
Large area forest aboveground biomass mapping by integrating GEDI and Sentinel-2

Chongguang Fan¹

Abstract

Forest aboveground biomass (AGB) is an important variable for carbon storage calculation and forest management. Remote sensing techniques are widely used to retrieve forest AGB in large areas. Here, a 10-m resolution forest AGB map of Northeast China for the year 2022 is produced by combining GEDI footprint aboveground biomass density (AGBD) and Sentinel-2 spectral bands using machine learning methods. Random Forest model (RMSE=45.45, MAE=34.71) outperforms Light GBM model and Neural Networks. Based on Random Forest model, the forest AGBD in Northeast China ranges from 16 to 299 Mg/ha, with a mean value of 118 Mg/ha. This study presents a common methodology for large area high-resolution forest AGB mapping which could be applied on various spatial and temporal scales.

1. Introduction

Forests cover more than thirty percent of the global land area (FAO, 2022), providing high ecosystem service value in terms of timber supply, climate regulation, soil and water conservation, and biodiversity maintenance. In addition, as one of the primary carbon reservoirs, forests are integral to the global carbon cycle and climate change mitigation. Forest aboveground biomass (AGB) refers to the weight of living vegetation above the soil, including stem, branch, foliage, etc. (Ashton, 2012), which directly reflects the status and changes of forest ecosystem and is a key indicator of carbon storage (Li et al., 2019). Consequently, accurate estimations of forest AGB are essential, not only for studying terrestrial carbon cycles in the context of climate change but also for formulating strategies for sustainable forest management (Tian et al., 2023).

¹College of Urban and Environmental Sciences, Peking University, Beijing, China. Correspondence to: Chongguang Fan <cgfan@stu.pku.edu.cn>.

This report is the outcome of the final project for the *Machine Learning for Earth and Environmental Sciences* course at the University of Lausanne, Autumn Semester 2023.

Forest AGB can be measured directly through harvesting and weighing methods, or indirectly by applying empirical allometric equations that utilize measurements such as diameter at breast height (DBH) and tree height (Lourenço et al., 2021). While these methods are considered precise, they are labor-intensive and thus limited to small-scale plots. Over the past few decades, advancements in remote sensing techniques have opened up possibilities for estimating and mapping forest AGB over large areas. The common methodology is to develop models that correlate measured biomass in plots with remote sensing metrics. These metrics include optical image band values, microwave radar's backscatter coefficients, LiDAR (light detection and ranging) metrics, etc. As the latest remote sensing technology, LiDAR is particularly effective due to its high penetration capability, which allows for the detailed detection of the vertical structure of trees (Yang et al., 2022). This capability to some extent mitigates the saturation problem encountered in high biomass areas when using optical and radar imagery, whereby the signal stops increasing beyond a certain biomass density, limiting the accuracy of biomass estimation (Luo et al., 2019).

Here, two remote sensing datasets are combined to map forest AGB in Northeast China for the year 2022: The Global Ecosystem Dynamics Investigation (GEDI) LiDAR-based L4A footprint aboveground biomass density (AGBD) product and The Harmonized Sentinel-2 MultiSpectral Instrument (MSI) L2A optical image product. We employ several machine learning methods to establish a relationship between the auto-labeled GEDI footprints' AGBD and corresponding Sentinel-2 band values as well as the geographic coordinates (latitude and longitude). After training, the model is capable of predicting AGBD for each forested pixel. In this way, we can finally produce a high resolution, wall-to-wall forest AGB map covering large areas.

2. Datasets

2.1. Study area

Northeast China, spanning from 38.72°N to 53.56°N and 115.52°E to 135.09°E, encompasses Heilongjiang, Jilin, and Liaoning Provinces, as well as four municipalities in the

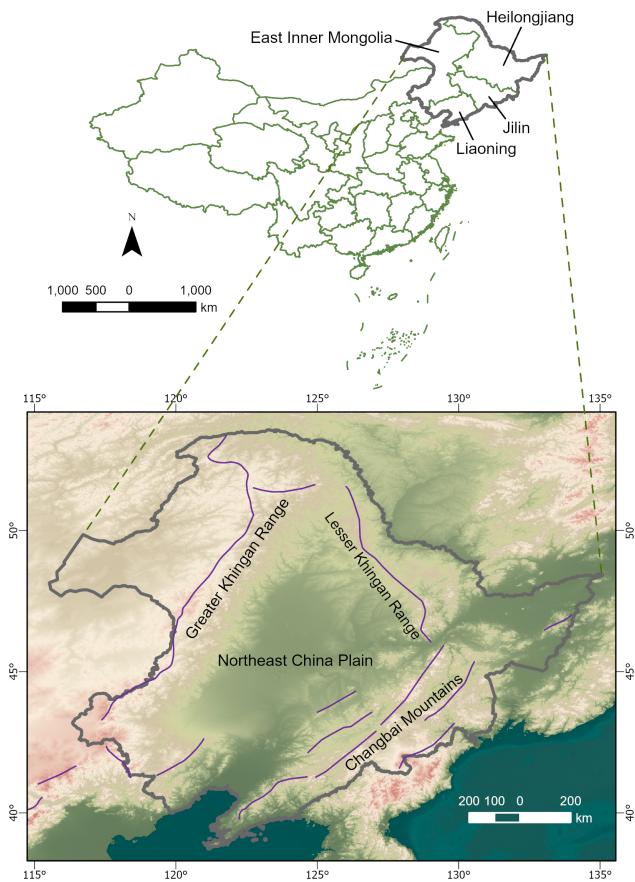


Figure 1. Location and topography of Northeast China.

eastern part of the Inner Mongolia Autonomous Region (Fig. 1). This region, covering an area of approximately $1.25 \times 10^6 \text{ km}^2$ (Wang et al., 2022), experiences primarily temperate monsoon and temperate continental climates. The mean annual temperature ranges from -5.6°C to 10°C and the annual precipitation ranges from 250 mm to 1200 mm (Qian et al., 2003). Forests, covering around 40% of the land area, are predominantly found in the Greater Khingan Range in the northwest, the Lesser Khingan Range in the northeast, and the Changbai Mountains in the southeast. The region is characterized by three main forest types: deciduous coniferous forests, mixed evergreen coniferous-deciduous broad-leaved forests, and deciduous broad-leaved forests (Qian et al., 2003). Given the significant impact of human activities and climate change on these forests, understanding the actual distribution of forest AGB is crucial for effective forest management and achieving carbon neutrality goals in Northeast China.

2.2. GEDI data processing

The GEDI L4A product for Northeast China, covering the period from April to September 2022, is accessed via NASA's Harmony API. Using a full waveform LiDAR system that operates from the International Space Station, GEDI produces measurements of forest structure and determines AGBD in 25-m radius footprints based on established models (Duncanson et al., 2022). Within the L4A product, several variables are selected for model training and analysis. These include "agbd", which provides the AGBD for each footprint and serves as a label for the model; "lat_lowestmode" and "lon_lowestmode", indicating the latitude and longitude coordinates of each footprint, essential for spatial analysis; as well as "algorithm_run_flag", "l2_quality_flag", and "l4_quality_flag". The latter three are critical for filtering out footprints of lower quality, ensuring the reliability and accuracy of the model's predictions.

Initially, GEDI footprints that have "algorithm_run_flag", "l2_quality_flag", or "l4_quality_flag" set to zero are excluded to ensure data quality. Next, we eliminate footprints not located within forested areas. This is achieved by overlaying the footprints onto a 10-m resolution land use land cover (LULC) map based on Sentinel-2 (Karra et al., 2021) for the year 2022. Specifically, a buffer zone with a radius of 25-m is established for central points of each footprint. A footprint is retained only if all pixels in the LULC map within the buffer zone are classified as "Trees". We choose the radius of 25-m instead of 12.5-m, which is the actual extant of footprints, for the reason that GEDI footprints have a horizontal geolocation error of 10.3-m (1σ) (Oliveira et al., 2023) and larger sampling area is more likely to overlap the true footprint location (Campbell et al., 2021).

At last, we address the issue of anomalously high AGBD values observed in a very small subset of the footprints, with some exceeding 1000 Mg/ha. These values are counterintuitive when compared to other AGBD products (Yang et al., 2023; Spawn et al., 2020). To maintain data integrity, we restrict our analysis to footprints with AGBD values within the 1st to the 99th percentile range. This decision is based on the need to exclude extreme outliers that are not representative of the typical biomass density in the region. After this filtration process, a total of 5,602,097 footprints remain, with AGBD values ranging from 15 to 320 Mg/ha.

2.3. Sentinel-2 data processing

The Harmonized Sentinel-2 MSI L2A product for Northeast China, covering the period from April to September 2022, is accessed via Google Earth Engine platform. This L2A product has already undergone atmospheric correction (Guerra-Hernández et al., 2022). To ensure data quality, the QA60 band is used to mask clouds, and a median composite method is applied to fill in the gaps. We utilize all 12 optical

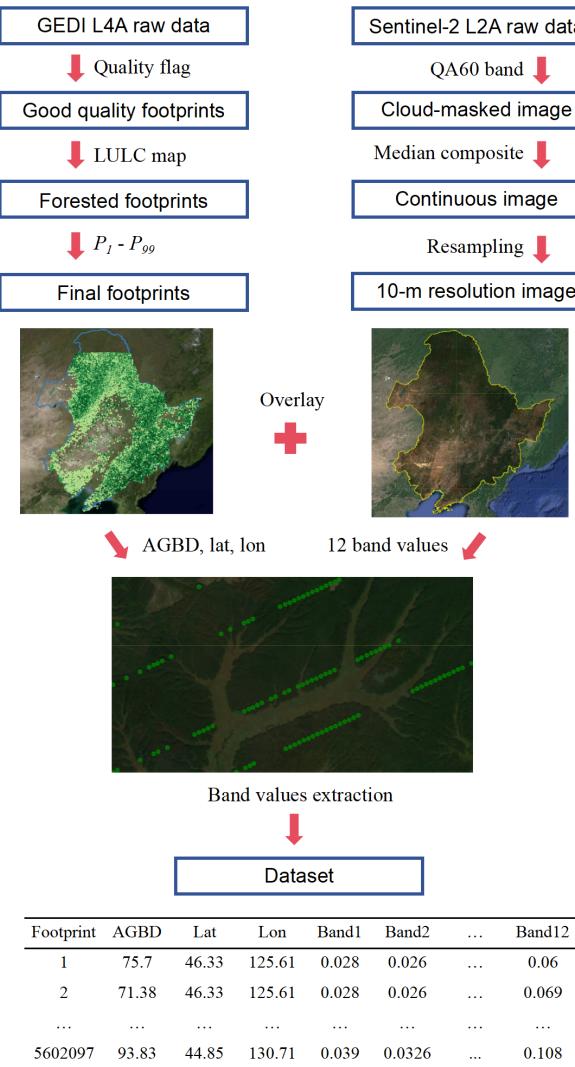


Figure 2. Overall workflow of data processing.

image bands as predictive variables for the model. These include bands B2, B3, B4, B8 with 10-m resolution, and bands B5, B6, B7, B8a, B11, B12 with 20-m resolution, as well as band B1, B9 with 60-m resolution. The bands with 20-m and 60-m resolution are resampled to 10-m in order to align all the bands (Lang et al., 2023).

In the final step, each GEDI footprint obtained in Section 2.2 is matched with Sentinel-2 band values. To achieve this, we calculate the average of all valid pixels within the 25-m radius buffer zone of each footprint. Here, a pixel is deemed “valid” if the buffer zone encompasses the pixel’s central point. As a result, we construct a dataset comprising 5,602,097 samples. Each sample includes the GEDI AGBD value as the label and a combination of latitude, longitude,

Table 1. Hyperparameter search for Light GBM model using Bayesian optimization (init_points=50, n_iter=100).

Hyperparameter	Range	Best
objective	quantile	–
alpha	0.5	–
n_estimators	50-2000	1261
max_depth	100-1000	756
num_leaves	1000-10000	6499
min_child_samples	10-200	52
colsample_bytree	0.5-1	0.64
subsample	0.5-1	0.88
learning_rate	0.01-0.5	0.04

and the 12 Sentinel-2 band values as predictive variables. The overall workflow of data processing is shown in Fig. 2.

3. Methods

3.1. Random Forest and Light GBM

Machine learning techniques have become increasingly popular for modeling the relationship between remote sensing features and AGB measurements, with the Random Forest algorithm being one of the most effective methods for handling high-dimensional data (Yang et al., 2023). The whole dataset is divided randomly to training set (80%) and test set (20%). During training, we only set “n_estimators” to 50 and keep other parameters at their default settings to allow the model to learn the characteristics in the training set as much as possible.

In contrast, we also apply Light Gradient Boosting Machine (GBM) model with several regularization hyperparameters. Light GBM is also a decision tree-based machine learning technique and it is highly efficient for handling large datasets (Ke et al., 2017). Bayesian optimization is used to select hyperparameters that yield the most accurate LightGBM model (Shendryk, 2022). The dataset is divided in the same manner as for the Random Forest model, and we use the best parameters in Table 1 to train Light GBM model.

3.2. Deep Neural Networks

Compared to classic machine learning methods, deep learning techniques such as Artificial Neural Network (ANN) could offer enhanced capabilities in detecting complex feature relationships. Multilayer Perceptron (MLP) is a basic form of ANN, which is composed of interconnected neurons (Gardner & Dorling, 1998). In our study, five dense layers constitute the MLP, with the number of neurons being 128, 64, 32, 16, 1 respectively. Except for the final output layer, which uses “Linear” as its activation function, all other layers use “ReLU” as their activation function. The dataset is randomly divided to training set (70%), validation set (15%)

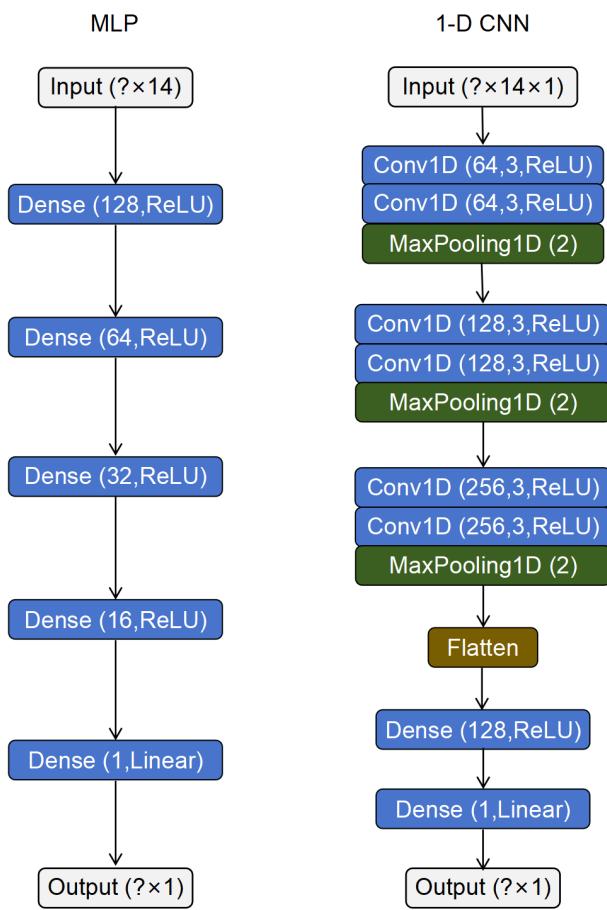


Figure 3. The architecture of MLP and 1-D CNN used in this study. Note that in the Dense layer, the parameters in parentheses represent units and activation function; in the Conv1D layer, the parameters represent filters, kernel size, and activation function; in the MaxPooling1D layer, the parameter represents pool size; in the Input and Output layers, the parameter represents data dimensions.

and test set (15%). Before training, the input variables are normalized using a standard scaler to improve model convergence. As for hyperparameters, we set optimizer as “Adam”, loss function as “MSE”, training epochs to 100, batch size to 256, and patience of callbacks to 10.

In addition, 1-Dimensional Convolutional Neural Network (1-D CNN) is applied. CNN is a type of ANN which is designed originally to process images (Kiranyaz et al., 2021), and we use the 1-D version because the pixel values corresponding to footprints are averaged in each of the 12 bands. The network comprises six convolutional layers, arranged in three pairs, with each pair followed by a maxpooling layer. The convolutional pairs feature 64, 128, and 256 filters, respectively, all with a kernel size of 3 and employing the ReLU activation function. The pooling layers have a

pool size of 2. Following the convolutional layer groups, there is a flatten layer to convert the data into a 1-D format, succeeded by a dense layer with 128 neurons using the ReLU activation function. The final output layer consists of a single neuron with a Linear activation function. The approach to dataset splitting, the scaling technique for input normalization, and the hyperparameters are consistent with those used in the MLP approach. The architecture of the two Neural Networks is shown in Fig. 3.

3.3. Evaluation metrics

Two metrics are employed in this study to evaluate the performance of models: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). RMSE is calculated as the square root of the average of the squared differences between the predicted and actual values (Equation 1) and is more sensitive to large errors. MAE is defined as the average of the absolute differences between the observed and predicted values (Equation 2), which assigns equal weight to all errors. Lower RMSE and MAE present higher model performance and combining the two metrics could lead to more comprehensive analysis of models’ errors.

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

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

4. Results

4.1. Model performance

The evaluation metrics of the four models used in this study is shown in Table 2 and the confusion plots clearly illustrate the varying performance of the models on the test set (Fig. 4). Random Forest model exhibits the lowest RMSE, demonstrating a good overall fit. Light GBM model records the lowest MAE, but it tends to underestimate AGBD within the range of 100-200 Mg/ha. MLP exhibits the highest RMSE and MAE, revealing significant issues with underestimation. 1-D CNN performs better than MLP, but tends to overestimate AGBD. Overall, Random Forest model demonstrates the best agreement between predictions from Sentinel-2 and GEDI reference. Therefore, we choose Random Forest model for further analysis of feature importance and for mapping the forest AGB.

4.2. Feature importance

We analyse both the intrinsic feature importance and permutation feature importance of the Random Forest model

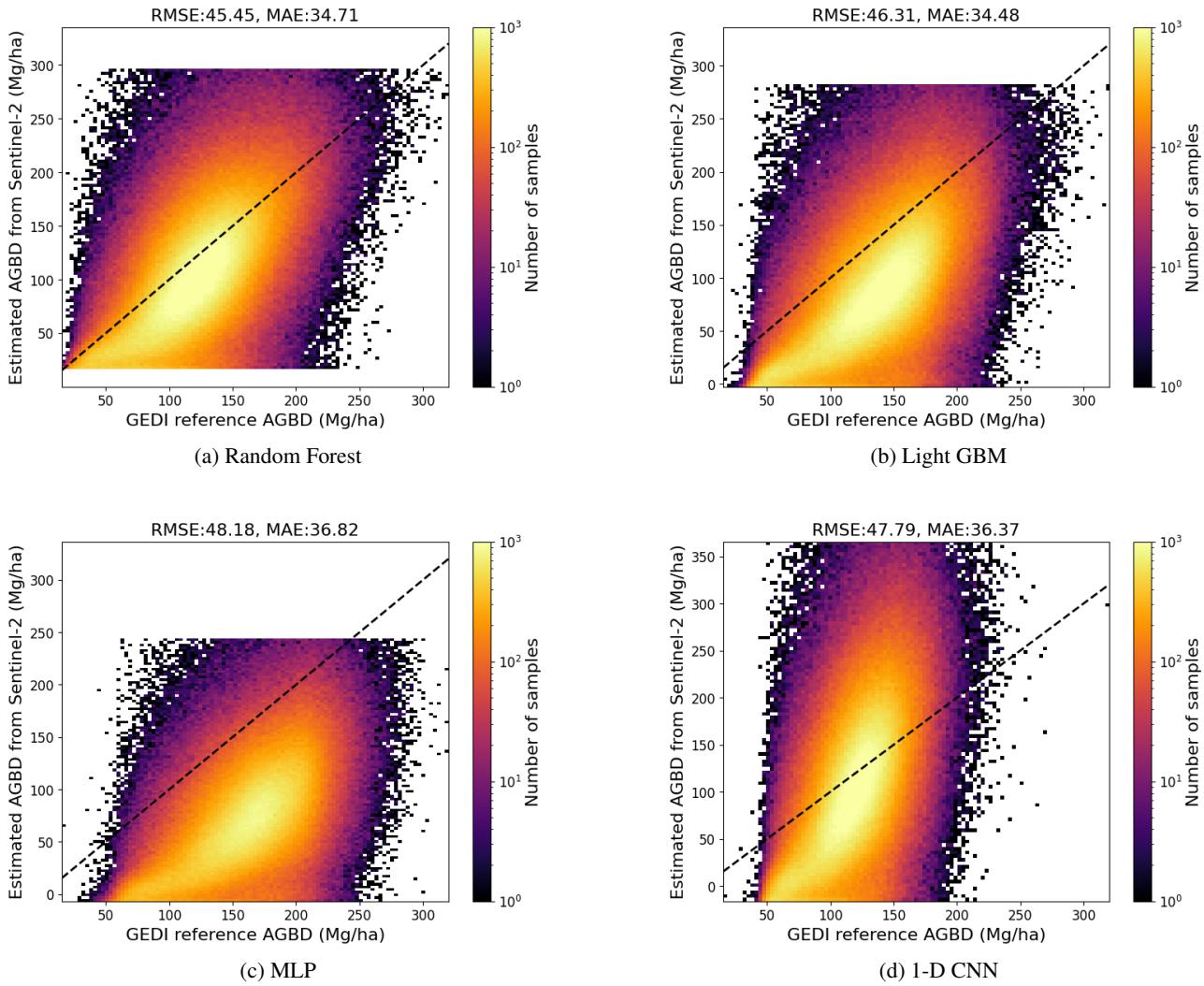


Figure 4. Confusion plots for the models on the test set.

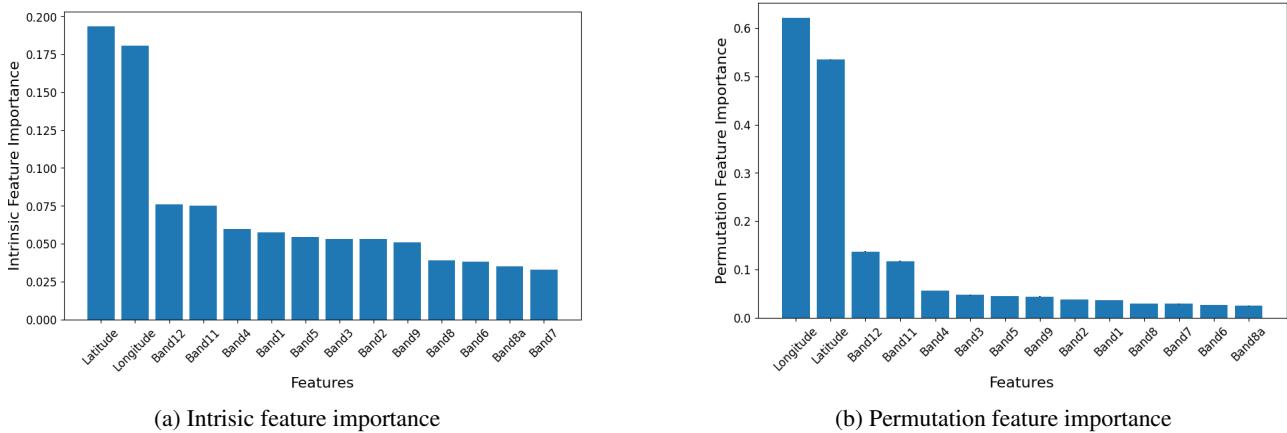


Figure 5. Feature importance scores for the Random Forest model.

Table 2. Model performance comparison on the train, validation and test set. Note that retraining Neural Networks may result in minor variations in evaluation metrics.

Model	Train set		Validation set		Test set	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
Random Forest	17.33	13.02	—	—	45.45	34.71
Light GBM	39.71	26.45	—	—	46.31	34.48
MLP	48	36.71	48.15	36.82	48.18	36.85
1-D CNN	46.18	35.18	47.7	36.27	47.76	36.29

(Fig. 5). Intrinsic feature importance score is calculated by quantifying the contribution of each feature in decreasing the Gini impurity when it's used to split nodes across all trees (Géron, 2019), while permutation feature importance score is calculated on the test set by shuffling feature values to evaluate their impact on model performance (Gagne et al., 2019). Although there are some minor differences in the rankings obtained from these two methods, the overall assessment of feature significance remains consistent: latitude and longitude are the most important features, followed by band 12 and band 11 values, while other features are less important. The high importance of coordinates of footprints indicates the model's ability to discern spatial patterns in AGBD. This observation resonates with the First Law of Geography, which posits that closer entities are more closely related than distant ones. In addition, band 12 and band 11 are short wave infrared bands. Their prominence among all the bands presents a strong correlation between AGBD and short wave infrared band values.

4.3. Forest AGBD map

Finally, we produce a 10-m resolution forest AGBD map of Northeast China for the year 2022 (Fig. 6) by employing the Random Forest model. For each forested pixel, its latitude and longitude coordinates along with the 12 Sentinel-2 band values are inputted into the model to predict AGBD. The non-forested pixels are marked with NAN values. The model's predictions for forest AGBD range from 16 to 299 Mg/ha, averaging at 118 Mg/ha. Notably, the Greater Khingan Range, Lesser Khingan Range, and Changbai mountains areas are identified as regions with higher AGBD values. In contrast, the central and western plains and hilly regions of Northeast China exhibit lower AGBD values. The Changbai Mountains in the southeast, benefiting from their lower latitude and position on the windward slope of the monsoon, exhibit the highest AGBD values in the region. This is attributed to the favorable conditions for forest growth provided by ample heat and abundant rainfall in these areas. Furthermore, we address the absence of GEDI AGBD observations in areas north of 51.6°N by utilizing predictions based on Sentinel-2 data. Overall, this map finely presents the distribution of forest AGB of Northeast China, providing

valuable references for forest management at various scales.

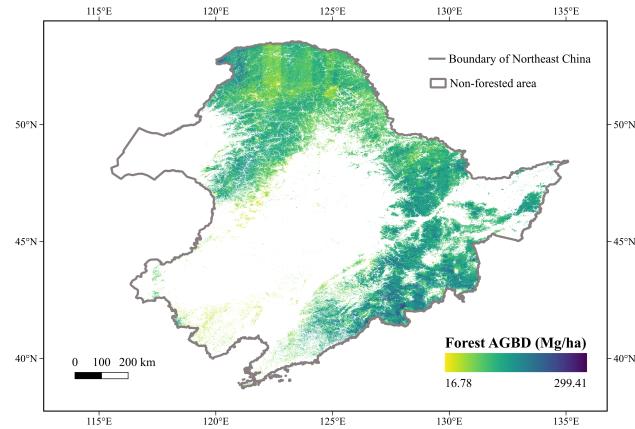


Figure 6. 10-m resolution forest aboveground biomass map of Northeast China for the year 2022.

5. Discussion and conclusion

While integrating sparse GEDI AGBD labels with continuous Sentinel-2 optical imagery enables the production of a high-resolution, comprehensive AGB map, this method presents several limitations. Firstly, a “strip effect” (Liu et al., 2022) is observable in the map, particularly in northern areas lacking GEDI footprints. This refers to uneven, band-like color variations, manifesting in our study as inconsistent AGBD values along boundary lines. This effect results from the use of Sentinel-2 image tiles composed at different times and may indicate the degradation of model performance in these areas (Lang et al., 2023). Secondly, the accuracy of the GEDI L4A footprint level AGBD product, which forms the model's label, is critical. Yet, this product tends to overestimate AGBD, as highlighted in Section 2.2, often showing values exceeding 200 Mg/ha. This overestimation might be linked to the absence of reference airborne LiDAR data during GEDI's AGBD model construction in Continental Asia (Dubayah et al., 2022). Thirdly, the mismatch in spatial scales between GEDI footprints (25-m diameter) and Sentinel-2 pixels (10/20/60-m) complicates

dataset alignment, potentially affecting model performance. Additionally, the geolocation error in GEDI footprints is another significant source of uncertainty. Fourthly, the method assumes consistent phenological characteristics of forest vegetation from April to September. However, variations in the acquisition times between GEDI footprints and Sentinel-2 pixels over these six months could introduce errors. Lastly, the combination of optical imagery and LiDAR does not completely eliminate the saturation problem of optical images in predicting AGBD. Therefore, this method may have greater errors in areas with high AGBD values.

In future developments, our aim is to optimize the modeling methodology to improve its robustness and accuracy. One approach involves incorporating additional auxiliary data as predictive variables, such as elevation, temperature, and precipitation. These variables are closely related to AGB and can provide more insights, despite the challenge that some datasets may not be available at the desired 10-m resolution. Other remote sensing datasets such as Sentinel-1 radar images could also be considered. Another crucial step is the creation of an uncertainty map, which can be accomplished by initializing the model multiple times with different parameters. What's more, to evaluate the quality of the AGB map more comprehensively, we could compare it with independent AGB measurements, such as estimations from field plots, LiDAR retrieval results, and previous AGB products. In addition, the methodology of this study could be applied on a broader spatio and temporal scale, given the fact that GEDI L4A product is available from April 2019 to March 2023 between 51.6°N and 51.6°S, while Sentinel-2 L2A product is available from March 2017 globally and is still upgrading. An annual time series AGB map spanning a larger area can be produced.

To conclude, this study successfully demonstrates the capability of combining GEDI footprint AGBD with Sentinel-2 spectral bands through machine learning techniques to generate a continuous and high-resolution forest AGB map in large areas. The potential to extend this methodology to various spatial and temporal scales can greatly assist in monitoring forests' dynamics and in formulating policies concerning carbon emission reduction and climate change mitigation. Furthermore, this research highlights the value of the integration of diverse data sources for enhanced environmental monitoring and management, which is particularly useful in an era with growing abundance and richness of remote sensing data.

Data and code availability

GEDI L4A data: https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=2056.

Sentinel-2 L2A data: <https://developers.google.com/earth-engine/datasets/>

[catalog/COPERNICUS_S2_SR_HARMONIZED](#).
Sentinel-2 LULC map: <https://livingatlas.arcgis.com/landcoverexplorer>.

The dataset after filtration and codes for data downloading, data processing, model training and map generation: https://github.com/ChguFan/2023_ML_EES/tree/main/FinalProject.

Appendix

Table A1. Abbreviations used in this study.

Abbreviation	Description
1-D CNN	1-Dimensional Convolutional Neural Network
AGB	Aboveground biomass
AGBD	Aboveground biomass density (Mg/ha)
ANN	Artificial Neural Network
DBH	Diameter at breast height
GEDI	Global Ecosystem Dynamics Investigation
LiDAR	Light detection and ranging
Light GBM	Light Gradient Boosting Machine
LULC	Land use land cover
MAE	Mean Absolute Error
MLP	Multilayer Perceptron
MSE	Mean Square Error
MSI	Multispectral Instrument
ReLU	Rectified Linear Unit
RMSE	Root Mean Square Error

References

- Ashton, P. M. S. *Managing forest carbon in a changing climate*. Springer Dordrecht, 2012. doi: 10.1007/978-94-007-2232-3. URL <https://doi.org/10.1007/978-94-007-2232-3>.
- Campbell, M. J., Dennison, P. E., Kerr, K. L., Brewer, S., and Anderegg, W. R. L. Scaled biomass estimation in woodland ecosystems: Testing the individual and combined capacities of satellite multispectral and lidar data. *Remote Sensing of Environment*, 262:112511, 9 2021. doi: 10.1016/j.rse.2021.112511. URL <https://doi.org/10.1016/j.rse.2021.112511>.
- Dubayah, R., Armston, J., Healey, S. P., Bruening, J. M., Patterson, P. L., Kellner, J. R., Duncanson, L., Saarela, S., Ståhl, G., Yang, Z., Tang, H., Blair, J. B., Fatoyinbo, L., Goetz, S. J., Hancock, S., Hansen, M. C., Hofton, M. A., Hurt, G. C., and Luthcke, S. B. GEDI launches a new era of biomass inference from space. *Environmental Research Letters*, 17(9):095001, 8 2022. doi: 10.1088/1748-9326/ac8694. URL <https://doi.org/10.1088/1748-9326/ac8694>.
- Duncanson, L., Kellner, J. R., Armston, J., Dubayah, R., Minor, D., Hancock, S., Healey, S. P., Patterson, P. L.,

- Saarela, S., Marselis, S., Silva, C. E., Bruening, J. M., Goetz, S. J., Tang, H., Hofton, M. A., Blair, B., Luthcke, S. B., Fatoyinbo, L., Abernethy, K., Alonso, A., Andersen, H., Aplin, P., Baker, T. R., Barbier, N., Bastin, J. F., Biber, P., Boeckx, P., Bogaert, J., Boschetti, L., Boucher, P., Boyd, D. S., Burslem, D. F. R. P., Calvo-Rodriguez, S., Chave, J., Chazdon, R. L., Clark, D. B., Clark, D. A., Cohen, W. B., Coomes, D. A., Corona, P., Cushman, K. C., Cutler, M. E. J., Dalling, J. W., Dalponte, M., Dash, J. P., de-Miguel, S., Deng, S., Ellis, P. W., Erasmus, B., Fekety, P. A., Fernández-Landa, A., Ferraz, A., Fischer, R., Fisher, A., García-Abril, A., Gobakken, T., Hacker, J., Heurich, M., Hill, R. A., Hopkinson, C., Huang, H., Hubbell, S. P., Hudak, A. T., Huth, A., Imbach, B., Jeffery, K. J., Katoh, M., Karsley, E., Kenfack, D., Kljun, N., Knapp, N., Král, K., Krůček, M., Labrière, N., Lewis, S. L., Longo, M., Lucas, R., Main, R., Manzanera, J. A., Martínez, R. V., Mathieu, R., Memiaghe, H. R., Meyer, V., Mendoza, A. M., Monerris, A., Montesano, P., Morsdorf, F., Næsset, E., Naidoo, L., Nilus, R., O'Brien, M. J., Orwig, D. A., Papathanassiou, K., Parker, G. G., Philipson, C. D., Phillips, O. L., Písek, J., Poulsen, J. R., Pretzsch, H., Rüdiger, C., Saatchi, S., Sánchez-Azofeifa, A., Sánchez-López, N., Scholes, R. J., Da Silva, C. A., Simard, M., Skidmore, A., Stereńczak, K., Tanase, M. A., Torresan, C., Valbuena, R., Verbeeck, H., Vrška, T., Wessels, K. J., White, J. C., White, L., Zahabu, E., and Zgraggen, C. Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission. *Remote Sensing of Environment*, 270:112845, 3 2022. doi: 10.1016/j.rse.2021.112845. URL <https://doi.org/10.1016/j.rse.2021.112845>.
- FAO. *The State of the World's Forests 2022. Forest pathways for green recovery and building inclusive, resilient and sustainable economies.* FAO, 2022. doi: 10.4060/cb9360en. URL <https://doi.org/10.4060/cb9360en>.
- Gagne, D. J., Haupt, S. E., Nychka, D., and Thompson, G. Interpretable deep learning for spatial analysis of severe hailstorms. *Monthly Weather Review*, 147(8):2827–2845, 7 2019. doi: 10.1175/mwr-d-18-0316.1. URL <https://doi.org/10.1175/mwr-d-18-0316.1>.
- Gardner, M. and Dorling, S. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14-15):2627–2636, 8 1998. doi: 10.1016/s1352-2310(97)00447-0. URL [https://doi.org/10.1016/s1352-2310\(97\)00447-0](https://doi.org/10.1016/s1352-2310(97)00447-0).
- Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition.* O'Reilly Media, Inc., 9 2019. URL <https://powerunit-ju.com/>
- wp-content/uploads/2021/04/Aurelien-Geron-Hands-On-Machine-Learning-with-Scikit-Learn-Keras-and-Tensorflow_-Concepts-Tools-and-Techniques-to-Build-Intelligent-Systems-OReilly-Media-2019.pdf.
- Guerra-Hernández, J., Narine, L. L., Pascual, A., González-Ferreiro, E., Botequim, B., Malambo, L., Neuenschwander, A. L., Popescu, S. C., and Godinho, S. Aboveground biomass mapping by integrating ICESat-2, SENTINEL-1, SENTINEL-2, ALOS2/PALSAR2, and topographic information in Mediterranean forests. *Giscience & Remote Sensing*, 59(1):1509–1533, 9 2022. doi: 10.1080/15481603.2022.2115599. URL <https://doi.org/10.1080/15481603.2022.2115599>.
- Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., Mathis, M., and Brumby, S. P. Global land use/land cover with sentinel 2 and deep learning. In *2021 IEEE international geoscience and remote sensing symposium IGARSS*, pp. 4704–4707. IEEE, 2021. doi: 10.1109/igarss47720.2021.9553499. URL <https://doi.org/10.1109/igarss47720.2021.9553499>.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30:3149–3157, 12 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.
- Kiranyaz, S., Avci, O., Abdeljaber, O., İnce, T., Gabbouj, M., and Inman, D. J. 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, 151:107398, 4 2021. doi: 10.1016/j.ymssp.2020.107398. URL <https://doi.org/10.1016/j.ymssp.2020.107398>.
- Lang, N., Jetz, W., Schindler, K., and Wegner, J. D. A high-resolution canopy height model of the Earth. *Nature Ecology and Evolution*, 7(11):1778–1789, 9 2023. doi: 10.1038/s41559-023-02206-6. URL <https://doi.org/10.1038/s41559-023-02206-6>.
- Li, Y., Li, C., Li, M., and Liu, Z. Influence of variable selection and forest type on forest aboveground biomass estimation using machine learning algorithms. *Forests*, 10(12):1073, 11 2019. doi: 10.3390/f10121073. URL <https://doi.org/10.3390/f10121073>.
- Liu, X., Su, Y., Hu, T., Yang, Q., Liu, B., Deng, Y., Tang, H., Tang, Z., Fang, J., and Guo, Q. Neural network guided interpolation for mapping canopy height of China's forests by integrating GEDI and ICESat-2

- data. *Remote Sensing of Environment*, 269:112844, 2 2022. doi: 10.1016/j.rse.2021.112844. URL <https://doi.org/10.1016/j.rse.2021.112844>.
- Lourenço, P., Godinho, S., Sousa, A., and Gonçalves, A. C. Estimating tree aboveground biomass using multispectral satellite-based data in Mediterranean agroforestry system using random forest algorithm. *Remote Sensing Applications: Society and Environment*, 23:100560, 8 2021. doi: 10.1016/j.rsase.2021.100560. URL <https://doi.org/10.1016/j.rsase.2021.100560>.
- Luo, S., Wang, C., Xi, X., Nie, S., Xiu, F., Chen, H. Y. H., Ma, D., Liu, J., Zou, J., Lin, Y., and Zhou, G. Estimating forest aboveground biomass using small-footprint full-waveform airborne LiDAR data. *International journal of applied earth observation and geoinformation*, 83:101922, 11 2019. doi: 10.1016/j.jag.2019.101922. URL <https://doi.org/10.1016/j.jag.2019.101922>.
- Oliveira, P. V. C., Zhang, X., Peterson, B., and Ometto, J. P. Using simulated GEDI waveforms to evaluate the effects of beam sensitivity and terrain slope on GEDI L2A relative height metrics over the Brazilian Amazon Forest. *Science of Remote Sensing*, 7:100083, 6 2023. doi: 10.1016/j.srs.2023.100083. URL <https://doi.org/10.1016/j.srs.2023.100083>.
- Qian, H., Yuan, X.-Y., and Chou, Y.-L. *Forest Vegetation of Northeast China*, pp. 181–230. Springer Netherlands, Dordrecht, 2003. ISBN 978-94-017-0143-3. doi: 10.1007/978-94-017-0143-3_6. URL https://doi.org/10.1007/978-94-017-0143-3_6.
- Shendryk, Y. Fusing GEDI with earth observation data for large area aboveground biomass mapping. *International journal of applied earth observation and geoinformation*, 115:103108, 12 2022. doi: 10.1016/j.jag.2022.103108. URL <https://doi.org/10.1016/j.jag.2022.103108>.
- Spawn, S., Sullivan, C., Lark, T. J., and Gibbs, H. K. Harmonized global maps of above and belowground biomass carbon density in the year 2010. *Scientific Data*, 7(1), 4 2020. doi: 10.1038/s41597-020-0444-4. URL <https://doi.org/10.1038/s41597-020-0444-4>.
- Tian, L., Wu, X. B., Yu, T., Li, M., Qian, C., Liao, L., and Fu, W. Review of Remote Sensing-Based Methods for Forest Aboveground Biomass Estimation: Progress, Challenges, and Prospects. *Forests*, 14(6):1086, 5 2023. doi: 10.3390/f14061086. URL <https://doi.org/10.3390/f14061086>.
- Wang, S., Xu, X., and Huang, L. Spatial and Temporal Variability of Soil Erosion in Northeast China from 2000 to 2020. *Remote Sensing*, 15(1):225, 12 2022. doi: 10.3390/rs15010225. URL <https://doi.org/10.3390/rs15010225>.
- Yang, Q., Su, Y., Hu, T., Jin, S., Liu, X., Niu, C., Z, L., Kelly, M., Wei, J., and Guo, Q. Allometry-based estimation of forest aboveground biomass combining LiDAR canopy height attributes and optical spectral indexes. *Forest Ecosystems*, 9:100059, 1 2022. doi: 10.1016/j.feces.2022.100059. URL <https://doi.org/10.1016/j.feces.2022.100059>.
- Yang, Q., Niu, C., Liu, X., Feng, Y., Ma, Q., Wang, X., Tang, H., and Guo, Q. Mapping high-resolution forest aboveground biomass of China using multisource remote sensing data. *Giscience & Remote Sensing*, 60(1), 4 2023. doi: 10.1080/15481603.2023.2203303. URL <https://doi.org/10.1080/15481603.2023.2203303>.