



Original Article

Large-scale Mapping of Landslide and Debris Flow using FlowR Model with Statistical and Machine Learning Methods

Do Minh Hien^{1,*}, Nguyen Van Hoang¹, Mai Le Dung¹,
Luong Huu Dung², Ngo Thi Thuy², Van Thi Hang²

¹*Vietnam Institute of Geosciences and Mineral Resources, 67 Chien Thang, Thanh Xuan, Hanoi, Vietnam*

²*Vietnam Institute of Meteorology, Hydrology and Climate Change,
23/62 Nguyen Chi Thanh, Dong Da, Hanoi, Vietnam*

Received 12 May 2022

Revised 20 June 2022; Accepted 19 August 2022

Abstract: The main purpose of this article is to establish a susceptibility zonation map of the landslides and debris flows in Phin Ngan commune, Bat Xat district, Lao Cai province on a large scale using statistical methods and machine learning combined with the FlowR model. First, the five Landslide Susceptibility Index (LSI) maps were established from two statistical models (Logistic Regression - LR, Discriminant Analysis - DA) and three machine learning models (Bayesian Network - BN, Artificial Neural Network - ANN, Support Vector Machine - SVM) were generated based on seven maps of landslide conditioning factors (slope, curvature, stream power index-SPI, topographic wetness index-TWI, sediment transportation index-STI, land use/land cover and weathering crust). Next, the five LSI maps will be evaluated for performance with the value of Area Under the Curve (AUC) according to the Receiver Operating Characteristic (ROC) curve. After that, a susceptibility map of debris flow established with FlowR software was combined with the five LSI maps created from five statistical and machine learning methods to generate a susceptibility zonation map of landslides and debris flows in the study area. The area percentage of the locations with landslides and debris flows located in the zones of susceptibility (very low, low, medium, high, very high), which were created from five combined methods: BN-FlowR, LR-FlowR, DA-FlowR, ANN-FlowR, and SVM-FlowR, were compared and evaluated. The results indicate that the integrated models have given outputs with good forecasting ability. They are also very useful in land-use planning as well as the prevention and mitigation of risks due to landslides and debris flows in the research area and other similar mountainous areas.

Keywords: Landslide, debris flow, FlowR, statistical model, machine learning model.

* Corresponding author.

E-mail address: hien_dm@yahoo.com

<https://doi.org/10.25073/2588-1094/vnuées.4872>

Thành lập bản đồ phân vùng nhạy cảm tai biến trượt lở và lũ bùn đá ở tỷ lệ lớn bằng các phương pháp thống kê, học máy kết hợp với mô hình FlowR

Đỗ Minh Hiền^{1,*}, Nguyễn Văn Hoàng¹, Mai Lê Dũng¹,
Lương Hữu Dũng², Ngô Thị Thủy², Văn Thị Hằng²

¹Viện Khoa học Địa chất và Khoáng sản, 67 Chiến Thắng, Thanh Xuân, Hà Nội, Việt Nam

²Viện Khoa học Khí tượng Thủy văn và Biến đổi Khí hậu,
23/62 Nguyễn Chí Thanh, Đống Đa, Hà Nội, Việt Nam

Nhận ngày 12 tháng 5 năm 2022

Chỉnh sửa ngày 20 tháng 6 năm 2022; Chấp nhận đăng ngày 19 tháng 8 năm 2022

Tóm tắt: Mục đích chính của bài báo này là thành lập bản đồ phân vùng nhạy cảm tai biến trượt lở và lũ bùn đá khu vực xã Phìn Ngan huyện Bát Xát tỉnh Lào Cai ở tỷ lệ lớn bằng các phương pháp thống kê, học máy kết hợp với mô hình FlowR. Đầu tiên, năm bản đồ nhạy cảm trượt lở (LSI) được thành lập từ hai mô hình thống kê (hồi quy logic-LR, phân tích biệt thức-DA và ba mô hình học máy (mạng Bayes-BN, mạng thần kinh nhân tạo-ANN, máy véc tơ hỗ trợ-SVM) được thành lập dựa trên bảy bản đồ tác nhân gây trượt lở (độ dốc, mặt cong địa hình, chỉ số năng lượng dòng chảy, chỉ số độ ẩm địa hình, chỉ số vận chuyển trầm tích, thấm phủ và vỏ phong hóa). Tiếp theo, năm bản đồ LSI này sẽ được đánh giá hiệu năng thực hiện mô hình bằng giá trị phân trăm diện tích (AUC) theo đặc tính đường cong nhận được (ROC). Sau đó, một bản đồ nhạy cảm lũ bùn đá được thành lập bằng phần mềm FlowR sẽ được kết hợp với năm bản đồ LSI đã tạo ra từ năm phương pháp thống kê và học máy để thành lập bản đồ phân vùng nhạy cảm tai biến trượt lở và lũ bùn đá cho khu vực nghiên cứu. Phân trăm diện tích các điểm trượt lở và lũ bùn nằm trong các lớp phân vùng nhạy cảm rất thấp, thấp, trung bình, cao, rất cao được tạo ra từ năm phương pháp kết hợp BN-FlowR, LR-FlowR, DA-FlowR, ANN-FlowR và SVM-FlowR được so sánh và đánh giá. Kết quả chỉ ra rằng các mô hình kết hợp đều đưa ra các kết quả với khả năng dự báo tốt, các kết quả này rất hữu ích trong công tác quy hoạch đất sử dụng và trợ giúp công tác phòng chống, giảm thiểu rủi ro do tai biến trượt lở và lũ bùn đá gây nên tại khu vực nghiên cứu và các khu vực miền núi khác.

Từ khóa: Trượt lở, lũ bùn đá, FlowR, mô hình thống kê, mô hình học máy.

1. Mở đầu

Hiện nay trên thế giới, công tác đánh giá rủi ro tai biến trượt lở, lũ bùn đá đang ngày càng được quan tâm. 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. Dữ liệu đầu vào thường được suy luận ra từ sự đánh giá trong suốt quá trình thực

* Tác giả liên hệ.

Địa chỉ email: hien_dm@yahoo.com

<https://doi.org/10.25073/2588-1094/vnuees.4872>

địa, có thể được hỗ trợ bởi các kết quả giải đoán từ ảnh máy bay. Những hệ phương pháp này cũng được gọi là các phương pháp đánh giá bởi chuyên gia [1], và 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 [2]. Các phương pháp định lượng có tính chặt chẽ hơn và chúng áp dụng các phương pháp phân tích thống kê (2 biến hoặc đa biến), các phương pháp học máy và các phương pháp tiên định dựa trên các mô hình vật lý như các mô hình ổn định sườn dốc và các mô hình thủy văn và các mô hình khác về địa chất công trình [2, 3]. Các phương pháp tiên định dựa trên mô hình vật lý và được xem có độ chính xác cao nhất. Tuy nhiên, giới hạn chính của phương pháp này là chúng chỉ thích hợp cho các khu vực mà ở đó chuyên động trượt lở không phức tạp và các đặc điểm địa chất, địa mạo ở đó là khá đồng nhất [4].

Các mô hình thống kê và học máy thường được coi là các phương pháp tiếp cận theo hướng dữ liệu, cả hai đều tập trung vào việc phân tích các yếu tố ảnh hưởng đến trượt lở đất bằng cách sử dụng bộ dữ liệu trượt đất trong quá khứ và hiện tại [5-7]. Các phương pháp được sử dụng phổ biến nhất trong hệ phương pháp thống kê hai biến là phân tích tỉ số tần suất - FR [8-13], mô hình trọng số bằng chứng - WoE [12, 14-22], mô hình giá trị thông tin - IV [23, 24], phương pháp chỉ số thống kê - SI [13, 25, 26].

Trong kỹ thuật thống kê đa biến, mô hình hồi quy logistic (LR) và phân tích biệt thức (DA) là hai phương pháp thống kê đa biến được sử dụng rộng rãi nhất trong đánh giá tính nhạy cảm với trượt lở đất bởi các nghiên cứu khác nhau ở các vùng khác nhau trên thế giới [9, 10, 22, 26-28].

Trong những nghiên cứu gần đây, các thuật toán học máy đã và đang được sử dụng để thực hiện các phân loại có giám sát cho phân tích nhạy cảm trượt lở, nghĩa là mô hình hóa sự phụ thuộc chưa biết giữa một số các tác nhân gây trượt lở (các tham số đầu vào) và sự hiện diện hay vắng mặt của trượt lở (đầu ra dạng nhị phân). Một số thuật toán học máy cơ bản như máy véc tơ hỗ trợ (SVM) [29-32], mạng thần kinh nhân tạo (ANN) [33-38], cây quyết định (DT) [29, 31], mạng Bayes (Bayesian Network) [39].

Để đánh giá tai biến lũ bùn đá, các phương pháp thường được sử dụng là mô phỏng dòng lan truyền vật liệu mà chủ yếu dựa vào các mô hình thực nghiệm [40, 41]. Các mô hình thực nghiệm kết nối khối lượng dòng bùn đá với khu vực trầm tích [42, 43], đưa đến một sự cải thiện cho việc đánh giá tai biến và một công cụ hữu ích cho phân vùng ví dụ như mô hình LAHARZ [42].

Mô hình 1D cũng đã được áp dụng mô phỏng dòng bùn đá [44, 45]. Rickenmann và nnk [46] đã thực hiện một nghiên cứu đánh giá các mô hình số hai chiều mô phỏng các kiểu trầm tích của các dòng bùn đá. Ba mô hình 2D được Rickenmann áp dụng là DFEM (Debris flow Finite Element Model), mô hình mô phỏng theo cách tiếp cận lưu biến học để xem xét các đặc trưng dòng bùn đá; mô hình FLO-2D là chương trình mô phỏng dòng bùn và dòng bùn đá dựa theo luật lưu biến học bậc hai; và một mô hình 2D là mô hình thể tích giới hạn bao gồm thuật toán lưu biến học Herchel-Bulkley (HB) đại diện cho dòng chất lỏng quán tính, sét, dòng bùn đá nhiều bùn với sự có mặt các hợp phần sét.

Các mô hình mô phỏng động năng khác như mô hình mô phỏng số là RAMMS-DF (rapid mass movement system-debris-flow) và DAN3D (dynamic analysis of landslides in three dimensions) thường được dùng để tính toán ngược các sự kiện dòng bùn đá tiêu biểu và để so sánh khoảng giá trị và độ nhạy của các tham số đầu vào đối với mô hình dòng Voellmy [47].

Một mô hình thực nghiệm đã được áp dụng phổ biến trong khoảng chục năm gần đây là mô hình FlowR (Flow path assessment of gravitational hazards at a Regional scale). Mô hình FlowR, mô hình thực nghiệm đánh giá dòng lan truyền của các loại tai biến liên quan đến trọng lực, [48-56], trong đó có dòng bùn đá được phát triển bởi đại học Lausanne. Mô hình này đã được áp dụng thành công ở nhiều khu vực khác nhau trên thế giới. Nó cũng thích hợp để đánh giá các tai biến tự nhiên khác như đá rơi đá đổ, lở tuyết và lũ [22, 50]. Ba tiêu chí quan trọng trong một tổ hợp liên quan đến việc bắt đầu dòng bùn đá: độ dốc địa hình, nguồn nước đầu vào và sự sẵn có của trầm tích. Hầu hết các dòng chảy mảnh vụn xảy ra từ địa hình có độ dốc lớn hơn

15°. Mặt cong địa hình được sử dụng để nhận dạng các rãnh xói. Với đặc điểm địa hình không thuận lợi như miền núi bị chia cắt mạnh, dốc, mạng lưới thoát nước dày đặc [50]. Bên cạnh hạn chế về địa hình, đây còn là khu vực hoạt động kiến tạo với cấu tạo địa chất phức tạp, có sự phức tạp của hệ thống đứt gãy, vỏ phong hóa dày. Ngoài ra, các yếu tố gây kích hoạt trượt lở đất như lượng mưa lớn cũng làm gia tăng các dạng trượt lở và lũ bùn đá. Trong nghiên cứu này, mô hình FlowR sẽ được áp dụng để thành lập bản đồ phân vùng cảnh báo về nguy cơ tai biến lũ bùn đá của khu vực nghiên cứu.

Ở Việt Nam, trong những năm gần đây, không ít các kết quả nghiên cứu về trượt lở đã được công bố trên các tạp chí quốc tế. Một số đã áp dụng các phương pháp chuyên gia, thống kê để thành lập bản đồ phân vùng nhạy cảm trượt lở [22, 57-59], mô hình kết hợp phương pháp chuyên gia (AHP), thống kê (WoE và LR) với mô hình FlowR để thành lập bản đồ nhạy cảm trượt lở, lũ bùn đá [22] và phương pháp học máy [39, 60-64] cho các khu vực nghiên cứu khác nhau tại Việt Nam. Trong công bố của [22], bản đồ nhạy cảm tai biến được tạo ra từ FlowR sẽ được tích hợp với bản đồ nhạy cảm trượt lở được thành lập theo 3 mô hình khác nhau AHP, WoE và LR từ 13 tác nhân trượt lở. Bản đồ nhạy cảm trượt lở được tích hợp từ các bản đồ LSI với bản đồ nhạy cảm được tạo ra từ FlowR sẽ khắc phục được những nhược điểm do không mô phỏng được mức độ lan truyền của các tai biến liên quan đến trọng lực từ các bản đồ nhạy cảm trượt lở đã thành lập từ 3 mô hình bởi 13 tác nhân. Điều này làm tăng độ chính xác của bản đồ nhạy cảm tích hợp khi so sánh các bản đồ trước và sau khi kết hợp trong phần thẩm định tính chính xác của bài báo.

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, với đặc điểm địa hình, địa chất phức tạp và sự gia tăng kiểu thời tiết cực đoan với đặc trưng mưa bất thường do biến đổi khí hậu toàn cầu cũng như sự không đồng bộ về dữ liệu đầu vào đã ảnh hưởng không nhỏ đến các bản đồ kết quả phân vùng

nguy cơ tai biến địa chất (trượt lở và lũ bùn đá). Các bản đồ này thường được thành lập ở tỷ lệ nhỏ và trung bình (từ tỷ lệ 1:500,000 đến 1:50,000) nên mang tính phức tạp và khái quát cao. Công việc thành lập bản đồ phân vùng nguy cơ trượt lở ở tỷ lệ lớn (1:10,000) đòi hỏi cần có được bộ cơ sở dữ liệu đầu vào ở mức độ tương đương thì kết quả mới có ý nghĩa về mặt dự báo các khu vực có tiềm năng dẫn đến các tai biến địa chất như trượt lở, lũ bùn đá trong tương lai.

Chính vì vậy, mục đích chính của bài báo này là áp dụng các phương pháp thống kê và học máy kết hợp với mô hình FlowR trong điều kiện dữ liệu đầu vào hạn chế (chủ yếu là dữ liệu được triết xuất từ bản đồ địa hình tỷ lệ 1:10,000 gồm các 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.

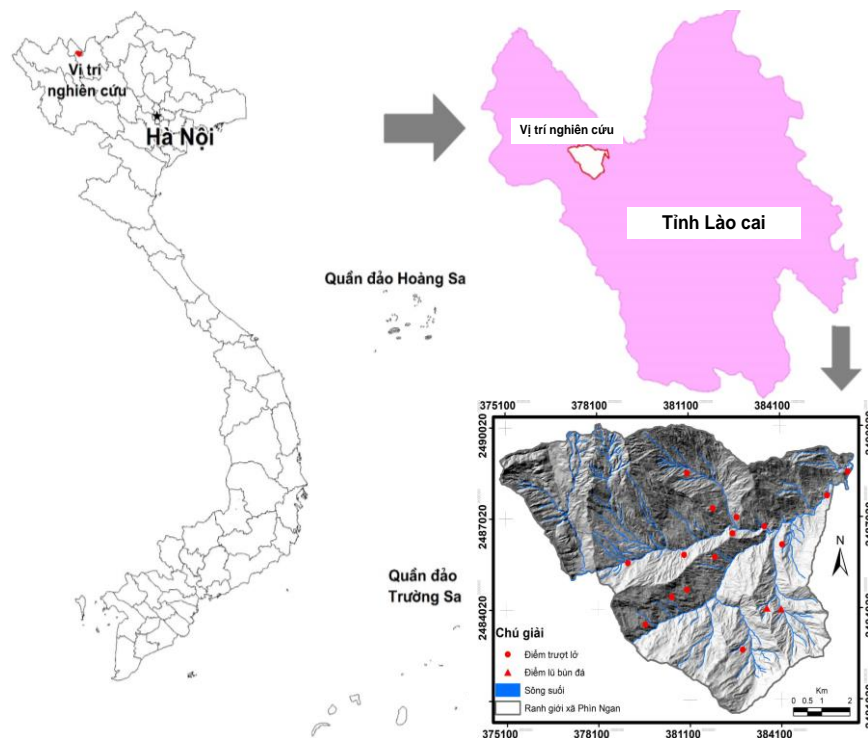
2. Khu vực nghiên cứu

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 km² 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.

Lũ quét, sạt lở đất là 2 loại hình thiên tai phổ biến nhất ở các tỉnh miền núi, trong đó có Lào Cai mỗi khi mùa mưa tới. Do địa hình núi cao, vực sâu lại có nhiều sông suối nên hiểm nguy thiên tai luôn rình rập bà con nơi đây. Hiện tại, trên địa bàn xã Phìn Ngan có 69 hộ dân trong vùng thiên tai nguy hiểm cần di dời [65]. Theo thông tin trên báo điện tử [66], số người chết do lũ quét và trượt lở đất trên địa bàn xã Phìn Ngan

từ năm 2004 cho đến năm 2020 là 29 người. Cụ thể số người chết do lũ quét và sạt lở đất năm

2004 là 23 người, năm 2016 là 3 người, 2018 là 1 người và ngày 6/8/2020 là 2 người.



Hình 1. Khu vực nghiên cứu xã Phìn Ngan, huyện Bát Xát, tỉnh Lào Cai.

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 trong phức hệ Po Sen với pha 1 (GD_i/PZ1_{ps}1) chủ yếu là diorit thạch anh, granodiorit và pha 2 (G/PZ1_{ps}2) là granit, granit biotit. Có một phần phía tây bắc diện tích nghiên cứu nằm trong hệ tầng Đá Đỉnh (NP_{dd}) 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.

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 m³), 3 điểm quy mô trung bình (200-1,000 m³) và 3 điểm quy mô lớn (1,000-20,000 m³). 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. Các loại đá này khi bị phong hóa mạnh và hoàn toàn sẽ tạo ra vật liệu phong hóa chứa nhiều sét, cát, mảnh vụn đá với độ gắn kết yếu, dễ dàng bị bão hòa nước [67].

Mục đích của nghiên cứu là thành lập bản đồ phân vùng nguy cơ tai biến trượt lở và lũ bùn đá ở tỷ lệ lớn 1:10,000 với sự hạn chế về mặt số liệu. Do đó, ngoài số liệu trượt lở như đã trình bày ở phần trước, 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. Do đó, 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 1m 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

Trượt lở hình thành bởi sự tương tác giữa các tác nhân địa hình và địa chất [68]. Một mô hình số độ cao (DEM) với kích thước pixel là 5x5m 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 10m 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 chia theo phương pháp phân loại Natural Break và được phân thành 5 nhóm với độ dốc thay đổi từ 0°-76.88° (Hình 2).

3.2. Mặt cong địa hình

Mặt cong địa hình sẽ cho ta biết được các hình dạng mặt lồi, mặt lõm và khu vực phẳng của địa hình. Nó cũng 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 [69, 70] (Hình 2).

3.3. Chỉ số độ ẩm địa hình-TWI

TWI (Hình 2) là một trong những tác nhân quan trọng ảnh hưởng đến sự mất ổn định của sườn [71, 72]. 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. TWI biểu thị sự phân bố độ ẩm của đất trong không gian [72, 73]. TWI là tác nhân liên quan Bản đồ TWI của 4 xã 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 [71, 74]. Tính toán chỉ số độ ẩm địa hình theo công thức (1) dưới đây được đề xuất bởi Pradhan và những người khác (nnk) [71]:

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

Trong công thức (1) và (2), A (m²) là tích lũy dòng chảy, b (m) là chiều rộng ô pixel mà nước chảy qua, β (radian) là độ dốc sườn.

3.4. Chỉ số năng lượng dòng chảy-SPI

SPI (Hình 2) 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 phụ phân bố của thực vật [71, 74]. SPI dùng để ước lượng năng lượng xói mòn của nước. Giá trị xói mòn càng cao ở dọc các kênh, rãnh thoát nước [72]. Chỉ số năng lượng dòng chảy được Pradhan và Kim [71] đề xuất theo công thức dưới đây:

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

3.5. Chỉ số vận chuyển trầm tích-STI

STI (Hình 2) 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 [70, 75]. Trong đó chiều dài sườn được tính toán theo công thức (3). Chiều dài sườn được tạo thành bởi sự kết hợp của độ dốc sườn (S) và chiều dài sườn (L) và được đề cập đầu tiên bởi

Moore và Burch [76]. Trong công thức (3), A_s là diện tích lưu vực và β là độ dốc.

$$LS = \left(\frac{A_s}{22.13} \right)^{0.6} \times \left(\frac{\sin\beta}{0.0896} \right)^{1.3} \quad (3)$$

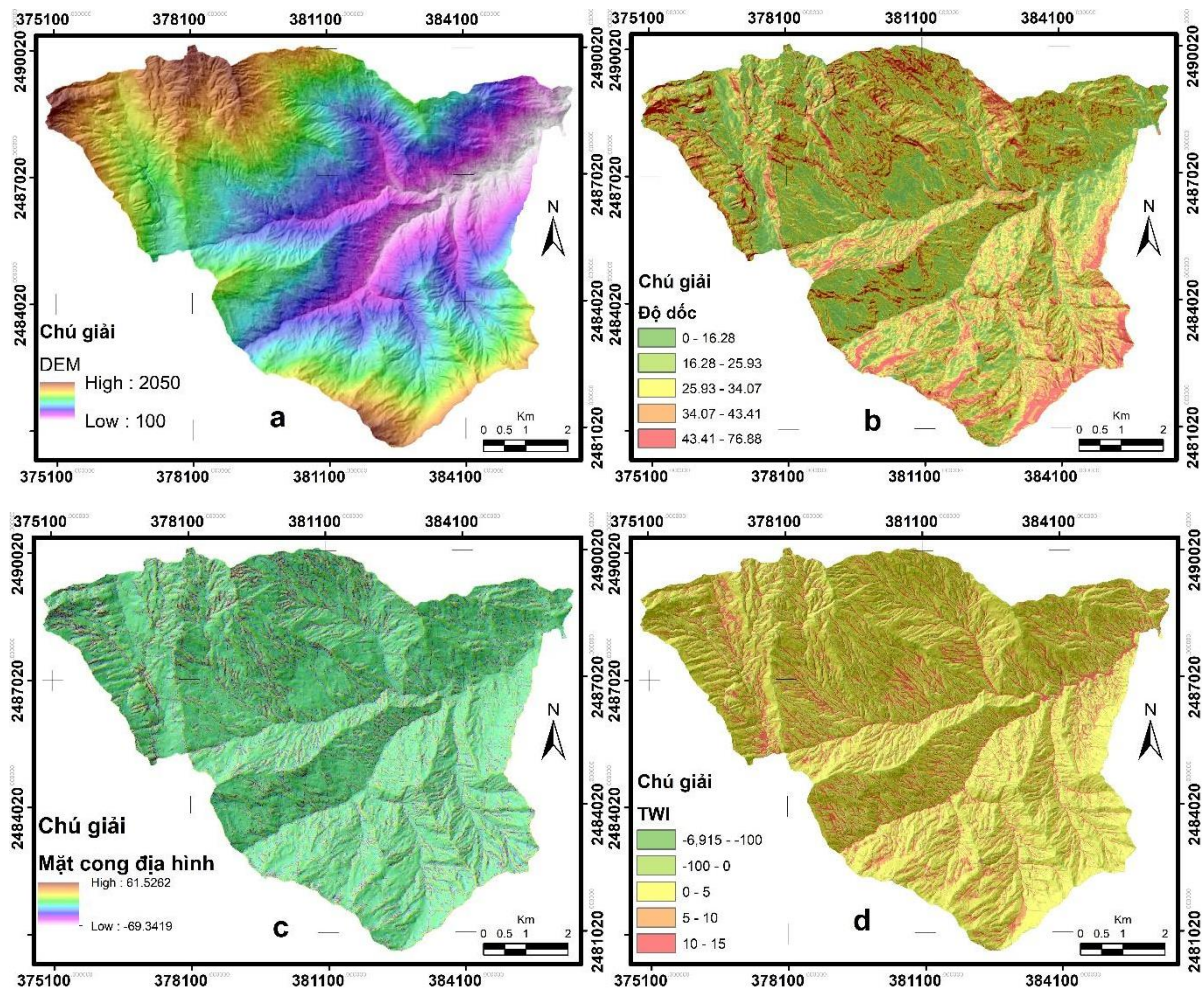
3.6. Thảm phủ

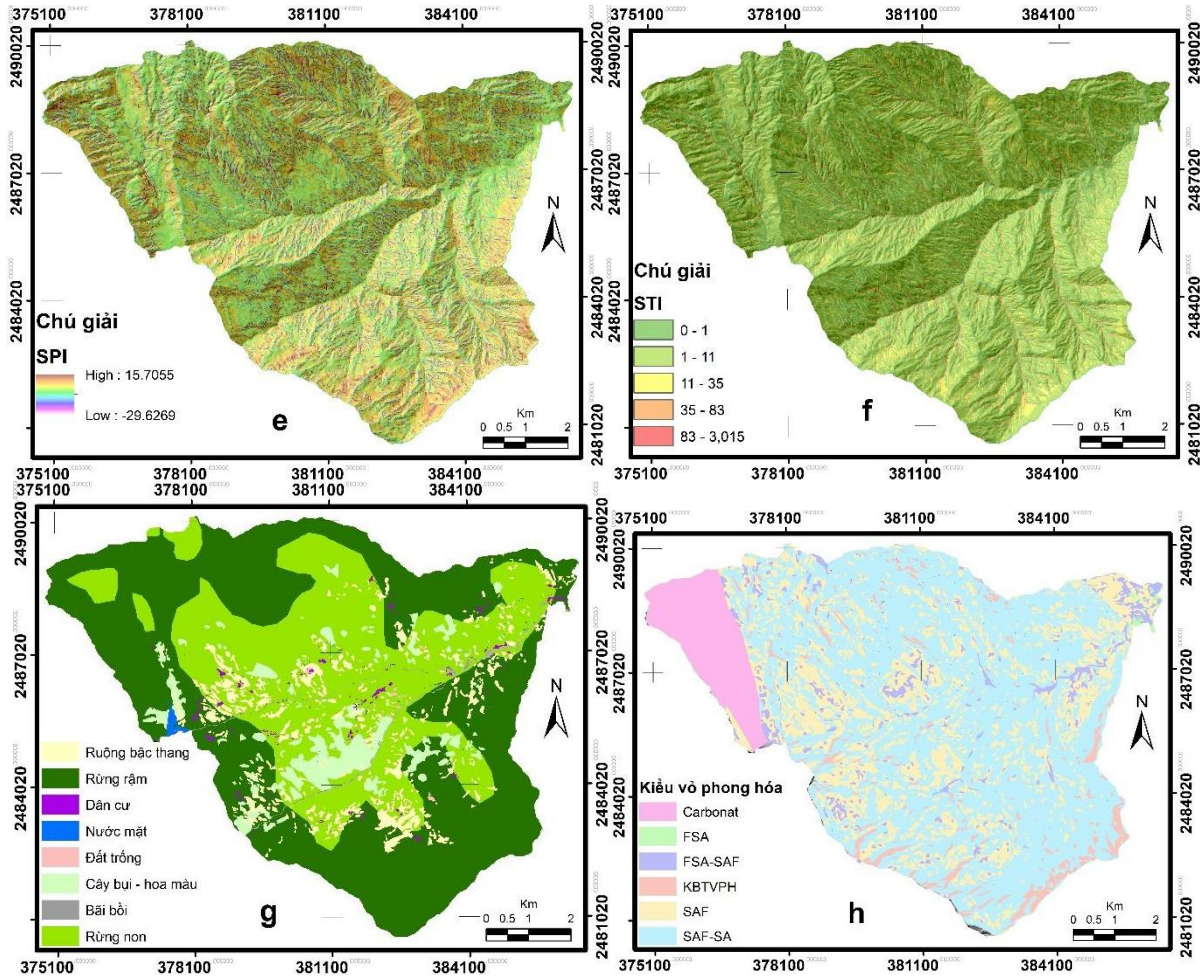
Bản đồ thảm phủ của khu vực nghiên cứu được thành lập từ ảnh vệ tinh phân giải cao được tải từ phần mềm Google Earth Pro với ngày thu nhận ảnh là tháng 11 năm 2020 với độ phân giải 1x1 m. Bản đồ này (Hình 2) được phân loại thành 8 lớp: rừng rậm, rừng non, đất trồng, cây bụi -

hoa màu, dân cư, nước mặt, ruộng bậc thang và bãi bồi.

3.7. Vô phong hóa

Bản đồ vô 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 vô phong hóa gồm: carbonat, FSA-vô phong hóa Ferrosialit, FSA-SFA-vô phong hóa Ferrosialit-Sialferit, SAF-vô phong hóa Sialferit, SAF-SA-vô phong hóa Sialferit-Sialit và KBTVPH (Hình 2).



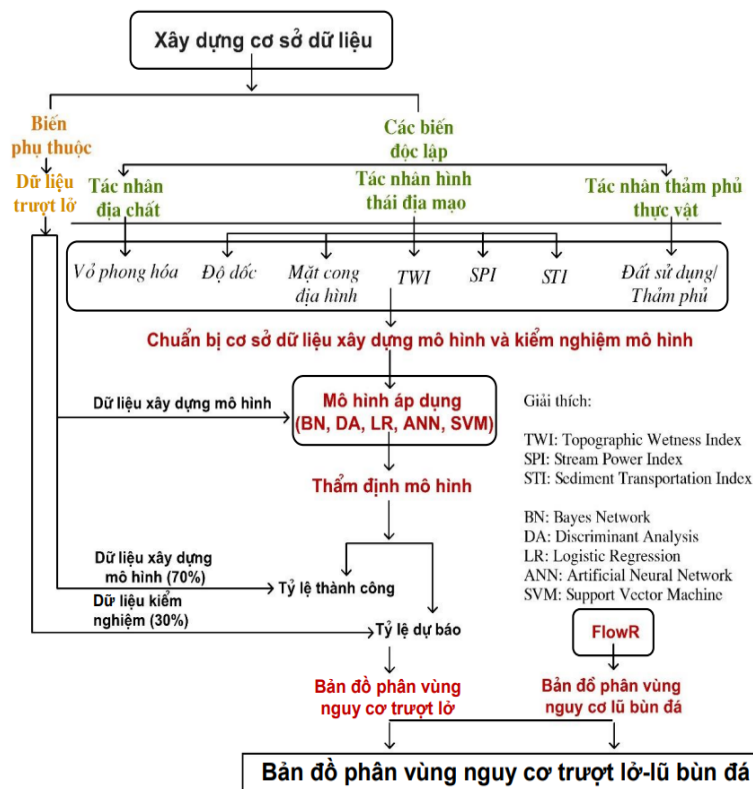


Hình 2. Các bản đồ tác nhân trượt lở khu vực xã Phin Ngan (a-DEM, b-độ dốc, c-mặt cong địa hình, d-TWI, e-SPI, f-STI, g-thảm phủ và h-vỏ phong hóa).

4. Hệ phương pháp

Phân tích nhạy cảm trượt lở trong nghiên cứu này gồm 6 bước chính (Hình 3): i) 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; ii) Phân chia dữ liệu trượt lở thành hai tập dữ liệu xây dựng mô hình (70%) và kiểm nghiệm mô hình (30%); iii) Áp dụng 5 mô hình thống kê và học máy (hồi quy logistic-Logistic Regression, phân tích biệt thức-Discriminant Analysis, mạng Bayes-Bayes Network, mạng thần kinh nhân tạo-

Artificial Neural Net và máy véc tơ hỗ trợ-Support Vector Machine) để thành lập bản đồ chỉ số nhạy cảm trượt lở (LSI); iv) Thẩm định mô hình để đánh giá tỷ lệ thành công và tỷ lệ dự báo của các bản đồ LSI; v) Thành lập bản đồ phân vùng nguy cơ lũ bùn đá bằng mô hình FlowR; và vi) Kết hợp bản đồ phân vùng nguy cơ trượt lở và bản đồ phân vùng nguy cơ lũ bùn đá để đưa ra bản đồ phân vùng nguy cơ tai biến trượt lở và lũ bùn đá của khu vực nghiên cứu ở tỷ lệ 1:10,000.



Hình 3. Hệ phương pháp thực hiện trong nghiên cứu này.

4.1. Mạng Bayes-BN

BN là một phương pháp hiệu quả để biểu diễn tri thức theo ảnh hưởng không chắc chắn [77], nó được giới thiệu đầu tiên bởi Friedman và nnk [78]. 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 [79]. BN thường sử dụng rộng rãi trong các mô hình hệ thống phức tạp [80]. 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 dưới đây:

$$P_B(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P_B(X_i | \prod_{x_i}) = \prod_{i=1}^n \theta_{x_i | \prod_{x_i}} \quad (4)$$

Trong đó, $X = (X_1, X_2, \dots, X_n)$ là các tác nhân gây 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.

4.2. Hồi quy logistic-LR

LR là một trong các phương pháp đa biến được sử dụng rộng rãi nhất trong đánh giá nhạy cảm trượt lở ở các khu vực khác nhau trên thế giới [22, 26-27, 81]. Cốt lõi 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 [9, 22, 26]. 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 sau [9, 22, 26, 82]:

$$P = 1/(1+e^{-Z}) \quad (5)$$

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 sau [9, 22, 26]:

$$Z = C_0 + C_1X_1 + C_2X_2 + \dots + C_nX_n \quad (6)$$

Trong công thức (6), 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.

4.3. Phân tích biệt thức-DA

Phương pháp phân tích biệt thức (DA) để phân loại mẫu thành các nhóm thay thế trên cơ sở một tập hợp các phép đo. Theo Lee và nnk [83], DA cho phép xác định sự khác biệt tối đa cho từng biến độc lập (ví dụ: yếu tố gây trượt lở) giữa nhóm trượt lở và nhóm không trượt lở và xác định trọng số cho các yếu tố này.

Mục đích của phân tích biệt thức là lập một phương trình tuyến tính (hàm biệt thức) để tách hai hoặc nhiều nhóm đối tượng liên quan đến một số biến đồng thời [84]. Hàm biệt thức có dạng:

$$D = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n, \quad (7)$$

Ở đó, D là giá trị biệt thức, X_i ($i = 1, 2, \dots, n$) là các biến độc lập, b_i ($i = 0, 1, 2, \dots, n$) là hệ số chính tắc chưa chuẩn hóa của hàm phân biệt đối với biến thứ i ; và n là số biến độc lập. Nếu một điểm trượt lở với tham số X_i cho kết quả là $D > 0$, thì nó được phân loại vào nhóm ổn định; nếu không, nó sẽ được xếp vào nhóm không ổn định. Hệ số của hàm biệt thức tuyến tính có thể được xác định bằng cách giải bài toán giá trị riêng tổng quát. Một số văn bản thống kê đa biến [84, 85] đề cập đến suy luận toán học để tính hệ số.

Hai giả định cơ bản về đặc tính thống kê của các yếu tố nguyên nhân trong một hàm biệt thức là: i) Mỗi nhóm được lấy từ một tập hợp có phân phối chuẩn đa biến; và ii) Không có biến nào có thể là sự kết hợp tuyến tính của các biến khác [84].

4.4. Mạng thần kinh nhân tạo-ANN

Là một thuật toán thường được sử dụng trong lập bản đồ tính nhạy cảm với trượt lở, mạng nơ-ron nhân tạo là các hàm phi tuyến chung mô phỏng hệ thống thần kinh của não người. Chúng độc lập với phân phối thống kê của dữ liệu đào tạo và có thể xử lý dữ liệu liên tục, phân loại và nhị phân [38]. Trong số nhiều loại mô hình mạng nơ-ron nhân tạo, mạng chuyển tiếp ba lớp được sử dụng phổ biến trong thành lập bản đồ nhạy cảm trượt lở [33, 34, 36, 37]. Số lượng tế bào thần kinh trên lớp ẩn được tính theo kinh nghiệm tùy theo các ứng dụng khác nhau. Như đã đề cập bởi Liu và nnk [86], số lượng tế bào thần kinh trên lớp ẩn nhiều gấp đôi số tế bào thần kinh đầu vào. Đối với mô hình mạng nơ-ron nhân tạo, dữ liệu đào tạo và các yếu tố gây ra trượt lở được chọn tùy vào mục đích nghiên cứu. Trong một mạng nơ-ron nhân tạo, một mạng chuyển tiếp 3 lớp gồm một lớp đầu vào, một lớp ẩn và một lớp đầu ra được sử dụng làm cấu trúc mạng.

Mục đích của mạng nơ-ron nhân tạo là xây dựng một mô hình của quá trình tạo dữ liệu, để mạng có thể tổng quát hóa và dự đoán kết quả đầu ra từ các đầu vào mà nó chưa từng thấy trước đây. Thuật toán học tập này là một mạng nơ-ron nhiều lớp, bao gồm một lớp đầu vào, các lớp ẩn và một lớp đầu ra. Các nơ-ron của lớp ẩn và lớp đầu ra xử lý đầu vào của chúng bằng cách nhân mỗi đầu vào với trọng số tương ứng, tính tổng tích, sau đó xử lý tổng bằng cách sử dụng hàm truyền phi tuyến để tạo ra kết quả. Một mạng nơ-ron nhân tạo “học” bằng cách điều chỉnh trọng số giữa các nơ-ron để đáp ứng các sai số giữa giá trị đầu ra thực tế và giá trị đầu ra mục tiêu. Vào cuối giai đoạn huấn luyện này, mạng nơ-ron cung cấp một mô hình có thể dự đoán giá trị mục tiêu từ một giá trị đầu vào nhất định [87].

Dựa trên sự sắp xếp mạng thần kinh và kiểu kết nối của các lớp, các mạng lưới thần kinh khác nhau có thể được tạo. Hai mô hình mạng thần kinh cơ bản hay áp dụng là multilayer perceptron-MLP và radial basic function-RBF là hai trong số các mạng thần kinh nổi tiếng với nhiều ứng dụng trong giải quyết vấn đề [88].

4.5. Máy véc tơ hỗ trợ-SVM

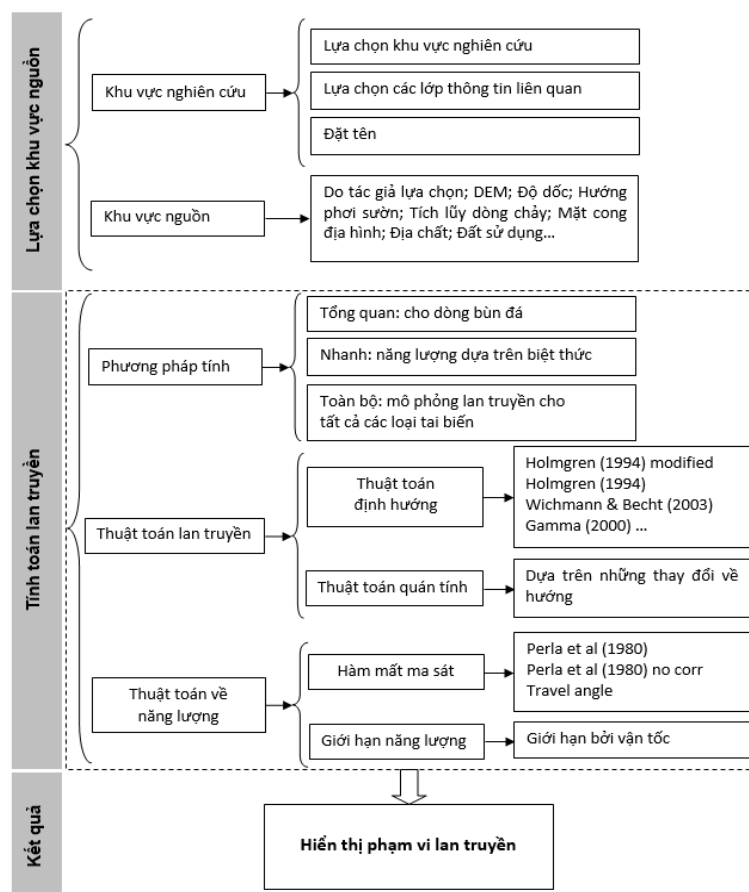
Máy vectơ hỗ trợ là kỹ thuật dựa trên hạt nhân không tham số được rút ra từ lý thuyết thống kê học [89]. Chúng đặc biệt hấp dẫn để giải các bài toán phân loại, hồi quy và ước lượng mật độ cao phi tuyến bằng cách kiểm soát độ phức tạp của mô hình [90-92]. SVM xem xét tập dữ liệu trượt lở đầu vào trong 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 [93]:

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

Trong đó, $X = (X_1, X_2, \dots, X_n)$ 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, α_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 [94, 95]. Để phân loại nhị phân cho trượt lở và không trượt lở, điều kiện để giải phương trình (8) được mô phỏng dưới đây:

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

Ở đó, $\varphi(X_i)$ là một hàm phi tuyến chia không gian đầu vào thành không gian thứ nguyên cao hơn, ω đại diện cho véc tơ trọng số [39].



Hình 4. Các quá trình thực hiện trong mô hình FlowR.

4.6. Mô hình FlowR

Trong nghiên cứu này, mô hình FlowR sẽ được áp dụng để thành lập bản đồ phân vùng nguy cơ lũ bùn đá của khu vực nghiên cứu.

Dữ liệu đầu vào của mô hình FlowR là DEM và các dữ liệu khác được định nghĩa bởi các tiêu chuẩn về hình thái địa mạo và của người sử dụng. Tập dữ liệu đầu vào có thể đại diện cho các loại thông tin không gian khác nhau như: độ dốc, hướng phơi sườn, tích lũy dòng chảy, mặt cong địa hình, đất sử dụng,... và các tham số khác do người dùng định nghĩa.

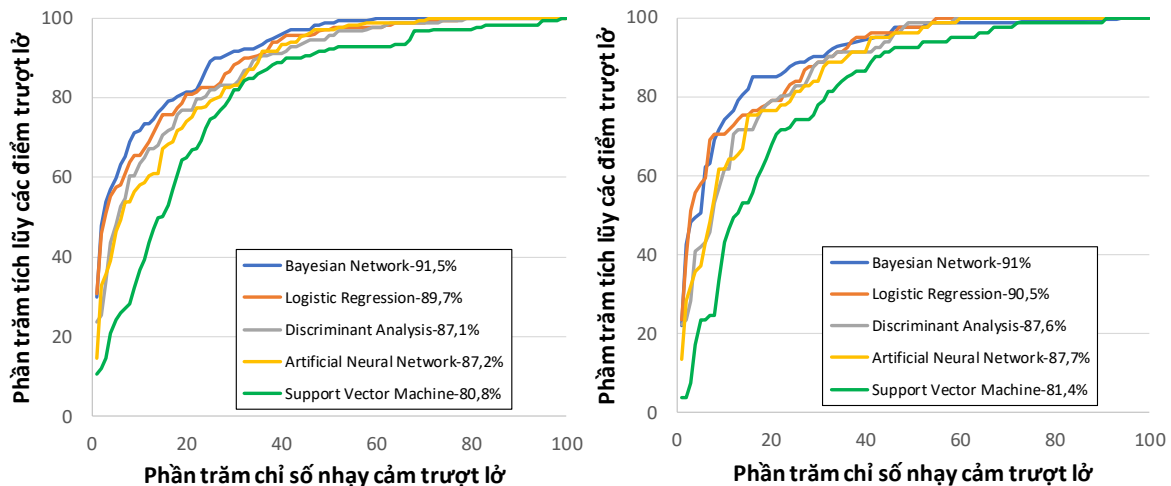
Mô hình gồm hai bước phân biệt dựa trên mô hình số độ cao DEM: i) Các khu vực nguồn được xác định đầu tiên bởi các tiêu chuẩn định nghĩa bởi người dùng và các dạng hình thái học; và sau đó ii) Các dòng bùn đá được lan truyền từ những nguồn này trên nền tảng các thuật toán định hướng dòng chảy và các luật ma sát [22, 50]. Hình 4 trình bày các quá trình xử lý trong mô hình FlowR [22].

4.7. Phương pháp đánh giá và so sánh

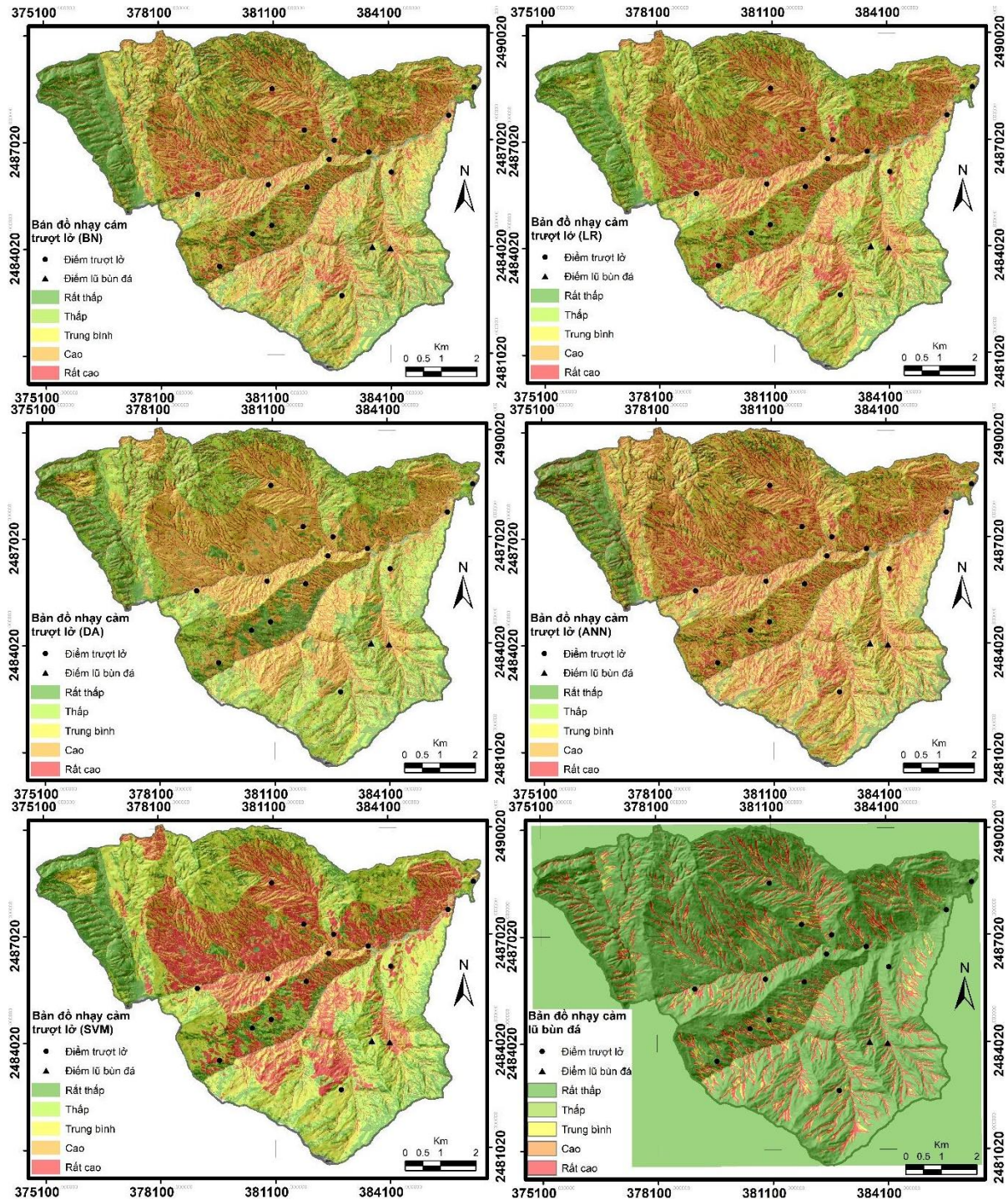
Có nhiều cách để đánh giá độ chính xác của các bản đồ phân vùng nhạy cảm trượt lở. Cách

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 [96]. Đánh giá kết quả dự báo là một trong những nhiệm vụ quan trọng nhất trong mô hình nhạy cảm trượt lở. 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 [12, 97, 98]. 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-square, đường cong đặc tính hoạt động (ROC),... đã được sử dụng để xác định độ chính xác của các mô hình trượt lở.

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”. 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 [29]. 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 [62].



Hình 5. Đường cong ROC cho tập dữ liệu xây dựng (5a) và kiểm tra mô hình (5b) của các phương pháp thống kê và học máy đã áp dụng cho khu vực nghiên cứu.



Hình 6. Các bản đồ nhạy cảm trượt lở được thành lập từ các phương pháp khác nhau (BN, LR, DA, ANN, SVM) và bản đồ nhạy cảm lũ bùn đá được thành lập từ mô hình FlowR.

5. Kết quả và thảo luận

5.1. Bản đồ nhạy cảm trượt lở

Trong nghiên cứu này, năm mô hình đã được sử dụng để thành lập các bản đồ nhạy cảm trượt lở. Các bản đồ này được thành lập dựa trên hai bước chính: i) Thành lập bản đồ chỉ số nhạy cảm trượt lở (LSI); và ii) Phân loại các bản đồ LSI thành các bản đồ nhạy cảm trượt lở (LSM). Ở bước đầu, các bản đồ LSI đã được tạo ra cho khu vực nghiên cứu mà mỗi ô pixel được ấn định một giá trị nhạy cảm. Trong bước hai, các bản đồ LSI sẽ được phân loại lại theo phương pháp quantile trong phần mềm ArcGIS 10.8 để thành lập các bản đồ LSM. Dựa trên cách phân khoảng của phương pháp này, các lớp nhạy cảm được xác định để thành lập các bản đồ LSM gồm: rất thấp, thấp, trung bình, cao và rất cao (Hình 6).

5.2. Thẩm định các mô hình trượt lở

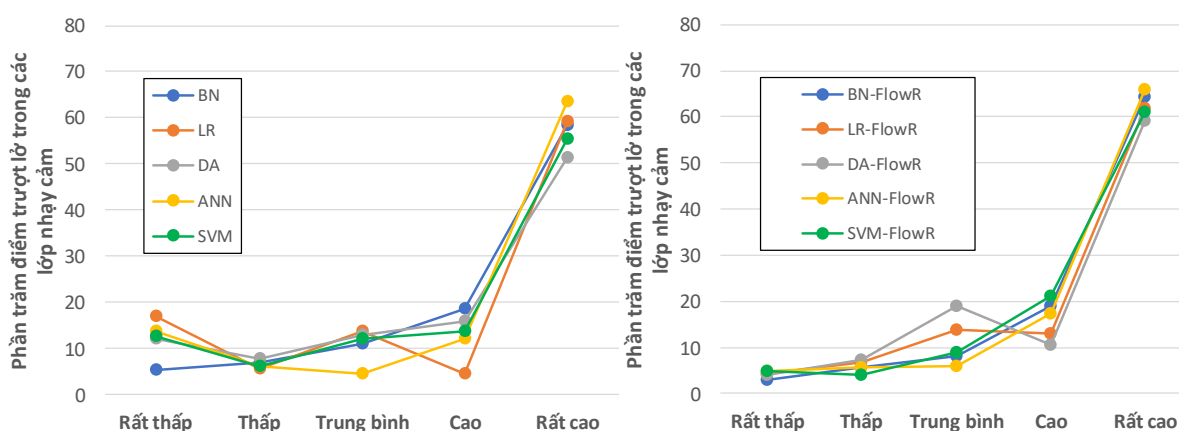
Để đánh giá và so sánh các mô hình dự báo, hai tập dữ liệu xây dựng mô hình và kiểm nghiệm mô hình đã được tạo ra. Đầu tiên, dữ liệu trượt lở được phân chia ngẫu nhiên thành hai tập dữ liệu này. Tập dữ liệu xây dựng mô hình chiếm 70% số pixel trượt lở (177 pixel) và tập dữ liệu kiểm nghiệm mô hình chiếm 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, ANN, DA 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= 91,5%), tiếp theo là mô hình LR (AUC= 89,7%), ANN (AUC= 87,2%), DA (AUC= 87,1%) và SVM (AUC= 80,8%). Sau đó, 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 (91%), tiếp theo LR (AUC= 90,5%), ANN (AUC= 87,7%), DA (AUC=87,6%) và SVM (AUC= 81,4%). So sánh với các kết quả thẩm định mô hình được công bố trước đây, giá trị AUC của phương pháp LR ở mức độ tương đương, không có sự chênh lệch

hiều. Giá trị AUC này trong các công bố của [9, 10, 23, 26, 31, 69, 70, 99] lần lượt là 86,37%, 90,34%, 74,2%, 80,9%, 81,1%, 76%, 77,51%, 84%. Đối với các kết quả thẩm định của mô hình BN, giá trị AUC được kiểm chứng từ tập dữ liệu xây dựng mô hình và tập dữ liệu kiểm nghiệm mô hình lần lượt trong các công bố của [39, 61] lần lượt là 97,1%, 91,5% và 90,3% và 91,5%. Như vậy, kết quả thẩm định của đề tài đối với tập dữ liệu xây dựng mô hình và tập dữ liệu kiểm nghiệm mô hình theo phương pháp BN có giá trị AUC là 91,5% và 91% cũng có giá trị tương đương với các công bố trước đây. Mô hình ANN áp dụng trong nghiên cứu này cũng có giá trị AUC không chênh lệch nhiều so với một số công bố trước đây của [69] là 89%, [100] là 74,29%. Giá trị AUC của phương pháp ANN áp dụng trong nghiên cứu này lần lượt là 87,2% và 87,7% cho tập dữ liệu xây dựng và kiểm tra mô hình. Các nghiên cứu áp dụng mô hình học máy SVM trước kia của [31, 39] đã đưa ra giá trị AUC lần lượt là 82,4%, 95,5%. Trong khi đó, một nghiên cứu của [32] đã kiểm nghiệm giá trị AUC của mô hình SVM theo hai phương pháp đường cong ROC, đường cong tỷ lệ thành công (SRC) và đường cong tỷ lệ dự báo (PRC) cho tập dữ liệu xây dựng mô hình và kiểm nghiệm mô hình có giá trị lần lượt là 96,9%, 91,8% và 85,51%, 82,46%. Cuối cùng, giá trị AUC của mô hình DA trong các nghiên cứu trước đây của [39, 101, 102] lần lượt là 88,4%, 89% và 92,2%, cũng có giá trị tương đương với kết quả nghiên cứu này.

Trong nghiên cứu này, cùng với dữ liệu trượt lở và lũ bùn đá, 7 bản đồ tác nhân được sử dụng để thành lập bản đồ nhạy cảm nguy cơ lũ bùn đá bao gồm: độ dốc, TWI, SPI, STI, mật độ địa hình, vỏ phong hóa và thảm phủ/đất sử dụng đã cho thấy khả năng tin cậy của các mô hình. Mặc dù chỉ có 7 tác nhân đầu vào, tuy nhiên, khi so sánh giá trị AUC để đánh giá độ chính xác của các mô hình trong nghiên cứu này với các nghiên cứu trước kia thì giá trị này là tương đương. Một số công bố đã áp dụng các mô hình học máy với số lượng các bản đồ tác nhân đầu vào nhiều hơn trong nghiên cứu này nhưng có giá trị AUC tương đương phải kể đến như [26, 39] đã sử dụng 15 bản đồ tác nhân; [30, 31, 69, 70] với 14 tác

nhân trượt lở; [22] với 13 bản đồ tác nhân; [99] với 12 bản đồ tác nhân. Ngoài ra, trong 7 bản đồ tác nhân đầu vào đã sử dụng trong nghiên cứu này thì có đến 6 bản đồ tác nhân trượt lở (độ dốc, mặt cong địa hình, TWI, SPI, STI, thảm phủ/đất sử dụng) có mặt hầu hết trong các công bố trước đây. Như vậy, mặc dù có sự hạn chế về mặt dữ liệu nhưng những tác nhân đã lựa chọn làm đầu vào để thành lập bản đồ nhạy cảm tai biến trượt lở và lũ bùn đá trong khu vực nghiên cứu có sự liên quan chặt chẽ đến các đặc điểm hình thái địa mạo và thảm phủ/đất sử dụng của khu vực

nghiên cứu. Từ đó có thể kết luận rằng, dữ liệu phản ánh các đặc điểm hình thái, địa mạo và thảm phủ/đất sử dụng là không thể thiếu trong việc thành lập các bản đồ nhạy cảm tai biến trượt lở và lũ bùn đá. Từ các kết quả so sánh với các nghiên cứu trước, có thể khẳng định rằng các kết quả tính toán trong nghiên cứu này đáng tin cậy và hữu ích để áp dụng cho khu vực xã Phìn Ngan nói chung và các khu vực miền núi khác của Việt Nam nói riêng trong công tác thành lập bản đồ nhạy cảm trượt lở ở tỷ lệ lớn.



Hình 7. Đồ thị thể hiện phần trăm các điểm trượt lở trên các bản đồ phân vùng nhạy cảm thành lập từ các phương pháp thống kê và học máy trước (hình trái) và sau khi được kết hợp với mô hình FlowR (hình phải).

5.3. Bản đồ nhạy cảm lũ bùn đá

Bảng 1. Tham số khai báo trong phần mềm FlowR

Dữ liệu nguồn	- DEM	≥ 300 m
	- Độ dốc	$\geq 15^0$
Thuật toán áp dụng	- Tích lũy dòng chảy	
	Thuật toán định hướng	Holmgren (1994) modified
	Thuật toán quán tính	weights
	Hàm mất năng lượng	Góc dịch chuyển (5 độ)
	Giới hạn năng lượng	Vận tốc <10 mps

Bản đồ nhạy cảm lũ bùn đá của khu vực nghiên cứu được thành lập bởi mô hình FlowR

theo quy trình Hình 4. Dữ liệu và tham số đã áp dụng để thành lập bản đồ nhạy cảm lũ bùn đá được trình bày trong Bảng 1. Bản đồ nhạy cảm lũ bùn đá của khu vực nghiên cứu được trình bày trong Hình 6.

Mô hình FlowR dựa vào một phương pháp phân bố thực nghiệm để thành lập bản đồ nhạy cảm tai biến. Điểm mạnh của mô hình FlowR đó là sự đơn giản về dữ liệu đầu vào. Chỉ cần dữ liệu DEM và một số dữ liệu khác do người dùng định nghĩa như dữ liệu định hướng dòng chảy, dữ liệu tích lũy dòng chảy, dữ liệu độ dốc hoặc địa chất hay thảm phủ và các thuật toán để thành lập các bản đồ nhạy cảm tai biến khác nhau cho những khu vực nghiên cứu khác nhau như lũ bùn đá, trượt lở hay lũ lụt. Các khu vực mô phỏng lan truyền tai biến từ các kết quả nói chung có diện phân bố lớn hơn các sự kiện quan sát được ngoài

thực địa, điều này rất có ý nghĩa trong thành lập bản đồ tính nhạy cảm. Thật vậy, bản đồ nên mô phỏng mọi sự kiện có thể xảy ra, ngay cả trường hợp xấu nhất. Bản đồ tính nhạy cảm được thành lập từ mô hình FlowR cung cấp một cái nhìn tổng quan để chỉ ra nơi cần tiến hành điều tra thực địa để thiết lập một bản đồ mức độ chi tiết [50].

Đánh giá độ chính xác của bản đồ nhạy cảm lũ bùn đá khu vực nghiên cứu trình bày trong 6 đã chỉ ra rằng hai điểm lũ bùn đá thu thập từ thực địa hoàn toàn nằm trong phần diện tích nguy cơ cao và rất cao của bản đồ này. Ngoài ra, một số điểm trượt lở cũng nằm trong phần diện tích có nguy cơ cao và rất cao ở bản đồ này. Như vậy, bản đồ nhạy cảm lũ bùn đá được thành lập từ mô hình FlowR đã cho thấy khả năng dự báo chính xác về loại tai biến này trong khu vực nghiên cứu.

5.4. Bản đồ phân vùng nhạy cảm trượt lở và lũ bùn đá

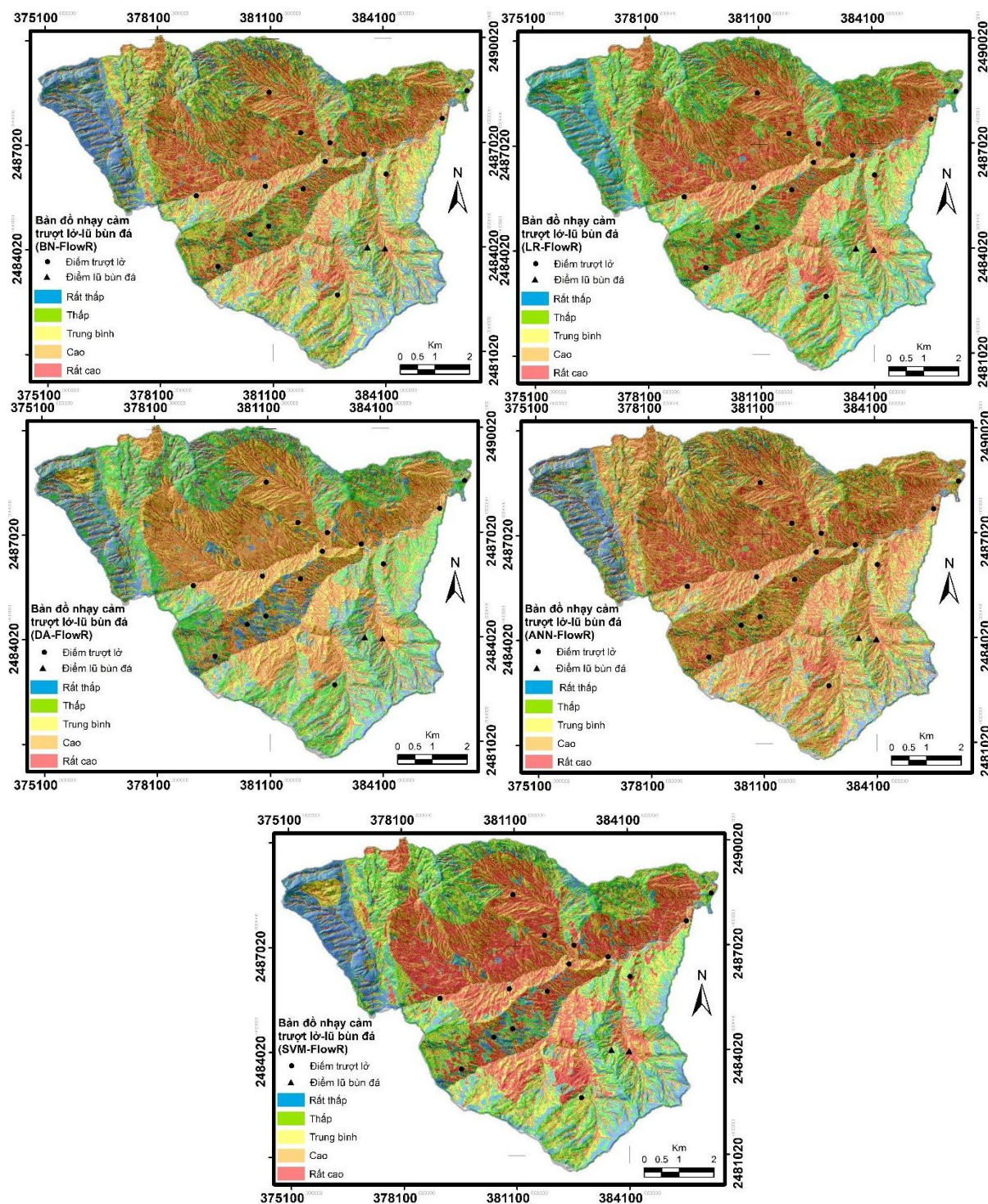
Bản đồ phân vùng nhạy cảm trượt lở và lũ bùn đá được tạo thành bằng việc kết hợp bản đồ LSM được thành lập từ năm mô hình thống kê và học máy với bản đồ phân vùng nhạy cảm lũ bùn đá được tạo ra từ mô hình FlowR. Nguyên tắc kết hợp hai bản đồ này được thực hiện trong phần mềm GIS, dựa trên ma trận kết hợp giữa các lớp có cùng mức độ rủi ro của một bản đồ LSM với bản đồ được tạo ra từ mô hình FlowR. Hình 8 là ma trận kết hợp để thành lập bản đồ nhạy cảm trượt lở và lũ bùn đá của khu vực nghiên cứu. Hình 9 là 5 bản đồ phân vùng nguy

cơ trượt lở và lũ bùn đá được tạo ra từ 5 mô hình kết hợp: BN-FlowR, LR-FlowR, DA-FlowR, ANN-FlowR và SVM-FlowR.

Để so sánh các kết quả, công việc tiếp theo là đánh giá phần trăm các điểm trượt ở mỗi lớp nhạy cảm trên các bản đồ nhạy cảm trượt lở được tạo ra từ 5 phương pháp học máy và từ 5 phương pháp kết hợp. Phần trăm của điểm trượt lở trong mỗi lớp nhạy cảm của các bản đồ kết quả được thể hiện trong Hình 7. Trong Hình 7, phần trăm các điểm trượt lở nằm trong các lớp nhạy cảm rất thấp, thấp trung bình và cao được thành tạo từ các mô hình BN, LR, DA, ANN, SVM có sự chênh lệch không nhiều. Tuy nhiên, có sự chênh lệch đáng kể về số lượng điểm trượt nằm trong lớp rất cao. Để đánh giá khả năng dự báo và so sánh các bản đồ kết quả, số lượng điểm trượt lở (phần trăm điểm trượt) nằm trong hai lớp nhạy cảm cao và rất cao cũng là một trong các tiêu chí đánh giá. Phần trăm điểm trượt lở trong hai lớp cao và rất cao của các phương pháp BN, LR, DA, ANN và SVM lần lượt là 76,92%, 63,56%, 67,21%, 75,71% và 69,23%. Phần trăm này đã tăng lên từ hơn 2% đến 11% khi mà đánh giá theo các mô hình kết hợp. Giá trị phần trăm các điểm trượt lở xuất hiện trong hai lớp nhạy cảm cao và rất cao của 5 mô hình kết hợp BN-FlowR, LR-FlowR, DA-FlowR, ANN-FlowR và SVM-FlowR lần lượt là 83,40%, 74,90%, 69,64% 83,40% và 82,19%. Như vậy, các mô hình kết hợp đã cải thiện khả năng dự báo của các bản đồ phân vùng nguy cơ tai biến trượt lở và lũ bùn đá của khu vực nghiên cứu.

Các lớp nhạy cảm tại biển lũ bùn đá	Các lớp nhạy cảm tại biển trượt lở					
	Rất thấp	Thấp	Trung bình	Cao	Rất cao	
Rất thấp	Rất thấp	Thấp	Trung bình	Cao	Rất cao	
Thấp	Thấp	Thấp	Trung bình	Cao	Rất cao	
Trung bình	Trung bình	Trung bình	Trung bình	Cao	Rất cao	
Cao	Cao	Cao	Cao	Cao	Rất cao	
Rất cao	Rất cao	Rất cao	Rất cao	Rất cao	Rất cao	

Hình 8. Ma trận tích hợp hai bản đồ nhạy cảm trượt lở và nhạy cảm lũ bùn đá.



Hình 9. Các bản đồ nhạy cảm trượt lở và lũ bùn đá của khu vực nghiên cứu được thành lập từ các mô hình kết hợp BN-FlowR, LR-FlowR, DA-FlowR, ANN-FlowR, SVM-FlowR.

6. Kết luận

Nghiên cứu này đã áp dụng các mô hình thống kê, học máy kết hợp với mô hình FlowR để thành lập bản đồ phân vùng nguy cơ trượt lở và lũ bùn đá ở tỷ lệ 1:10,000 cho khu vực nghiên cứu xã Phìn Ngan, huyện Bát Xát, tỉnh Lào Cai. Đầu tiên, 5 bản đồ phân vùng nguy cơ tai biến trượt lở được thành lập từ 5 mô hình BN, LR, DA, ANN và SVM. Tiếp theo, một bản đồ phân vùng nguy cơ tai biến lũ bùn đá được thành lập từ mô hình FlowR. Sau đó lần lượt 5 bản đồ phân vùng nguy cơ trượt lở sẽ được kết hợp với bản đồ phân vùng nguy cơ tai biến lũ bùn đá để tạo ra 5 bản đồ phân vùng nguy cơ tai biến trượt lở và lũ bùn đá của khu vực nghiên cứu. Kết quả thẩm định mô hình (Hình 5) cho các phương pháp thống kê và học máy cho thấy khả năng dự báo của từng mô hình là khá cao và không có nhiều khác biệt với các nghiên cứu đã công bố. Mô hình BN và LR có giá trị AUC gần tương đương nhau, lần lượt là 91,5% và 89,7%. Giá trị AUC của hai mô hình ANN và DA cũng xấp xỉ bằng nhau khi phân tích đường cong sử dụng tập dữ liệu xây dựng mô hình, giá trị này là 87,2% và 87,1%. Các mô hình này cũng có giá trị cao xấp xỉ nhau khi sử dụng tập dữ liệu kiểm tra mô hình để đánh giá giá trị AUC. Các giá trị này của mô hình BN, LR, ANN, DA lần lượt là 91%, 90,5%, 87,7% và 87,6%. Chỉ có mô hình SVM cho kết quả thẩm định thấp hơn cả, giá trị AUC cho tập dữ liệu xây dựng và kiểm tra mô hình đạt 80,8% và 81,4%. Mặc dù các giá trị AUC của các mô hình đều lớn hơn 80%. Tuy nhiên, giá trị này có thể cải thiện khi tập dữ liệu sử dụng cho mô hình (xây dựng mô hình và kiểm nghiệm mô hình) có số lượng nhiều hơn. Tức dữ liệu về tai biến trượt lở và lũ bùn đá nên thu thập và bổ sung trong các năm tiếp theo để kết quả mô hình dự báo tốt hơn. Chính vì vậy, để áp dụng hiệu quả hơn hệ phương pháp này cho các khu vực miền núi khác, thu thập bộ dữ liệu có tính đa thời gian để tăng hiệu quả dự báo cho các bản đồ nguy cơ tai biến trượt lở và lũ bùn đá là cần thiết. Ngoài ra, việc kết hợp các mô hình học máy trong thành lập bản đồ nguy cơ tai biến cũng là hướng nghiên cứu tiếp theo của tập thể tác giả để khắc phục nhược điểm của từng mô hình, từ đó tăng

độ chính xác và khả năng dự báo của các bản đồ kết quả.

Các bản đồ phân vùng nguy cơ trượt lở và lũ bùn đá được tạo ra từ các mô hình kết hợp BN-FlowR, LR-FlowR, ANN-FlowR, DA-FlowR và SVM-FlowR (Hình 9) cũng được so sánh bằng phần trăm điểm trượt lở có mặt trong từng lớp nguy cơ rất thấp, thấp, trung bình, cao, rất cao với các bản đồ phân vùng nguy cơ trượt lở được tạo ra từ 5 mô hình thống kê học máy BN, LR, ANN, DA và SVM. Kết quả cho thấy phần trăm điểm trượt trong các mô hình kết hợp đã tăng lên từ hơn 2% đến 11% so với giá trị này trong các mô hình chưa kết hợp.

Kết quả của nghiên cứu chỉ ra rằng các phương pháp thống kê và học máy hoàn toàn có thể áp dụng để thành lập bản đồ nguy cơ trượt lở ở tỷ lệ lớn 1:10,000 cho các khu vực hạn chế về dữ liệu đầu vào. Mô hình kết hợp giữa các phương pháp này với mô hình FlowR để thành lập bản đồ nguy cơ tai biến trượt lở và lũ bùn đá phù hợp để áp dụng ở tỷ lệ 1:10,000. Kết quả nghiên cứu sẽ góp phần giảm thiểu tai biến trượt lở và lũ bùn đá, hỗ trợ việc quy hoạch lãnh thổ trong khu vực nghiên cứu. Các phương pháp đã áp dụng trong nghiên cứu này có thể áp dụng cho các khu vực khác để thành lập bản đồ nguy cơ tai biến trượt lở và lũ bùn đá ở tỷ lệ tương tự.

Lời cảm ơn

Bài báo này là một trong những kết quả nghiên cứu của đề tài mã số TNMT.2021.02.09 do Bộ Tài nguyên và Môi trường đặt hàng Viện Khoa học Địa chất và Khoáng sản (VIGMR) chủ trì thực hiện.

Tài liệu tham khảo

- [1] E. Leroi, Landslide Hazard-risk Maps at Different Scales: Objectives, Tools and Developments, in: Proc VII Int Symp Landslides, Trondheim, June 1996, pp. 35-52, <https://doi.org/10.12691/jgg-5-4-2>.
- [2] P. Aleotti, R. Chowdhury, Landslide Hazard Assessment: Summary Review and New Perspectives, Bull Eng Geol Environ, Vol. 58,

- No. 1, 1999, pp. 21-44,
<https://doi.org/10.1007/s100640050066>.
- [3] J. Corominas, C. V. Westen, P. Frattini, L. Cascini, J. P. Malet, S. Fotopoulou, F. Catani, M. V. D. Eeckhaut, O. Mavrouli, F. Agliardi, K. Pitolakis, M. G. Winter, M. Pastor, S. Ferlisi, V. Tofani, J. Herv, J. T. Smith, Recommendations for the Quantitative Analysis of Landslide Risk, *Bull. Eng. Geol, Environ*, Vol. 73, 2014, pp. 209-263, <https://doi.org/10.1007/s10064-013-0538-8>.
- [4] C. J. V. Westen, M. T. J. Terlien, An Approach Towards Deterministic Landslide Hazard Analysis in GIS, A Case Study from Manizales (Colombia), *Earth Surf Process Landforms*, Vol. 21, No. 9, 1996, pp. 853-868, [https://doi.org/10.1002/\(SICI\)1096-9837\(199609\)21:9<853::AID-ESP676>3.0.CO;2-C](https://doi.org/10.1002/(SICI)1096-9837(199609)21:9<853::AID-ESP676>3.0.CO;2-C).
- [5] J. N. Goetz, A. Brenning, H. Petschko, P. Leopold, Evaluating Machine Learning and Statistical Prediction Techniques for Landslide Susceptibility Modeling, *Computers and Geosciences*, Vol. 81, 2015, pp. 1-11, <https://doi.org/10.1016/j.cageo.2015.04.007>.
- [6] J. L. Zezere, S. Pereira, R. Melo, S. C. Oliveira, R. A. C. Garcia, Mapping Landslide Susceptibility Using Data-driven Methods, *Science of Total Environment*, Vol. 589, 2017, pp. 250-267, <https://doi.org/10.1016/j.scitotenv.2017.02.188>.
- [7] Y. Huang, L. Zhao, Review on Landslide Susceptibility Mapping Using Support Vector Machines, *Catena*, Vol. 165, 2018, pp. 520-529, <https://doi.org/10.1016/j.catena.2018.03.003>.
- [8] S. Lee, K. Min, Statistical Analysis of Landslide Susceptibility at Yongin, Korean, *Environmental Geology*, Vol. 40, 2001, pp. 1095-1113, <https://doi.org/10.1007/s002540100310>.
- [9] S. Lee, T. Sambath, Landslide Susceptibility Mapping in the Damrei Romel Area, Cambodia Using Frequency Ratio and Logistic Regression Models, *Environ Geol*, Vol. 50, 2006, pp. 847-855, <https://doi.org/10.1007/s00254-006-0256-7>.
- [10] S. Lee, B. Pradhan, Landslide Hazard Mapping at Selangor, Malaysia Using 42 Frequency Ratio and Logistic Regression Models, *Landslides*, Vol. 4, 2007, pp. 33-41, <https://doi.org/10.1007/s10346-006-0047-y>.
- [11] A. Yalcin, S. Reis, A. C. Aydinoglu, T. Yomralioglu, 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*, Vol. 85, No. 3, 2011, pp. 274-287, <http://dx.doi.org/10.1016/j.catena.2011.01.014>.
- [12] M. Mohammady, H. R. Pourghasemi, B. Pradhan, Landslide Susceptibility Mapping At Golestan Province, Iran: A Comparison Between Frequency Ratio, Dempster-Shafer, and Weights of Evidence Models, *J. Asian Earth Sci*, Vol. 61, 2012, pp. 221-236, <http://dx.doi.org/10.1016/j.jseaes.2012.10.005>.
- [13] A. D. Regmi, K. C. Devkota, K. Yoshida, B. Pradhan, H. R. Pourghasemi, T. Kumamoto, A. Akgun, 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*, Vol. 7, No. 2, 2014, pp. 725-742, <http://dx.doi.org/10.1007/s12517-012-0807-z>.
- [14] R. K. Dahal, S. Hasegawa, A. Nonomura, M. Yamanaka, S. Dhakal, P. Paudyal, Predictive Modeling of Rainfall-Induced Landslide Hazard In The Lesser Himalaya of Nepal Based on Weights of Evidence, *Geomorphology*, Vol. 102, No. 3-4, 2008, pp. 496-510, <http://dx.doi.org/10.1016/j.geomorph.2008.05.041>.
- [15] R. K. Dahal, S. Hasegawa, A. Nonomura, M. Yamanaka, T. Masuda, K. Nishino, GIS-based Weights of Evidence Modeling of Rainfall-Induced Landslides in Small Catchments For Landslide Susceptibility Mapping, *Environ Geol*, Vol. 54, No. 2, 2008, pp. 311-324, <http://dx.doi.org/10.1007/s00254-007-0818-3>.
- [16] N. R. Regmi, J. R. Giardino, J. D. Vitek, Modeling Susceptibility to Landslides Using The Weight of Evidence Approach: Westen Colorado, USA, *Geomorphology*, Vol. 115, No. 1-2, 2010, pp. 172-187, <http://dx.doi.org/10.1016/j.geomorph.2009.10.002>.
- [17] H. J. Oh, S. Lee, Assessment of Ground Subsidence Using GIS and the Weights of Evidence Model, *Eng Geol*, Vol. 115, No. 1-2, 2010, pp. 36-48, <http://dx.doi.org/10.1016/j.enggeo.2010.06.015>.
- [18] C. Xu, X. Xu, F. C. Dai, J. Z. Xiao, X. B. Tan, R. M. Yuan, Landslide Hazard Mapping Using GIS and Weight of Evidence Model in Qingshui River Watershed of 2008 Wenchuan Earthquake Struck Region, *J. Earth Sci*, Vol. 23, No. 1, 2012, pp. 97-120, <http://dx.doi.org/10.1007/s12583-012-0236-7>.
- [19] H. R. Pourghasemi, B. Pradhan, C. Gokceoglu, M. Mohammadi, H. R. Moradi, Application of Weights of Evidence and Certainty Factor Models and Their Comparison In Landslide Susceptibility

- Mapping At Haraz Watershed, Iran, Arab J. Geosci, Vol. 6, No. 7, 2013, pp. 2351-2365, <http://dx.doi.org/10.1007/s12517-012-0532-7>.
- [20] E. R. Sujatha, P. Kumaravel, G. V. Rajamanickam, Assessing Landslide Susceptibility Using Bayesian Probability-Based Weight of Evidence Model, Bull Eng Geol Environ, Vol. 73, No. 1, 2014, pp. 147-161, <http://dx.doi.org/10.1007/s10064-013-0537-9>.
- [21] Q. Q. Wang, W. P. Li, M. L. Xing, Y. L. Wu, Y. B. Pei, D. D. Yang, H. Y. Bai, Landslide Susceptibility Mapping At Gongliu County, China Using Artificial Neural Network and Weight of Evidence Models, Geosci J, Vol. 20, No. 5, 2016, pp. 705-718, <http://dx.doi.org/10.1007/s12303-016-0003-3>.
- [22] M. D. Hien, K. L. Yin, Z. Z. Guo, 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, pp. 2449-2485, <https://doi.org/10.1080/19475705.2020.1846086>.
- [23] T. Chen, R. Q. Niu, X. P. Jia, A Comparison of Information Value and Logistic Regression Models In Landslide Susceptibility Mapping By Using GIS, Environ Earth Sci, Vol. 75, No. 10, 2016, pp. 867, <http://dx.doi.org/10.1007/s12665-016-5317-y>.
- [24] F. Mengistu, K. V. Suryabhadgavan, T. K. Raghuvanshi, E. Lewi, 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, Vol. 3, No. 1, 2019, pp. 1-14, <https://doi.org/10.21523/gcjl.19030101>.
- [25] G. F. Zhang, Y. X. Cai, Z. Zheng, J. W. Zhen, Y. L. Liu, K. Y. Huang, Integration of the Statistical Index Method and the Analytic Hierarchy Process Technique for the Assessment of Landslide Susceptibility in Huizhou, China, Catena, Vol. 142, 2016, pp. 233-244, <http://dx.doi.org/10.1016/j.catena.2016.03.028>.
- [26] Z. Wu, Y. Wu, Y. Yang, F. Chen, N. Zhang, Y. Ke, W. Li, A Comparative Study on the Landslide Susceptibility Mapping Using Logistic Regression And Statistical Index Models, Arab J. Geosci, Vol. 10, No. 8, 2017, pp. 1-17, <http://dx.doi.org/10.1007/s12517-017-2961-9>.
- [27] H. J. Oh, S. Lee, G. M. Soedradjat, Quantitative Landslide Susceptibility Mapping at Pemalang Area, Indonesia, Environmental Earth Sciences, Vol. 60, No. 6, 2010, pp. 1317-1328, <https://doi.org/10.1007/s12665-009-0272-5>.
- [28] B. Pradhan, 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, Vol. 18, No. 3, 2011, pp. 471-493, <http://dx.doi.org/10.1007/s10651-010-0147-7>.
- [29] B. Pradhan, A Comparative Study on the Predictive Ability of the Decision Tree, Support Vector Machine And Neuro-Fuzzy Models in Landslide Susceptibility Mapping Using GIS, Comput Geosci, Vol. 51, 2013, pp. 350-365, <http://dx.doi.org/10.1016/J.Cageo.2012.08.023>.
- [30] H. R. Pourghasemi, A. G. Jirandeh, B. Pradhan, C. Xu, C. Gokceoglu, Landslide Susceptibility Mapping Using Support Vector Machine and GIS at the Golestan Province, J Earth Syst Sci, Vol. 122, No. 2, 2013, pp. 349-369, <https://doi.org/10.1007/S12040-013-0282-2>.
- [31] H. Y. Hong, B. Pradhan, C. Xu, T. B. Dieu, Spatial Prediction of Landslide Hazard at the Yihuang Area (China) Using Two-Class Kernel Logistic Regression, Alternating Decision Tree and Support Vector Machines, Catena, Vol. 133, 2015, pp. 266-281, <http://dx.doi.org/10.1016/j.catena.2015.05.019>.
- [32] B. T. Dieu, H. Shahabi, A. Shirzadi, K. Chapi, B. Pradhan, W. Chen, K. Khosravi, M. Panahi, B. B. Ahmad, L. Saro, Land Subsidence Susceptibility Mapping in South Korea Using Machine Learning Algorithms, Sensors, Vol. 18, No. 8, 2018, pp. 1-20, <https://doi.org/10.3390/S1808246>.
- [33] L. Ermini, F. Catani, N. Casagli, Artificial Neural Networks Applied to Landslide Susceptibility Assessment, Geomorphology, Vol. 66, 2005, pp. 327-343, <https://doi.org/10.1016/j.geomorph.2004.09.025>.
- [34] H. A. Nefeslioglu, C. Gokceoglu, H. Sonmez, 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, Vol. 97, No. 3-4, 2008, pp. 171-91, <https://doi.org/10.1016/j.enggeo.2008.01.004>.
- [35] B. Pradhan, S. Lee, Landslide Risk Analysis Using Artificial Neural Network Model Focussing on Different Training Sites, Int J Physsci, Vol. 4, 2009, pp. 001-015, <https://doi.org/10.5897/ijps.9000343>.

- [36] B. Pradhan, S. Lee, Delineation Of Landslide Hazard Areas on Penang Island, Malaysia, By Using Frequency Ratio, Logistic Regression, and Artificial Neural Network Models, *Environmental Earth Sciences*, Vol. 60, No. 5, 2010, pp. 1037-1054, <https://doi.org/10.1007/s12665-009-0245-8>.
- [37] B. Pradhan, S. Lee, Landslide Susceptibility Assessment and Factor Effect Analysis: Backpropagation Artificial Neural Networks and Their Comparison with Frequency Ratio and Bivariate Logistic Regression Modelling, *Environ Modell Softw*, Vol. 25, 2010, pp. 747-59, <https://doi.org/10.1016/j.envsoft.2009.10.016>.
- [38] I. Yilmaz, Comparison of Landslide Susceptibility Mapping Methodologies For Koyulhisar, Turkey: Conditional Probability, Logistic Regression, Artificial Neural Networks, and support vector machine, *Environmental Earth Sciences*, Vol. 61, No. 4, 2010, pp. 821-836, <https://doi.org/10.1007/s12665-009-0394-9>.
- [39] P. T. Binh, B. Pradhan, T. D. Bui, I. Prakash, M. B. Dholakia, A Comparative Study of Different Machine Learning Methods for Landslide Susceptibility Assessment: A Case Study of Uttarakhand Area (India), *Environmental Modelling & Software*, Vol. 84, 2016, pp. 240-250, <https://doi.org/10.1016/j.envsoft.2016.07.005>.
- [40] J. Corominas, The Angle of Reach as Amobility Index For Small And Large Landslides, *Canadian Geotechnical Journal*, Vol. 33, No. 2, 1996, pp. 260–271, <https://doi.org/10.1139/t96-005>.
- [41] F. Legros, Themobility of Long-Runout Landslides, *Engineering Geology*, Vol. 63, 2002, pp. 301-331, [https://doi.org/10.1016/S0013-7952\(01\)00090-4](https://doi.org/10.1016/S0013-7952(01)00090-4).
- [42] R. M. Iverson, S. P. Schilling, J. W. Vallance, Objective Delineation of Lahar-Inundation Hazard Zones, *Geological Society of America Bulletin*, Vol. 110, No. 8, 1998, pp. 972-984, [https://doi.org/10.1130/0016-7606\(1998\)110<0972:ODOLIH>2.3.CO;2](https://doi.org/10.1130/0016-7606(1998)110<0972:ODOLIH>2.3.CO;2).
- [43] G. B. Crosta, S. Cucchiario, P. Frattini, Validation of Semi-Empirical Relationships for the Definition of Debris-flow Behavior In Granular Materials, in: Rickenmann, D., Chen, C. (Eds.), *3rd Int. Conf. on Debris-Flow Hazards Mitigation*, Millpress, Davos, 2003, pp. 821-831.
- [44] R. M. Iverson, The Physics of Debris Fows, *Review of Geophysics*, Vol. 35, No. 3, 1997, pp. 245-296, <https://doi.org/10.1029/97RG00426>.
- [45] M. Hürlimann, D. Rickenmann, V. Medina, A Bateman, Evaluation of Approaches to Calculate Debris-Flow Parameters for Hazard Assessment, *Engineering Geology*, Vol. 102, 2008, pp. 152-163, <https://doi.org/10.1016/j.enggeo.2008.03.012>.
- [46] D. Rickenmann, D. Laigle, B. Mc Ardell, J. Hübl, Comparison of 2D Debris-flow Simulation Models with field Events, *Computational Geosciences*, Vol. 10, No. 2, 2006, pp. 24-264, <https://doi.org/10.1007/s10596-005-9021-3>.
- [47] K. Schraml, B. Thomschitz, B. W. Mc Ardell, C. Graf, R. Kaitna, Modeling Debris-flow Runout Patterns on Two Alpine Fans With Different Dynamic Simulation Models, *Nat Hazards Earth Syst Sci*, Vol. 15, 2015, pp. 1483-1492, <https://doi.org/10.5194/nhess-15-1483-2015>.
- [48] J. Blahut, P. Horton, S. Sterlacchini, M. Jaboyedoff, Debris Fow Hazard Modelling on Medium Scale: Valtellina di Tirano, Italy, *Nat. Hazards Earth Syst Sci*, Vol. 10, 2010, pp. 2379-2390, <https://doi.org/10.5194/nhess-10-2379-2010>.
- [49] P. Horton, M. Jaboyedoff, M. Zimmermann, B. Mazotti, C. Longchamp, Flow-R, a Model for Debris Flow Susceptibility Mapping at A Regional Scale – Some Case Studies, *5th Int Conf on Debris-Flow Hazards Mitigation*, Padua, Italy, *Italian Journal of Engineering Geology and Environment*, 2011, pp. 875-884.
- [50] P. Horton, M. Jaboyedoff, B. Rudaz, M. Zimmermann, Flow-R, a Model for Susceptibility Mapping of Debris Flows and Other Gravitational Hazards At A Regional Scale, *Nat Hazards Earth Syst Sci*, Vol. 13, No. 4, 2013, pp. 869-885, <http://dx.doi.org/10.5194/nhess-13-869-2013>.
- [51] M. Jaboyedoff, Ch. Choffet, M. H. Derron, P. Horton, A. Loye, C. Longchamp, B. Mazotti, C. Michoud, A. Pedrazzini, Preliminary Slope Mass Movements Susceptibility Mapping Using DEM and LiDAR DEM, in: *Terrigenous Mass Movements: Detection, Modelling, Early Warning and Mitigation Using Geoinformation Technology*, Edited by: Pradhan, B. And Buchroithner, M., Springer-Verlag, Berlin Heidelberg, Germany, 2012, pp. 109-170, https://doi.org/10.1007/978-3-642-25495-6_5.
- [52] B. Q. Luna, J. Blahut, M. Kappes, S. O. Akbas, J. P. Malet, A. Remaitre, T. V. Asch, M. Jaboyedoff, Chapter 5: Methods for Debris Flow Hazard and Risk Assessment, in Van Asch T Ed, *Mountain Risks: From Prediction To Management and Governance, Advances in Natural and Technological Hazards Research*, Vol. 34, 2014, <http://dx.doi.org/10.1007/978-94-007-6769-05>.
- [53] B. Q. Luna, J. Blahut, C. Camera, C. Van Westen, T. Apuani, V. Jetten, S. Sterlacchini, Physically

- Based Dynamic Run-Out Modelling for Quantitative Debris Flow Risk Assessment: A Case Study in Tresenda, Northern Italy, *Environ Earth Sci*, Vol. 72, 2014, pp. 645-661, <http://dx.doi.org/10.1007/s12665-013-2986-7>.
- [54] B. Q. Luna, J. Blahut, T. V. Asch, C. V. Westen, M. Kappes, ASCHFLOW: A Dynamic Landslide Run-Out Model for Medium Scale Hazard Analysis, *Geoenviron Disasters*, Vol. 3, 2016, pp. 1-17, <http://dx.doi.org/10.1186/s40677-016-0064-7>.
- [55] S. H. Kang, S. R. Lee, N. Nikhil, J. Y. V. Park, D. H. Lee, Development of An Initiation Criterion for Debris Flows Based on Local Topographic Properties and Applicability Assessment at A Regional Scale, *Eng Geol*, Vol. 230, 2017, pp. 64-76, <http://dx.doi.org/10.1016/j.enggeo.2017.09.017>.
- [56] S. Kang, S. R. Lee, Debris Flow Susceptibility Assessment Based on An Empirical Approach in The Central Region of South Korea, *Geomorphology*, Vol. 308, 2018, pp. 1-12, <https://doi.org/10.1016/j.geomorph.2018.01.025>.
- [57] N. T. Long, F. D. Smedt, Application of an Analytical Hierarchical Process Approach for Landslide Susceptibility Mapping in A Luoi District, Thua Thien Hue Province Vietnam, *Environ Earth Sci*, Vol. 66, 2012, pp. 1739-1752, <http://dx.doi.org/10.1007/s12665-011-1397-x>.
- [58] L. Q. Hung, N. T. H. Van, M. D. Duc, L. T. C. Ha, P. V. Son, N. H. Khanh, L. T. Binh, 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*, Vol. 13, 2016, pp. 1285-1301, <https://doi.org/10.1007/s10346-015-0657-3>.
- [59] N. T. Long, F. D. Smedt, Analysis and Mapping of Rainfall-Induced Landslide Susceptibility in A Luoi District, Thua Thien Hue Province, Vietnam, *Water-MDPI*, Vol. 11, 2019, <https://doi.org/10.3390/w11010051>.
- [60] B. T. Dieu, B. Pradhan, O. Lofman, I. Revhaug, O. B. Dick, Spatial Prediction of Landslide Hazards in Hoabinh Province (Vietnam): A Comparative Assessment of the Efficacy of Evidential Belief Functions And Fuzzy Logic Models, *Catena*, Vol. 96, 2012, pp. 28-40, <https://doi.org/10.1016/j.catena.2012.04.001>.
- [61] B. T. Dieu, B. Pradhan, O. Lofman, I. Revhaug, O. B. Dick, Landslide Susceptibility Assessment in The Hoabinh Province of Vietnam: A Comparison of the Levenberg-Marquardt and Bayesian Regularized Neural Networks, *Geomorphology*, Vol. 171-172, 2012, pp. 12-29, <https://doi.org/10.1016/j.geomorph.2012.04.023>.
- [62] B. T. Dieu, P. T. Binh, Q. P. Nguyen, N. D. Hoang, 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*, 2016, pp. 1-21, <https://doi.org/10.1080/17538947.2016.1169561>.
- [63] P. T. Binh, B. Pradhan, B. T. Dieu, Spatial Prediction of Landslides Using A Hybrid Machine Learning Approach Based on Random Subspace and Classification and Regression Trees, *Geomorphology*, Vol. 303, 2018, pp. 256-270, <https://doi.org/10.1016/j.geomorph.2017.12.008>.
- [64] P. T. Binh, N. T. Thoi, C. C. Qi, T. V. Phong, J. Dou, H. S. Lanh, L. V. Hiep, I. Prakash., Coupling RBF Neural Network with Ensemble Learning Techniques for Landslide Susceptibility Mapping, *Catena*, Vol. 195, <https://doi.org/10.1016/j.catena.2020.104805>.
- [65] Vietnam Meteorological and Hydrological Administration, <http://vmha.gov.vn/public/kttv-voi-san-xuat-va-doi-song-106/nhieu-hiem-nguy-thien-tai-rinh-rap-nguoi-dan-vung-cao-lao-cai-12248.html> (accessed on: May 9th, 2022).
- [66] Laocai Radio and Television, <http://laocaitv.vn/luquet/sat-lo-dat-khien-2-nguoi-tu-vong-o-xa-phing-ngan-huyen-bat-xat> (accessed on: May 9th 2022).
- [67] N. X. Khien et al., Assess the Current Status of Geological Hazards in Ha Giang, Cao Bang, Tuyen Quang and Bac Kan Provinces, Identify Causes, Forecast and Propose Measures to Prevent and Minimize Consequences, Vietnam Institute of Geosciences and Mineral Resources, 2011 (in Vietnamese).
- [68] O. F. Althuwaynee, B. Pradhan, H. J. Park, J. H. Lee, A Novel Ensemble Bivariate Statistical Evidential Belief Function with Knowledge-Based Analytical Hierarchy Process and Multivariate Statistical Logistic Regression for Landslide Susceptibility Mapping, *Catena*, Vol. 114, 2014, pp. 21-36, <http://dx.doi.org/10.1016/j.catena.2013.10.011>.
- [69] E. Yesilnacar, T. Topal, Landslide Susceptibility Mapping: A Comparison of Logistic Regression and Neural Networks Methods In A Medium Scale Study, Hendek Region (Turkey), *Eng Geol*, Vol. 79, No. 3-4, 2005, pp. 251-266, <http://dx.doi.org/10.1016/j.enggeo.2005.02.002>.
- [70] H. Y. Hong, H. R. Pourghasemi, Z. S. Pourtaghi, Landslide Susceptibility Assessment in Lianhua County (China): A Comparison Between A

- Random Forest Data Mining Technique and Bivariate and Multivariate Statistical Models, *Geomorphology*, Vol. 259, 2016, pp. 105-118, <http://dx.doi.org/10.1016/j.geomorph.2016.02.012>.
- [71] A. M. S. Pradhan, Y. T. Kim, Relative Effect Method of Landslide Susceptibility Zonation in Weathered Granite Soil: A Case Study in Deokjeok-Ri Creek, *Nat Hazards*, Vol. 72, 2014, pp. 1189-1217, <http://dx.doi.org/10.1007/s11069-014-1065-z>.
- [72] A. M. S. Pradhan, Y. T. Kim, Spatial Data Analysis And Application of Evidential Belief Functions to Shallow Landslide Susceptibility Mapping At Mt, Umyeon, Seoul, Korea, *Bull Eng Geol Environ*, Vol. 76, No. 4, 2017, pp. 1263-1279, <http://dx.doi.org/10.1007/s10064-016-0919-x>.
- [73] K. Beven, M. Kirkby, 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*, Vol. 24, 1979, pp. 43-69.
- [74] I. Yilmaz, Landslide Susceptibility Mapping Using Frequency Ratio, Logistic Regression, Artificial Neural Networks and Their Comparison: A Case Study from Kat Landslides (Tokat-Turkey), *Comput Geosci*, Vol. 35, 2009, pp. 1125-1138, <https://doi.org/10.1016/j.cageo.2008.08.007>.
- [75] A. M. S. Pradhan, H. S. Kang, S. Lee, Y. T. Kim, Spatial Model Integration For Shallow Landslide Susceptibility and Its Runout Using A GIS-Based Approach in Yongin, *Geocarto Int*, 2016, <https://doi.org/10.1080/10106049.2016.1155658>.
- [76] I. D. Moore, G. J. Burch, Physical Basis of the Length-Slope Factor in the Universal Soil Loss Equation, *Soil Sci Soc Am J*, Vol. 50, No. 5, 1986, pp. 1294-1298, <https://doi.org/10.2136/sssaj1986.03615995005000050042x>.
- [77] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, 1988.
- [78] N. Friedman, D. Geiger, M. Goldszmidt, Bayesian Network Classifiers, *Machine Learning*, Vol. 29, 1997, pp. 131-163, <https://doi.org/10.1023/a:1007465528199>.
- [79] B. G. Marcot, J. D. Steventon, G. D. Sutheland, R. K. Mccann, Guidelines for Developing and Updating Bayesian Belief Networks Applied to Ecological Modeling and Conservation, *Can J for Res*, Vol. 36, No. 12, 2006, pp. 3063-3074, <https://doi.org/10.1139/x06-135>.
- [80] Y. Song, J. Gong, S. Gao, D. Wang, T. Cui, Y. Li, B. Wei, Susceptibility Assessment of Earthquake-Induced Landslides Using Bayesian Network: A Case Study in Beichuan China, *Comput. Geosci*, Vol. 42, 2012, pp. 189-199, <https://doi.org/10.1016/j.cageo.2011.09.011>.
- [81] B. Pradhan, 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*, Vol. 18, No. 3, 2011, pp. 471-493, <http://dx.doi.org/10.1007/s10651-010-0147-7>.
- [82] F. C. Dai, C. F. Lee, J. Li, Z. W. Xu, Assessment of Landslide Susceptibility on the Natural Terrain of Lantau Island, Hong Kong, *Environ Geol*, Vol. 40, 2001, pp. 381-391, <https://doi.org/10.1007/S002540000163>.
- [83] C. T. Lee, C. C. Huang, J. F. Lee, K. L. Pan, M. N. Lin, J. J. Dong, Statistical Approach to Storm Event-Induced Landslides Susceptibility, *Nat Hazard Earth Syst Sci*, Vol. 8, No. 4, 2008, pp. 941-960, <https://doi.org/10.5194/nhess-8-941-2008>.
- [84] W. R. Klecka, Discriminant Analysis, Sage University Paper Series on Quantitative Applications in the Social Sciences, Sage Publications, Beverly Hills and London, 1980, pp. 07-19.
- [85] J. C. Davis, Statistics and Data Analysis in Geology, (3rd Ed), John Wiley And Sons, New York, 2002.
- [86] C. Liu, W. Y. Li, H. B. Wu, P. Lu, K. Sang, W. W. Sun, W. Chen, Y. Hong, R. X. Li, Susceptibility Evaluation and Mapping of China's Landslides Based on Multi-Source Data, *Nat. Hazards*, Vol. 69, 2013, pp. 1477-1495, <https://doi.org/10.1007/S11069-013-0759-Y>.
- [87] S. Lee, J. H. Ryu, M. J. Lee, J. S. Won, The Application of Artificial Neural Networks to Landslide Susceptibility Mapping At Janghung, Korea, *Mathematical Geology*, Vol. 38, 2006, <https://doi.org/10.1007/s11004-005-9012-x>.
- [88] A. H. Fath, F. Madanifar, M. Abbasi, Implementation of Multilayer Perceptron (MLP) And Radial Basis Function (RBF) Neural Networks to Predict Solution Gas-Oil Ratio of Crude Oil Systems, *Petroleum*, Vol. 6, 2020, pp. 80-91, <https://doi.org/10.1016/j.petlm.2018.12.002>.
- [89] V. N. Vapnik, The Nature of Statistical Learning Theory, 2nd, Springer, New York, 1999.
- [90] B. Schölkopf, A. Smola, Learning with Kernels: Support Vectormachines, Regularization, Optimization, and Beyond, MIT Press, Cambridge, 2002.

- [91] J. M. Moguerza, A. Munoz, Support Vector Machines With Applications, *Stat Sci*, Vol. 21, 2006, pp. 322-336, <https://doi.org/10.1214/088342306000000493>.
- [92] V. Cherkassky, F. Mulier, *Learning From Data: Concepts, Theory, and Methods*, Wiley, New York, 2007.
- [93] T. Kavzoglu, E. K. Sahin, I. Colkesen, Landslide Susceptibility Mapping Using GIS-Based Multi-Criteria Decision Analysis, Support Vector Machines, and Logistic Regression, *Landslides*, Vol. 11, No. 3, 2014, pp. 425-439, <https://doi.org/10.1007/s10346-013-0391-7>.
- [94] J. A. Suykens, J. Vandewalle, Least Squares Support Vector Machine Classifiers, *Neural Process Lett*, Vol. 9, No. 3, 1999, pp. 293-300, <https://doi.org/10.1023/a:1018628609742>.
- [95] B. Dixon, N. Candade, Multispectral Landuse Classification Using Neural Networks and Support Vector Machines: One or the Other, or Both?, *Int J Remote Sens*, Vol. 29, No. 4, 2008, pp. 1185-1206, <https://doi.org/10.1080/01431160701294661>.
- [96] P. Kayastha, M. R. Dhital, F. D. Smedt F, Application of the Analytical Hierarchy Process (AHP) for Landslide Susceptibility Mapping: A Case Study from the Tinau Watershed, West Nepal, *Comput Geosci*, Vol. 52, 2013, pp. 398-408, <http://dx.doi.org/10.1016/j.cageo.2012.11.003>.
- [97] C. J. F. Chung, A. G. Fabbri, Validation of Spatial Prediction Models for Landslide Hazard Mapping, *Natural Hazards*, Vol. 30, No. 3, 2003, pp. 451-472, <https://doi.org/10.1023/b:nhaz.0000007172.62651.2b>.
- [98] H. R. Pourghasemi, H. R. Moradi, S. M. Fatemi Aghda, C. Gokceoglu, B. Pradhan, GIS-based Landslide Susceptibility Mapping with Probabilistic Likelihood Ratio and Spatial Multi-Criteria Evaluation Models (North of Tehran, Iran), *Arab J Geosci*, Vol. 7, No. 5, 2014, pp. 1857-1878, <http://dx.doi.org/10.1007/s12517-012-0825-x>.
- [99] R. Talaei, Landslide Susceptibility Zonation Mapping Using Logistic Regression and Its Validation in Hashtchin Region, Northwest of Iran, *Journal Geological Society of India*, Vol. 84, 2014, pp. 68-86, <https://doi.org/10.1007/s12594-014-0111-5>.
- [100] J. W. Choi, H. J. Oh, H. J. Lee, C. W. Lee, S. Lee, Combining Landslide Susceptibility Maps Obtained from Frequency Ratio, Logistic Regression and Artificial Neural Network Models Using ASTER Images and GIS, *Engineering Geology*, Vol. 124, 2012, pp. 12-23, <https://doi.org/10.1016/j.enggeo.2011.09.011>.
- [101] J. J. Dong, Y. Y. Tung, C. C. Chen, J. J. Liao, Y. W. Pan, Discriminant Analysis of Theo Geomorphic Characteristics and Stability of Landslide Dams, *Geomorphology*, Vol. 110, 2009, pp. 162-171, <https://doi.org/10.1016/j.geomorph.2009.04.004>.
- [102] S. W. He, P. Pan, L. Dai, H. J. Wang, J. P. Liu, Application of Kernel-Based Fisher Discriminant Analysis to Map Landslide Susceptibility In The Qinggan River Delta, Three Gorges, China, *Geomorphology*, Vol. 171-172, 2012, pp. 30-41, <https://doi.org/10.1016/j.geomorph.2012.04.024>.