Tích hợp dữ liệu viễn thám quang học và radar đa thời gian trong lập bản đồ vùng trồng cà phê hỗ trợ tuân thủ EUDR: Nghiên cứu điểm tại xã Nà Tấu, tại tỉnh Điện Biên

Nguyễn Trọng Cương[1], Vũ Văn Trường[*], Nguyễn Trọng Nghĩa, Nguyễn Trọng Đạo, Đinh Hoàng Huy, Đào Thế Anh, Hoàng Thanh Sơn
Trường Đại học Lâm nghiệp

Integrating multi-temporal optical and radar remote sensing data for coffee plantation in mapping to support EUDR compliance: A case study in Na Tau commune, Dien Bien province, Vietnam

Nguyen Trong Cuong, Vu Van Truong[*], Nguyen Trong Nghia, Nguyen Trong Dao, Dinh Hoang Huy, Dao The Anh, Hoang Thanh Son
Vietnam National University of Forestry
**Corresponding author: truongvv@vnuf.edu.vn*

https://doi.org/10.55250/jo.vnuf.15.t11.2026.050-059

Thông tin chung:
Ngày nhận bài: 05/06/2026
Ngày phản biện: 03/08/2026
Ngày quyết định đăng: 03/09/2026

Từ khóa:
Google Earth Engine, lập bản đồ cà phê, Quy định EUDR, random forest, tỉnh Điện Biên, viễn thám đa nguồn.

TÓM TẮT

Trong bối cảnh Quy định chống mất rừng và suy thoái rừng của Liên minh châu Âu (EUDR) bắt đầu có hiệu lực, việc xây dựng bản đồ vùng trồng cà phê có độ tin cậy cao đóng vai trò then chốt trong quản lý và truy xuất nguồn gốc nông sản. Nghiên cứu này tích hợp dữ liệu viễn thám đa nguồn miễn phí gồm: ảnh Sentinel-2, Sentinel-1, lớp xác suất Dynamic World và mô hình số độ cao SRTM trên nền tảng Google Earth Engine để phân loại lớp phủ đất và lập bản đồ vùng trồng cà phê tại xã Nà Tấu, tỉnh Điện Biên. Tổng cộng 3.657 mẫu tham chiếu thuộc sáu lớp phủ được thu thập từ điều tra thực địa kết hợp giải đoán ảnh độ phân giải cao, trong đó 80% số mẫu dùng để huấn luyện và 20% để kiểm định độc lập. Nghiên cứu đã sử dụng 46 biến dự báo đại diện cho đặc trưng phổ, cấu trúc tán, địa hình và xác suất lớp phủ để huấn luyện mô hình Random Forest (RF). Kết quả cho thấy mô hình đạt độ chính xác tổng thể 92,1% và hệ số Kappa 0,89. Thông tin phổ đa thời gian và xác suất lớp phủ là hai nhóm đặc trưng đóng góp lớn nhất vào hiệu quả phân loại, trong khi dữ liệu radar bổ sung thông tin về cấu trúc tán giúp giảm thiểu sự nhầm lẫn giữa các lớp phủ. Phương pháp tích hợp này có tính khả thi, dễ tái lập và mở rộng quy mô, cung cấp giải pháp kỹ thuật hiệu quả hỗ trợ xây dựng cơ sở dữ liệu vùng trồng và đáp ứng các yêu cầu tuân thủ EUDR tại Việt Nam.

Keywords:
Coffee classification mapping, Dien Bien province, EUDR compliance, Google Earth Engine, multi-source remote sensing, random forest.

ABSTRACT

In the context of the European Union Deforestation Regulation (EUDR), developing highly reliable coffee plantation maps is essential for spatial management and supply chain traceability. This study integrates free, multi-source remote sensing data—including Sentinel-2 and Sentinel-1 imagery, Dynamic World probability layers, and the SRTM digital elevation model— on the Google Earth Engine platform to classify land cover and map coffee plantations in Na Tau commune, Dien Bien province. A total of 3.657 reference samples representing six land cover classes were collected through field surveys and visual interpretation of high-resolution imagery, with 80% allocated for model training and 20% for independent validation. A predictor set of 46 variables capturing spectral, canopy structural, topographic, and land cover probability characteristics was used to train a RF classifier. The model achieved an overall accuracy of 92.1% with a Kappa coefficient of 0.89. Multi-temporal spectral information and land-cover probabilities were the primary drivers of classification performance, whereas radar data provided key structural insights that mitigated spectral confusion among classes. The integrated approach offers a scalable and reproducible framework for developing digital plantation databases and supporting EUDR compliance in Vietnam.

1. ĐẶT VẤN ĐỀ

Cà phê là một trong những mặt hàng xuất khẩu có giá trị và là một trong những ngành quan trọng nhất trong ngành nông nghiệp Việt Nam [1], trong đó thị trường Liên minh châu Âu (EU) chiếm tỷ trọng lớn với khoảng 33% giá trị xuất khẩu trong tháng 5/2026 [2]. Việc EU ban hành Quy định về sản xuất và thương mại các sản phẩm hàng hóa không gây mất rừng (Quy định (EU) 2023/1115) [3] đã đặt ra yêu cầu các sản phẩm cà phê xuất khẩu phải chứng minh không liên quan đến mất rừng sau ngày 31/12/2020, đồng thời phải cung cấp thông tin truy xuất nguồn gốc đến từng vị trí tọa độ của lô sản xuất. Trong bối cảnh đó, việc xây dựng cơ sở dữ liệu không gian và bản đồ vùng trồng cà phê có độ chính xác cao trở thành yêu cầu quan trọng nhằm hỗ trợ quản lý vùng nguyên liệu, truy xuất nguồn gốc và cung cấp dữ liệu đầu vào cho quá trình đánh giá tuân thủ EUDR.

Những năm gần đây, dữ liệu viễn thám miễn phí kết hợp với nền tảng Google Earth Engine (GEE) đã được ứng dụng rộng rãi trong lập bản đồ lớp phủ đất và cây trồng trên quy mô lớn [4] . Phổ biến có các dữ liệu Sentinel, trong đó Sentinel-2 cung cấp thông tin phổ và các chỉ số thực vật [5], dữ liệu Sentinel-1 cho phép trích xuất các đặc trưng cấu trúc tán cây và không bị ảnh hưởng bởi mây che phủ [6], trong khi Dynamic World cung cấp xác suất lớp phủ đất và dữ liệu DEM phản ánh điều kiện địa hình. Nhiều nghiên cứu đã chứng minh việc kết hợp dữ liệu Sentinel-1 và Sentinel-2 cùng các thuật toán học máy, giúp nâng cao độ chính xác phân loại vùng trồng cà phê so với sử dụng từng nguồn dữ liệu riêng lẻ [7].

Tuy nhiên, các nghiên cứu hiện nay chủ yếu tập trung vào tích hợp ảnh quang học và radar, trong khi việc kết hợp đồng thời Sentinel-1, Sentinel-2, Dynamic World và các tham số địa hình trong một mô hình phân loại còn hạn chế, đặc biệt đối với khu vực miền núi phía Bắc. Phần lớn các nghiên cứu sử dụng dữ liệu đơn thời điểm hoặc chưa khai thác đầy đủ đặc trưng đa thời gian giữa mùa mưa và mùa khô để tăng khả năng phân biệt cà phê với rừng và các cây trồng lâu năm khác [8].

Nghiên cứu này tiến hành tích hợp dữ liệu Sentinel-1, Sentinel-2, Dynamic World và mô hình số độ cao SRTM để xây dựng bộ dữ liệu gồm 46 biến dự báo, kết hợp thông tin phổ, radar, lớp phủ và địa hình nhằm phân loại các lớp phủ đất và xây dựng bản đồ vùng trồng cà phê trên nền tảng GEE bằng thuật toán RF, phục vụ xây dựng cơ sở dữ liệu vùng trồng, truy xuất nguồn gốc và hỗ trợ thực hiện EUDR

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Khu vực nghiên cứu

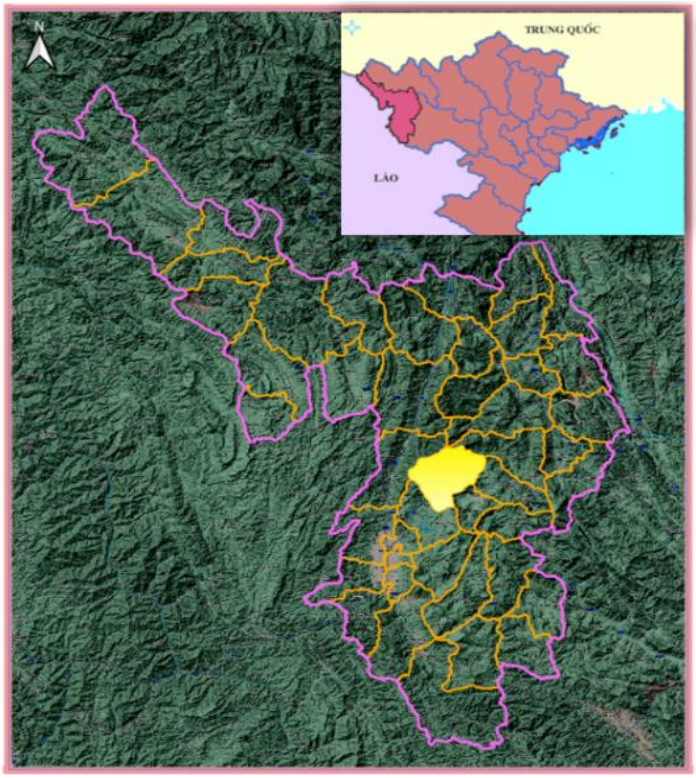
Nghiên cứu được thực hiện tại xã Nà Tấu, tỉnh Điện Biên là một trong những địa bàn trọng điểm về phát triển cây công nghiệp, đặc biệt là cà phê (Hình 1).

2.2. Dữ liệu nghiên cứu

2.2.1. Dữ liệu ảnh vệ tinh

Dữ liệu viễn thám đa nguồn được xây dựng bằng cách tích hợp ảnh quang học Sentinel-2 ảnh Sentinel-2 Level-2A (SR_HARMONIZED), ảnh radar Sentinel-1 GRD, lớp xác suất lớp phủ đất Dynamic World và các biến địa hình được trích xuất từ mô hình số độ cao SRTM (DEM). Toàn bộ ảnh viễn thám Sentinel-1 và Sentinel-2 được tách theo hai mùa, mùa khô (từ 01/4/2024 - 30/9/2025) và mùa mưa (từ 01/10/2024 - 30/4/2026), để đảm bảo đủ số lượng cảnh ảnh đặc biệt là vào mùa mưa do có nhiều mây che phủ. Có tổng số 189 ảnh Sentinel-1 và 115 ảnh Sentinel-2, 271 ảnh Dynamic World đã được lựa chọn theo từng mùa theo đúng khoảng thời gian phù hợp bằng hàm calendarRage.

Đối với ảnh Sentinel-2, ảnh tổ hợp mùa mưa được xây dựng từ các ảnh thu nhận bằng phương pháp Percentlite 25 (P_{25}), giúp ưu tiên các quan sát ổn định hơn và giảm ảnh hưởng của nhiễu khí quyển, với ngưỡng che phủ mây nhỏ hơn 40%. Ảnh tổ hợp mùa khô được xây dựng từ các ảnh thu nhận trong mùa khô năm 2025 bằng phương pháp trung vị (median) với ngưỡng che phủ mây nhỏ hơn 20%... Mây và mây ti được loại bỏ bằng lớp chất lượng QA60. Các vị trí bị thiếu dữ liệu trong ảnh mùa mưa được bổ sung bằng giá trị tương ứng từ ảnh mùa khô nhằm hạn chế hiện tượng mất dữ liệu do mây che phủ.



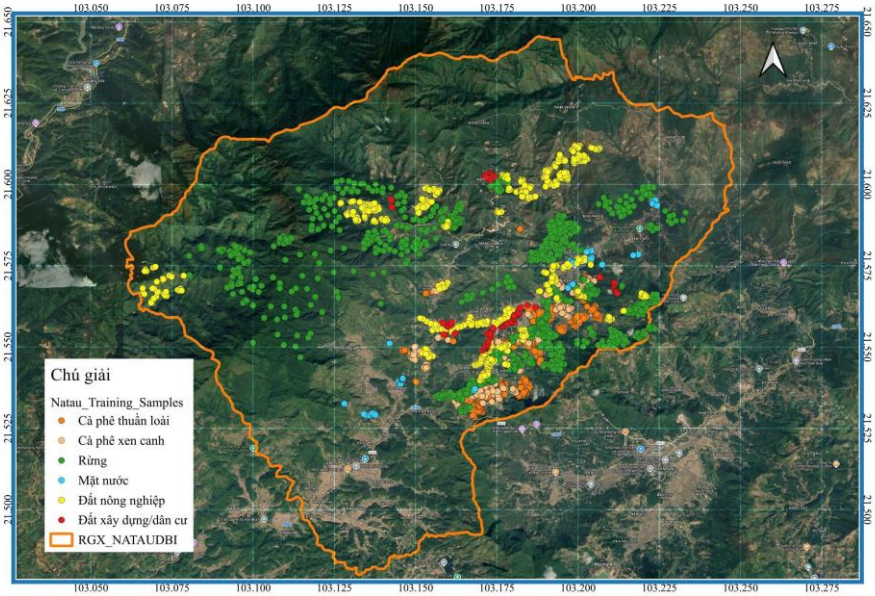
Hình 1. Vị trí khu vực nghiên cứu

Ảnh Sentinel-1 thu nhận ở chế độ Interferometric Wide Swath (IW) với hai phân cực VV và VH được xử lý riêng cho mùa mưa và mùa khô cùng thời điểm với ảnh Sentinel-2. Các lớp xác suất của Dynamic World được tổng hợp riêng cho mùa mưa và mùa khô nhằm trích xuất xác suất lớp cây gỗ (Trees) và đất canh tác (Crops). Đồng thời, các biến động xác suất giữa hai mùa cũng được tính toán để phản ánh sự thay đổi trạng thái lớp phủ. Các biến địa hình được xây dựng từ DEM SRTM bao gồm độ cao (Elevation), độ dốc (Slope), hướng phơi (Aspect), thành phần hướng Bắc (Northness), thành phần hướng Đông (Eastness), chỉ số vị trí địa hình (Topographic Position Index - TPI) và chỉ số độ gồ ghề địa hình (Terrain Ruggedness Index - TRI).

Toàn bộ các biến quang học, radar, Dynamic World và địa hình được tích hợp thành một tập dữ liệu đa nguồn gồm 46 biến dự báo gồm: B2_wet, B3_wet, B4_wet, B8_wet, B11_wet, B12_wet, B2_dry, B3_dry, B4_dry, B8_dry, B11_dry, B12_dry, NDVI_wet, NDVI_dry, NDVI_delta, NDMI_wet, NDMI_dry, NDMI_delta, NBR_wet, NBR_dry, NBR_delta, VV_wet, VV_dry, VV_delta, VH_wet, VH_dry, VH_delta, RVI_wet, RVI_dry, RVI_delta, VHVV_wet, VHVV_dry, VHVV_delta, Elevation, Slope, Aspect, Northness, Eastness, TPI, TRI, trees_wet, crops_wet, trees_dry, crops_dry, trees_delta, crops_delta. Các biến này được sử dụng để phân loại cà phê và các lớp phủ chính [8].

2.2.2. Dữ liệu điều tra thực địa

Với tổng cộng 3.657 mẫu tham chiếu được thu thập từ điều tra thực địa (tháng 5/2026) kết hợp với tạo mẫu ảnh trên Google Earth độ phân giải cao, đặc trưng cho sáu lớp phủ đất, bao gồm cà phê thuần loài (892 mẫu), cà phê xen canh (668 mẫu), các trạng thái rừng (1.089 mẫu), đất nông nghiệp (720 mẫu), mặt nước (125 mẫu) và đất xây dựng/dân cư (163 mẫu), các mẫu được lấy trải đều trong toàn bộ khu vực nghiên cứu để tránh hiện tượng quá khớp khi phân loại cũng như kiểm tra (Hình 2). Tập mẫu được chia ngẫu nhiên thành 80% mẫu huấn luyện, được cân bằng bằng phương pháp lấy mẫu ngẫu nhiên theo từng lớp, và 20% mẫu kiểm định độc lập.



Hình 2. Sơ đồ vị trí các mẫu phục vụ phân loại và đánh giá độ chính xác sau phân loại

2.3. Tiền xử lý và trích xuất đặc trưng ảnh

Các biến quang học bao gồm sáu kênh phổ Sentinel-2 (Blue, Green, Red, NIR, SWIR1 và SWIR2) cho cả hai mùa khô và mưa, cùng với các chỉ số thực vật và độ ẩm phổ biến gồm NDVI [9], NDMI [10] và NBR [11]. Các biến động theo mùa được xác định thông qua các chỉ số chênh lệch giữa mùa mưa và mùa khô (ΔNDVI, ΔNDMI và ΔNBR). Tương tự, đối với ảnh Sentinel-1 ảnh tổ hợp trung vị được xây dựng cho từng mùa, từ đó tính toán các biến radar bao gồm hệ số tán xạ ngược VV, VH, chỉ số Radar Vegetation Index (RVI) [12], tỷ số VH/VV và các chỉ số biến động theo mùa như ΔVV, ΔVH và ΔRVI. Giá trị của các chỉ số này được tính toán dựa trên ảnh trung vị (Median) trích xuất từ chuỗi ảnh vệ tinh trong suốt giai đoạn nghiên cứu, giúp đảm bảo tính đại diện cho các lớp phủ. Kết quả trích xuất tạo ra một bộ dữ liệu thuộc tính đa chiều cho mỗi pixel ảnh. Công thức toán học của các chỉ số thực vật thể hiện tại Bảng 1.

Bảng 1. Các chỉ số thực vật trích xuất từ ảnh Sentinel-2 và Sentinel-1

TT	Chỉ số	Công thức toán học	Tài liệu tham khảo
1	NDVI (Normalized Difference Vegetation Index)	$NDVI = \frac{(B8 - B4)}{(B8 + B4)}$	[9]
2	NDMI (Normalized Difference Moisture Index)	$NDMI = \frac{(B8 - B11)}{(B8 + B11)}$	[10]
3	NBR (Normalized Burn Ratio)	$NBR = \frac{(B8 - B12)}{(B8 + B12)}$	[11]
4	RVI (Radar Vegetation Index)	$RVI = \frac{4\ x\ VH}{(VV + VH\ )}$	[12]

Ghi chú: B4, B8, B11, B12, lần lượt là các kênh phổ Red, NIR, SWIR1 và SWIR2 trên ảnh Sentinel-2; VV, VH là các kênh phổ trên ảnh Radar Sentinel-1.

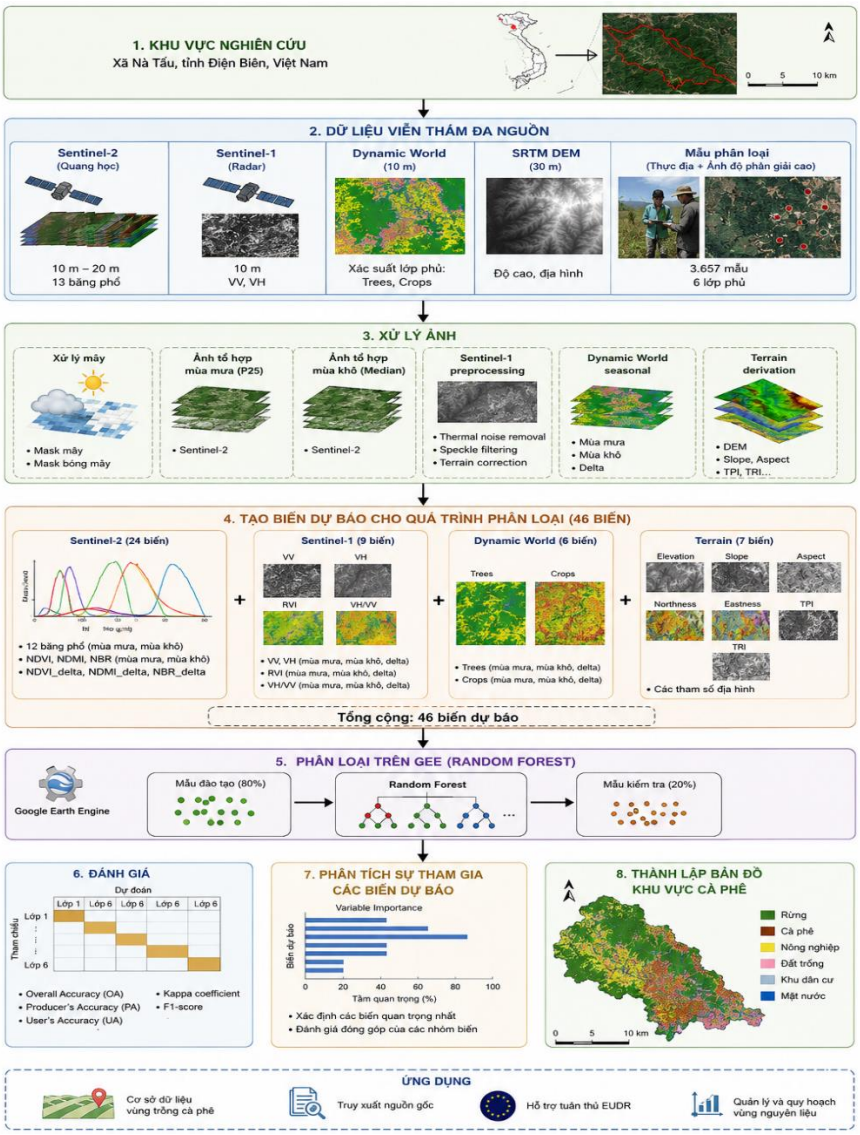
2.4. Phân loại bằng thuật toán RF và đánh giá độ chính xác

Thuật toán RF cho thấy hiệu quả hơn so với thuật toán phân loại thường được sử dụng vì có khả năng tìm ra thuộc tính nào quan trọng hơn so với những thuộc tính khác. Trên thực tế, nó còn có thể chỉ ra rằng một số thuộc tính là không có tác dụng trong cây quyết định [13] .

Trong nghiên cứu này, thuật toán RF được thực hiện trên nền tảng GEE với 500 cây quyết định, 6 biến được lựa chọn ngẫu nhiên tại mỗi nút phân tách, số lượng mẫu tối thiểu tại lá bằng 3 và tỷ lệ lấy mẫu bagging bằng 0,7. Mô hình được huấn luyện bằng tập dữ liệu gồm 46 biến dự báo và đánh giá bằng tập kiểm định độc lập. Chất lượng phân loại được đánh giá thông

qua ma trận sai số, độ chính xác tổng thể (OA), độ chính xác nhà sản xuất (PA), độ chính xác người sử dụng (UA) và hệ số Kappa. Ngoài ra, xác suất phân loại và bản đồ bất định cũng được xây dựng từ đầu ra đa xác suất của mô hình RF nhằm đánh giá mức độ tin cậy của kết quả phân loại. Sau cùng, việc kiểm chứng kết quả giải đoán các khu vực có cà phê được thực hiện thông qua 80 điểm cà phê ngoài thực địa vào tháng 5/2026 (38 điểm mẫu Cà phê thuần loài, 42 điểm cà phê xen canh) đã được thu thập độc lập với dữ liệu phục vụ giải đoán.

Toàn bộ quá trình nghiên cứu thể hiện chi tiết tại Hình 3.



Hình 3. Sơ đồ quá trình nghiên cứu

3. KẾT QUẢ NGHIÊN CỨU

3.1. Kết quả phân loại tổng thể

3.1.1. Kết quả phân loại các trạng thái sau khi huấn luyện

Kết quả đánh giá độ chính xác trên tập huấn luyện cho thấy mô hình phân loại đạt hiệu quả rất cao (Bảng 2). Độ chính xác tổng thể (OA) đạt 0,99 và hệ số Kappa đạt 0,98, chứng tỏ mức độ phù hợp giữa kết quả phân loại và dữ liệu tham chiếu là gần như hoàn hảo. Hầu hết các mẫu đều được phân loại đúng, thể hiện qua các giá trị lớn trên đường chéo chính của ma trận. Các lớp sử dụng đất như cà phê thuần, cà phê xen canh, rừng, đất nông nghiệp và dân cư đều có độ chính xác nhà sản xuất (PA) từ 0,97 - 0,99; riêng lớp mặt nước đạt PA là 0,95. Độ chính xác người dùng (UA) của các lớp cũng rất cao, dao động từ 0,98 - 1,00, cho thấy mức độ tin cậy lớn của kết quả

phân loại. Sai số nhầm lẫn giữa các lớp rất ít và chỉ xuất hiện ở một số trường hợp đơn lẻ như giữa rừng với cà phê hoặc đất nông nghiệp.

Kết quả này cho thấy mô hình đã học tốt đặc trưng phổ của các lớp và có khả năng phân biệt đối tượng rất hiệu quả trên bộ dữ liệu huấn luyện.

Bảng 2. Ma trận nhầm lẫn và các chỉ số độ chính xác phân loại của phương pháp Random Forest

Lớp	Cà phê thuần loài	Cà phê xen canh	Rừng	Mặt nước	Đất nông nghiệp	Dân cư	PA
Cà phê thuần loài	**707**	0	0	0	0	0	**1,00**
Cà phê xen canh	2	**542**	1	0	0	0	**0,99**
Rừng	3	2	**872**	0	1	0	**0,99**
Mặt nước	2	0	1	**94**	2	0	**0,94**
Đất nông nghiệp	3	0	1	0	**561**	0	**0,99**
Dân cư	2	2	0	0	1	**130**	**0,96**
UA	**0,98**	**0,99**	**0,99**	**1,00**	**0,99**	**1,00**	**OA= 0,99 Kappa = 0,98**

3.1.2. Đánh giá độ chính xác sau phân loại bằng dữ liệu kiểm tra

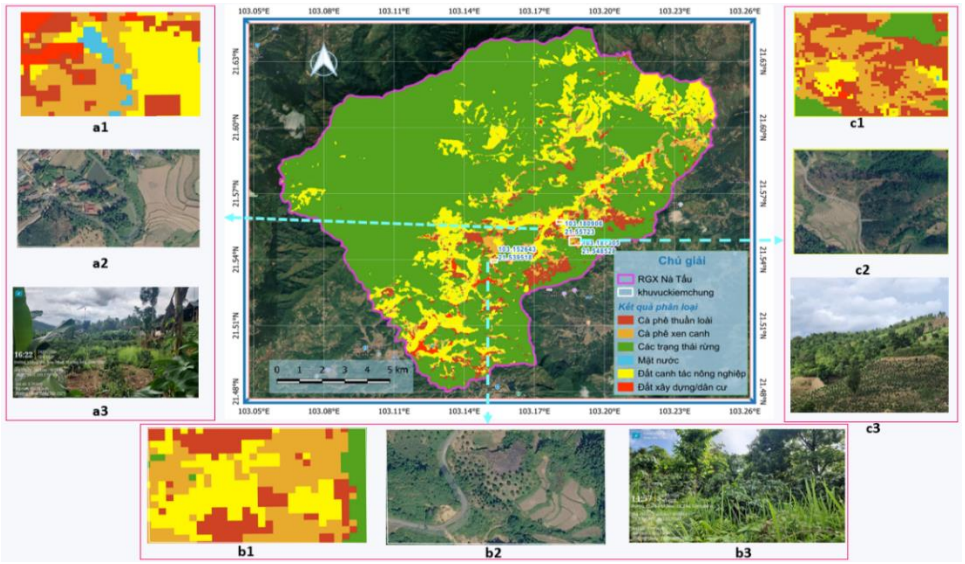
Từ 723 điểm kiểm tra của mô hình phân loại, nhóm tác giả tiến hành đánh giá độ chính xác của kết quả phân loại, ma trận đánh giá thể hiện trong Bảng 3.

Bảng 3. Ma trận nhầm lẫn và các chỉ số độ chính xác của dữ liệu kiểm chứng

Lớp	Cà phê thuần loài	Cà phê xen canh	Rừng	Mặt nước	Đất nông nghiệp	Dân cư	PA
Cà phê thuần loài	174	8	2	0	1	0	**0,94**
Cà phê xen canh	7	113	3	0	0	0	**0,91**
Rừng	1	3	201	0	1	0	**0,97**
Mặt nước	0	2	1	20	3	0	**0,76**
Đất nông nghiệp	6	1	3	0	145	0	**0,93**
Dân cư	1	2	0	0	1	24	**0,85**
UA%	**0,92**	**0,87**	**0,95**	**1,00**	**0,96**	**1,00**	**OA= 0,92 Kappa = 0,89**

Kết quả đánh giá cho thấy độ chính xác tổng thể (OA) đạt 92,1% và hệ số Kappa đạt 0,89, phản ánh mức độ phù hợp cao giữa kết quả phân loại và dữ liệu kiểm chứng. Trong đó, lớp rừng có độ chính xác cao nhất với PA đạt 0,97 và UA đạt 0,95; cà phê thuần loài, cà phê xen canh và đất nông nghiệp cũng đạt độ chính xác cao với PA lần lượt là 0,94; 0,91 và 0,93. Mặc dù lớp mặt nước và dân cư có PA thấp hơn, lần lượt đạt 0,76 và 0,85, nhưng UA của cả hai lớp đều đạt 1,00. Điều này cho thấy mô hình đã nhận dạng chính xác các đối tượng được phân loại là mặt nước và dân cư, hầu như không xảy ra sai số do gán nhầm từ các lớp khác. Tuy nhiên, một số điểm kiểm chứng thuộc hai lớp này lại bị phân loại sang đất nông nghiệp hoặc cà phê. Nguyên nhân xuất phát việc đối với khu dân cư miền núi và các ao, hồ có kích thước nhỏ hoặc bị che phủ một phần bởi thảm thực vật, khiến phổ phản xạ và đặc tính radar có sự tương đồng với các lớp phủ lân cận.

Kết quả kiểm chứng thực địa tại 80 điểm thực tế cho thấy, có 35/38 điểm cà phê thuần loài (độ chính xác thực địa đạt 92,1%) và 36/42 điểm cà phê xen canh trùng khớp giữa bản đồ phân loại và thực tế (độ chính xác thực địa đạt 85,7%) . Kết quả kiểm chứng thực địa cho phép hiệu chỉnh và nâng cao độ chính xác của bản đồ phân loại. Minh họa một số khu vực giữa kết quả phân loại và bản đồ Google Earth phát hành tháng 10/2025 và ảnh kiểm chứng ngoài thực địa thể hiện trong Hình 4.

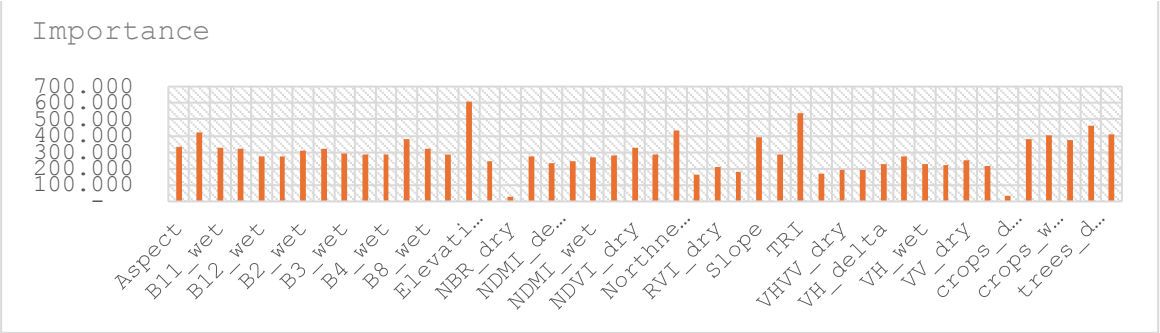


Hình 4. Kiểm chứng các kết quả phân loại trên ảnh Google Earth và ngoài thực địa

Hình 4a thể hiện kết quả phân loại giữa khu vực cà phê thuần loài và cà phê xen canh với lớp nước và đất nông nghiệp (trong đó: Hình 4a1 thể hiện kết quả phân loại của nghiên cứu; Hình 4a2 thể hiện ảnh GE; Hình 4a3 là ảnh chụp ngoài thực địa). Tương tự, Hình 4b thể hiện kết quả phân loại giữa cà phê và đất nông nghiệp; Hình 4c thể hiện kết quả phân loại giữa cà phê, đất nông nghiệp và rừng.

3.1.3. Đánh giá mức độ tham gia của các biến dự báo

Để làm rõ nguyên nhân giúp mô hình RF đạt độ chính xác cao, nghiên cứu tiến hành đánh giá mức độ quan trọng của 46 biến dự báo (Variable Importance) tại Hình 5.



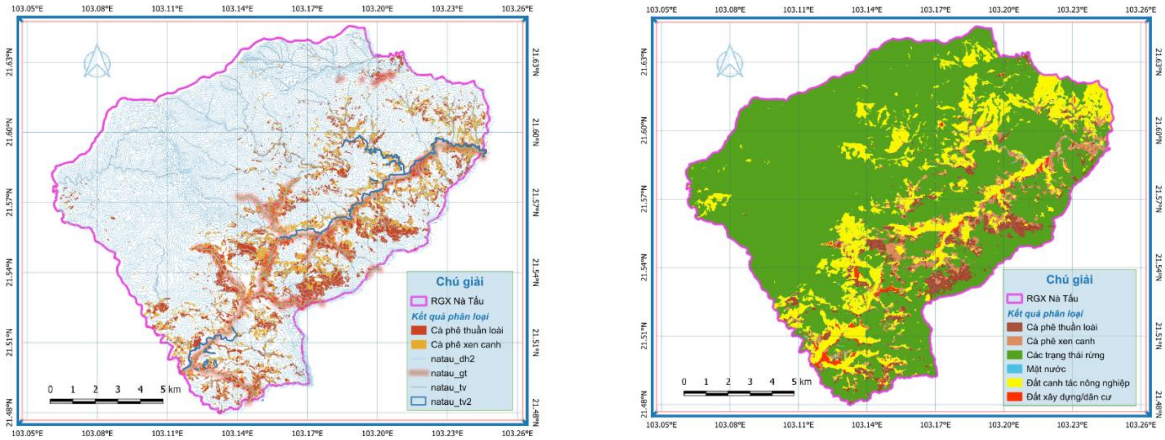
Hình 5. Mức độ tham gia của các biến dự báo trong mô hình phân loại

Kết quả cho thấy mô hình không phụ thuộc vào một nguồn dữ liệu riêng lẻ mà khai thác hiệu quả tính bổ trợ giữa dữ liệu quang học, radar, địa hình và xác suất lớp phủ. Trong đó, các biến địa hình, đặc biệt là Elevation, TRI, Slope và Northness có mức độ đóng góp cao nhất, phản ánh vai trò chi phối của điều kiện địa hình đối với sự phân bố các lớp phủ tại khu vực miền núi. Đối với dữ liệu Sentinel-2, các kênh SWIR (B11, B12), NIR (B8) cùng các chỉ số NDVI, NDMI và NBR, đặc biệt ở mùa khô và các biến động theo mùa (delta), đóng góp đáng kể trong việc phân biệt cà phê với rừng và đất nông nghiệp. Các biến Sentinel-1 (VV, VH, RVI và VH/VV) cung cấp thông tin bổ sung về cấu trúc tán và độ ẩm bề mặt, giúp giảm nhầm lẫn giữa các lớp có phổ phản xạ tương tự. Đáng chú ý, các lớp xác suất Trees và Crops của Dynamic World cũng nằm trong nhóm biến có mức độ quan trọng cao, cho thấy thông tin xác suất lớp phủ góp phần nâng cao khả năng nhận dạng các hệ thống canh tác cà phê. Kết quả này khẳng định rằng việc tích hợp đồng thời dữ liệu quang học, radar, địa hình và lớp phủ đa thời gian là yếu tố quan trọng giúp mô hình đạt độ chính xác phân loại cao trên tập dữ liệu kiểm chứng.

3.2. Xây dựng bản đồ phân bố cà phê tại khu vực nghiên cứu

Các kết quả phân loại và kiểm chứng cuối cùng để hiệu chỉnh và chồng xếp với các lớp bản đồ địa hình, bản đồ thủy văn và bản đồ giao thông để hiệu chỉnh và cho phép thành lập

được bản đồ các khu vực canh tác cà phê thuần và xen canh (Hình 6a) và bản đồ phân loại các lớp phủ (Hình 6b). Diện tích các lớp phủ chính của khu vực nghiên cứu thể hiện tại Bảng 4.



(a) **(b)**

Hình 6. Hình ảnh bản đồ các khu vực canh tác cà phê thuần và xen canh (a); kết quả phân loại các lớp phủ chính (b)

Bảng 4. Diện tích các lớp phủ tại khu vực nghiên cứu

TT	Trạng thái	Diện tích (ha)
1	Cà phê thuần loài	897,33
2	Cà phê xen canh	973,71
3	Các trạng thái rừng	14.134,70
4	Mặt nước	136,81
5	Đất nông nghiệp	3.678,99
6	Đất xây dựng/dân cư	187,86
	Tổng	**20.009,40**

Kết quả thống kê diện tích lớp phủ cho thấy các trạng thái rừng chiếm ưu thế với diện tích khoảng 14.134,70 ha, chiếm hơn 70% tổng diện tích toàn xã, đất nông nghiệp 3.678,99 ha (xấp xỉ 18,4%). Diện tích cà phê khoảng 1.871 ha (chiếm 9,35%), trong đó cà phê xen canh khoảng 973,71 ha và cà phê thuần loài khoảng 897,33 ha. Diện tích mặt nước khoảng 136,81 ha và diện tích đất xây dựng/dân cư khoảng 187,86 ha. Kết quả phân loại cũng cho thấy ở khu vực nghiên cứu cà phê phân bố chủ yếu ở các khu vực gần khu dân cư trong khi đất nông nghiệp phân bổ rải rác ở gần rừng, kể cả các khu vực xa khu dân cư và khu vực có địa hình đồi núi phức tạp.

4. THẢO LUẬN

Việc tích hợp dữ liệu viễn thám đa nguồn gồm Sentinel-1, Sentinel-2, Dynamic World và DEM giúp nâng cao đáng kể hiệu quả phân loại các trạng thái lớp phủ tại khu vực nghiên cứu. Mô hình RF đạt độ chính xác tổng thể (OA) 0,92 và hệ số Kappa 0,89, cao hơn các nghiên cứu trước đây chỉ sử dụng dữ liệu Sentinel-2 hoặc kết hợp Sentinel-1 và Sentinel-2 để lập bản đồ cà phê. Kết quả này cũng phù hợp các quan điểm về phân loại cà phê của Hunt và cộng sự (2020) [14], trong đó việc kết hợp dữ liệu quang học, SAR và các biến phụ trợ được xem là hướng tiếp cận hiệu quả nhất đối với lập bản đồ cà phê ở vùng nhiệt đới thường xuyên có mây che phủ. Đồng thời, so với nghiên cứu của Nguyễn Thị Thanh Hương và cộng sự (2020) (OA = 80.0%, K = 0.802) [15] hay của Maskell và cộng sự (2021) (OA = 72,5%, K = 0.675) [8] thì độ chính xác đạt được trong nghiên cứu này cũng cao hơn, cho thấy lợi ích rõ rệt của việc tích hợp thêm dữ liệu Dynamic World, địa hình và đặc trưng đa thời gian.

Phân tích mức độ quan trọng của các biến cho thấy hiệu quả của mô hình không đến từ một nguồn dữ liệu riêng lẻ mà chủ yếu từ sự bổ trợ giữa các nhóm biến quang học, radar, địa hình và xác suất lớp phủ. Đáng chú ý, các biến địa hình như Elevation, TRI, Slope và Northness có mức độ đóng góp cao, phản ánh vai trò chi phối của điều kiện địa hình đối với sự phân bố của Cà phê tại khu vực Điện Biên. Trong nhóm Sentinel-2, các

kênh SWIR (B11, B12) và NIR (B8) cùng các chỉ số NDVI, NDMI và NBR, đặc biệt ở mùa khô và các giá trị biến động theo mùa (delta) thể hiện khả năng mô tả tốt độ ẩm và trạng thái sinh trưởng của thảm thực vật, góp phần giảm sự nhầm lẫn giữa cà phê, rừng và đất nông nghiệp. Trong khi đó, Sentinel-1 cung cấp thông tin độc lập về cấu trúc tán và đặc tính tán xạ vi sóng, hỗ trợ phân biệt các đối tượng có phổ phản xạ tương đồng. Các lớp xác suất Trees và Crops của Dynamic World cũng nằm trong nhóm biến quan trọng, cho thấy việc sử dụng thông tin xác suất lớp phủ thay vì chỉ sử dụng nhãn phân loại giúp tăng khả năng nhận dạng các hệ thống canh tác cà phê. Kết quả này khẳng định rằng việc tích hợp đồng thời thông tin phổ, cấu trúc, địa hình và lớp phủ đa thời gian là yếu tố quyết định giúp nâng cao độ chính xác của mô hình… Các biến phổ, radar và xác suất lớp phủ đều có mức độ đóng góp đáng kể, chứng minh rằng mô hình khai thác đồng thời thông tin sinh học và thông tin ngữ cảnh để nhận dạng vùng trồng cà phê. Điều này giúp giảm nguy cơ mô hình chỉ học quy luật phân bố theo địa hình và nâng cao khả năng mở rộng sang các khu vực có điều kiện sinh thái tương tự.

Mặc dù đạt độ chính xác cao, ma trận phân loại cho thấy vẫn còn sự nhầm lẫn giữa cà phê xen canh, rừng và đất nông nghiệp. Nguyên nhân chủ yếu xuất phát từ sự tương đồng về cấu trúc tán và phổ phản xạ giữa các hệ thống canh tác xen canh với tái sinh sau nương rẫy, rừng nghèo kiệt. Đối với lớp mặt nước và dân cư, độ chính xác người dùng thấp hơn nhưng độ chính xác tổng thể rất cao cho thấy mô hình hầu như không gán nhầm các lớp khác thành mặt nước hoặc dân cư, trong khi vẫn bỏ sót một số đối tượng thực tế do kích thước nhỏ và hiện tượng pixel hỗn hợp ở độ phân giải 10 m. Điều này cho thấy sai số chủ yếu bắt nguồn từ đặc điểm của đối tượng và độ phân giải dữ liệu hơn là từ phương pháp phân loại.

Hiện nay, các mô hình học sâu như CNN, Vision Transformer (ViT), các phương pháp phân loại chuỗi thời gian như TWDTW và SITS-BERT hay các mô hình nền tảng như SatMAE và Presto đang trở thành xu hướng trong viễn thám, RF vẫn được lựa chọn trong nghiên cứu này vì phù hợp với mục tiêu và điều kiện dữ liệu. Bộ dữ liệu nghiên cứu gồm hơn 3.000 mẫu huấn luyện và 46 biến dự báo đã được trích xuất từ nhiều nguồn viễn thám, quy mô phù hợp với các thuật toán học máy dựa trên cây hơn là các mô hình học sâu đòi hỏi lượng mẫu lớn và hạ tầng GPU chuyên dụng. Hơn nữa, mục tiêu của nghiên cứu là đánh giá tiềm năng của phương pháp phân loại trực tiếp trên GEE bằng dữ liệu mở và cho kết quả nhanh chóng, trong khi các mô hình học sâu và mô hình nền tảng hiện vẫn chưa được hỗ trợ đầy đủ trên nền tảng này.

Đóng góp chính của nghiên cứu là xây dựng thành công quy trình tích hợp dữ liệu đa nguồn và đa thời gian hoàn toàn trên nền tảng GEE bằng các nguồn dữ liệu mở. Quy trình này không chỉ nâng cao độ chính xác phân loại mà còn bảo đảm tính đơn giản, khả năng tái lập và khả năng mở rộng trên phạm vi lớn. Đây là những yêu cầu quan trọng đối với các ứng dụng giám sát sử dụng đất và truy xuất nguồn gốc vùng trồng trong bối cảnh thực hiện Quy định chống mất rừng của Liên minh châu Âu (EUDR). Trong tương lai, việc mở rộng nghiên cứu bằng chuỗi ảnh đa thời gian, độ phân giải cao và các mô hình học sâu hoặc mô hình nền tảng có thể là hướng phát triển tiềm năng nhằm tiếp tục cải thiện khả năng tổng quát hóa và hỗ trợ xây dựng cơ sở dữ liệu vùng trồng cà phê phục vụ quản lý bền vững.

5. KẾT LUẬN

Nghiên cứu đã khẳng định tiềm năng của việc tích hợp dữ liệu viễn thám đa nguồn miễn phí gồm Sentinel-1, Sentinel-2, Dynamic World và dữ liệu địa hình SRTM trong phân loại lớp phủ đất và xây dựng bản đồ vùng trồng cà phê tại xã Nà Tấu, tỉnh Điện Biên trên nền tảng GEE. Việc tổ hợp 46 biến dự báo phản ánh đồng thời đặc trưng phổ, cấu trúc tán, địa hình và xác suất lớp phủ đã giúp mô hình RF đạt độ chính xác tổng thể cao (OA) 0,92 và hệ số Kappa 0,89, chứng minh hiệu quả của cách tiếp cận đa nguồn trong điều kiện địa hình miền núi và mây che phủ kéo dài. Một đóng góp quan trọng của nghiên cứu là đưa ra phương pháp xây dựng ảnh tổ hợp đa mùa sử dụng P25 cho mùa mưa và trung vị cho mùa khô, kết hợp với dữ liệu Sentinel-1 nhằm giảm ảnh hưởng của mây che phủ và nâng cao tính ổn định của dữ liệu đầu vào. Việc này không chỉ cải thiện hiệu quả phân loại mà còn có khả năng tái lập, chi phí thấp và dễ mở rộng trên nền tảng điện toán đám mây, tạo tiền đề để triển khai cho các khu vực trồng cà phê khác tại Việt Nam.

Kết quả nghiên cứu cung cấp cơ sở dữ liệu và cơ sở khoa học cho việc xây dựng cơ sở dữ

liệu không gian về vùng trồng cà phê phục vụ quản lý sản xuất, giám sát biến động sử dụng đất và truy xuất nguồn gốc tuân thủ EUDR. Tuy nhiên, bản đồ phân loại hiện mới phản ánh hiện trạng phân bố vùng trồng và chưa thể thay thế hoàn toàn cơ sở dữ liệu truy xuất nguồn gốc. Để hình thành một hệ thống truy xuất nguồn gốc hoàn chỉnh, cần tiếp tục tích hợp kết quả phân loại với dữ liệu địa chính, thông tin chủ sử dụng đất, thời gian trồng, sản lượng và cơ sở dữ liệu về biến động rừng sau năm 2020. Việc chuẩn hóa các lớp dữ liệu này sẽ cho phép xây dựng cơ sở dữ liệu số có khả năng xác định nguồn gốc từng vùng sản xuất, đánh giá rủi ro mất rừng và hỗ trợ thực hiện quy trình thẩm định (DDS) phục vụ tuân thủ EUDR.

Trong các nghiên cứu tiếp theo, cần mở rộng phạm vi nghiên cứu trên các khu vực khác ở Tây Bắc, đồng thời xây dựng chuỗi dữ liệu đa thời gian để theo dõi biến động vùng trồng và giám sát chuyển đổi sử dụng đất theo thời gian. Bên cạnh đó, việc tích hợp dữ liệu viễn thám độ phân giải cao hoặc siêu cao, ảnh UAV, dữ liệu LiDAR và các mô hình học sâu hoặc các mô hình tiên tiến khác sẽ là hướng nghiên cứu tiềm năng nhằm nâng cao khả năng phân loại cà phê xen canh và cải thiện tính tổng quát của mô hình.

Lời cảm ơn

Nghiên cứu này là sản phẩm của chương trình nghiên cứu giữa Trường Đại học Lâm nghiệp và tổ chức Forest Trends thông qua *"Mạng lưới nghiên cứu chính sách nông lâm nghiệp".* Nhóm tác giả trân trọng cảm ơn các tổ chức, cá nhân có liên quan và các nhà khoa học đã quan tâm, hỗ trợ và đóng góp ý kiến để hoàn thiện nghiên cứu.

TÀI LIỆU THAM KHẢO

[1]. Forest Trends và Vifoca (2025). Ngành Cà phê Việt nam, thực trạng và một số khía cạnh chính sách. https://www.mkresourcesgovernance.org/vi/documents/41dc8c1a-218a-4074-84a3-12d8d2f7f2ef/vn/nganh-ca-phe-viet-nam-thuc-trang-va-mot-so-khia-canh-chinh-sach.pdf.

[2]. Forest Trends (2026). Bản tin thị trường Cà phê tháng từ 4-11 tháng 6 năm 2026. https://www.mkresourcesgovernance.org/vi/documents/0df3d08c-5740-4960-8ba5-b424d82154a8/vn/ban-tin-thi-truong-ca-phe-4-1162026.pdf.

[3]. European Union (2023). Regulation (EU) 2023/1115 of the European Parliament and of the Council on the making available on the Union market and the export from the Union of certain commodities and products associated with deforestation and forest degradation. https://forestpolicy.org/policy-law/eu-regulation-deforestation-free-products-eudr.

[4]. Luis Carrasco, O'Neil Aneurin, R.Daniel Morton & Clare.S Rowland (2019). Evaluating Combinations of Temporally Aggregated Sentinel-1, Sentinel-2 and Landsat 8 for Land Cover Mapping with Google Earth Engine. Remote Sensing. 11(3): 21. DOI: 10.3390/rs11030288.

[5]. Ewa Grabska, Patrick Hostert, Dirk Pflugmacher & Katarzyna Ostapowicz (2019). Forest stand species mapping using the Sentinel-2 time series. Remote Sensing. 11(10): 1197. DOI: 10.3390/rs11101197.

[6]. Luojia Hu, Nan Xu, Jian Liang, Zhichao Li, Luzhen Chen & Feng Zhao (2020). Advancing the Mapping of Mangrove Forests at National-Scale Using Sentinel-1 and Sentinel-2 Time-Series Data with Google Earth Engine: A Case Study in China. Remote Sensing. 12(19): 3120. DOI: 10.3390/rs12193120.

[7]. Trần Quang Bảo, Nguyễn Trọng Cương, Lê Sỹ Doanh, Lã Nguyên Khang, Đinh Văn Tuyến & Lê Đức Nhật Minh (2023). Kết hợp ảnh vệ tinh Sentiel-2 và Google Earth Engine để xác định diện tích rừng chuyển đổi sang đất nông nghiệp giai đoạn 2016-2021: trường hợp nghiên cứu tại huyện Đăk Glong, tỉnh Đăk Nông. Tạp chí KH&CN Lâm nghiệp. 2: 54-64. DOI: 10.55250/jo.vnuf.2023.2.054-064.

[8]. Gina Maskell, Abel Chemura, Huong Nguyen, Christoph Gornott & Pinki Mondal (2021). Integration of Sentinel optical and radar data for mapping smallholder coffee production systems in Vietnam. Remote Sensing of Environment. 266. DOI: 10.1016/j.rse.2021.112709.

[9]. Nathalie Pettorelli (2013). The normalized difference vegetation index. Oxford University Press, USA.

[10]. Suming Jin & Steven A Sader (2005). Comparison of time series tasseled cap wetness and the normalized difference moisture index in detecting forest disturbances. Remote sensing of Environment. 94(3): 364-372.

[11]. Emanuele Alcaras, Domenica Costantino, Francesca Guastaferro, Claudio Parente & Massimiliano Pepe (2022). Normalized Burn Ratio Plus (NBR+): a new index for Sentinel-2 imagery. Remote Sensing. 14(7): 1727. DOI: 10.3390/rs14071727.

[12]. Yihyun Kim, Thomas Jackson, Rajat Bindlish, Hoonyol Lee & Sukyoung Hong (2011). Radar vegetation index for estimating the vegetation water content of rice and soybean. IEEE Geoscience and Remote Sensing Letters. 9(4): 564-568. DOI: 10.1109/LGRS.2011.2174772.

[13]. Hasan Ahmed Salman, Ali Kalakech & Amani Steiti (2024). Random Forest Algorithm Overview. Babylonian Journal of Machine Learning. 2024: 69-79. DOI: 10.58496/bjml/2024/007.

[14]. David A. Hunt, Karyn Tabor, Jennifer H. Hewson, Margot A. Wood, Louis Reymondin, Kellee Koenig, Mikaela Schmitt-Harsh & Forrest Follett (2020). Review of Remote Sensing Methods to Map Coffee Production Systems. Remote Sensing. 12(12): 2041. DOI: 10.3390/rs12122041.

[15]. Nguyen Thi Thanh Huong, Doan Trung Minh, Tomppo Erkki & McRoberts Ronald E. (2020). Land Use/Land Cover Mapping Using Multitemporal Sentinel-2 Imagery and Four Classification Methods—A Case Study from Dak Nong, Vietnam. Remote Sensing. 12(9): 1367. DOI: 10.3390/rs12091367.