

NGHIÊN CỨU PHÂN LOẠI LỚP PHỦ SỬ DỤNG ĐẤT TỪ ẢNH VỆ TINH SENTINEL - 2 DỰA TRÊN THUẬT TOÁN SVM

Nguyễn Thị Huyền Trang, Đặng Thanh Tùng
Phạm Thị Sao Mai, Hà Thị Thu Hồng
Trường Đại học Tài nguyên và Môi trường Hà Nội

Tóm tắt

Ảnh vệ tinh Sentinel - 2 là nguồn cung cấp miễn phí có độ phân giải về không gian và thời gian cao, dữ liệu này có khả năng đem lại hiệu quả trong việc phân loại các lớp phủ sử dụng đất phục vụ công tác giám sát, quản lý đất đai. Trên nền tảng dữ liệu điện toán đám mây Google Earth Engine (GEE), nghiên cứu này sử dụng thuật toán Support Vector Machine (SVM) áp dụng cho máy học để phân loại hiện trạng các lớp phủ bề mặt sử dụng đất tại khu vực quận Bắc Từ Liêm, Hà Nội. Kết quả của nghiên cứu đưa ra 4 loại lớp phủ sử dụng đất cơ bản, trong đó diện tích lớp phủ mặt nước là nhỏ nhất với 422,31 ha, diện tích lớn nhất là lớp phủ thực vật đạt 1876,97 ha. Độ chính xác phân loại đạt kết quả tốt với hệ số Kappa đạt 0,82.

Từ khóa: Sentinel - 2; Lớp phủ sử dụng đất; Google Earth Engine; Thuật toán SVM.

Abstract

Research classification of land use - land cover from Sentinel - 2 satellite images based on the SVM algorithm

The Sentinel - 2 satellite imagery is a free source with a high spatial and temporal resolution, this data has to be effective in classifying land use - land cover for monitoring, and land management. Based on the Google Earth Engine (GEE), this study uses the Support Vector Machine (SVM) algorithm applied to machine learning to classify the land use - land cover status in Bac Tu Liem district, Hanoi. The results of the study show that there are 4 basic types of land use - land cover, in which the water - bodies area is the smallest with 422.31 hectares and the largest area is the cover of Vegetation with 1876.97 hectares. The classification accuracy achieved good results with the Kappa coefficient reaching 0.82.

Keywords: Sentinel - 2; Land use - land cover; Google Earth Engine; SVM algorithms.

1. Mở đầu

Hiện trạng các lớp phủ sử dụng đất là tài liệu rất cần thiết cho việc thống kê, kiểm kê đất đai, đồng thời là thông tin thực sự hữu ích cho công tác giám sát thực hiện quy hoạch sử dụng đất. Việc ứng dụng công nghệ thông tin kết hợp với công nghệ viễn thám là một trong những giải pháp hiệu quả, nhanh chóng,

tiết kiệm chi phí để hỗ trợ thành lập bản đồ hiện trạng sử dụng đất và nó đặc biệt có ý nghĩa là có thể xác định ở những thời điểm gần như theo thời gian thực.

Một số thuật toán truyền thống trước đây như Maximum Likelihood Classifier (MLC), Minimum Distance Classifier (MDC), K - Nearest Neighbor (KNN) thường được ứng dụng trong các phần mềm

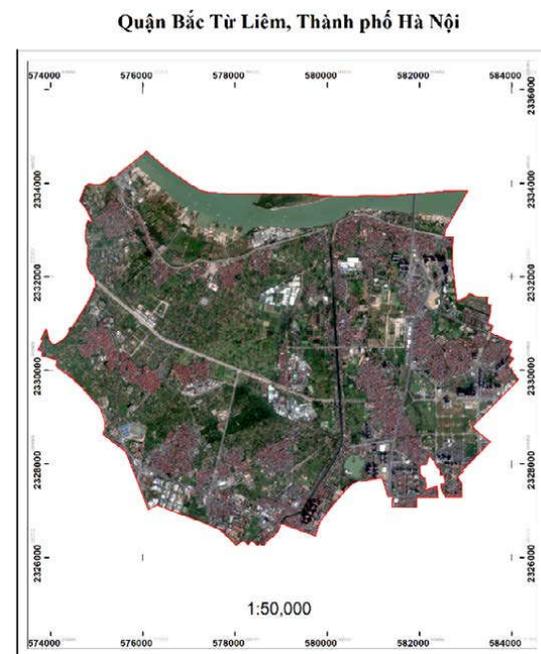
thương mại như ArcGIS, ErdasImagine, Envi, ER - Mapper,... [1 - 3]. Hiện nay, cách mạng 4.0 phát triển mạnh mẽ đã tạo ra nhiều ưu thế cho việc xây dựng hiện trạng các lớp phủ sử dụng đất từ ảnh vệ tinh bằng việc sử dụng các thuật toán học máy, trong đó có những thuật toán nổi bật như Support Vector Machine (SVM), Random Forest (RF) và Classification and regression tree (Cart),...[4 - 7].

Nhiều nghiên cứu trên thế giới và tại Việt Nam đã sử dụng các thuật toán học máy để phân loại các lớp phủ bề mặt từ ảnh vệ tinh, các kết quả nghiên cứu này đã chứng tỏ được hiệu quả của việc xác định hiện trạng lớp phủ bề mặt. Có thể kể đến một số nghiên cứu trên thế giới như của Lewis, R. J. (2000) [5], Breiman L. J. M. I. (2001) [6], Jia, K., và cộng sự (2014) [8]. Và cũng đã có những nghiên cứu tại Việt Nam như của Phạm Minh Hải và cộng sự (2019) [4], Bùi Thị Hồng Thắm, Trịnh Thị Hoài Thu (2020) [9], Đặng Thanh Tùng (2021) [10]. Các nghiên cứu trên thường dùng dữ liệu ảnh vệ tinh Landsat là ảnh vệ tinh có độ phân giải trung bình, cụ thể với ảnh Landsat 8 có độ phân giải không gian là 30 m và độ phân giải thời gian là 16 ngày. Bài báo này trình bày kết quả phân loại hiện trạng các lớp phủ sử dụng đất từ ảnh vệ tinh Sentinel - 2 có độ phân giải không gian 10 m và chu kỳ chụp lặp lại là 5 ngày, thử nghiệm tại khu vực quận Bắc Từ Liêm, thành phố Hà Nội. Các kết quả được phân tích, xử lý dựa nền tảng điện toán đám mây Google Earth Engine (GEE) và ứng dụng thuật toán học máy SVM.

2. Dữ liệu và khu vực nghiên cứu

Quận Bắc Từ Liêm được thành lập cuối năm 2013 từ việc chia tách huyện

Từ Liêm trước đây thành hai đơn vị hành chính cấp quận là Nam Từ Liêm và Bắc Từ Liêm [11]. Năm 2020, quận có diện tích 45,24 km², dân số năm 2020 là 340.605 người, mật độ 7.529 người/km², đơn vị hành chính gồm có 13 phường. Từ khi khu vực trở thành quận Bắc Từ Liêm cho đến nay, tốc độ phát triển cơ sở hạ tầng, kinh tế, dân cư,... đã có những biến động rõ rệt, kèm theo đó là các biến động nhanh chóng về các loại hình sử dụng đất. Hình 1 minh họa vị trí khu vực nghiên cứu.



Hình 1: Hình ảnh, vị trí khu vực nghiên cứu

Nghiên cứu này, nhóm tác giả sử dụng tư liệu ảnh vệ tinh Sentinel - 2 được chụp ngày 19 tháng 12 năm 2022. Ảnh vệ tinh được khai thác trực tuyến thông qua nền tảng GEE. Ảnh vệ tinh Sentinel - 2 gồm có ảnh Sentinel - 2A và Sentinel - 2B, tần suất quay lại vị trí cụ thể của mỗi vệ tinh Sentinel đơn lẻ này là 10 ngày. Việc kết hợp cả Sentinel - 2A và Sentinel - 2B tạo ra bộ ảnh vệ tinh Sentinel - 2 có độ phân giải thời

Nghiên cứu

gian là 5 ngày. Ảnh vệ tinh Sentinel - 2 được lưu giữ ở mức độ 12 bits. Do đó, dữ liệu ảnh này có độ phân giải bức xạ cao. Điều này đem lại khả năng phân biệt mức độ sáng trên ảnh có phạm vi

tiềm năng từ 0 - 4 095. Các thông tin, đặc điểm cơ bản về độ phân giải không gian, bước sóng các kênh ảnh của dữ liệu ảnh vệ tinh Sentinel - 2 được trình bày tại Bảng 1 dưới đây:

Bảng 1. Thông tin cơ bản về dữ liệu ảnh vệ tinh Sentinel - 2

Độ phân giải không gian (m)	Kênh ảnh	Sentinel - 2A	Sentinel - 2B
		Bước sóng (nm)	Bước sóng (nm)
10	2	492,4	492,1
	3	559,8	559,0
	4	664,6	664,9
	8	832,8	832,9
20	5	704,1	703,8
	6	740,5	739,1
	7	782,8	779,7
	8a	864,7	864,0
	11	1613,7	1610,4
60	12	2202,4	2185,7
	1	442,7	442,2
	9	945,1	943,2
	10	1373,5	1376,9

Dữ liệu ảnh vệ tinh Sentinel - 2 trong khu vực nghiên cứu được lựa chọn đảm bảo có chất lượng rõ ràng, độ phủ mây khoảng 0,05 %. Với các thông số như độ phân giải cao, chu kỳ thu nhận ảnh lặp lại là tương đối ngắn (5 ngày), độ rộng dải quét lên tới 290 km là rất lớn so với chiều rộng dải quét của Landsat 5 TM và Landsat 7 ETM+ là 185 km, SPOT - 5 là 120 km, các điều kiện này cho thấy ảnh vệ tinh Sentinel - 2 miễn phí là nguồn dữ liệu rất phù hợp để áp dụng cho nghiên cứu này.

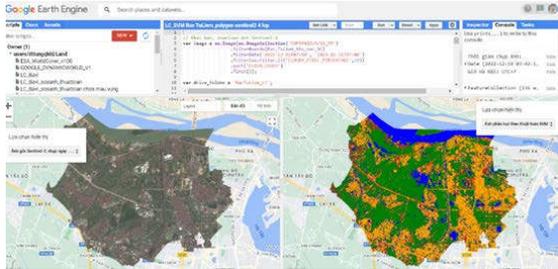
3. Phương pháp nghiên cứu

3.1. Tổng quan quy trình nghiên cứu

Đối với nghiên cứu này, thuật toán SVM được áp dụng để tính toán, phân loại lớp phủ sử dụng đất với dữ liệu đầu vào là ảnh vệ tinh Sentinel - 2. Đây là một trong những thuật toán được áp dụng nhiều và đã

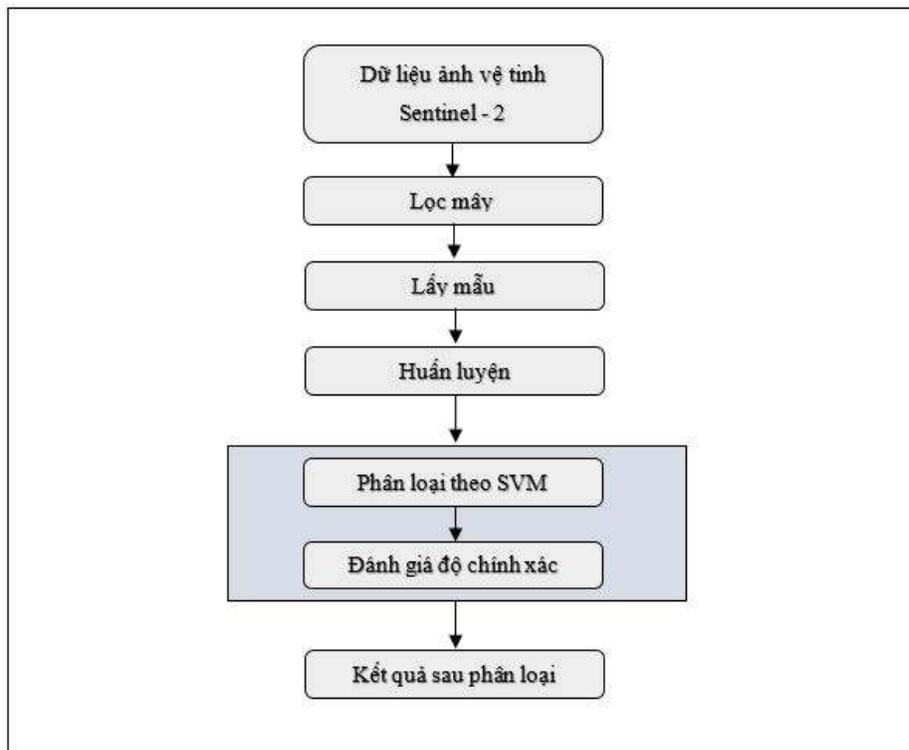
được chứng minh tính hiệu quả trong phân loại lớp phủ sử dụng đất từ dữ liệu ảnh vệ tinh [12, 13]. Khu vực Bắc Từ Liêm được lựa chọn thử nghiệm phân loại với 4 lớp phủ sử dụng đất cơ bản bao gồm: 1) Lớp phủ đất trống, 2) Lớp phủ mặt nước, 3) Lớp phủ thực vật, 4) Lớp phủ nhà - CTXD. Các lớp phủ được lựa chọn để phân loại là hiện trạng tại thời điểm thu nhận ảnh. Đặc điểm của lớp đất trống được lựa chọn là các khu vực như bãi đất, cát, bãi bồi, khu vực đang san lấp, các khu vực đất trống mới san lấp chuẩn bị xây dựng. Lớp phủ mặt nước gồm các khu vực ao hồ, sông suối, kênh mương. Lớp phủ thực vật bao gồm những khu vực có cây cối, bãi cỏ, ruộng vườn. Lớp phủ nhà - CTXD gồm các công trình xây dựng độc lập và các khu vực nhà cửa, công trình xây dựng tại làng mạc, khu chung cư, các tòa nhà cao tầng. Tổng số lượng điểm lấy mẫu là 136 điểm cho 4 lớp phủ của toàn bộ

khu vực nghiên cứu. Việc thực hiện phân loại các lớp phủ sử dụng đất được tiến hành trên nền tảng dữ liệu điện toán đám mây GEE và sử dụng ngôn ngữ lập trình JavaScript.



Hình 2: Thực hiện chương trình phân loại lớp phủ trên GEE

Quy trình thực hiện theo từng bước, các bước lần lượt bao gồm: Trước hết là lấy dữ liệu ảnh vệ tinh Sentinel - 2 của khu vực nghiên cứu; Sau đó lọc mây đảm bảo có dữ liệu tốt nhất có thể; Tiến hành việc huấn luyện để cho chương trình học và ghi nhận các thông tin về các mẫu cho từng loại lớp phủ; Phân loại các lớp phủ theo mẫu đã huấn luyện; Ghi nhận kết quả và đánh giá độ chính xác phân loại. Hình 3 minh họa tổng quan quy trình thực hiện nghiên cứu.

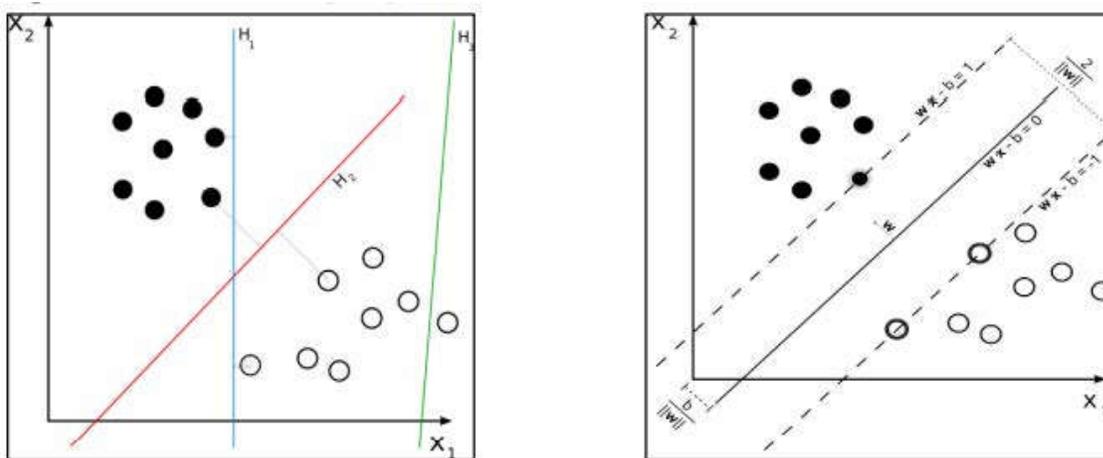


Hình 3: Sơ đồ quy trình phân loại các lớp phủ sử dụng đất

3.2. Thuật toán SVM

SVM là một thuật toán học máy được giám sát thường mang lại kết quả tốt trong quá trình phân loại và hồi quy. SVM phân chia các vector hỗ trợ nhằm phân loại các

điểm dữ liệu nhật ký một cách rõ ràng với mục tiêu tìm ra hai loại vectơ hỗ trợ độc lập có biên độ lớn nhất [14, 15]. Hình 4 dưới đây minh họa mô hình thuật toán SVM.



Hình 4: Mô hình phân loại các lớp theo thuật toán SVM

Với bộ dữ liệu mẫu được huấn luyện thuộc các lớp cho trước, thuật toán SVM xây dựng mô hình để phân loại các đối tượng vào các lớp cho trước đó. Mục tiêu của thuật toán này là tìm được một không gian F và siêu phẳng quyết định f trên F sao cho sai số phân loại là thấp nhất. Việc phân lớp được thực hiện qua hàm quyết định:

$$f(x) = \text{sign}(+ b)$$

Hàm f(x) này thu được bằng việc thay đổi vectơ chuẩn w, đây là vectơ để cực đại hóa khoảng cách tới các siêu phẳng (Optimal hyperplane). Các điểm dữ liệu cho trước nằm trên các siêu phẳng song song được gọi là Support Vector. Trong một không gian có nhiều điểm và các kí hiệu như sau:

y_i : là các lớp chứa các điểm dữ liệu x_i .

x_i : là vector thực nhiều chiều (p chiều).

Mỗi siêu phẳng đều có thể được viết dưới dạng một tập các điểm thỏa mãn $w \cdot x - b = 0$.

w: là 1 vector pháp tuyến của siêu phẳng.

$b/\|w\|$: xác định khoảng cách giữa gốc tọa độ và siêu phẳng theo hướng vectơ pháp tuyến w.

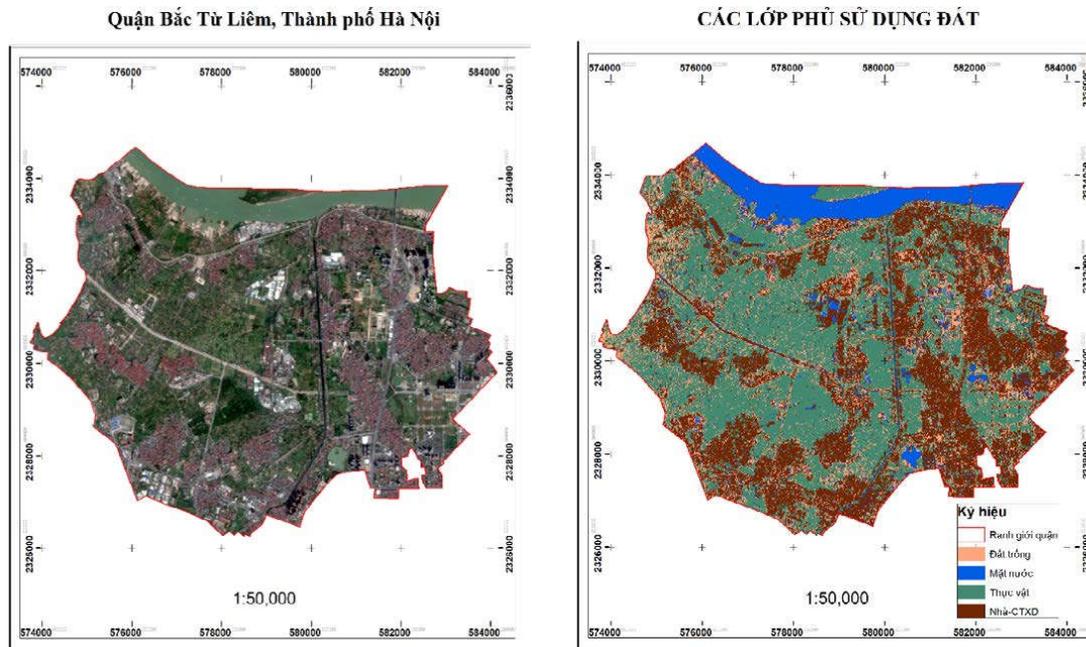
Trong Hình 4, ta có H2 là siêu phẳng cần tìm. Lúc này các siêu phẳng đó được xác định như sau: $w \cdot x - b = 1$ và $w \cdot x - b = -1$.

3.3. Phương pháp đánh giá độ chính xác

Phương pháp đánh giá độ chính xác phân loại các lớp phủ sử dụng đất trong nghiên cứu này được thực hiện thông qua ma trận nhầm lẫn (Confusion Matrix). Ma trận nhầm lẫn là phương pháp quan trọng và phổ biến được sử dụng để đánh giá độ chính xác, có thể mô tả độ chính xác của phân loại và chỉ ra sự nhầm lẫn giữa các lớp đối tượng. Các thống kê cơ bản cho ma trận nhầm lẫn trong đó có đánh giá về độ chính xác tổng thể (Overall Accuracy - OA) và hệ số Kappa. Trong đó OA có giá trị càng gần tới 1 thì càng có độ tin cậy cao, hệ số Kappa có giá trị từ 0,4 đến 0,6 được đánh giá là đạt kết quả trung bình, giá trị từ lớn hơn 0,6 đến 0,8 là tốt và hơn 0,8 đến 1,0 là rất tốt [15].

4. Kết quả nghiên cứu và thảo luận

Kết quả phân loại các lớp phủ sử dụng đất theo thuật toán SVM được minh họa tại Hình 5. Ảnh sau phân loại bao gồm 4 lớp phủ cơ bản là lớp phủ Đất trống, lớp phủ Mặt nước, lớp phủ Thực vật, lớp phủ Nhà - CTXD.

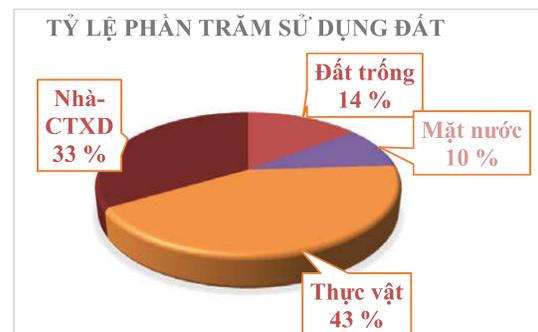


Ảnh vệ tinh Sentinel - 2

Ảnh phân loại các lớp phủ

Hình 5: Ảnh gốc và ảnh sau phân loại khu vực quận Bắc Từ Liêm, thành phố Hà Nội

Ảnh sau phân loại theo thuật toán SVM được thể hiện rõ ràng. Trong đó các lớp phủ chiếm đa số là lớp phủ Thực vật và lớp phủ Nhà - CTXD. Hai lớp phủ này có độ che phủ từ khoảng 30 - 40 % cho cho mỗi lớp. Các lớp phủ còn lại là lớp Đất trống và lớp Mặt nước có độ che phủ từ khoảng 10 % đến 15 % so với tổng diện tích tự nhiên. Kết quả của phân bố các lớp phủ như trên thể hiện sự phù hợp với đặc điểm tự nhiên, phân bố dân cư, tập quán canh tác, sản xuất, kinh doanh,... tại khu vực quận Bắc Từ Liêm, thành phố Hà Nội. Cụ thể lớp Đất trống có diện tích khoảng 638,5 ha, tương đương 14,5 % so với tổng diện tích tự nhiên. Lớp Mặt nước là 422,31 ha (9,5 %), lớp Thực vật 1876,9 ha (42,5 %), lớp Nhà - CTXD chiếm diện tích 1477,5 ha (33,5%). Biểu đồ tỷ lệ sử dụng đất được minh họa tại Hình 6.



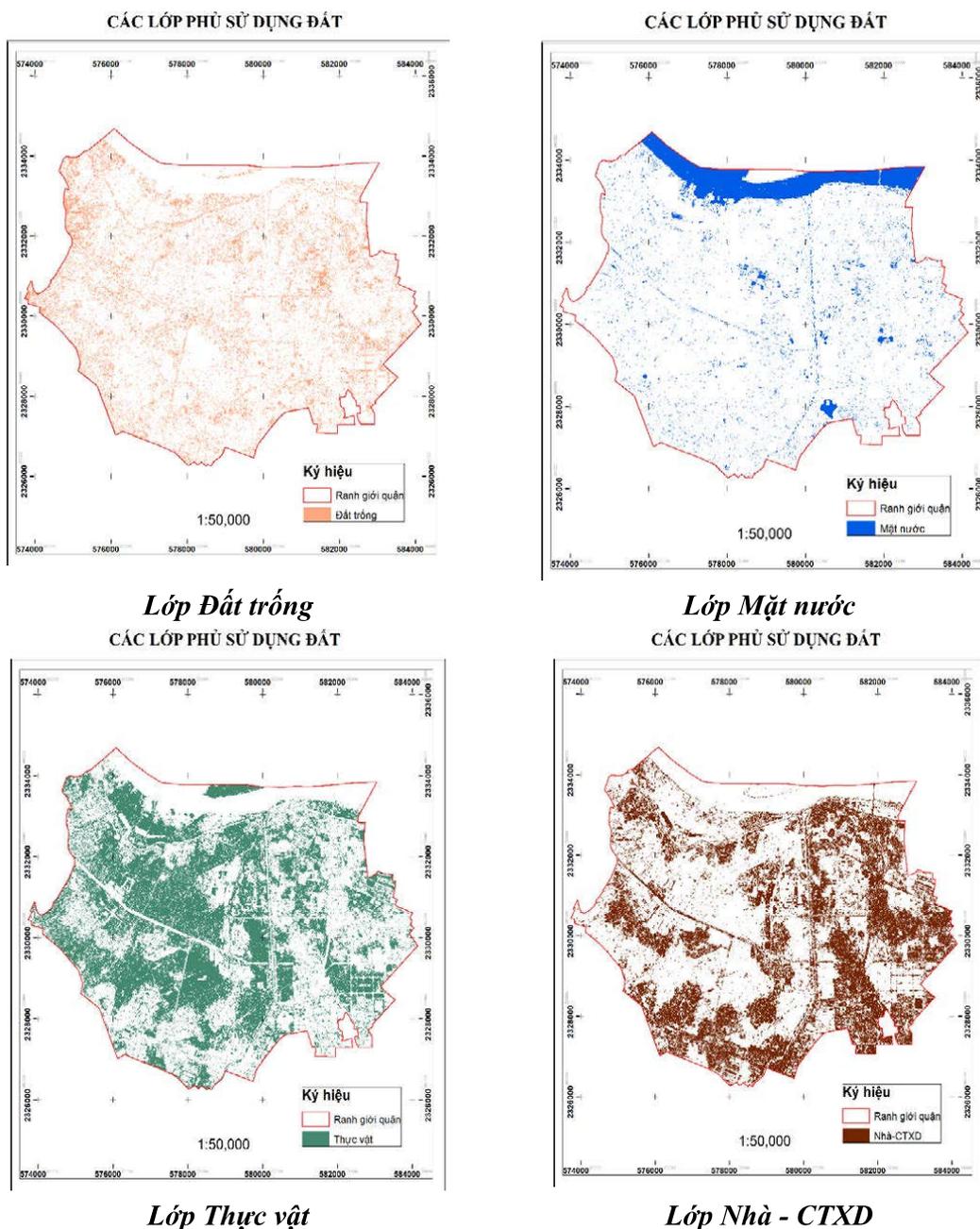
Hình 6: Biểu đồ cơ cấu sử dụng đất tại quận Bắc Từ Liêm, Thành phố Hà Nội

Kết quả phân loại các lớp phủ sử dụng đất cho thấy rằng lớp Đất trống hiện tại chiếm tỷ lệ không nhiều và phân bố không tập trung, chủ yếu là các khoảng đất trống nằm rải rác trên địa bàn 13 phường của quận Bắc Từ Liêm. Đây là một số khu đất đã được quy hoạch sử dụng vào nhiều mục đích khác nhau, trong đó có những khu đất hiện tại chưa thực hiện đúng tiến độ của dự án, ngoài ra còn có các mảnh đất, ô đất nhỏ lẻ nằm rải rác trong các khu dân cư chưa được người dân xây dựng công trình, nhà cửa. Ngoài ra các ô,

Nghiên cứu

mảnh đất trống trên còn là một số mảnh vườn để trống, chưa thực hiện canh tác, sản xuất,... Đặc biệt là lớp Mặt nước tại khu vực nghiên cứu có diện tích rất nhỏ, chỉ chiếm khoảng 10 % so với tổng diện tích tự nhiên, trong đó mặt nước của phần sông Hồng nằm trên địa bàn quận chiếm khoảng 80 - 90 % diện tích mặt nước của toàn bộ diện tích lớp Mặt nước được phân loại tại thời điểm tháng 12 năm 2022. Đối

với lớp Thực vật (42,5 %) và lớp Nhà - CTXD (33,5 %) là diện tích chiếm chủ yếu trong khu vực. Đây cũng là nguồn dữ liệu hiện trạng hữu ích có thể tham khảo giúp cho công tác quản lý nhà nước về đất đai, quy hoạch sử dụng đất được hiệu quả hơn. Hình 7 minh họa hiện trạng các lớp sử dụng đất trên địa bàn quận Bắc Từ Liêm ngày 19 tháng 12 năm 2022.



Hình 7: Các lớp phủ sử dụng đất tại quận Bắc Từ Liêm, Thành phố Hà Nội

Trong tổng số có 136 điểm mẫu đã xây dựng kiểm định và thành lập ma trận nhầm lẫn. Tỷ lệ điểm kiểm tra và tổng số điểm lấy mẫu lần lượt là 30 % và 70 %. Kết quả các độ chính xác sau phân loại được tính toán từ ma trận nhầm lẫn bao gồm các thông tin cơ bản là độ chính xác tổng thể (OA) là 0,88 và hệ số Kappa là 0,82. Có những nghiên cứu sử dụng số lượng mẫu phân loại cho mỗi lớp trong khoảng từ 10 đến 30 số lượng kênh ảnh sử dụng. Tuy nhiên cũng có nhiều kết quả nghiên cứu đã chứng minh rằng số lượng mẫu phân loại không cần thiết quá nhiều,

có thể số lượng mẫu chỉ cần bằng từ 2 đến 4 lần số lượng kênh ảnh sử dụng trong phân loại để có độ chính xác tổng thể và hệ số Kappa cao tới hơn 90 % [16, 17]. Với nghiên cứu này, khu vực có diện tích không quá lớn, lớp mặt nước, lớp thực vật có nhiều giá trị pixel đồng nhất và sử dụng 8 kênh ảnh Sentinel - 2 để phân loại, do đó số lượng mẫu sử dụng trung bình cho mỗi lớp là khoảng 34 mẫu. Số lượng mẫu này tương đương với hơn 4 lần số lượng kênh ảnh sử dụng và kết quả đạt độ chính xác cao. Kết quả phân tích ma trận nhầm lẫn thể hiện chi tiết tại Bảng 2.

Bảng 2. Ma trận nhầm lẫn (Confusion)

	Đất trống	Mặt nước	Thực vật	Nhà-CTXD	Tổng
Đất trống	12	4	8	3	27
Mặt nước	15	184	8	13	220
Thực vật	0	0	255	2	257
Nhà-CTXD	6	6	1	62	75
Tổng	33	194	272	80	579
OA: 88,60 %			Kappa: 0,82289		

Kết quả nghiên cứu cho thấy độ chính xác ảnh sau phân loại là rất cao nên kết quả này đảm bảo độ tin cậy để sử dụng cho các thống kê, phân tích về các lớp phủ sử dụng đất tại khu vực quận Bắc Từ Liêm, thành phố Hà Nội. Các mẫu huấn luyện và chất lượng ảnh có ảnh hưởng nhất định tới độ chính xác phân loại các lớp phủ sử dụng đất. Trong nghiên cứu này, chất lượng ảnh thu nhận có độ phủ mây rất thấp, hình ảnh và các thông tin vật lý rõ ràng. Do đó, với các điều kiện trên ta có ảnh kết quả sau phân loại là có độ tin cậy cao, đảm bảo để sử dụng cho các bước tính toán tiếp theo.

5. Kết luận và đề xuất

Khai thác, phân tích dữ liệu ảnh vệ tinh trực tuyến phục công tác quản lý đất đai đang ngày càng tỏ ra hiệu quả, nhanh

chóng và tiết kiệm chi phí. Ngoài ra áp dụng các thuật toán học máy trong phân loại lớp phủ sử dụng đất từ ảnh vệ tinh cũng góp phần tích cực vào công cuộc chuyển đổi số, cách mạng công nghiệp 4.0 nói chung hiện nay. Phương pháp nghiên cứu đã đưa ra các kết quả đảm bảo độ tin cậy và các dữ liệu phân tích thường là có thời gian gần đây nhất. Các kết quả này có độ chi tiết cao, có tính thời điểm và cho thấy khả năng đạt hiệu quả cao để phục vụ công tác giám sát quy hoạch sử dụng đất, theo dõi và dự báo biến động đất đai.

Các kết quả phân loại ảnh với các lớp phủ cơ bản là Đất trống, Mặt nước, Thực vật, Nhà - CTXD đã thể hiện được độ tin cậy, tuy nhiên để nên có các nghiên cứu sâu hơn và có sự so sánh với các loại bản đồ hiện trạng cùng thời điểm được thành

lập với các kết quả điều tra và bổ sung từ thực địa.

TÀI LIỆU THAM KHẢO

[1]. Fischer, Manfred M., and Arthur Getis, eds (2010). *Handbook of applied spatial analysis: Software tools, methods and applications*. Berlin: Springer.

[2]. MohanRajan, Sam Navin, Agilandeewari Loganathan and Prabukumar Manoharan (2020). *Survey on Land Use/Land Cover (LU/LC) change analysis in remote sensing and GIS environment: Techniques and Challenges*. Environmental Science and Pollution Research 27 (2020): 29900 - 29926.

[3]. Green, Edmund, et al., (2000). *Remote sensing: Handbook for tropical coastal management*. United Nations Educational, Scientific and Cultural Organization (UNESCO).

[4]. Phạm Minh Hải và cộng sự (2019). *Nghiên cứu cơ sở khoa học ứng dụng thuật toán Radom Forest trong phân loại ảnh vệ tinh Spot 6 với khu vực thực nghiệm tại tỉnh Cà Mau*. Kỷ yếu Hội nghị: Nghiên cứu cơ bản trong “Khoa học Trái đất và Môi trường”. 361.

[5]. Lewis, R. J. (2000). *An introduction to classification and regression tree (CART) analysis*. In Annual meeting of the society for academic emergency medicine in San Francisco, California (Vol. 14). Citeseer.

[6]. Breiman L. J. M. 1.1 (2001). *Random forests*. Kluwer Academic Publishers. Manufactured in The Netherlands. 45, 5.

[7]. Cortes C. and V. Vapnikl (1995). *Support-vector networks*. J Machine learning, Kluwer Academic Publishers, Boston. Manufactured in The Netherlands. 20, 273.

[8]. Jia, K., Wei, X., Gu, X., Yao, Y., Xie, X., & Li, B. (2014). *Land cover classification using Landsat 8 operational land imager data in Beijing, China*. Geocarto international, 29(8), 941 - 951.

[9]. Bùi Thị Hồng Thắm, Trịnh Thị Hoài Thu (2020). *Phân loại đối tượng chiết tách lớp phủ bề mặt tại khu vực công viên địa chất*

toàn cầu Non nước Cao Bằng dựa trên nền tảng điện toán đám mây. Tạp chí Khoa học Tài nguyên và Môi trường. 31, 65.

[10]. Đặng Thanh Tùng (2021). *Khai thác trực tuyến cơ sở dữ liệu ảnh vệ tinh, so sánh thuật toán học máy về phân loại lớp phủ trên nền Google Earth Engine*. Kỷ yếu Hội thảo Quốc gia: Giải pháp kết nối và chia sẻ hệ thống cơ sở dữ liệu phục vụ công tác đào tạo, quản lý lĩnh vực Tài nguyên và Môi trường. Nxb. Khoa học Tự nhiên và Công nghệ - Hà Nội, trang 192.

[11]. <https://baochinhphu.vn/tach-huyen-tu-liem-ha-noi-thanh-2-quan-102154360.htm>

[12]. Manevitz, L. M., & Yousef, M. (2001). *One - class SVMs for document classification*. Journal of machine Learning research, 2(Dec), 139 - 154.

[13]. Talukdar, S., Singha, P., Mahato, S., Pal, S., Liou, Y.A., & Rahman, A. (2020). *Land - use land - cover classification by machine learning classifiers for satellite observations - A review*. Remote Sensing, 12(7), 1135.

[14]. Vapnik V.N. and A. Y. Chervonenkis (1971). *On the uniform convergence of relative frequencies of events to their probabilities. Measures of complexity*. Springer, 11.

[15]. Visa S., B. Ramsay, A. L. Ralescu and E. J. M. Van Der Knaapl (2011). *Confusion matrix - based feature selection*. 710, 120.

[16]. Li, C., Wang, J., Wang, L., Hu, L., & 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.

[17]. Van Niel, T. G., McVicar, T. R., & Datt, B. (2005). *On the relationship between training sample size and data dimensionality: Monte Carlo analysis of broadband multi - temporal classification*. Remote sensing of environment, 98(4), 468 - 480.

BBT nhận bài: 06/02/2023; Phản biện xong: 14/02/2023; Chấp nhận đăng: 28/3/2023