Pontifícia Universidade Católica
do Rio de Janeiro

Paola Edith Ayma Quirita

An evaluation of Deep Learning techniques for forest parameters estimation in the Brazilian Legal Amazon from multi-source remote sensing imagery

Dissertação de Mestrado

Thesis presented to the Programa de Pós–graduação em Engenharia Elétrica, do Departamento de Engenharia Elétrica da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Engenharia Elétrica.

Advisor : Prof. Marco Aurelio Pacheco
Coadvisor: Prof. Manoela Rabello Kohler

Rio de Janeiro
October 2024

Pontifícia Universidade Católica
do Rio de Janeiro

Paola Edith Ayma Quirita

An evaluation of Deep Learning techniques for forest parameters estimation in the Brazilian Legal Amazon from multi-source remote sensing imagery

Thesis presented to the Programa de Pós–graduação em Engenharia Elétrica da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Engenharia Elétrica. Approved by the Examination Committee:

Prof. Marco Aurelio Pacheco
Advisor
Departamento de Engenharia Elétrica – PUC-Rio

Prof. Manoela Rabello Kohler
Coadvisor
ICA – PUC-Rio

Cristina Maria Bentz
CENPES – Petróleo Brasileiro

Pedro Marco Achanccaray Diaz
Technische Universität Braunschweig – TU BS

Jose David Bermudez Castro
McMaster

Vitor Bento de Sousa
ICA – PUC-Rio

Rio de Janeiro, October the 1st, 2024

Paola Edith Ayma Quirita

The author received her bachelor's degree in Electronic Engineering at the Universidad Nacional de San Antonio Abad del Cusco (Cusco, Perú) in 2019.

Bibliographic data

Ayma Quirita, Paola Edith

An evaluation of Deep Learning techniques for forest parameters estimation in the Brazilian Legal Amazon from multi-source remote sensing imagery / Paola Edith Ayma Quirita; advisor: Marco Aurelio Pacheco; coadvisor: Manoela Rabello Kohler. – 2024.

76 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2024.

Inclui bibliografia

1. Engenharia Elétrica – Teses. 2. Estimação de parâmetros florestais. 3. Amazônia brasileira. 4. Sensoriamento Remoto de Múltiplas Fontes. 5. Aprendizagem Profunda. 6. Redes Residuais. 7. Saída multimodal Dual UNet. 8. ViT-ConvNet híbrida. I. Pacheco, Marco. II. Kohler, Manoela. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. IV. Título.

CDD: 620.11

To my parents, Victor and Rina, for their support,
encouragement and guidance.

Acknowledgments

First of all, I would like to thank my parents, their teachings and thoughts were the basis of this work. To my brothers and sisters for their encouragement. To my nieces and nephews, who taught me the value of patience. To my love, Franklin, for your support and listening.

I am very grateful to my advisor, Prof. Marco Aurelio, for the opportunity, support and understanding. I would also like to thank Prof. Manoela Kohler, whose guidance and expertise have been invaluable throughout this project.

Finally, I wish to thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil (CNPq), and the Pontifical Catholic University of Rio de Janeiro (PUC-Rio), without them this work would not have been possible. Also, this study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Abstract

Ayma Quirita, Paola Edith; Pacheco, Marco (Advisor); Kohler, Manoela (Co-Advisor). **An evaluation of Deep Learning techniques for forest parameters estimation in the Brazilian Legal Amazon from multi-source remote sensing imagery**. Rio de Janeiro, 2024. 76p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

In recent years, estimating forest parameters such as Tree Height (CH) and AboveGround Biomass (AGB) has gained importance due to their essential role in understanding the global carbon cycle, mitigating climate change, and preventing biodiversity loss. Accurate inference of these parameters is crucial because they are key indicators of forest health and carbon storage capacity. The Brazilian Amazon, a vital tropical forest, plays a crucial role in absorbing as much carbon as is released through deforestation and degradation. Understanding and monitoring CH and AGB enable better management and conservation strategies and promote sustainable practices. Traditionally, these forest parameters have been estimated through ground-based methods, such as forest inventory plots, which involve physically measuring trees. While these methods are highly accurate, they are labor-intensive and often impractical for large-scale assessments due to the vast and inaccessible nature of forests. Additionally, the application of Machine Learning (ML) and Deep Learning (DL) techniques offers significant advantages over traditional methods, providing rapid and scalable solutions for estimating forest parameters across extensive areas. Moreover, they can integrate data from various sources, enhancing the robustness of the estimates. While many studies have utilized forest inventory plots, RS, and ML techniques, DL techniques remain underexplored in studies within the Brazilian Amazon. This study aims to evaluate DL techniques for estimating TH and AGB in dense tropical forests using various RS imagery, including Sentinel-1, ALOS-2/PALSAR-2, Sentinel-2, and GEDI. Three DL models were tested for CH estimation, where the best of the models achieve a R^2 of 0.751, an MAE of 4.068 meters, and an RMSE of 5.737 meters. Furthermore, various ML techniques were evaluated for AGB estimation, resulting in an R^2 of 0.648, an MAE of 48.842 Mg·ha^{-1}, and RMSE of 70.745 Mg·ha^{-1}.

Keywords

Forest parameters estimation; Brazilian Amazon; Multisource Remote Sensing; Deep learning; Residual Network; Multimodal Output Dual Unet; Hybrid ViT-ConvNet.

Resumo

Ayma Quirita, Paola Edith; Pacheco, Marco; Kohler, Manoela. **Avaliação de modelos de Deep Learning para estimação de parâmetros de floresta na Amazônia Brasileira Legal a partir de imagens de sensoriamento remoto**. Rio de Janeiro, 2024. 76p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Nos últimos anos, a estimativa de parâmetros florestais, como a altura das árvores (CH) e a biomassa acima do solo (AGB) tem ganhado muita importância devido ao seu papel essencial na compreensão do ciclo global do carbono, na mitigação das mudanças climáticas e na prevenção da perda de biodiversidade. A inferência precisa desses parâmetros é crucial porque eles são indicadores chave da saúde da floresta e da capacidade de armazenamento de carbono. A Amazônia brasileira, uma floresta tropical vital, desempenha um papel crucial na absorção de tanto carbono quanto o que é liberado pelo desmatamento e pela degradação. A compreensão e o monitoramento da CH e da AGB permitem melhores estratégias de gestão e conservação, e promovem práticas sustentáveis. Tradicionalmente, esses parâmetros florestais, têm sido estimados por meio de métodos de campo, como inventário florestal, que envolvem medição física das árvores. No entanto, esses métodos são altamente precisos, são trabalhosos e, muitas vezes, impraticáveis para avaliações em larga escala devido à natureza vasta e inacessível das florestas. Além disso, a aplicação de técnicas de aprendizado de máquina (ML) e aprendizado profundo (DL) oferece vantagens significativas em relação aos métodos tradicionais, fornecendo soluções rápidas e dimensionáveis para a estimação dos parâmetros florestais em áreas extensas. Além disso, esses métodos podem integrar dados de diferentes fontes, aumentando a robustez das estimativas. Embora muitos estudos tenham utilizado dados de inventário florestal, RS e técnicas de ML, as técnicas de DL permanecem pouco exploradas em estudos na Amazônia brasileira. Este estudo visa avaliar técnicas de DL para estimar a CH e a AGB em florestas tropicais densas usando diferentes imagens de RS, incluindo o Sentinel-1, ALOS-2/PALSAR-2, Sentinel-2 e GEDI. Três modelos de DL foram testados para a estimativa da CH, sendo que o melhor modelo alcançou um R^2 de 0.751, um MAE de 4.068 metros, e um RMSE 5.737 metros. Além disso, várias técnicas de ML foram avaliadas para a estimativa de AGB, resultando em um R^2 de 0.648, MAE de 48.842 $Mg\cdot\text{ha}^{-1}$, e RMSE of 70.745 $Mg\cdot\text{ha}^{-1}$.

Palavras-chave

Estimação de parâmetros florestais; Amazônia brasileira; Sensoriamento Remoto de Múltiplas Fontes; Aprendizagem Profunda; Redes Residuais; Saída multimodal Dual UNet; ViT-ConvNet híbrida.

Table of contents

List of figures

List of tables

List of Abreviations

CH – Canopy Height

AGB – Aboveground Biomass

ML – Machine Learning

DL – Deep Learning

RS – Remote Sensing

IAF – International Arrangement on Forest

REDD – Reducing Deforestation and Forest Degradation

BGB – BelowGround Biomass

SAR – Synthetic Aperture Radar

LiDAR – Light Detection and Ranging

RF – Random Forest

GEDI – Global Ecosystem Dynamics Investigation

ALS – Airborne Laser Scanning

CNN – Convolutional Neural Network

InSAR – Interferometric Synthetic Aperture Radar

ViT – Vision Transformers

TM – Thematic Mapper

HSI – Hyperspectral Imagery

EC – European Comission

ESA – European Space Agency

IW – Interferometric wide

GRD – Ground Range Detected

JAXA – Japan Aerospace Exploration Agency

MLP – MultiLayer Perceptron

LR – Linear Regression

RF – Random Forest

CEB – Convolutional Encoder Block

CDB – Convolution Decoder Block

SAA – Self-Aware Attention

TSA – Transformer Self-Attention

GSA – Global Spatial Attention

NLP – Natural Language Processing

MSA – Multi-head Self-Attention

KD – Knowledge Distillation

XGBoost – Extreme Gradient Boosting

GEE – Google Earth Engine

INPE – Instituto Nacional de Pesquisas Espaciais

RH – Relative Height

VI – Vegetation Index

AC – Acre

AP – Amapá

AM – Amazonas

MA – Maranhão

MT – Mato Grosso

PA – Pará

RR – Roraima

RO – Rondônia

TO – Tocantinss

NISAR – NASA-ISRO SAR

Quand tu regarderas le ciel, la nuit, puisque j'habiterai dans l'une d'elles, puisque je rirai dans l'une d'elles, alors ce sera pour toi comme si riaient toutes les étoiles. Tu auras, toi, des étoiles qui savent rire!

Antoine de Saint-Exupéry, *Le Petit Prince.*

1
INTRODUCTION

1.1
Motivation

Forests, characterized by their arboreal cover and diverse ecosystems, are of vital importance due to their significant ecological and environmental contributions. These natural habitats, dominated by trees and where animals, people, and other living organisms interact and benefit from, play an essential role as a water source. Forests not only enhance water quality and mitigate soil erosion [1], but they also regulate water cycles, thus providing habitats for numerous species, all of which contribute to the overall health and stability of ecosystems [2]. On the other hand, forests also act as an essential carbon sink, absorbing carbon dioxide from the atmosphere and helping mitigate climate change [3]. These characteristics, in addition to the complex relationships between species diversity and forest attributes such as age, height, and biomass, underscore the vital role of forests in maintaining ecological balance and biodiversity [4], as well as promoting global biological evolution and community succession [5] and mitigating climate change.

For these reasons, many governments worldwide have joined forces to combat global warming and conserve the Earth's forests. Some initiatives are the International Arrangement on Forests (IAF) [6], the Paris Agreement [7], and Reducing Emissions from Deforestation and Forest Degradation (REDD) [8]. The IAF, established in October 2000 by the United Nations Economic and Social Council, aims to promote forest management, conservation, and sustainable development and strengthen long-term political commitment to this end. Later, in 2017, the goal of increasing forest area by 3% globally by 2030 was included. On the other hand, the Paris Agreement is an international climate change treaty held in Paris in 2015. Its overall objective is to limit the increase in the global average temperature to well below 2°C above pre-industrial levels. Furthermore, as significant carbon absorbers, forests are key players in climate regulation. However, when destroyed or damaged, they can also be a source of greenhouse gas emissions. To address this, countries established the REDD+ framework as part of the Paris Agreement to protect forests, conserving and enhancing forest carbon stocks.

These arrangements represent the basis for conserving forests around the world. One of the countries that has become part of an initiative to conserve and protect forests and combat climate change is Brazil, which is the second country with the largest forest area in the world [9], corresponding to 59% of its surface, this area comprises 5.0 million Km^2, is home to 56% of Brazil's Indigenous population and host the world's greatest biodiversity [10]. This biodiversity is due to the pairing of plant and animal living organisms in similar geoclimatic conditions and a shared history of changes. In this sense, Brazil encompasses several climatic zones that lead to great ecological variations and form distinct biogeographical zones or biomes: Amazon forest, Pantanal, Cerrado, Caatinga, Pampas, and Atlantic forest [11].

On the other hand, in line to combat climate change and preserve the Earth's forests, researchers are studying the relationships between species diversity and forest attributes [4]. They are also working on methods to estimate and map forest parameters such as height, biomass, and carbon storage. While height measures the tree's vertical growth, biomass refers to the amount of living material in a given area [12] and can be influenced by the tree's height. Forest biomass consists of aboveground biomass (AGB), which includes stems, branches, seeds, and foliage, and belowground biomass (BGB), which comprises all live roots with a diameter greater than 2mm [13]. These parameters are key indicators of forest health and carbon storage capacity. Additionally, understanding and monitoring these parameters enable better management and conservation strategies, aiding in the fight against climate change, promoting sustainable practices, and preventing biodiversity loss.

With advances in science and technology, methods to estimate forest parameters have gone through several stages of development, from traditional field measurements to the use of remote sensing (RS) data [14]:

1. Traditional field measurements: This method involves physically measuring trees and constructing relationships between these measurements and forest biomass. Although traditional field measurements are the most accurate methods, they cannot be easily generalized to all areas.
2. RS-based methods: These methods leverage RS imagery and utilize machine learning (ML) and deep learning (DL) techniques to measure and monitor trees and forest biomass effectively.

In recent years, considerable studies worldwide have focused on using RS-based methods to estimate forest parameters efficiently by processing large RS imagery datasets. As mentioned before, each RS data has its own advantages and disadvantages. For example, optical and radar data are widely used to estimate AGB because they are suitable for extracting information about simple and homogeneous forests. However, optical data has limitations, like saturation, especially in forests with high biomass density, such as tropical regions [15], resulting in poor performance in estimating biomass [16]. On the other hand, synthetic aperture radar (SAR) data addresses the saturation optical data problem, as well as penetration aspects, especially with L-band SAR data. However, some limitations of SAR data are identifying uncertainties at different stages of the biomass estimation procedure [15]. While LiDAR can characterize vertically distributed attributes and remove data saturation, its limitation is data availability [16]. Therefore, integrating different sensor data can produce accurate estimates by improving the information extraction process [14]. In addition, these methods offer rapid and scalable solutions for estimating forest parameters over extensive areas. Moreover, based on the literature reviewed, several studies have been conducted using ML techniques to estimate both canopy height (CH) and AGB in diverse forest types. However, implementing DL techniques for AGB remains underexplored, and to the best of our knowledge, no studies have been reported using DL techniques in the Brazilian Amazon.

In this context, the objective of this study is to compare and evaluate DL techniques estimating forest parameters, specifically CH and AGB in the Brazilian legal Amazon territory by integrating multisource RS imagery, encompassing Sentinel-1, ALOS-2/PALSAR-2, Sentinel-2, and GEDI data. The significance of this area lies not only in its extensive forest cover and biodiversity but also in the fact that it is a tropical forest, and this forest type plays an important role in the global carbon cycle by absorbing as much carbon as is released through deforestation and degradation [17]. On the other hand, the use of GEDI data as ground truth represents a challenge, since GEDI presents problems of localization, identification of forests and non-forests, and above all because its data are not continuous, but rather are sparse, which allows the existence of areas without GEDI data.

1.2
Objectives

1.2.1
General Objective

The main objective of this research is to evaluate machine learning and deep learning techniques for forest parameters estimation in the Brazilian Legal Amazon region using optical, radar, and LiDAR remote sensing imagery.

1.2.2
Specific Objectives

The specific objectives of this work are the following:

1. Data fusion of multiple sensors at the same spatial resolution.
2. Analyze models developed for canopy tree height and adapt them for biomass estimation.
3. Evaluate the models for automatic tree height and aboveground biomass estimation.

1.3
Contributions

The main contributions of this research are:

- An evaluation of state-of-the-art Deep Learning techniques for automatic tree height estimation in the Brazilian Legal Amazon.
- An assessment of these methods for biomass estimation in the Brazilian Legal Amazon. As well as assessing the accuracy of these methods under sparse samples.

1.4
Organization of the remaining parts of this thesis

The remainder of this document is as follows:

Chapter 2 presents the related work available in the literature for tree height and aboveground biomass estimation using different approaches, from classic machine learning to deep learning models.

Chapter 3 introduces the fundamental concepts and theory underlying the methods used in this work for a better understanding of the models.

Chapter 4 explains the modifications of the deep learning methods used for tree height and aboveground biomass estimation.

Chapter 5 exhibits the study area and the dataset employed in this work, the experimental setup followed in the experiments, and summarizes the results obtained from each method.

Finally, Chapter 6 presents the conclusions derived from the experiments performed in this research and gives guidelines for further implementations in order to improve the proposed method and results.

2
RELATED WORKS

This chapter presents an overview of different works related to CH and AGB estimation techniques using RS imagery from one or several sources. Some of the works presented in this chapter focus on Brazilian states, especially those belonging to the Brazilian Amazon, as well as other areas studied in the literature.

In recent decades, significant advances have been made in estimating forest structural parameters, including TH and AGB. Different studies have explored the application of ML techniques using single and multisource RS imagery, some of them combined with forest inventory plots data to develop and validate biomass estimation models, most of them addressing the limitations of forest inventory plots.

2.0.1
Canopy height estimation

Many studies have been developed around the world to estimate canopy height. For example, Zhu et al. [18] used forest inventory data and optical and radar RS data, estimating the canopy height using classical ML techniques like Random Forest (RF) in Beijing, China. In addition, the Global Ecosystem Dynamics Investigation (GEDI) mission has provided LiDAR data, which has been integrated with Airborne Laser Scanning (ALS) to improve CH and AGB models [19], but also has been integrated with Sentinel-1 and Sentinel-2 [20], particularly in Mediterranean forests. In the same way, Stojanova et al. [21] used regression trees and ensemble models to model the vegetation height and canopy cover in Slovenia by integrating LiDAR and Landsat satellite data.

DL techniques have also been explored to develop CH estimations. Convolutional neural networks (CNNs) [22] have shown promise in predicting vegetation height using only RGB and NIR (near-infrared) bands of a very high spatial resolution from WorldView-2 satellite bands [23] in Russia, but also in forest mapping using Sentinel-1 Interferometric Synthetic Aperture Radar (InSAR) short time series [24], achieving high accuracy in deforestation detection and forest structure mapping in cloud-covered regions like the Amazon Basin. Similarly, Ge et al. [25] introduced an improved semi-supervised deep learning approach based on UNet [26], demonstrating its ability to model the relationship between forest structural parameters and satellite remote sensing imagery in Finland. This study used RS data from Sentinel-1, Sentinel-2, and ALS. Recent advancements include using Vision Transformers (ViTs) and self-supervised learning [27] using multispectral satellite imagery, which has shown significant improvements over traditional convolutional-based methods.

More specific studies developed across the Brazilian Amazon are presented by Bispo et al. [28], where the authors investigate the potential of TH estimation in Pará state from TanDEM-X InSAR data and a LiDAR digital terrain model for separating successional stages, primary or old growth, and secondary forest at different stages of succession, employing a maximum likelihood classification. Additionally, Sawada et al. [29] aimed to map TH in the central Amazon, using ICESat, MODIS, and terrain and rainfall data with a new method called Self-Organizing Relationships. Similarly, Qin et al. [30] implemented a robust decision tree algorithm to generate annual forest cover maps over the Brazilian Amazon using images from ALOS/PALSAR and MODIS, while ICESat imagery was used to assess the forest cover maps. Moreover, in the study presented in [31], the authors used a DL technique to map canopy palm cover over the Brazilian Amazon, more precisely, UNet architecture. In this study, airborne LiDAr data was employed.

2.0.2
AboveGround Biomass estimation

On the other hand, different studies worldwide also show the benefits of combined RS data and ML algorithms for AGB estimation. In the study realized by Wang et al. [32], the authors compare the performance of tree RS data, ICESat, MODIS, and ALOS/PALSAR, along with three ML algorithms using field measurements in the Khingan Mountains, China. Also, the study presented by Tang et al. [33] was carried out across China and used RS data from MODIS and spaceborne LiDAR, as well as climate factors and topographic data, employing six ML models to estimate AGB. Moreover, in the study from Saatchi et al. [34], the authors determined relationships between RS metrics and AGB from forest plots and used those relationships directly to estimate AGB over the entire Amazon basin, employed a decision tree and adopted a biomass classification approach to segment the image into different AGB ranges. The dataset included both optical and microwave RS sensors from MODIS, JERS-1, and QSCAT.

In addition, Kanmegne et al. [35] evaluated the potential of optical, SAR, and LiDAR data for AGB estimations in agroforestry systems in three countries in Africa, using field measurements and GEDI data as a reference source, while optical and SAR data were used as predictors of AGB to train different ML algorithms. In the same way, Wang et al. [36] used the same RS data as Kanmegne et al. [35] and estimated AGB in Yunnan Province in China; the authors also explored the differences when using the data individually and in their different combinations.

Similarly to CH estimation, recent works have explored the use of DL techniques for AGB estimation, such as the study released by Dong et al. [37]. In this study, the authors estimated AGB on bamboo forests in Taihuyuan, China, using data from WorldView-2 to compare different ML and DL algorithms, including a CNN and performed various tests on the input's window size. Moreover, Pascarella et al. [38] used a UNet architecture to produce a pixel-wise regression map. This study used Sentinel-2 imagery and AGB data from the European Space Agency's Climate Change Initiative Biomass project as ground truth to estimate AGB and carbon sequestration for the Astroni nature reserve in Italy. Another work that has used UNet-based architecture is the one presented by Dong et al. [39] and incorporates an additive spatial self-attention mechanism in the skip connections of the UNet. The RS data used in this study was from Sentinel-1, Sentinel-2, ALOS-2/PALSAR-2, and GEDI, and topographic data was also used.

Additionally, some studies have been carried out across the Brazilian Amazon. For instance, the study presented in De Almeida et al. [40], the authors compared the ability of different data sources like airborne LiDAR and hyperspectral imagery (HSI) and their combination, along with six regression methods, to improve AGB predictions. Also, field inventory data was used to obtain a reference of AGB, showing that models combining metrics from both datasets resulted in more accurate AGB estimations, regardless of the regression method. Other studies have focused on estimating AGB in forest degraded areas, involving ALS and Landsat 8 Operational Land Imager data [41]. Furthermore, Schuh et al. [42] investigate the potential and accuracy of classical ML techniques, such as RF and support vector machine, along with a generalized linear model to learn the relationships between LiDAR-derived metrics and forest inventory data.

Moreover, Bispo et al. [43] produced an AGB map for a region in the Brazilian Cerrado biome using a combination of vegetation inventory plots, airborne LiDAR data, and multispectral and radar satellite images from Landsat 8 and ALOS-2/PALSAR-2, with a combination of RF models and jackknife analyses. Similarly, the study developed by Ometto et al. [44] integrates multisource RS data and precipitation information, calibrating ALS data by forest inventory plots and using RF with inputs from PALSAR, vegetation indices obtained from MODIS and precipitation information to produce an AGB map for the Amazon biome. Another study, presented by De Jesus et al. [45], estimated AGB in a region in the Caatinga biome, using Sentinel-1 from different periods of the year and measurements of AGB by using allometric equations from nineteen sample plots. The authors extracted attributes from Sentinel-1 and used single and multiple linear regression.

Finally, the evaluation of spatial coverage data in the Brazilian Amazon by Tejada et al. [46] showed that even when Brazil has forest inventory data from different sources, they are not standardized and are not always available to the scientific community. Additionally, the synthesized biomass data maps show that the area covered by these data represents a small fraction of the Brazilian Amazon.

These studies collectively underscore the potential of integrating various remote sensing technologies and modeling approaches to enhance the accuracy of AGB estimation in the Amazon, which is vital for carbon stock assessment and climate change mitigation efforts. Also, they highlight that DL techniques remain underexplored in studies for AGB estimation within the Brazilian Amazon. In this sense, this study evaluates different DL techniques for CH estimation but also evaluates ML techniques for AGB estimation, leveraging the advantages from multisource RS imagery encompassing Sentinel-1, ALOS-2/PALSAR-2, Sentinel-2, and GEDI data.

3
THEORETICAL BACKGROUND

This chapter provides the fundamental theory for a proper understanding of the dataset, methods, and models used in the development of this study. We will start by defining biomass. Next, the basic concept of remote sensing, types of sensors, and satellite information. Then, we will explain the basics of CNN [22], ResNet [47], and UNet [26]. Finally, a description of the DL models used in the development of this study is presented, as well as the components of these models and the loss functions used.

3.1
Biomass

Biomass is the contraction of biological mass, this means biomass is the amount of living material in a given area or volume of Earth's surface [12], and can be transformed to energy through direct or indirect processes [48], e.g., it can be burned to produce heat (direct), or processed into biofuel (indirect). Forest biomass is composed of living mass and can be divided into AGB, where stems, stumps, branches, bark, seeds, and foliage make part of it, and BGB, which consists of all live roots with more than a 2 mm diameter. In addition, forest biomass is composed of dead mass of fine and coarse leaf litter associated with the soil. Brown et al. [13] define forest biomass as *"the total amount of aboveground living organic matter in trees expressed as oven-dry tons per unit area"* and is usually measured in units of mass per hectares, like $Mg \cdot ha^{-1}$ or $T \cdot ha^{-1}$.

AGB estimation can be relatively simple or very difficult; it mainly depends on the type of forest and the method used to make the estimations. Initially, AGB was estimated using field measurements for individual trees or sampling plots in specific areas. Then, with the advances in science and technology, AGB estimation used remote sensing (RS) data [14] covering more extensive areas, from regional to global scales. In this context, the methods used for AGB estimation are:

1. Traditional field measurements: This method involves physically measuring trees like height, diameter at breast height, or trunk weight and constructing relationships between these measurements and for-

est biomass by employing regression equations (non-destructive method, e.g., in [49, 50]) or by converting volume to biomass (destructive method, e.g., in [51, 52]). Although traditional field measurements are the most accurate method, they are impractical for large-scale assessments due to forests' vast and inaccessible nature. Also, traditional field measurements are often outdated, as well as costly and labor-intensive. Moreover, the use of regression equations is highly dependent on the type of forest, as shown by many studies [53, 54, 50, 51], making them site-specific and cannot be easily generalized to all areas.

2. RS-based methods: These methods leverage RS imagery and utilize ML and DL techniques to effectively measure and monitor trees and forest biomass. RS-based methods can capture spectral and textural information, providing broad coverage (e.g., in [55, 56]) and identifying complex patterns and relationships within the data that might be missed by traditional methods. Among the data used in this method are optical, radar, and light detection and ranging (LiDAR) data, each with its own advantages and disadvantages. Therefore, integrating different sensor data and leveraging the advantages of each sensor can produce accurate estimates by improving the information extraction process [14].

3.2
Remote Sensing

RS provides information about the Earth's surface and atmosphere, both on a local and global scale. This information is based on the radiation emitted, reflected, and/or transmitted by a sensor or instrument aboard satellites or aircraft. These satellites are located at a distance from the surface, enabling the acquisition of information that would otherwise be challenging or impossible to obtain [57].

The information acquired depends on the resolution of the sensors used. Resolution refers to the ability to distinguish finer details and the accuracy of the information that can be obtained from an image or data set. Four types of resolution are associated with RS, each influencing the interpretability of RS imagery [58, 59].

1. *Spatial resolution:* The determination of spatial resolution relies on the sensor's pixel size, denoting the ground area covered by each pixel in the image. Consequently, a smaller pixel size corresponds to a higher spatial resolution.

2. *Spectral resolution:* In RS, spectral resolution denotes the capacity of a sensor to discern finer wavelengths or color of light, which is determined by the number of bands and the wavelength interval that a detector can capture.

3. *Radiometric resolution:* It refers to the sensor's ability to detect slight differences in th e intensity of radiation or light, essentially determining the amount of information recorded in each pixel according to a number of bits.

4. *Temporal resolution:* Satellites must complete an orbit to revisit the same observation area. In this sense, temporal resolution is the interval of time between successive images over a given area.

On the other hand, sensors and Earth's surface interact in active and passive modes (see Fig. 3.1). In passive mode, sensors harness the natural energy emitted by the sun, which illuminates the Earth's surface and detects the reflection of this energy. Most passive sensors operate in the electromagnetic spectrum's visible, infrared, thermal, and microwave portions. These types of sensors capture diverse physical attributes, including land and sea surface temperature, vegetation properties, cloud and aerosol properties, and other physical characteristics. However, a notable limitation of this type of sensor is its inability to penetrate the dense cloud cover and, therefore, has limitations in observing areas such as the tropics. In contrast, in active mode, sensors emit their own energy source to illuminate objects on the Earth's surface and then measure the energy reflected back. Most active sensors operate in the microwave band of the electromagnetic spectrum, allowing the sensors to penetrate the atmosphere under most conditions and help measure forest structure, precipitation, and winds [60, 61].

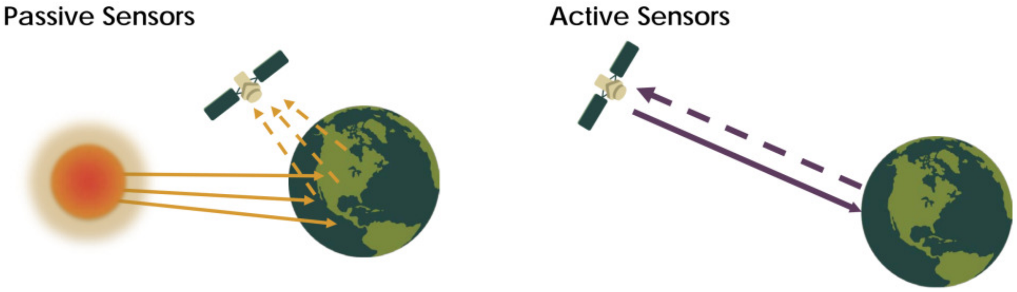


Figure 3.1: Diagram of passive and active sensors, taken from [61].

3.2.1
Sentinel-1

The Sentinel-1 belongs to the Copernicus mission, a joint initiative of the European Commission (EC) and the European Space Agency (ESA). There are two satellites, Sentinel-1A was launched on 3 April 2014 and Sentinel-1B was launched on 25 April 2016. The Sentinel-1 collects C-band SAR imagery and operates in four exclusive acquisition modes: Stripmap (SM) at 5 m by 5 m of spatial resolution (SR), Interferometric Wide swath (IW) at 5 m by 20 m of SR, Extra Wide swath (EW) at 20 m by 40 m SR, and Wave (WV) at 5 m by 5 m of SR. It also provides products distributed at three processing levels: Level 0, Level 1, and Level 2. In Level 1, it can generate either Single Look Complex (SLC) or Ground Range Detected (GRD) in a single (HH or VV), dual (HH/HV OR VV/VH), or partial dual polarization (only HH, HV, VV, or VH).

One application of Sentinel-1 is in land monitoring. The most suitable product in forestry is in IW with a cross-polarisation VH, which is more sensitive to changes in vegetation density and structure [62]. Moreover, combining multiple dates of VH and VV polarisations had a higher correlation with biomass than single date imagery [63].

3.2.2
Sentinel-2

The Sentinel-2 also belongs to the Copernicus mission from the ESA and was launched on June 23, 2015. The Sentinel-2 is based on a constellation of two identical satellites in the same orbit. Each satellite carries a multispectral imager providing optical observations in 13 spectral channels, including the visible spectrum, near-infrared, and short-wave infrared, as well as several red-edge bands; four spectral bands as 10 m (B2, B3, B4, B8), six bands at 20 m (B5, B6, B7, B8A, B11, B12) and three bands at 60 m (B1, B9, B10) of SR. In addition, the revisit frequency of each satellite is ten days, and the combined constellation revisit time is five days [64].

The Sentinel-2 satellite data provide different product levels. Level-1C applies radiometric and geometric corrections, including orthorectification and spatial registration. On the other hand, this satellite is the only one with three red-edge bands (B5, B6, and B7). These bands effectively monitor vegetation health information [65], while B1, B9, and B10 bands are related to atmosphere and water elements [66].

3.2.3
ALOS-2/PALSAR-2

The PALSAR-2 aboard the ALOS-2 was produced by the Japan Aerospace Exploration Agency (JAXA) and was launched on May 24, 2014, by the H-IIA launch vehicle. The PALSAR-2 is an L-band SAR radar that operates in active mode and has three data acquisition modes: Single mode (HH, HV, VH, or VV polarizations), Dual mode (HH+HV or VH+VV polarizations) and Full mode (HH+HV+VH+VV polarizations). Also, it has a revisit cycle of 14 days [67] and has a minimum spatial resolution of 3 m.

The PALSAR-2, as an active sensor, can provide satellite imagery both day and night, and because it operates in L-band, it is less affected by rain and clouds and has better penetration compared to the C-band. In addition, it can penetrate through vegetation, partially reaching the ground. In this context, it can obtain information about vegetation and the ground surface [67]. Furthermore, PALSAR-2 is sensitive to biomass, especially from low to moderate biomass up to 100-150 $Mg{\cdot}ha^{-1}$ on its single or dual acquisition mode [68]. It has been found that HH polarization is suitable for sparse areas, whereas HV polarization is suitable for dense areas [69] and has a higher sensitivity to AGB levels [70].

3.2.4
GEDI

The GEDI is a NASA Earth Ventures Instrument mission launched on December 05, 2018. GEDI is a full-waveform LiDAR instrument that performs detailed measurements of the 3D structure of the Earth's surface. The LiDAR waveforms quantify the vertical distribution of vegetation by recording the amount of laser energy reflected by plants such as stems, branches, and leaves at different heights above the ground [71]. GEDI provides sparse data and offers products at four levels: L1, L2, L3 and L4. L1 product is the raw wavefroms, while L2 contains information derived from the geolocated GEDI return waveforms, and the GEDI02-A product includes ground elevation, tree canopy height, and relative return energy metrics. L3 products are the gridded L2 metrics. Moreover, L4 provides AGB density estimation from L2 data using statistical models in areas where AGB density has already been estimated from forest inventories [72], both at 25 m diameter of resolution.

3.3
Machine Learning

ML is the subset of artificial intelligence that aims to build systems that learn and improve as they use data. In regression tasks, this technique predicts a continuous outcome (dependent variable) based on the values of the predictors (independent variables). There are four main types of machine learning [73]:

- Supervised ML: it uses labeled data to train the algorithm.
- Unsupervised ML: this type of ML employs a more independent approach by not using labeled data, which makes the algorithm learn to identify complex processes and patterns.
- Semi-supervised ML: during training it uses a smaller amount of labeled data, addressing the problem of not having enough labeled data.
- Reinforcement ML: it is similar to unsupervised ML but it works toward a set goal rather than looking at the data to discover patterns that may exist.

Among the ML algorithms used for regression are the following:

- Multi-Layer Perceptron (MLP): is an artificial neural network with multiple layers of neurons and activation functions.
- Linear Regression (LR): predicts numerical values by determining a linear relationship between the independent and dependent variables.
- Random Forest (RF) regression: combines multiple decision trees to create a single model that predicts the target value, which is obtained by averaging the predictions of all the trees.
- Gradient boosting: the model is built sequentially, combining several weak learners into strong learners that attempt to minimize the loss function.

3.4
Knowledge Distillation

Knowledge Distillation (KD) is an ML technique introduced by Hinton et al. [74] to transfer knowledge from a large pre-trained model called the teacher model to a smaller student model. The main objective is to train the student model to match the predictions made by the teacher model during training and generalize to new data after training.

KD typically uses two loss functions. The first operates on the final predictions from the student model and the ground truth data. The second loss is the distillation loss, measured between the student model's preliminary predictions and those of the teacher model. In this context, distillation loss measures the difference between the probability distribution of the teacher and student models' preliminary predictions.

KD methods try to mimic the output from the teacher model and the reasoning steps the teacher model takes to arrive at that output. Other KD methods focus on weights and activations in the network's hidden layers or the relationships between different network parts [75]. I) Response-based knowledge: The student model is trained to output predictions that match the teacher model's predictions, thus transferring information from the teacher model to the student model. II) Feature-based knowledge: This method focuses on the information transmitted at the intermediate layers to train the student model to learn the same features as the teacher network. III) Relation-based knowledge: In this approach, the student model is trained to emulate the teacher model's "thought process" because this method focuses on the relationships between different layers or feature maps representing the activations at different layers or locations.

Knowledge distillation methods can also be categorized by their impact on the teacher network. I) Offline distillation: The teacher network is already pre-trained and its model weights are frozen to prevent further changes. II) Online distillation: These schemes aim to train both the teacher and student networks simultaneously. III) Self-distillation: In self-distillation, one network acts as both teacher and student, refining its own predictions.

3.5
Convolutional Neural Networks

Convolutional Neural networks, also known as CNN or ConvNet, were introduced by LeCun et al. [22]. This network processes data arranged in a grid-like fashion, employing a mathematical operation called convolution, which allows the model to learn directly from images. ConvNets consist of a series of convolutional layers and pooling layers to extract the relevant features from the input image (see Fig. 3.2). Then, the fully connected layers use these features to make predictions [76, 77].

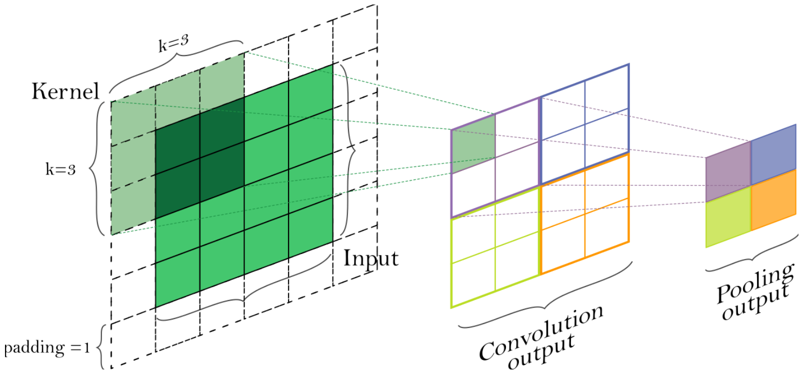


Figure 3.2: Schematic representation of convolution and pooling layers. The figure shows the convolution process with a kernel of 3×3 and a padding of 1, while the pooling process is with 2×2 filter and stride of 2.

- Convolutional layer: The convolutional layer performs convolution operation, extracting features from the input image; in this sense, the layer's output is a set of feature maps. In this layer, a dot product is computed between the input image and a small matrix of size M×M that slides over the input image; the small matrix is known as a filter. This layer ensures the spatial relationship between the pixels is preserved.
- Pooling layer: The pooling layer is used to reduce redundancy and spatial dimensions of the feature maps. This layer operates on each feature map independently and downsamples it by taking the maximum (max pooling) or average value (average pooling) of non-overlapping regions.
- Fully connected layer: The fully connected layer connects every previous neuron to every neuron in the current layer. For regression and classification tasks, it is typically used at the network's end to produce the final output. This layer helps the network to map the representation between the input and the output.

3.5.1
Residual Network

The Residual Network or ResNet is a DL model proposed by He et al. [47], which introduced the concept of residual learning to simplify the training of deep networks and address the vanishing gradient problem. The results demonstrated that ResNet with 50, 101, and 152 layers had lower error rates in image classification tasks compared to a 34-layer plain network. The key components of ResNet are shown in Fig 3.3:

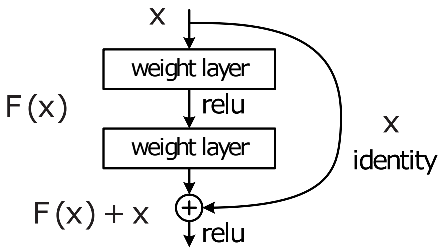


Figure 3.3: Schematic representation of the residual learning, This representation was taken from [47].

- Residual block: This block allows the input to skip one or more convolutional layers by creating a shortcut connection. The output of the skipped layer is then added to the original input.
- Identity Mapping: The identity mapping is performed by the residual block when the input is directly added to the output of the convolutions.
- Bottleneck Design: The bottleneck is a residual block with three convolutional layers: 1×1, 3×3, and 1×1. The 1×1 convolutional layers reduce and then restore the dimensionality, respectively.

3.5.2
UNet

The UNet architecture, introduced by Ronneberger et al. [26], is a U-shaped network designed for biomedical image segmentation. The results obtained by this network demonstrate that it can be trained end-to-end from very few images. The UNet architecture consists of (see Fig. 3.4):

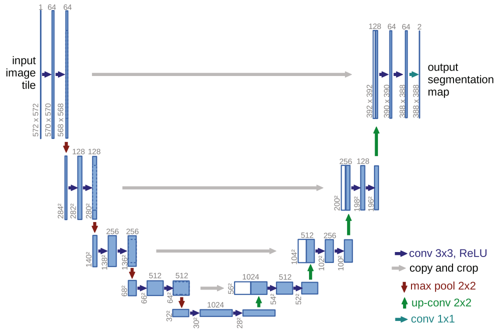


Figure 3.4: UNet architecture. Taken from [26]

- Encoder path: Also known as the contracting path, this path captures high-resolution, low-level features and follows the typical ConvNet architecture. Each downsampling step reduces the spatial resolution while increasing the number of feature channels.

- Decoder path: Also known as the expansive path, this path enables precise localization and allows the network to reconstruct a dense segmentation map. Each upsampling step increases the spatial resolution of the feature maps while reducing the number of channels.
- skip connection: Skip connections are used in the UNet to connect corresponding layers of the encoding and decoding paths. The feature map from the contracting path is concatenated to the corresponding upsampled feature map in the expansive path. This concatenation helps the network to incorporate multi-scale information for accurate segmentation.

3.6
Vision Transformer

ViT was introduced by Dosovitskiy et al. [78] for image classification tasks and was inspired by the transformer architecture proposed for natural language processing (NLP) by Vaswani et al. [79]. While transformers in NLP encode words in a sentence into a sequence of vectors to learn the relationship between the words using the self-attention mechanism, ViT represents an image as a series of smaller patches. By leveraging the self-attention mechanism, ViT captures both local and global relationships within images to predict class labels for the input image.

As mentioned before, ViT represents an image as a series of smaller patches. This is done by dividing the image into smaller patches. Then, these patches are linearly transformed into a vector using a learnable linear projection called patch embeddings. Next, positional embeddings are added to the patch embeddings to retain positional information. This image representation is the input to the Transformer encoder.

The Transformer encoder, shown in Fig. 3.5, is composed of I) Multi-head Self-Attention (MSA) to focus on important patches and capture both local and global relationships. II) MLP to introduce non-linearity and allow the model to learn complex patch relationships. III) Layer normalization and residual connections are applied before and after MSA and MLP blocks, respectively.

3.7
Self-Aware Attention module

The Self-Aware Attention module (SAA) acts as a bridge between the encoder and decoder. This placement improves the UNet model's ability to capture long-range contextual information, thus enhancing the encoder's

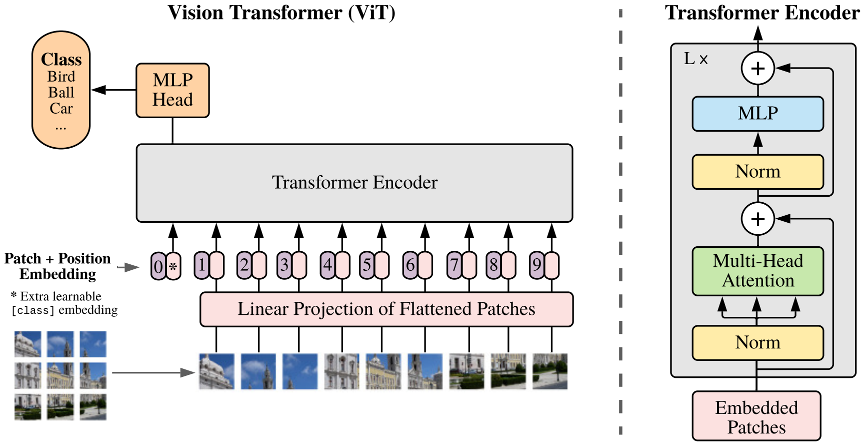


Figure 3.5: On the left is presented the ViT architecture, and on the right is the encoder. The image was taken from [78].

semantic feature representation. The SAA module comprises two separate self-attention mechanism as presented in Fig. 3.6:

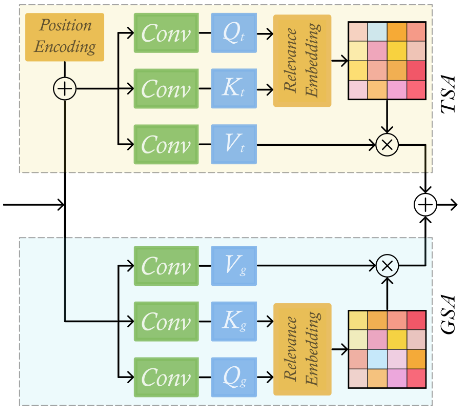


Figure 3.6: Self-Aware Attention module, adapted from [80].

1. *Transformer Self-Attention (TSA)*: TSA leverages Transformer's multi-head attention feature to capture semantic details from global representation subspaces. It employs learned positional encoding that can be shared across all attention layers for a given query/key-value sequence, integrating both absolute and relative positional information. The multi-head attention processes each attention head independently and then combines it through another embedding.

2. *Global Spatial Attention (GSA)*: This mechanism improves learned features with global context and encodes broader contextual positional information into local features. This improves intra-class compactness and optimizes feature representations.

3.8
Alpha Multimodal Dual Output UNet

Alpha Multimodal Dual Output UNet (α-2MDU) model is inspired by the study of Chen et al. [80]. Its architecture is based on UNet architecture and was developed for CH estimation from RS imagery. The UNet encoder path extracts features from the input image while reducing the resolution, and the decoder path reconstructs the image output from the extracted features (see Fig. 3.7).

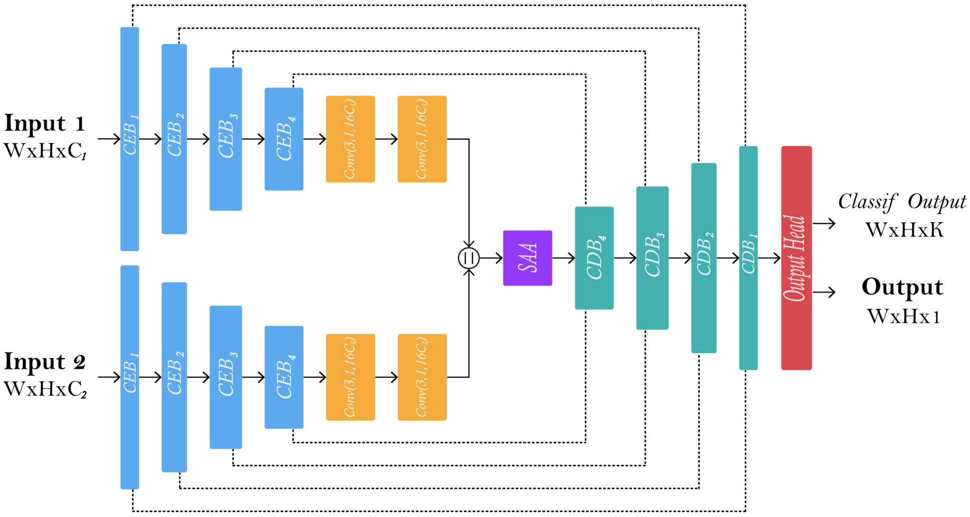


Figure 3.7: α-2MDU architecture, adapted from [81].

α-2MDU architecture [81] consists of two encoder paths for each RS data source, optical and SAR data, allowing the network to learn to extract relevant features and representations independently for each data source. Moreover, both encoders comprise four Convolutional-based Encoder Blocks (CEB) that are based on convolutional layers and, to preserve the information in the feature map, also use a convolutional layer with a stride of 2×2 instead of a pooling layer. Then, both encoders' outputs are concatenated channels-wise and fed into a self-aware attention (SAA) module [80] to learn non-local interactions among both encoders' features. Furthermore, the decoder path also comprises four Convolutional-based Decoder Blocks (CDB).

Another particularity of this model is that its output head performs two tasks: classification and regression tasks. The first task classifies the CH into k overlapping bins, and the second task outputs continuous height estimates (regression task).

3.9
Hybrid ViT-ConvNet

Hybrid ViT-ConvNet (Hy-TeC) architecture [81], shown in Fig 3.8, is based on a hybrid ViT-ConvNet model developed for the continuous predictions of CH. Addressing the dense prediction problem, the model employs a ViT architecture as the encoder and a ConvNet as the decoder. Then, the decoder is connected to the output head, which is the same as the α-2MDU model, performing the classification and regression tasks. Furthermore, this approach also integrates two pre-trained UNet-based models and optimizes the model in a teacher/student training paradigm. It is worth saying that the inputs of this model are according to the original resolution of 10m and 20m.

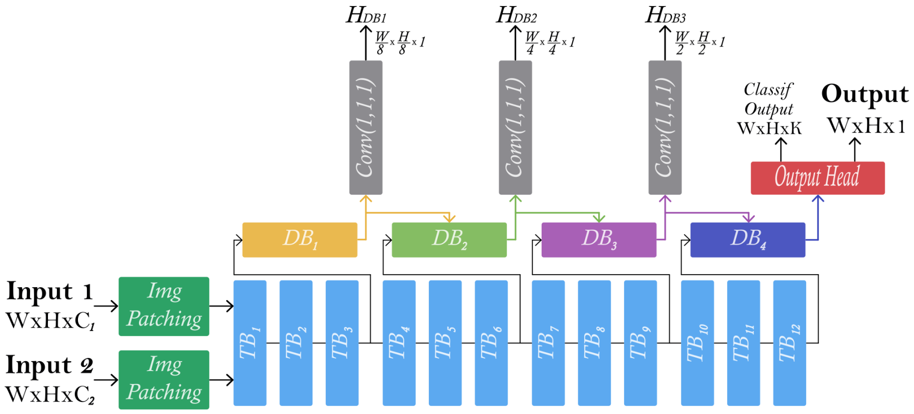


Figure 3.8: Hy-TeC architecture, adapted from [81].

Finally, acting as teacher models, two single-encoder UNets are pre-trained on each optical and SAR data source. This acts as a filtering criterion for the data that is transferred to the transformer model. Then, since the soft labels can not be computed, the filtered output of the teacher model is used to influence the transformer-based model.

3.10
Loss Function

The loss function is used to evaluate an ML algorithm's performance in predicting the expected outcome in a specific task by quantifying the distance between the target value and the predicted output of the ML algorithm. The objective during training is to minimize the value of the loss function.

Over the years, many loss functions have been developed to improve the performance of the ML and DL algorithms, which depends on the task for which the model is trained [82, 83]. The most common implementation in classification tasks is cross-entropy loss [84], which describes the distance between the actual and expected output. The smaller the value of cross entropy is, the closer the two probability distributions are. For regression tasks, some loss functions used are:

- Weighted Mean Square Error (MSE): this loss measures the average squared difference between the estimated values and truth values and gives more importance to a certain set of points compared to others.
- Huber loss: this loss function [85] combines the advantages of square loss and absolute loss using a parameter θ as a boundary judge between both losses. This loss function is robust to outliers and is less sensitive to examples with large errors.

  $$\ell_H(Y, Y') = \frac{1}{N}\sum_{i=1}^{N} \begin{cases} \frac{1}{2}(Y_i - Y_i')^2 \ , \ if \ |Y_i - Y_i'| \ < \ \theta \\ \theta \ * \ \left(|Y_i - Y_i'| - \frac{1}{2} \cdot \theta\right), \ otherwise \end{cases} \quad , \tag{3-1}$$

  where Y and Y' are the target and truth values respectively, and N is the number of samples
- Adaptive and robust regression loss: Unlike Huber loss, this adaptive loss [86] is a single function that can tune its parameters adaptively, α is a shape parameter that controls the robustness of the loss, and **c** is a scale parameter, and Y and Y' are the target and truth values respectively

  $$\ell_{RA}(Y,\ Y') = \ \frac{|\alpha - 2|}{\alpha}\left(\left(\frac{\left(\frac{Y - Y'}{c}\right)^2}{|\alpha - 2|} + 1\right)^{\frac{\alpha}{2}} - 1\right) \tag{3-2}$$

4
METHODS

This chapter describes in more detail the DL methods evaluated in this work for CH and AGB estimation: ResNet based, α-2MDU, and HyTec. The last two networks were proposed by Fayad et al. [81] and were originally proposed to estimate CH. Final implementation details, as well as the particular setups, are clarified in the following sections.

4.1
Residual Network based model

The ResNet-based model used in this work will be the baseline for CH estimation since it was implemented to estimate CH in the Brazilian legal Amazon during the competition "Universities for Goal 13" from Siemens Gamesa and the United Nations Sustainable Development Solutions Network, on its second edition in 2023. Additionally, this model also was evaluated for AGB estimation.

The architecture of the ResNet-based model is shown in Fig 4.1 and comprises a series of convolutional layers, ReLU activation functions, MaxPooling blocks, batch normalization, residual blocks, and an output block. Each MaxPooling block (see Fig. 4.2 reduces the number of channels by half and consists of a sequence of a convolutional layer, batch normalization, ReLU activation function and a MaxPooling layer. Residual blocks (see Fig. 4.2), on the other hand, consist of a convolutional layer, batch normalization, and ReLU activation function, followed by another convolutional layer and ending with a final ReLU activation function. The output head (see Fig. 4.2 generates a value for the central pixel of the image, achieved through a sequence involving adaptive average pooling, a flattened layer, a dropout layer, and a linear layer. Unlike other pooling methods, adaptive average pooling does not require a kernel size parameter; instead, the output size must be specified. Finally, the regression result is assigned to the patch central pixel with a sliding window approach.

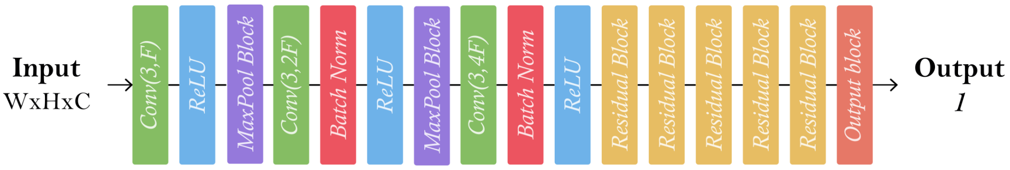


Figure 4.1: ResNet-based model architecture, where C represents the number of bands for the input to the model.

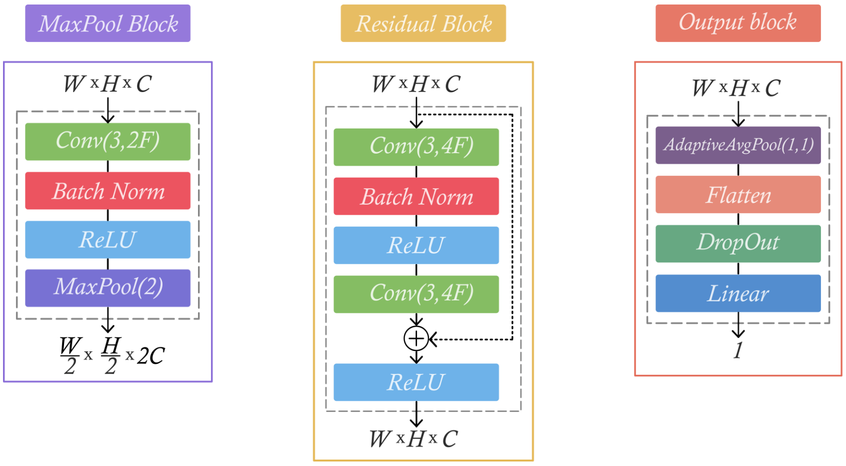


Figure 4.2: ResNet-based model blocks.

4.2
Alpha Multimodal Dual Output UNet

As mentioned in the previous chapter, each encoder path comprises four CEB blocks, and the decoder path comprises four CDB blocks. However, to the development of this study, we also adapted the architecture of the model to estimate CH and AGB (see Fig. 3.7) by increasing one CEB block on each encoder path; therefore, one CDB block was increased in the decoder block (see Fig. A.1). Additionally, the classes for the classification task are according to the type of estimation, CH or AGB. The regression results of this model architecture are in all the pixels of the patch. See the model's blocks in the appendix section A

4.3
Hybrid ViT-ConvNet

From the previous chapter, we can see that this model performs four outputs at different resolutions; three outputs are used for the knowledge distillation along with the mean of the outputs of the pre-trained models, and the last output is used for the output head. See the model's blocks in the appendix section A

The Hy-Tec architecture used in this thesis performs three outputs instead of four, as shown in Fig. 4.4. Two of the outputs are used for the

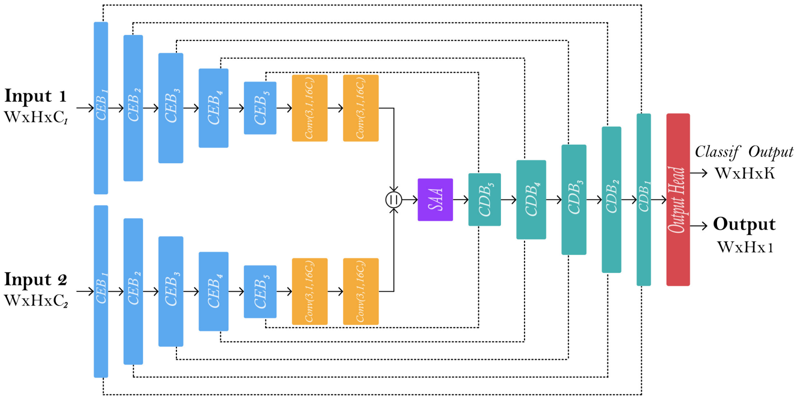


Figure 4.3: α-2MDU architecture with five CEB blocks and five CDB blocks, modified from [81]. C_1 and C_2 represent the number of bands for input 1 and input 2, respectively. k represents the number of classes for TH estimation and AGB estimation.

knowledge distillation, and the other one goes to the output head, where the classes for the classification task are according to the type of estimation, CH or AGB. Similarly to the α-2MDU model, the regression results are in all the patch pixels.

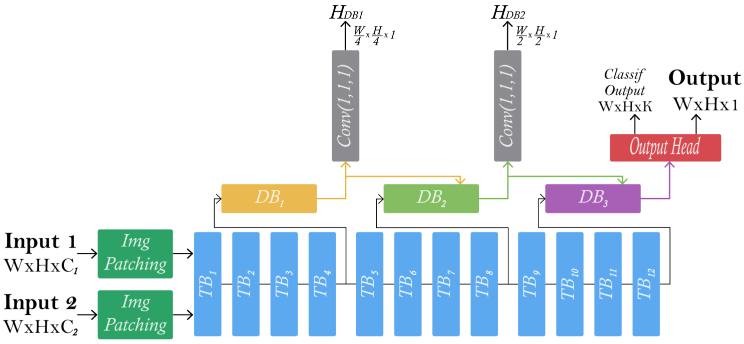


Figure 4.4: Hy-Tec architecture with three outputs, modified from [81]. C_1 and C_2 represent the number of bands used for input 1 and input 2, respectively and k represents the number of classes for CH estimation and AGB estimation.

5
EXPERIMENTAL ANALYSIS

This chapter presents the study area and the experimental implementation to evaluate the DL methods introduced in the previous chapter. Additionally, classical ML methods, such as RF, LR, MLP, and extreme gradient boosting (XGBoost), were evaluated for AGB estimation. First, the study area is presented. Next, the datasets used in the experiments are presented. Then, the experimental protocol followed for CH and AGB estimation is described, as well as the parameter setup. Finally, the results obtained in the experiments are reported and discussed.

5.1
Brazilian Legal Amazon

The area selected to evaluate the methods comprises the Brazilian legal Amazon (Fig. 5.1). Brazil has the second-largest forest area with the greatest biodiversity in the world [9, 11]. Brazil's Legal Amazon comprises eight states belonging to the Amazon basin: Acre, Amapá, Amazonas, Mato Grosso, Pará, Roraima, Rondônia, Tocantins, and the region west of the 44th meridian in the state of Maranhão. Its area corresponds to 59% of the Brazilian territory; this area is important due to its natural resources, ecosystems, hydrological cycle, and carbon storage.

Brazil has mostly a tropical climate (A zone), representing 81,4% of its territory, especially in the Brazilian Legal Amazon [87, 88]. In the Amazon basin, the rainy season takes place from October to May; however, the months with the most precipitation are from January to March, while the drier period occurs from June to September, especially in July and August [89].

5.2
Dataset

The datasets used are publicly available to the scientific community. The satellite datasets are available on Google Earth Engine (GEE) [1], and the forest mask dataset was developed by the Instituto Nacional de Pesquisas Espaciais (INPE) that is available in the Terrabrasilis platform [90].

[1] https://earthengine.google.com/, accessed on September 22nd, 2023

Figure 5.1: The image shows the boundaries of Brazil and its states; the boundaries of the Legal Amazon are shown in orange, and the forest mask for the Legal Amazon is in green.

The dataset comprises Sentinel-1, Sentinel-2, ALOS-2/PALSAR-2, and GEDI satellites for the year 2020, as well as the forest mask from INPE. This dataset was divided into dry and wet seasons for Sentinel-1 and Sentinel-2 ([63]), where the dry season was considered between June 1st to September 30th and the wet season from January 1st to March 31st and December. Also, a cloud coverage filter with de 60% and re-scale all data to a resolution of 30 meters were applied. Finally, Table 5.1 summarizes the information on the RS data used and the bands selected for each of them. The selection of the bands was following the explanation in the Remote Sensing section chapter 3. In addition, the data divided into the dry and wet seasons were concatenated channels-wise, getting the double of channels for those sensors. In this sense, the final image comprises all the rs data already explained, obtaining an image with 29 bands.

Data from Sentinel-1, Sentinel-2, and ALOS-2/PALSAR-2 are used as input to the models, and GEDI data is used as a ground-truth. Depending on whether the CH or AGB was estimated, we use the 28th channel or the 29th channel, respectively.

Table 5.1: Information about de RS data and the bands selected for each satellite.

Satellite	Data Information	Bands
Sentinel-1	Level-1 GRD, IW acquisition mode in the dry and wet season, and dual polarization bands.	VV, VH
Sentinel-2	Level-1C in the dry and wet season.	B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12
ALOS-2/ PALSAR-2	Single polarization bands with annual data.	HH, HV
GEDI	▷ CH: L2A product with a relative height (RH) of 98. ▷ AGB: L4A product.	

5.3
Methodology

The Fig. 5.2 provides an overview of the process employed during the study, each process is then described in the following sections.

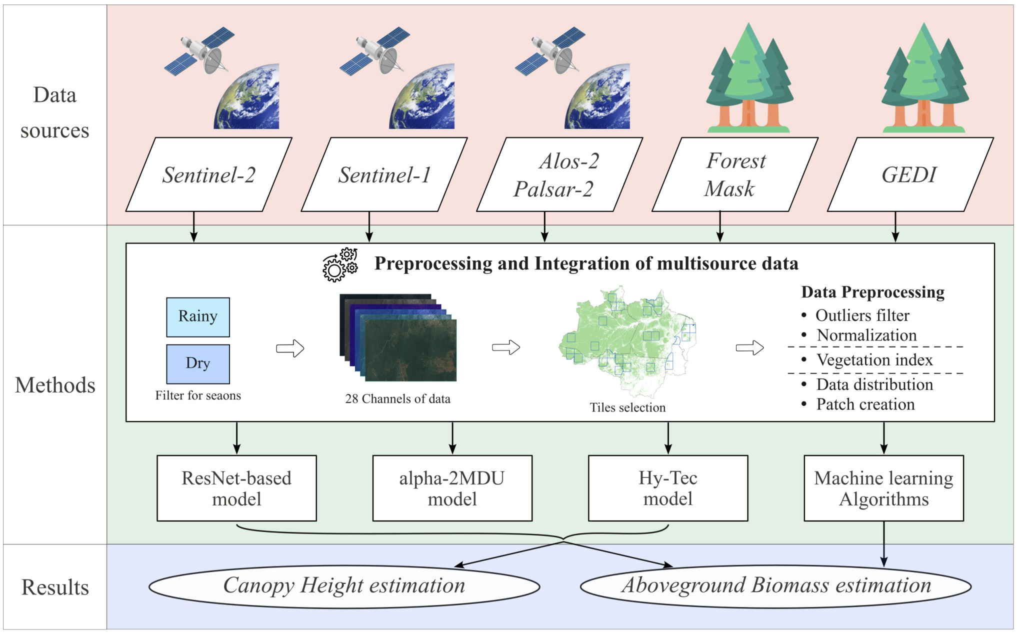


Figure 5.2: Flowchart showing the methodology used to estimate forest parameter in the Brazilian Legal Amazon.

5.3.1
Data Pre-processing

As the Legal Amazon is a very extensive area, and the dataset was downloaded in tiles from GEE (each tile with a size of 6400×6400 approximately), we used the forest mask to compute the forest percentage and select three tiles of each state of the Brazilian Legal Amazon, making a total of twenty-seven tiles (see Fig. 5.3). For states of Amazonas and Pará, tiles with more than 90% of the forest were randomly selected, and for states of Acre, Roraima, Matogrosso, Amapá, Rondonia, Maranhão and Tocantins, that have less than 90% of forest, tiles with the most percentage of forest were chosen (see Table 5.2). Then, these tiles were randomly divided into train, validation, and test with twelve (44.4%), six (22.2%), and nine (33.3%) tiles respectively, ensuring that for validation and test, tiles were of different states. Finally, a filter of outliers and normalization data was applied. For example, outliers can be caused by sensor errors, cloud cover, or shadows. Normalization data was made for each sensor and band individually with a standard scaler for CH estimation and a Min-Max scaler for AGB estimation, due to these type of normalization gave better results.

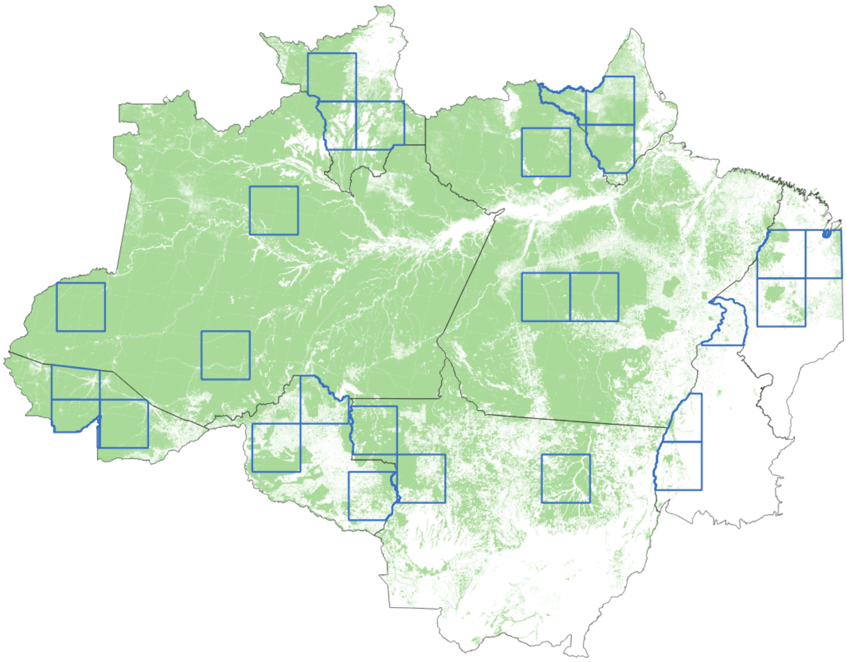


Figure 5.3: Brazilian legal Amazon. Blue boundaries are the tiles chosen for train, validation, and test according to the forest percentage.

Additionally, we observed that the distribution of GEDI data for both CH and AGB showed a wide range of values. For this reason, CH and AGB were limited to a certain range, selecting the values within that range. For CH data, trees with a height less than 88.5 meters were considered [91]. Then, due

Table 5.2: Table with the percentage of forest for each tile used.

State	**Tile 1 (%)**	**Tile 2 (%)**	**Tile 3 (%)**
Acre	61.41	60.19	96.34
Amapá	38.49	64.31	74.71
Amazonas	98.19	99.34	97.76
Maranhão	25.88	13.88	26.65
Mato Grosso	61.21	70.79	81.26
Pará	92.94	91.68	96.86
Roraima	50.20	59.23	40.38
Rondônia	61.85	39.12	86.15
Tocantins	7.24	5.14	8.29

to the small number of values above 45 meters, CH was limited to a range of 1 to 45 meters. Similarly, AGB data was limited to a range of 1 to 552.8 $Mg \cdot ha^{-1}$ [92] and due to the large number of values under 5 $Mg \cdot ha^{-1}$, AGB was limited to a range of 5 to 552.8 $Mg \cdot ha^{-1}$. GEDI data distribution for the different sets for both CH and AGB are presented in Fig. 5.4 and Fig. 5.5, respectively. Finally, the AGB data were intersected with the CH data to avoid data mismatches.

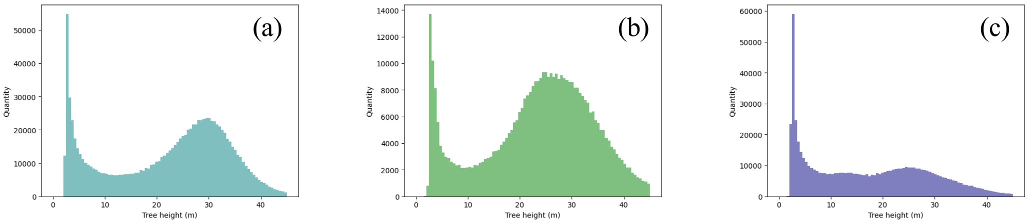


Figure 5.4: CH data distribution for (a) train, (b) validation, and (c) test.

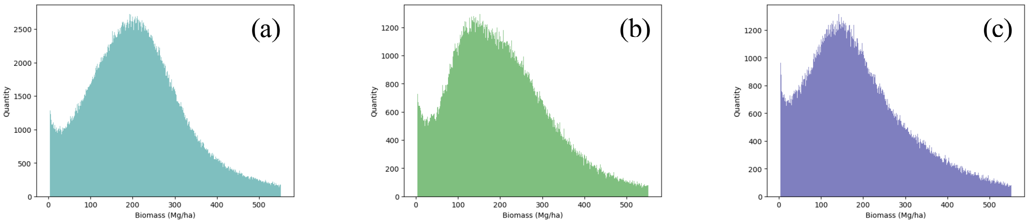


Figure 5.5: AGB data distribution for (a) train, (b) validation, and (c) test.

Moreover, since two of the DL models studied in the development of this work perform a classification task, CH data was clustered into nine classes according to the literature review. This data distribution of the classes is shown in Fig. 5.6(a). Whereas AGB data was clustered into ten classes chosen

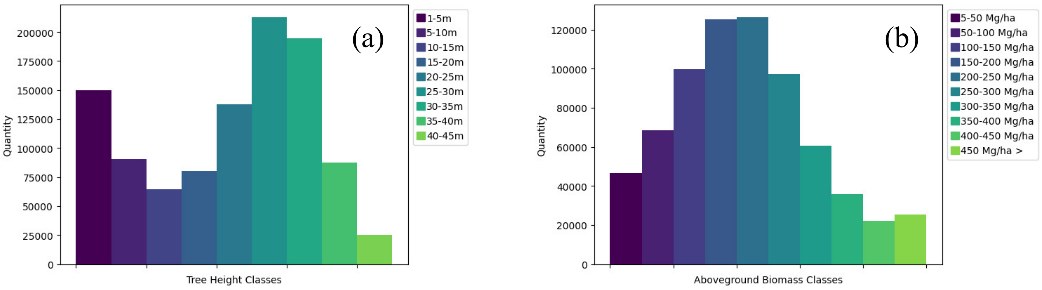


Figure 5.6: Class distribution for (a) CH and (b) AGB estimation from the training set.

experimentally, the data distribution of the classes is shown in Fig. 5.6(b). The weights for this task were computed as the inverse ratio of GEDI acquisitions belonging to a certain class to the total number of GEDI acquisitions over the train tiles.

On the other hand, different features and vegetation indices (VI) based on the data type were extracted for AGB estimation. While Table 5.3 displays the features extracted from the SAR datasets, Table 5.4 showcases the vegetation indices obtained from the optical dataset. Since the ALOS-2/PALSAR-2 data used in this work are annual, only one feature was extracted; in contrast, as Sentinel-1 was divided into dry and wet seasons, four features were extracted. In the same way, for Sentinel-2, ten VIs were extracted, five VIs for each season. These features and VIs were normalized and used for the ML algorithms and also as input to the DL models.

Table 5.3: Features extracted from SAR datasets (Sentinel-1(S1) and ALOS-2/PALSAR-2 (AP2)).

Variable Type	**Variable Name**	**Definition**
Polarization Information	VV, VH	Backscatter coefficients from S1
	HH, HV	Backscatter coefficients from AP2
Constructed features	VV - VH	Polarization Difference
	VH + VV	Polarization Sum
	HH/HV	Polarization Ratio

5.3.2
Patch generation

A common practice when working with remote sensing images is the extraction of patches, which are created following the sliding window overlay procedure. However, since the study focuses on TH estimation and AGB esti-

Table 5.4: Vegetation indexes from Optical data (Sentinel-2).

Variable Index	**Equation**	**Vegetation Index**
NDVI	$\frac{B_{NIR} - B_{red}}{B_{NIR} + B_{red}}$	Normalized Difference Vegetation Index
EVI	$2.5 \cdot \left(\frac{B_{NIR} - B_{red}}{(B_{NIR} + 6 \cdot B_{red} - 7.5 \cdot B_{blue}) + 1} \right)$	Enhanced Vegetation Index
SAVI	$\left(\frac{B_{NIR} - B_{red}}{B_{NIR} + B_{red} + L} \right) \cdot (1 + L), L = 0.25$	Soil Adjusted Vegetation Index
NDWI	$\frac{B_{green} - B_{NIR}}{B_{green} + B_{NIR}}$	Normalized Difference Water Index
CI	$\frac{B_{NIR} + 2 \cdot B_{SWIR}}{B_{blue} + B_{green} + B_{red}}$	Cloud Index

mation by evaluating different DL models. Two approaches were implemented for patch generation for training and validation.

In the first approach, patches of size 15×15 were created to facilitate the training of the ResNet-based model. For this purpose, pixels with GEDI information were identified, as GEDI is our ground-truth data. Then, these pixels were used as the central pixel for patch creation, and patches of size 15×15 were created around them. Next, we calculated the forest percentage from the patch forest mask to avoid oversampling and ensure consistency with tile selection; any patches with less than 90% forest coverage were filtered out. Finally, the image patch was rotated 90° clockwise and flipped over the vertical and horizontal axis. For CH estimation, the patches encompass RS data and the forest mask, with 632812 patches for the training set and 309454 for the validation set. In contrast, AGB estimation patches encompass RS data, the features and VIs extracted, and the forest mask, with 668998 patches for the training set and 331737 for the validation set (see Table 5.5.

For the second approach, we used TorchGeo[2] library, which is focused on geospatial datasets, randomly sampling patches of size 128×128 and 256×256 with 5000 samples for training and 3250 samples for validation. This library's advantage is that it samples the patches focusing on the pixels that have data in the mask, which is, in this case, GEDI data. This approach was used to train α-2MDU and Hytec models.

Finally, to test the models and ensure they use the same data and have the same distribution, we also use the TorchGeo library, with a grid sample of patches of 512×512 and sub-dividing it according to each model's patch size.

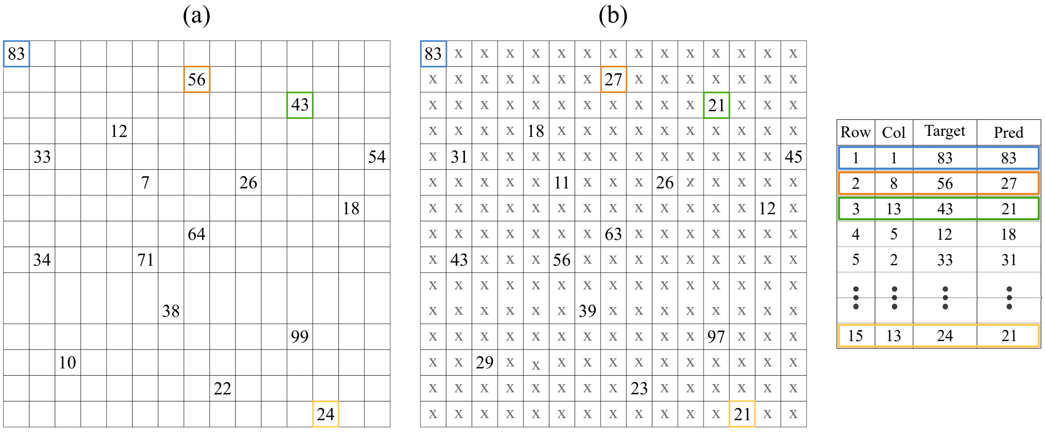
[2] https://TorchGeo.readthedocs.io/en/stable/

Table 5.5: Summary table of the data for each type of estimation and the number of patches generated for train and validation sets for the first approach of patch generation.

Estimation type	**Data**	**Patches per set**
CH estimation	▷ CH from GEDI	
	▷ Sentinel-1	Train: 632812
	▷ Sentinel-2	
	▷ ALOS-2/PALSAR-2	Validation: 309454
	▷ Forest mask	
AGB estimation	▷ AGB from GEDI	
	▷ Sentinel-1	
	▷ Sentinel-2	Train: 668998
	▷ ALOS-2/PALSAR-2	
	▷ Features extracted from SAR data	
	▷ VI from optical data	Validation: 331737
	▷ Forest mask	
	▷ CH from GEDI	

5.3.3 Loss function

Considering that GEDI data is not continuous in all the image or patch, the loss functions are computed using only the pixels with GEDI information, as can be seen in the Fig. 5.7.



(a)

83														
							56							
											43			
				12										
	33													54
					7				26					
													18	
							64							
	34				71									
						38								
											99			
		10												
								22						
												24		

(b)

83	x	x	x	x	x	x	x	x	x	x	x	x	x	x
x	x	x	x	x	x	x	27	x	x	x	x	x	x	x
x	x	x	x	x	x	x	x	x	x	x	21	x	x	x
x	x	x	x	18	x	x	x	x	x	x	x	x	x	x
x	31	x	x	x	x	x	x	x	x	x	x	x	x	45
x	x	x	x	x	11	x	x	x	26	x	x	x	x	x
x	x	x	x	x	x	x	x	x	x	x	x	x	12	x
x	x	x	x	x	x	x	63	x	x	x	x	x	x	x
x	43	x	x	x	56	x	x	x	x	x	x	x	x	x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
x	x	x	x	x	x	39	x	x	x	x	x	x	x	x
x	x	x	x	x	x	x	x	x	x	x	97	x	x	x
x	x	29	x	x	x	x	x	x	x	x	x	x	x	x
x	x	x	x	x	x	x	x	23	x	x	x	x	x	x
x	x	x	x	x	x	x	x	x	x	x	x	21	x	x

Row	Col	Target	Pred
1	1	83	83
2	8	56	27
3	13	43	21
4	5	12	18
5	2	33	31
⋮	⋮	⋮	⋮
15	13	24	21

Figure 5.7: Loss function procedure with GEDI data as a ground truth (a) and the output of the model (b).

5.4
Experimental Setup

As stated earlier in the preceding section, different DL models were evaluated in this study to estimate CH and AGB, as well as ML techniques to estimate AGB.

5.4.1
ResNet-based model

The input to this model for CH estimation is an image patch with twenty-six channels from Sentinel-1, Sentinel-2, and ALOS-2/PALSAR-2 (see Fig. 5.8 for CH configuration example). For ABG estimation, the input to the model is an image patch with forty-three channels from Sentinel-1, Sentinel-2, ALOS-2/PALSAR-2, the features and VIs extracted, the forest mask, and CH data. The output of this model is a pixel value that is compared with the central pixel value from the GEDI patch. In order to optimize this model, AdamW, RMS-Prop, and SDG optimizers were employed independently, with batch size equal to 512 with 80 epochs and early stopping to break after ten epochs without improvement. The loss functions included weighted mean square error and Huber loss separately.

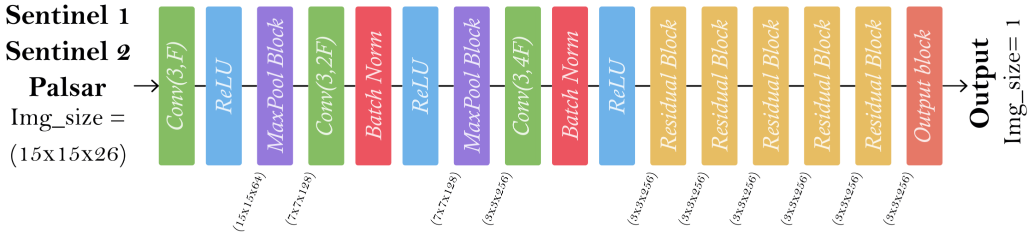


Figure 5.8: Architecture configuration of ResNet-based model for CH estimation.

5.4.2
α-2MDU model

As mentioned in the previous chapter, the architecture of this model comprises four CEB blocks and four CDB blocks, but we also adapted this model to comprise five CEB blocks and five CDB blocks, both with input patches of size 256×256. Since the model comprises two separate encoders and we made estimations CH and AGB, the configuration of the input to the encoders is as follows (see Fig. 5.9 for CH configuration example):

- for CH estimation, the input to the first encoder was the radar data from Sentinel-1 and ALOS-2/PALSAR-2 with six channels ($C_1 = 6$), and the input to the other encoder was the optical data from Sentinel-2 with twenty channels ($C_2 = 20$). Then, the classification sub-module has nine classes ($k = 9$), as mentioned in the data pre-processing section.

- for AGB estimation, the input to the encoders followed the same configuration as for CH estimation but also used the forest mask and CH data. Additionally, the features extracted from Sentinel-1 and ALOS-2/PALSAR-2 were concatenated channel-wise to the first input with 13 channels ($C_1 = 13$), and VIs from Sentinel-2 were concatenated channel-wise to the second input with 32 channels ($C_2 = 32$). Then, the classification sub-module has ten classes ($k = 10$), as mentioned in the data pre-processing section.

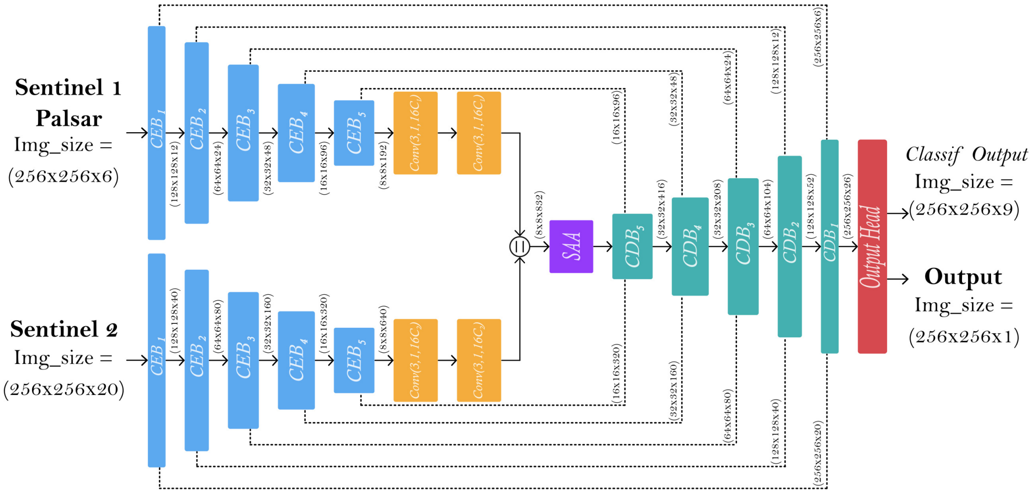


Figure 5.9: Architecture configuration of α-2MDU model with five CEB blocks and five CDB blocks for CH estimation.

Finally, the output of this model is an image of size 256×256 with one channel. To optimize this model, AdamW, RMS-Prop, and SDG optimizers were employed independently, with batch size equal to 12 with 80 epochs and early stopping to break after 10 epochs without improvement. As the output head performs two tasks, regression and classification, the total loss (ℓ_{CR}) is defined as:

$$\ell_{CR}(P, T, Y, Y') = \ell_{CE}(P, T) + \alpha \cdot \ell_R(Y, Y'), \tag{5-1}$$

where P and T are, respectively, the predicted and truth probabilities for each class and for each target and truth values used to calculate the classification loss that is a weighted cross-entropy loss (ℓ_{CE}), Y and Y' are the target and truth values respectively used to calculate the regression loss (ℓ_R) which is an adaptive and robust regression loss, and α is a scaling parameter used to balance both losses, ℓ_{CE} and ℓ_R, with a value in the range 0.02 to 4.5.

5.4.3
Hy-Tec model

This model was also adapted from [81] for CH estimation. The main difference relies on the quantity of outputs, getting three outputs instead of four. The model's input for processing is from Sentinel-2 for the dry season, separating it according to its native resolution. In this sense, the image for input one has a size of 128×128×4, and input two has a size of 128×128×6. Then, both embedded patches were concatenated and passed to the model encoder. As mentioned before, the model outputs three outputs with size 32×32×1, 64×64×1, and 128×128×1 (see Fig. 5.10(a)). Moreover, the model uses two pre-trained models for each data type using a single encoder UNet-based model (see Fig. 5.10(b). These outputs from the pre-trained models are downsampled to be used in the loss function with the corresponding output of the Hy-Tec model.

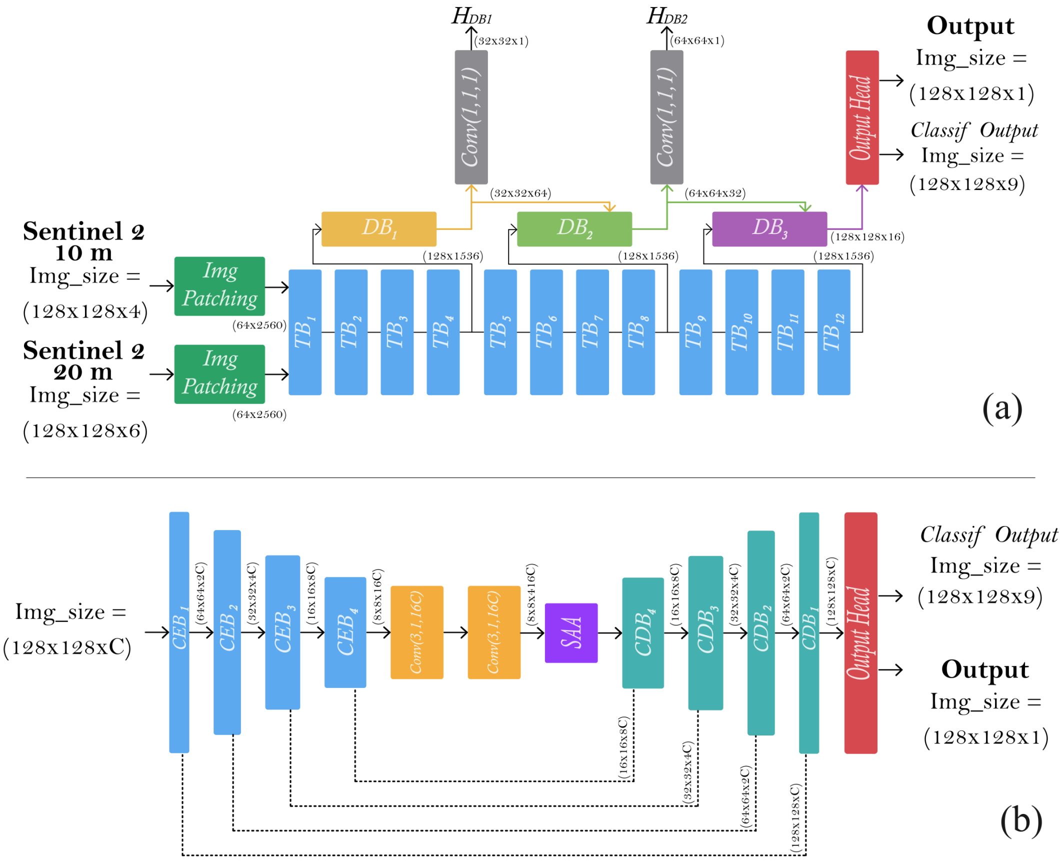


Figure 5.10: Hy-Tec architecture configuration for CH estimation: (a) Hy-Tec model and (b) Single encoder UNet-based model where C represents the number of bands used to pre-train the model.

To optimize this model, AdamW, RMSprop, and SDG optimizers were employed independently, with batch size equal to 8 with 80 epochs and early stopping to break after ten epochs without improvement. Finally, as the output head performs two tasks, regression and classification, the total loss (ℓ_{HT}) is defined as:

$$\ell_{HT} = \beta_1 \cdot \ell_{H_{DB1}} + \beta_2 \cdot \ell_{H_{DB2}} + \beta_3 \cdot \ell_{CR}, \tag{5-2}$$

where $\ell_{H_{DBi}}$ is the Huber loss for DB_i block where $i \in 1,2$ and ℓ_{CR} is the classification and regression loss from the third DB block, and $\beta_i, i \in 1,2,3$ are the scaling hyper-parameters where $\beta_{1,2} = 0.75$ and $\beta_3 = 1$.

$$\ell_{H_{DBi}}(Y_{dbi}, Y'_{dbi}) = \ell_H(Y_{dbi}, Y'_{dbi}),\ \ for\ i \in 1,2, \tag{5-3}$$

$$\ell_{CR}(P_{db3}, T_{db3}, Y_{db3}, Y'_{db3}) = \ell_{CE}(P_{db3}, T_{db3}) + \alpha \cdot \ell_{RA}(Y_{db3}, Y'_{db3}) \tag{5-4}$$

Additionally, in Equation 5-3, Y_{dbi} and Y'_{dbi} are the target and truth values respectively used to calculate the regression loss ($\ell_{H_{DBi}}$) for the outputs at a lower resolution. In Equation 5-4, P_{db3} and T_{db3} are, respectively, the predicted and truth probabilities for each class and for each target and truth values used to calculate the classification loss (ℓ_{CE}, weighted cross-entropy loss), Y_{db3} and Y'_{db3} are the target and truth values respectively used to calculate the regression loss (ℓ_{RA}, adaptive and robust regression loss), and α is a scaling parameter used to balance both losses, ℓ_{CE} and ℓ_{RA}, with a value in the range 0.002 to 1.

5.4.4
Machine Learning algorithms

The ML algorithms such as LR, RF, and MLP are from Scikit Learn[3] library, while XGBoost is from XGBoost[4]. These algorithms are implemented for AGB estimation, extracting features and vegetation index from the RS images, as presented in Table5.3 and Table5.4. Also, since Sentinel-1 and Sentinel-2 were divided into dry and wet seasons, a total of fifteen features and VI were extracted. Forest mask and CH data were also used.

The RF model contains three hyperparameters: the number of decision trees (n_estimators=50), the maximum depth of the decision trees (max_depth=20), the minimum number of samples in the leaf nodes (min_samples_leaf=5). For MLP, the hyperparameters were the hidden layers(hidden_layer_sizes=350), the optimizer (solver=adam), and the initial learning rate (learning_rate_init=0.0001), batch size (batch_size=128) and momentum for gradient descent (momentum=0.75). Finally, the hyperparameters for XGBoost were the maximum depth of a tree (max_depth=15), the step size shrinkage (eta=0.25), the maximum number of nodes (max_leaves=18), the subsample ratio (subsample=0.9), and regularization term (alpha=0.5) and number of trees (n_estimators=120).

[3] https://scikit-learn.org/stable/index.html

[4] https://xgboost.readthedocs.io/en/stable/index.html

5.5
Results

The experimental results are presented in the following subsections, first for CH estimation and then for AGB estimation.

5.5.1
Canopy height estimation

A summary of the results obtained by the different models for CH estimation is presented in Table 5.6 in terms of R^2, MAE and $RMSE$. This table shows the result obtained for each best model evaluated for CH estimation. Also, the two last results presented in Table 5.6 are those obtained from the pre-trained UNet-based model used to train the Hytec model; these results are for each data type, radar and optical. These results show that the α-2MDU model achieves the best performance compared to the other models with the higher R^2 and lowest MAE AND RMSE. As expected, the α-2MDU model has better results when using a multisource dataset, giving better estimations than when using one data type or a single data source. However, the model's performance with radar data is better than optical data, with a relatively high R^2 and low MAE and RMSE values.

Table 5.6: Results for CH estimation for the different DL models.

Model	**R2 ↑**	**MAE ↓**	**RMSE ↓**
ResNet-based	0.702	4.724	6.282
α-2MDU	**0.751**	**4.068**	**5.738**
Hy-Tec	0.619	5.078	7.097
UNet based (radar data)	0.668	4.808	6.655
UNet based (optical data)	0.656	4.873	6.772

Additionally, in Fig. 5.11, the results are displayed in a graphical way, showing the comparison distribution between CH from GEDI and the predictions made by the evaluated models. Moreover, Fig. 5.11 also presents two density plots for each model. The first one is plotted considering all the CH data, and the other one considers CH data above 10m because of the large amount of data below 10m. The CH distribution (left side plot) from Fig. 5.11(a) shows that ResNet-based model performance is better in the range of approximately 1 to 31m compared to the values above 31m, which have not-so-good performance. Moreover, this model presents two peaks near the 23 and 29m values. Similarly, in Fig. 5.11(c), we can observe that the Hy-Tec model performs similarly to the ResNet-based model. The model predictions decrease from approximately 32m, with a peak near 29m. Also, this model makes more predictions in the range of 2 to 5m. On the other hand, the α-2MDU model presents better predictions, Fig. 5.11(b); however, this model also fails in predicting values above 34m approximately, presenting a peak near 30m.

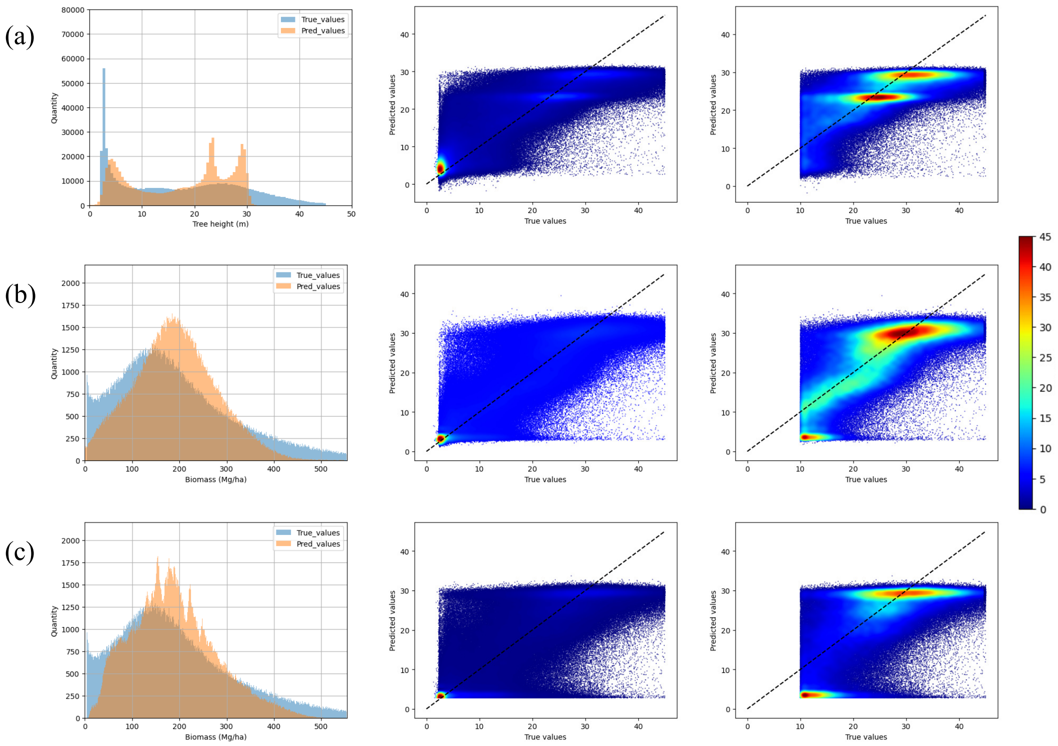


Figure 5.11: Distribution results from the different DL models tested for CH estimation: (a) ResNet-based model, (b) α-2MDU model and (c) Hy-Tec model.

As mentioned before, Fig. 5.11 presents two density plots for each model. The first density plot of each model shows a high concentration of the predicted values in the 1 to 10m interval, which does not allow us to see how the models perform for values above 5m. In this sense, the density plots on the right side present a broader view of the performance of the models. For the ResNet-based model, we can see that this model overestimates CH values below 5m, while for values above 31m, the model tends to underestimate. Moreover, the model has better predictions from 20 to 31m. On the other hand, the HyTec model predictions tend to be underestimated across the CH data. In contrast, α-2MDU model has more distributed predictions; however, some values from 5 to 15 are underestimated. Moreover, α-2MDU model was executed 10 times, using the same hyperparameters each time. In Table. 5.7, the mean and standard deviation results are presented, and we can observe that the predictions made by the model each time are not too different from the other.

Additionally, we evaluated the performance of α-2MDU for each state in

Table 5.7: Results for CH estimation for the different DL models.

	R2 ↑	**MAE** ↓	**RMSE** ↓
Mean	0.7254	4.1860	5.9163
Standard deviation	0.0128	0.1014	0.1426

the Brazilian Legal Amazon, as it is the best model. The results are shown in Table 5.8 and Fig. 5.12. We can observe from Table 5.8 that the model has better performance in Mato Grosso and Rondônia states in comparison to the other states. Also, the data distribution in Fig. 5.12(e) and Fig. 5.12(h), for Mato Grosso and Rondônia states respectively, show that these states have a similar distribution between them and total data. Moreover, the model fails to estimate CH above approximately 33 meters in both images. In contrast, the model did not perform as expected in the Amazonas state, where the negative value R^2 means that the model's predicted values for this state were poor compared to the other states.

Table 5.8: Results for CH estimation in the Brazilian Legal Amazon states.

State	**R2** ↑	**MAE** ↓	**RMSE** ↓
Acre (AC)	0.325	4.404	5.742
Amapá (AP)	0.037	6.375	7.841
Amazonas (AM)	−0.172	5.298	7.152
Maranhão (MA)	0.175	3.329	4.819
Mato Grosso (MT)	**0.771**	**3.657**	**5.108**
Pará (PA)	0.115	6.531	8.076
Roraima (RR)	0.401	4.743	6.437
Rondônia (RO)	**0.774**	**3.432**	**5.083**
Tocantins (TO)	0.483	2.962	4.505

5.5.2
Aboveground Biomass estimation

As mentioned in the previous section, for AGB estimation we used the same models as for CH estimation, and additionally we used four ML algorithms. The results of the models are summarized in Table 5.9 and Table 5.10 and are in terms of R2, MAE, and RMSE. Table 5.9 presents the results obtained by the DL models using RS data from Sentinel-1, Sentinel-2, and ALOS-2/PALSAR-2 and CH data from GEDI. Similarly, Table5.10 presents the result for the DL model using RS data, GEDI CH data, forest mask, as

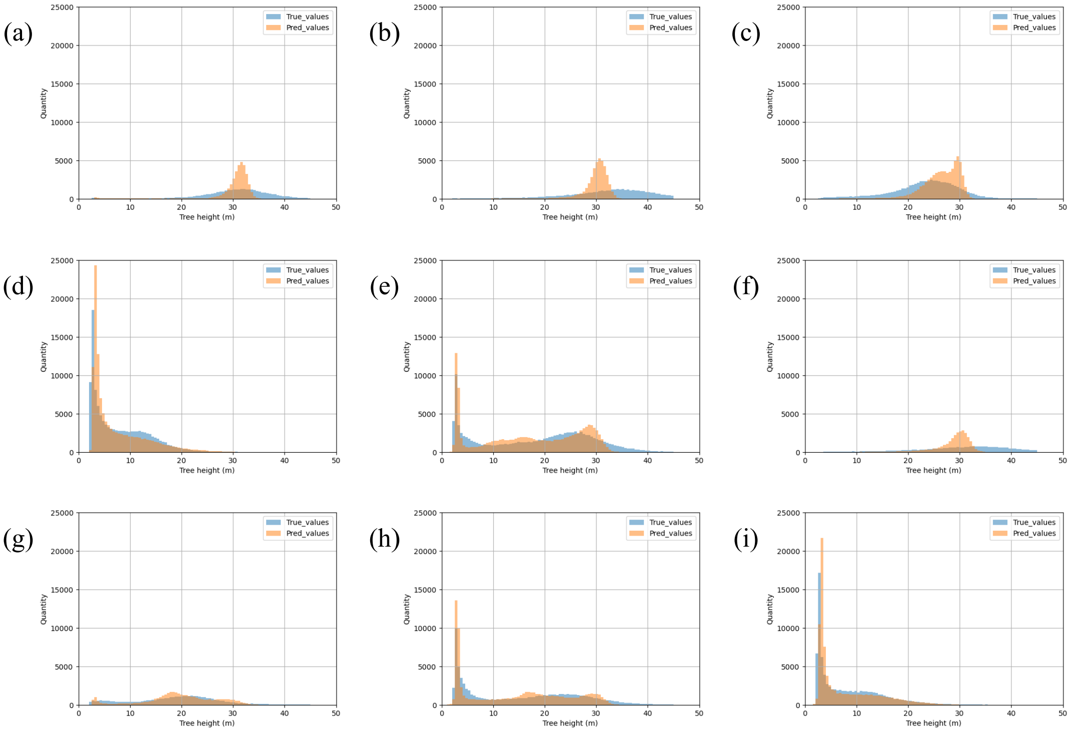


Figure 5.12: Distribution results for the CH estimation across the nine states of the Brazilian Legal Amazon: (a) AC, (b) AP, (c) AM, (d) MA, (e) MG, (f) PA, (g) RR, (h) RO and (i) TO.

well as Sentinel-1 and ALOS-2 / PALSAR-2 features and Sentinel-2 vegetation index. Also, Table5.10 presents the results from ML algorithms; these algorithms used Sentinel-1 and ALOS-2/PALSAR-2 features and Sentinel-2 vegetation index.

In Table 5.9, we observe that the best result was obtained from the α-2MDU model, using RS data and CH data as input to the model. However, this model did not achieve results as in Table 5.10, where the best DL model achieves a R^2 of 0.304. Moreover, ML algorithms achieve better results than DL models, where the best result was obtained by the XGBoost Regressor, closely followed by MLP, RF, and LR. It is important to mention that the Hy-Tec model performs poorly, as observed in previous experiments. These experiments demonstrated that the Hy-Tec model depends on UNet-based pretrained models' performance. Therefore trainings for AGB estimation was not considered until achieving reasonable performance with Unet-based models.

Fig. 5.13 presents the results for the ML and DL algorithms. The comparison distribution between the AGB from GEDI and the predictions made by the models is presented on the left side of each model. Here, we can observe that the four ML algorithms have problems in estimating biomass above $350 Mg \cdot ha^{-1}$, having almost null predictions for values above $450 Mg \cdot ha^{-1}$. Similarly, these models present problems in estimating values

Table 5.9: Results for aboveground biomass estimation for the different models using data from Sentinel-1, Sentinel-2, ALOS-2/PALSAR-2 and CH data from GEDI.

Model	**R2** $\uparrow$	**MAE** $\downarrow$	**RMSE** $\downarrow$
ResNet-based	-0.734	119.105	157.02
α-2MDU	**0.278**	**56.676**	**73.885**
UNet based (radar data)	$-2,101$	173.5	209.9
UNet based (optical data)	-0.408	112.6	141.5

Table 5.10: Results for aboveground biomass estimation for the different models using RS data, CH data from GEDI, forest mask, Sentinel-1 and ALOS-2/PALSAR-2 features and Sentinel-2 vegetation index.

Model	**R2** $\uparrow$	**MAE** $\downarrow$	**RMSE** $\downarrow$
ResNet-based	-0.694	115.2	155.2
α-2MDU	0.304	56.442	74.685
UNet based (radar data)	$-2,101$	173.5	209.9
UNet based (optical data)	-0.003	95.07	119.4
LR	0.633	51.634	72.194
MLP	0.646	49.178	70.981
RF	0.643	49.503	71.262
XGBoost Reg	**0.648**	**48.842**	**70.74**5

below $50Mg \cdot ha^{-1}$, which is more evident for MLP, RF, and XGBoost. On the other hand, in the images on the right side of each model in Fig. 5.13, we can observe that the four ML algorithms perform closer predictions for values above $300Mg \cdot ha^{-1}$ approximately and underestimate values above $400Mg \cdot ha^{-1}$. Furthermore, in Fig. 5.13 we can observe that the amount of data used to estimate AGB is considerably less than the data used to estimate CH, this is because the AGB data were filtered by overlapping CH and AGB data. This decrease in the amount of data possibly affected the performance of the DL models in comparison to the ML algorithms.

Similarly to the CH estimation, we evaluated the performance of the model, XGBoost Regressor, for each state in the Brazilian Legal Amazon since it is the best model. The results are presented in Table 5.11 and Fig. 5.14, where we can observe that the model performs well in all states and fails in estimating small AGB values. However, in Acre and Maranhão, its performance decreases. We can observe that in the state of Maranhão (see Fig. 5.14(d)), the presence of small values of AGB is predominant, and this could be the reason

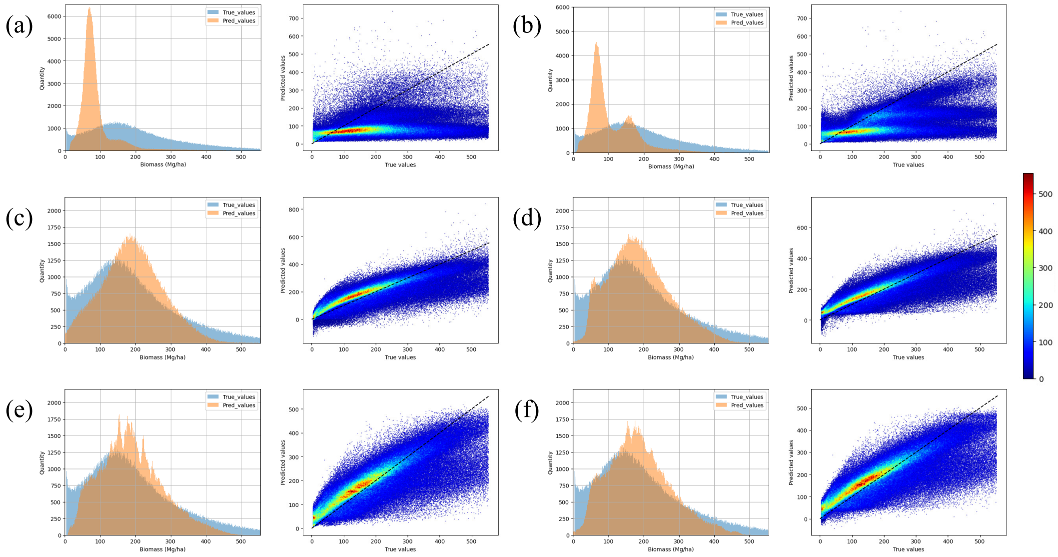


Figure 5.13: Results for the DL and ML models tested for AGB estimation: (a) ResNet-based model, (b) α-2MDU model, (c) LR, (d) MLP, (e) RF and (f) XGBoost Reg.

why the performance of the model is lower compared to the other states. Also, we can observe that Amazônas, Mato Grosso, Rondônia, and Roraima have similar distributions, presenting the same difficulty in estimating values over 350 $Mg \cdot ha^{-1}$.

Table 5.11: Results for AGB estimation in the Brazilian Legal Amazon states.

State	**R2** ↑	**MAE** ↓	**RMSE** ↓
Acre (AC)	0.279	70.944	98.086
Amapá (AP)	0.510	61.421	84.394
Amazonas (AM)	0.559	38.583	54.305
Maranhão (MA)	0.281	42.694	55.074
Mato Grosso (MT)	0.572	44.491	66.392
Pará (PA)	0.635	56.439	77.660
Roraima (RR)	0.591	37.657	54.154
Rondônia (RO)	0.497	50.499	73.707
Tocantins (TO)	0.404	47.613	69.694

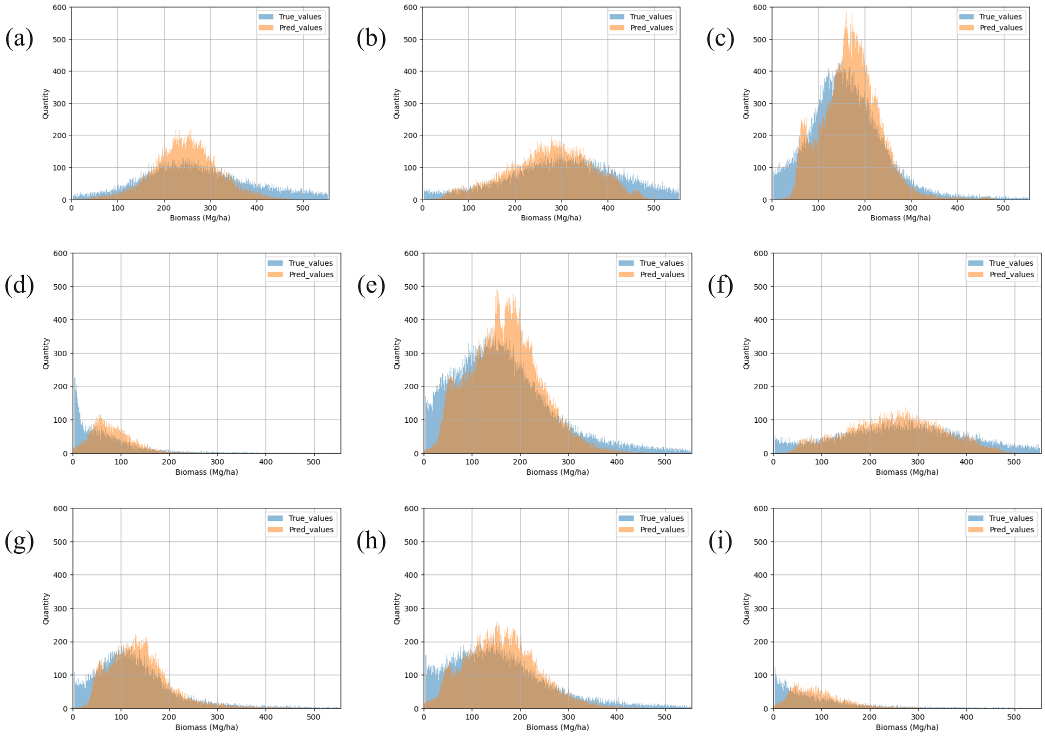


Figure 5.14: Distribution results for the AGB estimation across the nine states of the Brazilian Legal Amazon: (a) AC, (b) AP, (c) AM, (d) MA, (e) MG, (f) PA, (g) RR, (h) RO and (i) TO.

6
CONCLUSIONS

This work reported an evaluation of three state-of-the-art deep learning techniques for CH estimation: ResNet-based model, α-2MDU, and Hy-Tec. The ResNet-based model acts as a baseline for this estimation since it was used during a previous study in the same region. In addition, this work reported an evaluation of three state-of-the-art DL techniques for aboveground biomass estimation: ResNet-based model, α-2MDU, and Hy-Tec. In this case, the performance of these models was compared against four machine learning algorithms: LR, MLP, RF, and XGBoost, which are the most popular algorithms used for biomass estimation.

The experiments were carried out using public datasets across the nine states belonging to the Brazilian Legal Amazon: Amazonas, Pará, Acre, Roraima, Mato Grosso, Amapá, Rondônia, Maranhão and Tocantins. The references for CH and AGB used in the development of this work were from GEDI. The experimental analysis relied on twenty-seven remote sensing image tiles, each image tile comprising data from Sentinel-1, ALOS-2/PALSAR-2, Sentinel-2, and Forest mask from Prodes.

6.1
For Canopy Height Estimation

The DL models have demonstrated the capacity to estimate CH in a dense tropical forest. The best DL model is the α-2MDU, performing good results with an R2 of 0.751, an MAE of 4.068 meters, and an RMSE of 5.738 meters across the Brazilian Amazon. Moreover, in the analysis made for states, the best performance of the model was for Mato Grosso and Rondônia states, reaching better results than for all nine states. As expected, the use of multisource datasets and from different data types, as in this case optical and radar data, improves the model predictions.

6.2
For Aboveground Biomass Estimation

Using deep learning and remote sensing models alone did not demonstrate good results. This could be due to the small amount of data since when filtering the biomass data, the number of points with information was considerably reduced. Despite the addition of the bands with the features from sentinel-1 and ALOS-2/PALSAR-2 and the vegetation indexes from Sentinel-2, as well as the CH data and the forest mask did not achieve the result that we expected. In contrast, the machine learning algorithms presented better results, especially XGBoost, with an R2 of 0.633, an MAE of 49.775 Mg/ha, and an RMSE of 70.745 Mg/ha.

6.3
Future Directions

Future works are intended to include different kinds of data for AGB estimation, for instance, climatic data or topographic data, in order to improve the performance of the deep learning model for AGB estimation and evaluate the effects of this data addition. Moreover, once the NASA-ISRO SAR (NISAR) is launched we will be able to use its data, as this sensor operates in the L-band and penetrate the leaves and branches of the forest canopy, bouncing off the tree trunks and the ground providing detailed information about the rainforest as in the Amazon basin [92]. Another consideration is to evaluate other state-of-the-art methods.

We observed that a critical issue is the data distribution difference from the nine Brazilian Legal Amazon states. In this sense, a promising research is using domain adaptation techniques to improve the model prediction for the states across the Brazilian Amazon and mitigate the error predictions, for both CH and AGB estimation. Additionally, this could allow us to use the CH estimation as an input to the DL models for AGB estimation. Also, due to the different distributions across the train, validation, and test sets, it would be interesting to compare the distribution of these sets per state and analyze if this is the reason why the model is better for some states than others. this could allow us to use CH estimation as input to DL models for AGB estimation.

As two DL models perform a classification task, some other techniques for the distribution class, like percentiles, could be tested, but also using lower ranges for clustering AGB data. Moreover, using data of forest species in the classification task could be tested to see if the performance of the models improves.

Despite the good results obtained by the ML algorithms for AGB estimation with the features and vegetation indexes extracted, the DL models do not achieve the same results, in this sense it could be possible to test with some other vegetation indexes and analyzed which of them have a better relation with biomass.

Bibliography

[1] LIU, N.; CALDWELL, P. V.; DOBBS, G. R.; MINIAT, C. F.; BOLSTAD, P. V.; NELSON, S. A. ; SUN, G.. **Forested lands dominate drinking water supply in the conterminous united states**. Environmental Research Letters, 16(8):084008, 2021.

[2] MIGLIAVACCA, M.; MUSAVI, T.; MAHECHA, M. D.; NELSON, J. A.; KNAUER, J.; BALDOCCHI, D. D.; PEREZ-PRIEGO, O.; CHRISTIANSEN, R.; PETERS, J.; ANDERSON, K. ; OTHERS. **The three major axes of terrestrial ecosystem function**. Nature, 598(7881):468–472, 2021.

[3] BASTIN, J.-F.; FINEGOLD, Y.; GARCIA, C.; MOLLICONE, D.; REZENDE, M.; ROUTH, D.; ZOHNER, C. M. ; CROWTHER, T. W.. **The global tree restoration potential**. Science, 365(6448):76–79, 2019.

[4] PAN, Y.; MCCULLOUGH, K. ; HOLLINGER, D. Y.. **Forest biodiversity, relationships to structural and functional attributes, and stability in new england forests**. Forest Ecosystems, 5:1–12, 2018.

[5] LI, Y.; LI, M. ; WANG, Y.. **Forest aboveground biomass estimation and response to climate change based on remote sensing data**. Sustainability, 14(21):14222, 2022.

[6] UNITED NATIONS. **International arrangement on forests**, 2014. Accessed on January 01, 2024.

[7] UNITED NATIONS CLIMATE CHANGE. **The paris agreement**, 2020. Accessed on January 01, 2024.

[8] UNITED NATIONS CLIMATE CHANGE. **What is redd+?**, 2023. Accessed on January 01, 2024.

[9] RITCHIE, H.; ROSER, M.. **Forests and deforestation**. Our World in Data, 2021. https://ourworldindata.org/forests-and-deforestation.

[10] INSTITUTO DE PESQUISA ECONÔMICA APLICADA - IPEA. **O que é? amazônia legal**, 2008. Accessed on January 01, 2023.

[11] FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED STATES. **Biodiversidade brasileira**, 2017. Accessed on January 15, 2024.

[12] PARRESOL, B. R.. **Biomass**. Encyclopedia of Environmetrics, p. 196–198, 2002.

[13] BROWN, S.. **Estimating biomass and biomass change of tropical forests: a primer**, volumen 134. Food & Agriculture Org., 1997.

[14] LU, D.. **The potential and challenge of remote sensing-based biomass estimation**. International journal of remote sensing, 27(7):1297–1328, 2006.

[15] SINHA, S.; JEGANATHAN, C.; SHARMA, L. K. ; NATHAWAT, M. S.. **A review of radar remote sensing for biomass estimation**. International Journal of Environmental Science and Technology, 12(5):1779–1792, 2015.

[16] LU, D.; CHEN, Q.; WANG, G.; LIU, L.; LI, G. ; MORAN, E.. **A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems**. International Journal of Digital Earth, 9(1):63–105, 2016.

[17] MITCHARD, E. T.. **The tropical forest carbon cycle and climate change**. Nature, 559(7715):527–534, 2018.

[18] ZHU, Y.; FENG, Z.; LU, J. ; LIU, J.. **Estimation of forest biomass in beijing (china) using multisource remote sensing and forest inventory data**. Forests, 11(2):163, 2020.

[19] DORADO-RODA, I.; PASCUAL, A.; GODINHO, S.; SILVA, C. A.; BOTEQUIM, B.; RODRÍGUEZ-GONZÁLVEZ, P.; GONZÁLEZ-FERREIRO, E. ; GUERRA-HERNÁNDEZ, J.. **Assessing the accuracy of gedi data for canopy height and aboveground biomass estimates in mediterranean forests**. Remote Sensing, 13(12):2279, 2021.

[20] ALVITES, C.; O'SULLIVAN, H.; FRANCINI, S.; MARCHETTI, M.; SANTOPUOLI, G.; CHIRICI, G.; LASSERRE, B.; MARIGNANI, M. ; BAZZATO, E.. **High-resolution canopy height mapping: Integrating nasa's global ecosystem dynamics investigation (gedi) with multi-source remote sensing data**. Remote Sensing, 16(7):1281, 2024.

[21] STOJANOVA, D.; PANOV, P.; GJORGJIOSKI, V.; KOBLER, A. ; DŽEROSKI, S.. **Estimating vegetation height and canopy cover**

from remotely sensed data with machine learning. Ecological Informatics, 5(4):256–266, 2010.

[22] LECUN, Y.; BOTTOU, L.; BENGIO, Y. ; HAFFNER, P.. **Gradient-based learning applied to document recognition**. Proceedings of the IEEE, 86(11):2278–2324, 1998.

[23] ILLARIONOVA, S.; SHADRIN, D.; IGNATIEV, V.; SHAYAKHMETOV, S.; TREKIN, A. ; OSELEDETS, I.. **Estimation of the canopy height model from multispectral satellite imagery with convolutional neural networks**. IEEE Access, 10:34116–34132, 2022.

[24] DAL MOLIN JR, R.; RIZZOLI, P.. **Potential of convolutional neural networks for forest mapping using sentinel-1 interferometric short time series**. Remote Sensing, 14(6):1381, 2022.

[25] GE, S.; GU, H.; SU, W.; PRAKS, J. ; ANTROPOV, O.. **Improved semisupervised unet deep learning model for forest height mapping with satellite sar and optical data**. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15:5776–5787, 2022.

[26] RONNEBERGER, O.; FISCHER, P. ; BROX, T.. **U-net: Convolutional networks for biomedical image segmentation**. In: MEDICAL IMAGE COMPUTING AND COMPUTER-ASSISTED INTERVENTION–MICCAI 2015: 18TH INTERNATIONAL CONFERENCE, MUNICH, GERMANY, OCTOBER 5-9, 2015, PROCEEDINGS, PART III 18, p. 234–241. Springer, 2015.

[27] TOLAN, J.; YANG, H.-I.; NOSARZEWSKI, B.; COUAIRON, G.; VO, H.; BRANDT, J.; SPORE, J.; MAJUMDAR, S.; HAZIZA, D.; VAMARAJU, J. ; OTHERS. **Sub-meter resolution canopy height maps using self-supervised learning and a vision transformer trained on aerial and gedi lidar**. arXiv e-prints, p. arXiv–2304, 2023.

[28] BISPO, P. D. C.; PARDINI, M.; PAPATHANASSIOU, K. P.; KUGLER, F.; BALZTER, H.; RAINS, D.; DOS SANTOS, J. R.; RIZAEV, I. G.; TANSEY, K.; DOS SANTOS, M. N. ; OTHERS. **Mapping forest successional stages in the brazilian amazon using forest heights derived from tandem-x sar interferometry**. Remote Sensing of Environment, 232:111194, 2019.

[29] SAWADA, Y.; SUWA, R.; JINDO, K.; ENDO, T.; OKI, K.; SAWADA, H.; ARAI, E.; SHIMABUKURO, Y. E.; CELES, C. H. S.; CAMPOS, M. A. A.

; OTHERS. **A new 500-m resolution map of canopy height for amazon forest using spaceborne lidar and cloud-free modis imagery**. International Journal of Applied Earth Observation and Geoinformation, 43:92–101, 2015.

[30] QIN, Y.; XIAO, X.; TANG, H.; DUBAYAH, R.; DOUGHTY, R.; LIU, D.; LIU, F.; SHIMABUKURO, Y.; ARAI, E.; WANG, X. ; OTHERS. **Annual maps of forest cover in the brazilian amazon from analyses of palsar and modis images**. Earth System Science Data, 16(1):321–336, 2024.

[31] DALAGNOL, R.; WAGNER, F. H.; EMILIO, T.; STREHER, A. S.; GALVÃO, L. S.; OMETTO, J. P. ; ARAGAO, L. E.. **Canopy palm cover across the brazilian amazon forests mapped with airborne lidar data and deep learning**. Remote Sensing in Ecology and Conservation, 8(5):601–614, 2022.

[32] WANG, X.; LIU, C.; LV, G.; XU, J. ; CUI, G.. **Integrating multi-source remote sensing to assess forest aboveground biomass in the khingan mountains of north-eastern china using machine-learning algorithms**. Remote Sensing, 14(4):1039, 2022.

[33] TANG, Z.; XIA, X.; HUANG, Y.; LU, Y. ; GUO, Z.. **Estimation of national forest aboveground biomass from multi-source remotely sensed dataset with machine learning algorithms in china**. Remote Sensing, 14(21):5487, 2022.

[34] SAATCHI, S. S.; HOUGHTON, R. A.; DOS SANTOS ALVALA, R.; SOARES, J. V. ; YU, Y.. **Distribution of aboveground live biomass in the amazon basin**. Global change biology, 13(4):816–837, 2007.

[35] KANMEGNE TAMGA, D.; LATIFI, H.; ULLMANN, T.; BAUMHAUER, R.; BAYALA, J. ; THIEL, M.. **Estimation of aboveground biomass in agroforestry systems over three climatic regions in west africa using sentinel-1, sentinel-2, alos, and gedi data**. Sensors, 23(1):349, 2022.

[36] WANG, C.; ZHANG, W.; JI, Y.; MARINO, A.; LI, C.; WANG, L.; ZHAO, H. ; WANG, M.. **Estimation of aboveground biomass for different forest types using data from sentinel-1, sentinel-2, alos palsar-2, and gedi**. Forests, 15(1):215, 2024.

[37] DONG, L.; DU, H.; HAN, N.; LI, X.; ZHU, D.; MAO, F.; ZHANG, M.; ZHENG, J.; LIU, H.; HUANG, Z. ; OTHERS. **Application of convo-**

lutional neural network on lei bamboo above-ground-biomass (agb) estimation using worldview-2. Remote Sensing, 12(6):958, 2020.

[38] PASCARELLA, A. E.; GIACCO, G.; RIGIROLI, M.; MARRONE, S. ; SANSONE, C.. **Reuse: Regressive unet for carbon storage and above-ground biomass estimation.** Journal of Imaging, 9(3):61, 2023.

[39] DONG, W.; MITCHARD, E. T.; YU, H.; HANCOCK, S. ; RYAN, C. M.. **Forest aboveground biomass estimation using gedi and earth observation data through attention-based deep learning.** arXiv preprint arXiv:2311.03067, 2023.

[40] DE ALMEIDA, C. T.; GALVÃO, L. S.; OMETTO, J. P. H. B.; JACON, A. D.; DE SOUZA PEREIRA, F. R.; SATO, L. Y.; LOPES, A. P.; DE ALENCASTRO GRAÇA, P. M. L.; DE JESUS SILVA, C. V.; FERREIRA-FERREIRA, J. ; OTHERS. **Combining lidar and hyperspectral data for above-ground biomass modeling in the brazilian amazon using different regression algorithms.** Remote Sensing of Environment, 232:111323, 2019.

[41] SANTOS, E. G. D.; SHIMABUKURO, Y. E.; MENDES DE MOURA, Y.; GONÇALVES, F. G.; JORGE, A.; GASPARINI, K. A.; ARAI, E.; DUARTE, V. ; OMETTO, J. P.. **Multi-scale approach to estimating aboveground biomass in the brazilian amazon using landsat and lidar data.** International journal of remote sensing, 40(22):8635–8645, 2019.

[42] SCHUH, M.; FAVARIN, J. A. S.; MARCHESAN, J.; ALBA, E.; FERNANDO BERRA, E. ; SOARES PEREIRA, R.. **Machine learning and generalized linear model techniques to predict aboveground biomass in amazon rainforest using lidar data.** Journal of Applied Remote Sensing, 14(3):034518–034518, 2020.

[43] BISPO, P. D. C.; RODRÍGUEZ-VEIGA, P.; ZIMBRES, B.; DO COUTO DE MIRANDA, S.; HENRIQUE GIUSTI CEZARE, C.; FLEMING, S.; BALDACCHINO, F.; LOUIS, V.; RAINS, D.; GARCIA, M. ; OTHERS. **Woody aboveground biomass mapping of the brazilian savanna with a multi-sensor and machine learning approach.** Remote Sensing, 12(17):2685, 2020.

[44] OMETTO, J. P.; GORGENS, E. B.; DE SOUZA PEREIRA, F. R.; SATO, L.; DE ASSIS, M. L. R.; CANTINHO, R.; LONGO, M.; JACON, A. D. ; KELLER, M.. **A biomass map of the brazilian amazon from multisource remote sensing.** Scientific Data, 10(1):668, 2023.

[45] DE JESUS, J. B.; KUPLICH, T. M.; DE C. BARRETO, Í. D.; HILLEBRAND, F. L. ; DA ROSA, C. N.. **Estimation of aboveground biomass of arboreal species in the semi-arid region of brazil using sar (synthetic aperture radar) images**. Journal of Arid Land, 15(6):695–709, 2023.

[46] TEJADA, G.; GÖRGENS, E. B.; ESPÍRITO-SANTO, F. D. B.; CANTINHO, R. Z. ; OMETTO, J. P.. **Evaluating spatial coverage of data on the aboveground biomass in undisturbed forests in the brazilian amazon**. Carbon Balance and Management, 14:1–18, 2019.

[47] HE, K.; ZHANG, X.; REN, S. ; SUN, J.. **Deep residual learning for image recognition**. In: PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, p. 770–778, 2016.

[48] BRIDGWATER, T.. **Biomass for energy**. Journal of the Science of Food and Agriculture, 86(12):1755–1768, 2006.

[49] NELSON, B. W.; MESQUITA, R.; PEREIRA, J. L.; DE SOUZA, S. G. A.; BATISTA, G. T. ; COUTO, L. B.. **Allometric regressions for improved estimate of secondary forest biomass in the central amazon**. Forest ecology and management, 117(1-3):149–167, 1999.

[50] CHAVE, J.; ANDALO, C.; BROWN, S.; CAIRNS, M. A.; CHAMBERS, J. Q.; EAMUS, D.; FÖLSTER, H.; FROMARD, F.; HIGUCHI, N.; KIRA, T. ; OTHERS. **Tree allometry and improved estimation of carbon stocks and balance in tropical forests**. Oecologia, 145:87–99, 2005.

[51] BROWN, S.; GILLESPIE, A. J. ; LUGO, A. E.. **Biomass estimation methods for tropical forests with applications to forest inventory data**. Forest science, 35(4):881–902, 1989.

[52] BROWN, S.; LUGO, A. E.. **Aboveground biomass estimates for tropical moist forests of the brazilian amazon**. Interciencia. Caracas, 17(1):8–18, 1992.

[53] DJOMO, A. N.; CHIMI, C. D.. **Tree allometric equations for estimation of above, below and total biomass in a tropical moist forest: Case study with application to remote sensing**. Forest Ecology and Management, 391:184–193, 2017.

[54] SHERMAN, R. E.; FAHEY, T. J. ; MARTINEZ, P.. **Spatial patterns of biomass and aboveground net primary productivity in a**

mangrove ecosystem in the dominican republic. Ecosystems, p. 384–398, 2003.

[55] SU, H.; SHEN, W.; WANG, J.; ALI, A. ; LI, M.. **Machine learning and geostatistical approaches for estimating aboveground biomass in chinese subtropical forests**. Forest Ecosystems, 7:1–20, 2020.

[56] ZHANG, L.; SHAO, Z.; LIU, J. ; CHENG, Q.. **Deep learning based retrieval of forest aboveground biomass from combined lidar and landsat 8 data**. Remote Sensing, 11(12):1459, 2019.

[57] READ, J. M.; CHAMBERS, C. ; TORRADO, M.. **Remote sensing**. In: INTERNATIONAL ENCYCLOPEDIA OF HUMAN GEOGRAPHY, SECOND EDITION, p. 411–422. Elsevier, 2019.

[58] SABINS JR, F. F.; ELLIS, J. M.. **Remote sensing: Principles, interpretation, and applications**. Waveland Press, 2020.

[59] GEORGE JOSEPH, C. J.. **Fundamentals of Remote Sensing**. Universities Press, 2018.

[60] RUSTAMOV, R. B.; HASANOVA, S. ; ZEYNALOVA, M. H.. **Multipurposeful application of geospatial data**. BoD–Books on Demand, 2018.

[61] NATIONAL AERONAUTICS AND SPACE ADMINISTRATION - NASA. **What is remote sensing?**, 2021. Accessed on March 25, 2024.

[62] PROGRAMME OF THE EUROPEAN UNION, COPERNICUS AND EUROPEAN SPACE AGENCY. **Sentinel-1**, 2023. Accessed on March 25, 2024.

[63] CASTILLO, J. A. A.; APAN, A. A.; MARASENI, T. N. ; SALMO III, S. G.. **Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the philippines using sentinel imagery**. ISPRS Journal of Photogrammetry and Remote Sensing, 134:70–85, 2017.

[64] MARTIMORT, P.; ARINO, O.; BERGER, M.; BIASUTTI, R.; CARNICERO, B.; DEL BELLO, U.; FERNANDEZ, V.; GASCON, F.; GRECO, B.; SILVESTRIN, P. ; OTHERS. **Sentinel-2 optical high resolution mission for gmes operational services**. In: 2007 IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, p. 2677–2680. IEEE, 2007.

[65] HAN, J.; WEI, C.; CHEN, Y.; LIU, W.; SONG, P.; ZHANG, D.; WANG, A.; SONG, X.; WANG, X. ; HUANG, J.. **Mapping above-ground biomass of winter oilseed rape using high spatial resolution satellite data at parcel scale under waterlogging conditions**. Remote Sensing, 9(3):238, 2017.

[66] SENTINELHUB. **Sentinel-2 bands**, 2016. Accessed on March 25, 2024.

[67] JAPAN AEROSPACE EXPLORATION AGENCY - JAXA. **Palsar-2**, 2016. Accessed on March 25, 2024.

[68] FLORES-ANDERSON, A. I.; HERNDON, K. E.; THAPA, R. B. ; CHERRINGTON, E.. **The sar handbook: comprehensive methodologies for forest monitoring and biomass estimation**. Technical report, SERVIR, 2019.

[69] GOLSHANI, P.; MAGHSOUDI, Y. ; SOHRABI, H.. **Relating alos-2 palsar-2 parameters to biomass and structure of temperate broadleaf hyrcanian forests**. Journal of the Indian Society of Remote Sensing, 47:749–761, 2019.

[70] SHIMADA, M.; ITOH, T.; MOTOOKA, T.; WATANABE, M.; SHIRAISHI, T.; THAPA, R. ; LUCAS, R.. **New global forest/non-forest maps from alos palsar data (2007–2010)**. Remote Sensing of environment, 155:13–31, 2014.

[71] NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA). **Earth in the third dimension: First gedi data available**, 2020. Accessed on March 25, 2024.

[72] NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA) AND UNIVERSITY OF MERYLAND. **Mission overview**, 2018. Accessed on March 25, 2024.

[73] ORACLE; CHEN, M.. **What is machine learning?**, 2024. Accessed on March 25, 2024.

[74] HINTON, G.; VINYALS, O. ; DEAN, J.. **Distilling the knowledge in a neural network**. arXiv preprint arXiv:1503.02531, 2015.

[75] INTERNATIONAL BUSINESS MACHINES CORPORATION - IBM; BERGMANN, D.. **What is knowledge distillation?**, 2024. Accessed on May 30, 2024.

[76] GOODFELLOW, I.; BENGIO, Y. ; COURVILLE, A.. **Deep Learning**. Adaptive Computation and Machine Learning series. MIT Press, 2016.

[77] KRICHEN, M.. **Convolutional neural networks: A survey**. Computers, 12(8):151, 2023.

[78] DOSOVITSKIY, A.; BEYER, L.; KOLESNIKOV, A.; WEISSENBORN, D.; ZHAI, X.; UNTERTHINER, T.; DEHGHANI, M.; MINDERER, M.; HEIGOLD, G.; GELLY, S. ; OTHERS. **An image is worth 16x16 words: Transformers for image recognition at scale**. arXiv preprint arXiv:2010.11929, 2020.

[79] VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, Ł. ; POLOSUKHIN, I.. **Attention is all you need**. Advances in neural information processing systems, 30, 2017.

[80] CHEN, B.; LIU, Y.; ZHANG, Z.; LU, G. ; KONG, A. W. K.. **Transattunet: Multi-level attention-guided u-net with transformer for medical image segmentation**. IEEE Transactions on Emerging Topics in Computational Intelligence, 2023.

[81] FAYAD, I.; CIAIS, P.; SCHWARTZ, M.; WIGNERON, J.-P.; BAGHDADI, N.; DE TRUCHIS, A.; D'ASPREMONT, A.; FRAPPART, F.; SAATCHI, S.; SEAN, E. ; OTHERS. **Hy-tec: a hybrid vision transformer model for high-resolution and large-scale mapping of canopy height**. Remote Sensing of Environment, 302:113945, 2024.

[82] WANG, Q.; MA, Y.; ZHAO, K. ; TIAN, Y.. **A comprehensive survey of loss functions in machine learning**. Annals of Data Science, p. 1–26, 2020.

[83] TIAN, Y.; SU, D.; LAURIA, S. ; LIU, X.. **Recent advances on loss functions in deep learning for computer vision**. Neurocomputing, 497:129–158, 2022.

[84] RUBINSTEIN, R.. **The cross-entropy method for combinatorial and continuous optimization**. Methodology and computing in applied probability, 1:127–190, 1999.

[85] HUBER, P. J.. **Robust estimation of a location parameter**. In: BREAKTHROUGHS IN STATISTICS: METHODOLOGY AND DISTRIBUTION, p. 492–518. Springer, 1992.

[86] BARRON, J. T.. **A general and adaptive robust loss function**. In: PROCEEDINGS OF THE IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, p. 4331–4339, 2019.

[87] ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; GONÇALVES, J. D. M.; SPAROVEK, G. ; OTHERS. **Köppen's climate classification map for brazil**. Meteorologische zeitschrift, 22(6):711–728, 2013.

[88] DUBREUIL, V.; PECHUTTI FANTE, K.; PLANCHON, O. ; NETO, J. L. S.. **Les types de climats annuels au brésil: une application de la classification de köppen de 1961 à 2015**. EchoGéo, (41), 2017.

[89] LUIZ-SILVA, W.; OSCAR-JÚNIOR, A. C.; CAVALCANTI, I. F. A. ; TREISTMAN, F.. **An overview of precipitation climatology in brazil: space-time variability of frequency and intensity associated with atmospheric systems**. Hydrological Sciences Journal, 66(2):289–308, 2021.

[90] INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. COORDENAÇÃO GERAL DE OBSERVAÇÃO DA TERRA. PROGRAMA DE MONITORAMENTO DA AMAZÔNIA E DEMAIS BIOMAS. **Desmatamento – amazônia legal**, 2019. Accessed on September 22, 2023.

[91] GLOBO. **Pesquisadores chegam à árvore mais alta da amazônia, de 88,5 metros e cerca de 400 anos**, 2022. Accessed on September 22, 2023.

[92] CHEN, N.; TSENDBAZAR, N.-E.; SUAREZ, D. R.; SILVA-JUNIOR, C. H.; VERBESSELT, J. ; HEROLD, M.. **Revealing the spatial variation in biomass uptake rates of brazil's secondary forests**. ISPRS Journal of Photogrammetry and Remote Sensing, 208:233–244, 2024.

[92] NASA. **Nasa-isro radar mission to provide dynamic view of forests, wetlands**, 2023. Accessed on November 01, 2024.

A

Alpha MDU blocks

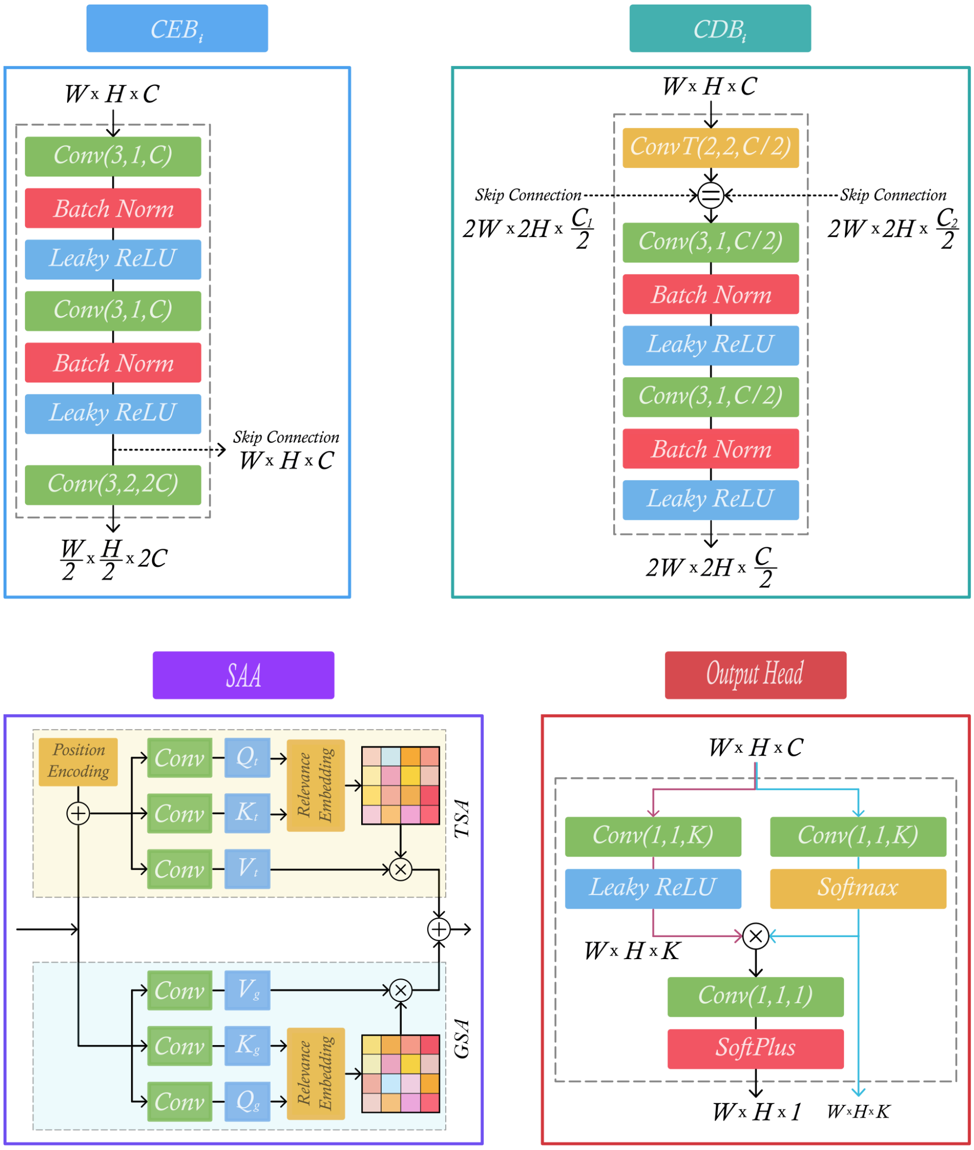


Figure A.1: α-2MDU model blocks: convolution encoder block (CEB), convolution decoder block (CDB), self-aware attentio module (SAA) and output head.

B
Hy-Tec U-Net blocks

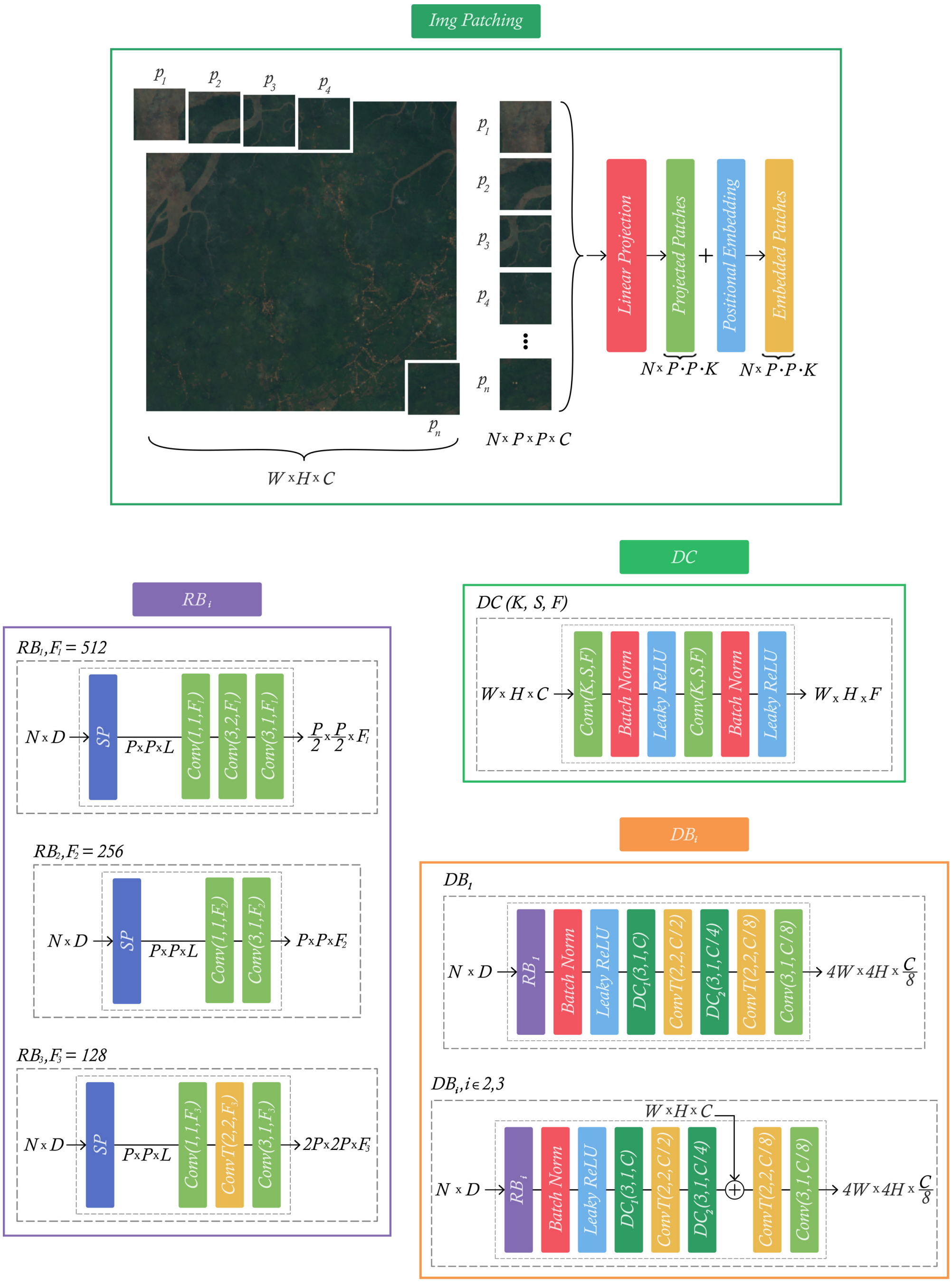


Figure B.1: Hy-Tec blocks: image patching, reprojection block (RB), decoder convolution block (DC) and decoder block (DB).

C
Published paper

The following paper entitled *'Canopy Height Estimation in the Brazilian Amazon using Multisource Remote Sensing Datasets'* has been published at the international congress Latin America Conference on Computational Intelligence 2024