



Original Article

Application of Satellite Images and Artificial Intelligence to Monitor Land Cover Changes in Hanoi Area During 2013-2023 Period

Dang Thanh Tung^{1,*}, Dinh Thi Thanh Huyen¹, Hoang Thi Thuy², Ta Minh Ngoc

¹*Hanoi University of Natural Resources and Environment, 41A Phu Dien, Bac Tu Liem, Hanoi, Vietnam*

²*Hanoi University of Mining and Geology, 18 Pho Vien, Bac Tu Liem, Hanoi, Vietnam*

Received 21 June 2023

Revised 13 October 2023; Accepted 12 December 2023

Abstract: Artificial intelligence (AI) and remote sensing technology have now increasingly improved their efficiency and reliability in monitoring the changes in land cover. With the amendment of the Vietnamese Law on Land in 2013 and the administrative boundary expansion of Hanoi, Hanoi experiences significant changes in land use and land cover for the last ten years. To monitor the actual land use changes in the area, this study used the Random Forest (RF) machine learning algorithm to classify the basic land covers, monitor, and analyze the spatial variation of land use and land cover in the 2013 to 2023 period. The study findings indicate a relatively high rate of expansion of construction zone area and a decrease in land cover related to water bodies and vegetated area. Water bodies decrease by an average of 0.8% annually, whereas the construction zone area increased by 7% of the total area.

Keywords: Monitoring the change, Land cover, Google Earth Engine, Random Forest algorithms.

* Corresponding author.

E-mail address: dttung.qldd@hunre.edu.vn

<https://doi.org/10.25073/2588-1094/vnuees.4962>

Nghiên cứu sử dụng ảnh vệ tinh, trí tuệ nhân tạo theo dõi biến động các lớp phủ bề mặt khu vực Hà Nội giai đoạn 2013-2023

Đặng Thanh Tùng^{1,*}, Đinh Thị Thanh Huyền¹, Hoàng Thị Thủy², Tạ Minh Ngọc¹

¹Trường Đại học Tài nguyên và Môi trường Hà Nội,
41A Đường Phú Diễn, Bắc Từ Liêm, Hà Nội, Việt Nam

²Trường Đại học Mỏ - Địa chất, 18 Phố Viên, Bắc Từ Liêm, Hà Nội, Việt Nam

Nhận ngày 21 tháng 6 năm 2023

Chỉnh sửa ngày 13 tháng 10 năm 2023; Chấp nhận đăng ngày 12 tháng 12 năm 2023

Tóm tắt: Trí tuệ nhân tạo (AI - Artificial Intelligence) và công nghệ viễn thám hiện nay đã ngày càng nâng cao hiệu quả và độ tin cậy trong theo dõi, giám sát biến động các lớp phủ bề mặt. Từ khi Luật đất đai bổ sung và sửa đổi năm 2013 và việc mở rộng địa giới hành chính đến nay, Hà Nội đã có nhiều thay đổi về sử dụng đất và các lớp phủ bề mặt. Để theo dõi về biến động các lớp phủ bề mặt thực tế trên địa bàn, nhóm nghiên cứu đã sử dụng thuật toán học máy Random Forest (RF) tiến hành phân loại các lớp phủ bề mặt, theo dõi và phân tích các biến động từ năm 2013 đến 2023. Kết quả của nghiên cứu đã cho thấy tốc độ mở rộng diện tích khu vực đất xây dựng tương đối cao và cùng với đó là sự suy giảm diện tích của các lớp phủ như là mặt nước, thảm cây xanh dày đặc.... Cụ thể lớp phủ mặt nước giảm trung bình khoảng 0,8% hàng năm. Trong khi đó lớp phủ đất xây dựng tăng khoảng 7% so với tổng diện tích tự nhiên.

Từ khóa: Theo dõi biến động, Lớp phủ sử dụng đất, Google Earth Engine, thuật toán Random Forest.

1. Mở đầu

Tình trạng biến động và quản lý sử dụng đất ở Hà Nội luôn là một vấn đề phức tạp và có nhiều yếu tố khác nhau tác động lên nó. Cùng với đó, các loại hình sử dụng đất có những biến động khác nhau tạo nên thực trạng biến động diện tích của các lớp phủ sử dụng đất trên địa bàn Hà Nội. Từ năm 2013 đến nay đã cho thấy diện tích mặt nước đang tiếp tục suy giảm nhanh chóng, ngược lại diện tích đất xây dựng có xu hướng tăng hàng năm.

Trong các phương pháp theo dõi biến động lớp phủ bề mặt, hiện nay phương pháp sử dụng dữ liệu ảnh vệ tinh đang được đánh giá là có hiệu

quả cao, có tính liên tục theo thời gian và có độ tin cậy đảm bảo. Đã có những nghiên cứu sử dụng một số thuật toán truyền thống như Maximum Likelihood Classifier (MLC), Minimum Distance Classifier (MDC), K-Nearest Neighbor (KNN), dựa trên các phần mềm thương mại để phân loại lớp phủ bề mặt từ ảnh vệ tinh [1-3]. Các nghiên cứu sử dụng hệ thống phần mềm này đã chứng minh được hiệu quả của mình có thể kể đến như nghiên cứu của tác giả Nguyễn T. N (2011, 2019) [4, 5]. Tuy nhiên, tại thời điểm này các nghiên cứu chưa có khả năng xử lý dữ liệu trực tuyến. Hiện nay, việc sử dụng nền tảng điện toán đám mây của Google

* Tác giả liên hệ.

Địa chỉ email: dttung.qldd@hunre.edu.vn

<https://doi.org/10.25073/2588-1094/vnuees.4962>

Earth Engine (GEE) đem lại khả năng xử lý dữ, phân tích liệu trực tuyến đáp ứng kịp thời các yêu cầu cung cấp thông tin, dữ liệu phục vụ theo dõi, giám sát biến động sử dụng đất tại từng thời điểm lựa chọn và qua các thời kỳ khác nhau [6, 7]. Trong nghiên cứu này, nhóm nghiên cứu sử dụng mô hình trí tuệ nhân tạo (AI - Artificial Intelligence) với thuật toán học máy (Machine Learning) Random Forest (RF) thực hiện phân loại, giám sát biến động các lớp phủ bề mặt tại khu vực Hà Nội giai đoạn 2013-2023. Các dữ liệu ảnh vệ tinh và các kết quả phân tích được thực hiện trên nền tảng điện toán đám mây GEE và sử dụng ngôn ngữ lập trình Javascript.

2. Dữ liệu, khu vực nghiên cứu

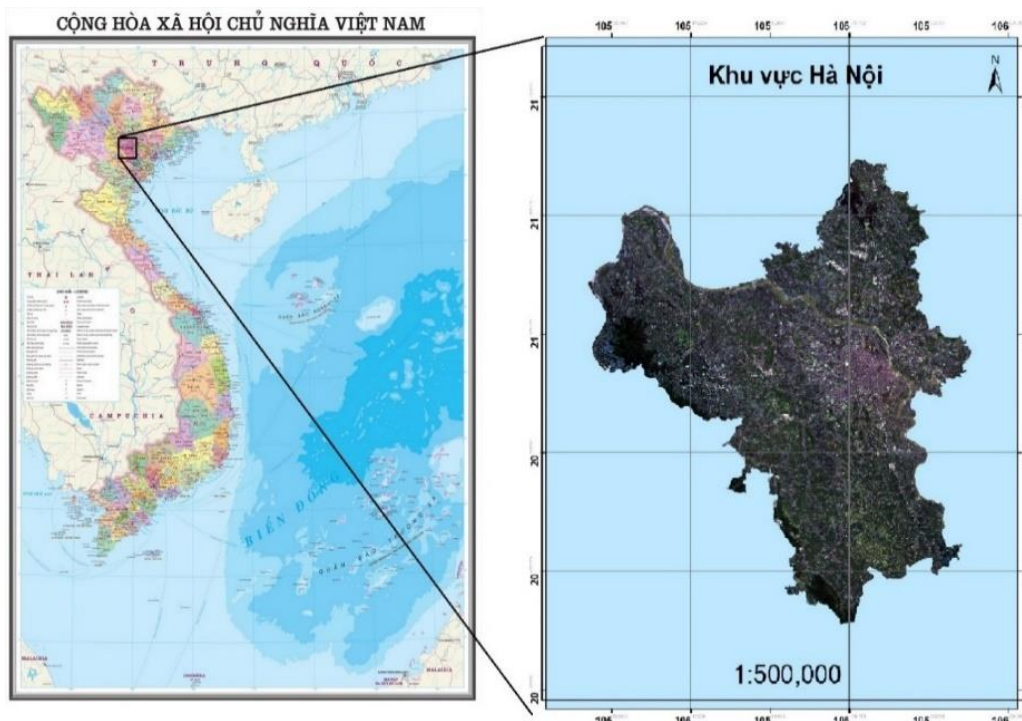
2.1. Khu vực nghiên cứu

Hà Nội có diện tích tự nhiên khoảng 334470,02 ha, dân số là hơn 7 triệu người; gồm 30 đơn vị hành chính cấp quận, huyện, thị xã, 577 xã, phường, thị trấn (Hình 1) [8]. Trong hơn

mười năm qua, Hà Nội mở rộng và phát triển, điều này cũng đã gây ra những biến động về các loại hình sử dụng đất, các loại lớp phủ như đất xây dựng, giao thông có sự gia tăng về diện tích, trong khi đó các lớp phủ như cây xanh, mặt nước,... có thể bị suy giảm đáng kể.

2.2. Dữ liệu nghiên cứu

Dữ liệu sử dụng trong nghiên cứu là tư liệu ảnh vệ tinh Landsat 8 và Sentinel 2 từ năm 2013 đến 2023 được khai thác trực tuyến trên nền tảng GEE. Các dữ liệu ảnh vệ tinh được lựa chọn là dữ liệu tổng hợp của các ảnh trong khoảng thời gian từ tháng 3 đến tháng 6 hàng năm, có chất lượng hình ảnh rõ ràng, độ phủ mây thấp đảm bảo chất lượng để phân tích, tính toán kết quả. Trong đó, từ năm 2013 đến 2018 sử dụng dữ liệu ảnh Landsat, từ năm 2019 đến 2023 dữ liệu sử dụng là ảnh Sentinel 2. Đây là các ảnh vệ tinh đã được hiệu chỉnh TOA – Top Of Atmosphere. Các thông tin cơ bản về dữ liệu ảnh vệ tinh được trình bày tại Bảng 1 dưới đây.



Hình 1. Khu vực nghiên cứu.

Bảng 1. Dữ liệu ảnh vệ tinh khu vực nghiên cứu

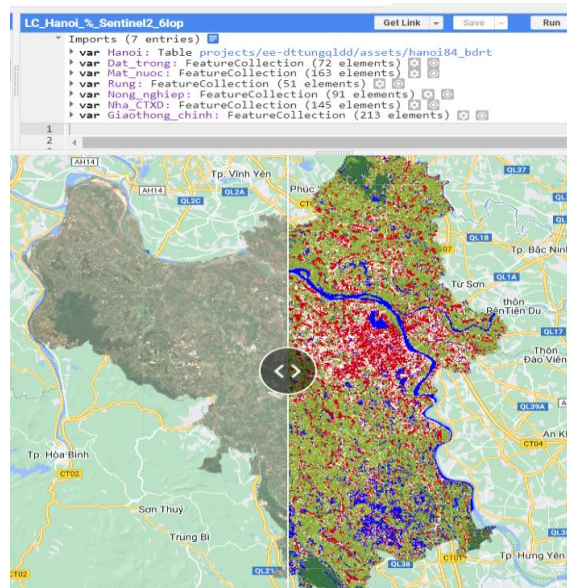
Thời gian	Loại ảnh	Độ phủ mây
2013	LandSat 8	0,05%
2014	LandSat 8	0,86%
2015	LandSat 8	2,68%
2016	LandSat 8	0,12%
2017	LandSat 8	1,03%
2018	LandSat 8	2,5%
2019	Sentinel 2	1,53%
2020	Sentinel 2	0,98%
2021	Sentinel 2	0,73%
2022	Sentinel 2	1,22%
2023	Sentinel 2	0,88%

3. Phương pháp nghiên cứu

3.1. Tổng quan về phương pháp nghiên cứu

Nghiên cứu sử dụng, thuật toán học máy RF để tính toán, phân loại lớp phủ bề mặt với dữ liệu đầu vào là ảnh vệ tinh Landsat 8, Sentinel 2 được thực hiện dựa trên nền tảng tảng dữ liệu điện toán đám mây GEE và ngôn ngữ lập trình JavaScript.

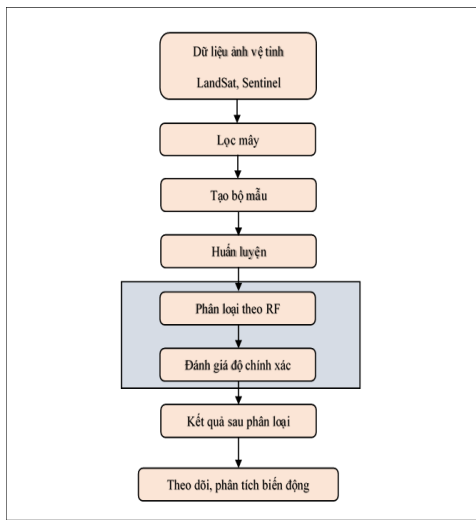
Trong nghiên cứu này, nhóm tác giả lựa chọn 6 lớp phủ để phân loại lần lượt bao gồm: i) Lớp phủ Đất trồng; ii) Lớp phủ Mặt nước; iii) Lớp phủ Thực vật dày đặc; iv) Lớp phủ Cây nông nghiệp; v) Lớp phủ Đất xây dựng; và vi) Lớp phủ Giao thông. Các lớp phủ được lựa chọn để phân loại là hiện trạng tại thời điểm thu nhận ảnh, trong đó với lớp Đất trồng các 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 dày đặc bao gồm những khu vực có cây xanh dày đặc như là khu vực cây xanh có bóng mát lớn, cây bóng mát trong các công viên, một số khu rừng như ở vườn quốc gia Ba Vì,... Lớp phủ Cây nông nghiệp bao gồm bãi cỏ, ruộng vườn tại thời điểm thu nhận ảnh đang có cây trồng. Lớp phủ Đất xây dựng gồm nhà cửa, 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. Lớp phủ Giao thông gồm các loại đường lớn, đường chính trong khu vực nghiên cứu.



Hình 2. Quá trình thực hiện phân loại lớp phủ bề mặt từ dữ liệu ảnh vệ tinh trên GEE.

Các kết quả phân loại dựa trên việc huấn luyện máy học, các dữ liệu huấn luyện được tạo ra từ các bộ mẫu, trong đó các mẫu cụ thể được lấy trên ảnh vệ tinh cho từng loại lớp phủ. Tổng số gồm 735 mẫu đã được lựa chọn phục vụ việc huấn luyện máy học. Các mẫu được lựa chọn là các khu vực dạng vùng (polygon) tương ứng với đặc điểm của từng loại lớp phủ trên ảnh vệ tinh. Trong đó, số lượng mẫu cụ thể cho từng loại lớp phủ như sau: i) Lớp phủ Đất trồng có 72 mẫu; ii) lớp phủ Mặt nước có 163 mẫu; iii) Lớp phủ Thực vật dày đặc 51 mẫu; iv) Lớp phủ cây nông nghiệp có 91 mẫu; v) Lớp phủ Đất xây dựng 145 mẫu; vi) Lớp phủ Giao thông có 213 mẫu. Nghiên cứu này sử dụng ngôn ngữ lập trình JavaScript để thực hiện phân loại lớp phủ bề mặt từ dữ liệu ảnh vệ tinh, với các lệnh lựa chọn các kênh ảnh như là `image.select(bands)` và các hàm lấy mẫu `sampleRegions` để thực hiện việc huấn luyện mẫu. Kết thúc quá trình huấn luyện, sử dụng hàm `ee.Classifier.randomForest()` để tiến hành phân loại các lớp phủ theo thuật toán RF trong khu vực nghiên cứu. Hình 2 thể hiện kết quả thực hiện của chương trình trên GEE. Các bước xử lý, phân loại ảnh bao gồm: thu thập dữ liệu ảnh vệ tinh dựa trên nền tảng dữ liệu điện toán đám mây

GEE; Lọc và lựa chọn ảnh có độ phủ mây là thấp nhất; Tạo bộ mẫu; Huấn luyện máy học; Phân loại các lớp phủ trên ảnh vệ tinh theo các thuật toán RF; Thu nhận kết quả ảnh phân loại; Đánh giá độ chính xác của kết quả sau phân loại. Sau khi có kết quả phân loại ảnh, tiến hành các bước phân tích, theo dõi biến động các lớp phủ sử dụng đất giai đoạn 2013-2023 tại khu vực nghiên cứu (Hình 3).



Hình 3. Sơ đồ quy trình phân loại ảnh.

3.2. Thuật toán RF

RF là một thuật toán máy học tích hợp có thể tích hợp nhiều cây quyết định và sau đó tạo thành một khu rừng. Thuật toán này kết hợp các tính năng ngẫu nhiên để tạo ra một cây. Phương pháp đóng bao được sử dụng để tạo các mẫu huấn luyện và mỗi tính năng đã chọn được rút ngẫu nhiên bằng cách thay thế N (kích thước của tập huấn luyện ban đầu). Sau đó, kết quả dự đoán cuối cùng thu được bằng cách kết hợp nhiều cây quyết định [1]. Công thức dưới đây thực hiện quyết định phân loại cuối cùng như sau:

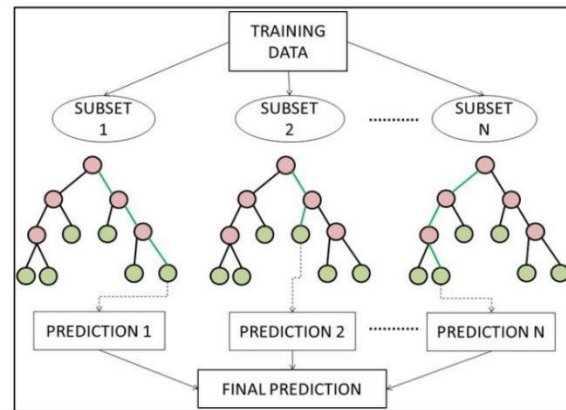
$$H(x) = \operatorname{argmax}_Y \sum_{i=1}^k I(h_i(x) = Y)$$

Trong đó:

$H(x)$ là mô hình kết hợp, h_i là mô hình phân loại của cây quyết định đơn, k hệ số đàn hồi;

Y là biến đầu ra (hoặc biến mục tiêu) và $I(\cdot)$ là hàm chỉ báo.

Công thức cho thấy rằng thuật toán RF sử dụng đa số các quyết định biểu quyết để xác định phân loại cuối cùng. Tham số điều chỉnh của thuật toán RF là số lượng cây và số lượng cây có thể được chọn theo kinh nghiệm của người xử lý tính toán. Trong các bài toán phân lớp dữ liệu thì thuật toán RF được sử dụng tương đối phổ biến. Thuật toán RF được đánh giá cao bởi tính chính xác của mô hình [9, 10]. Nhược điểm chính của thuật toán RF là khối lượng tính toán lớn [11, 12].



Hình 4. Mô hình phân loại theo thuật toán RF.

3.3. Phương pháp đánh giá độ chính xác

Ma trận sai lẫn (Confusion Matrix) được sử dụng để đánh giá độ chính xác kết quả phân loại các lớp phủ trong nghiên cứu này, nó có thể mô tả độ chính xác của phân loại và chỉ ra sự sai lẫn giữa các lớp đối tượng [13-15]. Trong đó, các thống kê cơ bản bao gồm đánh giá về độ chính xác tổng thể (Overall Accuracy - OA) và hệ số Kappa. 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 [10]. Thực hiện đánh giá độ chính xác ảnh sau phân loại bằng việc sử dụng hàm errorMatrix trên GEE, trong đó sử dụng 70% số lượng mẫu dùng để phân loại ảnh và 30% số lượng mẫu dùng để kiểm tra đánh giá [16, 17].

4. Kết quả nghiên cứu, thảo luận

Kết quả sau phân loại bao gồm 6 lớp phủ là lớp phủ Đất trồng, lớp phủ Mặt nước, lớp phủ Rừng, lớp phủ Cây nông nghiệp, lớp phủ Đất xây dựng, lớp phủ Giao thông. Nghiên cứu đã cho thấy sự biến động rõ rệt của các lớp phủ sử dụng đất tại khu vực nghiên cứu trong giai đoạn 2013-2023. Hình 5 minh họa sản phẩm các lớp phủ bề

mặt sau phân loại tại khu vực nghiên cứu theo thuật toán RF

Kết quả giám sát biến động lớp phủ mặt nước giai đoạn 2013-2023 tại khu vực Hà Nội cho thấy rằng có sự suy giảm đáng kể về diện tích mặt nước qua các giai đoạn và các khu vực khác nhau tại Hà Nội. Điều này cho thấy rằng sau khoảng 10 năm phát triển, lớp Mặt nước có diện tích giảm 8,80% tương đương khoảng 29406 ha.



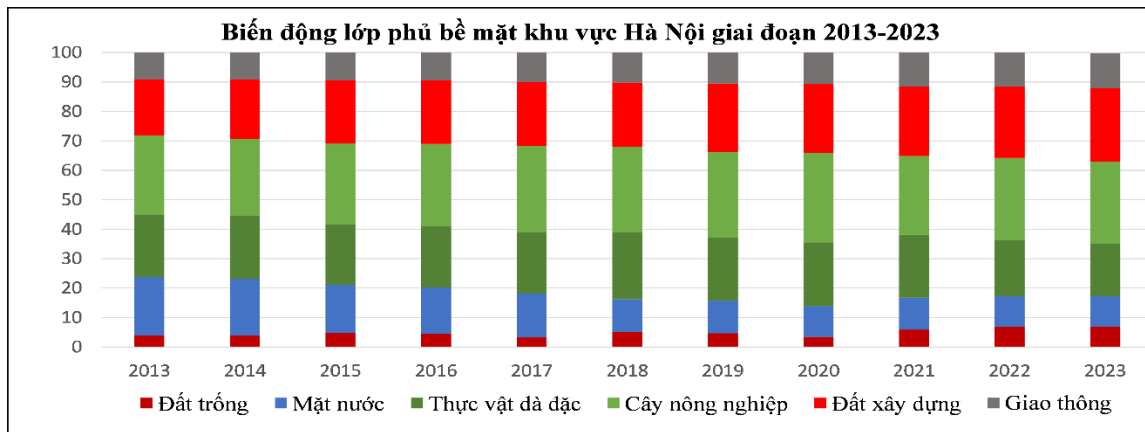
Hình 5. Các lớp phủ bề mặt tại Hà Nội giai đoạn 2013-2023.

Như vậy, đối với lớp phủ Mặt nước, mỗi năm giảm trung bình khoảng 0,8% so với tổng diện tích tự nhiên. Tổng thể, từ năm 2013 đến 2023, xu hướng giảm diện tích của lớp phủ là Mặt nước

là rõ ràng. Theo kết quả phân tích, trung bình hàng năm trong giai đoạn này, tốc độ suy giảm diện tích của lớp Mặt nước là tương đối cao trong giai đoạn từ 2013 đến 2018 có tốc độ giảm trung

bình khoảng gần 1%/năm, còn trong giai đoạn từ 2018 đến 2023, tốc độ suy giảm diện tích mặt nước tại Hà Nội đã giảm đáng kể chỉ còn khoảng gần 0,3%/năm. Việc suy giảm diện tích mặt nước tại Hà Nội cũng được JICA thống kê năm với kết quả trong vòng 15 năm, số lượng hồ tự nhiên của

Hà Nội (chưa mở rộng) đã giảm từ 40 hồ xuống còn 19 hồ. Cùng với 21 hồ mất tích, diện tích mặt nước hồ đã thu hẹp từ 850 ha xuống còn 547 ha. Hình 6 thể hiện biểu đồ và số liệu về biến động lớp phủ sử dụng mặt nước khu vực Hà Nội.



Hình 6. Kết quả theo dõi lớp phủ bề mặt khu vực Hà Nội giai đoạn 2013-2023.

Đối với các lớp phủ bề mặt khác tại khu vực Hà Nội giai đoạn 2013-2023 có sự tăng giảm khác biệt theo những thời điểm khác nhau. Tại khu vực nội thành các lớp phủ bề mặt có sự biến động tương đối nhỏ, ngược lại các khu vực xa trung tâm có biến động lớn hơn, nhất là các khu vực mới chuyển thành cấp. Kết quả về việc mở rộng khu vực đô thị này cũng tương đồng với một số công bố của nghiên cứu khác về phát triển đô thị ở Việt Nam [3, 18]. Trong đó đối với 2 lớp phủ là Đất xây dựng và lớp phủ Giao thông cho thấy có sự tương quan về tăng và giảm diện tích của hai lớp này trong giai đoạn nghiên cứu. Cụ thể, đối với khu vực quận Bắc Từ Liêm và Nam Từ Liêm có diện tích lớp Cây nông nghiệp có xu hướng giảm trung bình 0,91% một năm, ngược lại lớp phủ Đất xây dựng có xu hướng tăng 1,07% mỗi năm trong giai đoạn từ 2013 đến 2023. Như vậy, có thể thấy rằng sau thời gian khoảng 10 năm phát triển thành cấp quận các khu vực này có tốc độ đô thị hóa là tương đối cao và thể hiện rất rõ ràng bởi sự phát triển, mở rộng diện tích nhanh chóng của các khu vực khu dân cư, các khu chung cư, các khu vực nhà cửa thương mại,... và các khu vực đất canh tác nông

nghiệp có diện tích suy giảm rõ rệt. Nhất là dưới tác động của biến đổi khí hậu, xây dựng đô thị và các hoạt động đổ thải, san lấp ao hồ trái phép đang diễn biến phức tạp cũng là một trong những nguyên nhân gây suy giảm diện tích mặt nước một cách nhanh chóng tại khu vực nghiên cứu [19].

Kết quả phân tích biến động khu vực Hà Nội giai đoạn 2013-2023 đã cho thấy diện tích của lớp Đất xây dựng đã tăng khoảng 7%, lớp Giao thông tăng khoảng 3%, lớp Đất trồng tăng 3%, lớp Cây nông nghiệp tăng 2% trong khi đó lớp Thực phủ dày đặc giảm 6%, lớp mặt nước giảm 9%. Việc biến động lớp phủ Cây nông nghiệp là không có xu hướng thống nhất, có những năm diện tích lớp phủ này tăng so với năm trước, có những năm lại giảm so với năm trước đó 1-2%, tuy nhiên so sánh kết quả của năm 2013 với năm 2023 thì diện tích lớp phủ Cây nông nghiệp tăng khoảng hơn 2%. Kết quả biến động này có thể có nhiều nguyên nhân như là điều kiện thời tiết thay đổi khiến các đối tượng cây bụi, mùa màng mỗi năm mỗi khác, việc chuyển đổi cơ cấu kinh tế, sản xuất, chuyển đổi cây trồng, loại cây trồng của từng năm khác nhau,... đây là một trong những thực tế mà nghiên cứu này chưa đủ điều kiện để

xác định rõ. Các kết quả tăng giảm và chuyển đổi từ loại đất này sang loại đất khác được thể hiện

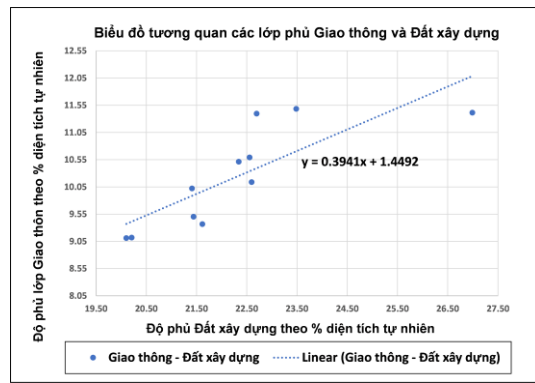
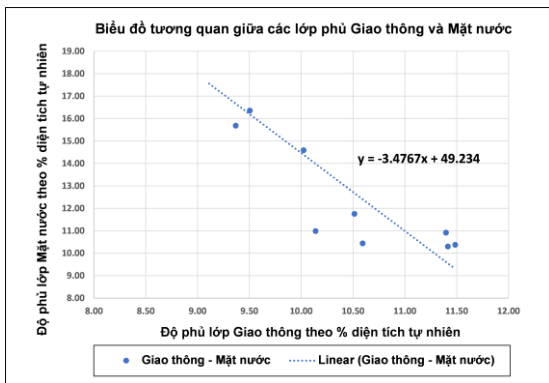
chi tiết trong Bảng 2 về ma trận biến động lớp phủ bề mặt [20-22] dưới đây.

Bảng 2. Ma trận biến động lớp phủ bề mặt khu vực Hà Nội 2013-2023 (đơn vị: m²)

2013-2023	Đất trống	Giao thông	Mặt nước	Đất xây dựng	Cây nông nghiệp	Thực phủ dày đặc	Tổng hàng
Đất trống	13341662	36960931	33873075	62190404	82084926	13928838	242379838
Giao thông	13297384	55031183	62027097	132610536	99005992	20754356	382726551
Mặt nước	10831063	17076448	235621366	27233879	43540948	11106081	345409788
Đất xây dựng	41152839	62416039	149124392	194842610	180959476	275255037	903750395
Cây nông nghiệp	44506325	108846431	117620627	216368859	399064717	81114589	967521550
Thực phủ dày đặc	12969241	25167101	42597219	37096226	77632003	311946745	507408538
Tổng cột	136098516	305498136	640863778	670342516	882288065	716105649	3344196662

Các biến động tại khu vực nghiên cứu cũng có các tương quan nhất định giữa các lớp phủ. Trong nghiên cứu này, nhóm tác giả sử dụng hệ số tương quan r Pearson để xác định các mối tương quan giữa các lớp phủ sử dụng đất [23]. Cụ thể và thể hiện rõ ràng nhất là sự tương quan thuận giữa lớp Giao thông và lớp Đất xây dựng, lớp phủ Mặt nước. Khi diện tích các khu vực Đất xây dựng tăng thì đồng thời lớp Giao thông cũng tăng, hai lớp này có tương quan thuận với hệ số r(Pearson)= 0,805 như vậy là có tương quan rất chặt chẽ và phù hợp với thực tế phát triển, đô thị

hóa của khu vực nghiên cứu trong những năm qua. Ngược lại với sự gia tăng về diện tích của lớp Giao thông là sự suy giảm nghiêm trọng của lớp phủ Mặt nước trong giai đoạn 2013-2023. Mối tương quan giữa hai lớp diện tích Giao thông và diện tích Mặt nước là tương quan nghịch với r(Pearson)=-0,899, có nghĩa là giao thông, xây dựng, đô thị hóa ngày càng phát triển thì đồng thời diện tích mặt nước và một số lớp khác cũng ngày càng suy giảm [24]. Các kết quả phân tích này được minh họa tại Hình 7.



Hình 7. Biểu đồ tương quan giữa lớp phủ Giao thông và Mặt nước, Đất xây dựng giai đoạn 2013-2023 tại khu vực Hà Nội.

Tập thể tác giả sử dụng ma trận sai lẫn (Confusion matrix) để đánh giá độ chính xác phân loại các lớp phủ bề mặt. 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 sai lẫn bao gồm các

thông tin cơ bản là độ chính xác tổng thể (OA) và hệ số Kappa được thể hiện trong Bảng 2 bao gồm kết quả đánh giá độ chính xác phân loại các lớp phủ cho giai đoạn 2013-2023. Trong đó, độ chính xác tổng thể OA thấp nhất là 0,91, cao nhất là 0,94, đồng thời hệ số Kappa cũng đạt từ 0,86 trở lên. Theo các đánh giá về giá trị của hệ số Kappa đạt từ trên 0,41 đến 0,60 là có độ chính xác trung bình; Kappa từ 0,61 đến 0,80 là có độ chính xác tốt; Kappa từ 0,81 đến 1,0 là rất tốt, còn dưới 0,40 là độ chính xác kém [25, 26]. 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 đưa ra các thống kê, phân tích về các lớp phủ sử dụng đất trong giai đoạn 2013-2023 tại khu vực Hà Nội.

Bảng 3. Độ chính xác phân loại ảnh

STT	Thời gian	OA	Kappa
1	2023	91,37%	86,00%
2	2022	94,70%	91,23%
3	2021	92,37%	87,64%
4	2020	92,95%	88,57%
5	2019	92,08%	87,15%
6	2018	94,01%	90,20%
7	2017	91,62%	86,36%
8	2016	93,01%	88,61%
9	2015	91,40%	86,00%
10	2014	92,60%	87,99%
11	2013	91,39%	86,09%

Độ chính xác phân loại ảnh phụ thuộc vào chất lượng ảnh và nhất là chất lượng của các mẫu huấn luyện [27, 28]. Trong nghiên cứu này, chất lượng ảnh thu nhận qua các năm là tương đối đồng đều, độ phủ mây rất thấp, hình ảnh và các thông tin vật lý rõ ràng nên theo kết quả đánh giá tất cả các ảnh sau phân loại từ 2013 đến 2023 đều có độ tin cậy cao.

4. Kết luận

Phương pháp nghiên cứu, sử dụng mô hình trí tuệ nhân tạo với thuật toán RF để phân loại các lớp phủ sử dụng đất là có chất lượng cao, kết

quả sau phân loại đảm bảo yêu cầu cho công tác theo dõi biến động sử dụng đất tại khu vực nghiên cứu. Sử dụng mô hình trí tuệ nhân tạo với thuật toán RF mang lại một số ưu điểm so với các phương pháp phân loại truyền thống như là có khả năng tích hợp đặc trưng nâng cao từ đó tăng khả năng phân loại chính xác; có thể xử lý những phi tuyến này một cách hiệu quả, giúp nó phù hợp với các lớp phủ bề mặt phức tạp và đa dạng; có khả năng tổng quát hóa tốt đối với dữ liệu chưa từng được thấy trước; tự động thực hiện việc chọn đặc trưng bằng cách đánh giá tầm quan trọng của từng biến đầu vào, việc này là tương đối khó khăn khi thực hiện bằng các phương pháp truyền thống; khả năng chống nhiễu giúp duy trì độ chính xác phân loại. Các kết quả phân loại đều đảm bảo độ chính xác cao với hệ số Kappa và OA tương ứng đạt từ 0,86 và 0,91 trở lên.

Kết quả cũng cho thấy rõ ràng về sự phát triển, đô thị hóa nhanh chóng của một số khu vực trên địa bàn Hà Nội. Đồng thời cũng chỉ ra được sự gia tăng diện tích các khu vực xây dựng, các loại đất giao thông tương đối nhanh với khoảng 7% so với tổng diện tích tự nhiên đối với Đất xây dựng trong giai đoạn 2013-2023 tại khu vực Hà Nội. Ngoài ra kết quả theo dõi cũng cho thấy tốc độ suy giảm lớp phủ mặt nước rất cao là gần 0,8%/năm, sau 10 năm theo dõi từ 2013 đến 2023, diện tích của lớp phủ mặt nước giảm khoảng 8,8%. Kết quả thực tế này có khả năng góp phần hiệu quả vào việc phục vụ công tác quản lý, quy hoạch sử dụng đất tại khu vực Hà Nội.

Lời cảm ơn

Xin trân trọng cảm ơn Khoa Quản lý đất đai, Trường Đại học Tài nguyên và Môi trường Hà Nội đã hỗ trợ nhóm tác giả thực hiện nội dung nghiên cứu này.

Tài liệu tham khảo

- [1] L. J. M. Breim, Random Forests, Kluwer Academic Publishers Manufactured in the Netherlands, Vol. 45, Iss. 5, 2001.

- [2] N. T. P. Thi et al., Applying Random Forest Approach in Forecasting Flash Flood Susceptibility Area in Lao Cai Region, *Journal of Mining and Earth Sciences*, Vol. 61, Iss. 5, 2020, pp. 30-42.
- [3] L. H. Trinh et al., Assessment and Prediction of Urban Land Use Changes of Hanoi City Using Remote Sensing and GIS Technique, *Natural Sciences and Technolog* Vol. 14, 2017, pp. 176-187.
- [4] T. N. Nguyen et al., Application of Remote Sensing and GIS in Studying Land Cover Changes in Thanh Tri District, Hanoi City in The Period 2001-2008, Doctoral Dissertation, 2011.
- [5] N. T. Luyen, Urban Land Cover Classification with Venus Satellite Images for Hanoi, Vietnam, Doctoral Dissertation, 2019.
- [6] G. Noel et al, Google Earth Engine: Planetary-scale Geospatial Analysis for Everyone, *Remote Sensing of Environment*, Vol. 202, 2017, pp. 18-27.
- [7] A. S. Mishra, Google Earth Engine: A Review of Its Applications and Future Directions, *Journal of the Indian Society of Remote Sensing*, Vol. 49, Iss. 1, 2021, pp. 107-124.
- [8] Hanoi Portal, An Overview of the Geography of Hanoi City, <https://hanoi.gov.vn/diachihanoi/-/hn/RtLibd2X8kEn/1001/124742/gioi-thieu-tong-quan-va-khai-quat-ve-ia-li-thanh-pho-hanoi.html#:~:text=Th%E1%BB%A7%20%C4%91%C3%B4%20H%C3%A0%20N%E1%BB%99i%20sau,x%C3%A3%2C%20ph%C6%B0%E1%BB%9Dng%2C%20th%E1%BB%8B%20tr%E1%BA%A,> 2023 (accessed on: April 10th, 2023).
- [9] G. Robin et al., Random Forests for Big Data, *Big Data Research*, Vol. 9, 2017, pp. 28-46.
- [10] B. Leo, Random Forests, *Machine Learning*, Vol. 45, 2001, pp. 5-32.
- [11] H. T. Kam, Random Decision Forests, *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, Montreal, QC, 1995, pp. 278-282.
- [12] S. Deliang et al., A Random Forest Model of Landslide Susceptibility Mapping Based on Hyperparameter Optimization Using Bayes Algorithm, *Geomorphology*, Vol. 362, 2020, pp. 107201.
- [13] L. Amalia et al., The Impact of Class Imbalance in Classification Performance Metrics Based on the Binary Confusion Matrix, *Pattern Recognition*, Vol. 91, 2019, pp. 216-231,
- [14] C. Alexis et al., Spatial Analysis of Remote Sensing Image Classification Accuracy, *Remote Sensing of Environment*, Vol. 127, 2012, pp. 237-246.
- [15] V. Sofia et al., Confusion Matrix-Based Feature Selection, Vol. 120, 2011, pp. 710.
- [16] P. Uday et al., Google Earth Engine Based Three Decadal Landsat Imagery Analysis for Mapping of Mangrove Forests and Its Surroundings in the Trat Province of Thai Land, 2018.
- [17] K. Karren, G. Lightbody, R. Yacamini, Power System Harmonic Analysis Using the Kalman Filter, in 2003 IEEE Power Engineering Society General Meeting (IEEE Cat. No. 03CH37491), Vol. 2, 2003, pp. 752-757.
- [18] L. V. Hao, L. T. P. Mi, Application of Remote Sensing and GIS to Monitor the Urbanization Process in Ho Chi Minh City in The Period 1989–2019, *Journal of Hydrometeorology*, Vol. 720, 2020, pp. 48-59.
- [19] Department of Water Resources Management, Symptoms of Ground Water Degradation in Hanoi City and Causes, [http://dwr.gov.vn/index.php?language=vi&nv=news&op=Khoa-hoc-Cong-nghe/cac-bieu-hien-suy-thoai-nuoc-duoi-dat-vung-thanh-pho-ha-noi-va-cac-nguyen-nhan-3036,](http://dwr.gov.vn/index.php?language=vi&nv=news&op=Khoa-hoc-Cong-nghe/cac-bieu-hien-suy-thoai-nuoc-duoi-dat-vung-thanh-pho-ha-noi-va-cac-nguyen-nhan-3036) 2023 (accessed on: April 10th, 2023).
- [20] E. F. Lambin, H. J. Geist, E. Lepers, Dynamics of Land-use and Land-cover Change in Tropical Regions, *Annual Review of Environment and Resources*, Vol. 28, Iss. 1, 2003, pp. 205-241.
- [21] T. Billie, F. L. Eric, A. Reenberg, The Emergence of Land Change Science for Global Environmental Change and Sustainability, *Proceedings of the National Academy of Sciences*, Vol. 104, Iss. 52, 2007, pp. 20666-20671.
- [22] N. T. Xuan, D. T. B. Hoa, Application of Remote Sensing and GIS for Analysis of Wetland for Building Multi-Benefit Solution to Sustainable Uses of Wetland in Quang Yen Commune, Quang Ninh Province, *Journal of Geodesy and Cartography*, Vol. 22, 2014, pp. 24-31.
- [23] M. K. Donald, Mathematical Contributions to the Theory of Evolution, VII, on the Correlation of Characters not Quantitatively Measurable, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 187, 1986, pp. 253.
- [24] P. M. Hai et al., Impact of Urban Expansion to Cultural Heritage: the Case Study of the Complex of Hue Monu-ments, *Journal of Geodesy and Cartography*, Vol. 40, 2019, pp. 42-49.

- [25] C. Jacob, A Coefficient of Agreement for Nominal Scales, Educational and Psychological Measurement, Vol. 20, No. 1, 1960, pp. 37-46.
- [26] M. W. P. David, Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation, Journal of Machine Learning Technologies, Vol. 2, Iss. 1, 2011, pp. 37-63.
- [27] F. Bin Siddique et al., Impact of Image Quality on Object Detection Performance: A Study on Influence of JPEG Compression, Noise, and Resolution, IEEE Transactions on Image Processing, 2019.
- [28] K. Nørvåg et al., The Importance of Data Quality for Classification Performance in Machine Learning, Journal of Intelligent Learning Systems and Applications, 2016.