Deep Regression Neural Network for Estimating Canopy Height in Vietnam's National Forests

***Abstract*—Estimating canopy height in forests is an important and necessary step in measuring forest health, biodiversity, and carbon storage. This paper studies the challenges and applicability of using deep neural network models to predict canopy height in the national forests of Vietnam, a country with a rapidly growing economy and a commitment to achieving a net-zero carbon footprint by 2050. We first argue that estimating canopy height in Vietnam presents significant challenges, particularly the need for a low-cost approach and the country's complex forest structures. Then, using wide-coverage and freely accessible Sentinel-2 data, besides GEDI, we systematically study the performance of the existing canopy height regression models in the context of Vietnam. Finally, we propose a new approach that can effectively take advantage of the vegetation indices with original Sentinel bands data and show promising results compared to the existing models on standard evaluation metrics such as Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). The results show the potential of a low-cost canopy-height estimation approach, taking a step towards sustainable forest management and environmental conservation in Vietnam.**

***Index Terms*—Canopy height estimation, vegetation structure, multispectral remote sensing**

I. Introduction

The advent of deep learning and its application in remote sensing has paved the way for essential and innovative environmental monitoring techniques for addressing contemporary global challenges such as climate change. In Northern Vietnam, where terrestrial ecosystems play a crucial role in maintaining biodiversity and mitigating climate effects, understanding these ecosystems through vegetation characteristics like canopy height is vital [1]. This research aims to understand the challenges of measuring canopy height in Vietnam's forests and develop a deep-learning estimation model using low-cost Sentinel-2 satellite imagery, contributing to Vietnam's commitment to achieving a net-zero carbon footprint by 2050, as declared at the COP26 World Leaders' Summit[1] in 2021.

Global efforts to combat climate change and promote biodiversity are underscored by the United Nations' forest goals, which include enhancing global carbon stocks and increasing forest cover [2]. Accurately mapping and monitoring forest structural variables like canopy height is crucial for ecosystem service assessment, carbon stock quantification, wildlife management, and fire modeling. However, traditional methods such as field-based measurements [3]–[5], while accurate, are often limited in scope and feasibility due to their time-consuming and labor-intensive nature, especially in remote and complex forest structures. Thus, there is a pressing need for novel remote sensing techniques that provide consistent and expansive data on morphological traits predictive of biodiversity and carbon stocks on a global scale.

Canopy height serves as a critical ecological indicator due to its correlation with species composition, climate, and site quality, and its ability to estimate stand age, successional stages, primary productivity, aboveground biomass, and biodiversity [6]–[8]. Understanding canopy height through remote sensing offers a promising pathway to capture these ecological metrics across extensive areas, facilitating better management and conservation practices. Leveraging this large-scale approach is particularly important for countries such as Vietnam, as it seeks to balance rapid economic growth with sustainable environmental management.

The quest for accurate canopy height data has traditionally relied on three main approaches: field-based observations, UAV-based sensing, and satellite remote sensing. Field inventories [9]–[11] provide detailed but geographically and temporally limited data. UAVs equipped with LiDAR [12]–[14] offer precise measurements but at a higher cost and reduced scale. In contrast, satellite remote sensing, particularly using platforms like Sentinel-2 [15], presents a viable alternative for Vietnam due to its wide coverage and free access to data.

In this study, we utilize Sentinel-2 data, which stands out for its capability to provide high-resolution imagery essential for detailed analysis of complex vegetative structures [15]–[17].

One of the primary reasons for choosing Sentinel-2 is its cost-effectiveness. The data provided by Sentinel-2 are freely accessible, which significantly reduces the financial barriers to obtaining high-quality satellite images for scientific research and practical applications in conservation and climate change mitigation. This accessibility is especially important for regions like Vietnam, where resource constraints might limit the feasibility of frequent and expansive environmental monitoring campaigns.

In this paper, we make the following contributions:

- We construct and process satellite data from Sentinel-2 and GEDI for forest regions in Vietnam. This serves as the foundation of our paper, and any subsequent research on canopy height estimation in Vietnam.
- Aiming to develop low-cost and effective canopy height estimation models for forest regions in Vietnam, we develop a neural model capable of learning additional structural information from the original images to achieve more accurate canopy height estimations. We also study the utilization of vegetation indices in addition to the

[1]https://unfccc.int/cop26/world-leaders-summit

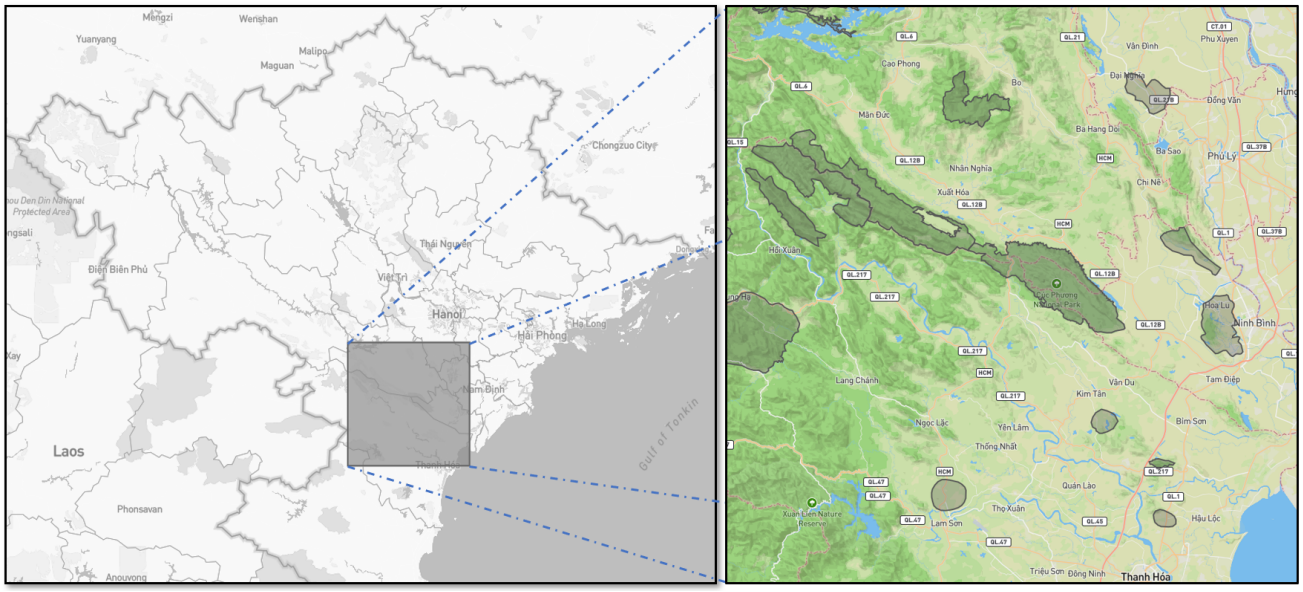


Fig. 1: The region of interest.

original data bands and further improve the prediction outcomes.

- Finally, we perform experiments to validate the effectiveness of our approach. Specifically, our approach outperforms several representative canopy-height estimation baselines across various metrics.

The contributions of our paper provide rigorous evaluations of deep learning techniques for forest monitoring and management in Vietnam, showing the potentials and limitations of using satellite data for environmental research and decision-making in this region.

II. Related work

A. Machine learning

Machine learning algorithms, such as Random Forest (RF) and Symbolic Regression (SR), are applied to estimate canopy height in the Bhitarkanika Wildlife Sanctuary (BWS) using Sentinel images and field data, achieving R^2 scores of 0.6 and 0.62 respectively [18]. Similarly, [19] utilized RF to predict canopy height in Doon Valley using data from ICESat-2 and Sentinel, with a model accuracy of R^2 of 0.84 and an RMSE of 4.48%. Another study employed a combination of GEDI LIDAR, Sentinel multispectral datasets, and SRTM data to predict forest canopy cover and height in tropical forests, yielding high accuracies with R^2 values of 0.86 for height and 0.87 for cover [20]. While machine learning algorithms demonstrate significant potential in estimating canopy height and cover, their accuracy is heavily reliant on the quality and diversity of the input data, which can be a limiting factor in regions with sparse or inconsistent data coverage.

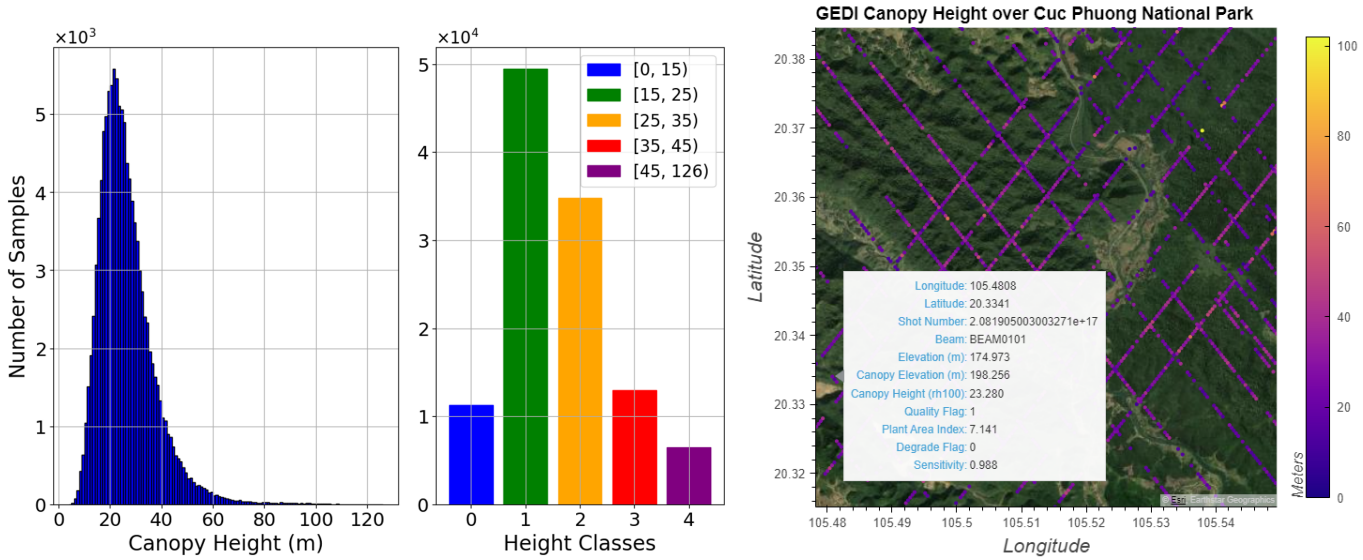
B. Deep learning

1) Convolutional Neural Network: Distinct from traditional machine learning methods that rely on manually extracted features, such as metrics derived from LiDAR waveforms, DL methods excel at processing raw signals directly, thereby simplifying the typically complex preprocessing steps [22].

CNNs, a staple in signal processing, are renowned for their efficacy across various signal types, including one-dimensional time series and two-dimensional images [23], [24]. The convolutional layer, fundamental to CNN architectures, adeptly handles the spatial or temporal auto-correlation inherent in many signal types. In the specific case of GEDI's univariate waveform signals, which travel from the atmosphere to the Earth's surface, one-dimensional CNNs (1D-CNNs) are effectively utilized to capture the auto-correlation along the wave dimension. Alternatively, reconfiguring the waveform into a two-dimensional format allows the application of two-dimensional CNNs (2D-CNNs). This method, while less conventional for waveform data, is particularly advantageous for GEDI signals due to their high sparsity, helping to preserve and highlight the structural information within the signal, thereby enhancing internal signal contrast [25].

Following the foundational introduction of CNNs for managing spatial and temporal correlations in signal data, their practical effectiveness has been further validated through diverse applications in remote sensing for environmental monitoring. Notably, CNNs [22], [26], [27] have adeptly processed raw waveform and optical data across various datasets, substantially enhancing the accuracy and efficiency of forest biophysical parameter estimations, such as canopy heights and biomass. These studies collectively demonstrate the power of CNNs to improve signal contrast and extract valuable insights from complex data, making them invaluable tools in the advancement of geospatial analysis.

2) Vision Transformer: In the domain of aerial imagery, ViTs have been applied with notable success [31]–[34]. Despite their strengths, the application of ViTs in generating



(a) Height distribution over the ROI

(b) The GEDI footprints

Fig. 2: The reference GEDI (LiDAR-derived) data

comprehensive canopy height maps from high-resolution, airborne LiDAR data faces challenges, particularly due to the scarcity of such data. This scarcity can hinder the models' generalizability to new geographies, especially in data-poor regions [35]. To overcome these limitations, self-supervised learning (SSL) techniques like the SSL DINOv2 approach have been pivotal in advancing the state-of-the-art across various vision tasks, including image classification and segmentation [36], [37]. Furthermore, to reduce dependency on SSL, [38] employed knowledge distillation from a U-Net CNN teacher model to produce a 10-m canopy height model (CHM) of Ghana, utilizing a combination of Sentinel-1, Sentinel-2, and aerial LiDAR data, thus enhancing the robustness and applicability of ViTs in practical scenarios.

III. Study Area and Datasets

Northern Vietnam boasts rich biodiversity, and within its verdant landscapes lie two of its most renowned national parks: Cuc Phuong and Pu Luong. These parks serve as invaluable repositories of ecological diversity and natural heritage.

Cuc Phuong National Park, established in 1962, stands as Vietnam's first national park and is recognized as one of the country's most significant conservation areas. Encompassing an area of approximately 222 square kilometers, Cuc Phuong is renowned for its lush tropical rainforests, limestone karst formations, and diverse wildlife. Meanwhile, Pu Luong Nature Reserve, situated further to the northwest, covers a sprawling area of over 200 square kilometers, encompassing a landscape characterized by rugged mountains, deep valleys, and cascading rice terraces. In our research article, we will utilize comprehensive datasets surrounding these two primary forested areas, as specifically described in Figure 1.

A. The Remote Sensing Data

The study utilized remote sensing datasets comprising multispectral imagery, as well as satellite-derived LiDAR data, covering a buffer zone spanning approximately 110×110 km^2, delineated by the southwest coordinates (19.805063, 104.99981) and the northeast coordinates (20.800241, 106.04804).

The multispectral imagery utilized in our study is sourced from the Sentinel-2 (S2) mission[2], with the available data products include Level-1C and Level-2A. Within the scope of our analysis, we specifically employ Level-2A data. The Sentinel-2 mission offers a total of 12 spectral bands, with pixel sizes varying between 10, 20, or 60 meters. Among these, four bands have a spatial resolution of 10 meters, including B2: Blue, B3: Green, B4: Red, and B8: Near Infrared (NIR). Additionally, there is one band, B9: Short Wave Infrared, with a 60-meter resolution, while the remaining bands have a resolution of 20 meters, comprising B1: Ultra Blue, B5-B6-B7: Red edge, B8A: narrow NIR, and B11-B12: Short Wave Infrared (SWIR).

The satellite imagery was acquired on February 12, 2024, with a maximum cloud coverage value of 3%. The imagery obtained at a 10-meter resolution has dimensions of 10980×10980 pixels.

The satellite LiDAR data. We utilize canopy height reference data obtained from the Global Ecosystem Dynamics Investigation[3] (GEDI) to generate a continuous height map with a spatial resolution of 10 meters. The GEDI instrument comprises three lasers, producing eight ground tracks spaced approximately 600 meters apart in the cross-track direction relative to the flight path, and around 735 meters of spacing

[2]https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2
[3]https://gedi.umd.edu/

Vegetation Index	Formula
Normalized Difference Vegetation Index (NDVI)	**(B08-B04) / (B08+B04)**
Normalized Difference Index (NDVI45)	**(B05-B04) / (B05+B04)**
Green Normalized Difference Vegetation Index (GNDVI)	**(B08 - B03) / (B08 + B03)**
Vegetation Index green (VIgreen)	**(B03-B04) / (B03+B04)**
Red Edge Normalized Difference Vegetation Index (RENDVI)	**(B07-B04) / (B07+B04)**
Normalized Difference Infrared Index (NDII)	**(B8A-B11) / (B8A+B11)**
Specific Leaf Area Vegetation Index (SLAVI)	**B8A / (B05+B 12)**

TABLE I: Description of Sentinel 2 (S2) vegetation indices.

in the zonal direction (parallel to lines of latitude). Each beam transect includes footprint samples with an approximate spacing of 25 meters along the track, sampled at intervals of roughly 60 meters. The data products are categorized into various levels, reflecting the extent of post-collection processing, including lower-level products (L1 & L2) and higher-level products (L3 & L4).

Our study primarily relies on L2B data, providing canopy height and profile metrics. We access the data through the Earth Science Data Systems (ESDS) Program, covering three years from January 1, 2021, to April 5, 2024. Within our area of interest, the dataset consists of 96.5 GB distributed across 174 data granule files in .h5 format. Figure 2b depicts the data footprints within Cuc Phuong - Pu Luong national forests.

B. Preprocessing of GEDI Data

After data collection, it is necessary to conduct filtering and removal of inaccurate and poor-quality data points due to potential interference from weather conditions such as rain, fog, or sunlight, among other factors, which may render some waveforms unusable for providing information on the vertical forest structure. Three primary indices were chosen for this purpose: (1) the *quality_flag*, where a value of 0 indicates poor quality and a value of 1 signifies that the laser shot meets specific criteria based on factors like energy, sensitivity, amplitude, and real-time surface tracking quality; (2) the *degrade_flag*, which has a value greater than zero if the shot occurs during a degrade period, and zero otherwise; and (3) the Sensitivity layer, using a threshold of 0.95. Additionally, footprints were selected based on the Plant Area Index (PAI) to ensure that data points lacking vegetation are excluded. Within the Region of Interest (ROI), a total of 1,780,874 footprints were obtained, out of which 114,943 points met the required quality standards, accounting for 6.5% of the total. This filtered dataset will serve as the basis for training and evaluation in subsequent experiments.

C. Sentinel 2 (S2) vegetation indices

In addition to the surface reflectance (data bands), we derived 7 vegetation indices from S2 data. These indices, along with their respective formulas, are detailed in Table I. As part of the computation process, bands with lower resolutions (20 and 60m) will initially be upsampled to 10m resolution, followed by standard calculations. We will incorporate these indices into our training data configuration for their substantial vegetation-related insights. Consequently, in our experiments, we will explore and evaluate the effects of integrating supplementary indices on the canopy height model.

IV. Methodology

A. Network Architecture

Given the image patches $\mathbf{x} \in \mathrm{R}^{H \times W \times C}$, where H and W denote the patch's height and width respectively, and C represents the number of channels in the image. In the dataset, the labels associated with the patches consist of matrices $\mathbf{y}$ having the same dimensions as the input image patch. Due to data sparsity within the matrix, only certain elements possess values, while the remainder are NaN (Not a Number). The objective is to construct a deep regression network that takes image patches as input and outputs a matrix representing the corresponding height.

Our proposed network is a tailored and refined version derived from the Xception framework [40] (the detailed structure of the network is depicted in Figure 3). It comprises three primary elements: Entry flow (a PointwiseBlock); succeeded by Middle flow with eight identical separable convolutions (SepConv) blocks and subsequent feature aggregation layers; and lastly, a Prediction flow for generating the final regression forecasts. Throughout the process of developing the network architecture, we recognized the importance of the model's capacity to comprehend the structure and terrain depicted in the original image. Consequently, integrating two feature streams behind the SepConv block with each other using the coefficient σ enhances the emphasis on critical features more effectively. Moreover, combining them with the original features (post Pointwise block) aids in assimilating structural features from the original image.

The entry flow is a PointwiseBlock, gradually increasing the channel depth of the data cube to 728 channels. The block consists of three layers, each layer comprising convolution, batch normalization, and ReLU activation, respectively.

The middle flow begin with a SepConv Block. This block starts with the activation function ReLU and is subsequently succeeded by a depthwise separable convolution layer (as illustrated in Figure 3, adapted from [40]). Within this layer, the overall 2D convolution using a 3D kernel is broken down

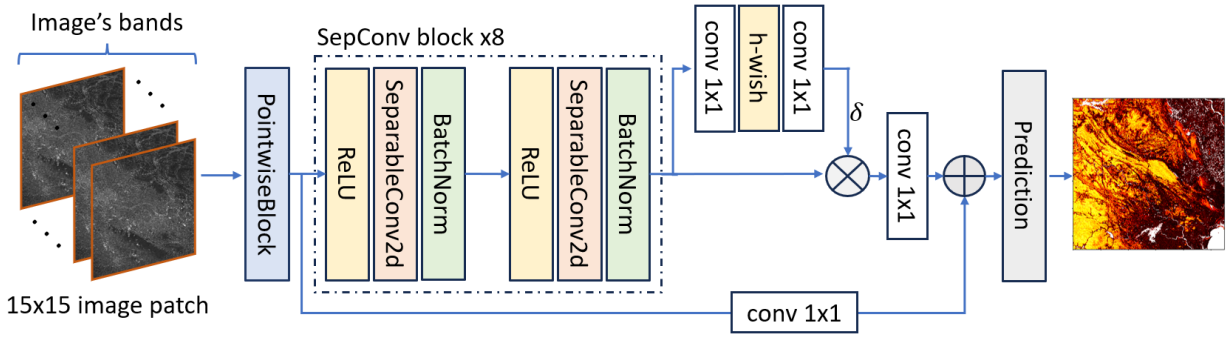


Fig. 3: Overall architecture of the canopy height model.

into 3×3 2D kernels applied to each input channel, alongside a 1D kernel (a linear combination) that combines the results from all channels. This decomposition separates spatial and inter-channel correlations, thus reducing the number of parameters that need to be trained. The parameter count is further minimized by utilizing the activation maps of the spatial kernels (3×3) as input to all subsequent point-wise kernels. Following the separable convolution is a batch normalization (BatchNorm), which normalizes the data cube for a batch (training instances are processed through the network in small batches).

Subsequently, for the feature aggregation segment, we adopt an approach inspired by the structure of the Squeeze and Excitation-ResNet module [41]–[43]. Here, the input to the SepConv blocks is bifurcated into two directions, with the first direction remaining unaltered. In the second direction, the embedding undergoes a 1×1 convolution layer, followed by an h-wish layer, and again a *conv* 1×1. Subsequently, these two directions are amalgamated via element-wise multiplication with a coefficient σ for the second direction. The resulting output then passes through a *conv* 1×1 and is ultimately aggregated with the features obtained from the Entry flow (after undergoing a 1×1 convolution). The resulting final feature possesses a size of $(H, W, 728)$.

The prediction flow is the final layer of our network, which consists of a pointwise convolution that merges the 728 activation maps into a singular canopy height value for each pixel.

B. Loss Functions

Given that we're dealing with a regression task involving continuous height values as targets, our approach involves minimizing the loss function. In this context, we use Gaussian Negative Log-likelihood Loss (GNLL) for learning strategy.

The model assigns each pixel's output as a conditional Gaussian probability distribution representing potential canopy heights, based on the input data:

$$\mathcal{L}_{\text{GNLL}} = \frac{1}{n} \sum_{i=1}^{n} \frac{\left(\hat{\mu}\left(\mathbf{x}_i\right) - \mathbf{y}_i\right)^2}{2\hat{\sigma}^2\left(\mathbf{x}_i\right)} + \frac{1}{2} \log \hat{\sigma}^2\left(\mathbf{x}_i\right).$$

Where $\mathbf{y}_i$ represents the actual canopy height for the i-th pixel, $\hat{\mu}\left(\mathbf{x}_i\right)$ and $\hat{\sigma}^2\left(\mathbf{x}_i\right)$ denotes the estimated mean and variance of the Gaussian distribution for the i-th pixel, n denotes the number of data points.

Minimizing this loss function during training enables the model to make accurate predictions of canopy heights while accounting for prediction uncertainty captured by the variance term.

V. EXPERIMENTAL RESULTS

A. Dataset Configurations and Experimental Settings

	Cuc Phuong - Pu Luong	Pu Hoat - Pu Huong
Train	55,467	171,275
Test	6,169	17,132

TABLE II: Dataset statistic

1) Training and Testing Datasets: The initial satellite image dataset comprises 12 bands with varying resolutions of 10m, 20m, and 60m, respectively. Consequently, in order to train the neural network model effectively, it is necessary to standardize the image resolutions. We opt to standardize them to 10m resolution, thus the initial step involves upscaling the 20m and 60m bands to match this resolution using bicubic interpolation. The resulting dataset is an image tensor with dimensions (10980, 10980, 12). Furthermore, we generate a corresponding matrix that delineates the coordinates of each pixel within the image. Subsequently, the resulting image is partitioned into patches of size 15×15. However, given the sparse nature of the reference data, specifically GEDI, it is imperative to exclude patches devoid of pixels labeled with canopy height. Table II illustrates the count of patches within the training and testing sets (in a 9:1 ratio) post-processing. These represent the images that will be directly inputted into the model during the training phase.

2) Experimental Settings: Our fine-tuned model, based on the architecture as outlined in IV-A, was trained on the dataset we collected. The model was trained on 6 NVIDIA RTX A5000 devices with 24GB RAM each, training spanned 1000 epochs, employing a batch size of 512. The initial learning rate coefficient was set to 4e-4, and at epochs 200, 400, and 700, it was halved.

In comparison to baseline models, we opted for 3 pre-trained models from [39], namely: FT_ALL_CB,

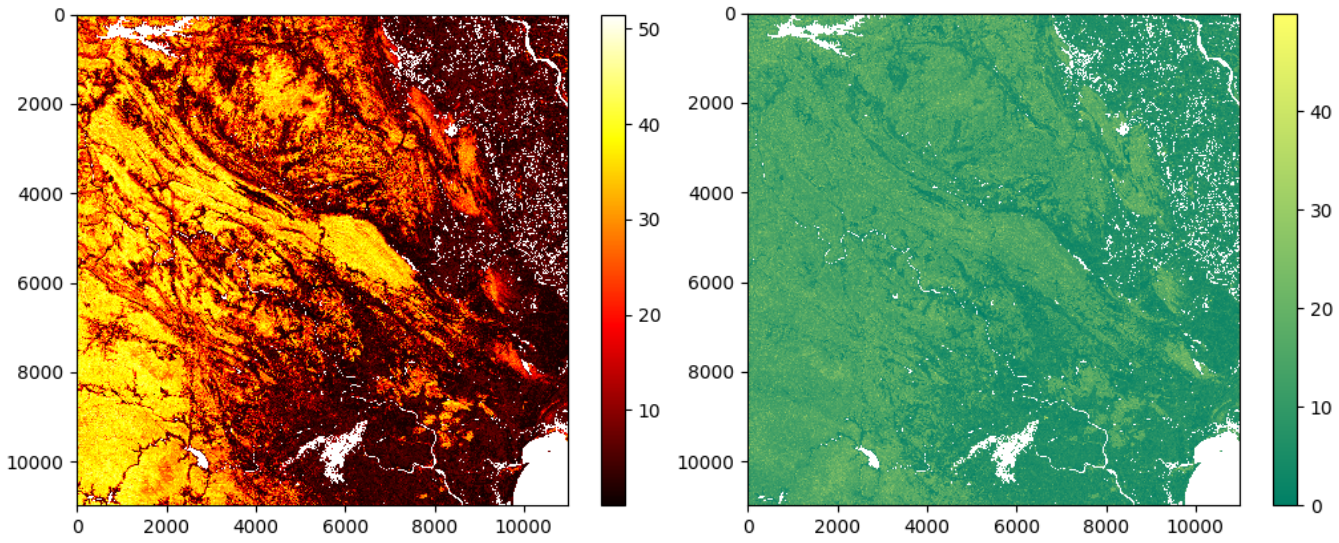


Fig. 4: The canopy height (left) and its variance (right) predictions.

Method	Cuc Phuong - Pu Luong			Pu Hoat - Pu Huong		
	MAPE	RMSE	MAE	MAPE	RMSE	MAE
FT_ALL_CB	0.6982	12.8011	9.2544	0.5937	12.2581	9.4707
FT_Lm_SRCB	0.7833	12.1848	9.0029	0.5848	12.2261	9.2667
ST_geoshift_IB	0.5231	12.1373	8.8225	0.8002	12.8162	9.6150
7 indices (ours)	0.4129	11.8678	8.1286	0.5972	11.9923	8.4850
12 bands + 7 indices (ours)	0.3510	**10.1131**	7.2952	0.3029	**9.9108**	7.6512
12 bands (ours)	**0.3244**	10.1822	**7.1172**	**0.2862**	9.9720	**7.5441**

TABLE III: Canopy-height estimation errors

FT_Lm_SRCB, and ST_geoshift_IB (the specific configurations of these models are elaborated in the code[4]). These models underwent training and evaluation on a global dataset.

B. Comparison with Existing Methods

The table III presents comparison results between different methods across two regions: Cuc Phuong - Pu Luong and Pu Hoat - Pu Huong. Each method is evaluated based on three metrics: Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE).

Our methodology employs three different data configurations: one exclusively containing vegetation indices, another combining 12 bands with indices, and the last one comprising solely bands. The findings from our analysis reveal that our fine-tuned models significantly outperform others in terms of MAPE, RMSE, and MAE across both regions. Notably, the '12 bands + 7 indices' approach demonstrates the lowest RMSE values, while using only bands produces the most favorable outcomes concerning MAPE and MAE.

From the results obtained from the three proposed configurations, it becomes evident that utilizing only indices leads to some enhancement compared to baseline models, albeit not significantly. Combining bands and indices slightly improves over the case of using bands alone in terms of RMSE, however, it does not yield significantly superior results. Nevertheless, we can still discern the potential for further development of the model in terms of minimizing errors in the future.

C. Qualitative Study

We conducted predictions based on satellite imagery inputs of the Cuc Phuong - Pu Luong area, with thermal maps illustrating the results as shown in Figure 4. The visual representation indicates that the model can relatively well capture information about the image structure. In areas corresponding to lakes or rivers, the result maps appear white, indicating extremely high values. Additionally, we can discern populated areas and road sections represented by deep red indicating low height values. Forested areas, indicated by a range of bright red to yellow, represent heights from 20 - 40m. These observations demonstrate that the model has been able to learn relatively effectively and can provide forecasts across the entire AOI.

[4]https://github.com/langnico/global-canopy-height-model.git

REFERENCES

[1] Mai, D., & Yen, M. (2018). PRELIMINARY ASSESSMENT OF BIODIVERSITY / ECOSYSTEMS VULNERABILITY BY CLIMATE

CHANGE AND SUGGESTION A SYSTEM OF MITIGATION MEASURES FOR THEM, CASE STUDY: HANOI CITY, VIETNAM. EPH - International Journal of Biological & Pharmaceutical Science. https://doi.org/10.53555/eijbps.v4i1.25.

[2] Michael Köhl, Rodel Lasco, Miguel Cifuentes, Örjan Jonsson, Kari T. Korhonen, Philip Mundhenk, Jose de Jesus Navar, Graham Stinson, Changes in forest production, biomass and carbon: Results from the 2015 UN FAO Global Forest Resource Assessment, Forest Ecology and Management, Volume 352, 2015, Pages 21-34, ISSN 0378-1127, https://doi.org/10.1016/j.foreco.2015.05.036.

[3] Corte, A., Rex, F., Almeida, D., Sanquetta, C., Silva, C., Moura, M., Wilkinson, B., Zambrano, A., Neto, E., Veras, H., Moraes, A., Klauberg, C., Mohan, M., Cardil, A., & Broadbent, E. (2020). Measuring Individual Tree Diameter and Height Using GatorEye High-Density UAV-Lidar in an Integrated Crop-Livestock-Forest System. Remote. Sens., 12, 863. https://doi.org/10.3390/rs12050863.

[4] Kotivuori, E., Kukkonen, M., Mehtätalo, L., Maltamo, M., Korhonen, L., & Packalen, P. (2020). Forest inventories for small areas using drone imagery without in-situ field measurements. Remote Sensing of Environment, 237, 111404. https://doi.org/10.1016/j.rse.2019.111404.

[5] Woo, H., Acuna, M., Choi, B., & Han, S. (2021). FIELD: A Software Tool That Integrates Harvester Data and Allometric Equations for a Dynamic Estimation of Forest Harvesting Residues. Forests. https://doi.org/10.3390/f12070834.

[6] Sun, Z., Sonsuthi, A., Jucker, T., Ali, A., Cao, M., Liu, F., Cao, G., Hu, T., Ma, Q., Guo, Q., & Lin, L. (2023). Top Canopy Height and Stem Size Variation Enhance Aboveground Biomass across Spatial Scales in Seasonal Tropical Forests. Plants, 12. https://doi.org/10.3390/plants12061343.

[7] Vargas-Larreta, B., López-Martínez, J., González, E., Corral-Rivas, J., & Hernández, F. (2020). Assessing above-ground biomass-functional diversity relationships in temperate forests in northern Mexico. Forest Ecosystems, 8, 1-14. https://doi.org/10.21203/rs.3.rs-42734/v2.

[8] Lang, N., Jetz, W., Schindler, K. et al. A high-resolution canopy height model of the Earth. Nat Ecol Evol 7, 1778–1789 (2023). https://doi.org/10.1038/s41559-023-02206-6

[9] Garrido, A., Gobakken, T., Ørka, H., Næsset, E., & Bollandsås, O. (2020). Reuse of field data in ALS-assisted forest inventory. Silva Fennica, 54. https://doi.org/10.14214/sf.10272.

[10] Bont, L., Hill, A., Waser, L., Bürgi, A., Ginzler, C., & Blattert, C. (2020). Airborne-laser-scanning-derived auxiliary information discriminating between broadleaf and conifer trees improves the accuracy of models for predicting timber volume in mixed and heterogeneously structured forests. Forest Ecology and Management. https://doi.org/10.1016/j.foreco.2019.117856.

[11] Shang, C., Coops, N., Wulder, M., White, J., & Hermosilla, T. (2020). Update and spatial extension of strategic forest inventories using time series remote sensing and modeling. Int. J. Appl. Earth Obs. Geoinformation, 84. https://doi.org/10.1016/j.jag.2019.101956.

[12] Kovanič Ľ, Topitzer B, Peťovský P, Blišťan P, Gergeľová MB, Blišťanová M. Review of Photogrammetric and Lidar Applications of UAV. Applied Sciences. 2023; 13(11):6732. https://doi.org/10.3390/app13116732

[13] H. Sier, X. Yu, I. Catalano, J. P. Queralta, Z. Zou and T. Westerlund, "UAV Tracking with Lidar as a Camera Sensor in GNSS-Denied Environments," 2023 International Conference on Localization and GNSS (ICL-GNSS), Castellón, Spain, 2023, pp. 1-7, doi: 10.1109/ICL-GNSS57829.2023.10148919.

[14] I. Catalano, H. Sier, X. Yu, T. Westerlund and J. P. Queralta, "UAV Tracking with Solid-State Lidars: Dynamic Multi-Frequency Scan Integration," 2023 21st International Conference on Advanced Robotics (ICAR), Abu Dhabi, United Arab Emirates, 2023, pp. 417-424, doi: 10.1109/ICAR58858.2023.10406884.

[15] Segarra J, Buchaillot ML, Araus JL, Kefauver SC. Remote Sensing for Precision Agriculture: Sentinel-2 Improved Features and Applications. Agronomy. 2020; 10(5):641. https://doi.org/10.3390/agronomy10050641

[16] Naghdyzadegan Jahromi, M., Zand-Parsa, S., Doosthosseini, A., Razzaghi, F., Jamshidi, S. (2022). Enhancing Vegetation Indices from Sentinel-2 Using Multispectral UAV Data, Google Earth Engine and Machine Learning. In: Bozorg-Haddad, O., Zolghadr-Asli, B. (eds) Computational Intelligence for Water and Environmental Sciences. Studies in Computational Intelligence, vol 1043. Springer, Singapore. https://doi.org/10.1007/978-981-19-2519-1_24

[17] Naghdyzadegan Jahromi, M., Zand-Parsa, S., Doosthosseini, A., Razzaghi, F., Jamshidi, S. (2022). Enhancing Vegetation Indices from Sentinel-2 Using Multispectral UAV Data, Google Earth Engine and Machine Learning. In: Bozorg-Haddad, O., Zolghadr-Asli, B. (eds) Computational Intelligence for Water and Environmental Sciences. Studies in Computational Intelligence, vol 1043. Springer, Singapore. https://doi.org/10.1007/978-981-19-2519-1_24

[18] Ghosh SM, Behera MD, Paramanik S. Canopy Height Estimation Using Sentinel Series Images through Machine Learning Models in a Mangrove Forest. Remote Sensing. 2020; 12(9):1519. https://doi.org/10.3390/rs12091519

[19] Nandy, Subrata & Srinet, Ritika & Padalia, Hitendra. (2021). Mapping Forest Height and Aboveground Biomass by Integrating ICESat-2, Sentinel-1 and Sentinel-2 Data Using Random Forest Algorithm in Northwest Himalayan Foothills of India. Geophysical Research Letters. 48. 10.1029/2021GL093799. https://doi.org/10.1029/2021GL093799

[20] Chere, Z., Zewdie, W. & Biru, D. Machine learning for modeling forest canopy height and cover from multi-sensor data in Northwestern Ethiopia. Environ Monit Assess 195, 1452 (2023). https://doi.org/10.1007/s10661-023-12066-z

[21] Reichstein, M., Camps-Valls, G., Stevens, B. et al. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204 (2019). https://doi.org/10.1038/s41586-019-0912-1

[22] Fayad, Ibrahim & Ienco, Dino & Baghdadi, Nicolas & Gaetano, Raffaele & Alcarde Alvares, Clayton & Stape, Jose & Scolforo, Henrique & le Maire, Guerric. (2021). A CNN-based approach for the estimation of canopy heights and wood volume from GEDI waveforms. Remote Sensing of Environment. 265. 16. 10.1016/j.rse.2021.112652.

[23] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature 521, 436–444 (2015). https://doi.org/10.1038/nature14539

[24] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 2261-2269, doi: 10.1109/CVPR.2017.243.

[25] Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. IEEE Geoscience and Remote Sensing Magazine, 5(4), 8-36. https://doi.org/10.1109/MGRS.2017.2762307

[26] Mahesh, Ragini & Hänsch, Ronny. (2023). Deep Learning for Forest Canopy Height Estimation from SAR. 10.1109/IGARSS52108.2023.10281899.

[27] S. Oehmcke, T. Nyegaard-Signori, K. Grogan and F. Gieseke, "Estimating Forest Canopy Height With Multi-Spectral and Multi-Temporal Imagery Using Deep Learning," 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 2021, pp. 4915-4924, doi: 10.1109/BigData52589.2021.9672018.

[28] Malambo L, Popescu S. Image to Image Deep Learning for Enhanced Vegetation Height Modeling in Texas. Remote Sensing. 2023; 15(22):5391. https://doi.org/10.3390/rs15225391

[29] R. Ranftl, A. Bochkovskiy and V. Koltun, "Vision Transformers for Dense Prediction," 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 12159-12168, doi: 10.1109/ICCV48922.2021.01196.

[30] Dosovitskiy, Alexey & Beyer, Lucas & Kolesnikov, Alexander & Weissenborn, Dirk & Zhai, Xiaohua & Unterthiner, Thomas & Dehghani, Mostafa & Minderer, Matthias & Heigold, Georg & Gelly, Sylvain & Uszkoreit, Jakob & Houlsby, Neil. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, doi: https://doi.org/10.48550/arXiv.2010.11929

[31] Xu, Z., Zhang, W., Zhang, T., Yang, Z., Li, J., 2021. Efficient transformer for remote sensing image segmentation. Remote Sens. 13 https://doi.org/10.3390/rs13183585

[32] Wang, W., Tang, C., Wang, X., Zheng, B., 2022. A ViT-based multiscale feature fusion approach for remote sensing image segmentation. IEEE Geosci. Remote Sens. Lett. 19, 1–5. https://doi.org/10.1109/LGRS.2022.3187135

[33] Gibril, M.B.A., Shafri, H.Z.M., Al-Ruzouq, R., Shanableh, A., Nahas, F., Al Mansoori, S., 2023. Large-scale date palm tree segmentation from multiscale UAV-based and aerial images using deep vision transformers. Drones 7. https://doi.org/10.3390/ drones7020093.

[34] Reed, Colorado & Gupta, Ritwik & Li, Shufan & Brockman, Sarah & Funk, Christopher & Clipp, Brian & Candido, Salvatore & Uyttendaele, Matt & Darrell, Trevor. (2022). Scale-MAE: A Scale-Aware Masked Autoencoder for Multiscale Geospatial Representation Learning. 10.48550/arXiv.2212.14532.

[35] Schacher, A., Roger, E., Williams, K.J., Stenson, M.P., Sparrow, B., Lacey, J., 2023. Usespecific considerations for optimizing data quality trade-offs in citizen science: recommendations from a targeted literature review to improve the usability and utility for the calibration and validation of remotely sensed products. Remote Sens. 15 https://doi.org/10.3390/rs15051407.

[36] Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.Y., Li, S.W., Misra, I., Rabbat, M., Sharma, V., Synnaeve, G., Xu, H., Jegou, H., Mairal, J., Labatut, P., Joulin, A., Bojanowski, P., 2023. Dinov2: learning robust visual features without supervision arXiv:2304.07193. URL: https://arxiv.org/abs/2304.07193

[37] Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Bouchareb, Y.S.E., Dauphin, Y.N., Keysers, D., Neumann, M., Ciss´ e, M., Quinn, J., 2021. Continental-scale building detection from high resolution satellite imagery. CoRR abs/2107.12283. URL: https://doi.org/10.48550/arXiv.2107.12283

[38] Fayad, I., Ciais, P., Schwartz, M., Wigneron, J.P., Baghdadi, N., de Truchis, A., d'Aspremont, A., Frappart, F., Saatchi, S., Pellissier-Tanon, A., Bazzi, H., 2023. Vision transformers, a new approach for high-resolution and large-scale mapping of canopy heights arXiv:2304.11487. URL: https://doi.org/10.48550/arXiv.2304.11487

[39] Lang, N., Jetz, W., Schindler, K., & Wegner, J. D. (2023). A high-resolution canopy height model of the Earth. Nature Ecology & Evolution, 1-12.

[40] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017 pp. 1800-1807. doi: 10.1109/CVPR.2017.195

[41] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018 pp. 4510-4520. doi: 10.1109/CVPR.2018.00474

[42] J. Hu, L. Shen and G. Sun, "Squeeze-and-Excitation Networks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 7132-7141, doi: 10.1109/CVPR.2018.00745.

[43] Yang, C., Qiao, S., Yu, Q., Yuan, X., Zhu, Y., Yuille, A.L., Adam, H., & Chen, L. (2022). MOAT: Alternating Mobile Convolution and Attention Brings Strong Vision Models. ArXiv, abs/2210.01820.