

DOI:10.22144/ctu.jvn.2022.254

SỬ DỤNG ẢNH VỆ TINH SENTINEL 2 VÀ THUẬT TOÁN HỌC MÁY THÀNH LẬP BẢN ĐỒ HIỆN TRẠNG RỪNG TẠI HUYỆN BÙ ĐĂNG, TỈNH BÌNH PHƯỚC

Nguyễn Thanh Tuấn^{1*}, Phan Văn Tuấn¹, Nguyễn Văn Quý¹ và Hoàng Thị Phương Nhung²

¹Phân hiệu Trường Đại học Lâm nghiệp tại Đồng Nai

²Chi cục Kiểm lâm tỉnh Bình Phước

*Người chịu trách nhiệm về bài viết: Nguyen Thanh Tuan (email: nttuan@vnfu2.edu.vn)

Thông tin chung:

Ngày nhận bài: 27/08/2022

Ngày nhận bài sửa: 28/09/2022

Ngày duyệt đăng: 11/10/2022

Title:

Using Sentinel-2 satellite images and machine learning algorithms for forest mapping in Bu Dang district, Binh Phuoc province

Từ khóa:

Hiện trạng sử dụng đất, mạng nhân tạo, rừng ngẫu nhiên, thuật toán phân loại, vector hỗ trợ

Keywords:

Land use, neural networks, random forest, classification algorithms, support vector machine

ABSTRACT

This research used three machine learning algorithms-MLC (neural networks-NNET, random forest-RF, and support vector machine-SVM) for classification of Sentinel 2 imagery into forest categories in Bu Dang district, Binh Phuoc province. Ten types of land use and land cover were discriminated for the whole area, in which 6 forest classes were classified in the study area. The highest accuracy was found using RF with overall accuracy of 90% and Kappa of 0,86, followed by the NNET and SVM algorithm. Among the Sentinel-2 spectral bands, the SWIR B11, followed by the red-edge B5, B6, narrow NIR B8A, SWIR B12, and the green B3 contributed the most to the accuracy of the forest types of classification. In conclusion, the results of this study demonstrated that the random forest algorithm is well-suited for vegetation mapping using Sentinel imagery, which can be applied to large-scale forest type's classifications.

TÓM TẮT

Nghiên cứu sử dụng 3 thuật toán phân loại học máy - MLC (mạng thần kinh nhân tạo - NNET, rừng ngẫu nhiên - RF và véc tơ hỗ trợ - SVM) để phân loại hiện trạng rừng của huyện Bù Đăng, tỉnh Bình Phước từ ảnh vệ tinh Sentinel-2. Mười loại hiện trạng sử dụng đất và lớp phủ trong đó có 6 lớp hiện trạng rừng đã được phân loại trong khu vực nghiên cứu. Phương pháp RF có độ chính xác cao nhất với độ chính xác toàn cục bằng 90% và hệ số Kappa bằng 0,86, tiếp theo là NNET và SVM. Trong số các kênh ảnh của vệ tinh Sentinel 2, kênh sóng ngắn hồng ngoại (kênh 11), tiếp theo là kênh rìa đỏ (kênh 5 và 6), kênh gần hồng ngoại hẹp (8A), kênh sóng ngắn hồng ngoại (kênh 12) và kênh xanh lục (kênh 3) đóng góp nhiều nhất vào độ chính xác của kết quả phân loại. Kết quả của nghiên cứu này đã chứng minh rằng thuật toán RF phù hợp cho việc lập bản đồ thảm thực vật từ ảnh vệ tinh Sentinel, và có thể được áp dụng được cho phân loại bản đồ hiện trạng rừng ở quy mô không gian rộng.

1. GIỚI THIỆU

Xây dựng bản đồ hiện trạng rừng là một nhiệm vụ quan trọng, cung cấp cơ sở dữ liệu cho công tác quản lý và phát triển rừng (Gebhardt et al., 2014;

Trung & Huong, 2019). Đặc biệt, kiểm kê hàng năm hoặc định kỳ hiện trạng rừng còn dùng để xác định tình trạng suy thoái do thiên tai hoặc do tác động của con người, từ đó đề xuất những chiến lược quản lý

phù hợp (Luong et al., 2017; Furuya et al., 2020). Kiểm kê rừng có thể thực hiện bằng các phương pháp đo trên mặt đất, sử dụng ảnh vệ tinh, hoặc sử dụng kết hợp cả 2 phương pháp. Việc kiểm kê rừng bằng các phương pháp truyền thống đo đếm trực tiếp thường tốn kém và khó tiếp cận ở những nơi vùng núi hoặc bị địa hình chia cắt. Ở những nơi như vậy, việc sử dụng ảnh vệ tinh là cần thiết và đem lại hiệu quả cao hơn (Li et al., 2013). Hiện nay, tư liệu viễn thám được khai thác từ nhiều loại vệ tinh khác nhau như Landsat, Sentinel, Spot,... và được ứng dụng rộng rãi trong nhiều lĩnh vực như quản lý môi trường, lâm nghiệp, khoáng sản (Volke & Abarca, 2020). Trong đó, sử dụng ảnh vệ tinh để thành lập bản đồ hiện trạng rừng là một trong những thành tựu nổi bật trong quản lý rừng (Huong & Trung, 2018). Trong số các loại ảnh vệ tinh được sử dụng để phân loại lớp phủ mặt đất nói chung và hiện trạng rừng nói riêng, ảnh Sentinel được sử dụng nhiều do được cung cấp miễn phí, chu kỳ chụp ngắn, đa kênh phổ và độ phân giải không gian khá cao (Duẩn và ctv., 2019; Trung & Huong, 2019).

Một trong những vấn đề đặt ra trong việc sử dụng ảnh vệ tinh phân loại lớp phủ mặt đất là độ chính xác của kết quả nhận được. Trong đó, kết quả phân loại các đối tượng từ ảnh vệ tinh phụ thuộc nhiều yếu tố như cảm biến, dữ liệu huấn luyện, độ chính xác dữ liệu đánh giá, số lớp, đặc biệt là các thuật toán sử dụng trong công tác xử lý và phân loại ảnh (Tùng và ctv., 2021). Hiện nay, có nhiều phương pháp được áp dụng nhằm cải thiện độ chính xác phân loại ảnh vệ tinh và thành lập các bản đồ lớp phủ mặt đất (Lu & Weng, 2007). Ngoài các phương pháp thường được dùng phổ biến như phân loại có kiểm định và phân loại định hướng đối tượng, gần đây phương pháp cây quyết định (Decision Tree), mạng thần kinh nhân tạo (Neural Network), véc-tơ hỗ trợ (Support Vector Machine), láng giềng gần nhất (k-Nearest Neighbor),... được sử dụng nhiều trong các nghiên cứu về phân loại đối tượng từ ảnh vệ tinh (Apte & Patravali, 2015; Mora et al., 2017). Những phương pháp này được ứng dụng trong các mô hình bằng máy hay gọi là học máy (Machine Learning). Trong số các phương pháp học máy, Support Vector Machine (SVM) và Random Forest (RF) đã nổi lên như là các phương pháp tiềm năng để phân loại ảnh vệ tinh (Noi & Kappas, 2017; Hiền và ctv., 2022).

Tuy nhiên, việc sử dụng thuật toán học máy để phân loại ảnh vệ tinh hiện nay chưa có nhiều tác giả đề cập, một số ít nghiên cứu chủ yếu được thực hiện ở rừng ôn đới hoặc rừng trồng. Cho đến nay rất ít các nghiên cứu thực hiện tại các khu rừng bị tác

động như ở Việt Nam, vì vậy đây vẫn là một chủ đề đang bỏ ngõ và cần được tiếp cận để có thể nâng cao hiệu quả trong phân loại ảnh vệ tinh xây dựng các bản đồ hiện trạng rừng (Huong & Trung, 2018). Vì vậy, cần có những nghiên cứu đánh giá khả năng sử dụng ảnh vệ tinh kết hợp với thuật toán học máy trong việc phân loại các loại rừng tạo tiền đề cho việc thành lập bản đồ chi tiết các loại trạng thái rừng, phục vụ công tác kiểm kê rừng ở nước ta. Trong nghiên cứu này, ảnh Sentinel 2 của khu vực huyện Bù Đăng, tỉnh Bình Phước được sử dụng làm đối tượng nghiên cứu với mục tiêu (i) đánh giá độ chính xác của các thuật toán học máy trong mô hình phân loại các hiện trạng rừng và đất lâm nghiệp từ ảnh Sentinel 2; (ii) kiểm tra mức độ phù hợp của các kênh ảnh Sentinel 2 trong các mô hình phân loại; và (iii) thành lập bản đồ hiện trạng rừng và đất lâm nghiệp cho huyện Bù Đăng, tỉnh Bình Phước.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Khu vực nghiên cứu

Nghiên cứu được thực hiện tại Huyện Bù Đăng, tỉnh Bình Phước. Khu vực này nằm trong vành đai khí hậu nhiệt đới gió mùa, cận xích đạo nên có nền nhiệt cao quanh năm, ít gió bão và không có mùa đông lạnh. Khí hậu phân hóa thành hai mùa rõ rệt, mùa mưa từ tháng 5 đến tháng 11 và mùa khô từ tháng 11 đến tháng 4 năm sau, lượng mưa bình quân khoảng 2.400 mm/năm. Nhiệt độ bình quân năm khoảng 24°C. Mùa mưa có lượng mưa chiếm 85 - 90% tổng lượng mưa cả năm. Tổng số giờ nắng trong năm từ 2.400 - 2.500 giờ. Số giờ nắng bình quân trong ngày từ 6,2 đến 6,6 giờ (Ban quản lý Rừng phòng hộ Bù Đăng, 2021).

2.2. Dữ liệu điều tra thực địa

Lấy mẫu thông qua điều tra thực địa được tiến hành vào tháng 8 đến tháng 12 năm 2020 dựa theo phương pháp của Trung và Huong (2019). Để đảm bảo tính khách quan và tăng độ chính xác, bản đồ hiện trạng rừng năm 2016 của tỉnh Bình Phước và ảnh Google Earth tại thời điểm nghiên cứu được sử dụng để xác định tọa độ các điểm mẫu điều tra trên ảnh. Cụ thể, mỗi một tầng là một hiện trạng rừng và đất lâm nghiệp khác nhau, trong mỗi tầng các mẫu sẽ được lấy ngẫu nhiên trên toàn bộ diện tích nghiên cứu (mẫu ngẫu nhiên phân tầng - stratified random sampling), sao cho khoảng cách giữa các mẫu trong một tầng lớn hơn 500 m. Trong đó, hiện trạng rừng và đất lâm nghiệp được phân chia dựa vào thông tư 33/2018/TT-BNNPTNT về điều tra, kiểm kê và theo dõi diễn biến rừng của Bộ Nông nghiệp và Phát triển nông thôn. Khu vực nghiên cứu được phân chia thành 10 loại hiện trạng rừng và đất lâm nghiệp. Với

các điểm mẫu như đất trống, nông nghiệp, đất khác (đất xây dựng công trình cơ sở hạ tầng phục vụ sản xuất nông lâm nghiệp), mặt nước, rừng trồng và rừng hỗn giao được khoanh vẽ ngoài thực địa bằng máy định vị toàn cầu cầm tay- GPS Garmin 64 (diện tích tối thiểu các vùng là 100 m²).

Dựa theo phương pháp của Thị và Bảo (2014) đối với các điểm mẫu thuộc các hiện trạng rừng tự nhiên lá rộng thường xanh, các ô tiêu chuẩn diện tích

500 m² được lập để phân loại theo trữ lượng (giàu, trung bình, nghèo và kiệt) và khoanh vẽ vùng bao quanh ô tiêu chuẩn ngoài hiện trường. Tổng cộng có 203 mẫu được thu thập (146 vùng mẫu và 57 ô tiêu chuẩn), sau đó toàn bộ vùng mẫu và ô tiêu chuẩn được chuyển về dữ liệu raster với kích thước 10 × 10 m. Số lượng và diện tích các mẫu cho từng hiện trạng rừng và đất lâm nghiệp sau khi khoanh vẽ ở thực địa được thể hiện ở Bảng 1.

Bảng 1. Số lượng mẫu khóa ảnh được sử dụng trong nghiên cứu

STT	Hiện trạng rừng và đất lâm nghiệp	Mã hiện trạng	Số lượng mẫu	Diện tích mẫu (số pixel)
1	Đất trống	DT	24	2.513
2	Mặt nước	MN	15	1.278
3	Đất khác	DKH	11	347
4	Hỗn giao	HG	22	12.744
5	Nông nghiệp	NN	8	392
6	Rừng trồng	RTG	49	16.236
7	Rừng tre nứa	RTN	17	3.776
8	Rừng lá rộng thường xanh trung bình	TXB	27	8.164
9	Rừng lá rộng thường xanh nghèo	TXN	18	3.999
10	Rừng lá rộng thường xanh kiệt	TXK	12	321

Trong mỗi ô tiêu chuẩn được lập ở các hiện trạng rừng lá rộng thường xanh 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 ($D_{1.3}$) đo bằng thước dây, chiều cao vút ngọn (H) được xác định bằng thước đo cao Blume-leiss. Tên loài cây gỗ được xác định ở hiện trường bằng phương pháp so sánh hình thái dựa trên các tài liệu gồm cây cỏ Việt Nam (Hộ, 2003), cây gỗ Việt Nam (Hợp, 2002), tên khoa học của loài cây 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 $D_{1.3} > 5$ cm.

Cuối cùng tính toán trữ lượng của mỗi ô tiêu chuẩn để phân chia ra các hiện trạng rừng lá rộng thường xanh khác nhau dựa vào dựa vào Thông tư 33/2018/TT-BNNPTNT: thường xanh giàu ($M \geq 200$ m³/ha), thường xanh trung bình ($100 \leq M < 200$ m³/ha), thường xanh nghèo ($50 \leq M < 100$ m³/ha) và thường xanh kiệt ($M < 50$ m³/ha). Trữ lượng (M : m³/ha) của ô tiêu chuẩn được tính dựa vào tổng thể tích của tất cả các cây (V_i) trong ô tiêu chuẩn tương ứng sau khi quy đổi ra đơn vị là ha.

$$V_i = 0,785 \times D_{1.3}^2 \times H_{vn} \times f \quad (1)$$

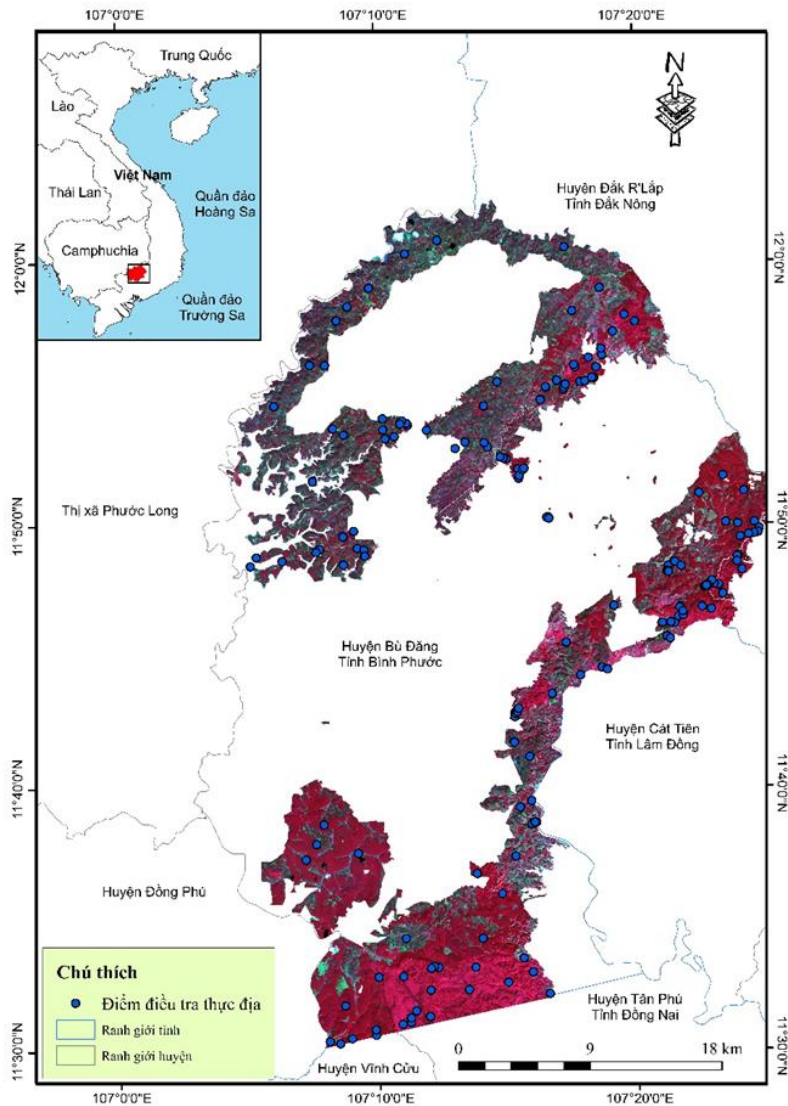
Trong đó, $D_{1.3}$ là đường kính tại vị trí 1,3 m, H_{vn} là chiều cao vút ngọn của cây rừng và f là hình số bằng 0,4826 (Hinh, 2012).

2.3. Thu thập và tiền xử lý dữ liệu ảnh viễn thám

Ảnh Sentinel 2 sử dụng trong nghiên cứu này được tải từ trang web <https://earthexplorer.usgs.gov>. Sau đó tiến hành lọc ảnh theo thời gian chụp từ tháng 8 đến tháng 12 năm 2020 để trùng với thời gian điều tra lấy mẫu ngoài thực địa và độ che phủ của mây nhỏ hơn 1%. Kết quả đã lọc được 2 ảnh Sentinel cấp 1C là L1C_T48PYT và L1C_T48PYU, ngày thu nhận là 18 tháng 11 năm 2020 với độ che phủ của mây là 0.0584 để đưa vào phân loại thành lập bản đồ hiện trạng rừng và đất lâm nghiệp của khu vực nghiên cứu. Trong tổng số 13 kênh của ảnh Sentinel 2 thu thập được, 10 kênh ảnh (kênh 2, 3, 4, 5, 6, 7, 8, 8A, 11 và 12) được lựa chọn để đưa vào phân loại. Trong đó, các kênh 1, 9 và 10 không được đưa vào phân loại vì những kênh này được sử dụng để hiệu chỉnh khí quyển (Puletti et al., 2018).

Tiếp theo, các bước tiền xử lý ảnh được thực hiện trước khi tiến hành phân loại các hiện trạng rừng và đất lâm nghiệp, bao gồm:

- Cắt ảnh khu vực nghiên cứu: Sử dụng bản đồ quy hoạch 3 loại rừng để xác định ranh giới rừng và đất lâm nghiệp của huyện Bù Đăng năm 2020 (Hình 1).



Hình 1. Bản đồ các điểm điều tra thực địa tại huyện Bù Đăng, tỉnh Bình Phước trên nền ảnh Sentinel 2 (tổ hợp kênh 8, 4, 3)

– Quá trình hiệu chỉnh ảnh hưởng khí quyển cho ảnh Sentinel 2 (cấp 1C) bằng phương pháp loại trừ điểm đen (DOS - Dark Object Subtract) được đề xuất bởi Chavez (1988) trên phần mềm Qgis 3.18.

– Các kênh ảnh có độ phân giải không gian khác nhau từ 10-20 (m). Vì vậy, các kênh ảnh được chuyển độ phân giải không gian về cùng kích thước 10 (m) bằng thuật toán láng giềng gần nhất trước khi tính chỉ số thực vật NDVI (Tucker, 1979):

$$NDVI = \frac{(Kênh\ 8 - Kênh\ 4)}{(Kênh\ 8 + Kênh\ 4)} \quad (2)$$

2.4. Các thuật toán phân loại được sử dụng trong nghiên cứu

Các kênh ảnh và chỉ số thực vật được chuẩn hóa về phân bố chuẩn trước khi tiến hành sử dụng các thuật toán phân loại bằng hàm normImage trong R (Singh & Singh, 2020; Izonin et al., 2022). Bước tiếp theo tiến hành phân loại các hiện trạng rừng và đất lâm nghiệp dựa vào ba thuật toán học máy (RF, SVM và NNET):

– Thuật toán rừng ngẫu nhiên (RF) là một số lượng lớn các cây quyết định được xác định hoàn toàn ngẫu nhiên từ các biến đầu vào để xác định giá

trị đầu ra. Theo Ho (1995), thuật toán lấy mẫu cho phương pháp RF như sau:

Một tập mẫu $X = x_1, \dots, x_n$ với các giá trị phân hồi $Y = y_1, \dots, y_n$, lấy giá trị trung bình (B lần), chọn một mẫu ngẫu nhiên từ bộ mẫu phù hợp với cây quyết định. Lập $b = 1 \dots B$; n mẫu từ giá trị tọa độ (X, Y) hay gọi là (X_b, Y_b) . Lốp dữ liệu hay kết quả của hồi quy f_b của biến X_b, Y_b .

Sau khi lấy mẫu, các phép tính toán cho các mẫu là ẩn số x' có thể được thực hiện bằng cách lấy trung bình các giá trị nội suy từ tất cả các cây hồi quy riêng lẻ của biến x' hoặc lấy giá trị từ đa số của các mẫu trong cây quyết định:

$$f = \frac{1}{B} \sum_{b=1}^B f_b(x') \quad (3)$$

Trong đó, hai tham số cần được xác định trong thuật toán phân loại này là n_{tree} (số lượng cây được phát triển) và m_{try} (số lượng biến để phân chia tại mỗi node). Số n_{tree} được lựa chọn phụ thuộc vào khoảng thời gian xử lý ngắn nhất để kết quả đạt được độ sai số thấp nhất. Nghiên cứu thử nghiệm giá trị n_{tree} chạy từ 1 đến 1.000 cây và m_{try} biến động từ số biến độc lập tối thiểu bằng 1 đến số biến độc lập tối đa được sử dụng trong phân loại là 11 biến (10 kênh phổ và chỉ số NDVI).

– Thuật toán véc-tơ hỗ trợ (SVM) là thuật toán 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 (Nguyen et al., 2020). Mẫu huấn luyện ban đầu: (x_i, y_i) , ($i=1,2,\dots,n$), 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 (Chen & Hay, 2011):

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

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, xác định W và b dựa vào sai số nhỏ nhất của hồi quy dựa vào công thức:

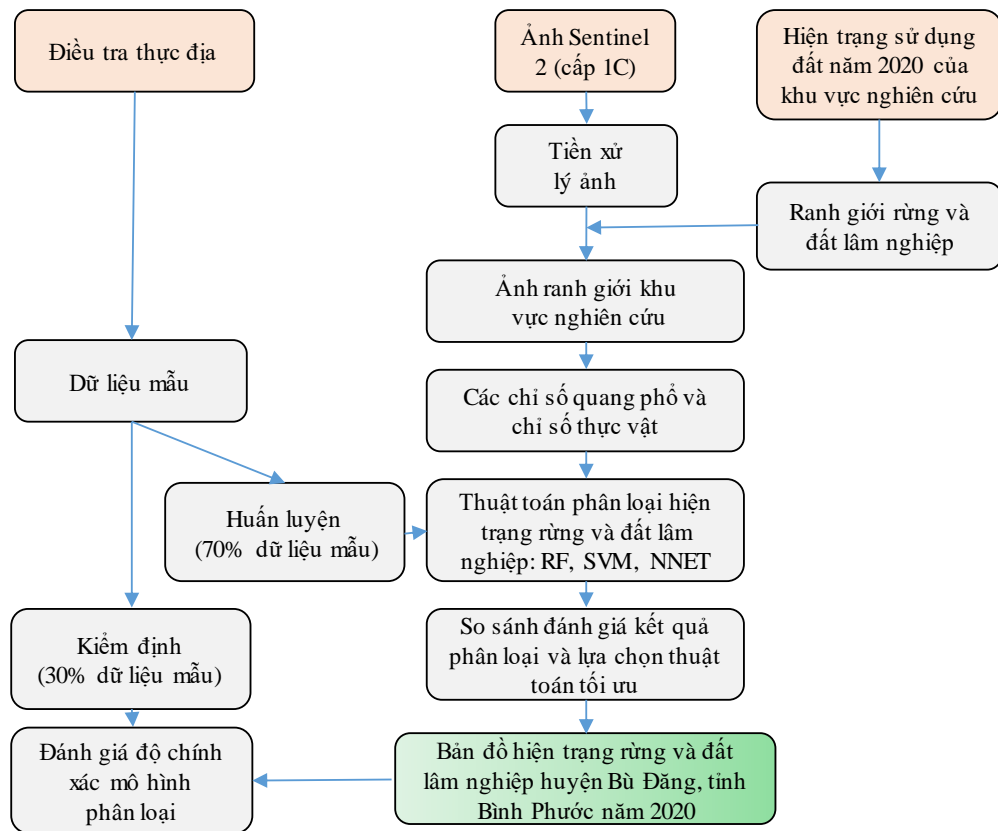
$$\begin{aligned} & \text{minimize } \frac{1}{2} \|W\|^2 + C \sum_{i=1}^n \xi_i + \xi_i^* \\ & \text{subject to } \begin{cases} y_i - f(x_i) \leq \varepsilon + \xi_i^* \\ f(x_i) - y_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0, i = 1, \dots, n \end{cases} \quad (6) \end{aligned}$$

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. Dựa vào nghiên cứu của Trung and Huang (2019), C được thử nghiệm với các giá trị $2^{-1}, 2^0, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6, 2^7, 2^8$.

– Mạng 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 nơ-ron 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 Sentinel 2, lớp ẩn và lớp đầu ra là biến cần dự đoán các lớp hiện trạng rừng và đất lâm nghiệp. Trong đó, lớp ẩn gồm các nơ-ron nhận dữ liệu đầu vào từ các nơ-ron ở lớp trước đó và chuyển đổi các dữ liệu đầu vào này cho các lớp xử lý tiếp theo. Một điểm quan trọng trong việc thiết lập một mạng NNET là làm cho kết quả của mô hình đạt độ chính xác cao và cấu trúc của mạng là thích hợp nhất. 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 nơ-ron và lớp ẩn được lựa chọn sao cho sai số của mô hình là nhỏ nhất. Nghiên cứu thử nghiệm số lượng lớp ẩn từ 1 cho đến 11, số lượng nơ-ron trong lớp ẩn được xác định dựa vào công thức của Shi and Zhang (2012):

$$i = \sqrt{j + m} + R \quad (7)$$

Trong đó, i là số nơ-ron trong lớp ẩn, m là số nơ-ron trong lớp dữ liệu đầu vào j là số nơ-ron trong lớp dữ liệu đầu ra R có giá trị từ 0 đến 10.



Hình 2. Các bước thành lập bản đồ hiện trạng rừng và đất lâm nghiệp từ ảnh vệ tinh Sentinel 2 cho huyện Bù Đăng, tỉnh Bình Phước

2.5. Phương pháp đánh giá độ chính xác

2.5.1. Đánh giá độ chính xác kết quả phân loại ảnh

Để đánh giá độ chính xác của kết quả phân loại, nghiên cứu đã chia ngẫu nhiên dữ liệu ban đầu thành 2 phần độc lập, một phần được dùng để tạo lập mô hình (70% dữ liệu ban đầu), còn lại 30% dữ liệu không tham gia lập mô hình dùng để đánh giá độ chính xác phân loại. Trong đó, bộ dữ liệu sử dụng cho phân loại và đánh giá được dùng chung cho cả 3 thuật toán học máy (RF, NNET và SVM).

2.5.2. Đánh giá độ chính xác mô hình huấn luyện

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 liệu 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.

Ngoài các chỉ số được đề xuất bởi Story and Congalton (1986) như độ chính xác toàn cục – OA (Accuracy) và hệ số Kappa (K), nghiên cứu còn sử dụng hệ số Precision, Recall và F1 score để so sánh độ chính xác giữa các thuật toán phân loại (RF, NNET và SVM), từ đó lựa chọn thuật toán tối ưu để tiến hành phân loại hiện trạng rừng và đất lâm nghiệp. Cuối cùng, chất lượng của mô hình phân loại hiện trạng rừng và đất lâm nghiệp tại khu vực nghiên cứu được đánh giá thông qua độ chính xác người dùng (Ua) và độ chính xác người thực hiện (Pa) (Story & Congalton, 1986).

Toàn bộ các thuật toán học máy được huấn luyện và đánh giá độ chính xác bằng gói “caret” trên phần mềm R 4.0 (R Core Team, 2021). Các bước thành lập bản đồ hiện trạng rừng và đất lâm nghiệp từ ảnh vệ tinh Sentinel 2 cho huyện Bù Đăng, tỉnh Bình Phước được thể hiện ở Hình 2.

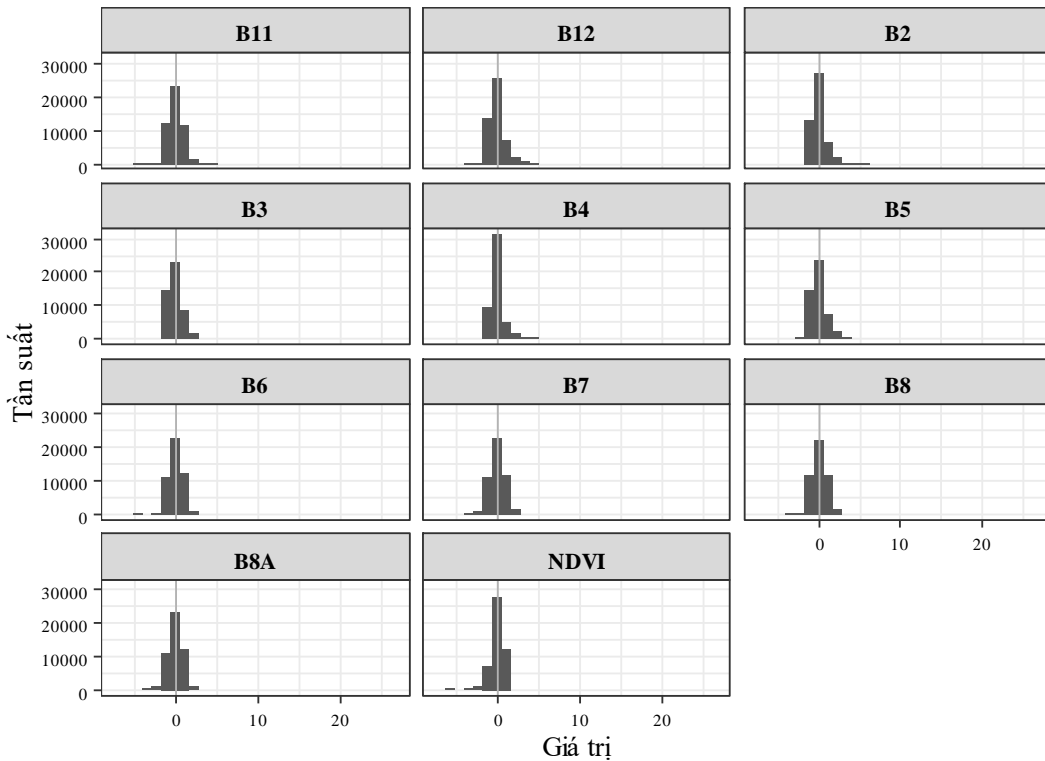
3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Chuẩn hóa các kênh ảnh và chỉ số NDVI

Giá trị phổ của các kênh ảnh và chỉ số NDVI ở các mẫu sau khi chuẩn hóa được thể hiện ở Hình 3.

Cụ thể, 2 kênh sóng ngắn hồng ngoại (B11 và B12) và kênh rìa đỏ 1 (B5) giá trị phổ sau khi chuẩn hóa nằm trong khoảng từ -5 đến 7. Kênh rìa đỏ 2 (B6), gần hồng ngoại hẹp (B7 và B8A) và gần hồng ngoại (B8) giá trị phổ từ -4 đến 4. Biến động cao nhất là

kênh xanh nước biển (B2) từ -2 đến 26, sau đó đến kênh xanh lục (B3) từ -3 đến 16 và kênh đỏ (B4) từ -2 đến 12. Đối với chỉ số NDVI sau khi chuẩn hóa giá trị từ -1 đến 1.



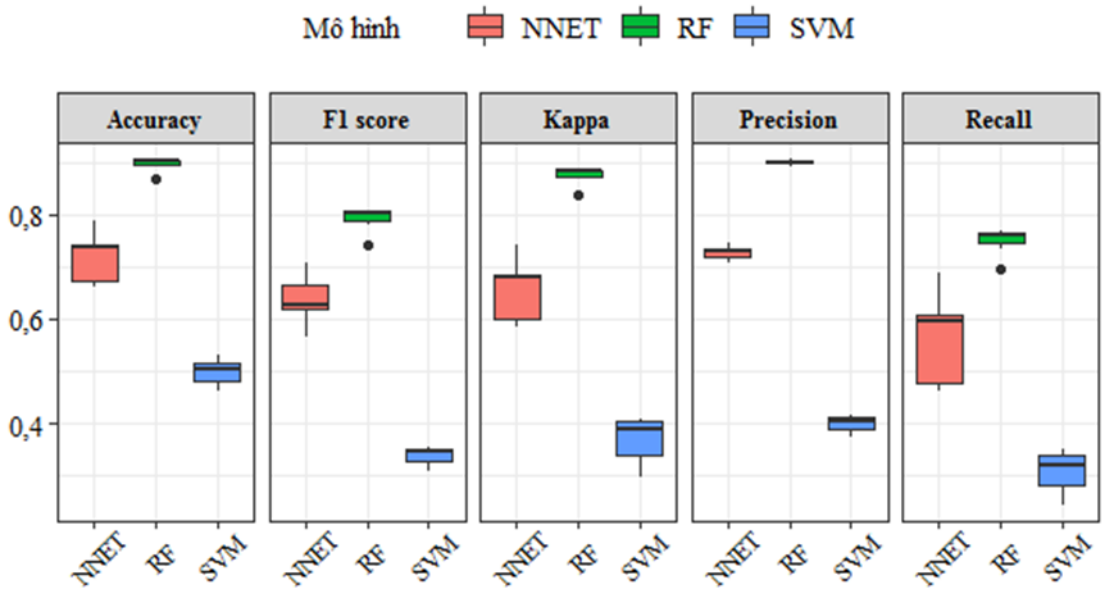
Hình 3. Phân bố giá trị phổ của các kênh ảnh và chỉ số NDVI sau khi chuẩn hóa

3.2. Lựa chọn thuật toán phân loại

Trong số các thuật toán phân loại được sử dụng trong nghiên cứu này, thuật toán rừng ngẫu nhiên (RF) ở cả 5 chỉ số (Accuracy, F1 score, Kappa, Precision và Recall) đều thể hiện cao nhất, tiếp đến là thuật toán NNET và thấp nhất ở thuật toán SVM (Hình 4). Ngoài ra, sự sai khác giữa các thuật toán có ý nghĩa thống kê với $p < 0,05$. Cụ thể, giá trị trung bình chỉ số Accuracy, F1 score, Kappa, Precision và Recall của thuật toán RF lần lượt là: 0,9; 0,79; 0,87; 0,90 và 0,75; của NNET là: 0,72; 0,64; 0,66; 0,72 và 0,57; của SVM là: 0,50; 0,34; 0,37; 0,40 và 0,31. Do vậy, nghiên cứu sử dụng thuật toán rừng ngẫu nhiên để tiến hành phân loại các hiện trạng rừng và đất lâm nghiệp tại huyện Bù Đăng, tỉnh Bình Phước.

Trong những năm gần đây, RF được sử dụng khá phổ biến trong mô hình phân loại vì tính ưu việt của nó so với các thuật toán khác do thuật toán RF

ít bị ảnh hưởng tự tương quan giữa các biến và sự đa chiều của các biến thuộc tính đa không gian (Novack et al., 2011). Lowe and Kulkarni (2015) chỉ ra rằng thuật toán RF có độ chính xác toàn cục tới 96,25% và cao hơn so với SVM, NNET lần lượt là 10% và 20%. Tương tự, độ chính xác của thuật toán RF trong các phân loại lớp phủ từ ảnh viễn thám cũng được tìm thấy trong các nghiên cứu của Immitzer et al. (2012), Lim et al. (2019), Meng et al. (2020), Hiền và ctv. (2022) với độ chính xác trên 90%. Mặc dù, độ chính xác của mô hình RF trong nghiên cứu này là 90% cao hơn so với Noi and Kappas (2018) khi phân loại 6 lớp thảm phủ khu vực đồng bằng Sông Hồng bằng ảnh vệ tinh Sentinel 2 với độ chính xác là 83,5%. Tuy nhiên, thấp hơn so với của Sothe et al. (2017) khi phân loại thảm phủ rừng cận nhiệt đới ở Đông Nam Brazil với độ chính xác từ 92,6 đến 97,4%.

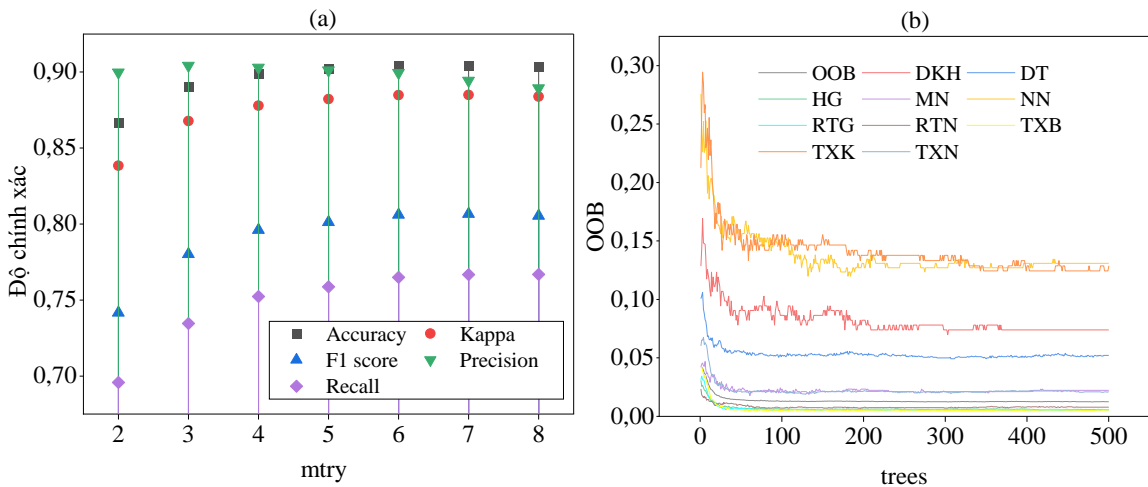


Hình 4. Độ chính xác của các thuật toán phân loại

3.3. Đặc điểm các tham số của mô hình phân loại RF

Dựa vào đánh giá lỗi sai số (OOB) và độ chính xác để lựa chọn tham số ntree và mtry tối ưu nhất cho mô hình RF. Khi cho thuật toán chạy mô hình với số ntree tăng từ 1 đến 500 cây thì tìm ra được sai số OOB của chúng. Xu hướng chung cho thấy với số ntree càng tăng thì sai số của mô hình càng thấp. Trong đó, sai số của tất cả các phân lớp tương đối

cao trong khoảng từ 0 đến 100 cây, từ 100-250 cây mức độ sai số ổn định hơn và từ 250 - 500 thì sai số nhìn chung ít biến động và ổn định (Hình 5b). Vì vậy, thông số ntree được lựa chọn là 250. Đối với tham số mtry, khi sử dụng số mtry tăng thì các tiêu chí đánh giá độ chính xác của mô hình phân loại có xu hướng tăng dần và đạt đến giá trị tối đa tại mtry bằng 6, sau đó có xu hướng giảm dần (Hình 5a). Vì vậy đề tài lựa chọn sử dụng thông số mtry cho độ chính xác cao nhất là mtry bằng 6.



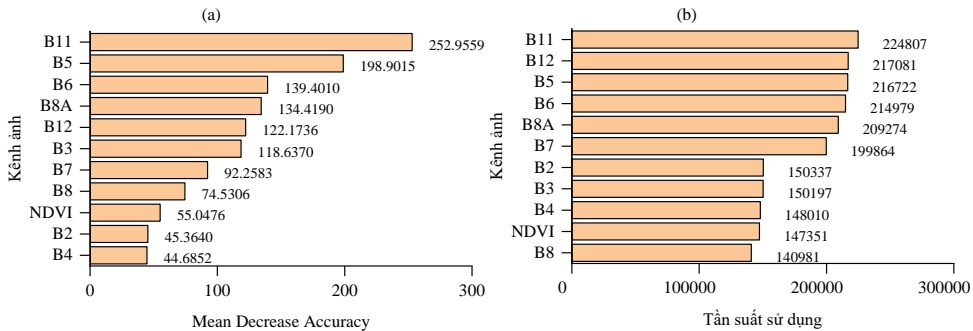
Hình 5. Lựa chọn tham số cho mô hình rừng ngẫu nhiên

3.4. Vai trò của các kênh ảnh và chỉ số NDVI trong mô hình phân loại RF

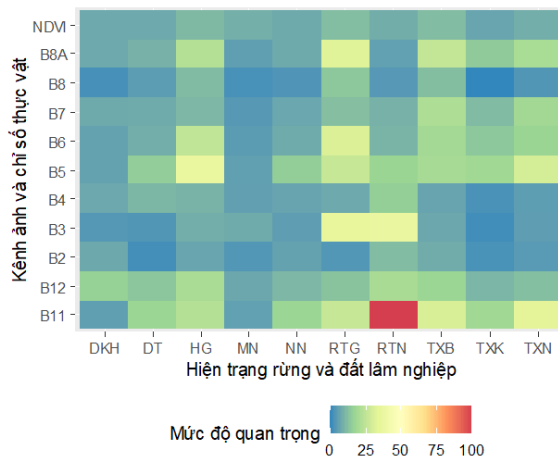
Hình 6a cho thấy kênh sóng ngắn hồng ngoại 1, 2 (kênh 11, 12), kênh rìa đỏ 1, 2 (kênh 5, 6), kênh gần hồng ngoại hẹp 2 (kênh 8A) và kênh xanh lục (kênh 3) thể hiện mức độ ảnh hưởng cao nhất đến độ chính xác của kết quả phân loại. Mặt khác, tần suất sử dụng trong mô hình phân loại RF của các kênh ảnh này cũng cao hơn với các kênh còn lại (Hình 6b). Tương tự kết quả trong nghiên cứu này, rất nhiều tác giả đã nhấn mạnh tầm quan trọng của cả vùng kênh sóng ngắn hồng ngoại và rìa đỏ đối với phân loại rừng và lĩnh vực nông lâm nghiệp và môi trường (Schultz et al., 2015; Ramoelo et al., 2015; Immitzer et al., 2016; Waśniewski et al., 2020). Đặc biệt, Waśniewski et al. (2020) cho thấy kênh 5 có tầm quan trọng nhiều hơn rất nhiều so với kênh 6 khi thành lập bản đồ rừng mưa ở Gabon và kênh 6 có độ quan trọng gấp 1.4 lần so với kênh 5. Kênh rìa đỏ ít nhạy cảm hơn với các tác động của khí quyển và đất, có thể cung cấp thông tin từ sự kết hợp của dải phổ hồng ngoại gần và dải phổ nhìn thấy. Ngoài

ra, phản xạ xung quanh rìa đỏ nhạy cảm hơn với các đặc tính của lá như hàm lượng chất diệp lục, nitrogen, LAI, và sinh khối. Trong khi đó, dải sóng gần hồng ngoại (kênh 8A) và sóng ngắn hồng ngoại (kênh 11 và 12) ảnh hưởng bởi hàm lượng nước, lignin, tinh bột và nito của lá cây (Fassnacht et al., 2016). Do vậy, kênh 5, 6, 8A, 11 và 12 có tầm quan trọng trong việc phân loại các đối tượng thực vật (Mutanga & Skidmore, 2007; Sothe et al., 2017; Waśniewski et al., 2020).

Ngược lại, kênh 2,4 và chỉ số NDVI có vai trò thấp nhất trong mô hình dự đoán và cũng có tần suất sử dụng trong mô hình phân loại thấp hơn so với các kênh ảnh còn lại. Đồng nhất với kết quả của nghiên cứu này, nghiên cứu phân loại thành phần các loài cây ở New Caledonia (Pouteau et al., 2018) và lập bản đồ biến động rừng nhiệt đới ở Brazil (Silveira et al., 2018) cũng cho thấy NDVI có mức độ quan trọng thấp. Giá trị NDVI thấp trong mô hình phân loại là do sự giống nhau về độ dày của tán giữa các hiện trạng rừng, điều này đã được tìm thấy tương tự trong kết quả nghiên cứu của Xi et al. (2021).



Hình 6. Mức độ quan trọng và tần suất sử dụng của các biến phân loại



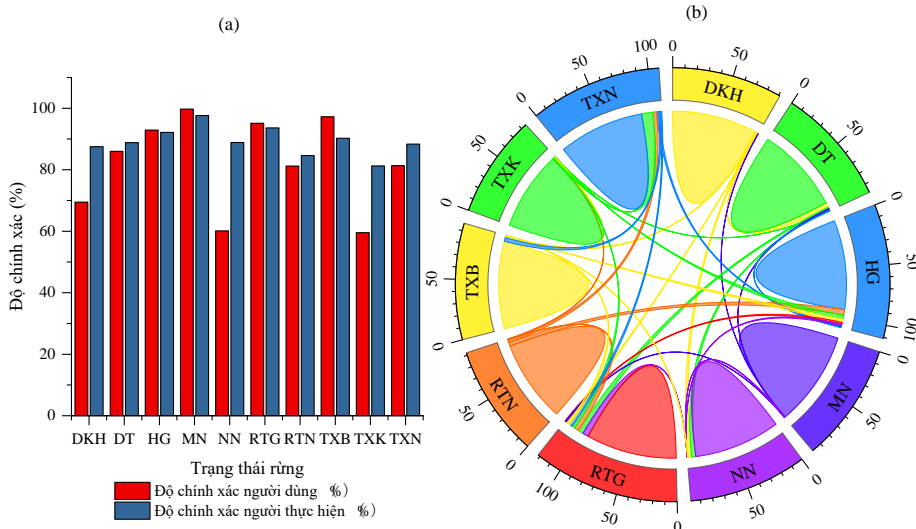
Hình 7. Tầm quan trọng của các kênh ảnh trong phân loại các lớp đối tượng

Vai trò của các kênh ảnh và chỉ số thực vật trong việc phân loại các lớp đối tượng ở Hình 7 cho thấy: kênh 11 có tầm quan trọng cao nhất đối với việc phân loại các hiện trạng rừng lá rộng thường xanh (trung bình, nghèo và kiệt), rừng tre nửa, nông nghiệp và đất trống. Trong khi đó, kênh 3 có mức độ quan trọng cao nhất đối với việc phân loại rừng trồng, kênh 5 có vai trò quan trọng trong việc phân loại hiện trạng rừng hỗn giao, chỉ số NDVI có vai trò trong việc phân loại mặt nước. Cuối cùng, kênh

12 có mức độ quan trọng cao nhất đối với phân loại lớp hiện trạng đất khác.

3.5. Độ chính xác của mô hình phân loại

Hình 8a cho thấy rằng ở mô hình phân loại các hiện trạng rừng và đất lâm nghiệp bằng mô hình RF hầu hết độ chính xác thành phần (người dùng và người thực hiện) đều đạt trên 80% ngoại trừ hiện trạng rừng thường xanh kiệt với UA=59,54% và đất khác với UA=69,47%.



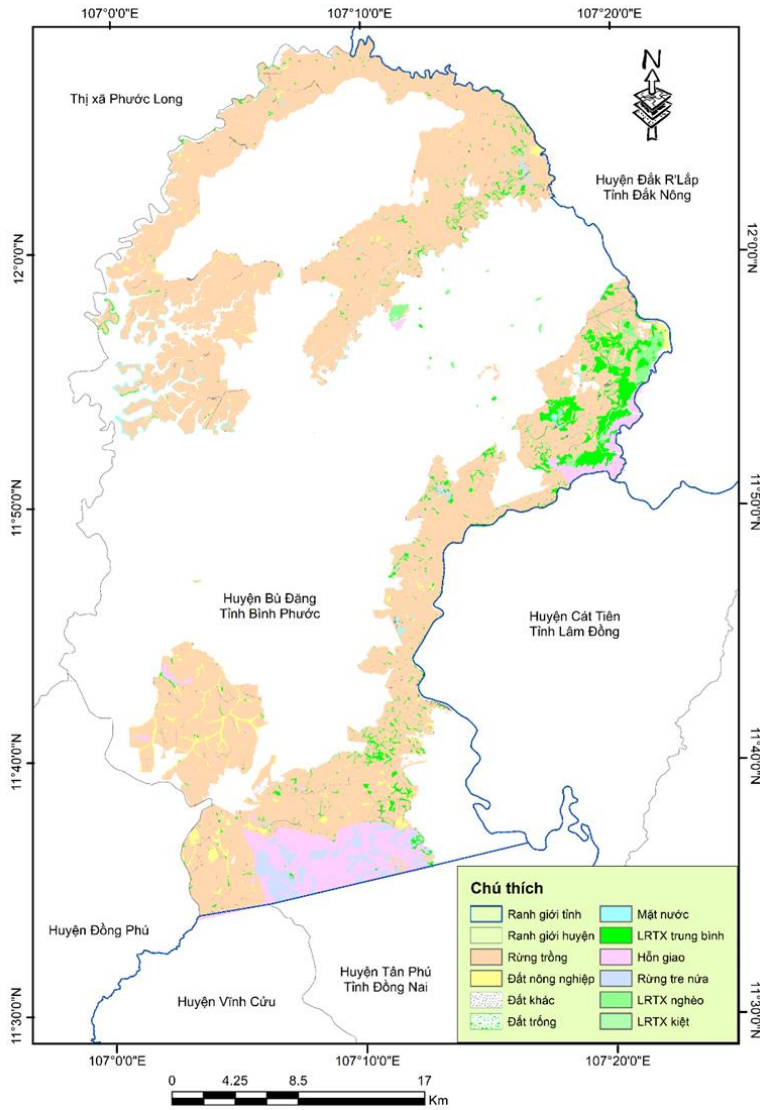
Hình 8. Độ chính xác của mô hình phân loại các lớp rừng và đất lâm nghiệp bằng phương pháp rừng ngẫu nhiên (RF)

Sự nhầm lẫn giữa các hiện trạng rừng và đất lâm nghiệp trong mô hình phân loại được thể hiện ở sơ đồ dây cung 8b. Kết quả cho thấy rừng trồng có kết quả phân loại nhầm lẫn với hầu hết các hiện trạng rừng và đất lâm nghiệp khác. Ngược lại, mặt nước chỉ nhầm lẫn với rừng trồng, đất nông nghiệp, đất trống và đất khác. Trong khi đó các hiện trạng rừng lá rộng thường xanh, rừng trồng, rừng tre nửa và rừng hỗn giao đều có sự nhầm lẫn với nhau trong mô hình phân loại. Tuy vậy, mức độ nhầm lẫn trong mô hình phân loại giữa các lớp hiện trạng rừng và đất lâm nghiệp là rất nhỏ (độ lớn của đường dây cung là hẹp). Kết quả này cho thấy rằng độ chính xác của mô hình phân loại kiểu rừng phức tạp hơn các loại lớp phủ. Điều này là do sự không đồng nhất của các loại rừng và điều kiện môi trường (Wasniewski et al., 2020).

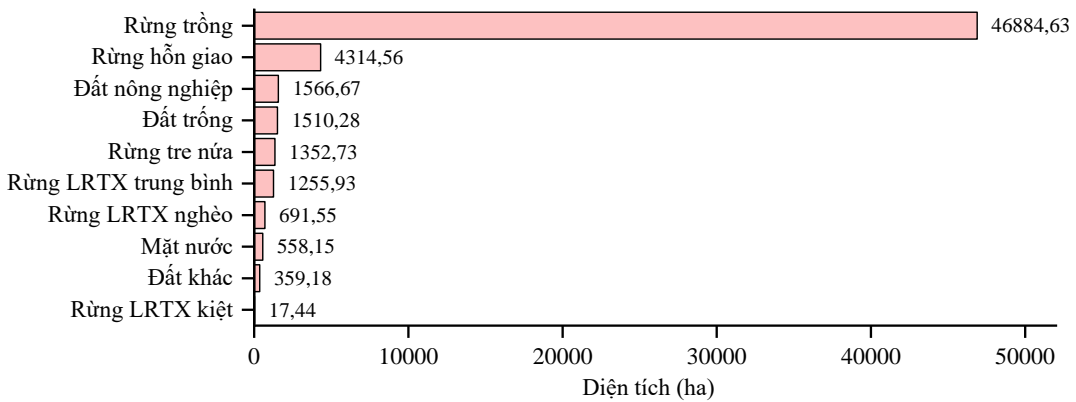
3.6. Bản đồ hiện trạng rừng và đất lâm nghiệp năm 2020 của huyện Bù Đăng, tỉnh Bình Phước

Kết quả thành lập bản đồ hiện trạng rừng và đất lâm nghiệp năm 2020 của huyện Bù Đăng, tỉnh Bình

Phước từ ảnh vệ tinh Sentinel 2 bằng thuật toán RF được thể hiện ở Hình 9 và Hình 10. Tổng diện tích rừng và đất lâm nghiệp năm 2020 của khu vực nghiên cứu là 58.511,12 (ha), với diện tích rừng trồng là 46.884,63 ha (chiếm tới 80,13%). Rừng lá rộng thường xanh chủ yếu phân bố ở khu vực giáp ranh giới với huyện Đăk R’Lấp, tỉnh Đăk Nông. Tổng diện tích hiện trạng rừng này là 1.964,92 (ha), trong đó rừng trung bình là 1.255,93 (ha), nghèo là 691,55 (ha) và kiệt là 17,44 (ha). Rừng hỗn giao (gỗ và tre nửa) và tre nửa thuần loài chủ yếu tập trung ở phía nam khu vực nghiên cứu giáp với huyện Vĩnh Cửu và Tân Phú của tỉnh Đồng Nai, với tổng diện tích là 4.314,56 và 1.352,73 (ha). Đất trồng, đất nông nghiệp, đất khác và mặt nước nằm rải rác với diện tích của từng loại là: 1.510,28; 1.566,67; 359,18 và 558,15 (ha).



Hình 9. Bản đồ hiện trạng rừng và đất lâm nghiệp năm 2020 của huyện Bu Đang, tỉnh Bình Phước



Hình 10. Diện tích hiện trạng rừng và đất lâm nghiệp năm 2020 của huyện Bu Đang, tỉnh Bình Phước

4. KẾT LUẬN

So với thuật toán SVM và NNET, thuật toán RF có độ chính xác vượt trội (OA=90%, K=0,87) trong mô hình phân loại 10 lớp hiện trạng rừng và đất lâm nghiệp tại huyện Bù Đăng, tỉnh Bình Phước từ ảnh Sentinel 2. Kết quả của nghiên cứu cho thấy tiềm năng sử dụng ảnh Sentinel 2 và thuật toán RF trong xây dựng các loại bản đồ hiện trạng rừng và đất lâm nghiệp. Ngoài ra, kênh sóng ngắn hồng ngoại (kênh 11, 12), kênh rìa đỏ (kênh 5, 6), kênh gần hồng ngoại hẹp (kênh 8A) và kênh xanh lục (kênh 3) có vai trò quan trọng trong mô hình phân loại.

Tuy nghiên cứu đã đạt được những kết quả nhất định, nhưng chỉ dừng lại trong việc sử dụng giá trị phổ các kênh ảnh trong phân loại các đối tượng mà chưa khai thác các chỉ số cấu trúc (texture measures)

của chúng như: giá trị trung bình (mean), mức biến động (variance), độ đồng nhất (homogeneity), độ tương phản (contrast),... Vì vậy, để nâng cao độ chính xác, những nghiên cứu tiếp theo nên sử dụng thêm các yếu tố cấu trúc của các kênh ảnh và mô hình số độ cao (DEM) trong phân loại kiểu rừng, đặc biệt là giữa các hiện trạng rừng lá rộng thường xanh.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Phân hiệu trường Đại học Lâm nghiệp tại Đồng Nai trong đề tài nghiên cứu khoa học của sinh viên mã số 687-QĐ-PHĐHLN. Ngoài ra, nhóm nghiên cứu xin chân thành cảm ơn Chi cục Kiểm lâm Bình Phước và Hạt kiểm lâm huyện Bù Đăng đã hỗ trợ và cung cấp dữ liệu tham chiếu trong quá trình điều tra thực địa.

TÀI LIỆU THAM KHẢO

- Apte, K. S., & Patravali, D. S. (2015). Development of back propagation neural network model for extracting the feature from a satellite image using curvelet transform. *Int. J. Eng. Res. Gen. Sci*, 3, 226-236.
- Ban quản lý rừng phòng hộ Bù Đăng. (2021). *Phương án quản lý rừng bền vững giai đoạn 2021-2030 của Ban quản lý rừng phòng hộ Bù Đăng*.
- Chavez, P. S. (1988). An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote sensing of environment*, 24(3), 459-479.
- Chen, G., & Hay, G. J. (2011). 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*, 77(7), 733-741.
- Duẩn, P. V., Doanh, L. S., Thìn, V. T., Thị, N. V., Khiên, H. V., Dũng, P. T., Tuyền, Đ. V. (2019). Đánh giá khả năng khai thác ảnh vệ tinh quang học miễn phí phục vụ giám sát lớp phủ mặt đất tại Việt Nam. *Tạp chí khoa học và công nghệ lâm nghiệp*, 3, 65-75
- Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., ... & Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186, 64-87.
- Furuya, D. E. G., Aguiar, J. A. F., Estrabis, N. V., Pinheiro, M. M. F., Furuya, M. T. G., Pereira, D. R., ... & Ramos, A. P. M. (2020). A Machine Learning Approach for Mapping Forest Vegetation in Riparian Zones in an Atlantic Biome Environment Using Sentinel-2 Imagery. *Remote Sensing*, 12(24), 4086.
- Gebhardt, S., Wehrmann, T., Ruiz, M. A. M., Maeda, P., Bishop, J., Schramm, M., ... & Schmidt, M. (2014). MAD-MEX: Automatic wall-to-wall land cover monitoring for the Mexican REDD-MRV program using all Landsat data. *Remote Sensing*, 6(5), 3923-3943.
- Hiền, L. T. D., Luân, P. H., Tuyết, H. T., & Toàn, Đ. Q. (2022). Phân loại lớp đất phủ sử dụng mô hình Random Forest kết hợp chỉ số thực vật NDVI và yếu tố địa hình: một nghiên cứu điển hình tại tỉnh Thanh Hóa, Việt Nam. *Science of The Earth & Environment*, 6(1), 507-520.
- Hình, V. T. (2012). Xây dựng biểu thể tích bằng phương pháp chặt ngã cho rừng tự nhiên ở Việt Nam. *Nhà xuất bản Nông nghiệp, Hà Nội*.
- Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (Vol. 1, pp. 278-282). IEEE.
- Hộ, P. H. (2003). *Cây cỏ Việt Nam*. Nhà xuất bản trẻ, Hà Nội.
- Hợp, T. (2002). *Cây rừng Việt Nam*. Nhà xuất bản Nông nghiệp, Hà Nội.
- Hương, N. T. T., & Trung, D. M. (2018). Áp dụng thuật toán phân loại Random Forest để xây dựng bản đồ sử dụng đất/thảm phủ tỉnh Đắk Lắk dựa vào ảnh vệ tinh Landsat 8 OLI. *Tạp chí Nông nghiệp & Phát triển nông thôn*, 13, 122-129.
- Immitzer, M., Atzberger, C., & Koukal, T. (2012). Tree species classification with random forest using very high spatial resolution 8-band WorldView-2 satellite data. *Remote sensing*, 4(9), 2661-2693.
- Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First experience with Sentinel-2 data for crop

- and tree species classifications in central Europe. *Remote sensing*, 8(3), 166.
- Izonin, I., Tkachenko, R., Shakhovska, N., Ilchyshyn, B., & Singh, K. K. (2022). A Two-Step Data Normalization Approach for Improving Classification Accuracy in the MedicalDiagnosis Domain. *Mathematics*, 10(11), 1942-1960.
- Li, M., Im, J., & Beier, C. (2013). Machine learning approaches for forest classification and change analysis using multi-temporal Landsat TM images over Huntington Wildlife Forest. *GIScience & Remote Sensing*, 50(4), 361-384.
- Lim, J., Kim, K. M., & Jin, R. (2019). Tree species classification using hyperion and sentinel-2 data with machine learning in South Korea and China. *ISPRS International Journal of Geo-Information*, 8(3), 150.
- Lowe, B., & Kulkarni, A. (2015). Multispectral image analysis using random forest. *International Journal on Soft Computing (IJSC)*, 6(1),1-13
- Luong, N. V., Tateishi, R., Kondoh, A., Anh, N. D., & Thanh, N. (2017). Land cover mapping in Yok Don National Park, Central Highlands of Viet Nam using Landsat 8 OLI images. *Vietnam Journal of Earth Sciences*, 39(4), 393-406.
- Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International journal of Remote sensing*, 28(5), 823-870.
- Meng, Y., Cao, B., Mao, P., Dong, C., Cao, X., Qi, L., ... & Wu, Y. (2020). Tree species distribution change study in Mount Tai based on Landsat remote sensing image data. *Forests*, 11(2), 130.
- Mora, A., Santos, T. M., Łukasik, S., Silva, J. M., Falcão, A. J., Fonseca, J. M., & Ribeiro, R. A. (2017). Land cover classification from multispectral data using computational intelligence tools: A comparative study. *Information*, 8(4), 147-158
- Mutanga, O., & Skidmore, A. K. (2007). Red edge shift and biochemical content in grass canopies. *ISPRS journal of Photogrammetry and Remote Sensing*, 62(1), 34-42.
- Noi, P.T, & Kappas, M. (2017). Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors*, 18(1), 1-18.
- Novack, T., Esch, T., Kux, H., & Stilla, U. (2011). Machine learning comparison between WorldView-2 and QuickBird-2-simulated imagery regarding object-based urban land cover classification. *remote sensing*, 3(10), 2263-2282.
- Nguyen, T. T. P., Mac Duy, H., Duong, T. N., & Nghiem, T. D. (2020). Forecast of Hourly Tropospheric Ozone Concentration in Quang Ninh using MLP and SVM. *VNU Journal of Science: Earth and Environmental Sciences*, 36(3), 46-54
- Pouteau, R., Gillespie, T. W., & Birnbaum, P. (2018). Predicting tropical tree species richness from normalized difference vegetation index time series: The devil is perhaps not in the detail. *Remote Sensing*, 10(5), 698.
- Puletti, N., Chianucci, F., & Castaldi, C. (2018). Use of Sentinel-2 for forest classification in Mediterranean environments. *Ann. Silv. Res*, 42(1), 32-38.
- R Core Team. (2021). *The R Project for Statistical Computing*. Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org/>.
- Ramoelo, A., Cho, M., Mathieu, R., & Skidmore, A. K. (2015). Potential of Sentinel-2 spectral configuration to assess rangeland quality. *Journal of applied remote sensing*, 9(1), 094096.
- Schultz, B., Immitzer, M., Roberto Formaggio, A., Del'Arco Sanches, I., José Barreto Luiz, A., & Atzberger, C. (2015). Self-guided segmentation and classification of multi-temporal Landsat 8 images for crop type mapping in Southeastern Brazil. *Remote Sensing*, 7(11), 14482-14508.
- Silveira, E. M., Bueno, I. T., Acerbi-Junior, F. W., Mello, J. M., Scolforo, J. R. S., & Wulder, M. A. (2018). Using spatial features to reduce the impact of seasonality for detecting tropical forest changes from Landsat time series. *Remote Sensing*, 10(6), 808.
- Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97, 105524.
- Sothe, C., Almeida, C. M. D., Liesenberg, V., & Schimalski, M. B. (2017). Evaluating Sentinel-2 and Landsat-8 data to map successional forest stages in a subtropical forest in Southern Brazil. *Remote Sensing*, 9(8), 838.
- Story, M., & Congalton, R. G. (1986). Accuracy assessment: A user's perspective: *Photogrammetric Engineering and remote sensing*, 52(3), 397-399.
- Thị, N. V., & Bảo, T. Q. (2014). Ứng dụng kỹ thuật phân loại ảnh hưởng đối tượng nhằm phân loại trạng thái rừng theo thông tư 34. *Tạp chí khoa học lâm nghiệp*, 2, 3343 - 3353.
- Trung, D. M., & Huong, N. T. (2019). Đánh giá độ chính xác của các phương pháp phân loại thảm phủ dựa vào ảnh vệ tinh Sentinel-2. *Kỷ yếu hội thảo gis toàn quốc*, 733-744.

- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, 8(2), 127-150.
- Tùng, Đ. T., Tung, N. T., Thủy, H. T., Nhàn, T. T. T., Hằng, Đ. T., Hải, V. N., & Dương N. D. (2021). Khai thác trực tuyến cơ sở dữ liệu ảnh vệ tinh, so sánh thuật toán học máy về phân loại lớp phủ trên nền google earth engine. *Kỷ yếu hội thảo khoa học quốc gia giải pháp kết nối và chia sẻ hệ thống cơ sở dữ liệu phục vụ công tác đào tạo, quản lý lĩnh vực tài nguyên và môi trường*, 1-11.
- Volke, M. I., & Abarca-Del-Rio, R. (2020). Comparison of machine learning classification algorithms for land cover change in a coastal area affected by the 2010 Earthquake and Tsunami in Chile. *Natural Hazards and Earth System Sciences Discussions*, 1-14.
- Waśniewski, A., Hościło, A., Zagajewski, B., & Moukétou-Tarazewicz, D. (2020). Assessment of Sentinel-2 satellite images and random forest classifier for rainforest mapping in Gabon. *Forests*, 11(9), 941.
- Xi, Y., Ren, C., Tian, Q., Ren, Y., Dong, X., & Zhang, Z. (2021). Exploitation of time series sentinel-2 data and different machine learning algorithms for detailed tree species classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 7589-7603.
- Shi, Y., & Zhang, J. Y. (2012, June). The application of artificial neural network model in estimation of single tree volume growth. In *2012 2nd International Conference on Remote Sensing, Environment and Transportation Engineering* (pp. 1-6). IEEE.