

Thành lập bản đồ nhạy cảm tai biến trượt lở tỷ lệ lớn bằng các mô hình thống kê và học máy

Đỗ Minh Hiền^{1*}, Nguyễn Văn Hoàng², Mai Lê Dũng¹, Phạm Đức Trọng¹,
Nguyễn Cao Cường¹, Bùi Trọng Tấn¹

¹Viện Khoa học Địa chất và Khoáng sản, 67 Chiến Thắng, Văn Quán, Hà Đông, Hà Nội

²Trường Đại học Warszawa, Ba Lan

* Tác giả liên hệ: hien_dm@yahoo.com

Ngày gửi bài:
01/02/2025

Ngày chấp nhận đăng:
24/4/2025

Từ khóa: Trượt lở, mô hình thống kê, mô hình học máy, Phước Thành.

Tóm tắt: Mục đích chính của bài báo này là thành lập bản đồ phân vùng nhạy cảm tai biến trượt lở khu vực xã Phước Thành huyện Phước Sơn tỉnh Quảng Nam ở tỷ lệ lớn bằng các phương pháp thống kê và học máy. 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, thảm 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 phần trăm diện tích (AUC) theo đặc tính đường cong nhận được (ROC). Kết quả chỉ ra rằng các mô hình 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 gây nên tại khu vực nghiên cứu và các khu vực miền núi khác.

1. Giới thiệu

Hiện nay trên thế giới, công tác đánh giá rủi ro tai biến trượt lở 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 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 (Leroi, 1996), 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 (Aleotti và nnk, 1999). 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 (Aleotti và nnk, 1999; Corominas và nnk, 2014). 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 chuyển động trượt lở không phức tạp và các đặc điểm địa chất, địa mạo ở đó là khá đồng nhất (Westen và nnk, 1996).

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 (Goetz và nnk, 2015; Zezere, và nnk, 2017; Huang và nnk, 2018). 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 (Lee và Min, 2001; Lee và Sambath, 2006; Lee và Pradhan, 2007; Yacin và nnk, 2011; Mohammady và nnk, 2012; Regmi và nnk, 2014), mô hình trọng số bằng chứng - WoE (Mohammady và nnk, 2012; Dahal và nnk, 2008a; Dahal và nnk, 2008b; Regmi và nnk, 2009; Oh và Lee, 2010; Xu và nnk, 2012; Pourghasemi và nnk, 2013; Sujatha và nnk, 2014; Wang và nnk, 2016; Hien và nnk, 2020), mô hình giá trị thông tin - IV (Chen và nnk, 2016; Mengistu và nnk, 2019), phương pháp chỉ số thống kê - SI (Regmi và nnk, 2014; Zhang và nnk, 2016; Wu và nnk, 2017).

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 để đánh giá tính nhạy cảm tai biến trượt lở ở các vùng khác nhau trên thế giới (Lee và Sambath, 2006; Lee và Pradhan, 2007; Hien và nnk, 2020; Wu và nnk, 2017; Oh và nnk, 2010; Pradhan, 2011).

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) (Pradhan, 2012; Pourghasemi và nnk, 2013; Hong và nnk, 2015; Diệu và nnk, 2018), mạng thần kinh nhân tạo (ANN) (Ermini, và nnk, 2005; Nefeslioglu và nnk, 2008; Pradhan và Lee, 2009; Pradhan và nnk, 2010; Yilmaz, 2010), cây quyết định (DT) (Pradhan, 2013; Hong và nnk, 2015), mạng Bayes (Bayesian Network) (Bình và nnk, 2016).

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 đã áp dụng các phương pháp chuyên gia, thống kê để thành lập bản đồ phân vùng nhạy cảm trượt lở (Hiền, và nnk, 2020; Long và Smedt, 2012; Hùng và nnk, 2016; Long và Smedt, 2019), 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 (Hiền, và nnk, 2020)

và phương pháp học máy (Bình và nnk, 2016; Diệu và nnk, 2012a; Diệu và nnk, 2012b; Diệu và nnk, 2016; Bình và nnk, 2018; Bình và nnk, 2020) cho các khu vực nghiên cứu khác nhau tại Việt Nam.

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, đã đạt được những kết quả đáng khích lệ. 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 trượt lở. 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 phác họa và khái quát cao. Công việc thành lập bản đồ phân vùng nguy cơ trượt lở ở tỷ lệ lớn (1:5.000) 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 tai biến trượt lở trong tương lai.

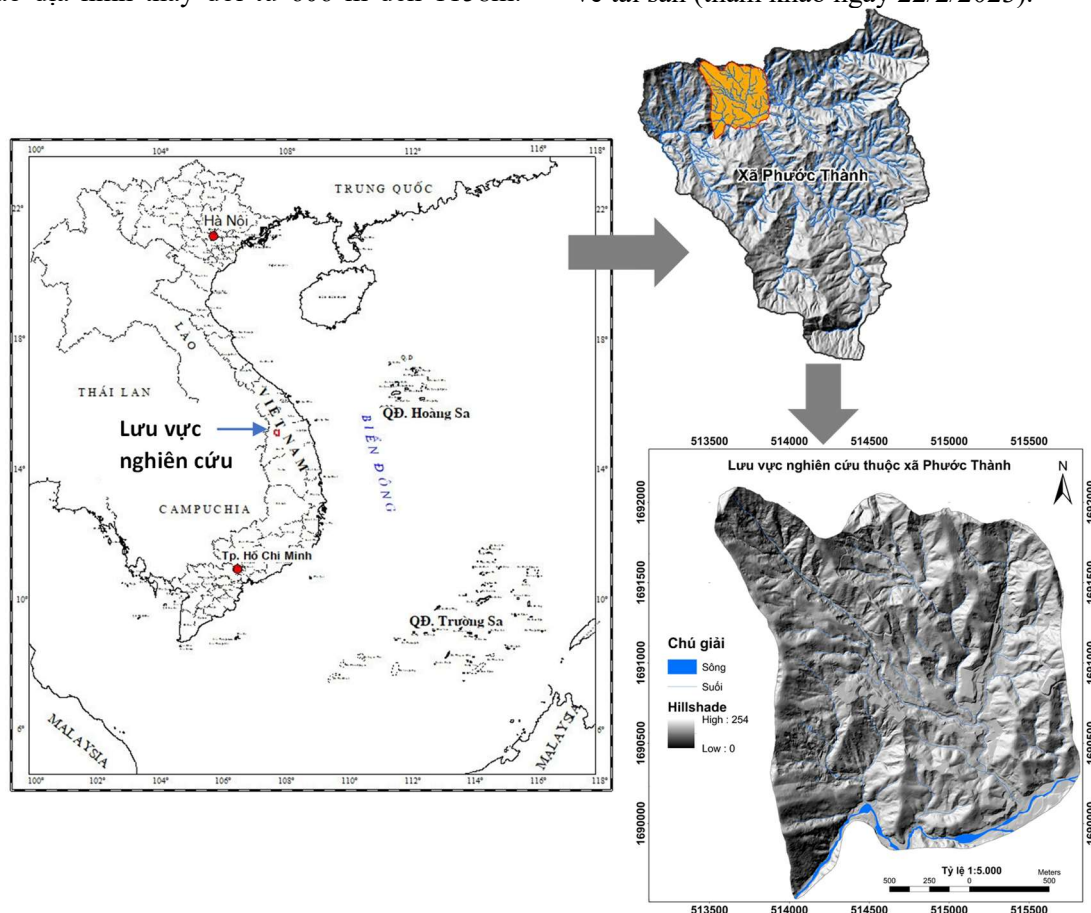
Chính vì vậy, mục đích chính của bài báo này là áp dụng các phương pháp thống kê và học máy với điều kiện dữ liệu đầu vào hạn chế (chủ yếu là dữ liệu được triểu xuất từ bản đồ địa hình tỷ lệ 1:5.000 gồm các bản đồ độ dốc, mặt cong địa hình, chỉ số độ ẩm địa hình (TWI), chỉ số năng lượng dòng chảy (SPI), chỉ số vận chuyển trầm tích (STI), bản đồ vơ 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 ở tỷ lệ 1:5.000.

2. Khu vực nghiên cứu

Phước Thành là xã vùng cao của huyện miền núi Phước Sơn tỉnh Quảng Nam. Xã Phước Thành cách thị trấn Khâm Đức 44 Km về phía đông nam, phía Đông giáp xã Trà Leng (Nam Trà My), phía Tây giáp xã Phước Lộc, Phía Nam Giáp xã Mường Hoong huyện ĐakGlei (Kon Tum), Phía Bắc giáp xã Phước Kim. Tổng diện tích đất tự nhiên khoảng 62 km², nhưng đất nông nghiệp chỉ chiếm 6 km² (trong đó đất canh tác lúa nước chỉ có 48 ha). Địa hình phức tạp phần lớn là đồi núi bao quanh, chia cách các thôn bản, dân cư thưa

thót phân bố không đồng đều (tham khảo ngày 22/2/2023). Lưu vực nghiên cứu (hình 1) thuộc xã Phước Thành có diện tích khoảng 11.5 km² giữa 107°52'35" đến 107°53'50" kinh độ Đông và 15°16'35" đến 15°18'3" vĩ độ Bắc, có địa hình phức tạp, vùng núi cao có độ chia cắt và độ dốc lớn. Độ cao địa hình thay đổi từ 606 m đến 1158m.

Theo báo cáo hàng năm thì xã Phước Thành là một trong các xã bị ảnh hưởng nặng nề bởi tai biến trượt lở và lũ bùn đá. Theo số liệu thống kê năm 2020, trượt lở kèm theo lũ bùn đá đã khiến 9 người ở xã Phước Lộc tử vong, 4 người đang mất tích, hàng trăm hộ dân ở 2 xã Phước Thành, Phước Lộc bị thiệt hại nặng nề về tài sản (tham khảo ngày 22/2/2023).



Hình 1. Vị trí nghiên cứu

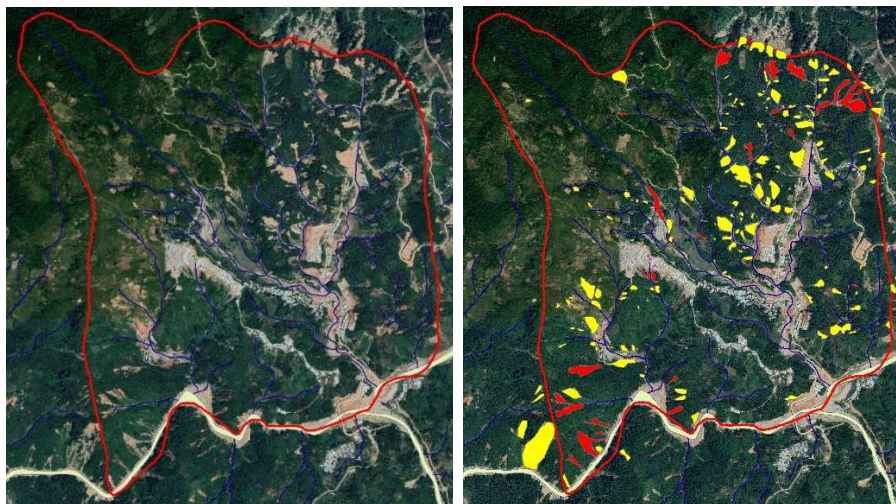
Về khí hậu, Phước Thành là một xã của tỉnh Quảng Nam nên cũng có đặc điểm chung của khí hậu nhiệt đới, chỉ có 2 mùa là mùa mưa và mùa khô. Nhiệt độ trung bình năm 25,6 °C, Mùa đông nhiệt độ vùng đồng bằng có thể xuống dưới 12 °C và nhiệt độ vùng núi thậm chí còn thấp hơn. Độ ẩm trung bình trong không khí đạt 84%. Lượng mưa trung bình 2000-2500mm. Mùa mưa thường kéo dài từ tháng 10 đến tháng 12, mùa khô kéo dài từ tháng 2 đến tháng 8, tháng 1 và tháng 9 là các tháng chuyển tiếp với đặc trưng là thời tiết hay nhiều loạn và khá nhiều mưa (tham khảo ngày 22/2/2023).

Về địa chất, lưu vực nghiên cứu với diện

tích chủ yếu phức hệ Bà Nà ở phía bắc bao gồm cả pha 1 (γT_2bn_1 : granit biotit, granit hai mica, granosyenit biotit có muscovit) và pha 2 (γT_2bn_2 : granit hai mica, granit biotit, granit alaskit) và hệ tầng Khâm Đức nằm ở phía đông và nam bao gồm tập 1 (NP- $\epsilon_1 kd_1$: plagiogneis amphibol-hai pyroxen, plagiogneis hai pyroxen, đá phiến thạch anh-plagioclas-hai pyroxen, gneis-biotit-diopsid, đá phiến plagioclas-amphibol) và tập 2 (NP- $\epsilon_1 kd_2$: plagiogneis biotit, plagiogneis biotit-granat, đá phiến thạch anh-biotit, đá phiến thạch anh-biotit-silimanit-granat-graphit, đá phiến hai mica-granat-disthen-graphit) (VIGMR). Trong lưu vực nghiên cứu các diện lộ đá gốc

không nhiều, chủ yếu lộ dọc theo các lưu vực sông suối. Đất đá ở khu vực này bị phong hóa mạnh, các lớp phong hóa dày, có độ gắn kết kém, lẫn nhiều mảnh vụn thạch anh và sét nên dễ dàng bị bão hòa nước. Cùng với đặc điểm bất lợi về địa hình như độ chia cắt lớn, sườn dốc, các tai biến như trượt lở, lũ bùn đá thường xuyên xảy ra khi mùa mưa đến.

3. Dữ liệu



Hình 2. Ảnh Google Earth (trái) và các điểm trượt lở (màu vàng), trượt dạng dòng (màu đỏ) thu thập từ thực địa và giải đoán ảnh viễn thám (phải)

Các điểm trượt đều xảy ra chủ yếu trong đới phong hóa hoàn toàn, phong hóa mạnh và trung bình. Tiếp theo, thông qua giải đoán ảnh vệ tinh phân giải cao Google Earth (ảnh chụp ngày 10/01/2021), số lượng điểm trượt lở và dòng bùn đá đã tăng lên đáng kể (tổng số điểm cả thực địa và giải đoán là 184 điểm, trong đó có 30 điểm dòng bùn đá) với diện tích điểm nhỏ nhất là 40 m² và lớn nhất là hơn 24.000 m² với tổng diện tích trượt lở và dòng bùn đá trong toàn lưu vực nghiên cứu khoảng 230.000 m². Sơ đồ vị trí khoanh định các điểm trượt lở và dòng bùn đá được trình bày trong hình 2.

3.2. Các dữ liệu khác

Mục đích của nghiên cứu là thành lập bản đồ phân vùng nguy cơ tai biến trượt lở ở tỷ lệ lớn 1:5.000 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

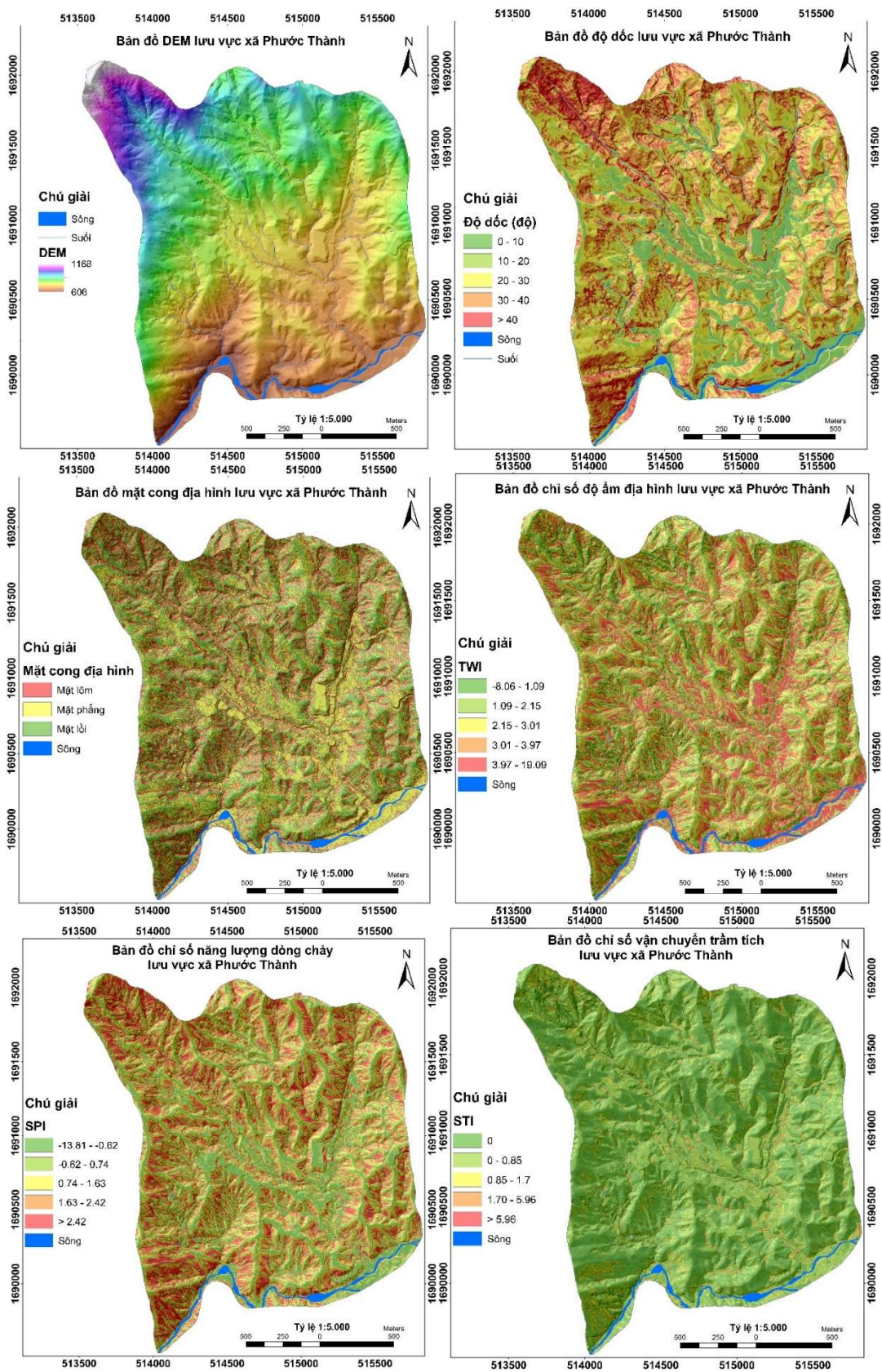
3.1. Dữ liệu trượt lở

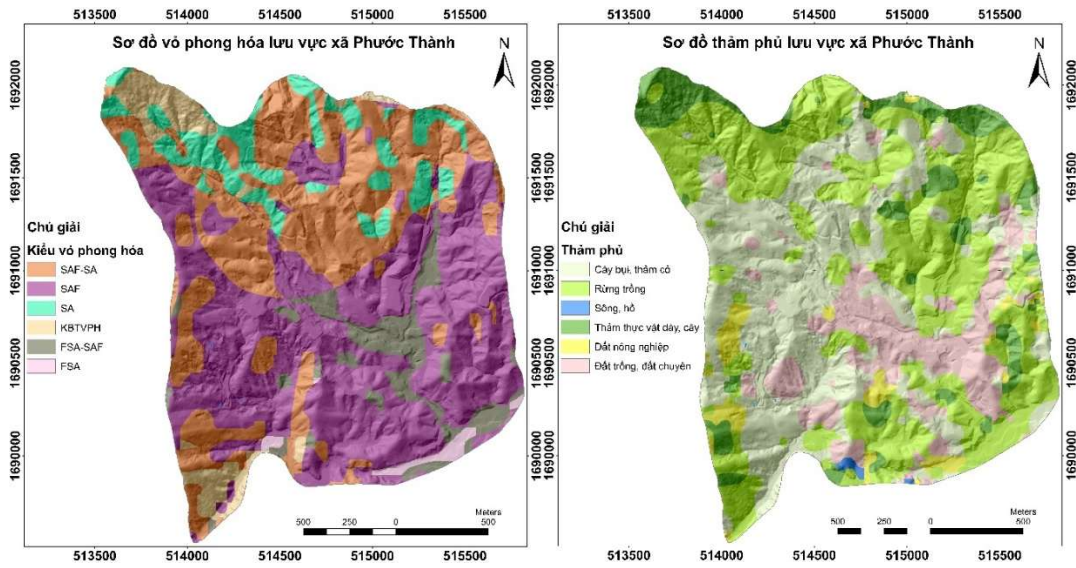
Khảo sát thực địa được tiến hành trong thời gian từ 15 tháng 4 đến 30 tháng 4 năm 2022. Với số lượng điểm trượt lở khảo sát thực địa theo tuyến giao thông chính thì thu thập được 21 điểm trượt lở, trong đó chủ yếu là kiểu trượt hỗn hợp với 11 điểm, trượt xoay có 3 điểm, trượt chảy có 7 điểm.

5 bản đồ tác nhân đầu được thành lập từ bản đồ địa hình tỷ lệ 1:5.000 (bản đồ địa hình ở tỷ lệ 1:5.000 vào năm 2021) 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.

3.3. Độ dốc

Trượt lở hình thành bởi sự tương tác giữa các tác nhân địa hình và địa chất. Một mô hình số độ cao (DEM) với kích thước pixel là 2x2m đã được tạo ra từ bản đồ địa hình của khu vực nghiên cứu tỷ lệ 1:5.000 với khoảng cách các đường bình độ là 5m 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 0⁰-72.5⁰ (Hình 3).





Hình 3. Các bản đồ tác nhân gây trượt lở lưu vực xã Phước Thành

3.4. 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 chuyên động xuống phía dưới (Yesilnacar và nnk, 2005; Hong và nnk, 2016) (Hình 3).

3.5. Chỉ số độ ẩm địa hình-TWI

TWI (Hình 3) là một trong những tác nhân quan trọng ảnh hưởng đến sự mất ổn định của sườn (Pradhan và Kim, 2014; Pradhan và Kim, 2017). 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 (Pradhan và Kim, 2017; Beven và nnk, 1979). 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 (Pradhan và Kim, 2014; Yilmaz, 2009). Tính toán chỉ số độ ẩm địa hình theo công thức (1) dưới đây được đề xuất bởi Pradhan và Kim (2014):

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

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.

3.6. Chỉ số năng lượng dòng chảy-SPI

SPI (Hình 3) 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 (Pradhan và Kim, 2014; Yilmaz, 2009). SPI dùng để ước lượng năng lượng xói mòn của nước. Giá trị xói mòn càng cao ở dọc các kênh, rãnh thoát nước (Pradhan và Kim, 2017). Chỉ số năng lượng dòng chảy được Pradhan và Kim (2014) đề xuất theo công thức dưới đây:

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

3.7. Chỉ số vận chuyển trầm tích-STI

STI (Hình 3) 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 (Hong và nnk, 2016; Pradhan và nnk, 2016). 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 (1986). Trong công thức (3), As là diện tích lưu vực và β là độ dốc.

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

3.8. Thảm phủ

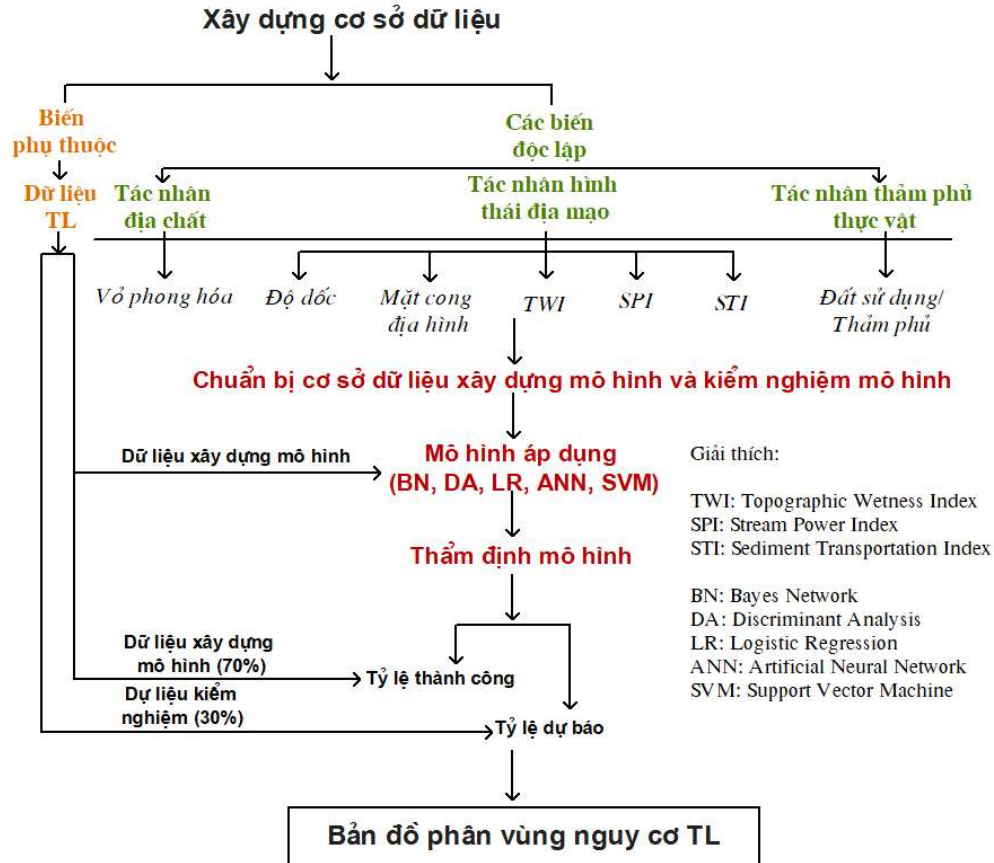
Bản đồ thảm phủ của khu vực nghiên cứu được thu thập từ các đề tài, dự án trước. Bản

đồ này được phân loại thành 5 lớp: rừng non hỗn hợp, đất trồng cây công nghiệp, rừng tự nhiên, đất trống và cây bụi, đất không che phủ.

3.9. 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 các kiểu vỏ phong hóa gồm: SAF-SA-vỏ phong hóa Sialferit-Sialit, SAF-vỏ phong hóa Sialferit, SA-vỏ phong hóa Sialit, FSA-SAF-vỏ phong hóa Ferosialit-Sialferit, FSA-vỏ phong hóa Ferosialit và KBTVPH (Hình 3).



Hình 4. Hệ phương pháp thực hiện trong nghiên cứu này

4. Phương pháp

Phân tích nhạy cảm trượt lở trong nghiên cứu này gồm 6 bước chính (Hình 4): 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; và v) Bản đồ phân vùng nguy cơ trượt lở của khu vực nghiên cứu ở tỷ lệ 1:5.000.

4.1. Mạng Bayes-BN

BN là một phương pháp hiệu quả để biểu diễn tri thức theo ảnh hưởng không chắc chắn (Pearl, 1988), nó được giới thiệu đầu tiên bởi Friedman và nnk (Friedman và nnk, 1997). 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 (Marrcot và nnk, 2006). BN thường sử dụng rộng rãi trong các mô hình hệ thống phức tạp (Song và nnk, 2012). 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:

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

Trong đó, $X = (X_1, X_2, \dots, X_n)$ là các tác nhân gây trượt lở, $P_B(X | \prod_{i=1}^n X_i) = \theta_{X_i | \prod_{j=1}^{i-1} X_j}$ 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 giá nhạy cảm trượt lở ở các khu vực khác nhau trên thế giới (Oh và nnk, 2010; Pradhan, 2011; Wu và nnk, 2017; Hiền và nnk, 2020). 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 (Lee và Sambath, 2006; Wu và nnk, 2017; Hiền và nnk, 2020). 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 (Dai và nnk, 2001; Lee và Sambath, 2006; Wu và nnk, 2017; Hiền và nnk, 2020):

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

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 (Lee và Sambath, 2006; Wu và nnk, 2017; Hiền và nnk, 2020):

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

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 (Lee và nnk, 2008), 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 (Klecka, 1980). 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 (Klecka, 1980; Davis, 2002) 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 (Klecka, 1980).

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

neuron 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 (Yilmaz, 2010). Trong số nhiều loại mô hình mạng neuron nhân tạo, mạng chuyển tiếp ba lớp được sử dụng phổ biến trong thành lập bản đồ nhạy cảm trượt lở (Ermini và nnk, 2005; Nefeslioglu và nnk, 2008; Pradhan và Lee, 2010a; Pradhan và Lee, 2010b). 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 (Liu và nnk, 2013), số lượng tế bào thần kinh trên lớp ẩn nhiều gấp đôi số tế bào thần kinh

đầu vào. Đối với mô hình mạng nơron 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 nơron nhân tạo, một mạng chuyên 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 nơron 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 nơron nhiều lớp, bao gồm một lớp đầu vào, các lớp ẩn và một lớp đầu ra. Các nơron 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 nơron nhân tạo “học” bằng cách điều chỉnh trọng số giữa các nơron 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 nơron cung cấp một mô hình có thể dự đoán giá trị mục tiêu từ một giá trị đầu vào nhất định (Lee và nnk, 2006).

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 (Fath và nnk, 2020).

4.5. Máy véc tơ hỗ trợ-SVM

Máy vector hỗ trợ là kỹ thuật dựa trên hạt nhân không tham số được rút ra từ lý thuyết thống kê học (Vapnik, 1999). 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 (Schölkopf và Smola, 2002; Moguerza và Munoz, 2006; Cherkassky và Mulier, 2007). 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

(Kavzoglu và nnk, 2014):

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

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 (Suykens và Vandewalle, 1999; Dixon và Candale, 2008). Để phân loại nhị phân cho trượt lở và không trượt lở, điều kiện để giải phương trình (8) được mô phỏng dưới đây:

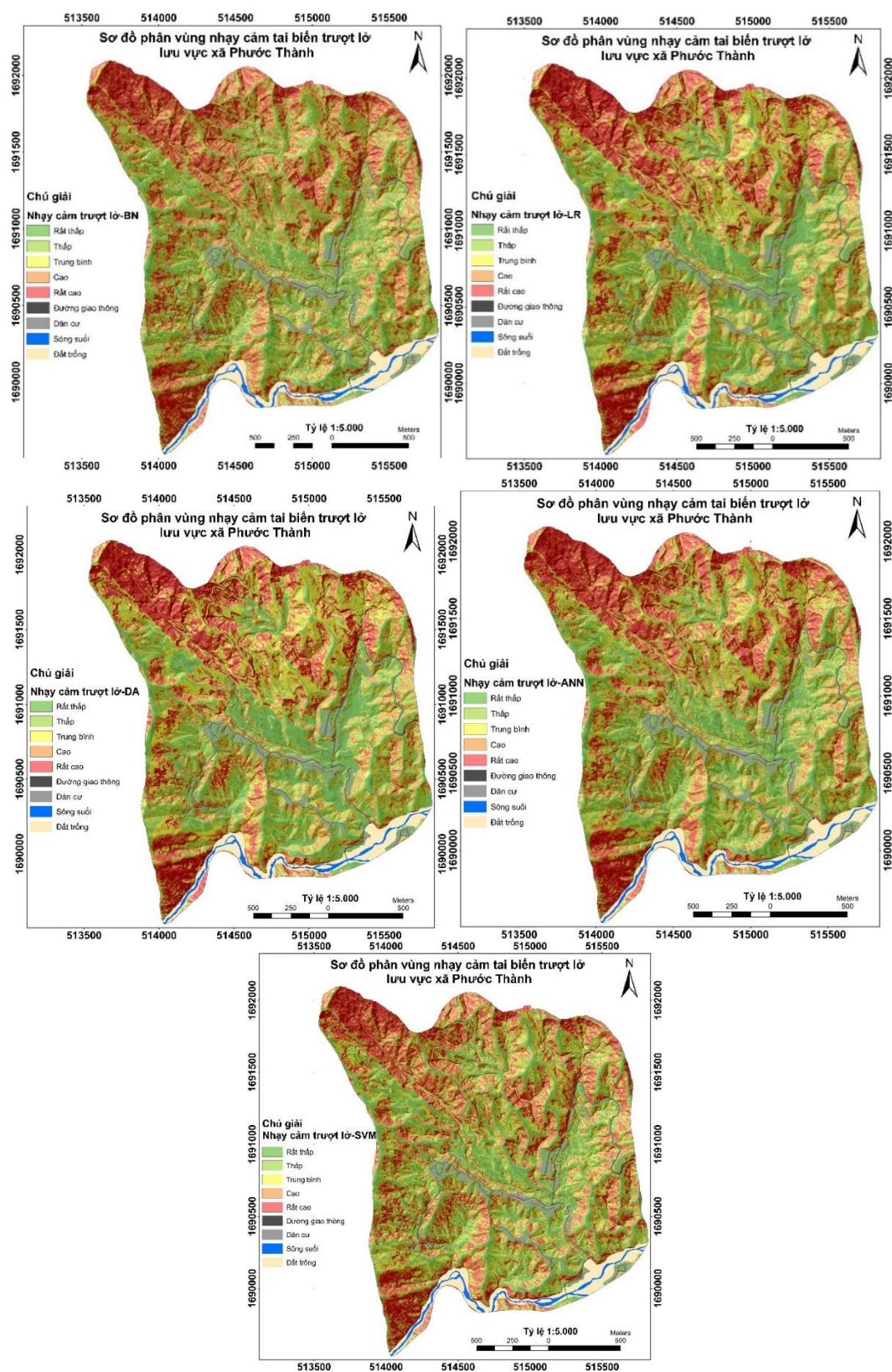
$$\begin{cases} \omega^T \varphi(X_i) + c \geq 1, & \text{if } Y_j = +1 (\text{landslide}) \\ \omega^T \varphi(X_i) + c \leq -1, & \text{if } Y_j = -1 (\text{non-landslide}) \end{cases} \quad (9)$$

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 (Binh và nnk, 2016).

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 5).



Hình 5. Các bản đồ nhạy cảm trượt lở được thành lập từ các phương pháp khác nhau (BN, LR, DA, ANN, SVM) lưu vực xã Phước Thành

5.2. Thẩm định các mô hình trượt lở

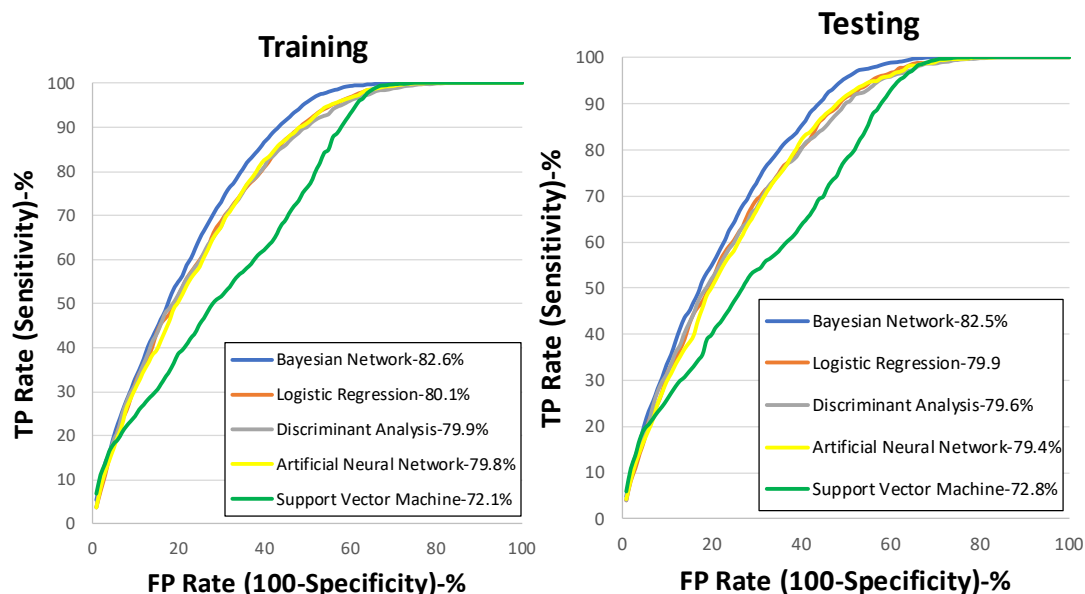
Có nhiều cách để đánh giá độ chính xác của các bản đồ phân vùng trượt lở. Cách tốt nhất là thẩm định ngoài thực địa nhưng việc này rất khó. Tuy nhiên có thể thẩm định dựa trên các thông tin về các điểm trượt lở đã xảy ra, đặc biệt cho các khu vực rủi ro cao và rất cao (Kayastha và nnk, 2013). Đánh giá kết quả dự báo là một trong những nhiệm vụ quan trọng nhất trong mô hình 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 (Chung và Fabbri, 2003; Mohammady và nnk, 2012; Pourghasemi và nnk, 2014). 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 (Diệu và nnk, 2016).

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 và tập dữ liệu kiểm nghiệm mô hình chiếm 30% số pixel còn lại. 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.

Đường cong ROC đánh giá độ chính xác của 5 mô hình học máy được thể hiện trong hình 5. Trong đó mô hình BN có độ chính xác cao nhất với giá trị AUC là 82.6% và 82.5% cho dữ liệu xây dựng mô hình và dữ liệu kiểm tra. Tiếp theo 3 mô hình LR, DA và ANN có giá trị AUC cho dữ liệu xây dựng mô hình và cho tập dữ liệu kiểm tra gần tương tự nhau. Lần lượt các giá trị AUC cho dữ liệu xây dựng mô hình của 3 phương pháp này là 80.1%, 79.9% và 79.8% và giá trị AUC cho dữ liệu kiểm tra là 79.9%, 79.6% và 79.4%. Cuối cùng là mô hình SVM có giá trị AUC cho tập dữ liệu xây dựng mô hình và dữ liệu kiểm tra là 72.1% và 72.8%.



Hình 6. Đường con ROC cho tập dữ liệu đào tạo (training) và kiểm tra (testing) của 5 mô hình thống kê và học máy đã áp dụng cho lưu vực xã Phước Thành

5.3. Thảo luận

Trong nghiên cứu này, cùng với dữ liệu trượt lở, 7 bản đồ tác nhân được sử dụng để thành lập bản đồ nguy cơ trượt lở bao gồm: độ dốc, TWI, SPI, STI, mặt cong 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 (Bình và nnk, 2016; Wu và nnk, 2017) đã sử dụng 15 bản đồ tác nhân; (Yesilnacar và Topal, 2005; Pourghasemi và nnk, 2013; Hong và nnk, 2015; Hong và nnk, 2016) với 14 tác nhân trượt lở; (Hiển và nnk, 2020) với 13 bản đồ tác nhân; (Talaiei, 2014) 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 nguy cơ trượt lở 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 bản đồ nguy cơ trượt lở. 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 Phước Thành nói chung và các khu vực miền núi khác của Việt Nam nói riêng trong công tác thành lập bản đồ nguy cơ trượt lở ở tỷ lệ lớn.

6. Kết luận

Nghiên cứu này đã áp dụng các mô hình thống kê và học máy để thành lập bản đồ phân vùng nguy cơ tai biến trượt lở ở tỷ lệ 1:5.000 cho khu vực xã Phước Thành, huyện Phước Sơn, tỉnh Quảng Nam. 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. Sau đó các kết quả được thẩm định để đánh giá độ chính xác (Hình 6) cho các phương pháp thống kê và học máy. Kết quả 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. Trong đó mô hình BN có độ chính xác cao nhất với giá trị AUC là 82.6% và 82.5% cho dữ liệu xây dựng mô hình và dữ liệu kiểm tra. Tiếp theo 3 mô hình LR, DA và ANN có giá trị AUC cho dữ liệu xây dựng mô hình và cho tập dữ liệu kiểm tra gần tương tự nhau. Lần lượt các giá trị AUC cho dữ liệu xây dựng mô hình của 3 phương pháp này là 80.1%, 79.9% và 79.8% và giá trị AUC cho dữ liệu kiểm tra là 79.9%, 79.6% và 79.4%. Cuối cùng là mô hình SVM có giá trị AUC cho tập dữ liệu xây dựng mô hình và dữ liệu kiểm tra là 72.1% và 72.8%. Mặc dù các giá trị AUC của các mô hình đều dao động từ hơn 70% đến hơn 80%. Tuy nhiên, giá trị này có thể cải thiện khi dữ liệu đầu vào được tăng lên và được cập nhật với độ chính xác cao hơn (thảm phủ, vỏ phong hóa...). Ngoài ra, dữ liệu về tai biến trượt lở cũng 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 là cần thiết. Ngoài ra, việc kết hợp các mô hình học máy trong thành lập bản đồ nguy cơ trượt lở 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.

Kết quả của nghiên cứu chỉ ra rằng các mô hình thống kê và học máy hoàn toàn có thể áp dụng để thành lập bản đồ nguy cơ trượt lở ở tỷ lệ lớn 1:5.000 cho các khu vực hạn chế về dữ liệu đầu vào. Kết quả nghiên cứu sẽ góp phần nâng cao khả năng phòng tránh và giảm thiểu tai biến trượt lở, 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 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. Đồng thời, tác giả xin chân thành cảm ơn các ý kiến góp ý của các phản biện mà đã góp phần nâng cao chất lượng của bài báo.

Đóng góp của các tác giả

Đỗ Minh Hiền chịu trách nhiệm lên ý tưởng, lựa chọn phương pháp, chuẩn bị, phân tích, tính toán, xử lý số liệu, viết bài báo, và biên tập cuối cùng. Nguyễn Văn Hoàng, Mai Lê Dũng, Phạm Đức Trọng, Nguyễn Cao Cường, Bùi Trọng Tấn tham gia công tác thực địa, thu thập thông tin về tai biến địa chất trong khu vực nghiên cứu.

Tài liệu tham khảo

- A. D. Regmi, K. C. Devkota, K. Yoshida, B. Pradhan, H. R. Pourghasemi, T. Kumamoto, A. Akgun, 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.*, Vol. 7, No. 2, 2014, pp. 725-742, <http://dx.doi.org/10.1007/s12517-012-0807-z>.
- A. H. Fath, F. Madanifar, M. Abbasi, 2020. 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>.
- A. M. S. Pradhan, H. S. Kang, S. Lee, Y. T. Kim, 2016. 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>.
- A. M. S. Pradhan, Y. T. Kim, 2014. 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>.
- A. M. S. Pradhan, Y. T. Kim, 2017. 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>.
- A. Yalcin, S. Reis, A. C. Aydinoglu, T. Yomralioglu, 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*, Vol. 85, No. 3, 2011, pp. 274-287, <http://dx.doi.org/10.1016/j.catena.2011.01.014>.
- B. Dixon, N. Candade, 2008. 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>.
- B. G. Marroto, J. D. Steventon, G. D. Sutheland, R. K. Mccann, 2006. 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>.
- B. Pradhan, 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, *Comput Geosci*, Vol. 51, 2013, pp. 350-365, <http://dx.doi.org/10.1016/J.Cageo.2012.08.023>.
- B. Pradhan, 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*, Vol. 18, No. 3, 2011, pp. 471-493, <http://dx.doi.org/10.1007/s10651-010-0147-7>.
- B. Pradhan, 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*, Vol. 18, No. 3, 2011, pp. 471-493, <http://dx.doi.org/10.1007/s10651-010-0147-7>.
- B. Pradhan, S. Lee, 2010a. 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, 2010a, pp. 1037-1054, <https://doi.org/10.1007/s12665-009-0245-8>.
- B. Pradhan, S. Lee, 2009. 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>.
- B. Pradhan, S. Lee, 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*, Vol. 25, 2010b, pp. 747-59, <https://doi.org/10.1016/j.envsoft.2009.10.016>.
- B. Q. Luna, J. Blahut, C. Camera, C. Van Westen, T. Apuani, V. Jetten, S. Sterlacchini, 2014.** 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>.
- B. Q. Luna, J. Blahut, M. Kappes, S. O. Akbas, J. P. Malet, A. Remaitre, T. V. Asch, M. Jaboyedoff, 2014.** 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>.
- B. Q. Luna, J. Blahut, T. V. Asch, C. V. Westen, M. Kappes, 2016.** 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>.
- B. Schölkopf, A. Smola, 2002.** Learning with Kernels: Support Vectormachines, Regularization, Optimization, and Beyond, *MIT Press, Cambridge*, 2002.
- B. T. Dieu, B. Pradhan, O. Lofman, I. Revhaug, O. B. Dick, 2012a.** Landslide Susceptibility Assessment in The Hoabinh Province of Vietnam: A Comparison of the Levenberg-Marquardt and Bayesian Regularized Neural Networks, *Geomorphology*, Vol. 171-172, 2012b, pp. 12-29, <https://doi.org/10.1016/j.geomorph.2012.04.023>.
- B. T. Dieu, B. Pradhan, O. Lofman, I. Revhaug, O. B. Dick, 2012b.** 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, 2012a, pp. 28-40, <https://doi.org/10.1016/j.catena.2012.04.001>.
- B. T. Dieu, H. Shahabi, A. Shirzadi, K. Chapi, B. Pradhan, W. Chen, K. Khosravi, M. Panahi, B. B. Ahmad, L. Saro, 2018.** 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>.
- B. T. Dieu, P. T. Binh, Q. P. Nguyen, N. D. Hoang, 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*, 2016, pp. 1-21, <https://doi.org/10.1080/17538947.2016.1169561>.
- Bản đồ địa chất thu thập từ Đề án trượt lở quốc gia do Viện Khoa học Địa chất và Khoáng sản (VIGMR) chủ trì thực hiện (2012-2020).**
- C. J. F. Chung, A. G. Fabbri, 2003.** 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>.
- C. J. V. Westen, M. T. J. Terlien, 1996.** 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).
- C. Liu, W. Y. Li, H. B. Wu, P. Lu, K. Sang, W. W. Sun, W. Chen, Y. Hong, R. X. Li, 2013.** 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>.
- C. T. Lee, C. C. Huang, J. F. Lee, K. L. Pan, M. N. Lin, J. J. Dong, 2008.** 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>.
- C. Xu, X. Xu, F. C. Dai, J. Z. Xiao, X. B. Tan, R. M. Yuan, 2012.** 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>.
- D. Rickenmann, D. Laigle, B. McArdell, J. Hübl, 2006.** 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>.
- E. Leroi, 1996.** 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>.

- E. R. Sujatha, P. Kumaravel, G. V. Rajamanickam, 2014.** 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>.
- E. Yesilnacar, T. Topal, 2005.** 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>.
- F. C. Dai, C. F Lee, J. Li, Z. W. Xu, 2001.** 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>.
- F. Legros, 2002.** 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).
- F. Mengistu, K. V. Suryabhadgavan, T. K. Raghuvanshi, E. Lewi, 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*, Vol. 3, No. 1, 2019, pp. 1-14, <https://doi.org/10.21523/gcj1.19030101>.
- G. B. Crosta, S. Cucciaro, P. Frattini, 2003** 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.
- G. F. Zhang, Y. X. Cai, Z. Zheng, J. W. Zhen, Y. L. Liu, K. Y. Huang, 2016.** 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>.
- H. A. Nefeslioglu, C. Gokceoglu, H. Sonmez, 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*, Vol. 97, No. 3-4, 2008, pp. 171-91, <https://doi.org/10.1016/j.enggeo.2008.01.004>.
- H. J. Oh, S. Lee, 2010.** 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>.
- H. J. Oh, S. Lee, G. M. Soedradjat, 2010.** 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>.
- H. R. Pourghasemi, A. G. Jirandeh, B. Pradhan, C. Xu, C. Gokceoglu, 2013.** 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>.
- H. R. Pourghasemi, B. Pradhan, C. Gokceoglu, M. Mohammadi, H. R. Moradi, 2013.** 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>.
- H. R. Pourghasemi, H. R. Moradi, S. M. Fatemi Aghda, C. Gokceoglu, B. Pradhan, 2014.** 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>.
- H. Y. Hong, B. Pradhan, C. Xu, T. B. Dieu, 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*, Vol. 133, 2015, pp. 266-281, <http://dx.doi.org/10.1016/j.catena.2015.05.019>.
- H. Y. Hong, H. R. Pourghasemi, Z. S. Pourtaghi, 2016.** 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>.
<http://phuocthanh.phuocson.quangnam.gov.vn/Default.aspx?tabid=1532&language=en-US> (tham khảo ngày 22/2/2023).
- <https://laodong.vn/xa-hoi/khan-cap-khoi-phuc-sat-lo-duong-vao-2-xa-bi-co-lap-o-quang-nam-861388.ldo> (tham khảo ngày 22/2/2023).
- https://quangnam.gov.vn/webcenter/portal/ubn/d/pages_tin-tuc/chi-tiet?dDocName=PORTAL072221 (tham khảo ngày 22/2/2023).

- I. D. Moore, G. J. Burch, 1986.** 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>.
- I. Yilmaz, 2010.** 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>.
- I. Yilmaz, 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*, Vol. 35, 2009, pp. 1125–1138, <https://doi.org/10.1016/j.cageo.2008.08.007>.
- J. A. Suykens, J. Vandewalle, 1999.** Least Squares Support Vector Machine Classifiers, *Neural Process Lett*, Vol. 9, No. 3, 1999, pp. 293–300, <https://doi.org/10.1023/a:1018628609742>.
- J. Blahut, P. Horton, S. Sterlacchini, M. Jaboyedoff, 2010.** Debris Flow 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>.
- J. C. Davis, 2002.** Statistics and Data Analysis in Geology, (3rd Ed), *John Wiley And Sons*, New York, 2002.
- 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. Pitilakis, M. G. Winter, M. Pastor, S. Ferlisi, V. Tofani, J. Hervé, J. T. Smith, 2014.** 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>.
- J. Corominas, 1996.** The Angle of Reach as a Mobility Index For Small And Large Landslides, Canadian. *Geotechnical Journal*, Vol. 33, No. 2, 1996, pp. 260–271, <https://doi.org/10.1139/t96-005>.
- J. L. Zezere, S. Pereira, R. Melo, S. C. Oliveira, R. A. C. Garcia, 2017.** 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>.
- J. M. Moguerza, A. Munoz, 2006.** Support Vector Machines With Applications, *Stat Sci*, Vol. 21, 2006, pp. 322–336, <https://doi.org/10.1214/088342306000000493>.
- J. N. Goetz, A. Brenning, H. Petschko, P. Leopold, 2015.** 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>.
- J. Pearl, 1988.** Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, *Morgan Kaufmann*, 1988.
- K. Beven, M. Kirkby, 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*, Vol. 24, 1979, pp. 43–69.
- K. Schraml, B. Thomschitz, B. W. McArdell, C. Graf, R. Kaitna, 2015.** 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>.
- L. Ermini, F. Catani, N. Casagli, 2005.** 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>.
- L. Q. Hung, N. T. H. Van, M. D. Duc, L. T. C. Ha, P. V. Son, N. H. Khanh, L. T. Binh, 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*, Vol. 13, 2016, pp. 1285–1301, <https://doi.org/10.1007/s10346-015-0657-3>.
- M. D. Hien, K. L. Yin, Z. Z. Guo, 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, pp. 2449–2485, 2020, <https://doi.org/10.1080/19475705.2020.1846086>.
- M. Hürlimann, D. Rickenmann, V. Medina, A. Bateman, 2008.** 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>.
- M. Jaboyedoff, Ch. Choffet, M. H. Derron, P. Horton, A. Loye, C. Longchamp, B. Mazotti, C. Michoud, 2012.** 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.
- M. Mohammady, H. R. Pourghasemi, B. Pradhan, 2012.** 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>.
- N. Friedman, D. Geiger, M. Goldszmidt, 1997.** Bayesian Network Classifiers, *Machine Learning*, Vol. 29, 1997, pp. 131-163, <https://doi.org/10.1023/a:1007465528199>.
- N. R. Regmi, J. R. Giardino, J. D. Vitek, 2010.** Modeling Susceptibility to Landslides Using The Weight of Evidence Approach: Western Colorado, USA, *Geomorphology*, Vol. 115, No. 1-2, 2010, pp. 172-187, <http://dx.doi.org/10.1016/j.geomorph.2009.10.002>.
- N. T. Long, F. D. Smedt, 2019.** 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>.
- N. T. Long, F. D. Smedt, 2012.** 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>.
- P. Aleotti, R. Chowdhury, 1999.** 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>.
- P. Horton, M. Jaboyedoff, B. Rudaz, M. Zimmermann, 2013.** 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>.
- P. Horton, M. Jaboyedoff, M. Zimmermann, B. Mazotti, C. Longchamp, 2011.** 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.
- P. Kayastha, M. R. Dhital, F. D. Smedt F, 2012.** 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>.
- P. T. Binh, B. Pradhan, B. T. Dieu, 2018.** 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>.
- P. T. Binh, B. Pradhan, T. D. Bui, I. Prakash, M. B. Dholakia, 2016.** 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>.
- P. T. Binh, N. T. Thoi, C. C. Qi, T. V. Phong, J. Dou, H. S. Lanh, L. V. Hiep, I. Prakash., 2020.** Coupling RBF Neural Network with Ensemble Learning Techniques for Landslide Susceptibility Mapping, *Catena*, Vol. 195, <https://doi.org/10.1016/j.catena.2020.104805>.
- Q. Q. Wang, W. P. Li, M. L. Xing, Y. L. Wu, Y. B. Pei, D. D. Yang, H. Y. Bai, 2016.** 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>.
- R. K. Dahal, S. Hasegawa, A. Nonomura, M. Yamanaka, S. Dhakal, P. Paudyal, 2008.** 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>.
- R. K. Dahal, S. Hasegawa, A. Nonomura, M. Yamanaka, T. Masuda, K. Nishino, 2008.** 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>.
- R. M. Iverson, S. P. Schilling, J. W. Vallance, 1998.** Objective Delineation of Lahar-Inundation Hazard Zones, *Geological Society of America Bulletin*, Vol. 110, No. 8, 1998, pp. 972-984, [https://doi.org/10.1130/00167606\(1998\)110<0972:ODOLIH>2.3.CO;2](https://doi.org/10.1130/00167606(1998)110<0972:ODOLIH>2.3.CO;2).

- R. M. Iverson, 1997.** The Physics of Debris Flows, *Review of Geophysics*, Vol. 35, No. 3, 1997, pp. 245-296, <https://doi.org/10.1029/97RG00426>.
- R. Talaei, 2014.** 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>.
- S. H. Kang, S. R. Lee, N. Nikhil, J. Y. V. Park, D. H. Lee, 2017.** 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, <https://doi.org/10.1016/j.enggeo.2017.09.017>.
- S. Kang, S. R. Lee, 2018.** 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>.
- S. Lee, B. Pradhan, 2007.** 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>.
- S. Lee, J. H. Ryu, M. J. Lee, J. S. Won, 2006.** 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>.
- S. Lee, K. Min, 2001.** Statistical Analysis of Landslide Susceptibility at Yongin, Korean, *Environmental Geology*, Vol. 40, 2001, pp. 1095-1113, <https://doi.org/10.1007/s002540100310>.
- S. Lee, T. Sambath, 2006.** 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>.
- T. Chen, R. Q. Niu, X. P. Jia, 2016.** 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>.
- T. Kavzoglu, E. K. Sahin, I. Colkesen, 2014.** 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>.
- V. Cherkassky, F. Mulier, 2007.** Learning From Data: Concepts, Theory, and Methods, Wiley, New York, 2007.
- V. N. Vapnik, 1999.** The Nature of Statistical Learning Theory, 2nd, Springer, New York, 1999.
- W. R. Klecka, 1980.** Discriminant Analysis, Sage University Paper Series on Quantitative Applications in the Social Sciences, Sage Publications, Beverly Hills and London, 1980, pp. 07-19.
- Y. Huang, L. Zhao, 2018.** 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>.
- Y. Song, J. Gong, S. Gao, D. Wang, T. Cui, Y. Li, B. Wei, 2012.** 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>.
- Z. Wu, Y. Wu, Y. Yang, F. Chen, N. Zhang, Y. Ke, W. Li, 2017.** 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>.

Summary
Using statistical and machine learning models for large-scale landslide hazard susceptibility mapping

Do Minh Hien^{*1}, Nguyen Van Hoang², Mai Le Dung¹, Pham Duc Trong¹,
Nguyen Cao Cuong¹, Bui Trong Tan¹

¹ *Vietnam Institute of Geosciences and Mineral Resources.
67 Chien Thang, Van Quan, Ha Dong, Ha Noi*

² *University of Warszawa, Poland*

Corresponding author: hien_dm@yahoo.com

The main purpose of this article is to establish a susceptibility zonation map of the landslides in Phuoc Thanh commune, Phuoc Son district, Quang Nam province on a large scale using statistical methods and machine learning models. 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. 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, statistical model, machine learning model, Phuoc Thanh.*