

NGHIÊN CỨU ỨNG DỤNG KỸ THUẬT HỌC MÁY KẾT HỢP HỆ THỐNG THÔNG TIN ĐỊA LÝ ĐỂ XÂY DỰNG BẢN ĐỒ NHẠY CẢM SẠT LỞ ĐẤT

Trần Xuân Biên, Lưu Thùy Dương
Trường Đại học Tài nguyên và Môi trường Hà Nội

Tóm tắt

Sạt lở đất hàng năm đã gây thiệt hại lớn về tài sản, cơ sở hạ tầng và con người ở nhiều vùng miền núi của Việt Nam. Các nghiên cứu về sạt lở đất đã được quan tâm trong công tác quản lý, phòng chống thiên tai. Nghiên cứu này trình bày phương pháp học máy, đó là phân loại hàm cơ sở xuyên tâm (RBFC) để tạo bản đồ nhạy cảm sạt lở đất tại huyện Mường Chà của tỉnh Điện Biên thuộc vùng núi Tây Bắc, nơi thường xuyên bị ảnh hưởng bởi hệ thống đứt gãy kiến tạo phát triển rất mạnh mẽ. Trong các nghiên cứu mô hình, 12 yếu tố ảnh hưởng đã được lựa chọn dựa trên địa hình và điều kiện địa của khu vực. Để xác nhận hiệu suất của mô hình, các chỉ số thống kê bao gồm đường cong ROC/AUC và các chỉ số thống kê khác nhau được sử dụng. Kết quả nghiên cứu cho thấy rằng các mô hình phân loại hàm cơ sở xuyên tâm (RBFC) có độ chính xác cao trong xây dựng bản đồ dự báo không gian sạt lở đất, với $AUC_{\text{đào tạo}} = 0,931$, $AUC_{\text{kiểm chứng}} = 0,857$. Nghiên cứu này hữu ích cho việc xây dựng bản đồ nhạy cảm sạt lở đất với mục tiêu xác định các khu vực, vùng dễ bị sạt lở đất để quản lý rủi ro.

Từ khóa: Sạt lở đất; Phân loại hàm cơ sở xuyên tâm; Mường Chà; Việt Nam.

Abstract

Study on the application of machine learning technique and geographical information systems to build landslide susceptibility maps

Landslides have caused great damage to property, infrastructure and people in many mountainous areas of Vietnam. Studies on landslides have been addressed in the management and prevention of natural disasters. This study presents a machine learning method, which is a radial basis function classifier (RBFC) to create a landslide susceptibility map in Muong Cha district of Dien Bien province in the Northwest mountainous region, which is frequently affected by landslides. In the model, 12 influential factors were selected based on the topography and geographical conditions of the study area. To confirm the performance of the model, statistical indicators including ROC/AUC curves and various statistical indicators were used. The results showed that the RBFC models have high accuracy in building landslide spatial prediction maps, with $AUC_{\text{train}} = 0,931$, $AUC_{\text{test}} = 0,857$. This study is useful for building landslide susceptibility maps with the aim of identifying landslide prone areas for risk management.

Keywords: Landslide; Radial basis function classification; Muong Cha; Vietnam.

1. Mở đầu

Sạt lở đất được coi là một trong những thiên tai nguy hiểm, gây thiệt hại lớn về tài sản, cơ sở hạ tầng và con người trên toàn thế giới [1]. Các vụ sạt lở đất xảy ra ngày càng nhiều và có tính chất nguy hiểm, đặc biệt khu vực Châu Á có địa hình hiểm trở và địa chất phức tạp [2]. Hơn nữa, với sự biến đổi của khí hậu ngày càng gia tăng trên toàn cầu thì tác động của thiên tai như: lũ lụt, sạt lở đất,... ngày càng diễn ra thường xuyên và khắc nghiệt hơn, đòi hỏi sự quan tâm hơn nữa của chính phủ và các tổ chức khác. Mặc dù không thể ngăn chặn các vụ sạt lở đất, nhưng dự đoán chính xác tính nhạy cảm của sạt lở đất có thể giúp giảm thiệt hại đáng kể.

Nguyên nhân sạt lở đất chủ yếu xảy ra do sự tác động của yếu tố tự nhiên và yếu tố hoạt động của con người. Việc lựa chọn các yếu tố ảnh hưởng đến trượt lở chủ yếu dựa vào dữ liệu có sẵn và kinh nghiệm của chuyên gia đối với vùng nghiên cứu dẫn đến kết quả không phản ánh đúng bản chất. Tính nhạy cảm của sạt lở đất được đánh giá định lượng hoặc định tính của một khu vực có sự phân bố sạt lở đất theo không gian, nơi có khả năng xảy ra sạt lở đất [3]. Đây là thước đo xác suất sạt lở đất trong tương lai.

Nghiên cứu về lập bản đồ nhạy cảm sạt lở đất đã được các nhà khoa học phân tích với nhiều phương pháp thống kê khác nhau. Các phương pháp thống kê như: Phương pháp gần trọng số, tỷ số tần suất, phân cấp thứ bậc và lấy ý kiến chuyên gia là các phương pháp mang tính chủ quan và có độ chính xác không cao. Vì vậy, việc nghiên cứu và phát triển các

công cụ dự báo sử dụng các phương pháp mới là cần thiết. Trong những năm gần đây, các thuật toán về trí tuệ nhân tạo và học máy được áp dụng hiệu quả trong dự báo. Các phương pháp này dựa trên các thuật toán toán học hiệu quả để phân tích và dự đoán [4]. Một số phương pháp ML phổ biến được sử dụng để đánh giá tính nhạy cảm với lũ lụt là [5]: Máy vector hỗ trợ (SVM) [6]; Hồi quy logistic (LR) [7]; Hệ thống suy luận mờ thần kinh thích ứng (ANFIS) [8]. Cho đến nay, ở Việt Nam chưa có nhiều mô hình máy học áp dụng cho khu vực để đánh giá và lập bản đồ nhạy cảm sạt lở đất. Cần có nghiên cứu để lựa chọn các mô hình thích hợp để xác định và lập bản đồ chính xác các khu vực dễ bị sạt lở đất. Với mục tiêu này, nhóm tác giả đã áp dụng mô hình máy học (RBFC). Đây là một mạng nơron có giám sát xem xét một bài toán xấp xỉ trong không gian đa chiều được sử dụng để trả lời các câu hỏi như nội suy và nhận dạng [9]. Kết quả được xác nhận và đánh giá hiệu suất mô hình đối với lập bản đồ tính nhạy cảm sạt lở đất.

Vùng núi Tây Bắc là một trong những khu vực dễ bị sạt lở nhất ở Việt Nam. Khu vực này thường xuyên bị ảnh hưởng bởi hệ thống đứt gãy kiến tạo phát triển mạnh mẽ và địa chất phong hóa dẫn đến sạt lở đất thường xuyên. Việc áp dụng mô hình học máy RBFC cho lập bản đồ nhạy cảm sạt lở đất sẽ mang lại hiệu quả cho việc quy hoạch và quản lý rủi ro.

2. Khu vực nghiên cứu

Mường Chà là huyện nằm ở trung tâm tỉnh Điện Biên và có diện tích là 1.199,42 km². Khu vực có phía Đông giáp huyện Tủa Chùa và huyện Tuần Giáo.

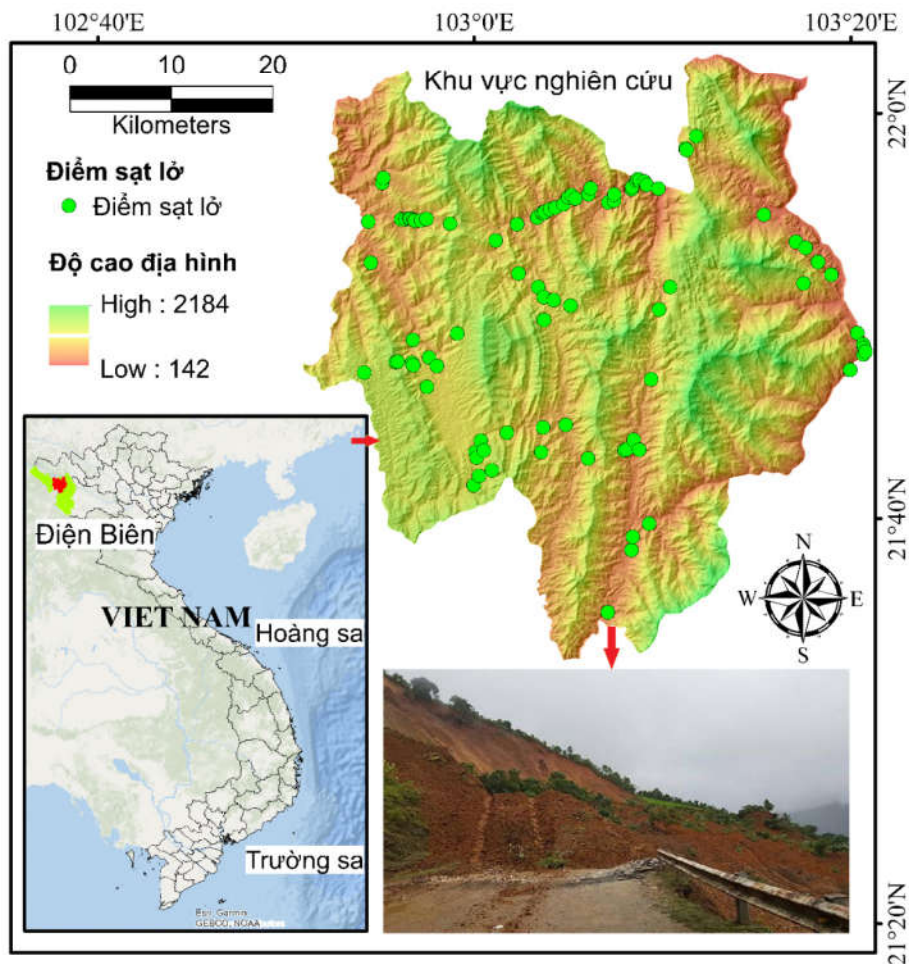
Nghiên cứu

Phía Tây giáp huyện Nậm Pồ và Lào. Phía Nam giáp thành phố Điện Biên Phủ và các huyện Điện Biên, Mường Ảng. Phía Bắc giáp thị xã Mường Lay và tỉnh Lai Châu.

Địa hình khu vực bị chia cắt phức tạp do có nhiều núi cao và khe sâu tạo thành với độ dốc từ 160 - 450 m, độ cao trung bình so với mặt nước biển từ 350 đến 1.500 m và nghiêng dần theo hướng Tây Bắc - Đông Nam. Địa chất chủ yếu là đá nứt nẻ, đập vỡ mảnh, với chiều dày phong hóa thường thay đổi từ 12 m đến 20 m và được phân thành các nhóm khác nhau bao gồm: Nhóm Cretaceous, Devonian, Neoproterozoic, Paleogene, Permian, Triassic, Silurian. Hệ thống sông suối khu vực khá đa dạng bao gồm các sông

như: Nậm Mực, Nậm He, Nậm Chim, Nậm Pồ và nhiều dòng suối nhỏ: Suối Nậm Piền, suối Huổi Chá, suối Ma Thì Hồ, suối Năm Khăn, suối Đề Cua Tử,...

Mường Chà có khí hậu nhiệt đới gió mùa, nhiệt độ trung bình trong năm là từ 18 - 25 °C. Độ ẩm không khí trung bình từ 83 - 85 %. Lượng mưa trung bình cả năm là từ 1.600 mm đến 2.400 mm. Mùa mưa tập trung chủ yếu từ tháng 4 đến tháng 9, mùa khô bắt đầu từ tháng 10 đến tháng 4 năm sau. Với lượng mưa tập trung vào mùa mưa nên thường xuyên xảy ra sạt lở đất và lũ lụt trên các sông suối gây ảnh hưởng lớn đến đời sống và kinh tế của con người.



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

3. Thu thập và phân tích dữ liệu***3.1. Hiện trạng sạt lở đất khu vực nghiên cứu***

Trong quá trình xây dựng mô hình, dữ liệu lịch sử về các vụ sạt lở đất là rất quan trọng [10]. Khu vực nghiên cứu thường xuyên bị ảnh hưởng bởi các trận mưa lớn trong thời gian ngắn gây ra nhiều trận sạt lở đất. Hầu hết các điểm trượt lở đều trượt lở trong vỏ phong hóa, đặc biệt là taluy đường với các kiểu trượt như tịnh tiến, trượt ngang, rãnh xói, trượt hỗn hợp. Các điểm trượt lở này sẽ còn nguy cơ tiếp diễn. Trong nghiên cứu này, vị trí không gian của các điểm sạt lở đất đã được thu thập từ Sở Tài nguyên và Môi trường tỉnh Điện Biên và điều tra thực địa kết hợp với giải đoán hình ảnh vệ tinh Google Earth (Hình 1). Sau khi kiểm chứng, một bản đồ kiểm kê với 92 điểm sạt lở đất đã được tổng hợp lại dựa trên phần mềm GIS. Dữ liệu sau đó được chia ngẫu nhiên 64 (70 %) điểm sạt lở để phục vụ công tác xây dựng bản đồ nguy cơ sạt lở đất và 28 (30 %) điểm sạt lở để đánh giá độ chính xác bản đồ đã xây dựng [11].

3.2. Các yếu tố ảnh hưởng đến sạt lở đất

Lựa chọn các yếu tố ảnh hưởng thích hợp để đánh giá sạt lở đất là vấn đề quan trọng. Các yếu tố gây sạt lở đất được phân chia thành 2 loại bao gồm: Yếu tố tự nhiên và yếu tố tác động của con người. Tổng cộng có 12 yếu tố bao gồm: Độ cao địa hình, góc mái dốc, hướng mái dốc, hình dáng bề mặt, bao phủ thực vật (NDVI), khoảng cách đứt gãy, địa chất, khoảng cách đến sông suối, khoảng cách đến đường giao thông, tích lũy dòng chảy, độ ẩm địa hình (TWI), sức mạnh dòng chảy

(SPI) đã được lựa chọn để phân tích và mô hình hóa. Trong nghiên cứu này, mô hình số độ cao (DEM) với độ phân giải 30×30 m được xây dựng từ bản đồ địa hình tỷ lệ 1:50.000 và được trích xuất các yếu tố địa mạo (độ cao địa hình, góc mái dốc, hướng mái dốc, hình dáng bề mặt) và các yếu tố thủy văn (khoảng cách đến sông suối, tích lũy dòng chảy, sức mạnh dòng chảy). 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.

Độ cao địa hình, góc mái dốc, hướng mái dốc, hình dáng bề mặt [12] là yếu tố cần thiết để nghiên cứu tính nhạy cảm sạt lở đất. Các yếu tố này được trích xuất từ dữ liệu DEM và được phân loại thành từng lớp cụ thể. Trong nghiên cứu này, độ cao địa hình được chia thành 9 loại: 142 - 458; 458 - 626; 626 - 779; 779 - 922; 922 - 1063; 1063 - 1206; 1206 - 1371; 1371 - 1586; 1586 - 2184 (Hình 3.a). Góc mái dốc của khu vực nghiên cứu được phân loại lại thành 5 loại như sau: 0 - 12,863767; 12,863767 - 20,973533; 20,973533 - 28,524005; 28,524005 - 37,193066; 37,193066 - 71,310013 (Hình 2.b). Hướng mái dốc được chia thành 9 loại như sau: Phẳng, Bắc, Đông bắc, Đông, Đông nam, Nam, Tây nam, Tây và Tây bắc (Hình 2.c). Hình dáng bề mặt được chia thành 3 loại như sau: Lõm (< -0.05), mặt bằng ($-0.05 - 0.05$) và lồi (> 0.05).

Bản đồ bao phủ thực vật (NDVI) có thể ước tính định lượng sự phát triển và sinh khối của thảm thực vật bằng cách đo độ phản xạ bề mặt [13]. Giá trị NDVI được phân loại lại thành năm loại: -0,016024 - 0,126959; 0,126959 - 0,168266; 0,168266

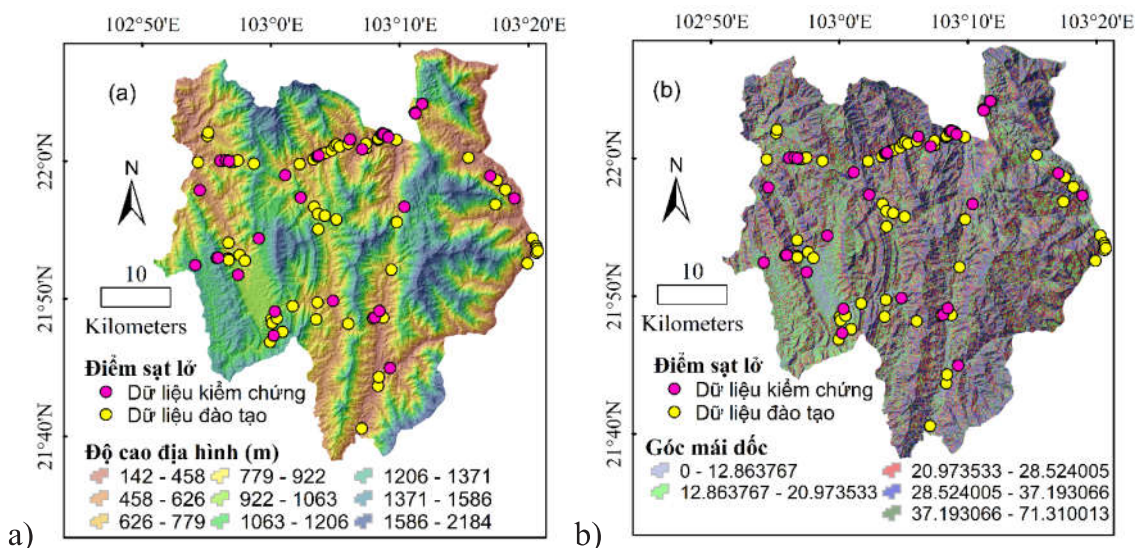
Nghiên cứu

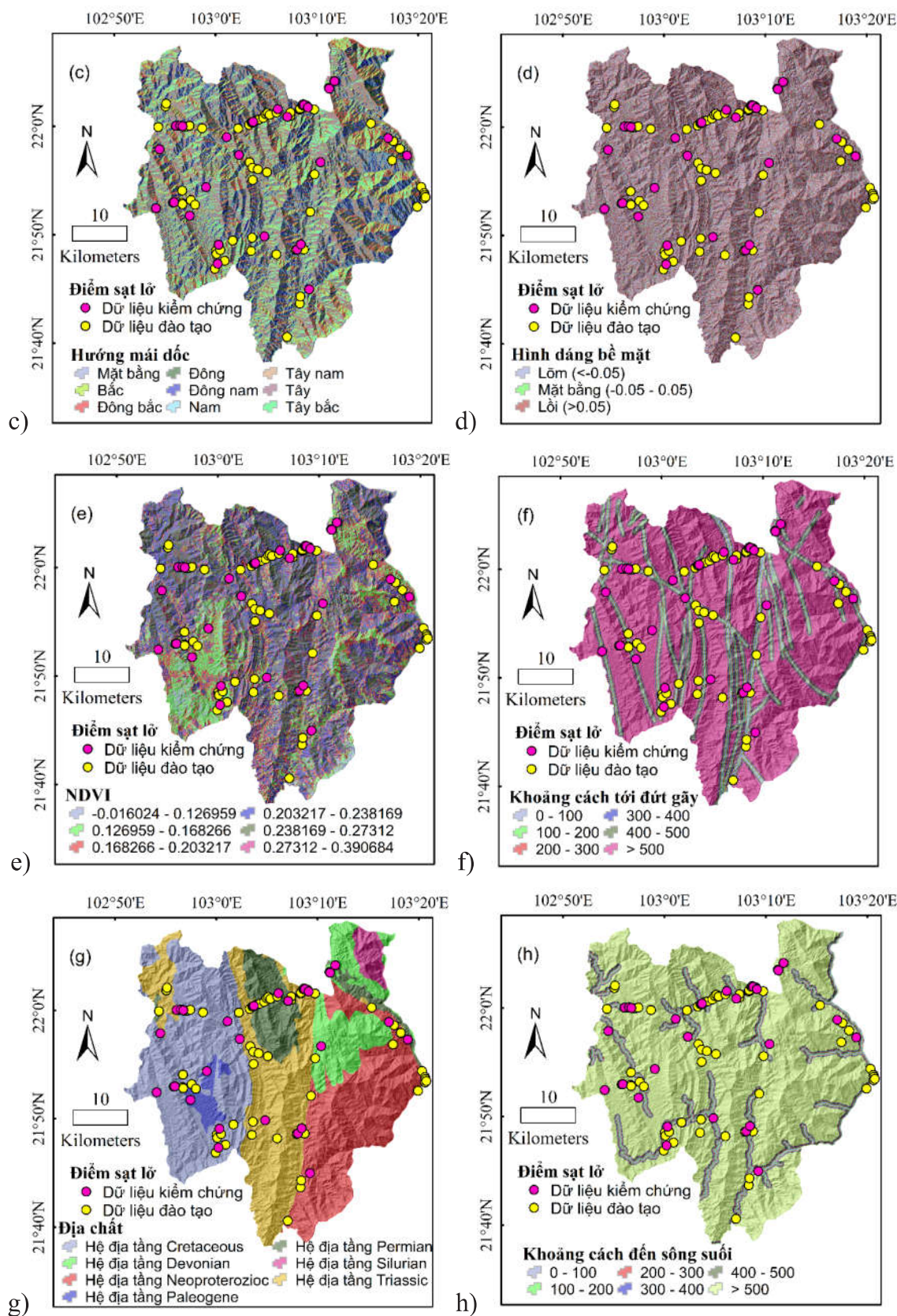
- 0,203217; 0,203217 - 0,238169; 0,238169 - 0,27312; 0,27312 - 0,390684 (Hình 2.e). Đứt gãy là sự chuyển động kiến tạo gây sự mất liên tục trong đất đá [14]. Bản đồ đứt gãy được trích xuất từ bản đồ địa chất với tỷ lệ 1:200.000 và được chia thành 6 lớp: 0 - 100, 100 - 200, 200 - 300, 300 - 400, 400 - 500, > 500 (Hình 2.f). Tại khu vực này, bản đồ địa chất được thu thập từ việc điều tra và nguồn dữ liệu quốc gia và được chia 7 nhóm: Cretaceous, Devonian, Neoproterozoic, Paleogene, Permian, Silurian, Triassic (Hình 2.g).

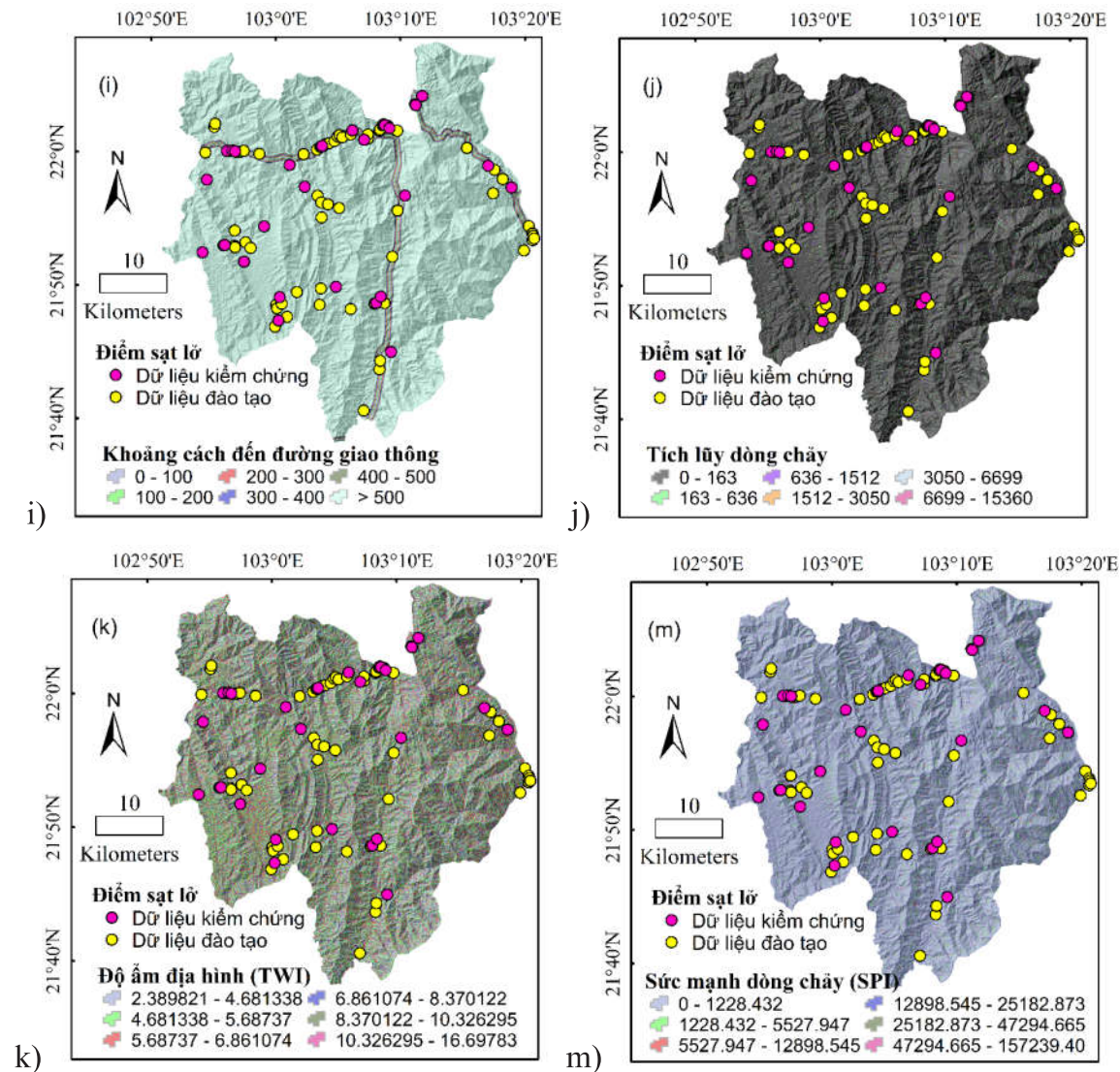
Khoảng cách đến sông và khoảng cách với đường giao thông là những yếu tố quan trọng trong việc lập mô hình tính nhạy cảm với trượt lở đất [15]. Con sông làm cho 2 bên bờ bị xói mòn. Quá trình này làm cho mái dốc dễ bị hỏng hóc hơn. Khoảng cách đến các con sông được chia thành 5 vùng đệm với khoảng cách 100 m (Hình 2.h). Các điều kiện tự nhiên của mái dốc bị phá hủy trong quá trình xây dựng các cơ sở giao thông vận tải và các hoạt động kỹ thuật khác của

con người. Khoảng cách đến đường giao thông được nhóm thành 5 vùng đệm với khoảng cách 100 m (Hình 2.i).

Tích lũy dòng chảy phản ánh lực xói mòn của dòng nước mặt trên mặt đất và được phân thành 6 loại [16]: 0 - 163; 163 - 636; 636 - 1512; 1512 - 3050; 3050 - 6699; 6699 - 15360 (Hình 2.j). TWI ảnh hưởng đến vật liệu mái dốc bằng cách ảnh hưởng đến độ ẩm của đất và dòng chảy của nước ngầm. Trong khu vực nghiên cứu, TWI cũng được chia thành 6 loại: 2,389821 - 4,681338; 4,681338 - 5,68737; 5,68737 - 6,861074; 6,861074 - 8,370122; 8,370122 - 10,326295; 10,326295 - 16,69783 (Hình 2.k). SPI là một yếu tố thủy văn quan trọng, cho thấy quá trình xói mòn dòng chảy tiềm tàng [17]. Trong khu vực nghiên cứu, SPI cũng được chia thành 6 loại: 0 - 12289,432; 1288,432 - 5527,947; 5527,947 - 12898,545; 12898,545 - 25182,873; 25182,873 - 47294,665; 47294,665 - 157239,40 (Hình 2.m).







Hình 2: (a) Độ cao địa hình, (b) Góc mái dốc, (c) Hướng mái dốc, (d) Hình dáng bề mặt, (e) Bao phủ thực vật (NDVI), (f) Khoảng cách đứt gãy, (g) Địa chất, (h) Khoảng cách đến sông suối, (i) Khoảng cách đến đường giao thông, (j) Tích lũy dòng chảy, (k) Độ ẩm địa hình (TWI) khoảng cách đến sông suối, (m) Sức mạnh dòng chảy (SPI)

4. Phương pháp nghiên cứu

4.1. Phân loại hàm cơ sở xuyên tâm (RBFC)

Phân loại hàm cơ sở xuyên tâm RBFC là một mạng nơron có giám sát xem xét một bài toán xấp xỉ trong không gian đa chiều. Trong quá trình học tập này, mạng đang tìm kiếm một bề mặt trong không gian đa chiều, cho phép so sánh tốt hơn tập dữ liệu đào tạo. Tương ứng, dữ liệu thử nghiệm có thể được nội suy bằng cách

sử dụng bề mặt đa chiều. Mạng được cấu tạo bởi ba lớp: Lớp đầu tiên là lớp đầu vào, lớp thứ hai là lớp mặt nạ, và lớp cuối cùng là lớp đầu ra [9]. Mỗi lớp được nhóm theo các phần tử tạo nên các đầu vào và đầu ra. Các phần tử của mỗi lớp được liên kết để truyền thông tin (các phần tử của mỗi lớp không liên quan với nhau).

Trong quá trình truyền thông tin, hàm Gauss được sử dụng làm hàm cơ sở xuyên tâm sau:

$$\phi(x) = \exp(-|x-c|^2 / 2\sigma^2), \sigma > 0 \quad (1)$$

Trong đó: $\phi(x)$ là dữ liệu đầu ra được xác định là các lớp sạt lở hoặc không sạt lở. Phần tử trong lớp ẩn nơi hàm kích hoạt được áp dụng để phân tích mối quan hệ giữa các biến đầu vào và đầu ra, $x = (x_1, \dots, x_n)$ là vector dữ liệu đầu vào của các yếu tố sạt lở đất được liên kết với yếu tố trong lớp ẩn, c được suy ra là tâm của hàm cơ sở và σ là bán kính của hàm cơ sở.

4.2. Phương pháp xác thực

Để đánh giá mô hiệu suất của mô hình, các phương pháp xác thực phổ biến đã được áp dụng trong nghiên cứu này bao gồm: Khu vực đường dưới cong ROC (AUC) và các chỉ số thống kê khác nhau. Đường cong AUC được sử dụng như một tiêu chí để đo độ chính xác dự đoán của mô hình. Các kết quả cụ thể được khái quát và đáng tin cậy. Hai giá trị được sử dụng để xây dựng đường cong ROC: độ nhạy và độ đặc hiệu [18]. Hiệu suất của các mô hình được phân tích định lượng bằng cách sử dụng diện tích dưới đường cong AUC. Giá trị AUC thay đổi giữa “0,5 - 1”, với giá trị bằng 1 cho biết phân loại tốt nhất, trong khi 0,5 tương ứng với các mô hình không chính xác [19]. Giá trị AUC được tính theo công thức:

$$AUC = \frac{(\sum TP + \sum TN)}{(P + N)} \quad (2)$$

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

Các chỉ số thống kê khác nhau như: Độ chính xác (ACC), độ nhạy (SST), độ

đặc hiệu (SPF), giá trị dự đoán tích cực (PPV), giá trị dự đoán âm (NPV), Kappa Index (K), lỗi bình phương trung bình gốc (RMSE), lỗi tuyệt đối (MAE) đã được sử dụng [20]. Với PPV và NPV là các giá trị dự đoán được phân loại chính xác và không chính xác sạt lở đất gây ra. Giá trị K là một biện pháp thống kê để đo lường sự đồng thuận ngẫu nhiên giữa các yếu tố. Giá trị K gần 1 thể hiện độ tin cậy tốt hơn. Giá trị ACC là tỷ số giữa tỷ lệ số lần dự đoán đúng và tổng số lần dự đoán và RMSE thể hiện sự khác biệt giữa quan sát dữ liệu và ước tính dữ liệu.

Công thức tính toán các chỉ số thống kê như sau:

$$SST = \frac{TP}{TP + FN} \quad (3)$$

$$SPF = \frac{TN}{TN + FP} \quad (4)$$

$$PPV = \frac{TP}{FP + TP} \quad (5)$$

$$NPV = \frac{TN}{FN + TN} \quad (6)$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$K = \frac{P_p - P_{exp}}{1 - P_{exp}} \quad (8)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (X_{predicted} - X_{actual})^2} \quad (9)$$

$$MAE = \frac{1}{n} \sum |X_{predicted} - X_{actual}| \quad (10)$$

Trong đó, FP và FN là tỷ lệ phân loại không chính xác sạt lở và không sạt lở, 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

Đánh giá hiệu suất của mô hình RBFC được thể hiện trong (Bảng 1). Đối với dữ liệu đào tạo, kết quả cho thấy tương đối tốt với giá trị PPV (80,56 %), NPV (89,29 %), SST (90,63 %), SPF

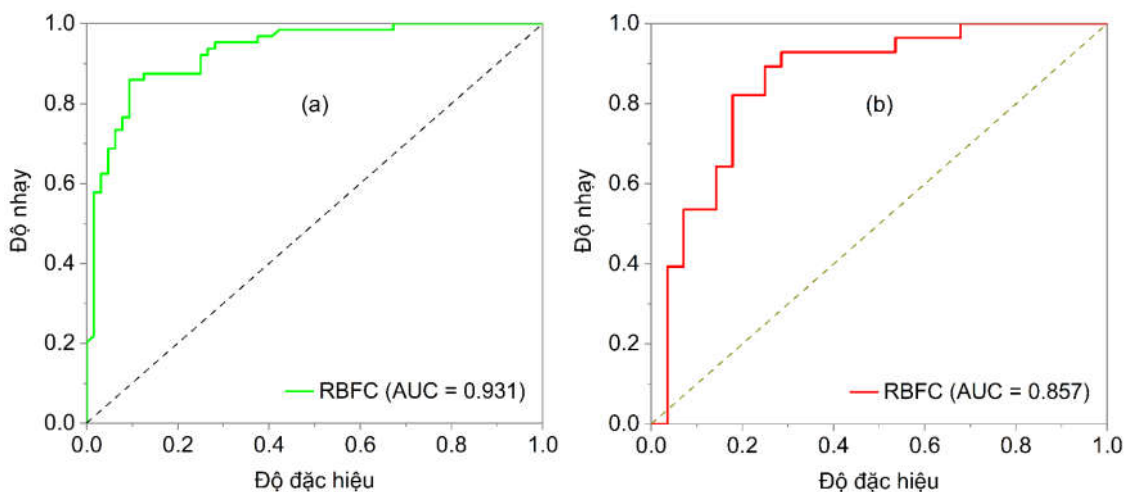
(78,13 %) và ACC (84,38 %). Kết quả dữ liệu kiểm chứng thấp hơn dữ liệu đào tạo, tuy nhiên giá trị hiệu suất đảm bảo với PPV (71,88 %), NPV (79,17 %), SST (82,14 %), SPF (67,86 %) và ACC (75 %). Về giá trị K, với dữ liệu đào tạo có giá trị cao nhất là K (0,69) và dữ liệu kiểm chứng là K (0,50).

Bảng 1. Hiệu suất của mô hình

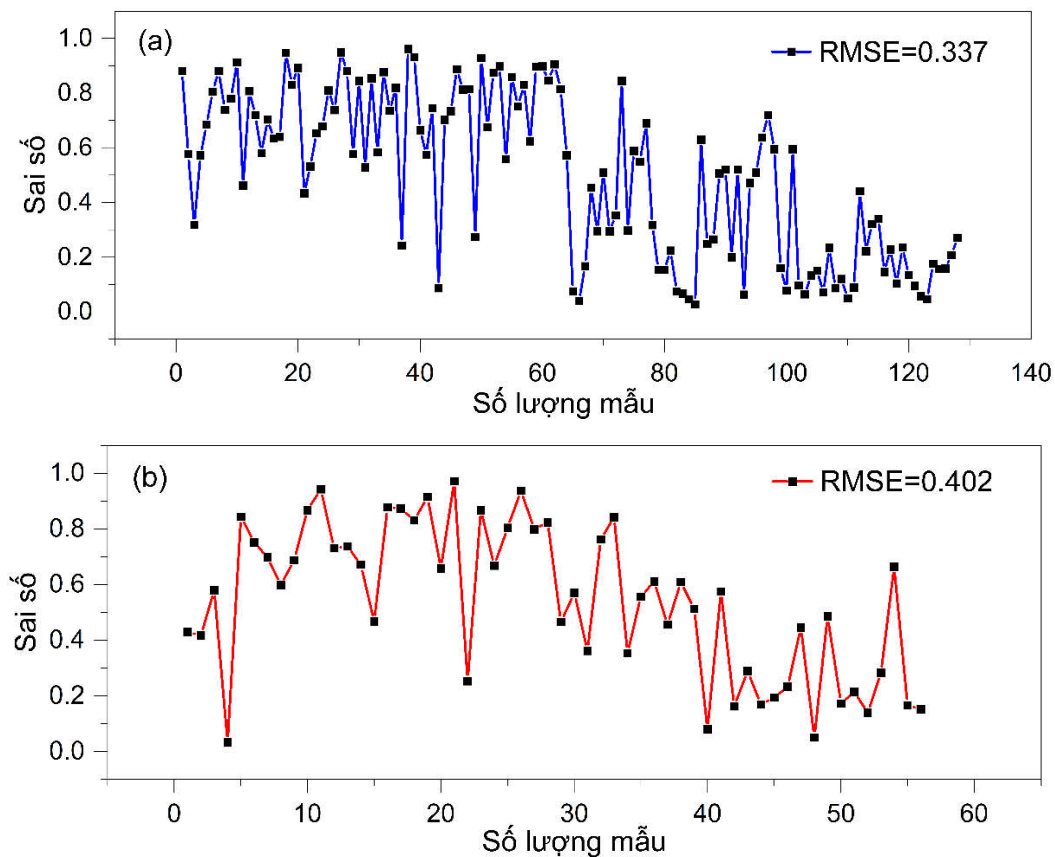
STT	Tham số	Dữ liệu đào tạo	Dữ liệu kiểm chứng
1	TP	58	23
2	TN	50	19
3	FP	14	9
4	FN	6	5
5	PPV (%)	80,56	71,88
6	NPV (%)	89,29	79,17
7	SST (%)	90,63	82,14
8	SPF (%)	78,13	67,86
9	ACC (%)	84,38	75,00
10	K	0,69	0,50
11	MAE	0,28	0,34

Kết quả đường cong ROC/AUC chỉ ra rằng, mô hình phân loại hàm cơ sở xuyên tâm RBFC có kết quả dự đoán đào tạo tốt với $AUC_{\text{đào tạo}} = 0,931$ và kết quả kiểm chứng tương ứng là $AUC_{\text{Kiểm chứng}} = 0,857$. Với giá trị RMSE (Hình 4), mô hình

RBFC có giá trị RMSE với dữ liệu đào tạo (0,337) và dữ liệu xác thực (0,402). Dựa trên những kết quả này, có thể khẳng định rằng hiệu suất của bộ dữ liệu đào tạo tốt hơn bộ dữ liệu kiểm chứng.

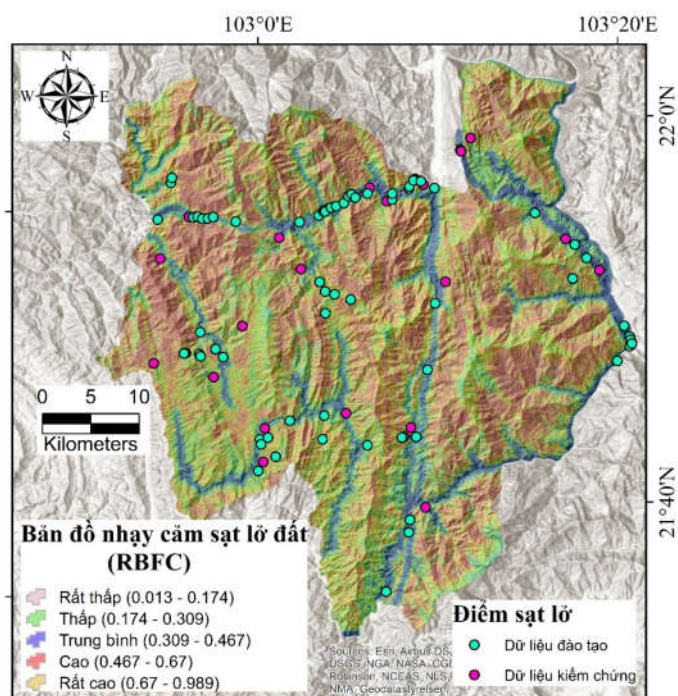


Hình 3: Giá trị AUC của mô hình: (a) Bộ dữ liệu đào tạo, (b) Bộ dữ liệu kiểm chứng



Hình 4: Giá trị lỗi bình phương của mô hình: (a) Giá trị đào tạo; (b) Giá trị kiểm chứng

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



Hình 5: Bản đồ nhạy cảm sạt lở đất khu vực nghiên cứu

Bản đồ nhạy cảm sạt lở đất được xây dựng bằng mô hình RBFC với năm lớp: Rất thấp, thấp, trung bình, cao và rất cao (Hình 5). Sự phân bố của từng lớp nhạy cảm trên các bản đồ được thể hiện trong (Bảng 2). Trong bản đồ được xây dựng bởi mô hình RBFC, tỷ lệ phần trăm trên mỗi lớp nhạy cảm là 31,31 % khu vực nghiên cứu ở mức độ rất thấp, 19,58 % ở

mức độ trung bình và 7,67 % ở mức độ rất cao. Việc xác thực bản đồ sử dụng tỷ lệ tần số, tỷ lệ phần trăm sạt lở đất trên mỗi lớp nhạy cảm cũng thu được kết quả như trong Bảng 2. Kết quả kiểm chứng cho thấy hầu hết các điểm sạt lở đất được quan sát ở mức cao và rất cao. Như vậy, có thể nói bản đồ nhạy cảm sạt lở đất có độ tin cậy cao.

Bảng 2. Phần trăm các vụ sạt lở đất và tỷ số tần suất của các lớp nhạy cảm sạt lở đất

STT	Tỷ số tần suất	Tỷ lệ phần trăm của lớp nhạy cảm	Tỷ lệ phần trăm sạt lở trên mỗi lớp nhạy cảm
Rất thấp (0,013 - 0,174)	0,114	31,31 %	3,57 %
Thấp (0,174 - 0,309)	0,13	27,51 %	3,57 %
Trung bình (0,309 - 0,467)	1,094	19,58 %	21,43 %
Cao (0,467 - 0,67)	1,806	13,84 %	25 %
Rất cao (0,67 - 0,989)	5,98	7,76 %	46,43 %

6. Kết luận

Xác định các khu vực nhạy cảm sạt lở đất được coi là vấn đề quan trọng nhất để quản lý rủi ro và quy hoạch sử dụng đất. Mặc dù có một số phương pháp khác nhau được phát triển và áp dụng để dự báo nhạy cảm sạt lở đất trên khắp thế giới, việc xây dựng bản đồ nhạy cảm sạt lở đất sử dụng các phương pháp phù hợp cho một khu vực cụ thể vẫn là một chủ đề được các nhà nghiên cứu quan tâm. Trong nghiên cứu này, mô hình RBFC đã được áp dụng để lập bản đồ nhạy cảm sạt lở đất vì đây là những mô hình máy học tiên tiến và hiệu quả để dự báo và đánh giá tai biến thiên nhiên. Kết quả phân tích thống kê cho thấy mô hình RBFC có hiệu suất tốt ($AUC_{\text{đào tạo}} = 0,931$, $AUC_{\text{kiểm chứng}} = 0,857$). Bản đồ phân vùng nhạy cảm cho thấy rằng, mô hình ADT đã phân định khoảng 25 % và 46,43 % diện tích khu vực nghiên cứu vào các khu vực có mức độ nhạy cảm cao và rất cao trong. Đánh giá kết quả cho

thấy bản đồ nhạy cảm sạt lở đất có độ tin cậy cao và có thể ứng sử dụng trong quản lý, quy hoạch sử dụng đất. Tuy nhiên, với mức độ phức tạp của sạt lở đất và các yếu tố ảnh hưởng, cần phải so sánh nhiều phương pháp mô hình hơn và có thể xác định các nhóm đặc điểm và yếu tố khác nhau bằng cách sử dụng các kỹ thuật khác nhau để có thể đưa ra các quan điểm khác nhau về việc lựa chọn và cải thiện hiệu suất của các mô hình học máy.

TÀI LIỆU THAM KHẢO

- [1]. Haque, U, Da Silva, P.F, et al. (2019). *The human cost of global warming: Deadly landslides and their triggers (1995 - 2014)*. Science of the Total Environment, 682.
- [2]. Reichenbach, P, Rossi, M, et al. (2018). *A review of statistically - based landslide susceptibility models*. Earth - science reviews, 180.
- [3]. Flentje, P. N, Miner, A, et al. (2007). *Guidelines for landslide susceptibility, hazard and risk zoning for land use planning*.
- [4]. Darema, F (2004). *Dynamic data driven applications systems: A new*

paradigm for application simulations and measurements. International conference on computational science. Springer.

[5]. Agatonovic - Kustrin, S, Beresford, R. (2000). *Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research*. Journal of Pharmaceutical and Biomedical analysis, 22(5).

[6]. Wang, H, Hu, D. (2005). *Comparison of SVM and LS - SVM for regression*. International conference on neural networks and brain. IEEE.

[7]. Pham, B. T, Tien Bui, D. and Prakash, I. (2017). *Landslide susceptibility assessment using bagging ensemble based alternating decision trees, logistic regression and J48 decision trees methods: A comparative study*. Geotechnical and Geological Engineering, 35(6).

[8]. Karaboga, D, Kaya, E. (2019). *Adaptive network based fuzzy inference system (ANFIS) training approaches: A comprehensive survey*. Artificial Intelligence Review, 52(4).

[9]. Pham, B. T, Phong, T. V et al. (2020). *A comparative study of kernel logistic regression, radial basis function classifier, multinomial naïve bayes and logistic model tree for flash flood susceptibility mapping*. Water, 12(1).

[10]. Van Westen, C, Van Asch, T. W and Soeters, R. (2006). *Landslide hazard and risk zonation - why is it still so difficult?* Bulletin of Engineering geology and the Environment, 65(2).

[11]. Bui, K - T. T, Tien Bui, D et al. (2018). *A novel hybrid artificial intelligent approach based on neural fuzzy inference model and particle swarm optimization for horizontal displacement modeling of hydropower dam*. Neural Computing and Applications, 29(12).

[12]. Luu, C, Nguyen, D. D et al. (2022). *Using decision tree J48 based machine learning algorithm for flood susceptibility mapping: A case study in Quang Binh province, Vietnam*. In CIGOS 2021, Emerging

Technologies and Applications for Green Infrastructure. Springer. p. 1927 - 1935.

[13]. Tran, H. T, Campbell, J. B et al. (2017). *Monitoring drought vulnerability using multispectral indices observed from sequential remote sensing (Case study: Tuy Phong, Binh Thuan, Vietnam)*. GIScience & Remote Sensing, 54(2).

[14]. Phong, T. V, Pham, B. T et al. (2021). *Groundwater potential mapping using GIS - based hybrid artificial intelligence methods*. Groundwater, 59(5).

[15]. Bui, D.T, Lofman, O et al. (2011). *Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression*. Natural hazards, 59(3).

[16]. Chen, W, Fan, L et al. (2019). *Spatial prediction of landslides using hybrid integration of artificial intelligence algorithms with frequency ratio and index of entropy in Nanzheng county, China*. Applied Sciences, 10(1)

[17]. Chapi, K, Singh, V. P et al. (2017). *A novel hybrid artificial intelligence approach for flood susceptibility assessment*. Environmental modelling & software, 95.

[18]. Tien Bui, D, Shirzadi, A et al. (2019). *A novel ensemble artificial intelligence approach for gully erosion mapping in a semi - arid watershed (Iran)*. Sensors, 19(11).

[19]. Danh, N.T, Ngô, Đ.V, Dũng, T. Q. (2017). *Ứng dụng phương pháp hồi quy logistic xác định tổ hợp tối ưu các yếu tố ảnh hưởng và xây dựng bản đồ tai biến trượt lở đất huyện Khánh Vĩnh, tỉnh Khánh Hòa*. Science and Technology Development Journal, 20(K4 - 2017).

[20]. Pham, B.T, Nguyen, M. D et al. (2019). *A novel artificial intelligence approach based on multi - layer perceptron neural network and biogeography - based optimization for predicting coefficient of consolidation of soil*. Catena, 173.

Ngày nhận bài: 24/6/2022; Ngày chấp nhận
đăng: 27/9/2022