

ỨNG DỤNG MACHINE LEARNING THEO DÕI BIẾN ĐỘNG SỬ DỤNG ĐẤT GIAI ĐOẠN 2010 - 2024 TẠI THÀNH PHỐ TÂN AN, TỈNH LONG AN

Nguyễn Kim Hoa^{1,2}, Nguyễn Thị Hải¹, Nguyễn Quang Thi³, Nguyễn Hữu Ngự^{1,*}

¹Trường Đại học Nông Lâm, Đại học Huế

²Trường Đại học Tài nguyên và Môi trường thành phố Hồ Chí Minh

³Trường Đại học Nông Lâm, Đại học Thái Nguyên

*Email: nguyenuhuunguwx@huaf.edu.vn

TÓM TẮT

Nghiên cứu này ứng dụng thuật toán học máy Random Forest để phân loại và theo dõi sự thay đổi lớp phủ sử dụng đất tại thành phố Tân An, tỉnh Long An trong giai đoạn 2010 - 2024. Dữ liệu đầu vào gồm: Ảnh vệ tinh Landsat (các năm 2010, 2015, 2020), Sentinel-2 (năm 2024), dữ liệu mô hình số độ cao DEM và các chỉ số như: NDVI, NDWI, NDBI, BSI, LSWI, MNDWI, AWEI. Các mẫu huấn luyện được xây dựng gồm 6 lớp sử dụng đất chính gồm: Đất trống, mặt nước, cây lâu năm, cây hàng năm, khu dân cư và giao thông. Kết quả phân loại đạt độ chính xác cao với tổng độ chính xác tổng thể (OA) trên 91% và hệ số Kappa trên 0,89 cho từng năm. Biến động sử dụng đất thể hiện xu hướng đô thị hóa rõ rệt, với diện tích đất dân cư và giao thông tăng mạnh trong khi đất nông nghiệp giảm. Giai đoạn 2010 - 2024, đất dân cư tăng hơn 1.600 ha, chủ yếu chuyển đổi từ đất cây lâu năm và cây hàng năm. Trong khi đó, mặt nước biến động nhẹ, còn đất trống giảm mạnh qua các năm. Các bản đồ lớp phủ theo từng giai đoạn, bản đồ biến động không gian và ma trận chuyển đổi sử dụng đất được xây dựng và biên tập trên phần mềm GIS đã cung cấp cái nhìn trực quan về quá trình chuyển đổi sử dụng đất. Kết quả cho thấy xu hướng chuyển đổi rõ rệt từ đất nông nghiệp sang đất xây dựng, phản ánh tốc độ đô thị hóa mạnh mẽ của thành phố Tân An trong hơn một thập kỷ qua. Nghiên cứu góp phần cung cấp cơ sở khoa học cho công tác đánh giá và định hướng quy hoạch sử dụng đất tại địa phương.

Từ khóa: Biến động sử dụng đất, Google Earth Engine, học máy, Random Forest, Tân An.

1. ĐẶT VẤN ĐỀ

Trong bối cảnh đô thị hóa đang diễn ra ngày càng mạnh mẽ, việc theo dõi và đánh giá biến động sử dụng đất (SDĐ) đóng vai trò quan trọng nhằm đảm bảo quản lý tài nguyên hiệu quả và định hướng phát triển bền vững. Tại Việt Nam, đặc biệt là ở các đô thị cấp tỉnh như thành phố Tân An tỉnh Long An, sự gia tăng dân số và nhu cầu phát triển cơ sở hạ tầng đã thúc đẩy quá trình chuyển đổi mạnh mẽ từ đất nông nghiệp sang đất phi nông nghiệp, gây ra nhiều thách thức trong công tác quy hoạch và bảo vệ đất đai [1].

Trước đây, các phương pháp truyền thống và kiểm kê đất đai định kỳ hiện nay chủ yếu dựa trên dữ liệu hành chính và bản đồ hiện trạng, vì vậy còn

nhiều hạn chế trong việc cập nhật kịp thời và thể hiện chi tiết biến động theo không gian [2]. Công nghệ GIS và viễn thám đã và đang phát triển mạnh, các phương pháp phân tích, đánh giá về biến động SDĐ, về đô thị hóa cũng ngày càng trở nên thuận tiện và có hiệu quả. Hiện nay, đã có những nghiên cứu trên thế giới sử dụng các thuật toán học máy Machine Learning (ML) để phân loại lớp phủ dựa trên các phần mềm thương mại chuyên dụng để phân loại lớp phủ bề mặt đất từ ảnh vệ tinh, như nghiên cứu của Swapan Talukdar và cs (2020) [3], Brian W Szuster và cs (2011) [4], Abbas Taati và cs (2015) [5], Subhra Swetanisha và cs (2022) [6]. Ngoài ra, nghiên cứu của Phạm Anh Tuấn (2024) [7] cũng đã sử dụng ML là thuật

toán rừng ngẫu nhiên Random Forest (RF) và ngôn ngữ lập trình Java Script để xây dựng chương trình tự động hóa việc phân loại, giám sát biến động các lớp phủ SĐĐ dựa trên nền tảng dữ liệu điện toán đám mây Google Earth Engine (GEE) từ ảnh vệ tinh Sentinel-2. Nghiên cứu của Nguyễn Thị Huyền Trang và cs (2023) [8] sử dụng thuật toán học máy Support Vector Machine (SVM) để phân loại, xác định các lớp phủ SĐĐ tại khu vực quận Bắc Từ Liêm, thành phố Hà Nội năm 2023. Kết quả của các nghiên cứu này đã sử dụng ML để phân loại được một số lớp phủ bề mặt của khu vực nghiên cứu. Tuy nhiên, để nâng cao hiệu quả hơn nữa, cần có kết quả phân tích của nhiều năm liên tục để theo dõi, giám sát biến động lớp phủ SĐĐ theo các giai đoạn khác nhau.

Ngoài ra, một trong những nền tảng ứng dụng hữu ích khác có thể kể đến là GEE, đây là nền tảng dữ liệu điện toán đám mây, có khả năng xử lý dữ liệu viễn thám online, đáp ứng kịp thời các yêu cầu cung cấp thông tin, dữ liệu phục vụ công tác giám sát biến động SĐĐ [9]. Gần đây, các thuật toán học máy như RF đã được chứng minh là có khả năng phân loại ảnh vệ tinh hiệu quả, đặc biệt trong môi trường dữ liệu đa chiều và phức tạp [10]. Thuật toán này không những có khả năng kháng nhiễu tốt, mà còn có thể đánh giá được mức độ

quan trọng của từng biến đầu vào, phù hợp với dữ liệu từ ảnh Landsat, Sentinel-2, chỉ số NDVI, NDWI, NDBI, cùng các yếu tố địa hình như mô hình số độ cao (viết tắt là DEM) [11].

Xuất phát từ yêu cầu thực tiễn và khả năng ứng dụng cao của công nghệ mới, nghiên cứu này được thực hiện nhằm: (1) Ứng dụng thuật toán RF để phân loại lớp phủ SĐĐ từ dữ liệu viễn thám đa nguồn; (2) Phân tích biến động SĐĐ giai đoạn 2010 - 2024 tại thành phố Tân An, nơi đang chịu áp lực lớn từ quá trình đô thị hóa.

2. VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Khu vực nghiên cứu

Thành phố Tân An được xem là một trong những đô thị vệ tinh của thành phố Hồ Chí Minh và là đô thị cửa ngõ của vùng đồng bằng sông Cửu Long, có vị trí trọng yếu về an ninh quốc phòng với sự giao lưu thuận lợi bằng các tuyến giao thông quan trọng như Quốc lộ 1A, đường cao tốc thành phố Hồ Chí Minh - Trung Lương, Quốc lộ 62 và sông Vàm Cỏ Tây. Thành phố Tân An là một trong những khu vực phát triển kinh tế - xã hội trọng điểm của tỉnh Long An, nằm cách thành phố Hồ Chí Minh khoảng 47 km về phía Tây Nam theo tuyến Quốc lộ 1A (Hình 1).



Hình 1. Sơ đồ vị trí khu vực nghiên cứu

2.2. Vật liệu nghiên cứu

Nghiên cứu sử dụng tổng hợp các nguồn dữ liệu viễn thám, địa hình và dữ liệu huấn luyện thực địa như sau:

Ảnh vệ tinh quang học đa thời gian:

- Landsat 5 TM: Sử dụng cho năm 2010, độ phân giải không gian 30 m, gồm 6 kênh phản xạ (B1 - B5, B7), dữ liệu từ bộ *LANDSAT/LT05/C02/T1_L2*.

- Landsat 8 OLI: Áp dụng cho các năm 2015 và 2020, độ phân giải 30 m, kênh phổ từ B2 - B7, trích xuất từ bộ *LANDSAT/LC08/C02/T1_L2*.

- Sentinel-2 MSI: Sử dụng cho năm 2024, độ phân giải 10 - 20 m, dữ liệu hiệu chỉnh khí quyển từ bộ *COPERNICUS/S2_SR*, gồm các kênh B2 - B12.

Để đảm bảo chất lượng, ảnh được lựa chọn trong hai mùa (mùa khô: 1/1 - 30/4 và mùa mưa: 1/7 - 31/10) với điều kiện mây thấp (< 10% với Sentinel-2 và Landsat 8, < 20% với Landsat 5). Ảnh trong mỗi mùa được tổng hợp theo trung vị để tạo ảnh đại diện cho từng năm.

Chỉ số phổ và dữ liệu độ cao:

Từ ảnh vệ tinh, các chỉ số được tính toán nhằm tăng khả năng phân biệt lớp phủ đất:

- Chỉ số thực vật:

+ Normalized Difference Vegetation Index (NDVI).

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

+ Soil Adjusted Vegetation Index (SAVI).

$$SAVI = \frac{(NIR - Red)(1 + L)}{NIR + Red + L}, L = 0,5$$

Trong đó: L là hằng số hiệu chỉnh ảnh hưởng nền đất (0 - 1; 0,5 thường dùng khi che phủ trung bình).

- Chỉ số nước:

+ Normalized Difference Water Index (NDWI)

$$NDWI = \frac{Green - NIR}{Green + NIR}$$

+ Modified Normalized Difference Water Index (MNDWI)

$$MNDWI = \frac{Green - SWIR1}{Green + SWIR1}$$

+ Automated Water Extraction Index for shadowed areas (AWEI)

$$AWEI_{sh} = 4(Green - SWIR1) - (0,25NIR + 2,75SWIR2)$$

- Chỉ số xây dựng và đất trống:

+ Normalized Difference Built-up Index (NDBI)

$$NDBI = \frac{SWIR1 - NIR}{SWIR1 + NIR}$$

+ Bare Soil Index (BSI)

$$BSI = \frac{(SWIR1 + Red) - (NIR + Blue)}{(SWIR1 + Red) + (NIR + Blue)}$$

+ Land Surface Water Index (LSWI)

$$LSWI = \frac{NIR - SWIR1}{NIR + SWIR1}$$

Ngoài ra, DEM từ bộ *USGS/SRTMGL1_003* (30 m) được tích hợp nhằm bổ sung thông tin độ cao phục vụ phân loại, đặc biệt hữu ích ở các vùng trũng hoặc xây dựng ven kênh, rạch.

Dữ liệu mẫu huấn luyện và kiểm tra:

- Dữ liệu mẫu được thu thập và kiểm tra trực quan dựa trên ảnh Google Earth theo từng năm trong giai đoạn nghiên cứu.

- Tổng số mẫu sử dụng là 3.350 điểm, phân bố đều cho 6 lớp SDD: Giao thông (GT) có 800 mẫu. Các lớp còn lại (đất trống, nước, cây lâu năm (CLN), cây hàng năm (CHN)): Trung bình 500 mẫu/lớp. Riêng lớp dân cư là 550 mẫu.

- Mẫu được chia ngẫu nhiên thành 02 tập: 70% huấn luyện, 30% cho kiểm định mô hình phân loại.

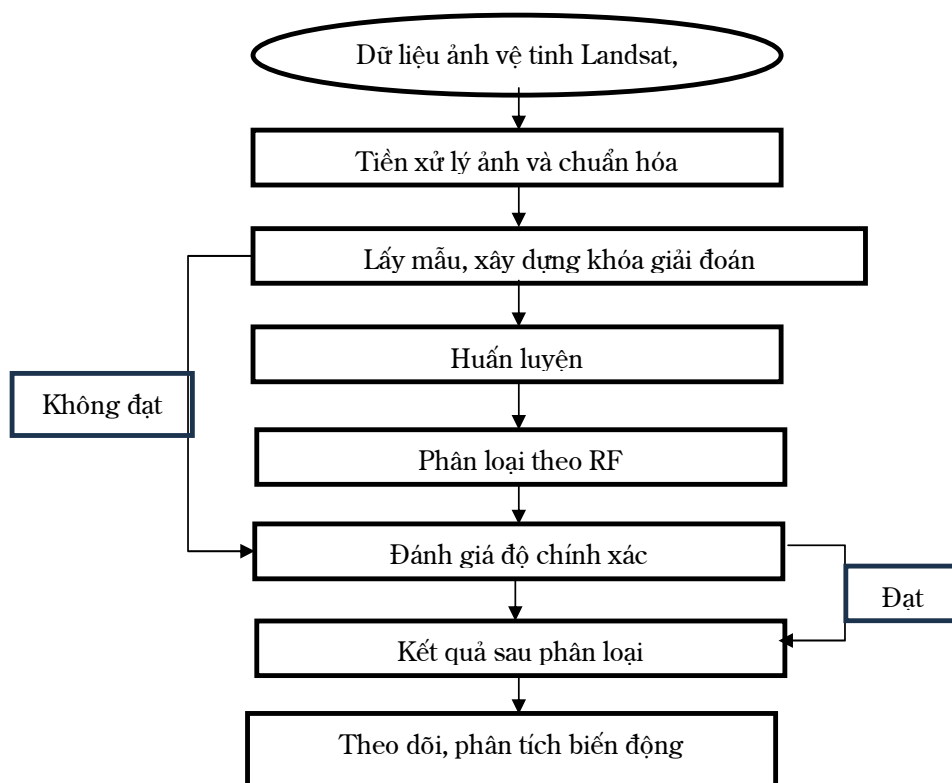
2.3. Phương pháp nghiên cứu

2.3.1. Quy trình nghiên cứu

Nghiên cứu sử dụng dữ liệu ảnh vệ tinh được cung cấp miễn phí thông qua nền tảng GEE và ứng dụng thuật toán RF để tính toán, phân loại lớp phủ bề mặt cho dữ liệu đầu vào là ảnh vệ tinh Landsat

5 TM, Landsat 8 OLI và Sentinel-2. Dữ liệu ảnh vệ tinh Sentinel-2 sử dụng trong nghiên cứu là các ảnh đã được hiệu chỉnh khí quyển, các ảnh có lượng mây che phủ thấp, được hiệu chỉnh hình học trên hệ quy chiếu WGS84. Việc phân loại ảnh được thực hiện trực tuyến trên nền tảng dữ liệu điện toán đám mây GEE. Nghiên cứu lựa chọn phân loại ảnh theo 6 lớp phủ bề mặt tương đương với các lớp: 1) Lớp phủ đất trống, 2) Lớp phủ mặt nước, 3) Lớp phủ CLN, 4) Lớp phủ CHN, 5) Lớp phủ dân cư, 6) Lớp phủ giao thông. Việc lấy mẫu để huấn luyện và thực hiện phân loại được lựa chọn cho từng lớp. Trong đó, đối với lớp phủ đất trống, các điểm lấy mẫu phân loại được 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 CLN bao gồm những khu vực có cây cối lớn, dày đặc, lâu năm. Lớp phủ CHN gồm các khu vực như hoa màu, bãi cỏ, ruộng vườn tại thời điểm thu nhận ảnh đang có cây trồng hàng năm. Lớp phủ dân cư gồm nhà cửa, công trình xây dựng độc lập và các khu vực nhà cửa, công trình xây dựng tại thôn, xóm, khu chung cư, các tòa nhà cao tầng. Lớp phủ giao thông gồm các hệ thống tuyến giao thông đường bộ.



Hình 2. Sơ đồ quy trình thực hiện

Việc thực hiện phân loại các lớp phủ SĐĐ được thực hiện tự động dựa trên việc lập trình bằng ngôn ngữ lập trình JavaScript. Các bước thực hiện được tiến hành theo sơ đồ quy trình ở hình 2. Cụ thể: Trích xuất dữ liệu ảnh vệ tinh của khu vực nghiên cứu dựa trên nền tảng dữ liệu điện toán đám mây GEE; tiền xử lý dữ liệu ảnh và chuẩn hóa đảm bảo có dữ liệu tốt nhất có thể; tiến hành việc

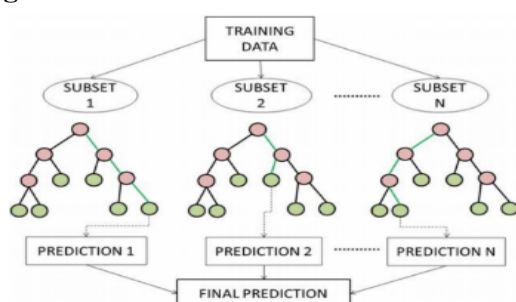
tạo mẫu, huấn luyện để cho chương trình học và ghi nhận các thông tin về các mẫu cho từng loại lớp phủ; phân loại các lớp phủ theo mẫu đã huấn luyện bằng thuật toán RF; thu nhận kết quả và đánh giá độ chính xác phân loại ảnh theo thuật toán RF; sau khi có kết quả phân loại ảnh, đảm bảo độ chính xác theo yêu cầu, tiến hành các bước phân tích, theo dõi biến động các lớp phủ SĐĐ trong giai đoạn 2010 - 2024 tại khu vực nghiên cứu.

2.3.2. Phương pháp sử dụng thuật toán RF

RF là một thuật toán học máy thuộc nhóm các phương pháp ensemble learning, được phát triển bởi Breiman (2001) [10]. Thuật toán này hoạt động dựa trên việc xây dựng một tập hợp nhiều cây quyết định (decision trees) và kết hợp kết quả của chúng thông qua cơ chế bỏ phiếu đa số (majority voting) nhằm nâng cao độ chính xác và giảm thiểu hiện tượng quá khớp (overfitting). Trong bài toán phân loại SDD, mỗi cây trong rừng được huấn luyện từ một tập con ngẫu nhiên của dữ liệu huấn luyện (thông qua phương pháp bootstrap sampling), đồng thời tại mỗi nút phân chia, RF chỉ chọn một tập con ngẫu nhiên các thuộc tính để quyết định chia nhánh. Cách tiếp cận “đóng bao” (bagging) này giúp giảm phương sai của bộ phân loại và hạn chế hiện tượng quá khớp. Sau đó, kết quả dự đoán cuối cùng cho một điểm dữ liệu thu được bằng cách kết hợp nhiều cây quyết định [12]. Cụ thể, quyết định phân loại được viết dưới dạng:

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

Trong đó: $H(x)$ là mô hình kết hợp, h_i là mô hình phân loại của cây quyết định đơn lẻ, Y là biến đầu ra (hoặc biến mục tiêu), $I(.)$ là hàm chỉ báo. Công thức cho thấy rằng, RF sử dụng đa số các quyết định biểu quyết để xác định phân loại cuối cùng.



Hình 3. Mô hình phân loại theo thuật toán RF

Nguồn: Đỗ Lan Phương và cs (2024) [9]

Thuật toán RF được đánh giá cao nhờ độ chính xác và tính ổn định; các tham số quan trọng thường được chọn theo kinh nghiệm gồm: Số cây (ntree) là 150, vì đường cong OA/Kappa bão hòa quanh 120 - 160 cây; tầng > 150 chỉ cải thiện rất nhỏ nhưng tốn thời gian, trong khi 150 cho kết quả ổn định qua nhiều lần chia train/test, số đặc trưng

ngẫu nhiên tại mỗi nút (mtry) là 3 theo quy tắc $mtry \approx \sqrt{p}$, $p \approx 14$ biến đầu vào và đã kiểm chứng bằng tuning, giúp giảm tương quan giữa cây, giữ cân bằng bias - variance và đạt OA tương đương cấu hình lớn hơn. Các giá trị này nên được lựa chọn thông qua kiểm định chéo hoặc thử nghiệm thực nghiệm [13], [14]. Nhược điểm của RF là chi phí tính toán tăng theo số cây và độ sâu, song việc huấn luyện có thể song song hóa nên thời gian thực thi thường vẫn hiệu quả. Trong nghiên cứu này, đã hiện thực RF trên nền tảng điện toán đám mây GEE bằng JavaScript, để tiền xử lý, trích xuất đặc trưng và phân loại ảnh Landsat và Sentinel-2. Kết quả tạo ra các bản đồ lớp phủ gồm sáu lớp (đất trống, mặt nước, CLN, CHN, khu dân cư, giao thông) cho khu vực thành phố Tân An.

2.3.3. Phương pháp đánh giá độ chính xác phân loại ảnh

Trong nghiên cứu này, độ chính xác của kết quả phân loại lớp phủ SDD được xác định bằng ma trận nhầm lẫn (confusion matrix) - bảng đối chiếu giữa nhãn tham chiếu và nhãn dự đoán theo từng lớp, qua đó nhận diện các kiểu nhầm lẫn và mức độ phù hợp của mô hình [15]. Từ ma trận này, tính các chỉ số chính gồm độ chính xác tổng thể (Overall Accuracy, OA) và hệ số Kappa (Cohen's Kappa). Theo thường lệ, hệ số Kappa trong khoảng từ 0,4 - 0,6 phản ánh mức độ phù hợp trung bình, giá trị từ lớn hơn 0,6 - 0,8 cho thấy sự phù hợp tốt và hơn 0,8 - 1,0 thể hiện mức gần như hoàn hảo [14]. Dữ liệu mẫu được phân chia theo tỷ lệ 70% cho huấn luyện và 30% cho kiểm định, bảo đảm đánh giá ngoài mẫu một cách khách quan [16].

2.3.4. Phương pháp phân tích, theo dõi biến động các lớp phủ SDD

Kết quả các lớp phủ SDD từ việc phân loại ảnh vệ tinh tại khu vực nghiên cứu được sử dụng làm dữ liệu đầu vào để phân tích, theo dõi các biến động tại khu vực trong giai đoạn 2010 - 2024. Việc phân tích, theo dõi biến động lớp phủ SDD được thực hiện với phần mềm mã nguồn mở QGIS. Các lớp dữ liệu được tính toán tổng hợp và tiến hành chồng xếp không gian để tìm ra các thay đổi qua các giai đoạn từ năm 2010 đến năm 2024.

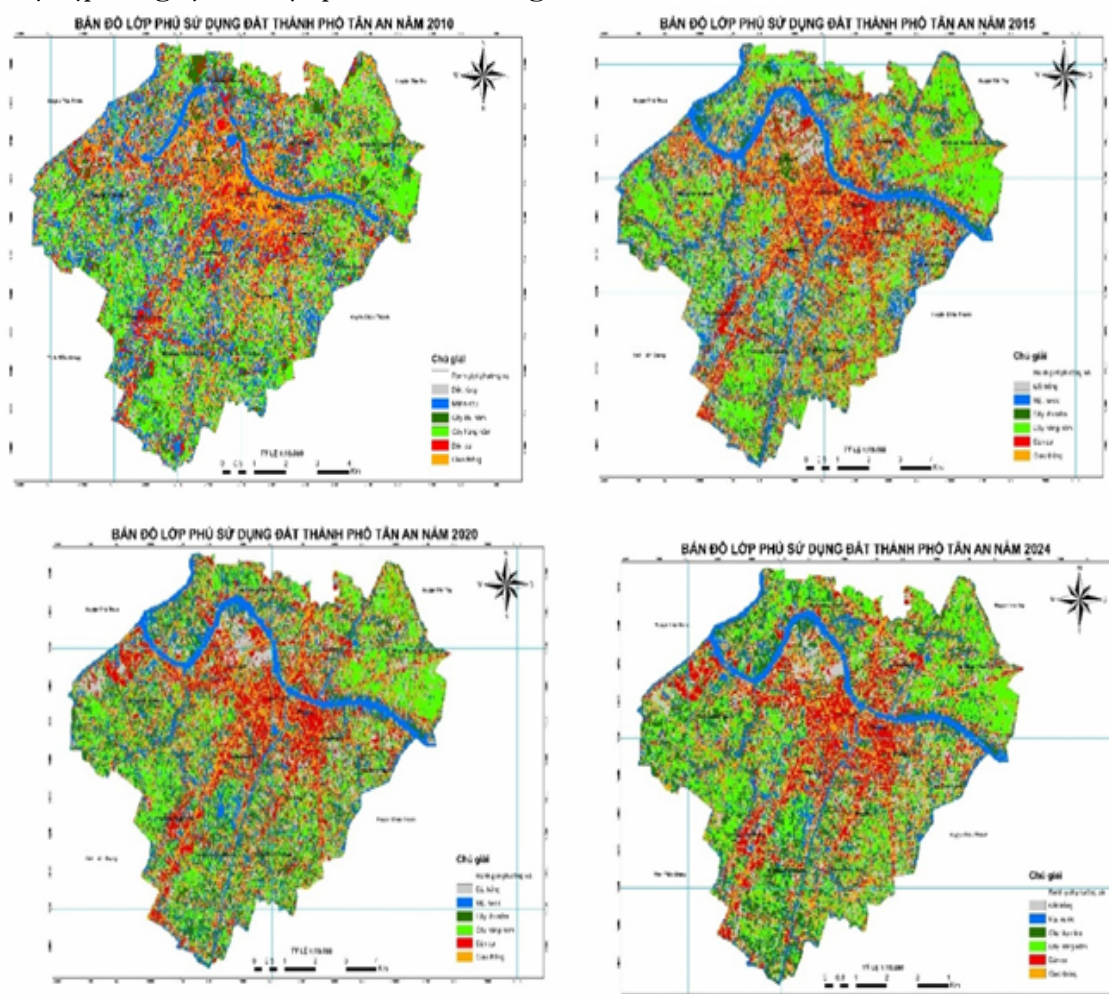
3. KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

3.1. Kết quả phân loại các lớp phủ SĐĐ giai đoạn 2010 - 2024 tại thành phố Tân An

Sau khi xử lý dữ liệu, phân loại ảnh, kết quả là sản phẩm ảnh sau khi phân loại theo thuật toán RF. Mỗi một ảnh sau phân loại bao gồm 6 loại lớp phủ SĐĐ bao gồm: Lớp phủ đất trống, lớp phủ mặt nước, lớp phủ CLN, lớp phủ CHN, lớp phủ dân cư, lớp phủ giao thông. Hình 4 thể hiện sản phẩm của lớp phủ SĐĐ sau phân loại của khu vực nghiên cứu theo thuật toán RF.

Kết quả phân loại lớp phủ SĐĐ giai đoạn 2010 - 2024 ở hình 4 cho thấy, xu hướng đô thị hóa mở rộng từ trung tâm thành phố Tân An ra các khu vực ven phía Nam và Đông Bắc. Năm 2010, đất nông nghiệp (CHN và CLN) vẫn chiếm tỷ lệ lớn, đặc biệt tập trung tại khu vực phía Nam và Đông

Bắc thành phố. Đất dân cư chủ yếu phân bố quanh khu vực trung tâm và các tuyến giao thông chính như Quốc lộ 1 và Quốc lộ 62. Tuy nhiên, đến năm 2024, diện tích đất phi nông nghiệp, đặc biệt là đất dân cư và đất giao thông đã mở rộng rõ rệt, bao phủ phần lớn trung tâm thành phố và lan tỏa mạnh về phía Tây Nam (phường 3, phường 4, phường Tân Khánh và phía Đông Bắc (phường Khánh Hậu, xã Lợi Bình Nhơn). Quá trình này cho thấy sự lan rộng không gian đô thị theo xu hướng “đa tâm”, phản ánh đặc điểm phát triển đô thị hoá điển hình ở các đô thị vệ tinh vùng thành phố Hồ Chí Minh. Nguyên nhân sự tăng giảm trên, ngoài việc do tình trạng sử dụng đất thực tế mỗi năm một khác, còn có thể do tại các thời điểm thu nhận ảnh, chất lượng ảnh cũng tạo nên những tính toán có thể tồn tại sai lệch trong phân loại.



Hình 4. Kết quả phân loại lớp phủ SĐĐ giai đoạn 2010 - 2024 theo thuật toán RF

Bảng 1. Độ chính xác phân loại ảnh

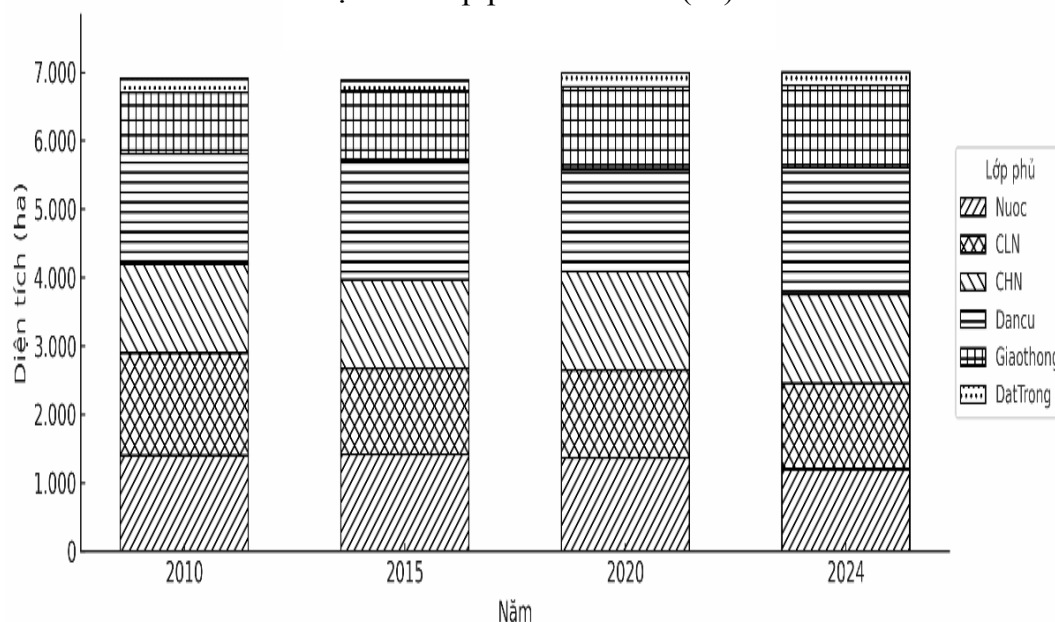
Năm	OA	Kappa
2010	91,3 %	0,89
2015	92,0%	0,90
2020	92,7%	0,91
2024	93,1%	0,92

Kết quả đánh giá độ chính xác phân loại ảnh ở bảng 1 cho thấy, các bản đồ lớp phủ SĐĐ cho 4 năm đều đạt độ chính xác cao, với OA lần lượt là 91,3%, 92,0%, 92,7% và 93,1%; đồng thời hệ số Kappa cũng đạt từ 0,89 - 0,92. Lỗi chủ yếu nằm ở ranh giới giữa CHN và CLN, một phần do tính chất cạnh tác hỗn hợp và chuyển đổi thời vụ. Kết quả 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 SĐĐ trong giai đoạn 2010 - 2024 tại khu vực thành phố Tân An, tỉnh Long An.

3.2. Kết quả thống kê diện tích lớp phủ SĐĐ qua các năm

Kết quả thống kê diện tích lớp phủ SĐĐ qua các năm tại hình 5 cho thấy: Đất dân cư tăng từ 1.560 ha (năm 2010) lên 2.030 ha (năm 2024), tương ứng tăng 470 ha ($\approx 30\%$). Sự gia tăng này cho thấy quá trình đô thị hóa đang diễn ra mạnh mẽ và ổn định theo thời gian. Đất giao thông tăng từ 870 ha lên 1.200 ha, phản ánh sự mở rộng và nâng cấp hạ tầng kỹ thuật phục vụ đô thị hóa. CHN giảm mạnh nhất, từ 1.410 ha xuống còn 960 ha, mất khoảng 450 ha ($\approx 32\%$), cho thấy đất nông nghiệp dễ bị chuyển đổi sang các mục đích phi nông nghiệp. Đất trống và mặt nước giảm nhẹ hoặc duy trì tương đối ổn định, trong khi CLN duy trì với mức giảm không đáng kể, chứng tỏ tính ổn định hơn của hình thức SĐĐ này.

Diện tích lớp phủ theo năm (ha)



Hình 5. Kết quả thống kê diện tích lớp phủ SĐĐ qua các năm

3.3. Kết quả chuyển đổi SĐĐ qua các giai đoạn

Sau khi chồng xếp các lớp hiện trạng, đã tiến hành tính toán và xác định được ma trận biến động SĐĐ của thành phố Tân An trong giai đoạn 2010 - 2024. Kết quả được thể hiện ở bảng 2.

Dữ liệu bảng ma trận chuyển đổi SĐĐ cho thấy, khu vực nghiên cứu đang có những biến động rất rõ ràng, đặc biệt là theo hướng đô thị hóa và phát triển hạ tầng. Bảng 2 cho thấy đất trống là loại bị chuyển đổi nhiều nhất, giảm tới 163,17 ha. Cụ thể, có tới 838,08 ha đất trống được chuyển sang đất CHN, 655,20 ha chuyển sang đất CLN, 693,18 ha sang đất giao thông, ngoài ra còn hơn 464 ha chuyển thành đất dân cư. Điều này cho thấy, đất trống đang được tận dụng tối đa để phục vụ sản xuất, xây dựng và mở rộng cơ sở hạ tầng. Tiếp theo là đất mặt nước, cũng giảm đáng kể với 172,35 ha bị mất đi. Nhiều diện tích mặt nước đã bị chuyển sang các mục đích khác, trong đó 696,42

ha chuyển thành đất CLN, 423,90 ha thành đất trống và gần 461,70 ha thành đất CHN. Việc này có thể liên quan đến san lấp ao hồ để phục vụ sản xuất hoặc phát triển đô thị, nhưng cũng tiềm ẩn rủi ro về môi trường nếu không được kiểm soát kỹ. Trong khi đó, đất CLN gần như giữ ổn định, chỉ tăng nhẹ 8,37 ha. Loại đất này vừa nhận thêm từ các nhóm khác như mặt nước và đất trống, vừa chuyển bớt sang đất CHN (851,31 ha) và đất dân cư (301,14 ha). Điều này cho thấy, có sự linh hoạt trong canh tác, có thể do người dân luân canh hoặc chuyển đổi cơ cấu cây trồng để phù hợp với điều kiện và thị trường. Đất CHN tăng đáng kể, thêm 104,40 ha. Đây là loại đất nhận được nhiều chuyển đổi nhất: Từ đất trống (838,08 ha), đất CLN (851,31 ha) và cả đất mặt nước (461,70 ha). Việc mở rộng diện tích trồng CHN cho thấy, người dân đang đầu tư nhiều hơn cho sản xuất nông nghiệp ngắn hạn, linh hoạt và phù hợp với canh tác mùa vụ.

Bảng 2. Ma trận chuyển đổi SĐĐ tổng hợp giai đoạn 2010 - 2024

Đơn vị tính: ha.

Từ/đến	Đất trống	Mặt nước	CLN	CHN	Dân cư	Giao thông	Tăng (inflow)	Giảm (outflow)	Net (mức tăng/giảm ròng)
Đất trống	1.065,69	476,64	655,20	838,08	464,67	693,18	2.964,60	3.127,77	-163,17
Mặt nước	423,90	1.611,81	696,42	461,70	233,01	544,05	2.186,73	2.359,08	-172,35
CLN	626,22	573,93	1.052,91	851,31	301,14	607,68	2.968,65	2.960,28	+8,37
CHN	715,41	460,08	825,03	2.299,32	162,81	629,91	2.897,64	2.793,24	+104,40
Dân cư	498,60	172,53	243,81	187,83	1.304,19	710,01	2.032,65	1.812,78	+219,87
Giao thông	700,47	503,55	548,19	558,72	871,02	1.528,29	3.184,83	3.181,95	+2,88

Đất dân cư tăng mạnh nhất trong tất cả các loại, tăng tới 219,87 ha. Gần 1.304 ha đất CHN đã được chuyển thành đất ở, cùng với 464,67 ha từ đất trống và 301,14 ha từ đất CLN. Sự gia tăng mạnh mẽ này phản ánh quá trình đô thị hóa đang diễn ra nhanh chóng, kéo theo nhu cầu lớn về nhà ở, khu

dân cư và các công trình xây dựng. Cuối cùng là đất giao thông, tuy chỉ tăng nhẹ 2,88 ha, nhưng cũng nhận thêm nhiều diện tích từ các loại đất khác như: 710,01 ha từ đất CHN, 693,18 ha từ đất trống và 607,68 ha từ đất CLN. Đây là dấu hiệu cho thấy cơ sở hạ tầng, đặc biệt là đường sá cũng đang

được mở rộng để phục vụ quá trình đô thị hóa. Nhìn chung, bảng 2 cho thấy bức tranh quen thuộc tại các đô thị đang phát triển: Đất trống, đất nông nghiệp và mặt nước dần dần bị thu hẹp để nhường chỗ cho nhà ở và giao thông. Đây là xu thế tất yếu, nhưng cũng đặt ra thách thức lớn về quy hoạch SDD sao cho hợp lý - vừa đáp ứng được nhu cầu phát triển, vừa đảm bảo cân bằng sinh thái và môi trường sống lâu dài cho người dân.

3.4. Thảo luận

Nghiên cứu đã ứng dụng hiệu quả thuật toán RF trên nền tảng GEE kết hợp với ảnh Landsat, Sentinel-2, DEM và các chỉ số phổ (NDVI, NDWI, NDBI, SAVI, MNDWI, AWEI...) để phân loại lớp phủ SDD và đánh giá biến động SDD trong giai đoạn 2010 - 2024. Độ chính xác phân loại cao ($OA > 91\%$, $Kappa > 0,89$), cùng với hệ thống bản đồ và ma trận chuyển đổi được biên tập chi tiết đã cho thấy, xu hướng chuyển đổi rõ rệt từ đất nông nghiệp sang đất xây dựng phản ánh đúng thực trạng đô thị hóa tại địa phương.

So với kết quả nghiên cứu của Phạm Anh Tuấn (2024) [7] tại huyện Gia Lâm, thành phố Hà Nội, cũng ứng dụng trí tuệ nhân tạo để phân loại và dự báo biến động SDD, có thể thấy sự tương đồng trong việc sử dụng ảnh Sentinel-2, nền tảng GEE và thuật toán RF. Tuy nhiên, nghiên cứu của Phạm Anh Tuấn (2024) [7] mở rộng thêm khả năng dự báo xu hướng đến năm 2028 thông qua mô hình MOLUSCE trên QGIS, đồng thời chú trọng đề xuất các giải pháp quản lý đất đai. Khác với hướng dự báo, nghiên cứu tại thành phố Tân An ưu tiên phân tích chuỗi thời gian dài, từ năm 2010 - 2024 (14 năm), phản ánh biến động thực tế bằng dữ liệu phong phú và quy trình xử lý trọn vẹn (tiền xử lý $\rightarrow$ hợp ảnh $\rightarrow$ RF $\rightarrow$ kiểm định $\rightarrow$ bản đồ biến động).

Tương tự, nghiên cứu của Đỗ Lan Phương và cs (2024) [9] tại quận Bắc Từ Liêm, thành phố Hà Nội cũng sử dụng thuật toán RF trên nền tảng

GEE để phân loại lớp phủ từ ảnh Sentinel-2 giai đoạn 2019 - 2023. Tuy nhiên, nghiên cứu này chỉ phân tích 5 lớp phủ cơ bản, không tích hợp ảnh Landsat hay dữ liệu địa hình và chỉ lựa chọn ảnh trong mùa khô. Do đó, khả năng phản ánh chi tiết biến động không gian - thời gian còn hạn chế so với kết quả nghiên cứu tại thành phố Tân An. Dù vậy, kết quả vẫn cho thấy xu hướng đô thị hóa với diện tích đất dân cư tăng 5,27% và thực vật dày giảm 5,32%, tương đồng với xu thế biến động đất đai tại thành phố Tân An.

Từ việc so sánh với các kết quả nghiên cứu khác, có thể khẳng định rằng, nghiên cứu tại thành phố Tân An có độ bao phủ tốt hơn về thời gian, độ sâu dữ liệu và tính toàn diện trong phương pháp xử lý, từ đó tăng cường độ tin cậy và khả năng ứng dụng trong quy hoạch SDD và giám sát đô thị hóa ở quy mô địa phương.

4. KẾT LUẬN

Phương pháp ứng dụng thuật toán RF đã cho thấy hiệu quả vượt trội trong phân loại lớp phủ SDD tại thành phố Tân An, với độ chính xác cao và khả năng xử lý nhanh hơn so với các phương pháp truyền thống. Kết quả cho thấy, xu hướng mở rộng mạnh mẽ đất phi nông nghiệp (dân cư, giao thông) trong giai đoạn 2010 - 2024, chủ yếu tại các phường ven đô như: Phường 5, phường Tân Khánh, xã Bình Tâm và xã Hướng Thọ Phú. Diện tích đất nông nghiệp (CHN, CLN) suy giảm rõ rệt, đặc biệt dọc theo các trục giao thông và ven sông Vàm Cỏ Tây. Xu thế này phản ánh quá trình đô thị hóa nhanh nhưng cũng đặt ra thách thức trong bảo vệ quỹ đất nông nghiệp và phát triển bền vững. Mặc dù kết quả đạt được có độ tin cậy cao, các nghiên cứu tiếp theo cần tích hợp khảo sát thực địa và đối chiếu tài liệu để tăng cường tính xác thực và ứng dụng trong quản lý, quy hoạch SDD.

LỜI CẢM ƠN

Bài báo được thực hiện dưới sự hỗ trợ của đề tài “Ứng dụng Machine Learning theo dõi biến

động sử dụng đất giai đoạn 2010 - 2024 tại thành phố Tân An, tỉnh Long An". Mã số: CT.2025.12 do Trường Đại học Tài nguyên và Môi trường thành phố Hồ Chí Minh chủ trì.

TÀI LIỆU THAM KHẢO

1. Bộ Tài nguyên và Môi trường (2020). *Báo cáo quy hoạch sử dụng đất quốc gia giai đoạn 2021 - 2030, tầm nhìn đến năm 2050*.
2. Nguyễn Văn Cường, Trần Thị Mai Hương (2019). Ứng dụng viễn thám trong kiểm kê đất đai. *Tạp chí Khoa học Đất*, (58), 45 – 52.
3. Swapan Talukdar, Pankaj Singha, Susanta Mahato, Swades Pal, Yuei-An Liou and Atiqur Rahman (2020). Land-use land-cover classification by machine learning classifiers for satellite observations - A review. *Remote Sensing*, 12(7), 1135. <https://doi.org/10.3390/rs12071135>.
4. Brian W Szuster, Qi Chen and Michael Borger (2011). A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. *Applied Geography*, 31(2): 525 - 532.
5. Abbas Taati, Fereydoon Sarmadian, Amin Mousavi, Chamran Taghati Hossien Pour and Amir Hossein Esmaili Shahir (2015). Land use classification using support vector machine and maximum likelihood algorithms by Landsat 5 TM images. *Walailak Journal of Science*, 12(8): 681 - 687.
6. Subhra Swetanisha, Amiya Ranjan Panda and Dayal Kumar Behera (2022). Land use/land cover classification using machine learning models. *International Journal of Electrical Computer Engineering (IJECE)*, 12(2): 2040 - 2046.
7. Phạm Anh Tuấn (2024). Theo dõi biến động sử dụng đất huyện Gia Lâm, thành phố Hà Nội giai đoạn 2018 – 2023. *Tạp chí Khoa học Tài nguyên và Môi trường, Trường Đại học Tài nguyên và Môi trường Hà Nội*, 52, 89 - 97.
8. Nguyễn Thị Huyền Trang, Đặng Thanh Tùng, Phạm Thị Sao Mai, Hà Thị Thu Hồng (2023). Nghiên cứu phân loại lớp phủ sử dụng đất từ ảnh vệ tinh Sentinel-2 dựa trên thuật toán SVM. *Science Journal of Natural Resources*, 45, 74 - 82.
9. Đỗ Lan Phương, Hoàng Thị Thủy, Đỗ Như Hiệp (2024). Nghiên cứu ứng dụng trí tuệ nhân tạo theo dõi biến động lớp phủ sử dụng đất khu vực quận Bắc Từ Liêm, Hà Nội giai đoạn 2019 - 2023. *Tạp chí Trắc địa - Bản đồ*, 01, trang 23 - 30.
10. Breiman, L (2001). Random forests. *Machine Learning*, 45(1), 5 - 32. <https://doi.org/10.1023/A:1010933404324>.
11. Thanh Noi Phan, Verena Kuch and Lukas W. Lehnert (2020). Land cover classification using Google Earth Engine and Random Forest classifier-the role of image composition. *Remote Sensing*, 12(15), 2411. <https://doi.org/10.3390/rs12152411>.
12. Frank Thonfeld, Stefanie Steinbach, Javier Muro, Fridah Kiriimi (2020), Long-term land use/land cover change assessment of the kilombo catchment in Tanzania using Random Forest classification and robust change vector analysis. *Remote Sensing*, 12(7), 1057. <https://doi.org/10.3390/rs12071057>.
13. J Ross Quinlan (1996). Bagging, boosting and C4. 5. *Aai/laai*, 1, 725 - 730.
14. Giang Thị Phương Thảo, Phạm Thị Thu Hương, Phạm Việt Hòa, Nguyễn An Bình (2021). Đánh giá độ chính xác trong phân loại lớp phủ dựa trên thuật toán học máy và dữ liệu viễn thám thông qua Google Earth Engine: Áp dụng tại tỉnh Đắk Lắk. *Tạp chí Khoa học Tài nguyên và Môi trường, Trường Đại học Tài nguyên và Môi trường Hà Nội*, 46, 55 - 65.
15. Am Hay (1988). The derivation of global estimates from a confusion matrix. *International Journal of Remote Sensing*, 9(8): 1395 - 1398.
16. Ahmed M Abdel-Zaher and Ayman M Eldeib (2016). Breast cancer classification using deep belief networks. *Expert Systems with Applications*, 46, 139 - 144.

**APPLYING MACHINE LEARNING TO TRACK LAND USE CHANGES IN TAN AN CITY,
LONG AN PROVINCE IN THE PERIOD OF 2010 - 2024**

Nguyen Kim Hoa^{1,2}, Nguyen Thi Hai¹, Nguyen Quang Thi³, Nguyen Huu Ngu¹

¹ *University of Agriculture and Forestry, Hue University*

² *Ho Chi Minh University of Natural Resources and Environment*

³ *Thai Nguyen University of Agriculture and Forestry*

Abstract

In this study, land use cover changes in Tan An city, Long An province, between 2010 and 2024 are classified and tracked using the Random Forest machine learning algorithm. The input data consists of Sentinel-2 (2024), Landsat satellite images (2010, 2015, 2020), DEM terrain data and indices like NDVI, NDWI, NDBI, BSI, LSWI, MNDWI and AWEL. The training models are built including 6 main land use classes (vacant land, water surface, perennial trees, annual trees, residential areas and traffic). The classification results are highly accurate with a total accuracy (OA) of over 91% and a Kappa coefficient of over 0.89 for each year. The land use dynamics indicate a pronounced urbanization trend - residential and transportation areas expanded markedly, while agricultural land declined. In the period of 2010 - 2024, residential land will increase by more than 1.600 hectares, mainly converted from land for perennial trees and annual trees. Meanwhile, the water surface fluctuates slightly, while the vacant land decreases sharply over the years. Phased overlay maps, spatial variation maps and land-use conversion matrices built and edited on GIS software provided a visual view of the land-use transition process. The results show a clear trend of conversion from agricultural land to construction land, reflecting the strong urbanization rate of Tan An city over the past decade. The study contributes to providing a scientific basis for the assessment and orientation of land use planning in the locality.

Keywords: *Land use change, Google Earth Engine, machine learning, Random Forest, Tan An.*

Ngày nhận bài: 20/6/2025

Ngày chuyển phản biện: 26/6/2025

Ngày thông qua phản biện: 18/7/2025

Ngày duyệt đăng: 31/7/2025