

NGHIÊN CỨU MỘT SỐ PHƯƠNG PHÁP HỌC MÁY TRONG THÀNH LẬP BẢN ĐỒ LỚP PHỦ BỀ MẶT TỈNH CÀ MAU TRÊN NỀN TẢNG GOOGLE EARTH ENGINE

TRẦN VÂN ANH^(1,4), TRẦN HỒNG HẠNH^(1,4), LÊ THANH NGHỊ^(1,4)
PHẠM THI THANH HÒA^(1,4), TRẦN QUÝ ANH⁽²⁾, NGUYỄN THỊ VÂN ANH⁽³⁾
PHẠM THỊ THU⁽²⁾, TRẦN QUÝ THỌ⁽³⁾

⁽¹⁾Bộ môn Đo ảnh và Viễn thám, Trường Đại học Mở - Địa chất

⁽²⁾Công ty TNHH MTV Trắc địa Bản đồ, CN miền Trung, Cục Bản đồ-Bộ Tổng Tham mưu

⁽³⁾Cục Đo đạc Bản đồ và hệ Thông tin Địa lý Việt Nam

⁽⁴⁾Nhóm nghiên cứu Công nghệ Địa tin học trong Khoa học Trái đất (GES),
Trường Đại học Mở - Địa chất

Tóm tắt:

Trong bài báo này, chúng tôi đã xây dựng bản đồ lớp phủ đất tỉnh Cà Mau, Việt Nam sử dụng 3 phương pháp phân loại khác nhau là rừng ngẫu nhiên (RF), Máy véc tơ hỗ trợ (SVM) và Gradient boosting (Gboost). Khu vực nghiên cứu là Cà Mau thuộc Đồng bằng sông Cửu Long, Việt Nam. Ảnh vệ tinh dùng để phân loại là ảnh Sentinel-2 đa thời gian từ tháng 1 đến tháng 12 năm 2021. Số ảnh giai đoạn này sau khi loại bỏ mây còn lại 17 ảnh. Phương pháp lọc trung bình đã được sử dụng để tạo ra một hình ảnh duy nhất trong khoảng thời gian này để sử dụng cho phân loại. Công cụ để thực hiện việc phân loại là nền tảng Google Earth Engine. Các mẫu phân loại được lấy dựa trên bản đồ sử dụng đất tỉnh Cà Mau năm 2014 và ảnh Google Earth. Số lượng mẫu lấy để phân loại cho cả 3 phương pháp là gần 4000 pixel và số lượng mẫu để đánh giá độ chính xác là 3000 pixel. Sai số tổng thể (OA) của SVM là 79,5%, hệ số Kappa là 0,72 trong khi phương pháp Gboost đạt 85,6% với Kappa là 0,79 và RF có OA là 86,5%, Kappa đạt 0,81. Ảnh phân loại theo phương pháp RF có độ chính xác cao nhất đã được chọn để xây dựng bản đồ lớp phủ bề mặt tỷ lệ 1:50.000 của tỉnh Cà Mau.

Từ khóa: Lớp phủ bề mặt, RF, SVM, Gboost, Cà Mau, GEE.

1. Giới thiệu

Cà mau là một tỉnh nằm ở cực Nam của Tổ quốc với ba mặt giáp biển. Trong những năm gần đây, do biến đổi khí hậu và nước biển dâng, tình trạng sụt lún, xói lở bờ sông, bờ

biển diễn ra nghiêm trọng. Chính vì vậy, quy hoạch đất nông nghiệp cũng phải thay đổi cho phù hợp và thích ứng với tình hình biến đổi khí hậu. Hiện nay, công nghệ viễn thám đã trở nên phổ biến và giúp các nhà khoa học xây

dựng các loại hình sử dụng đất, thảm phủ, quy hoạch hay các lĩnh vực khác một cách dễ dàng. Để có thể xây dựng được các loại bản đồ này, điều quan trọng nhất là thuật toán phân loại phục vụ cho việc tách các lớp dữ liệu sử dụng đất khác nhau. Trong những năm gần đây, các phương pháp học máy đã trở nên phổ biến và chúng đã được chứng minh là có khả năng và có độ chính xác tốt. Một trong những thuật toán học máy đầu tiên cần được kể tên là rừng ngẫu nhiên (RF) [1]. Sự phổ biến của thuật toán này là nó có thể được sử dụng cho cả mục đích phân loại và hồi quy, do đó có thể được sử dụng với các biến phân loại và liên tục. Do tính linh hoạt này, RF đã được sử dụng trong nhiều ngành khoa học Trái đất. Các ứng dụng bao gồm mô hình hóa lớp phủ rừng [2], thành lập bản đồ sử dụng đất [3], bản đồ lớp phủ đất [4] và lập bản đồ lớp phủ sử dụng đất bằng kết hợp giữa học đồng bộ của RF và phân loại hướng đối tượng [5].

Phương pháp phân loại Máy vector hỗ trợ (SVM) là một thuật toán học máy dựa trên lý thuyết học thống kê đã được sử dụng rộng rãi trong cộng đồng viễn thám. Chúng dường như hoạt động tốt hơn hầu hết các bộ phân loại thông thường [6], [7], [8]. Hơn nữa, SVM thậm chí còn vượt trội so với một số phương pháp nhận dạng mẫu mới, chẳng hạn như mạng nơ ron [6], [9]. Trong nghiên cứu của Foody và Mathur đã khẳng định việc lựa chọn mẫu trong phân loại SVM là một việc quan trọng sẽ ảnh hưởng đến độ chính xác phân loại [9].

Phương pháp Gradient boosting (Gboost) là một thuật toán lần đầu tiên được mô tả bởi Schapire, 1999 [10], thuật toán này có thể được hiểu là một thuật toán tối ưu hóa trên một hàm chi phí phù hợp. Một trong những ứng dụng viễn thám trên cơ sở cải tiến Gboost là XGBoost được thực hiện bởi [10] sử dụng tối

ưu hóa tham số Bayes trên dữ liệu WorldView-3 có độ phân giải rất cao. Họ nhận thấy rằng Xgboost vượt trội hơn SVM 0,3%. Xgboost cũng được phát hiện là tốt hơn một chút so với RF 0,2% trong một nghiên cứu phân loại LULC sáu lớp gần đây sử dụng dữ liệu có độ phân giải cao từ RapidEye [11].

Ba phương pháp học máy này đều cho độ chính xác khá tốt theo như các bài báo đã khảo sát ở trên. Tuy nhiên, khu vực nghiên cứu là khu vực khá phức tạp do có nhiều kiểu thảm thực vật và thay đổi theo mùa. Vì vậy, nghiên cứu này muốn so sánh khả năng phân loại của ba thuật toán học máy phổ biến RF, SVM và GBoost để thành lập bản đồ lớp phủ đất cho khu vực nghiên cứu Cà Mau, Việt Nam sử dụng ảnh Sentinel-2.

Gần đây với sự ra mắt của Google Earth Engine (GEE) - một nền tảng điện toán đám mây để phân tích không gian địa lý, giúp dễ dàng truy cập các tài nguyên tính toán hiệu suất cao để xử lý dữ liệu vệ tinh đa thời gian trực tuyến [12]. Tính sẵn có của dữ liệu ảnh vệ tinh và các công cụ xử lý ảnh trên GEE cho phép người dùng có thể thực hiện các thuật toán học máy như RF, SVM hay Gboost một cách nhanh chóng. Mục tiêu chính của bài báo là ứng dụng các nguồn ảnh trên GEE như Sentinel-2 và các thuật toán sẵn đã đề cập ở trên để phân loại lớp phủ khu vực Cà Mau phục vụ thành lập bản đồ lớp phủ bề mặt.

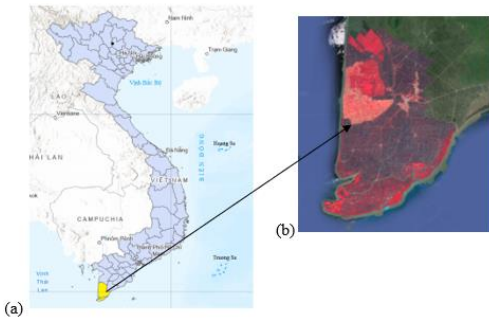
2. Khu vực và dữ liệu nghiên cứu

2.1. Khu vực nghiên cứu

Tỉnh Cà Mau là vùng đất tận cùng của Tổ quốc với 3 mặt giáp biển, có vị trí địa lý: Phía Đông giáp Biển Đông với bờ biển dài 107 km. Phía Tây và Nam giáp Vịnh Thái Lan với bờ biển dài 147 km. Bắc giáp tỉnh Bạc Liêu và tỉnh Kiên Giang. Cà Mau là vùng đất trũng, thường xuyên bị ngập lụt. Cà Mau có 5 nhóm

đất chính gồm: đất phèn, đất than bùn, đất phù sa, đất mặn và đất kênh rạch.

(Công thông tin điện tử Cà Mau).



Hình 1: (a) Vị trí Cà Mau ở Việt Nam; (b) Tổ hợp màu giả của hình ảnh Sentinel-2 trên nền tảng điện toán đám mây Google Earth Engine

2.2. Ảnh vệ tinh

Để thành lập bản đồ lớp phủ đất, chúng tôi đã chọn ảnh Sentinel-2. Sentinel-2 là một vệ tinh đa phổ, hỗ trợ các nghiên cứu giám sát đất đai của Copernicus, bao gồm giám sát thảm thực vật, lớp phủ đất và nước. Hình ảnh Sentinel-2 có độ phân giải không gian là 10 m và 20 m với chu kỳ lặp lại là 12 ngày. Thời gian thu thập hình ảnh là từ tháng 1 năm 2021 đến tháng 12 năm 2021.

3. Phương pháp nghiên cứu

3.1. Máy véc tơ hỗ trợ (SVM)

Máy véc tơ hỗ trợ (SVM) lần đầu tiên được mô tả trong [13] dựa trên công trình của Vapnik (1982) và là một kỹ thuật học có giám sát thường được sử dụng trong nhiều ứng dụng viễn thám. Thuật toán SVM tìm ra sự giảm thiểu tối ưu, tức là ranh giới quyết định, của các đầu ra của bộ phân loại rõ ràng trong một không gian bài toán. Ranh giới quyết định này được gọi là một siêu phẳng và nó phân tách vấn đề phân loại thành một tập hợp các lớp được xác định trước phù hợp với dữ liệu huấn luyện. Thuật toán trải qua một quy trình lặp để tìm ranh giới siêu phẳng tối ưu trong không gian bộ phân loại n chiều để phân biệt các mẫu

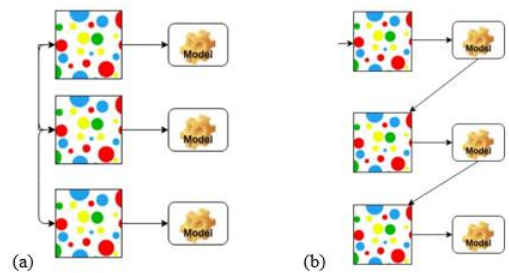
trong dữ liệu huấn luyện, sau đó áp dụng cấu hình tương tự cho một tập hợp dữ liệu đánh giá riêng biệt. Các kích thước trong ngữ cảnh này là số dải quang phổ và các vectơ là các pixel riêng lẻ trong tập hợp nhiều kênh. Có các nhân khác nhau mà qua đó ranh giới siêu phẳng có thể được xác định. Có thể tìm thấy mô tả toán học chi tiết về thuật toán này trong Cortes và Vapnik (1995) [13].

3.2. Rừng ngẫu nhiên (RF)

Rừng ngẫu nhiên là một thuật toán bao gồm nhiều cây quyết định đơn lẻ hoạt động giống như các hiệp hội. Mỗi cây riêng lẻ trong Rừng ngẫu nhiên đưa ra dự đoán về lớp và lớp có nhiều phiếu bình chọn nhất sẽ trở thành dự đoán của mô hình đó (Hình 2a) [14], [15].

Như tên gọi của nó, Random Forest (RF) dựa trên: (1) Random nghĩa là Tính ngẫu nhiên; (2) Rừng tương ứng với nhiều cây quyết định.

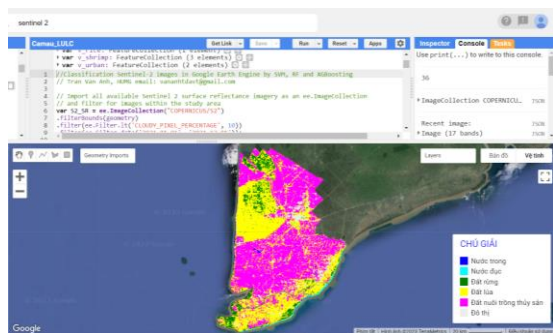
Đơn vị của RF là thuật toán cây quyết định với hàng trăm cây. Mỗi cây quyết định được tạo ngẫu nhiên từ lấy mẫu ngẫu nhiên và chỉ sử dụng một phần nhỏ các tính năng ngẫu nhiên từ tất cả các biến trong dữ liệu. Rừng ngẫu nhiên khi thực hiện tại mỗi lần tách cây, mô hình chỉ xem xét một tập hợp con nhỏ của các đối tượng hơn là tất cả các đối tượng của mô hình.



Hình 2: (a) Sơ đồ hình thành RF (Các cây được bố trí song song), (b) Sơ đồ hình thành Gboost (Các cây được bố trí một cách tuần tự)

3.3. Phương pháp Gboost

Ý tưởng cơ bản của phương pháp Tăng cường độ dốc (GBoost) là tạo ra một loạt các mô hình yếu học cách bổ sung cho nhau. Nói cách khác, trong Boosting, các mô hình sau sẽ cố gắng học hỏi để hạn chế những sai lầm của các mô hình trước (dữ liệu mà mô hình trước dự đoán sai), việc này tạo thành một chuỗi các mô hình mà cái sau sẽ tốt hơn cái trước bởi trọng số được cập nhật qua mỗi mô hình (cụ thể ở đây là trọng số của những dữ liệu dự đoán đúng sẽ không đổi, còn trọng số của những dữ liệu dự đoán sai sẽ được tăng thêm). Chúng ta sẽ lấy kết quả của mô hình cuối cùng trong chuỗi mô hình này làm kết quả trả về. Hình 2(b) mô tả phương pháp Gboost với chuỗi các mô hình tuần tự. Chi tiết thuật toán Gboost được trình bày trong nghiên cứu [16].



Hình 3: Giao diện viết code thực thi lệnh của GEE

3.4. Công cụ và các bước

Công cụ được sử dụng trong nghiên cứu này là Google Earth Engine (GEE). GEE hoạt động thông qua giao diện trực tuyến ứng dụng JavaScript (API) hoặc Python được gọi là trình chỉnh sửa mã. Trên giao diện này, người dùng có thể viết và chạy các tập lệnh để chia sẻ và lặp lại các quy trình xử lý và phân tích dữ liệu không gian địa lý. Code Editor giúp người dùng thực hiện tất cả các chức năng có trong Earth Engine. Dữ liệu đầu vào có thể khai thác trực tiếp trên các trang WEB dữ liệu

mà không cần tải xuống máy tính[14]. Hình 3 là giao diện của GEE.

Để tiến hành phân loại ảnh trên GEE cần trải qua một số bước cơ bản [14]. Sử dụng ngôn ngữ lập trình Javascript, chúng tôi đã xử lý dữ liệu theo lưu đồ (Hình 4) để đưa ra ba cách phân loại cho khu vực Cà Mau.

Đối với phân loại ảnh thì mẫu phân loại cần được lựa chọn rất cẩn thận. Mẫu trong nghiên cứu này lấy trên bản đồ sử dụng đất tỷ lệ 1:50.000 năm 2014 và ảnh google Earth. Các kênh ảnh được sử dụng để phân loại là Kênh 1 đến Kênh 8 và NDVI. Chúng tôi đã lựa chọn 3 phương pháp phân loại là SVM, RF và GBoost. Cả ba phương pháp đều có thể sử dụng cùng một bộ mẫu để phân loại và đánh giá độ chính xác nên bộ mẫu này được lấy cùng nhau. Gần 3.386 pixel được lấy mẫu cho 6 lớp khác nhau bao gồm: nước trong (bao gồm nước biển và sông lớn), nước đục (gồm nước ao hồ và nước biển gần bờ), rừng, đất lúa, đất nuôi trồng thủy sản, đất đô thị và 3.000 pixel được lấy mẫu để đánh giá độ chính xác của 3 phương pháp phân loại.

4. Kết quả và thảo luận

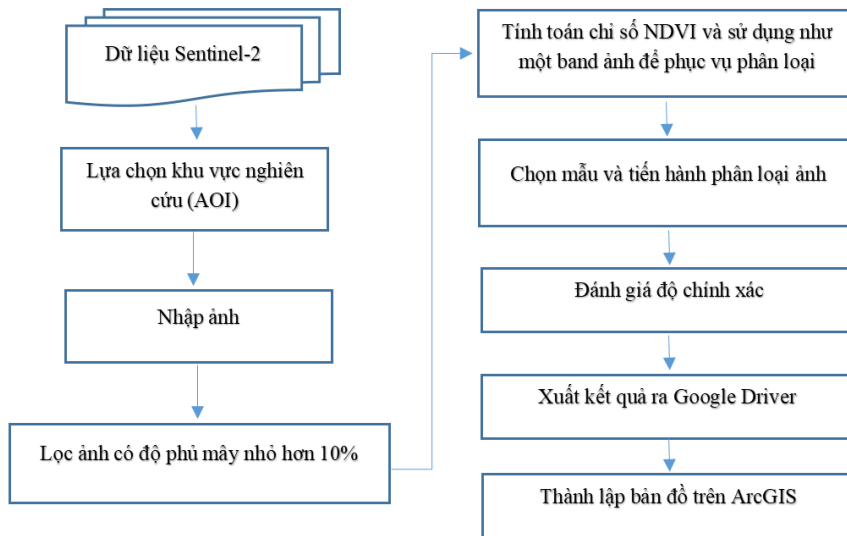
Theo sơ đồ xử lý ảnh Hình 4, kết quả phân loại khu vực Cà Mau từ ba phương pháp được thể hiện trong Hình 5. Các bảng 1, 2 và 3 trình bày kết quả đánh giá độ chính xác cho ba phương pháp. Kết quả cho thấy với 3 phương pháp phân loại được lựa chọn thì RF và GBoost đều cho độ chính xác tốt hơn so với SVM. Bảng 1, 2 và 3 thể hiện kết quả đánh giá độ chính xác bao gồm ma trận sai số, sai số tổng thể (OA), hệ số Kappa, độ chính xác của Người sử dụng (UA) và độ chính xác của Nhà sản xuất (PA). Với lớp rừng, cả hai phương pháp RF và GBoost đều có độ chính xác phân loại cao hơn so với SVM. Nhìn vào ma trận nhầm lẫn chúng ta dễ dàng nhận thấy khả năng

tách Rừng với lớp Ruộng lúa của SVM kém hơn rất nhiều khi có tới 433 pixel bị lẫn vào lớp Ruộng lúa trong khi với 2 phương pháp còn lại chỉ là 206 pixel với Gboost và 199 pixel với RF. Đối với lớp nuôi trồng thủy sản, cả ba phương pháp đều có sai số tương tự nhau. Khu vực nghiên cứu là vùng đất trũng với nhiều đầm nuôi tôm nên khó phân biệt giữa ruộng nuôi tôm và ruộng lúa nước, vì vậy, dễ dàng nhận thấy với cả 3 cách phân loại này, sự nhầm lẫn giữa ruộng nuôi tôm và ruộng lúa nước vẫn khá cao. Tương tự đối với các lớp nuôi trồng thủy sản, khu vực Cà Mau có đất đô thị không nhiều, chủ yếu xen lẫn với đất trồng trọt nên việc phân loại cũng có sự nhầm lẫn với các lớp khác. Giá trị UA của đô

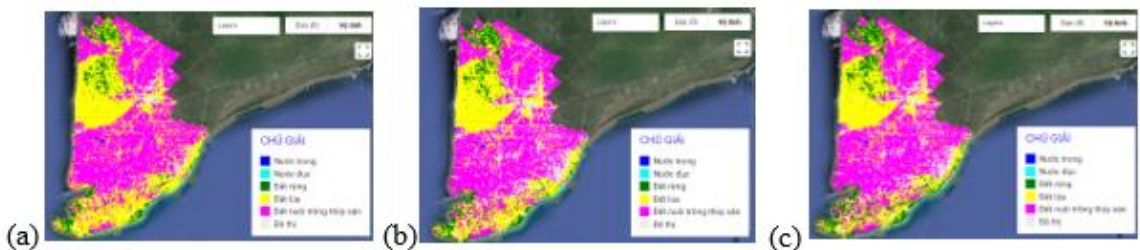
thị nhỏ hơn 40% cho cả ba phương pháp, điều này cho thấy khả năng phân loại khu vực đô thị khi lẫn nhiều thực phủ vẫn còn là một vấn đề đáng quan tâm.

Sai số tổng thể (OA) của SVM là 79,5%, hệ số kappa là 0,72 trong khi phương pháp Gboost đạt 85,6% và Kappa: 0,79 và RF có OA: 86,5% và Kappa: 0,81.

Qua các đánh giá trên, kết quả phân loại từ RF được chọn vì có sai số tổng thể OA và hệ số Kappa cao nhất. Kết hợp với điều tra thực địa, chúng tôi đã thành lập được bản đồ tham phủ tỉnh Cà Mau trên phần mềm ArcGIS 10.4 (Hình 6).



Hình 4: Lưu đồ xử lý ảnh phục vụ phân loại ảnh trên GEE



Hình 5: (a) Phân loại SVM, (b) Phân loại RF, (c) Phân loại Gboost

Bảng 1: Độ chính xác của phân loại bằng SVM

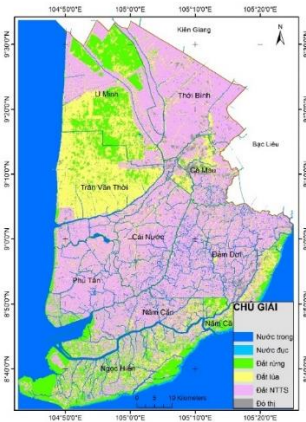
Độ chính xác				Ma trận nhầm lẫn					
Các lớp học	UA (%)	PA (%)		Nước trong	Nước đục	Rừng	Ruộng lúa	Đất nuôi trồng thủy sản	Đô thị
Nước trong	89,6	95,2	Nước trong	640	0	0	0	32	0
Nước đục	99,7	88,9	Nước đục	43	401	0	0	2	5
Rừng	99,9	75,2	Rừng	0	0	1319	433	0	1
Ruộng lúa	31,7	99,6	Ruộng lúa	0	0	1	254	0	0
Đất nuôi trồng thủy sản	80,9	47,6	Đất nuôi trồng thủy sản	22	1	0	112	174	56
Đô thị	34,0	66,6	Đô thị	9	0	0	0	7	32
Độ chính xác tổng thể: 79,5% và Kappa: 0,72									

Bảng 2: Độ chính xác của phân loại bằng GBoost với 300 cây

Độ chính xác				Ma trận nhầm lẫn					
Các lớp phân loại	UA (%)	PA (%)		Nước trong	Nước đục	Rừng	Ruộng lúa	Đất nuôi trồng thủy sản	Đô thị
Nước trong	88,3	96,8	Nước trong	651	0	0	0	21	0
Nước đục	98,6	83,1	Nước đục	67	375	0	0	7	2
Rừng	98,9	86,6	Rừng	0	0	1542	206	1	41
Ruộng lúa	45,3	97,6	Ruộng lúa	0	0	6	249	0	0
Đất nuôi trồng thủy sản	81,4	49,3	Đất nuôi trồng thủy sản	19	5	11	94	180	56
Đô thị	36,7	75,0	Đô thị	0	0	0	0	12	36
Độ chính xác tổng thể: 85,6% và Kappa: 0,79									

Bảng 3: Độ chính xác của phân loại bằng RF với 300 cây

Độ chính xác				Ma trận nhầm lẫn					
Các lớp phân loại	UA (%)	PA (%)		Nước trong	Nước đục	Rừng	Ruộng lúa	Đất nuôi trồng thủy sản	Đô thị
Nước trong	91,3	95,6	Nước trong	643	0	0	0	29	0
Nước đục	99,7	88,4	Nước đục	43	399	0	1	7	1
Rừng	99,8	88,8	Rừng	0	0	1558	190	1	4
Ruộng lúa	45,2	99,2	Ruộng lúa	0	0	2	253	0	0
Đất nuôi trồng thủy sản	78,4	48,7	Đất nuôi trồng thủy sản	18	1	0	115	178	53
Đô thị	38,3	75,0	Đô thị	0	0	0	0	12	36
Độ chính xác tổng thể: 86,5% và Kappa: 0,81									



Hình 6: Bản đồ lớp phủ bề mặt tỉnh Cà Mau

5. Kết luận

Qua sử dụng ảnh vệ tinh Sentinel-2 để xây dựng bản đồ lớp phủ bề mặt khu vực tỉnh Cà Mau, Việt Nam, chúng tôi nhận thấy ảnh Sentinel-2 có độ phân giải khá tốt để có thể xây dựng bản đồ lớp phủ tỷ lệ 1:50.000.

Trong ba phương pháp phân loại ảnh được lựa chọn là SVM, GBoost và RF thì phương pháp GBoost và RF cho độ chính xác gần như nhau nhưng RF có độ chính xác nhỉnh hơn một chút với OA là 86.5% và Kappa là 0.81. SVM là phương pháp có độ chính xác thấp nhất trong 3 phương pháp và đặc biệt với phương pháp này, độ phân biệt giữa hai đối tượng giống nhau là rừng và ruộng lúa kém hơn nhiều so với phương pháp RF và Gboost. ○

Lời cảm ơn

Nhóm nghiên cứu chân thành cảm ơn sự hỗ trợ về mặt tài chính từ đề tài Nghiên cứu khoa học của Bộ Giáo dục và Đào tạo Việt Nam, mã số: B2022-MDA-13.

Tài liệu tham khảo

- [1]. Breiman, L.: Random forests. *Machine learning* 45, 5-32 (2001).
- [2]. Betts, M.G., Wolf, C., Ripple, W.J., Phalan, B., Millers, K.A., Duarte, A., Butchart, S.H., Levi, T.: Global forest loss

disproportionately erodes biodiversity in intact landscapes. *Nature* 547, 441-444 (2017).

- [3]. Araki, S., Shima, M., Yamamoto, K.: Spatiotemporal land use random forest model for estimating metropolitan NO₂ exposure in Japan. *Science of The Total Environment* 634, 1269-1277 (2018).

- [4]. Nitze, I., Barrett, B., Cawkwell, F.: Temporal optimisation of image acquisition for land cover classification with Random Forest and MODIS time-series. *International Journal of Applied Earth Observation and Geoinformation* 34, 136-146 (2015).

- [5]. Kavzoglu, T.: Object-oriented random forest for high resolution land cover mapping using quickbird-2 imagery. *Handbook of neural computation*, pp. 607-619. Elsevier (2017).

- [6]. Huang, C., Davis, L., Townshend, J.: An assessment of support vector machines for land cover classification. *International Journal of remote sensing* 23, 725-749 (2002).

- [7]. Kavzoglu, T., Colkesen, I.: A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation* 11, 352-359 (2009).

- [8]. Su, L., Huang, Y.: Support vector machine (svm) classification: comparison of linkage techniques using a clustering-based method for training data selection. *GIScience & Remote Sensing* 46, 411-423 (2009).
- [9]. Foody, G.M., Mathur, A.: Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. *Remote Sensing of Environment* 93, 107-117 (2004).
- [10]. Schapire, R.E.: A brief introduction to boosting. In: *Ijcai*, pp. 1401-1406. Citeseer, (Year).
- [11]. Man, C.D., Nguyen, T.T., Bui, H.Q., Lasko, K., Nguyen, T.N.T.: Improvement of land-cover classification over frequently cloud-covered areas using Landsat 8 time-series composites and an ensemble of supervised classifiers. *International Journal of Remote Sensing* 39, 1243-1255 (2018).
- [12]. Tran, V.A., Le, T.L., Nguyen, N.H., Le, T.N., Tran, H.H.: Monitoring Vegetation Cover Changes by Sentinel-1 Radar Images Using Random Forest Classification Method. *Inżynieria Mineralna* (2021).
- [13]. Cortes, C., Vapnik, V.: Support-vector networks. *Machine learning* 20, 273-297 (1995).
- [14]. Van Anh, T., Hanh, T. H., Nga, N. Q., Nghi, L. T., Quang, T. X., Dong, K. T., & Anh, T. T.. Determination of Illegal Signs of Coal Mining Expansion in Thai Nguyen Province, Vietnam from a Combination of Radar and Optical Imagery, Springer Cham (2023).
- [15]. Van Anh, T., Hang, L., Duong, N.: Determination of land cover change by multi-temporal radar imagery. In: 40th Asian Conference on Remote Sensing, ACRS 2019: Progress of Remote Sensing Technology for Smart Future. Asian Association on Remote Sensing, (2019).
- [16]. Mason, D.C., Giustarini, L., Garcia-Pintado, J., Cloke, H.L.: Detection of flooded urban areas in high resolution Synthetic Aperture Radar images using double scattering. *International Journal of Applied Earth Observation and Geoinformation* 28, 150-159 (2014).○

Summary

Research on machine learning methods for land cover mapping of Ca Mau province using Google Earth Engine

Tran Van Anh, Tran Hong Hanh, Lê Thanh Nghi, Pham Thi Thanh Hoa

Department of Photometry and Remote Sensing, Hanoi University of Mining and Geology

Tran Quy Anh, Pham Thi Thu

Middle region Branch of Survey and Aerial Mapping One Member Limited Liability Company

Tran Quy Tho, Nguyen Thi Van Anh

Vietnam Department Of Survey, Mapping and Geographic Information

In this paper, we present a land cover map of Ca Mau province in Vietnam, which was generated using three different classification methods: Random forest (RF), Support Vector Machine (SVM), and Gradient Boost (Gboost). The study area is a lowland region

with complex land cover in the Mekong Delta. To create the map, we used multi-temporal Sentinel-2 satellite images from January to December 2021, with 17 images remaining after cloud removal. A mean filtering method was employed to generate a single image for classification. The Google Earth Engine platform was used to perform the classifications, with sample data drawn from the 2014 land use map of Ca Mau province and Google Earth images. We collected nearly 4000 pixels for classification for all three methods, with an additional 3000 pixels used to evaluate accuracy. The SVM method produced an overall error (OA) of 79.5% and a kappa coefficient of 0.72, while the Gboost method achieved 85.6% accuracy with a Kappa of 0.79. The RF method had the highest accuracy, with an OA of 86.5% and Kappa of 0.81. We selected the land cover map generated by the RF method using the most accurate image for a scale of 1:50.000.○

Keywords: Land cover, RF, SVM, Gboost, Camau, GEE.

ẢNH HƯỞNG CỦA TƯ THẾ VỆ TINH.....

(Tiếp theo trang 7)

Summary

Effects of satellite attitude on PPP accuracy when using code products

Nguyen Ngoc Lau

Department of Geomatics Engineering, Ho Chi Minh City University of Technology, Vietnam

Vietnam National University Ho Chi Minh City, Vietnam

Nguyen Thi Thanh Huong

Institute of Geodesy and Cartography, Vietnam

The product package of CODE supports PPP-AR GPS+GALILEO with satellite orbit every 5 minutes and satellite attitude every 15 minutes. However, these products do not give us the best PPP accuracy [18]. To find out if this problem is caused by satellite attitude, we used PPP to process 90 days of data from January 1, 2022, to March 31, 2022, for 4 permanent GNSS stations in Southeast Asia according to 2 options to replace the CODE satellite attitude file: using the nominal satellite attitude and using the GFZ's satellite attitude file with an interval of 30 seconds. The results show that the PPP accuracy of the two options is the same, and the improvement when using the CODE satellite attitude file is (12, 27, 8)%, respectively, in the North, East and Elevation components. Accuracy is also similar to when using CNES products. Therefore, we conclude that using satellite attitude files with an interval of 15 minutes is inappropriate because it reduces the accuracy of PPP.○

Keywords: ITRF2014, ITRF2020, PPP, CORS, Vietnam.