

Ứng dụng trí tuệ nhân tạo và viễn thám trong đánh giá thoái hóa đất hiện đại

Nguyễn Phi Sơn¹, Đinh Ngọc Phú², Nguyễn Thanh Thủy^{3*}, Trần Xuân Trường⁴

¹*Viện Khoa học Đo đạc và Bản đồ*

²*Văn phòng đăng ký đất đai Hà Nội*

³*Viện Dữ liệu không gian*

⁴*Trường Đại học Mỏ - Địa chất*

*Email tác giả liên hệ: ntthuydostic@gmail.com

<https://doi.org/10.67484/mgc.13>

Tóm tắt: Đánh giá thoái hóa đất theo quy định hiện hành dựa trên điều tra, lấy mẫu thực địa và khoan vẽ theo đơn vị đất nên khó đáp ứng yêu cầu giám sát liên tục, chi tiết và trên diện rộng. Nghiên cứu này luận giải cơ sở khoa học của việc ứng dụng trí tuệ nhân tạo (Artificial Intelligence - AI) kết hợp viễn thám để suy giải các chỉ số, hệ số phục vụ đánh giá 06 loại hình thoái hóa đất. Kết quả chính gồm: làm rõ bản chất bài toán dưới dạng tương quan phi tuyến giữa đặc trưng phổ, điều kiện tự nhiên và mẫu phân tích đất; đề xuất khung biến đặc trưng đa nguồn cùng dạng tổng quát của bộ mẫu huấn luyện; xác lập quy trình huấn luyện, đánh giá mô hình và lập bản đồ phù hợp với khung pháp lý 06 loại hình. Trên cơ sở tổng hợp bằng chứng quốc tế, nghiên cứu chỉ ra các họ mô hình phù hợp với từng loại hình thoái hóa, trong đó các mô hình rừng ngẫu nhiên và tăng cường gradient thường đạt hiệu suất cao và ổn định với dữ liệu đa nguồn, làm cơ sở triển khai theo hướng tự động hóa, cập nhật thường xuyên và áp dụng theo quy mô tỉnh, vùng và quốc gia.

Từ khóa: Thoái hóa đất, Đánh giá thoái hóa đất, Trí tuệ nhân tạo, Viễn thám, Học máy, Học sâu.

Ngày nhận bài: 25/04/2026 Ngày sửa lại: 18/05/2026 Ngày chấp nhận đăng: 19/05/2026 Ngày xuất bản: 30/08/2026

Application of artificial intelligence and remote sensing in modern land degradation assessment

Phi Son Nguyen¹, Ngoc Phu Dinh², Thanh Thuy Nguyen^{3*}, Xuan Truong Tran⁴

¹*The Vietnam Institute of Surveying and Mapping*

²*Hanoi Department of Land Registration*

³*Spatial Data Institute*

⁴*Hanoi University of Mining and Geology*

*Corresponding Author Email: ntthuydostic@gmail.com

Abstract: Land degradation assessment under current Vietnamese regulations relies on field surveys, soil sampling and parcel-based delineation, which struggle to meet the demand for continuous, detailed and large-area monitoring. This study clarifies the scientific basis for integrating artificial intelligence (AI) with remote sensing to retrieve the indices and coefficients used in assessing the six statutory types of land degradation. The main contributions are: framing the task as a nonlinear correlation between spectral features, natural conditions and laboratory soil samples; proposing a multi-source feature set together with a general form of the training sample; and establishing a model training, evaluation and mapping workflow consistent with the six-type legal framework. Drawing on international evidence, the study identifies the model families suited to each degradation type, in which random forest and gradient boosting models generally achieve high and stable performance with multi-source data, providing a basis for automated, regularly updated assessment at provincial, regional and national scales.

Keywords: Land degradation, Land degradation assessment, Artificial intelligence, Remote sensing, Machine learning, Deep learning.

Submission received: 25/04/2026

Revised: 18/05/2026

Accepted: 19/05/2026

Published: 30/08/2026

1. Giới thiệu

Đánh giá thoái hóa đất đã được quy định trong Luật Đất đai 2024 [1] và các văn bản dưới Luật. Phương pháp đánh giá, chỉ tiêu đánh giá đã quy định tại văn bản hợp nhất số 47/VBHN-BNNMT [2], hiện nay các địa phương trong cả nước đang áp dụng để định kỳ đánh giá thoái hóa đất. Số liệu về thoái hóa đất được xác định là một trong các chỉ tiêu thống kê đất đai quốc gia, được thực hiện ở 3 cấp: tỉnh, vùng và quốc gia. Thoái hóa đất được đánh giá theo 06 loại hình gồm: suy giảm độ phì; xói mòn đất do mưa; kết von, đá ong hóa; khô hạn, hoang mạc hóa, sa mạc hóa; phèn hóa; mặn hóa.

Quy trình kỹ thuật trong văn bản hợp nhất số 47/VBHN-BNNMT [2] về bản chất là điều tra thực địa, lấy mẫu, phân tích phòng thí nghiệm, khoan vùng theo bản đồ đơn vị đất, rất chuẩn

mức nhưng khó đáp ứng yêu cầu “giám sát liên tục, chi tiết, diện rộng” do mỗi giá trị phân tích mẫu đại diện cho một đơn vị đất. Cách tiếp cận này chủ yếu dựa vào điều tra thực địa, phân tích mẫu và chồng xếp bản đồ theo ngưỡng phân cấp cố định nên mang tính kinh nghiệm và tương đối (cao - thấp, nặng - nhẹ), chưa phản ánh đúng tính biến thiên không gian liên tục của môi trường đất, dẫn đến sai số đại diện, chưa phản ánh đầy đủ các quan hệ phi tuyến và tương tác phức tạp giữa điều kiện tự nhiên - môi trường tác động đến sự phân bố mức độ suy thoái trong từng đơn vị đất (coi trong một đơn vị đất có sự đồng nhất tương đối).

Khi sử dụng ảnh viễn thám và các dữ liệu về điều kiện tự nhiên ở mức độ chi tiết phù hợp làm đầu hiệu tương quan với giá trị thoái hóa đất tại điểm lấy mẫu đất, thông qua các thuật toán học máy, học sâu để thiết lập mô hình huấn luyện sẽ suy giải được các pixel đất còn lại của toàn bộ khu vực đánh giá. Điều này tạo ra giá trị liên tục về mức độ suy thoái của mỗi loại hình theo không gian một cách dễ dàng.

Phương pháp ứng dụng trí tuệ nhân tạo và viễn thám trong đánh giá thoái hóa đất hiện đại được phát triển mang tính “*thay thế/tăng sự tác động của điều kiện tự nhiên/suy giải chi tiết dựa trên tương quan đa tác động*” chứ không nhất thiết bỏ hoàn toàn phương pháp theo quy định đánh giá thoái hóa đất hiện hành. Cách tiếp cận này giải quyết được các hạn chế như:

Trước hết, cách tiếp cận này khắc phục hạn chế về độ phủ không gian và tính đại diện của mẫu. Điều tra truyền thống luôn bị ràng buộc bởi mật độ và vị trí điểm lấy mẫu nên khó đại diện cho tính dị thể của đất ở quy mô rộng, trong khi ảnh vệ tinh cho phép phủ kín diện rộng, tạo bề mặt liên tục và giảm rủi ro đánh giá lệch do nội suy từ ít điểm điều tra. Cùng với đó là khả năng nâng cao tần suất cập nhật theo thời gian, bởi độ phì, xói mòn, kết von, khô hạn, mặn hóa và phèn hóa cùng các yếu tố liên quan như độ ẩm đất, che phủ thực vật, canh tác, khí tượng và thủy văn đều biến động theo mùa vụ và theo năm. Ảnh đa thời gian vì vậy cho phép theo dõi xu thế biến động cũng như phát hiện điểm nóng theo thời gian gần thực thay vì chỉ phản ánh một thời điểm điều tra.

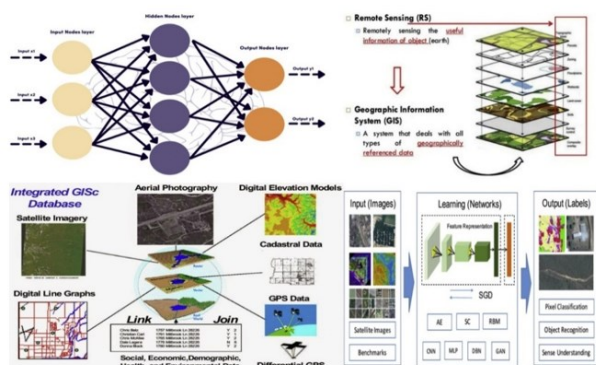
Quan trọng hơn, học máy và học sâu cho phép mô hình hóa quan hệ phi tuyến, đa biến vốn là bản chất của thoái hóa đất, khi mỗi loại hình là kết quả tổ hợp của thổ nhưỡng, địa hình, khí hậu, canh tác, lớp phủ và thủy văn. Quan hệ giữa đặc trưng phổ hoặc chuỗi thời gian với chỉ tiêu đất thường phi tuyến và có tương tác, trong khi các mô hình học máy, học sâu xử lý tốt dữ liệu đa nguồn, đa chiều để ước lượng chỉ số, hệ số cho từng loại hình; nhiều nghiên cứu cho thấy học máy kết hợp viễn thám có thể xây dựng mô hình ước lượng thoái hóa đất ở quy mô tỉnh, vùng với độ tin cậy cao [3]. Đồng thời, việc tích hợp viễn thám, GIS và học máy, học sâu giúp chuẩn hóa quy trình theo nguyên tắc cùng dữ liệu đầu vào cho cùng kết quả, kèm theo định lượng độ tin cậy và khả năng so sánh theo chuỗi thời gian.

Ứng dụng AI kết hợp viễn thám trong đánh giá thoái hóa đất tạo ra bước chuyển từ cách tiếp cận kinh nghiệm (knowledge-based) sang cách tiếp cận dựa trên dữ liệu (data-driven), với các ưu điểm cốt lõi như: bao phủ rộng, tính đến các tác động đa nguồn, mô hình hóa phi tuyến, tự động hóa cao, cập nhật nhanh và độ chính xác đồng nhất hơn, qua đó đáp ứng hiệu quả yêu cầu giám sát tài nguyên đất trong bối cảnh chuyển đổi số (Hình 1).

Trên thế giới, ảnh vệ tinh cùng dữ liệu điều kiện tự nhiên - môi trường đã được sử dụng rộng rãi để huấn luyện các mô hình học máy (Machine Learning - ML) và học sâu (Deep Learning - DL) nhằm suy giải các chỉ số đánh giá và loại hình thoái hóa đất. Các bài tổng quan [4-6] cho thấy xu hướng dịch chuyển từ nội suy tuyến tính sang lập bản đồ liên tục bằng học máy, khai thác dữ liệu chuỗi thời gian, ảnh đất trống tổng hợp, tổ hợp ảnh radar khẩu độ tổng hợp (Synthetic Aperture Radar - SAR) với ảnh quang học và các mô hình có khả năng giải thích (Explainable AI - XAI).

Phân tích sâu hơn các công trình đã công bố cho thấy mỗi họ mô hình có ưu thế, hạn chế và điều kiện áp dụng riêng. Các mô hình tổ hợp cây quyết định như rừng ngẫu nhiên (Random Forest - RF) và tăng cường gradient (Extreme Gradient Boosting - XGBoost) được sử dụng phổ biến nhất do xử lý tốt dữ liệu đa nguồn, ít nhạy với nhiễu và cho phép xếp hạng tầm quan trọng của biến; trong đó XGBoost thường vượt trội với dữ liệu mất cân bằng hoặc bài toán phân loại còn RF ổn

định và dễ hiệu chỉnh hơn. Mạng nơ-ron nhân tạo (Artificial Neural Network - ANN) và mạng nơ-ron tích chập (Convolutional Neural Network - CNN) phù hợp khi cần khai thác cấu trúc không gian hoặc quan hệ phi tuyến phức tạp nhưng đòi hỏi số lượng mẫu lớn và chi phí tính toán cao. Hồi quy vector hỗ trợ (Support Vector Regression - SVR) lại tỏ ra cạnh tranh trong các bài toán hồi quy với cỡ mẫu vừa phải. Việc lựa chọn mô hình vì vậy phụ thuộc vào loại hình thoái hóa, đặc tính dữ liệu và mục tiêu đầu ra, được minh chứng qua các nhóm nghiên cứu theo từng loại hình dưới đây.



Hình 1. Minh họa các giai đoạn suy giải chỉ số, hệ số phục vụ đánh giá thoái hóa đất

Đối với mặn hóa, các nghiên cứu khai thác kết hợp kênh phổ vùng nhìn thấy, cận hồng ngoại và hồng ngoại sóng ngắn (VIS-NIR-SWIR), các chỉ số mặn, mô hình số độ cao (Digital Elevation Model - DEM), biến khí hậu và nhân độ dẫn điện. Trên ảnh Landsat OLI với 13 chỉ số mặn và 241 mẫu đất, mô hình ANN cho diện tích dưới đường cong AUC đạt 0,921, cao hơn RF [7]; với 26 biến môi trường trích từ Landsat-8, XGBoost vượt RF cả về R^2 và RMSE (R^2 khoảng 0,83), đồng thời phân tích SHAP cho thấy mức đóng góp của chỉ số thực vật và kênh

nhìn thấy thay đổi theo từng khu vực [9]. Các nghiên cứu tại UAE [8], Kazakhstan [10] và vùng bán khô hạn [11] cũng cố nhận định rằng nhóm mô hình tổ hợp và ANN cho kết quả tốt nhưng phụ thuộc vào mật độ mẫu và mức độ che phủ thực vật.

Đối với carbon hữu cơ trong đất (Soil Organic Carbon - SOC) và chất hữu cơ trong đất (Soil Organic Matter - SOM), hướng tiếp cận hồi quy chiếm ưu thế. Mô hình SVR cho kết quả tốt nhất khi tổ hợp chỉ số phổ từ Sentinel-2 và Sentinel-3, đạt $R^2 = 0,627$ và $RMSE = 0,620\%$ [3]; việc hợp nhất dữ liệu đa phổ với radar kết hợp học chủ động và học tổ hợp giúp nâng cao độ chính xác dự báo SOC đất nông nghiệp [13]. Các mô hình CNN cải tiến khai thác tốt cấu trúc không gian để lập bản đồ SOM [14], [15], trong khi việc theo dõi động thái SOM dài hạn bằng MODIS [12] và xử lý điều kiện bề mặt đất trống được xác định là điều kiện tiên quyết để mô hình SOC ổn định [16].

Đối với pH, độ phì, xói mòn và sa mạc hóa, việc tổ hợp dữ liệu radar - quang học cùng kiểm soát ảnh hưởng của lớp phủ giữ vai trò quyết định. Bản đồ pH quy mô châu lục được lập từ dữ liệu LUCAS, Sentinel-1/2 và PALSAR [17]; loại bỏ thảm phủ thực vật cải thiện rõ rệt dự báo độ phì [18]; đất thoái hóa do xói mòn được nhận diện từ Sentinel-2 [19]; mô hình RF kết hợp chuỗi Markov theo dõi sa mạc hóa [20]. Đáng chú ý, khi so sánh trực tiếp RF, XGBoost, Naïve Bayes (NB) và K-Nearest Neighbors (KNN) trên 06 chỉ số đa phổ tại Turkmenistan, XGBoost đạt độ chính xác tổng thể cao nhất 96,33%, với chỉ số đất trống (Bare Soil Index - BSI) là chỉ thị then chốt [21]. Bảng chứng tiêu biểu theo từng loại hình được tổng hợp tại Bảng 1.

Qua tổng quan các nghiên cứu gần đây cho thấy trí tuệ nhân tạo, đặc biệt là học máy và học sâu kết hợp ảnh viễn thám, dữ liệu môi trường và mẫu đất thực địa, đã trở thành hướng tiếp cận trung tâm trong lập bản đồ, giám sát và dự báo động các dạng thoái hóa đất ở nhiều quy mô khác nhau. Tuy nhiên phần lớn các nghiên cứu quốc tế được công bố mới chỉ giải quyết riêng lẻ từng loại hình thoái hóa (mặn hóa, suy giảm hữu cơ, sa mạc hóa, xói mòn...) trên những vùng lãnh thổ cụ thể, sử dụng một nhóm cảm biến nhất định và chưa gắn với một khung phân loại pháp lý thống nhất. Tính mới của nghiên cứu này không nằm ở việc đề xuất một thuật toán đơn lẻ mà ở chỗ luận giải có hệ thống cơ sở khoa học cho việc tích hợp đồng thời ba nhóm yếu tố - dữ liệu viễn thám đa nguồn (quang học, radar và các chỉ số phổ dẫn xuất), các điều kiện tự nhiên hình thành đất, cùng quy trình đánh giá thoái hóa đất theo quy định hiện hành của Việt Nam với đầy đủ 06 loại hình - trong một khung phương pháp luận thống nhất. Đây là khoảng trống mà các nghiên cứu đơn loại hình ở nước ngoài chưa giải quyết và cũng là điểm khác biệt so với các đánh giá thoái hóa đất theo phương pháp truyền thống đã triển khai trong nước. Trên cơ sở đó, nghiên cứu này

luận giải ba vấn đề khoa học cốt lõi. Thứ nhất là khả năng và mức độ chính xác mà các đặc trưng phổ từ ảnh vệ tinh đa phổ và ảnh radar có thể phản ánh định lượng sự biến đổi của các chỉ tiêu độ phì nhiêu đất, khô hạn đất, kết von đất, nhiễm mặn và nhiễm phèn. Thứ hai là giải pháp tích hợp hiệu quả các nguồn dữ liệu không đồng nhất - gồm các kênh phổ, chỉ số phổ, điều kiện tự nhiên và mẫu đất thực địa - nhằm nâng cao độ chính xác suy giải các chỉ số thoái hóa. Thứ ba là cơ sở lựa chọn mô hình học máy hoặc học sâu có khả năng khai thác tốt nhất dữ liệu viễn thám và GIS để ước lượng chính xác các chỉ số thoái hóa đất.

Bảng 1. Tổng hợp một số mô hình học máy, học sâu tiêu biểu theo loại hình thoái hóa đất từ các nghiên cứu quốc tế

TT	Loại hình thoái hóa	Nghiên cứu	Dữ liệu, cảm biến chính	Mô hình cho kết quả tốt và hiệu suất tiêu biểu
1	Mặn hóa	[7]	Landsat OLI; 13 chỉ số mặn; 241 mẫu đất	Mạng nơ-ron nhân tạo (ANN) đạt diện tích dưới đường cong AUC = 0,921, cao hơn rừng ngẫu nhiên (RF)
2	Mặn hóa	[9]	Landsat-8; 26 biến môi trường; phân tích SHAP	XGBoost vượt RF cả về R^2 và RMSE ($R^2 \approx 0,83$); diễn giải đóng góp biến bằng học máy giải thích được
3	Suy giảm độ phì (SOC)	[3]	Tổ hợp chỉ số phổ Sentinel-2 và Sentinel-3	Hồi quy vector hỗ trợ (SVR) cho kết quả tốt nhất: $R^2 = 0,627$; RMSE = 0,620%
4	Suy giảm độ phì (SOM)	[14], [15]	Sentinel-2; mạng nơ-ron tích chập (CNN)	CNN cải tiến khai thác cấu trúc không gian để lập bản đồ chất hữu cơ đất
5	Khô hạn, sa mạc hóa	[21]	Sentinel-2; 06 chỉ số đa phổ (NDVI, SAVI, NDMI, BSI, EVI, LST)	XGBoost đạt độ chính xác tổng thể cao nhất 96,33% (vượt RF, NB, KNN); chỉ số đất trống BSI là chỉ thị then chốt
6	pH, độ phì, xói mòn	[17], [18], [19]	Sentinel-1/2, PALSAR, dữ liệu LUCAS	Tổ hợp radar - quang học lập bản đồ pH quy mô châu lục; loại bỏ thảm phủ cải thiện dự báo độ phì

2. Phương pháp nghiên cứu

2.1. Xác định các chỉ số, hệ số và loại hình thoái hóa cần ứng dụng mô hình học máy, học sâu

Qua phân tích vai trò của từng chỉ số, hệ số để định lượng mức độ thoái hóa của 06 loại hình, các mô hình học máy/học sâu được áp dụng vào như sau:

Mô hình đánh giá độ phì và suy giảm độ phì đất: cần 06 chỉ số pH_{KCl} , %OM, %CEC, %N, % P_2O_5 , % K_2O . Đây là các thành phần hóa học đất, vì vậy cả 06 chỉ số đều cần đến ứng dụng học máy, học sâu, viễn thám và các điều kiện tự nhiên khác để dự báo và suy giải.

Mô hình đánh giá xói mòn đất: hiện đang quy định sử dụng phương trình mất đất phổ dụng cải tiến (Revised Universal Soil Loss Equation - RUSLE) để định lượng mức độ mất đất trên 1 ha trong 1 năm [22], biểu diễn theo công thức (1):

$$A = R \cdot K \cdot LS \cdot C \cdot P \quad (1)$$

Trong đó: A là mức độ mất đất (tấn/ha/năm); R là hệ số xói mòn do mưa; K là hệ số kháng xói của đất; LS là hệ số địa hình; C là hệ số lớp phủ thực vật; P là hệ số biện pháp canh tác. Trong 05 hệ số sẽ có 02 hệ số cần đến ứng dụng học máy/học sâu để suy giải, đó là lớp dữ liệu K và lớp dữ liệu P. Đối với hệ số R, LS, C được suy giải dựa trên các phần mềm GIS để tạo ra lớp dữ liệu liên tục.

Mô hình đánh giá kết von, đá ong hóa: dựa trên 03 chỉ số gồm hình dạng, kích thước và số lượng hạt kết von trong đất. Đây là thuộc tính cấu trúc của đất, do vậy cần ứng dụng học máy/học sâu để suy giải thông qua các đặc trưng từ ảnh vệ tinh, đặc biệt là ảnh radar kết hợp các đặc trưng là điều kiện tự nhiên tác động lên sự hình thành của kết von, đá ong hóa.

Mô hình đánh giá khô hạn, hoang mạc hóa, sa mạc hóa: đều dựa trên các chỉ số đo của các trạm khí tượng như lượng mưa, lượng bốc hơi, nhiệt độ, độ ẩm, thảm thực vật. Việc nội suy để tạo lớp dữ liệu liên tục được thực hiện dựa trên ArcGIS. Đối với phân loại thảm thực vật sẽ ứng

dụng học sâu để suy giải kiểu thực vật quyết định đến loại hoang mạc hóa, sa mạc hóa.

Mô hình đánh giá đất bị nhiễm mặn: phụ thuộc nhiều vào các biến đặc trưng tác động đến quá trình nhiễm mặn và phân bố độ mặn, vì vậy cần ứng dụng học máy/học sâu để suy giải thông qua các đặc trưng từ ảnh vệ tinh kết hợp điều kiện tự nhiên hình thành tác nhân mặn và quá trình nhiễm mặn.

Mô hình đánh giá đất bị nhiễm phèn: phụ thuộc nhiều vào các biến đặc trưng tác động đến quá trình nhiễm phèn và phân bố độ phèn, vì vậy cần ứng dụng học máy/học sâu để suy giải thông qua các đặc trưng từ ảnh vệ tinh kết hợp điều kiện tự nhiên hình thành tác nhân phèn và quá trình nhiễm phèn.

2.2. Bản chất khoa học của việc ứng dụng học máy, học sâu

Về bản chất khoa học, việc ứng dụng học máy, học sâu để suy giải, dự đoán thoái hóa đất từ mẫu phân tích đất (nhân) và biến đặc trưng (ảnh vệ tinh, điều kiện tự nhiên...) chính là một bài toán tương quan phi tuyến. Làm rõ được bản chất chính là cơ sở cho việc thiết lập bộ dữ liệu huấn luyện có tương quan tốt với biến kết quả của mô hình học máy, học sâu. Mô hình sẽ học đúng bản chất của đối tượng đó là phân bố liên tục trong không gian, có sự tương quan “láng giềng”. Bản chất của tương quan thể hiện ở các khía cạnh sau:

(1) *Học ánh xạ.* Mục tiêu của mô hình là học một hàm thể hiện qua công thức (2):

$$f: X \rightarrow Y \quad (2)$$

Trong đó: X là các đặc trưng đầu vào (kênh phổ VIS, Red-edge, NIR, SWIR; radar VV, VH, texture; chỉ số ảnh NDVI, MSI, BSI, LSWI...); biến điều kiện tự nhiên DEM, slope, TWI, khí hậu, LULC...; Y là nhân - giá trị mẫu thực địa (kết quả phân tích phòng thí nghiệm về pH, OM, CEC, EC...).

(2) *Tính vật lý - sinh thái.* Học máy, học sâu không “đo trực tiếp” thuộc tính đất mà khai thác các mối liên hệ gián tiếp.

Trước hết là tương quan của phổ phản xạ với các biểu hiện tính chất đất: kênh Red-edge nhạy với hàm lượng diệp lục - chỉ thị gián tiếp cho mức sinh trưởng của thực vật và qua đó phản ánh độ phì, trong khi kênh SWIR nhạy cảm với khoáng vật và độ ẩm vốn liên quan đến %OM, %CEC cũng như suất phản xạ của bề mặt đất trống. Tiếp đến, ảnh radar phản ánh tốt cấu trúc đất khi các kênh phân cực VV, VH mô tả độ nhám, độ ẩm và cấu trúc bề mặt, còn đặc trưng kết cấu (texture - GLCM) được trích xuất từ quan hệ không gian giữa các pixel để mô tả tính không đồng nhất của bề mặt đất. Cuối cùng, các biến điều kiện tự nhiên đóng vai trò tác nhân của quá trình hình thành và biến đổi đất: DEM và độ dốc chi phối quá trình rửa trôi, tích tụ; các biến khí hậu (P, PET, SPEI) tác động đến phong hóa; còn khí tượng và biến đổi khí hậu ảnh hưởng đến lớp phủ bề mặt, đến dấu hiệu carbon hữu cơ, cấu trúc đất và thành phần NPK trong đất.

Học máy, học sâu thực chất học chuỗi quan hệ: điều kiện tự nhiên, địa hình, khí hậu... tác động đến quá trình đất, quá trình đất biểu hiện ở phổ ảnh, từ đó định lượng thoái hóa đất dựa trên nhân.

(3) *Tính thống kê - dữ liệu.* Tính chất thống kê - dữ liệu thể hiện ở ba khía cạnh.

Khía cạnh thứ nhất là quá trình chuyển từ “điểm mẫu đất” sang “bề mặt liên tục”, trong đó dữ liệu đầu vào gồm hàng trăm đến hàng nghìn điểm mẫu đất còn đầu ra là bản đồ liên tục - chính là quá trình khái quát hóa không gian. Khía cạnh thứ hai là sự chuyển đổi từ cách làm dựa trên tri thức và kinh nghiệm chuyên gia sang cách làm dựa trên dữ liệu: nếu quy trình truyền thống gán giá trị theo quy tắc và ngưỡng cố định thì học máy, học sâu học trực tiếp từ dữ liệu (data-driven) mà không cần giả định tuyến tính hay ngưỡng cứng. Khía cạnh thứ ba là khả năng khai thác quan hệ phi tuyến và đa biến, bởi các mô hình học máy, học sâu vốn là mô hình phi tuyến, có khả năng mô tả tương tác giữa các biến và học cấu trúc phân bố không gian của đối tượng.

(4) *Tính không gian.* Tính không gian được thể hiện ở việc tiếp cận theo pixel thay vì polygon: mỗi pixel tương ứng với một vector đặc trưng, không còn giả định đồng nhất theo khoảng đất. Đồng thời, mô hình tính đến tự tương quan không gian: các điểm gần nhau có xu hướng giống nhau.

(5) *Tính bất định.* Do quan hệ giữa phổ ảnh vệ tinh/radar với thuộc tính đất là gián tiếp,

không duy nhất, do ảnh hưởng bởi thảm phủ, độ ẩm, điều kiện chiếu xạ... nên học máy, học sâu thực chất là ước lượng xác suất, không phải đo trực tiếp. Vì vậy việc lựa chọn mô hình và đánh giá hiệu suất là cần thiết để thu nhận được kết quả đầu ra đáp ứng mục đích cao nhất.

2.3. Cơ sở đề xuất bộ dữ liệu đặc trưng

Dựa trên phân tích bản chất của việc ứng dụng học máy, học sâu trong suy giải, dự đoán các chỉ số, hệ số của 06 loại hình thoái hóa đất thành phần, nghiên cứu đề xuất danh mục biến đặc trưng. Giá trị của chỉ số, hệ số tại vị trí lấy mẫu chính là giá trị nhãn. Mỗi bộ mẫu huấn luyện có dạng tổng quát theo công thức (3):

$$Y_i = f(B_i, T_i, I_i, E_i, C_i) + \varepsilon_i \quad (3)$$

Trong đó: Y_i là giá trị nhãn; f là hàm phi tuyến; B_i là các đặc trưng kênh phổ, tán xạ radar; I_i là các đặc trưng chỉ số phổ; T_i là các đặc trưng địa hình; E_i là các đặc trưng kinh tế - xã hội - nhân sinh; C_i là các đặc trưng về khí hậu và biến đổi khí hậu; ε_i là sai số ngẫu nhiên.

Từ các nghiên cứu [13-16], một số nhóm biến đặc trưng được ví dụ tại Bảng 2.

Việc lựa chọn biến đặc trưng đầu vào không thực hiện một cách tùy ý mà tuân theo bốn tiêu chí có cơ sở khoa học. Thứ nhất là tính liên quan vật lý - sinh thái, nghĩa là mỗi biến phải gắn với một cơ chế hình thành hoặc biến đổi của loại hình thoái hóa tương ứng (ví dụ chỉ số mặn và độ dẫn điện cho mặn hóa, chỉ số thực vật và độ ẩm cho suy giảm độ phì) nhằm bảo đảm khả năng diễn giải của mô hình. Thứ hai là khả năng quan trắc và tính sẵn có của dữ liệu, ưu tiên các biến trích xuất ổn định từ ảnh vệ tinh đa thời gian, mô hình số độ cao và dữ liệu khí hậu có độ phủ liên tục. Thứ ba là kiểm soát hiện tượng đa cộng tuyến giữa các biến phổ và chỉ số dẫn xuất nhằm tránh dư thừa thông tin và giảm rủi ro quá khớp. Thứ tư là mức độ đóng góp thực tế của biến vào kết quả dự báo, được đánh giá hậu nghiệm thông qua các phương pháp diễn giải mô hình như độ quan trọng đặc trưng và giá trị SHAP, trên cơ sở đó loại bỏ các biến ít ý nghĩa và giữ lại tập biến tối ưu cho từng loại hình thoái hóa.

Bảng 2. Ví dụ các nhóm đặc trưng đầu vào cho mô hình học máy, học sâu

TT	Nhóm đặc trưng	Mô tả và ví dụ biến điển hình
1	Kênh phổ vệ tinh	Sử dụng Sentinel-2: các kênh B2–B8A, SWIR (B11, B12); Red-edge hữu ích với thảm thực vật/cấu trúc lá - có quan hệ đồng biến với hàm lượng OM, N-P-K.
2	Chỉ số phổ	Đánh giá độ chua pHKCl: RVI, NDVI, SI1, SI2, SI3, SSSI, NDSI, BI, MSI, VH/VV...; Đánh giá hàm lượng chất hữu cơ tổng số (OM%): NDVI, NDHI, TVI, GNDVI, BI, BI2, CI, CI1, B2-B8, B8A, B11, B12...
3	Đặc trưng địa hình	DEM, độ dốc (slope), hướng dốc (aspect); chỉ số ẩm ướt địa hình (Topographic Wetness Index - TWI); chỉ số dòng chảy tích lũy...
4	Khí hậu và biến đổi khí hậu	Lượng mưa tích lũy; nhiệt độ; bốc thoát hơi (ET); độ ẩm; chỉ số khô hạn (SPI/SPEI); nhiệt độ trung bình năm; tổng lượng mưa trung bình năm; tần suất nhiệt độ cực đoan; tần suất mưa cực đoan...
5	Kinh tế - xã hội - nhân sinh	Biến gián tiếp: mật độ dân cư; mức độ đô thị hóa; chuyển đổi mục đích sử dụng đất (Land Use/Land Cover - LULC)...

2.4. Xây dựng và chuẩn hóa bộ mẫu huấn luyện

Quy trình xây dựng và chuẩn hóa bộ mẫu huấn luyện gồm các nội dung chính sau:

Tiêu chuẩn về mẫu: Một mẫu luôn có dạng tổng quát theo công thức (4):

$$Sample = \{ID, X, y, metadata\} \quad (4)$$

Trong đó: ID là mã mẫu; X là vector hoặc tensor (đại lượng nhiều chiều dùng để biểu diễn dữ liệu) đặc trưng; y là nhãn (giá trị phân tích mẫu thực đo); metadata là thông tin phụ (tọa độ, thời gian...).

Số lượng mẫu: không có ngưỡng tuyệt đối, tùy thuộc vào bài toán suy giải, mức độ đa dạng của loại đặc trưng, loại mô hình học máy/học sâu, mục tiêu đầu ra và mức độ chính xác mong muốn để định lượng số lượng mẫu.

Lựa chọn mẫu: trong mô hình học máy, học sâu có giám sát các mẫu đều có nhãn.

Chia tệp huấn luyện: có thể chia mẫu huấn luyện theo nhiều cách - chia theo dạng đầu vào của mẫu; chia theo đơn vị không gian của mẫu; chia theo dạng nhãn đầu ra; chia theo vai trò trong quá trình xây dựng mô hình.

Chuẩn hóa mẫu: các mẫu đưa vào mô hình có cấu trúc thống nhất, giá trị nhất quán, chất lượng tin cậy và khả năng so sánh được giữa các mẫu với nhau. Cần chuẩn hóa về nhãn đầu ra, về không gian, về thời gian, về giá trị số liệu, về chất lượng dữ liệu, về cách mã hóa dữ liệu, về phân bố bộ mẫu và theo yêu cầu của mô hình.

Việc chuẩn hóa toàn bộ quy trình tạo mẫu nhằm bảo đảm mô hình học từ dữ liệu đúng, sạch, đồng nhất và có khả năng khái quát hóa.

3. Kết quả nghiên cứu và thảo luận

3.1. Kết quả nghiên cứu

3.1.1. Đề xuất quy trình huấn luyện và đánh giá mô hình

Huấn luyện mô hình để đánh giá thoái hóa đất không chỉ là chọn thuật toán mà là một quy trình tổng thể gồm: thiết kế biến đầu vào có ý nghĩa vật lý, chuẩn hóa dữ liệu viễn thám, kiểm soát tự tương quan không gian, lựa chọn và đánh giá mô hình đúng cách, đồng thời đảm bảo khả năng tổng quát hóa và tuân thủ quy định chuyên ngành. Quy trình tổng thể được trình bày tại Hình 2.

Việc huấn luyện và đánh giá mô hình học máy phân loại vừa có phần chung vừa phụ thuộc vào từng loại mô hình (cách tối ưu, dữ liệu đầu vào, kỹ thuật đánh giá nâng cao); đặc biệt trong học máy không gian, cách đánh giá còn phụ thuộc mạnh vào đặc tính không gian của dữ liệu. Các nội dung đánh giá mô hình bao gồm:

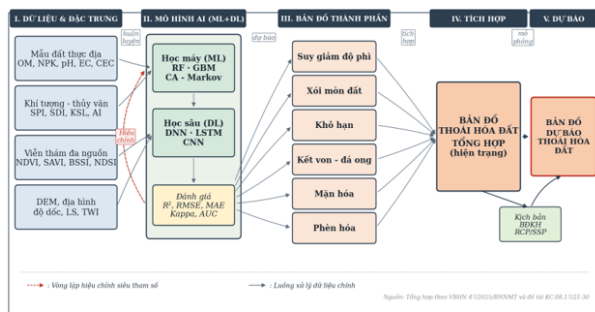
Các đại lượng định lượng đo lường hiệu suất của mô hình học máy: phụ thuộc vào loại mô hình học máy để chọn chỉ số đánh giá. Học máy hồi quy sử dụng chỉ số RMSE, MAE, R^2 ; học máy phân loại sử dụng chỉ số F1, Precision, Recall, Confusion Matrix; học máy phân đoạn sử dụng chỉ số mIoU, Boundary-F1.

Đánh giá theo không gian: có thể phân tích hoặc đánh giá riêng biệt cho từng vùng thay vì toàn bộ dữ liệu chung, áp dụng đánh giá chéo.

Phân tích sai số: vẽ biểu đồ phần dư theo một biến hoặc nhóm yếu tố cụ thể để kiểm tra sai số của mô hình có phân bố hợp lý hay không.

Diễn giải mô hình: kiểm tra mối quan hệ giữa các biến có phù hợp với quy luật vật lý hay không (ví dụ: OM tăng khi NDVI tăng thì hợp lý).

Trên cơ sở các đại lượng đánh giá nêu trên, việc lựa chọn mô hình cho từng loại hình thoái hóa dựa vào ba nhóm tiêu chí. Trước hết là hiệu suất dự báo ổn định trên tập kiểm tra độc lập, đo bằng RMSE, MAE và R^2 đối với bài toán hồi quy hoặc F1, độ chính xác tổng thể và ma trận nhầm lẫn đối với bài toán phân loại. Tiếp đến là khả năng diễn giải (XAI), bởi mô hình phục vụ quản lý nhà nước cần cung cấp được bằng chứng về vai trò của từng biến đầu vào thông qua độ quan trọng đặc trưng và giá trị SHAP, qua đó kiểm chứng tính hợp lý vật lý của kết quả. Cuối cùng là khả năng triển khai theo quy mô tỉnh - vùng - quốc gia với chi phí tính toán và yêu cầu dữ liệu phù hợp với điều kiện thực tế.



Hình 2. Sơ đồ khối của ứng dụng học máy, học sâu trong đánh giá thoái hóa đất hiện đại

Khả năng tổng quát hóa của mô hình được đặc biệt chú trọng do điều kiện tự nhiên của Việt Nam phân hóa mạnh giữa các vùng. Mô hình huấn luyện trên một vùng có thể suy giảm độ chính xác khi áp dụng sang vùng khác do hiện tượng dịch chuyển miền dữ liệu (domain shift) về thổ nhưỡng, khí hậu và tập quán canh tác. Để hạn chế rủi ro này, quy trình áp dụng kiểm định chéo theo không gian (spatial cross-validation) thay vì chia mẫu ngẫu nhiên thuần túy, đồng thời đánh giá độ ổn định của

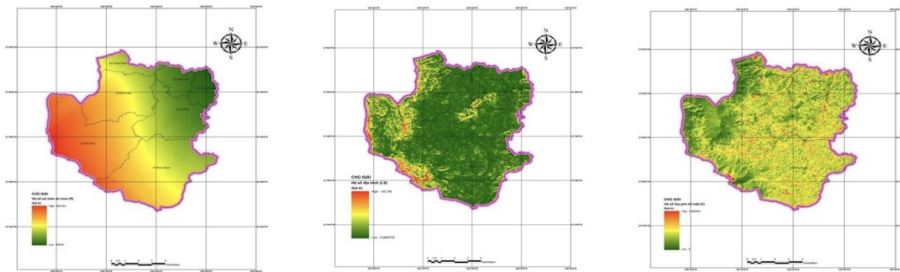
mô hình trên các vùng sinh thái khác nhau trước khi mở rộng phạm vi. Đây cũng là lý do bộ biên đầu vào ưu tiên các đặc trưng có ý nghĩa vật lý chung cho nhiều vùng nhằm nâng cao khả năng chuyển giao của mô hình.

3.1.2. Xây dựng bản đồ các chỉ số, hệ số

Bản đồ về các chỉ số, hệ số được xây dựng dựa trên lớp dữ liệu đầu ra của mô hình học máy, học sâu. Đầu ra thường là giá trị liên tục theo từng pixel, biểu thị bằng thang màu liên tục; khi thành lập bản đồ chuyên đề sẽ phân cấp thành các nhóm (mức độ thoái hóa). Nếu đầu ra là phân loại biểu thị bằng raster phân loại (Hình 3), sau đó sẽ được vector hóa theo phân vùng đồng nhất hoặc polygon hóa.

3.1.3. Xây dựng bản đồ thoái hóa đất theo loại hình và bản đồ thoái hóa đất tổng hợp

Sau khi có đủ dữ liệu về các chỉ số, hệ số sẽ thực hiện chồng xếp dữ liệu để thành lập bản đồ thoái hóa đất theo loại hình, gồm: bản đồ độ phì và suy giảm độ phì; bản đồ xói mòn đất do mưa; bản đồ khô hạn, hoang mạc hóa, sa mạc hóa; bản đồ kết von, đá ong hóa; bản đồ mặn hóa; bản đồ phèn hóa. Từ các bản đồ thoái hóa đất thành phần sẽ được tổng hợp theo phương pháp tổng hợp mức thoái hóa đất. Việc tổng hợp số liệu và lập báo cáo kết quả điều tra, đánh giá thoái hóa đất tuân thủ đúng quy định tại văn bản hợp nhất số 47/VBHN-BNNMT [2].



Hình 3. Ví dụ về các lớp dữ liệu đầu ra của mô hình học máy, học sâu ở định dạng Raster

3.1.4. Ví dụ minh họa quy trình đánh giá cho loại hình mặn hóa

Để minh họa cụ thể cách vận hành của quy trình đề xuất, phần này phân tích một ví dụ ứng dụng cho loại hình mặn hóa, dựa trên dữ liệu và kết quả đã công bố trong các nghiên cứu quốc tế chứ không phải số liệu thực nghiệm của nhóm tác giả. Mặn hóa được lựa chọn vì đây là loại hình có quan hệ phi tuyến rõ rệt giữa đặc trưng phổ và mức độ nhiễm mặn, đồng thời có nhiều bộ dữ liệu tham chiếu đáng tin cậy.

Về dữ liệu đầu vào, các nghiên cứu điển hình khai thác các kênh phổ vùng nhìn thấy, cận hồng ngoại và hồng ngoại sóng ngắn (VIS-NIR-SWIR), tập hợp các chỉ số mặn dẫn xuất, mô hình số độ cao (DEM) và các biến khí hậu phản ánh cân bằng nước. Biến mục tiêu (nhãn) là độ dẫn điện (EC) đo từ mẫu đất thực địa. Quy trình xử lý tuần tự gồm trích xuất đặc trưng từ ảnh, gán nhãn EC theo điểm mẫu, chia tập huấn luyện và kiểm tra theo nguyên tắc kiểm soát tự tương quan không gian, huấn luyện các mô hình ứng viên (RF, XGBoost, ANN) và đánh giá trên tập độc lập.

Kết quả tham chiếu cho thấy năng lực thực tế của cách tiếp cận này. Trên ảnh Landsat OLI với 13 chỉ số mặn và 241 mẫu đất, mô hình ANN đạt diện tích dưới đường cong (AUC) 0,921, cao hơn mô hình RF [7]. Trong một nghiên cứu khác sử dụng 26 biến môi trường trích từ ảnh Landsat-8, mô hình XGBoost vượt RF cả về hệ số xác định và sai số toàn phương trung bình, với R^2 đạt khoảng 0,83; đồng thời phân tích SHAP chỉ ra mức đóng góp của nhóm chỉ số thực vật và các kênh nhìn thấy thay đổi theo từng khu vực [9]. Ví dụ này cho thấy quy trình đề xuất không chỉ khả thi về mặt kỹ thuật mà còn cung cấp được cả kết quả định lượng lẫn cơ chế diễn giải, qua đó đáp ứng yêu cầu về độ tin cậy và tính minh bạch trong đánh giá thoái hóa đất.

3.2. Thảo luận

So sánh với phương pháp truyền thống được quy định tại văn bản hợp nhất số 47/VBHN-BNNMT [2], cách tiếp cận ứng dụng AI - viễn thám đề xuất trong nghiên cứu này mang lại các giá trị bổ sung quan trọng. Phương pháp truyền thống dựa trên nguyên tắc khoanh đất đồng nhất và ngưỡng phân cấp cố định, phù hợp với tính chuẩn mực pháp lý nhưng hạn chế trong việc phản ánh tính biến thiên không gian liên tục. Trong khi đó, phương pháp dựa trên dữ liệu cho phép tạo

ra bề mặt liên tục theo pixel, đồng thời tích hợp đa nguồn dữ liệu vệ tinh quang học, radar và biến điều kiện tự nhiên. Sự khác biệt giữa hai cách tiếp cận trên các tiêu chí cốt lõi được trình bày tại Bảng 3.

Bảng 3. So sánh phương pháp truyền thống và phương pháp tích hợp AI - viễn thám trong đánh giá thoái hóa đất

TT	Tiêu chí so sánh	Phương pháp truyền thống (VBHN 47/VBHN-BNNMT)	Phương pháp tích hợp AI - viễn thám
1	Đơn vị không gian	Khoanh theo đơn vị đất được coi là đồng nhất	Suy giải liên tục theo pixel, tạo bề mặt liên tục
2	Cơ sở phân cấp	Ngưỡng phân cấp cố định, mang tính kinh nghiệm	Học quan hệ phi tuyến, đa biến trực tiếp từ dữ liệu
3	Nguồn dữ liệu	Điều tra thực địa, lấy mẫu, phân tích phòng thí nghiệm	Tích hợp ảnh quang học, radar, điều kiện tự nhiên và mẫu đất
4	Độ phủ và tần suất cập nhật	Phụ thuộc mật độ điểm mẫu, cập nhật chậm	Phủ kín diện rộng, cập nhật theo ảnh đa thời gian
5	Cơ chế đánh giá độ tin cậy	Định tính theo cấp (cao - thấp, nặng - nhẹ)	Định lượng bằng RMSE, R^2 , AUC và độ chính xác tổng thể
6	Khả năng diễn giải	Dựa trên quy tắc và kinh nghiệm chuyên gia	Diễn giải bằng độ quan trọng đặc trưng và SHAP (XAI)
7	Tính chuẩn mực pháp lý	Tuân thủ trực tiếp quy định hiện hành	Cần chuẩn hóa quy trình để tích hợp vào khung pháp lý

Tổng hợp kinh nghiệm quốc tế cho thấy một quy luật tương đối nhất quán giữa các loại hình thoái hóa. Nhóm mô hình cây tăng cường gradient (XGBoost) và mô hình tổ hợp (RF) thường đạt hiệu suất cao và ổn định nhất khi làm việc với đầu vào đa nguồn, trong khi mạng nơ-ron nhân tạo (ANN) tỏ ra hiệu quả đối với bài toán nhiệm vụ vốn có quan hệ phi tuyến mạnh. Các bằng chứng định lượng tiêu biểu theo từng loại hình - từ diện tích dưới đường cong trong lập bản đồ mặn, hệ số xác định trong dự báo carbon hữu cơ đến độ chính xác tổng thể trong phân loại sa mạc hóa - đã được tổng hợp tại Bảng 1 và cho thấy mức hiệu suất đủ tin cậy để hỗ trợ công tác đánh giá thoái hóa đất. Một điểm chung khác là vai trò ngày càng được coi trọng của khả năng giải thích được (Explainable AI - XAI), khi phân tích SHAP giúp xác định nhóm biến phổ và chỉ số thực vật giữ vai trò quyết định trong từng bối cảnh cụ thể. Đối với điều kiện Việt Nam, việc lựa chọn mô hình do đó cần cân nhắc đồng thời giữa độ chính xác, tính giải thích được và khả năng triển khai theo quy mô tỉnh - vùng - quốc gia.

Hạn chế chính của hướng tiếp cận này là sự phụ thuộc vào chất lượng và mật độ mẫu nhân (mẫu phân tích đất), cũng như yêu cầu chuẩn hóa nghiêm ngặt về không gian, thời gian và phương pháp phân tích phòng thí nghiệm. Bên cạnh đó, các vùng có lớp phủ thực vật dày hoặc địa hình phức tạp có thể làm giảm khả năng suy giải của mô hình từ dữ liệu viễn thám. Do vậy, AI và viễn thám được đề xuất đóng vai trò công cụ hỗ trợ, không thay thế hoàn toàn điều tra thực địa.

4. Kết luận

Ứng dụng trí tuệ nhân tạo kết hợp viễn thám là hướng tiếp cận hiện đại và phù hợp cho đánh giá thoái hóa đất trong bối cảnh chuyển đổi số và hướng tiếp cận “dựa vào dữ liệu”. So với phương pháp truyền thống, cách tiếp cận này cho phép tích hợp đa nguồn dữ liệu, mô hình hóa các quan hệ phi tuyến, mở rộng phạm vi không gian, tăng tần suất cập nhật và tạo ra các lớp thông tin liên tục theo pixel. Đặc biệt, kết quả đánh giá thoái hóa đất đã tính đến tác động ảnh hưởng của đa nguồn đến quy mô, phân bố và mức độ của từng loại hình thoái hóa đất. Trên cơ sở dữ liệu ảnh vệ tinh, điều kiện tự nhiên và mẫu phân tích đất, các mô hình học máy, học sâu có thể suy giải nhiều chỉ số, hệ số liên quan đến 06 loại hình thoái hóa đất theo quy định.

AI và viễn thám không thay thế hoàn toàn điều tra thực địa mà đóng vai trò công cụ hỗ trợ, giúp

nâng cao hiệu quả điều tra, giám sát và xây dựng cơ sở dữ liệu thoái hóa đất theo hướng chi tiết, khách quan và cập nhật thường xuyên hơn. Trong thời gian tới, cần tiếp tục hoàn thiện cơ sở dữ liệu mẫu đất, chuẩn hóa quy trình xử lý dữ liệu và tăng cường kiểm định thực địa để nâng cao khả năng ứng dụng của các mô hình trong quản lý, giám sát tài nguyên đất.

Lời cảm ơn: Bài báo được công bố dựa trên kết quả thực hiện đề tài cấp Quốc gia mã số KC.08.17/21-30.

Cam kết của các tác giả

Tất cả các tác giả có tên trong bài báo cam kết sự đồng thuận và không có xung đột lợi ích trong công bố khoa học tại bài báo này.

Tài liệu tham khảo

- [1] Quốc hội nước Cộng hòa xã hội chủ nghĩa Việt Nam (2024), *Luật Đất đai số 31/2024/QH15* ngày 18/01/2024.
- [2] Bộ Nông nghiệp và Môi trường (2025), *Văn bản hợp nhất số 47/VBHN-BNNMT quy định kỹ thuật điều tra, đánh giá đất đai; kỹ thuật bảo vệ, cải tạo, phục hồi đất*.
- [3] Abdoli, P., Khanmirzaei, A., Hamzeh, S., Rezaei, S., & Moghimi, S. (2023), Use of remote sensing data to predict soil organic carbon in some agricultural soils of Iran, *Remote Sensing Applications: Society and Environment*, 30, 100969.
- [4] Odebiri, O., Odindi, J., & Mutanga, O. (2021), Basic and deep learning models in remote sensing of soil organic carbon estimation: A brief review, *International Journal of Applied Earth Observation and Geoinformation*, 102, 102389.
- [5] Diaz-Gonzalez, F. A., Vuelas, J., Correa, C. A., Vallejo, V. E., & Patino, D. (2022), Machine learning and remote sensing techniques applied to estimate soil indicators - Review, *Ecological Indicators*, 135, 108517. <https://doi.org/10.1016/j.ecolind.2021.108517>
- [6] Wadoux, A. M. J.-C., Minasny, B., & McBratney, A. B. (2020), Machine learning for digital soil mapping: Applications, challenges and suggested solutions, *Earth-Science Reviews*, 210, 103359. <https://doi.org/10.1016/j.earscirev.2020.103359>
- [7] Sarkar, S. K., Rudra, R. R., Sohan, A. R., Das, P. C., Ekram, K. M. M., Talukdar, S., Rahman, A., Alam, E., Islam, M. K., & Islam, A. R. M. T. (2023), Coupling of machine learning and remote sensing for soil salinity mapping in coastal area of Bangladesh, *Scientific Reports*, 13(1), 17056. <https://doi.org/10.1038/s41598-023-44132-4>
- [8] Kaplan, G., Gašparović, M., Alqasemi, A. S., Aldhaheri, A., Abuelgasim, A., & Ibrahim, M. (2023), Soil salinity prediction using machine learning and Sentinel-2 remote sensing data in hyper-arid areas, *Physics and Chemistry of the Earth, Parts A/B/C*, 130, 103400.
- [9] Aksoy, S., Sertel, E., Roscher, R., Tanik, A., & Hamzehpour, N. (2024), Assessment of soil salinity using explainable machine learning methods and Landsat 8 images, *International Journal of Applied Earth Observation and Geoinformation*, 130, 103879.
- [10] Amirgaliyev, Y., Mukhamediev, R., Merembayev, T., Kuchin, Y., Ataniyazova, A., & Omarova, P. (2024), Remote sensing and machine learning algorithms to predict soil salinity in southern Kazakhstan, *Discover Sustainability*, 5(1), 230.
- [11] Chaaou, A., Ait-Ichou, H., El Hachemy, S., Chikhaoui, M., Naimi, M., Hssaisoune, M., El Hafyani, M., Ait Brahim, Y., & Bouchaou, L. (2025), Mapping soil salinity using machine learning and remote sensing data in semi-arid croplands, *Frontiers in Soil Science*, 5, 1553400. <https://doi.org/10.3389/fsoil.2025.1553400>
- [12] Chen, D., Chang, N., Xiao, J., Zhou, Q., & Wu, W. (2019), Mapping dynamics of soil organic matter in croplands with MODIS data and machine learning algorithms, *Science of the Total Environment*, 669, 844-855. <https://doi.org/10.1016/j.scitotenv.2019.03.151>
- [13] Nguyen, T. T., Pham, T. D., Nguyen, C. T., Delfos, J., Archibald, R., Dang, K. B., Hoang, N. B., Guo, W., & Ngo, H. H. (2022), A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion, *Science of the Total Environment*, 804, 150187.
- [14] Zeng, P., Song, X., Yang, H., Wei, N., & Du, L. (2022), Digital soil mapping of soil organic matter with deep learning algorithms, *ISPRS International Journal of Geo-Information*, 11(5), 299. <https://doi.org/10.3390/ijgi11050299>
- [15] Ma, L., Zhao, L., Cao, L., Li, D., Chen, G., & Han, Y. (2022), Inversion of soil organic matter content based on improved convolutional neural network, *Sensors*, 22(20), 7777.

- <https://doi.org/10.3390/s22207777>
- [16] Dvorakova, K., Heiden, U., Pepers, K., Staats, G., van Os, G., & van Wesemael, B. (2023), Improving soil organic carbon predictions from a Sentinel-2 soil composite by assessing surface conditions and uncertainties, *Geoderma*, 429, 116128. <https://doi.org/10.1016/j.geoderma.2022.116128>
- [17] Geng, Y., Zhou, T., Zhang, Z., Cui, B., Sun, J., Zeng, L., Yang, R., Wu, N., Liu, T., Pan, J., Si, B., & Lausch, A. (2024), Continental-scale mapping of soil pH with SAR-optical fusion based on long-term earth observation data in Google Earth Engine, *Ecological Indicators*, 166, 112421. <https://doi.org/10.1016/j.ecolind.2024.112421>
- [18] Winzeler, H. E., Mancini, M., Blackstock, J. M., Libohova, Z., Owens, P. R., Ashworth, A. J., Miller, D. M., & Silva, S. H. G. (2024), Vegetation masking of remote sensing data aids machine learning for soil fertility prediction, *Remote Sensing*, 16(17), 3297.
- [19] Žížala, D., Juřicová, A., Zádorová, T., Zelenková, K., & Minařík, R. (2019), Mapping soil degradation using remote sensing data and ancillary data: South-East Moravia, Czech Republic, *European Journal of Remote Sensing*, 52(Suppl. 1), 108-122.
- [20] Hashemi, Z., Sodaieizadeh, H., Mokhtari, M. H., Hakimzadeh Ardakani, M. A., & Kamali Aliabadi, K. (2024), Monitoring and forecasting desertification and land degradation using remote sensing and machine learning techniques in Sistan plain, Iran, *Journal of African Earth Sciences*, 218, 105375. <https://doi.org/10.1016/j.jafrearsci.2024.105375>
- [21] Berdyayev, A., Al-Masnay, Y. A., Juliev, M., & Abuduwaili, J. (2024), Desertification monitoring using machine learning techniques with multiple indicators derived from Sentinel-2 in Turkmenistan, *Remote Sensing*, 16(23), 4525. <https://doi.org/10.3390/rs16234525>
- [22] Renard, K. G., Foster, G. R., Weesies, G. A., McCool, D. K., & Yoder, D. C. (1997), *Predicting soil erosion by water: A guide to conservation planning with the Revised Universal Soil Loss Equation (RUSLE)*, USDA Agriculture Handbook No. 703.