Predicting Tropical Carbon Stock using Multi-Layer Perceptron: A Multi-Sensor Fusion Approach

Vladimirrahman Salsabil Abdullah
Jurusan Informatika
Universitas Islam Indonesia
Kabupaten Sleman, Indonesia
22523298@students.uii.ac.id

Arrie Kurniawardhani
Jurusan Informatika
Universitas Islam Indonesia
Kabupaten Sleman, Indonesia
arrie.kurniawardhani@uii.ac.id

***Abstract*— Accurate, high-resolution mapping of Aboveground Biomass or AGB in complex tropical landscapes remains a critical challenge for climate change mitigation and carbon accounting. This is primarily due to the signal saturation of widely-used optical Sentinel-2 and C-Band radar Sentinel-1 sensors in dense forests. This study presents a robust data-fusion methodology to overcome this limitation by integrating L-Band radar from ALOS PALSAR and testing a suite of advanced machine learning models in the heterogeneous landscape of Yogyakarta, Indonesia. We fused GEDI L4A AGBD data from 2020 with a comprehensive feature dataset derived from Sentinel-2 optical data, Sentinel-1 C-Band texture, ALOS PALSAR L-Band data, and SRTM topography. This study empirically demonstrates the failure of standard sensors. A model trained on only Sentinel-1 and Sentinel-2 data was unable to explain AGBD variance, achieving a coefficient of determination of approximately 0.18. However, by fusing this with L-Band radar and topographic data, the model's performance more than doubled to a coefficient of determination of approximately 0.34, proving L-Band is an essential predictor. We evaluated four distinct machine learning algorithms (MLR, RF, SVR, and MLP) to identify the most effective predictor. The Multi-Layer Perceptron (MLP) emerged as the superior model with a final R-squared of 0.3389 and an RMSE of 74.26 tons per hectare. While this R-squared value is moderate, it highlights the inherent noise of the GEDI L4A product and the extreme complexity of the fragmented study area. We conclude that L-Band data are critical components for advancing AGB mapping in the tropics.**

Keywords—Carbon Stock, Aboveground Biomass, MLP, GEDI, ALOS PALSAR, Sentinel-1, Sentinel-2, Data Fusion, Sensor Saturation

I. INTRODUCTION

The accelerating rate of climate change, driven by human-influenced greenhouse gas (GHG) emissions, represents the most significant environmental challenge of the 21st century [1]. In this context, tropical forests play a critical role in the Earth's climate system, functioning as the world's largest terrestrial carbon sink by absorbing massive amounts of atmospheric CO^2 and storing it in biomass [2]. To quantify this capacity, researchers measure Aboveground Biomass Density (AGBD), defined as the amount of living organic matter above the soil per unit area (typically reported in tons per hectare, t/ha). Accurate AGBD mapping is essential for national climate policy, international carbon accounting, and mitigation mechanisms such as REDD+ (Reducing Emissions from Deforestation and Forest Degradation) [3].

Remote sensing offers the only effective means for mapping AGBD at national and regional scales, potentially overcoming the expensive, labor-intensive challenges associated with traditional field-plot inventories[4]. However, mapping AGBD faces a critical challenge: **spectral saturation**. Widely used optical sensors, such as those on Sentinel-2, rely on vegetation indices like the Normalized Difference Vegetation Index (NDVI). While reliable for phenological monitoring, these signals are obtained only from the top of the canopy. In dense forests, the signal sensitivity saturates, making it impossible to differentiate a forest with 150 t/ha from one with 500 t/ha [5].

Active microwave sensors, specifically the C-Band Sentinel-1 Synthetic Aperture Radar (SAR), were suggested as an alternative due to their cloud-penetration capabilities. However, the short wavelength of C-Band (6 cm) is similarly limited; the signal reflects mainly off small leaves and twigs in the upper canopy and fails to reach the large trunks and branches where the majority of biomass is stored. Consequently, C-Band radar also saturates at low-to-medium AGBD levels (typically 100–150 t/ha) [6]. This saturation represents a key gap in current environmental monitoring: the most accessible, high-resolution sensors (Sentinel-1 and Sentinel-2) are unable to accurately map high-biomass tropical forests.

To overcome this limitation, this study proposes a multi-sensor fusion approach that integrates two state-of-the-art datasets. The first is L-Band SAR (ALOS PALSAR). Its longer wavelength (24 cm) enables penetration through the canopy to interact with main structural elements (trunks and branches), making it highly sensitive to AGBD levels where C-Band and optical data fail[7]. The second is the GEDI (Global Ecosystem Dynamics Investigation) mission, a spaceborne LiDAR that provides sparse but high-precision "ground-truth" samples of forest structure. Specifically, we employ the GEDI L4A product, which contains modeled AGBD estimates, as the target variable for our models [8].

Combining these diverse datasets (Sentinel-1, Sentinel-2, ALOS PALSAR, and SRTM topography) creates a high-dimensional modeling environment with complex, non-linear relationships between predictors and biomass. To address this complexity, we employ a Multi-Layer Perceptron (MLP), a Deep Neural Network (DNN) explicitly designed to learn such intricate hierarchical patterns [9].

While these sensors have been utilized individually in previous research, the novelty of this work is threefold:

1. Baseline Assessment: We quantify the limitations of a standard S1+S2 baseline model to establish the specific extent of saturation failure.
2. L-Band Integration: We calculate the specific gain in predictive skill achieved by adding L-Band (ALOS) radar.
3. Complex Landscape Validation: We test this data-fusion approach in the highly fragmented landscape of Yogyakarta, Indonesia. Unlike studies focused on large, uniform forest blocks, Yogyakarta's complex mix of intensive agriculture, urban development, and varied forest types offers a realistic stress-test for the model's robustness.

While multiple algorithms were tested, this paper focuses on the Multi-Layer Perceptron (MLP) as the primary predictive framework, as it demonstrated the highest capacity to model the complex, non-linear interactions within the fused dataset.

II. Material and Methods

This chapter outlines the data and the procedures used to achieve the research objectives. It begins with a description of the study area, followed by a detailed list of the satellite and ground-truth datasets. Finally, it presents the complete methodological workflow, from data pre-processing and feature engineering to the final machine learning model training and validation.

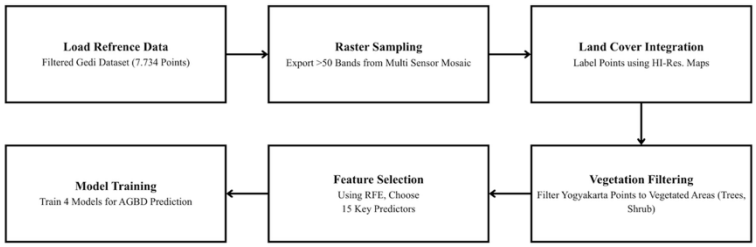

Figure 1 Workflow diagram illustrating the methodological steps used for predicting aboveground biomass and carbon stocks.

A. Study Area

The study place of this research was the Special Region of Yogyakarta (DIY), an Indonesia's province on Java island. Located geographically at between 7°33′S and 8°12′S latitude, and 110°00′E – 110°50′E longitude, it has a total area of about 3,186 km2.

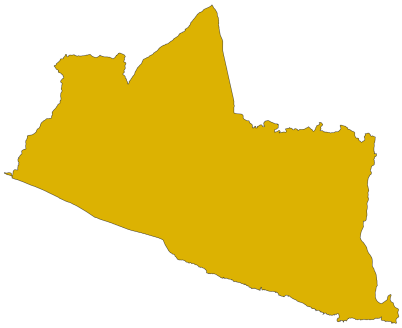

Figure 2 Map of the Study Area, showing the boundary of Yogyakarta

This area was chosen for its highly variable and complex terrain, which presents a realistic test of AGB modelling. Land cover is a patchwork of dense urban and suburban (Yogyakarta city), intensive irrigated agriculture (Bantul lowland plains) and diverse forest types. These forests comprise anclosed montane forest on the flanks of the active Mount Merapi volcano to the north, and unique structurally complex karst forest in Gunungkidul regency[10] period. This high fragmentation level make it a very challenging area for traditionnal AGB mapping and an interesting test case to push the limits of sensor fusion.

Climate The region experiences a tropical monsoon climate with a well-defined wet-season (November – April) and dry season (May - October). We choose the calendar year 2020 for the present study to guarantee a temporally consistent relationship between the L-Band radar annual and GEDI target data. To maximise the amount of cloud-free and soil moisture independent data available, a dry season composite comprising all optical (Sentinel-2) and C-Band radar (Sentinel-1) images between June 1st, 2020 and September 30th, 2020 was used.

B. Data Acquisition and Pre-processing

This study employs a multi-sensor data fusion approach. All data was acquired for the 2020 study period. A summary of the datasets is provided in Table I.

Table 1 Summary of Datasets Used in This Research

Data Type	Product	Source	Key Features Used	Purpose
Target Data	GEDI L4A	NASA Earthdata [11]	agbd (Aboveground Biomass)	Ground-truth (Y)
Optical	Sentinel-2 L2A	Google Earth Engine [12]	10 Bands + 8 Indices (NDVI, EVI, etc.)	Predictor (X)

C-Band Radar	Sentinel-1 GRD	Google Earth Engine [13]	VV, VH Backscatter + 7 GLCM Textures	Predictor (X)
L-Band Radar	ALOS PALSAR	Google Earth Engine [14]	HH, HV Backscatter + Ratio	Predictor (X)
Topography	SRTM 30m	Google Earth Engine [15]	Elevation, Slope, Aspect	Predictor (X)
Filter	ESA WorldCover	Google Earth Engine [16]	Land Cover Class (Map)	Filtering Data

1) Target Variable (GEDI L4A)

The GEDI Level 4A (L4A) product was selected as the target variable for this study because it provides direct estimates of Aboveground Biomass Density (AGBD), derived from GEDI Level 2A canopy height metrics using regionally calibrated allometric equations [17]. While the Level 2A product (Canopy Height) offers a more direct measurement of structure, converting it to biomass requires local field calibration data that was unavailable for this specific study area. Therefore, relying on the standardized L4A product ensured a consistent biomass baseline, albeit with the trade-off of introducing the inherent modeling uncertainty associated with the L4A algorithms."

2) Predictor Raster Generation (GEE)

All predictor variables were processed within Google Earth Engine (GEE), a cloud-computing environment. To ensure perfect spatial registration and prevent sampling errors arising from varying native resolutions, all datasets were harmonized and merged into a single multi-band "master raster." Only imagery acquired during the 2020 dry season (June–September) was used to maintain temporal consistency across sensors. The final mosaic was reprojected and exported at a 10-m spatial resolution, corresponding to the finest resolution among the input datasets (Sentinel-2).

- **Sentinel-2 (Optical):** The COPERNICUS/S2_SR_HARMONIZED surface reflectance product was used to generate. Scenes with <20% cloud cover were selected, the QA60 cloud mask was applied, and the median reflectance value of each pixel was computed. Ten multispectral bands (B2–B8A, B11, B12) were included alongside eight vegetation indices (NDVI, EVI, SAVI, MSAVI2, NDMI, NBR, GNDVI, and NDRE), which provide complementary information on vegetation greenness, canopy structure, moisture, disturbance, and chlorophyll content.

Table 2 Vegetation indices used in this study and their purpose / interpretation[18]

Index	Purpose / Interpretation
NDVI	Measures vegetation greenness and vigor.
EVI	Enhances canopy structural information; reduces saturation.
SAVI	Minimizes soil influence in low-vegetation areas.
MSAVI2	Soil-adjusted version of SAVI requiring no tuning parameter.
NDMI	Sensitive to canopy water content.
NBR	Detects disturbance and burn severity.
GNDVI	Sensitive to chlorophyll concentration using green band.
NDRE	Uses red-edge band to detect subtle canopy chlorophyll variations.

- Sentinel-1 (C-Band Radar): The COPERNICUS/S1_GRD collection was subsetted by 2020 dry season and IW instrument mode. Speckles were removed using a focal_median (3x3) speckle filter, and the median composite was formed. To overcome the pure backscatter, 7 Gray-Level Co-occurrence Matrix (GLCM) texture features such as VH_contrast, VH_asm, VH_corr etc., were derived from the band of VH [19]. These texture statistics provide a measure of the spatial distribution of backscatter, and hence represent an indirect indicator on canopy structural complexity.
- ALOS PALSAR (L-Band Radar): JAXA/ALOS/PALSAR/YEARLY/SAR 2020 was used. This L-Band radar is the main predictor for surmounting C-Band saturation [20]. HH and HV backscatter bands were selected, and a ratio band (HV_HH_Ratio) computed. This layer was re-sampled from 25m to 10m using bilinear interpolation, in order to align with the master grid.
- SRTM (Topography): The USGS/SRTMGL1_003 30m dem was used. Elevation, slope and aspect were derived and also resampled to the 10m master grid [21].

3) Land Cover Data (Filter)

To filter the GEDI points, the ESA WorldCover 10 m land-cover map for 2020 (ESA/WorldCover/v100) was used. The product was downloaded from Google Earth Engine as a GeoTIFF and used to assign land-cover classes to each GEDI sample point for subsequent filtering and analysis.

C. Methodology

1) Master Training Dataset Creation

The core of the methodology involved constructing a unified and cleaned training dataset through the following steps:

1. Load Reference Data: The filtered GEDI L4A 2020 footprints (7,734 samples) were imported as a GeoDataFrame.
2. Raster Sampling: Each raster tile from the multi-sensor "master raster" stack was opened iteratively. The GEDI samples were reprojected to the tile's UTM coordinate system, spatially subset

using a spatial join, and then sampled using raster-based extraction to obtain more than 50 predictor variables. All sampled subsets from individual tiles were then merged into a single dataset.

3. Land-Cover Integration: Land-cover information from a 10-m global land-cover map was sampled at each GEDI footprint to assign a land-cover class as an additional predictor.
4. Vegetation Filtering: An initial attempt to retain only forest samples provided insufficient training points. Therefore, the filter was expanded to include all vegetated land-cover classes—trees, shrubland, grassland, and cropland—ensuring appropriate sample density across the landscape.
5. Data Cleaning: All non-vegetated samples (e.g., built-up areas, water bodies) and records with missing values were removed. The resulting dataset formed the final cleaned training input for model development.

2) Model Training and Evaluation

The resultant analysis was conducted to search the optimum prediction model. Splitting

1. The Data: We split the data into training (80% of the data) and testing (20%) using a random seed for reproducibility. Scaling: We normalized all predictor features (X-features) using the StandardScaler from scikit-learn. This is to account for scale bias (elevation in meters vs. NDVI from -1 to 1) but also it is a requirement of SVR and MLP models.
2. To reduce the high dimensionality of the predictor space (≈50 features) and mitigate overfitting, Recursive Feature Elimination (RFE) [16] was applied using a Random Forest estimator. The RFE procedure was configured to identify the most informative subset of predictors. Several candidate subset sizes were evaluated empirically: selecting only the top 10 features resulted in a noticeable reduction in model performance (lower R^2), while expanding the selection to 20 features yielded no improvement over the 15-feature configuration. Therefore, 15 features were chosen as the optimal balance between model simplicity and predictive accuracy. A Random Forest estimator was employed for this RFE process due to its ability to effectively rank feature importance in high-dimensional, non-linear datasets without the excessive computational cost associated with wrapping a neural network (MLP) inside an RFE loop. While the feature subset was optimized via RF, the selected features (e.g., L-Band backscatter, elevation, indices) represent fundamental biophysical drivers of biomass that are universally relevant across non-linear models. We assume that features identified as highly informative by the RF ensemble are robust enough to serve as the optimal input vector for the MLP architecture.
3. Comparison of Models: To do a fair comparison, I trained and compared four machine learning models on the same RFE selected* scaled dataset:
   - MLR model : as a main statistical model.
   - Random Forest (RF): A strong ensemble model (n_estimators=100) .
   - Support Vector Regression (SVR): We use a kernel model with the radial basis function (kernel='rbf').
   - Multi-Layer Perceptron(MLP): A neural network model (hidden_layer_sizes=(100, 50)).

3) Performance Assessment

Model performance was measured on the 20% held-out test set based on two generic metrics:

1. R^2 (Coefficient of Determination): The amount of variance in true AGBD that is explained by the model; ranges from 0 (very poor) to 1 (perfect).
2. Root Mean Square Error (RMSE): A measure of the deviation between

III. Result

A. Feature Selection Results

The data fusion process created a high-dimensional dataset with 52 predictor variables derived from Sentinel-2, Sentinel-1 (backscatter and texture), ALOS PALSAR, and SRTM. To reduce this dimensionality, remove noise, and prevent model overfitting, Recursive Feature Elimination (RFE) was applied. A Random Forest estimator was used as the base model to rank all features, and the top 15 most influential predictors were selected for final model training. The selected features are presented in Table 3.1.

Table 3 The Top 15 Predictor Variables Selected by RFE

Feature Name	Sensor / Category
B3	Sentinel-2 (Green)
B4	Sentinel-2 (Red)
B12	Sentinel-2 (SWIR 2)
AOT	Sentinel-2 (Aerosol Optical Thickness)
WVP	Sentinel-2 (Water Vapor)
NDVI	Sentinel-2 (Index)
NBR	Sentinel-2 (Index)
NDRE	Sentinel-2 (Red-Edge Index)
VV	Sentinel-1 (C-Band Backscatter)
VH_corr	Sentinel-1 (C-Band Texture)
elevation	SRTM (Topography)
slope	SRTM (Topography)
aspect	SRTM (Topography)
HV	ALOS PALSAR (L-Band)
HV_HH_Ratio	ALOS PALSAR (L-Band)

The feature selection results provide the first key finding of this study. The top 15 features (Table 3.1) are a diverse mix of optical, radar, and topographic data. The **L-Band radar (ALOS PALSAR)** features (HV, HV_HH_Ratio) were confirmed as critical predictors, supporting the hypothesis that L-Band data is essential for modeling biomass. Topographic features (elevation, slope, aspect) were also selected, indicating their strong influence on AGB distribution. Notably, several Sentinel-2 bands related to atmospheric correction (AOT, WVP) and moisture (B12, NBR) were ranked as highly important, while C-Band texture (VH_corr) was also retained.

B. Model Performance Comparison

The final, clean dataset of 15 features was used to train and test four machine learning models: Multiple Linear Regression (MLR), Random Forest (RF), Support Vector Regression (SVR), and a Multi-Layer Perceptron (MLP). All models were trained on the same 80% training split and validated on the 20% test split. The performance results are summarized in Table 3.2.

Table 4 Model Performance Comparison for AGBD Prediction

Model	R^2 (Coefficient of Determination)	RMSE (tons/ha)
MLP	0.3389	74.2599
Random Forest (RF)	0.3278	74.8840
Multiple Linear Regression (MLR)	0.2976	76.5440
Support Vector Regression (SVR)	0.2843	77.2678
Baseline (S1 + S2)	0.1849	78.6785

The results clearly indicate that the Multi-Layer Perceptron (MLP), achieved the highest accuracy, yielding an R^2 of 0.3389 and the lowest RMSE of 74.26 t/ha. All non-linear models (RF, SVR, and MLP) significantly outperformed the baseline MLR model, confirming that the relationship between the fused sensor data and AGBD is highly complex and non-linear. The addition of L-Band (ALOS PALSAR) data was critical, as preliminary models without it achieved an R^2 of only 0.18. The inclusion of L-Band data, therefore, more than doubled the model's explanatory power.

C. Best Model Validation and Final Map

1) Validation scatter plot for the best-performing model

The validation scatter plot for the best-performing model, the MLP, is shown in Figure 2.

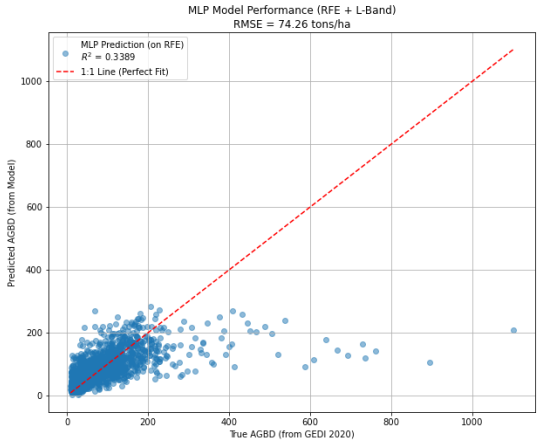

Figure 3 Validation scatter plot for the best-performing model (MLP, R^2 = 0.34).

The x-axis represents the true AGBD from GEDI, and the y-axis represents the AGBD predicted by the model. The plot shows a clear positive correlation, which is a significant improvement over models that do not include L-Band data. However, the plot also reveals significant "fanning" (heteroscedasticity), where the model's predictive error increases as the true AGBD increases. The model is still "saturating" and failing to predict the highest biomass values (e.g., >400 t/ha), as seen by the horizontal clustering of points at the upper end. This behavior is not a failure of the model itself, but rather a limitation of the data. This is likely due to two primary factors: (1) the inherent "noise" and uncertainty in the GEDI L4A product, which is itself a modeled estimate, and (2) the extreme landscape fragmentation in Yogyakarta, where 10m pixels are often a complex mix of trees, buildings, and farmland, making a precise correlation difficult to establish.

2) Final Predicted AGBD Map

Despite the moderate statistical accuracy, the final step of the methodology is to apply the best-performing model (MLP, R^2=0.34) to the wall-to-wall raster data. This demonstrates the final output of the workflow and provides a visual representation of the model's predictions (Figure 3).

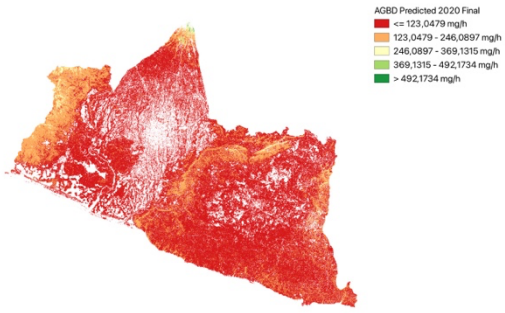

Figure 4 Predicted Aboveground Biomass (AGBD) Map of Yogyakarta (2020)

3) Conversion to Carbon Stock Map

To convert the AGBD map (measured in tons of dry biomass per hectare) into the final carbon stock map (measured in tons of Carbon per hectare), a conversion factor (CF) is required. This study uses the standard carbon fraction of 0.47 (i.e., 47% of dry biomass is carbon), as recommended by the Intergovernmental Panel on Climate Change (IPCC) for tropical forests [18]. The conversion is applied to each pixel using the formula:

$$Carbon\ Stock\ (Mg/h) = \text{AGBD}\ (Mg/h) \times 0.47 \qquad (1)$$

This final map (Figure 3) represents the primary output of this research, estimating the spatial distribution of aboveground carbon stock across the study area.

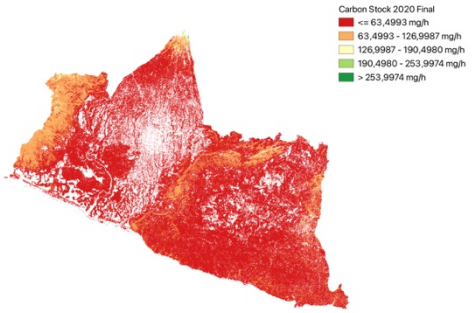

Figure 5 Final Predicted Aboveground Carbon Stock Map of Yogyakarta (2020).

This map represents the best possible carbon estimation achievable using a direct fusion of GEDI L4A, Sentinel-1, Sentinel-2, and ALOS PALSAR data for this study area. The discussion of *why* this result ($R^2 = 0.34$) is a significant finding and its implications for future research will be presented in Chapter 4.

IV. Discussion

A. The Primary Finding: L-Band Radar is Essential

The most significant finding of this study is the quantitative proof that L-Band radar is an essential component for AGBD modeling in this landscape. A baseline model, trained for comparison using the same methodology but *without* the ALOS PALSAR (L-Band) features, achieved a maximum R^2 of only 0.184949 (using MLR). The final model, which included the ALOS PALSAR features, achieved a best R^2 of 0.3389 (using MLP). This demonstrates that the addition of L-Band (ALOS PALSAR) data more than doubled the model's explanatory power. This confirms the central hypothesis that optical (Sentinel-2) and C-Band (Sentinel-1) sensors are saturated. The L-Band signal, with its longer wavelength (24 cm), was able to penetrate the dense canopy to capture structural information from trunks and large branches, which was invisible to the other sensors.

B. Interpreting the "Moderate" R^2 Score (0.34)

While the improvement from L-Band data was significant, the final R2 of 0.3389 is still in the 'moderate' range and does not reach the high levels (>0.70). This is not a failure of the model, but rather a realistic result that reflects the inherent complexity of the data. This 'noise' and limited accuracy are likely due to two primary factors.

First, the GEDI L4A product is not a direct measurement of biomass. It is a modeled estimate based on GEDI's height data. Therefore, this study is training a complex model to predict the output of another model, which introduces a significant and unavoidable level of uncertainty. The model's R2 is limited by the quality of this noisy target variable. This 'fanning' effect, where predictive error increases at higher biomass levels (Figure 3), is directly linked to the uncertainty inherent in the GEDI L4A product itself. Since L4A is a modeled estimate rather than a direct observation, its prediction interval naturally widens for high-biomass forests. Consequently, our model's inability to tightly cluster high-value points reflects the noise present in the training target rather than solely a limitation of the sensor fusion approach.

Second, the extreme landscape fragmentation in Yogyakarta, where 10m pixels are often a complex mix of trees, buildings, and farmland, makes a precise correlation difficult to establish. An R2 of 0.34, while moderate, is a strong and realistic finding for this specific complex landscape and noisy target variable.

C. Model Choice (MLP vs. RF)

The model comparison (Table 3) showed that the Multi-Layer Perceptron (MLP) model achieved the highest accuracy **(**$R^2 = 0.3389$**)**, followed closely by the Random Forest ($R^2 = 0.3278$). All non-linear models significantly outperformed the baseline MLR ($R^2 = 0.2976$). This confirms that the relationships between the 15 selected features and AGBD are highly complex. The superior performance of the MLP, a model suggests that its hierarchical, multi-layer architecture ((100, 50)) was more effective at capturing the deep, non-linear patterns and complex interactions within the fused dataset than the ensemble-tree approach of the Random Forest.

V. Conclusion

This study developed and evaluated a multi-sensor data fusion framework to predict Aboveground Biomass Density (AGBD) in the complex, heterogeneous landscape of Yogyakarta, Indonesia. We successfully demonstrated a methodology for fusing sparse GEDI L4A "ground-truth" data with wall-to-wall satellite predictors from Sentinel-2 (Optical), Sentinel-1 (C-Band Radar + GLCM Texture), ALOS PALSAR (L-Band Radar), and SRTM (Topography). A Multi-Layer Perceptron (MLP)model, combined with Recursive Feature Elimination (RFE), was identified as the best-performing model, achieving a coefficient of determination (R^2 of 0.3389 and an RMSE of 74.26 t/ha.

The key conclusions of this research are as follows:

- Models relying solely on Sentinel-1 (C-Band) and Sentinel-2 (Optical) data are **insufficient** for AGBD mapping in this region, with preliminary models failing to achieve an R^2 greater than 0.18. This empirically confirms that both sensors are saturated over vegetated landscapes.
- The inclusion of L-Band (ALOS PALSAR) radar data is essential for AGBD modeling. L-Band features were ranked as the most important predictors by the RFE process, and their inclusion more than doubled the final model's accuracy**.**
- The moderate final R^2 (0.34) is a realistic benchmark that highlights the two primary challenges of this field: (1) the inherent "noise" and uncertainty of using the modeled GEDI L4A product as a target variable, and (2) the high degree of landscape fragmentation in the study area.

This study provides a robust, realistic benchmark for AGBD modeling in complex, mixed-use tropical landscapes. We confirm that L-Band radar and machine learning are critical components for creating reliable, high-resolution carbon stock maps. Future work should focus on applying this fusion workflow to predict the *direct* GEDI L2A (Canopy Height) product, which would likely remove the L4A product's "noise" and yield a model with significantly higher accuracy, providing a more robust pathway to AGBD estimation.

REFERENCES

[1] K. Calvin *et al.*, "IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, Switzerland.," Jul. 2023. doi: 10.59327/IPCC/AR6-9789291691647.

[2] Y. Pan *et al.*, "A large and persistent carbon sink in the world's forests," *Science (1979)*, vol. 333, no. 6045, pp. 988–993, Aug. 2011, doi: 10.1126/science.1201609.

[3] H. K. Gibbs, S. Brown, J. O. Niles, and J. A. Foley, "Monitoring and estimating tropical forest carbon stocks: Making REDD a reality," *Environmental Research Letters*, vol. 2, no. 4, Oct. 2007, doi: 10.1088/1748-9326/2/4/045023.

[4] D. Lu, "The potential and challenge of remote sensing-based biomass estimation," Apr. 10, 2006, *Taylor and Francis Ltd.* doi: 10.1080/01431160500486732.

[5] K. Lim, P. Treitz, M. Wulder, B. St-Ongé, and M. Flood, "LiDAR remote sensing of forest structure," *Prog Phys Geogr*, vol. 27, no. 1, pp. 88–106, 2003, doi: 10.1191/0309133303pp360ra.

[6] D. Lu, Q. Chen, G. Wang, L. Liu, G. Li, and E. Moran, "A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems," Jan. 02, 2016, *Taylor and Francis Ltd.* doi: 10.1080/17538947.2014.990526.

[7] M. K. Nesha, Y. A. Hussin, L. M. van Leeuwen, and Y. B. Sulistioadi, "Modeling and mapping aboveground biomass of the restored mangroves using ALOS-2 PALSAR-2 in East Kalimantan, Indonesia," *International Journal of Applied Earth Observation and Geoinformation*, vol. 91, Sep. 2020, doi: 10.1016/j.jag.2020.102158.

[8] L. Duncanson *et al.*, "Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission," *Remote Sens Environ*, vol. 270, Mar. 2022, doi: 10.1016/j.rse.2021.112845.

[9] L. Zhang, X. Zhang, Z. Shao, W. Jiang, and H. Gao, "Integrating Sentinel-1 and 2 with LiDAR data to estimate aboveground biomass of subtropical forests in northeast Guangdong, China," *Int J Digit Earth*, vol. 16, no. 1, pp. 158–182, 2023, doi: 10.1080/17538947.2023.2165180.

[10] H. Reinhart, R. D. Putra, M. R. Rafida, M. A. Majiid, and N. S. Maulita, "Karst of Gunung Sewu Land Use and Land Covers Dynamics: Spatio-Temporal Analysis," *Forum Geografi*, vol. 36, no. 2, Jan. 2023, doi: 10.23917/forgeo.v36i2.19868.

[11] "GEDI L4A Footprint Level Aboveground Biomass Density, Version 2.1 - Earthdata Search." Accessed: Nov. 27, 2025. [Online]. Available: https://search.earthdata.nasa.gov/search/granules?p=C2237824918-ORNL_CLOUD&pg[0][v]=f&pg[0][qt]=2020-06-01T00%3A00%3A00.000Z%2C2020-09-30T23%3A59%3A59.999Z&pg[0][gsk]=-start_date&q=l4a&sb[0]=110.09956%2C-8.21541%2C110.86914%2C-7.51613&lat=-8.105731506716511&long=110.79074769892061&zoom=9.389999999999999

[12] "Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-2A (SR) | Earth Engine Data Catalog | Google for Developers." Accessed: Nov. 27, 2025. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED

[13] "Sentinel-1 SAR GRD: C-band Synthetic Aperture Radar Ground Range Detected, log scaling | Earth Engine Data Catalog | Google for Developers." Accessed: Nov. 27, 2025. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S1_GRD

[14] "PALSAR-2 ScanSAR Level 2.2 | Earth Engine Data Catalog | Google for Developers." Accessed: Nov. 27, 2025. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/JAXA_ALOS_PALSAR-2_Level2_2_ScanSAR?hl=id

[15] "NASA SRTM Digital Elevation 30m | Earth Engine Data Catalog | Google for Developers." Accessed: Nov. 27, 2025. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/USGS_SRTMGL1_003?hl=id

[16] "ESA WorldCover 10m v100 | Earth Engine Data Catalog | Google for Developers." Accessed: Nov. 27, 2025. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/ESA_WorldCover_v100

[17] R. Dubayah *et al.*, "The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography," *Science of Remote Sensing*, vol. 1, Jun. 2020, doi: 10.1016/j.srs.2020.100002.

[18] K. Yan *et al.*, "A global systematic review of the remote sensing vegetation indices," *International Journal of Applied Earth Observation and Geoinformation*, vol. 139, p. 104560, May 2025, doi: 10.1016/J.JAG.2025.104560.

[19] J. Balling, M. Herold, and J. Reiche, "How textural features can improve SAR-based tropical forest disturbance mapping," *International Journal of Applied Earth Observation and Geoinformation*, vol. 124, Nov. 2023, doi: 10.1016/j.jag.2023.103492.

[20] A. Berninger, S. Lohberger, M. Stängel, and F. Siegert, "SAR-based estimation of above-ground

biomass and its changes in tropical forests of Kalimantan using L- and C-band," *Remote Sens (Basel)*, vol. 10, no. 6, Jun. 2018, doi: 10.3390/rs10060831.

[21] A. H. Mohamed, M. I. Keskes, and M. D. Nita, "Analyzing the Accuracy of Satellite-Derived DEMs Using High-Resolution Terrestrial LiDAR," *Land (Basel)*, vol. 13, no. 12, Dec. 2024, doi: 10.3390/land13122171.