VNU Journal of Science: Earth and Environmental Sciences

Journal homepage: https://js.vnu.edu.vn/EES

Original Article

Applied Machine Learning Algorithms and Landsat 8 for Estimating Aboveground Carbon Stock in Evergreen Broadleaf Forest in Binh Phuoc Province

Nguyen Thanh Tuan[1,*], Nguyen Van Phu[1], Nguyen Van Quy[1], Hoang Thi Phuong Nhung[2]

[1]*Vietnam National University of Forestry at Dongnai, Tran Phu, Trang Bom, Dong Nai, Vietnam*

[2]*Binh Phuoc Sub-Department of Forest Protection, 673 Phu Rieng Do, Dong Xoai, Binh Phuoc, Vietnam*

Received 16 July 2022
Revised 17 August 2022; Accepted 30 August 2022

Abstract: The assessment of carbon stocks is one of the key measurements to support climate change mitigation policies. The research applied Landsat 8 satellite imagery combined with field-measurements using four machine learning methods (random forest - RF, artificial neural networks - NNET, support vector machines – SVM, and linear regression - LM) to estimate aboveground carbon in evergreen broadleaf forest in Binh Phuoc province. The field sample plots were randomly divided into training (96 plots) and testing (24 plots) data. The results showed that RF yielded the greatest precision with an R^2 value above 0,9 and RMSE below 6 ton/ha on the training data, with an R^2 value of 0,41 and RMSE of 11,04 ton/ha on the testing data. The estimate of forest carbon stock increased distinctly from the mean value of 59,80 ton/ha in the very poor forest to 87,78 ton/ha in the rich forest. The results found in the present study demonstrated that Landsat 8 imagery in conjunction with RF has the appropriate to estimate aboveground carbon stock in evergreen broadleaf forest-leaved in Binh Phuoc province.

Keywords: Random forest, aboveground carbon, REDD+, forest carbon estimation.

* Corresponding author.
E-mail address: nttuan@vnuf2.edu.vn

https://doi.org/10.25073/2588-1094/vnuees.4890

Ứng dụng một số thuật toán máy học và ảnh Landsat 8 ước tính trữ lượng carbon trên mặt đất rừng tự nhiên lá rộng thường xanh tại tỉnh Bình Phước

Nguyễn Thanh Tuấn[1,*], Nguyễn Văn Phú[1],
Nguyễn Văn Quý[1], Hoàng Thị Phương Nhung[2]

[1]*Trường Đại học Lâm nghiệp Phân hiệu tại tỉnh Đồng Nai, Trần Phú, Trảng Bom, Đồng Nai, Việt Nam*
[2]*Chi cục Kiểm lâm tỉnh Bình Phước, 673 Phú Riềng Đỏ, Đồng Xoài, Bình Phước, Việt Nam*

Nhận ngày 16 tháng 7 năm 2022
Chỉnh sửa ngày 17 tháng 8 năm 2022; Chấp nhận đăng ngày 30 tháng 8 năm 2022

Tóm tắt: Đánh giá trữ lượng carbon là một trong những phép đo quan trọng để hỗ trợ các chính sách giảm thiểu biến đổi khí hậu. Nghiên cứu sử dụng ảnh Landsat 8 và dữ liệu điều tra thực địa kết hợp với 4 thuật toán học máy (rừng ngẫu nhiên, mạng nơron nhân tạo, véctơ hỗ trợ và hồi quy tuyến tính) để ước tính trữ lượng carbon trên mặt đất của rừng tự nhiên lá rộng thường xanh tại tỉnh Bình Phước. Các ô mẫu điều tra thực địa được chia ngẫu nhiên thành 2 phần dữ liệu huấn luyện (96 ô tiêu chuẩn) và kiểm định (24 ô tiêu chuẩn). Kết quả chỉ ra rằng mô hình rừng ngẫu nhiên mang lại độ chính xác cao nhất với giá trị R^2 trên 0,9 và sai số RMSE dưới 6 tấn/ha ở phần dữ liệu huấn luyện, R^2 bằng 0,41 và RMSE bằng 11,04 tấn/ha ở phần dữ liệu kiểm định. Trữ lượng carbon trung bình từ mô hình dự đoán tăng rõ rệt từ rừng nghèo kiệt (59,8 tấn/ha) đến rừng giàu (87,78 tấn/ha). Kết quả tìm thấy trong nghiên cứu này đã chứng tỏ rằng ảnh vệ tinh Landsat 8 kết hợp với thuật toán học máy rừng ngẫu nhiên phù hợp để ước tính trữ lượng carbon trên mặt đất các trạng thái rừng tự nhiên rừng lá rộng thường xanh tại tỉnh Bình Phước.

Từ khóa: Rừng ngẫu nhiên, Carbon trên mặt đất, REDD+, dự đoán carbon rừng.

1. Mở đầu

Mất rừng và suy thoái rừng nhiệt đới do mở rộng đất nông nghiệp, chuyển đổi sang đất đồng cỏ, khai thác gỗ, cháy rừng và các nguyên nhân khác chiếm 11% lượng phát thải khí nhà kính trên thế giới [1]. Sau nhiều năm đàm phán, công ước khung của Liên hợp quốc về biến đổi khí hậu (UNFCCC) từng bước công nhận cơ chế giảm nhẹ phát thải từ mất rừng và suy thoái rừng, thông qua bảo tồn, quản lý bền vững và tăng cường trữ lượng carbon rừng ở vùng nhiệt đới và cận nhiệt đới (REDD+). Trong đó, hệ thống giám sát rừng quốc gia được sử dụng để đánh giá mức độ vận hành của các hoạt động REDD+. Hệ thống giám sát rừng quốc gia cung cấp thông tin về trữ lượng carbon rừng và thông tin này được sử dụng để chỉ ra các yếu tố phát thải (EF) [2]. Hiện nay ở nước ta, điều tra thực địa kết hợp với công nghệ viễn thám được sử dụng phổ biến và có độ chính xác cao trong việc xác định trữ lượng carbon cho các trạng thái rừng lá rộng thường xanh ở huyện Kim Bôi, tỉnh Hòa Bình [3], rừng rụng lá ở Tây Nguyên [4], rừng ngập mặn ở

* Tác giả liên hệ.
Địa chỉ email: nttuan@vnuf2.edu.vn

https://doi.org/10.25073/2588-1094/vnuees.4890

huyện Cần Giờ, Thành Phố Hồ Chí Minh và tỉnh Quảng Ninh [5, 6]. Tuy nhiên, việc đo lường carbon rừng cho từng đối tượng rừng cụ thể gặp rất nhiều khó khăn và thách thức do sự biến động lớn giữa các kiểu rừng, trạng thái rừng, loài cây ưu thế và tuổi lâm phần [3, 7]. Do vậy, cần phải có những nghiên cứu về khả năng tích lũy carbon của từng đối tượng rừng cho mỗi địa phương cụ thể để làm cơ sở cho việc hỗ trợ tài chính của chương trình REDD+.

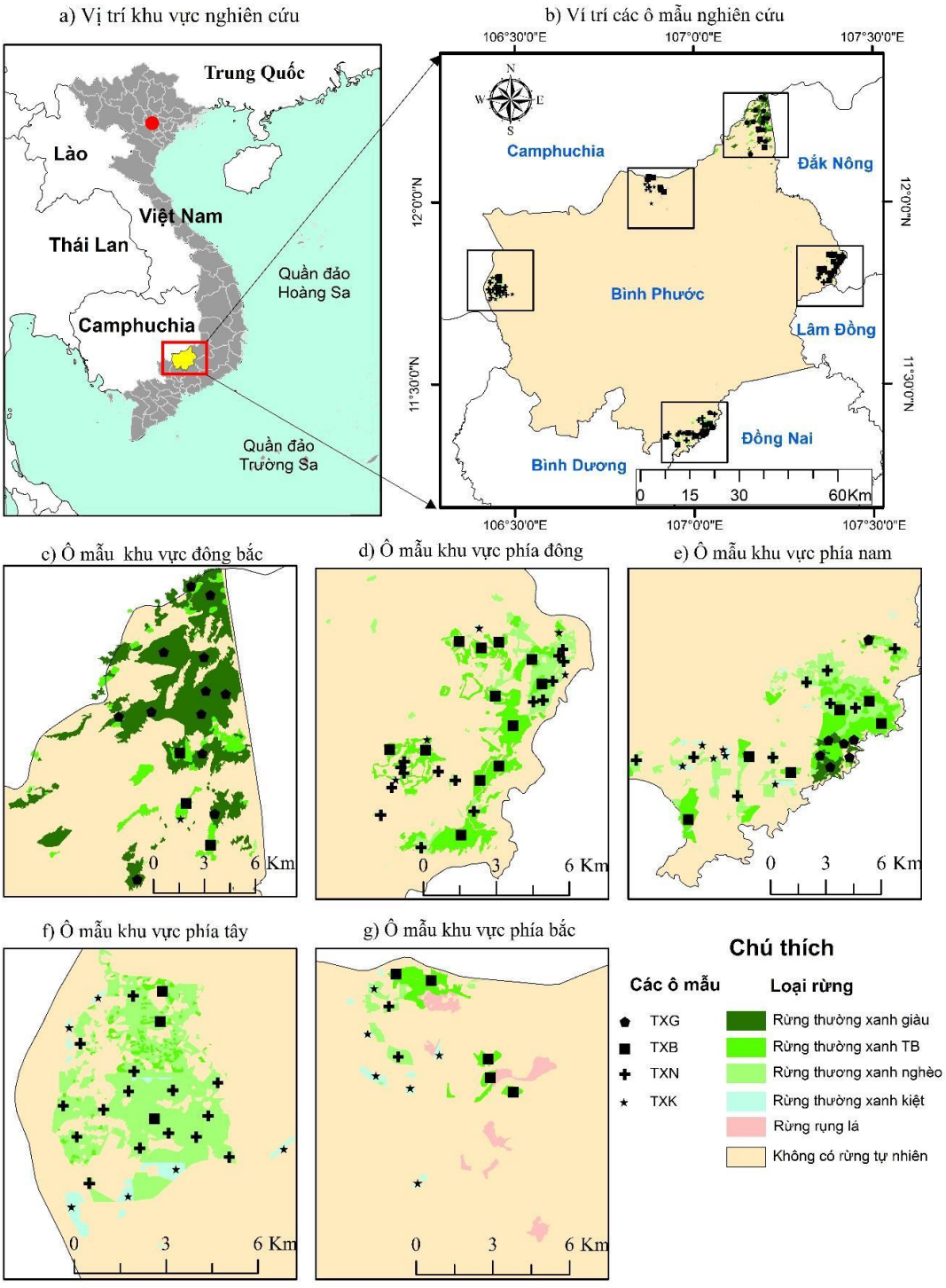
Landsat 8 đã được Mỹ phóng thành công lên quỹ đạo vào ngày 11/02/2013, cung cấp các ảnh có độ phân giải trung bình (từ 15-100 m), phủ kín ở các vùng cực cũng như những vùng địa hình khác nhau trên trái đất. Ảnh vệ tinh Landsat 8 đã được sử dụng rộng rãi trong việc xác định trữ lượng carbon rừng tự nhiên [3, 8, 9]. Một trong những vấn đề trọng tâm của ứng dụng viễn thám trong xác định trữ lượng carbon là xác định các thuật toán phù hợp để xây dựng các mô hình ước lượng [10]. Hiện tại có rất nhiều thuật toán đã được phát triển cho việc xác định trữ lượng carbon rừng từ ảnh viễn thám, bao gồm 2 loại chính: tham số (hồi quy đơn biến, đa biến và phi tuyến) và phi tham số (thuật toán láng giềng gần nhất, mạng lưới nơron, thuật toán rừng ngẫu nhiên, thuật toán véc tơ hỗ trợ). Thuật toán tham số thể hiện mối quan hệ giữa biến phụ thuộc (sinh khối) và biến độc lập (giá trị phản xạ phổ của các kênh ảnh, chỉ số thực vật,…) bằng các mô hình hồi quy tuyến tính đơn biến, đa biến hoặc hàm phi tuyến. Ngược lại, các thuật toán phi tham số không ấn định trước một cách rõ ràng cấu trúc mô hình vì vậy nó có tính linh hoạt hơn. Do tính linh hoạt của các thuật toán phi tham số nên chúng được đánh giá là chuyên nghiệp hơn trong việc tạo ra các mô hình ước tính sinh khối cho các đối tượng rừng có cấu trúc phức tạp [11]. Mặt khác, khả năng tổng quát hóa và chống nhiễu của các phương pháp học máy phi tham số đã được chứng minh trong trường hợp dữ liệu mẫu ít [12].

Theo số liệu kiểm kê rừng năm 2020, tổng diện tích rừng tự nhiên của tỉnh Bình Phước là 56.148,55 ha, trong đó trạng thái rừng tự nhiên lá rộng thường xanh chiếm 14.269,61 ha [13]. Do vậy, xác định khả năng hấp thụ carbon của rừng lá rộng thường xanh tại tỉnh Bình Phước là việc làm có ý nghĩa, nó cung cấp thông tin về trữ lượng carbon cho hệ thống giám sát rừng quốc gia của REDD+. Từ những lý do trên, nghiên cứu được tiến hành với mục tiêu bao gồm: i) Đánh giá độ chính xác của một số thuật toán học máy để ước tính khả năng tích lũy carbon trên mặt đất của rừng; và ii) Xây dựng bản đồ tích lũy carbon các trạng thái rừng tự nhiên lá rộng thường xanh tại tỉnh Bình Phước.

2. Phương pháp nghiên cứu

2.1. Phương pháp điều tra ngoại nghiệp

Căn cứ vào bản đồ hiện trạng rừng của tỉnh Bình Phước năm 2020, nghiên cứu tiến hành xây dựng các tuyến điều tra và xác định tọa độ theo các tuyến. Sau đó tiến hành sơ thám và tham vấn ý kiến của cán bộ quản lý rừng địa bàn nhằm điều chỉnh tuyến cho phù hợp với điều kiện thực tế đi qua các trạng thái rừng lá rộng thường xanh tại khu vực nghiên cứu. Tuyến điều tra được xác định đại diện và đảm bảo bao quát toàn bộ các dạng địa hình, các trạng thái rừng lá rộng thường xanh. Khoảng cách giữa các tuyến điều tra cách nhau ít nhất 1000 m. Trên tuyến bố trí ô mẫu đại diện cho các trạng thái rừng lá rộng thường xanh tại khu vực nghiên cứu. Các trạng thái rừng tự nhiên lá rộng thường xanh được phân chia theo thông tư 33 của Bộ Nông nghiệp và Phát triển Nông thôn quy định về điều tra, kiểm kê và theo dõi diễn biến tài nguyên rừng. Cụ thể, trạng thái rừng thường xanh giàu có trữ lượng lớn hơn 200 m^3/ha, trung bình từ 100 đến 200 m^3/ha, nghèo từ 50 đến 100 m^3/ha và kiệt từ 10 đến 50 m^3/ha. Dựa theo khảo sát sơ bộ hiện trạng trên tuyến, nghiên cứu tiến hành thiết lập các ô tiêu chuẩn có diện tích 500 m^2 (25 x 20 m). Thời gian tiến hành điều tra trên các ô tiêu chuẩn từ tháng 1 đến tháng 8 năm 2020. Vị trí mỗi ô mẫu được xác định bằng máy định vị GPS. Tổng cộng 120 ô tiêu chuẩn đã được lập, trong đó rừng lá rộng thường xanh giàu 19 ô mẫu, trung bình 30 ô mẫu, nghèo 47 ô mẫu và kiệt là 24 ô mẫu (Hình 1).



Hình 1. Vị trí các ô mẫu tại khu vực nghiên cứu
(TXG, TXB, TXN và TXK: ô mẫu thuộc rừng lá rộng thường xanh giàu, trung bình, nghèo và kiệt).

Trong mỗi ô mẫu, tiến hành đo đếm và thu thập các thông tin về thành phần loài, số lượng cá thể của mỗi loài; các chỉ tiêu sinh trưởng như đường kính tại vị trí 1,3 m (*D*) bằng thước dây, chiều cao vút ngọn (*H*) được xác định bằng thước đo cao Blume-leiss. Tên loài cây gỗ được xác định bằng phương pháp so sánh hình thái của Ho (2003) [14] dựa trên các tài liệu bao gồm: cây cỏ Việt Nam [14], cây gỗ Việt Nam [15] và tên khoa học được hiệu chỉnh bởi Kew Science (http://www.plantsoftheworldonline.org), World flora online (http://104.198.148.243). Trong ô tiêu chuẩn chỉ thu thập thông tin của những cây rừng có D>5 cm.

2.2. Phương pháp ước tính sinh khối và khả năng tích lũy carbon trên mặt đất

Nghiên cứu dựa vào phương trình sinh khối các trạng thái rừng tự nhiên tại khu vực Nam bộ và Tây Nguyên của tác giả Bảo Huy (2016) [16] để xác định sinh khối khô trên mặt đất cho các ô mẫu. Cụ thể, phương trình tương quan sinh khối trên mặt đất:

$$AGB = 0{,}647261 \times (\rho \times D^2 \times H)^{0{,}931} \quad (1)$$

Trong đó: *D* là đường kính ngang ngực (cm), *H* là chiều cao và ρ khối lượng riêng của từng loài cây (g cm^{-3}). Khối lượng riêng của từng loài cây dựa vào dữ liệu của Zanne và cộng sự (2009) [17].

Cuối cùng, nghiên cứu sử dụng hệ số chuyển đổi carbon từ sinh khối khô theo IPCC (2006) [18] áp dụng chung cho các trạng thái rừng tự nhiên ở Việt Nam là 0,47.

2.3. Tiền xử lý dữ liệu ảnh viễn thám

Tư liệu viễn thám được lựa chọn để ước tính lượng carbon tích lũy ở các trạng thái rừng là ảnh Landsat 8 với các thông số kỹ thuật ở Bảng 1.

Bảng 1. Các thông số kỹ thuật của ảnh Landsat 8 sử dụng trong nghiên cứu

Ký hiệu ảnh	Thời gian chụp	Độ che phủ của mây (%)	Kênh ảnh	Độ phân giải (m)
LC08_L2SP_124052_ 20200115_20200824_02_T1	2020-01-15	0,2	Kênh 1 (Aerosol ven biển - Coastal aerosol)	30
			Kênh 2 (xanh - Blue)	30
			Kênh 3 (xanh - Green)	30
			Kênh 4 (đỏ -Red)	30
			Kênh 5 (cận hồng ngoại -NIR)	30
			Kênh 6 (hồng ngoại sóng ngắn thứ nhất - SWIR1)	30
LC08_L2SP_125052_ 20200106_20200823_02_T1	2020-01-06	0,65	Kênh 7 (hồng ngoại sóng ngắn thứ hai - SWIR2)	30
			Kênh 8 (toàn sắc- Panchromatic)	15
			Kênh 9 (kênh Cirrus)	30
			Kênh 10 (kênh hồng ngoại nhiệt 1- TIRS 1)	100
			Kênh 11 (kênh hồng ngoại nhiệt 2- TIRS 2)	100

Nghiên cứu sử dụng ảnh Landsat 8 đã được xử lý ở mức 2 (L2) được tải từ trang web: https://earthexplorer.usgs.gov. Ảnh Landsat 8 mức 2 cung cấp giá trị phản xạ trên mặt đất tại từng pixel (BOA). Dữ liệu phản xạ này của ảnh Landsat 8 đã được USGS hiệu chỉnh bằng phương pháp LaSRC nhằm loại bỏ ảnh hưởng của khí quyển như tán xạ, hấp thụ khí quyển, từ đó được dùng để ước tính giá trị phản xạ bề mặt. Mặt khác, ảnh Landsat 8 đã được xử lý ở mức

trực ảnh nghĩa là đã hiệu chỉnh biến dạng bởi chênh cao địa hình và được đăng ký trong hệ tọa độ WGS-84. Khi sử dụng để chỉnh lý, hiệu chỉnh bản đồ địa hình sẽ không cần phải nắn ảnh mà chỉ cần tính chuyển đổi về hệ tọa độ VN-2000. Sau khi đưa ảnh về tọa độ VN-2000, tiến hành chuyển đổi giá trị cấp độ xám (DN) của ảnh Landsat 8 thành giá trị phản xạ phổ của vật thể bằng công thức:

$$\rho\lambda=(M_\rho Q_{cal}+A_\rho)/\cos(\theta_{sz}) \quad (2)$$

$\rho\lambda$: phản xạ ở tầng trên của khí quyển

Q_{cal}: giá trị số trên ảnh (DN)

M_ρ: giá trị REFLECTANCE_MULT_BAND_x

A_ρ: giá trị REFLECTANCE_ADD_BAND_x

θ_{sz}: góc thiên đỉnh (góc cao) của mặt trời (độ)

Trong nghiên cứu này, hiệu chỉnh ảnh hưởng khí quyển cho ảnh Landsat 8 được thực hiện trên phần mềm Qgis 3.18. Ngoài ra các kênh ảnh có độ phân giải không gian khác nhau từ 15-100 m, vì vậy nghiên cứu tiến hành chuyển độ phân giải không gian của các kênh ảnh về cùng kích thước 15 m trước khi tiến hành trích xuất các giá trị quang phổ. Các kênh ảnh Landsat 8 được sử dụng để ước tính trữ lượng carbon được thể hiện ở Bảng 1.

Cuối cùng, ảnh khu vực nghiên cứu được cắt dựa vào ranh giới các trạng thái rừng lá rộng thường xanh trích xuất từ bản đồ hiện trạng rừng năm 2020 của tỉnh Bình Phước.

2.4. Tính toán các chỉ số quang phổ từ ảnh viễn thám

Ngoài giá trị phản xạ phổ ở các kênh ảnh, nghiên cứu còn sử dụng 7 chỉ số quang phổ để xây dựng mô hình ước tính trữ lượng carbon của các trạng thái rừng nghiên cứu (Bảng 2).

Bảng 2. Các chỉ số quang phổ được tính từ các kênh ảnh Landsat 8

Chỉ số	Công thức tính	Tham khảo
Chỉ số thực vật	NDVI = (NIR − red)/(NIR + red)	J. W. Rouse (1974) [19]
Chỉ số thực vật tăng cường	EVI = (NIR − red)/(NIR + 6 (red) – 7,5 (blue) + 1)	Huete (1997) [20]
Chỉ số thực vật có hiệu chỉnh phản xạ mặt đất	SAVI = ((NIR − red)/(NIR + red + 0,5)) (1,5)	Huete (1988) [21]
SAVI hiệu chỉnh	MSAVI = (2 (NIR) + 1 − sqrt ((2 (NIR) + 1)2 − 8 (NIR − red)))/2	Rondeaux (1996) [22]
Chỉ số độ ẩm khác biệt	NDMI = (NIR − SWIR1)/(NIR + SWIR1)	Gao (1996) [23]
Tỉ số cháy	NBR = (NIR − SWIR2)/(NIR + SWIR2)	Miller (2007) [24]
Tỉ số cháy hiệu chỉnh	NBR2 = (SWIR1 − SWIR2)/(SWIR1 + SWIR2)	Miller (2007) [24]

Trong đó, đối với ảnh Landsat 8: blue (Kênh 2), red (Kênh 4), NIR (Kênh 5), SWIR1 (kênh 6), SWIR2 (Kênh 7).

2.5. Lựa chọn các biến dự đoán và thuật toán học máy để ước tính trữ lượng carbon

Nhiều biến xác định từ ảnh vệ tinh đã được sử dụng trong mô hình ước tính sinh khối. Tuy nhiên, không phải tất cả các biến đều hữu ích trong việc xây dựng mô hình dự đoán. Do vậy, chúng ta cần tiến hành loại bỏ những biến có tự tương quan cao ($|R|>0,8$) để giảm tính đa cộng tuyến giữa các biến trước khi xây dựng mô hình ước tính tích lũy carbon [25]. Mặt khác, các biến được chuẩn hóa (Căn bậc 2, mũ, Logarit,...) về phân bố chuẩn trước khi phân tích ở các bước tiếp theo. Tương quan giữa các biến được thể hiện ở phụ lục 01. Cuối cùng, 5 biến số (giá trị các kênh ảnh 1,5, 6, 10 và chỉ số NBR2) được dữ lại để xây dựng mô hình ước tính trữ lượng carbon rừng.

Sau khi lựa chọn các biến dùng để dự đoán, nghiên cứu sử dụng phương pháp tham số (hồi quy tuyến tính đa biến - LME) và phương pháp phi tham số (Rừng ngẫu nhiên - RF, Mạng nhân tạo - NNET và véctơ hỗ trợ - SVM) trong thuật toán học máy để ước tính trữ lượng carbon của các trạng thái rừng.

Mô hình hồi quy đa biến được lựa chọn dựa vào chỉ số lý thuyết thông tin AIC. Trong đó, các biến độc lập được bổ sung dần dần (từng biến

một) vào phương trình hồi quy. Mô hình hồi quy tuyến tính đa biến số được biểu diễn qua phương trình (3):

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + + \beta_k x_{ki} + \varepsilon_i \quad (3)$$

Trong đó y_i là trữ lượng carbon rừng (tấn/ha), x_i là các biến số dùng để dự đoán. Chúng ta có nhiều biến x (x_1, x_2,… đến x_k), và mỗi biến có một thông số β_j (j = 1, 2,…, k) cần phải ước tính. Phương pháp ước tính β_j cũng chủ yếu dựa vào phương pháp bình phương nhỏ nhất.

Phương pháp máy vec-tơ hỗ trợ (SVM) là thuật toán thống kê dựa trên phương pháp hạt nhân (kernel) để chuyển hồi quy phi tuyến sang tuyến tính trong không gian nhiều chiều, có thể sử dụng cho thuật toán phân lớp đối tượng hoặc hồi quy [26]. Chẳng hạn mẫu huấn luyện ban đầu: (x_i, y_i), (i=1,2…n), trong đó x_i là đa biến số đầu vào, y_i là đầu ra vô hướng và n là số mẫu huấn luyện. Với phương pháp SVM mô hình ban đầu sẽ được chuyển thành mô hình tuyến tính trong không gian nhiều chiều mới [27]:

$$y = f(x) = \langle w.\varphi(x) \rangle + b = \sum_{i=1}^{n} w_i \varphi_i(x) + b \quad (4)$$

Trong đó w là trọng số của vector, b là độ dịch. φ là hàm phi tuyến chuyển đổi từ không gian đầu vào thành không gian nhiều chiều mới. Thay vì xác định chính xác dạng hàm của φ, chúng ta sử dụng hàm thức hạt nhân như sau:

$$K(x_i, x) = \langle \varphi(x_i).\varphi(x) \rangle \quad (5)$$

Thông thường hàm thức hạt nhân bao gồm hàm tuyến tính, đa thức bậc cao và hàm cơ sở bán kính. Mặt khác, chúng ta cần xác định w và b dựa vào sai số nhỏ nhất của hồi quy dựa vào công thức:

$$minimize \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{n} \xi_i + \xi_i^* \quad (6)$$

$$subject\ to\{y_i - f(x_i) \le \varepsilon + \xi_i^*\ f(x_i) - y_i \le \varepsilon + \xi_i\ \xi_i, \xi_i^* \ge 0, i = 1, \dots, n$$

Để lựa chọn hàm thức hạt nhân và giá trị C tối ưu cho mô hình SVM, nghiên cứu tiến hành chạy thử nghiệm giá trị C và các hàm thức hạt nhân khác nhau để chọn ra giá trị phù hợp với sai số của mô hình là nhỏ nhất.

Mạng nơron nhân tạo (NNET) là mô hình trong đó giá trị đầu ra được kết nối với các biến đầu vào thông qua mạng lưới (mẫu – kết quả) được huấn luyện từ trước. Cấu trúc chung của một mạng nơron nhân tạo gồm 3 thành phần đó là: lớp dữ liệu đầu vào, lớp ẩn và lớp đầu ra. Trong nghiên cứu này lớp dữ liệu đầu vào là các giá trị phổ của ảnh Landsat 8, lớp ẩn và lớp đầu ra là biến cần dự đoán trữ lượng carbon. Trong đó, lớp ẩn gồm các nơron nhận dữ liệu đầu vào từ các nơron ở lớp trước đó và chuyển đổi các dữ liệu đầu vào này cho các lớp xử lý tiếp theo [28]. Một điểm quan trọng trong việc thiết lập một mạng NNET là làm sao cho kết quả của mô hình đạt độ chính xác cao và cấu trúc của mạng là thích hợp nhất. Nếu có quá nhiều lớp và nơ-ron được sử dụng thì mạng sẽ có rất nhiều hệ số tự do, điều này có thể dẫn đến việc làm nhiễu thông tin đầu vào. Ngược lại, nếu có quá ít lớp và nơ-ron được sử dụng trong mạng thì có thể mạng sẽ không đủ khả năng biểu diễn hết tương quan giữa các biến đầu vào và đầu ra trong mạng. Vì vậy, một mạng NNET tối ưu cần phải xác định có số lượng lớp, số lượng nơ-ron trong từng lớp và hàm chuyển hoá thông tin, kỹ thuật "luyện mạng" thích hợp. Số lượng Nơron và lớp ẩn được lựa chọn sao cho sai số của mô hình là nhỏ nhất.

Mô hình rừng ngẫu nhiên (RF) là một số lượng lớn các cây hồi quy được xác định hoàn toàn ngẫu nhiên từ các biến đầu vào (có thể là biến liên tục hoặc rời rạc) để xác định giá trị đầu ra. Các giá trị đầu ra sau đó được xác định bằng trung bình cộng kết quả đầu ra từ tất cả các cây hồi quy. Hai tham số cần được xác định trong thuật toán phân loại này là ntree (số lượng cây được phát triển) và mtry (số lượng biến để phân chia tại mỗi nút phân nhánh). Số ntree được lựa chọn phụ thuộc vào khoảng thời gian xử lý ngắn nhất để kết quả đạt được độ sai số thấp nhất, ntree chạy từ 1 đến 1000 cây và mtry biến động từ số biến độc lập tối thiểu bằng 1 đến số biến độc lập tối đa được sử dụng trong phân loại.

2.6. Đánh giá độ chính xác các mô hình dự báo

Nghiên cứu sử dụng dữ liệu của 24 ô tiêu chuẩn thuộc khu vực ranh giới phía đông giáp với tỉnh Lâm Đồng làm dữ liệu kiểm định, dữ liệu này không tham gia xây dựng mô hình. Còn lại 96 ô tiêu chuẩn dùng để huấn luyện mô hình. Ngoài ra, nghiên cứu sử dụng phương pháp đánh giá chéo so sánh độ chính xác của các mô hình

huấn luyện. Đối với phương pháp này, dữ được chia ngẫu nhiên thành K phần (K=10). Sau đó huấn luyện mô hình K lần, mỗi lần huấn luyện sẽ chọn 1 phần làm dữ liệu đánh giá và K-1 phần còn lại làm dữ liệu huấn luyện. Kết quả đánh giá mô hình cuối cùng sẽ là trung bình cộng kết quả đánh giá của K lần huấn luyện. Giá trị của K được gán cố định bằng 10, một giá trị thường được sử dụng và đã được chứng minh là cho sai số nhỏ, phương sai thấp. Kết quả đánh giá chéo dựa vào ba tiêu chí là hệ số tương quan R^2, căn của sai số toàn phương trung bình (RMSE) và sai số trung bình tuyệt đối (MAE). Giá trị của R^2 được sử dụng để kiểm định mối tương quan tuyến tính giữa giá trị thực tế và giá trị dự đoán. Trong khi đó, RMSE và MAE là hai tiêu chí được sử dụng để đánh giá sự khác biệt giữa giá trị thực tế và giá trị tính toán từ mô hình. Về mặt định lượng, giá trị của R^2 càng gần 1, giá trị của RMSE và MAE càng gần 0 thì sẽ thể hiện độ chính xác cao hơn của mô hình học máy.

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(y_i - y_{pred}\right)^2} \quad (7)$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left(y_i - y_{pred}\right) \quad (8)$$

$$R^2 = \frac{\sum_{i=1}^{m}\left(y_{pred} - \bar{y}\right)^2}{\sum_{i=1}^{m}(y_i - \bar{y})^2} \quad (9)$$

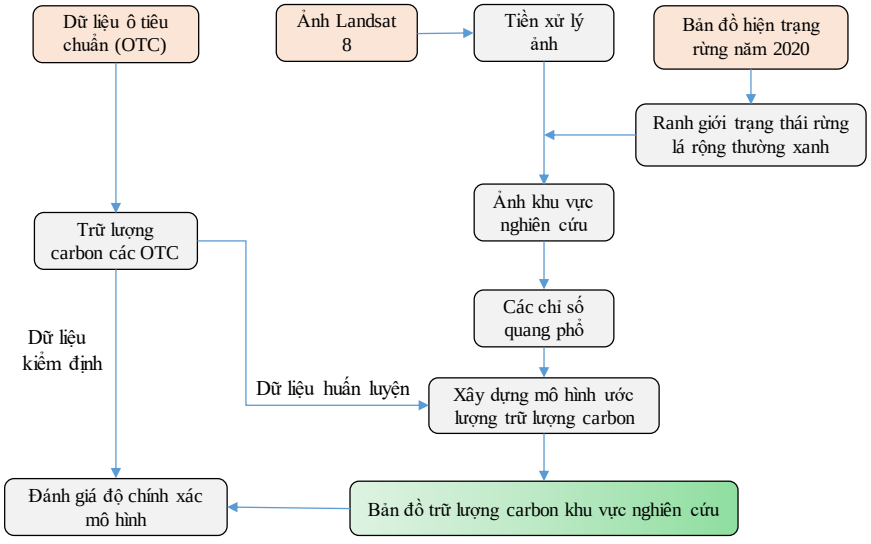
Trong đó: m là giá trị của mẫu kiểm định, y_i giá trị đo thực tế, y_{pred} giá trị được ước lượng từ mô hình, $\bar{y}$ giá trị đo thực tế trung bình.

Cuối cùng, nghiên cứu sử dụng tiêu chuẩn t so sánh giá trị R^2, RSME và MAE giữa 4 mô hình toán máy học để lựa chọn mô hình tối ưu nhất với R^2 lớn nhất, RSME và MAE nhỏ nhất với mức ý nghĩa thống kê p<0,05.

Toàn bộ các thuật toán học máy được huấn luyện và đánh giá độ chính xác bằng gói "caret" chạy trên R 4.0.2.

2.7. Xây dựng bản đồ tích lũy carbon rừng lá rộng thường xanh tại khu vực nghiên cứu

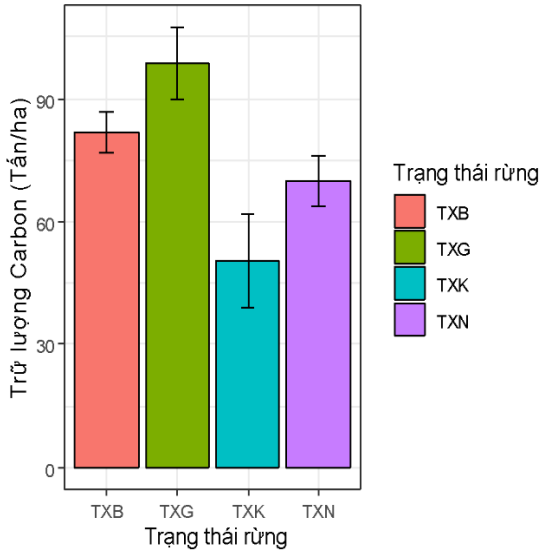
Sử dụng thuật toán học máy tối ưu đã được lựa chọn để tiến hành ước lượng trữ lượng carbon cho các trạng thái rừng lá rộng thường xanh của tỉnh Bình Phước dựa vào các biến đầu vào là giá trị quang phổ của ảnh Landsat 8. Toàn bộ các bước xây dựng bản đồ tích lũy carbon rừng được thể hiện ở Hình 2.



Hình 2. Sơ đồ mô tả phương pháp xây dựng bản đồ trữ lượng carbon các trạng thái rừng lá rộng thường xanh tại tỉnh Bình Phước năm 2020 từ ảnh Landsat 8 và số liệu điều tra thực địa.

3. Kết quả nghiên cứu và thảo luận

3.1. Trữ lượng carbon các trạng thái rừng lá rộng thường xanh

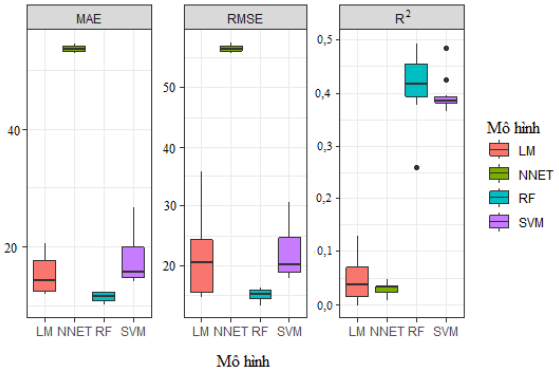


Hình 3: Trữ lượng carbon của các trạng thái rừng trong các ô mẫu nghiên cứu (TXG, TXB, TXN và TXK: ô mẫu thuộc rừng lá rộng thường xanh giàu, trung bình, nghèo và nghèo kiệt).

Khả năng tích trữ carbon trên mặt đất của trạng thái rừng giàu là 98,57 (tấn/ha), rừng trung bình là 81,88 (tấn/ha), rừng nghèo là 69,88 (tấn/ha) và rừng nghèo kiệt là 50,37 (tấn/ha). Mức biến động carbon tích trữ trên mặt đất của trạng thái rừng giàu từ 87,80 (tấn/ha) đến 118,96 (tấn/ha), rừng trung bình từ 68,49 (tấn/ha) đến 93,00 (tấn/ha), rừng nghèo từ 53,34 (tấn/ha) đến 86,31 (tấn/ha) và rừng nghèo kiệt từ 13,58 (tấn/ha) đến 68,03 (tấn/ha) (Hình 3). Các trạng thái rừng thứ sinh ở hiện tại là kết quả của quá trình khai thác chọn rừng tự nhiên từ những năm 1980 với cường độ khác nhau [29]. Cụ thể, quá trình khai thác chọn với cường độ nhỏ hơn 10% trữ lượng gọi là khai thác cường độ thấp, trữ lượng rừng còn lại lớn và sau thời gian phục hồi những trạng thái rừng này gọi là rừng giàu (trữ lượng lớn hơn 200 m^3/ha). Rừng trung bình trữ lượng từ 100-200 m^3/ha (cường độ khai thác chọn 10-30% và có thời gian phục hồi rừng từ 20-30 năm). Rừng nghèo trữ lượng từ 50-100 m^3/ha (cường độ khai thác 30-50% và có thời gian phục hồi rừng từ 10-20 năm). Rừng nghèo kiệt trữ lượng dưới 10-50 m^3/ha (rừng bị khai thác cường độ 30-50% và có thời gian phục hồi từ 5-10 năm). Như vậy, khả năng tích trữ carbon của rừng lá rộng thường xanh chịu ảnh hưởng của quá trình tác động của con người và thời gian phục hồi của rừng.

3.2. Lựa chọn mô hình ước tính trữ lượng carbon

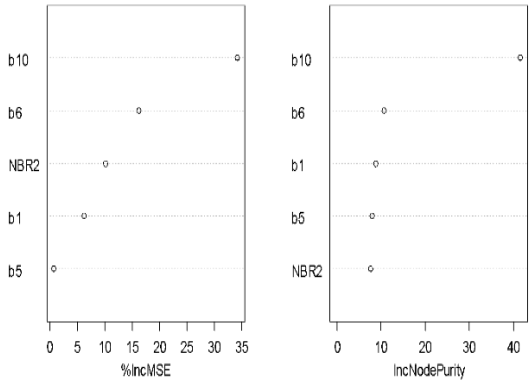
Mô hình rừng ngẫu nhiên (RF) có sai số MAE và RMSE nhỏ hơn so với 3 mô hình (hồi quy tuyến tính: LM; mạng nơ ron: NNET; Vecto hỗ trợ: SVM) với sự sai khác có ý nghĩa thống kê (p<0,05). Về hệ số tương quan (Rsquared) mô hình rừng ngẫu nhiên cũng cao hơn so với 3 mô hình còn lại với hệ số tương quan R^2=0,42 (Hình 4).



Hình 4. So sánh độ chính xác của 4 mô hình toán học máy bằng dữ liệu kiểm chứng (MAE: sai số trung bình tuyệt đối; RMSE: sai số toàn phương trung bình; R^2: hệ số tương quan; LM: hồi quy tuyến tính; NNET: mạng thần kinh nhân tạo, RF: rừng ngẫu nhiên, SVM: véc tơ hỗ trợ).

Về tổng thể, sai số của RF tương đối cao trong khoảng từ 0 đến 100 cây; từ 100-250 cây mức độ sai số ổn định hơn và từ 250 cây trở đi thì sai số nhìn chung ít biến động và ổn định. Việc lựa chọn số ntree phù hợp cho quá trình phân loại phụ thuộc vào mức độ ổn định và bão hòa của sai số cũng như thời gian ít nhất để đạt được sai số là thấp nhất, giảm thiểu được thời gian tính toán. Số lượng cây quyết định càng ít thì thời gian tính toán càng thấp và ngược lại. Kết

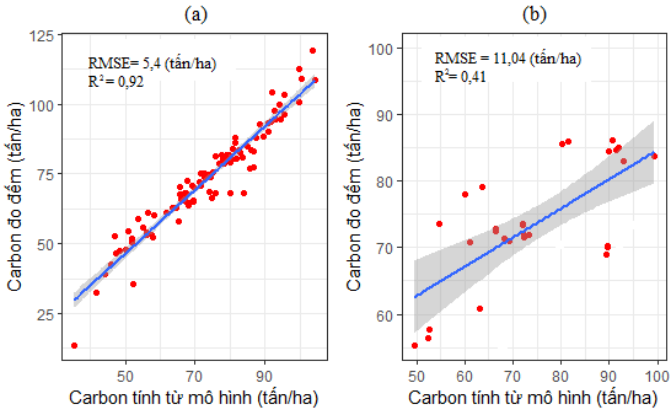
quả thể hiện ở phụ lục 02 cho thấy ở khoảng ntree = 250 thì mức độ sai số gần như bão hòa và biến động rất ít nên xác định ntree của mô hình RF trong nghiên cứu này là 250 cây.



Hình 5. Mức độ quan trọng của các biến dự đoán (b1, b5, b6, b10 giá trị các kênh 1, 5, 6 và 10 của ảnh Landsat 8; NBR2: chỉ số hỏa hoạn hiệu chỉnh; %IncMSE: mức tăng lên của sai số trung phương tương đối; IncNodePurity: mức tăng lên tính đồng nhất của dữ liệu).

Đối với tham số mtry của mô hình rừng ngẫu nhiên, kết quả ở phụ lục 03 cho thấy khi sử dụng 1 biến (1 kênh ảnh) tại mỗi nút phân chia (mtry) thì sai số của phần kiểm chứng RMSE là 0,775, còn nếu sử dụng mtry=2 thì sai số RMSE là 0,76 và mtry=4 tương ứng sai số RMSE là 0,757. Việc sử dụng mtry=2 và mtry=4 thì sai số RMSE chỉ lệch 0,003, đây là tỉ lệ chênh lệch cực kì nhỏ nên khi sử dụng mtry=2 hoặc 4 đều mang lại kết quả có ý nghĩa thống kê. Trong nghiên cứu này sử dụng mtry=4 là thông số mà kết quả mong đợi.

Ảnh hưởng của các kênh ảnh đến độ chính xác kết quả ước tính trữ lượng carbon trên mặt đất của các trạng thái rừng lá rộng thường xanh được xác định thông qua mức độ quan trọng của các biến trong mô hình rừng ngẫu nhiên (Hình 5). Kết quả cho thấy, mức độ quan trọng của các kênh ảnh trong mô hình rừng ngẫu nhiên ước tính trữ lượng carbon của rừng lần lượt là: kênh 10, 6, chỉ số NBR2 và kênh 1. Trong khi đó, kênh 5 có vai trò thấp nhất trong mô hình RF. Vì vậy, mô hình RF tối ưu trong nghiên cứu này là ntree=250, mtry=4 và sử dụng các kênh phổ 1, 5, 6 và 10 cùng với chỉ số NBR2.



Hình 6. Kết quả mô hình rừng ngẫu nhiên cho phần huấn luyện (a) và phần kiểm định (b) (RMSE: sai số toàn phương trung bình, R^2: hệ số tương quan).

Mô hình RF cho 2 phần huấn luyện và kiểm định được thể hiện ở hình 06. Từ hình 06 cho thấy khả năng dự báo của mô hình là tương đối sát với dữ liệu carbon đo thực tế. Giá trị tương quan thu được của phần huấn luyện R^2=0,92 và của phần kiểm định là 0,41. Vì vậy, trong 4 mô hình được đưa vào thử nghiệm thì mô hình RF cho việc dự báo carbon rừng là khả thi. Hệ số tương quan dành cho phần kiểm định chưa thực sự cao như phần huấn luyện. Tuy vậy, độ chính xác của phần kiểm định trong nghiên cứu này vẫn cao hơn của Safari và cộng sự (2017) [30] khi sử dụng ảnh Landsat 8 và thuật máy học ước tính carbon của rừng Sồi tại tỉnh Kermanshah,

Iran với R^2 từ 0,27 đến 0,35 cho phần dữ liệu kiểm định. Tương tự, Dube và Mutanga (2015) [31] khi sử dụng ảnh Landsat ETM + để ước tính sinh khối trên mặt đất cho rừng tự nhiên ở tỉnh KwaZulu-Natal, Nam Phi với R^2 bằng 0,32. Xét về sai số tuyệt đối của mô hình dự đoán trong nghiên cứu này là 11,04 (tấn/ha) cũng thấp hơn so với kết quả của Gizachew và cộng sự (2016) [8] với RMSE bằng 56,00 (tấn/ha) khi dùng mô hình tuyến tính ước tính trữ lượng carbon từ ảnh Landsat 8 cho rừng miombo của Tanzania. Thuật toán RF được sử dụng rộng rãi và có độ chính xác cao trong xây dựng các mô hình ước tính sinh khối và carbon rừng từ ảnh vệ tinh. Liu và cộng sự (2017) [32] thông qua nghiên cứu cho thấy khả năng của các thuật toán phi tham số máy học như RF trong việc ước tính sinh khối cũng như trữ lượng carbon từ dữ liệu viễn thám với độ chính xác cải thiện đáng kể so với mô hình tuyến tính. Kết quả tương tự về độ chính xác của mô hình RF cũng được tìm thấy trong nghiên cứu của Dube và Mutanga (2015) [31], khi sử dụng ảnh Landsat 8 để đánh giá sinh khối trên mặt đất tại uMgeni catchment, Nam Châu Phi. Dang và cộng sự (2019) [33] khi sử dụng ảnh Sentinel ước tính sinh khối trên mặt đất tại Vườn Quốc Gia York Đôn tại Việt Nam cũng cho thấy độ chính xác cao của mô hình RF trong kết quả dự đoán. Độ chính xác của kết quả dự đoán phụ thuộc vào đặc trưng cấu trúc của rừng.Các trạng thái rừng loài cây đồng nhất hoặc chỉcó một số loài ưu thế thì mô hình dự đoán sai số thường thấp hơn, chẳng hạn Safari và cộng sự (2017) [30] khi sử dụng ảnh Landsat 8 để tính trữ lượng carbon của rừng Sồi tại Iran thì RMSE chỉ bằng 2,94 tấn/ha cho mô hình RF. Điều này được lý giải là do khu vực rừng thưa mật độ sinh khối thấp, cấu trúc rừng đồng nhất với loài cây Sồi ưu thế trong cả khu vực nghiên cứu. Do khu vực nghiên cứu có nhiều trạng thái rừng lá rộng thường xanh khác nhau khác nhau với cấu trúc và thành phần loài cây đa dạng dẫn đến sai số của mô hình với RMSE bằng 11,04 (tấn/ha).

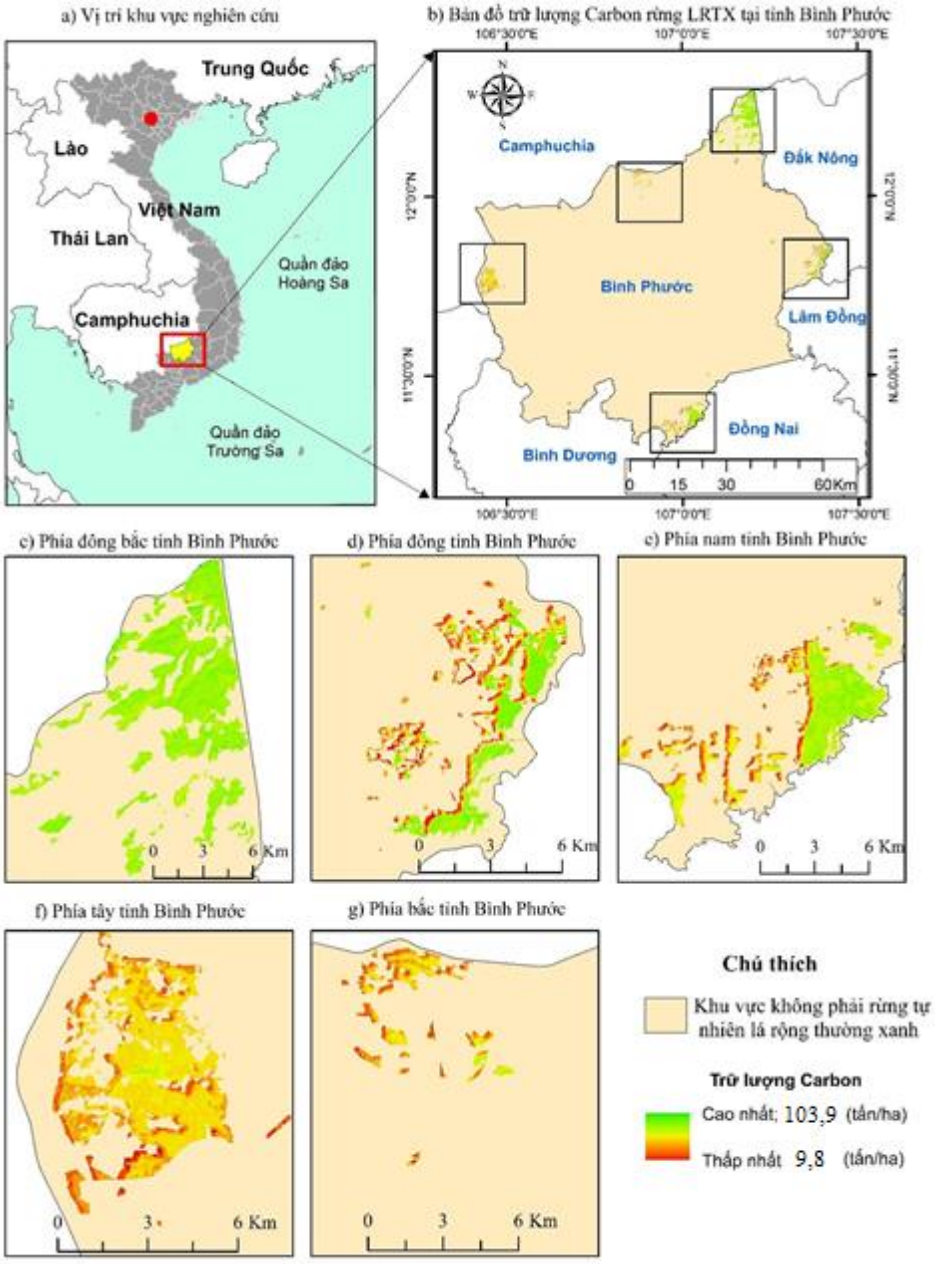
3.3. Xây dựng bản đồ tích lũy carbon các trạng thái rừng lá rộng thường xanh tại khu vực nghiên cứu

Thông qua mô hình RF giữa trữ lượng và các giá trị kênh ảnh ta có bản đồ ước tính trữ lượng carbon khu vực nghiên cứu (Hình 7). Từ kết quả mô hình cho ta thấy trữ lượng carbon lớn nhất tính từ ảnh đạt 104 tấn/ha tập trung ở khu vực giáp biên giới Campuchia và tỉnh Đăk Nông thuộc các lâm phần rừng lá rộng thường xanh giàu của Vườn quốc gia Bù Gia Mập. Ngược lại, khu vực phía tây và phía bắc trữ lượng carbon thấp, nơi tập trung các trạng thái rừng nghèo kiệt. Trong khi đó, các trạng thái rừng tự nhiên giáp tỉnh Lâm Đồng và Đồng Nai có trữ lượng carbon ở mức dưới 80 (tấn/ha), ở đây có các trạng thái rừng trung bình, nghèo và kiệt. Trữ lượng carbon trung bình tính từ ảnh Landsat 8 bằng thuật toán RF cho các trạng thái rừng lần lượt là: 87,78 tấn/ha (TXG), 77,58 tấn/ha (TXB), 70,29 tấn/ha (TXN) và 59,80 tấn/ha (TXK). Tổng trữ lượng carbon tích lũy trong các trạng thái rừng lá rộng thường xanh tại tỉnh Bình Phước năm 2020 là: 606,42 nghìn tấn (Bảng 3).

Bảng 3. Trữ lượng carbon tích lũy trong các trạng thái rừng lá rộng thường xanh tại tỉnh Bình Phước ước tính từ ảnh Landsat 8 bằng thuật toán học rừng ngẫu nhiên

Trạng thái rừng	Diện tích (ha)	Trữ lượng carbon tính trên 1 ha (tấn/ha)	Tổng trữ lượng carbon (nghìn tấn)
Rừng lá rộng thường xanh giàu	5254,11	87,78	247,18
Rừng lá rộng thường xanh trung bình	4126,68	77,58	174,91
Rừng lá rộng thường xanh nghèo	4300,39	70,29	166,84
Rừng lá rộng thường xanh nghèo kiệt	524	59,80	17,49
Tổng	14205,18		606,42

Trữ lượng carbon của các trạng thái rừng tại tỉnh Bình Phước từ 9,8 tấn/ha đến 103,9 tấn/ha, tương tự so với các trạng thái rừng lá rộng thường xanh tại tỉnh Đăk Nông với trữ lượng carbon trên mặt đất của trạng thái rừng nghèo là 34,39 tấn/ha và rừng giàu là 107,73 tấn/ha [34], hay tại Khu bảo tồn Xuân Liên, Thanh Hóa với trữ lượng carbon ước tính từ 42,78-118,35 tấn/ha [35]. Tuy vậy, nó thấp hơn so với các trạng thái rừng tự nhiên tại lưu vực sông Vụ Gia Thu Bồnnằm trên các tỉnh Đà Nẵng, Kon Tum và Quảng Nam với trữ lượng carbon từ 18,2 tấn/ha đến 166,8 tấn/ha [36].



Hình 7. Bản đồ ước tính trữ lượng carbon các trạng thái rừng lá rộng thường xanh tại tỉnh Bình Phước.

4. Kết luận

Nghiên cứu sử dụng 4 thuật toán học máy tính (rừng ngẫu nhiên, véctơ hỗ trợ, mô hình nơ-ron nhân tạo, hồi quy tuyến tính) để dự đoán trữ lượng carbon trên mặt đất của rừng tự nhiên lá rộng thường xanh tại tỉnh Bình Phước từ ảnh vệ tinh Landsat 8. Kết quả chỉ ra rằng mô hình rừng ngẫu nhiên khả thi trong việc ước tính trữ lượng carbon thông qua giá trị các kênh phổ với hệ số tương quan cho dữ liệu huấn luyện là 0,92 và dữ liệu kiểm định là 0,41. Trong đó, các kênh 10, 6, 1 và chỉ số NBR2 của ảnh Landsat 8 là biến số quan trọng trong mô hình ước tính trữ lượng carbon rừng lá rộng thường xanh tại khu vực nghiên cứu. Sai số của mô hình huấn luyện cho dữ liệu kiểm định là 11,04 tấn/ha. Thông qua mô hình rừng ngẫu nhiên (RF) trữ lượng carbon các trạng thái rừng lá rộng thường xanh khu vực nghiên cứu được ước tính từ ảnh Landsat 8 với giá trị từ 9,8 đến 103,9 tấn/ha. Cuối cùng, nghiên cứu đề xuất một trong những hướng nghiên cứu trong tương lai trong việc sử dụng mô hình thuật toán máy học kết hợp với những ảnh vệ tinh có độ phân giải cao hơn như SPOT, ALOS PALSAR,… để nâng cao độ chính xác trong ước tính trữ lượng carbon cho các trạng thái rừng khác nhau tại Việt Nam.

Tài liệu tham khảo

[1] https://www.fao.org/newsroom/detail/Unlocking-the-carbon-storage-potential-of-forests-through-better-data/en (accessed on: August 15th, 2022).

[2] V. H. Than, L. T. Truong, V.T. Hien, REDD+ Training Material, RECOFTC, 2017, pp. 1-7 (in Vietnamese).

[3] T. Q. Bao, Indentifying Carbon Sequestration of Forest Through Remote Sensing in Kim Boi District, Hoa Binh Province, Vietnam, Journal of Forestry Science and Technology, No. 1, 2013, pp. 14-21 (in Vietnamese).

[4] V. L. Nguyen, Estimation of Biomass for Calculating Carbon Storage and CO2 Sequestration Using Remote Sensing Technology in Yok Don National Park, Central Highlands of Vietnam, Journal of Vietnamese Environment, Vol. 3, No. 1, 2012, pp. 14-18, https://doi.org/10.13141/jve.vol3.no1.pp14-18 (in Vietnamese).

[5] T. D. Pham, N. N. Le, N. T. Ha, L. V. Nguyen, J. Xa, N. Yokoya, T. T. To, H. X. Trinh, L. Q. Kieu, W. Takeuchi, Estimating Mangrove Above-Ground Biomass Using Extreme Gradient Boosting Decision Trees Algorithm with Fused Sentinel-2 and ALOS-2 PALSAR-2 Data in Can Gio Biosphere Reserve, Vietnam, Remote Sensing, Vol. 12, No. 5, 2020, pp. 777-789, https://doi.org/10.3390/rs12050777.

[6] H. H. Nguyen, T. T. H. Nguyen, Above-Ground Biomass Estimation Models of Mangrove Forests Based on Remote Sensing and Field-Surveyed Data: Implications for C-PFES Implementation in Quang Ninh Province, Vietnam, Regional Studies in Marine Science, Vol. 48, 2021, pp. 101985, https://doi.org/10.1016/j.rsma.2021.101985.

[7] R. Erika, M. Herold, L. Kooistra, D. Murdiyarso, L. Verchot, Assessing Capacities of non-Annex I Countries for National Forest Monitoring in The Context of REDD+, Environmental Science & Policy, Vol. 19, 2012, pp. 33-48, https://doi.org/10.1016/j.envsci.2012.01.005.

[8] B. Gizachew, S. Solberg, E. Næsset, T. Gobakken, O. M. Bollandsås, J. Breidenbach, E. W. Mauya. Mapping and Estimating the Total Living Biomass and Carbon in Low-Biomass Woodlands Using Landsat 8 CDR Data, Carbon Balance and Management, Vol. 11, No. 1, 2026, pp. 1-14, https://doi.org/10.1186/s13021-016-0055-8.

[9] M. A. Marelign, D. T. Mekonen, Estimating and Mapping Woodland Biomass and Carbon Using Landsat 8 Vegetation Index: A Case Study in Dirmaga Watershed, Ethiopia, Computational Ecology and Software, Vol. 12, No. 2, 2022, pp. 67-79.

[10] J. Zhou, Z. Zhou, Q. Zhao, Z. Han, P. Wang, J. Xu, Y. Dian, Valuation of Different Algorithms for Estimating the Growing Stock Volume of Pinus Massoniana Plantations Using Spectral and Spatial Information from a SPOT6 Image, Forests, Vol. 11, No. 5, pp. 540-554, https://doi.org/10.3390/f11050540.

[11] P. V. Duan, V. T. Thin, Determinate Estimate Biomass and Volume Forests from Satellite Images, Journal of Forestry Science and Technology, No. 3, 2015, pp. 17-28 (in Vietnamese).

[12] D. T. A. Ngoc, N. T. T. Van, N. P. Anh, Monitoring of Rice Paddy and Estimating Biomass Based on Machine Learning Algorithms to Multi-temporal Sentinel-1a Data, Journal of Geodesy and Cartography, No. 49, 2021, pp. 52-64 (in Vietnamese).

[13] Binh Phuoc Sub-Department of Forest Protection, Survey and Assessment of Forest Area and Forest Status in Binh Phuoc Province, 2020, pp. 1-145 (in Vietnamese).

[14] P. H. Ho, An Illustrated Flora of Vietnam, Young Publisher, Hanoi, 2003 (in Vietnamese).

[15] T. Hop, Trees of Vietnam, Agricultural Publisher, Hanoi, 2002 (in Vietnamese).

[16] B. Huy, K. Kralicek, K. P. Poudel, V. T. Phuong, P. V. Khoa, N. D. Hung, H. Temesgen, Allometric Equations for Estimating Tree Above-Ground Biomass in Evergreen Broadleaf Forests of Vietnam, Forest Ecology and Management, Vol. 382, 2016, pp. 193-205, https://doi.org/10.1016/j.foreco.2016.10.021.

[17] A. E. Zanne, G. L. Gonzalez, D. A. Coomes, J. Ilic, S. Jansen, S. L. Lewis, R. B. Miller, N. G. Swenson, M. C. Wiemann, J. Chave. Data from: Towards a Worldwide Wood Economics Spectrum, https://doi.org/10.5061/dryad.234.

[18] IPCC, Guidelines for National Greenhouse Gas Inventories, Prepared by the National Greenhouse Gas Inventories Programme, Eggleston HS, Buendia L, Miwa K, Ngara T, Tanabe K, Editors, Published: IGES, Japan, 2006.

[19] J. R. Jr, R. H. Haas, J. A. Schell, D. W. Deering, Monitoring Vegetation Systems in the Great Plains with ERTS, the Proceedings of a Symposium Held by Goddard Space Flight Center at Washington, 1974, pp. 309-318.

[20] A. R. Huete, A Comparison of Vegetation Indices Global Set of TM Images for EOS-MODIS, Remote Sensing of Environment, Vol. 59, No. 3, 1997, pp. 440-451, https://doi.org/10.1016/S0034-4257(96)00112-5.

[21] A. R. Huete, A Soil-Adjusted Vegetation Index (Savi), Remote Sensing of Environment, Vol. 25, No. 3, 1988, pp. 295-309, https://doi.org/10.1016/0034-4257(88)90106-X.

[22] G. Rondeaux, M. Steven, F. Baret, Optimization of Soil-Adjusted Vegetation Indices, Remote Sensing of Environment, Vol. 55, No. 2, 1996, pp. 95-107, https://doi.org/10.1016/0034-4257(95)00186-7.

[23] B. C. Gao, NDWI- a Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space. Remote Sensing of Environment, Vol. 58, No. 3, 1996, pp. 257-266, https://doi.org/10.1016/S0034-4257(96)00067-3.

[24] J. D. Miller, A. E. Thode, Quantifying Burn Severity in a Heterogeneous Landscape with a Relative Version of the Delta Normalized Burn Ratio (dNBR), Remote Sensing of Environment, Vol. 109, No. 1, 2007, pp. 66-80, https://doi.org/10.1016/j.rse.2006.12.006.

[25] N. Ahmed, C. Atzberger, W. Zewdie, Species Distribution Modelling Performance and its Implication for Sentinel-2-Based Prediction of Invasive Prosopis juliflora in Lower Awash River Basin, Ethiopia, Ecological Processes, Vol. 10, No. 1, 2021, pp. 1-16, https://doi.org/10.1186/s13717-021-00285-6.

[26] T. T. P. Nguyen, M. H. Duy, T. N. Duong, T. D. Nghiem, Forecast of Hourly Tropospheric Ozone Concentration in Quang Ninh using MLP and SVM, VNU Journal of Science: Earth and Environmental Sciences, Vol. 36, No. 3, 2020, pp. 46-54, https://doi.org/10.25073/2588-1094/vnuees.4604.

[27] G. Chen, G. J. Hay, A Support Vector Regression Approach to Estimate Forest Biophysical Parameters at the Object Level using Airborne Lidar Transects and Quickbird Data, Photogrammetric Engineering & Remote Sensing, Vol. 77, No. 7, 2011, pp 733-741, https://doi.org/10.14358/PERS.77.7.733.

[28] L. H. An, S. H. Dang, Determination of the Mineral Volumes for the Pre-Cenozoic Magmatic Basement Rocks of Cuu Long Basin from Well Log Data via Using the Artificial Neural Networks, VNU Journal of Science: Earth and Environmental Sciences, Vol. 30, No. 1, 2014, pp. 1-11.

[29] V. D. Hai, T. V. Do, D. T. Trieu, T. Sato, O. Kozan, Carbon Stocks in Tropical Evergreen Broadleaf Forests in Central Highland, Vietnam, International Forestry Review, Vol. 17, No. 1, 2015, pp. 20-29, https://doi.org/10.1505/146554815814725086.

[30] A. Safari, H. Sohrabi, S. Powell, S. Shataee, A Comparative Assessment of Multi-temporal Landsat 8 and Machine Learning Algorithms for Estimating Aboveground Carbon Stock in Coppice Oak Forests, International Journal of Remote Sensing, Vol. 38, No. 22, 2017, pp. 6407-6432, https://doi.org/10.1080/01431161.2017.1356488.

[31] T. Dube, O. Mutanga, Evaluating the Utility of the Medium-Spatial Resolution Landsat 8 Multispectral Sensor in Quantifying Aboveground Biomass in uMgeni Catchment, South Africa, ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 101, 2015, pp. 36-46, https://doi.org/10.1016/j.isprsjprs.2014.11.001.

[32] K. Liu, J. Wang, W. Zeng, J. Song, Comparison and Evaluation of three Methods for Estimating Forest Above Ground Biomass using TM and GLAS Data, Remote Sensing, Vol. 9, No. 4, 2017, pp. 341-353, https://doi.org/10.3390/rs9040341.

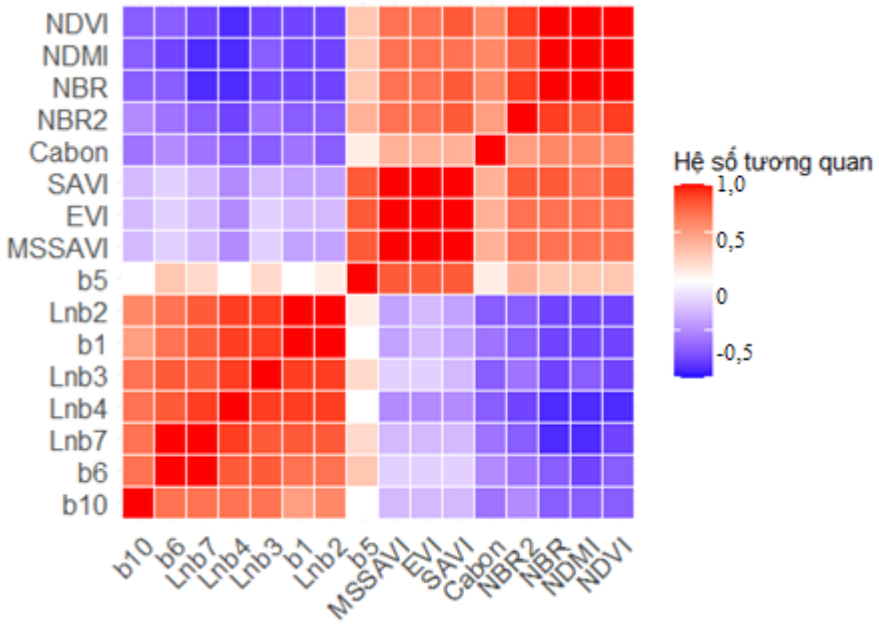
[33] A. T. Dang, S. Nandy, R. Srinet, N. V. Luong, S. Ghosh, A. S. Kumar, Forest Aboveground Biomass Estimation using Machine Learning Regression Algorithm in Yok Don National Park, Vietnam. Ecological Informatics, Vol. 50, 2019, pp. 24-32, https://doi.org/10.1016/j.ecoinf.2018.12.010.

[34] N. V. Hop, N. V. Quy, B. H. Quoc, N. T. Luong, Relationship between Plant Biodiversity and Carbon Stocks in Evergreen Broad-leaved Forests in the Central Highlands, Journal of Forestry Science and Technology, No. 11, 2021, pp. 59-69.

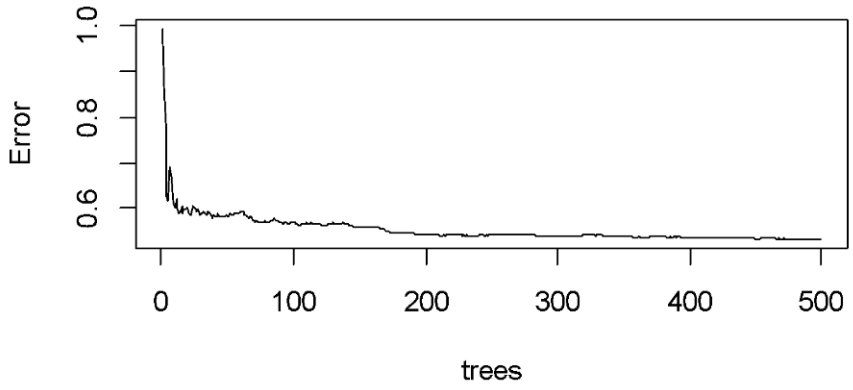
[35] T. D. Nguyen, M. Kappas, Estimating the Above-ground Biomass of an Evergreen Broadleaf forest in Xuan Lien Nature Reserve, Thanh Hoa, Vietnam, using SPOT-6 data and the Random forest Algorithm, International Journal of Forestry Research, Vol. 2020, 2020, pp. 1-13, https://doi.org/10.1155/2020/4216160.

[36] V. Avitabile, M. Schultz, N. Herold, S. D. Bruin, A. K. Pratihast, C. P. Manh, M. Herold, Carbon Emissions from Land Cover Change in Central Vietnam, Carbon Management, Vol. 7, No. 6, 2016, pp. 333-346, https://doi.org/10.1080/17583004.2016.1254009.

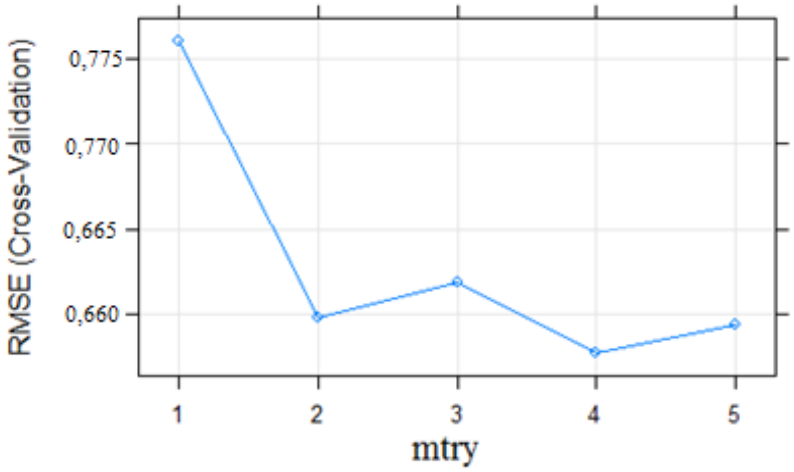
Phụ lục



Phụ lục 1. Hệ số tương quan giữa các kênh ảnh với trữ lượng Carbon
(b1, b5, b6, b10: giá trị kênh 1, 5, 6 và 10 của ảnh Landsat 8; Lnb2, Lnb3, Lnb4, Lnb7: giá trị logarit tự nhiên của kênh 2, 3, 4 và 7 của ảnh Landsat 8; NDVI: cỉ số thực vật; EVI: Chỉ số thực vật tăng cường, SAVI: chỉ số thực vật có hiệu chỉnh phản xạ mặt đất, MSAVI: chỉ số SAVI hiệu chỉnh; NDMI: chỉ số độ ẩm khác biệt; NBR và NBR2 chỉ số hỏa hoạn 1 và 2).



Phụ lục 2. Biến thiên sai số của mô hình RF dựa vào số lượng các cây ntree (Error: sai số toàn phương của mô hình rừng ngẫu nhiên, trees: số lượng cây để xây dựng mô hình).



Phụ lục 3. Mức độ sai số RMSE dữ liệu kiểm chứng với sự thay đổi giá trị của mtry (RMSE sai số toàn phương của dữ liệu kiểm định, mtry là số lượng biến dùng để chạy mô hình).