the predictive modeling of forest attributes across the entire test area. These intricate cartographic representations provide a granular perspective on model predictions at a spatial resolution of 10 m. The three models show varying degrees of accuracy in their estimations compared to observed data.

GBTA consistently provides predictions that closely align with the actual data, particularly excelling in capturing the upper range of values such as taller tree heights, larger DBH, higher volumes, and BA. RF generally achieves accuracy comparable to GBTA but tends to produce more moderate predictions. It often smooths out extremes and may slightly underpredict in areas with exceptionally large trees. CART, on the other hand, tends to underpredict, especially in regions with higher observed values, showing a more conservative approach, particularly in predicting larger trees in terms of H, DBH, and Vol. Overall, GBTA is the most accurate visually for environments with high variability; RF offers reliable but more moderate predictions, and CART is best suited for scenarios where lower, conservative estimates are preferable.

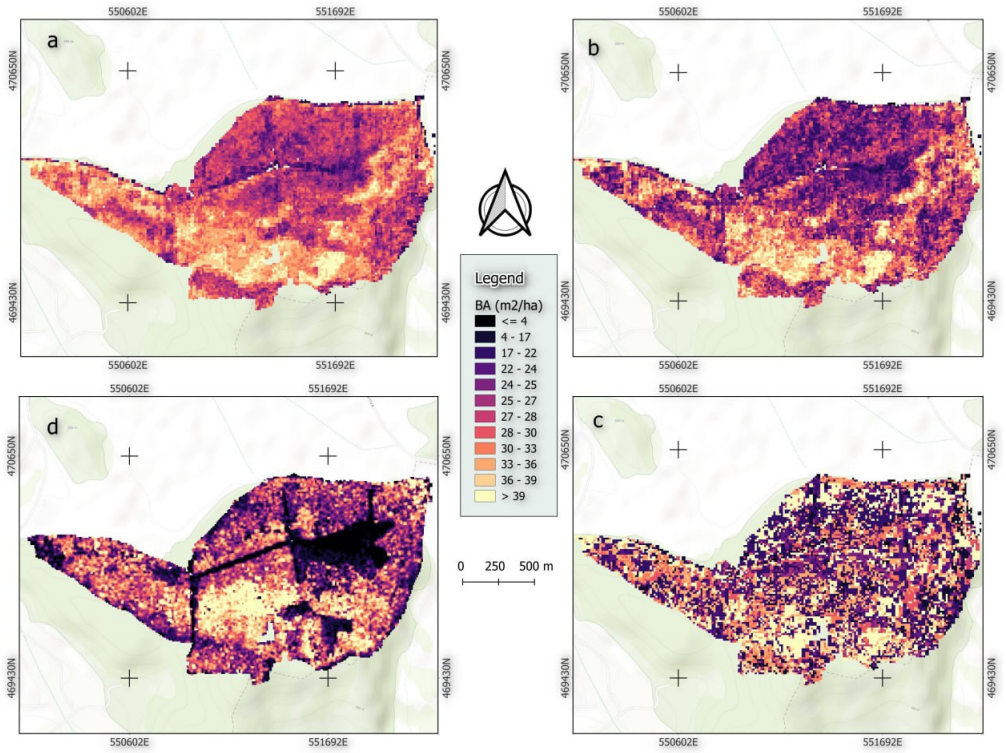

Figure 7. Forest Characteristic Mapping for the BA using (**a**) RF, (**b**) GBTA, (**c**) CART, (**d**) Field measurements.

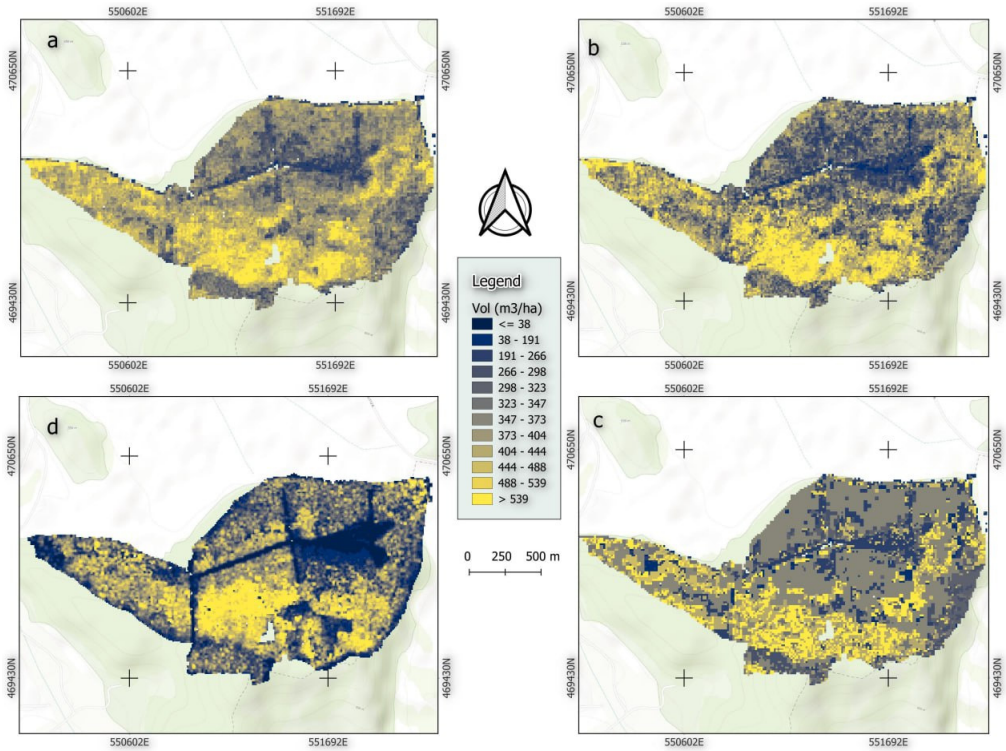

Figure 8. Forest Characteristic Mapping for the Vol using (**a**) RF, (**b**) GBTA, (**c**) CART, (**d**) Field measurements.

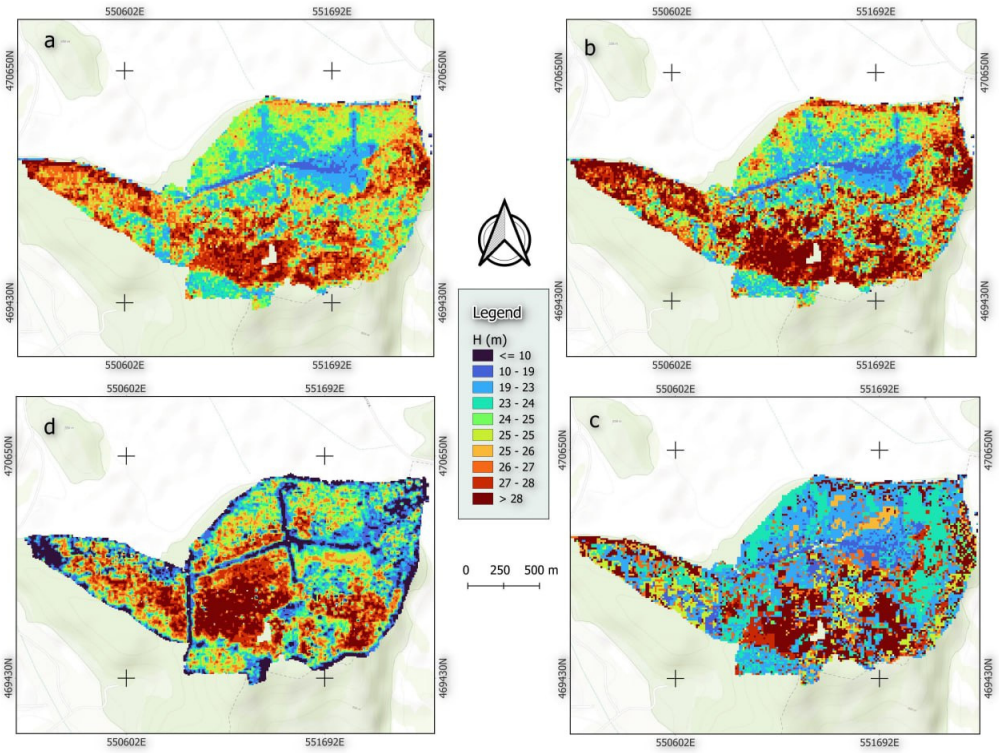

Figure 9. Forest Characteristic Mapping for the H using (**a**) RF, (**b**) GBTA, (**c**) CART, (**d**) Field measurements.

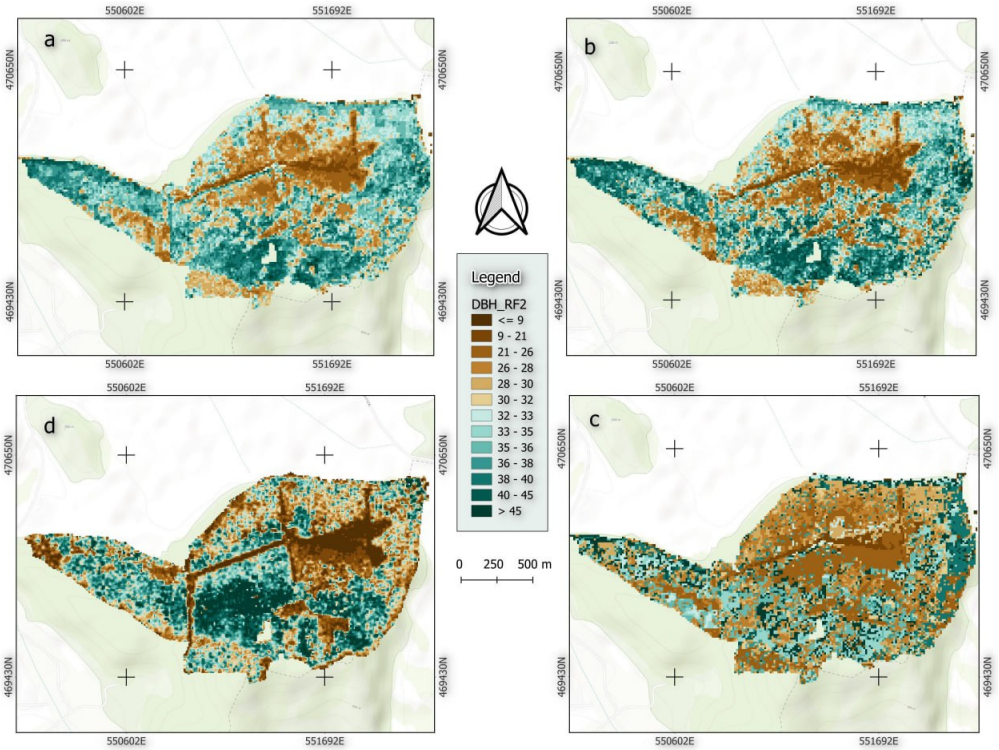

Figure 10. Forest Characteristic Mapping for the DBH using (**a**) RF, (**b**) GBTA, (**c**) CART, (**d**) Field measurements.

3.2.2. Evaluation of Predictive Accuracy in Forest Attribute Estimation Across Different Resolutions

The accuracy of these algorithms in predicting the independent dataset when varying the pixel size from low to high is presented in Table 2. This comprehensive assessment

allowed for an optimal determination of the balance between detail and precision for optimal prediction.

Table 2. Performance of forest attributes estimation models with independent datasets at different resolutions.

Attribute	Algorithm	Resolution	R^2	RMSE	rRMSE (%)	MAE
DBH (cm)	RF	10	0.285	9.200	0.288	7.921
		50	0.297	6.935	0.212	6.241
		100	0.578	5.248	0.162	4.483
	GBTA	10	0.278	9.218	0.293	7.885
		50	0.312	6.037	0.186	7.377
		100	0.596	4.219	0.138	4.326
	CART	10	0.244	9.179	0.306	7.754
		50	0.220	8.974	0.310	4.752
		100	0.577	4.982	0.155	3.498
H	RF	10	0.207	6.062	0.245	5.254
		50	0.419	3.359	0.135	2.910
		100	0.504	4.300	0.173	3.865
	GBTA	10	0.201	6.091	0.242	5.418
		50	0.466	3.299	0.131	3.028
		100	0.555	4.155	0.165	4.103
	CART	10	0.176	6.270	0.260	5.192
		50	0.349	3.484	0.135	2.837
		100	0.417	4.507	0.175	3.723
Vol	RF	10	0.234	120.943	0.297	102.554
		50	0.215	59.666	0.149	49.566
		100	0.286	66.809	0.1278	56.783
	GBTA	10	0.222	134.598	0.344	100.610
		50	0.367	87.015	0.229	39.646
		100	0.388	64.431	0.1531	44.834
	CART	10	0.217	120.067	0.294	109.795
		50	0.061	50.781	0.148	65.647
		100	0.360	59.374	0.1405	53.118
BA	RF	10	0.286	6.592	0.217	5.433
		50	0.343	6.203	0.202	5.424
		100	0.194	7.897	0.260	7.046
	GBTA	10	0.281	9.285	0.323	5.723
		50	0.358	6.626	0.227	6.600
		100	0.153	8.171	0.283	8.587
	CART	10	0.351	6.993	0.228	7.484
		50	0.256	7.392	0.238	5.614
		100	0.102	9.451	0.309	7.430

At the 10 m resolution, the accuracy of the models is generally low. This finer resolution tends to introduce more variability and noise into the data, which the models struggle to handle effectively. Consequently, the coefficient of determination values are lower, suggesting that the models account for less variance in the data. Furthermore, the error metrics are higher, indicating reduced predictive performance.

Figures 11–14 show the scatterplots for the model's performance within the test area. When examining DBH and H, both RF and GBTA algorithms demonstrate significant improvement in R^2 at 100 m (about 55% of the dataset was accurately predicted) with the lowest RMSE and rRMSE values, suggesting that higher pixel sizes lead to better predictive accuracy for these attributes. The CART algorithm follows a similar trend, with the best performance observed at 100 m for both DBH and Height at 41.7%. This proves that for

predicting DBH and Height, higher resolutions (100 m) generally provide more accurate and precise results across all three algorithms, particularly for RF and GBTA.

For Vol, the RF and CART algorithms show slight improvements in R^2 at 100 m, but this comes with significantly higher RMSE and rRMSE. In contrast, the GBTA algorithm demonstrates superior performance, with notable improvements in accuracy and decreasing errors at both 50 m and 100 m resolutions. For BA, the optimal accuracy and precision were at 50 m resolution for RF and GBTA algorithms, with a significant improvement in error metrics. The CART algorithm performs best at 10 m but shows a decline in accuracy at higher resolutions. This indicates that the models struggle particularly with predicting Vol and BA effectively, even when considering larger resolutions.

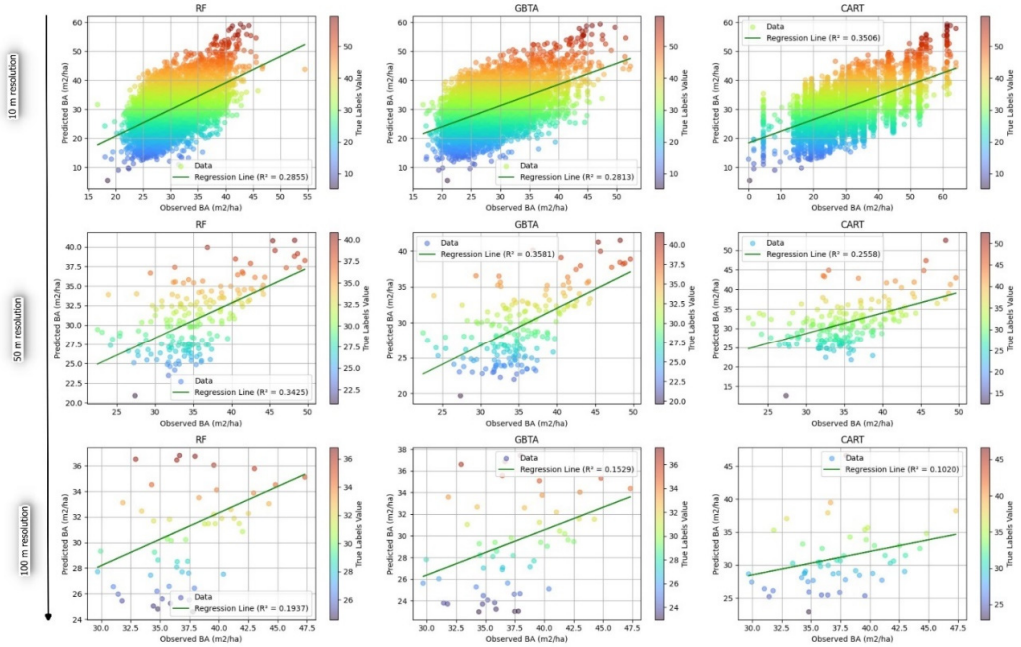

Figure 11. Scatterplots for the model's performance with the test area for BA prediction under different resolutions.

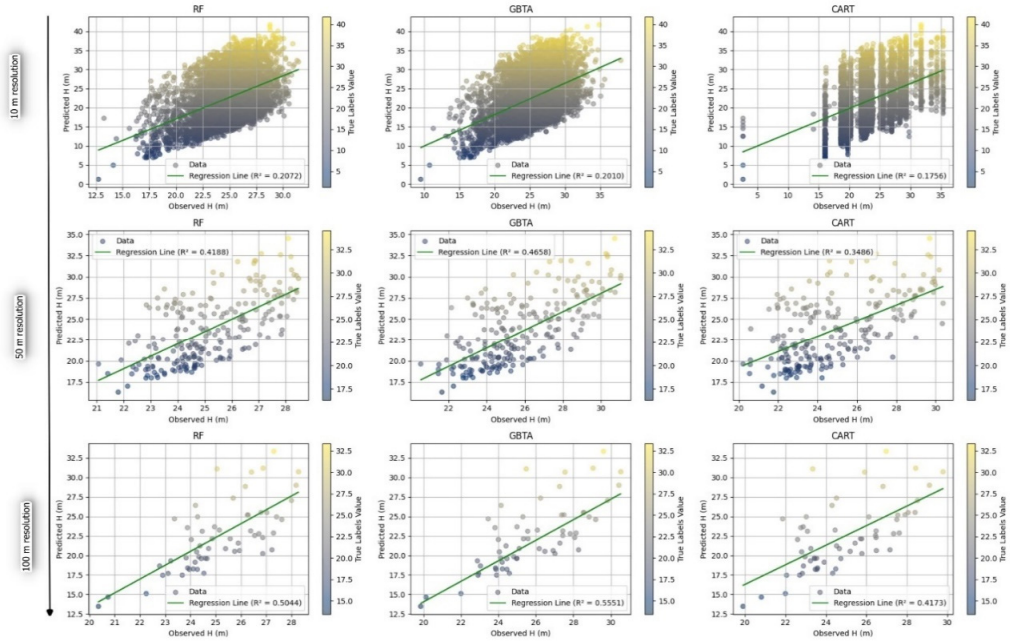

Figure 12. Scatterplots for the model's performance with the test area for H prediction under different resolutions.

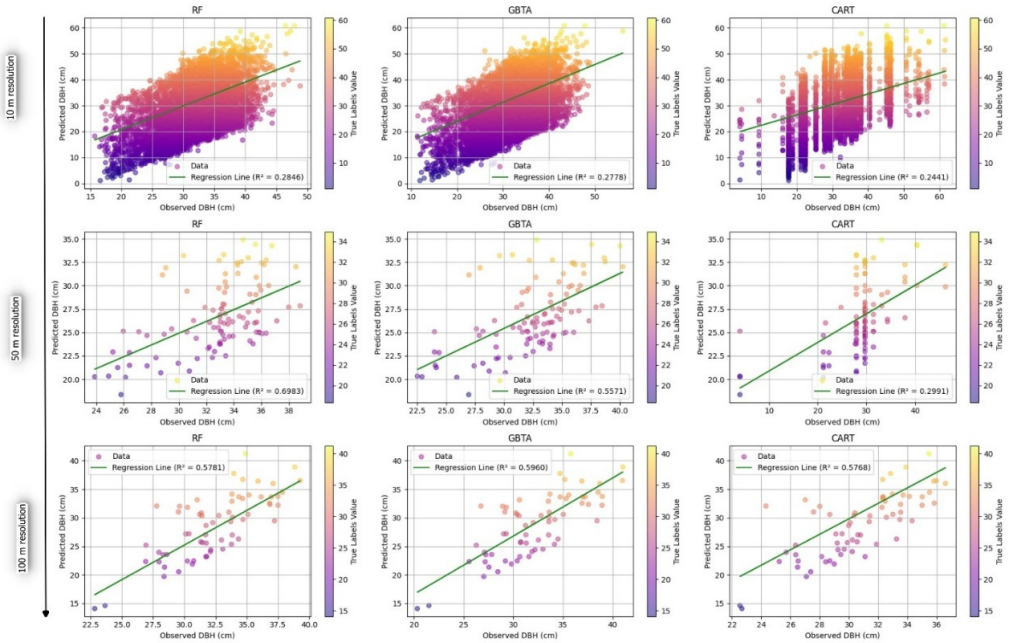

Figure 13. Scatterplots for the model's performance with the test area for DBH prediction under different resolutions.

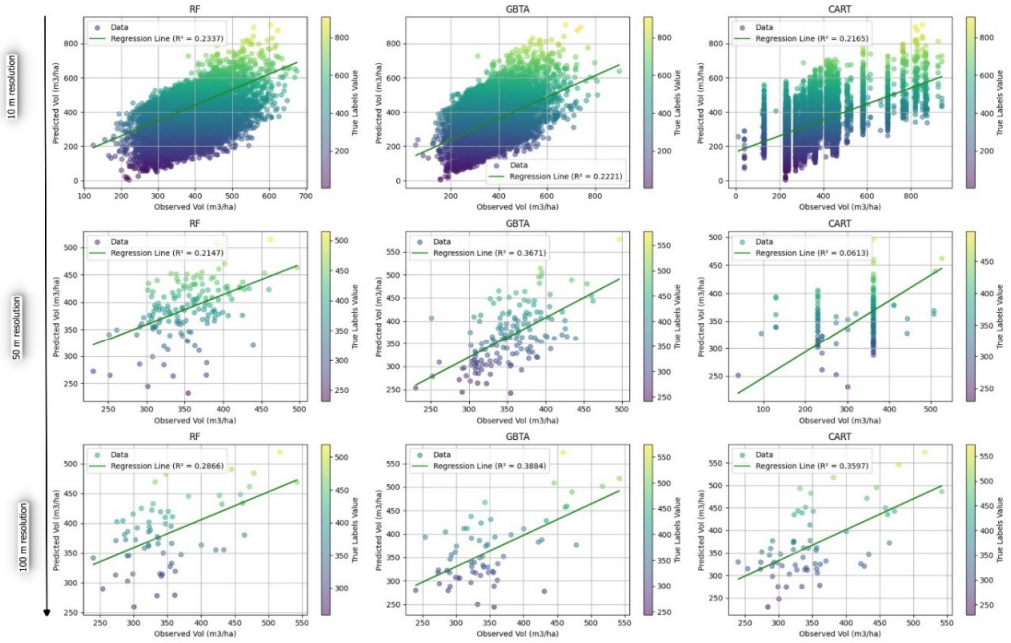

Figure 14. Scatterplots for the model's performance with the test area for Vol prediction under different resolutions.

3.2.3. Results of Evaluation of Model-Predicted Volume Accuracy Against NFI Data

The comparison of model-predicted volume against NFI data (Figure 15) highlights GBTA as the most accurate model (R^2 = 0.915), effectively capturing volume variations due to its iterative error correction. RF also performed well (R^2 = 0.842) but showed slight underestimation at higher values, likely due to its averaging nature. In contrast, CART had the lowest accuracy (R^2 = 0.656), struggling with continuous variability due to its hierarchical partitioning.

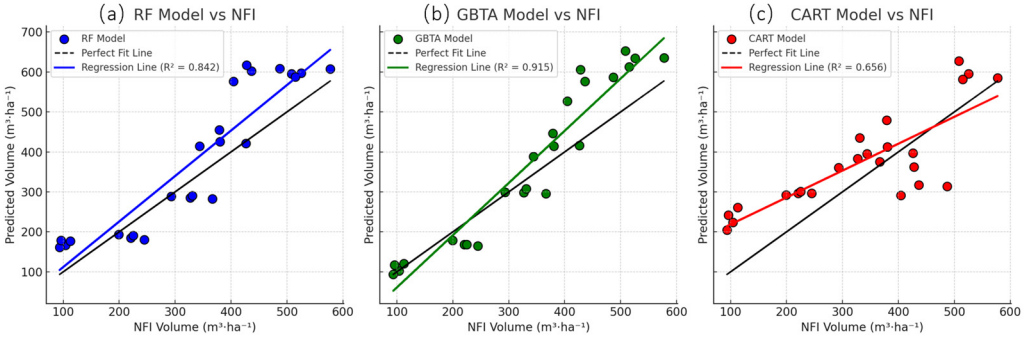

Figure 15. Comparison between NFI volume measurements at national and regional levels against predicted models, using (**a**) RF, (**b**) GBTA, (**c**) CART.

These results validate that the independent field data used in modeling was enough to produce a model comparable with the structured sampling approach of NFI. Moreover, the evaluation emphasizes GBTA's suitability for complex forest volume predictions, while RF remains a reliable alternative. The findings underscore the importance of selecting models based on attribute complexity to enhance large-scale forest inventory accuracy.

4. Discussion

Romania's forests, recognized for their exceptional diversity and critical contributions to carbon sequestration, biodiversity conservation, and sustainable resource management [1,64], face challenges to be monitored efficiently at a national scale. Traditional ground-based inventories, although highly accurate, are resource-intensive and insufficient for large-scale, frequent assessments [7,65]. Against this backdrop, our study aimed to model and map key forest attributes, including Vol, BA, DBH, and H, by integrating Sentinel-2 remote sensing data with advanced ML techniques. This approach was designed not only to overcome the limitations of conventional methods but also to generate comprehensive, high-resolution maps that support sustainable forest management and policy-making [4,17].

Our research focused on comparing the performance of three prominent ML algorithms, RF, GBTA, and CART, in predicting forest attributes across Romania's diverse temperate ecosystems. Each algorithm has unique strengths and limitations. GBTA, for instance, consistently delivered high accuracy, especially for complex and variable attributes like standing volume and DBH. When optimized to a complexity level of 2500 trees, GBTA achieved an R^2 of 0.93 for volume estimation, outperforming the other models in both accuracy and error metrics [38,75]. This superior performance can be attributed to GBTA's iterative boosting mechanism, which refines predictions by focusing on previously mispredicted samples and thereby effectively capturing nonlinear relationships inherent in forest structure data [76].

RF also demonstrated robust predictive capabilities, explaining over 80% of the variance for most forest attributes. The ensemble approach of RF, which combines the predictions of multiple decision trees through bagging [34], provides stability and resilience against overfitting. However, our results indicated that RF tended to underpredict extreme values, a tendency that has been noted in previous studies [31,32]. This limitation is likely due to RF's inherent smoothing effect, which, while beneficial in reducing variance, may obscure important variations in areas with highly heterogeneous forest conditions. In contrast, CART, known for its simplicity and interpretability, excelled in predicting BA by achieving the lowest RMSE values among the models. Nonetheless, CART's performance was less robust for attributes like DBH and H, where its tendency to partition data into discrete categories led to conservative and sometimes biased estimates [33,35]. This behavior aligns

with earlier observations that CART models often require post-processing adjustments to capture the continuous nature of complex environmental variables accurately [77,78]. In this regard, the limitations of CART suggest that its inherent structure may not be well-suited for capturing continuous variations in all forest attributes; alternative improvement strategies or model adjustments might be needed to improve its performance for variables such as DBH and H [79].

A critical objective of our study was to evaluate the influence of model complexity on predictive performance. Our findings clearly indicate that increasing the number of trees in GBTA markedly improves its accuracy. The boosting mechanism inherent to GBTA allows the model to iteratively correct its errors, thus progressively reducing prediction discrepancies [75]. Conversely, our experiments with RF revealed that beyond a certain threshold, adding more trees did not significantly enhance performance. This observation is consistent with previous research, which suggests that while RF is robust against overfitting, its performance tends to plateau once an optimal number of trees is reached [34,80]. Although our evaluation of complexity was primarily based on the number of trees, it is important to note that other factors—such as maximum tree depth, the number of features considered per split, and overall computational cost—also contribute to model complexity. Future studies may benefit from incorporating these additional indicators to provide a more nuanced understanding of the trade-offs between model complexity and predictive performance. The manual grid search for hyperparameter tuning, although time-consuming due to GEE limitations [71,72], proved effective in optimizing our models, ensuring a robust balance between computational efficiency and predictive accuracy.

Spatial resolution emerged as another significant factor influencing model performance. Our sensitivity analysis on a unique, fully inventoried forest of 168 hectares in the Lempes area revealed that finer resolutions (e.g., 10 m) introduce greater variability and noise, which can overwhelm ML models and reduce their predictive accuracy for attributes such as DBH and H [81,82]. In contrast, coarser resolutions (e.g., 100 m) tend to average out noise and capture broader trends, leading to improved R^2 values and lower error metrics for these attributes. However, this benefit comes with the trade-off of losing fine-scale detail, which may be critical for certain management applications. Notably, for attributes like standing volume and basal area, our models appeared less sensitive to changes in spatial resolution, likely because these metrics depend more on aggregated forest structure rather than fine-scale variability [83].

Another challenge faced during this study was the inherent variability in training data and topographic complexity across Romania's diverse forest ecosystems. The heterogeneity in species composition, management practices, and terrain creates complex patterns that can be difficult for ML models to generalize. Our extensive field campaigns, which collected detailed measurements from over 1300 circular plots, helped mitigate these issues by providing a representative sample of Romania's forests [57,58]. Despite these efforts, variations in local conditions still introduced some degree of uncertainty in model predictions. For example, while RF models showed relatively stable performance in predicting H, their error metrics increased in areas with steep slopes or mixed species stands, a phenomenon also observed in studies [23,84]. Moreover, uncertainties arising from sensor errors, environmental heterogeneity, and inherent limitations in remote sensing data interpretation further contribute to the overall prediction uncertainty [85]. These aspects should be taken into account when generalizing the model outcomes to areas with conditions that may fall outside the training distribution.

The broader applicability of our findings extends well beyond Romania. Many temperate forest ecosystems around the world share similar challenges in terms of monitoring and management. The methodologies developed in this study, particularly the integration

of Sentinel-2 data with advanced ML algorithms, offer a scalable framework that can be adapted to other regions facing similar environmental and logistical constraints [86,87]. The ability to generate accurate, large-scale maps of forest attributes is invaluable for supporting decision-making processes related to timber harvesting, carbon accounting, and biodiversity conservation. Moreover, the cost-effectiveness and efficiency of our approach reduce reliance on resource-intensive ground surveys, thereby facilitating more frequent and comprehensive forest monitoring [18,19].

In conclusion, this study underscores the transformative potential of integrating ML with remote sensing for national-scale forest mapping. By leveraging Sentinel-2 imagery and rigorous ground measurements, we developed a robust framework capable of accurately predicting critical forest attributes such as standing volume, basal area, DBH, and tree height. Among the ML algorithms tested, GBTA and RF emerged as particularly effective, with GBTA demonstrating superior performance for complex attributes when tuned to higher model complexity. CART, while offering excellent interpretability and high performance for basal area, was less effective for capturing continuous variations in DBH and H. Future work should further address the uncertainties and limitations highlighted here by exploring alternative algorithms (e.g., deep learning or ensemble techniques like XGBoost, incorporating additional complexity metrics, and investigating model-specific adjustments to improve performance, particularly for CART in predicting continuous variables.

5. Conclusions

Our findings have significant implications for forest management and policy. The scalable and cost-effective nature of the proposed approach not only enhances the accuracy of forest inventories but also supports more informed decision-making for sustainable resource management. As Romania and other temperate forest regions continue to grapple with the challenges of forest degradation and climate change, the adoption of advanced ML and remote sensing methodologies will be critical in ensuring the long-term health and productivity of these vital ecosystems.

Future work should focus on integrating additional remote sensing data sources, such as LiDAR or hyperspectral imagery, and exploring automated hyperparameter optimization techniques to further refine model performance. Such advancements will likely lead to even more precise and comprehensive forest attribute maps, reinforcing the role of modern technologies in sustainable forest management globally.

Ultimately, this study represents a significant step forward in harnessing the power of ML and remote sensing for forest monitoring. It provides a robust framework that not only addresses the current gaps in national forest inventories but also offers a pathway for applying similar methodologies to other regions, thereby contributing to global efforts in forest conservation and sustainable resource management.

Author Contributions: Conceptualization, M.I.K. and M.D.N.; methodology, M.I.K. and M.D.N.; software, M.I.K.; validation, M.I.K. and A.H.M.; formal analysis, S.A.B. and M.D.N.; investigation, M.I.K.; resources, S.A.B. and M.D.N.; data curation, M.I.K. and M.D.N.; writing—original draft preparation, M.I.K., A.H.M. and M.D.N.; writing—review and editing, S.A.B. and M.D.N.; visualization, M.I.K.; supervision, M.D.N.; project administration, M.D.N.; funding acquisition, S.A.B. and M.D.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Transilvania University of Brasov Ph.D. program.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

Acknowledgments: We are grateful to Forest Design team for supporting the field data collection.

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

Appendix A

Table A1. Descriptive statistics of the national scale dataset used for machine learning.

Stastic	BA (m^2/ha)	DBH (cm)	H (m)	Vol (m^3/ha)
Mean	30.554	32.195	24.564	389.643
Standard Error	0.388	0.347	0.171	5.921
Median	28.296	30.100	25.100	359.864
Mode	26.800	28.000	20.000	375.900
Standard Deviation	13.929	12.446	6.118	212.330
Range	95.180	79.738	38.500	1554.389
Minimum	0.010	1.000	1.500	0.013
Maximum	95.190	80.738	40.000	1554.402
1st quartile	22.100	25.360	21.440	253.381
3rd quartile	36.413	37.590	28.260	480.499
CV %	45.589	38.659	24.908	54.494

Table A2. Descriptive statistics of the independent dataset used to validate the national product in the test area.

Statistic	BA (m^2/ha)	DBH (cm)	H (m)	Vol (m^3/ha)
Mean	43.892	27.903	19.650	343.095
Standard Error	0.287	0.149	0.113	2.743
Median	41.188	27.513	19.197	315.663
Mode	43.800	31.300	22.700	269.850
Standard Deviation	16.791	8.727	6.589	160.335
Range	162.425	58.426	35.733	1723.700
Minimum	0.400	9.514	3.100	1.175
Maximum	162.825	67.940	38.833	1724.875
1st quartile	33.300	21.914	14.800	244.844
3rd quartile	51.180	32.762	24.150	404.219
CV %	38.256	31.277	33.532	46.732

Table A3. Vegetation indices, band used, formula, and description.

Index	Bands Used	Formula	Description, Applications, and Rationale
Normalized Difference Vegetation Index (NDVI)	NIR (B8) Red (B4)	$\mathrm{NDVI} = \frac{(NIR-Red)}{(NIR+Red)}$	Measures vegetation health by comparing NIR reflectance (healthy vegetation) with Red reflectance (chlorophyll absorption). Chosen for its widespread use in assessing vegetation cover and health.
Shadow Index (SI)	Blue (B2) Green (B3) Red (B4)	$\mathrm{SI} = (1-Blue) \times (1-Green) \times (1-Red)$	Custom index to detect shadowed areas in forests using visible bands. Selected to differentiate shadows from water and dark surfaces.
Soil-Adjusted Vegetation Index (SAVI)	NIR (B8) Red (B4)	$\mathrm{SAVI} = \frac{(NIR-Red)}{(NIR+Red+L)} \times (1+L)$	Minimizes soil brightness influence, improving vegetation detection in areas with sparse cover. Useful for agricultural fields and degraded lands.
Enhanced Vegetation Index (EVI)	Blue (B2) Red (B4) NIR (B8)	$\mathrm{EVI} = \frac{(NIR-Red)\times 2.5}{(NIR+6Red-7.5Blue+1)}$	Enhances sensitivity to dense vegetation, reducing soil and atmospheric effects. Effective in monitoring forest canopy health.
Bare Soil Index (BI)	SWIR1 (B11) SWIR2 (B12) NIR (B8)	$\mathrm{BI} = \frac{(SWIR1-SWIR2-NIR)}{(SWIR1+SWIR2+NIR)} + 2.5$	Differentiates bare soil from vegetation, useful in detecting exposed soils and erosion-prone areas. Selected for monitoring land degradation.
Normalized Difference Infrared Index (NDII)	IR (B8) NIR (B8)	$\mathrm{NDII} = \frac{(IR-NIR)}{(IR+NIR)}$	Assesses water content in vegetation. Chosen for its ability to monitor drought stress and moisture levels in forests.

Table A4. Scheirer–Ray–Hare Test Results: Effect of Algorithm and Model Complexity on Performance Metrics.

Factor	RMSE			MAE			rRMSE			R^2		
	DF	H	*p*-Value	DF	H	*p*-Value	DF	H	*p*-Value	DF	H	*p*-Value
Algorithm	2	0.434	0.815	2	4.718	0.095	2	1.301	0.522	2	3.785	0.153
N^0 of trees	3	1.830	0.610	3	1.453	0.691	3	7.492	0.063	3	10.792	0.015
Interaction (N^0 of trees * Algorithm)	6	1.387	0.711	6	3.117	0.379	6	4.533	0.217	6	11.502	0.014

Figure A1. Distribution of plot clusters containing the number of plots used for training.

References

1. Kozak, D. How to Regenerate a Forest. European Investment Bank. 2023. Available online: https://www.eib.org/en/stories/romania-sustainable-forestry (accessed on 30 January 2025).
2. Tomppo, E.; Gschwantner, T.; Lawrence, M.; McRoberts, R.E. (Eds.) *National Forest Inventories*; Springer: Dordrecht, The Netherlands, 2010; ISBN 978-90-481-3232-4.
3. Bratu, I.A.; Câmpu, V.R.; Budău, R.; Stanciu, M.A.; Enescu, C.M. Subsidies for Forest Environment and Climate: A Viable Solution for Forest Conservation in Romania? *Forests* **2024**, *15*, 1533. [CrossRef]
4. Marin, G.; Strimbu, V.C.; Abrudan, I.V.; Strimbu, B.M. Regional Variability of the Romanian Main Tree Species Growth Using National Forest Inventory Increment Cores. *Forests* **2020**, *11*, 409. [CrossRef]
5. Bouriaud, O.; Don, A.; Janssens, I.A.; Marin, G.; Schulze, E.D. Effects of Forest Management on Biomass Stocks in Romanian Beech Forests. *For. Ecosyst.* **2019**, *6*, 19. [CrossRef]
6. Tudoran, G.M.; Cicșa, A.; Boroeanu, M.; Dobre, A.C.; Pascu, I.S. Forest Dynamics after Five Decades of Management in the Romanian Carpathians. *Forests* **2021**, *12*, 783. [CrossRef]
7. Inventarul Forestier National. *What Is IFN?* Inventarul Forestier National: Ilfov, Romania, 2018. Available online: https://roifn.ro/site/en/ (accessed on 2 February 2025).

8. Lausch, A.; Blaschke, T.; Haase, D.; Herzog, F.; Syrbe, R.-U.; Tischendorf, L.; Walz, U. Understanding and Quantifying Landscape Structure—A Review on Relevant Process Characteristics, Data Models and Landscape Metrics. *Ecol. Model.* **2015**, *295*, 31–41. [CrossRef]
9. Nagendra, H.; Rocchini, D.; Ghate, R. Beyond Parks as Monoliths: Spatially Differentiating Park-People Relationships in the Tadoba Andhari Tiger Reserve in India. *Biol. Conserv.* **2010**, *143*, 2900–2908. [CrossRef]
10. Chave, J.; Réjou-Méchain, M.; Búrquez, A.; Chidumayo, E.; Colgan, M.S.; Delitti, W.B.C.; Duque, A.; Eid, T.; Fearnside, P.M.; Goodman, R.C.; et al. Improved Allometric Models to Estimate the Aboveground Biomass of Tropical Trees. *Glob. Change Biol.* **2014**, *20*, 3177–3190. [CrossRef]
11. García, M.; Riaño, D.; Chuvieco, E.; Danson, F.M. Estimating Biomass Carbon Stocks for a Mediterranean Forest in Central Spain Using LiDAR Height and Intensity Data. *Remote Sens. Environ.* **2010**, *114*, 816–830. [CrossRef]
12. Florea, S.C.; Dutca, I.; Nita, M.D. Tradeoffs and Limitations in Determining Tree Characteristics Using 3D Pointclouds from Terrestrial Laser Scanning: A Comparison of Reconstruction Algorithms on European Bech (*Fagus sylvatica* L.) Trees. *Ann. For. Res.* **2024**, *67*, 185–199. [CrossRef]
13. Buendia, C.; Tanabe, E.; Kranjc, K.; Baasansuren, A.; Fukuda, J.; Ngarize, M.; Osako, S.; Pyrozhenko, A. *Efinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories Task Force on National Greenhouse Gas Inventories*; IPCC: Geneva, Switzerland, 2019; ISBN 978-4-88788-232-4.
14. Rahlf, J.; Hauglin, M.; Astrup, R.; Breidenbach, J. Timber Volume Estimation Based on Airborne Laser Scanning—Comparing the Use of National Forest Inventory and Forest Management Inventory Data. *Ann. For. Sci.* **2021**, *78*, 49. [CrossRef]
15. Olofsson, P.; Kuemmerle, T.; Griffiths, P.; Knorn, J.; Baccini, A.; Gancz, V.; Blujdea, V.; Houghton, R.A.; Abrudan, I.V.; Woodcock, C.E. Carbon Implications of Forest Restitution in Post-Socialist Romania. *Environ. Res. Lett.* **2011**, *6*, 45202. [CrossRef]
16. Munteanu, C.; Kuemmerle, T.; Boltiziar, M.; Butsic, V.; Gimmi, U.; Halada, L.; Kaim, D.; Király, G.; Konkoly-Gyuró, É.; Kozak, J.; et al. Forest and Agricultural Land Change in the Carpathian Region—A Meta-Analysis of Long-Term Patterns and Drivers of Change. *Land Use Policy* **2014**, *38*, 685–697. [CrossRef]
17. Prăvălie, R.; Niculiță, M.; Roșca, B.; Marin, G.; Dumitrașcu, M.; Patriche, C.; Birsan, M.-V.; Nita, I.-A.; Tișcovschi, A.; Sîrodoev, I.; et al. Machine Learning-Based Prediction and Assessment of Recent Dynamics of Forest Net Primary Productivity in Romania. *J. Environ. Manag.* **2023**, *334*, 117513. [CrossRef]
18. Dogaru, L. Timber Tracking Information System, a Necessity in Forest Protection and Sustainability. SUMAL Information System, an Initiative in Romania. In *International Conference Interdisciplinarity in Engineering*; Springer: Cham, Switzerland, 2023; pp. 63–71. Available online: https://link.springer.com/chapter/10.1007/978-3-031-54671-6_5 (accessed on 2 February 2025).
19. Dubovik, O.; Schuster, G.L.; Xu, F.; Hu, Y.; Bösch, H.; Landgraf, J.; Li, Z. Grand Challenges in Satellite Remote Sensing. *Front. Remote Sens.* **2021**, *2*, 619818. [CrossRef]
20. Avram, S.; Ontel, I.; Gheorghe, C.; Rodino, S.; Roșca, S. Applying a Complex Integrated Method for Mapping and Assessment of the Degraded Ecosystem Hotspots from Romania. *Int. J. Environ. Res. Public Health* **2021**, *18*, 11416. [CrossRef]
21. Keskes, M.I.; Nita, M.D. Developing an AI Tool for Forest Monitoring: Introducing SylvaMind AI. *Bull. Transilv. Univ. Bras. Ser. II For. Wood Ind. Agric. Food Eng.* **2024**, *17*, 39–54. [CrossRef]
22. Liu, J.; Zuo, Y.; Wang, N.; Yuan, F.; Zhu, X.; Zhang, L.; Zhang, J.; Sun, Y.; Guo, Z.; Guo, Y.; et al. Comparative Analysis of Two Machine Learning Algorithms in Predicting Site-Level Net Ecosystem Exchange in Major Biomes. *Remote Sens.* **2021**, *13*, 2242. [CrossRef]
23. Ghosh, S.M.; Behera, M.D.; Paramanik, S. Canopy Height Estimation Using Sentinel Series Images through Machine Learning Models in a Mangrove Forest. *Remote Sens.* **2020**, *12*, 1519. [CrossRef]
24. Tang, Z.; Xia, X.; Huang, Y.; Lu, Y.; Guo, Z. Estimation of National Forest Aboveground Biomass from Multi-Source Remotely Sensed Dataset with Machine Learning Algorithms in China. *Remote Sens.* **2022**, *14*, 5487. [CrossRef]
25. Chen, M.; Qiu, X.; Zeng, W.; Peng, D. Combining Sample Plot Stratification and Machine Learning Algorithms to Improve Forest Aboveground Carbon Density Estimation in Northeast China Using Airborne LiDAR Data. *Remote Sens.* **2022**, *14*, 1477. [CrossRef]
26. Li, C.; Zhou, L.; Xu, W. Estimating Aboveground Biomass Using Sentinel-2 MSI Data and Ensemble Algorithms for Grassland in the Shengjin Lake Wetland, China. *Remote Sens.* **2021**, *13*, 1595. [CrossRef]
27. Hu, T.; Sun, Y.; Jia, W.; Li, D.; Zou, M.; Zhang, M. Study on the Estimation of Forest Volume Based on Multi-Source Data. *Sensors* **2021**, *21*, 7796. [CrossRef] [PubMed]
28. McRoberts, R.E. A Two-Step Nearest Neighbors Algorithm Using Satellite Imagery for Predicting Forest Structure within Species Composition Classes. *Remote Sens. Environ.* **2009**, *113*, 532–545. [CrossRef]
29. Cutler, D.R.; Edwards, T.C.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. Random Forests for Classification in Ecology. *Ecology* **2007**, *88*, 2783–2792. [CrossRef]
30. Witten, I.H.; Frank, E.; Hall, M.A. *Data Mining: Practical Machine Learning Tools and Techniques*; Elsevier: Amsterdam, The Netherlands, 2010; ISBN 9780123748560.

31. Jeong, J.H.; Resop, J.P.; Mueller, N.D.; Fleisher, D.H.; Yun, K.; Butler, E.E.; Timlin, D.J.; Shim, K.-M.; Gerber, J.S.; Reddy, V.R.; et al. Random Forests for Global and Regional Crop Yield Predictions. *PLoS ONE* **2016**, *11*, e0156571. [CrossRef]
32. Luo, Y.; Qi, S.; Liao, K.; Zhang, S.; Hu, B.; Tian, Y. Mapping the Forest Height by Fusion of ICESat-2 and Multi-Source Remote Sensing Imagery and Topographic Information: A Case Study in Jiangxi Province, China. *Forests* **2023**, *14*, 454. [CrossRef]
33. Mola, F.; Miele, R. Evolutionary Algorithms for Classification and Regression Trees. In *Data Analysis, Classification and the Forward Search, Proceedings of the Meeting of the Classification and Data Analysis Group (CLADAG) of the Italian Statistical Society, University of Parma, Parma, Italy, 6–8 June 2005*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 255–262. [CrossRef]
34. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
35. Salzberg, S.L. Book Review: C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. *Mach. Learn.* **1994**, *16*, 235–240. [CrossRef]
36. Franklin, J. The Elements of Statistical Learning: Data Mining, Inference and Prediction. *Math. Intell.* **2005**, *27*, 83–85. [CrossRef]
37. Zhang, N.; Chen, M.; Yang, F.; Yang, C.; Yang, P.; Gao, Y.; Shang, Y.; Peng, D. Forest Height Mapping Using Feature Selection and Machine Learning by Integrating Multi-Source Satellite Data in Baoding City, North China. *Remote Sens.* **2022**, *14*, 4434. [CrossRef]
38. Sprangers, O.; Schelter, S.; de Rijke, M. Probabilistic Gradient Boosting Machines for Large-Scale Probabilistic Regression. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Singapore, 14–18 August 2021. [CrossRef]
39. Ibrahim, S.; Balzter, H.; Tansey, K. Machine learning feature importance selection for predicting aboveground biomass in African savannah with landsat 8 and ALOS PALSAR data. *Mach. Learn. Appl.* **2024**, *16*, 100561. [CrossRef]
40. Chirici, G.; Barbati, A.; Corona, P.; Marchetti, M.; Travaglini, D.; Maselli, F.; Bertini, R. Non-Parametric and Parametric Methods Using Satellite Images for Estimating Growing Stock Volume in Alpine and Mediterranean Forest Ecosystems. *Remote Sens. Environ.* **2008**, *112*, 2686–2700. [CrossRef]
41. Vafaei, S.; Soosani, J.; Adeli, K.; Fadaei, H.; Naghavi, H.; Pham, T.; Tien Bui, D. Improving Accuracy Estimation of Forest Aboveground Biomass Based on Incorporation of ALOS-2 PALSAR-2 and Sentinel-2A Imagery and Machine Learning: A Case Study of the Hyrcanian Forest Area (Iran). *Remote Sens.* **2018**, *10*, 172. [CrossRef]
42. Peng, X.; Zhao, A.; Chen, Y.; Chen, Q.; Liu, H.; Wang, J.; Li, H. Comparison of Modeling Algorithms for Forest Canopy Structures Based on UAV-LiDAR: A Case Study in Tropical China. *Forests* **2020**, *11*, 1324. [CrossRef]
43. Rubbens, P.; Brodie, S.; Cordier, T.; Destro Barcellos, D.; Devos, P.; Fernandes-Salvador, J.A.; Fincham, J.I.; Gomes, A.; Handegard, N.O.; Howell, K.; et al. Machine Learning in Marine Ecology: An Overview of Techniques and Applications. *ICES J. Mar. Sci.* **2023**, *80*, 1829–1853. [CrossRef]
44. Stăncioiu, P.T.; Niță, M.D.; Fedorca, M. Capercaillie (*Tetrao urogallus*) Habitat in Romania A Landscape Perspective Revealed by Cold War Spy Satellite Images. *Sci. Total Environ.* **2021**, *781*, 146763. [CrossRef]
45. Saha, T.K.; Pal, S.; Talukdar, S.; Debanshi, S.; Khatun, R.; Singha, P.; Mandal, I. How Far Spatial Resolution Affects the Ensemble Machine Learning Based Flood Susceptibility Prediction in Data Sparse Region. *J. Environ. Manag.* **2021**, *297*, 113344. [CrossRef]
46. Smith, A.M.; Capinha, C.; Kramer, A.M. Predicting Species Distributions with Environmental Time Series Data and Deep Learning. *bioRxiv* **2022**. [CrossRef]
47. Ota, T.; Ogawa, M.; Shimizu, K.; Kajisa, T.; Mizoue, N.; Yoshida, S.; Takao, G.; Hirata, Y.; Furuya, N.; Sano, T.; et al. Aboveground Biomass Estimation Using Structure from Motion Approach with Aerial Photographs in a Seasonal Tropical Forest. *Forests* **2015**, *6*, 3882–3898. [CrossRef]
48. Cracknell, M.J.; Reading, A.M. Geological Mapping Using Remote Sensing Data: A Comparison of Five Machine Learning Algorithms, Their Response to Variations in the Spatial Distribution of Training Data and the Use of Explicit Spatial Information. *Comput. Geosci.* **2014**, *63*, 22–33. [CrossRef]
49. Lang, N.; Schindler, K.; Wegner, J.D. Country-Wide High-Resolution Vegetation Height Mapping with Sentinel-2. *Remote Sens. Environ.* **2019**, *233*, 111347. [CrossRef]
50. 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]
51. Burzykowski, T.; Geubbelmans, M.; Rousseau, A.-J.; Valkenborg, D. Validation of Machine Learning Algorithms. *Am. J. Orthod. Dentofac. Orthop.* **2023**, *164*, 295–297. [CrossRef]
52. Spârchez, G.; Târziu, D.R.; Dincă, L. *Pedologie Cu Elemente de Geologie Și Geomorfologie*; Universitatea Transilvania din Brașov: Brașov, Romania, 2013.
53. USGS. *Shuttle Radar Topography Mission*; USGS: Reston, VA, USA, 2003. Available online: https://www.usgs.gov/publications/shuttle-radar-topography-mission-srtm-0 (accessed on 2 February 2025).
54. Nita, M.D.; Munteanu, C.; Gutman, G.; Abrudan, I.V.; Radeloff, V.C. Widespread Forest Cutting in the Aftermath of World War II Captured by Broad-Scale Historical Corona Spy Satellite Photography. *Remote Sens. Environ.* **2018**, *204*, 322–332. [CrossRef]

55. García-Duro, J.; Ciceu, A.; Chivulescu, S.; Badea, O.; Tanase, M.A.; Aponte, C. Shifts in Forest Species Composition and Abundance under Climate Change Scenarios in Southern Carpathian Romanian Temperate Forests. *Forests* **2021**, *12*, 1434. [CrossRef]
56. Stăncioiu, P.T.; Niță, M.D.; Lazăr, G.E. Forestland Connectivity in Romania—Implications for Policy and Management. *Land Use Policy* **2018**, *76*, 487–499. [CrossRef]
57. Munteanu, C.; Nita, M.D.; Abrudan, I.V.; Radeloff, V.C. Historical Forest Management in Romania Is Imposing Strong Legacies on Contemporary Forests and Their Management. *For. Ecol. Manag.* **2016**, *361*, 179–193. [CrossRef]
58. Niță, M.D. Testing Forestry Digital Twinning Workflow Based on Mobile LiDAR Scanner and AI Platform. *Forests* **2021**, *12*, 1576. [CrossRef]
59. Giurgiu, V.; Decei, I.; Draghiciu, D. *Metode si Tabele Dendrometrice*; Editura Ceres: Bucharest, Romania, 2004.
60. Nguyen, Q.H.; Ly, H.B.; Ho, L.S.; Al-Ansari, N.; Van Le, H.; Tran, V.Q.; Prakash, I.; Pham, B.T. Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. *Math. Probl. Eng.* **2021**, *2021*, 4832864. [CrossRef]
61. Schneider, F.D.; Longo, M.; Paul-Limoges, E.; Scholl, V.M.; Schmid, B.; Morsdorf, F.; Pavlick, R.P.; Schimel, D.S.; Schaepman, M.E.; Moorcroft, P.R. Remote Sensing-Based Forest Modeling Reveals Positive Effects of Functional Diversity on Productivity at Local Spatial Scale. *J. Geophys. Res. Biogeosci.* **2023**, *128*, e2023JG007421. [CrossRef]
62. Qin, Y.; Xiao, X.; Dong, J.; Zhou, Y.; Wang, J.; Doughty, R.B.; Chen, Y.; Zou, Z.; Moore, B. Annual Dynamics of Forest Areas in South America during 2007–2010 at 50-m Spatial Resolution. *Remote Sens. Environ.* **2017**, *201*, 73–87. [CrossRef]
63. 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]
64. Baban, G.; Daniel Niță, M. Measuring Forest Height from Space. Opportunities and Limitations Observed in Natural Forests. *Measurement* **2023**, *211*, 112593. [CrossRef]
65. Nita, M.-D.; Clinciu, I. Hydrological Mapping of the Vegetation Using Remote Sensing Products. *Bull. Transilv. Univ. Bras. Ser. II For. Wood Ind. Agric. Food Eng.* **2010**, *3*, 73–78.
66. Le Toan, T.; Quegan, S.; Davidson, M.W.J.; Balzter, H.; Paillou, P.; Papathanassiou, K.; Plummer, S.; Rocca, F.; Saatchi, S.; Shugart, H.; et al. The BIOMASS Mission: Mapping Global Forest Biomass to Better Understand the Terrestrial Carbon Cycle. *Remote Sens. Environ.* **2011**, *115*, 2850–2860. [CrossRef]
67. Naik, P.; Dalponte, M.; Bruzzone, L. Prediction of Forest Aboveground Biomass Using Multitemporal Multispectral Remote Sensing Data. *Remote Sens.* **2021**, *13*, 1282. [CrossRef]
68. Potapov, P.V.; Turubanova, S.A.; Tyukavina, A.; Krylov, A.M.; McCarty, J.L.; Radeloff, V.C.; Hansen, M.C. Eastern Europe's Forest Cover Dynamics from 1985 to 2012 Quantified from the Full Landsat Archive. *Remote Sens. Environ.* **2015**, *159*, 28–43. [CrossRef]
69. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Routledge: London, UK, 2017; ISBN 9781315139470.
70. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer: New York, NY, USA, 2009; ISBN 978-0-387-84857-0.
71. Bergstra, J.; Bengio, Y. Random Search for Hyper-Parameter Optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
72. Snoek, J.; Larochelle, H.; Adams, R.P. Practical Bayesian Optimization of Machine Learning Algorithms. *arXiv* **2012**, arXiv:1206.2944.
73. Scheirer, C.J.; Ray, W.S.; Hare, N. The Analysis of Ranked Data Derived from Completely Randomized Factorial Designs. *Biometrics* **1976**, *32*, 429. [CrossRef]
74. Calvin, D. *Choosing and Using Statistics: A Biologist's Guide*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2011; Volume 147.
75. 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; ACM: New York, NY, USA, 2016; pp. 785–794.
76. Nemeth, M.; Borkin, D.; Michalconok, G. The Comparison of Machine-Learning Methods XGBoost and LightGBM to Predict Energy Development. In *Computational Statistics and Mathematical Modeling Methods in Intelligent Systems, Proceedings of the 3rd Computational Methods in Systems and Software, Szczecin, Poland, 2019*; Springer: Cham, Switzerland, 2019; pp. 208–215. [CrossRef]
77. Lundberg, S.M.; Erion, G.G.; Lee, S.-I. Consistent Individualized Feature Attribution for Tree Ensembles. *arXiv* **2018**, arXiv:1802.03888.
78. Wager, S.; Athey, S. Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests. *J. Am. Stat. Assoc.* **2018**, *113*, 1228–1242. [CrossRef]
79. Vorovencii, I. Assessing Various Scenarios of Multi-Temporal Sentinel-2 Imagery, Topographic Data, Texture Features, and Machine Learning Algorithms for Tree Species Identification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 15373–15392. [CrossRef]
80. Rodriguez-Galiano, V.F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J.P. An Assessment of the Effectiveness of a Random Forest Classifier for Land-Cover Classification. *ISPRS J. Photogramm. Remote Sens.* **2012**, *67*, 93–104. [CrossRef]
81. Persson, H. Estimation of Boreal Forest Attributes from Very High Resolution Pléiades Data. *Remote Sens.* **2016**, *8*, 736. [CrossRef]

82. Chen, L.; Wang, Y.; Ren, C.; Zhang, B.; Wang, Z. Optimal Combination of Predictors and Algorithms for Forest Above-Ground Biomass Mapping from Sentinel and SRTM Data. *Remote Sens.* **2019**, *11*, 414. [CrossRef]
83. Fassnacht, F.E.; White, J.C.; Wulder, M.A.; Næsset, E. Remote Sensing in Forestry: Current Challenges, Considerations and Directions. *For. Int. J. For. Res.* **2024**, *97*, 11–37. [CrossRef]
84. Chere, Z.; Zewdie, W.; Biru, D. Machine Learning for Modeling Forest Canopy Height and Cover from Multi-Sensor Data in Northwestern Ethiopia. *Environ. Monit. Assess.* **2023**, *195*, 1452. [CrossRef]
85. Mohamed, A.H.; Keskes, M.I.; Nita, M.D. Analyzing the Accuracy of Satellite-Derived DEMs Using High-Resolution Terrestrial LiDAR. *Land* **2024**, *13*, 2171. [CrossRef]
86. Erb, K.-H.; Kastner, T.; Luyssaert, S.; Houghton, R.A.; Kuemmerle, T.; Olofsson, P.; Haberl, H. Bias in the Attribution of Forest Carbon Sinks. *Nat. Clim. Chang.* **2013**, *3*, 854–856. [CrossRef]
87. Munteanu, C.; Senf, C.; Nita, M.D.; Sabatini, F.M.; Oeser, J.; Seidl, R.; Kuemmerle, T. Using Historical Spy Satellite Photographs and Recent Remote Sensing Data to Identify High-conservation-value Forests. *Conserv. Biol.* **2022**, *36*, e13820. [CrossRef] [PubMed]