

Ước tính sinh khối trên mặt đất bằng viễn thám tích hợp và thuật toán học máy:  
nghiên cứu điểm tại Khu dự trữ sinh quyển miền Tây Nghệ An, Việt Nam

Đỗ Thị Nhung<sup>1</sup>, Nguyễn Đình Hùng<sup>2</sup>, Phạm Ngọc Hải<sup>2</sup>,  
Đặng Đỗ Lâm Phương<sup>1</sup>, Nguyễn Thị Diễm My<sup>1</sup>, Phạm Văn Mạnh<sup>1\*</sup>

<sup>1</sup>Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội

<sup>2</sup>Viện Điều tra, Quy hoạch Rừng, Bộ Nông nghiệp và Phát triển Nông thôn

Estimation of Above ground biomass using a combination  
of integrated Remote sensing and Machine learning algorithms: a case study  
in the Western Nghe An Biosphere Reserve, Vietnam

Do Thi Nhung<sup>1</sup>, Nguyen Dinh Hung<sup>2</sup>, Pham Ngoc Hai<sup>2</sup>,  
Dang Do Lam Phuong<sup>1</sup>, Nguyen Thi Diem My<sup>1</sup>, Pham Van Manh<sup>1\*</sup>

<sup>1</sup>University of Science, Vietnam National University Hanoi

<sup>2</sup>Forest Inventory and Planning Institute (FIPI), Ministry of Agriculture and Rural Development

\*Corresponding author: manh10101984@gmail.com

<https://doi.org/10.55250/jo.vnuf.12.4.2023.081-092>

**TÓM TẮT**

Ước tính sinh khối trên mặt đất (AGB) từ dữ liệu viễn thám quang học và Radar khẩu độ tổng hợp (SAR) là một cách tiếp cận thiết thực để theo dõi chất lượng rừng tại khu dự trữ sinh quyển dài hạn. Sinh khối của rừng có ý nghĩa quan trọng thể hiện khả năng lưu trữ Carbon lớn và góp phần làm giảm biến đổi khí hậu toàn cầu. Nghiên cứu này trình bày phương pháp tích hợp dựa trên dữ liệu viễn thám đa nguồn và thuật toán học máy định lượng AGB và sự phân bố không gian của các kiểu rừng trong Khu dự trữ sinh quyển miền Tây Nghệ An, Việt Nam. Dữ liệu khảo sát thực địa năm 2022 được sử dụng với 169 ô mẫu được thu thập, trong đó 118 ô tham gia mô hình học máy để ước tính AGB và 51 ô còn lại được sử dụng để xác nhận kết quả. Hệ số xác định ( $R^2$ ), sai số bình phương trung bình (RMSE) và sai số trung bình tuyệt đối (MAE) được sử dụng để đánh giá và xác nhận hiệu suất của mô hình. Kết quả cho thấy cả 3 chỉ tiêu của mô hình đều tốt với RMSE và MAE có mức sai số dưới 30 Mg/ha và  $R^2$  khoảng 0,81 cho ước tính AGB. Nghiên cứu này cung cấp cái nhìn sâu sắc mới về mô hình ước tính AGB dựa trên công nghệ viễn thám đa nguồn cho rừng nhiệt đới tại Khu dự trữ sinh quyển thế giới thông qua phân tích toàn diện về dữ liệu viễn thám và mô hình học máy.

**ABSTRACT**

Estimating above-ground biomass (AGB) through the utilization of optical remote sensing and synthetic aperture radar (SAR) data presents a practical approach for the long-term monitoring of forest quality in a biosphere reserve. The significance of forest biomass lies in its capacity for substantial carbon storage and its contribution to mitigating global climate change. This study introduces an integrated methodology that relies on multi-source remote sensing data and machine learning algorithms to quantify above-ground biomass (AGB) and analyze the spatial distribution of diverse forest types within the Western Nghe An Biosphere Reserve in Vietnam. A total of 169 sample plots were collected during on-site surveys conducted in 2022. Out of these, 118 plots were employed for machine learning modeling to estimate AGB, while the remaining 51 plots were reserved for result validation. The model's performance was evaluated and confirmed using metrics such as the coefficient of determination ( $R^2$ ), root mean square error (RMSE), and mean absolute error (MAE). The results demonstrate the effectiveness of the model, with RMSE and MAE exhibiting errors below 30 Mg/ha and an  $R^2$  value of approximately 0.81 for AGB estimation. Through a comprehensive analysis of remote sensing data and machine learning models, this study provides a fresh and insightful perspective on AGB estimation models based on multi-source remote sensing technology for tropical forests within the World Biosphere Reserve.

**Thông tin chung:**

Ngày nhận bài: 16/06/2023

Ngày phản biện: 19/07/2023

Ngày quyết định đăng: 02/08/2023

**Từ khóa:**

học máy, Khu dự trữ sinh quyển thế giới, Sinh khối trên mặt đất, Tây Nghệ An, viễn thám, Việt Nam.

**Keywords:**

above ground biomass, machine learning, remote sensing, Vietnam, Western Nghe An Biosphere Reserve, World biosphere reserve.

## 1. ĐẶT VẤN ĐỀ

Hệ sinh thái rừng có ý nghĩa to lớn là hệ sinh thái tự nhiên lớn nhất và quan trọng nhất trên đất liền. Hệ sinh thái rừng đóng một vai trò quan trọng trong việc duy trì cân bằng sinh thái toàn cầu và tạo điều kiện thuận lợi cho quá trình tiến hóa sinh học và diễn thế cộng đồng trên toàn thế giới [1]. Ước tính chính xác sinh khối rừng là rất quan trọng để đánh giá khả năng hấp thụ Carbon và cân bằng Carbon của các hệ sinh thái rừng, đặc biệt khi nghiên cứu chu trình Carbon trên quy mô lớn [2]. Các phép đo thực địa truyền thống và các phương pháp viễn thám thường được sử dụng để ước tính sinh khối trên mặt đất của rừng [3]. Mặc dù các phép đo thực địa cung cấp các phép tính AGB chính xác cho các lâm phần nhỏ, nhưng tính khả thi của chúng giảm đi khi áp dụng cho quy mô khu vực do chi phí cao, cường độ lao động và thời gian [4]. Dữ liệu viễn thám, đặc biệt là SAR tán xạ ngược, đã được chứng minh là có hiệu quả trong việc đo lường và giám sát sinh khối rừng trong khu vực [5]. Các cảm biến bị động và chủ động đều được sử dụng để ước tính AGB, dữ liệu SAR tán xạ ngược thể hiện khả năng xuyên qua tán rừng và tương tác với các thành phần sinh khối chính như thân cây và cành cây [6]. Cường độ tán xạ ngược của SAR tăng khi sinh khối rừng cao hơn và độ nhạy của nó thay đổi tùy theo bước sóng [7]. Các bước sóng dài hơn thể hiện các giá trị bão hòa tán xạ cao hơn và mối tương quan mạnh hơn giữa SAR tán xạ ngược và sinh khối, làm cho chúng phù hợp hơn để ước tính sinh khối. Tuy nhiên, việc sử dụng dữ liệu SAR bước sóng dài hơn thường không thực tế do chi phí liên quan đến các vệ tinh thương mại cung cấp dữ liệu này. Vệ tinh Sentinel-1A của Cơ quan Vũ trụ châu Âu cung cấp dữ liệu SAR với bước sóng băng tần C (C-band) có độ phân giải không gian cao, truy cập miễn phí trên toàn thế giới [8].

Tuy nhiên, việc chỉ sử dụng duy nhất dữ liệu này là không đủ để ước tính chính xác khối lượng sinh khối của rừng do sự bão hòa và khả năng đâm xuyên thấp của băng tần C [9]. Ngoài dữ liệu SAR, dữ liệu viễn thám quang học (Landsat, SPOT, WorldView, Pléiades, Sentinel-2) và các sản phẩm tạo ra từ dữ liệu này như chỉ số thực vật và hình ảnh cấu trúc, đã được chứng minh rằng chúng có mối liên hệ

chặt chẽ với khối lượng sinh khối [8]. Nghiên cứu trước đây đã chỉ ra rằng việc kết hợp dữ liệu từ các nguồn khác nhau để ước tính AGB sẽ mang lại kết quả chính xác hơn so với việc sử dụng dữ liệu từ một nguồn duy nhất, đặc biệt là ở các khu vực nhiệt đới nơi có cấu trúc rừng phức tạp và đa dạng thành phần loài cây [10]. Sentinel-1 và Sentinel-2 là vệ tinh của chương trình giám sát môi trường mang tên Copernicus do Cơ quan Hàng không Vũ trụ châu Âu (ESA) thực hiện với sứ mệnh quan sát trái đất nhằm mục đích giám sát môi trường và an ninh Toàn cầu. Tuy nhiên, việc kết hợp giữa Sentinel-2 và Sentinel-1 vẫn chưa được nghiên cứu đầy đủ trong mô hình ước tính AGB tại các khu rừng nhiệt đới ẩm, đặc biệt tại các Khu dự trữ sinh quyển thế giới. Bên cạnh việc chọn dữ liệu viễn thám phù hợp, việc lựa chọn thuật toán tối ưu cũng là rất quan trọng trong việc thiết lập mô hình ước tính AGB. Phương pháp hồi quy thống kê truyền thống đơn giản và dễ tính toán, nhưng không thể mô tả một cách hiệu quả mối quan hệ phi tuyến phức tạp giữa AGB và dữ liệu viễn thám [8]. Để cải thiện khả năng ước lượng phi tuyến tính, các phương pháp học máy (decision tree, artificial neural network, k-nearest neighbor, support vector machine) đã được áp dụng để ước tính AGB từ dữ liệu viễn thám [11]. Các nghiên cứu trước đây đã chỉ ra rằng các thuật toán dựa trên cây quyết định (decision tree) đạt hiệu suất cao trong việc ước tính khối lượng sinh khối [12, 13]. Hơn nữa, các thuật toán học máy thường có nhiều tham số có thể điều chỉnh, đóng vai trò quan trọng trong việc điều chỉnh mô hình, tuy nhiên, việc điều chỉnh tham số này thường bị bỏ qua. Nhiều nghiên cứu cũng chỉ ra rằng độ nhạy của các tham số trong mô hình học máy dựa trên cây quyết định là khác nhau và quá trình điều chỉnh tham số đó có ý nghĩa quan trọng đối với hiệu suất của mô hình tính toán.

Trong nghiên cứu này, một phương pháp tích hợp dựa trên dữ liệu Sentinel-2A, Sentinel-1A SAR và sử dụng phương pháp hồi quy kết hợp với mô hình học máy Cubist để định lượng AGB và sự phân bố không gian của các kiểu rừng đặc trưng tại Khu dự trữ sinh quyển miền Tây Nghệ An – một trong những Khu dự trữ sinh quyển thế giới và là khu rừng tự nhiên lớn nhất ở Việt

Nam. Mục tiêu của nghiên cứu bao gồm một số bước cần thiết: (i) xác định mô hình phù hợp nhất bằng cách sử dụng dữ liệu AGB được đo trên thực địa và đánh giá hiệu suất của mô hình bằng hệ số xác định ( $R^2$ ), sai số bình phương trung bình (RMSE) và sai số trung bình tuyệt đối (MAE); (ii) ước tính và thành lập bản đồ phân bố không gian AGB.

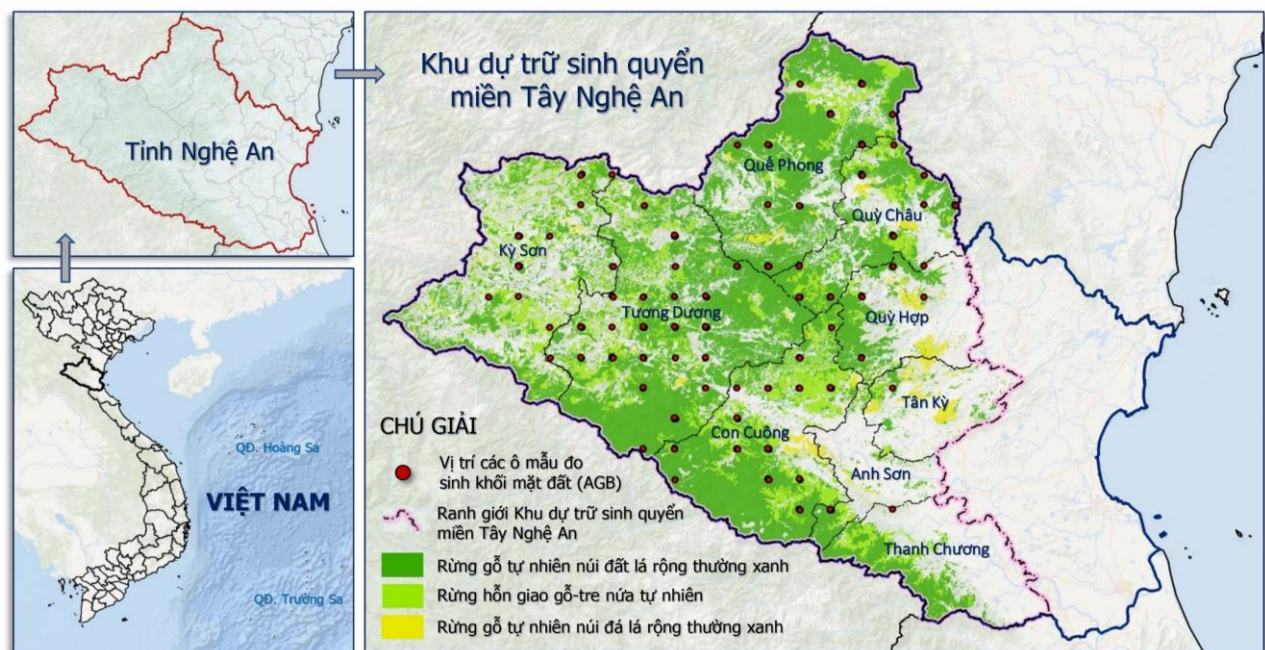
## 2. PHƯƠNG PHÁP NGHIÊN CỨU

### 2.1. Khu vực nghiên cứu

Khu dự trữ sinh quyển miền Tây Nghệ An là một khu vực đặc biệt và quan trọng về mặt sinh thái và đa dạng sinh học nằm ở tỉnh Nghệ An, Việt Nam. Khu dự trữ này được UNESCO công nhận là Khu dự trữ sinh quyển thế giới vào năm 2007. Khu dự trữ sinh quyển miền Tây Nghệ An được xem như một hành lang xanh quan trọng, kết nối ba vùng lõi là Vườn quốc gia Pù Mát, Khu bảo tồn thiên nhiên Pù Huông và Khu bảo tồn thiên nhiên Pù Hoạt. Điều này tạo ra sự liên tục về môi trường sống và sinh cảnh, góp phần hiệu quả trong việc bảo tồn đa dạng sinh học và giảm thiểu tác động tiêu cực của sự chia cắt do

hoạt động kinh tế của con người. Khu dự trữ sinh quyển miền Tây Nghệ An bao gồm 9 huyện: Kỳ Sơn, Tương Dương, Con Cuông, Thanh Chương, Quỳnh Hợp, Quỳnh Châu, Quế Phong, Tân Kỳ và Anh Sơn (Hình 1).

Địa hình của Khu dự trữ sinh quyển miền Tây Nghệ An được đặc trưng là núi cao với độ cao từ 140 m - 2.711 m so với mực nước biển. Khu dự trữ sinh quyển miền Tây Nghệ An là một vùng nhiệt đới gió mùa với lượng mưa trung bình hàng năm khoảng 1.372 mm và nhiệt độ trung bình hàng năm là 21,4°C. Khu vực nghiên cứu vẫn giữ được nguồn tài nguyên rừng tự nhiên phong phú và đây là nguồn dự trữ Carbon rừng quan trọng, rất cần thiết cho quản lý rừng quốc gia và mang lại lợi ích cho việc đánh giá, giám sát và bảo vệ. Khu dự trữ sinh quyển miền Tây Nghệ An với các hệ sinh thái rừng chính như: (i) Rừng gỗ tự nhiên núi đất lá rộng thường xanh, (ii) Rừng gỗ tự nhiên núi đá lá rộng thường xanh, (iii) Rừng hỗn giao gỗ-tre nửa tự nhiên.



Hình 1. Vị trí của khu vực nghiên cứu và phân bố không gian của ô mẫu AGB thực địa

### 2.2. Dữ liệu khảo sát thực địa

Bộ dữ liệu ô mẫu AGB năm 2022 của Viện Điều tra, Quy hoạch rừng (FIPI), đã được sử dụng trong nghiên cứu này (Hình 1). Các ô mẫu thành lập được tiến hành đáp ứng các tiêu chí sau: (1) Tính đại diện của các loại rừng trong

Khu dự trữ sinh quyển miền Tây Nghệ An và điều kiện địa hình tương ứng; (2) Độ che phủ của một số kích cỡ cây khác nhau; và (3) Thiết lập trên những loại rừng ít bị xáo trộn, nơi có nhiều kích thước cây lớn. Các ô mẫu có kích thước 30×33,3 m (0,1 ha) được thiết lập và trong

các góc của ô mẫu, bốn ô mẫu phụ được bố trí với kích thước 5×5 m (0,0025 ha). Đối với cây gỗ, đo đường kính ngang ngực (DBH) của cây có  $DBH \geq 6$  cm và với cây nhỏ hơn 6 cm đo ở chiều cao  $\geq 1,3$  m thì chỉ đếm số cây trong ô phụ. Các giá thông số này được sử dụng để tính toán AGB dựa trên các phương trình sinh khối được xây dựng cho các loại rừng ở Khu dự trữ sinh quyển miền Tây Nghệ An. Giá trị AGB lấy từ số liệu thực địa tại các ô mẫu dao động từ 23,94 Mg/ha đến 929,96 Mg/ha. Toàn bộ các ô mẫu thực địa được phân tầng theo tỷ lệ kiểu rừng nhiệt đới và được chia thành hai phần, một phần tham gia mô hình huấn luyện (chọn khoảng 70% số ô mẫu) và phần còn lại để kiểm chứng và đánh giá độ chính xác của kết quả dự đoán (khoảng 30% số ô mẫu). Cần lưu ý rằng AGB thực địa từ các ô mẫu được sử dụng làm biến đầu vào (biến phụ thuộc) của mô hình ước tính AGB, trong khi các chỉ số lý sinh khác nhau được tính toán từ hình ảnh Sentinel-2A, Sentinel-1A SAR và dữ liệu địa hình (biến độc lập). Như vậy, đã thu được 169 ô mẫu, trong đó có 89 ô mẫu thuộc kiểu rừng gỗ tự nhiên núi đất lá rộng thường xanh, 14 ô mẫu thuộc kiểu rừng gỗ tự nhiên núi đá lá rộng thường xanh và 66 ô mẫu thuộc kiểu rừng hỗn giao gỗ-tre nửa tự nhiên.

### **2.3. Các biến dự đoán từ Sentinel-1A và Sentinel-2A**

Dữ liệu SAR được sử dụng trong nghiên cứu này được thu thập từ vệ tinh Sentinel-1A, băng tần C (C-band) với tần số trung tâm là 5.405 GHz, chiều rộng dải 250 km, độ phân giải không gian 5×20 m. Trong nghiên cứu này, hình ảnh Level-1 Ground Range Detected (GRD) của Sentinel-1A thu nhận tháng 4/2022 với hai phân cực (VV: truyền dọc–thu dọc và VH: truyền dọc–thu ngang). Dữ liệu Sentinel-1A SAR được xử lý trước bằng hộp công cụ (S1TBX) được nhúng trong gói phần mềm SNAP từ cơ quan Vũ trụ Châu Âu phát triển (<https://step.esa.int/>). Các giai đoạn chính của quá trình tiền xử lý bao gồm: (i) Hiệu chuẩn đo tán xạ để chuyển đổi các giá trị của hình ảnh thô (DN) sang hệ số tán xạ ngược của Radar ( $\sigma_{naught}$ ); (ii) Sử dụng

bộ lọc thích ứng Frost tăng cường với kích thước cửa sổ di chuyển (7×7) được áp dụng để lọc nhiễu hạt và làm mịn hệ số tán xạ ngược của dữ liệu; (iii) Hiệu chỉnh địa hình bằng cách sử dụng dữ liệu mô hình độ cao ở độ phân giải không gian 10 m để sửa các biến dạng hình học. Ảnh vệ tinh Sentinel-2A đa phổ thu nhận tháng 4/2022 với độ phủ mây dưới 3%, được hiệu chỉnh các hiệu ứng bức xạ/khí quyển nhằm loại bỏ các nhiễu bằng cách sử dụng quy trình xử lý ATCOR được tích hợp trong gói phần mềm Catalyst Professional (<https://catalyst.earth/>). Quy trình xử lý bao gồm ba phần chính như: (i) hiệu chỉnh đỉnh phản xạ ở tầng khí quyển; (ii) loại bỏ mây mù trên ảnh và (iii) chuyển đổi giá trị bức xạ về giá trị phản xạ bề mặt. Cuối cùng, cả hai loại dữ liệu Sentinel-1A và Sentinel-2A đều được hiệu chỉnh về hệ tọa độ VN2000 múi 48. Mặc dù hạn chế của băng tần C trong việc tương tác với các thành phần sâu hơn của rừng đã được đề cập, một số công trình đã điều tra các ứng dụng tiềm năng của Sentinel-1A với nhiều chỉ số hình ảnh quang học [5], [8]. Nghiên cứu này là phần tiếp theo của nghiên cứu về băng tần C từ Sentinel-1A SAR trong ước tính AGB tại các khu rừng nhiệt đới. Ngoài ra, các hình ảnh quang học cung cấp thông tin hữu ích về điều kiện rừng nhiệt đới bằng cách biến đổi các dải quang phổ để tăng cường sự đóng góp của các đặc tính thực vật hoặc các thành phần lý sinh của lá và thân.

Nhiều loại chỉ số thực vật khác nhau về biến đổi và các mục tiêu cần thiết đã được sử dụng trong nghiên cứu này. Các chỉ số chiết tách từ dữ liệu ảnh Sentinel-2A và Sentinel-1A SAR được xem xét. Ngoài ra, các chỉ số được tính toán từ địa hình cũng được sử dụng trong nghiên cứu là độ cao (elevation), độ dốc (slope) và hướng dốc (aspect). Các chỉ số này được thành lập từ mô hình số độ cao (DEM) có độ phân giải không gian 10 m, được thu thập từ Bộ Tài nguyên và Môi trường. Bảng 1 tổng hợp các biến dự đoán được sử dụng trong nghiên cứu này để ước tính AGB từ dữ liệu Sentinel-1A SAR, Sentinel-2A và các chỉ số từ địa hình của Khu dự trữ sinh quyển miền Tây Nghệ An.

**Bảng 1. Các biến dự đoán từ dữ liệu Sentinel-1A SAR, Sentinel-2A và địa hình**

	Biến độc lập	Tên/Phương trình	Trích dẫn
Chỉ số quang phổ từ ảnh Sentinel-2A	Chỉ số sắc tố không nhạy cảm với cấu trúc (Structurally Independent Pigment Index – SIPI)	$SIPI = \frac{(NIR - BLUE)}{(NIR - RED)}$	[14]
	Chỉ số thực vật khác biệt chuẩn hóa (Normalized Difference Vegetation Index – NDVI)	$NDVI = \frac{(NIR - RED)}{(NIR + RED)}$	
	Tỷ lệ đơn gian cận hồng ngoại/xanh lục (Simple Ratio Near Infrared/Green – SRNG)	$SRNG = \frac{NIR}{GREEN}$	[15]
	Tỷ lệ đơn gian cận hồng ngoại/đỏ (Simple Ratio Near Infrared/Red – SRNR)	$SRNR = \frac{NIR}{RED}$	[16]
	Chỉ số thực vật có hiệu chỉnh phân xạ mặt đất (Soil Adjusted Vegetation Index – SAVI)	$SAVI = \frac{1.25 \times (NIR - RED)}{NIR + RED + 0.25}$	[17]
	Chỉ số thực vật chuyển đổi có hiệu chỉnh phân xạ mặt đất (Transformed Soil Adjusted Vegetation Index – TSAVI)	$TSAVI = \frac{a \times (NIR - a \times RED - b)}{NIR + RED \cdot 0.25}$	[18]
	Chỉ số thực vật vuông góc (Perpendicular Vegetation Index - PVI)	$PVI = (\alpha \times NIR) - (\beta \times RED)$	[19]
	Chỉ số thực vật có dải động rộng (Wide Dynamic Range Vegetation Index – WDRVI)	$WDRVI = \frac{0.1NIR - RED}{0.1NIR + RED}$	[20]
	Chỉ số thực vật khác biệt chuẩn hóa xanh lam (Green Normalized Difference Vegetation Index – GNDVI)	$GNDVI = \frac{(NIR - GREEN)}{(NIR + GREEN)}$	[21]
	Chỉ số thực vật khác biệt được chuẩn hóa lại (Renormalized Difference Vegetation Index – RDVI)	$RDVI = \frac{NIR - RED}{\sqrt{NIR + RED}}$	[22]
Chỉ số tán xạ ngược từ ảnh Sentinel-1A	Phân cực truyền dọc – nhận ngang (Vertical Transmit-Horizontal Receive - VH)	Sigma <sub>θ</sub> _VH (dB)	[5]
	Phân cực truyền dọc – nhận dọc (Vertical Transmit-Vertical Receive - VV)	Sigma <sub>θ</sub> _VV (dB)	
	Chỉ số phân cực tổng (Polarization addition - ADD)	VH+VV (dB)	
	Chỉ số phân cực trung bình (Polarization average - AVERAGE)	(VH+VV)/2 (dB)	
	Chỉ số phân cực tích (Polarization multiply - MULTI)	VH×VV (dB)	
Chỉ số từ địa hình	Độ cao (DEM 10m)	ELEVATION	
	Độ dốc (DEM 10m)	SLOPE	
	Hướng dốc (DEM 10m)	ASPECT	

**2.4. Bối cảnh của thuật toán học máy được sử dụng trong nghiên cứu**

Kể từ nghiên cứu đầu tiên của Quinlan [23], Cubist đã được sử dụng rộng rãi để giải quyết nhiều vấn đề trong ứng dụng hỏi quy dựa trên kỹ thuật từ cây quyết định (decision tree), hỏi quy tuyến tính (linear regression) và hỏi quy từng phần (piecewise regression) để tạo ra một mô hình dự đoán mạnh mẽ. Trong nghiên cứu, mô hình Cubist được lựa chọn dựa trên các tài liệu nghiên cứu trước đây như [24], Cubist là một thuật toán dựa trên các nguyên tắc được sử dụng để xây dựng các mô hình dự báo dựa trên việc phân tích dữ liệu đầu vào, các mô hình vẫn được kết hợp bằng cách sử dụng kết hợp tuyến tính của hai mô hình được thể hiện (Phương trình 1).

$$\hat{y}_{par} = a\hat{y}_k + (1 - a)\hat{y}_p \tag{1}$$

Trong đó:  $\hat{y}_k$  là dự đoán từ mô hình hiện tại và  $\hat{y}_p$  là mô hình gốc phía trên trong nhánh mô hình. So với các cây của mô hình, Cubist tính toán các tỷ lệ lẫn bằng cách sử dụng một phương trình thay thế với hệ số a thể hiện độ mịn của mô hình, với  $e_p$  là giá trị phần dư mô hình gốc,  $e_k$  là giá trị tương tự đối với mô hình dự đoán. Đầu tiên, mô hình xác định quy trình làm mịn phương sai giữa hai bộ phần dư của  $e_p, e_k$ . Đây là thước đo tổng thể về mối quan hệ tuyến tính giữa hai bộ phần dư. Nếu độ đo sự biến thiên cùng nhau của hai biến ngẫu nhiên lớn (hiệp phương sai), điều này ngụ ý rằng phần dư thường có cùng dấu và độ lớn tương đối, trong khi giá trị gần bằng 0 sẽ cho thấy không có mối

quan hệ tuyến tính giữa hai mô hình. Hệ số làm mịn ( $a$ ) được sử dụng bởi Cubist (Phương trình 2).

$$a = \frac{\text{Var}(e_p) - \text{Cov}(e_k, e_p)}{\text{Var}(e_p - e_k)} \quad (2)$$

Mô hình học máy Cubist cũng cho thấy sự khác biệt về phương sai giữa hai mô hình. Nguyên tắc là hiệu phương sai của sai số mô hình gốc lớn hơn tương quan giữa sai số của các mô hình dự đoán và mô hình gốc, khi đó phương pháp làm mịn có xu hướng gán trọng số cho mô hình dự đoán với trọng số lớn hơn mô hình gốc và ngược lại. Cuối cùng, mô hình có hệ số RMSE nhỏ nhất sẽ có trọng số cao hơn trong mô hình làm mịn. Khi các mô hình có cùng RMSE, chúng sẽ có trọng lượng như nhau trong quá trình làm mịn bất kể mối tương quan nào. Do đó, Cubist có lợi thế hơn các mô hình khác về (i) các kỹ thuật được sử dụng để làm mịn mô hình tuyến tính, tạo quy tắc và cắt tỉa; (ii) một quy trình tăng cường tùy chọn được sử dụng; và (iii) các dự đoán do mô hình tạo ra sử dụng các điểm lân cận từ tập dữ liệu huấn luyện. Một ưu điểm khác của Cubist đối với dự đoán là việc sử dụng các cây nhánh và hiệu chỉnh dựa trên trường hợp trong bộ dữ liệu phi tuyến tính, điều thường xảy ra với dữ liệu viễn thám [25]. Mô hình Cubist có khả năng tự động loại bỏ các biến không quan trọng hoặc tạo biến mới từ các biến ban đầu để cải thiện hiệu suất dự đoán của mô hình nhờ vào đặc điểm của thuật toán mà mô hình sử dụng [24]. Do đó, Cubist có thể được hiểu ngắn gọn và dễ dàng giải thích kết hợp với các bộ dữ liệu nghiên cứu được chia thành các trường dữ liệu độc lập trong nghiên cứu này.

Nghiên cứu này giảm thiểu sự phức tạp của các mô hình thực nghiệm và lựa chọn mô hình phù hợp nhất để tính toán AGB trích xuất từ dữ liệu ảnh vệ tinh. Điều này đạt được bằng cách ước tính sai số dự đoán mong muốn và cung cấp chỉ báo độ chính xác của mô hình dự đoán. Trong nghiên cứu này, phương pháp xác thực chéo k-fold đã được sử dụng, trong đó dữ liệu được chia ngẫu nhiên thành 10 lần, đây là con số tối ưu để xác thực chéo như đã được đề xuất trong các nghiên cứu [26]. Đầu tiên, dữ liệu

được chia ngẫu nhiên thành 10 tập hợp con, với một trong các tập hợp con được sử dụng làm tập dữ liệu xác nhận và chín tập hợp con còn lại được sử dụng để huấn luyện. Việc huấn luyện được lặp lại 10 lần bằng cách sử dụng tất cả các tập con huấn luyện. Ở lần lặp lại cuối cùng, một bộ xác nhận đã được sử dụng để xác thực một lần. Để đánh giá hiệu quả của mô hình học máy trong mô hình, nghiên cứu này sử dụng các chỉ số đánh giá hiệu suất của mô hình để hiệu chuẩn và xác nhận được đánh giá thông qua ước tính Hệ số xác định ( $R^2$ ), Sai số bình phương trung bình gốc (RMSE) và Sai số tuyệt đối trung bình (MAE).

Hệ số xác định ( $R^2$ ) rất hữu ích trong việc giải thích các biến dự báo độc lập (Phương trình 3).

$$R^2 = 1 - \frac{\sum e^2}{\sum y^2} = \frac{\beta_2 \sum y_t x_{2t} + \beta_3 \sum y_t x_{3t}}{\sum y_t^2} \quad (3)$$

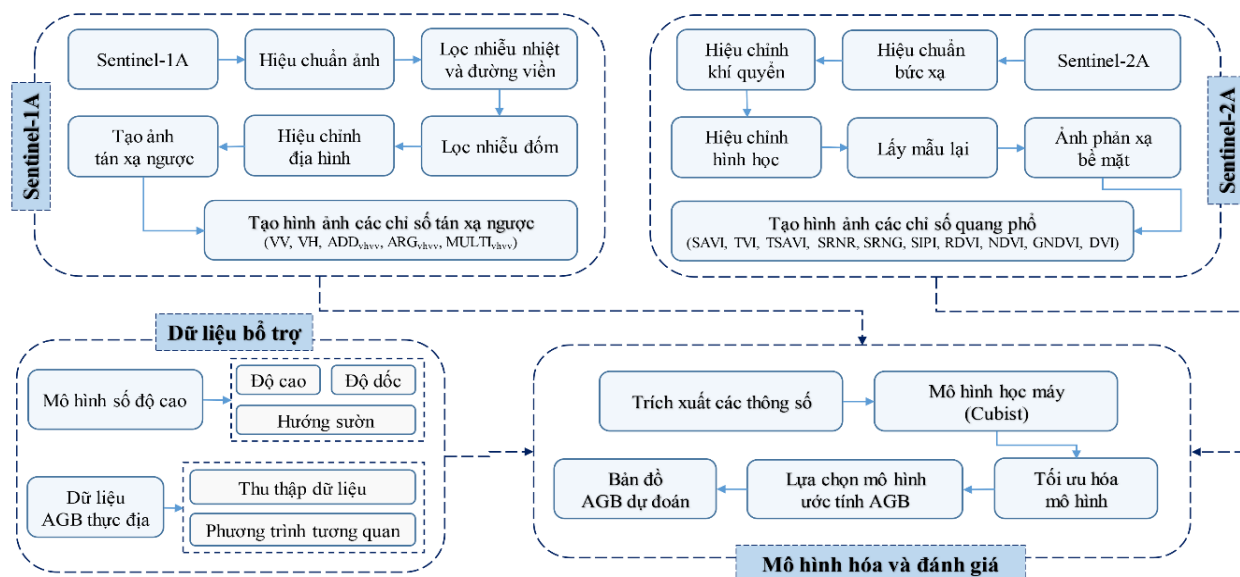
Sai số bình phương trung bình gốc (RMSE) là hàm mục tiêu tối ưu của quy trình tối ưu hóa và để phản ánh độ chính xác của mô hình ước tính bằng cách so sánh dữ liệu dự đoán với dữ liệu thực địa (Phương trình 4). Hạn chế của giá trị này là nó nhạy cảm với sai số lớn nên cần thực hiện sàng lọc sơ bộ dữ liệu đầu vào để loại bỏ bất kỳ giá trị ngoại lệ nào.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n}} \quad (4)$$

Sai số tuyệt đối trung bình (MAE) cũng cung cấp lỗi dự đoán trung bình với điểm số có hướng âm, nghĩa là giá trị càng thấp càng tốt (Phương trình 5).

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (5)$$

Giá trị  $R^2$  cao hơn và các giá trị RMSE, MAE thấp hơn cho thấy hiệu suất ước tính AGB của mô hình tốt hơn. Để tối ưu hóa việc ước tính AGB, giá trị thông số này đã được điều chỉnh trong mỗi lần học máy và được phân chia ngẫu nhiên (70% dữ liệu để huấn luyện và 30% dữ liệu còn lại để xác nhận) từ bộ dữ liệu gốc. Dữ liệu huấn luyện (118 mẫu) và dữ liệu kiểm chứng (51 mẫu) được tách ngẫu nhiên từ 169 mẫu AGB thực địa.



Hình 2. Sơ đồ quy trình nghiên cứu ước tính AGB áp dụng khu vực nghiên cứu

### 3. KẾT QUẢ VÀ THẢO LUẬN

#### 3.1. Ước tính Sinh khối trên mặt đất (AGB) dựa trên mô hình học máy Cubist

Công việc này nhằm mục đích tìm ra mô hình phù hợp nhất nhưng để đảm bảo mô hình này không bị quá khớp. Mô hình có thể đạt được thông qua việc ước tính lỗi dự đoán kỳ vọng. Lựa chọn các thông số tối ưu có tác động đáng kể đến hiệu suất của các mô hình hoặc hội quy thông qua việc loại bỏ các biến không liên quan và có thể thúc đẩy hoạt động một cách hiệu quả. Nghiên cứu sử dụng tổng số 118 mẫu hiện

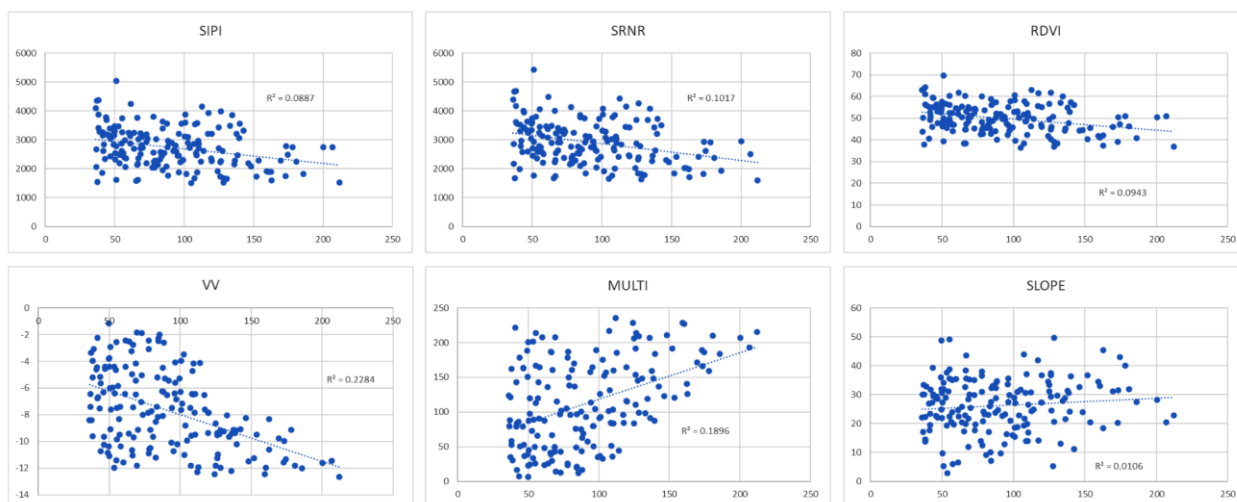
trường để đào tạo trong mô hình học máy Cubist để tính toán AGB và 51 mẫu còn lại để đánh giá hiệu suất và kết quả dự đoán. Kết quả của mô hình học máy đã được tối ưu hóa với quy trình điều chỉnh bằng cách xác nhận chéo 10 lần theo chu kỳ. Mô hình học máy Cubist đã tự động loại bỏ những biến không quan trọng và đưa ra các kết quả ước tính AGB giảm sự phức tạp hóa của các mô hình. Từ đó, đưa ra lựa chọn các biến chính là một trong những yếu tố quan trọng để ước tính sinh khối trên mặt đất.

Bảng 2. Kết quả thống kê các chỉ số đánh giá hiệu suất của các mô hình Cubist và xác thực bằng cách sử dụng bộ dữ liệu kiểm chứng

Lựa chọn mô hình	Thông số tối ưu	Hiệu suất mô hình			Xác thực		
		CV-RMSE	CV-R <sup>2</sup>	CV-MAE	RMSE	R <sup>2</sup>	MAE
1	SIPI, SRNR, RDVI, VV, MULTI, SLOPE	29,99	0,83	26,02	32,35	0,81	28,21
2	SRNR, SIPI	31,66	0,81	27,20	32,94	0,79	28,54
3	SAVI, NDVI, TVI, SRNG, RDVI, SLOPE, GNDVI, ADD	31,68	0,81	27,46	33,14	0,77	28,77
4	SAVI, NDVI, TVI, SRNG, GNDVI, ADD	31,81	0,82	27,30	33,38	0,80	28,78
5	SAVI, TVI, SRNG, VV	33,33	0,67	29,90	34,67	0,64	31,59
6	NDVI, TVI, SRNG, ADD	32,72	0,76	29,23	34,07	0,72	31,12
7	TVI, SAVI, VV	32,78	0,67	29,39	33,99	0,65	31,51
8	NDVI, TVI, SRNG, ADD	32,09	0,82	28,87	33,96	0,80	30,16
9	SAVI, TVI, SRNG, GNDVI, ADD	32,71	0,67	29,26	34,44	0,66	30,97
10	NDVI, SRNG, TVI, RDVI, SLOPE, GNDVI, ADD	32,08	0,82	28,85	33,55	0,78	30,18

Bảng 2 thể hiện thống kê các chỉ số đánh giá hiệu suất từ các mô hình Cubist và kết quả đánh giá độ chính xác của từng mô hình ước tính AGB so với bộ dữ liệu kiểm chứng thu được từ các ô mẫu. Dựa trên kết quả của các mô hình được tối ưu hóa, CV-RMSE dao động từ 29,99 Mg/ha đến 33,33 Mg/ha, CV-R<sup>2</sup> dao động từ 0,67 đến 0,83 và CV-MAE dao động từ 26,02 Mg/ha đến 29,90 Mg/ha. Trong khi đó, giá trị AGB được ước tính từ dữ liệu ảnh viễn thám đa nguồn và dữ liệu từ địa hình có giá trị lớn hơn không đáng kể so với hiệu suất của mô hình được tính toán (Bảng 2). Độ chính xác ước tính bằng cách sử dụng mô hình học máy với các số

$$AGB = 54,79 + (0,07 \times SIPI) - (0,06 \times SRNR) - (1,40 \times RDVI) - (1,97 \times VV) - (0,04 \times MULTI) + (0,22 \times SLOPE) \tag{6}$$



Hình 3. Biểu đồ phân tán các thông số tối ưu (SIPI, SRNR, RDVI, VV, MULTI, SLOPE) được chọn bởi mô hình học máy Cubist so với AGB của các ô mẫu

### 3.2. Ước tính sinh khối trên mặt đất tại Khu dự trữ sinh quyển miền Tây Nghệ An

Trong số các mô hình khác nhau, mô hình với sáu biến ((SIPI, SRNR, RDVI, VV, MULTI, SLOPE) đã thực hiện dự đoán tối ưu cho ước tính AGB và lập bản đồ của Khu dự trữ sinh quyển miền Tây Nghệ An. Dựa trên mô hình học máy Cubist được xác định, bản đồ ước tính AGB được thành lập. Các khu vực không có rừng như đất trồng, đất nông nghiệp, nước, rừng không có thảm thực vật đã được loại bỏ khỏi kết quả của bản đồ AGB. Bản đồ AGB được hiển thị theo năm lớp (Hình 4) cho thấy: (i) dưới 50 (Mg/ha), (ii) từ 50 đến 100 (Mg/ha), từ 100 đến 150 (Mg/ha), từ 150 đến 200 (Mg/ha), trên 200 (Mg/ha).

Phân bố không gian AGB dự đoán phù hợp

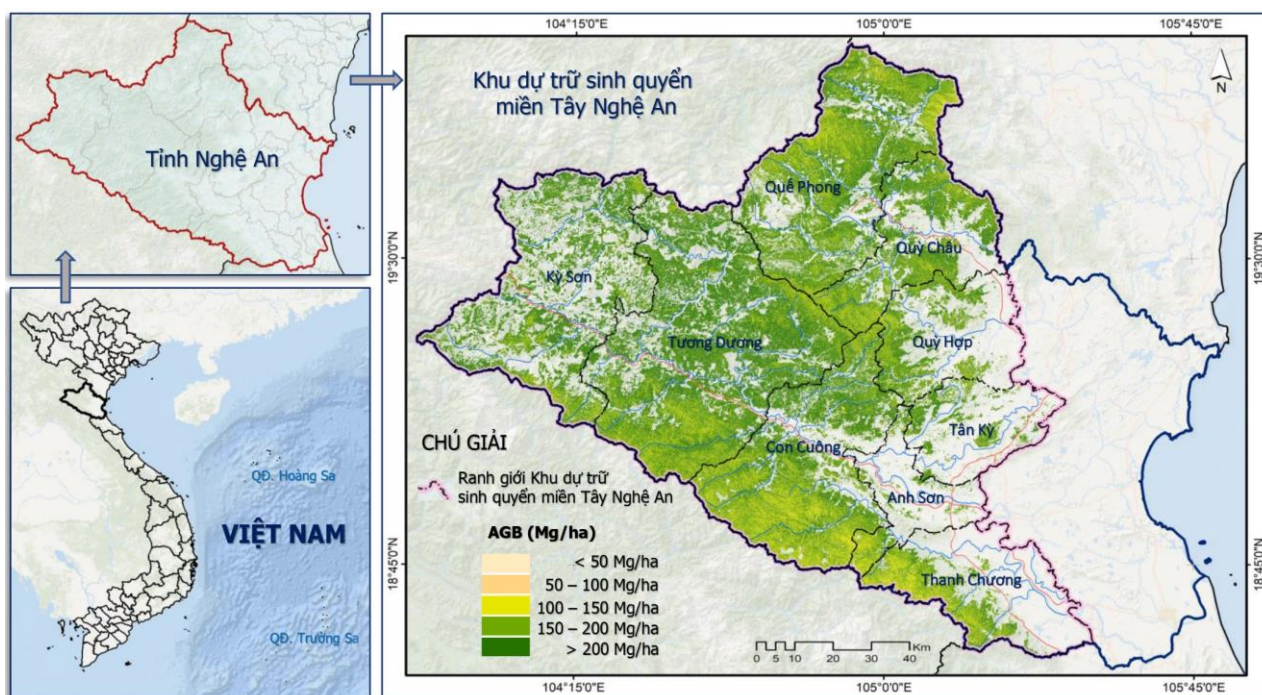
liệu kết hợp khác nhau là tương đối tốt. Nhìn chung, AGB ước tính trong các mô hình đều có giá trị CV-RMSE, CV-R<sup>2</sup> và CV-MAE chấp nhận được. Tuy nhiên, mô hình 1 có hiệu suất tối ưu với sáu biến được lựa chọn (SIPI, SRNR, RDVI, VV, MULTI, SLOPE), với các giá trị của hệ số R<sup>2</sup>=0.81, RMSE=32,35 Mg/ha, và MAE=28,21 Mg/ha là tốt nhất trong số các mô hình. Do đó, kết quả mô hình 1 (Phương trình 6) được áp dụng để ước tính AGB từ dữ liệu ảnh viễn thám đa nguồn kết hợp với dữ liệu từ địa hình tại Khu dự trữ sinh quyển miền Tây Nghệ An, Việt Nam.

với thực tế quan sát tại Khu dự trữ sinh quyển miền Tây Nghệ An, Việt Nam. Các khu vực màu xanh lục đậm thể hiện AGB có giá trị lớn hơn trên 200 Mg/ha trên mỗi pixel so với các khu vực màu xanh lục vàng (từ 150-200 Mg/ha) và màu vàng nghệ đậm từ 100-150 Mg/ha. Điều này có liên quan đến thực tế là khu vực này phân bố các cây cao có đường kính lớn được tìm thấy ở Vườn Quốc gia Pù Mát (thuộc huyện Tương Dương, Con Cuông và Anh Sơn) và hai Khu Bảo tồn thiên nhiên Pù Hường (thuộc huyện Quế Phong, Quỳnh Châu, Quỳnh Hợp, Tương Dương và Con Cuông), Khu Bảo tồn thiên nhiên Pù Hoạt (huyện Quế Phong). Các khu vực này được bao phủ phần lớn bởi Rừng gỗ tự nhiên núi đất lá rộng thường xanh và Rừng hỗn giao gỗ - tre nứa tự nhiên, được bảo vệ nghiêm ngặt.



Ngược lại, các khu vực được thể hiện bằng màu vàng đậm từ 50-100 Mg/ha và vàng nhạt dưới 50 Mg/ha có giá trị AGB trung bình và thấp, tương ứng. Những màu này (đặc biệt là trong

diện tích Rừng gỗ tự nhiên núi đá lá rộng thường xanh) xuất hiện trên các đỉnh núi và rặng núi ở phía Bắc và phía Tây Nam của khu vực nghiên cứu (Hình 4).



**Hình 4. Bản đồ phân bố không gian AGB tại Khu dự trữ sinh quyển miền Tây Nghệ An**

Giá trị ước tính AGB (Bảng 3) phần lớn là trên 200 Mg/ha với diện tích 352.145,91 ha, chủ yếu tập trung ở kiểu Rừng gỗ tự nhiên núi đất lá rộng thường xanh (233.850,09 ha) và Rừng hỗn giao gỗ - tre nửa tự nhiên (6.317,02 ha), phần còn lại thuộc về kiểu Rừng gỗ tự nhiên núi đá lá rộng thường xanh (11.978,80 ha). Diện tích 268.628,28 ha với giá trị AGB từ 150-200 Mg/ha tập trung chủ yếu ở Rừng gỗ tự nhiên núi

đất lá rộng thường xanh (206.270,81 ha) và Rừng hỗn giao gỗ - tre nửa tự nhiên (54.639,34 ha). Không gian trong rừng hỗn giao hẹp hơn so với rừng gỗ tự nhiên bởi cây trưởng thành và cây tre nửa xếp chồng lên nhau. Sinh khối trên mặt đất có sự phân bố không đều, với một phần lớn tập trung ở vùng gốc cây gỗ và các vùng có sự tương tác giữa cây gỗ và cây.

**Bảng 3. Kết quả thống kê các loại diện tích rừng tự nhiên được phân thành các khoảng ước tính AGB**

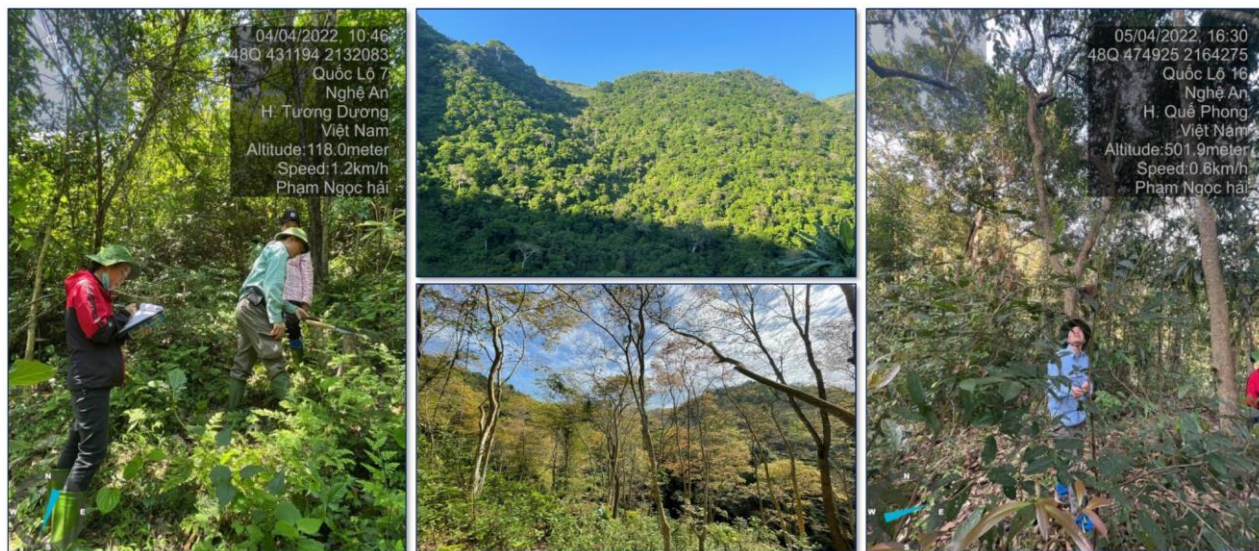
	Diện tích rừng tự nhiên có mật độ					Tổng (ha)
	0-50 Mg/ha	50-100 Mg/ha	100-150 Mg/ha	150-200 Mg/ha	>200 Mg/ha	
Rừng gỗ tự nhiên núi đất lá rộng thường xanh	136,06	6.176,2	87.890,89	206.270,81	233.850,09	534.324,05
Rừng gỗ tự nhiên núi đá lá rộng thường xanh	0,83	930,14	3.766,13	7.718,13	11.978,80	24.394,03
Rừng hỗn giao gỗ-tre nửa tự nhiên	28,67	934,34	13.833,44	54.639,34	106.317,02	175.752,81
<b>Tổng (ha)</b>	<b>165,56</b>	<b>8.040,68</b>	<b>105.490,46</b>	<b>268.628,28</b>	<b>352.145,91</b>	<b>734.470,89</b>

Khu dự trữ sinh quyển miền Tây Nghệ An có tính đa dạng sinh học rất cao đại diện cho hầu hết kiểu rừng của rừng mưa nhiệt đới, các sinh cảnh sống đa dạng bao gồm: núi, đất ngập nước, suối và sinh cảnh khác. Đây là khu vực duy nhất

của miền Bắc của Việt Nam còn lại một diện tích lớn rừng nguyên sinh đang được bảo vệ tốt, đặc biệt là khu vực dọc biên giới Việt - Lào. Hệ sinh thái rừng nơi đây là một trong những hệ sinh thái đa dạng nhất trên toàn quốc và có giá

trị dịch vụ hệ sinh thái cao nhất. Việc đánh giá nhanh chóng và chính xác khối lượng sinh khối và xác định các khu vực loài phải bảo tồn chính là điều kiện tiên quyết để hỗ trợ giữ gìn và phục hồi tài nguyên rừng và bảo tồn đa dạng sinh học của các Khu dự trữ sinh quyển thế giới [27], [28]. Ngoài ra, rừng tự nhiên là hệ sinh thái cơ bản, giúp giảm phát thải Carbon trong khí quyển thông qua quá trình cô lập Carbon. Khu dự trữ sinh quyển miền Tây Nghệ An là một phần của điểm nóng đa dạng sinh học với tư cách là một hệ sinh thái rừng tự nhiên với mức độ đặc hữu và sự phong phú về loài đáng kể. Việc đánh giá sinh khối mặt đất trong hệ sinh thái rừng nhiệt đới gặp nhiều khó khăn do sự khác biệt trong địa bản và đối tượng nghiên cứu, cũng như hạn

chế trong thu thập dữ liệu. Điều này dẫn đến việc sử dụng các chỉ số và phương pháp đánh giá không thống nhất và chưa có một phương pháp ước tính chung. Do đó, nghiên cứu này sử dụng một phương pháp kết hợp dữ liệu sinh khối thực địa và đặc tính quang phổ và đặc điểm tán xạ ngược của ảnh SAR để khắc phục khó khăn trong việc đánh giá sinh khối mặt đất của hệ sinh thái rừng nhiệt đới. Dữ liệu thực địa từ các ô kiểm kê rừng của Viện Điều tra, Quy hoạch Rừng (FIPI) đã được sử dụng để tạo bộ tham số đầu vào cho mô hình Cubist. Bộ dữ liệu này đặc biệt có sự lâu dài và khả năng nhân rộng cho khu vực lớn (Hình 5). Do đó, dữ liệu AGB kiểm kê có độ tin cậy cao và việc đo lường tại chỗ đảm bảo sự chính xác trong quá trình mô hình hóa.



**Hình 5. Hình ảnh thực địa ô kiểm kê rừng tại Khu dự trữ sinh quyển miền Tây Nghệ An**

Lập bản đồ AGB là mối quan tâm lớn trên phạm vi toàn cầu và đối với nhiều nước đang phát triển vì đây là một nhiệm vụ đầy thách thức do thiếu dữ liệu điều tra thực địa. Đối với một hệ sinh thái rừng tự nhiên nhất định, những bản đồ này có thể được sử dụng để giám sát rừng, suy thoái rừng và các ngành liên quan đến rừng khác, chẳng hạn như bảo tồn, quản lý bền vững và tăng cường lưu trữ Carbon... Trong đó, sự tích lũy Carbon trong các hệ sinh thái rừng tự nhiên chịu ảnh hưởng của mật độ cây, loài cây và tuổi cây. Do đó, việc bảo vệ khu dự trữ Carbon rộng lớn trong các hệ sinh thái rừng tự nhiên tại Khu dự trữ sinh quyển miền Tây Nghệ An, Việt Nam nói riêng và hệ sinh thái rừng nhiệt đới khắp châu Á nói chung, là rất quan trọng để giảm thiểu việc phát thải khí Carbon vào khí quyển [28, 29]. Tích hợp dữ liệu viễn

thám đa nguồn (Sentinel-2A và Sentinel-1A SAR) có tính vượt trội trong việc ước tính AGB, bởi độ phân giải cao, dải quang phổ cạnh đỏ đặc trưng và thông tin quang phổ đầy đủ [29]. Những kết quả này đã chứng minh tiềm năng của các dải được xác định ưu việt trong việc tăng cường khả năng ước tính AGB của các cảm biến từ xa. Ngoài ra, một số nghiên cứu đã chứng minh rằng sự sẵn có dải quang phổ của Sentinel-2A và khả năng trong giám sát thảm thực vật thông qua các chỉ số thực vật [30], [31]. Tính sẵn có của dữ liệu Sentinel-2A và Sentinel-1A SAR rất quan trọng đối với ước tính AGB của thảm thực vật là các hệ sinh thái rừng, đặc biệt là tại các khu vực ở cấp độ quy mô lớn.

#### **4. KẾT LUẬN**

Khu dự trữ sinh quyển miền Tây Nghệ An là

nơi có diện tích rừng nguyên sinh rất lớn, vùng bảo tồn thiên nhiên còn giữ được vẻ hoang sơ, nguyên sinh, có thảm thực vật phong phú, đa dạng bậc nhất ở Việt Nam. Kết quả ước tính từ phương pháp được đề xuất cho thấy sự thay đổi không gian của sinh khối dao động từ 29,5 Mg/ha đến 1004,5 Mg/ha. Kết quả này phù hợp với các ô mẫu thực địa, với mối tương quan đáng kể giữa các giá trị ước tính và giá trị quan sát được. Vì AGB là một chỉ số ước tính quan trọng trong quản lý rừng bền vững, nên việc ước tính kịp thời chỉ số này là rất quan trọng để giám sát những thay đổi bề mặt hoặc khả năng mất mát và suy thoái của hệ sinh thái rừng tự nhiên trong khu vực. Từ đó, ta có cái nhìn rõ ràng hơn về tình trạng và xu hướng của hệ sinh thái, có thể đưa ra các biện pháp bảo vệ và phục hồi phù hợp.

Mô hình Cubist đã cải thiện đáng kể hiệu suất của hồi quy thông qua so sánh với các chức năng được định chuẩn bằng cách sử dụng các chỉ báo thống kê chung với các kết hợp tính năng khác nhau. Các giá trị tốt nhất được tìm thấy tại  $RMSE = 32,35$  Mg/ha,  $MAE = 28,21$  Mg/ha và  $R^2 = 0,81$ . Một khái niệm cơ bản khác là việc lựa chọn tính năng đóng một vai trò thiết yếu trong việc xác định các biến dự đoán quan trọng nhất trước khi chạy bất kỳ phương pháp hồi quy nào. Nghiên cứu này đã điều tra các ứng dụng tiềm năng của cả bộ dữ liệu viễn thám đa nguồn (Quang học và Radar), nhận thấy rằng sự kết hợp của cả hai loại dữ liệu là rất quan trọng trong việc loại bỏ hiệu ứng bão hòa và cải thiện độ chính xác của ước tính. Thông tin tán xạ ngược từ dữ liệu Radar và chỉ số thực vật được đánh giá để xác định cấu trúc những thay đổi trong trạng thái cây và AGB liên quan. Ngoài ra, dữ liệu viễn thám đa nguồn với độ phân giải không gian tốt hơn và thông tin quang phổ phong phú góp phần cải thiện độ chính xác của mô hình AGB và khắc phục các vấn đề về bão hòa.

Nghiên cứu này nghiên cứu phương pháp học máy và đã trở thành một chủ đề thịnh hành vì các ứng dụng rộng rãi của nó trong hầu hết các lĩnh vực nghiên cứu. Sự gia tăng năng lực tính toán và nhiều nền tảng cảm biến từ xa đã làm cho dữ liệu được định vị địa lý trở nên sẵn có để phân tích không gian. Để hiểu về dòng Carbon, học máy chủ yếu được áp dụng cho các phương pháp được sử dụng trong hồi quy hệ số phản xạ quang phổ và tán xạ ngược của hình ảnh vệ tinh đối với các phép đo AGB thực địa.

Nghiên cứu này nhằm mục đích nghiên cứu tích hợp dữ liệu viễn thám đa nguồn, từ dữ liệu địa hình và thuật toán học máy để hợp nhất dữ liệu và mô hình hóa không gian và thời gian để ước tính sinh khối ở Khu dự trữ sinh quyển miền Tây Nghệ An, Việt Nam. Những phát hiện có liên quan thực tế và phương pháp luận là hợp lý về mặt khoa học. Nghiên cứu này là một cách tiếp cận và đóng góp vào kiến thức toàn cầu trong lĩnh vực ước tính AGB của thảm thực vật. Do quá trình sinh trưởng của thảm thực vật thể hiện những đặc điểm vùng khác nhau cùng với sự thay đổi của điều kiện tự nhiên từng vùng nên mô hình ước tính sinh khối trên mặt đất của thảm thực vật được xây dựng trong nghiên cứu này có phù hợp với các vùng khác hay không cần được nghiên cứu và kiểm chứng thêm.

### Lời cảm ơn

Nghiên cứu này được hỗ trợ bởi Đề tài khoa học công nghệ tiềm năng cấp Bộ Nông nghiệp và Phát triển nông thôn: “*Nghiên cứu ứng dụng công nghệ viễn thám đa nguồn, miễn phí để thử nghiệm tính sinh khối rừng*”.

### TÀI LIỆU THAM KHẢO

- [1]. Trogisch S., Liu X., Rutten G., Xue K., Bauhus J., Brose U., Bu W., Cesarz S., Chesters D., Connolly J., Cui X., Eisenhauer N., Guo L., Haider S., Härdtle W., Kunz M., Liu L., Ma Z., Neumann S. & Bruehlheide H. (2021). The significance of tree-tree interactions for forest ecosystem functioning. *Basic and Applied Ecology*. 55: 33–52. <https://doi.org/10.1016/j.baae.2021.02.003>.
- [2]. Singh C., Karan S. K., Sardar P. & Samadder S. R. (2022). Remote sensing-based biomass estimation of dry deciduous tropical forest using machine learning and ensemble analysis. *Journal of Environmental Management*. 308: 114639. <https://doi.org/10.1016/j.jenvman.2022.114639>
- [3]. Li Z., Bi S., Hao S. & Cui Y. (2022). Aboveground biomass estimation in forests with random forest and Monte Carlo-based uncertainty analysis. *Ecological Indicators*. 142: 109246. <https://doi.org/10.1016/j.ecolind.2022.109246>.
- [4]. Brede B., Terryn L., Barbier N., Bartholomeus H. M., Bartolo R., Calders K., Derroire G., Krishna Moorthy S. M., Lau A., Levick S. R., Raunonen P., Verbeeck H., Wang D., Whiteside T., van der Zee J. & Herold, M. (2022). Non-destructive estimation of individual tree biomass: Allometric models, terrestrial and UAV laser scanning. *Remote Sensing of Environment*. 280: 113180. <https://doi.org/10.1016/j.rse.2022.113180>.
- [5]. Pham Van Manh, Pham Minh Tam, Du Vu Viet Quan, Bui Quang Thanh, Tran Van Anh, Pham Minh Hai, & Nguyen Ngoc Thach (2019). Integrating Sentinel-1A SAR data and GIS to estimate aboveground biomass and carbon accumulation for tropical forest types in Thuan Chau district, Vietnam. *Remote Sensing Applications: Society and Environment*. 14: 148–157. <https://doi.org/10.1016/j.rsase.2019.03.003>
- [6]. Shi J., Zhang W., Marino A., Zeng P., Ji Y., Zhao H., Huang G. & Wang M. (2023). Forest total and

- component biomass retrieval via GA-SVR algorithm and quad-polarimetric SAR data. *International Journal of Applied Earth Observation and Geoinformation*. 118: 103275. <https://doi.org/10.1016/j.jag.2023.103275>.
- [7]. Tian L., Wu X., Tao Y., Li M., Qian C., Liao L. & Fu W. (2023). Review of Remote Sensing-Based Methods for Forest Aboveground Biomass Estimation: Progress, Challenges, and Prospects. *Forests*. 14(6): 1086. <https://doi.org/10.3390/f14061086>.
- [8]. Pham Minh Hai, Do Thi Hoai, Pham Van Manh & Bui Quang Thanh (2020). Mangrove forest classification and aboveground biomass estimation using an atom search algorithm and adaptive neuro-fuzzy inference system. *PLOS ONE*. 15(5). Article 5. <https://doi.org/10.1371/journal.pone.0233110>.
- [9]. Ghosh S. M. & Behera M. D. (2018). Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest. *Applied Geography*. 96: 29–40. <https://doi.org/10.1016/j.apgeog.2018.05.011>.
- [10]. P. Zhao, P., Lu, D., Wang, G., Liu, L., Li, D., Zhu, J. & Yu S. (2016). Forest aboveground biomass estimation in Zhejiang Province using the integration of Landsat TM and ALOS PALSAR data. *International Journal of Applied Earth Observation and Geoinformation*. 53: 1–15. <https://doi.org/10.1016/j.jag.2016.08.007>.
- [11]. S. Talebiesfandarani, S., & Shamsoddini, A. (2022). Global-scale biomass estimation based on machine learning and deep learning methods. *Remote Sensing Applications: Society and Environment*. 28: 100868. <https://doi.org/10.1016/j.rsase.2022.100868>
- [12]. Ghosh S. M. & Behera M. D. (2018). Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest. *Applied Geography*. 96: 29–40. <https://doi.org/10.1016/j.apgeog.2018.05.011>.
- [13]. Li Y., Li M., Li C. & Liu Z. (2020). Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms. *Scientific Reports*. 10(1): 9952. <https://doi.org/10.1038/s41598-020-67024-3>.
- [15]. Rama Rao, N., Garg, P. K., Ghosh, S. K., & Dadhwal, V. K. (2008). Estimation of leaf total chlorophyll and nitrogen concentrations using hyperspectral satellite imagery. *The Journal of Agricultural Science*. 146(1): 65–75. <https://doi.org/10.1017/S0021859607007514>.
- [16]. Gitelson A. A. (2004). Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. *Journal of Plant Physiology*. 161(2): 165–173. <https://doi.org/10.1078/0176-1617-01176>.
- [17]. Blackburn G. A. (1998). Spectral indices for estimating photosynthetic pigment concentrations: A test using senescent tree leaves. *International Journal of Remote Sensing*. 19(4): 657–675. <https://doi.org/10.1080/014311698215919>.
- [18]. Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment*. 80(1): 76–87. [https://doi.org/10.1016/S0034-4257\(01\)00289-9](https://doi.org/10.1016/S0034-4257(01)00289-9).
- [19]. Broge N. H. & Leblanc E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment*. 76(2): 156–172. [https://doi.org/10.1016/S0034-4257\(00\)00197-8](https://doi.org/10.1016/S0034-4257(00)00197-8).
- [20]. Komiyama A., Ong J. E. & Pongparn S. (2008). Allometry, biomass, and productivity of mangrove forests: A review. *Aquatic Botany*. 89(2): 128–137. <https://doi.org/10.1016/j.aquabot.2007.12.006>.
- [21]. Buschmann C. & Nagel E. (1993). In vivo spectroscopy and internal optics of leaves as basis for remote sensing of vegetation. *International Journal of Remote Sensing*. 14(4): 711–722. <https://doi.org/10.1080/01431169308904370>.
- [22]. Roujean J.-L. & Breon F.-M. (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment*. 51(3): 375–384. [https://doi.org/10.1016/0034-4257\(94\)00114-3](https://doi.org/10.1016/0034-4257(94)00114-3).
- [23]. Quinlan J. R. (1992). Learning with continuous classes.
- [24]. Kuhn M., & Johnson K. (2013). Regression Trees and Rule-Based Models. In M. Kuhn & K. Johnson, *Applied Predictive Modeling* (pp. 173–220). Springer New York. [https://doi.org/10.1007/978-1-4614-6849-3\\_8](https://doi.org/10.1007/978-1-4614-6849-3_8).
- [25]. Chen S. S., Kimirei I. A., Yu C., Shen Q. & Gao Q. (2022). Assessment of urban river water pollution with urbanization in East Africa. *Environmental Science and Pollution Research*. <https://doi.org/10.1007/s11356-021-18082-1>.
- [26]. Do Thi Nhung, Nguyen Thi Diem My, Ghimire, J., Vu Kim Chi, Dang Do Lam Phuong, Pham Sy Liem & Pham Van Manh (2023). Assessing surface water pollution in Hanoi, Vietnam, using remote sensing and machine learning algorithms. *Environmental Science and Pollution Research*. <https://doi.org/10.1007/s11356-023-28127-2>.
- [27]. Nguyễn Quốc Hiệu, Phùng Văn Khoa, Nguyễn Hải Hòa, Tôn Thất Minh, Trần Cường & Trần Văn Nam (2018). Sử dụng chỉ số viễn thám để phát hiện suy thoái rừng tại khu dự trữ sinh quyển thế giới Langbiang, Lâm Đồng. *Tạp chí Khoa học Lâm nghiệp*. (3): 67-82.
- [28]. Ha Tri Son, Pham Duy Quang, Nguyen Hai Hoa, & Vu Van Truong (2022). Using sentinel-2 data for above-ground mangrove biomass and carbon stocks mapping over 7 years (2016–2022) in Tien Yen, Quang Ninh Province. *Journal of Forestry Science and Technology*. 14: 153–165. <https://doi.org/10.55250/jo.vnuf.2022.14.153-165>.
- [29]. Haas J. & Ban Y. (2017). Sentinel-1A SAR and sentinel-2A MSI data fusion for urban ecosystem service mapping. *Remote Sensing Applications: Society and Environment*. 8: 41–53. <https://doi.org/10.1016/j.rsase.2017.07.006>.
- [30]. Shoko C. & Mutanga O. (2017). Examining the strength of the newly-launched Sentinel 2 MSI sensor in detecting and discriminating subtle differences between C3 and C4 grass species. *ISPRS Journal of Photogrammetry and Remote Sensing*. 129: 32–40. <https://doi.org/10.1016/j.isprsjprs.2017.04.016>.
- [31]. Ma T., Hu Y., Wang J., Beckline M., Pang D., Chen L., Ni X. & Li X. (2023). A Novel Vegetation Index Approach Using Sentinel-2 Data and Random Forest Algorithm for Estimating Forest Stock Volume in the Helan Mountains, Ningxia, China. *Remote Sensing*, 15(7), 1853. <https://doi.org/10.3390/rs15071853>.