TRƯỜNG ĐẠI HỌC SÀI GÒN
TẠP CHÍ KHOA HỌC
TRƯỜNG ĐẠI HỌC SÀI GÒN
ISSN 1859-3208
e-ISSN 3093-3056
Số 103, Tập 01/2026

SAIGON UNIVERSITY
SCIENTIFIC JOURNAL
OF SAIGON UNIVERSITY
No. 103, Vol 01/2026

Email: tcdhsg@sgu.edu.vn ; Website: http://sj.sgu.edu.vn/

XÁC ĐỊNH TRỌNG SỐ CÁC NHÂN TỐ GÂY XÓI MÒN ĐẤT Ở KHU VỰC ĐỒI NÚI TỈNH NGHỆ AN

Quantifying the Relative Importance of Factors Controlling Soil Erosion in the Mountainous Region of Nghe An Province, Vietnam

Trần Thị Tuyến
Trường Đại học Sài Gòn

Tóm tắt

Xói mòn đất là một trong những quá trình gây suy thoái đất, mất độ phì và đe dọa phát triển bền vững ở khu vực đồi núi. Nghiên cứu này ứng dụng mô hình học máy XGBoost kết hợp với SHAP nhằm xác định trọng số các nhân tố tự nhiên và nhân sinh tác động đến xói mòn đất ở khu vực miền núi tỉnh Nghệ An. Kết quả cho thấy thảm phủ/sử dụng đất có hệ số cao nhất (0,12) và địa chất (0,10), các yếu tố có hệ số thấp hơn là độ dốc và độ cao (0,06), lượng mưa (0,05) và hướng sườn (0,05). Các biến khác như khoảng cách đến đứt gãy, NDVI và độ cong địa hình có hệ số thấp nhất (tương ứng là 0,03 và 0,02). Phân tích ma trận tương quan Pearson cho thấy hiện tượng đa cộng tuyến không nghiêm trọng, chứng minh rằng các biến đều cung cấp thông tin bổ sung cho mô hình. Kết quả nghiên cứu khẳng định tầm quan trọng của thảm phủ và nền địa chất trong kiểm soát quá trình xói mòn, đồng thời gợi ý các định hướng quản lý đất bền vững cho khu vực đồi núi Nghệ An.

***Từ khóa**: Xói mòn đất; XGBoost; SHAP; Học máy; Miền núi Nghệ An.*

Abstract

Soil erosion is one of the geomorphological - environmental processes that strongly affect ecosystems and sustainable development in mountainous regions. This study applies a machine learning model (XGBoost) in combination with the Shapley Additive Explanations (SHAP) method to evaluate the relative importance (weights) of natural and anthropogenic factors influencing soil erosion in the mountainous areas of Nghe An Province. The results reveal that Land Cover (0.12) and Geological conditions (0.10) are the most important factors, followed by Slope and Elevation (0.06), and Rainfall and Aspect (0.05). Other factors, such as distance to faults, NDVI, and curvature, show relatively lower impacts (0.02 - 0.03). These findings highlight the dominant roles of surface cover and geological conditions in controlling soil erosion risk, while also demonstrating the potential of SHAP for explaining predictive models of soil erosion hazard.

***Keywords:** Soil erosion; XGBoost; SHAP; Machine learning; Mountainous Nghe An.*

1. Đặt vấn đề

Xói mòn đất là một trong những nguyên nhân chính gây thoái hóa đất, làm suy giảm độ phì nhiêu, ảnh hưởng đến an ninh lương thực và gây mất ổn định hệ sinh thái trên toàn cầu (Pimentel, 2006). Ở Việt Nam, các khu vực đồi núi chiếm phần lớn diện tích lãnh thổ và có nguy cơ xói mòn nghiêm trọng do địa hình dốc, mưa nhiều, tập trung theo mùa và sử dụng đất chưa hợp lý (Nguyen et al., 2023). Nghệ An là tỉnh có trên 80% diện tích là đồi núi, hoạt

Tác giả liên hệ: tttuyen1982@sgu.edu.vn

động sản xuất nông - lâm nghiệp trên địa hình dốc phổ biến, phát triển hạ tầng diễn ra mạnh mẽ dẫn đến quá trình xói mòn diễn biến phức tạp và khó kiểm soát.

Các nghiên cứu trước đây đã khẳng định các nhóm nhân tố ảnh hưởng đến xói mòn bao gồm: (i) khí hậu, đặc biệt là lượng và cường độ mưa (Wischmeier & Smith, 1978); (ii) địa hình, gồm các yếu tố: độ dốc, chiều dài sườn, độ cong địa hình (Morgan, 2009) ; (iii) thảm phủ và hình thức canh tác (Morgan, 2009; Nguyễn Văn Khiết, 2014); (iv) đặc điểm đất và địa chất (Morgan, 2009; Guoguang et al., 2020); (v) Các hoạt động nhân sinh (sản xuất nông - lâm nghiệp, làm đường giao thông,…). Các mô hình thực nghiệm như RUSLE đã được sử dụng trong nhiều nghiên cứu nhưng khó giải thích vai trò tương đối của từng nhân tố (Ranzi et al., 2012). Gần đây, các mô hình học máy (machine learning) đã chứng minh ưu thế trong dự báo không gian xói mòn nhờ khả năng xử lý dữ liệu lớn và phi tuyến tính. Đặc biệt, phương pháp SHAP (Shapley Additive Explanations) cho phép giải thích minh bạch kết quả dự báo bằng cách lượng hóa đóng góp của từng biến đầu vào (Lundberg & Lee, 2017). Một số nghiên cứu gần đây tại Việt Nam cũng khẳng định hiệu quả của học máy trong lập bản đồ nguy cơ xói mòn (Nguyen et al., 2024). Xu hướng nghiên cứu hiện nay đang dịch chuyển từ các mô hình thực nghiệm truyền thống sang các mô hình học máy có khả năng giải thích cao, phù hợp cho việc quản lý rủi ro xói mòn đất ở các khu vực trên thế giới cũng như ở Việt Nam.

Tuy nhiên, hầu hết các nghiên cứu trong và ngoài nước mới chỉ tập trung vào việc lập bản đồ nguy cơ xói mòn hoặc so sánh độ chính xác giữa các mô hình dự báo, mà chưa đi sâu vào việc định lượng mức độ đóng góp tương đối của từng nhân tố tự nhiên và nhân sinh. Các công trình ứng dụng học máy trong nghiên cứu xói mòn đất tại Nghệ An còn hạn chế, dừng lại ở thử nghiệm một số mô hình như MLP, AdaBoost hay Gradient Boosting (Nguyen et al., 2024), trong khi các thuật toán mạnh như XGBoost kết hợp với phương pháp phân tích giá trị Shapley (SHAP) vẫn chưa được khai thác đầy đủ (Chen & Guestrin, 2016; Lundberg & Lee, 2017). Cho đến nay, trong nghiên cứu xói mòn đất, vẫn còn thiếu các nghiên cứu định lượng có khả năng làm rõ vai trò tương đối của từng nhân tố, đặc biệt trong điều kiện địa hình phức tạp ở miền núi Nghệ An. Nghiên cứu này bổ sung khoảng trống đó bằng việc ứng dụng mô hình học máy XGBoost kết hợp phương pháp giải thích SHAP nhằm định lượng và xếp hạng mức độ ảnh hưởng của các nhân tố tự nhiên và nhân sinh đến quá trình xói mòn đất. Sự kết hợp giữa XGBoost và SHAP không chỉ đảm bảo độ chính xác cao trong dự báo mà còn mang lại khả năng giải thích được vai trò của các yếu tố gây xói mòn, hỗ trợ cho việc ưu tiên kiểm soát tác động của từng yếu tố và quản lý rủi ro xói mòn ở khu vực nghiên cứu.

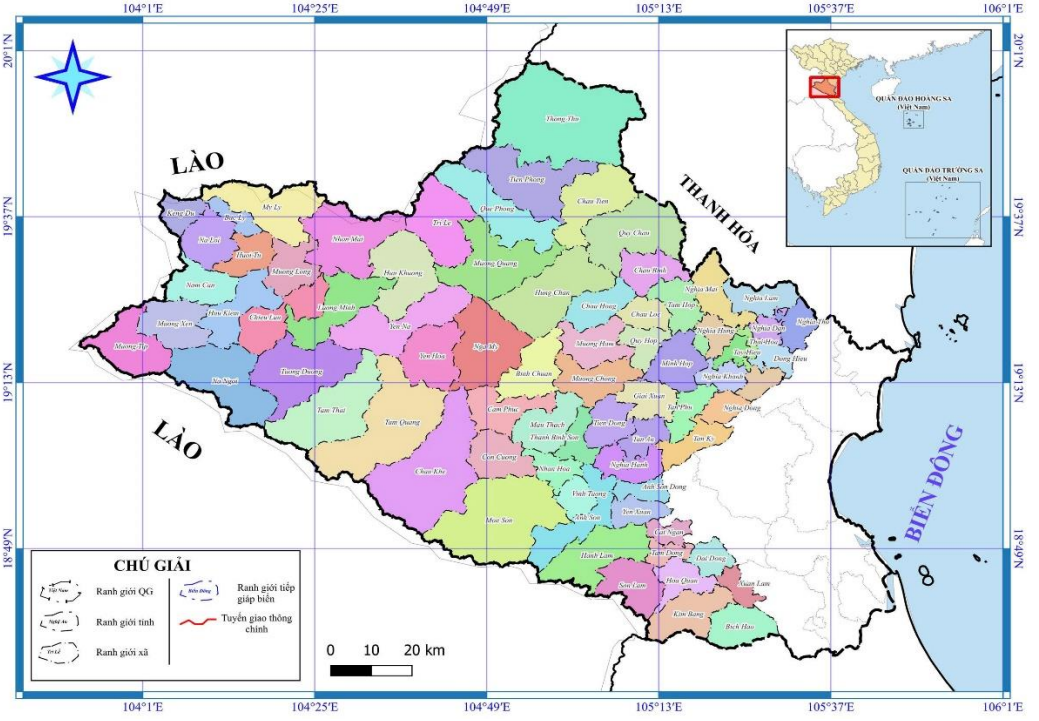
2. Dữ liệu và phương pháp nghiên cứu

2.1. Khái quát về khu vực nghiên cứu

Khu vực nghiên cứu nằm trên địa miền núi phía Tây tỉnh Nghệ An, thuộc Bắc Trung Bộ. Đây là vùng chuyển tiếp giữa vùng núi Trường Sơn Bắc xuống đồng bằng ven biển, có địa hình phân hóa phức tạp, chia cắt mạnh và độ dốc lớn. Địa hình khu vực chủ yếu là đồi núi trung bình và thấp, độ cao biến thiên từ 50 m đến trên 1.500 m, hướng nghiêng chung từ Tây sang Đông. Các dãy núi xen kẽ thung lũng hẹp, sườn dốc phổ biến từ 15° đến 35°, nhiều nơi vượt quá 40°. Về địa chất, khu vực nghiên cứu chủ yếu được cấu tạo bởi các đá biến chất, đá phiến sét, đá vôi, cát kết và granit có mức độ phong hóa khác nhau. Các thành tạo đá mềm yếu, dễ phong hóa như đá phiến sét và sa

phiến thạch phân bố rộng, là yếu tố nền quan trọng làm gia tăng nguy cơ xói mòn. Các loại đất được tạo thành chủ yếu thuộc nhóm đất đất feralit đỏ vàng hình thành trên đá axit và đất xám bạc màu. Tầng phong hóa ở các sườn dốc thường dày nhưng kết cấu rời rạc, dễ bị phá vỡ và cuốn trôi trong điều kiện mưa lớn hoặc khi thảm phủ suy giảm. Vì vậy, khu vực này thường xuyên xảy ra xói mòn đất và trượt lở sườn dốc.

Khí hậu khu vực nghiên cứu mang đặc trưng của vùng nhiệt đới gió mùa, thể hiện rõ sự phân hóa lượng mưa theo mùa và theo địa hình, đồng thời chịu tác động của gió mùa Tây Nam khô nóng. Nhiệt độ trung bình năm dao động từ 22 - 25°C, lượng mưa trung bình đạt 1.800 - 2.500 mm/năm, mưa tập trung chủ yếu từ tháng 6 đến tháng 10. Cường độ mưa cao kết hợp với địa hình dốc và bề mặt chia cắt mạnh đã tạo điều kiện thuận lợi cho quá trình xói mòn đất xảy ra mạnh trong mùa mưa. Thêm vào đó, mạng lưới sông suối dày đặc, bao gồm các nhánh chính của sông Cả, sông Hiếu và sông Giăng có dạng lưu vực hẹp, lòng sông dốc và tốc độ dòng chảy lớn dễ gây xói mòn. Trong khi đó, kết quả phân tích ảnh Sentinel-2 năm 2023 cho thấy giá trị NDVI trung bình toàn khu vực đạt 0,52, phản ánh mức độ che phủ trung bình. Tuy nhiên, các khu vực có thảm phủ thấp (NDVI < 0,4) chiếm khoảng 21,4% diện tích nghiên cứu, tập trung chủ yếu tại các huyện Quế Phong, Quỳ Hợp, Nghĩa Đàn và Tân Kỳ. Đây cũng là những nơi có độ dốc trung bình trên 25°, lượng mưa hằng năm cao trên 2.000 mm và nền địa chất yếu (đá phiến sét, cát kết phong hóa), khiến nguy cơ xói mòn đất đặc biệt nghiêm trọng (Nguyen et al., 2024). Như vậy, mặc dù tỷ lệ che phủ rừng của Nghệ An ở mức cao, nhưng sự suy giảm rừng tự nhiên và phân bố không đồng đều của thảm phủ đã tạo nên sự khác biệt lớn về mức độ xói mòn đất giữa các khu vực, đặc biệt tại những nơi có địa hình dốc, mưa lớn và lớp phủ thưa thớt.



Hình 1. Bản đồ khu vực nghiên cứu

2.2. Dữ liệu sử dụng

Các lớp dữ liệu đầu vào được xây dựng từ nhiều nguồn khác nhau, bao gồm:

(1) Lượng mưa trung bình năm và cường độ mưa cực đoan (thu thập từ số liệu khí tượng thủy văn và bộ dữ liệu khí hậu toàn cầu);

(2) Các yếu tố địa hình như độ dốc, độ dài sườn, độ cao tuyệt đối và chỉ số cong địa hình (curvature) được trích xuất từ mô hình số độ cao (DEM);

(3) Thảm phủ và sử dụng đất được xác định thông qua phân tích bản đồ hiện trạng sử dụng đất (LULC);

(4) Chỉ số NDVI được tính toán từ ảnh viễn thám Sentinel 2;

(5) Đặc điểm địa chất (thạch học) và thổ nhưỡng (các loại đất theo thành phần đá mẹ) kèm theo thông tin về thành phần cơ giới, độ dày tầng đất mặt và loại đá mẹ, được khai thác từ bản đồ đất và địa chất tỷ lệ 1:50.000 (Sở Tài nguyên và Môi trường tỉnh Nghệ An, 2024).

Tập dữ liệu huấn luyện được xây dựng bằng cách tích hợp các lớp dữ liệu không gian theo đơn vị pixel/lưới phân tích, trong đó: Biến đầu vào gồm các nhân tố khí hậu, địa hình, thảm phủ, NDVI và địa chất; Biến đầu ra là mức độ/nguy cơ xói mòn đất (được xác định từ dữ liệu thực nghiệm/bản đồ xói mòn). Dữ liệu được chuẩn hóa và kiểm tra tính nhất quán, loại bỏ giá trị thiếu và ngoại lai; Dữ liệu hiện trạng gồm điểm xói mòn được chia thành tập huấn luyện và tập kiểm định theo tỷ lệ phù hợp, tương ứng là 70:30 (Pham et al., 2026). Quy trình này đảm bảo dữ liệu đầu vào có độ tin cậy cao và phù hợp cho việc huấn luyện mô hình học máy.

2.3. Phương pháp sử dụng

- Mô hình XGBoost và SHAP

Mô hình XGBoost

Trong nghiên cứu này, mô hình XGBoost được sử dụng để dự báo nguy cơ xói mòn đất dựa trên các yếu tố gây xói mòn, gồm: lượng mưa, độ dốc, thảm phủ, NDVI và đặc điểm đất. Mô hình được xây dựng theo cách học từng bước, trong đó mỗi bước sẽ cải thiện kết quả dự báo bằng cách giảm sai số còn lại từ bước trước. XGBoost có khả năng xử lý tốt các mối quan hệ phức tạp giữa các yếu tố đầu vào, ví dụ như sự kết hợp giữa độ dốc, lượng mưa và thảm phủ trong việc ảnh hưởng đến xói mòn đất (Chen & Guestrin, 2016). Dữ liệu được chia thành tập huấn luyện và tập kiểm định theo tỉ lệ 70:30 để xây dựng và đánh giá mô hình. Kết quả dự báo từ XGBoost sau đó được sử dụng để phân tích vai trò của từng yếu tố thông qua phương pháp SHAP, từ đó làm rõ mức độ ảnh hưởng của các nhân tố đến nguy cơ xói mòn đất.

Mô hình SHAP

Trong nghiên cứu này, phương pháp SHAP được áp dụng nhằm làm rõ mức độ đóng góp của từng nhân tố gây xói mòn vào kết quả dự báo nguy cơ xói mòn đất của mô hình XGBoost. Dựa trên nền tảng của lý thuyết trò chơi hợp tác, mỗi biến đầu vào được xem như một "tác nhân" tham gia vào quá trình hình thành giá trị dự báo, trong đó giá trị Shapley phản ánh phần đóng góp riêng biệt của từng nhân tố đối với kết quả cuối cùng (Lundberg & Lee, 2017). Liu et al. (2019) cũng cho thấy xói mòn đất có thể thúc đẩy sự vận chuyển ngang nhanh của các nguyên tố vết trong các vùng karst, nhấn mạnh thêm ý nghĩa môi trường của quá trình xói mòn. Giá trị Shapley được tính toán dựa trên nhiều tổ hợp khác nhau của các biến đầu vào, cho phép lượng hóa mức độ đóng góp riêng của từng yếu tố, từ đó xác định tầm quan trọng tương đối giữa các nhân tố gây xói mòn. Cách tiếp cận này đặc biệt phù hợp trong bối cảnh các yếu tố

như độ dốc, lượng mưa, thảm phủ và địa chất có mối quan hệ phi tuyến và tương tác phức tạp. Trên cơ sở đó, giá trị SHAP trung bình tuyệt đối của từng biến được sử dụng để xếp hạng mức độ ảnh hưởng và xác định trọng số của các nhân tố trong mô hình (Nohara et al., 2019).

Việc tích hợp SHAP với mô hình XGBoost không chỉ nâng cao khả năng dự báo mà còn giúp giải thích rõ cơ chế tác động của từng nhân tố đến xói mòn đất. Điều này góp phần khắc phục hạn chế của các mô hình học máy dạng "hộp đen", đồng thời cung cấp cơ sở khoa học đáng tin cậy cho việc xác định ưu tiên quản lý và đề xuất các giải pháp giảm thiểu xói mòn phù hợp với điều kiện tự nhiên của khu vực nghiên cứu. Thông qua việc tính toán các giá trị Shapley, phương pháp này định lượng sự đóng góp biên của mỗi biến bằng cách so sánh đầu ra của mô hình trên tất cả các tổ hợp có thể có của các tập con đặc trưng. Khung lý thuyết này đảm bảo tính công bằng trong việc phân bổ và cung cấp một thước đo khách quan về cách các biến riêng lẻ ảnh hưởng đến hành vi dự đoán (Lundberg & Lee, 2017).

- *Phân tích tương quan giữa các nhân tố*

Để đánh giá mối quan hệ tuyến tính giữa các nhân tố, nghiên cứu tính toán hệ số tương quan Pearson:

$$r_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$$

trong đó Cov(X,Y) là hiệp phương sai, σX, σY là độ lệch chuẩn. Giá trị r nằm trong [-1; 1]. Đây là thước đo thống kê phản ánh mức độ và chiều hướng quan hệ tuyến tính giữa hai biến. Giá trị của r dao động từ -1 đến +1: Nếu r>0: mối quan hệ thuận chiều (một biến tăng thì biến kia cũng tăng); nếu r<0: mối quan hệ nghịch chiều (một biến tăng thì biến kia giảm); nếu r càng gần 1 thì mối quan hệ tuyến tính càng chặt chẽ; và r gần 0 cho thấy mối quan hệ yếu hoặc không đáng kể.

3. Kết quả nghiên cứu

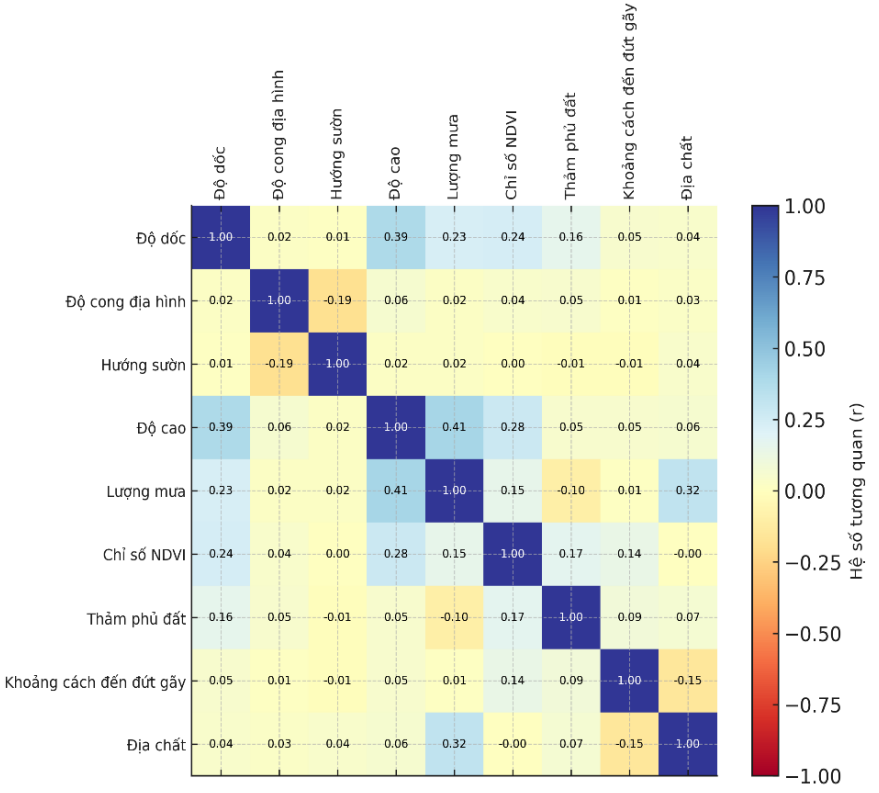
3.1. Phân tích tương quan giữa các nhân tố gây xói mòn đất ở miền núi tỉnh Nghệ An

Kết quả phân tích hệ số tương quan Pearson giữa các biến đầu vào được trình bày trong Hình 2. Ma trận này cho thấy mối quan hệ tuyến tính giữa các nhân tố tự nhiên và nhân sinh ảnh hưởng đến xói mòn đất tại khu vực nghiên cứu. Nhìn chung, các giá trị hệ số tương quan |r| đều nhỏ hơn 0,5 phản ánh rằng hiện tượng đa cộng tuyến giữa các biến là không đáng kể. Điều này cho phép sử dụng toàn bộ tập biến đầu vào trong mô hình học máy mà không gây sai lệch trong quá trình huấn luyện.

Trong số các cặp biến, độ dốc (Slope) và độ cao (Elevation) thể hiện mối tương quan thuận rõ rệt (r ≈ 0,39), phù hợp với đặc điểm địa hình vùng núi Nghệ An, nơi các khu vực có cao độ lớn thường đi kèm với sườn dốc cao. Bên cạnh đó, lượng mưa (Rainfall) và độ cao (Elevation) có tương quan trung bình (r ≈ 0,41), cho thấy ảnh hưởng của quy luật địa hình đến phân bố mưa: các vùng cao thường có lượng mưa lớn hơn do tác động nâng của khối khí ẩm. Một số mối liên hệ yếu khác bao gồm NDVI - Elevation (r ≈ 0,28) và Geological - Rainfall (r ≈ 0,32), phản ánh tác động gián tiếp của điều kiện địa hình và khí hậu đến thảm thực vật và đặc điểm phong hóa đá mẹ. Ngược lại, các biến như độ cong địa hình (Curvature), hướng sườn (Aspect) và khoảng cách đến đứt gãy (Distance to fsaults) có hệ số tương quan rất thấp (|r| < 0,1), chứng tỏ tính độc lập cao và ít bị chi phối bởi các yếu tố khác.

Như vậy, ma trận tương quan khẳng

định rằng các biến đầu vào có mức độ độc lập tương đối, đảm bảo tính tin cậy của mô hình XGBoost trong việc xác định vai trò riêng của từng nhân tố.

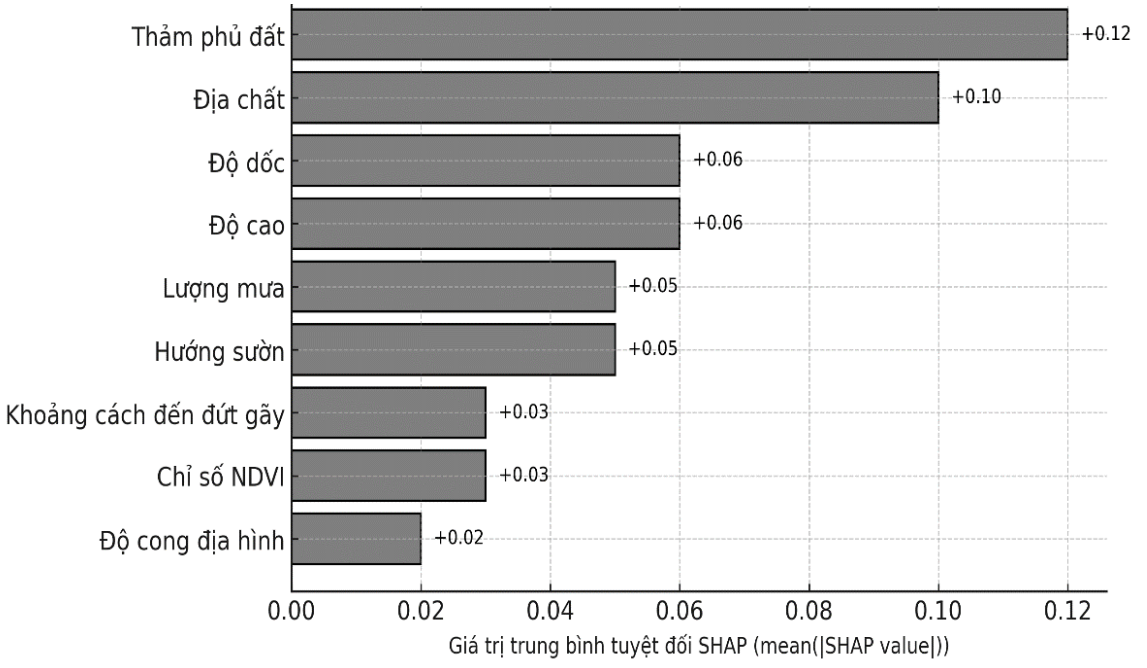


Hình 2. Ma trận tương quan giữa các yếu tố gây xói mòn đất ở đồi núi tỉnh Nghệ An

3.2. Phân tích vai trò các nhân tố theo phương pháp SHAP

Kết quả phân tích SHAP (Hình 3) cho phép định lượng mức độ đóng góp của từng biến đầu vào vào kết quả dự báo xói mòn đất. Trong số các biến, thảm phủ/sử dụng đất có giá trị trung bình tuyệt đối của SHAP cao nhất (0,12), thể hiện vai trò chi phối của lớp phủ bề mặt trong việc điều tiết dòng chảy, giữ đất và giảm xói mòn. Kết quả này phản ánh thực tế rằng các khu vực được che phủ bởi rừng tự nhiên hoặc cây trồng lâu năm có mức xói mòn thấp hơn đáng kể so với vùng đất trống, nương rẫy hoặc xây dựng. Yếu tố địa chất đứng thứ hai (0,10), cho thấy tác động nền tảng của loại đá mẹ đến khả năng chống xói mòn. Các khu vực có đá trầm tích phong hóa mạnh, đất sét hoặc đá phiến dễ bị rửa trôi hơn so với vùng có đá cứng. Các biến độ dốc và độ cao có giá trị bằng nhau (0,06) phản ánh ảnh hưởng của yếu tố địa hình. Độ dốc cao làm gia tăng vận tốc dòng chảy bề mặt và năng lượng phá hủy, trong khi độ cao có liên hệ đến cả chế độ mưa và quá trình rửa trôi. Lượng mưa (0,05) đóng vai trò yếu tố kích hoạt, quyết định cường độ dòng chảy và năng lượng tách hạt đất. Hướng sườn (0,05) tác động gián tiếp đến sự ổn định lớp phủ thông qua sự khác biệt về chiếu xạ mặt trời và độ ẩm. Các yếu tố khoảng cách đến đứt gãy (0,03) và NDVI (0,03) có vai trò hỗ trợ. NDVI phản ánh mật độ che phủ tại thời điểm quan sát, trong khi lớp phủ biểu thị

trạng thái sử dụng đất lâu dài, nên kết quả SHAP thấp hơn là hợp lý. Độ cong địa hình có giá trị nhỏ nhất (0,02), cho thấy yếu tố này ít ảnh hưởng trong điều kiện địa hình trung bình đến lớn, và có thể chỉ quan trọng ở quy mô vi địa hình.



Hình 3. Trọng số các nhân tố gây xói mòn đất trong mô hình SHAP

Như vậy, mô hình SHAP chỉ ra rằng thảm phủ bề mặt và nền địa chất là hai nhân tố chi phối chính đến nguy cơ xói mòn, trong khi nhóm địa hình và khí hậu đóng vai trò điều kiện thúc đẩy. Kết quả của nghiên cứu phù hợp với nhiều công trình trước đây về xói mòn đất. Theo Pimentel (2006) và Morgan (2009), thảm phủ thực vật là yếu tố quan trọng nhất trong việc kiểm soát xói mòn, bởi nó giúp bảo vệ bề mặt đất và duy trì cấu trúc đất. Phát hiện rằng thảm phủ có trọng số cao nhất trong mô hình SHAP củng cố luận điểm này, đồng thời nhấn mạnh tầm quan trọng của công tác phục hồi rừng và che phủ đất tại khu vực miền núi Nghệ An. Tương tự, yếu tố địa chất (Geological) được xác định là có vai trò nền tảng, phù hợp với các nghiên cứu quốc tế như của Guoguang et al. (2020), trong đó loại đá mẹ và mức độ phong hóa được xem là nguyên nhân chính dẫn đến sự khác biệt về cường độ xói mòn giữa các khu vực. Nhóm yếu tố địa hình và khí hậu có tầm quan trọng trung bình, nhưng đóng vai trò kích hoạt và điều kiện hóa. Lượng mưa và độ dốc là hai biến chủ chốt trong mô hình RUSLE và các mô hình học máy hiện đại, điều này tương đồng với kết quả của một số nghiên cứu đã công bố (Wischmeier and Smith, 1978; Nguyen et al., 2024).

Bảng 1. Giá trị trung bình tuyệt đối SHAP vai trò của các yếu tố gây xói mòn

STT	Yếu tố	Giá trị trung bình tuyệt đối SHAP	Xếp hạng tầm quan trọng	Diễn giải vai trò chính
1	Thảm phủ	0,12	1	Yếu tố chi phối, quyết định khả năng che phủ và bảo vệ đất trước dòng chảy bề mặt.
2	Địa chất	0,10	2	Quyết định tính chất vật lý – cơ học của đất, khả năng phong hóa và chống xói mòn.
3	Độ dốc	0,06	3	Làm tăng tốc độ dòng chảy và năng lượng phá hủy của nước mưa.
4	Độ cao	0,06	4	Liên quan đến chế độ mưa và phân bố địa hình, ảnh hưởng gián tiếp đến xói mòn.
5	Lượng mưa	0,05	5	Yếu tố kích hoạt, điều khiển cường độ dòng chảy và rửa trôi đất.
6	Hướng sườn	0,05	6	Tác động đến độ ẩm và bức xạ, qua đó ảnh hưởng đến lớp phủ thực vật.
7	Khoảng cách đến đứt gãy	0,03	7	Khu vực gần đứt gãy thường có cấu trúc đất yếu, dễ bị rửa trôi hơn.
8	Chỉ số NDVI	0,03	8	Phản ánh mật độ che phủ tại thời điểm quan sát, có tương quan cao với Land Cover.
9	Độ cong địa hình	0,02	9	Ảnh hưởng nhỏ, chỉ quan trọng ở vi địa hình hoặc vùng tích tụ – rửa trôi cục bộ.

Kết quả của hai phương pháp Pearson và SHAP bổ trợ lẫn nhau: Pearson khẳng định các biến có tính độc lập cao, còn SHAP định lượng mức ảnh hưởng riêng biệt. Những phát hiện này góp phần làm rõ cơ chế tương tác giữa các nhân tố tự nhiên và nhân sinh, đồng thời cung cấp bằng chứng định lượng đầu tiên về trọng số tương đối của các nhân tố gây xói mòn đất tại miền núi Nghệ An. Kết quả nghiên cứu có ý nghĩa thực tiễn quan trọng, giúp các nhà quản lý xác định ưu tiên phục hồi thảm phủ, quy hoạch sử dụng đất hợp lý và kiểm soát hoạt động nhân sinh tại các vùng có nền địa chất yếu hoặc độ dốc cao. Điều này gợi ý rằng trong công tác quản lý và quy hoạch, cần đặc biệt chú trọng đến việc quản lý sử dụng đất hợp lí và bảo vệ rừng ở

khu vực địa hình dốc, đồng thời kết hợp đánh giá đặc điểm địa chất và chế độ mưa để xây dựng các biện pháp giảm thiểu rủi ro hiệu quả.

4. Kết luận

Nghiên cứu đã ứng dụng mô hình học máy XGBoost kết hợp với phương pháp SHAP để phân tích và lượng hóa mức độ ảnh hưởng của các nhân tố chính đến nguy cơ xói mòn đất tại khu vực đồi núi tỉnh Nghệ An. Kết quả cho thấy có chín nhân tố tác động đến xói mòn đất, trong đó lớp phủ đất và đặc điểm địa chất là hai yếu tố quan trọng nhất, tiếp theo là độ dốc, cao độ và lượng mưa. Các yếu tố khác như hướng sườn, khoảng cách đến đứt gãy, NDVI và độ cong địa hình cũng có vai trò nhất định nhưng ở mức thấp hơn. Phân tích hệ số tương quan Pearson cho thấy mối quan hệ tuyến tính đáng kể giữa một số yếu tố địa hình (độ dốc, cao độ) và điều kiện khí hậu (lượng mưa) với hiện tượng xói mòn, củng cố tính hợp lý của kết quả mô hình.

Việc kết hợp giữa XGBoost và SHAP không chỉ giúp đánh giá chính xác nguy cơ xói mòn đất mà còn làm rõ đóng góp riêng lẻ của từng nhân tố, cung cấp cơ sở khoa học quan trọng cho công tác quản lý và quy hoạch sử dụng đất bền vững ở Nghệ An và khu vực Bắc Trung Bộ. Những phát hiện này có thể hỗ trợ các nhà quản lý địa phương trong việc ưu tiên giải pháp bảo vệ đất, đặc biệt là các biện pháp duy trì và phục hồi lớp phủ rừng cũng như kiểm soát các hoạt động nhân sinh trên nền địa chất và địa hình nhạy cảm. Mặc dù nghiên cứu đã đạt được những kết quả quan trọng trong việc lượng hóa tác động của các nhân tố đến xói mòn đất bằng mô hình XGBoost kết hợp với SHAP, nhưng mới tập trung vào các yếu tố tự nhiên cơ bản (địa hình, thảm phủ, khí hậu, địa chất) mà chưa tích hợp sâu các nhân tố nhân sinh như hoạt động canh tác nông nghiệp, khai thác rừng hoặc xây dựng hạ tầng. Trong các nghiên cứu tiếp theo, các hoạt động nhân sinh sẽ được khảo sát và phân tích cụ thể hơn.

Lời cảm ơn

Bài báo là sản phẩm của đề tài nghiên cứu mã số CS.2025.A4.049 được tài trợ bởi Trường Đại học Sài Gòn

TÀI LIỆU THAM KHẢO

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Guoguang, C., Hongying, L., Jinquan, C., Xiaodong, Z. H. A. N. G., & Long, Z. (2020). Geological influence factors of soil erosion in Changting County, Fujian Province and the countermeasures to prevent and control. *Hydrogeology & Engineering Geology*, *47*(6), 26-35.

Liu, M., Zhang, Q., Ge, S., Mason, R. P., Luo, Y., He, Y., Xie, H., Sa, R., Chen, L., & Wang, X. (2019). Rapid increase in the lateral transport of trace elements induced by soil erosion in major karst regions in China. *Environmental Science & Technology*, *53*(8), 4206-4214. https://doi.org/10.1021/acs.est.8b06143

Lundberg, S. M., & Lee, S.I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30* (pp. 4765-4774). Curran Associates, Inc.

Morgan, R. P. C. (2009). *Soil erosion and conservation*. John Wiley & Sons.

Nguyen, C. Q., Tran, T. T., Nguyen, T. T. T., Nguyen, T. H. T., Astarkhanova, T. S., Van Vu, L., Dau, K. T., Nguyen, H. N., Pham, G. H., Nguyen, D. D., Prakash, I., & Pham, B. (2024). *Mapping of soil erosion susceptibility using advanced machine learning models at Nghe An, Vietnam.* Journal of Hydroinformatics, *26(1),* 72-87.

Nguyen, T. T. H., Tran, T. T., Sarzhanovna, A. T., Thuy, H. T., Luong, V. V., Du, T. D., Tai, D. K., The, H. A., Thanh, N. N., Duong, P. T., Ha, V. T. T., & Khanh, V. T. N. (2023). Potential risks of soil erosion in North-Central Vietnam using remote sensing and GIS. *Revista Brasileira de Engenharia Agrícola e Ambiental, 27*(11), 910-916. https://doi.org/10.1590/1807-1929/agriambi.v27n11p910-916

Nguyễn Văn Khiết. (2014). Nghiên cứu xác định vai trò của một số yếu tố liên quan đến xói mòn đất ở nước ta. *Tạp chí Khoa học Lâm nghiệp*, (*1),* 3145-3153.

Nohara, Y., Matsumoto, K., Soejima, H., & Nakashima, N. (2019). Explanation of machine learning models using improved Shapley additive explanation. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics* (p. 546). Association for Computing Machinery.

Pham, G. H., Pham, B. T. T., Hoang, K. O. T., Tran, T. T., & Hoang Nguyen Duc Chi. (2026). GIS based soil erosion susceptibility assessment using deep learning models: A case study in the mountainous region of Nghe An, Vietnam. *Journal of Science and Transport Technology*, *6*(1), 67-86.

Pimentel, D. (2006). Soil erosion: A food and environmental threat. *Environment, Development and Sustainability, 8(*1), 119-137.

Ranzi, R., Le, T. H., & Rulli, M. C. (2012). A RUSLE approach to model suspended sediment load in the Lo River (Vietnam): Effects of reservoirs and land use changes. *Journal of Hydrology, 422-423*, 17-29.

Wischmeier, W. H., & Smith, D. D. (1978). *Predicting rainfall erosion losses: A guide to conservation planning* (Agriculture Handbook No. 537). U.S. Department of Agriculture, Science and Education Administration.

Ngày nhận bài: 06/10/2025 Ngày chấp nhận đăng: 05/8/2026