CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Carbon stock estimation at scale from aerial and satellite imagery
Authors	To, Alex;Pham, Hoang Quoc Viet;Nguyen, Quang H.;Davis, Joseph G.;O'Sullivan, Barry;Pan, Shan L.;Nguyen, Hoang D.
Publication date	30-07-2024
Original Citation	To, A., Pham, H. Q. V., Nguyen, Q. H., Davis, J. G., O'Sullivan, B., Pan, S. L. and Nguyen, H. D. (2024) 'Carbon stock estimation at scale from aerial and satellite imagery', 2024 IEEE Conference on Artificial Intelligence (IEEE CAI 2024), Singapore, 25-27 June, pp. 292-299. https://doi.org/10.1109/CAI59869.2024.00064
Type of publication	Conference item
Link to publisher's version	https://ieeecai.org/2024/ - 10.1109/CAI59869.2024.00064
Rights	© 2024, IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Download date	2026-11-01 14:14:07
Item downloaded from	https://hdl.handle.net/10468/15967



University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Carbon Stock Estimation at Scale from Aerial and Satellite Imagery

Alex To
University of Sydney
Sydney, Australia
duto3894@uni.sydney.edu.au

Hoang Quoc Viet Pham
University College Cork
Cork, Ireland
123121186@umail.ucc.ie

Quang H. Nguyen
Reliable Machine Learning Group
Hanoi, Vietnam
nh.quang313@gmail.com

Joseph G. Davis
University of Sydney
Sydney, Australia
joseph.davis@sydney.edu.au

Barry O'Sullivan
Insight Centre for Data Analytics
University College Cork, Ireland
b.osullivan@cs.ucc.ie

Shan L Pan
Institute of Climate Risk and Response
University of New South Wales, Australia
shan.pan@unsw.edu.au

Hoang D. Nguyen*
University College Cork
Cork, Ireland
hn@cs.ucc.ie

***Abstract*—In the ongoing efforts to mitigate climate change effect, the capability to reliably estimate forest carbon stock on a global scale is vital to support sustainable development. This entails the investigation of tree coverage from diverse forest ecosystems worldwide, necessitating a substantial volume of high-resolution images. This paper integrates a variety of remote sensing data sources, from aerial to satellite imagery, for the training and development of our AI system. Given the heterogeneous nature of these data sources, we develop a standardization method to ensure consistent image size and resolution between source platforms. Our harmonized dataset includes 86,088 training images and 21,768 validation images, each with a high resolution of 1.194 m^2 per pixel. We introduce a novel technique for tree semantic segmentation which offers a more effective alternative to traditional individual tree crown delineation for large-scale tree coverage estimation. To assess the adaptability of our AI models, we conducted experiments on a hand-annotated satellite image test set and achieved a *High Vegetation* IoU score of 45.73%. Building on these findings, we present an interactive web-based Geographic Information System for navigating high vegetation segmented satellite images and estimating carbon stock on a global scale.**

***Index Terms*—remote sensing, tree semantic segmentation, aerial imagery, satellite imagery, domain adaptation**

I. Introduction

Carbon dioxide (CO_2) is a primary greenhouse gas, and its increased concentration in the atmosphere is a leading contributor to global warming and climate change. Forests play a crucial role in mitigating these effects by acting as carbon sinks. Understanding and estimating the carbon stock of forests is vital for several environmental and economic reasons such as 1) Climate change mitigation [1] ; 2) Carbon credits trading [2]; 3) Sustainable forest management [3]; 4) Conservation planning [4]; and 5) Environmental policy development [5]; among others.

Measuring tree coverage is essential to support forest carbon stock estimation [6]. Numerous studies have been conducted to measure tree coverage from aerial or satellite imagery and, in some cases, in conjunction with ground-based surveys but mostly on a regional scale [7]–[9].

Despite the advances in deep learning and remote sensing, accurate global tree coverage estimation faces a number of challenges. Firstly, the training data sources are heterogeneous. For example, data collected from independent studies may vary in terms of labeling, image quality, size, and resolution. Secondly, there are discrepancies between training and test data distributions, influenced by factors such as different vegetation types, weather conditions, and other environmental variables. Last but not least, the volume of unseen images required to map the world is enormous. To put this into context, covering just Ireland, which has an area of 84,421 km^2, requires approximately 1.1 million images at a resolution of 1.194 m^2 per pixel, each of size 256 $\times$ 256 pixels.

To address these challenges, we propose harmonizing disparate datasets to maximize the amount of available training data. Additionally, we explore tree semantic segmentation as an alternative to individual tree crowns (ITC) delineation as ITC delineation does not capture tree shapes reliably especially in images with lower resolution or regions with dense canopies. Our contributions are as follows:

1) We standardized and harmonized diverse datasets from different platforms (e.g. aerial and satellite imagery) with multi-modalities (e.g. LiDAR and RGB) and automatically generated a large amount of semantic segmentation masks. We also hand annotated aerial and satellite images to validate our methods. We have made our datasets publicly available to support future research.
2) We conducted extensive experiments on the newly established datasets and reported our baselines.
3) Building on these findings, we introduce a web-based GIS designed for navigating satellite images. Our system harnesses state-of-the-art semantic segmentation models, trained on diverse datasets but perform vegetation semantic segmentation on satellite images and estimate forest carbon stock on a global map.

* Corresponding author.



(a) LiDAR 3D point cloud (b) RGB Image (c) Segmented RGB Image

Fig. 1: The classified 3D LiDAR points (a) are extracted and then projected onto the corresponding RGB image (b) to create the semantic segmentation mask (c). In the mask, the areas highlighted indicate regions classified as *High Vegetation*.

II. Related Work

Numerous large-scale studies have effectively investigated tree density, canopy cover, and the delineation of individual trees using aerial [7], [10]–[19] and satellite imagery [9], [20], [21]. We applied a search strategy similar to [22] to collect articles related to tree detection, ITC delineation and tree semantic segmentation from aerial and satellite imagery and selected 86 relevant articles. From these studies, we compiled a list of publicly accessible datasets. Our selection criteria are limited to datasets that are either freely available for download or can be obtained through reasonable efforts by contacting the authors. The compiled datasets are listed and compared in Table I.

Following the success of early convolutional neural networks (CNNs) like LeNet and AlexNet, there has been widespread adoption of CNNs in the field of remote sensing. Tree coverage estimation in remote sensing typically falls into three categories: tree detection, ITC delineation and tree semantic segmentation. There is a broad range of models discussed in the literature. These include LeNet [23] for tree detection, Resnet50 [7], [14], [24], U-Net [21], Mask R-CNN [18], [25], GoogLeNet [26], MT-EDv3 [9], AlexNet, VGG16, Resnet18, Resnet152 [16], Faster R-CNN [27] for ITC delineation and DeepLabv3+ [20] , U-Net [12], [28], Strong-Weak Faster-RCNN [17], as well as Single Shot Detector (SSD) [13] for tree semantic segmentation.

III. Data Sources

We utilized data from 7157 tiles across 37 sites within the National Ecological Observation Network (NEON), which provides an extensive array of airborne remote sensing data products. These include Discrete Return LiDAR Point Cloud [37], High-Resolution Orthorectified RGB Camera Imagery [38], Ecosystem Structure (Canopy Height Model) [39], Vegetation Structure [40], Vegetation Indices [41], among others. NEON's airborne surveys cover each site extensively, typically spanning over at least 10,000 hectares. A key feature of NEON's data is the consistent use of the Universal Transverse Mercator (UTM) coordinate system across all products, along with the standard organization into 1km $\times$ 1km tiles, which enables effective cross-referencing.

With the focus on tree semantic segmentation, we utilized two key data sources from NEON i.e. the High-Resolution Orthorectified RGB Camera Imagery and the Discrete Return LiDAR Point Cloud. The RGB images offer detailed visual information which is essential for identifying tree features. Meanwhile, the LiDAR data, comprising a 3D classified point cloud, is processed and projected onto a 2D plane. This projection is used to create ground truth semantic segmentation masks corresponding to the RGB images.

In conjunction with the aerial imagery data from NEON, we manually annotated 1,246 satellite images collected from publicly available GIS services. These satellite images complement our aerial images offering a global perspective with their wider availability. Satellite imagery often differs from aerial imagery in terms of color, resolution, and contrast. This diversity in image sources enhances our dataset with a broader range of visual characteristics.

A. NEON Discrete Return LiDAR Point Cloud

LiDAR, which stands for Light Detection and Ranging, has been used for laser distance measurements since the 1960s. LiDAR works by sending out a laser pulse that travels through the air, hits objects like trees, ground, or buildings, and then some of this light bounces back. The time it takes for this light to return is measured to determine distances.

The NEON LiDAR system gathers data in the form of a 3D point cloud with an average density of 5 pts m^{-2}. Each point within this cloud is classified in accordance with the American Society for Photogrammetry and Remote Sensing (ASPRS) definition and is associated with UTM coordinates and intensity values. For generating the semantic segmentation masks, we utilized the points classified under the *High Vegetation* category. The *High Vegetation* class in LiDAR is defined by ASPRS as trees being taller than 1.3 meters.

Dataset Name	Modality			Platform	Specification				Annotation		
	RGB	LiDAR	Other		No of Images	Size (px)	Res (m/px)	Geo	BBox	Instance Seg	Semantic Seg
Biomassters [29]			x	Satellite	310k	256 x 256	10∼60				
Reforestree [30]	x			Aerial	4663	400 x 400	0.02	x	x		
IDTReeS 2020 [31]	x	x		Aerial	85	200 x 200	0.1	x	x		
TreeSatAI [32]	x			Aerial	50,381	300 x 300	0.2	x			
LoveDA [15]	x			Satellite	5,987	1024 x 1024	0.3				x
Larch Casebearer [33]	x			Aerial	1,543	1500 x 1500		x	x		
Sahara and Sahel [21]	x		x	Satellite	11,128		0.5	x		x	
SWISSIMAGE 10 cm [34]	x			Aerial	42,700	10k x 10k	0.1∼2	x			
Auto Arborist [35]	x			Aerial	2,637,208	512 x 512	0.05				
NEON Crowns [36]	x	x		Aerial	11,000	10k x 10k	0.1	x	x		
Ours	**x**	**x**		**Multi**	**107,856**	**256 x 256**	**0.1∼1**	**x**			**x**

TABLE I: A list of publicly available datasets for tree detection from aerial or satellite imagery.

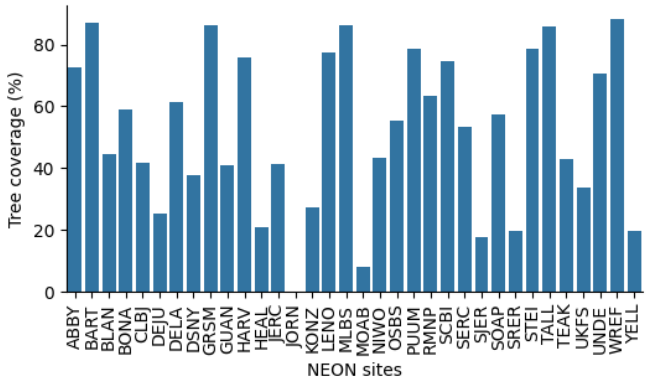


Fig. 2: Tree coverage is defined as the ratio of the area covered by tree canopies to the total area, represented as a percentage. The sites exhibit a balanced distribution of low, medium, and high tree coverage, with the exception of JORN, which has almost no tree coverage.

B. NEON High-Resolution Orthorectified RGB Camera Imagery

NEON provides high-resolution orthorectified camera images that are assembled into mosaics, organized in 1km × 1km tiles. This arrangement maintains a uniform spatial grid, ensuring a spatial resolution of at least 0.1 meters per pixel. Consequently, this results in image sizes of 10000 × 10000 pixels. These images are provided in GeoTIFF format with UTM coordinates, enabling cross-referencing with the LiDAR product.

C. Public GIS Satellite Imagery

We annotated a set of 1246 images from publicly available GIS services, each measuring 256 × 256 pixels, with two classes: *Background (class 0)* and *High Vegetation (class 1)*. For each image, two annotators were assigned to ensure the accuracy of the process. We also excluded areas that appeared to be grassland to align with common practices in carbon stock estimation, where grassland is usually not considered [42].

IV. Data Preprocessing

Our processing pipeline is designed to achieve two main tasks: 1) Creating semantic segmentation masks by extracting classified LiDAR 3D point clouds and projecting them onto a 2D plane, ensuring alignment with the corresponding RGB images; 2) Standardizing the resolution and image size of images from both NEON aerial and GIS satellite images to maintain consistency. To ensure consistency in datum, projection, and units of measure between data sources, we transform all their Coordinate Reference Systems (CRS) to EPSG:4326 using the Python library *pyproj* before any preprocessing steps.

In the first task, we resized RGB images to a dimension of 1000 × 1000 pixels. Given that each tile corresponds to a 1 km ×1 km area, this resizing results in images with a resolution of 1 m^2 per pixel. We projected the 3D LiDAR point cloud onto a 2D plane. The LiDAR 3D point cloud from NEON typically exhibits a density ranging from 5 to 30 points m^{-2}. Therefore, we selected the highest point within each square meter as the representative classification. Points not classified as *High Vegetation* were excluded. From these selected points, we created binary semantic segmentation masks for the corresponding RGB images. In these masks, each pixel is categorized as either *Background* (class 0) or *High Vegetation* (class 1). (Fig. 3).

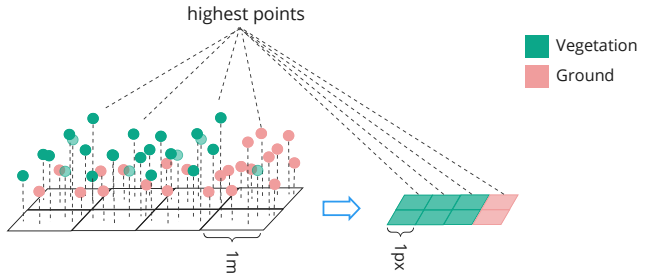


Fig. 3: Projection of a classified LiDAR 3D point cloud onto a 2D plane

In the second task, we rescaled the NEON images and masks obtained from task 1 from a resolution of 1 m^2 per pixel to 1.194 m^2 per pixel to align with the resolution of our satellite images. For this step, we simply resized the images

with nearest neighbor resampling. We employed a sliding window technique, moving from the top left to the bottom right of each image to crop it into patches of 256 $\times$ 256 pixels. When utilizing the sliding window approach, if the last patch on the right or at the bottom does not fit the 256 $\times$ 256 pixel size, we backtrack from the right or bottom edge to ensure each patch is consistently 256 $\times$ 256 pixels. The cropping process resulted in a total of 107,856 patches.

V. Experiments

We structured our experiments to evaluate three key aspects:

1) Model Fitting to NEON Data: this aspect estimates the model's capacity for learning and fitting to the NEON dataset.
2) Single-Site Domain Adaptation: also known as "Train on All Sites, Test on One Site", this part of experiment assesses the model's ability to adapt to an unseen site from within NEON network.
3) Aerial to Satellite Domain Adaptation: this assesses the model's ability to learn from aerial imagery and effectively generalize to satellite imagery, representing a shift between source platforms.

A. Model Fitting to NEON Data

To test the model's capacity in fitting to the NEON dataset, we used all 107,856 images from all 37 NEON sites for training and validation. We allocated approximately 80% of the 107,856 patches, amounting to 86,088 images, to the training set and assigned the remaining 20%, which is 21,768 images, to the validation set. This distribution was stratified by site and tile so that all patches from the same tile were placed in the same set, either training or validation.

B. Single-Site Domain Adaptation (Train on All Sites, Test on One Site)

We employed the approach of Leave-One-Out cross validation at site level to test the model's ability to adapt to an unseen site. During a "site iteration", one site was left out as the test set which is entirely unseen during the training phase, while the remaining sites were stratified into training and validation sets.

C. Aerial to Satellite Domain Adaptation

We used the full model trained in Section.V-A *Model Fitting to NEON Data* to test on the satellite images. Fig. 5 illustrates our data split strategies.

VI. Setup

A. Models and Hyperparameters

DeepLabv3+ [43] with a Resnet50 backbone [44], pre-trained on ImageNet [45] was selected as our baseline model. The models were trained using Stochastic Gradient Descent (SGD) with an initial learning rate of 0.01, momentum set to 0.9, and weight decay of 0.0005. The learning rate followed a polynomial decay rate over the training iterations. We trained the models with a batch size of 32 for 20 epochs. The models were developed in Pytorch and all the experiments were conducted on a system with Ubuntu 22.04 powered by an Intel Core i9 20 cores with 256GB memory and two NVIDIA 3090 GPUs with 24 GB GPU memory each.

B. Data Augmentation

To enhance the robustness of our models, we incorporated a series of data augmentation techniques including: Random Brightness, Random Contrast, Color Conversion from RGB to HSV, Random Saturation, Random Hue, Color Conversion from HSV back to BGR, Random Channel Swapping, Random Resizing, and Random Cropping. Each of these augmentation methods was applied to the training samples with a probability of 0.5.

C. Evaluation Metrics

We used Intersection over Union (IoU) metric, also known as Jaccard Index, to evaluate the model performance. IoU is defined as the area of the intersection over union of the predicted segmentation and the ground truth. We report both IoU for *Background* and *High Vegetation* class as well as the mean IoU (mIoU).

VII. Results

A. Model Fitting to NEON Data

Overall, the "Full Model" trained on all NEON sites, classifies *High Vegetation* and *Background* with good results, yielding IoU scores of 83.41% and 81.72%, respectively, and a mean IoU of 82.57% on the validation set (Table II).

High Vegetation IoU (%)	Background IoU (%)	mIoU (%)
83.41	81.72	82.57

TABLE II: The model was trained using data from all NEON sites (Section. V-A). The reported results are based on the validation set.

B. Single-Site Domain Adaptation

During Leave One Out training, the model's performance on the training folds remained nearly constant and comparable to that of the Full Model. This outcome was anticipated, as the quantity of images excluded from each individual site constitutes only about 3 to 5% of the total image count. The model's performance on unseen test sites varied, with some sites showing increased *High Vegetation* IoU and others experiencing a decrease. (Fig. 4).

The JORN site exhibits the most substantial decrease in performance metrics, with its *High Vegetation* IoU scores of only 8.77%. This outcome presents a noteworthy case, as JORN is known to have virtually no tree coverage. Its inclusion in the dataset was deliberate, aimed at evaluating the model's adaptability in scenarios characterized by either extremely dense or virtually nonexistent tree coverage. On the other hand, some test sites achieve higher *High Vegetation* IoU compared to the training performance for e.g. BART, BLAN, DELA, GRSM etc.

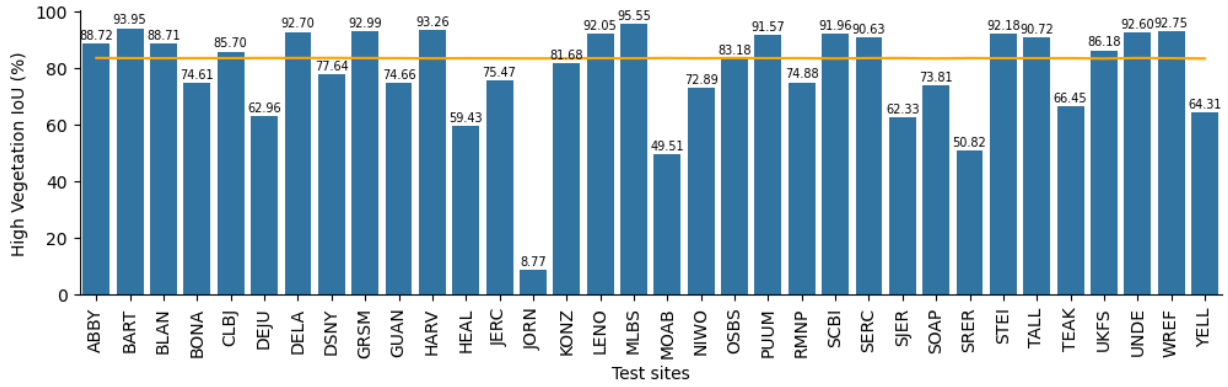


Fig. 4: Train on All Sites, Test on One Site: The line plot represents the training performance across all sites and the bar plot represents the performance on the unseen test excluded during training.

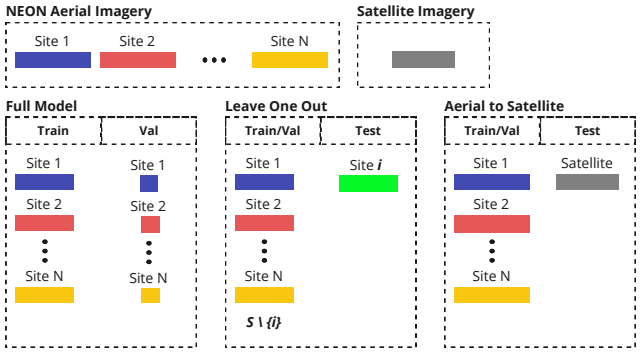


Fig. 5: Data split strategies for Full Model, Single-Site Domain Adaptation and Aerial to Satellite Domain Adaptation models.

C. Aerial to Satellite Domain Adaptation

Applying the Full Model to our hand-annotated GIS satellite imagery, we observe a decrease in *High Vegetation* IoU, which dropped to 45.73%. Despite this decline, the results remain promising for practical applications.

High Vegetation IoU (%)	Background IoU (%)	mIoU (%)
45.73	84.54	65.14

TABLE III: The model was trained using data from all NEON sites (Section. V-A). The reported results are based on the GIS satellite imagery set.

VIII. AI-Powered Geographic Information System for Tree Coverage and Carbon Stock Estimation

Based on the promising results of our models, we introduce a publicly available GIS for tree coverage and carbon stock estimation.

A. Carbon Stock Estimation Workflow

Current methodologies for estimating carbon stock on a large scale primarily involve using allometric equations [46]–[51] that establish a relationship between the structure and species of trees and their estimated carbon stock through empirical experiments. These equations estimate carbon stock based on field data collected from small sample plots, which is then extrapolated to site and regional levels [42]. We adopt a similar approach for our end-to-end GIS, assuming a relationship exists between forest tree species, tree coverage and their carbon stock.

Fig. 7 illustrates our end-to-end carbon stock estimation workflow. The inputs include RGB satellite imagery to estimate forest tree coverage, and other modalities such as LiDAR and Canopy Height Model are used to calculate forest structure. Additionally, Eco-regions metadata provides genus and species information. These inputs are then combined with allometric equations to output carbon stock estimation. When other modalities and eco-regions metadata are not available, the system defaults to conservative allometric equations.

B. System Design

Our application consists of four main blocks, a Semantic Segmentation back-end (develop with PyTorch), a Web API back-end (developed with Python FastAPI), a Geo-spatial Indexing and Caching back-end (developed with MongoDB) and a Web User Interface (developed with React) as shown in Figure 8. The Semantic Segmentation back-end hosts our models for tree semantic segmentation. The Web API back-end serves the upper layers (i.e Geo-spatial Indexing and Caching and WebUI) and communicates with the Semantic Segmentation back-end for inferencing. The Geo-spatial Indexing and Caching back-end stores the semantic segmentation masks to save on computation. The masks are indexed based on their geographic location leveraging the geospatial capabilities of MongoDB.

C. User Interface

The user interface of the system presents an interactive map, allowing users to seamlessly navigate and explore real-time tree semantic segmentation as shown in Figure 6. As they pan and zoom across the map, the segmented representations of trees dynamically update, providing an intuitive visualization. To facilitate quick navigation, a search box is prominently



Fig. 6: The GIS interface allows the user to navigate satellite images and perform tree semantic segmentation in real-time. The user can select an area and view the estimated carbon stock within the selected area.

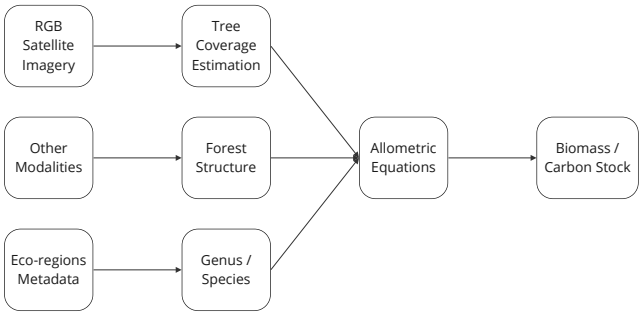


Fig. 7: Our GIS carbon stock estimation workflow.

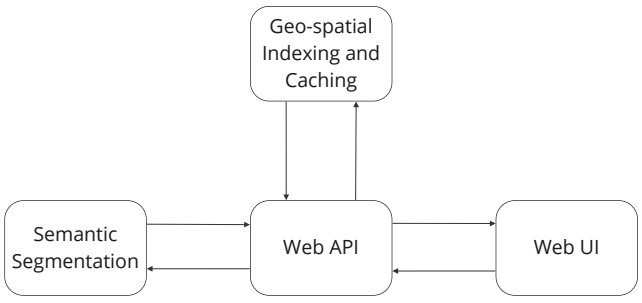


Fig. 8: Our GIS system architecture.

positioned at the top left, enabling users to instantly jump to specific locations by entering relevant keywords or coordinates. Additionally, a toolbox is available on the right, offering a suite of tools for various tasks. The user can delineate specific areas of interest using point-based selections, rectangular regions, and other geometric shapes. Once areas are selected using these tools, they are highlighted on the map, and additional information is computed on-the-fly, such as the estimated biomass and carbon stock.

IX. Discussion

Our Full Model provides an indication of the appropriate model size necessary for learning from forest environments, such as those within the NEON network. During Leave One Out training, some sites such as JORN exhibit a significant decrease in *High Vegetation* IoU, often resulting from a shift in forest characteristics from those in training data. In this case, JORN is reported to have almost no tree coverage as derived from LiDAR data.

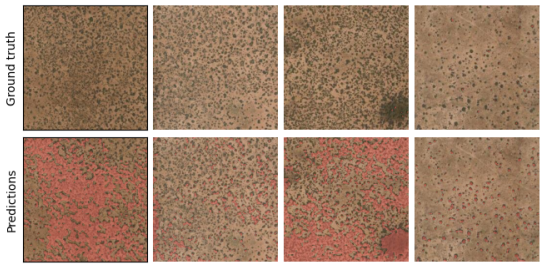


Fig. 9: In the comparison of ground truth (first row) versus predictions (second row) for JORN site, the images contain no ground truth masks, even though some areas visually resemble trees or vegetation.

Upon visual inspection of the predicted masks against the ground truth for JORN site (Fig. 9), it is observed that numerous false positives likely arose from discrepancies between LiDAR data and RGB data in the definition of trees i.e. LiDAR data does not classify small trees or bushes as *High Vegetation* due to the threshold of tree height. On the other hand, large bushes can visually resemble a large canopy from a bird's eye view, which can be challenging to distinguish even for human experts. This poses a challenge in predicting tree coverage using only RGB imagery. While RGB images offer visual details, they lack the structural information about trees that LiDAR data can provide.

By contrast, sites such as MLBS achieve 95.55% *High Vegetation* IoU and "near-perfect" predictions in most samples (Fig. 10). In this case, the explanation for this is that MLBS site has very high tree coverage (85%, Fig. 2) and monoculture forests such as MLBS appear uniform from an aerial perspective due to the predominance of a single tree genus and species. Therefore, in training, the model only needs to learn the unique features of this single tree genus and species, in order to be able to effectively predict across the entire forest due to the low biodiversity. Within the NEON network, there

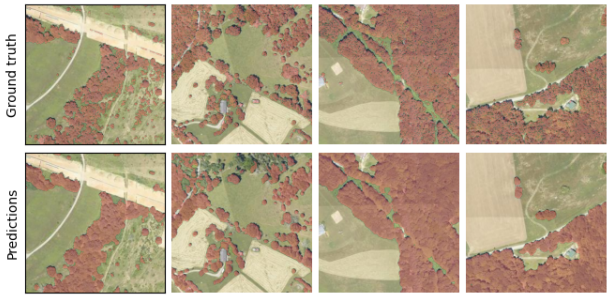


Fig. 10: Test sites such as MLBS achieved "near-perfect" predictions.

is a high correlation between tree coverage and test time *High Vegetation* IoU with a Pearson correlation coefficient of 0.88.

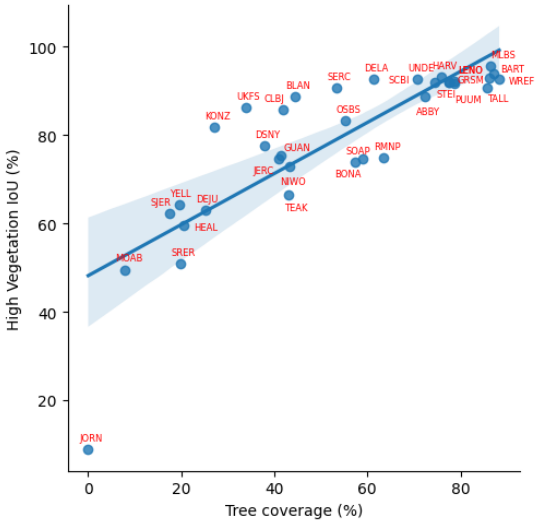


Fig. 11: Tree coverage vs test time *Vegetation* IoU. Pearson correlation coefficient = 0.88.

We use our AI models to develop a GIS for carbon stock estimation. Currently, our system operates in a two-step process, initially predicting tree coverage and then integrating this with data from various modalities and allometric equations for carbon stock calculation. As part of our future work, we aim to enhance our datasets and models to support multi-modality training, enabling a more direct approach where the model takes in all modalities as input and directly outputs carbon stock estimates.

X. Conclusions

In this study, we introduce a new dataset from aerial and satellite imagery for tree semantic segmentation of its various baseline models. We also introduce an end to end GIS for tree coverage and carbon stock estimation. Our works demonstrate the potential of AI in environmental analysis, offering valuable tools for sustainability research. We anticipate our dataset and baselines will be a valuable contribution to the remote sensing community to tackle tree coverage and carbon stock estimation at scale. Our current work primarily focuses on RGB imagery due to its greater availability from satellite, and aerial sources. However, future directions will include exploration of additional modalities, such as infrared, multispectral or hyperspectral bands, to enhance model performance.

Data availability

Our code and analysis are released on GitHub https://github.com/SustainableML/2024_ieee_cai. Data are available upon request.

Acknowledgment

This publication has emanated from research conducted with financial support of Science Foundation Ireland under Grant 12/RC/2289-P2, and 18/CRT/6223.

References

[1] B. W. Griscom, J. Adams, P. W. Ellis, R. A. Houghton, G. Lomax, D. A. Miteva, W. H. Schlesinger, D. Shoch, J. V. Siikamäki, P. Smith *et al.*, "Natural climate solutions," *Proceedings of the National Academy of Sciences*, vol. 114, no. 44, pp. 11 645–11 650, 2017.

[2] L. P. Koh, Y. Zeng, T. V. Sarira, and K. Siman, "Carbon prospecting in tropical forests for climate change mitigation," *Nature Communications*, vol. 12, no. 1, p. 1271, 2021.

[3] N. K. Phat, W. Knorr, and S. Kim, "Appropriate measures for conservation of terrestrial carbon stocks—analysis of trends of forest management in southeast asia," *Forest Ecology and Management*, vol. 191, no. 1-3, pp. 283–299, 2004.

[4] C. Soto-Navarro, C. Ravilious, A. Arnell, X. De Lamo, M. Harfoot, S. Hill, O. Wearn, M. Santoro, A. Bouvet, S. Mermoz *et al.*, "Mapping co-benefits for carbon storage and biodiversity to inform conservation policy and action," *Philosophical Transactions of the Royal Society B*, vol. 375, no. 1794, p. 20190128, 2020.

[5] J. I. Ajani, H. Keith, M. Blakers, B. G. Mackey, and H. P. King, "Comprehensive carbon stock and flow accounting: a national framework to support climate change mitigation policy," *Ecological Economics*, vol. 89, pp. 61–72, 2013.

[6] A. Fujimoto, C. Haga, T. Matsui, T. Machimura, K. Hayashi, S. Sugita, and H. Takagi, "An end to end process development for uav-sfm based forest monitoring: Individual tree detection, species classification and carbon dynamics simulation," *Forests*, vol. 10, no. 8, p. 680, 2019.

[7] B. G. Weinstein, S. Marconi, S. Bohlman, A. Zare, and E. White, "Individual tree-crown detection in rgb imagery using semi-supervised deep learning neural networks," *Remote Sensing*, vol. 11, no. 11, p. 1309, 2019.

[8] L. Duncanson, J. R. Kellner, J. Armston, R. Dubayah, D. M. Minor, S. Hancock, S. P. Healey, P. L. Patterson, S. Saarela, S. Marselis *et al.*, "Aboveground biomass density models for nasa's global ecosystem dynamics investigation (gedi) lidar mission," *Remote Sensing of Environment*, vol. 270, p. 112845, 2022.

[9] G. Lassalle, M. P. Ferreira, L. E. C. La Rosa, and C. R. de Souza Filho, "Deep learning-based individual tree crown delineation in mangrove forests using very-high-resolution satellite imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 189, pp. 220–235, 2022.

[10] T. W. Crowther, H. B. Glick, K. R. Covey, C. Bettigole, D. S. Maynard, S. M. Thomas, J. R. Smith, G. Hintler, M. C. Duguid, G. Amatulli *et al.*, "Mapping tree density at a global scale," *Nature*, vol. 525, no. 7568, pp. 201–205, 2015.

[11] B. G. Weinstein, S. Marconi, S. A. Bohlman, A. Zare, A. Singh, S. J. Graves, and E. P. White, "A remote sensing derived data set of 100 million individual tree crowns for the national ecological observatory network," *Elife*, vol. 10, p. e62922, 2021.

[12] F. Schiefer, T. Kattenborn, A. Frick, J. Frey, P. Schall, B. Koch, and S. Schmidtlein, "Mapping forest tree species in high resolution uav-based rgb-imagery by means of convolutional neural networks," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 170, pp. 205–215, 2020.

[13] A.-I. Pleșoianu, M.-S. Stupariu, I. Șandric, I. Pătru-Stupariu, and L. Drăguț, "Individual tree-crown detection and species classification in very high-resolution remote sensing imagery using a deep learning ensemble model," *Remote Sensing*, vol. 12, no. 15, p. 2426, 2020.

[14] M. Aubry-Kientz, A. Laybros, B. Weinstein, J. G. Ball, T. Jackson, D. Coomes, and G. Vincent, "Multisensor data fusion for improved segmentation of individual tree crowns in dense tropical forests," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3927–3936, 2021.

[15] J. Wang, Z. Zheng, A. Ma, X. Lu, and Y. Zhong, "Loveda: A remote sensing land-cover dataset for domain adaptive semantic segmentation," *arXiv preprint arXiv:2110.08733*, 2021.

[16] M. Onishi and T. Ise, "Explainable identification and mapping of trees using uav rgb image and deep learning," *Scientific reports*, vol. 11, no. 1, p. 903, 2021.

[17] Y. Wang, G. Yang, and H. Lu, "Domain adaptive tree crown detection using high-resolution remote sensing images," *Journal of Applied Remote Sensing*, vol. 16, no. 4, pp. 044 505–044 505, 2022.

[18] J. G. Ball, S. H. Hickman, T. D. Jackson, X. J. Koay, J. Hirst, W. Jay, M. Archer, M. Aubry-Kientz, G. Vincent, and D. A. Coomes, "Accurate delineation of individual tree crowns in tropical forests from aerial rgb imagery using mask r-cnn," *Remote Sensing in Ecology and Conservation*, 2023.

[19] S. Li, M. Brandt, R. Fensholt, A. Kariryaa, C. Igel, F. Gieseke, T. Nord-Larsen, S. Oehmcke, A. H. Carlsen, S. Junttila *et al.*, "Deep learning enables image-based tree counting, crown segmentation, and height prediction at national scale," *PNAS nexus*, vol. 2, no. 4, p. pgad076, 2023.

[20] M. Liu, B. Fu, S. Xie, H. He, F. Lan, Y. Li, P. Lou, and D. Fan, "Comparison of multi-source satellite images for classifying marsh vegetation using deeplabv3 plus deep learning algorithm," *Ecological Indicators*, vol. 125, p. 107562, 2021.

[21] M. Brandt, C. J. Tucker, A. Kariryaa, K. Rasmussen, C. Abel, J. Small, J. Chave, L. V. Rasmussen, P. Hiernaux, A. A. Diouf *et al.*, "An unexpectedly large count of trees in the west african sahara and sahel," *Nature*, vol. 587, no. 7832, pp. 78–82, 2020.

[22] H. Zhao, J. Morgenroth, G. Pearse, and J. Schindler, "A systematic review of individual tree crown detection and delineation with convolutional neural networks (cnn)," *Current Forestry Reports*, pp. 1–22, 2023.

[23] W. Li, H. Fu, L. Yu, and A. Cracknell, "Deep learning based oil palm tree detection and counting for high-resolution remote sensing images," *Remote sensing*, vol. 9, no. 1, p. 22, 2016.

[24] Y. Gan, Q. Wang, and A. Iio, "Tree crown detection and delineation in a temperate deciduous forest from uav rgb imagery using deep learning approaches: Effects of spatial resolution and species characteristics," *Remote Sensing*, vol. 15, no. 3, p. 778, 2023.

[25] Z. Hao, L. Lin, C. J. Post, E. A. Mikhailova, M. Li, Y. Chen, K. Yu, and J. Liu, "Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (mask r-cnn)," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 178, pp. 112–123, 2021.

[26] S. Yan, L. Jing, and H. Wang, "A new individual tree species recognition method based on a convolutional neural network and high-spatial resolution remote sensing imagery," *Remote Sensing*, vol. 13, no. 3, p. 479, 2021.

[27] M. Beloiu, L. Heinzmann, N. Rehush, A. Gessler, and V. C. Griess, "Individual tree-crown detection and species identification in heterogeneous forests using aerial rgb imagery and deep learning," *Remote Sensing*, vol. 15, no. 5, p. 1463, 2023.

[28] T. Kattenborn, J. Eichel, and F. E. Fassnacht, "Convolutional neural networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution uav imagery," *Scientific reports*, vol. 9, no. 1, p. 17656, 2019.

[29] A. Nascetti, R. Yadav, K. Brodt, Q. Qu, H. Fan, Y. Shendryk, I. Shah, and C. Chung, "Biomassters: A benchmark dataset for forest biomass estimation using multi-modal satellite time-series," in *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2023.

[30] G. Reiersen, D. Dao, B. Lütjens, K. Klemmer, K. Amara, A. Steinegger, C. Zhang, and X. Zhu, "Reforestree: A dataset for estimating tropical forest carbon stock with deep learning and aerial imagery," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 11, 2022, pp. 12 119–12 125.

[31] University of Florida, "Idtrees 2020," 2020. [Online]. Available: https://idtrees.org/products/

[32] S. Ahlswede, C. Schulz, C. Gava, P. Helber, B. Bischke, M. Förster, F. Arias, J. Hees, B. Demir, and B. Kleinschmit, "Treesatai benchmark archive: A multi-sensor, multi-label dataset for tree species classification in remote sensing," *Earth System Science Data Discussions*, vol. 2022, pp. 1–22, 2022.

[33] Swedish Forest Agency - National Forest Data Lab, "Forest damages – larch casebearer 1.0," 2021. [Online]. Available: https://lila.science/datasets/forest-damages-larch-casebearer/

[34] Federal Office of Topography - swisstopo, "The orthophoto mosaic swissimage 10 cm," 2023. [Online]. Available: https://www.swisstopo.admin.ch/en/geodata/images/ortho/swissimage10.html

[35] S. Beery, G. Wu, T. Edwards, F. Pavetic, B. Majewski, S. Mukherjee, S. Chan, J. Morgan, V. Rathod, and J. Huang, "The auto arborist dataset: A large-scale benchmark for multiview urban forest monitoring under domain shift," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 21 294–21 307.

[36] B. G. Weinstein, S. Marconi, S. Bohlman, A. Zare, A. Singh, S. J. Graves, and E. White, "Neon crowns: a remote sensing derived dataset of 100 million individual tree crowns," *BioRxiv*, pp. 2020–09, 2020.

[37] National Ecological Observatory Network (NEON), "Discrete return lidar point cloud (dp1.30003.001)," 2023. [Online]. Available: https://data.neonscience.org/data-products/DP1.30003.001/RELEASE-2023

[38] ——, "High-resolution orthorectified camera imagery (dp1.30010.001)," 2023. [Online]. Available: https://data.neonscience.org/data-products/DP1.30010.001/RELEASE-2023

[39] ——, "Ecosystem structure (dp3.30015.001)," 2023. [Online]. Available: https://data.neonscience.org/data-products/DP3.30015.001/RELEASE-2023

[40] ——, "Vegetation structure (dp1.10098.001)," 2023. [Online]. Available: https://data.neonscience.org/data-products/DP1.10098.001/RELEASE-2023

[41] ——, "Vegetation indices - spectrometer - mosaic (dp3.30026.001)," 2023. [Online]. Available: https://data.neonscience.org/data-products/DP3.30026.001/RELEASE-2023

[42] K. Cushman, S. Saatchi, R. E. McRoberts, K. J. Anderson-Teixeira, N. A. Bourg, B. Chapman, S. M. McMahon, and C. Mulverhill, "Small field plots can cause substantial uncertainty in gridded aboveground biomass products from airborne lidar data," *Remote Sensing*, vol. 15, no. 14, p. 3509, 2023.

[43] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[45] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[46] G. B. West, J. H. Brown, and B. J. Enquist, "A general model for the origin of allometric scaling laws in biology," *Science*, vol. 276, no. 5309, pp. 122–126, 1997.

[47] B. W. Nelson, R. Mesquita, J. L. Pereira, S. G. A. De Souza, G. T. Batista, and L. B. Couto, "Allometric regressions for improved estimate of secondary forest biomass in the central amazon," *Forest ecology and management*, vol. 117, no. 1-3, pp. 149–167, 1999.

[48] T. M. Araújo, N. Higuchi, and J. A. de Carvalho Júnior, "Comparison of formulae for biomass content determination in a tropical rain forest site in the state of pará, brazil," *Forest ecology and management*, vol. 117, no. 1-3, pp. 43–52, 1999.

[49] T. Verwijst and B. Telenius, "Biomass estimation procedures in short rotation forestry," *Forest ecology and management*, vol. 121, no. 1-2, pp. 137–146, 1999.

[50] C.-W. Xiao and R. Ceulemans, "Allometric relationships for below- and aboveground biomass of young scots pines," *Forest ecology and management*, vol. 203, no. 1-3, pp. 177–186, 2004.

[51] J. Chave, C. Andalo, S. Brown, M. A. Cairns, J. Q. Chambers, D. Eamus, H. Fölster, F. Fromard, N. Higuchi, T. Kira *et al.*, "Tree allometry and improved estimation of carbon stocks and balance in tropical forests," *Oecologia*, vol. 145, pp. 87–99, 2005.