

# SỬ DỤNG PHƯƠNG PHÁP TỶ SỐ TẦN SUẤT VÀ CÁC PHƯƠNG PHÁP HỌC MÁY ĐỂ THÀNH LẬP BẢN ĐỒ NHẠY CẢM TRƯỢT LỞ. KHU VỰC THỬ NGHIỆM: XÃ PHÌN NGAN, TỈNH LÀO CAI

Đỗ Minh Hiền<sup>1</sup>, Nguyễn Văn Hoàng<sup>1</sup>, Mai Lê Dũng<sup>1</sup>  
Ngô Thị Hương<sup>1</sup>, Lương Hữu Dũng<sup>2</sup>, Nguyễn Bình Phong<sup>3</sup>

<sup>1</sup>Viện Khoa học Địa chất và Khoáng sản

<sup>2</sup>Viện Khoa học Khí tượng Thủy văn và Biến đổi Khí hậu

<sup>3</sup>Trường Đại học Tài nguyên và Môi trường Hà Nội

## Tóm tắt

Đánh giá tính nhạy cảm với tai biến trượt lở khu vực xã Phìn Ngan, tỉnh Lào Cai, Việt Nam đã được thực hiện bằng cách áp dụng ba mô hình học máy là hồi quy logic (LR), mạng Bayes (BN), máy véc tơ hỗ trợ (SVM) và phương pháp thống kê tỷ số tần suất (FR) mà FR được sử dụng để tính toán các giá trị trọng số của mỗi lớp tham số trong các bản đồ tác nhân. Các bản đồ có trọng số này sau đó được kết hợp với bản đồ trượt lở để đánh giá mối quan hệ không gian của chúng. Tiếp theo, các mô hình học máy sẽ được áp dụng để tính toán mức độ quan trọng của từng bản đồ tác nhân gây trượt lở. Hiệu suất của các mô hình học máy đã được đánh giá bằng cách sử dụng đường cong đặc tính hoạt động thu được (ROC) và diện tích dưới đường cong (AUC). Phân tích và so sánh kết quả cho thấy cả 3 mô hình đều cho kết quả tốt khi đánh giá tính nhạy cảm với trượt lở đất (AUC = 87,2 - 97,5 %). Tuy nhiên, mô hình BN (AUC = 97,5 %) có hiệu suất tốt nhất so với các mô hình trượt lở khác, tiếp theo là mô hình LR (AUC = 94,6 %) và mô hình SVM (AUC = 87,2 %). Kết quả chỉ ra rằng các mô hình đã cho kết quả đầu ra với khả năng dự báo tốt. Chúng cũng rất hữu ích trong việc hỗ trợ lập kế hoạch sử dụng đất, phòng ngừa và giảm thiểu rủi ro do sạt lở đất trong khu vực nghiên cứu.

**Từ khóa:** Bản đồ tính nhạy cảm trượt lở; Tỷ lệ tần suất; Hồi quy logistic; Mạng bayes; Máy vector hỗ trợ.

## Abstract

**Landslide susceptibility mapping by using frequency ratio method and machine learning models: A case study of Phin Ngan Commune, Lao Cai Province, Vietnam**

Landslide susceptibility assessment in Phin Ngan commune, Lao Cai province, Vietnam has been carried out by applying three machine learning (ML) models (Logistic Regression - LR, Bayesian Network - BN and Support Vector Machines - SVM), and the Frequency Ratio (FR) method. First, FR was applied to calculate the weighting values of each parameter classes in the factor maps. Second, ML models were applied to calculate the importance of each landslide related factor map. Then, the landslide susceptibility index (LSI) maps were generated by combining the importance values of factor maps obtained from ML models and the parameter classes that was assigned the weighting values of factor maps created by FR model. Next, the performance of

*these ML methods has been evaluated using receiver operating characteristic (ROC) curve and area under the curve (AUC). Analysis and comparison of the results showed that all three ML models performed well for landslide susceptibility assessment (AUC = 87.2 % - 97.5 %). The BN model (AUC = 97.5 %) showed the best performance in comparison to other landslide models (LR model: AUC = 94.6 % and the SVM model: AUC = 87.2 %). The results indicated that the models have given outputs with good forecasting ability. They are also very useful in supporting land - use planning, the prevention and mitigation of risks due to landslides in the research area.*

**Keywords:** Landslide susceptibility mapping; Frequency ratio; Logistic regression; Bayesian network; Support vector machines.

### **1. Giới thiệu**

Trong những năm gần đây, công tác điều tra, quản lý và phòng chống rủi ro tai biến trượt lở, lũ bùn đá đang ngày càng được quan tâm nghiên cứu ở Việt Nam. Phân loại nghiên cứu về đánh giá rủi ro và tai biến trượt lở được phân chia thành hai hệ phương pháp chính, các phương pháp định tính và các phương pháp định lượng. Các phương pháp định tính nói chung dựa trên đánh giá của một hay nhiều chuyên gia thực hiện đánh giá tai biến [33]. Phương pháp chuyên gia có thể phân chia thành hai loại: Phân tích đặc điểm địa mạo từ thực địa và sự kết hợp hay chồng chập các bản đồ chỉ số theo trọng số [1]. Các phương pháp định lượng có tính chặt chẽ hơn gồm các phương pháp phân tích thống kê, các phương pháp học máy và các phương pháp tiên định dựa trên các mô hình ổn định sườn, mô hình thủy văn và các mô hình khác về địa chất công trình [1, 11]. Giới hạn của phương pháp tiên định là chỉ thích hợp cho các khu vực mà ở đó chuyển động trượt lở không phức tạp và các đặc điểm địa chất, địa mạo ở đó là khá đồng nhất [63].

Mô hình thống kê và học máy là các phương pháp tiếp cận theo hướng dữ liệu, cả hai đều tập trung vào việc phân tích

các tác nhân gây trượt lở đất bằng cách sử dụng bộ dữ liệu trượt đất trong quá khứ và hiện tại [22, 26, 72]. Phương pháp được sử dụng phổ biến nhất trong thống kê hai biến là tỷ số tần suất (FR) [30, 31, 32, 39, 56, 68]. Các phương pháp khác như mô hình trọng số bằng chứng (WoE) [12, 13, 23, 39, 42, 44, 57, 60, 65, 67], giá trị thông tin (IV) [7, 37] và chỉ số thống kê (SI) [56, 66, 73]. Đối với thống kê đa biến, phương pháp hồi quy logic (LR) và phân tích biệt thức (DA) là hai phương pháp được sử dụng nhiều nhất [23, 32, 31, 42, 54, 66].

Khoảng mười năm gần đây, các thuật toán học máy đã và đang được áp dụng rộng rãi trong nghiên cứu tai biến nói chung và trượt lở nói riêng như máy véc tơ hỗ trợ (SVM) [16, 25, 45, 55], mạng thần kinh nhân tạo (ANN) [20, 41, 51, 52, 53, 71], cây quyết định (DT) [25, 54] và mạng Bayes (BN) [3].

Ở Việt Nam, không ít các kết quả nghiên cứu về trượt lở đã được công bố trên các tạp chí quốc tế. Một số nghiên cứu ở các khu vực khác nhau đã áp dụng các phương pháp chuyên gia, thống kê để thành lập bản đồ phân vùng nhạy cảm trượt lở [27, 34, 35], các mô hình kết hợp, chuyên gia và thống kê [23] và các mô hình học máy [3, 4, 5, 15, 17, 18].

## Nghiên cứu

Mặc dù việc nghiên cứu và quản lý tai biến địa chất đặc biệt là trượt lở, lũ bùn đá, lũ quét ở Việt Nam đã được quan tâm chú ý với hàng loạt các đề án nghiên cứu lớn. Tuy nhiên, các bản đồ kết quả thường được thành lập ở tỷ lệ nhỏ và trung bình (Từ tỷ lệ 1:500.000 đến 1:50.000) nên mang tính phác họa và khái quát cao. Do đó, việc thành lập bản đồ phân vùng nguy cơ trượt lở ở tỷ lệ lớn (1:10.000) là cấp bách và mang ý nghĩa về mặt dự báo cho các khu vực có tiềm năng dẫn đến các tai biến địa chất như trượt lở, lũ bùn đá trong tương lai.

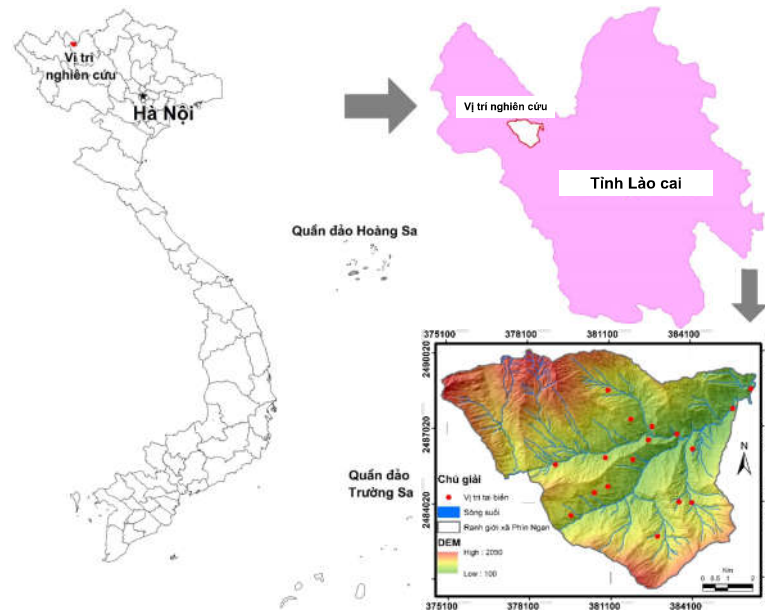
Do đó, mục đích chính của nghiên cứu này là thành lập bản đồ nhạy cảm trượt lở ở tỷ lệ 1:10.000 với khả năng dự báo tốt để hỗ trợ cho chính quyền địa phương trong việc ra quyết định liên quan đến công tác quy hoạch, phòng chống và giảm thiểu rủi ro. Để thực hiện mục tiêu này, nghiên cứu đã áp dụng mô hình kết hợp giữa phương pháp thống kê tỷ số tần suất (FR) và ba phương pháp học máy là mạng Bayes (BN), hồi quy logic (LR) và máy véc tơ hỗ trợ (SVM) để thành lập bản đồ nhạy cảm trượt lở. Các bản đồ đầu vào được sử dụng chủ yếu là dữ liệu được triết xuất từ bản đồ địa hình tỷ lệ 1:10.000 bao gồm: Bản đồ độ dốc, mặt cong địa hình, chỉ số độ ẩm địa hình (TWI), chỉ số năng lượng dòng chảy (SPI), chỉ số vận chuyển trầm tích (STI), bản đồ vỏ phong hóa và bản đồ thảm phủ được thành lập từ ảnh vệ tinh phân giải cao Google Earth để thành lập bản đồ phân vùng nguy cơ tai biến trượt lở và lũ bùn đá ở tỷ lệ 1:10.000.

## **2. Khu vực nghiên cứu**

Khu vực nghiên cứu (Hình 1) là Xã Phìn Ngan nằm ở phía Nam của huyện Bát Xát, cách huyện lỵ khoảng 8 km về phía Nam và cách thành phố Lào Cai khoảng 16 km về phía Tây bắc. Diện tích xã Phìn Ngan xấp xỉ 66 km<sup>2</sup> giữa 103°47'10" đến 103°53'56" kinh độ Đông và 22°30'14" đến 22°25'03" vĩ độ Bắc và có địa hình phức tạp, vùng núi cao có độ chia cắt lớn, thung lũng hẹp, khe sâu, độ dốc lớn. Vùng thấp là nơi tập trung các dải đồi thấp, thoải địa hình tương đối bằng phẳng. Độ cao địa hình thay đổi từ 100 đến 2050 m.

Đặc điểm khí hậu của khu vực nghiên cứu là khí hậu nhiệt đới, nóng ẩm mưa nhiều. Trong đó, địa hình vùng núi cao mang tính chất của khí hậu cận nhiệt đới và ôn đới ẩm. Mùa nóng từ tháng 5 đến tháng 10, mùa lạnh từ tháng 11 đến tháng 4 năm sau, nhiệt độ trung bình cả năm cao nhất 17 °C, thấp nhất 14 °C; vùng thấp thì có nhiệt độ trung bình cả năm cao hơn so với vùng cao. Lượng mưa trung bình năm của xã Phìn Ngan khoảng 1500 - 2000 mm. Lượng mưa tập trung chủ yếu từ tháng 5 đến tháng 9 với lượng mưa thay đổi từ 150 mm đến 500 mm (theo số liệu mưa trạm Lào Cai).

Về địa chất, gần như toàn bộ diện tích xã Phìn Ngan nằm trọn trong phức hệ Po Sen với pha 1 (GD<sub>i</sub>/PZ1<sub>ps</sub>1) chủ yếu là diorit thạch anh, granodiorit và pha 2 (G/PZ1<sub>ps</sub>2) là granit, granit biotit. Có một phần phía Tây bắc diện tích nghiên cứu nằm trong hệ tầng Đá Đỉnh (NP<sub>dd</sub>) gồm đá hoa, đá hoa dolomit, dolomit, đá hoa tremolit. Trong khu vực nghiên cứu rất khó quan sát được các đá gốc ngoại trừ một vài vị trí lộ dọc theo các khu vực sông, suối.



**Hình 1: Khu vực nghiên cứu xã Phìn Ngan, tỉnh Lào Cai**

### 3. Dữ liệu

Trong khu vực nghiên cứu đã ghi nhận được 17 điểm trượt lở trong đó có 3 điểm là giải đoán viễn thám, 14 điểm là khảo sát thực địa. Trong 14 điểm này thì 10 điểm trượt lở được thu thập từ đề án trượt lở quốc gia do Viện Khoa học Địa chất và Khoáng sản thực hiện. Trong đó, có 9 điểm là trượt xoay, 1 điểm là trượt hỗn hợp, với 4 điểm quy mô nhỏ (< 200 m<sup>3</sup>), 3 điểm quy mô trung bình (200 - 1000 m<sup>3</sup>) và 3 điểm quy mô lớn (1.000 - 20.000 m<sup>3</sup>). Còn lại 4 điểm mới gồm 2 điểm trượt lở mới và 2 điểm lũ bùn đá đã được thu thập thêm trong quá trình khảo sát thực địa tháng 12 năm 2021. Các điểm trượt lở đều xảy ra ở đới phong hóa mạnh và phong hóa hoàn toàn, lớp vỏ phong hóa dày (từ 5 - 25 m) trên nền đá gốc là đá granit màu xám, xám đen, đá granit biotit, đá granodiorit và diorit thạch anh.

Để thành lập bản đồ nhạy cảm trượt lở, ngoài số liệu trượt lở như đã trình bày ở trên, lựa chọn các bản đồ tác nhân gây trượt lở để làm dữ liệu tính toán phải đảm

bảo về mặt tỷ lệ tương ứng với bản đồ đầu ra. Như vậy, một số bản đồ tác nhân như bản đồ địa chất hay thạch học, khoảng cách đến đứt gãy được loại trừ do các bản đồ này ở tỷ lệ 1:50.000 nên không đảm bảo được tính chi tiết của dữ liệu. Tác nhân kích hoạt như lượng mưa trung bình năm cũng không sử dụng làm dữ liệu đầu vào vì mức độ chi tiết của nó không đảm bảo để tính toán. Như vậy, yếu tố dữ liệu đầu vào để thành lập bản đồ nhạy cảm trượt lở ở tỷ lệ lớn cần tập trung vào 3 nhóm: (1) Nhóm các yếu tố thể hiện yếu tố thành phần vật chất, trong bài báo này đã lựa chọn là bản đồ vỏ phong hóa; (2) Nhóm yếu tố hình thái địa mạo thể hiện được các chỉ số địa hình thuận lợi cho tai biến trượt lở như chỉ số độ ẩm địa hình (TWI), chỉ số năng lượng dòng chảy (SPI), chỉ số vận chuyển trầm tích (STI) và mặt cong địa hình thể hiện mặt lồi, mặt lõm, mặt phẳng địa hình; (3) Yếu tố liên quan đến kích hoạt tai biến, trong bài báo này không lựa chọn lượng mưa vì lý do không đủ chi tiết, do đó, tác nhân tham phủ đã được

## **Nghiên cứu**

lựa chọn. Như vậy, 7 tác nhân chính được xác định cho vùng nghiên cứu là độ dốc, mặt cong địa hình, TWI, SPI, STI, thảm phủ và vỏ phong hóa. Trong đó 5 bản đồ tác nhân đầu được thành lập từ bản đồ địa hình tỷ lệ 1:10.000 của khu vực nghiên cứu. Bản đồ thảm phủ được thành lập từ ảnh vệ tinh phân giải cao Google Earth với độ phân giải 1 m và bản đồ vỏ phong hóa được thành lập từ công tác thu thập, phân tích mẫu hóa kết hợp với phân tích độ dốc địa hình ở tỷ lệ 1:10.000.

### **3.1. Độ dốc**

Một mô hình số độ cao (Hình 2a) (DEM) với kích thước pixel là  $5 \times 5$  m đã được tạo ra từ bản đồ địa hình của khu vực nghiên cứu tỷ lệ 1:10.000 với khoảng cách các đường bình độ là 10 m bằng phần mềm ArcGIS 10.8. Bản đồ độ dốc của khu vực nghiên cứu được triết xuất từ DEM sử dụng hàm độ dốc trong phần mềm ArcGIS 10.8. Giá trị độ dốc của tại khu vực nghiên cứu được phân thành 5 nhóm ( $0 - 10^\circ$ ;  $10^\circ - 20^\circ$ ;  $20^\circ - 30^\circ$ ;  $30^\circ - 40^\circ$  và  $> 40^\circ$ ) với độ dốc thay đổi từ  $0^\circ - 76.88^\circ$  (Hình 2b).

### **3.2. Mặt cong địa hình**

Mặt cong địa hình thể hiện tốc độ xói mòn, tốc độ dòng chảy, sự lắng đọng và kiểm soát sự thay đổi về vận tốc của các khối vật liệu chuyển động xuống phía dưới. Nó cho ta biết được các hình dạng mặt lõm, mặt lồi và khu vực phẳng của địa hình [24, 69] (Hình 2c).

### **3.3. Chỉ số độ ẩm địa hình - TWI**

TWI (Hình 2d) là một trong những tác nhân quan trọng ảnh hưởng đến sự mất ổn định của sườn [48, 49]. Dòng chảy tác động bất lợi tới sự ổn định bằng việc xói mòn các sườn. TWI được mô tả ảnh hưởng của địa hình đối với vị trí và kích

thước của các khu vực nguồn bão hòa tạo ra dòng chảy. Bản đồ TWI của khu vực nghiên cứu được phân thành 5 lớp để phục vụ cho các bước tính toán tiếp theo. Ngoài ra, TWI có thể chỉ ra độ ẩm của đất, các khu vực bão hòa, chiều sâu mực nước ngầm và dòng tích lũy [50, 70]. Tính toán chỉ số độ ẩm địa hình theo công thức (1) dưới đây được đề xuất bởi Pradhan và cộng sự [50]:

$$TWI = \log_e \left( \frac{A}{b \tan \beta} \right) \quad (1)$$

Trong công thức (1) và (2), A ( $m^2$ ) là tích lũy dòng chảy, b (m) là chiều rộng ô pixel mà nước chảy qua,  $\beta$  (radian) là độ dốc sườn.

### **3.4. Chỉ số năng lượng dòng chảy - SPI**

SPI (Hình 2e) có thể cho ta hiểu năng lượng xói mòn tiềm năng của các dòng chảy trên bề mặt địa hình, chiều dày của đất và sự phân bố của thực vật [50, 70]. SPI dùng để ước lượng năng lượng xói mòn của nước. Giá trị xói mòn càng cao ở dọc các kênh, rãnh thoát nước [49]. SPI được đề xuất bởi Pradhan và cộng sự (2014) [48] theo công thức sau:

$$SPI = A \tan \beta / b \quad (2)$$

### **3.5. Chỉ số vận chuyển trầm tích - STI**

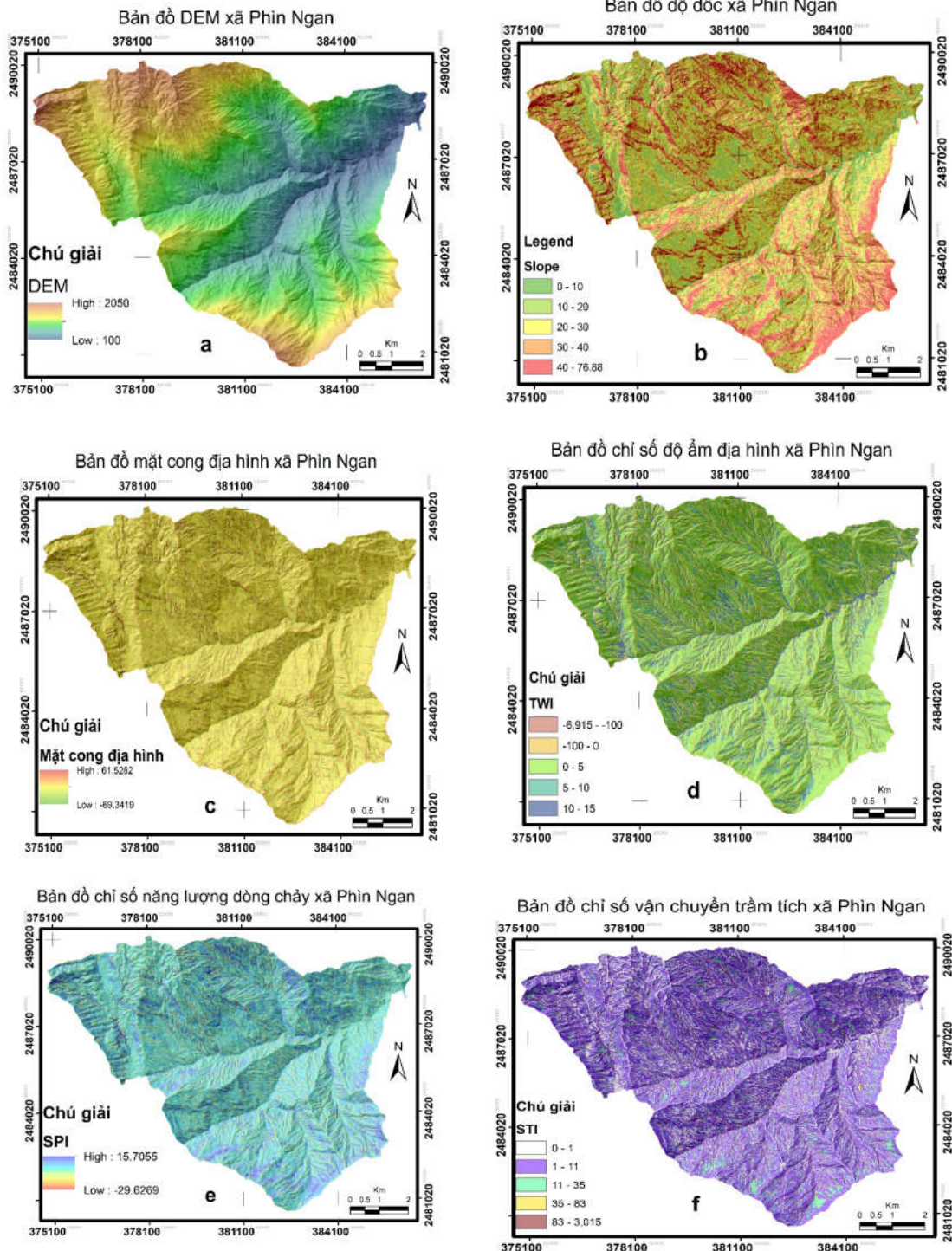
STI (Hình 2f) được tính toán bằng cách kết hợp các yếu tố độ dốc (chiều dài và độ dốc). STI đặc trưng cho quá trình xói mòn và bồi tụ [49]. STI được tính toán theo công thức (3) do Burrough and McDonnell (1998) [6] đề xuất. Trong công thức (3),  $A_s$  là diện tích lưu vực và  $\beta$  là độ dốc.

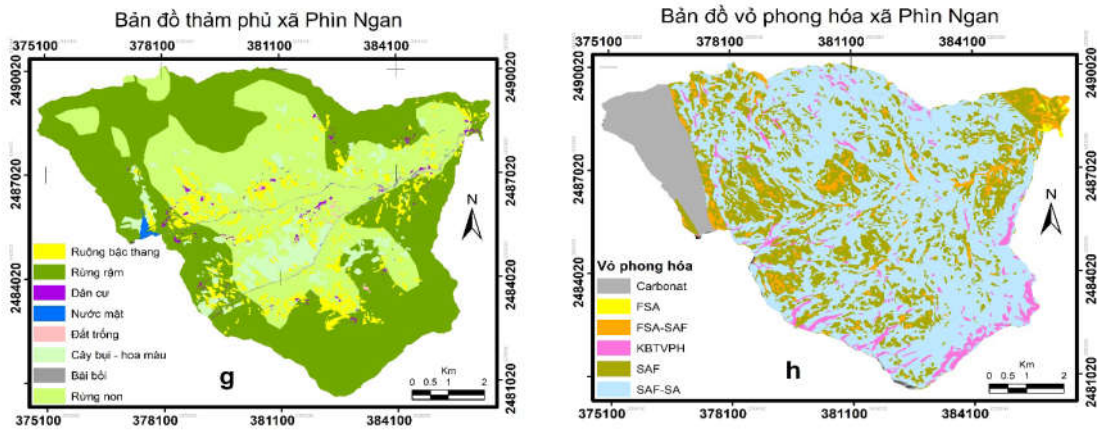
$$STI = (A_s / 22.13)^{0.6} \times \sin(\beta / 0.0896)^{1.3} \quad (3)$$

**3.6. Thảm phủ/Đất sử dụng - LULC**

Bản đồ thảm phủ của khu vực nghiên cứu được thành lập từ ảnh vệ tinh phân giải cao được tải từ phần mềm Google Earth Pro với ngày thu nhận ảnh là tháng 11 năm 2020 với độ phân giải  $1 \times 1$  m. Bản

đồ này (Hình 2g) được phân loại thành 8 lớp: Rừng rậm, rừng non, đất trống, cây bụi - hoa màu, dân cư, nước mặt, ruộng bậc thang và bãi bồi. Nếu không có sẵn dữ liệu thảm phủ thì có thể sử dụng bản đồ sử dụng đất để thay thế.





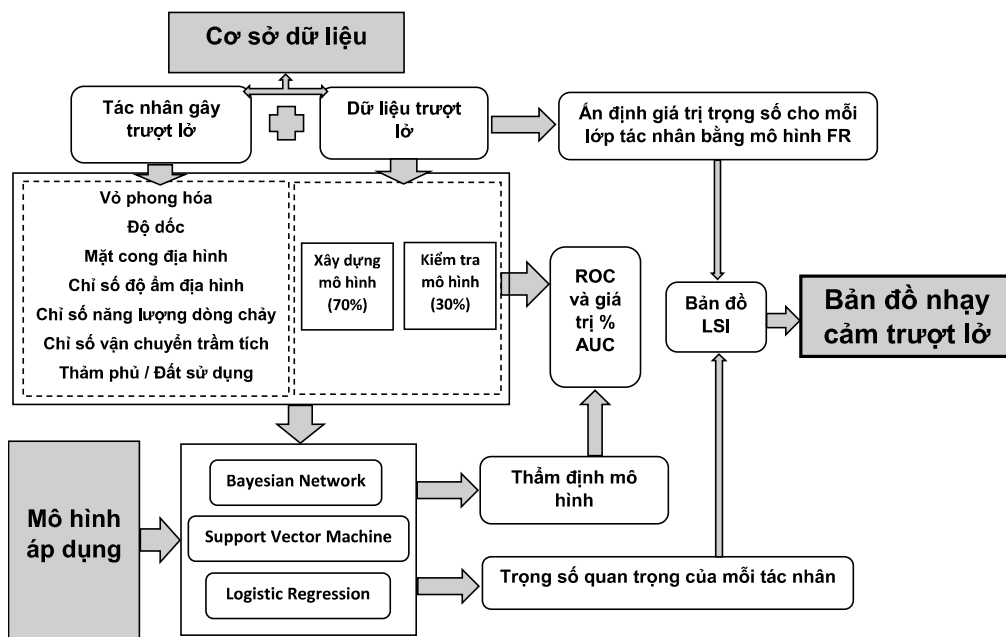
**Hình 2: Các bản đồ tác nhân trượt lở (a - DEM, b - Độ dốc, c - Mặt cong địa hình, d - TWI, e - SPI, f - STI, g - LULC và h - Vỏ phong hóa)**

**3.7. Vỏ phong hóa**

Bản đồ vỏ phong hóa đã được phân loại dựa trên đặc tính của các loại đá gốc và các đặc tính cơ học, khoáng vật học và hóa học của đất. Bản đồ này được thu thập từ Viện Khoa học Địa chất và Khoáng sản và được phân

thành 6 kiểu vỏ phong hóa gồm: Carbonat, FSA - Vỏ phong hóa Ferosialit, FSA - SFA - Vỏ phong hóa Ferosialit - Sialferit, SAF - Vỏ phong hóa Sialferit, SAF - SA - Vỏ phong hóa Sialferit - Sialit và KBTVPH (không bảo tồn vỏ phong hóa) (Hình 2h).

**4. Hệ phương pháp**



**Hình 3: Quy trình thành lập bản đồ nhạy cảm trượt lở trong nghiên cứu này**

Phân tích nhạy cảm trượt lở trong nghiên cứu này gồm 6 bước chính (Hình 3): (1) Xây dựng cơ sở dữ liệu trượt lở và lựa chọn các bản đồ tác nhân trượt lở phù hợp với tỷ lệ nghiên cứu; (2) Phân

chia dữ liệu trượt lở thành hai tập dữ liệu xây dựng mô hình (70%) và kiểm nghiệm mô hình (30%); (3) Áp dụng mô hình FR để tính trọng số cho các lớp bản đồ tác nhân (4) áp dụng 3 mô hình học máy (Hồi

quy logistic - Logistic regression, mạng Bayes - Bayes Network và máy véc tơ hỗ trợ - Support vector machine) tính toán hệ số quan trọng cho mỗi bản đồ tác nhân; (4) Thẩm định mô hình để đánh giá tỷ lệ thành công và tỷ lệ dự báo của các bản đồ LSI; (5) Thành lập bản đồ chỉ số nhạy cảm trượt lở - LSI bằng việc kết hợp các bản đồ đã gán trọng số bằng mô hình FR và các trọng số quan trọng từ mô hình học máy; (6) Bản đồ nhạy cảm trượt lở của khu vực nghiên cứu ở tỷ lệ 1:10.000.

Phương pháp FR được sử dụng để tính toán mối quan hệ không gian giữa điểm trượt lở và mỗi tác nhân gây ra trượt lở. Giá trị FR được suy luận từ mối quan hệ không gian giữa phân bố điểm trượt với mỗi tác nhân liên quan đến trượt lở. Nếu tỷ số này nhỏ hơn 1 thì nó có sự tương quan thấp và nếu giá trị lớn hơn 1 thì có sự tương quan cao hơn. Bản đồ LSI trong công thức (4) được tính toán cho tổng các bản đồ đã gán trọng số của các tác nhân liên quan trượt lở [9, 24, 31, 68].

**4.1. Phương pháp tỷ số tần suất (FR)**

$$LSI = Fr_1 + Fr_2 + Fr_3 + \dots + Fr_n \quad (4)$$

Trong đó, Fr là bản đồ mỗi tác nhân đã gán hệ số. Để tính toán trọng số của từng tác nhân theo phương pháp FR, đầu tiên cần tính toán phần trăm diện tích các lớp, tiếp theo tính toán phần trăm diện tích các điểm trượt trong mỗi lớp thuộc tính của bản đồ tác nhân. Trọng số của từng lớp sẽ được tính bằng thương giữa phần trăm điểm trượt với phần trăm diện tích của từng lớp của bản đồ tác nhân. Công thức 4-a, 4-b, 4-c sẽ trình bày chi tiết tính toán trọng số theo phương pháp FR.

$$\% TL = \frac{DTTLTL \times 100}{\sum DTTL} \quad (4a)$$

$$\% DTL = \frac{DTL \times 100}{\sum DTL} \quad (4b)$$

$$Trọng\ số = \frac{\% TL}{\% DTL} \quad (4c)$$

Trong công thức (4-a), (4-b) và (4-c), TL: Trượt lở; DTTLTL: Diện tích trượt lở trong một lớp;  $\sum DTTL$ : Tổng diện tích trượt lở; DTL: Diện tích của một lớp;  $\sum DTL$ : Tổng diện tích lớp (diện tích toàn vùng nghiên cứu hay diện tích bản đồ tác nhân).

**4.2. Mạng Bayes (BN)**

BN là một phương pháp hiệu quả để biểu diễn tri thức theo ảnh hưởng không chắc chắn [43], được giới thiệu đầu tiên bởi [21]. Phương pháp này dựa trên lý thuyết Bayes để biểu thị mối tương quan về xác suất và đồ thị và mối quan hệ giữa các biến [36]. BN thường sử dụng rộng rãi trong các mô hình hệ thống phức tạp [59]. Để dự báo trượt lở, BN được áp dụng để phân tích ảnh hưởng của các tác nhân tới quá trình trượt lở và được đánh giá qua sự nhạy cảm về xuất hiện trượt lở. Xác suất kết hợp các sự kiện trượt lở với tập tác nhân gây ra sử dụng phương pháp BN được thể hiện bằng công thức (5) dưới đây:

$$P_B(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P_B(X_i | \prod_{X_i}) = \prod_{i=1}^n \theta_{X_i | \prod_{X_i}} \quad (5)$$

Trong đó,  $X = (X_1, X_2, \dots, X_n)$  là các tác nhân liên quan đến trượt lở,  $P_B(X | \prod_{X_i}) = \theta_{X_i | \prod_{X_i}}$  là phân bố xác suất chung so với tác nhân  $X_i$ , n là số lượng các tác nhân gây trượt lở.

### **4.3. Hồi quy logic (LR)**

LR là một trong những phương pháp được sử dụng phổ biến nhất trong đánh giá nguy cơ trượt lở ở các khu vực khác nhau trên thế giới [23, 42, 54, 66]. Nhân của mô hình hồi quy logistic là mối quan hệ hồi quy đa biến giữa một biến phụ thuộc và các biến độc lập khác, đóng vai trò quan trọng trong việc dự đoán sự có hoặc không có của một sự kiện trượt lở dựa trên các giá trị của một tập hợp các biến liên quan. Ưu điểm của hồi quy logistic là bằng cách thêm một hàm ràng buộc thích hợp vào mô hình hồi quy tuyến tính thông thường, các biến có thể liên tục hoặc rời rạc, hoặc bất kỳ sự kết hợp nào của cả hai loại và chúng không nhất thiết phải có phân phối chuẩn [23, 32, 66]. Mối quan hệ giữa sự xuất hiện trượt lở và sự phụ thuộc của nó vào một số biến số được xác định bằng phương trình (6) sau [14, 23, 32, 66]:

$$P = 1 / (1 + e^{-Z}) \quad (6)$$

Trong đó, P là xác suất xuất hiện trượt lở và Z là sự kết hợp tuyến tính. Mô hình hồi quy logistic được thể hiện theo công thức (7) sau [23, 32, 66]:

$$Z = C_0 + C_1X_1 + C_2X_2 + \dots + C_nX_n \quad (7)$$

Trong công thức (7), Z là sự kết hợp tuyến tính,  $C_0$  là hệ số chặn của mô hình hồi quy,  $C_1, C_2, \dots, C_n$  là các hệ số góc của mô hình hồi quy và  $X_1, X_2, \dots, X_n$  là các biến độc lập.

### **4.4. Máy véc tơ hỗ trợ (SVM)**

Máy vectơ hỗ trợ là kỹ thuật dựa trên hạt nhân không tham số được rút ra từ lý thuyết thống kê học [64]. Chúng đặc biệt hấp dẫn để giải các bài toán phân loại, hồi quy và ước lượng mật độ cao phi tuyến

bằng cách kiểm soát độ phức tạp của mô hình [8, 38, 58]. SVM xem xét tập dữ liệu trượt lở đầu vào trong một không gian đa chiều, sau đó mặt siêu phẳng tối ưu với tối đa khoảng cách biên sẽ được xác định để tách biệt thành hai lớp: Lớp có trượt lở và lớp không có trượt lở. Mặt siêu phẳng tối ưu được định nghĩa bởi một số véc tơ hỗ trợ và có thể xác định bằng cách giải vấn đề tối ưu sau [28]:

$$f(X) = \text{sign} \left[ \sum_{i=1}^n \alpha_i Y_j k(X, X_i) + c \right] \quad (8)$$

Trong công thức (8),  $X = (X_1, X_2, \dots, X_n)$  là véc tơ của các tác nhân gây ra trượt lở và  $Y_j = (Y_1, Y_2)$  là véc tơ của biến phụ thuộc (trượt lở và không trượt lở), c là phần bù so với gốc của siêu phẳng, n là số lượng các yếu tố ảnh hưởng đến trượt lở,  $\alpha_i$  là hằng số thực dương,  $k(X, X_i)$  là hàm nhân mà có thể là 4 loại thuật toán chính được sử dụng để phân loại là linear, polynomial, radial basis function (RBF) và sigmoid [19, 61]. Để phân loại nhị phân cho trượt lở và không trượt lở, điều kiện để giải phương trình (8) được mô phỏng bởi công thức (9) dưới đây:

$$\begin{cases} \omega^T \varphi(X_i) + c \geq 1, & \text{if } Y_j = +1 (\text{landslide}) \\ \omega^T \varphi(X_i) + c \leq -1, & \text{if } Y_j = -1 (\text{non-landslide}) \end{cases} \quad (9)$$

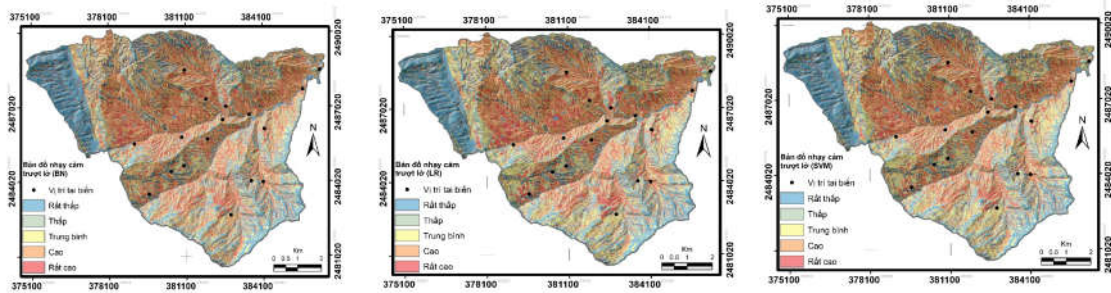
Trong đó,  $\varphi(X_i)$  là một hàm phi tuyến chia không gian đầu vào thành không gian thứ nguyên cao hơn,  $\omega$  đại diện cho véc tơ trọng số [3].

### **4.5. Đánh giá và so sánh các phương pháp**

Có nhiều cách để đánh giá độ chính xác của các bản đồ phân vùng nguy cơ trượt lở. Cách tốt nhất là thẩm định ngoài thực địa nhưng việc này rất khó. Tuy nhiên có thể thẩm định dựa trên các

thông tin về các điểm trượt lở đã xảy ra, đặc biệt cho các khu vực rủi ro cao và rất cao [29]. Đánh giá kết quả dự báo là một trong những nhiệm vụ quan trọng nhất, nếu thiếu công việc này, các kết quả của mô hình dự báo thiếu tính ứng dụng và không có nhiều ý nghĩa khoa học [10, 39,

46). Một số công cụ thống kê và toán học như tính toán mật độ trượt lở, đường cong tỷ lệ thành công (SRC) và đường cong tỷ lệ dự báo (PRC), kiểm chứng chi - quare, đường cong đặc tính hoạt động (ROC),... đã được sử dụng để xác định độ chính xác của các mô hình trượt lở.



**Hình 4: Các bản đồ nhạy cảm trượt lở được tạo ra từ ba mô hình học máy (BN, LR, SVM)**

Trong nghiên cứu này, ROC đã được sử dụng để đánh giá chất lượng của các mô hình. Đường cong ROC được xây dựng bằng đồ thị các cặp chỉ số thống kê của “sensitivity” và “100 - specificity”. Trong bài báo này, cặp chỉ số thống kê này được biểu thị bằng “phần trăm tích lũy các điểm trượt lở” và “phần trăm chỉ số nhạy cảm trượt lở”. Giá trị AUC là phần diện tích dưới đường cong ROC được sử dụng để đánh giá chất lượng của các mô hình đã áp dụng [55]. Chỉ số AUC càng cao thì hiệu năng của các mô hình trượt lở càng tốt và khi mà giá trị AUC xấp xỉ 1 thì hiệu năng của mô hình là hoàn hảo [15].

**5. Kết quả và thảo luận**

**5.1. Các bản đồ nhạy cảm trượt lở**

Trong nghiên cứu này, ba mô hình học máy đã được sử dụng để thành lập bản đồ nhạy cảm trượt lở. Các bản đồ tác nhân trượt lở đã được kiểm tra tính đa cộng tuyến trước khi được sử dụng làm đầu vào dữ liệu cho các mô hình học máy. Kết quả phân tích tính đa cộng tuyến của các

tác nhân liên quan đến trượt lở được thể hiện theo các chỉ số “Tolerance - TOL” và chỉ số “Variance Inflation Factor - VIF” (Bảng 1). Nếu giá trị TOL nhỏ hơn 0,2, nó thể hiện tính đa cộng tuyến của dữ liệu và nếu giá trị TOL nhỏ hơn 0,1, thì thể hiện sự đa cộng tuyến rất nghiêm trọng. Nếu VIF lớn hơn 10 thì nó chỉ ra tính đa cộng tuyến của dữ liệu [28]. Trong nghiên cứu này, tất cả các giá trị TOL đều lớn hơn 0,1 và tất cả các giá trị VIF đều nhỏ hơn 10. Điều này thể hiện dữ liệu không có tính đa cộng tuyến và thỏa mãn cho các tính toán của mô hình. Trong Bảng 1, curvature nghĩa là mật cong địa hình, slope là độ dốc, SPI là chỉ số năng lượng dòng chảy, STI là chỉ số vận chuyển trầm tích, TWI là chỉ số độ ẩm địa hình, weathering crust là vỏ phong hóa và LULC là thảm phủ/ đất sử dụng.

Những bản đồ nhạy cảm trượt lở (LSM) được thành lập dựa trên hai bước chính: (1) Thành lập các bản đồ LSI và (2) Phân loại các bản đồ LSI thành các bản đồ LSM. Trong bước đầu, các bản

## Nghiên cứu

đồ LSI được tạo ra cho khu vực nghiên cứu, trên đó mỗi ô pixel được ấn định một giá trị nhạy cảm. Trong bước hai, các bản đồ LSI sẽ được phân loại theo phương pháp phân loại quantile trong phần mềm

ArcGIS 10.8 để thành lập bản đồ LSM. Dựa trên sự phân loại của phương pháp này, các lớp nhạy cảm được xác định để tạo thành các bản đồ LSM gồm: Rất thấp, thấp, trung bình, cao và rất cao (Hình 4).

**Bảng 1. Kết quả phân tích tính đa cộng tuyến của dữ liệu**

Mô hình	Các hệ số không chuẩn hóa		Các hệ số chuẩn hóa	Hệ số t	Giá trị Sig	Thống kê cộng tuyến	
	Hệ số B	Sai số chuẩn	Beta			Chỉ số TOL	Chỉ số VIF
1 (Hằng số)	-.003	.000		-8.326	.000		
Mặt cong địa hình	.000	.000	.013	10.727	.000	.999	1.001
Độ dốc	.000	.000	.007	5.738	.000	.944	1.059
SPI	.000	.000	.009	6.781	.000	.961	1.041
STI	.000	.000	.028	22.112	.000	.971	1.030
TWI	.000	.000	.001	.982	.326	.998	1.002
Vỏ phong hóa	.000	.000	.005	4.067	.000	.959	1.043
LULC	.001	.000	.021	16.803	.000	.933	1.072

### **5.2. Thẩm định mô hình**

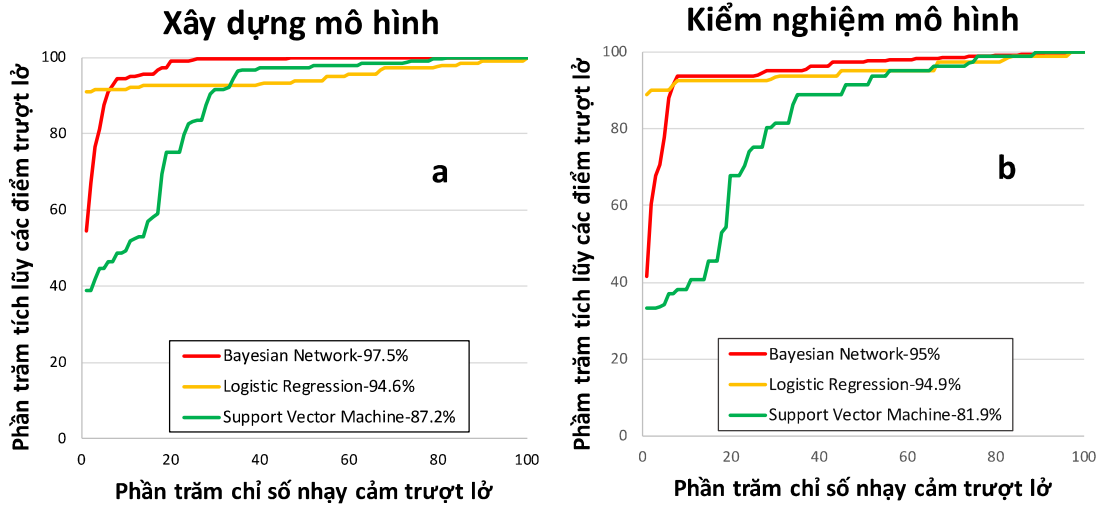
Để đánh giá và so sánh các mô hình dự báo, hai tập dữ liệu xây dựng mô hình và kiểm nghiệm mô hình đã được tạo ra. Đầu tiên, dữ liệu trượt lở được phân chia ngẫu nhiên thành hai tập dữ liệu này. Tập dữ liệu xây dựng mô hình chiếm 70 % số pixel trượt lở (177 pixel) và tập dữ liệu kiểm nghiệm mô hình chiếm 30 % số pixel còn lại (81 pixel). Sau đó, hai tập dữ liệu này sẽ được sử dụng để đánh giá hiệu năng của từng mô hình. Kết quả tính toán chỉ ra rằng tất cả các mô hình đều có khả năng dự báo tốt. Trong đó, mô hình BN có khả năng dự báo tốt nhất, tiếp theo là mô hình LR và SVM. Phân tích đường cong ROC sử dụng tập dữ liệu xây dựng mô hình thể hiện trong Hình 5a. Giá trị AUC lớn nhất là mô hình BN (AUC = 97,5 %), tiếp theo là mô hình LR (AUC = 94,6 %) và mô hình SVM (AUC = 87,2 %). Tiếp theo, phân tích đường cong ROC cho tập dữ liệu kiểm tra mô hình (Hình 5b) cho ta kết quả mô hình BN có giá trị AUC cao nhất (95 %),

tiếp theo là mô hình LR với AUC là 94,9 % và mô hình SVM (AUC = 81,9 %).

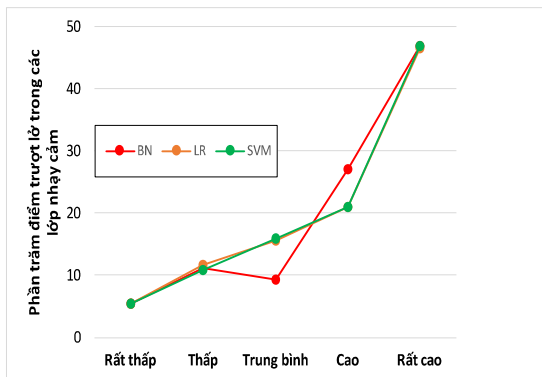
So sánh các kết quả nghiên cứu này với một số nghiên cứu đã công bố, nhóm tác giả có một số nhận định sau: (1) Đối với kết quả thẩm định của mô hình BN, các giá trị AUC được thẩm định từ tập dữ liệu xây dựng mô hình và tập dữ liệu kiểm nghiệm mô hình là 97,5 % và 95 %, có giá trị tương đương với các công bố trước kia của các tác giả [3, 18] với các giá trị này lần lượt cho các công bố này là 97,1 %; 91,5 % và 90,3 %; 91,5 %; (2) Các nghiên cứu đã áp dụng mô hình SVM trước kia [3, 25] có giá trị AUC lần lượt là 82,4 % và 95,5 %. Một nghiên cứu khác của [15] đã kiểm nghiệm các giá trị của mô hình SVM bởi 2 phương pháp, một là đường cong ROC và hai là cặp phương pháp đường cong tỷ lệ thành công (SRC) và đường cong tỷ lệ dự báo (PRC) cho hai tập dữ liệu xây dựng mô hình và kiểm nghiệm mô hình đã chỉ ra các giá trị AUC lần lượt là 96,9 %; 91,8 % và 85,51 %;

82,46 %; (3) Đối với mô hình LR, giá trị AUC của nghiên cứu này tương tự với nhiều công bố trước kia [7, 3, 25, 31, 32,

62, 66, 69] với các giá trị AUC lần lượt là 86,37 %; 90,34 %; 74,2 %; 80,9 %; 81,1 %; 76 %; 77,51 % và 84 %.



**Hình 5: Đường cong ROC cho tập dữ liệu xây dựng mô hình (5a) và kiểm nghiệm mô hình (5b) của các mô hình học máy**



**Hình 6: Biểu đồ thể hiện phần trăm điểm trượt lở quan sát được trong các lớp nguy hiểm trượt lở từ các bản đồ nguy hiểm được thành lập từ ba mô hình học máy BN, LR và SVM**

Trong nghiên cứu này, ngoài dữ liệu trượt lở và lũ bùn đá đã sử dụng, bảy bản đồ tác nhân liên quan đến tai biến gồm: Độ dốc, mặt cong địa hình, TWI, SPI, STI, thảm phủ và vỏ phong hóa. Tuy nhiên, với độ chính xác của các mô hình như đã phân tích ở trên cho thấy rằng các kết quả tính toán trong nghiên cứu này là đáng tin cậy và hữu ích để áp dụng cho khu vực nghiên cứu xã Phìn Ngan nói chung và các khu vực miền núi khác của Việt Nam nói riêng

trong việc thành lập các bản đồ nguy hiểm trượt lở ở tỷ lệ lớn.

## 6. Kết luận

Nghiên cứu này áp dụng một mô hình thống kê và ba mô hình học máy để tạo ra các bản đồ nguy hiểm với tai biến trượt lở tỷ lệ 1:10.000 cho khu vực nghiên cứu xã Phìn Ngan, huyện Bát Xát, tỉnh Lào Cai. Đầu tiên, giá trị trọng số của mỗi tác nhân liên quan đến trượt lở được tính toán dựa trên mô hình FR. Thứ hai, dữ liệu đại diện cho mối quan hệ giữa các lớp bản đồ tác nhân trượt lở đất và các điểm trượt lở được sử dụng làm dữ liệu đầu vào cho các mô hình học máy để tính toán giá trị quan trọng của từng tác nhân này. Sau đó, ba bản đồ nguy hiểm trượt lở được thiết lập từ ba mô hình học máy (BN, LR và SVM), được thiết lập bằng cách kết hợp trọng số của các lớp bản đồ tác nhân đã tính toán theo mô hình FR với giá trị quan trọng của mỗi tác nhân liên quan đến trượt lở được tính toán từ các mô hình ML. Cuối cùng, bản đồ LSM của khu vực nghiên cứu được thành lập theo năm

## Nghiên cứu

mức độ rủi ro: Rất thấp, thấp, trung bình, cao và rất cao.

Kết quả đánh giá mô hình cho thấy khả năng dự báo của từng mô hình khá cao và không có sự chênh lệch nhiều. Các mô hình BN và LR (Hình 5a) có giá trị AUC gần như tương tự, lần lượt là 97,5 % và 94,6 %. Giá trị AUC của mô hình còn lại SVM (Hình 5a) cũng có giá trị là 87,2 %. Các mô hình này cũng có giá trị AUC cao ngang bằng khi sử dụng tập dữ liệu thử nghiệm (Hình 5b). Giá trị của mô hình BN là 95 % và LR có giá trị 94,9 %. Mô hình SVM có giá trị AUC cho tập dữ liệu thử nghiệm mô hình 81,9 %.

Các bản đồ nhạy cảm trượt lở được tạo ra từ các mô hình BN, LR và SVM cũng được so sánh theo tỷ lệ trượt lở có trong từng mức độ nhạy cảm: Rất thấp, thấp, trung bình, cao và rất cao (Hình 6). Kết quả cho thấy tỷ lệ trượt lở quan sát được ở các lớp cao và rất cao của 3 mô hình cũng cho thấy sự khác biệt rất ít. Tỷ lệ này của các phương pháp BN, LR và SVM lần lượt là 74,03 %; 67,44 % và 67,83 %.

Kết quả của nghiên cứu cho thấy có thể kết hợp các phương pháp thống kê và máy học để thành lập bản đồ nhạy cảm trượt lở ở tỷ lệ lớn 1:10.000 cho các khu vực miền núi, nơi thường xuyên xảy ra tai biến trượt lở. Kết quả nghiên cứu sẽ góp phần giảm thiểu nguy cơ tai biến trượt lở và hỗ trợ quy hoạch lãnh thổ trong khu vực nghiên cứu. Các mô hình được áp dụng trong nghiên cứu này có thể được áp dụng cho các khu vực khác để thiết lập bản đồ nhạy cảm trượt lở ở tỷ lệ tương tự.

**Lời cảm ơn:** Nghiên cứu này được hỗ trợ bởi Đề tài “Nghiên cứu cơ sở khoa học để lựa chọn hệ phương pháp, mô hình phân vùng cảnh báo chi tiết và xác định ngưỡng mưa kích hoạt tại các khu vực nhạy cảm về trượt, sạt lở đất đá, lũ bùn

đá, lũ ống, lũ quét khu vực miền núi, trung du Việt Nam”. Mã số: TNMT.2021.02.09, do Bộ Tài nguyên và Môi trường cấp kinh phí, Viện Khoa học Địa chất và Khoáng sản Việt Nam chủ trì thực hiện.

## TÀI LIỆU THAM KHẢO

- [1]. Aleotti P, Chowdhury R. (1999). *Landslide hazard assessment: Summary review and new perspectives*. Bull Eng Geol Environ, 58(1):21 - 44. <https://doi.org/10.1007/s100640050066>.
- [2]. Beven K, Kirkby M. (1979). *Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant (A physically based, variable contributing area model of basin hydrology)*. Hydrol. Sci. J, 24, 43-69.
- [3]. Binh PT, Pradhan B, Bui TD, Prakash I, Dholakia MB. (2016). *A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India)*. Environmental modelling & software, 84, 240-250. <https://doi.org/10.1016/j.envsoft.2016.07.005>.
- [4]. Binh PT, Pradhan B, Dieu BT. (2018). *Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees*. Geomorphology, 303, 256-270. <https://doi.org/10.1016/j.geomorph.2017.12.008>.
- [5]. Binh PT, Thoi NT, Qi CC, Phong TV, Dou J, Lan HS, Hiep LV, Prakash I. (2020). *Coupling RBF neural network with ensemble learning techniques for landslide susceptibility mapping*. Catena, 195, 104805. <https://doi.org/10.1016/j.catena.2020.104805>.
- [6]. Burrough PA and Mc Donnell RA. (1998). *Principles of Geographical Information Systems*. Oxford, Oxford University Press, 330 pp.
- [7]. Chen T, Niu, RQ, Jia XP. (2016). *A comparison of information value and logistic regression models in landslide susceptibility mapping by using GIS*. Environ Earth Sci 75: 867. <http://dx.doi.org/10.1007/s12665-016-5317-y>.
- [8]. Cherkassky V, Mulier F. (2007). *Learning from data: Concepts, theory and methods*. Wiley, New York.

- [9]. Choi JW, Oh HJ, Lee HJ, Lee CW, Lee S. (2012). *Combining landslide susceptibility maps obtained from frequency ratio, logistic regression and artificial neural network models using ASTER images and GIS*. Engineering Geology 124: 12-23. Doi: 10.1016/j.enggeo.2011.09.011.
- [10]. Chung CJF, Fabbri AG. (2003). *Validation of spatial prediction models for landslide hazard mapping*. Natural Hazards 30: 451 - 472. <https://doi.org/10.1023/B:NHAZ.0000007172.62651.2b>.
- [11]. Corominas J, Westen CV, Frattini P, Cascini L, Malet JP, Fotopoulou S, Catani F, Van Den Eeckhaut M, Mavrouli O, Agliardi F, Pitilakis K, Winter MG, Pastor M, Ferlisi S, Tofani V, Hervas J, Smith JT. (2014). *Recommendations for the quantitative analysis of landslide risk*. Bull. Eng. Geol. Environ, 73, 209 - 263. <https://doi.org/10.1007/s10064-013-0538-8>.
- [12]. Dahal RK, Hasegawa S, Nonomura A, Yamanaka M, Dhakal S, Paudyal P. (2008a). *Predictive modeling of rainfall - induced landslide hazard in the Lesser Himalaya of Nepal based on weights of evidence*. Geomorphology 102: 496-510. <http://dx.doi.org/10.1016/j.geo-morph.2008.05.041>.
- [13]. Dahal RK, Hasegawa S, Nonomura A, Yamanaka M, Masuda T, Nishino K. (2008b). *GIS - based weights of evidence modeling of rainfall-induced landslides in small catchments for landslide susceptibility mapping*. Environ Geol 54: 311-324. <http://dx.doi.org/10.1007/s00254-007-0818-3>.
- [14]. Dai FC, Lee CF, Li J, Xu ZW. (2001). *Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong*. Environmental Geology 40 (3): 381-391. Doi: 10.1007/s002540000163.
- [15]. Dieu BT, Binh PT, Nguyen QP, Hoang ND. (2016). *Spatial prediction of rainfall - induced shallow landslides using hybrid integration approach of least - squares, support vector machines and differential evolution optimization: A case study in Central Vietnam*. Int. J. Digital Earth, 1-21. <https://doi.org/10.1080/17538947.2016.1169561>.
- [16]. Dieu BT, Shahabi H, Shirzadi A, Chapi K, Pradhan B, Chen W, Khosravi K, Panahi M, Ahmad BB, Saro L. (2018). *Land subsidence susceptibility mapping in South Korea using machine learning algorithms*. Sensors, 18, 2464. Doi:10.3390/s1808246.
- [17]. Dieu TB, Pradhan B, Lofman O, Revhaug I, Dick OB. (2012a). *Spatial prediction of landslide hazards in Hoabinh province (Vietnam): A comparative assessment of the efficacy of evidential belief functions and fuzzy logic models*. Catena 96: 28-40. <https://doi.org/10.1016/j.catena.2012.04.001>.
- [18]. Dieu TB, Pradhan B, Lofman O, Revhaug I, Dick OB. (2012b). *Landslide susceptibility assessment in the Hoabinh province of Vietnam: A comparison of the levenberg - marquardt and Bayesian regularized neural networks*. Geomorphology 171-172: 12-29. <https://doi.org/10.1016/j.geomorph.2012.04.023>.
- [19]. Dixon B, Candade N. (2008). *Multispectral landuse classification using neural networks and support vector machines: One or the other, or both?* Int. J. Remote Sens, 29 (4), 1185-1206. <https://doi.org/10.1080/01431160701294661>.
- [20]. Ermini L, Catani F, Casagli N. (2005). *Artificial neural networks applied to landslide susceptibility assessment*. Geomorphology, 66:327-343. Doi: 10.1016/j.geomorph.2004.09.025.
- [21]. Friedman N, Geiger D, Goldszmidt M. (1997). *Bayesian network classifiers*. Machine Learning, 29, 131-163. <https://doi.org/10.1023/A:1007465528199>.
- [22]. Goetz JN, Brenning A, Petschko H, Leopold P. (2015). *Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling*. Comput. Geosci, <https://doi.org/10.1016/j.cageo.2015.04.007>.
- [23]. Hien M.D, Yin KL, Guo ZZ. (2020). *A comparative study on the integrative ability of the analytical hierarchy process, weights of evidence and logistic regression methods with the Flow - R model for landslide susceptibility assessment*. Geomatics, Natural Hazard and Risk, Vol 11, No 1, 2449-2485. <https://doi.org/10.1080/19475705.2020.1846086>.
- [24]. Hong HY, Pourghasemi HR, Pourtaghi ZS. (2016). *Landslide susceptibility assessment in Lianhua County (China): A comparison between a*

*random forest data mining technique and bivariate and multivariate statistical models. Geomorphology* 259: 105-118. <http://dx.doi.org/10.1016/j.geomorph.2016.02.012>.

[25]. Hong HY, Pradhan B, Xu C, Dieu TB. (2015). *Spatial prediction of landslide hazard at the Yihuang area (China) using two - class kernel logistic regression, alternating decision tree and support vector machines. Catena* 133: 266-281. <http://dx.doi.org/10.1016/j.catena.2015.05.019>.

[26]. Huang Y, Zhao L. (2018). *Review on landslide susceptibility mapping using support vector machines. Catena*, <https://doi.org/10.1016/j.catena.2018.03.003>.

[27]. Hung LQ, Van NTH, Duc MD, Ha LTC, Son PV, Khanh NH, Binh LT. (2016). *Landslide susceptibility mapping by combining the analytical hierarchy process and weighted linear combination methods: A case study in the upper Lo River catchment (Vietnam). Landslides*, 13:1285-1301. Doi 10.1007/s10346-015-0657-3.

[28]. Kavzoglu T, Sahin EK, Colkesen I. (2014). *Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines and logistic regression. Landslides* 11: 425-439. <https://doi.org/10.1007/s10346-013-0391-7>.

[29]. Kayastha P, Dhital MR, Smedt FDe. (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. <http://dx.doi.org/10.1016/j.cageo.2012.11.003>.

[30]. Lee S, Min K. (2001). *Statistical analysis of landslide susceptibility at Yongin, Korean. Environ Geol.* 40:1095-113. <https://doi.org/10.1007/s002540100310>.

[31]. Lee S, Pradhan B. (2007). *Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. Landslides* 4: 33-41. Doi:10.1007/s10346-006-0047-y.

[32]. Lee S, Sambath T. (2006). *Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. Environ Geol* 50, 847-855. Doi:10.1007/s00254-006-0256-7.

[33]. Leroi E. (1996). *Landslide hazard - risk maps at different scales: Objectives,*

*tools and developments. In: Proc VII Int Symp Landslides, Trondheim, 1: 35-52. Doi: 10.12691/jgg-5-4-2.*

[34]. Long NT, De Smedt F. (2019). *Analysis and mapping of rainfall - induced landslide susceptibility in A Luoi district, Thua Thien Hue Province, Vietnam. Water-MDPI, 11, 51. Doi:10.3390/w11010051.*

[35]. Long NT, Smedt FDe. (2012). *Application of an analytical hierarchical process approach for landslide susceptibility mapping in A Luoi district, Thua Thien Hue Province, Vietnam. Environ Earth Sci* 66: 1739-1752. <http://dx.doi.org/10.1007/s12665-011-1397-x>.

[36]. Marriot BG, Steventon JD, Sutheland GD, McCann RK. (2006). *Guidelines for developing and updating Bayesian belief networks applied to eological modeling and conservation. Can.J. For. Res,* 36 (12), 3063-3074. <https://doi.org/10.1139/x06-135>.

[37]. Mengistu F, Suryabhagavan KV, Raghuvanshi TK, Lewi E. (2019). *Landslide hazard zonation and slope instability assessment using optical and InSAR data: A case study from Gidole town and its surrounding areas, southern Ethiopia. Remote Sensing of Land,* 3(1):1 - 14. Doi:10.21523/gcj1.19030101.

[38]. Moguerza JM, Munoz A. (2006). *Support vector machines with applications. Stat Sci,* 21:322-336. doi:10.1214/088342306000000493.

[39]. Mohammady M, Pourghasemi HR, Pradhan B. (2012). *Landslide susceptibility mapping at Golestan province, Iran: A comparison between frequency ratio, Dempster - Shafer and weights of evidence models. Journal of Asian Earth Sciences* 61: 221-236. <http://dx.doi.org/10.1016/j.jseaes.2012.10.005>.

[40]. Moore ID, Burch GJ. (1986). *Physical basis of the length - slope factor in the universal soil loss equation. Soil Sci Soc Am J,* 50(5):1294-1298. <https://doi.org/10.2136/sssaj1986.03615995005000050042x>.

[41]. Nefeslioglu HA, Gokceoglu C, Sonmez H. (2008). *An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility*

- maps. Eng Geol, 97(3-4):171-91. <https://doi.org/10.1016/j.enggeo.2008.01.004>.
- [42]. Oh HJ, Lee S, Soedradjat GM. (2010). *Quantitative landslide susceptibility mapping at Pemalang area, Indonesia*. Environmental Earth Sciences, 60(6):1317-1328. <https://doi.org/10.1007/s12665-009-0272-5>.
- [43]. Pearl J. (2014). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann.
- [44]. Pourghasemi HR, Pradhan B, Gokceoglu C, Mohammadi M, Moradi HR. (2013a). *Application of weights of evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed, Iran*. Arab J Geosci 6: 2351-2365. <http://dx.doi.org/10.1007/s12517-012-0532-7>.
- [45]. Pourghasemi HR, Jirandeh AG, Pradhan B, Xu C, Gokceoglu C. (2013b). *Landslide susceptibility mapping using support vector machine and GIS at the Golestan province, Iran*. J. Earth Syst. Sci 122, No 2: 349-369. Doi: 10.1007/s12040-013-0282-2.
- [46]. Pourghasemi HR, Moradi HR, Fatemi Aghda SM, Gokceoglu C, Pradhan B. (2014). *GIS - based landslide susceptibility mapping with probabilistic likelihood ratio and spatial multi-criteria evaluation models (North of Tehran, Iran)*. Arab J Geosci 7: 1857-1878. <http://dx.doi.org/10.1007/s12517-012-0825-x>.
- [47]. Pradhan AMS, Kang HS, Lee S, Kim YT. (2016). *Spatial model integration for shallow landslide susceptibility and its runout using a GIS - based approach in Yongin*. Geocarto Int. Doi:10.1080/10106049.2016.1155658.
- [48]. Pradhan AMS, Kim YT. (2014). *Relative effect method of landslide susceptibility zonation in weathered granite soil: A case study in Deokjeok-ri Creek, South Korea*. Nat hazards 72: 1189-1217. <http://dx.doi.org/10.1007/s11069-014-1065-z>.
- [49]. Pradhan AMS, KimYT. (2017). *Spatial data analysis and application of evidential belief functions to shallow landslide susceptibility mapping at Mt. Umyeon, Seoul, Korea*. Bull Eng Geol Environ 76: 1263-1279. <http://dx.doi.org/10.1007/s10064-016-0919-x>.
- [50]. Pradhan B, Abokharima MH, Jebur MN, Tehrani MS. (2014). *Land subsidence susceptibility mapping at Kinta Valley (Malaysia) using the evidential belief function model in GIS*. Nat hazards, 73: 1019-1042. Doi 10.1007/s11069-014-1128-1.
- [51]. Pradhan B, Lee S. (2009). *Landslide risk analysis using artificial neural network model focussing on different training sites*. Int J PhysSci, 4:001-015. <https://doi.org/10.5897/IJPS.9000343>.
- [52]. Pradhan B, Lee S. (2010a). *Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression and artificial neural network models*. Environmental Earth Sciences, 60(5):1037-1054. <https://doi.org/10.1007/s12665-009-0245-8>.
- [53]. Pradhan B, Lee S. (2010b). *Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling*. Environ ModellSoftw, 25:747-59. Doi: 10.1016/j.envsoft.2009.10.016.
- [54]. Pradhan B. (2011). *Manifestation of an advanced fuzzy logic model coupled with Geo - information techniques to landslide susceptibility mapping and their comparison with logistic regression modelling*. Environ Ecol Stat 18: 471-493. <http://dx.doi.org/10.1007/s10651-010-0147-7>.
- [55]. Pradhan B. (2013). *A comparative study on the predictive ability of the decision tree, support vector machine and neuro - fuzzy models in landslide susceptibility mapping using GIS*. Computers & Geosciences 51: 350-365. <http://dx.doi.org/10.1016/j.cageo.2012.08.023>.
- [56]. Regmi AD, Devkota KC, Yoshida K, Pradhan B, Pourghasemi HR, Kumamoto T, Akgun A. (2014). *Application of frequency ratio, statistical index and weights of evidence models and their comparison in landslide susceptibility mapping in central Nepal Himalaya*. Arab J Geosci 7: 725-742. <http://dx.doi.org/10.1007/s12517-012-0807-z>.
- [57]. Regmi NR, Giardino JR, Vitek JD. (2010). *Modeling susceptibility to landslides using the weight of evidence approach: Western Colorado, USA*. Geomorphology

115: 172-187. <http://dx.doi.org/10.1016/j.geomorph.2009.10.002>.

[58]. Schölkopf B, Smola A. (2002). *Learning with kernels: Support vectormachines, regularization, optimization and beyond*. MIT Press, Cambridge.

[59]. Song Y, Gong J, Gao S, Wang D, Cui T, Li Y, Wei B. (2012). *Susceptibility assessment of earthquake-induced landslides using Bayesian network: A case study in Beichuan, China*. *Comput. Geosci*, 42 (0), 189-199. <https://doi.org/10.1016/j.cageo.2011.09.011>.

[60]. Sujatha ER and Rajamanickam G.V. (2015). *Landslide hazard and risk mapping using weighted linear combination applied to Tevankarai sream watershed, Kodaikkanal, India*. *Human and Ecological Risk Assessment: An International Journal*. 21(6): 1445-1461. <http://dx.doi.org/10.1007/s10064-013-0537-9>.

[61]. Suykens JA, Vandewalle J. (1999). *Least squares support vector machine classifiers*. *Neural Process. Lett*, 9(3), 293-300. <https://doi.org/10.1023/A:1018628609742>.

[62]. Talaei R. (2014). *Landslide susceptibility zonation mapping using logistic regression and its validation in Hashichin region, Northwest of Iran*. *Journal Geological Society of India*, 84: 68 - 86. <https://doi.org/10.1007/s12594-014-0111-5>.

[63]. Van Westen CJ, Terlien MTJ. (1996). *An approach towards deterministic landslide hazard analysis in GIS. A case study from Manizales (Colombia)*. *Earth Surface Processes and Landforms* 21, 853 - 868. [https://doi.org/10.1002/\(SICI\)1096-9837\(19609\)21:9<853:AID-ESP676>3.0.CO;2-C](https://doi.org/10.1002/(SICI)1096-9837(19609)21:9<853:AID-ESP676>3.0.CO;2-C).

[64]. Vapnik VN. (1999). *The Nature of statistical learning theory*. 2<sup>nd</sup> ed. Springer, New York.

[65]. Wang QQ, Li WP, Xing ML, Wu YL, Pei YB, Yang DD, Bai HY. (2016). *Landslide susceptibility mapping at Gongliu county, China using artificial neural network and weight of evidence models*. *Geosciences Journal*, vol 20. No 5: 705-718. <http://dx.doi.org/10.1007/s12303-016-0003-3>.

[66]. Wu ZY, Wu YL, Yang YT, Chen FW, Zhang N, Ke YT. (2017). *A comparative study on the landslide susceptibility mapping using logistic regression and statistical index*

*models*. *Arab J Geosci*, 10: 187. <http://dx.doi.org/10.1007/s12517-017-2961-9>.

[67]. Xu C, Xu X, Dai FC, Xiao JZ, Tan XB, Yuan RM. (2012). *Landslide hazard mapping using GIS and weight of evidence model in Qingshui river watershed of 2008 Wenchuan earthquake struck region*. *Journal of Earth Science*, vol 23. No 1: 97-120. <http://dx.doi.org/10.1007/s12583-012-0236-7>.

[68]. Yalcin A, Reis S, Aydinoglu AC, Yomralioglu T. (2011). *A GIS - based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey*. *Catena* 85: 274-287. <http://dx.doi.org/10.1016/j.catena.2011.01.014>.

[69]. Yesilnacar E, Topal T. (2005). *Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey)*. *Engineering Geology* 79: 251-266. <http://dx.doi.org/10.1016/j.enggeo.2005.02.002>.

[70]. Yilmaz I. (2009). *Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat - Turkey)*. *Comput Geosci*, 35:1125-1138. <https://doi.org/10.1016/j.cageo.2008.08.007>

[71]. Yilmaz I. (2010). *Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: Conditional probability, logistic regression, artificial neural networks and support vector machine*. *Environmental Earth Sciences*, 61(4):821-836. Doi: 10.1007/s12665-009-0394-9.

[72]. Zezere JL, Pereira S, Melo R, Oliveira SC, Garcia RAC. (2017). *Mapping landslide susceptibility using data - driven methods*. *Science of the Total Environment* 589: 250 - 267. <https://doi.org/10.1016/j.scitotenv.2017.02.188>.

[73]. Zhang GF, Cai YX, Zheng Z, Zhen JW, Liu YL, Huang KY. (2016). *Integration of the statistical index method and the analytic hierarchy process technique for the assessment of landslide susceptibility in Huizhou, China*. *Catena* 142: 233-244. <http://dx.doi.org/10.1016/j.catena.2016.03.028>.

Ngày nhận bài: 22/8/2022; Ngày chấp nhận đăng: 27/9/2022