

TRƯỜNG ĐẠI HỌC MỎ - ĐỊA CHẤT
KHOA TRẮC ĐỊA - BẢN ĐỒ VÀ QUẢN LÝ ĐẤT ĐAI
BỘ MÔN ĐO ẢNH VÀ VIỄN THĂM



HỘI NGHỊ KHOA HỌC QUỐC GIA
CÔNG NGHỆ ĐỊA KHÔNG GIAN VÌ SỰ PHÁT TRIỂN BỀN VỮNG
(NCGSD 2024)



NHÀ XUẤT BẢN KHOA HỌC TỰ NHIÊN VÀ CÔNG NGHỆ

MỤC LỤC

Địa không gian thông minh trong hành trình thực hiện các mục tiêu phát triển bền vững	1
<i>Võ Chí Mỹ*</i>	
Giao dịch bất động sản - Thực trạng và biện pháp hạn chế rủi ro về mặt pháp lý	6
<i>Nguyễn Thị Dung*</i>	
Ảnh hưởng của mật độ điểm đo trực tiếp đến độ chính xác nội suy dị thường trọng lực bước sóng dài khi tính độ sâu từ dị thường trọng lực	12
<i>Nguyễn Văn Sáng*</i>	
Kết hợp đa chỉ số phổ trong chiết tách đất xây dựng ở khu vực Hà Nội từ ảnh vệ tinh Sentinel-2	18
<i>Nguyễn Văn Trung*, Lê Thị Thu Hà, Nguyễn Thành Len, Lê Thị Minh Phương</i>	
Phân tích tự tương quan không gian trong nghiên cứu kiểu mẫu phân bố bệnh sốt xuất huyết Dengue: Thử nghiệm tại Thành phố Hồ Chí Minh	28
<i>Vũ Danh Tuyên*, Vương Thị Hòe, Phạm Thị Thu Hương, Nguyễn Tiến Thành</i>	
Sử dụng dữ liệu vệ tinh Sentinel-2 thành lập bản đồ sinh vật đáy khu vực quần đảo Côn Đảo, tỉnh Bà Rịa - Vũng Tàu, Việt Nam	34
<i>Nguyễn Văn Trung*, Lê Thị Thu Hà, Đoàn Thị Nam Phương, Phạm Thị Hương, Nguyễn Quang Minh, Phạm Vọng Thành</i>	
Nghiên cứu giải pháp tự động chiết xuất đối tượng đường giao thông trên ảnh vệ tinh Sentinel- 2 MSI	44
<i>Lê Văn Phú, Trịnh Lê Hùng*, Tô Thị Phương</i>	
Phát hiện cháy rừng đang xảy ra sử dụng dữ liệu vệ tinh Sentinel-3 SLSTR khu vực xã Khánh Bình Tây, huyện Trần Văn Thời, tỉnh Cà Mau	51
<i>Đoàn Thị Nam Phương*, Nguyễn Văn Trung, Lê Thị Thu Hà, Nguyễn Quang Minh</i>	
Công nghệ SBAS xử lý dữ liệu lớn cho xác định biến dạng bề mặt địa hình vùng Thành phố Hồ Chí Minh	59
<i>Nguyễn Minh Hải*</i>	
Nghiên cứu phân bố không gian gió vùng biển vịnh Bắc Bộ giai đoạn 1981 - 2020 từ dữ liệu viễn thám	65
<i>Nguyễn Ngọc Tuấn, Nguyễn Hoàng Minh, Đỗ Thị Phương Thảo*</i>	
Nghiên cứu ứng dụng tư liệu ảnh Sentinel-5P trong giám sát chất lượng không khí tại các đô thị lớn	72
<i>Lê Minh Hằng, Đỗ Thị Hoài*</i>	
Thành lập mô hình lún để chẩn đoán Nhà máy lọc dầu Dung Quất	82
<i>Nguyễn Thị Kim Thanh*, Nguyễn Hà</i>	
Nghiên cứu ứng dụng phương pháp phân loại hướng đối tượng và mô hình học máy trong xây dựng bản đồ lớp phủ từ ảnh viễn thám quang học	90
<i>Khúc Thành Đông*, Trần Đình Trọng, Trần Văn Anh, Đỗ Quang Vinh, Hà Trung Khiên, Đào Duy Toàn, Trương Xuân Quang</i>	
Xác định lún đất khu vực Cà Mau bằng chuỗi ảnh radar trên nền tảng điện toán đám mây Google Colab	97
<i>Hà Trung Khiên, Trần Văn Anh*, Khúc Thành Đông, Nguyễn Đình Huy</i>	
Ứng dụng công nghệ GIS và viễn thám hỗ trợ phân tích, quản lý tài nguyên rừng trong điều kiện biến đổi khí hậu ở tỉnh Đắk Lắk	106
<i>Nguyễn Sách Thành*, Trịnh Lê Hùng, Lê Văn Phú, Nguyễn Văn Trung</i>	

Ứng dụng công nghệ địa tin học trong đánh giá hạn hán ở Quảng Nam	114
<i>Phạm Thị Thanh Hoà*, Trần Hồng Hạnh, Lưu Thị Diệu Chinh</i>	
Nghiên cứu ứng dụng mô hình LSTM (Long Short-Term Memory) trong dự báo độ mặn từ chuỗi dữ liệu đa biến tại trạm Đại Ngãi, Sóc Trăng	122
<i>Bùi Duy Quỳnh*, Hà Thị Hằng, Lưu Thị Diệu Chinh, Trần Xuân Trường</i>	
Kiểm định giả thuyết thống kê trong bình sai lưới quan trắc chuyển dịch ngang đập công trình thủy điện	131
<i>Trần Trung Anh*, Nguyễn Quang Hà</i>	
Phân tích, lọc và hiển thị dữ liệu đám mây điểm LiDAR với thư viện Pandas và Matplotlib của Python	138
<i>Nguyễn Thị Hữu Phương*, Phạm Thị Hải Vân, Đào Thị Hồng Thắm</i>	
Nghiên cứu xác định sự biến động độ sâu ven đảo xa bờ sử dụng dữ liệu viễn thám Sentinel-2 trên nền tảng Google Earth Engine	146
<i>Nguyễn Như Hùng*, Lê Văn Phú, Nguyễn Văn Sơn</i>	
Nghiên cứu một số thuật toán huấn luyện mạng nơron nhân tạo trong dự báo độ lún công trình thủy điện	154
<i>Phạm Quốc Khánh*</i>	
Công nghệ máy bay không người lái và thuật toán bán tự động trong việc xác định hệ thống khe nứt chính trên mô đá ốp lát	163
<i>Phạm Văn Việt*, Nguyễn Anh Tuấn, Trần Trung Anh, Phạm Văn Hoà</i>	
Phương trình LiDAR tính tọa độ điểm đo trong công nghệ bay quét laser hàng không	172
<i>Trần Trung Anh*, Trần Hồng Hạnh, Lưu Hải Âu, Quách Mạnh Tuấn</i>	
Lập bản đồ biến động hệ thực vật vùng đất ven biển khu vực Khánh Hòa bằng ảnh Landsat đa thời gian trên nền tảng Google Earth Engine	179
<i>Trần Thanh Hà*</i>	
Quan trắc trượt lở đất đá theo thời gian thực ở tầng khai thác mỏ lộ thiên dựa trên công nghệ GNSS/CORS	188
<i>Phạm Công Khải*, Nguyễn Văn Hải</i>	
Dự báo thay đổi lớp phủ bề mặt bằng thuật toán học máy	199
<i>Trần Thị Hoà*, Trần Thị Ngọc, Trần Đình Trí</i>	
Xây dựng cơ sở dữ liệu GIS phục vụ công tác tìm kiếm và cứu nạn khu vực vịnh Bắc Bộ	207
<i>Nguyễn Quang Minh*, Nguyễn Văn Trung, Bùi Tiến Diệu</i>	
Phân tích lượng mưa tích lũy và nguy cơ trượt lở đất: Nghiên cứu và mô phỏng số tại Km10+950, Quốc lộ 34, tỉnh Hà Giang	221
<i>Đỗ Văn Đăng*, Nguyễn Châu Lân, Hà Thị Hằng, Bùi Duy Quỳnh, Dương Công Hiếu, Lưu Thị Diệu Chinh</i>	
Nghiên cứu một số vấn đề chuyển đổi tọa độ từ WGS-84 về hệ tọa độ trắc địa sử dụng tại Việt Nam	228
<i>Hoàng Ngọc Hà*</i>	

Dự báo thay đổi lớp phủ bề mặt bằng thuật toán học máy

Trần Thị Hòa^{1*}, Trần Thị Ngọc¹, Trần Đình Trí²

¹Trường Đại học Mở - Địa chất

²Hội Trắc địa - Bản đồ - Viễn thám Việt Nam

TÓM TẮT

Lớp phủ mặt đất là bề mặt hình thái vật lý của lớp vỏ Trái đất, bao gồm thực vật, nước, bề mặt tự nhiên và những đối tượng nhân tạo. Những hoạt động diễn ra trên bề mặt Trái đất từ các hoạt động tự nhiên cho đến hoạt động phát triển kinh tế - xã hội, đô thị hóa đều dẫn đến sự thay đổi đáng kể của lớp phủ bề mặt. Tài nguyên đất là một trong những nguồn tài nguyên quan trọng và có ảnh hưởng sâu sắc đến môi trường nói chung và mục tiêu phát triển bền vững nói riêng, vì vậy, việc theo dõi và giám sát sự thay đổi của lớp phủ bề mặt, cũng như phân tích dự báo xu hướng thay đổi là một nhiệm vụ quan trọng. Các thuật toán học máy hiện nay đã có những đóng góp đáng kể trong việc tận dụng khả năng của máy tính và đảm bảo tính khách quan trong phân tích các vấn đề liên quan đến dự đoán. Bài báo xem xét khả năng ứng dụng của một số thuật toán học máy (Machine Learning) khác nhau trong dự đoán biến động lớp phủ bề mặt, cụ thể: cây quyết định (Decision Trees), rừng ngẫu nhiên (Random Forest) và vector máy hỗ trợ (SVM).

Từ khóa: Biến động lớp phủ, dự đoán, học máy, Random Forest, SVM.

1. Đặt vấn đề

Lớp phủ bề mặt là một thuật ngữ chung dùng để chỉ các vật liệu hoặc các yếu tố tự nhiên và nhân tạo bao phủ bề mặt Trái đất, bao gồm: thực vật (rừng, đồng cỏ, hay cây bụi); nước (sông suối, ao hồ hay biển); đất canh tác (đất trồng lúa, đất trồng màu hay đất trồng cây ăn quả - cây lâu năm); đất trống (cát, hoang mạc, đất không có cây cỏ); đất đô thị và công nghiệp (đường xá, khu vực xây dựng, khu vực có công trình hạ tầng) và khu vực bị băng tuyết bao phủ quanh năm (Giri, 2020). Những yếu tố này phân bố xen kẽ nhau tạo nên sự đa dạng và hình thái đặc trưng của mỗi vùng địa lý trên bề mặt Trái đất. Sự phân bố của các lớp phủ đóng vai trò quan trọng trong hệ sinh thái và môi trường.

Sự thay đổi của lớp phủ bề mặt có thể diễn ra dưới nhiều hình thức khác nhau, ví dụ sự thay mặt phân bố không gian hay diện tích hay sự biến đổi về cấu trúc - thành phần của một hay nhiều loại lớp phủ theo thời gian. Ví dụ như: cùng là lớp phủ thực vật nhưng từ rừng chuyển sang đồng cỏ hay sự thay đổi từ lớp phủ thực vật sang lớp phủ đất đô thị và công nghiệp. Nguyên nhân dẫn đến sự thay đổi lớp phủ bề mặt chia thành hai nhóm - yếu tố tự nhiên khách quan và yếu tố chủ quan - con người. Các yếu tố tự nhiên có thể gây ra sự thay đổi về lớp phủ một cách từ từ như sự thay đổi về khí hậu (nhiệt độ và lượng mưa) hoặc đột ngột như sự xuất hiện của các hiện tượng thiên tai như lũ lụt, hạn hán, cháy rừng hay động đất (Gregorio & Jansen, 2005) (Giri, 2020). Bên cạnh đó, các yếu tố về con người có thể kể đến như quá trình khai thác tài nguyên thiên nhiên, công nghiệp hóa, đô thị hóa, canh tác nông nghiệp hay các chính sách quản lý tài nguyên. Nhìn chung, những thay đổi về lớp phủ đều có những ảnh hưởng nhất định đến môi trường và xã hội, trong đó bao gồm giảm đa dạng sinh học, giảm chất lượng nước và không khí, biến đổi khí hậu và sự phát triển kinh tế vùng. Vì thế, công tác giám sát và quản lý lớp phủ bề mặt là công tác quan trọng đảm bảo sự phát triển bền vững và bảo vệ môi trường.

Những nghiên cứu về biến động lớp phủ thường chủ yếu liên quan đến sự thay đổi về mặt phân bố không gian (diện tích) ứng dụng công nghệ hiện đại như viễn thám, viễn thám đa thời gian, kết hợp với hệ thống tin địa lý (GIS) (Giri, 2020). Một số nghiên cứu nổi bật có thể kể đến như nghiên cứu về sự biến động rừng nhiệt đới amazon sử dụng dữ liệu vệ tinh được cung cấp bởi Cơ quan Vũ trụ hàng không quốc gia Hoa Kỳ (NASA) và Cơ quan Vũ trụ liên minh châu Âu (ESA) (Hansen et al., 2013). Nghiên cứu đã chỉ tốc độ phá rừng nhanh trong những năm gần đây thông qua sự mất mát phần lớn diện tích rừng và sự thay thế bởi đất trồng, đất nông nghiệp và chăn nuôi. Sự thay đổi này làm phá vỡ cấu trúc bền vững của hệ sinh thái rừng tự nhiên, gây ảnh hưởng tiêu cực đến đa dạng sinh học và thậm chí làm ảnh hưởng đến khí hậu toàn cầu. Hay nghiên cứu về ảnh hưởng của hiện tượng hoang mạc hóa ở châu Phi (Herrmann & Hutchinson, 2005) dẫn đến sự sụt giảm đáng kể của đất canh tác và lan rộng của sa mạc. Liên quan đến đô thị hóa, các rất nhiều nghiên cứu từ ảnh vệ tinh chỉ ra rằng tốc độ mở rộng đô thị nhanh ở những thành phố lớn đã làm gia tăng nhanh chóng các khu vực xây dựng, giảm thiểu diện tích đất nông nghiệp và các vùng xanh (Seto, Güneralp & Hutyrá, 2012) (Phạm & Nguyễn, 2023) (Nguyễn & Đặng, 2021) (Trần T. T., 2020). Những thay đổi trên có tác động lớn đến suy thoái môi trường, ảnh hưởng đến sinh kế, hiện tượng ngập lụt cục bộ và ô nhiễm không khí. Vì vậy, việc dự đoán sự thay đổi về lớp phủ bề mặt là

* Tác giả liên hệ

Email: hoatran106@gmail.com

điều cần thiết và có ý nghĩa quan trọng trong nhiều khía cạnh khác nhau, đặc biệt liên quan đến quản lý tài nguyên thiên nhiên và quy hoạch - phát triển bền vững.

Trong những năm gần đây, việc sử dụng trí tuệ nhân tạo (AI) mang lại nhiều vượt trội so với mô hình GIS truyền thống (Kumar & Sharma, 2023). Ví dụ như mô hình AI có khả năng xử lý và phân tích lượng lớn dữ liệu từ nhiều nguồn trong thời gian ngắn và phát hiện nhanh chóng thay đổi với thời gian thực. Bên cạnh đó, các thuật toán học sâu cho phép khả năng nhận diện và phân loại lớp phủ bề mặt với độ chính xác cao hơn (Zhang, Li, & Wang, 2023) (Johnson & Lee, 2023) (García & Rodríguez, 2022). Các mô hình AI có khả năng tự động cải thiện độ chính xác qua thời gian thông qua việc học từ dữ liệu thực tế. Ngoài ra, mô hình AI rất linh hoạt, có thể áp dụng được trong nhiều ngữ cảnh khác nhau, trong khi mô hình truyền thống thường bị giới hạn (Zhou & Liu, 2022). Ví dụ như không thể dùng mô hình dự báo sa mạc hóa truyền thống vào giám sát đô thị được, yếu tố can thiệp của con người cũng là một trong những nguyên nhân gây sai sót trong mô hình truyền thống mà AI có thể giảm thiểu được thông qua khả năng ứng biến và tự động hóa (Patel & Desai, 2021).

Nội dung của bài báo đi sâu vào tìm hiểu ưu nhược điểm và khả năng ứng dụng của một số thuật toán học sâu trong mô hình trí tuệ nhân tạo trong dự báo sự biến động của lớp phủ trong tương lai gần. Các thuật toán này khá thông dụng trong phân loại lớp phủ tự động như cây quyết định (Decision Trees), rừng ngẫu nhiên (Random Forest) hay máy hỗ trợ vector (SVM). Qua đó, đề xuất quy trình ứng dụng các thuật toán này trong dự báo biến động lớp phủ với việc xác định các tham số đầu vào vào nguồn tư liệu khai thác.

2. Cơ sở lý thuyết và phương pháp nghiên cứu

2.1. Lớp phủ bề mặt và sự thay đổi về lớp phủ bề mặt

Bề mặt tự nhiên của Trái đất là một tập hợp đa dạng các dạng địa hình và hệ sinh thái, bao gồm rừng, núi, sa mạc, đại dương và sông ngòi. Sự phân bố xen kẽ về mặt không gian và mối tác động tương quan với khí hậu và con người đã tạo ra những hình thái đặc trưng khác nhau về địa hình và sinh thái ở các khu vực địa lý khác nhau mà thể hiện bên ngoài của nó là lớp phủ bề mặt. Lớp phủ bề mặt Trái đất bao gồm hình thái của các thành phần tự nhiên và nhân tạo như đất, cây cối, đá, băng và các công trình xây dựng. Việc nghiên cứu lớp phủ bề mặt Trái đất là một phần quan trọng trong các lĩnh vực khoa học như địa lý, sinh thái học, nông nghiệp và môi trường.

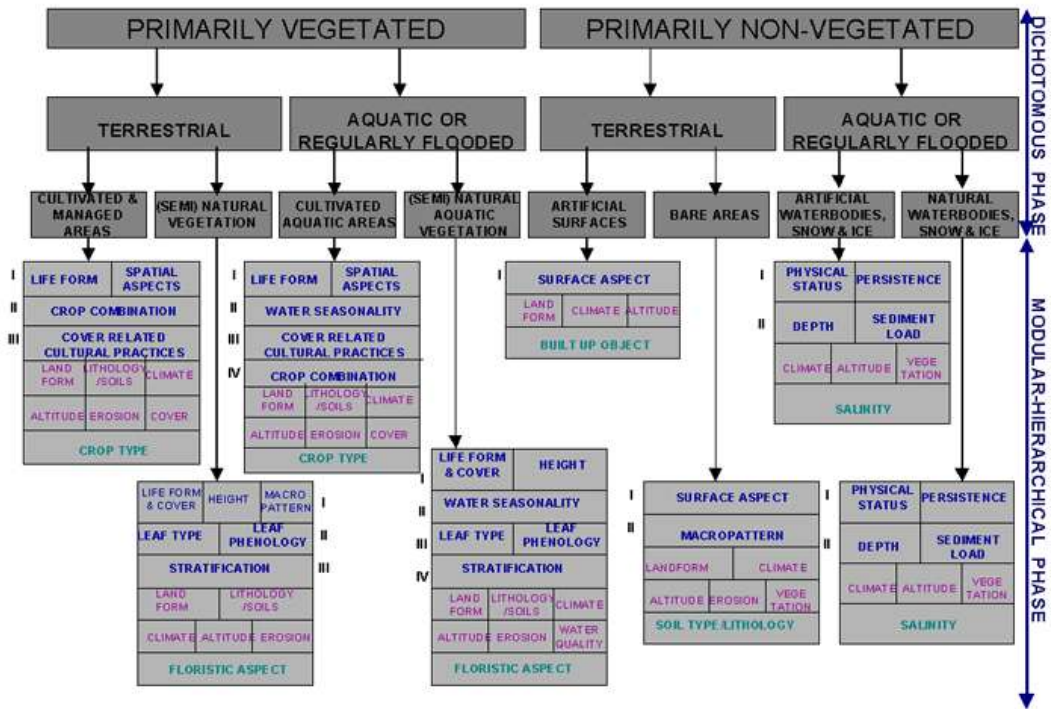
Mỗi một lĩnh vực nghiên cứu đều có cách thức phân loại lớp phủ bề mặt Trái đất với mức độ chi tiết khác nhau. Ví dụ như phân loại theo hệ sinh thái, lớp phủ được chia thành 05 nhóm: rừng, đồng cỏ, sa mạc, băng và nước, Phương pháp phân loại theo mục đích sử dụng, lớp phủ bề mặt được chia thành 04 nhóm lớn, bao gồm đất nông nghiệp, đô thị, rừng và đất không sử dụng. Còn đối với đặc điểm địa hình, lớp phủ chia thành 03 nhóm: địa hình cao, địa hình thấp và địa hình đặc biệt như hang động hay hẻm núi. Tổ chức Nông Lương thế giới (FAO) có cách phân chia theo ý nghĩa sử dụng của đất (Thackway & Atyeo, 2006); USGS - Cục khảo sát địa chất Hoa Kỳ có phương pháp riêng để phân loại lớp phủ bề mặt dựa trên đặc điểm đối tượng tự nhiên xuất hiện trên bề mặt lớp phủ. Tương tự, Cơ quan Vũ trụ châu Âu (ESA) cũng lựa chọn phương pháp phân chia lớp phủ của tổ chức liên hợp quốc dựa trên loại thực vật có trên bề mặt (the UN LCCS) (Gregorio & Jansen, 2005). Bên cạnh đó, để phù hợp với đặc điểm địa lý của từng khu vực cũng như mục đích sử dụng, mỗi quốc gia đều xây dựng hệ phân loại lớp phủ - lớp sử dụng đất khác nhau.

Về cơ bản, việc phân loại lớp phủ bề mặt nhằm mục đích để theo dõi, giám sát những tác động của tự nhiên và con người được thuận tiện và dễ dàng hơn. Qua đó, công tác bảo vệ, quản lý và quy hoạch sẽ đạt được.

Bảng 1 và Hình 1 thể hiện hệ thống phân loại mà FAO hiện đang sử dụng hiện nay trong phân loại lớp phủ bề mặt và lớp hiện trạng sử dụng đất đối với các tài liệu và bản đồ công bố bởi tổ chức này. Đây cũng là hệ thống phân loại khá phổ biến được áp dụng trong các nghiên cứu về biến động hay hiện trạng lớp phủ hiện nay. Giai đoạn phân đôi (dichotomous phase) xác định ba cấp phân loại dựa theo thực vật (Bảng 1), còn giai đoạn phân cấp (modular - hierarchical phase) (Gregorio & Jansen, 2005) thì mỗi lớp ở cấp 3 của giai đoạn 1 sẽ chia thành nhiều loại hơn nữa phụ thuộc vào tính chất của môi trường (như khí hậu, loại đất, khả năng xói mòn) và tính chất kỹ thuật, ví dụ như trong đất nông nghiệp có đất trồng lúa hay đất trồng màu. Đặc điểm này cho phép những nghiên cứu về sao không bị giới hạn về số lượng lớp phủ mà tùy thuộc vào khu vực, các nhà nghiên cứu có thể bổ sung thêm định nghĩa của lớp phủ bề mặt mới.

Bảng 1. Hệ phân loại phân đôi UN LCCS (Dichotomous Phase) (Thackway & Atyeo, 2006)

Cấp 1	Cấp 2	Cấp 3
Đất có thực vật	Thực vật trên cạn	Thực vật trên cạn có quản lý
		Thực vật tự nhiên hoặc bán tự nhiên trên cạn
	Thực vật ngập nước hoặc thường xuyên ngập nước	Thực vật vùng trồng ngập nước Thực vật vùng tự nhiên và bán tự nhiên
Đất gần như không có thực vật	Trên cạn	Bề mặt nhân tạo Đất trống
		Hồ nhân tạo, băng và tuyết
	Vùng ngập nước hoặc thường xuyên ngập nước	Vùng nước tự nhiên, băng và tuyết phủ theo mùa



Hình 1. Hệ thống phân loại theo giai đoạn: giai đoạn phân đôi và giai đoạn phân cấp (Gregorio & Jansen, 2005)

Trong nghiên cứu này, nhóm tác giả chủ yếu đi sâu vào mô hình phân loại của ESA, có kế thừa phương pháp phân loại của UN LCCS. Cụ thể gồm 11 lớp, trình bày ở Bảng 2.

Bảng 2. Hệ thống phân loại của ESA (ESA)

Giá trị	Màu sắc	Tên tiếng Anh	Tên tiếng Việt
10	#006400	Tree cover	Thực phủ
20	#ffbb22	Shrubland	Cây bụi
30	#ffff4c	Grassland	Đồng cỏ
40	#f096ff	Cropland	Đất nông nghiệp
50	#fa0000	Built-up	Đất xây dựng
60	#b4b4b4	Bare/sparse vegetation	Đất trống/ít thực vật
70	#f0f0f0	Snow and ice	Tuyết và băng
80	#0064c8	Permanent water bodies	Nước mặt
90	#0096a0	Herbaceous wetland	Thực vật thân thảo ngập nước
95	#00cf75	Mangroves	Rừng ngập mặn
100	#fae6a0	Moss and lichen	Đất ẩm có rêu và địa y

2.2. Một số thuật toán học máy trong dự đoán thay đổi lớp phủ bề mặt

Học máy và học sâu là hai lĩnh vực con của trí tuệ nhân tạo AI. Các thuật toán học sâu sử dụng các mạng nơron nhân tạo để biểu diễn dữ liệu phức tạp và trừu tượng. Những mạng nơron này thường nhiều tầng và lớp tự động tìm ra các đặc trưng dữ liệu và hạn chế xử lý dữ liệu thủ công. Tuy nhiên, đối với học sâu, dữ liệu cần lớn về cả dung lượng và thời gian để đạt hiệu suất tốt, đồng thời yêu cầu tài nguyên tính toán mạnh mẽ. Với lĩnh vực học máy, các thuật toán thường nhắm đến việc máy tính có thể tự học từ các đặc trưng lựa chọn thủ công mà không cần lập trình rõ ràng từng nhiệm vụ cụ thể. Các thuật toán học máy thường đơn giản và ít tham số hơn nhưng cần kiến thức chuyên gia trong lựa chọn đặc trưng. Chính vì vậy, học máy phù hợp với yêu cầu ít dữ liệu huấn luyện, không cần máy tính cấu hình cao, dễ dàng triển khai và hiệu quả trong các ứng dụng với dữ liệu thực tế. Cụ thể đối với dự đoán thay đổi lớp phủ bề mặt, dữ liệu đầu vào là ảnh vệ tinh đa thời gian nhưng không liên tục, cần có đánh giá của các chuyên gia trong trích xuất đặc trưng từng lớp phủ. Vì vậy, sử dụng học máy là một lựa chọn hợp lý. Dưới đây, các tác giả trình bày ba thuật toán học máy được ứng dụng phổ biến nhất trong phân loại lớp phủ bề mặt, về đề xuất quy trình ứng dụng với đặc điểm cụ thể của khu vực nghiên cứu.

2.2.1. Thuật toán cây quyết định (Decision Tree)

Thuật toán cây quyết định (DT) là một thuật toán học máy mạnh mẽ được sử dụng hiệu quả trong phân loại và dự đoán thay đổi lớp phủ từ dữ liệu ảnh vệ tinh (Pal & Mather, 2003). DT hoạt động bằng cách chia dữ liệu thành các nhánh dựa trên các tiêu chí quyết định, giúp phân loại và dự đoán các loại lớp phủ bề mặt khác nhau. Trong đó mỗi nút trong cây quyết định đại diện cho một thuộc tính của dữ liệu và mỗi nhánh là kết quả của một thuộc tính. Các quyết định này được tiến hành thực hiện cho đến khi đạt được sự thống nhất một dự đoán cụ thể ở lá cây (Breiman, Friedman, Stone & Olshen, 1984).

Công thức của thuật toán này được mô tả như sau (Quinlan, 1986):

- Entropy (S): đo lường mức độ hỗn loạn của tập dữ liệu:

$$H(S) = -\sum_{i=1}^c p_i \log_2 p_i \quad (1)$$

trong đó: p_i là tỷ lệ phần tử thuộc lớp thứ i trong tập dữ liệu S.

- Information Gain (IG): đo lường mức độ giảm hỗn loạn sau khi chia tách tập dữ liệu theo một thuộc tính A.

$$IG(S, A) = H(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} H(S_v) \quad (2)$$

trong đó: S_v là tập con của A với thuộc tính A có giá trị v.

Một trong những ưu điểm nổi bật của DT là tính diễn giải và trực quan giúp cho các nhà nghiên cứu phân định các yếu tố quyết định đến sự thay đổi của lớp phủ bề mặt thông qua xác định yếu tố hồi quy để ra quyết định phân loại.

2.2.2. Thuật toán rừng ngẫu nhiên (Random Forest)

Rừng ngẫu nhiên (RF) là một tập hợp của nhiều cây quyết định được huấn luyện độc lập trên các mẫu dữ liệu khác nhau. Kết quả cuối cùng được quyết định bằng cách lấy trung bình hoặc phiếu đa số của các cây quyết định. Điều này giúp giảm thiểu hiện tượng quá khớp dữ liệu (overfitting) so với việc sử dụng một cây đơn lẻ (Breiman, 2001).

Công thức của thuật toán này được mô tả như sau (Breiman, 2001):

- Bootstrap sampling: mỗi cây trong rừng được huấn luyện trên một mẫu dữ liệu bootstrap (chọn mẫu mới thay thế từ tập dữ liệu huấn luyện).

- Random feature selection: tại mỗi nút chia tách, một tập con ngẫu nhiên của các thuộc tính được chọn để tìm kiếm thuộc tính tốt nhất để chia tách.

- Ensemble Prediction: dự đoán tập hợp:

Đối với bài toán phân loại, kết quả được xác định bằng cách lấy đa phiếu từ các cây:

$$\hat{y} = \text{mode}(\{T_i(x) \mid i = 1, 2, \dots, N\}) \quad (3)$$

trong đó: $T_i(x)$ là dự đoán của cây thứ i cho mẫu x.

Đối với bài toán hồi quy, kết quả được xác định bằng cách lấy trị trung bình của các dự đoán:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N T_i(x) \quad (4)$$

trong đó, $T_i(x)$ là dự đoán của cây thứ i cho mẫu x.

So với thuật toán cây quyết định, RF linh hoạt hơn và có độ chính xác cao do có tính chất tổng hợp của nhiều cây quyết định. Bên cạnh đó, RF có thể xử lý hiệu quả các bộ dữ liệu lớn và phức tạp, bao gồm cả dữ liệu số và dữ liệu phân loại. Ngoài ra, việc sử dụng nhiều cây quyết định cũng giúp giảm thiểu hiện tượng overfitting, mang lại kết quả dự đoán ổn định hơn đặc biệt với dữ liệu vệ tinh khi mà các pixel nhiễu khó phân loại tương đối phổ biến (Belgiu & Drăguț, 2016) (Rodriguez-Galiano, Ghimire, Chica-Olmo, & Rigol-Sanchez, 2012). Đặc biệt, RF có thể cung cấp thông tin về tầm quan trọng của các biến đầu vào, giúp hiểu rõ hơn về các yếu tố ảnh hưởng đến dự đoán.

2.2.1. Thuật toán vector máy hỗ trợ (Support Vector Machine)

Vector máy hỗ trợ (SVM) là một thuật toán học máy có giám sát, chủ yếu được sử dụng cho các bài toán phân loại và hồi quy. SVM hoạt động bằng cách tìm kiếm một siêu phẳng (hyperplane) tối ưu để phân chia các lớp dữ liệu trong không gian nhiều chiều sao cho khoảng cách từ siêu phẳng đó đến các điểm dữ liệu gần nhất của mỗi lớp là lớn nhất, hay còn gọi là biên an toàn (margin) (Vapnik, 1998).

Về công thức của thuật toán SVM bao gồm các thành phần:

- Siêu phẳng phân tách có dạng:

$$w \cdot x + b = 0 \quad (5)$$

trong đó: w là vector trọng số và b là bias.

- Margin: khoảng cách từ siêu phẳng đến điểm dữ liệu gần nhất là: $\frac{1}{|w|}$

- Hàm mục tiêu:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (6)$$

Với các ràng buộc:

$$y_i(w \cdot x_i + b) \geq 1, \quad \forall i \quad (7)$$

- Hàm mục tiêu với soft margin (trong trường hợp SVM có lỗi cho phép):

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (8)$$

Với các ràng buộc:

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i \quad (9)$$

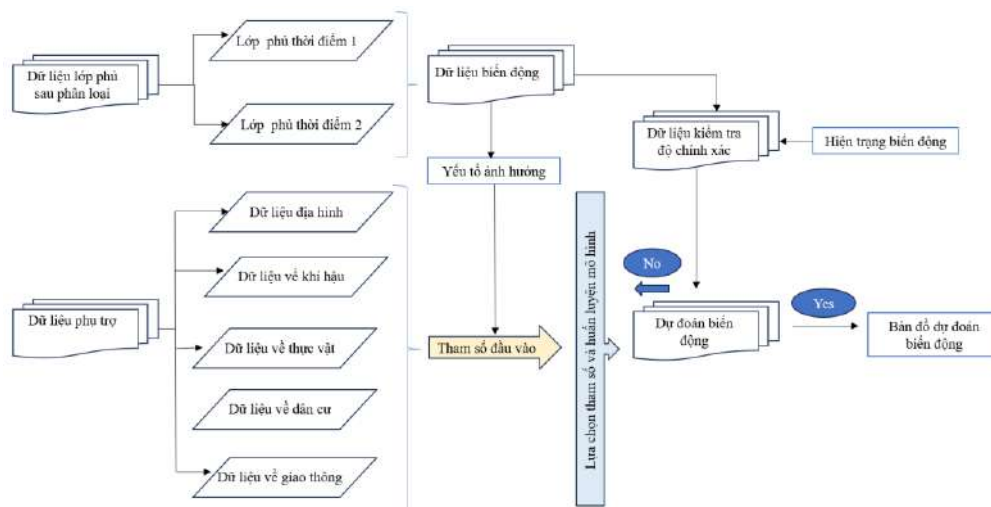
trong đó: ξ_i là biến lỗi cho phép và CCC là tham số điều chỉnh mức độ cho phép lỗi.

- Kernel Trick: SVM có thể sử dụng hàm kernel để ánh xạ dữ liệu vào không gian cao hơn, để dễ dàng phân tách. Một số kernel thường được dùng: Kernel tuyến tính; Kernel đa thức; RBF kernel.

SVM có hiệu quả khi làm việc với dữ liệu có nhiều thuộc tính hay đặc trưng và có thể tìm ra các mô hình phân loại trong không gian nhiều chiều (ví dụ thêm về chiều thời gian). Bên cạnh đó, SVM có khả năng tổng quát hóa tốt, đặc biệt với các bộ dữ liệu nhỏ đến vừa phải và có thể giảm thiểu hiện tượng overfitting (Melgani & Bruzzone, 2004) (Mountrakis, Im, & Ogole, 2011) (Huang, Davis & Townshend, 2002). Ngoài ra, bằng cách sử dụng các hàm kernel ánh xạ dữ liệu vào không gian cao hơn, SVM có thể phân tách dữ liệu dễ dàng hơn.

2.3. Đề xuất quy trình dự đoán biến động lớp phủ bề mặt

Việc dự đoán biến động lớp phủ bề mặt tương đối phức tạp do quá trình này có thể bị tác động bởi những yếu tố tự nhiên và nhân tạo, đôi khi là sự tương tác qua lại giữa các yếu tố này nên để có thể triển khai được mô hình học máy, cần nhiều dữ liệu tham số đầu vào trong đó vai trò chính vẫn là hiện trạng lớp phủ ở các thời điểm gián đoạn thu nhận hoặc cập nhật dữ liệu. Quy trình đề xuất được mô hình hóa ở Hình 2.



Hình 2. Quy trình đề xuất ứng dụng thuật toán học máy trong dự đoán biến động

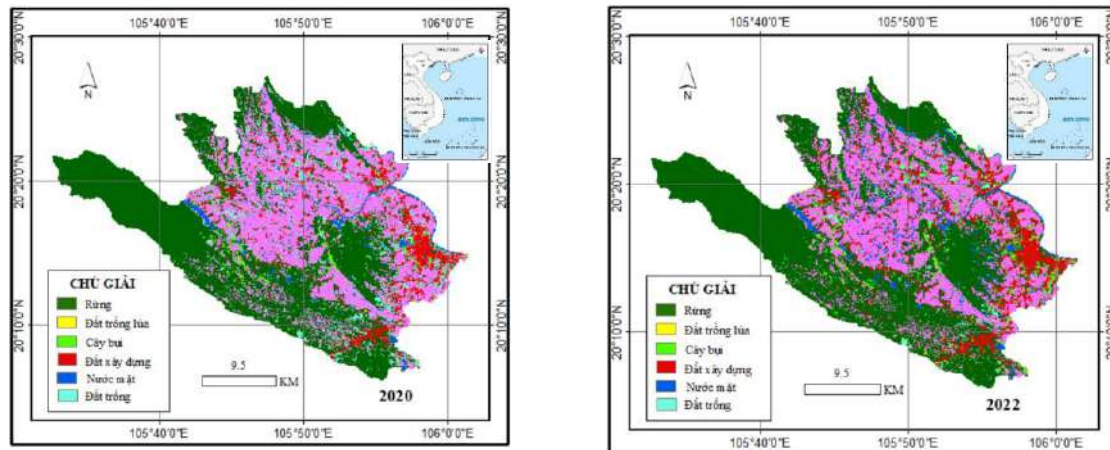
Trong đó, dữ liệu đầu vào bao gồm: dữ liệu ảnh vệ tinh để thu thập thông tin về hiện trạng lớp phủ; dữ liệu địa lý chung với thông tin về địa hình, khí hậu và các yếu tố tự nhiên khác (loại đất) và dữ liệu kinh tế - xã hội (dân cư, tốc độ đô thị hóa). Các dữ liệu này được chuẩn hóa về mặt hình học và độ phân giải, phân loại và phân để trích xuất các đặc trưng.

Các thông tin về phân loại và đặc trưng sau đó được lọc để xử lý giá trị nhiễu và tạo bộ dữ liệu mẫu chuẩn. Bộ dữ liệu mẫu được chia thành các tập huấn luyện và kiểm tra để đánh giá độ chính xác của mô hình.

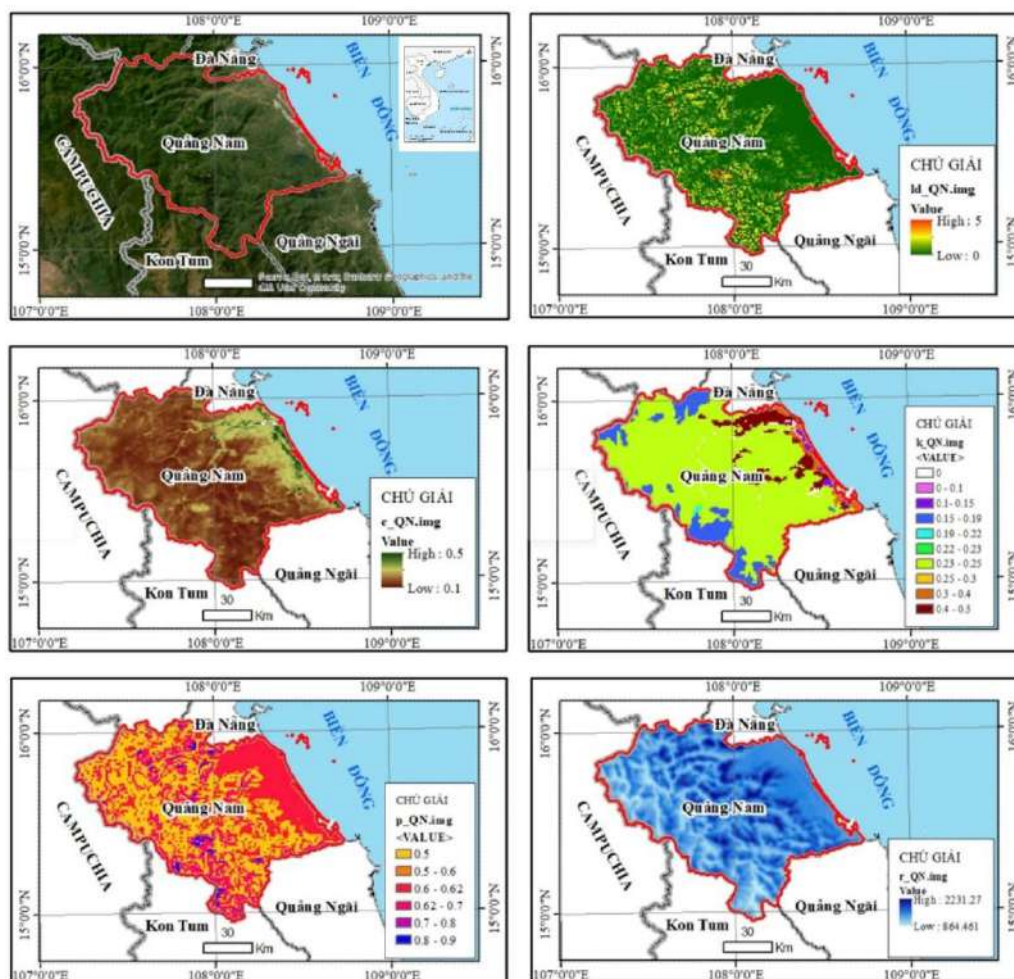
Trong quá trình huấn luyện, các tham số sẽ được tinh chỉnh để tối ưu hóa hiệu suất của mô hình. Việc đánh giá hiệu suất của mô hình được tiến hành thông qua đối sánh với tập dữ liệu kiểm tra thông qua các chỉ số như độ chính xác, độ đặc hiệu và F1 score. Ngoài ra, để đảm bảo mô hình không bị overfitting có khả năng tổng quát hóa tốt, các tác giả đề xuất sử dụng phương pháp k-fold. Khi mô hình đạt đến hiệu suất mong muốn, tiến hành triển khai mô hình vào hệ thống và dự đoán thay đổi đối với bộ dữ liệu mới, tiếp tục đối sánh để nâng cao chất lượng của mô hình. Bên cạnh việc giám sát hoạt động của mô hình, việc cập nhật dữ liệu mới định kỳ cũng đóng vai trò quan trọng trong duy trì độ chính xác và hiệu suất của dự đoán.

3. Kết quả và thảo luận

Hiện tại, nghiên cứu đang trong quá trình lựa chọn khu vực nghiên cứu và xây dựng bộ dữ liệu mẫu. Trong đó, theo dự kiến khu vực thực nghiệm ở hai khu vực: khu vực có tốc độ đô thị hóa cao và khu vực thường xảy ra các tai biến tự nhiên như xói lở. Hình 3 là bản đồ hiện trạng lớp phủ một trong hai khu vực dự kiến thực nghiệm. Hình 4 là khu vực tỉnh Quảng Ngãi với các tham số hiệu chỉnh phục vụ xói mòn đất.



Hình 3. Bản đồ hiện trạng lớp phủ 2020 và 2022 khu vực tỉnh Ninh Bình



Hình 4. Khu vực thực nghiệm tỉnh Quảng Ngãi với tham số hiệu chỉnh chuẩn hóa cho mô hình biến động do xói mòn đất với các chỉ số ls và P liên quan đến địa hình, c liên quan đến lớp phủ và lớp hiện trạng sử dụng đất, k liên quan đến loại đất, và r liên quan đến khí hậu (Trần, Trần & Đoàn, 2023)

Do điều kiện thu thập và khai thác dữ liệu ở mỗi khu vực có những hạn chế khác nhau, điều này cũng phản ánh thực tế trong những khó khăn của việc thu thập và xử lý dữ liệu ở Việt Nam nói chung. Tuy nhiên, với những dữ liệu hiện có, các tác giả sẽ tiếp tục tập trung về mặt làm hài hòa dữ liệu và tiến hành phát triển các thuật toán trong điều kiện hạn chế với các thông số đầu vào.

4. Kết luận

Nghiên cứu đã đề cập tới lý thuyết chung về lớp phủ bề mặt và ý nghĩa của việc theo dõi giám sát của lớp phủ bề mặt. Thêm vào đó, nghiên cứu cũng tiến hành phân tích ưu và nhược điểm của các thuật toán học máy phổ biến trong phân loại và dự đoán biến động lớp phủ bề mặt. Các thuật toán cây quyết định, rừng ngẫu nhiên và vector máy hỗ trợ đều là những thuật toán có tính tổng quát và trực quan có khả năng áp dụng đối với nghiên cứu về lớp phủ và biến động lớp phủ. Những thuật toán này dựa vào việc chiết xuất đặc trưng và chuyên gia phân tích để tạo ra các thông tin đặc trưng cho bộ dữ liệu mẫu, giúp máy tính học tốt hơn và phù hợp với khả năng tính toán trên máy tính cá nhân. Bên cạnh đó, với bộ dữ liệu không liên tục và không đủ lớn, các thuật toán này là lựa chọn phù hợp so với các thuật toán về mạng nơron nhân tạo. Nghiên cứu đã xác định khu vực thực nghiệm và đầu vào dữ liệu dự kiến. Trong thời gian tới, nghiên cứu sẽ tiến hành thiết lập bộ dữ liệu mẫu và huấn luyện mô hình. Kết quả về hiệu suất và độ chính xác của các thuật toán sẽ được đem ra so sánh để đề xuất thuật toán phù hợp có khả năng áp dụng ở những khu vực khác ở Việt Nam.

Lời cảm ơn

Bài báo là kết quả tìm hiểu và nghiên cứu của các tác giả khi thực hiện đề tài cấp cơ sở trường Đại học Mở Địa Chất, mã số: T24-02.

Một số hình ảnh trong bài viết như Hình 1 và Hình 4 được sử dụng tuân thủ theo quy định Copyright.

Tài liệu tham khảo

- Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*(114), 24-31.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and Regression Trees*. . Wadsworth International Group.
- ESA. (n.d.). Worldwide land cover mapping. Retrieved from <https://esa-worldcover.org/en>
- García, L., & Rodríguez, A. (2022). Predicting land cover changes with support vector machines: A case study in South America. *Journal of Environmental Management*, 308, 114528.
- Giri, C. P. (2020). *Remote Sensing of Land Use and Land Cover*. CRC Press.
- Gregorio, A. D., & Jansen, L. J. (2005). *Land Cover Classification System - Classification concepts and user manual*. Rome: FAO. Retrieved from <https://www.fao.org/4/x0596e/x0596e01f.htm#:~:text=The%20Land%20Cover%20Classification%20System,and%20Semi%2DNatural%20Terrestrial%20Vegetation>
- Hansen, M. C., Potapov, P. V., Hancher, M., Turubanova, S. A., Tyukavina, A. .., & Townshend, J. R. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160), 850-853.
- Herrmann, S. M., & Hutchinson, C. F. (2005). The changing contexts of the desertification debate. *Journal of Arid Environments*, 63(3), 538-555.
- Huang, C., Davis, L. S., & Townshend, J. R. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4), 725-749.
- Johnson, R., & Lee, S. (2023). Random forest-based models for forecasting land use and land cover changes. *Environmental Research Letters*, 18(3).
- Kumar, P., & Sharma, A. (2023). Machine learning approaches for land cover change detection: A review. *Environmental Monitoring and Assessment*, 195(2), 345.
- Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8), 1779-1790.
- Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247-259.
- Nguyễn, V. H., & Đăng, T. T. (2021). Đánh giá biến động lớp phủ mặt đất tại thành phố Đà Nẵng bằng ảnh viễn thám Landsat. *Tạp chí Môi trường*.
- Pal, M., & Mather, P. M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, 86(4), 554-565.
- Patel, D., & Desai, M. (2021). Machine learning models for predicting agricultural land cover changes: A comparative study. *Computers and Electronics in Agriculture*, 192.
- Phạm, M. T., & Nguyễn, T. L. (2023). Ứng dụng viễn thám trong đánh giá biến động lớp phủ đất tại vùng đồng bằng sông Hồng. *Tạp chí Khoa học và Công nghệ*.

- Quinlan, J. R. (1986). Induction of Decision Trees. *Machine Learning*, 1(1), 81-106.
- Rodriguez-Galiano, V. F., Ghimire, B. R., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*(67), 93-104.
- Seto, K. C., Güneralp, B., & Hutyra, L. R. (2012). Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences*, 109(4), 16083-16088.
- Thackway, R., & Atyeo, C. (2006). *Classifying Australia LandCover*.
- Trần, T. H., Trần, T. H., & Đoàn, C. L. (2023). A novel approach of neural networks and USLE in smart soil erosion modeling, case study in Southern coastal of Vietnam. *International conference GIS IDEAS 2023* (pp. 369-379). Ha Noi: Publishing house for Science and Technology.
- Trần, T. T. (2020). Phân tích biến động lớp phủ đô thị bằng dữ liệu ảnh vệ tinh tại thành phố Hồ Chí Minh. *Tạp chí Khoa học và Phát triển*.
- Vapnik, V. (1998). *Statistical Learning Theory*. Wiley.
- Zhang, H., Li, Y., & Wang, X. (2023). Application of deep learning in predicting land cover changes using satellite imagery. *Remote Sensing*, 15(1), 1-15.
- Zhou, F., & Liu, Y. (2022). Predicting land cover changes in arid regions using machine learning algorithms. *Journal of Arid Environments*, 199.

ABSTRACT

Predicting land cover changes using Machine learning algorithms

Tran Thi Hoa^{1*}, Tran Thi Ngoc¹, Tran Dinh Tri²

¹*Hanoi University of Mining and Geology*

²*Vietnam Association of Geography, Cartography and Remotesensing*

Land cover refers to the physical surface morphology of the Earth's crust, encompassing vegetation, water, natural surfaces, and artificial objects. Activities occurring on the Earth's surface, ranging from natural processes to socio-economic development and urbanization, lead to significant changes in land cover. Soil resources are among the critical resources with profound impacts on the environment in general and sustainable development goals in particular. Therefore, monitoring and observing changes in land cover, as well as analyzing and forecasting trends, are crucial tasks. Current machine learning algorithms have made significant contributions by leveraging computational power and ensuring objectivity in analyzing related predictive issues. This paper examines the potential applications of various machine learning algorithms in predicting land cover changes, specifically: Decision Trees, Random Forests, and Support Vector Machines (SVM).

Keywords: Land cover changes, predicting, Machine Learning, Random Forest, SVM.