

Development of decided tree model and alternating decision tree models for spatial prediction of landslides at Muong Nhe district, Dien Bien Province, Vietnam

Nguyen Duc Dam^{1*}, Le Van Hiep¹, Nguyen Thanh Tuan¹, Tran Van Phong²,
Binh Thai Pham¹

¹University of Transport Technology, Hanoi 100000, Vietnam

²Institute of Geological Sciences, Hanoi 100000, Vietnam

Article info

Type of article:

Original research paper

*Corresponding author:

E-mail address:

damnd@utt.edu.vn

Received:

November 30, 2021

Accepted:

January 03, 2022

Published:

March 11, 2022

Abstract: Landslide spatial prediction map is a useful tool for effective management of landslides of an area. In this study, we adopted machine approach based on alternating decision tree (ADT) and decision tree (DT) to predict landslides spatially at Muong district, Dien Bien Province, Vietnam. In this study, a total of 159 landslide past landslides were identified and 12 conditioning factors including: slope, aspect, curvature, elevation, distance to faults, NDVI, flow accumulation, TWI, SPI, geological, distance to rivers, distance to roads were used to build spatial database for landslide spatial modeling. Validation and comparison of the models was done using statistical indicators including ROC / AUC curves. The results show that performance of both ADT and DT is good for spatial prediction of landslides but ADT model (AUC_{train} = 0.928, AUC_{test} = 0.887) is better than DT model (AUC_{train} = 0.915, AUC_{test} = 0.800). Map of landslide spatial prediction at Muong Nhe District can be used in land use planning for better landslide hazard management.

Keywords: Landslide, Alternating Decision Tree, Decision Tree, Dien Bien, Viet Nam, Landslide susceptibility map.

Phát triển mô hình học máy cây quyết định và cây quyết định xen kẽ thành lập bản đồ dự báo không gian sạt lở đất tại huyện Mường Nhé, tỉnh Điện Biên, Việt Nam

Nguyễn Đức Đảm¹, Lê Văn Hiệp¹, Nguyễn Thanh Tuấn¹, Trần Văn Phong², Phạm Thái Bình¹

¹Trường Đại học Công nghệ GTVT, Hà Nội, Việt Nam

²Viện Địa chất, Viện Hàn lâm Khoa học và Công nghệ Việt Nam, Hà Nội, Việt Nam

Thông tin bài viết

Dạng bài viết:

Bài báo khoa học

*Tác giả liên hệ:

Địa chỉ E-mail:
damnd@utt.edu.vn

Ngày nộp bài:

30/11/2021

Ngày chấp nhận:

03/01/2022

Ngày đăng bài:

11/3/2022

Tóm tắt: Bản đồ dự báo không gian sạt lở đất (Bản đồ nhạy cảm sạt lở đất) là một công cụ hữu ích để quản lý hiệu quả sạt lở đất của một khu vực. Trong nghiên cứu này, chúng tôi áp dụng phương pháp tiếp cận máy học dựa trên hai thuật toán cây quyết định xen kẽ (ADT) và cây quyết định (DT) để lập bản đồ dự báo không gian sạt lở đất tại huyện Mường Nhé, tỉnh Điện Biên. Trong nghiên cứu này, 159 vị trí sạt lở đã được xác định và 12 yếu tố ảnh hưởng bao gồm: góc mái dốc, hướng mái dốc, hình dáng bề mặt địa hình, độ cao địa hình, khoảng cách đứt gãy, bao phủ thực vật (NDVI), tích lũy dòng chảy, độ ẩm địa hình (TWI), sức mạnh dòng chảy (SPI), địa chất, khoảng cách đến sông suối, khoảng cách đến đường giao thông đã được sử dụng để xây dựng cơ sở dữ liệu sử dụng cho các mô hình sạt lở đất. Việc đánh giá và so sánh độ chính xác của các mô hình được thực hiện sử dụng các chỉ số thống kê bao gồm đường cong ROC/AUC. Kết quả nghiên cứu cho thấy rằng các mô hình cây quyết định ADT và DT có độ chính xác cao trong xây dựng bản đồ dự báo không gian sạt lở đất, trong đó mô hình ADT ($AUC_{train} = 0.928$, $AUC_{test} = 0.887$) có độ chính xác cao hơn so với mô hình DT ($AUC_{train} = 0.915$, $AUC_{test} = 0.800$). Bản đồ dự báo không gian sạt lở đất huyện Mường Nhé được xây dựng có thể được sử dụng trong quy hoạch sử dụng đất nhằm quản lý tốt hơn thiên tai sạt lở đất tại khu vực nghiên cứu.

Từ khóa: Sạt lở đất, cây quyết định, Cây quyết định xem kẽ, Điện Biên, Việt Nam, Bản đồ dự báo không gian sạt lở đất.

1. Giới thiệu

Sạt lở đất là một trong những thảm họa thiên nhiên phổ biến trên thế giới, gây ra thiệt hại lớn về con người, kinh tế và môi trường [1]. Tác động của

sạt lở đất được dự đoán sẽ tăng trong các năm tiếp theo với sự phát triển nhanh chóng của đô thị hóa, nạn phá rừng và biến đổi khí hậu [1]. Đặc biệt khu vực châu Á được đánh giá là nơi xảy ra nhiều

vụ sạt lở nhất chiếm 65.2% và có xu hướng xảy ra cao hơn vào mùa mưa [1]. Ở Việt Nam, khoảng 22% sự cố lở đất xảy ra trong giai đoạn tháng 7 đến tháng 9 với lượng mưa lớn từ 300 - 700 mm mỗi tháng gây ra thiệt hại lớn về tài sản. Một trong những biện pháp chính để giảm giảm thiểu thiệt hại do sạt lở đất là lập bản đồ dự báo không gian sạt lở đất. Bản đồ dự báo không gian sạt lở đất là một công cụ hữu ích để khoanh vùng nguy cơ cao xảy ra mất ổn định mái dốc, dự báo các vụ sạt lở đất có thể xảy ra trong tương lai để từ đó đưa ra các biện pháp phòng tránh cũng như giảm thiểu tối đa hậu quả của thiên tai này để lại [2].

Nghiên cứu về sạt lở đất đã được thực hiện ở rất nhiều khu vực trên thế giới sử dụng nhiều phương pháp khác nhau. Các phương pháp truyền thống như: mô hình hồi quy Logistic [2], tỷ số tần suất (FR) [1], phương pháp phân tích thứ bậc (AHP) [1], chỉ số thống kê [1], mô hình trọng số dẫn chứng (WOE) [1]... các mô hình này được đánh giá là có kết quả không chắc chắn và thiếu độ tin cậy. Trong những năm gần đây, trí tuệ nhân tạo trong đó có học máy đã và đang được áp dụng hiệu quả trong việc dự báo, đánh giá sạt lở đất ở nhiều khu vực trên thế giới và cho những kết quả rất triển vọng. Một số mô hình đã được áp dụng Logistic Model Tree, Random Forest, và Classification and Regression Tree cho việc dự báo không gian sạt lở đất tại khu vực Long County, Trung Quốc [3]. Kết quả của nghiên cứu này chỉ ra rằng, các mô hình được áp dụng cho các kết quả dự báo chính xác cao với độ chính xác trên 80% trong đó mô hình Rừng ngẫu nhiên có độ chính xác cao hơn so với hai mô hình còn lại. Các thuật toán điển hình đã được áp dụng cho sạt lở đất có thể kể đến như Máy véc tơ hỗ trợ [4], Mạng thần kinh nhân tạo [5], Cây quyết định [6], Hệ thống suy luận mờ thích ứng [7]... Các nghiên cứu này cho thấy các mô hình học máy được dùng để dự báo sạt lở đất có độ chính xác cao và được thực hiện ở nhiều khu vực trên thế giới.

Trong nghiên cứu này, nhóm tác giả đã sử dụng mô hình học máy phổ biến hiện nay: Cây quyết định (DT) và cây quyết định xen kẽ (ADT)

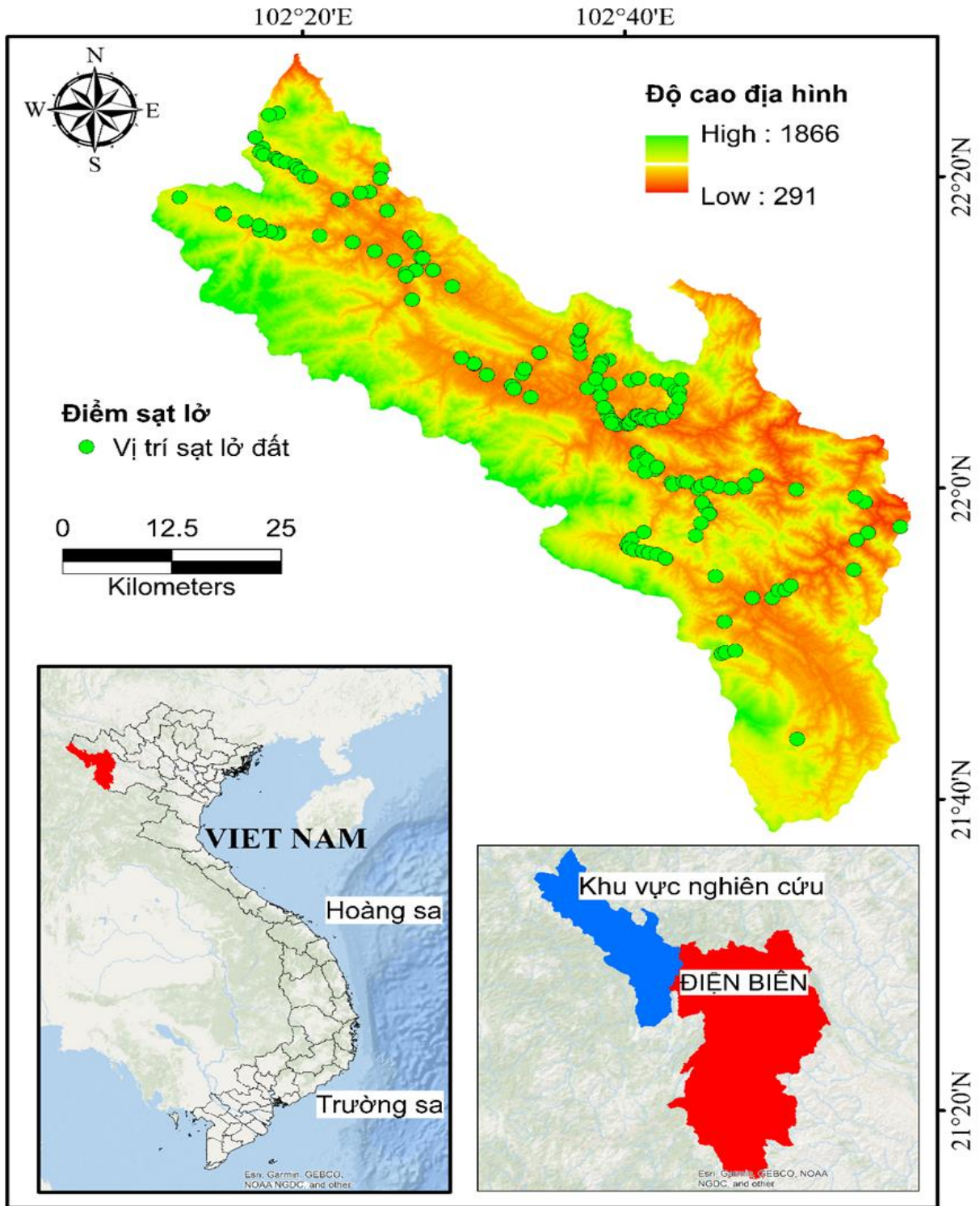
trong việc đánh giá và phân vùng nguy cơ sạt lở đất. Mô hình cây quyết định về cơ bản là thuật toán học có giám sát và được áp dụng cho cả 2 bài toán phân loại (Classification) và hồi quy (Regression), tuy nhiên bài toán phân loại được sử dụng nhiều hơn [1]. Khu vực được lựa chọn nghiên cứu là huyện Mường Nhé, tỉnh Điện Biên, Việt Nam khu vực này có địa hình đồi núi hiểm trở và thường xuyên hứng chịu những trận mưa lớn gây hậu quả sạt lở đất nghiêm trọng.

2. Khu vực nghiên cứu

Mường Nhé là một huyện miền núi tây bắc Việt Nam có tọa độ địa lý kinh độ 102°8' Đông, vĩ độ 22°44' (Hình 1) Bắc với tổng diện tích vào khoảng 2507,9 km². Phía Tây Bắc giáp tỉnh Vân Nam, Trung Quốc. Phía Tây và Tây Nam giáp Lào. Phía Nam giáp huyện Mường Chà tỉnh Điện Biên. Phía Đông và Đông Bắc giáp huyện Mường Tè tỉnh Lai Châu.

Về địa hình, do ảnh hưởng của các hoạt động kiến tạo nên địa hình của Mường Nhé rất phức tạp, chủ yếu là đồi núi dốc, hiểm trở và chia cắt mạnh bởi sông suối. Được cấu tạo bởi những dãy núi chạy dài theo hướng Tây Bắc - Đông Nam với độ cao biến đổi từ 291m đến hơn 1.866m. Địa hình thấp dần từ Bắc xuống Nam và nghiêng dần từ Tây sang Đông. Khu vực nghiên cứu thường xuyên bị ảnh hưởng bởi hệ thống đứt gãy kiến tạo phát triển rất mạnh mẽ bao gồm hệ thống đứt gãy có phương Tây Bắc - Đông Nam [8] và có địa chất được phân thành các nhóm khác nhau bao gồm: nhóm Cretaceous, Paleogene, Permian, Triassic.

Mường Nhé nằm trong vùng khí hậu nhiệt đới gió mùa, mùa đông lạnh khô, mùa hè nóng ẩm, mưa nhiều. Nhiệt độ trung bình từ 22°C - 23°C, có độ ẩm trung bình hàng năm khoảng 81 - 84%. Lượng mưa trung bình hàng năm từ 1400 - 1500mm nhưng không đều theo thời gian, lượng mưa tập trung vào tháng 5 - 9 chiếm 75 - 92% lượng mưa cả năm nên thường xuyên xảy ra sạt lở đất và mưa lũ trên các sông suối gây ảnh hưởng rất lớn đến kinh tế, đời sống của con người.



Hình 1. Vị trí khu vực nghiên cứu và hiện trạng sạt lở đất

3. Thu thập và phân tích kết quả

3.1. Hiện trạng sạt lở đất khu vực nghiên cứu

Hiện trạng sạt lở đất là điều kiện cần thiết để dự đoán sạt lở đất trong tương lai [1]. Trong nghiên cứu này, vị trí không gian của các điểm sạt lở đất đã xảy ra trong quá khứ đã được lấy từ tài

liệu lưu trữ lịch sử của Sở Tài nguyên và Môi trường tỉnh Điện Biên và điều tra thực địa kết hợp với giải đoán hình ảnh vệ tinh sử dụng Google Earth (Hình 2). Sau khi kiểm chứng, các điểm sạt lở được tổng hợp lại dựa trên phần mềm GIS và các khối trượt được coi là một điểm và hiển thị trên bản đồ hiện trạng. Tổng cộng có 159 điểm sạt lở

đất đã xảy ra trong quá khứ tại khu vực và được sử dụng để nghiên cứu bản đồ dự báo không gian sạt lở đất. Trong đó, chọn ngẫu nhiên 111 (70%) điểm sạt lở để phục vụ công tác xây dựng bản đồ



nguy cơ sạt lở đất và 48 (30%) điểm sạt lở để phục vụ công tác kiểm tra và đánh giá độ chính xác bản đồ đã xây dựng [2].



Nguồn:<http://muongnhe.gov.vn/>

Hình 2. Hình ảnh sạt lở đất tại huyện Mường Nhé

3.2. Các yếu tố ảnh hưởng đến sạt lở đất

Việc lựa chọn các tham số nguyên nhân gây ra sạt lở đất là một trong những bước quan trọng để đánh giá nguy cơ của sạt lở đất khu vực [4]. Khảo sát điều tra quá trình sạt lở đất xảy ra trong quá khứ chỉ ra rằng các tham số tác động đến quá trình sạt lở đất có liên quan đến địa hình địa mạo, địa chất thủy văn, quá trình sử dụng đất, điều kiện khí hậu và các nguyên nhân gây ra bởi con người như quá trình cắt xẻ bờ dốc để xây dựng các tuyến đường. Trong nghiên cứu này, có tổng cộng 12 tham số được đánh giá và lựa chọn để xây dựng bản đồ phân vùng sạt lở đất tại Mường Nhé bao gồm: góc mái dốc, hướng mái dốc, hình dáng bề mặt địa hình, độ cao địa hình, khoảng cách đứt gãy, độ bao phủ thực vật (NDVI), tích lũy dòng chảy, độ ẩm địa hình (TWI), sức mạnh dòng chảy (SPI), địa chất, khoảng cách đến sông suối, khoảng cách đến đường giao thông. Số liệu đầu vào để xây dựng bản đồ dự báo bao gồm bản đồ địa chất, đứt gãy được thu thập từ việc điều tra và nguồn dữ liệu quốc gia (<http://www.idm.gov.vn/1P1NPIT/vi-VN/Ban-Do-Dia-Chat.aspx>) với tỷ lệ 1: 200.000. Ngoài ra, bản đồ góc mái dốc, hướng mái dốc, hình dáng bề mặt

địa hình, độ cao địa hình số DEM với độ phân giải 30m đã thu thập được dữ liệu từ Earthexplorer (<https://earthexplorer.usgs.gov>). Mối quan hệ không gian giữa sạt lở đất và yếu tố ảnh hưởng đến sạt lở được phân tích bằng tỷ số tần suất (FR) theo Bảng 1.

3.2.1. Góc mái dốc

Góc mái dốc là một trong những yếu tố quan trọng ảnh hưởng tới quá trình xảy ra sạt lở đất [9]. Sạt lở đất thường xảy ra trên các mái dốc có góc dốc từ 15 -54° và thường ít khi xảy ra trên các mái dốc có độ dốc nhỏ 0 – 10°. Bản đồ này được xây dựng với các lớp khác nhau bao gồm 9 lớp: 0 - 8.332718, 8.332719 - 13.97747, 13.97747 - 18.81582, 18.81582 - 23.11658, 23.11658 - 27.14853, 27.14853 - 31.18049, 31.18049 - 35.75004, 35.75004 - 41.93238, 41.93238 - 68.54333 (Hình 3.a). Phân tích tỷ số tần suất (FR) cho thấy rằng sạt lở xảy ra lớn nhất ở lớp 6 (27.14853 - 31.18049) với FR = 1.73.

3.2.2. Hướng mái dốc

Hướng mái dốc là yếu tố quan trọng tác động đến quá trình xảy ra sạt lở đất bởi vì nó ảnh hưởng tới độ ẩm của vật liệu hình thành mái dốc

[8]. Trong nghiên cứu này, bản đồ hướng mái dốc được trích xuất từ mô hình địa hình số DEM 30m với các lớp khác nhau bao gồm: Mặt bằng, Bắc, Đông bắc, Đông, Đông nam, Nam, Tây nam, Tây, Tây bắc (Hình 3.b). Phân tích tỷ số tần suất (FR) cho thấy rằng sạt lở xảy ra lớn nhất ở lớp Tây Nam với $FR = 1.84$.

3.2.3. Hình dáng bề mặt địa hình

Hình dáng bề mặt địa hình ảnh hưởng đến quá trình xảy ra sạt lở đất bởi vì dòng nước và quá trình tích tụ nước mặt phụ thuộc đáng kể vào hình dáng bề mặt của địa hình [2]. Sạt lở đất thường xảy ra ở khu vực có bề mặt địa hình lõm hơn các khu vực địa hình bằng phẳng và địa hình lồi do nước mặt thường tích tụ ở địa hình lõm nhiều hơn. Trong nghiên cứu này, bản đồ hình dáng bề mặt địa hình được trích xuất từ mô hình độ cao số DEM 30m với 3 lớp như lõm (< -0.05), mặt bằng ($-0.05 - 0.05$) và lồi (> 0.05) (Hình 3.c). Phân tích tỷ số tần suất (FR) cho thấy rằng sạt lở xảy ra lớn nhất ở lớp 1 với $FR = 1.208$.

3.2.4. Độ cao địa hình

Độ cao địa hình ảnh hưởng tới quá trình xảy ra sạt lở đất do ở các độ cao địa hình khác nhau mức độ phong hóa của các loại đất đá trên mái dốc là khác nhau [10]. Trong nghiên cứu này, bản đồ độ cao địa hình được trích xuất từ mô hình độ cao số DEM 30m và được chia thành 9 lớp: 291 - 570, 570 - 679, 679 - 786, 786 - 894, 894 - 1008, 1008 - 1139, 1139 - 1290, 1290 - 1464, 1290 - 1464 (Hình 3.d). Kết quả phân tích tỷ số tần suất (FR) cho thấy rằng sạt lở xảy ra lớn nhất ở lớp 291 - 570m với $FR = 4.785$.

3.2.5. Bao phủ thực vật (NDVI)

Sạt lở đất có liên quan chặt chẽ với bao phủ thực vật. Các khu vực có mật độ che phủ thấp sẽ gây ra sạt lở lớn hơn với khu vực có độ che phủ thực vật cao [3]. Bản đồ bao phủ thực vật (NDVI) được lấy từ hình ảnh dữ liệu vệ tinh với và được chia thành 6 lớp: 0.00492 - 0.15494, 0.15494 - 0.20495, 0.20495 - 0.24603, 0.24603 - 0.28353, 0.28353 - 0.32283, 0.32283 - 0.46213 (Hình 3.e). Kết quả phân tích tỷ số tần suất (FR) cho thấy rằng sạt lở xảy ra lớn nhất ở lớp 0.15494 - 0.20495 với $FR = 3.59$.

Giá trị NDVI được tính toán bằng phương trình sau:

$$NDVI = (NIR - R)/(NIR + R) \quad (1)$$

Trong đó, NIR là giá trị hồng ngoại và R là phần màu đỏ tương ứng.

3.2.6. Địa chất

Ảnh hưởng của điều kiện địa chất, kiến tạo được coi là một nhân tố cơ bản gây ra quá trình sạt lở đất, đặc biệt thành phần thạch học là một trong những nhân tố quan trọng nhất ảnh hưởng đến sự ổn định của sườn. Các đá có độ bền thấp dễ có xu hướng phong hoá thành các vật liệu kém bền vững [11]. Bản đồ địa chất được thu thập từ nguồn dữ liệu quốc gia với tỷ lệ 1:200.000. Các lớp địa chất khu vực bao gồm: Nhóm Cretaceous, Paleogene, Permian, Triassic (Hình 3.f). Kết quả phân tích tỷ số tần suất (FR) cho thấy rằng sạt lở xảy ra lớn nhất ở lớp 2 (Paleogene) với $FR = 4.31$.

3.2.7. Khoảng cách đứt gãy

Đứt gãy là sản phẩm của các chuyển động kiến tạo gây sự mất liên tục trong đất đá trên mái dốc, vì vậy, ảnh hưởng đến quá trình xảy ra sạt lở đất [12]. Trong nghiên cứu này, khoảng cách tới các đứt gãy được lựa chọn là yếu tố nguyên nhân ảnh hưởng tới quá trình sạt lở đất. Hệ thống các đứt gãy được trích xuất từ bản đồ địa chất với tỷ lệ 1: 200.000. Bản đồ khoảng cách tới các đứt gãy được thành lập với 6 lớp: 0 - 100, 100 - 200, 200 - 300, 300 - 400, 400 - 500, > 500 (Hình 3.g). Phân tích tỷ số tần suất (FR) cho thấy rằng sạt lở xảy ra lớn nhất ở lớp 400 - 500m với $FR = 1.855$.

3.2.8. Tích lũy dòng chảy

Tích lũy dòng chảy được một số nhà nghiên cứu coi là một yếu tố điều hòa quan trọng để lập bản đồ nhạy cảm sạt lở đất. Nó được sử dụng để xác định dòng chảy hoặc dòng chảy tiềm năng của sông và suối [1]. Bản đồ tích lũy dòng chảy được thành lập bằng phần mềm Arcgis và được chia thành 6 lớp: 0 -103, 103 - 400, 400 - 930, 930 - 1868, 1868 - 3675, 3675 - 8956 (Hình 3.h). Phân tích tỷ số tần suất (FR) cho thấy rằng sạt lở xảy ra lớn nhất ở lớp 0 -103 với $FR = 1.02$.

3.2.9. Sức mạnh dòng chảy (SPI)

Sức mạnh dòng chảy (SPI) là một số liệu liên

quan đến vận tốc của dòng chảy và gây ra xói mòn cho vùng sông, suối. Sức mạnh dòng chảy càng lớn thì tốc độ xói mòn và ảnh hưởng càng lớn [4]. Bản đồ sức mạnh dòng chảy được thành lập bằng phần mềm ArcGIS và được chia thành 6 lớp: 0 - 502.9279, 502.9279 - 2514.6396, 2514.6396 - 5783.6711, 5783.6711 - 11064.4141, 11064.4141 - 21374.4365, 21374.4365 - 64274.7731 (Hình 3.i). Phân tích tỷ số tần suất (FR) cho thấy rằng sạt lở xảy ra lớn nhất ở lớp 0 - 502.9279 với FR = 1.01.

Giá trị SPI được tính toán bằng phương trình:

$$TWI = A_s \cdot \tan\beta \tag{2}$$

Trong đó, A_s là diện tích khu vực thu nước, β là độ dốc địa hình tính bằng độ.

3.2.10. Độ ẩm địa hình (TWI)

Độ ẩm địa hình (TWI) là một số liệu dựa trên tỷ lệ hoặc vùng lưu vực đến góc dốc [13]. Nó cung cấp một thước đo độ ẩm của đất có sự liên quan tích cực với sự xuất hiện sạt lở đất. Bản đồ sức mạnh dòng chảy được thành lập bằng phần mềm ArcGIS và được chia thành 6 lớp: 2.5029 - 4.7776, 4.7776 - 5.7298, 5.7298 - 6.8407, 6.8407 - 8.2689, 8.2689 - 10.1733, 10.1733 - 16.0452 (Hình 3.j). Phân tích tỷ số tần suất (FR) cho thấy rằng sạt lở xảy ra lớn nhất ở lớp 5.7298 - 6.8407 với FR = 1.14.

Giá trị TWI có thể được xây dựng như sau:

$$TWI = \ln\left(\frac{A_s}{\tan\beta}\right) \tag{3}$$

Trong đó, A_s là diện tích khu vực thu nước,

β là độ dốc địa hình tính bằng độ.

3.2.11. Khoảng cách đến sông suối

Sông suối ảnh hưởng tới quá trình xảy ra sạt lở đất do các mái dốc gần sông suối thường có độ ẩm cao hơn các khu vực khác [8]. Ngoài ra, các dòng nước ở khu vực có sông suối có tác động cơ học trực tiếp tới đất đá của mái dốc. Trong nghiên cứu này, khoảng cách tới sông suối được lựa chọn để phân tích mối quan hệ với quá trình xảy ra sạt lở đất. Hệ thống sông suối được trích xuất từ bản đồ địa hình tỷ lệ 1: 50000. Bản đồ khoảng cách tới sông suối được xây dựng thành 6 lớp: 0 - 100, 100 - 200, 200 - 300, 300 - 400, 400 - 500, > 500 (Hình 3.k). Kết quả phân tích tỷ số tần suất (FR) cho thấy rằng sạt lở xảy ra lớn nhất ở lớp 100 - 200m với FR = 7.96.

3.2.12. Khoảng cách đến đường giao thông

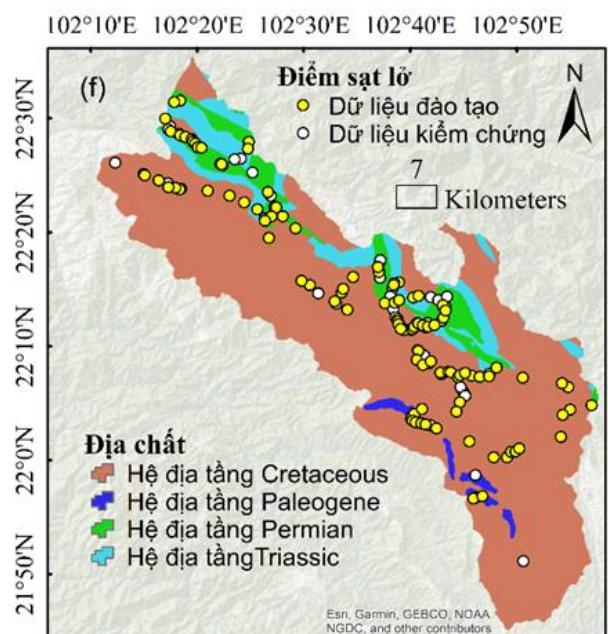
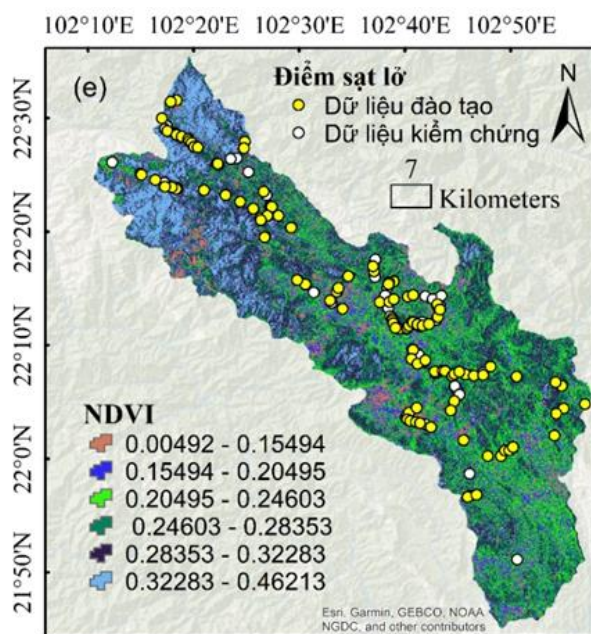
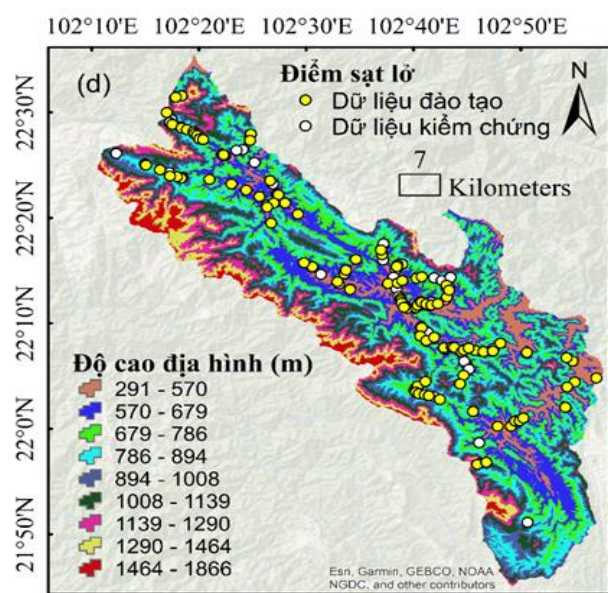
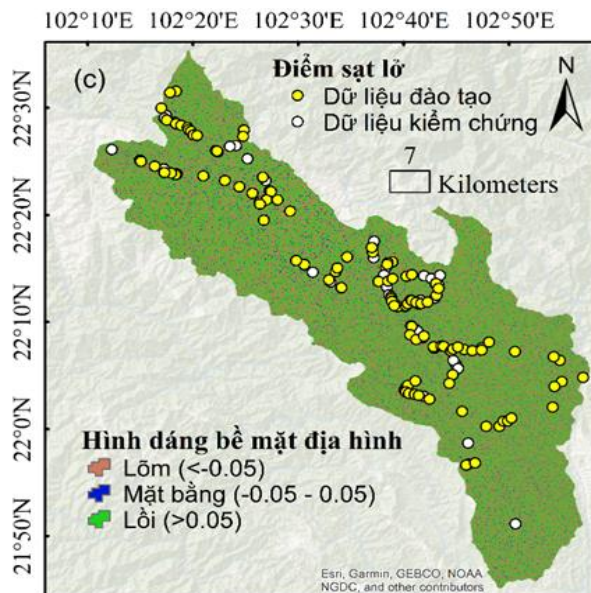
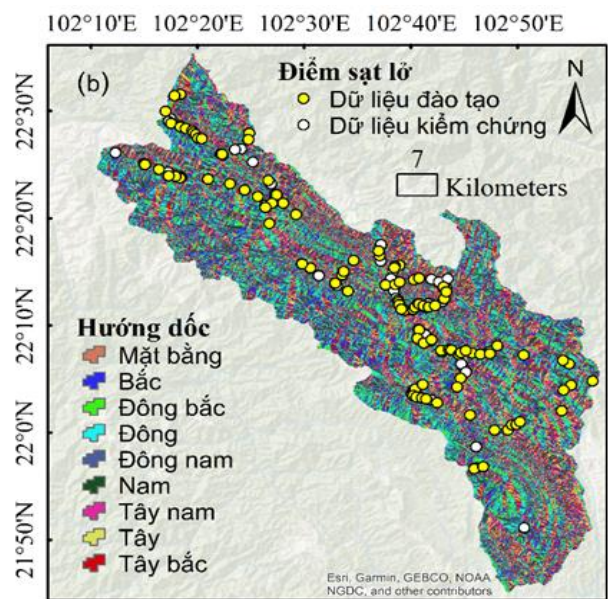
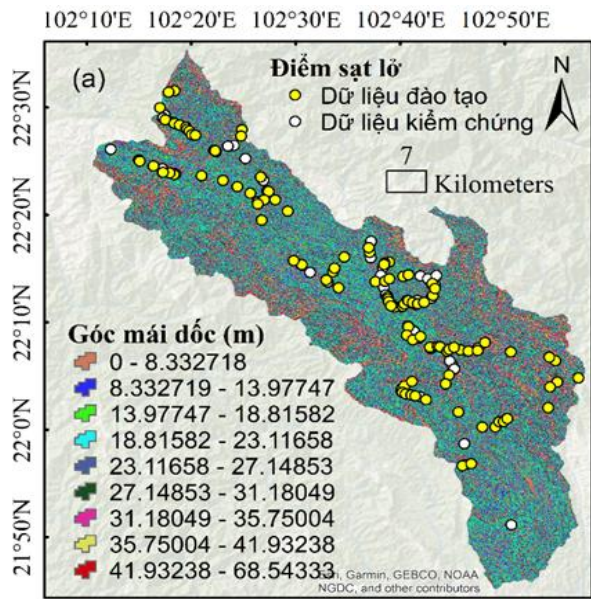
Quá trình cắt mái dốc để xây dựng các tuyến đường giao thông thường tác động trực tiếp tới mái dốc, làm mất tính liên tục của đất đá của mái dốc, tạo ra khu vực tích tụ nước làm giảm cường độ của đất đá trên mái dốc ảnh hưởng tới quá trình xảy ra sạt lở đất [8]. Trong nghiên cứu này, khoảng cách tới đường giao thông được lựa chọn để đánh giá sự ảnh hưởng của đường giao thông tới quá trình xảy ra sạt lở đất. Hệ thống đường giao thông được trích xuất từ bản đồ địa hình tỷ lệ 1:50000. Bản đồ khoảng cách tới sông suối được xây dựng thành 6 lớp: 0 - 100, 100 - 200, 200 - 300, 300 - 400, 400 - 500, > 500 (Hình 3.m). Kết quả phân tích tỷ số tần suất (FR) cho thấy rằng sạt lở xảy ra lớn nhất ở lớp 300 - 400m với FR = 3.66.

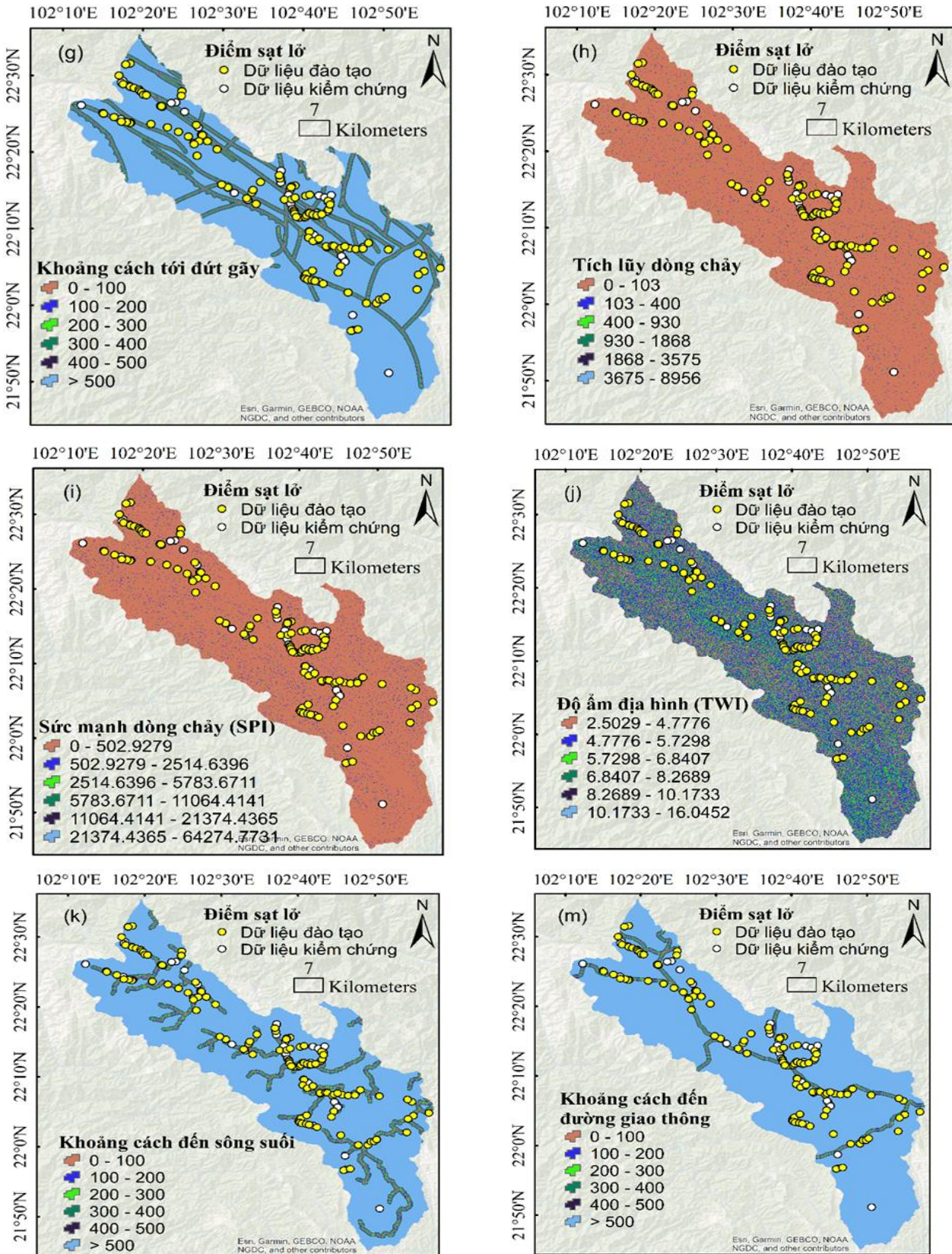
Bảng 1. Tỷ số tần suất của các yếu tố ảnh hưởng

Các yếu tố	Số lớp	Các lớp	Số điểm ảnh của các lớp	Số điểm ảnh sạt lở đất	Phần trăm điểm ảnh các lớp	Phần trăm điểm ảnh sạt lở đất	Tỷ số tần suất
Góc mái dốc (°)	1	0 - 8.332718	204130	3	7.35	2.7	0.37
	2	8.332719 - 13.97747	330548	11	11.91	9.91	0.83
	3	13.97747 - 18.8158	420493	19	15.15	17.12	1.13
	4	18.8158 - 23.11658	459772	18	16.56	16.22	0.98
	5	23.11658 - 27.14853	451661	20	16.27	18.02	1.11
	6	27.14853 - 31.18049	391143	27	14.09	24.32	1.73
	7	31.18049 - 35.75004	291349	7	10.5	6.31	0.60

	8	35.75004 - 41.93238	170631	5	6.15	4.5	0.73
	9	41.93238 - 68.54333	56259	1	2.03	0.9	0.45
Hướng mái dốc	1	Mặt bằng	172	0	0.01	0	0
	2	Bắc	407675	14	14.63	12.61	0.86
	3	Đông bắc	397730	8	14.27	7.21	0.50
	4	Đông	352861	10	12.66	9.01	0.71
	5	Đông nam	336300	15	12.07	13.51	1.12
	6	Nam	351381	20	12.61	18.02	1.43
	7	Tây nam	340648	25	12.22	22.52	1.84
	8	Tây	296405	12	10.64	10.81	1.02
	9	Tây bắc	303495	7	10.89	6.31	0.58
Hình dáng bề mặt	1	Lõm (<-0.05)	1350685	65	48.47	58.56	1.208
	2	Mặt bằng (-0.05 - 0.05)	130906	6	4.7	5.41	1.151
	3	Lồi (>0.05)	1305076	40	46.83	36.04	0.77
Độ cao địa hình (m)	1	291 - 570	257069	49	9.22	44.14	4.785
	2	570 - 679	511684	24	18.36	21.62	1.178
	3	679 - 786	484500	21	17.39	18.92	1.088
	4	786 - 894	438012	11	15.72	9.91	0.630
	5	894 - 1008	370718	1	13.3	0.9	0.068
	6	1008 - 1139	289068	5	10.37	4.5	0.434
	7	1139 - 1290	209225	0	7.51	0	0
	8	1290 - 1464	146986	0	5.27	0	0
	9	1464 - 1866	79405	0	2.85	0	0
NDVI	1	0.00492 - 0.15494	162956	12	5.85	10.81	1.848
	2	0.15494 - 0.20495	370883	53	13.31	47.75	3.587
	3	0.20495 - 0.24603	583470	32	20.94	28.83	1.376
	4	0.24603 - 0.28353	697029	9	25.01	8.11	0.324
	5	0.28353 - 0.32283	622304	3	22.33	2.7	0.121
	6	0.32283 - 0.46213	350025	2	12.56	1.8	0.143
Địa chất	1	Hệ địa tầng Cretaceous	2266388	69	81.34	62.16	0.76
	2	Hệ địa tầng Paleogene	46644	8	1.67	7.21	4.31
	3	Hệ địa tầng Permian	215775	25	7.74	22.52	2.91
	4	Hệ địa tầng Triassic	257420	9	9.24	8.11	0.88
Khoảng cách đứt gãy (m)	1	0 - 100	136579	5	4.9	4.5	0.919
	2	100 - 200	133670	9	4.8	8.11	1.690
	3	200 - 300	130962	8	4.7	7.21	1.534
	4	300 - 400	126785	7	4.55	6.31	1.386
	5	400 - 500	121805	9	4.37	8.11	1.855
	6	> 500	2136866	73	76.68	65.77	0.858

Tích lũy dòng chảy	1	0 - 103	2688232	109	96.47	98.2	1.018
	2	103 - 400	68217	2	2.45	1.8	0.736
	3	400 - 930	20363	0	0.73	0	0
	4	930 - 1868	7455	0	0.27	0	0
	5	1868 - 3675	1995	0	0.07	0	0
	6	3675 - 8956	405	0	0.01	0	0
SPI	1	0 - 502.9279	2649881	107	95.09	96.4	1.013
	2	502.9279 - 2514.6396	108414	4	3.89	3.6	0.926
	3	2514.6396 - 5783.6711	20376	0	0.73	0	0
	4	5783.6711 - 11064.414	6145	0	0.22	0	0
	5	11064.414 - 21374.436	1616	0	0.06	0	0
	6	21374.436 - 64274.773	235	0	0.01	0	0
TWI	1	2.5029 - 4.7776	682857	24	24.5	21.62	0.88
	2	4.7776 - 5.7298	985194	43	35.35	38.74	1.10
	3	5.7298 - 6.8407	615762	28	22.1	25.23	1.14
	4	6.8407 - 8.2689	289670	11	10.39	9.91	0.95
	5	8.2689 - 10.1733	142533	4	5.11	3.6	0.70
	6	10.1733 - 16.0452	70651	1	2.54	0.9	0.36
Khoảng cách đến sông suối (m)	1	0 - 100	102483	29	3.68	26.13	7.10
	2	100 - 200	94657	30	3.4	27.03	7.96
	3	200 - 300	92179	6	3.31	5.41	1.63
	4	300 - 400	92764	3	3.33	2.7	0.81
	5	400 - 500	85294	2	3.06	1.8	0.59
	6	> 500	2319290	41	83.23	36.94	0.44
Khoảng cách đến đường giao thông (m)	1	0 - 100	42287	6	1.52	5.41	3.562
	2	100 - 200	41661	3	1.5	2.7	1.808
	3	200 - 300	41453	8	1.49	7.21	4.845
	4	300 - 400	41125	6	1.48	5.41	3.663
	5	400 - 500	40650	4	1.46	3.6	2.470
	6	> 500	2579491	84	92.57	75.68	0.818





Hình 3. (a) Góc mái dốc, (b) Hướng mái dốc, (c) Hình dáng bề mặt địa hình, (d) Độ cao địa hình, (e) Bao phủ thực vật (NDVI), (f) Địa chất, (g) Khoảng cách đút gậy, (h) Tích lũy dòng chảy, (i) Sức mạnh dòng chảy (SPI), (j) Độ ẩm địa hình (TWI), (k) Khoảng cách đến sông suối, (m) Khoảng cách đến đường giao thông

4. Phương pháp nghiên cứu

4.1. Cây quyết định (DT)

Cây quyết định còn được gọi là cây phân loại phân cấp là một thuật toán quan trọng trong học máy. Một số thuật toán của cây quyết định đã được đề xuất như: ID3 áp dụng cho bài toán phân loại [14], CART áp dụng cho cả bài toán phân loại và hồi quy [15]. Gần đây, Quinlan [1] đã giới thiệu thuật toán phân loại C4.5 là phiên bản nâng cao của thuật toán ID3. Kể từ đó, C4.5 đã được coi là một mô hình tiêu chuẩn trong phân loại có giám sát. Một mô tả chi tiết về thuật toán C4.5 có thể được nhìn thấy ở Quinlan [1]. Có hai bước trong việc xây dựng cây quyết định: Xây dựng cây và tỉa cây [16]. Bước đầu tiên của cây quy trình là tìm hệ số đầu vào với tập dữ liệu đào tạo, sau đó chọn nút gọi là gốc nút. Tập dữ liệu đào tạo được phân chia dựa trên các giá trị ban đầu và được tạo nút phụ và tỷ lệ tăng được ước tính cho mỗi nút phụ. Các biến có tỷ lệ tăng cao nhất được chọn và phân vùng đệ quy của bộ dữ liệu đào tạo được tiếp tục cho đến khi tất cả các trường hợp trong tập dữ liệu đào tạo được gán để mất các nút hoặc không có các biến hoặc dữ liệu đào tạo có thể được chia thêm. Do đó, việc cắt tỉa được thực hiện bằng cách loại bỏ các nút không cần thiết nhưng với độ chính xác được bảo đảm [6].

Các tiêu chí ước tính được sử dụng để xác định các biến đầu vào được dựa trên việc giảm entropy và công thức sau đây đã khắc phục thông tin thu được [1]:

$$\text{Entropy}(E) = - \sum_{i=1}^n P_i \log_2 P_i \quad (4)$$

Trong đó n là số lớp trong miền của tập dữ liệu E; P_i là tỷ lệ của số lượng các thành phần lớp i về tổng số dữ liệu được đặt E.

Thông tin thu được được sử dụng để đo mức giảm dự kiến ở mức phân cấp thấp hơn ngay lập tức, trong đó các bộ dữ liệu được tinh chế bằng một thuộc tính hỗ trợ khác [17]. Thông tin thu được có thể được viết như sau:

$$\text{Gain}(E, A) =$$

$$\text{Entropy}(E) - \sum_{v \in \text{Values}(A)} \frac{|E_v|}{|E|} \text{Entropy}(E_v)^2 \quad (5)$$

Trong đó, các giá trị (A) cung cấp tên miền thuộc tính hỗ trợ (A), E_v biểu thị tập hợp con của E; giá trị tương ứng (A) là v cho mỗi bản ghi, $|E_v|$ and $|E|$ chỉ ra E_v và E, tương ứng.

4.2. Cây quyết định xen kẽ (ADT)

Cây quyết định xen kẽ là sự kết hợp giữa cây quyết định và thuật toán thúc đẩy, và được đề xuất bởi [18]. Mô hình ADT có độ chính xác cao hơn cây mô hình tiêu chuẩn trong các bài toán phân loại [19]. Nói chung, mô hình ADT bao gồm hai loại nút: nút bộ chia và nút dự đoán. Nút bộ tách phân chia dữ liệu dựa trên các giá trị thuộc tính đã chọn và nút dự đoán bao gồm điểm số được sử dụng để đưa ra dự đoán [13]. Một thước đo cơ bản ánh xạ từ cá thể sang số thực bao gồm một dự đoán c_1 , một điều kiện cơ sở c_2 và hai số thực: a và b. Dự đoán là a khi $c_1 \cap c_2$ hoặc b khi $c_1 \cap \neg c_2$. Giá trị của a và b được tính bằng các công thức sau.

$$a = \frac{1}{2} \ln \frac{W_+(c_1 \cap c_2)}{W_-(c_1 \cap c_2)}, b = \frac{1}{2} \ln \frac{W_+(c_1 \cap \neg c_2)}{W_-(c_1 \cap \neg c_2)} \quad (6)$$

Trong đó c_1 và c_2 tốt nhất được chọn bằng cách thu nhỏ $Z_t(c_1, c_2)$ và được định nghĩa là:

$$\begin{aligned} Z_t(c_1, c_2) &= 2 \left(\sqrt{W_+(c_1 \cap c_2) \cdot W_-(c_1 \cap c_2)} \right. \\ &\quad \left. + \sqrt{W_+(c_1 \cap \neg c_2) \cdot W_-(c_1 \cap \neg c_2)} \right) \\ &\quad + W(-C_2) \end{aligned} \quad (7)$$

Giả sử M là các quy tắc thiết lập cơ sở, sau đó một quy tắc mới có thể được định nghĩa là $M_t + 1 = M_t + r_t(x)$ hiển thị hai giá trị dự đoán (a và b) tại mỗi thời điểm sau của cây và x là một tập các trường hợp. Sự phân loại có thể được coi là dấu hiệu tích lũy các giá trị dự đoán trong $M_t + 1$.

$$\text{class}(x) = \text{sign} \left(\sum_{i=1}^T r_t(x) \right) \quad (8)$$

4.3. Phương pháp xác nhận

4.3.1. Đường cong đặc hiệu (ROC/AUC)

Một trong những chỉ số thống kê đánh giá phổ biến và nổi tiếng nhất là các đặc điểm hoạt động của người nhận AUC. Các kết quả cụ thể của AUC có thể được khái quát và đáng tin cậy. Phương pháp AUC là một phương thức dựa trên xác suất có thể đánh giá các thuật toán ở nhiều ngưỡng [3]. Biểu đồ AUC bao gồm hai trục x và y. Mỗi trục đại diện cho các đặc điểm của tính đặc hiệu và độ nhạy tương ứng tương ứng [2]. Diện tích dưới đường cong AUC thường được sử dụng như một tiêu chí để đo độ chính xác dự đoán của các thuật toán. Giá trị AUC thay đổi giữa "0,5 - 1". Giá trị 0,5 cho thấy hiệu suất kém của các thuật toán trong việc dự đoán sự nhạy cảm với các mối nguy hiểm sạt lở. Giá trị tiệm cận 1 cho thấy hiệu suất mạnh mẽ của các thuật toán trong việc dự đoán độ nhạy cảm với sạt lở đất. Nói chung, giá trị AUC càng gần 1 độ chính xác của thuật toán càng cao, trong khi gần 0,5 độ chính xác của thuật toán càng thấp hơn [20]. Phương trình của phương pháp đánh giá bằng phương pháp AUC như sau:

$$AUC = \frac{(\sum TP + \sum TN)}{(P + N)} \quad (9)$$

Trong đó, TP là số lượng sạt lở đất được phân loại chính xác, TN là số lượng không sạt lở đất phân loại chính xác, P là tổng số lượng sạt lở đất và N là tổng số không sạt lở đất.

4.3.2. Chỉ số thống kê

Các chỉ số thống kê được sử dụng để đánh giá hiệu suất của các mô hình bao gồm: giá trị dự đoán tích cực (PPV), giá trị dự đoán âm (NPV), độ nhạy (SST), độ đặc hiệu (SPF), độ chính xác (ACC), Kappa Index (K), căn của sai số toàn phương trung bình gốc (RMSE), sai số tuyệt đối trung bình (MAE) [21]. Chỉ báo K là một biện pháp thống kê hiệu quả giúp đo lường sự đồng thuận ngẫu nhiên giữa các yếu tố phân loại. K thay đổi giữa 1 và 0. Nếu các giá trị K gần gũi với 1, nó cho thấy độ tin cậy cao và độ tin cậy của thuật toán trong việc dự đoán sự nhạy cảm sạt lở đất. Tiêu chí ACC ước tính tỷ lệ hoặc dự báo chính xác để

dự báo toàn bộ sạt lở [8]. RMSE cho biết sự khác biệt giữa dữ liệu được quan sát và dữ liệu ước tính. MAE là một phạm vi sai số giữa các quan sát nhị phân. Các giá trị cao hơn của SPF, PPV, NPV, ACC, SST, K và các giá trị thấp hơn của RMSE và MAE cho biết hiệu suất cao hơn của mô hình trong việc dự đoán sự nhạy cảm sạt lở đất. Mô tả chi tiết và tính toán các chỉ số được trình bày trong các tài liệu được công bố [49, 53].

$$SST = \frac{TP}{TP + FN} \quad (10)$$

$$SPF = \frac{TN}{TN + FP} \quad (11)$$

$$PPV = \frac{TP}{FP + TP} \quad (12)$$

$$NPV = \frac{TN}{FN + TN} \quad (13)$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (14)$$

$$K = \frac{P_p - P_{exp}}{1 - P_{exp}} \quad (15)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_{predicted} - X_{actual})^2} \quad (16)$$

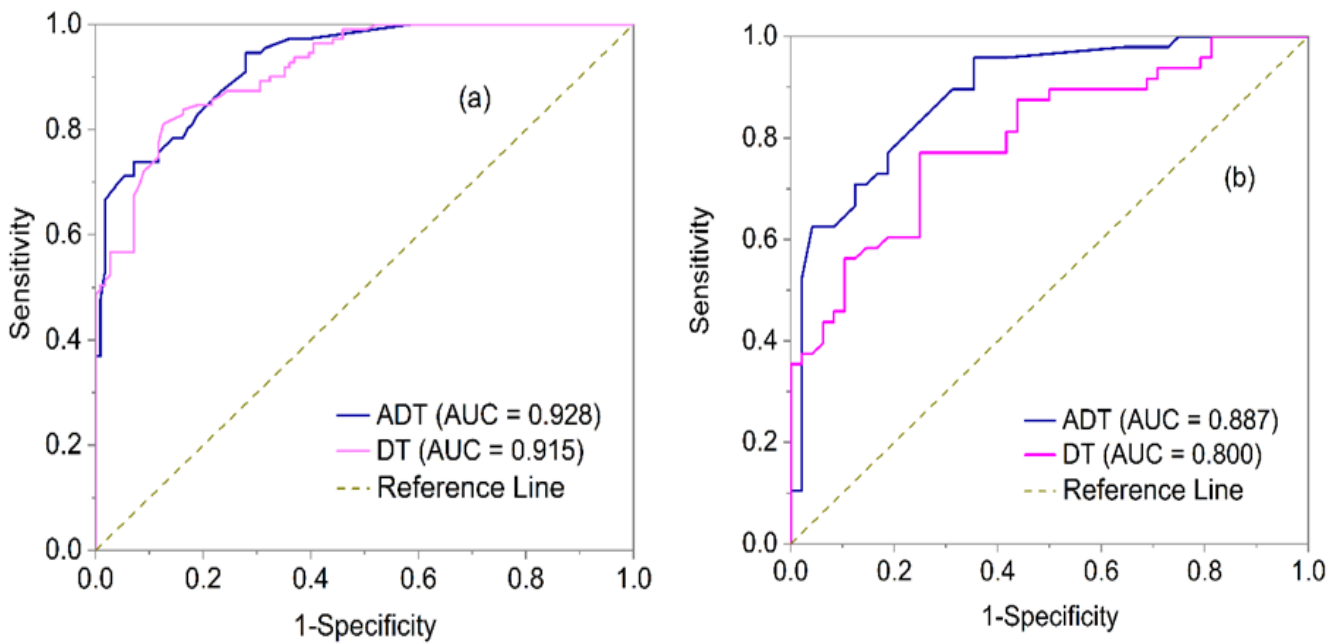
$$MAE = \frac{1}{n} \sum |X_{predicted} - X_{actual}| \quad (17)$$

Trong đó, TP là số lượng sạt lở đất được phân loại chính xác, TN là số lượng không sạt lở đất phân loại chính xác, FN là số lượng sạt lở đất phân loại giả và FP là số lượng không sạt lở đất phân loại giả, n tổng số lượng mẫu trong tập dữ liệu, $X_{predicted}$ và X_{actual} là các giá trị dự đoán và thực tế trong bộ dữ liệu, P_p là số lượng sạt lở quan sát được và P_{exp} là số lượng sạt lở ước tính.

5. Kết quả và thảo luận

5.1. Phân tích kết quả mô hình

Trong nghiên cứu, đường cong đặc hiệu AUC đã được sử dụng để phân tích dữ liệu đào tạo của từng mô hình và hiệu suất. Hai mô hình ADT và DT đã được đào tạo với bộ dữ liệu đào tạo và kết quả được hiển thị trong (Hình 4.a). Bằng cách so sánh giá trị AUC, mô hình ADT cho thấy hiệu suất tốt nhất với (AUC = 0.928) theo sau là mô hình DT (AUC = 0.915). Kiểm chứng hai mô hình (Hình 4.b) cũng cho thấy mô hình ADT (AUC = 0.887) có kết quả cao hơn DT (AUC = 0.800).



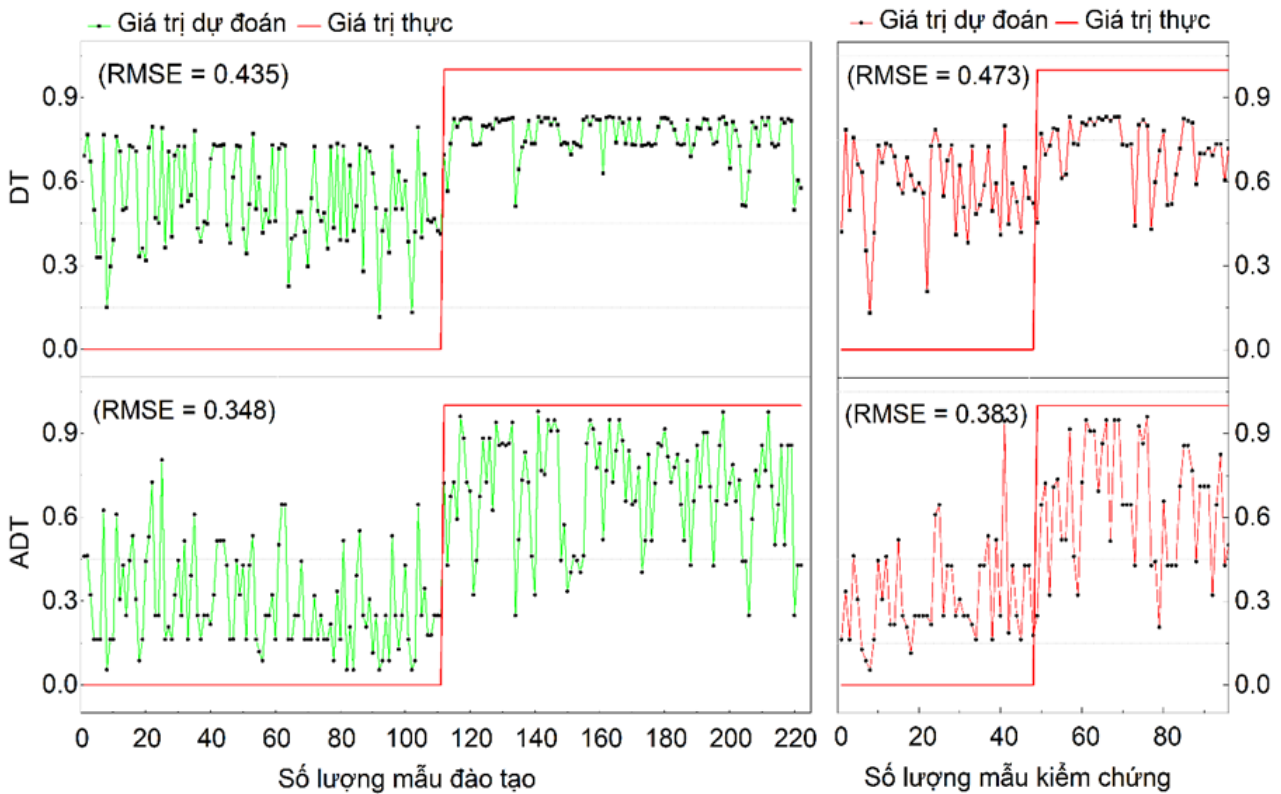
Hình 4. Giá trị AUC của các mô hình khác nhau: (a) Bộ dữ liệu đào tạo, (b) Bộ dữ liệu kiểm chứng

Kết quả phân tích chỉ số thống kê được trình bày trong Bảng 2. Phân tích kết quả cho bộ dữ liệu đào tạo cho thấy mô hình ADT có chỉ số tốt hơn mô hình DT với giá trị: PPV = 82.88%, NPV = 80.18%, SST = 80.70%, SPF = 82.41%, ACC = 81.53% và K = 0.63. Kết quả bộ dữ liệu kiểm chứng cũng cho giá trị cao hơn với PPV = 87.50% có giá trị cao nhất, tiếp theo là NPV = 70.83%, SST

= 75.00%, SPF = 85.00%, ACC = 79.17% và K = 0.58. Giá trị căn của sai số toàn phương trung bình gốc RMSE và sai số tuyệt đối trung bình MAE của mô hình ADT cho cả dữ liệu đào tạo, kiểm chứng đều nhỏ hơn mô hình DT có thể nhận thấy trong Bảng 2 và Hình 5. Phân tích kết quả cho thấy mô hình ADT có hiệu suất đào tạo và kiểm chứng tốt hơn mô hình DT.

Bảng 2. Hiệu suất của mô hình

STT	Tham số	Dữ liệu đào tạo		Dữ liệu kiểm chứng	
		ADT	DT	ADT	DT
1	TP	92	54	42	13
2	TN	89	111	34	45
3	FP	19	57	6	35
4	FN	22	0	14	3
5	PPV (%)	82.88	48.65	87.50	27.08
6	NPV (%)	80.18	100.00	70.83	93.75
7	SST (%)	80.70	100.00	75.00	81.25
8	SPF (%)	82.41	66.07	85.00	56.25
9	ACC (%)	81.53	74.32	79.17	60.42
10	K	0.63	0.49	0.58	0.208
11	MAE	0.30	0.39	0.33	0.429

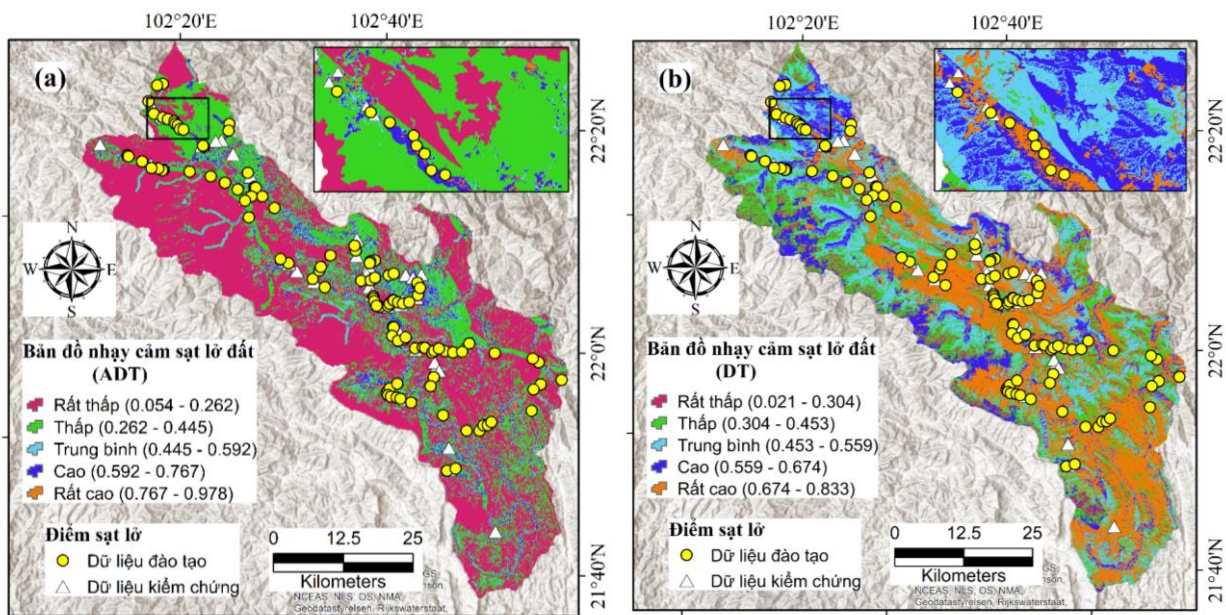


Hình 5. Giá trị căn của sai số toàn phương trung bình gốc (RMSE) của mô hình ADT và DT

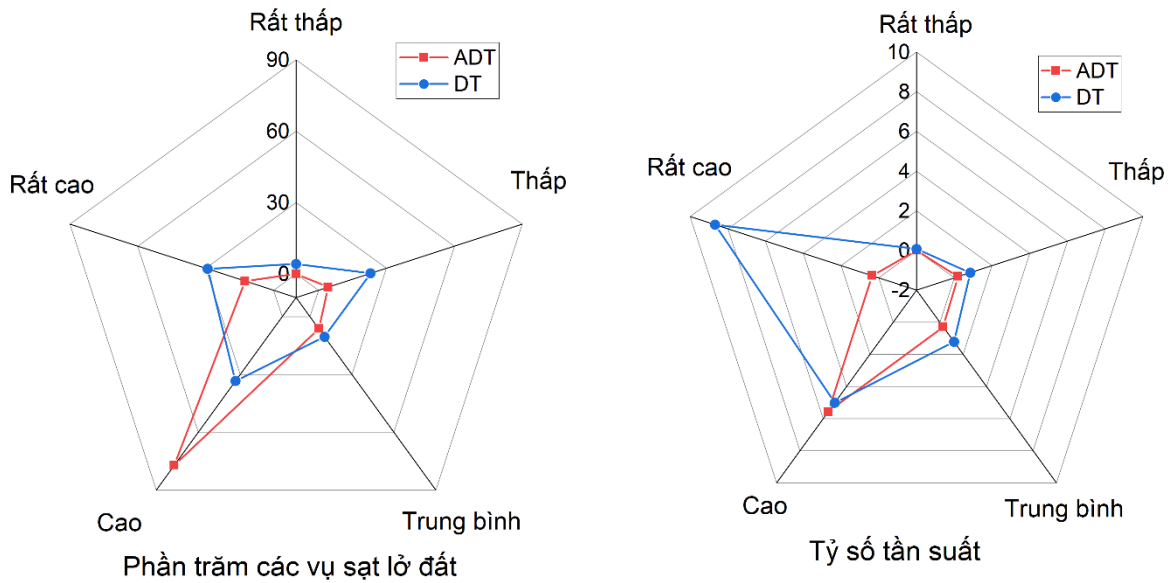
5.2. Đánh giá độ tin cậy của bản đồ phân vùng sạt lở đất khu vực nghiên cứu

Để đánh giá độ tin cậy của bản đồ dự báo không gian sạt lở đất khu vực nghiên cứu, 30% các vụ sạt lở đất chưa được sử dụng trong việc xây dựng

mô hình và bản đồ nhạy cảm sạt lở đất được sử dụng để kiểm chứng. Bằng việc chồng lấn các vụ sạt lở đất này lên các lớp nhạy cảm sạt lở đất của bản đồ dự báo không gian sạt lở đất. Bản đồ phân vùng nhạy cảm được chia thành 5 lớp theo Hình 6.a và Hình 6.b



Hình 6. Bản đồ nhạy cảm sạt lở đất khu vực nghiên cứu: (a) mô hình ADT và (b) mô hình DT



Hình 7. Phần trăm các vụ sạt lở đất và tỷ số tần suất của các lớp nhạy cảm sạt lở đất

bao gồm: Rất thấp, thấp, trung bình, cao và rất cao. Phương pháp "Natural break" trong phần mềm ArcGIS 10.8 đã được sử dụng để phân chia các lớp [54]. Kết quả phân tích được thể hiện trên (Hình 7) cho thấy, phần trăm các vụ sạt lở đất của mô hình ADT có giá trị lần lượt là 77.08% và 12.5% nằm trong khu vực cao và rất cao trong khi chỉ có 6.25%, 4.17%, 0% các vụ sạt lở đất được xác định nằm trong khu vực nhạy cảm trung bình, thấp và rất thấp. Đối với mô hình DT chỉ có 33.33%, 29.17% nằm trong khu vực cao và rất cao, các giá 10.41%, 22.92%, 4.17% nằm trong khu vực nhạy cảm trung bình, thấp và rất thấp. Những kết quả phân tích trên cho thấy rằng, bản đồ nhạy cảm sạt lở được xây dựng bằng mô hình ADT có kết quả tốt hơn so với mô hình còn lại và bản đồ nhạy cảm có thể sử dụng tốt để dự báo sạt lở đất tại khu vực nghiên cứu.

6. Kết luận

Trong nghiên cứu hiện tại, hai mô hình đã được áp dụng đó là mô hình ADT và DT để lập bản đồ dự báo không gian sạt lở đất cho khu vực huyện Mường Nhé, tỉnh Điện Biên. Kết quả phân tích thống kê dữ liệu mô hình ADT ($AUC_{train} = 0.928$, $AUC_{test} = 0.887$) có hiệu suất đào tạo, kiểm chứng cao hơn so với mô hình DT ($AUC_{train} = 0.915$, $AUC_{test} = 0.800$). Bản đồ dự

báo không gian sạt lở đất cho thấy rằng mô hình ADT đã phân định khoảng 77.08% và 12.5% diện tích khu vực nghiên cứu vào các khu vực có xác suất xảy ra sạt lở đất cao và rất cao trong khi đó mô hình DT chỉ có kết quả tương ứng là: 33.33%, 29.17%. Đánh giá kết quả cho thấy mô hình ADT là công cụ có độ chính xác cao trong việc xây dựng bản đồ dự báo không gian sạt lở đất, mô hình này cũng có thể được áp dụng cho các khu vực khác.

Tài liệu tham khảo

- [1]. S. Chakraborty and R. Pradhan. (2012). Development of GIS based landslide information system for the region of East Sikkim. *International Journal of Computer Applications*, 49(7).
- [2]. D. Kanungo, M. Arora, S. Sarkar and R. Gupta. (2012). Landslide Susceptibility Zonation (LSZ) Mapping—A Review.
- [3]. M. Ercanoglu, O. Kasmer and N. Temiz. (2008). Adaptation and comparison of expert opinion to analytical hierarchy process for landslide susceptibility mapping. *Bulletin of Engineering Geology and the Environment*, 67(4), 565-578.
- [4]. B.T. Pham, T. Van Phong, T. Nguyen-Thoi, P.T. Trinh, Q.C. Tran, L.S. Ho, S.K. Singh, T.T.T. Duyen, L.T. Nguyen and H.Q. Le.

- (2020). GIS-based ensemble soft computing models for landslide susceptibility mapping. *Advances in Space Research*, 66(6), 1303-1320.
- [5]. C.J. Van Westen, N. Rengers, M. Terlien and R. Soeters. (1997). Prediction of the occurrence of slope instability phenomena through GIS-based hazard zonation. *Geologische Rundschau*, 86(2), 404-414.
- [6]. A. Erener, A. Mutlu and H.S. Düzgün. (2016). A comparative study for landslide susceptibility mapping using GIS-based multi-criteria decision analysis (MCDA), logistic regression (LR) and association rule mining (ARM). *Engineering geology*, 203, 45-55.
- [7]. S. Mondal and R. Maiti. (2013). Integrating the analytical hierarchy process (AHP) and the frequency ratio (FR) model in landslide susceptibility mapping of Shiv-khola watershed, Darjeeling Himalaya. *International Journal of Disaster Risk Science*, 4(4), 200-212.
- [8]. P. Kayastha, M.R. Dhital and F. De Smedt. (2013). Application of the analytical hierarchy process (AHP) for landslide susceptibility mapping: A case study from the Tinau watershed, west Nepal. *Computers & Geosciences*, 52, 398-408.
- [9]. M. Gholami, E.N. Ghachkanlu, K. Khosravi and S. Pirasteh. (2019). Landslide prediction capability by comparison of frequency ratio, fuzzy gamma and landslide index method. *Journal of Earth System Science*, 128(2), 1-22.
- [10]. S. Ma, H. Qiu, S. Hu, Y. Pei, W. Yang, D. Yang and M. Cao. (2020). Quantitative assessment of landslide susceptibility on the Loess Plateau in China. *Physical Geography*, 41(6), 489-516.
- [11]. W. Chen, X. Xie, J. Wang, B. Pradhan, H. Hong, D.T. Bui, Z. Duan and J. Ma. (2017). A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena*, 151, 147-160.
- [12]. H. Hong, J. Liu, A.-X. Zhu, H. Shahabi, B.T. Pham, W. Chen, B. Pradhan and D.T. Bui. (2017). A novel hybrid integration model using support vector machines and random subspace for weather-triggered landslide susceptibility assessment in the Wuning area (China). *Environmental Earth Sciences*, 76(19), 652.
- [13]. B. Kalantar, B. Pradhan, S.A. Naghibi, A. Motevalli and S. Mansor. (2018). Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). *Geomatics, Natural Hazards and Risk*, 9(1), 49-69.
- [14]. A. Aditian, T. Kubota and Y. Shinohara. (2018). Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. *Geomorphology*, 318, 101-111.
- [15]. H. Hong, J. Liu, D.T. Bui, B. Pradhan, T.D. Acharya, B.T. Pham, A.-X. Zhu, W. Chen and B.B. Ahmad. (2018). Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). *Catena*, 163, 399-413.
- [16]. W. Chen, X. Xie, J. Peng, J. Wang, Z. Duan and H. Hong. (2017). GIS-based landslide susceptibility modelling: a comparative assessment of kernel logistic regression, Naïve-Bayes tree, and alternating decision tree models. *Geomatics, Natural Hazards and Risk*, 8(2), 950-973.
- [17]. I.N. Aghdam, B. Pradhan and M. Panahi. (2017). Landslide susceptibility assessment using a novel hybrid model of statistical bivariate methods (FR and WOE) and adaptive neuro-fuzzy inference system (ANFIS) at southern Zagros Mountains in Iran. *Environmental Earth Sciences*, 76(6), 237.

- [18]. C. Polykretis, C. Chalkias and M. Ferentinou. (2017). Adaptive neuro-fuzzy inference system (ANFIS) modeling for landslide susceptibility assessment in a Mediterranean hilly area. *Bulletin of Engineering Geology and the Environment*, 1-15.
- [19]. A. Priyam, G. Abhijeeta, A. Rathee and S. Srivastava. (2013). Comparative analysis of decision tree classification algorithms. *International Journal of current engineering and technology*, 3(2), 334-337.
- [20]. N. Van Hung, H.Q. Vinhand N.V. Huong. (2016). Tectono-structural system and geodynamic features of Northwest Vietnam in the late Cenozoic period. *Vietnam Journal of Earth Sciences*, 38(1), 38-45.
- [21]. N. Van Liem, N.P. Dat, B.T. Dieu, V. Van Phai, P.T. Trinh, H.Q. Vinhand T. Van Phong. (2016). Assessment of geomorphic processes and active tectonics in Con Voi mountain range area (Northern Vietnam) using the hypsometric curve analysis method. *Vietnam Journal of Earth Sciences*, 38(2), 202-216.
- [22]. D.T. Bui, P.-T.T. Ngo, T.D. Pham, A. Jaafari, N.Q. Minh, P.V. Hoa and P. Samui. (2019). A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. *Catena*, 179, 184-196.
- [23]. B.T. Pham, D. Tien Bui, P. Indra and M. Dholakia. (2015). Landslide susceptibility assessment at a part of Uttarakhand Himalaya, India using GIS-based statistical approach of frequency ratio method. *Int J Eng Res Technol*, 4(11), 338-344.
- [24]. P.T. Nguyen, T.T. Tuyen, A. Shirzadi, B.T. Pham, H. Shahabi, E. Omidvar, A. Amini, H. Entezami, I. Prakash and T.V. Phong. (2019). Development of a novel hybrid intelligence approach for landslide spatial prediction. *Applied Sciences*, 9(14), 2824.
- [25]. L. Ayalew, H. Yamagishi and N. Ugawa. (2004). Landslide susceptibility mapping using GIS-based weighted linear combination, the case in Tsugawa area of Agano River, Niigata Prefecture, Japan. *Landslides*, 1(1), 73-81.
- [26]. S. Lee and J.A. Talib. (2005). Probabilistic landslide susceptibility and factor effect analysis. *Environmental Geology*, 47(7), 982-990.
- [27]. E. Yesilnacar and T. Topal. (2005). Landslide susceptibility mapping: a comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). *Engineering Geology*, 79(3-4), 251-266.
- [28]. D.T. Bui, T.C. Ho, I. Revhaug, B. Pradhan and D.B. Nguyen. (2014). Landslide susceptibility mapping along the national road 32 of Vietnam using GIS-based J48 decision tree classifier and its ensembles. In *Cartography from pole to pole*, Springer: pp 303-317.
- [29]. Q. Meng, F. Miao, J. Zhen, X. Wang, A. Wang, Y. Peng and Q. Fan. (2016). GIS-based landslide susceptibility mapping with logistic regression, analytical hierarchy process, and combined fuzzy and support vector machine methods: a case study from Wolong Giant Panda Natural Reserve, China. *Bulletin of Engineering Geology and the Environment*, 75(3), 923-944.
- [30]. D.T. Bui, P. Tsangaratos, P.-T.T. Ngo, T.D. Pham and B.T. Pham. (2019). Flash flood susceptibility modeling using an optimized fuzzy rule based feature selection technique and tree based ensemble methods. *Science of the total environment*, 668, 1038-1054.
- [31]. H.R. Pourghasemi, B. Pradhan and C. Gokceoglu. (2012). Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. *Natural hazards*, 63(2), 965-996.
- [32]. P.R. Kadavi, C.-W. Lee and S. Lee. (2018). Application of ensemble-based machine learning models to landslide susceptibility mapping. *Remote Sensing*, 10(8), 1252.
- [33]. C.P. Poudyal, C. Chang, H.-J. Oh and S. Lee. (2010). Landslide susceptibility maps

- comparing frequency ratio and artificial neural networks: a case study from the Nepal Himalaya. *Environmental Earth Sciences*, 61(5), 1049-1064.
- [34]. J.P. Wilson and J.C. Gallant. (2000). Terrain analysis: principles and applications. *John Wiley & Sons*.
- [35]. B.T. Pham, T. Nguyen-Thoi, C. Qi, T. Van Phong, J. Dou, L.S. Ho, H. Van Le and I. Prakash. (2020). Coupling RBF neural network with ensemble learning techniques for landslide susceptibility mapping. *Catena*, 195, 104805.
- [36]. B.T. Pham, T.V. Phong, T. Nguyen-Thoi, K. Parial, S. K. Singh, H.-B. Ly, K.T. Nguyen, L.S. Ho, H.V. Le and I. Prakash. (2020). Ensemble modeling of landslide susceptibility using random subspace learner and different decision tree classifiers. *Geocarto International*, 1-23.
- [37]. J.R. Quinlan. (1987). Simplifying decision trees. *International journal of man-machine studies*, 27(3), 221-234.
- [38]. D. Steinberg and P. Colla. (2009). CART: classification and regression trees. *The top ten algorithms in data mining*, 9, 179.
- [39]. J.R. Quinlan. (2014). C4. 5: programs for machine learning. *Elsevier*.
- [40]. D.T. Bui, B. Pradhan, I. Revhaug and C.T. Tran. (2014). A comparative assessment between the application of fuzzy unordered rules induction algorithm and J48 decision tree models in spatial prediction of shallow landslides at Lang Son City, Vietnam. In *Remote sensing applications in environmental research*, Springer: pp 87-111.
- [41]. M. Ahmadlou, M. Karimi, S. Alizadeh, A. Shirzadi, D. Parvinnejhad, H. Shahabi and M. Panahi. (2019). Flood susceptibility assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and BAT algorithms (BA). *Geocarto International*, 34(11), 1252-1272.
- [42]. J.R. Quinlan. (1986). Induction of decision trees. *Machine learning*, 1(1), 81-106.
- [43]. Y. Freund and L. Mason. (1999). In The alternating decision tree learning algorithm. *icml*, Citeseer: pp 124-133.
- [44]. H.K. Sok, M.P.-L. Ooi, Y.C. Kuang and S. Demidenko. (2016). Multivariate alternating decision trees. *Pattern Recognition*, 50, 195-209.
- [45]. A. Shirzadi, K. Soliamani, M. Habibnejhad, A. Kaviani, K. Chapi, H. Shahabi, W. Chen, K. Khosravi, B. Thai Pham and B. Pradhan. (2018). Novel GIS based machine learning algorithms for shallow landslide susceptibility mapping. *Sensors*, 18(11), 3777.
- [46]. M. Avand, S. Janizadeh, D. Tien Bui, V.H. Pham, P.T.T. Ngo and V.-H. Nhu. (2020). A tree-based intelligence ensemble approach for spatial prediction of potential groundwater. *International Journal of Digital Earth*, 13(12), 1408-1429.
- [47]. A. Jaafari. (2018). LiDAR-supported prediction of slope failures using an integrated ensemble weights-of-evidence and analytical hierarchy process. *Environmental Earth Sciences*, 77(2), 42.
- [48]. W. Chen, H. Hong, M. Panahi, H. Shahabi, Y. Wang, A. Shirzadi, S. Pirasteh, A.A. Alesheikh, K. Khosravi, S. Panahi, F. Rezaie, S. Li, A. Jaafari, D.T. Bui and B. Bin Ahmad. (2019). Spatial Prediction of Landslide Susceptibility Using GIS-Based Data Mining Techniques of ANFIS with Whale Optimization Algorithm (WOA) and Grey Wolf Optimizer (GWO). *Applied Sciences*, 9(18), 3755.
- [49]. V.-H. Nhu, A. Mohammadi, H. Shahabi, B.B. Ahmad, N. Al-Ansari, A. Shirzadi, M. Geertsema, V. R. Kress, S. Karimzadeh and K.J.F. Valizadeh Kamran. (2020). Landslide Detection and Susceptibility Modeling on Cameron Highlands (Malaysia): A Comparison between Random Forest, Logistic Regression and Logistic Model Tree Algorithms. 11(8), 830.
- [50]. D.-H. Lee, Y.-T. Kim and S.-R. J.R.S. Lee. (2020). Shallow Landslide Susceptibility Models Based on Artificial Neural Networks

- Considering the Factor Selection Method and Various Non-Linear Activation Functions. 12(7), 1194.
- [51]. R. Prăvălie and R. Costache. (2014). The analysis of the susceptibility of the flash-floods' genesis in the area of the hydrographical basin of Bâsca Chiojdului river. *Forum geografic*, XIII(1), 39-49.
- [52]. P. De Rosa, A. Fredduzzi and C. Cencetti. (2019). Stream Power Determination in GIS: An Index to Evaluate the Most 'Sensitive' Points of a River. *Water*, 11(6), 1145.
- [53]. P. Yariyan, S. Janizadeh, T. Van Phong, H.D. Nguyen, R. Costache, H. Van Le, B.T. Pham, B. Pradhan and J.P.J.W.R.M. Tiefenbacher. (2020). Improvement of best first decision trees using bagging and dagging ensembles for flood probability mapping. 34(9), 3037-3053.
- [54]. S. Roy, S. Pandit, M. Papia, M.M. Rahman, J.C. Otto Rehder Ocampo, M.A. Razi, P. Fraile-Jurado, N. Ahmed, M. Al-Amin Hoque, M.M. Hasan, J. Yeasmin and M.S. Hossain. (2021). Coastal erosion risk assessment in the dynamic estuary: The Meghna estuary case of Bangladesh coast. *International Journal of Disaster Risk Reduction*, 61, 102364.