Assessment of Machine Learning Models for Predicting Aboveground Biomass in the Indian Subcontinent

Shailja Mamgain, Bhoomika Ghale, Harish Chandra Karnatak, Arijit Roy
shailja309@gmail.com, bhoomikaghale20@gmail.com, harish@iirs.gov.in, arijitroy@iirs.gov.in

Indian Institute of Remote Sensing, Indian Space Research Organization (ISRO), Department of Space, 4, Kalidas Road, Dehradun 248001, India

Keywords: Aboveground biomass; Machine learning models; Random Forest; Vegetation indices; Biomass distribution.

Abstract

Understanding the distribution of aboveground biomass (AGB) is vital for evaluating carbon stocks & ecosystem dynamics, especially in regions with diverse landscapes like Indian subcontinent. This study evaluates three machine learning models—Random Forest (RF), Gradient Tree Boosting (GTB), & Classification and Regression Trees (CART)—for predicting AGB across the subcontinent. Independent variable in these models is AGB, while dependent variables include a range of vegetation & topographic layers: Normalized Difference Vegetation Index, Enhanced Vegetation Index, Leaf Area Index, Fraction of Photosynthetically Active Radiation, land cover, elevation, aspect, slope, & hillshade. These predictors are essential for capturing ecological & topographical characteristics that influence biomass distribution. The models were evaluated using coefficient of determination (R^2) & Pearson's correlation coefficient (r) to assess predictive accuracy. RF emerged as most accurate, with an R^2 value of 0.834 & r value of 0.913, effectively capturing the spatial variability in AGB across subcontinent's diverse ecosystems, which was then used to predict AGB for 2023. The predictions reveal significant spatial variation in biomass density, reflecting region's diverse ecological zones & land-use patterns. In India, high biomass densities are found in Himalayan foothills, northeastern states, & Western Ghats, while arid regions like Rajasthan & Gujarat have lower values. Pakistan generally exhibits low biomass densities, with higher values near the northern border with India. Nepal & Bhutan show high densities in their forested regions, particularly in the mid-hills, high mountains, & Eastern Himalaya. Bangladesh has moderate to low biomass densities. In Sri Lanka, central highlands & southwestern rainforests have highest biomass densities, while the more arid northern & eastern regions exhibit lower values. This study highlights the importance of using robust machine learning models like RF to accurately capture spatial patterns of biomass distribution, which is crucial for forest management, carbon accounting, & biodiversity conservation in the Indian subcontinent.

1. INTRODUCTION

Aboveground biomass (AGB) is considered to be one of the important predictors in the estimation of carbon stocks, ecosystem dynamics, and sustainable forest management (Saatchi et. al., 2007). In the climate change scenario, forests play a significant role not only as a potential carbon sequestration site but also in maintaining biodiversity and managing water cycles (Malla et. al., 2023). AGB quantifies the total above-ground mass of living plant material, which includes trees, shrubs, and other vegetation, that reflects the carbon storage potential of forest ecosystems (Das et. al., 2021). Hence, estimating accurate biomass is a crucial concern in forest management, climate change mitigation strategies, and biodiversity conservation.

The diversity of landscape, from the Himalaya to the arid plains of Rajasthan, or from the tropical rainforests of the Western Ghats to the coastal mangroves of Sundarbans, make the Indian subcontinent an unlike challenge and opportunity for AGB estimation (Das et. al., 2024). These variabilities in geography result in a diversity of ecological zones with varied types of vegetation and biomass densities. The traditional methods of estimation of biomass, based on ground surveys, are tedious and extensive, especially in such heterogeneous landscapes. Remote sensing combined with machine learning techniques offers a powerful alternative for generating spatially continuous biomass estimates (Vorster et. al., 2020).

Machine learning models have gained prominence in ecological applications due to their ability to handle large datasets with complex interactions (Vorster et. al., 2020, Nguyen et. al., 2020; Fararoda et. al., 2021). Three of the widely used machine learning models, namely Random Forest (RF), Gradient Tree Boosting (GTB), and Classification and Regression Trees (CART) have been applied to predict AGB for Indian subcontinent, so that the difference in the approaches towards tree-based decision-making can be compared from all of these methods: ensemble learning in RF, boosting methods in GTB, and a much simpler and more interpretable method in CART. A comparison of these models will bring out the most robust tool for biomass estimation across the varied ecosystems of this subcontinent (Fararoda et. al., 2021).

The commonly used predictors of AGB in various machine learning models are the vegetation indices, with Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Leaf Area Index (LAI), and Fraction of Photosynthetically Active Radiation (FPAR) constituting some of them (Hofhansl et. al., 2020). These indices provide information on vegetation greenness, canopy structure, and photosynthetic activity, which are critical proxies for biomass. Additionally, elevation, aspect, slope, and hillshade are some of the topographic factors that may significantly influence biomass distribution through changing local microclimates, moisture availability, and characteristics of the soil.

In this study, machine learning models including RF, GTB, and CART models have been used for predicting the AGB over six countries in the Indian subcontinent, namely, India, Pakistan, Nepal, Bhutan, Bangladesh, and Sri Lanka, using both vegetation indices and topographic variables. This performance is assessed by the coefficient of determination (R^2), and Pearson's correlation coefficient for identification of the most accurate and reliable model for biomass estimation in this diversified region.

2. STUDY AREA

The Indian subcontinent ranges from India, Pakistan, Nepal, Bhutan, Bangladesh, and Sri Lanka (Figure 1). These regions showcase vast ecological and climatic variations, mainly due to differences in topography, latitude, and monsoon-driven precipitation patterns (Roy, 2014). The subcontinent can broadly be divided into the following geographic regions:

• **Himalayan Region:** The northern boundary of the subcontinent is defined by the Himalaya, the tallest mountain range in the world. This region includes the countries of Nepal, Bhutan, and parts of northern India. The Himalaya contain vast tracts of forest, primarily temperate and subalpine, which are significant carbon sinks.

• **Western Ghats:** Along the western coast of India, lies Western Ghats. It consists of tropical and subtropical rainforests that constitute a biodiversity hotspot. High biomass density is found in this region, especially in montane forests.

• **Eastern Ghats and Deccan Plateau:** The Eastern Ghats run parallel to the Bay of Bengal on the eastern coast of India. Vegetation types range from dry deciduous to moist deciduous forests. Dry forests and scrublands dominate the very large Deccan Plateau, which covers much of southern India.

• **Thar Desert and Arid Regions:** The Thar Desert in western India and southeastern Pakistan is characterized by low biomass density due to its arid conditions. Low biomasses, in turn, are indicative of low precipitation and sparse vegetation, showing region's relatively low carbon sequestration potential.

• **Sundarbans**: The Sundarbans mangrove forest constitutes one of the largest mangrove ecosystems on the India-Bangladesh border. It provides vital ecosystem services, including coastal protection and carbon sequestration, even though it has relatively low biomass compared with most terrestrial forests.

• **Sri Lanka:** This island nation encompasses varied ecosystems ranging from the central highlands' tropical rainforests to the dry zone forests found in the northern and eastern regions.

These distinct ecological zones offer both new opportunities and challenges to biomass estimation as these ranges from various climatic conditions, topography, to vegetation cover. AGB prediction across various landscapes is crucial in understanding regional carbon stocks for guiding conservation strategy inputs.

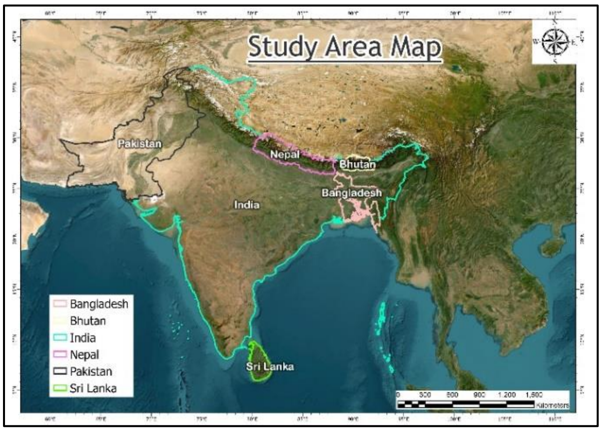

Figure 1. Study area map- Indian Subcontinent

3. DATA USED

In this study, satellite derived vegetation and topographic information have been utilized to predict AGB in the Indian subcontinent (Table 1). The datasets could be broadly categorized into two types: vegetation indices and topographic variables.

3.1 Global Aboveground and Belowground Biomass Carbon Density Maps

Global Aboveground and Belowground Biomass Carbon Density Map for the year 2010 has been used for providing the training inputs and testing data for the machine learning models (Figure 2). This dataset provides temporally consistent and harmonized global maps of aboveground and belowground biomass carbon density at a 300-m spatial resolution.

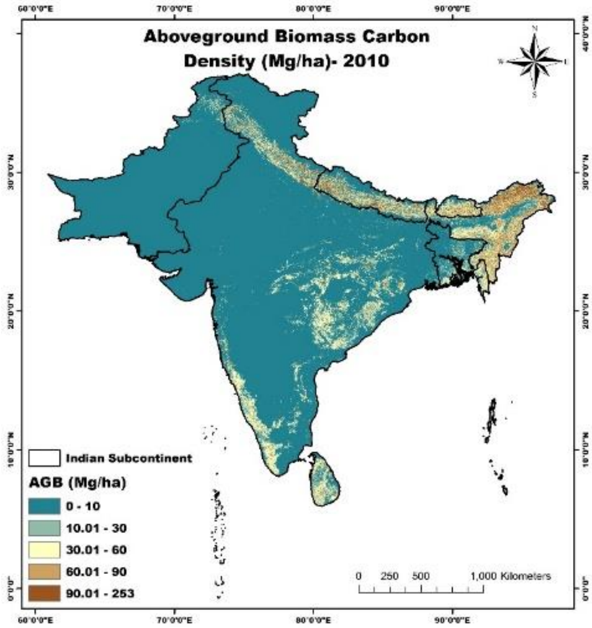

Figure 2. Aboveground Biomass Carbon Density Map-2010

3.2 Vegetation Indices

• **Normalized Difference Vegetation Index (NDVI):** NDVI is a widely used proxy for vegetation greenness and biomass. In this study MODIS NDVI product has been used. NDVI measures the difference between near-infrared and red reflectance, which gives information on the health and density of vegetation (Huete et. al., 1997).

• **Enhanced Vegetation Index (EVI):** Also derived from MODIS, EVI improves upon NDVI by reducing sensitivity to atmospheric conditions and saturation in dense vegetation. It provides additional information on canopy structure (Tucker et. al., 1986).

• **Leaf Area Index (LAI):** LAI is a measure of the leaf area per unit ground area and gives information regarding canopy density and amount of cover of the vegetation. It is another essential variable for estimating AGB (Chen et. al., 1996).

• **Fraction of Photosynthetically Active Radiation (FPAR):** FPAR measures the fraction of incoming solar radiation absorbed by vegetation for photosynthesis. The variable is strongly correlated with vegetation productivity and biomass (Di Bella et. al., 2004).

3.3 Topographic Variables

• **Elevation**: Elevation data were obtained from the Shuttle Radar Topography Mission (SRTM). Elevation conditions affect the temperature and precipitation profile and soil-related properties, which, in turn, affect the distribution of vegetation.

• **Slope and Aspect**: Slope and aspect were derived from the elevation data. Slope affects soil moisture retention and erosion, while aspect affects exposure to solar radiation.

• **Hillshade**: Hillshade, derived from elevation, accounts for the shading effects of the terrain. It gives information about potential changes in microclimate due to topography.

These datasets were preprocessed using standard techniques to ensure consistency and accuracy. The spatial resolution of the data ranged from 250 meters for MODIS-derived variables to 90 meters for topographic variables.

Table 1. Details of the datasets used in the study

S. no.	Variable Name	Spatial resolution (m)	Temporal resolution	Source	Link
1	**Global Aboveground and Belowground Biomass Carbon Density Maps**	300	static (2010)	NASA	https://developers.google.com/earth-engine/datasets/catalog/NASA_ORNL_biomass_carbon_density_v1
2	**NDVI**	250	16-day composite	MODIS	https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD13Q
3	**EVI**	250	16-day composite	MODIS	https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD13Q1
4	**Fraction of Photosynthetically Active Radiation**	500	8-day composite	MODIS	https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD15A2H
5	**Leaf Area Index**	500	8-day composite	MODIS	https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD15A2H
6	**Land Cover Type**	500	yearly	MODIS	https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MCD12Q1
7	**Elevation**	90	--	SRTM	https://developers.google.com/earth-engine/datasets/catalog/CGIAR_SRTM90_V4
8	**Slope**	90	--	SRTM	https://developers.google.com/earth-engine/datasets/catalog/CGIAR_SRTM90_V4
9	**Aspect**	90	--	SRTM	https://developers.google.com/earth-engine/datasets/catalog/CGIAR_SRTM90_V4
10	**Hillshade**	90	--	SRTM	https://developers.google.com/earth-engine/datasets/catalog/CGIAR_SRTM90_V4

4. METHODOLOGY

4.1 Overview of Machine Learning Models

Three machine learning models- RF, GTB and CART were used in this study to predict AGB over the Indian subcontinent (Figure 3). The application of all these models can be associated with a tree-based method but differ in its approach to decision-making and prediction.

4.1.1 Random Forest (RF)

RF is an ensemble learning method which builds a set of decision trees during the training process (Breiman, 2001). Each tree is trained on a random subset of the data, and the final prediction is made by averaging the predictions of all trees. RF greatly reduces the possibility of overfitting, which makes it a more robust model for complex ecological data.

4.1.2 Gradient Tree Boosting (GTB)

GTB is yet another ensemble technique that develops trees sequentially (Friedman, 2001). In this respect, the focus of the model at every iteration is on rectifying the mistakes made by the previously developed trees. Finally, the best model is an additive combination of these trees weighted with the accuracy of each tree while predicting AGB. This technique of CART is particularly efficient for complex nonlinear relationships among the variables

4.1.3 Classification and Regression Trees (CART)

CART is a single decision tree model that splits the data based on the most informative variables (Quinlan, 1986). It is a far simpler and interpretable model compared to RF and GTB but tends to overfit the complex data. CART is useful for identifying the most important variables in predicting AGB but may struggle with large, heterogeneous datasets like those used in this study.

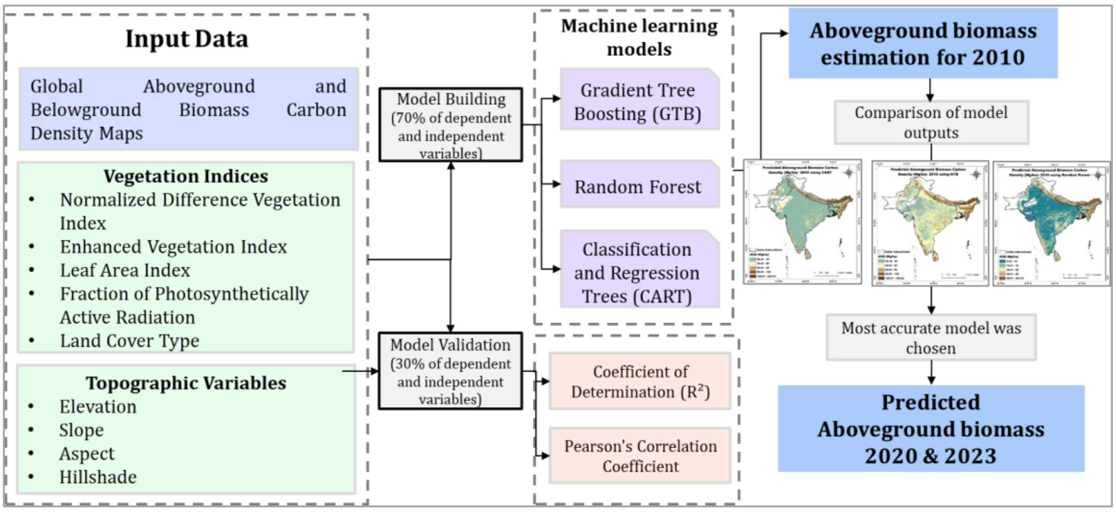

Figure 3. Methodology flow chart for the study

4.2 Model Training and Evaluation

The entire dataset was split into training (70%) and testing (30%) subsets. Every model was trained on the training set, and hyperparameters were optimized using cross-validation for the model. The hyperparameters included the number of trees, depth of trees, and the learning rate (for GTB). Two different metrics applied to evaluate the performance of every model:

• Coefficient of Determination (R^2): R^2 measures the proportion of variance in the dependent variable (AGB) explained by the model. A higher R^2 indicates better predictive accuracy (Tjur, 2009).
• Pearson's Correlation Coefficient: This metric assesses the linear correlation between predicted and actual AGB values. A higher Pearson's correlation coefficient indicates a stronger agreement between the model's predictions and the true values (Cohen et. al., 2009).

The model with the highest R^2 and Pearson's correlation coefficient was selected as the best model for predicting AGB across the subcontinent.

5. RESULTS

The performances of different machine learning models were evaluated in terms of their ability to predict AGB across the Indian subcontinent.

5.1 Random Forest (RF)

The RF model achieved the highest accuracy among the three models, with an R^2 of 0.834 and a Pearson correlation coefficient of 0.913. The model was particularly effective in regions with high biomass densities, such as the Western Ghats and northeastern India. However, it slightly underestimated biomass in the Himalayan region, likely due to the complex terrain and variation in vegetation types (Figure 4).

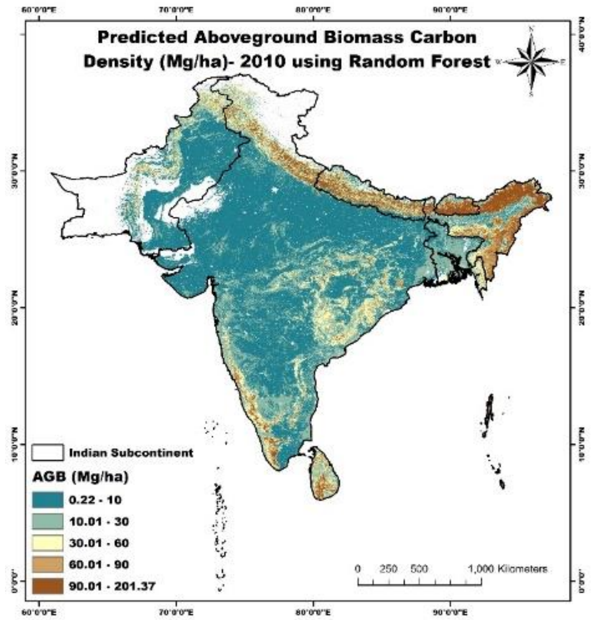

Figure 4. Predicted Aboveground Biomass Carbon Density Map-2010 using Random Forest

The spatial distribution of predicted AGB closely matched known patterns of biomass distribution, with high AGB values in forested regions and low values in arid zones.

5.2 Gradient Tree Boosting (GTB)

While GTB captured the overall patterns of biomass distribution, it struggled in regions with extreme variability, such as the transition zones between the Western Ghats and the Deccan Plateau (Figure 5). In these regions, GTB tended to overestimate biomass, particularly in areas with sparse vegetation cover. The R^2 values for the GTB model were 0.783, and its Pearson correlation coefficient was 0.885.

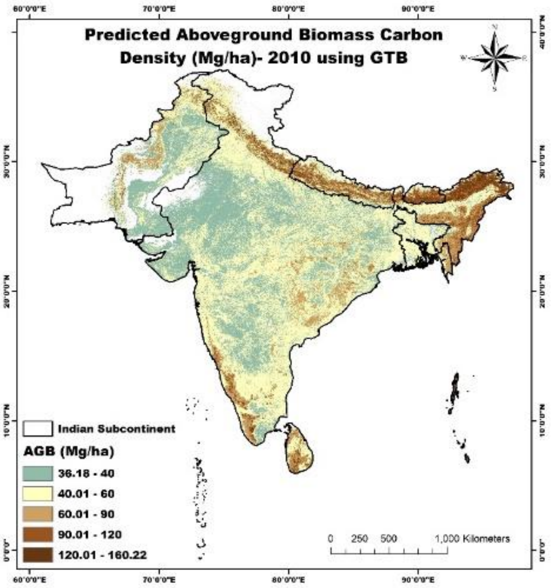

Figure 5. Predicted Aboveground Biomass Carbon Density Map-2010 using GTB

Despite these limitations, GTB showed good predictive performance in regions with moderate to high biomass density, such as the tropical forests of Sri Lanka and the Sundarbans mangrove forests.

5.3 Classification and Regression Trees (CART)

The CART model was the least accurate, with an R^2 of 0.646 and a Pearson correlation coefficient of 0.803. CART's simplicity resulted in a model that was unable to capture the complexity of the relationships between the vegetation indices, topographic variables, and AGB. The model tended to overfit the training data, leading to poor generalization on the test set.

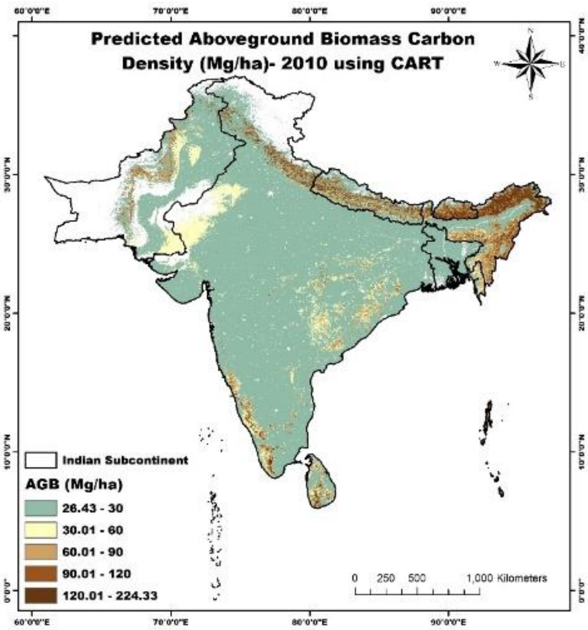

Figure 6. Predicted Aboveground Biomass Carbon Density Map-2010 using CART

CART's predictions were particularly inaccurate in the Himalayan region, where it failed to account for the influence of elevation and aspect on vegetation distribution (Figure 6). However, CART performed reasonably well in flat, homogenous regions, such as the Thar Desert and the Indo-Gangetic plains.

Table 1. Accuracy assessment of machine learning models for AGB estimation

Machine learning model	R-square value	Pearson's correlation
Random Forest	0.834	0.913
GTB (Gradient Tree Boost)	0.783	0.885
CART (Classification and Regression Trees)	0.646	0.803

5.4 Comparison with existing dataset

Visual comparison of RF predicted AGB has also been done by prediction of AGB Carbon density for the 2020 and comparing it with already available ESA CCI AGB data for the same year (Figure 7). The comparison shows that the RF predicted AGB carbon density is better able to account for the spatial heterogeneity of AGB distribution in the Indian Subcontinent especially in the regions of low biomass density such as arid regions of Rajasthan, central India and parts of Pakistan.

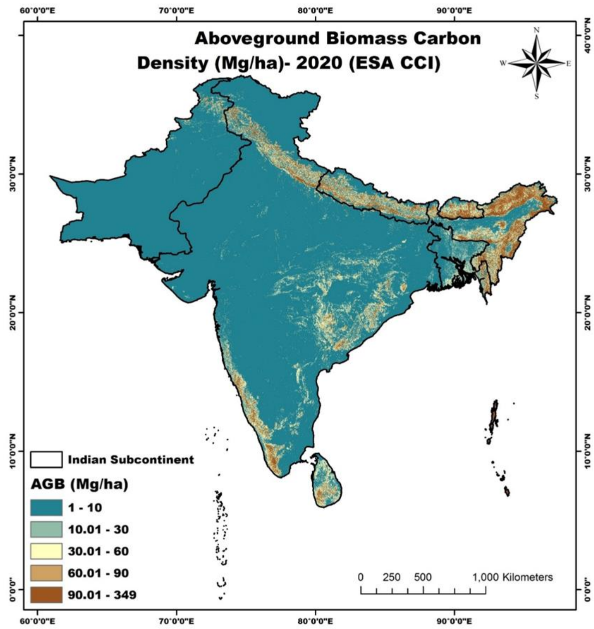

Figure 7. ESA CCI Aboveground Biomass Carbon Density (Mg/ha) for 2020

5.5 AGB prediction for 2023

AGB has been predicted for the recent year i.e. 2023 which shows high biomass density (60 - 349 Mg/ha) in the Himalayan foothills, northeastern India, and the Western Ghats, having dense forest cover with high rainfall. While moderate biomass density (30 - 60 Mg/ha) is found in central India and Sri Lanka's highlands with mixed forests and plantations. Low biomass density (<10 Mg/ha) present in arid regions like Rajasthan, Gujarat, and parts of Pakistan, showing sparse vegetation or grasslands and agricultural regions.

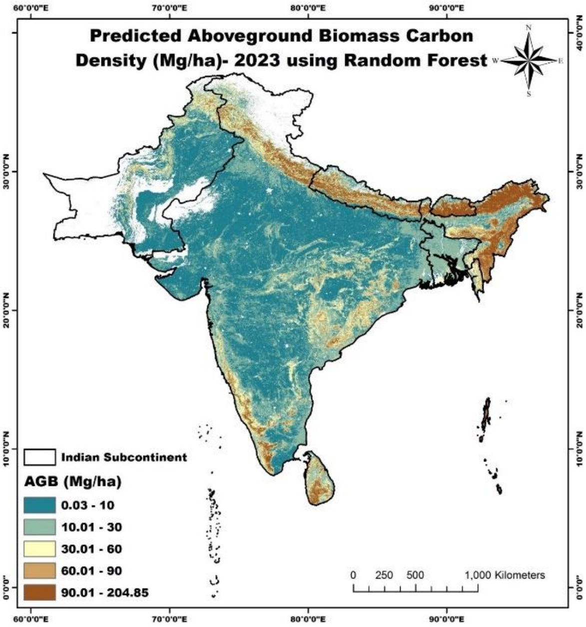

Figure 8. Predicted Aboveground Biomass Carbon Density Map-2023 using Random Forest

6. DISCUSSION

The results of this study indicate that RF is the most robust model for predicting AGB across the Indian subcontinent. The ensemble learning approach used by RF allowed it to capture the complex relationships between vegetation indices, topographic variables, and biomass. In contrast, GTB and CART, while useful in certain contexts, struggled with the region’s ecological variability.

6.1 Model Comparison

RF outperformed both GTB and CART in terms of predictive accuracy, particularly in regions with high biomass density. The ability of RF to handle large datasets and complex interactions between variables makes it well-suited for biomass prediction in heterogeneous landscapes. GTB, while effective in capturing moderate-to-high biomass regions, tended to overestimate biomass in transition zones, suggesting that it may require further tuning for such regions. CART, on the other hand, was too simplistic to handle the complexity of biomass distribution in the subcontinent, particularly in mountainous regions.

6.2 Ecological Insights

The spatial patterns of predicted AGB provide valuable insights into the ecological dynamics of the Indian subcontinent. The high biomass densities in the Western Ghats, northeastern India, and the Sundarbans are consistent with the known carbon sequestration potential of these regions (Ramachandra & Bharath, 2020; Rodda et. al., 2022; Manoj et. al., 2024). Conversely, the low biomass densities in the Thar Desert and the Indo-Gangetic plains highlight the challenges of biomass accumulation in arid and agriculturally dominated landscapes (Rani & Paul, 2023).

The underestimation of biomass in the Himalayan region by all three models suggests that further research is needed to account for the complex interactions between elevation, aspect, and vegetation types in mountainous ecosystems (Anees et. al., 2024). Incorporating additional variables, such as soil moisture and temperature, may improve the accuracy of AGB predictions in these regions (Ahirwal et. al., 2021).

6.3 Implications for Conservation and Policy

Accurate AGB estimation is critical for informing conservation strategies and carbon management policies (Law et. al., 2015). The findings of this study highlight the potential of RF to provide reliable biomass estimates across diverse landscapes, which can be used to prioritize conservation efforts in high-biomass regions. In addition, the ability to predict AGB in arid and semi-arid regions, where biomass is low but ecologically significant, is essential for understanding the subcontinent’s carbon dynamics.

7. CONCLUSION

This study evaluated the performance of three machine learning models—Random Forest, Gradient Tree Boosting, and Classification and Regression Trees—in predicting aboveground biomass across the Indian subcontinent. The results indicate that RF is the most accurate and reliable model for biomass estimation, followed by GTB and CART. The spatial patterns of predicted AGB are consistent with known ecological dynamics, with high biomass densities in forested regions and low biomass densities in arid zones. The findings of this study have important implications for forest management, carbon sequestration strategies, and biodiversity conservation in the Indian subcontinent. By providing accurate, spatially continuous biomass estimates, machine learning models like RF can inform policy decisions aimed at mitigating climate change and preserving the region’s ecological integrity.

DATA AVAILABILITY

The study utilizes publicly available satellite derived products which are available at Google Earth Engine platform. The links are provided in Table 1 for accessing each product.

REFERENCES

Ahirwal, J., Nath, A., Brahma, B., Deb, S., Sahoo, U. K., & Nath, A. J. (2021). Patterns and driving factors of biomass carbon and soil organic carbon stock in the Indian Himalayan region. *Science of the Total Environment*, *770*, 145292.

Anees, S. A., Mehmood, K., Khan, W. R., Sajjad, M., Alahmadi, T. A., Alharbi, S. A., & Luo, M. (2024). Integration of machine learning and remote sensing for above ground biomass estimation through Landsat-9 and field data in temperate forests of the Himalayan region. *Ecological Informatics*, *82*, 102732.

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.

Chen, J. M., & Cihlar, J. (1996). Retrieving leaf area index of boreal conifer forests using Landsat TM images. *Remote sensing of Environment*, *55*(2), 153-162.

Cohen, I., Huang, Y., Chen, J., Benesty, J., Benesty, J., Chen, J., ... & Cohen, I. (2009). Pearson correlation coefficient. *Noise reduction in speech processing*, 1-4.

Das, B., Bordoloi, R., Deka, S., Paul, A., Pandey, P. K., Singha, L. B., ... & Mishra, M. (2021). Above ground biomass carbon assessment using field, satellite data and model based integrated approach to predict the carbon sequestration potential of major land use sector of Arunachal Himalaya, India. *Carbon Management*, *12*(2), 201-214.

Das, B., Patnaik, S. K., Bordoloi, R., Paul, A., & Tripathi, O. P. (2024). Prediction of forest aboveground biomass using an integrated approach of space-based parameters, and forest inventory data. *Geology, Ecology, and Landscapes*, *8*(3), 381-393.

Di Bella, C. M., Paruelo, J. M., Becerra, J. E., Bacour, C., & Baret, F. (2004). Effect of senescent leaves on NDVI-based estimates of f APAR: Experimental and modelling evidences. *International Journal of Remote Sensing*, *25*(23), 5415-5427.

Fararoda, R., Reddy, R. S., Rajashekar, G., Chand, T. K., Jha, C. S., & Dadhwal, V. K. (2021). Improving forest above ground biomass estimates over Indian forests using multi source data sets with machine learning algorithm. *Ecological Informatics*, *65*, 101392.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.

Hofhansl, F., Chacón-Madrigal, E., Fuchslueger, L., Jenking, D., Morera-Beita, A., Plutzar, C., ... & Wanek, W. (2020). Climatic and edaphic controls over tropical forest diversity and vegetation carbon storage. *Scientific Reports*, *10*(1), 5066.

Huete, A. R., Liu, H. Q., Batchily, K. V., & Van Leeuwen, W. J. D. A. (1997). A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote sensing of environment*, *59*(3), 440-451.

Law, E. A., Bryan, B. A., Torabi, N., Bekessy, S. A., McAlpine, C. A., & Wilson, K. A. (2015). Measurement matters in managing landscape carbon. *Ecosystem Services*, *13*, 6-15.

Malla, R., Neupane, P. R., & Köhl, M. (2023). Assessment of above ground biomass and soil organic carbon in the forests of Nepal under climate change scenario. *Frontiers in Forests and Global Change*, *6*, 1209232.

Manoj, K., Arumugam, T., & Prakash, A. (2024). A comparative study on carbon sequestration potential of disturbed and undisturbed mangrove ecosystems in Kannur district, Kerala, South India. *Results in Engineering*, *21*, 101716.

Nguyen, T. D., & Kappas, M. (2020). Estimating the Aboveground Biomass of an Evergreen Broadleaf Forest in Xuan Lien Nature Reserve, Thanh Hoa, Vietnam, Using SPOT-6 Data and the Random Forest Algorithm. *International Journal of Forestry Research*, *2020*(1), 4216160.

Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, *1*, 81-106.

Ramachandra, T. V., & Bharath, S. (2020). Carbon sequestration potential of the forest ecosystems in the Western Ghats, a global biodiversity hotspot. *Natural Resources Research*, *29*(4), 2753-2771.

Rani, J., & Paul, B. (2023). Challenges in arid region reclamation with special reference to Indian Thar Desert—its conservation and remediation techniques. *International Journal of Environmental Science and Technology*, *20*(11), 12753-12774.

Rodda, S. R., Thumaty, K. C., Fararoda, R., Jha, C. S., & Dadhwal, V. K. (2022). Unique characteristics of ecosystem CO2 exchange in Sundarban mangrove forest and their relationship with environmental factors. *Estuarine, Coastal and Shelf Science*, *267*, 107764.

Roy, A. B. (2014). Indian subcontinent. In *A reference model in Earth system and environmental sciences*. Published by Elsevier.

Saatchi, S. S., Houghton, R. A., Dos Santos Alvala, R. C., Soares, J. V., & Yu, Y. (2007). Distribution of aboveground live biomass in the Amazon basin. *Global change biology*, *13*(4), 816-837.

Tjur, T. (2009). Coefficients of determination in logistic regression models—A new proposal: The coefficient of discrimination. *The American Statistician*, *63*(4), 366-372.

Tucker, C. J., Fung, I. Y., Keeling, C. D., & Gammon, R. H. (1986). Relationship between atmospheric CO2 variations and a satellite-derived vegetation index. Nature, 319(6050), 195-199.

Vorster, A. G., Evangelista, P. H., Stovall, A. E., & Ex, S. (2020). Variability and uncertainty in forest biomass estimates from the tree to landscape scale: the role of allometric equations. *Carbon Balance and Management*, *15*, 1-20.