

XÂY DỰNG BẢN ĐỒ KHÔNG GIAN XANH BẰNG THUẬT TOÁN RANDOM FOREST (RF) TRÊN NỀN TẢNG ĐIỆN TOÁN Đám MÂY CỦA GOOGLE EARTH ENGINE – THỬ NGHIỆM TẠI HUYỆN BA VÌ THÀNH PHỐ HÀ NỘI

Phạm Văn Duẩn, Hoàng Văn Khiên, Nguyễn Văn Tùng

Trường Đại học Lâm nghiệp

<https://doi.org/10.55250/jo.vnuf.2022.6.058-067>

TÓM TẮT

Bản đồ phân bố không gian xanh (KGX) rất quan trọng đối với công tác quản lý, nhưng việc xây dựng bản đồ này thường mất nhiều thời gian và công sức, làm cho công tác quản lý KGX gặp khó khăn. Căn cứ vào 440 điểm mẫu điều tra mặt đất, ảnh Sentinel-2, thuật toán Random forest (RF) trên nền tảng điện toán đám mây của Google Earth Engine (GEE), nghiên cứu tiến hành phân loại, xây dựng bản đồ KGX tại huyện Ba Vì, thành phố Hà Nội. Tổng số 13 biến đầu vào từ ảnh Sentinel-2 đã được tính toán thử nghiệm để phân loại xây dựng bản đồ KGX. Khi sử dụng từng kênh ảnh để phân loại thì kết quả chỉ đạt độ chính xác từ thấp đến vừa phải, nhưng khi sử dụng nhiều kênh ảnh để phân loại thì kết quả có độ chính xác cao. Từ đó, xây dựng được bản đồ KGX với độ phân giải không gian 10 m cho toàn huyện và xác định tại Ba Vì có 12.124 ha KGX (chiếm 28,7% diện tích tự nhiên của huyện) phân bố tại tất cả 31 xã, thị trấn của huyện. Kết quả cũng cho thấy đang xuất hiện sự thiếu hụt diện tích KGX tại các xã không có rừng trên địa bàn huyện. Độ chính xác của mô hình tốt nhất thu được là 95,7% và tư liệu ảnh Sentinel-2 mức 2A luôn có sẵn và được cung cấp miễn phí cho thấy tiềm năng của phương pháp lập bản đồ KGX trên nền tảng GEE được nghiên cứu đề xuất.

Từ khóa: Ba Vì, Google Earth Engine (GEE), không gian xanh đô thị, Random forest (RF), Sentinel-2.

1. ĐẶT VẤN ĐỀ

Không gian xanh (KGX) là các bề mặt và không gian nơi thực vật phát triển, chiếm giữ và bao phủ (Taylor & Hochuli, 2017). KGX đã đóng một vai trò quan trọng trong quy hoạch và hoạch định chính sách tại đô thị (Meng et al., 2018). KGX bao gồm công viên, vườn, cây xanh đường phố, rừng và cây xanh tại các địa điểm lịch sử... góp phần cải thiện chất lượng cuộc sống đô thị (Sathyakumar et al., 2020). Nhiều tài liệu đã chứng minh rằng KGX có ảnh hưởng đến cả sức khỏe về thể chất và tinh thần của con người (Shin et al., 2020; Wiese et al., 2021; Xie et al., 2020). Để quản lý KGX trong đô thị thì bản đồ phân bố KGX rất quan trọng, bản đồ này có thể cung cấp thông tin cơ bản về chất lượng môi trường tự nhiên xung quanh các khu dân cư đô thị và khả năng tiếp cận của người dân đối với KGX. Nhưng việc xây dựng được các bản đồ phân bố KGX thường mất nhiều thời gian và công sức, điều đó làm cho công tác quản lý KGX tại đô thị gặp nhiều khó khăn.

Hiện nay, các bộ dữ liệu vệ tinh miễn phí với độ phân giải không gian và thời gian cao, nền tảng điện toán đám mây và các thuật toán máy học cung cấp các khả năng mới để lập bản đồ hiện trạng thảm thực vật nói chung và hiện trạng KGX đô thị nói riêng. Google Earth Engine

(GEE) là một nền tảng điện toán đám mây lưu trữ kho ảnh vệ tinh không lồ (Landsat; MODIS; Sentinel...) cho phép tạo mẫu, phân tích và trực quan hóa dữ liệu không gian địa lý ở các quy mô khác nhau thông qua xử lý song song, giúp giảm thời gian tính toán. Nhiều nghiên cứu đã sử dụng GEE để xây dựng bản đồ hiện trạng sử dụng đất (Li et al., 2020), giám sát suy thoái rừng (Bullock et al., 2020), bảo tồn thảm thực vật (Pratico et al., 2021), lập bản đồ lũ (Coltin et al., 2016) và các loại ứng dụng liên quan đến môi trường và quản lý tài nguyên thiên nhiên khác. GEE đã được sử dụng để lập bản đồ không gian xanh trên quy mô toàn cầu và quốc gia. Huang và cộng sự (2018) đánh giá tác động của hình thái đô thị đến cấu trúc cảnh quan của không gian xanh đô thị ở Trung Quốc bằng cách sử dụng hình ảnh Landsat dựa trên Google Earth Engine. Huang và cộng sự (2021) lập bản đồ phạm vi phân bố của không gian xanh đô thị ở 1.039 thành phố trên toàn cầu bằng cách sử dụng ảnh vệ tinh. Duan và cộng sự (2019) tìm hiểu sự phân bố theo không gian của các khu rừng đô thị ở Trung Quốc bằng cách sử dụng hình ảnh Sentinel-2 với Google Earth Engine. Atasoy (2018) giám sát không gian xanh đô thị và sự phân mảnh cảnh quan bằng cách sử dụng viễn thám.

Theo Quyết định số 1259/QĐ-TTg ngày 26/7/2011 của Thủ tướng Chính phủ, đến năm 2030 và tầm nhìn đến năm 2050 xây dựng Hà Nội trở thành một thành phố phát triển bền vững về môi trường (Thành phố xanh). Nói đến không gian xanh tại Hà Nội thường liên tưởng đến cây xanh trong các công viên, vườn hoa, cây xanh trên đường phố tại các quận nội thành. Các nghiên cứu chưa quan tâm nhiều đến không gian tự nhiên như diện tích rừng và đặc biệt, thiếu vắng các nghiên cứu về vấn đề này tại các thị xã, thị trấn ngoại thành là nơi sẽ có nhịp điệu phát triển rất mạnh trong thời gian tới.

Mục tiêu của nghiên cứu này là sử dụng ảnh Sentinel-2 thông qua thuật toán Random Forest (RF) trên nền tảng điện toán đám mây của Google Earth Engine để xây dựng bản đồ phân bố không gian xanh tại huyện Ba Vì – thành phố Hà Nội.



(a) Các điểm sử dụng để xây dựng mô hình



(b) Các điểm sử dụng để kiểm chứng mô hình

Hình 1. Vị trí các điểm điều tra thực địa tại Ba Vì

(3) Lớp bản đồ ranh giới hành chính các xã và huyện Ba Vì được kế thừa từ Phòng Tài nguyên và Môi trường của huyện.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Tư liệu sử dụng

Tư liệu chủ yếu được sử dụng trong nghiên cứu gồm:

(1) Ảnh vệ tinh Sentinel-2 được nhà sản xuất hiệu chỉnh bức xạ, khí quyển và hình học, đưa về giá trị phản xạ mặt đất, hệ tọa độ UTM/WGS84. Các kênh trên ảnh có ba độ phân giải không gian: 10 m, 20 m và 60 m với tổng cộng 12 dải quang phổ. Chu kỳ chụp lặp lại một vị trí là 10 ngày.

(2) Hệ thống 440 điểm điều tra thực địa trong giai đoạn từ tháng 6/2021 – tháng 11/2021 tại huyện Ba Vì. Trong đó: đối tượng rừng và cây lâu năm điều tra 150 điểm mẫu; đối tượng mặt nước điều tra 89 điểm mẫu; khu dân cư, đất xây dựng điều tra 84 điểm mẫu; đất nông nghiệp, đất trống điều tra 117 điểm mẫu. Vị trí các điểm điều tra thực địa minh họa tại hình 1.

2.2. Phương pháp nghiên cứu

- Phương pháp thu thập và xử lý số liệu ngoại nghiệp:

Vị trí các điểm điều tra thực địa được đưa lên

bản đồ số ở hệ tọa độ UTM và chuyển vào GEE với các giá trị thuộc tính được gán mã như sau: đối tượng rừng và cây lâu năm gán mã 1; đối tượng mặt nước gán mã 2; khu dân cư, đất xây dựng gán mã 3; đất nông nghiệp, đất trống gán mã 4.

Các điểm điều tra thực địa sau khi đưa vào GEE được lập trình chia ngẫu nhiên thành 2 phần: phần thứ nhất gồm 2/3 số điểm điều tra được sử dụng để thử nghiệm xác định ngưỡng xây dựng bản đồ không gian xanh; phần thứ 2 gồm 1/3 số điểm còn lại được sử dụng để đánh giá độ chính xác của bản đồ không gian xanh tạo ra.

- Phương pháp xử lý và trích xuất thông tin từ ảnh vệ tinh:

Lựa chọn toàn bộ các cảnh ảnh chụp khu vực nghiên cứu từ 01/01/2021 - 31/12/2021 có tỷ lệ che phủ mây dưới 20%. Trên mỗi ảnh được lựa chọn, tiến hành lọc bỏ khu vực có mây từ đó tạo ra một tập hợp các ảnh không có mây (chỗ có mây đã bị đục thủng) tại khu vực nghiên cứu. Tập hợp hình ảnh trên được thu gọn thành một hình ảnh duy nhất, chứa thông tin của các hình ảnh đã chọn, để thực hiện việc trích xuất và phân loại. Giá trị pixel của hình ảnh mới được tính toán bằng trung bình giá trị pixel của tất cả các hình ảnh tại cùng một vị trí bằng hàm Median. Ảnh mới tạo ra này gọi là ảnh Trung bình được sử dụng để thực hiện các bước nghiên cứu tiếp theo.

Trên ảnh Trung bình, tất cả các band ảnh được đưa về cùng độ phân giải không gian là 10 m, hệ tọa độ UTM và tính chỉ số thực vật khác biệt chuẩn hóa NDVI theo công thức sau:

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (\text{Bannari et al., 1995})$$

Trong đó:

RED là giá trị phổ của kênh đỏ (Band 4);

NIR là giá trị phổ kênh cận hồng ngoại (Band 8).

Trong nhiều nghiên cứu về sự thay đổi và phân loại thảm thực vật, hầu hết đều tập trung vào mối quan hệ giữa thảm thực vật và chỉ số thực vật (VIs), trong đó NDVI là chỉ số thực vật được sử dụng phổ biến nhất. Trong nghiên cứu này, 14 giá trị trên ảnh Trung bình gồm: giá trị chỉ số thực vật NDVI và 12 giá trị của 12 band ảnh được sử dụng để thăm dò phương pháp xây dựng bản đồ không gian xanh tại khu vực. Toàn

bộ quá trình xử lý, tính toán trích xuất thông tin từ ảnh vệ tinh được lập trình thực hiện trên GEE.

- Lựa chọn thuật toán xây dựng mô hình:

RF là thuật toán phi tham số do Breiman đề xuất (Breiman, 2001) được sử dụng trong rất nhiều bài toán phân loại. Nhiều nghiên cứu trước đây đã chỉ ra rằng RF hiệu quả và chính xác hơn trong việc phân loại các loại đất đô thị so với các phương pháp tiếp cận máy học khác như máy vector hỗ trợ (SVM) và mạng nơ-ron nhân tạo (ANN). Thuật toán RF là bộ phân loại kết hợp một tập hợp các cây quyết định để đưa ra quyết định phân loại tốt hơn. Mỗi cây quyết định được tạo bằng cách chọn ngẫu nhiên một tập hợp con các mẫu huấn luyện từ tập dữ liệu gốc dựa trên nguyên tắc sử dụng phương pháp lấy mẫu lại Bootstrap. Về nguyên tắc, càng nhiều cây quyết định thì độ chính xác phân loại càng cao. Tuy nhiên, theo các nghiên cứu, kết quả phân loại tối ưu khi số cây quyết định khoảng 60 cây. Số cây quyết định lớn hơn cũng không cải thiện đáng kể độ chính xác của kết quả phân loại. Do đó, trong nghiên cứu này, số cây quyết định được đặt mặc định là 60 cây.

- Phương pháp lựa chọn các biến đầu vào để xây dựng mô hình theo các bước sau: (1) Sử dụng giá trị từng kênh ảnh (của 12 kênh ảnh) hoặc giá trị chỉ số thực vật NDVI của ảnh Trung bình làm biến đầu vào để phân loại xây dựng bản đồ không gian xanh tại khu vực nghiên cứu và đánh giá độ chính xác của mô hình; (2) Trên cơ sở độ chính xác của mô hình đơn biến, lựa chọn tổ hợp các biến đầu vào để phân loại xây dựng bản đồ không gian xanh tại khu vực nghiên cứu và đánh giá độ chính xác của mô hình; (3) Trên cơ sở bước (1) và bước (2) xác định các biến đầu vào tối ưu sử dụng để xây dựng bản đồ không gian xanh tại khu vực nghiên cứu.

- Phương pháp đánh giá độ chính xác mô hình:

Kết quả phân loại được đánh giá bằng phương pháp sử dụng ma trận sai số, là một định dạng tiêu chuẩn để đánh giá độ chính xác của phân loại. Ma trận sai số là ma trận thể hiện sự sai khác và trùng khớp giữa kết quả giải đoán so với thực tế, trong đó dữ liệu thực tế căn cứ vào 1/3 số điểm thực địa không tham gia xây dựng mô hình phân loại. Từ ma trận sai số, tính được độ chính xác tổng thể, độ chính xác phân loại từng đối tượng và chỉ số Kappa (K) là chỉ số

tổng hợp để đánh giá độ chính xác của kết quả phân loại. Chỉ số Kappa (K) được tính toán theo công thức:

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

Trong đó:

r là số lượng cột trong ma trận sai số;

X_{ii} là số lượng Pixel quan sát được tại hàng i, cột i trên đường chéo chính;

X_{i+} là tổng Pixel quan sát tại hàng i;

X_{+i} là tổng số Pixel quan sát tại cột i;

N là tổng số Pixel quan sát được trong ma trận sai số.

Chỉ số Kappa nằm giữa 0 và 1 và được đánh giá như sau: (1) $K > 0,8$: kết quả phân loại có độ chính xác cao; (2) $K < 0,4$: kết quả phân loại có độ chính xác thấp; (3) Giá trị K còn lại được đánh giá là kết quả phân loại có độ chính xác vừa phải.

Bảng 1. Kết quả đánh giá độ chính xác phân loại xây dựng bản đồ không gian xanh tại khu vực nghiên cứu bằng mô hình đơn biến

TT	Biến đầu vào	Độ chính xác (%)					Chỉ số Kappa (K)
		Tổng thể	Rừng, cây lâu năm	Mặt nước	Khu dân cư, đất xây dựng	Đất trồng, đất nông nghiệp	
1	NDVI	70,2	87,5	66,9	44,8	67,9	0,597
2	B1	49,6	66,7	22,2	50,0	36,4	0,314
3	B2	56,0	75,0	33,3	54,5	42,2	0,398
4	B3	51,1	71,7	23,8	40,7	39,4	0,328
5	B4	53,9	69,4	22,2	67,9	45,9	0,375
6	B5	58,9	70,8	37,5	63,2	59,5	0,442
7	B6	58,9	56,9	85,7	38,0	44,4	0,436
8	B7	64,5	62,2	85,7	69,2	42,8	0,521
9	B8	58,2	57,1	96,8	47,1	35,3	0,438
10	B8A	55,3	54,3	93,5	54,2	27,5	0,396
11	B9	57,4	57,4	75,8	56,5	42,1	0,424
12	B11	59,6	81,8	76,3	48,1	34,9	0,461
13	B12	75,2	79,6	90,3	76,9	54,3	0,663
	Nhỏ nhất	49,6	54,3	22,2	38,0	27,5	0,314
	Lớn nhất	75,2	87,5	96,8	76,9	67,9	0,663

Chỉ số Kappa của các mô hình đơn biến biến động từ 0,314 - 0,663, chứng tỏ kết quả phân loại có độ chính xác từ thấp đến vừa phải. Không có biến đầu vào nào cho kết quả phân loại có độ chính xác cao ($K > 0,8$).

Xét theo các biến đầu vào:

- Khi sử dụng giá trị các kênh ảnh: B1, B2, B3, B4 và B8A làm biến đầu vào để phân loại xây dựng bản đồ không gian xanh tại khu vực nghiên cứu thì kết quả phân loại có độ chính xác thấp ($K < 0,4$).

Căn cứ vào giá trị chỉ số Kappa lựa chọn mô hình tốt nhất là mô hình có chỉ số Kappa cao nhất và số biến đầu vào là ít nhất.

3. KẾT QUẢ VÀ THẢO LUẬN

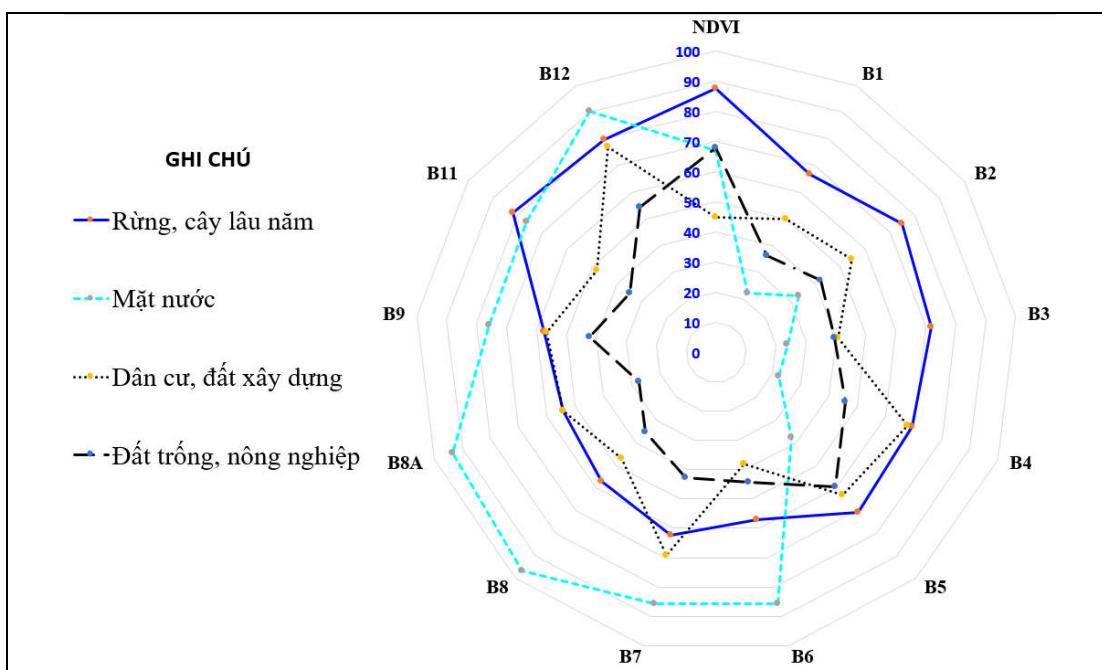
3.1. Kết quả nghiên cứu

3.1.1. Kết quả phân loại và đánh giá độ chính xác của các mô hình đơn biến

Trong nghiên cứu này, tổng số 12 biến đầu vào từ ảnh Sentinel-2 đã được tính toán để phân loại xây dựng bản đồ không gian xanh, bao gồm 12 biến là giá trị của các kênh ảnh Sentinel-2 xử lý ở mức 2A và 1 biến là giá trị chỉ số thực vật khác biệt chuẩn hóa (NDVI). Căn cứ vào các điểm điều tra mặt đất, giá trị các kênh ảnh tương ứng, sử dụng thuật toán RF trên nền tảng điện toán đám mây của GEE tiến hành phân loại xây dựng bản đồ không gian xanh tại khu vực nghiên cứu. Kết quả đánh giá độ chính xác của các mô hình đơn biến tại bảng 1.

- Khi sử dụng các kênh còn lại hoặc kênh chỉ số thực vật khác biệt chuẩn hóa NDVI làm biến đầu vào để phân loại xây dựng bản đồ không gian xanh tại khu vực nghiên cứu thì kết quả phân loại có độ chính xác vừa phải ($0,4 < K < 0,8$).

Biểu đồ độ chính xác phân loại nhóm đối tượng: rừng và cây lâu năm; mặt nước; khu dân cư, đất xây dựng; đất trồng, đất nông nghiệp theo các biến đầu vào tại hình 2 cho thấy:



Hình 2. Biểu đồ độ chính xác phân loại nhóm đối tượng theo biến đầu vào

- Với đối tượng rừng và cây lâu năm (Không gian xanh), kênh chỉ số thực vật NDVI cho kết quả phân loại có độ chính xác cao nhất, tiếp theo đến kênh 11.

- Với đối tượng là mặt nước, kênh 8 cho kết quả phân loại có độ chính xác cao nhất, tiếp theo đến kênh 8A.

- Với đối tượng khu dân cư, đất xây dựng, kênh 12 cho kết quả phân loại có độ chính xác cao nhất, tiếp theo đến kênh 7.

- Với đối tượng đất trống, đất nông nghiệp, kênh chỉ số thực vật NDVI cho kết quả phân loại có độ chính xác cao nhất, tiếp theo đến kênh 5.

3.1.2. Kết quả phân loại và đánh giá độ chính xác của các mô hình đa biến

Kết quả phân loại và đánh giá độ chính xác

của các mô hình đơn biến cho phép lựa chọn một số nhóm biến đầu vào để phân loại, xây dựng bản đồ không gian xanh tại khu vực như sau: (1) Trường hợp biến đầu vào là tập hợp các đơn biến cho kết quả phân loại đối tượng có độ chính xác cao nhất, đề xuất 2 nhóm biến đầu vào gồm: B8, B12, NDVI hoặc B5, B7, B8, B8A, B11, B12, NDVI; (2) Trường hợp biến đầu vào là tập hợp các đơn biến mà kết quả phân loại có độ chính xác vừa phải ($0,4 < K < 0,8$) gồm: B5, B6, B7, B8, B9, B11, B12, NDVI; (3) Trường hợp biến đầu vào là tập hợp tất cả các đơn biến: B1, B2, B3, B4, B5, B6, B7, B8, B8A, B9, B11, B12, NDVI.

Kết quả phân loại và đánh giá độ chính xác của các mô hình đa biến tại bảng 2.

Bảng 2. Kết quả đánh giá độ chính xác phân loại xây dựng bản đồ không gian xanh tại khu vực nghiên cứu bằng mô hình đa biến

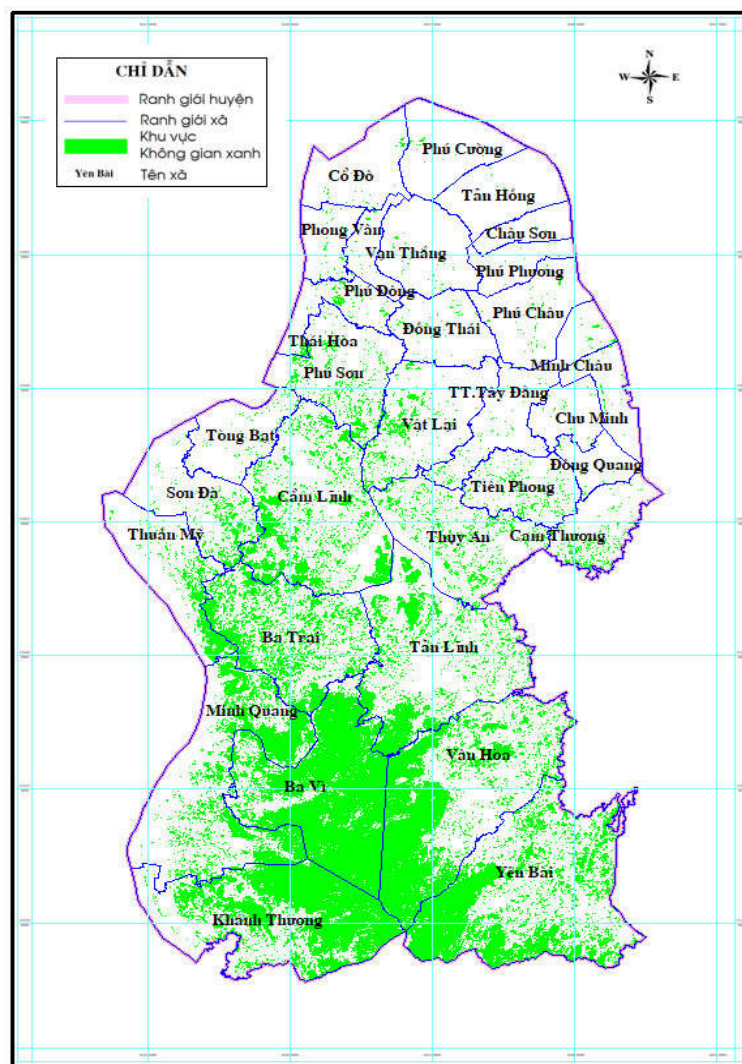
TT	Biến đầu vào	Độ chính xác (%)					Chỉ số Kappa (K)
		Tổng thể	Rừng, cây lâu năm	Mặt nước	Khu dân cư, đất xây dựng	Đất trống, đất nông nghiệp	
1	B8, B12, NDVI	88,6	95,5	96,8	81,8	78,1	0,847
2	B5, B7, B8, B8A, B11, B12, NDVI	92,2	97,8	96,8	87,5	84,4	0,895
3	B5, B6, B7, B8, B9, B11, B12, NDVI	92,9	97,8	96,8	90,3	84,8	0,904
4	B1, B2, B3, B4, B5, B6, B7, B8, B8A, B9, B11, B12, NDVI	95,7	97,9	96,8	99,0	88,6	0,942
	Nhỏ nhất	88,6	95,5	96,8	81,8	78,1	0,847
	Lớn nhất	95,7	97,9	96,8	99,0	88,6	0,942

Chỉ số Kappa của các mô hình đa biến từ 0,847 - 0,942, chứng tỏ kết quả phân loại theo các mô hình đa biến đều có độ chính xác cao ($K > 0,8$). Trong đó, mô hình với biến đầu vào là giá trị các kênh B8, B12, NDVI có chỉ số Kappa thấp nhất ($K = 0,847$) và mô hình với biến đầu vào là tất cả 13 kênh B1, B2, B3, B4, B5, B6, B7, B8, B8A, B9, B11, B12, NDVI có chỉ số Kappa cao nhất ($K = 0,942$).

Xem xét độ chính xác phân loại nhóm đối tượng: rừng và cây lâu năm; mặt nước; khu dân cư, đất xây dựng; đất trồng, đất nông nghiệp của các mô hình đa biến cho thấy: Độ chính xác phân loại các đối tượng thấp nhất khi biến đầu vào là giá trị các kênh B8, B12, NDVI. Độ chính xác phân loại các đối tượng cao nhất khi biến đầu vào là giá trị các kênh B1, B2, B3, B4, B5, B6, B7, B8, B8A, B9, B11, B12, NDVI.

Các mô hình đa biến đều cho thấy độ chính xác phân loại với nhóm đối tượng đất trồng, đất nông nghiệp là thấp nhất, trong khi đó độ chính xác phân loại với 3 nhóm đối tượng còn lại tương tự nhau và cao hơn. Điều này có thể lý giải là do nhóm đất nông nghiệp, đất trồng mang tính mùa vụ, thay đổi liên tục nên trên ảnh trung bình giá trị kênh ảnh biến động lớn, dẫn đến kết quả phân loại cho sai số lớn.

Như vậy, mô hình tối ưu được lựa chọn để phân loại và xây dựng bản đồ không gian xanh tại khu vực là mô hình với biến đầu vào là giá trị của 13 kênh gồm: B1, B2, B3, B4, B5, B6, B7, B8, B8A, B9, B11, B12, NDVI. Kết quả sử dụng mô hình để phân loại và xây dựng bản đồ không gian xanh tại huyện Ba Vì thể hiện tại hình 3.



Hình 3. Ảnh bản đồ không gian xanh tại Ba Vì chạy theo mô hình tối ưu

Bảng 3. Kết quả phân loại xác định diện tích không gian xanh tại huyện Ba Vì

TT	Xã	Diện tích (ha)			Dân số (người)	Diện tích KGX cần thiết (ha)
		Tự nhiên	Không gian xanh			
			Tổng	Có rừng		
1	Thị trấn Tây Đằng	1.228	42	0	16.208	41
2	Xã Ba Trại	2.032	869	133	14.681	37
3	Xã Ba Vì	2.551	2.237	2.177	2.230	6
4	Xã Cẩm Lĩnh	2.659	755	326	13.051	33
5	Xã Cam Thượng	814	93	0	7.326	18
6	Xã Châu Sơn	355	3	0	4.830	12
7	Xã Chu Minh	508	8	0	6.184	15
8	Xã Cổ Đô	880	11	0	8.460	21
9	Xã Đông Quang	393	9	0	5.239	13
10	Xã Đồng Thái	827	23	0	14.165	35
11	Xã Khánh Thượng	2.769	1.660	1.306	8.729	22
12	Xã Minh Châu	527	14	0	6.517	16
13	Xã Minh Quang	2.841	1.025	704	14.097	35
14	Xã Phong Vân	485	13	0	7.646	19
15	Xã Phú Châu	989	21	0	11.626	29
16	Xã Phú Cường	940	8	0	6.432	16
17	Xã Phú Đông	363	8	0	6.232	16
18	Xã Phú Phương	432	2	0	6.521	16
19	Xã Phú Sơn	1.345	190	73	9.732	24
20	Xã Sơn Đà	1.220	164	56	9.313	23
21	Xã Tân Hồng	880	3	0	13.773	34
22	Xã Tân Lĩnh	2.786	796	321	16.279	41
23	Xã Thái Hòa	571	47	0	8.169	20
24	Xã Thuần Mỹ	1.229	217	123	7.150	18
25	Xã Thụy An	1.636	262	59	9.736	24
26	Xã Tiên Phong	875	81	0	8.538	21
27	Xã Tòng Bạt	824	31	0	10.092	25
28	Xã Vân Hòa	3.255	1.696	1.255	12.084	30
29	Xã Vạn Thắng	991	12	0	16.000	40
30	Xã Vật Lại	1.445	163	56	14.553	36
31	Xã Yên Bài	3.537	1.662	1.111	7.436	19
Tổng		42.187	12.124	7.699	303.029	758

Diện tích không gian xanh theo đơn vị hành chính tính toán từ bản đồ (hình 3) thống kê tại bảng 3.

Tại bảng 3, tổng diện tích không gian xanh của toàn huyện là 12.124 ha (chiếm 28,7% diện tích tự nhiên) chi tiết đến từng xã được tính toán và xây dựng từ mô hình. Diện tích tự nhiên, diện tích rừng, dân số của các xã được kế thừa từ Phòng Tài nguyên và Môi trường; Phòng Kinh tế của huyện. Diện tích không gian xanh cần thiết của xã được xác định theo số dân với mức 25 m² cây xanh/người.

Kết quả cho thấy, huyện Ba Vì có 31 đơn vị hành chính xã, thị trấn (13 xã có rừng và 18 xã,

thị trấn không có rừng) và diện tích rừng chiếm trên 63,5% diện tích không gian xanh toàn huyện. Trong đó:

- Tại các xã có rừng, diện tích không gian xanh tính toán được từ mô hình đều lớn hơn so với diện tích rừng, thể hiện sự phù hợp và tính chính xác của kết quả phân loại không gian xanh từ mô hình. Tại tất cả các xã có rừng, diện tích không gian xanh hiện tại đều lớn hơn diện tích không gian xanh cần thiết.

- Tại 18 xã/thị trấn không có rừng thì 5/18 xã có diện tích không gian xanh hiện tại lớn hơn diện tích không gian xanh cần thiết; 13/18 xã có diện tích không gian xanh hiện tại nhỏ hơn diện

tích không gian xanh cần thiết. Điều này chứng tỏ tại các xã không có rừng trên địa bàn huyện đang xuất hiện sự thiếu hụt diện tích không gian xanh.

3.2. Thảo luận

Mục tiêu chính của nghiên cứu là thử nghiệm khả năng sử dụng ảnh vệ tinh Sentinel-2, thuật toán Random Forest (RF), trên nền tảng điện toán đám mây của Google Earth Engine để phân loại và xây dựng bản đồ không gian xanh tại khu vực cụ thể là huyện Ba Vì – thành phố Hà Nội. Kết quả đạt được cho phép đưa ra một số thảo luận sau:

Trong nghiên cứu, Google Earth Engine được sử dụng để chuẩn bị, sàng lọc và xử lý dữ liệu (ảnh vệ tinh, số liệu thực địa) theo yêu cầu của nghiên cứu, đồng thời tính toán các chỉ số liên quan và tiến hành phân tích kết quả thu được. Ngoài ra, nền tảng này cũng được sử dụng để xử lý dữ liệu đã phân loại và đánh giá độ chính xác của nó. Phương pháp này không chỉ tiết kiệm rất nhiều thời gian tải xuống và không gian lưu trữ so với phương pháp truyền thống mà còn hiệu quả hơn nhiều trong việc xử lý và phân tích dữ liệu.

Từ kết quả nghiên cứu, bản đồ không gian xanh với độ phân giải không gian 10 m đã được tạo ra cho huyện Ba Vì, thành phố Hà Nội. Nền tảng GEE được sử dụng mở ra cơ hội cập nhật nhanh chóng và tự động bản đồ không gian xanh tại khu vực nghiên cứu cho thời gian tiếp theo. Giá trị của 12 kênh ảnh và chỉ số thực vật NDVI từ hình ảnh Sentinel-2 được đưa vào bộ phân loại RF để phân biệt các loại đối tượng trên mặt đất. Độ chính xác trung bình của mô hình tốt nhất thu được là 95,7% cho thấy tiềm năng cao của phương pháp lập bản đồ không gian xanh tự động trên nền tảng GEE được đề xuất.

Độ chính xác của phân loại và xây dựng bản đồ không gian xanh đô thị chủ yếu bị ảnh hưởng bởi số lượng và chất lượng của ảnh vệ tinh, cũng như số lượng và tính đại diện của dữ liệu điều tra mặt đất, các đặc điểm của đối tượng cần phân loại. Kết hợp nhiều loại ảnh vệ tinh khác nhau, một mặt nâng cao được số lượng ảnh trong một khoảng thời gian nhất định, mặt khác chúng sẽ bổ sung dữ liệu cho nhau từ đó có thể cải thiện

độ chính xác phân loại (Zhong et al., 2019).

Vấn đề xây dựng và cập nhật bản đồ không gian xanh đô thị hàng năm đã và đang được đặt ra và ngày càng yêu cầu độ chi tiết cao hơn, để thực hiện được điều này việc sử dụng ảnh vệ tinh với độ phân giải không gian và thời gian cao là giải pháp khả thi. Tuy nhiên, đi kèm với yêu cầu đó là việc cần phải xử lý một số lượng lớn ảnh vệ tinh nên việc sử dụng nền tảng điện toán đám mây (như Google Earth Engine) là phù hợp. Với GEE có thể nhanh chóng truy xuất vào các ảnh vệ tinh mới nhất và xử lý hàng loạt dữ liệu ảnh mà các phần mềm xử lý ảnh truyền thống khó thực hiện được trong cùng khoảng thời gian.

Do tư liệu ảnh Sentinel-2 mức 2A luôn có sẵn và được cung cấp miễn phí nên các kỹ thuật và mô hình trong nghiên cứu này có thể được nhân rộng cho các khu vực khác với các đặc tính tương tự.

Một mô hình phân loại và xây dựng bản đồ không gian xanh có thể phải sử dụng rất nhiều biến đầu vào (tại nghiên cứu này có mô hình sử dụng đến 13 biến đầu vào) nên nhiều khi các hàm tương quan hoặc thuật toán phi tham số khác (như K-nn) khó áp dụng. Nghiên cứu cung cấp thêm phương pháp sử dụng thuật toán RF đã được nhiều nhà khoa học trên thế giới sử dụng để xây dựng mô hình phân loại hiện trạng sử dụng đất để phân loại và xây dựng bản đồ không gian xanh tại huyện Ba Vì, thành phố Hà Nội với lợi thế của thuật toán là có thể sử dụng nhiều biến đầu vào nhằm nâng cao độ chính xác để xây dựng các mô hình phân loại hiện trạng sử dụng đất.

4. KẾT LUẬN

Kết quả đánh giá độ chính xác phân loại xây dựng bản đồ phân bố không gian xanh của các mô hình đơn biến cho kết quả phân loại có độ chính xác từ thấp đến vừa phải, chỉ số Kappa của các mô hình từ 0,314 - 0,663. Trong đó:

- Sử dụng các kênh: B1, B2, B3, B4 và B8A để phân loại cho độ chính xác thấp; sử dụng các kênh B5, B6, B7, B8, B9, B11, B12 và NDVI để phân loại cho độ chính xác vừa phải.

- Với đối tượng rừng và cây lâu năm; đất trống, đất nông nghiệp, kênh NDVI cho kết quả

phân loại có độ chính xác cao nhất. Đối tượng là mặt nước, kênh 8 cho kết quả phân loại có độ chính xác cao nhất. Với đối tượng khu dân cư, đất xây dựng thì kênh 12 cho kết quả phân loại có độ chính xác cao nhất.

Kết quả đánh giá độ chính xác phân loại xây dựng bản đồ phân bố không gian xanh của các mô hình đa biến đều cho kết quả phân loại có độ chính xác cao ($K > 0,8$), chỉ số Kappa của các mô hình từ 0,847 - 0,942.

Các mô hình đa biến đều cho độ chính xác phân loại với nhóm đối tượng đất trồng, đất nông nghiệp thấp nhất, độ chính xác phân loại với 3 nhóm đối tượng còn lại tương tự nhau và cao hơn.

Mô hình tối ưu được lựa chọn để phân loại và xây dựng bản đồ không gian xanh tại khu vực là mô hình với biến đầu vào gồm 13 kênh: B1, B2, B3, B4, B5, B6, B7, B8, B8A, B9, B11, B12, NDVI.

Kết quả xác định tại Ba Vì có 12.124 ha không gian xanh (chiếm 28,7% diện tích tự nhiên) và phân bố tại tất cả 31 xã, thị trấn của huyện. Trong đó, tại các xã có rừng, diện tích không gian xanh tính toán được từ mô hình đều lớn hơn so với diện tích rừng và lớn hơn diện tích không gian xanh cần thiết. Tại các xã không có rừng thì trên 70% số xã có diện tích không gian xanh hiện tại nhỏ hơn diện tích không gian xanh cần thiết, chứng tỏ tại các xã không có rừng trên địa bàn huyện đang xuất hiện sự thiếu hụt diện tích không gian xanh.

TÀI LIỆU THAM KHẢO

1. Atasoy. M (2018). *Monitoring the urban green spaces and landscape fragmentation using remote sensing: A case study in Osmaniye, Turkey*. Environmental Monitoring and Assessment 190(12): 713.
2. Bannari. A, Morin. D, Bonn. F, Huete. A (1995). *A review of vegetation indices*. Remote Sensing Reviews, Vol 13, pp 95-120.
3. Breiman. L (2001). *Random forests*. Machine Learning vol. 45(1): pp 5-32.
4. Bullock. E.L, Woodcock. C.E, Olofsson. P (2020). *Monitoring tropical forest degradation using spectral unmixing and Landsat time series analysis*. Remote Sensing of Environment. Vol 238, 110968.
5. Coltin. B, McMichael. S, Smith. T, Fong. T (2016). *Automatic boosted flood mapping from satellite data*. International Journal of Remote Sensing. Vol 37, pp:

993-1015.

6. Duan. Q, Tan. M, Guo. Y, Wang. X, Xin. L (2019). *Understanding the spatial distribution of urban forests in China using Sentinel-2 images with Google Earth Engine*. Forests 10, 729.
7. Huang. C, Yang. J, Jiang. P (2018). *Assessing impacts of urban form on landscape structure of urban green spaces in China using Landsat images based on Google Earth Engine*. Remote Sensing, 10(10): 1569.
8. Huang. C, Yang. J, Clinton. N, Yu. L, Huang. H, Dronova. I, Jin. J (2021). *Mapping the maximum extents of urban green spaces in 1039 cities using dense satellite images*. Environmental Research Letters, Vol 16 (6).
9. Li. Q, Qiu. C, Ma. L, Schmitt. M, Zhu. X.X (2020). *Mapping the land cover of Africa at 10 m resolution from multi-source remote sensing data with Google Earth Engine*. Remote Sensing, Vol 12(4), 602.
10. Meng. Q, Zhang. L, Sun. Z, Meng. F, Wang. L, Sun. Y (2018). *Characterizing spatial and temporal trends of surface urban heat island effect in an urban main built-up area: A 12-year case study in Beijing, China*. Remote Sensing of Environment. Vol 204, pp: 826-837.
11. Praticò. S, Solano. F, Di Fazio. S, Modica. G (2021). *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(4), 586.
12. Sathyakumar. V, Ramsankaran. R, Bardhan. R (2020). *Geospatial approach for assessing spatiotemporal dynamics of urban green space distribution among neighbourhoods: A demonstration in Mumbai*. Urban Forestry and Urban Greening. Vol 48, 126585.
13. Shin. J.C, Kwan. M.P, Grigsby-Toussaint. D.S (2020). *Do Spatial Boundaries Matter for Exploring the Impact of Community Green Spaces on Health?* Environmental Research and Public Health. Vol 17(20): 7529.
14. Taylor. L, Hochuli. D.F (2017). *Defining greenspace: Multiple uses across multiple disciplines*. Landscape and Urban Planning. Vol 158, pp: 25-38.
15. Wiese. D; Stroup. A.M, Maiti. A, Harris. G, Lynch. S.M, Vucetic. S, Gutierrez-Velez. V.H, Henry. K.A (2021). *Measuring Neighborhood Landscapes: Associations between a Neighborhood's Landscape Characteristics and Colon Cancer Survival*. Environmental Research and Public Health. Vol 18, 4728.
16. Xie. Y, Xiang. H, Di. N, Mao. Z, Hou. J, Liu. X, Huo. W, Yang. B, Dong. G, Wang. C (2020). *Association between residential greenness and sleep quality in Chinese rural population*. Environment International. Vol 145, 106100.
17. Zhong. L, Hu. L, Zhou. H (2019). *Deep learning based multi-temporal crop classification*. Remote Sensing of Environment. Vol 221, pp: 430-443.

MAPPING OF GREEN SPACE DISTRIBUTION WITH RANDOM FOREST ALGORITHM (RF) ON THE GOOGLE EARTH ENGINE CLOUD PLATFORM – CASE STUDY IN BA VI DISTRICT, HANOI CITY

Pham Van Duan, Hoang Van Khien, Nguyen Van Tung
Vietnam National University of Forestry

SUMMARY

Distribution map of green space is very important for management, but this mapping often takes a lot of time and effort, making it difficult to manage green spaces. Based on 440 ground survey sample points, Sentinel-2 images, and random forest (RF) algorithm on Google Earth Engine (GEE) cloud computing platform, the research conducted classification and mapping of Green space distribution in Ba Vi district, Hanoi City. A total of 13 input variables from Sentinel-2 images were calculated and tested to classify and build green space maps. When using each image channel for classification, the results only have low to moderate accuracy, but when using multiple image channels for classification, the results have high accuracy. As a result, a map of green space (with a spatial resolution of 10 m) was built for the whole district and determined that in Ba Vi district, there are 12,124 hectares of green space (accounting for 28.7% of the district's natural area) distributed in 31 communes and towns. The results also show that there is a lack of green space in communes without forests in the district. The best model accuracy obtained was 95.7% and freely available Sentinel-2 level 2A image material demonstrates the potential of the green space mapping method on the GEE platform proposed by the study.

Keywords: Ba Vi, Google Earth Engine (GEE), random forest (RF); Sentinel-2, urban green spaces.

Ngày nhận bài : 15/8/2022

Ngày phản biện : 17/9/2022

Ngày quyết định đăng : 27/9/2022