Phát hiện mất rừng tự động sử dụng thuật toán học máy và ảnh Sentinel-2 trên Google Earth Engine: Nghiên cứu tại tỉnh Cà Mau

Vũ Văn Thái[1], Hoàng Việt Anh[1], Tạ Minh Huy[1], Dương Văn Vinh[2]*
[1]Công ty TNHH Tư vấn và Phát Triển Đồng Xanh
[2]Chi cục Kiểm lâm tỉnh Cà Mau
Email tác giả liên hệ: anh.hv@gfd.com.vn

doi.org/10.5281/zenodo.17068824

Tóm tắt:

Nghiên cứu này trình bày việc xây dựng và thử nghiệm các mô hình học máy nhằm tự động phát hiện mất rừng tại tỉnh Cà Mau. Toàn bộ quy trình, từ tiền xử lý dữ liệu ảnh vệ tinh Sentinel-2 (Phản xạ bề mặt - L2A) huấn luyện và đánh giá mô hình được thực hiện trên nền tảng điện toán đám mây Google Earth Engine (GEE). Nghiên cứu đã sử dụng kỹ thuật lọc mây để có được ảnh với mây che phủ dưới 40%, đồng thời trích xuất 10 kênh phổ và 8 chỉ số thực vật, trong đó các chỉ số delta (ΔNDVI, ΔEVI, ΔNBR, ΔSAVI) phục vụ huấn luyện và đánh giá mô hình. Ba thuật toán học máy được sử dụng bao gồm Random Forest, Support Vector Machine (SVM) và Gradient Boosting Trees (GBT). Nghiên cứu sử dụng 2.500 điểm mẫu được thu thập với tỷ lệ cân bằng 50% mất rừng và 50% không mất rừng, sau khi trích xuất các đặc trưng cho các điểm mẫu tiến hành chia dữ liệu huấn luyện chiếm 80% (train và validation), 20% dữ liệu kiểm chứng (test), đảm bảo duy trì tỷ lệ 50:50 trong mỗi tập. Kết quả đánh giá cho thấy mô hình Gradient Boosting Trees đạt hiệu suất cao nhất với độ chính xác tổng thể 85,40%, recall 72,00%, precision 98,36% và F1-score 83,14%. Với recall đạt mức khả quan và precision rất cao, hệ thống cho thấy độ tin cậy cao khi cảnh báo mất rừng, đồng thời phát hiện được phần lớn các trường hợp mất rừng thực tế. Nghiên cứu đã xây dựng thành công một công cụ phát hiện mất rừng tự động trên GEE, cung cấp thông tin hỗ trợ hiệu quả cho công tác quản lý rừng tại địa phương. Tuy nhiên, cần nghiên cứu tiếp về tích hợp dữ liệu radar Sentinel-1 để khắc phục hạn chế về mây che phủ trong mùa mưa và nâng cao hơn nữa khả năng phát hiện.

Từ khóa: *machine learning, random forest, sentinel-2, Cà Mau, phát hiện mất rừng*
Ngày nhận bài: 12/08/2025 Ngày sửa lại: 22/08/2025 Ngày chấp nhận đăng: 23/08/2025 Ngày xuất bản: 30/08/2025

Automated Forest Loss Detection Using Machine Learning Algorithms and Sentinel-2 Imagery on Google Earth Engine: A Study in Ca Mau Province, Vietnam

Vu Van Thai[1], Hoang Viet Anh[1], Ta Minh Huy[1], Duong Van Vinh[2]*
[1]Green Field Consulting & Development Ltd.,
[2]Provincial Forest Protection Department
Corresponding Author Email: anh.hv@gfd.com.vn

Abstract:

This study presents the development and testing of machine learning models for automated forest loss detection in Ca Mau Province. The entire workflow, from preprocessing Sentinel-2 satellite imagery (Surface Reflectance - L2A) to model training and evaluation, was implemented on the Google Earth Engine (GEE) cloud computing platform. The study applied cloud filtering techniques to obtain images with cloud cover below 40%, while extracting 10 spectral bands and 8 vegetation indices, with delta indices (ΔNDVI, ΔEVI, ΔNBR, ΔSAVI) serving for model training and evaluation. Three machine learning algorithms were employed including Random Forest, Support Vector Machine (SVM), and Gradient Boosting Trees (GBT). The study utilized 2,500 sample points collected with a balanced ratio of 50% forest loss and 50% non-forest loss, and after extracting features for these sample points, the data was divided into 80% for training (train and validation) and 20% for testing, maintaining the 50:50 ratio in each subset. Evaluation results showed that the Gradient Boosting Trees model achieved the highest performance with an overall accuracy of 85.40%, recall of 72.00%, precision of 98.36%, and F1-score of 83.14%. With considerable recall and very high precision, the system demonstrates high reliability when alerting forest loss while detecting the majority of actual forest loss cases. The study successfully developed an automated forest loss detection tool on GEE, providing effective information to support local forest management efforts. However, further research on integrating Sentinel-1 radar data is needed to overcome cloud cover limitations during the rainy season and further enhance detection capabilities.

Keywords*: machine learning, random forest, sentinel 2, Ca Mau, Vietnam, forest loss detection*
Submission received: 12/08/2025 Revised: 22/08/2025 Accepted: 23/08/2025 Published: 30/08/2025

1. Giới thiệu

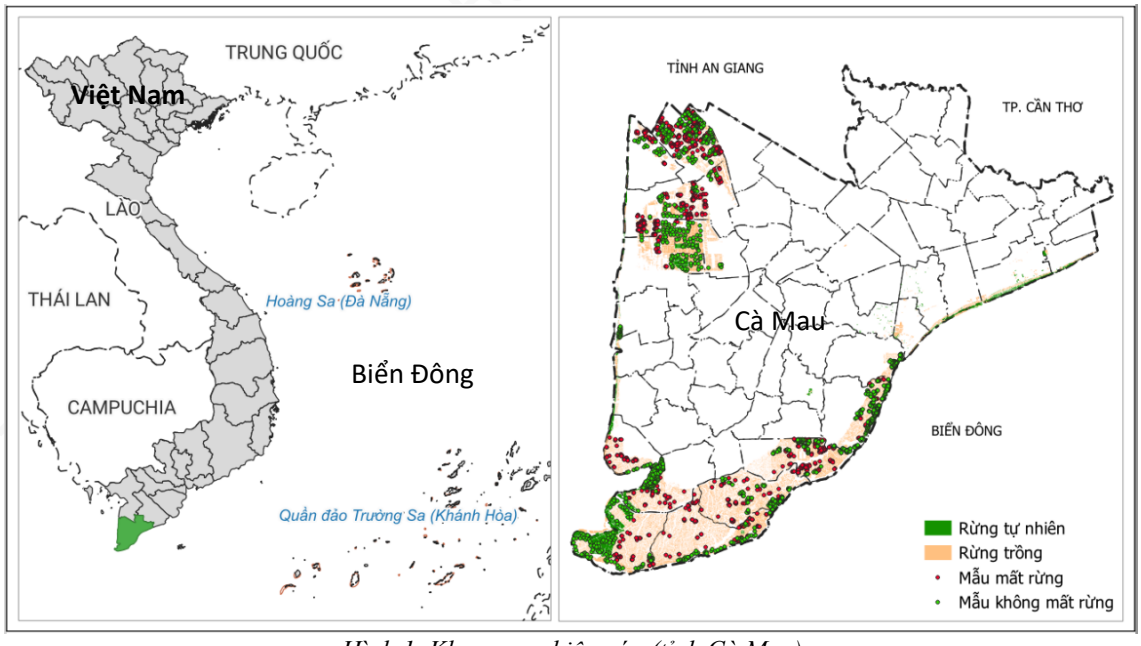
Giám sát biến động rừng đóng vai trò quan trọng trong nỗ lực quản lý tài nguyên bền vững, bảo tồn đa dạng sinh học và giảm thiểu tác động của biến đổi khí hậu. Tuy nhiên, các phương pháp giám sát truyền thống đang bộc lộ những tồn tại như việc điều tra thực địa, mặc dù cung cấp dữ liệu ở mức độ chi tiết cao, nhưng lại rất tốn kém về chi phí, nhân lực, mất nhiều thời gian và không khả thi để triển khai trên các khu vực rộng lớn hoặc khó tiếp cận [1, 2]. Các phương pháp tiếp cận dựa trên viễn thám quang học, tuy đã khắc phục được vấn đề về quy mô không gian, nhưng lại phải đối mặt với thách thức về mây che phủ, đặc biệt là các vùng nhiệt đới. Vào mùa mưa, tỷ lệ mây trên ảnh là rất lớn, do vậy làm giảm khả năng giám sát trên ảnh quang học [3-5]. Cà Mau là tỉnh thuộc khu vực Tây Nam Bộ, với 3 mặt giáp biển do vậy ảnh hưởng bởi yếu tố thời tiết là dẫn lớn đến chất lượng ảnh vệ tinh [6]. Sự khác biệt lớn về cấu trúc tán lá, đặc điểm phản xạ phổ và động thái thay đổi giữa hai loại hình rừng này đòi hỏi các phương pháp phân tích chuyên biệt, khiến cho việc áp dụng một mô hình giám sát chung trở nên kém chính xác. Do đó, cần xây dựng một giải pháp phát hiện mất rừng tự động, chính xác cao, phù hợp với điều kiện phức tạp tại Cà Mau và các tỉnh ven biển. Chúng tôi sử dụng thuật toán học máy để mô hình tự học đặc điểm mất rừng từ các chỉ số trích xuất trên ảnh qua đó mở ra cơ hội khắc phục hạn chế của phương pháp cũ [7-11]. Mô hình huấn luyện kỹ lưỡng sẽ cung cấp dữ liệu mất rừng kịp thời, hỗ trợ quản lý bền vững hệ sinh thái rừng ngập mặn tại tỉnh Cà Mau.

2. Phương pháp nghiên cứu

2.1. Khu vực nghiên cứu và dữ liệu

2.1.1. Khu vực nghiên cứu

Khu vực nghiên cứu là tỉnh Cà Mau, một tỉnh thuộc vùng sinh thái Tây Nam Bộ, phía Nam Việt Nam. Đối tượng nghiên cứu là toàn bộ diện tích có rừng và đất lâm nghiệp của tỉnh. Tính đến 31/12/2024, diện tích rừng tỉnh Cà Mau (cũ) là 74.103,71 ha, trong đó rừng tự nhiên là 11.389,03 ha; rừng trồng: 62.714,68 ha, tỷ lệ độ che phủ là 14,05 % [12].

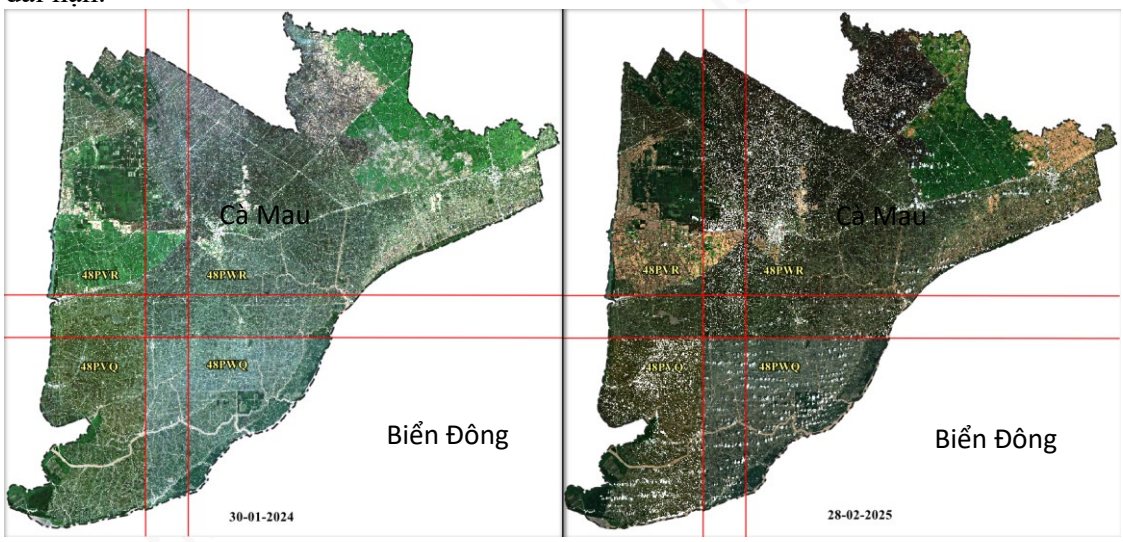


Hình 1. Khu vực nghiên cứu (tỉnh Cà Mau)

2.1.2. Dữ liệu ảnh vệ tinh

Trong nghiên cứu này, dữ liệu viễn thám được sử dụng là ảnh đa phổ Sentinel-2, là một vệ tinh thuộc chương trình Copernicus của Cơ quan Vũ trụ Châu Âu (ESA). Chùm vệ tinh Sentinel 2 gồm 2 vệ tinh song song Sentinel-2A phóng năm 2014, Sentinel-2B phóng năm 2017, ngoài ra có vệ tinh Sentinel 2C phóng năm 2024 nhằm thay thế khi Sentinel 2A ngừng hoạt động. Cả 3 vệ tinh hoạt động trên cùng một quỹ đạo và được trang bị bộ cảm biến đa phổ (Multispectral Instrument - MSI). Bộ cảm biến này thu nhận dữ liệu ở 13 kênh phổ trải dài từ vùng nhìn thấy (Visible) đến hồng ngoại sóng ngắn (SWIR) với các độ phân giải không gian lần lượt là 10m, 20m và 60m. Sự kết hợp của chùm vệ tinh mang lại tần suất chụp lặp lại là 5 ngày, đây là một ưu thế vượt trội cho các ứng dụng giám sát biến động [13]. Tại Việt Nam, dữ liệu Sentinel-2 đã được ứng dụng hiệu quả trong nhiều bài toán lâm nghiệp, từ việc thành lập bản đồ hiện trạng rừng, phân loại các kiểu rừng khác nhau như rừng ngập mặn và rừng trồng, cho đến giám sát biến động diện tích rừng và ước tính sinh khối [14-16].

Chúng tôi chọn dữ liệu Sentinel-2 vì độ phân giải 10 m ở kênh RGB và NIR giúp phát hiện mất rừng quy mô nhỏ (từ 0.5 ha), chi tiết hơn Landsat [17]. Tần suất chụp 5 ngày tăng cơ hội lấy ảnh sạch mây ở vùng nhiệt đới như Cà Mau, đảm bảo dữ liệu liên tục [18]. Ngoài ra, kênh Red-Edge nhạy với diệp lục giúp phân biệt trạng thái rừng, nâng cao độ chính xác mô hình [19, 20]. Toàn bộ dữ liệu Sentinel-2 được cung cấp miễn phí, tạo điều kiện thuận lợi cho các nghiên cứu trên phạm vi không gian rộng và theo dõi trong dài hạn.



Hình 2. Ảnh Sentinel-2 tỉnh Cà Mau tại thời điểm 30-01-2024 và 28-02-2025

Nghiên cứu sử dụng bộ dữ liệu Level-2A (L2A): Phản xạ bề mặt (Surface Reflectance - SR). Đây là dữ liệu L1C đã được xử lý qua một thuật toán hiệu chỉnh khí quyển. Giá trị pixel của nó biểu diễn năng lượng bức xạ phản xạ từ bề mặt Trái đất, sau khi đã loại bỏ các tác động gây nhiễu của khí quyển. Trong Google Earth Engine GEE, bộ dữ liệu này được ký hiệu là COPERNICUS/S2_SR_HARMONIZED. Trong nghiên cứu này chúng tôi có thực hiện việc lọc mây và mây mờ, tuy nhiên không sử dụng nhiều kỳ ảnh để bù mây. Nghiên cứu lựa chọn Sentinel-2 thuộc 5 kỳ ảnh từ 2023 – 2025 để phục vụ huấn luyện mô hình (kỳ 1: 04-02-2023 và 23-08-2023; kỳ 2: 04-02-2023 và 30-01-2024; kỳ 3: 23-08-2023 và 30-01-2024; kỳ 4: 30-01-2024 và 04-04-2024; kỳ 5: 30-01-2024 và 28-02-2025), và 1 kỳ ảnh để phục vụ đánh giá mô hình kỳ ảnh 20-01-2022 và 11-03-2023. Các kỳ ảnh thuộc tỉnh Cà Mau trên GEE thuộc 4 cảnh ảnh bao gồm 48PVR,

48PWR, 48PWQ, 48PVQ (Thể hiện tại Hình 2 theo thứ tự từ trái sang phải, từ trên xuống dưới).

2.1.3. Dữ liệu mất rừng, không mất rừng

Bộ mẫu các điểm mất rừng, không mất rừng được thu thập từ 2 nguồn chính: i) Dựa trên các điểm mất rừng, bản đồ hiện trạng rừng từ cơ sở dữ liệu theo dõi diễn biến rừng và đất lâm nghiệp (FRMS 4.0) được Chi cục kiểm lâm tỉnh Cà Mau cập nhật hàng năm và được kiểm chứng lại bằng phương pháp chuyên gia; ii) Dựa trên các điểm điều tra thực địa của nhóm nghiên cứu thực hiện vào tháng 10/2024, trong đó các điểm mất rừng đã được xác định tọa độ trực tiếp trên thực địa thông qua chụp ảnh bằng thiết bị bay không người lái (drone) có gắn thiết bị GPS RTK (Real-Time Kinematic) độ chính xác cao.

Các điểm dữ liệu mất rừng được thu thập và đánh nhãn phù hợp với thông tin các kỳ ảnh Sentinel-2. Nghiên cứu đã thu thập được 1.000 điểm mất rừng, 1.000 điểm không mất rừng tương ứng với 5 kỳ ảnh phục vụ huấn luyện mô hình. Dữ liệu để đánh giá mô hình gồm 250 điểm mất rừng, và 250 điểm không mất rừng trên 1 kỳ ảnh độc lập. Diện tích các điểm mẫu nằm trên các lô rừng có diện tích từ 0,5 ha trở lên.

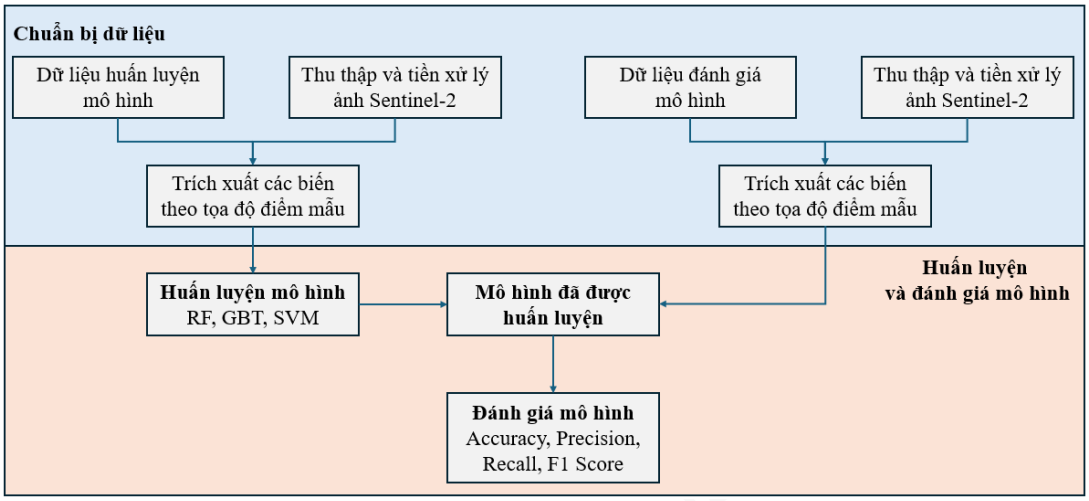
Hình 3. Mẫu dữ liệu mất rừng (màu đỏ) và mẫu dữ liệu không mất rừng (màu xanh da trời) trên ảnh Sentinel-2 giữa kỳ đầu (30-01-2024) và kỳ sau (28-02-2025)

2.2. Phương pháp trích xuất dữ liệu từ ảnh vệ tinh và xây dựng mô hình

2.2.1. Công cụ, luồng xử lý

Toàn bộ quy trình phân tích trong nghiên cứu này, từ tiền xử lý dữ liệu đến tạo ra sản phẩm cuối cùng, đều được thực hiện trên nền tảng điện toán đám mây. GEE là nền tảng điện toán đám mây được phát triển bởi Google, cung cấp khả năng xử lý và phân tích dữ liệu địa không gian ở quy mô hành tinh [21]. GEE được sử dụng để truy cập và lọc kho lưu trữ ảnh Sentinel-2, tạo và quản lý các bộ mẫu huấn luyện (bao gồm cả lớp "mất rừng" và "không mất rừng"), tách xuất các đặc trưng phổ và chỉ số thực vật từ chuỗi ảnh thời gian. Sau đó, quá trình huấn luyện mô hình học máy, lưu trữ bộ phân loại đã được tối ưu, và đánh giá độ chính xác thông qua ma trận nhầm lẫn đều được tiến hành trong cùng một môi trường. Nền tảng cung cấp sẵn các thuật toán học máy như Random Forest, SVM và Gradient Boosting Trees thông qua module ee.Classifier, cho phép huấn luyện và triển khai mô hình thực tiếp trên đám mây mà không bị giới hạn bởi tài nguyên tính toán cục bộ [22].

Luồng xử lý bắt đầu từ việc thu thập dữ liệu Sentinel-2 từ kho lưu trữ của Copernicus thông qua GEE Data Catalog, sau đó thực hiện các bước tiền xử lý bao gồm lọc mây, tạo ảnh tổng hợp và tính toán các chỉ số thực vật. Dữ liệu mẫu được thu thập từ khảo sát thực địa và phân tích ảnh độ phân giải cao, được chia thành tập huấn luyện (2023-2025) và tập kiểm tra độc lập (2022 - 2023) để đảm bảo đánh giá khách quan khả năng tổng quát hóa của mô hình [23].



Hình 4. Sơ đồ thực hiện quy trình chuẩn bị, chia tệp dữ liệu, huấn luyện mô hình và đánh giá mô hình

2.2.2. Thu thập và tiền xử lý ảnh

Dữ liệu ảnh Sentinel-2 Level-2A được thu thập từ bộ sưu tập "COPERNICUS/S2_SR_HARMONIZED" trên GEE cho khu vực tỉnh Cà Mau (8°35'-9°35'N, 104°43'-105°20'E) trong giai đoạn 2022-2025. Việc lựa chọn Level-2A đảm bảo dữ liệu đã được hiệu chỉnh khí quyển, cung cấp giá trị phản xạ bề mặt phù hợp cho phân tích thực vật [24]. Quy trình lọc mây được thực hiện theo hai bước chính. Kênh QA60 được sử dụng để xác định và loại bỏ pixel mây thông qua phép toán bit-masking với bit 10 (mây mờ) và bit 11 (mây dày). Sau đó, áp dụng ngưỡng xác suất mây 40% từ kênh MSK_CLDPRB được áp dụng để lọc thêm các pixel mây không được phát hiện [25]. Hàm *updateMask()* trong GEE được sử dụng để áp dụng cloud mask lên từng ảnh trong collection.

2.2.3. Trích xuất đặc trưng phổ và chỉ số viễn thám

Việc trích xuất đặc trưng phổ từ dữ liệu Sentinel-2 được thực hiện thông qua hai nhóm biến chính: giá trị phản xạ gốc của các kênh phổ và các chỉ số thực vật tính toán. Mười kênh phổ từ vùng nhìn thấy đến hồng ngoại sóng ngắn được sử dụng để nắm bắt đặc trưng phổ đa dạng của rừng ngập mặn [26]. Các Kênh Red Edge (B5, B6, B7) đặc biệt có giá trị trong việc phát hiện sớm sự suy thoái rừng do nhạy cảm với thay đổi hàm lượng chlorophyll và cấu trúc tán lá [27]. Kênh SWIR (B11, B12) phản ánh hàm lượng nước trong thực vật, yếu tố quan trọng để phân biệt rừng ngập mặn khỏe mạnh với rừng bị stress hoặc chết khô do xâm nhập mặn [28].

Chúng tôi tính toán chỉ số thực vật để nhấn mạnh sự khác biệt phổ giữa rừng và không rừng, giảm nhiễu môi trường. NDVI [29] và EVI [30] đánh giá sinh khối, với EVI ưu tiên cho rừng ngập mặn dày. SAVI [31] điều chỉnh nền đất ẩm ven biển, còn NBR [32] phát hiện thay đổi cấu trúc do chặt phá hoặc chuyển đổi thủy sản [23]. Chỉ số delta (ΔNDVI, ΔEVI, ΔNBR, ΔSAVI [33-35]) đo lường biến đổi giữa các kỳ ảnh, giúp mô hình phân biệt biến động tự nhiên (thủy triều, mùa vụ) với mất rừng thực tế tại Cà Mau.

Việc triển khai trên GEE thông qua các hàm tối ưu như *normalizedDifference()* và *expression()* đảm bảo xử lý hiệu quả cho khối lượng dữ liệu lớn ở quy mô tỉnh.

2.2. Phương pháp huấn luyện mô hình

2.2.1. Chia tệp dữ liệu

Việc phân chia dữ liệu được thiết kế theo nguyên tắc độc lập hoàn toàn giữa tập huấn luyện và tập kiểm tra để đảm bảo đánh giá khách quan hiệu suất mô hình [36]. Nghiên cứu tiến hành phân chia dữ liệu thành 2 tệp độc lập. Trong mỗi tệp dữ liệu đều gồm 50% là dữ liệu mất rừng và 50% là dữ liệu không mất rừng.

(1) Tập huấn luyện và hiệu chỉnh mô hình (train và validation) gồm 2.000 điểm thu thập trong giai đoạn 2023-2025, dữ liệu được phân chia với tỷ lệ giữa train và validation là 80:20). Nghiên cứu áp dụng phương pháp xác thực chéo (cross-validation k-fold) nhằm đảm bảo toàn bộ dung lượng mẫu lần lượt được huấn luyện và đánh giá, tối ưu hóa việc sử dụng dữ liệu [37]. Nghiên cứu áp dụng k=5, kết quả cuối cùng là trung bình của 5 lần đánh giá, qua đó giúp đánh giá độ tin cậy về hiệu suất của mô hình và phát hiện overfitting [36].

(2) Tập đánh giá mô hình (test) gồm 500 điểm cho kỷ ảnh 2022 – 2023. Tệp này hoàn toàn độc lập chỉ sử dụng cho đánh giá mô hình trên 1 ảnh mới hoàn toàn không sử dụng trong quá trình huấn luyện.

Bảng 1. Phân bố dữ liệu mẫu theo tập và lớp phân loại

Tập dữ liệu	Thời gian	Tổng số	Không có rừng	Mất rừng	Mục đích
Huấn luyện (train)	2023-2025	1500	750	750	Train mô hình
Đánh giá (Validation)	2023-2025	500	250	250	Điều chỉnh tham số
Kiểm chứng (Test)	2022 - 2023	500	250	250	Đánh giá hiệu suất mô hình
Tổng cộng		**2500**	**1250**	**1250**	

2.2.2. Huấn luyện mô hình

Trong nghiên cứu chúng tôi sử dụng 3 mô hình học máy bao gồm Random Forest (RF), Gradient Boosting Trees (GBT), Support Vector Machine (SVM).

Random Forest là thuật toán được lựa chọn ưu tiên cho bài toán phát hiện mất rừng do khả năng xử lý hiệu quả dữ liệu đa chiều và độ chính xác cao trong phân loại ảnh viễn thám [38]. Thuật toán hoạt động bằng cách xây dựng nhiều cây quyết định độc lập và tổng hợp kết quả dự đoán, giúp giảm overfitting và tăng độ ổn định của mô hình. Trong GEE, RF được triển khai thông qua hàm *ee.Classifier.smileRandomForest()* với tham số chính là số lượng cây (numberOfTrees) được đặt ở 100, giá trị được chứng minh hiệu quả cho phân loại rừng nhiệt đới [39]. Ưu điểm nổi bật của RF trong nghiên cứu này là khả năng đánh giá tầm quan trọng của các biến (variable importance), giúp xác định chỉ số thực vật nào đóng vai trò quan trọng nhất trong phát hiện mất rừng ngập mặn, từ đó tối ưu hóa việc thu thập dữ liệu cho các nghiên cứu tiếp theo.

GBT được triển khai trên GEE qua hàm *ee.Classifier.smileGradientTreeBoost()*, áp dụng kỹ thuật boosting để xây dựng mô hình mạnh từ chuỗi các mô hình yếu [40]. GBT đặc biệt hiệu quả trong việc nắm bắt các mối quan hệ phi tuyến phức tạp giữa các chỉ số thực vật và trạng thái rừng, phù hợp cho môi trường rừng ngập mặn với sự biến động cao do ảnh hưởng của thủy triều và mùa vụ [23]. Tham số learning rate được đặt ở 0.05 để cân bằng giữa độ chính xác và tốc độ huấn luyện, trong khi số lượng cây được giới hạn ở 150 để tránh overfitting trên GEE.

Support Vector Machine được triển khai trên GEE thông qua *ee.Classifier.libsvm()*, phù hợp cho việc phân tách dữ liệu phi tuyến tính trong không gian đặc trưng nhiều chiều [41]. SVM có ưu điểm hoạt động hiệu quả với số lượng mẫu huấn luyện hạn chế, quan trọng trong bối cảnh thu thập dữ liệu thực địa tại vùng rừng ngập mặn khó tiếp cận. Thuật toán tìm kiếm siêu phẳng tối ưu phân tách hai lớp với margin lớn nhất, giúp tăng khả năng tổng quát hóa cho dữ liệu mới. Trong nghiên cứu về rừng ngập mặn, SVM đã được chứng minh hiệu quả trong việc phân biệt các giai đoạn suy thoái rừng khác nhau, từ rừng nguyên sinh đến rừng thứ sinh và đất trống [42].

Việc huấn luyện tất cả các mô hình được thực hiện hoàn toàn trên nền tảng đám mây của GEE, tận dụng khả năng xử lý song song và tránh giới hạn bộ nhớ của máy tính cục bộ [21]. Quy trình huấn luyện được thực hiện thông qua các bước: (1) Chuẩn bị dữ liệu (training collection) với hàm *sampleRegions()* để trích xuất giá trị pixel tại các điểm mẫu; (2) Huấn luyện mô hình với hàm train() sử dụng danh sách biến đầu vào và nhãn phân loại; (3) Áp dụng mô hình cho toàn bộ ảnh với hàm *classify()* để tạo bản đồ phân loại. Việc so sánh hiệu suất giữa các mô hình được thực hiện thông qua hàm *confusionMatrix()* và *accuracy()* cho phép lựa chọn mô hình tối ưu.

2.3. Phương pháp đánh giá độ chính xác

Nghiên cứu này áp dụng phương pháp đánh giá độ chính xác của mô hình phân lớp nhị phân, sử dụng trên 4 chỉ số chính từ mô hình nhầm lẫn (confusion matrix) bao gồm Độ chính xác tổng thể (accuracy), độ chính xác dự đoán mục tiêu (precision), Độ nhạy – tỷ lệ phát hiện đúng mục tiêu (recall) và Trung bình điều hòa giữa Precision và Recall (F1 score). Điều này nhằm đảm bảo tính toàn diện và mức độ đáng tin cậy của mô hình [43, 44]. Accuracy thể hiện tỷ lệ phân loại đúng tổng thể, được tính bằng tổng số điểm phân loại đúng chia cho tổng số điểm kiểm tra [45]. Tuy nhiên, trong bài toán mất cân bằng dữ liệu, accuracy không phản ánh đầy đủ hiệu suất mô hình [46].

Precision (User's Accuracy) đo lường độ tin cậy của kết quả phân loại, thể hiện tỷ lệ pixel được phân loại chính xác trong tổng số pixel được gán vào lớp mất rừng [47]. Recall (Producer's Accuracy) phản ánh khả năng phát hiện đầy đủ các điểm mất rừng thực tế, được tính bằng tỷ lệ giữa số điểm mất rừng phát hiện đúng với tổng số điểm mất rừng thực địa [48]. Trong giám sát rừng, recall cao đặc biệt quan trọng để tránh bỏ sót các khu vực mất rừng [49].

F1-score là trung bình điều hòa của precision và recall, cung cấp đánh giá cân bằng về hiệu suất mô hình, đặc biệt hữu ích khi chi phí của false positive và false negative tương đương nhau [50]. Mô hình phát hiện mất rừng được coi là hiệu quả khi đạt F1-score trên 0,80, với recall tối thiểu 0,75 để đảm bảo phát hiện đầy đủ các khu vực mất rừng [51]. Việc kết hợp đánh giá đa chỉ số cho phép đánh giá toàn diện hiệu suất mô hình trong bối cảnh ứng dụng cụ thể [52].

Trong nghiên cứu này dự đoán mục tiêu (positive) là dự đoán mất rừng, dự đoán negative là dự đoán không mất rừng. Tổng hợp công thức tính các chỉ số đánh giá độ chính xác được xác định ở bảng 2 dưới đây.

Bảng 2. Tổng hợp công thức tính toán đánh giá độ chính xác trong nghiên cứu

	Mô hình dự đoán là mất rừng (Positive)	**Mô hình dự đoán là không mất rừng (Negative)**	
Thực tế là mất rừng	True positive (TP)	False negative (FN) (Lỗi loại II)	
Thực tế là không mất rừng (có rừng)	False positive (FP) (Lỗi loại I)	True negative (TN)	
Độ chính xác tổng thể (Accuracy)	$\frac{(TP + TN)}{(TP + TN + FP + FN)}$		(1)
Độ chính xác của dự đoán mục tiêu (Precision)	$\frac{(TP)}{(TP + FP)}$		(2)
Độ nhạy – Tỷ lệ phát hiện đúng mục tiêu (Recall)	$\frac{(TP)}{(TP + FN)}$		(3)
Trung bình điều hòa giữa Precision và Recall (F1 Score)	$\frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}$		(4)

3. Kết quả nghiên cứu và thảo luận

3.1. Kết quả nghiên cứu

3.1.1. Kết quả huấn luyện mô hình

Kết quả huấn luyện mô hình cho thấy cả ba thuật toán đều đạt hiệu suất huấn luyện tốt. Random Forest đạt độ chính xác trung bình cao nhất (97,99%), tiếp theo là GBT (97,94%) và SVM (97,73%). Hệ số Kappa tương ứng là 95,97%, 95,87% và 95,45%, khẳng định mức độ phù hợp rất tốt của các mô hình. Độ lệch chuẩn thấp giữa các fold (±1%) chứng tỏ tính ổn định cao trong quá trình huấn luyện. Cả ba mô hình đều vượt ngưỡng 96% về accuracy và 93% về Kappa ở mọi fold, cho thấy khả năng học tập đặc trưng phổ và chỉ số thực vật để phân biệt mất rừng/không mất rừng hiệu quả.

Bảng 3. Kết quả đánh giá huấn luyện mô hình

Fold	**Gradient Boosting Trees**		**Random Forest**		**Support Vector Machine**	
	Accuracy (%)	**Kappa**	**Accuracy (%)**	**Kappa**	**Accuracy (%)**	**Kappa**
1	98,68	0,9735	98,41	0,9683	98,68	0,9735
2	97,09	0,9414	97,62	0,9520	97,09	0,9414
3	98,15	0,9629	98,15	0,9629	97,88	0,9576
4	99,21	0,9841	98,94	0,9788	98,68	0,9735
5	96,58	0,9316	96,84	0,9368	96,32	0,9263
Mean ± SD	**97,94 ± 1,03**	**0,9587 ± 0,0207**	**97,99 ± 0,85**	**0,9597 ± 0,0169**	**97,73 ± 0,98**	**0,9545 ± 0,0196**

3.1.2. Phân tích ma trận nhầm lẫn của kết quả kiểm chứng

Sau khi huấn luyện, nghiên cứu đã tiến hành đánh giá 3 thuật toán học máy trên tệp dữ liệu kiểm chứng độc lập (test) gồm 250 điểm mất rừng, 250 điểm không mất rừng được thu thập trên 1 cặp ảnh mới vào thời gian 2022 – 2023. Ma trận nhầm lẫn của mô hình được trình bày tại Bảng 4.

Bảng 4. Ma trận nhầm lẫn của các mô hình học máy

Mô hình	TP	TN	FP	FN	Tổng dự đoán mất rừng	Tổng dự đoán có rừng
Gradient Boosting Trees (GBT)	180	247	3	70	183	317
Random Forest	162	248	2	88	164	336
Support Vector Machine (SVM)	147	247	3	103	150	350

Ghi chú: TP - True Positive (phát hiện đúng mất rừng); TN - True Negative (phát hiện đúng có rừng); FP - False Positive (nhận diện sai là mất rừng); FN - False Negative (bỏ sót mất rừng).

3.1.3. Phân tích kết quả kiểm chứng mô hình

Mô hình GBT đạt kết quả tốt nhất với 180/250 điểm mất rừng được phát hiện đúng (TP=180) và bỏ sót 70 điểm (FN=70). Mô hình chỉ có 3/250 điểm có rừng bị nhận diện nhầm là mất rừng (FP=3), thể hiện độ tin cậy cao trong việc cảnh báo. Tỷ lệ phát hiện đúng mất rừng đạt 72%, một cải thiện đáng kể so với nhiều nghiên cứu trước đây.

Mô hình Random Forest cho kết quả tốt thứ 2 với phát hiện 162/250 điểm mất rừng (TP=162), bỏ sót 88 điểm (FN=88). Với chỉ 2 cảnh báo sai (FP=2), mô hình thể hiện tính ổn định cao và độ chính xác dự đoán xuất sắc.

Mô hình SVM có hiệu suất thấp hơn nhưng vẫn đạt kết quả khả quan với 147/250 điểm mất rừng được phát hiện (TP=147), bỏ sót 103 điểm (FN=103). Với 3 cảnh báo sai (FP=3/250), mô hình vẫn duy trì được độ tin cậy cao khi phát hiện mất rừng, mặc dù khả năng phát hiện toàn diện còn hạn chế hơn so với hai mô hình còn lại.

Cả ba mô hình đều cho tỷ lệ cảnh báo sai rất thấp (FP ≤ 3), điều này đặc biệt quan trọng trong ứng dụng thực tế vì giúp tránh lãng phí nguồn lực khi kiểm tra thực địa các cảnh báo nhầm lẫn.

3.1.4. Các chỉ số đánh giá hiệu suất

Mô hình GBT cho hiệu suất cao nhất với độ chính xác tổng thể 85,40% và F1-score 83,14%, cao hơn đáng kể so với Random Forest (82,00% và 78,26%) và SVM (78,80% và 73,50%). Tất cả các mô hình đều đạt Precision trên 98%, chứng tỏ khi mô hình dự đoán một điểm là mất rừng, khả năng dự đoán đúng rất cao. Sự khác biệt giữa các mô hình là chỉ số Recall - khả năng phát hiện đầy đủ các điểm mất rừng thực tế. Mô hình GBT với Recall 72,00% phát hiện được gần 3/4 số điểm mất rừng, trong khi Random Forest phát hiện được 64,80% và SVM đạt 58,80%. Điều này cho thấy GBT có khả năng cân bằng tốt nhất giữa việc giảm cảnh báo sai và phát hiện đầy đủ các trường hợp mất rừng. F1-Score của GBT là 83,14%, Random Forest là 78,26% và SVM với 73,50%. Kết quả xác định GBT là mô hình phù hợp nhất cho triển khai thực tế tại Cà Mau.

Bảng 5. Kết quả đánh giá hiệu suất tổng hợp các mô hình

Mô hình	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Gradient Boosting Trees	85,40	98,36	72,00	83,14
Random Forest	82,00	98,78	64,80	78,26
Support Vector Machine	78,80	98,00	58,80	73,50

Kết quả nghiên cứu cho thấy không có sự chênh lệch đáng kể về độ chính xác giữa 3 mô hình khi huấn luyện (Bảng 6). Random Forest cho kết quả cao nhất 97,99%, cao hơn 0,05% so với GBT và 0,26% so với SVM, qua đó cho thấy giai đoạn huấn luyện mô hình cho kết quả tốt. Kết quả kiểm chứng mô hình cho thấy GBT giảm ít nhất với -12,54%, Random Forest là -15,99% và -18,93% ở SVM. Kết quả này phản ánh mức độ phức tạp khi nghiên cứu mất rừng trên đối tượng chủ yếu là rừng ngập mặn. Trong nghiên cứu sử dụng tệp test độc lập trên một kỳ ảnh hoàn toàn khác (2022-2023), do vậy có thể đã có

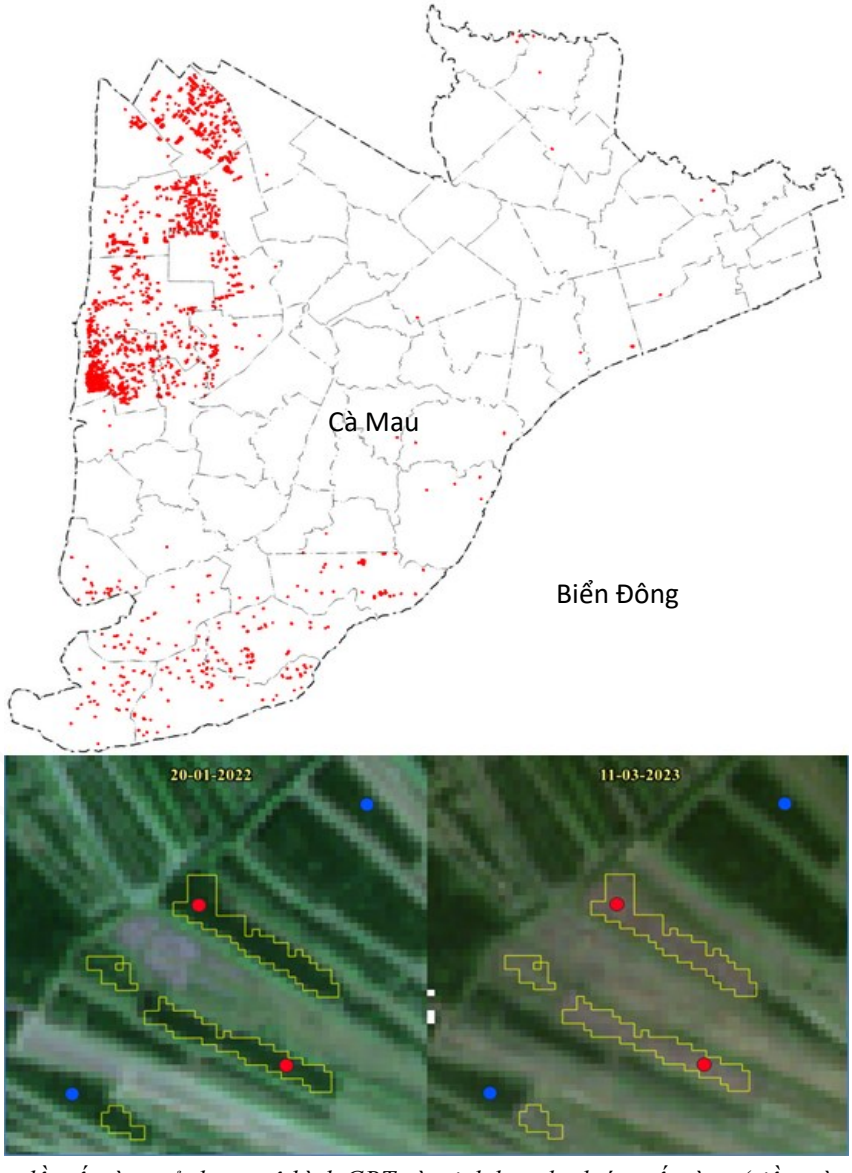
những yếu tố môi trường đã ảnh hưởng đến kết quả kiểm chứng. Đồng thời, nghiên cứu cần tiếp tục tăng cường dung lượng mẫu cho huấn luyện mô hình và đưa thêm các biến số mới nhằm nâng cao độ chính xác dự đoán của mô hình.

Bảng 6. So sánh độ chính xác giữa kết quả huấn luyện và kiểm chứng mô hình

Mô hình	Huấn luyện (%)	Kiểm chứng (%)	Chênh lệch (%)
Gradient Boosting Trees	97,94	85,40	-12,54
Random Forest	97,99	82,00	-15,99
Support Vector Machine	97,73	78,80	-18,93

3.1.5. Bản đồ mất rừng tại khu vực nghiên cứu

Nghiên cứu sử dụng mô hình GBT để xây dựng bản đồ mất rừng cho khu vực nghiên cứu sử dụng kỳ ảnh 20-01-2022 và 11-03-2023. Bản đồ này đã được vector hóa để tiện cho công tác xác minh, đối chiếu với dữ liệu mất rừng trong quá khứ. Nhóm nghiên cứu cũng đã phát triển thành công công cụ tạo ra bản đồ mất rừng ngày một cách tự động trên nền tảng GEE.



Hình 5. Bản đồ mất rừng sử dụng mô hình GBT và minh họa dự đoán mất rừng (viền màu vàng, điểm mẫu kiểm chứng mất rừng màu đỏ, và điểm không mất rừng màu xanh da trời).

3.2. Thảo luận

Nghiên cứu đã xây dựng thành công hệ thống phát hiện mất rừng tự động với mô hình Gradient Boosting Trees đạt độ chính xác 85,40% và F1-Score 83,14%. Đặc biệt là chỉ số precision cao (>98%) ở tất cả các mô hình, cho thấy độ tin cậy cao khi hệ thống cảnh báo mất rừng. Độ chính xác của nghiên cứu này cao hơn so với nghiên cứu tại tỉnh Thừa Thiên Huế khi chỉ chỉ sử dụng chỉ số chênh lệch chỉ số NDVI giữa 2 kỳ (70%) [34]. Các chỉ số về tỷ lệ phát hiện đúng điểm mất rừng, tỷ lệ bỏ sót (recall 72%) và F1-Score (83,14%) cho thấy mô hình GBT là đáng tin cậy và có thể đưa vào ứng dụng trong thực tế. Việc triển khai toàn bộ quy trình phát hiện mất rừng trên nền tảng GEE chứng minh tính khả thi của giám sát rừng quy mô lớn với chi phí thấp, phù hợp với điều kiện của các địa phương.

Kết quả xác thực chéo (k-fold cross validation) cho thấy sự ổn định cao trong quá trình huấn luyện với độ lệch chuẩn thấp (~1%). Tuy nhiên, chênh lệch giữa độ chính xác huấn luyện (~98%) và kiểm chứng (78,8 - 85,4%) cho thấy thách thức trong việc tổng quát hóa mô hình cho dữ liệu mới. Điều này phù hợp với đặc thù phức tạp của rừng ngập mặn Cà Mau chịu ảnh hưởng mạnh bởi thủy triều và biến động mùa vụ [51].

Hạn chế của nghiên cứu là chưa khắc phục được hoàn toàn ảnh hưởng của tỷ lệ mây lớn trong mùa mưa, dẫn tới việc có thể phát hiện rừng chậm. Nghiên cứu hiện đang dừng lại ở việc đánh giá mất rừng theo điểm, chưa đánh giá được diện tích thiệt hại do mô hình dự đoán và diện tích thực tế. Tỷ lệ bỏ sót điểm mất rừng còn ở mức trung bình (28%). Do đó, chúng tôi dự kiến tiếp tục nghiên cứu này và tích hợp Sentinel-1 radar có xuyên mây nhằm khắc phục các hạn chế về mây trên ảnh Sentinel-2 [53, 54]. Phương pháp này được kỳ vọng sẽ cải thiện cả độ chính xác và khả năng phát hiện kịp thời mất rừng, đặc biệt trong điều kiện thời tiết bất lợi của vùng nhiệt đới gió mùa.

4. Kết luận

Nghiên cứu đã xây dựng thành công hệ thống phát hiện mất rừng tự động tại Cà Mau sử dụng ảnh Sentinel-2 và thuật toán học máy trên nền tảng Google Earth Engine. Gradient Boosting Trees được xác định là thuật toán phù hợp nhất với độ chính xác 85,40%, recall 72,00%, precision 98,36%và F1-score 83,14%. Mặc dù recall còn hạn chế nhưng precision rất cao cho thấy độ tin cậy của hệ thống khi cảnh báo mất rừng, giúp tối ưu nguồn lực kiểm tra thực địa.

Việc áp dụng k-fold cross-validation đã cung cấp đánh giá toàn diện về hiệu suất mô hình, với độ chính xác huấn luyện ổn định (~98%) qua các fold. Tuy nhiên, hạn chế về mây che phủ trong mùa mưa và khả năng phát hiện chưa toàn diện (recall 72,00%) cần được cải thiện trong các nghiên cứu tiếp theo. Do đó, cần tiếp tục nghiên cứu tích hợp dữ liệu radar Sentinel-1 để khắc phục vấn đề mây che phủ và nâng cao khả năng phát hiện mất rừng.

Kết quả nghiên cứu cung cấp công cụ hữu ích cho Chi cục Kiểm lâm Cà Mau trong công tác giám sát và quản lý rừng, với khả năng phát hiện 72% các trường hợp mất rừng và độ chính xác cảnh báo trên 98%. Phương pháp được phát triển có thể nhân rộng cho các tỉnh ven biển có điều kiện sinh thái tương tự, góp phần vào nỗ lực bảo vệ và phát triển bền vững tài nguyên rừng ngập mặn tại Việt Nam.

Lời cảm ơn

Nhóm tác giả xin trân trọng cảm ơn Ủy ban Nhân dân tỉnh Cà Mau và Sở Khoa học và Công nghệ tỉnh Cà Mau đã phê duyệt, cấp kinh phí và tạo điều kiện thuận lợi cho việc thực hiện nghiên cứu này. Nghiên cứu là một phần kết quả của đề tài cấp tỉnh "Ứng dụng công nghệ viễn thám, ảnh vệ tinh để quản lý, giám sát tài nguyên rừng và sạt lở ven biển

trên địa bàn tỉnh Cà Mau". Chúng tôi cung xin gửi lời cảm ơn Chi cục Kiểm lâm tỉnh Cà Mau đã hỗ trợ chúng tôi trong quá trình khảo sát, thu thập dữ liệu điểm mẫu trên địa bàn tỉnh.

Cam kết của các tác giả

Tất cả các tác giả có tên trong bài báo cam kết sự đồng thuận và không có xung đột lợi ích trong công bố khoa học tại bài báo này.

Tài liệu tham khảo

[1] Knehtl, M., Petkovska, V., and Urbanič, G., "Is it time to eliminate field surveys from hydromorphological assessments of rivers?—Comparison between a field survey and a remote sensing approach," Ecohydrology, vol. 11, no. 2, p. e1924, 2018.
[2] Mumby, P. J., Green, E. P., Edwards, A. J., and Clark, C. D., "The cost-effectiveness of remote sensing for tropical coastal resources assessment and management," Journal of Environmental Management, vol. 55, no. 3, pp. 157-166, 1999.
[3] Asner, G., "Cloud cover in Landsat observations of the Brazilian Amazon," International Journal of Remote Sensing, vol. 22, pp. 3855-3862, 2001.
[4] Eberhardt, I. D. et al., "Cloud Cover Assessment for Operational Crop Monitoring Systems in Tropical Areas," Remote Sensing, vol. 8, no. 3, 2016.
[5] He, J. et al., "Recent advances and challenges in monitoring and modeling of disturbances in tropical moist forests," Frontiers in Remote Sensing, Mini Review vol. Volume 5 - 2024, 2024.
[6] Thanh Son, N., Chen, C.-F., Chang, N.-B., Chen, C.-R., Chang, L.-Y., and Bui, X.-T., "Mangrove Mapping and Change Detection in Ca Mau Peninsula, Vietnam, Using Landsat Data and Object-Based Image Analysis," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, pp. 503-510, 2015.
[7] Chitra, N. T., Anusha, R., Kumar, S., Chandana, D. S., Harika, C., and Kumar, V. U., Satellite Imagery for Deforestation Prediction using Deep Learning. pp. 522-525, 2021
[8] Nichols, K. and Hosein, P., Estimating Deforestation using Machine Learning Algorithms. pp. 82-87, 2021
[9] Saha, S., Bhattacharjee, S., Shit, P. K., Sengupta, N., and Bera, B., "Deforestation probability assessment using integrated machine learning algorithms of Eastern Himalayan foothills (India)," Resources, Conservation & Recycling Advances, vol. 14, p. 200077, 2022.
[10] Torres, D. L. et al., "Deforestation Detection with Fully Convolutional Networks in the Amazon Forest from Landsat-8 and Sentinel-2 Images," Remote Sensing, vol. 13, no. 24, 2021.
[11] Vo, Q. T., Oppelt, N., Leinenkugel, P., and Kuenzer, C., "Remote Sensing in Mapping Mangrove Ecosystems — An Object-Based Approach," Remote Sensing, vol. 5, no. 1, pp. 183-201, 2013.
[12] Ủy ban nhân dân tỉnh Cà Mau, (2025), Quyết định về việc công bố hiện trạng rừng tỉnh Cà Mau năm 2024.
[13] Drusch, M. et al., "Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services," Remote Sensing of Environment, vol. 120, pp. 25-36, 2012.
[14] Pham, T.-D., Xia, J., Baier, G., Le, N., and Yokoya, N., Mangrove Species Mapping Using Sentinel-1 and Sentinel-2 Data in North Vietnam. 2019
[15] Hoang, T. T., Truong, V. T., Hayashi, M., Tadono, T., and Nasahara, K. N., "New JAXA High-Resolution Land Use/Land Cover Map for Vietnam Aiming for Natural Forest and Plantation Forest Monitoring," Remote Sensing, vol. 12, no. 17, 2020.
[16] Nguyen, H.-H., Bich, N. T. N., Nghia, N. H., Lan, T. T. N., Nguyen, T. T. H., and Dang, H. V., "Estimation of changes in above-ground biomass and carbon stocks of mangrove forests using sentinel-2a in Thai Thuy district, Thai Binh province during 2015 - 2019," Vietnam Journal of Science and Technology, vol. 60, no. 1, pp. 73-91, 2022.
[17] Immitzer, M., Vuolo, F., and Atzberger, C., "First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe," Remote Sensing, vol. 8, p. 166, 2016.
[18] Hoscilo, A. and Lewandowska, A., "Mapping Forest Type and Tree Species on a Regional Scale Using Multi-Temporal Sentinel-2 Data," Remote Sensing, vol. 11, p. 929, 2019.
[19] Schuster, C., Förster, M., and Kleinschmit, B., "Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data," International Journal of Remote Sensing, vol. 33, pp. 5583-5599, 2012.

[20] Pradhan, S., Tiwari, K., and Dhar, A., "Evaluation of Sentinel 2 Red Edge Channel for Enhancing Land Use Classification," 2021, pp. 79-89.
[21] Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R., "Google Earth Engine: Planetary-scale geospatial analysis for everyone," Remote Sensing of Environment, vol. 202, pp. 18-27, 2017.
[22] Praticò, S., Solano, F., Di Fazio, S., and Modica, G., "Machine Learning Classification of Mediterranean Forest Habitats in Google Earth Engine Based on Seasonal Sentinel-2 Time-Series and Input Image Composition Optimisation," Remote Sensing, vol. 13, no. 4, 2021.
[23] Chen, B. et al., "A mangrove forest map of China in 2015: Analysis of time series Landsat 7/8 and Sentinel-1A imagery in Google Earth Engine cloud computing platform," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 131, pp. 104-120, 2017.
[24] Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., and Gascon, F., Sen2Cor for Sentinel-2. p. 3, 2017
[25] Skakun, S. et al., "Cloud Mask Intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2," Remote Sensing of Environment, vol. 274, p. 112990, 2022.
[26] European Space, A., "Sentinel-2 User Handbook," in "ESA Standard Document," European Space Agency2015.
[27] Delegido, J., Verrelst, J., Alonso, L., and Moreno, J., "Evaluation of Sentinel-2 Red-Edge Bands for Empirical Estimation of Green LAI and Chlorophyll Content," Sensors, vol. 11, no. 7, pp. 7063-7081, 2011.
[28] Wang, L., Silván-Cárdenas, J., and Sousa, W., "Neural Network Classification of Mangrove Species from Multi-seasonal Ikonos Imagery," Photogrammetric Engineering & Remote Sensing, vol. 74, pp. 921-927, 2008.
[29] Rouse, J., Haas, R., Schell, J., and Deering, D., "Monitoring Vegetation Systems in the Great Plains with ERTS," NASA Special Publication, vol. 1, 1974.
[30] Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and Ferreira, L. G., "Overview of the radiometric and biophysical performance of the MODIS vegetation indices," Remote Sensing of Environment, vol. 83, no. 1, pp. 195-213, 2002.
[31] Huete, A. R., "A soil-adjusted vegetation index (SAVI)," Remote Sensing of Environment, vol. 25, no. 3, pp. 295-309, 1988.
[32] Key, C. and Benson, N., "Landscape Assessment: Ground measure of severity, the Composite Burn Index; and Remote sensing of severity, the Normalized Burn Ratio," 2006, pp. LA 1-51.
[33] Lu, D., Mausel, P., Brondízio, E., and Moran, E., "Change Detection Techniques," International Journal of Remote Sensing, vol. 25, 2004.
[34] Thái, V. V., Nguyễn, H.-H. H., Lê Thị Quỳnh, Ngọc, N. M., Hiếu, T. P. H., and Liêm, N. D., "Sử dụng chỉ số viễn thám phát hiện mất rừng trên ảnh Sentinel-2 tại tỉnh Thừa Thiên-Huế," Tạp chí Khoa học và Công nghệ Lâm nghiệp, vol. 4, 2021.
[35] Miller, J. D. and Thode, A. E., "Quantifying burn severity in a heterogeneous landscape with a relative version of the delta Normalized Burn Ratio (dNBR)," Remote Sensing of Environment, vol. 109, no. 1, pp. 66-80, 2007.
[36] Lyons, M. B., Keith, D. A., Phinn, S. R., Mason, T. J., and Elith, J., "A comparison of resampling methods for remote sensing classification and accuracy assessment," Remote Sensing of Environment, vol. 208, pp. 145-153, 2018.
[37] A. Ramezan, C., A. Warner, T., and E. Maxwell, A., "Evaluation of Sampling and Cross-Validation Tuning Strategies for Regional-Scale Machine Learning Classification," Remote Sensing, vol. 11, no. 2. doi: 10.3390/rs11020185
[38] Belgiu, M. and Drăguţ, L., "Random forest in remote sensing: A review of applications and future directions," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 114, pp. 24-31, 2016.
[39] Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., and Rigol-Sanchez, J. P., "An assessment of the effectiveness of a random forest classifier for land-cover classification," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 67, pp. 93-104, 2012.
[40] Friedman, J., "Greedy Function Approximation: A Gradient Boosting Machine," The Annals of Statistics, vol. 29, 2000.
[41] Mountrakis, G., Im, J., and Ogole, C., "Support vector machines in remote sensing: A review," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 66, no. 3, pp. 247-259, 2011.

[42] Pham, T. D., Yokoya, N., Bui, D. T., Yoshino, K., and Friess, D. A., "Remote Sensing Approaches for Monitoring Mangrove Species, Structure, and Biomass: Opportunities and Challenges," Remote Sensing, vol. 11, no. 3, 2019.
[43] Rainio, O., Teuho, J., and Klén, R., "Evaluation metrics and statistical tests for machine learning," Scientific Reports, vol. 14, 2024.
[44] Congalton, R. and Green, K., Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, Third Edition. 2019
[45] Foody, G. M., "Status of land cover classification accuracy assessment," Remote Sensing of Environment, vol. 80, no. 1, pp. 185-201, 2002.
[46] He, H. and Garcia, E. A., "Learning from Imbalanced Data," Knowledge and Data Engineering, IEEE Transactions on, vol. 21, pp. 1263-1284, 2009.
[47] Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A., "Good practices for estimating area and assessing accuracy of land change," Remote Sensing of Environment, vol. 148, pp. 42-57, 2014.
[48] Story, M. and Congalton, R. G., "Accuracy assessment: a user's perspective," Photogrammetric Engineering, vol. 52, no. 3, pp. 397-399, 1986.
[49] Hansen, M. C. et al., "High-Resolution Global Maps of 21st-Century Forest Cover Change," Science (New York, N.Y.), vol. 342, pp. 850-853, 2013.
[50] Sokolova, M. and Lapalme, G., "A systematic analysis of performance measures for classification tasks," Information Processing & Management, vol. 45, pp. 427-437, 2009.
[51] Bullock, E. L., Woodcock, C. E., and Olofsson, P., "Monitoring tropical forest degradation using spectral unmixing and Landsat time series analysis," Remote Sensing of Environment, vol. 238, p. 110968, 2020.
[52] Pontius, R. and Millones, M., "Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment," International Journal of Remote Sensing, vol. 32, pp. 4407-4429, 2011.
[53] Dascălu, A., Catalão, J., and Navarro, A., "Detecting Deforestation Using Logistic Analysis and Sentinel-1 Multitemporal Backscatter Data," Remote Sensing, vol. 15, no. 2, 2023.
[54] Vargas, C., Itoh, T., Tsuji, S., Koide, T., Hirose, K., and Okonogi, H., "Automatic Deforestation Detection Methodology Using Sentinel-1," in IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, 2019, pp. 6590-6593.

Article © 2024 by Magazine of Geodesy - Cartography is licensed under CC BY 4.0