

SO SÁNH HIỆU QUẢ PHÂN LOẠI THẨM PHỦ TỈNH ĐẮK LẮK TỪ DỮ LIỆU ẢNH VIỄN THÁM BẰNG THUẬT TOÁN HỌC MÁY

Hồ Đình Bảo, Nguyễn Thế Hiền, Lê Thị Mai Trang, Nguyễn Thị Thanh Hương

Khoa Nông Lâm nghiệp, Trường Đại học Tây Nguyên

TÓM TẮT

Rừng ở Đắk Lắk rất đa dạng và phong phú, nhưng đã bị suy giảm nghiêm trọng do sự mở rộng phát triển kinh tế - xã hội, dẫn đến thay đổi thảm phủ sử dụng đất. Nhằm quản lý hiệu quả sự thay đổi này, nghiên cứu sử dụng dữ liệu ảnh vệ tinh Landsat 5 và các thuật toán học máy để phân loại bốn loại thảm phủ chính, gồm: Mặt nước, Dân cư, Nông nghiệp và Rừng. Ba thuật toán được áp dụng gồm Phân loại cây hồi quy (CART), Rừng ngẫu nhiên (RF), và Máy véc-tơ hỗ trợ (SVM), với quá trình phân loại thực hiện trên nền tảng Google Earth Engine (GEE). Tổng cộng 701 điểm mẫu được thu thập để huấn luyện, kiểm tra độc lập và đánh giá độ tin cậy phân loại hiện trạng. Kết quả cho thấy, thuật toán RF đạt độ chính xác tổng thể cao nhất (92,75%) và hệ số Kappa 0,90, vượt trội so với CART (90%, Kappa 0,86) và SVM (80,7%, Kappa 0,72). RF thể hiện ưu thế trong việc phân biệt các loại thảm phủ, đặc biệt giảm nhầm lẫn giữa nông nghiệp và rừng thưa. Trong khi đó, SVM gặp khó khăn với các lớp tín hiệu quang phổ chồng chéo như nông nghiệp và dân cư. CART đạt hiệu suất gần tương đương RF, là lựa chọn đáng tin cậy. Nghiên cứu khẳng định RF là phương pháp tối ưu trong phân loại thảm phủ sử dụng đất tại Đắk Lắk, cung cấp cơ sở khoa học quan trọng cho quản lý tài nguyên rừng và phát triển bền vững.

Từ khóa: Phân loại thảm phủ, viễn thám, thuật toán học máy, tỉnh Đắk Lắk

COMPARISON OF LAND COVER CLASSIFICATION PERFORMANCE IN DAK LAK PROVINCE FROM REMOTE SENSING DATA USING MACHINE LEARNING ALGORITHMS

Ho Dinh Bao, Nguyen The Hien, Nguyen Thi Thanh Huong

Faculty of Agriculture and Forestry, Tay Nguyen University

ABSTRACT

The forests in Dak Lak are highly diverse and abundant but have been significantly degraded due to the expansion of socio-economic development, leading to changes in land use and land cover. The study utilized Landsat 5 satellite imagery and machine learning algorithms to classify four main land cover types: Water, Urban Areas, Agriculture, and Forests to effectively manage these changes. Three algorithms were applied, including Classification and Regression Tree (CART), Random Forest (RF), and Support Vector Machine (SVM), with the classification process conducted on the Google Earth Engine (GEE) platform. A total of 701 sample points were collected for training, independent testing and validation of land cover classification. The results showed that the RF algorithm achieved the highest overall accuracy (92.75%) and a Kappa coefficient of 0.90, outperforming CART (90%, Kappa = 0.86) and SVM (80.7%, Kappa = 0.72). RF demonstrated superior capability in distinguishing land cover types, particularly in reducing confusion between agriculture and sparse forests. In contrast, SVM faced difficulties with overlapping spectral signals, such as those of agriculture and urban areas. CART, with performance close to RF, proved to be a reliable alternative. The study confirmed RF as the optimal method for land cover classification in Dak Lak, providing crucial scientific evidence for sustainable forest resource management and development planning.

Keywords: Land cover classification, remote sensing, machine learning algorithm, Dak Lak province

I. ĐẶT VẤN ĐỀ

Rừng và đất đai là hai trong số những tài nguyên thiên nhiên quan trọng, không chỉ góp phần duy trì sự cân bằng sinh thái mà còn cung cấp các dịch vụ hệ sinh thái và đảm bảo sinh kế cho con người (Jenkins và Schaap, 2018). Tuy nhiên, trong những thập kỷ gần đây, các áp lực từ khai thác gỗ trái phép, chuyển đổi mục đích sử dụng đất cho nông nghiệp, và quá trình đô thị hóa đã làm suy giảm đáng kể diện tích rừng và thay đổi cấu trúc sử dụng đất (Khalid *et al.*, 2019; Phan *et al.*, 2023). Điều này đặt ra yêu cầu cấp thiết về giám sát và quản lý tài nguyên thông qua các công cụ khoa học và công nghệ tiên tiến.

Bản đồ phân bố không gian của thảm phủ và sử dụng đất (LULC) đóng vai trò nền tảng trong nhiều lĩnh vực nghiên cứu, bao gồm quản lý tài nguyên, chính sách môi trường, và quy hoạch phát triển bền vững (Esfandeh *et al.*, 2022; Yao *et al.*, 2022). Nhờ sự phát triển mạnh mẽ của công nghệ viễn thám, các nguồn dữ liệu vệ tinh như Landsat, Sentinel, và MODIS đã trở thành công cụ hữu ích để xây dựng các bản đồ LULC với độ chính xác cao (Pande, 2022). Ảnh vệ tinh cung cấp thông tin chi tiết và liên tục theo thời gian, giúp theo dõi những biến động trong thảm phủ rừng và hiện trạng sử dụng đất trên quy mô lớn (Kuang *et al.*, 2018).

Tuy nhiên, việc ứng dụng công nghệ viễn thám trong xây dựng bản đồ LULC cũng đối mặt với nhiều thách thức, bao gồm độ chính xác của dữ liệu trong các khu vực địa lý phức tạp, khả năng xử lý khối lượng dữ liệu lớn, và yêu cầu về các thuật toán phân tích hiện đại. “Các thuật toán tiên tiến, bao gồm mạng nơ-ron nhân tạo (NN), Random Forest (RF), và Support Vector Machine (SVM), đang đóng vai trò quan trọng trong việc nâng cao độ chính xác phân loại, đòi hỏi sự kết hợp dữ liệu vệ tinh với khảo sát thực

địa nhằm đảm bảo độ tin cậy (Nguyễn Thanh Tuấn *et al.*, 2022; Mutale *et al.*, 2024). Đối với các khu vực quy mô lớn, lập bản đồ sử dụng đất có độ phân giải cao cần một lượng lớn dữ liệu để phân loại trong các kỹ thuật không gian địa lý truyền thống, vì vậy, việc chọn nền tảng Google Earth Engine (GEE) để phân loại giúp việc phân loại toàn bộ khu vực rộng lớn trở nên đơn giản. Kết quả khả năng lưu trữ khổng lồ, xử lý mạnh mẽ và khả năng sử dụng nhiều kỹ thuật khác nhau đều rất cần thiết (Xie *et al.*, 2019). Với sự ra đời của Google Earth Engine, những nhu cầu này đã được đáp ứng và loại công nghệ này đã được cung cấp miễn phí cho tất cả mọi người (Gorelick *et al.*, 2017; Sidhu *et al.*, 2018; Kolli *et al.*, 2020).

Các nhân tố ảnh hưởng đến độ chính xác phân loại thường gồm loại cảm biến, nguồn dữ liệu để phân loại và đánh giá độ chính xác kết quả phân loại, số lớp và phương pháp phân loại (Nguyen, 2011; Manandhar *et al.*, 2009). Trong các yếu tố này, việc lựa chọn một thuật toán phù hợp đạt được độ chính xác phân loại có thể chấp nhận được với thời gian xử lý tối thiểu có thể rất quan trọng (Lu và Weng, 2007). Nhiều phương pháp đã được đề xuất để xây dựng bản đồ thảm phủ dựa vào ảnh vệ tinh (Waske và Braun, 2009; Li *et al.*, 2014), cả phương pháp có kiểm định và không kiểm định, cả phương pháp tham số và phi tham số.

Với các thách thức về suy thoái tài nguyên và biến đổi khí hậu tại Đắk Lắk, nghiên cứu này không chỉ cung cấp bản đồ LULC chi tiết mà còn mở ra hướng ứng dụng công nghệ số và viễn thám trong quản lý tài nguyên tại các khu vực nhiệt đới. Các kết quả nghiên cứu sẽ cung cấp cơ sở khoa học vững chắc cho việc ra quyết định trong quản lý tài nguyên và chính sách phát triển, đồng thời góp phần vào bảo tồn hệ sinh thái và đảm bảo phát triển bền vững trong tương lai.

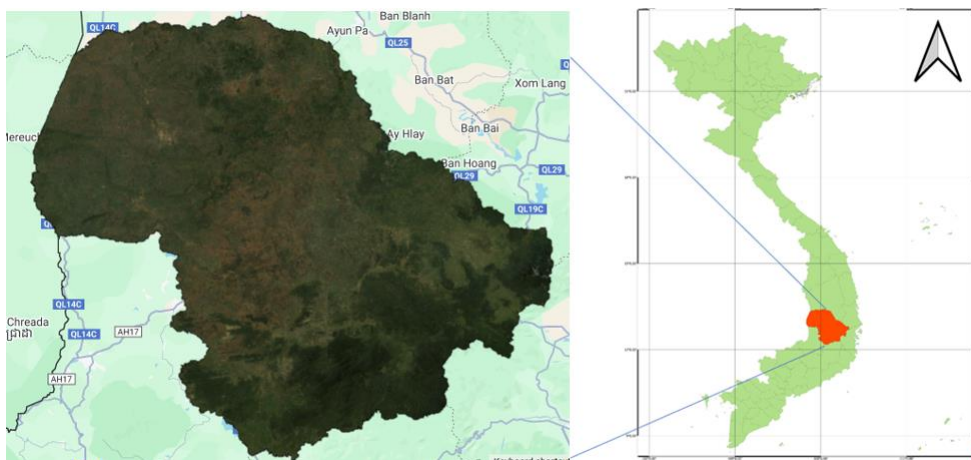
II. VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Vật liệu nghiên cứu

2.1.1. Đặc điểm khu vực nghiên cứu

Tỉnh Đắk Lắk nằm ở trung tâm vùng Tây Nguyên của Việt Nam, với tọa độ địa lý khoảng từ 107°28'57" đến 108°59'37" kinh độ Đông và từ 12°09'45" đến 13°25'06" vĩ độ Bắc (hình 1). Tỉnh có diện tích khoảng 13.085 km², chiếm khoảng 3,9% diện tích cả nước. Địa hình khu vực nghiên cứu chủ yếu là đồi núi. Độ cao ở khu vực này dao động từ 400 m đến 2.442 m (đỉnh Chư Yang Sin). Một dải cao nguyên bằng phẳng nằm ở phần giữa của tỉnh, chiếm khoảng 50% diện tích. Thời tiết ở Đắk Lắk được chia

thành hai mùa rõ rệt. Mùa mưa kéo dài từ tháng 5 đến tháng 10, kèm theo gió Tây Nam thịnh hành. Các tháng có lượng mưa cao nhất là tháng 7, tháng 8, và tháng 9, chiếm 80 - 90% tổng lượng mưa hàng năm. Tuy nhiên, ở khu vực phía Đông bị ảnh hưởng bởi dãy Trường Sơn Đông, mùa mưa kéo dài đến tháng 11. Mùa khô kéo dài từ tháng 11 đến tháng 4 năm sau. Tỉnh Đắk Lắk có tổng diện tích tự nhiên là 1.312.537 ha. Theo Quyết định công bố số liệu hiện trạng rừng và đất lâm nghiệp trên địa bàn tỉnh Đắk Lắk, đến cuối năm 2023, tỉnh Đắk Lắk có diện tích đất rừng là 497.235,2 ha, bao gồm 411.930,9 ha rừng tự nhiên và 85.304,3 ha rừng trồng. Tỷ lệ che phủ rừng đạt 38,04%.



Hình 1. Bản đồ khu vực nghiên cứu

Rừng ở Đắk Lắk được phân bố đều khắp các huyện trong tỉnh, đặc biệt là dọc biên giới của tỉnh với Campuchia. Rừng ở Đắk Lắk phong phú và đa dạng, thường có cấu trúc nhiều tầng với các loài cây, cây bụi, và thảm thực vật thân thảo đa dạng. Những khu rừng này có chức năng bảo tồn cao và chứa đựng nhiều loài có giá trị kinh tế cũng như ý nghĩa khoa học. Do điều kiện địa lý thuận lợi, khả năng tái sinh rừng ở đây có mật độ tương đối cao. Vì vậy, rừng đóng vai trò quan trọng trong việc kiểm soát xói mòn đất, điều tiết nguồn nước, giảm thiểu thiên tai và giảm thiểu tác động của biến đổi khí hậu.

2.1.2. Nguồn dữ liệu ảnh viễn thám

Sử dụng ảnh viễn thám trong quá khứ giúp theo dõi biến động sử dụng đất và lớp phủ đất theo thời gian, từ đó đánh giá tác động của con người và thiên tai đến tài nguyên. Ảnh vệ tinh cung cấp dữ liệu lịch sử đồng nhất, tiết kiệm thời gian, chi phí và hỗ trợ xây dựng cơ sở dữ liệu dài hạn. Đây là công cụ hữu ích để phân tích xu hướng, dự báo tương lai và đưa ra các giải pháp quản lý bền vững. Nghiên cứu này sử dụng ảnh vệ tinh Landsat 5 để phân loại thảm phủ sử dụng đất trong quá khứ nhằm phân tích biến động sử dụng đất, mất rừng... theo thời gian. Thời điểm

lựa chọn nghiên cứu là năm 1990 khi Việt Nam bắt đầu hội nhập với kinh tế quốc tế.

Ảnh Landsat 5 được thu thập kể từ năm 1984 đến năm 2013. Trong thời gian này, Landsat 5 đã thu thập hơn 700.000 hình ảnh phục vụ cho việc nghiên cứu, quan trắc các hiện tượng như biến đổi khí hậu, thực hành nông nghiệp, phát triển và đô thị hóa các thành phố, sự tiến hóa của hệ sinh thái, cũng như nhu cầu ngày càng tăng đối với quản lý tài nguyên thiên nhiên.

Cảm biến Landsat Thematic Mapper (TM) tạo ra các hình ảnh bao gồm sáu dải phổ với độ phân giải không gian 30 mét cho các dải 1 - 5 và 7, và một dải nhiệt (dải 6), chi tiết các kênh ảnh thể hiện ở bảng 1. Kích thước cảnh ảnh ước tính khoảng 170 km theo hướng Bắc - Nam và 183 km theo hướng Đông - Tây. Ảnh Landsat 5 TM không thể phân giải các ngôi nhà hoặc cây cối riêng lẻ, nhưng có thể ghi lại các khu vực nơi nhà cửa đã được xây dựng hoặc rừng đã bị chặt phá.

Bảng 1. Đặc điểm kênh phổ ảnh vệ tinh Landsat 5 TM (Theo USGS)

| Kênh ảnh | Tên kênh ảnh | Bước sóng (μm) | Độ phân giải (m) |
|----------|--------------------------------|-----------------------------|------------------|
| Band 1 | Blue - Xanh dương | 0,45 - 0,52 | 30 |
| Band 2 | Green - Xanh lá | 0,52 - 0,60 | 30 |
| Band 3 | Red - Đỏ | 0,63 - 0,69 | 30 |
| Band 4 | Near-Infrared - Cận hồng ngoại | 0,76 - 0,90 | 30 |
| Band 5 | Near-Infrared - Cận hồng ngoại | 1,55 - 1,75 | 30 |
| Band 6 | Thermal - Nhiệt | 10,40 - 12,50 | 120 |
| Band 7 | Mid-Infrared - Hồng ngoại giữa | 2,08 - 2,35 | 30 |

Nghiên cứu này sử dụng các ảnh vệ tinh được chụp vào năm 1990 tạo thành ảnh kết hợp không mây để loại bỏ hạn chế của loại ảnh vệ tinh quang học và sử dụng để phân loại thảm phủ nhằm đánh giá khả năng phân loại thảm phủ trong quá khứ của các thuật toán khác nhau sử dụng cho việc quản lý thay đổi thảm phủ. Sử dụng kỹ thuật mặt nạ mây (cloud masking) đối với kênh QA để loại bỏ các pixel bị ảnh hưởng bởi mây và bóng mây của các ảnh Landsat 5. Sau đó, các ảnh được ghép lại bằng cách áp dụng các phép tổng hợp thời gian bằng giá trị trung vị (Median) trên nền tảng GEE. Kết quả là một ảnh tổng hợp không mây, đại diện rõ ràng cho bề mặt đất trong khoảng thời gian thu thập ảnh.

2.2. Phương pháp nghiên cứu

Các loại ảnh quang học sử dụng miễn phí như Landsat, MODIS, Sentinel có hạn chế về độ phân giải không gian nên việc phân loại chi tiết với các loại thảm phủ/sử dụng đất là không khả

thi. Trong nghiên cứu này, đối tượng chính để phân tích giới hạn trong loại hình sử dụng đất với một số loại thảm phủ chính là Mặt nước, Dân cư, Nông nghiệp và Rừng sử dụng ảnh vệ tinh Landsat 5.

Để xây dựng phân bố không gian các loại thảm phủ, sử dụng ảnh vệ tinh đã được tiền xử lý để làm dữ liệu đầu vào cho các phương pháp phân loại. Tiến hành thu thập các điểm mẫu giải đoán, sử dụng các thuật toán phân loại và đánh giá kết quả phân loại ảnh vệ tinh để lập bản đồ phân bố không gian các loại thảm phủ.

2.2.1. Thu thập điểm giải đoán

Dữ liệu mẫu điển hình cho từng loại thảm phủ được dùng để phân loại và kiểm định trong nghiên cứu này được thu thập dựa trên các nguồn: i) Điểm mẫu được lấy trên ảnh có độ phân giải siêu cao Google Earth Engine và ảnh Landsat 5; ii) kinh nghiệm; iii) các nguồn dữ liệu có sẵn. Ít nhất có 30 điểm mẫu/kiểu thảm

phủ được thu thập để đảm bảo độ tin cậy thống kê phân loại ảnh.

Việc thu thập và giải đoán vùng mẫu được sử dụng kết hợp một cách linh hoạt và hỗ trợ cho nhau trong quá trình thực hiện. Dựa vào các dữ liệu bản đồ và số liệu sẵn có, cũng như các ảnh có độ phân giải cao và siêu cao đang có hiện nay, bộ mẫu được thu thập phân bố đều trên toàn bộ diện tích tương ứng với diện tích các loại LULC, để đảm bảo tính chính xác, tránh khoanh nhầm vào các loại khác, mẫu được thu thập tại trung tâm các loại LULC. Tùy vào điều kiện đồng nhất các lớp phủ, số lượng mẫu sẽ khác nhau. Có 701 điểm mẫu được phân bố trên toàn diện tích của tỉnh được sử dụng để thực hiện việc phân loại và kiểm định kết quả phân loại cho các lớp phủ.

2.2.2. Các thuật toán phân loại ảnh

Sau khi thu thập dữ liệu huấn luyện và tập hợp các thuộc tính lưu trữ nhãn lớp đã biết cùng các giá trị số của các biến dự đoán, bước tiếp theo là khởi tạo các bộ phân loại và định nghĩa các tham số của chúng. Trên nền tảng Google Earth Engine (GEE), có gói “Classifier” quản lý việc phân loại có giám sát bằng các thuật toán học máy. Trong nghiên cứu này, chúng tôi đã sử dụng 3 bộ phân loại ảnh gồm Phân loại cây hồi quy (Classification and Regression Tree - CART), Rừng ngẫu nhiên (Random Forest - RF) Máy véc-tơ hỗ trợ (Support Vector Machine - SVM). Trong bước này, chúng tôi huấn luyện các thuật toán trên dữ liệu huấn luyện đã xây dựng, sau đó áp dụng các thuật toán này để phân loại tạo ra bản đồ thảm phủ của khu vực nghiên cứu.

a) Phân loại cây hồi quy (Classification and Regression Tree - CART)

Thuật toán CART là một thuật toán học máy có giám sát được áp dụng trong hệ thống phân loại dựa trên cây quyết định (Decision Tree) được sử dụng các mẫu huấn luyện để xác định, nhận dạng và phân loại đối tượng trên ảnh viễn thám.

CART được sử dụng phổ biến trong phân loại ảnh viễn thám và còn được gọi là cây phân loại và hồi quy (Breiman, 1984).

CART là một thuật toán mạnh mẽ trong học máy, phù hợp cho cả phân loại và hồi quy. Điểm nổi bật của CART là khả năng chia nhỏ dữ liệu một cách có hệ thống để tối ưu hóa kết quả dự đoán. Cấu trúc cây dễ hiểu và dễ triển khai, đặc biệt trong các bài toán cần đưa ra quyết định logic hoặc trực quan.

Trong nghiên cứu này tập huấn luyện được trích ngẫu nhiên 80% điểm mẫu trong tổng số 701 điểm mẫu sử dụng hàm Random cho 04 loại thảm phủ chính là Mặt nước, Dân cư, Nông nghiệp và Rừng. Thuật toán CART được áp dụng với số lượng nút lá tối đa trên mỗi cây quyết định là không giới hạn (null) và số lượng lá tối thiểu là 01 ứng với các nút mà tập huấn luyện chứa ít nhất 01 điểm dữ liệu thuộc lớp này.

b) Phân loại rừng ngẫu nhiên (Random Forest - RF)

Thuật toán RF được giới thiệu bởi Breiman vào năm 2001, là một phương pháp học máy tổ hợp, tạo ra một “rừng” cây quyết định (ntree) để cải thiện độ chính xác và tính ổn định trong dự đoán. Thuật toán này được sử dụng rộng rãi trong phân loại, hồi quy và chọn lọc đặc trưng nhờ khả năng xử lý dữ liệu lớn, dữ liệu có nhiều chiều và các mối quan hệ phức tạp giữa các biến.

Phân loại RF yêu cầu lựa chọn ba tham số: ntree (số lượng cây cần phát triển), mtry (số lượng biến để phân tách mỗi nút), và tầm quan trọng của biến (số lượng biến/dải phổ ảnh hưởng đến hiệu suất của mô hình). Nguyễn Thị Thanh Hương và đồng tác giả (2018) chỉ ra rằng trong phạm vi $1 \leq ntree \leq 500$, $ntree = 300$ cho thấy kết quả phù hợp nhất. Ngoài ra, Breiman (2001) cũng lưu ý rằng việc sử dụng số lượng cây lớn hơn số cần thiết có thể không cần thiết, nhưng cũng không gây hại, vì mối quan hệ giữa độ chính xác và ntree có tính chất tiệm cận. Trong nghiên cứu này, thuật toán RF được sử dụng với $ntree = 100$, số lượng biến được sử dụng cho

mỗi lần phân tách là căn bậc hai của số lượng biến (4), không giới hạn số lượng nốt lá trên mỗi cây và số lượng lá tối thiểu là 01 ứng với các nốt mà tập huấn luyện chứa ít nhất 01 điểm dữ liệu thuộc lớp này.

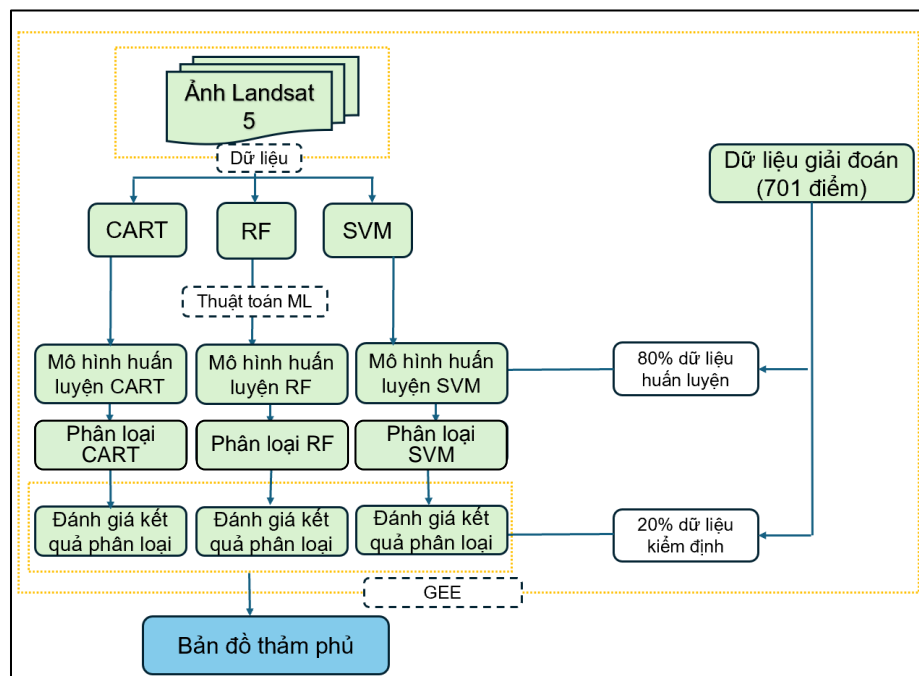
c) Phân loại Support Vector Machine

Support Vector Machines (SVM), được Vapnik phát triển vào năm 1979 và cải tiến với khái niệm siêu phẳng lề mềm năm 1995, là một phương pháp phân loại mạnh mẽ và linh hoạt. Ban đầu là bộ phân loại nhị phân, SVM tối ưu hóa khoảng cách phân tách giữa hai lớp bằng siêu phẳng (Vapnik, 2006), giúp xử lý dữ liệu không thể phân tách hoàn toàn (Tso và Mather, 2009). Hiện nay, SVM ngày càng phổ biến trong lĩnh vực viễn thám nhờ hiệu suất cao và khả năng cải thiện phân loại so với các phương pháp truyền thống như maximum likelihood.

Trong nghiên cứu này, SVM được sử dụng để phân loại dựa tập huấn luyện, các dữ liệu sẽ được phân loại theo bộ phân loại nhị phân. Các giá trị có được gán là 0 hoặc 1. Đầu ra của bộ phân loại của một lớp sẽ khớp với dữ liệu được xác định thuộc lớp đó.

d) Phân tích dữ liệu

Trong nền tảng Google Earth Engine (GEE), các bản đồ phân loại LULC (Land Use Land Cover) được tạo ra bằng cách sử dụng ảnh vệ tinh Landsat 5. Các thuật toán phân loại CART, RF và SVM được sử dụng để phân loại có giám định ảnh vệ tinh thành các loại thảm phủ khác nhau dựa trên các điểm mẫu đầu vào để huấn luyện cho 03 thuật toán áp dụng trong nghiên cứu. Các điểm ảnh bị nhiễu do mây che phủ đã được loại bỏ khỏi tất cả các ảnh bằng phương pháp mặt nạ đám mây và tạo ảnh kết hợp được cung cấp trên nền tảng GEE (Rahman *et al.*, 2020). Các mẫu huấn luyện và kiểm tra được tạo trên nền tảng GEE để huấn luyện các thuật toán phân loại. Theo Thomas và đồng tác giả (2004), mỗi lớp phân loại cần có tối thiểu 50 mẫu huấn luyện. LULC được chia thành bốn lớp chính: Mặt nước, Dân cư, Nông nghiệp và Rừng. Nghiên cứu này sử dụng 561 điểm mẫu phân loại để sử dụng làm tập huấn luyện. Trung bình 140 điểm mẫu huấn luyện cho một lớp phân loại. Quy trình nghiên cứu được thể hiện theo sơ đồ nghiên cứu ở hình 2.



Hình 2. Sơ đồ các bước nghiên cứu

e) Đánh giá kết quả phân loại

Đánh giá độ chính xác của kết quả phân loại được xem là bước quan trọng trước khi kết quả phân loại được sử dụng cho các phân tích tiếp theo (Congalton và Green, 1999; Nguyễn Thị Thanh Hương, 2011). Sử dụng các chỉ tiêu đánh giá như hướng dẫn của Congalton và Green (1999) như: sử dụng bảng ma trận để đánh giá độ chính xác dựa trên tiêu chí độ chính xác chung (Overall accuracy - OA), độ chính xác người sản xuất (Producer accuracy - PA) và độ chính xác người sử dụng (User accuracy - UA). Ngoài ra, việc đánh giá mức độ quan hệ giữa thực tế và kết quả phân loại còn dựa vào chỉ số Kappa. Hệ số Kappa dao động trong phạm vi từ 0 đến 1 và biểu thị sự giảm theo tỷ lệ về sai số được thực hiện bằng một yếu tố phân loại hoàn toàn ngẫu nhiên.

Việc đánh giá độ chính xác cho từng lớp phủ và của toàn bộ ảnh phân loại được thực hiện dựa trên bảng ma trận sai số theo các tiêu chí đề nghị của Congalton và đồng tác giả (1999) như sau:

Độ chính xác tổng thể:

$$OA = \frac{\sum_{i=1}^r X_{ii}}{N} \times 100$$

Độ chính xác người sử dụng:

$$UA = \frac{X_{ii}}{X_{i+}} \times 100$$

Độ chính xác người sản xuất PA:

$$PA = \frac{X_{ii}}{X_{+i}} \times 100$$

Hệ số Kappa (K):

$$K = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \times x_{+i})}$$

- Trong đó:* N là tổng số pixel lấy mẫu;
 r là số lớp đối tượng phân loại;
 x_{ii} là số pixel đúng trong lớp thứ i ($i = 1, 2, \dots, r$);
 x_{i+} là tổng pixel lớp thứ i của mẫu (loại thực tế/tổng giá trị theo hàng);
 x_{+i} là tổng pixel của lớp thứ i sau phân loại (loại giải đoán/tổng giá trị theo cột).

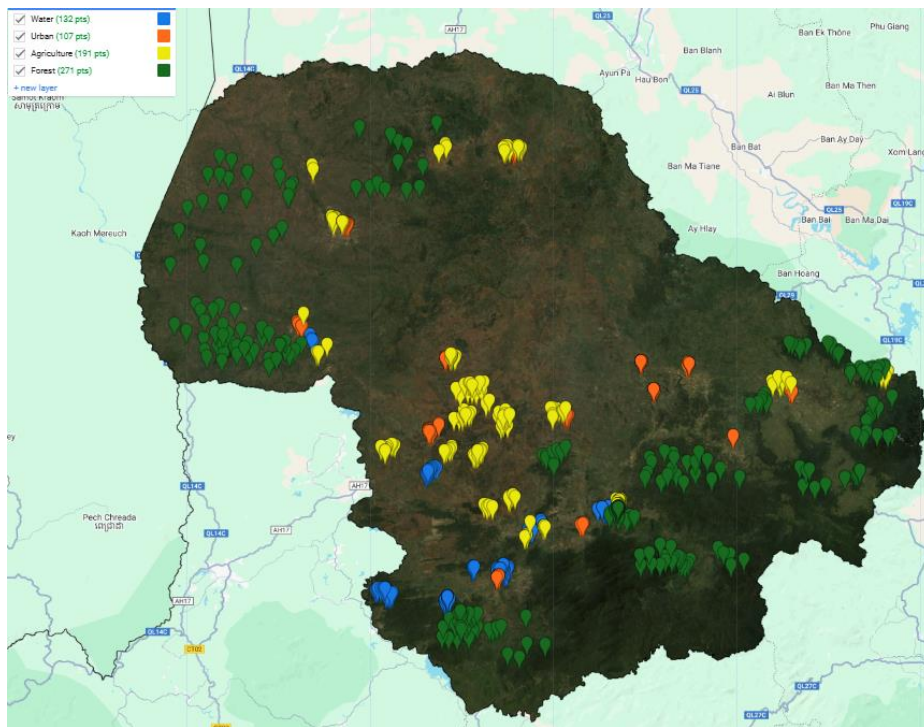
Trong nghiên cứu này, 20% tổng số điểm mẫu độc lập trong tổng số 701 điểm được sử dụng để đánh giá độ chính xác của kết quả phân loại cho từng thuật toán phân loại khác nhau, 80% điểm mẫu được dùng để huấn luyện (Ouchra *et al.*, 2023). Dựa trên các tiêu chí đánh giá độ chính xác của các thuật toán phân loại khác nhau để xác định phương pháp tốt nhất để phân loại các loại thảm phủ bằng ảnh vệ tinh Landsat 5 phục vụ cho các phân tích tiếp theo liên quan đến thảm phủ sử dụng đất, đặc biệt là thảm phủ rừng.

III. KẾT QUẢ VÀ THẢO LUẬN

3.1. Thu thập mẫu giải đoán, xử lý ảnh viễn thám

Dựa trên ảnh vệ tinh, bản đồ kiểm kê rừng hàng năm đã thu thập được 701 điểm giải đoán thuộc 04 lớp thảm phủ và sử dụng đất trong đó có lớp Mặt nước (165), Dân cư (150), Nông nghiệp (242) và Rừng (305). Các điểm mẫu được chọn đại diện cho từng kiểu thảm phủ trong nền tảng Google Earth Engine với 80% số lượng điểm mẫu tương ứng với 561 điểm được sử dụng cho phân loại và 20% còn lại (140 điểm) được sử dụng để đánh giá độ chính xác của kết quả phân loại một cách độc lập. Phân bố các điểm mẫu thể hiện ở hình 3 với các điểm màu xanh lam tương ứng với mặt nước, màu đỏ ứng với dân cư, màu vàng là đất nông nghiệp và màu xanh lá thể hiện các điểm mẫu là rừng.

Ảnh vệ tinh Landsat 5 chụp năm 1990 được xử lý loại bỏ mây và tạo bộ ảnh kết hợp giữa các cảnh ảnh theo thời gian nhằm tạo ảnh không mây. Do tính chất thảm phủ bị ảnh hưởng bởi mùa mưa và mùa khô đặc biệt là thảm phủ rừng khô cây họ Dầu bị rụng lá vào mùa khô dễ gây nhầm lẫn với các lớp thảm phủ khác nên thời gian thu thập ảnh được giới hạn trong mùa mưa để tránh nhầm lẫn trong quá trình phân loại.



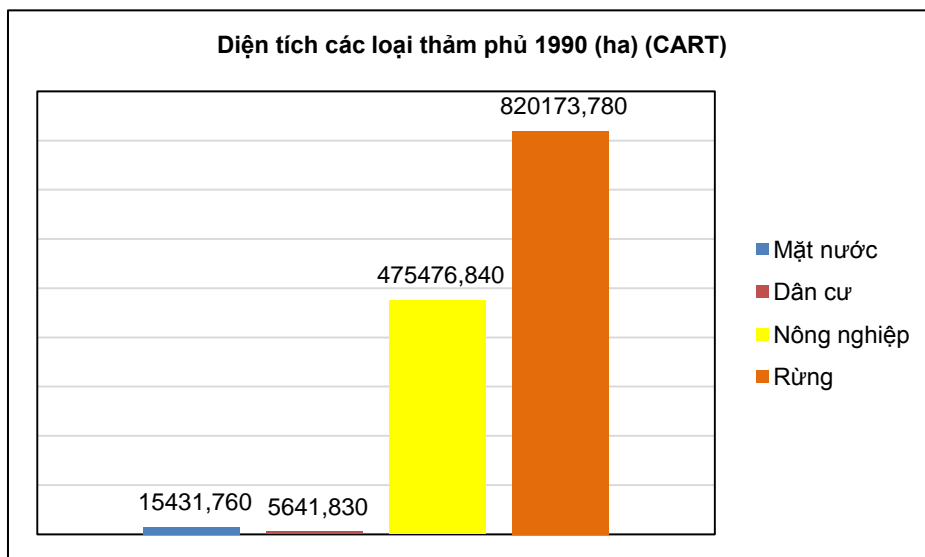
Hình 3. Vị trí các điểm mẫu phân loại ảnh

3.2. Kết quả phân loại thảm phủ

3.2.1. Phân loại thảm phủ bằng thuật toán CART

Sử dụng 80% điểm mẫu phân loại ở trên để tiến hành phân loại cho 4 lớp phủ chính bằng ảnh vệ tinh Landsat 5 ở thời điểm năm 1990 cho thấy, với độ chính xác toàn cục là 90% và hệ số

Kappa = 0,86, thuật toán CART cho kết quả phân loại có độ chính xác cao và có ý nghĩa về mặt thống kê. Diện tích các loại thảm phủ lần lượt là: Mặt nước 15.431,76 ha (chiếm 1,17%), Dân cư 5.641.83 ha (chiếm 0,43%), Nông nghiệp 475.476,84 ha (chiếm 36,11%) và Rừng 820.173,78 ha (chiếm 62,29%) (hình 4).



Hình 4. Diện tích các loại thảm phủ theo thuật toán CART

Dựa trên kết quả ma trận sai số phân loại ảnh sử dụng thuật toán CART (bảng 2) cho thấy, hiệu suất phân loại tổng thể khá cao với hệ số Kappa đạt 0,86, phản ánh sự phù hợp đáng kể giữa kết quả phân loại và dữ liệu thực tế. Độ chính xác của người sử dụng (User's Accuracy - UA's) dao động từ 0,83 đến 1, trong đó lớp Mặt nước đạt độ chính xác cao nhất (UA = 1), tiếp theo là Rừng (UA = 0,91), Dân cư (UA = 0,89),

và thấp nhất là Nông nghiệp (UA = 0,83). Về độ chính xác của nhà sản xuất (Producer's Accuracy - PA's), lớp Rừng đạt giá trị cao nhất (PA = 0,97), trong khi lớp Nông nghiệp có PA thấp nhất (0,77), cho thấy sự nhầm lẫn đáng kể giữa lớp này với các lớp khác, đặc biệt là lớp Nông nghiệp với các lớp khác đặc biệt là rừng Khộp thưa cây họ Dầu bị rụng lá vào mùa khô.

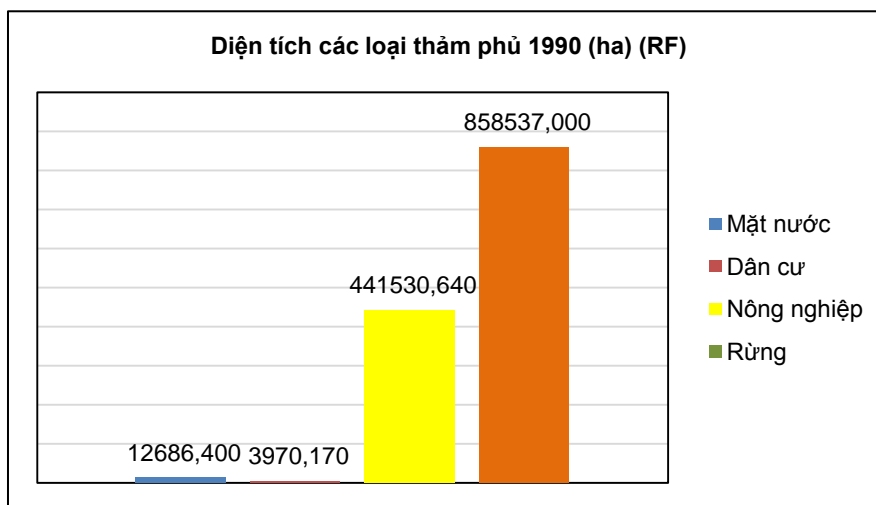
Bảng 2. Ma trận đánh giá độ chính xác và mức độ lẫn giữa các lớp phân loại bằng thuật toán CART

| | Mặt nước | Dân cư | Nông nghiệp | Rừng | UA's |
|------------------------------|----------|--------|-------------|------|-------------|
| Mặt nước | 23 | 0 | 0 | 0 | 1,00 |
| Dân cư | 0 | 24 | 3 | 0 | 0,89 |
| Nông nghiệp | 1 | 3 | 30 | 2 | 0,83 |
| Rừng | 0 | 0 | 6 | 58 | 0,91 |
| PA's | 0,96 | 0,89 | 0,77 | 0,7 | 0,90 |
| Độ chính xác toàn cục | | | | | 0,90 |
| Hệ số Kappa | | | | | 0,86 |

3.2.2. Phân loại thảm phủ bằng thuật toán Random Forest

Phân loại bốn loại thảm phủ chính bằng thuật toán RF cho thấy có độ chính xác toàn cục là 92,75% và hệ số Kappa = 0,90. Diện tích các loại thảm phủ lần lượt là: Mặt nước 12.686,40

ha (chiếm 0,96%), Dân cư 3.970,17 ha (chiếm 0,30%), Nông nghiệp 441.530,64 ha (chiếm 33,53%) và Rừng 858.537,00 ha (chiếm 65,20%). Thống kê diện tích các loại thảm phủ theo phân loại RF được thể hiện ở hình 5.



Hình 5. Diện tích các loại thảm phủ theo thuật toán RF

Phương pháp phân loại ảnh bằng thuật toán RF có hiệu suất phân loại tổng thể cao với hệ số Kappa đạt 0,88, phản ánh sự phù hợp đáng kể giữa kết quả phân loại và dữ liệu thực tế. Độ chính xác của người sử dụng (User's Accuracy - UA's) dao động từ 0,81 đến 1 (bảng 3), trong đó lớp Mặt nước đạt độ chính xác cao nhất (UA = 1,00), tiếp theo là Dân cư (UA = 0,96), Rừng

(UA = 0,92), và thấp nhất là Nông nghiệp (UA = 0,81). Về độ chính xác của nhà sản xuất (Producer's Accuracy - PA's), lớp Mặt nước vẫn đạt giá trị cao nhất (PA = 1,00), trong khi lớp Nông nghiệp có PA thấp nhất (0,83), cho thấy vẫn có sự nhầm lẫn đáng kể giữa lớp này với các lớp khác, đặc biệt là lớp Nông nghiệp với các lớp khác nhất là rừng Khộp thưa cây họ Dầu.

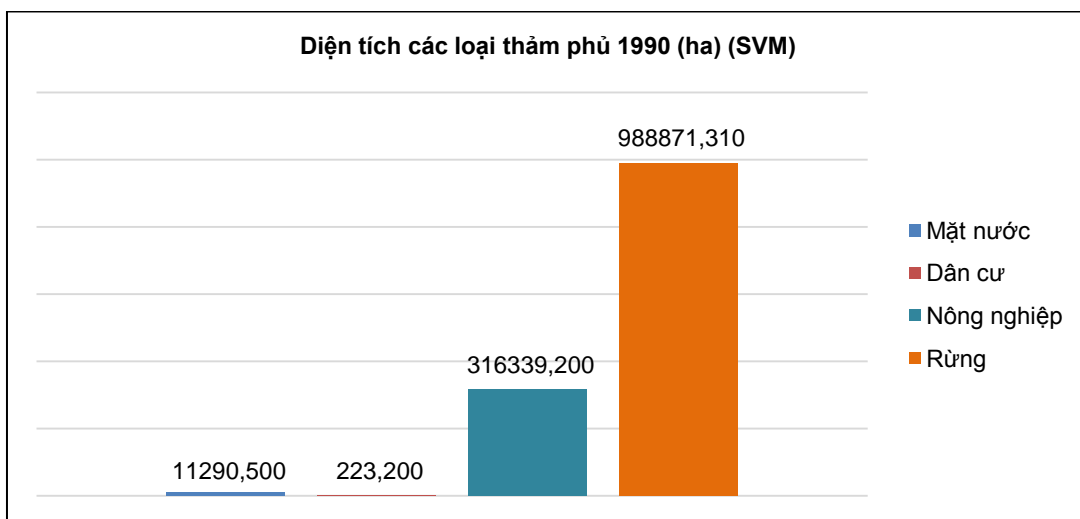
Bảng 3. Ma trận đánh giá độ chính xác và mức độ lẫn giữa các lớp phân loại theo thuật toán RF

| | Mặt nước | Dân cư | Nông nghiệp | Rừng | UA's |
|------------------------------|----------|--------|-------------|------|-------------|
| Mặt nước | 23 | 0 | 0 | 0 | 1,00 |
| Dân cư | 0 | 26 | 1 | 0 | 0,96 |
| Nông nghiệp | 0 | 4 | 29 | 3 | 0,81 |
| Rừng | 0 | 0 | 5 | 59 | 0,92 |
| PA's | 1,00 | 0,87 | 0,83 | 0,95 | 0,91 |
| Độ chính xác toàn cục | | | | | 0,91 |
| Hệ số Kappa | | | | | 0,88 |

3.2.3. Phân loại thảm phủ bằng thuật toán SVM

Kết quả phân loại bằng thuật toán SVM cho thấy độ chính xác phân loại bằng phương pháp này có độ chính xác toàn cục là 80,7% và hệ số Kappa = 0,72. Diện tích các loại thảm phủ phân loại bằng thuật toán SVM lần lượt là: Mặt nước 11.290,50 ha (chiếm 0,86%), Dân cư 223,20 ha (chiếm 0,02%), Nông nghiệp 316.339,20 ha

(chiếm 24,02%) và Rừng 988.871,31 ha (chiếm 75,10%). Với kết quả diện tích thảm phủ dân cư rất thấp, cho thấy sự không phù hợp của phương pháp SVM khi sử dụng để phân loại lớp phủ trong trường hợp nghiên cứu này. Thống kê diện tích các loại thảm phủ chính của tỉnh Đắk Lắk phân loại bằng thuật toán SVM được thể hiện ở hình 6.



Hình 6. Diện tích các loại thảm phủ theo thuật toán SVM

Ma trận sai số (bảng 4) của phương pháp phân loại ảnh bằng thuật toán SVM cho thấy, độ chính xác của người sử dụng (User's Accuracy - UA's) dao động từ 0,26 đến 1, trong đó lớp Mặt nước đạt độ chính xác cao nhất (UA = 1,00), tiếp theo là Rừng (UA = 0,92), Nông nghiệp (UA = 0,89),

và thấp nhất là Dân cư (UA = 0,26). Về độ chính xác của nhà sản xuất (Producer's Accuracy - PA's), lớp Mặt nước vẫn đạt giá trị cao nhất (PA = 1,00), trong khi lớp Nông nghiệp có PA thấp nhất (0,56), cho thấy vẫn có sự nhầm lẫn đáng kể giữa lớp Nông nghiệp với lớp Dân cư.

Bảng 4. Ma trận đánh giá độ chính xác và mức độ lẫn giữa các lớp phân loại theo thuật toán RF

| | Mặt nước | Dân cư | Nông nghiệp | Rừng | UA's |
|------------------------------|----------|--------|-------------|------|-------------|
| Mặt nước | 23 | 0 | 0 | 0 | 1,00 |
| Dân cư | 0 | 7 | 20 | 0 | 0,26 |
| Nông nghiệp | 0 | 0 | 32 | 4 | 0,89 |
| Rừng | 0 | 0 | 5 | 59 | 0,92 |
| PA's | 1,00 | 1,00 | 0,56 | 0,94 | 0,81 |
| Độ chính xác toàn cục | | | | | 0,81 |
| Hệ số Kappa | | | | | 0,72 |

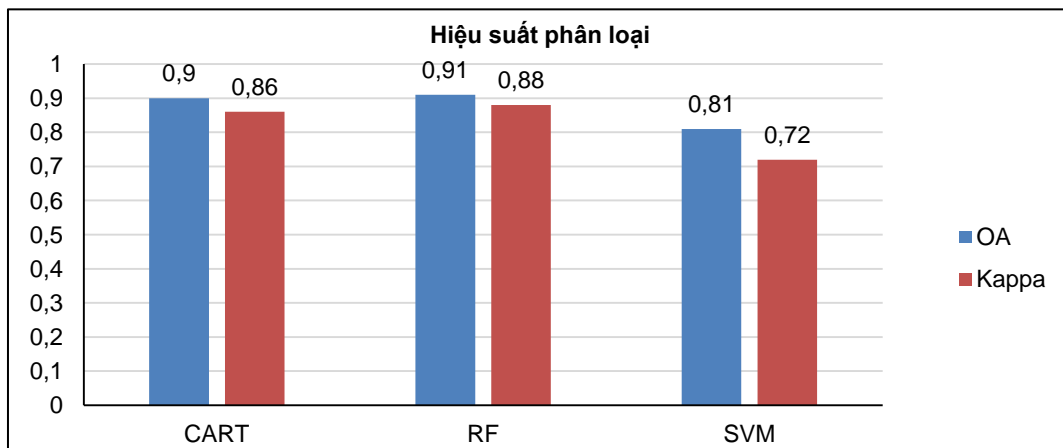
3.3. Đánh giá độ chính xác của các thuật toán phân loại

Đánh giá độ chính xác của phân loại được tiến hành nhằm đánh giá hiệu suất của các thuật toán phân loại khác nhau trong phân loại thảm phủ sử dụng đất. Các đánh giá bao gồm độ chính xác của người sản xuất và người sử dụng cho từng lớp thảm phủ cũng như độ chính xác tổng thể của phân loại và hệ số Kappa.

3.3.1. Độ chính xác toàn cục (OA)

Kết quả phân loại cho thấy, sự khác biệt đáng kể về độ chính xác tổng thể của ba phương pháp phân loại sử dụng các thuật toán khác nhau.

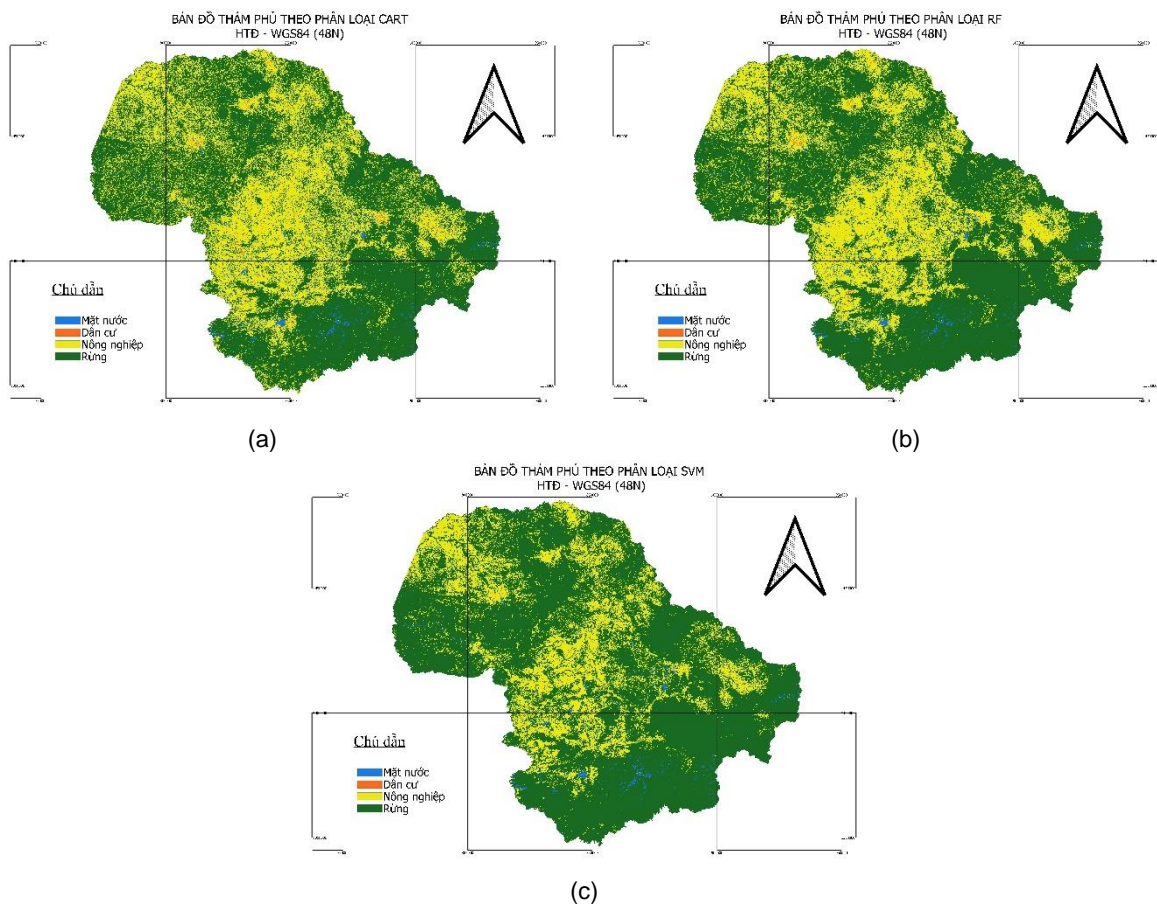
Random Forest là phương pháp đạt độ chính xác cao nhất với độ chính xác tổng thể 92,75% và hệ số Kappa là 0,90, thể hiện khả năng xử lý dữ liệu phức tạp và phân loại chính xác. CART đạt độ chính xác 90% và hệ số Kappa là 0,86, cho thấy độ tin cậy và ý nghĩa thống kê của phương pháp. SVM, ngược lại, có độ chính xác thấp nhất, chỉ đạt 80,7%, với hệ số Kappa là 0,72, phản ánh những hạn chế khi xử lý dữ liệu của phương pháp này so với CART và RF đối với dữ liệu ảnh Landsat 5 cho khu vực tỉnh Đắk Lắk. Khác biệt về hiệu suất phân loại của các phương pháp phân loại khác nhau được thể hiện ở hình 7.



Hình 7. Hiệu suất phân loại theo các thuật toán khác nhau

Kết quả so sánh hiệu suất phân loại cho thấy, mặc dù không có sự sai khác lớn trong hiệu suất phân loại bằng hai thuật toán RF và CART nhưng thuật toán RF vẫn cho thấy hiệu suất phân loại cao hơn so với CART và cao hơn hẳn so với phân loại bằng thuật toán SVM trong trường hợp nghiên cứu này. Kết quả này cho thấy thuật toán RF cho kết quả phân loại có độ chính xác cao hơn rõ rệt so với thuật toán SVM, điều này tương đồng với kết quả nghiên cứu của Kulkarni và đồng tác giả (2015). Tuy nhiên,

cũng có sự khác biệt so với các nghiên cứu của các tác giả Phan và Kappas (2018), Nguyễn Thị Thanh Hương và đồng tác giả (2019), Patil và Panhalkar (2023). Điều này có thể do ảnh hưởng của loại dữ liệu đầu vào (ảnh vệ tinh), số lượng thảm phủ sử dụng đất được phân loại cũng như kiểu thảm phủ ở các vùng địa lý khác nhau. Bản đồ phân bố các loại thảm phủ khác nhau tỉnh Đắk Lắk năm 1990 phân loại từ ảnh vệ tinh Landsat 5 bằng các thuật toán phân loại khác nhau được thể hiện ở hình 8.



Hình 8. Bản đồ thảm phủ, sử dụng đất theo 3 phương pháp phân loại: (a) CART, (b) RF, (c) SVM

3.3.2. Độ chính xác của người sử dụng (UA) và nhà sản xuất (PA)

Độ chính xác của người sử dụng đo lường mức độ chính xác trong việc phân loại các điểm ảnh được gán nhãn bởi dữ liệu tham chiếu. Độ chính xác của người sản xuất cho biết tỷ lệ các pixel ở lớp đó được phân loại chính xác. Kết quả phân

tích độ chính xác của người sử dụng cho từng thuật toán được trình bày trong bảng 5. Đối với lớp Mặt nước, các thuật toán RF và SVM đạt điểm độ chính xác của người sử dụng và người sản xuất hoàn hảo là 1,00 cho thấy việc phân loại thảm phủ mặt nước rất chính xác. Lớp Dân cư và Nông nghiệp dễ có sự nhầm lẫn với nhau

khi sử dụng thuật toán SVM. Đối với thảm phủ là Rừng thì độ chính xác trong phân loại các lớp phủ này khá cao và không có sự chênh lệch đáng kể giữa các thuật toán phân loại.

Bảng 5. Đánh giá khả năng phân biệt của từng loại thảm phủ

| Thảm phủ | Độ chính xác người sử dụng | | | Độ chính xác người sản xuất | | |
|-------------|----------------------------|------|------|-----------------------------|------|------|
| | CART | RF | SVM | CART | RF | SVM |
| Mặt nước | 1,00 | 1,00 | 1,00 | 0,96 | 1,00 | 1,00 |
| Dân cư | 0,89 | 0,96 | 0,26 | 0,89 | 0,87 | 1,00 |
| Nông nghiệp | 0,83 | 0,81 | 0,89 | 0,77 | 0,83 | 0,56 |
| Rừng | 0,91 | 0,92 | 0,92 | 0,97 | 0,95 | 0,94 |

Đánh giá độ chính xác của kết quả phân loại bằng các thuật toán khác nhau thông qua hệ số Kappa cho thấy phương pháp phân loại bằng thuật toán RF vẫn là phương pháp cho kết quả có độ chính xác cao nhất trong các thuật toán sử dụng ($K = 0,88$). Các phương pháp có độ tin cậy theo sau lần lượt là CART ($K = 0,86$) và SVM ($K = 0,72$). Qua đó cho thấy phương pháp phân loại thảm phủ bằng thuật toán RF là phương pháp ổn định và có độ tin cậy cao hơn so với thuật toán CART và SVM.

3.3.3. Thách thức và nhầm lẫn chính

CART và RF gặp khó khăn trong việc tách biệt khu vực nông nghiệp với rừng thưa, đặc biệt ở những vùng có thay đổi theo mùa, như rừng Khộp rụng lá vào mùa khô.

SVM có sự nhầm lẫn lớn giữa các khu dân cư và vùng nông nghiệp, cũng như PA thấp đối với khu vực nông nghiệp. Điều này cho thấy những hạn chế của thuật toán này trong việc xử lý tín hiệu quang phổ chồng chéo.

IV. KẾT LUẬN

Các kết quả phân loại bằng 3 thuật toán khác nhau bao gồm CART, RF và SVM cho thấy RF là phương pháp hiệu quả nhất trong phân loại thảm phủ tại tỉnh Đắk Lắk, với độ chính xác tổng thể cao nhất và hiệu suất vượt trội ở tất cả các chỉ số quan trọng. CART là một lựa chọn đáng tin cậy với hiệu suất gần tương đương.

Trong khi đó, SVM mặc dù được sử dụng rộng rãi, cho thấy những hạn chế đáng kể trong xử lý dữ liệu phức tạp và phân biệt các lớp thảm phủ chồng chéo.

Kết quả này nhấn mạnh tầm quan trọng của việc lựa chọn thuật toán phân loại phù hợp với đặc điểm khu vực nghiên cứu và dữ liệu. Đối với các ứng dụng đòi hỏi độ chính xác cao, RF là lựa chọn tối ưu, trong khi CART cung cấp một giải pháp hiệu quả đối với khu vực tỉnh Đắk Lắk. SVM có thể cần tối ưu hóa tham số và các kỹ thuật hỗ trợ để cải thiện hiệu suất trong các bối cảnh tương tự. Vì vậy, cần có các nghiên cứu tiếp theo trong việc sử dụng các tham số khác ngoài dải quang phổ ảnh vệ tinh để phân loại thảm phủ cho kết quả tốt hơn khi sử dụng các thuật toán học máy. Ngoài ra cần nghiên cứu thống kê tương quan giữa 03 thuật toán học máy và kết quả phân loại hiện trạng để phân tích được kết quả mỗi thuật toán phù hợp cho loại hình hiện trạng nào tại vùng nghiên cứu.

Lời cảm ơn: Kết quả nghiên cứu này là một phần của đề tài khoa học và công nghệ cấp tỉnh, tỉnh Đắk Lắk. Chúng tôi xin chân thành cảm ơn Sở Khoa học và Công nghệ tỉnh Đắk Lắk đã hỗ trợ tài chính và tạo điều kiện thuận lợi trong suốt quá trình thực hiện nghiên cứu. Đồng thời, chúng tôi cũng gửi lời cảm ơn sâu sắc đến các đồng nghiệp đã nhiệt tình hỗ trợ trong việc thu thập dữ liệu thực địa.

TÀI LIỆU THAM KHẢO

1. Breiman L. and Ihaka R., 1984. Nonlineardiscriminant analysis via scaling and ACE. Department of Statistics, University of California, Technical Report. 40, 1.
2. Breiman L., 2001. Random Forests. Machine Learning Journal Paper, 45(1), p. 5 - 32.
3. Congalton R.G., and Green K., 1999. Assessing the accuracy of remotely sensed data: Principles and practices. Lewis Publishers. Boca Raton FL.
4. Esfandeh S., Danehkar A., Salmanmahiny A., Sadeghi S.M.M., Marcu M.V., 2022. Climate change risk of urban growth and land use/land cover conversion: An In-Depth Review of the Recent Research in Iran Sustainability, 14(1). [https:// doi. org/ 10. 3390/ su140 10338](https://doi.org/10.3390/su14010338).
5. Gorelick N., Hancher M., Dixon M., Ilyushchenko S., Thau D., Moore R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens Environ 202:18 - 27. [https:// doi. org/ 10. 1016/j. rse. 2017. 06. 031](https://doi.org/10.1016/j.rse.2017.06.031).
6. Jenkins M. and Schaap B., 2018. Forest Ecosystem Services. The thirteenth session of United Nation Forum on Forests. Apr 2018.
7. Khalid F., Taj M.B., Jamil A., Kamal H., Afzal T., Iqbal M.J., Khan T., Ashiq M., Raheel A., Sharif M., and Tirmizi S.A., 2019. Multiple Impacts of Illegal Logging: A key to Deforestation Over the Globe. Biomedical Journal of Scientific & Technical Research, August, 2019, Volume 20, 5, pp 15430 - 15435. DOI: 10.26717/BJSTR.2019.20.003519.
8. Kolli M.K., Opp C., Karthe D., Groll M., 2020. Mapping of major land-use changes in the Kolleru Lake freshwater ecosystem by using Landsat satellite images in Google earth engine. Water, 12(9). [https:// doi. org/ 10. 3390/ w1209 2493](https://doi.org/10.3390/w12092493).
9. Kuang W., Yang T., Yan F., 2018. Examining urban land-cover characteristics and ecological regulation during the construction of Xiong'an New District, Hebei Province. China J Geograph Sci 28(1):109 - 123. [https:// doi. org/ 10. 1007/ s11442 - 018 - 1462 - 4](https://doi.org/10.1007/s11442-018-1462-4).
10. Li C., Wang J., Wang L., Hu L., Gong P., Li C., Gong P., 2014. Comparison of Classification Algorithms and Training Sample Sizes in Urban Land Classification with Landsat Thematic Mapper Imagery. Remote Sensing, 6(2), 964 - 983. <https://doi.org/10.3390/rs6020964>.
11. Lu D., Weng Q.A., 2007. Survey of image classification methods and techniques for improving classification performance. Int. J. Remote Sens., 28, 823 - 870.
12. Mutale B., Withanage N.C., Mishra P.K., Shen J., Abdelrahman K., Fnais M.S., 2024. A performance evaluation of random forest, artificial neural network, and support vector machine learning algorithms to predict spatio-temporal land use-land cover dynamics: a case from lusaka and colombo. Front. Environ. Sci., 11 September 2024, Sec. Land Use Dynamics, Volume 12 - 2024. <https://doi.org/10.3389/fenvs.2024.1431645>.
13. Nguyen, H. T. T., 2011. Forestry Remote Sensing: Using multi-data sources for inventory of natural broadleaved evergreen forests in the Central Highlands of Vietnam. 152pg. ISBN: 978 - 3 - 8454 - 3589 - 3. Lambert Academic Publishing, Germany.
14. Nguyen H. T. T., Doan T. M., Radeloff V., 2018. Applying random forest classification to map land use/land cover using Landsat 8 OLI, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-3/W4, 363 - 367, <https://doi.org/10.5194/isprs-archives-XLII-3-W4-363-2018>, 2018.
15. Nguyễn Thanh Tuấn, Phan Văn Tuấn, Nguyễn Văn Quý, Hoàng Thị Phương Nhung, 2022. 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. Tạp chí Khoa học Trường Đại học Cần Thơ, Tập 58, số 6B (2022): 150 - 163. DOI:10.22144/ctu.jvn.2022.254.
16. Ouchra H., Belangour A., Erraissi A. 2023. Machine learning algorithms for satellite image classification using Google Earth Engine and Landsat satellite data: Morocco case study. IEEE Access, volume 11, 2023. Digital Object Identifier 10.1109/ACCESS.2023.3293828.

17. Pande C.B., 2022. Land use/land cover and change detection map-ping in Rahuri watershed area (MS), India using the Google Earth Engine and machine learning approach. *Geocarto Int* 37(26):13860 - 13880. <https://doi.org/10.1080/10106049.2022.2086622>.
18. Phan Trọng Thế, Mai Thị Huyền, Nguyễn Bá Long, 2023. Đánh giá dự báo biến động sử dụng đất tại huyện Nhơn Trạch, tỉnh Đồng Nai trên cơ sở ứng dụng hệ thống thông tin địa lý (GIS) và chuỗi Markov. *Tạp chí Khoa học và công nghệ Lâm nghiệp* số 2 - 2023. <https://doi.org/10.55250/jo.vnuf.2023.2.146-155>.
19. Phan, T.N., & Kappas, M. (2018). Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors*, 18(1), 18 - 37, [Doi:10.3390/s18010018](https://doi.org/10.3390/s18010018).
20. Sidhu N., Pebesma E., Camara G., 2018. Using Google earth engine to detect land cover change: Singapore as a use case. *Eur J Remote Sens* 51(1):486 - 500. <https://doi.org/10.1080/22797254.2018.1451782>.
21. Thomas M.L., Ralph W. Kiefer J.C., 2004. Remote Sensing and Image Interpretation (Fifth Edition). *Geogr. J.* 2004, 146, 448 - 449.
22. Tso B., Mather P.M., 2009. Classification methods for remotely sensed data (Second edition ed.). Boca Raton: CRC.
23. Vapnik V., 1995. The nature of statistical learning theory. Springer-Verlag New York, Inc.
24. Vapnik V., 2006. Estimation of Dependences Based on Empirical Data (Information Science and Statistics): Springer.
25. Waske, B., & Braun, M., 2009. Classifier ensembles for land cover mapping using multitemporal SAR imagery. *ISPRS J. Photogramm. Remote Sens.*, 64, 450 - 457.
26. Xie S., Liu L., Zhang X., Yang J., Chen X., Gao Y., 2019. Auto-matic land-cover mapping using Landsat time-series data based on Google earth engine. *Remote Sens*, 11(24). <https://doi.org/10.3390/rs11243023>.
27. Yao Y., Yan X., Luo P., Liang Y., Ren S., Hu Y. *et al.*, 2022. Classifying land-use patterns by integrating time-series electricity data and high-spatial resolution remote sensing imagery. *Int J Appl Earth Obs Geoinf* 106:102664. <https://doi.org/10.1016/j.jag.2021.102664>.
28. Website: <https://www.usgs.gov/landsat-missions/landsat-5>. Truy cập ngày 15/11/2024.

Email tác giả liên hệ: hdbao@ttn.edu.vn

Ngày nhận bài: 19/12/2024

Ngày phản biện đánh giá và sửa chữa: 29/12/2024; 30/12/2024;

Ngày duyệt đăng: 07/01/2025