

Ứng dụng thuật toán cây quyết định CHAID xác định các yếu tố và ngưỡng ảnh hưởng đến sạt lở đất tầng nông tại xã Trung Chải và xã Tú Lệ, tỉnh Lào Cai

Kiều Thị Dương, Bùi Xuân Dũng*, Nguyễn Thị Thanh An,
Đỗ Anh Tuấn, Nguyễn Khánh Linh, Phạm Minh Thiện

Trường Đại học Lâm nghiệp

Application of the CHAID decision tree algorithm to identify key factors and thresholds influencing shallow landslide in Trung Chai and Tu Le communes, Lao Cai province

Kieu Thi Duong, Bui Xuan Dung*, Nguyen Thi Thanh An,
Do Anh Tuan, Nguyen Khanh Linh, Pham Minh Thien

Vietnam National University of Forestry

*Corresponding author: dungbx@vnuf.edu.vn

<https://doi.org/10.55250/jo.vnuf.15.7.2026.083-091>

TÓM TẮT

Sạt lở đất đang gia tăng cả về tần suất và mức độ nghiêm trọng tại các vùng núi phía Bắc Việt Nam. Các nghiên cứu định lượng để phục vụ công tác dự báo và phòng ngừa sạt lở hiệu quả là thực sự cần thiết. Nghiên cứu ứng dụng thuật toán CHAID trên phần mềm SPSS 20.0 nhằm xác định các yếu tố và ngưỡng giá trị ảnh hưởng đến cấp độ sạt lở đất tầng nông tại xã Trung Chải và xã Tú Lệ, tỉnh Lào Cai. Dữ liệu được thu thập từ 60 ô tiêu chuẩn (OTC) diện tích 100 m² mỗi ô. Kết quả cho thấy tại xã Trung Chải, độ dốc, bề dày tầng đất và chiều cao dưới cành lần lượt là các yếu tố ảnh hưởng mạnh nhất. Tại xã Tú Lệ, mật độ cây cao và độ che phủ cây bụi, thảm tươi là các biến quan trọng nhất. Phân tích gộp 60 OTC với kiểm chứng chéo 10-fold cho thấy độ dốc là yếu tố ảnh hưởng mạnh nhất, sau đó là mật độ cây cao. Theo đó, khi độ dốc $\leq 33^\circ$ tương ứng cấp sạt lở thấp, $34^\circ\text{--}37^\circ$ cấp trung bình và $> 37^\circ$ cấp cao. Trong nhóm độ dốc thấp ($\leq 33^\circ$), các OTC có cây gỗ hiện diện có tỷ lệ cấp sạt lở thấp cao hơn. Mô hình đạt độ chính xác phân loại tổng thể 73,3%. Đây được coi là nghiên cứu đầu tiên áp dụng CHAID ở quy mô thực địa tại vùng núi phía Bắc Việt Nam. Kết quả nghiên cứu bước đầu tạo cơ sở khoa học cho công tác dự báo, và xây dựng biện pháp phòng ngừa sạt lở đất tầng nông tại khu vực nghiên cứu.

ABSTRACT

Landslides are increasing in both frequency and severity in the mountainous regions of Northern Vietnam. Quantitative research to support effective landslide forecasting and prevention is therefore essential. This study applied the CHAID algorithm using SPSS 20.0 to identify the key factors and their threshold values associated with shallow landslide severity in Trung Chai and Tu Le communes, Lao Cai province. Data were collected from 60 field plots, each covering an area of 100 m². The results showed that in Trung Chai commune, slope gradient, soil depth, and crown base height were the most influential factors, respectively. In Tu Le commune, tree density and understory vegetation cover were the most important variables. Combined analysis of all 60 sample plots with 10-fold cross-validation indicated that slope gradient was the strongest influencing factor, followed by tree density. Accordingly, slopes of $\leq 33^\circ$ were mostly associated with low landslide severity, $34^\circ\text{--}37^\circ$ with a moderate level, and $> 37^\circ$ with high severity. Within the low-slope group ($\leq 33^\circ$), plots with woody tree presence had a higher proportion of low landslide severity. The overall classification accuracy was 73.3%. This study can be considered the first field-scale application of the CHAID algorithm in the mountainous regions of Northern Vietnam. The initial results provide a scientific basis for forecasting and developing preventive measures against shallow landslides in the study area.

Thông tin chung:

Ngày nhận bài: 08/05/2026

Ngày phản biện: 28/07/2026

Ngày quyết định đăng: 24/08/2026

Keywords:

CHAID, Lào Cai, sạt lở đất, yếu tố ảnh hưởng.

Từ khóa:

CHAID, influencing factors, landslide, Lao Cai.

1. ĐẶT VẤN ĐỀ

Sạt lở đất tầng nông được hiểu là loại hình tai biến xảy ra trong lớp vật liệu nằm phía trên đá gốc, bao gồm đất, saprolite (đá phong hóa) và colluvium (trầm tích dốc) [1]. Loại tai biến này thường xảy ra với lớp đất mỏng (chưa đến 2m) được kích hoạt bởi những trận mưa lớn, kéo dài tạo ra sức phá hoại rất lớn [2]. Ở Việt Nam, loại hình tai biến này xuất hiện phổ biến tại vùng núi phía Bắc, nơi có địa hình dốc, mưa lớn và điều kiện phong hóa mạnh. Những năm gần đây, do ảnh hưởng của biến đổi khí hậu cùng với các hoạt động tác động của con người, tần suất và mức độ nghiêm trọng của sạt lở đất ngày càng tăng, gây tổn thất lớn về người và tài sản, đe dọa sự phát triển kinh tế - xã hội bền vững của khu vực. Việc xác định các yếu tố ảnh hưởng và ngưỡng giá trị tới hạn của các yếu tố đó đối với cấp độ sạt lở đất là rất cần thiết, nhằm xây dựng cơ sở khoa học cho công tác dự báo, cảnh báo sớm và phòng ngừa sạt lở đất hiệu quả. Tuy nhiên, phần lớn các nghiên cứu về sạt lở đất gặp nhiều khó khăn bởi sạt lở đất chịu tác động bởi mối quan hệ tương tác phức tạp giữa nhiều yếu tố.

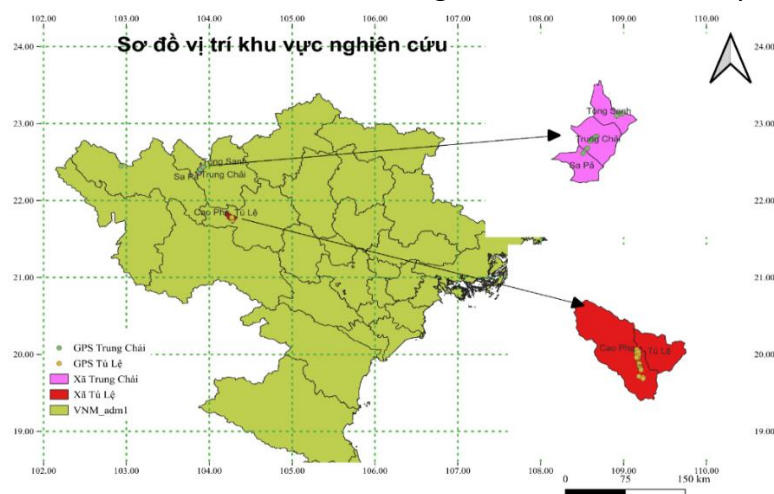
Trong các yếu tố ảnh hưởng đến sạt lở đất, độ dốc đất được coi là yếu tố địa hình then chốt; trong khi độ dốc kiểm soát chuyển động của các vật liệu thì che phủ thảm thực đóng vai trò quan trọng trong việc tăng cường độ ổn định sườn dốc thông qua hệ rễ và giảm xói mòn bề mặt [3-5]. Nghiên cứu của Celtek và cộng sự (2020); Nagarajan và cộng sự (2000) tại Ấn Độ khẳng định khả năng sạt lở tăng lên khi độ dốc tăng, nhưng ở những nơi có độ dốc trên 80° thường xuất hiện sạt lở kiểu đá rơi (rockfall-type landslides) còn các vụ sạt lở kiểu khối bùn trượt

(mudslide-type landslides) thường xảy ra trên các sườn có độ dốc lớn hơn 35° [3, 6]. Ở Việt Nam mặc dù đã có nghiên cứu về sạt lở, tuy nhiên còn thiếu các nghiên cứu định lượng nhằm xác định các yếu tố then chốt ảnh hưởng đến sạt lở và ngưỡng tác động tới hạn của từng yếu tố đối với cấp độ sạt lở đất. Bên cạnh đó, việc ứng dụng các phương pháp phân tích dữ liệu hiện đại, như mô hình cây quyết định, trong nghiên cứu sạt lở đất vẫn còn hạn chế. Xuất phát từ khoảng trống nghiên cứu nêu trên, nghiên cứu này ứng dụng thuật toán cây quyết định CHAID trên phần mềm SPSS 20 (phiên bản được xem là dùng ổn định nhất hiện nay) nhằm mục tiêu: (1) Xác định một số yếu tố ảnh hưởng rõ rệt nhất đến cấp độ sạt lở đất; (2) Tìm ra ngưỡng giới hạn tối ưu của từng yếu tố ảnh hưởng đến sạt lở tại xã Trung Chải và xã Tú Lệ, tỉnh Lào Cai. Kết quả nghiên cứu góp phần cung cấp cơ sở khoa học quan trọng cho việc dự báo và thực thi các biện pháp giảm thiểu nguy cơ sạt lở đất tầng nông tại khu vực nghiên cứu nói riêng và ở các khu vực miền núi Việt Nam nói chung.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Địa điểm nghiên cứu

Nghiên cứu được thực hiện tại xã Tú Lệ và xã Trung Chải, tỉnh Lào Cai. Xã Tú Lệ trước đây thuộc huyện Văn Chấn, tỉnh Yên Bái, nay được sát nhập với xã Cao Phạ huyện Mù Cang Chải thành xã mới có tên gọi là xã Tú Lệ thuộc tỉnh Lào Cai. Xã có địa hình chia cắt, độ cao tuyệt đối từ 800-2.000 m, lượng mưa từ 1.500-2.500 mm. Xã Trung Chải, Sapa, Lào Cai nơi địa hình chia cắt, độ dốc lớn, có độ cao từ 1.000- 1.500 m, lượng mưa trung bình từ 1.800- 2.700 mm. Vị trí các OTC và khu vực nghiên cứu được trình bày ở Hình 1.



Hình 1. Sơ đồ vị trí khu vực nghiên cứu

2.2. Phương pháp nghiên cứu

a) Phương pháp thu thập, kế thừa tài liệu:

Số liệu về loại đất, bề dày tầng đất cho từng OTC được xuất từ bản đồ đất Việt Nam theo định dạng Shapefile bằng phần mềm QGIS, trên hệ tọa độ WGS-84 do Viện Thổ nhưỡng Nông hóa Việt Nam xây dựng năm 2000, lượng mưa trung bình ngày cho từng OTC được trích xuất từ bộ dữ liệu tái phân tích ERA5-Land thông qua nền tảng Google Earth Engine.

b) Phương pháp điều tra thực địa

b1. Thiết kế mẫu và thu thập số liệu:

Nghiên cứu được thực hiện trên 60 OTC diện tích 100 m²/ô, phân bố tại hai xã Trung Chải và Tú Lệ, tỉnh Lào Cai (mỗi khu vực 30 OTC). Thiết kế mẫu áp dụng lấy mẫu phân tầng kết hợp có chủ đích (stratified purposive sampling) nhằm đảm bảo tính đại diện cho các điều kiện địa hình, thảm thực vật và hiện trạng sạt lở. Cụ thể, khu vực nghiên cứu được phân tầng theo ba tiêu chí chính: (i) lớp độ dốc (thấp–trung bình–cao), (ii) trạng thái thảm thực vật (rừng kín, rừng thưa, cây bụi/thảm tươi, đất trống), (iii) dấu hiệu sạt lở quan sát được (không/nhẹ–trung bình–mạnh). Trong mỗi tổ hợp phân tầng, vị trí OTC được chọn ngẫu nhiên có kiểm soát (randomized within strata) nhằm giảm sai lệch chọn mẫu và tránh tập trung điểm đo theo đường giao thông hoặc khu vực dễ tiếp cận. Khoảng cách tối thiểu giữa các OTC được duy trì (≥ 50 –100 m) để hạn chế sự tương quan không gian ở quy mô điểm đo. Tại mỗi OTC, thu thập 14 biến độc lập gồm: Địa hình: độ dốc ($^{\circ}$), độ cao (m); Đất: bề dày tầng đất (cm), độ ẩm (%), pH, độ chặt; Khí hậu: lượng mưa trung bình ngày; Thảm thực vật: mật độ cây cao (cây/ha), độ tàn che (%), tỷ lệ che phủ cây bụi, thảm tươi (%), tỷ lệ che phủ vật rơi rụng (%), Hvn (m), Hdc (m), D1.3 (cm). Các phép đo hiện trường tuân thủ quy trình lâm học tiêu chuẩn; các biến không gian, khí hậu được trích xuất từ dữ liệu thứ cấp đã được chuẩn hóa hệ tọa độ và kiểm tra nhất quán.

b.2. Xác định biến phụ thuộc (cấp độ sạt lở):

Biến phụ thuộc là cấp độ sạt lở đất tầng nông, phân thành 3 mức: cấp 1 (thấp), cấp 2 (trung bình), cấp 3 (cao). Phân loại được xây dựng theo chỉ số bán định lượng (semi-

quantitative index) dựa trên các tiêu chí thực địa có thể quan sát và đo đếm, bao gồm: Mật độ và chiều rộng vết nứt bề mặt; Thể tích/diện tích vật liệu trượt; Mức độ xói mòn và bóc tách lớp đất mặt; Dấu hiệu mất ổn định sườn dốc (trượt khối, trượt nông, bùn chảy...). Mỗi tiêu chí được chấm điểm theo các mức chuẩn hóa (từ 0–3), sau đó tổng hợp thành chỉ số tổng và quy đổi sang 3 cấp sạt lở theo ngưỡng định trước. Để giảm chủ quan, quy trình đánh giá được chuẩn hóa bằng phiếu hướng dẫn hiện trường, thực hiện đánh giá lặp giữa các thành viên trong nhóm và đối chiếu nhanh tại hiện trường. Cách tiếp cận này giúp nâng cao tính nhất quán và độ tin cậy của biến phụ thuộc trong điều kiện khó thu thập số liệu định lượng trực tiếp. Phân bố mẫu theo cấp sạt lở bao gồm: 14 OTC cấp sạt lở 2 chiếm 23,3% mẫu, cấp sạt lở 1 và 3 mỗi cấp gồm 23 OTC tương đương mỗi cấp là 38,3%.

c) Phương pháp xử lý số liệu

Phân tích cây quyết định CHAID được áp dụng nhằm xác định các yếu tố ảnh hưởng và ngưỡng tác động đến sạt lở. Dữ liệu đầu vào bao gồm 60 OTC được điều tra tại 2 khu vực nghiên cứu, mỗi khu vực gồm 30 OTC. Trong đó, biến phụ thuộc là cấp sạt lở được phân thành ba mức 1 – 3 tương ứng với cấp thấp, trung bình và cao; 14 biến độc lập gồm các biến đặc trưng cho địa hình, thảm thực vật và đất như đã mô tả ở phần điều tra thực địa. Trước khi xây dựng mô hình, 14 biến độc lập đầu vào đã được kiểm tra đặc điểm, trong đó có cả biến liên tục và các biến phân loại. Thuật toán CHAID có khả năng xử lý đồng thời biến liên tục và biến phân loại. Các biến liên tục được thuật toán tự động phân nhóm thành các khoảng dựa trên kiểm định χ^2 để xác định ngưỡng phân tách tối ưu tại mỗi nút, do đó không cần thực hiện biến đổi hay chuẩn hóa trước.

Phân tích được thực hiện qua ba lần tính toán: hai lần riêng cho từng khu vực và một lần tổng hợp (60 OTC), nhằm so sánh đặc điểm phân tách theo không gian và xác định quy luật chung. Thuật toán CHAID sử dụng kiểm định χ^2 để chọn biến phân tách tại mỗi nút dựa trên p-value; chỉ các biến có ý nghĩa thống kê ($p \leq 0,05$) được giữ lại. Để hạn chế phân tách quá mức,

các tham số được thiết lập thống nhất: số quan sát tối thiểu tại nút cha là 6 và tại nút con là 2 (Minimum Cases in Parent/Child Node).

Độ ổn định và khả năng khái quát hóa của mô hình được đánh giá bằng kết quả kiểm chứng chéo (10-fold cross-validation). Dữ liệu được chia thành 10 phần; mỗi lần lặp sử dụng 9 phần để huấn luyện và 1 phần để kiểm tra, lặp lại đến khi mỗi phần được dùng làm tập kiểm tra một lần. Kết quả cây quyết định được diễn giải thông qua các chỉ số biến quan trọng (importance), giá trị P hiệu chỉnh (Adj. P value),

ước tính rủi ro (risk estimate), sai số chuẩn (standard error) và độ chính xác phân loại (classification accuracy).

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Các yếu tố ảnh hưởng đến sạt lở đất tại xã Trung Chải, tỉnh Lào Cai

Trước khi phân tích chi tiết các yếu tố ảnh hưởng đến sạt lở đất ở từng địa phương, Bảng 1 dưới đây mô tả khái quát một số đặc trưng thống kê của các yếu tố ảnh hưởng của toàn bộ 2 khu vực nghiên cứu.

Bảng 1. Thống kê mô tả các biến độc lập

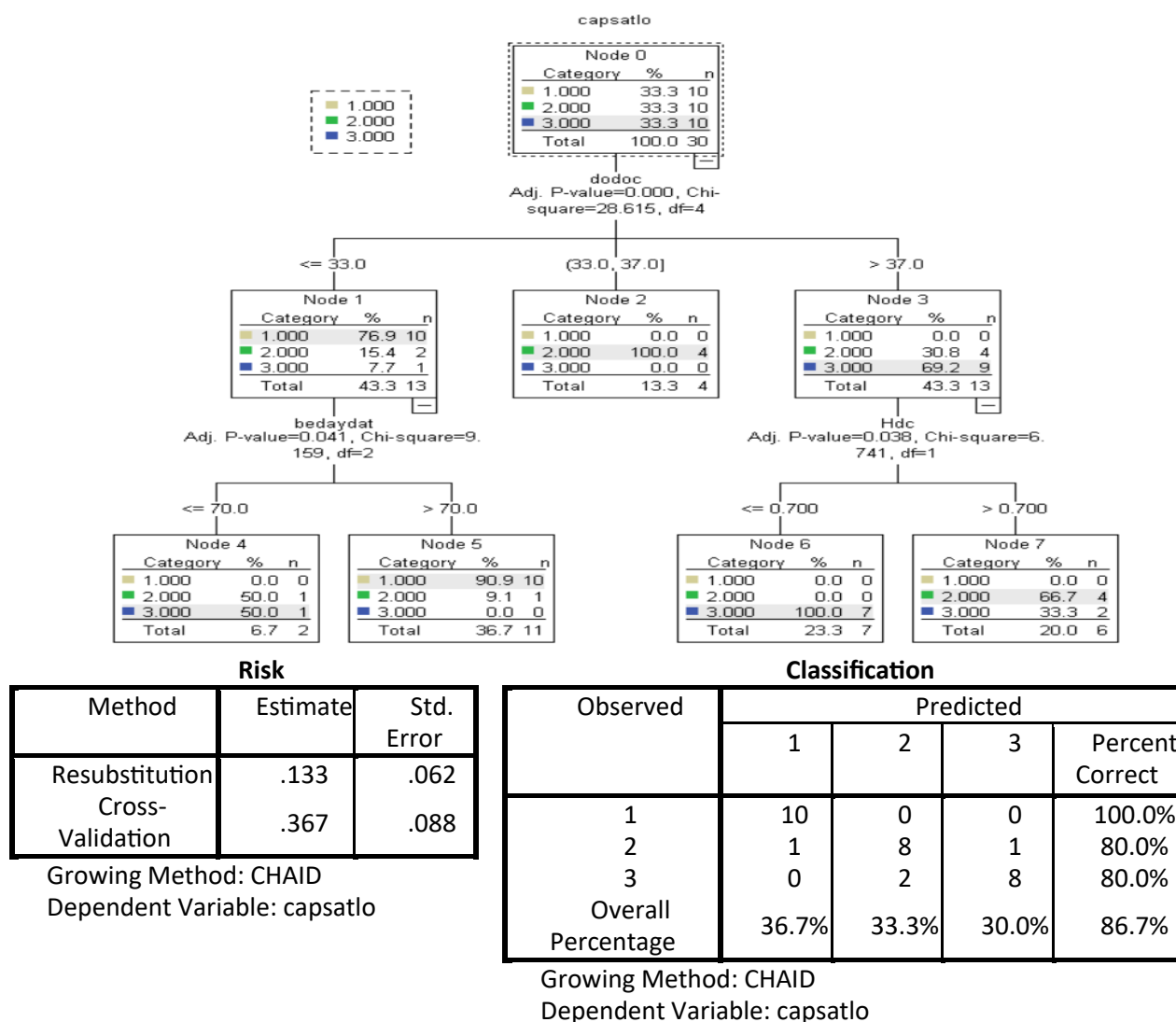
Chỉ tiêu	N	Min	Max	Mean	Median	Std. Deviation (SD)	Skewness	Kurtosis
Bề dày tầng đất (cm)	60	70	150	113,58	120	26,15	-,25	-1,38
Độ dốc (độ)	60	14	61	36,0	35	9,71	,31	0,17
Độ cao (m)	60	178	1236	895,6	957,5	244,43	-1,35	1,80
Lượng mưa TB ngày (mm)	60	5,60	7,2	6,8	6,97	0,57	-1,04	-0,19
Mật độ cây cao (cây/ha)	60	0	4.000	720,4	11	993,57	1,54	1,98
D1.3 (cm)	60	0	25,8	6,5	4,75	7,26	,63	-0,78
Hvn (m)	60	0	23,3	4,8	2,6	5,69	,94	0,28
Hdc (m)	60	0	8,2	2,08	0,775	2,55	,93	-0,37
Độ tàn che (%)	60	0	100	37,8	15	39,13	,27	-1,78
Che phủ thảm tươi (%)	60	0	100	46,5	52,5	33,26	-,11	-1,33
Tỷ lệ che phủ vật rơi rụng (%)	60	0	100	42,3	30	39,34	,31	-1,58
Độ ẩm đất (%)	60	0	78,0	43,3	45,5	23,78	-,10	-1,51
Độ pH đất	60	5,3	6,9	6,2	6,3	0,30	-,87	1,80
Độ chặt đất (mm)	60	6,4	22,6	14,3	14,3	3,57	-,01	-0,38

Kết quả thống kê mô tả cho thấy các biến ảnh hưởng đến sạt lở đất trên 60 OTC có mức độ biến thiên và đặc điểm phân bố khá khác nhau. Các biến địa hình và đất đai thể hiện sự biến động rõ rệt, điển hình như độ dốc (Mean = 36°, SD = 9,71) thể hiện mức độ biến thiên khá lớn hoặc bề dày tầng đất (Mean = 113,58 cm, SD = 26,15). Biến độ cao có biên độ rộng từ 178m đến 1236m, phản ánh sự đa dạng về địa hình của hai khu vực nghiên cứu. Biến lượng mưa trung bình ngày có biên độ biến thiên rất hẹp (5,6 – 7,2 mm), cho thấy yếu tố này thay đổi không nhiều giữa hai khu vực nghiên cứu.

Đáng chú ý nhóm biến thảm thực vật có độ lệch (Skewness) và độ nhọn (Kurtosis) lớn, phân phối lệch mạnh, điển hình là mật độ cây cao (Mean = 720,4 cây/ha, Median = 11) và độ

tàn che (Median = 15%), Hvn, Hdc và D1.3 đều có giá trị trung vị thấp hơn nhiều so với giá trị trung bình. Sự biến động này xuất phát từ việc nghiên cứu bao gồm cả các OTC có thực vật và không có thực vật (đất trống), dẫn đến sự phân bố không đều và tồn tại nhiều giá trị bằng 0 ở các biến cấu trúc rừng.

Phân tích CHAID được thực hiện trên 30 OTC tại xã Trung Chải, tỉnh Lào Cai với thiết lập Minimum Cases in Parent Node = 6 và Minimum Cases in Child Node = 2. Từ 14 biến độc lập ban đầu, mô hình chỉ ra 3 biến có ý nghĩa thống kê cao nhất có ảnh hưởng rõ rệt đến cấp độ sạt lở là độ dốc, bề dày tầng đất và chiều cao dưới cành (Hdc) của tầng cây cao. Kết quả được thể hiện ở cây quyết định CHAID (Hình 2).



Hình 2. Cây quyết định CHAID xác định các yếu tố ảnh hưởng đến cấp độ sạt lở đất tại xã Trung Chải, Lào Cai

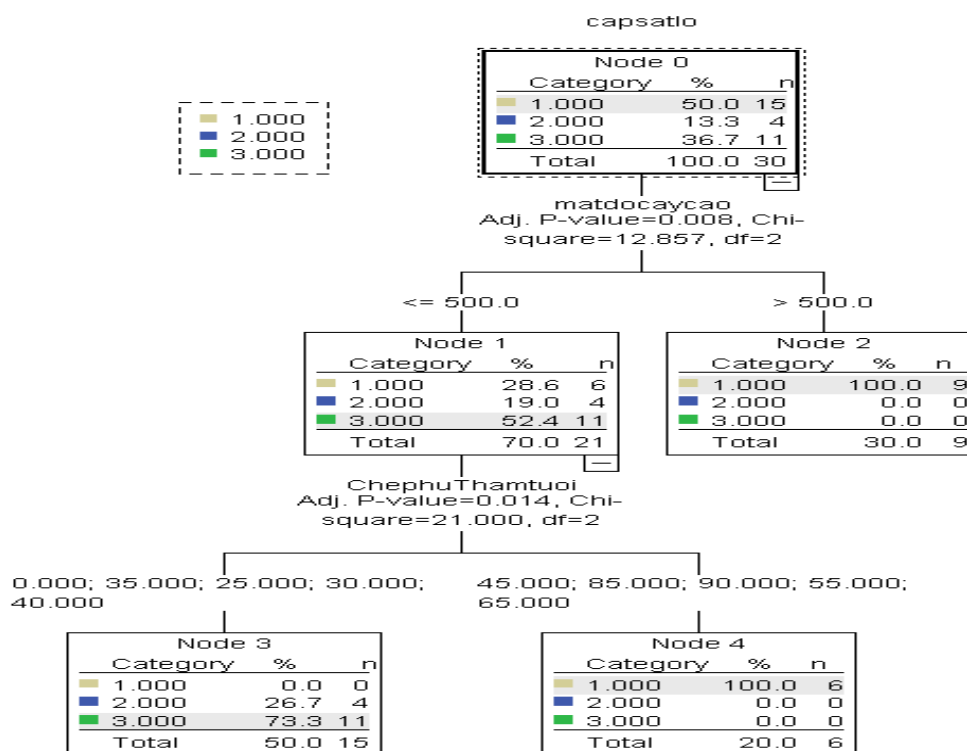
Kết quả phân tích cho thấy độ dốc (*dodoc*) là yếu tố ảnh hưởng mạnh nhất đến cấp độ sạt lở đất, đứng ở nút gốc cây quyết định với giá trị Chi-square = 28,615 ($p = 0,000$) có ý nghĩa thống kê. Cụ thể, khi độ dốc $\leq 33^\circ$ thì 76,9% các ô thuộc cấp sạt lở thấp (cấp 1); trong khoảng độ dốc từ 34° đến 37° thì 100% các ô thuộc cấp sạt lở 2; khi độ dốc $> 37^\circ$ thì cấp sạt lở cao (cấp 3) chiếm tỷ lệ 69,2%. Trong nhánh độ dốc thấp ($\leq 33^\circ$), yếu tố tiếp theo có ảnh hưởng đến sạt lở là bề dày tầng đất (*bedaydat*), trong đó bề dày đất > 70 cm cho thấy 90,9% sạt lở thuộc cấp 1. Ngược lại, trong cùng cấp độ dốc thấp này nếu bề dày tầng đất dưới 70 cm, thì chủ yếu các OTC thuộc cấp sạt lở 2, 3. Ở một nhánh độ dốc cao ($> 37^\circ$), yếu tố quyết định đến sạt lở là độ cao dưới cành của tầng cây cao (*Hdc*) là yếu tố có ý nghĩa thống kê (Adj. P-value = 0,038). Cụ thể, khi $Hdc \leq 0,7$ m, 100% các ô thuộc cấp sạt

lở 3 (rủi ro rất cao). Ngược lại, khi $Hdc > 0,7$ m tỷ lệ sạt lở cấp 3 giảm mạnh chỉ còn 33,3%, trong khi 66,7% chỉ ở cấp 2. Kết quả này cho thấy vai trò quan trọng của tầng cây cao: khi chiều cao dưới cành lớn hơn 0,7 m đã góp phần giảm đáng kể nguy cơ sạt lở trên các sườn dốc rất cao, thông qua cơ chế gia cố đất bằng hệ rễ và giảm động năng của hạt mưa tác động đến mặt đất. Mô hình đạt độ chính xác phân loại 86,7% trên dữ liệu 30 OTC, tuy nhiên sai số theo kiểm chứng chéo 10-fold là 0,367 (tương ứng với khả năng dự báo chính xác khoảng 63,3%) và tồn tại hiện tượng quá khớp (overfitting) nhẹ do cỡ mẫu nhỏ nên độ tin cậy chỉ ở mức trên trung bình. Kết quả nghiên cứu chỉ rõ độ dốc là yếu tố có ảnh hưởng mạnh nhất và ổn định nhất đến cấp độ sạt lở đất tại khu vực nghiên cứu. Kết quả này hoàn toàn phù hợp với các nghiên cứu trước đây trên thế giới và trong

nước. Khi độ dốc > 30–35° thường được coi là ngưỡng nguy cơ cao, điều này được đề cập trong nghiên cứu của Çellek và cộng sự (2020), Viet Du và cộng sự (2023) [3, 7]. Kết quả mô hình phân tích điểm ở xã Trung Chải có thể làm cơ sở khoa học tham khảo cho các khu vực miền núi phía Bắc có điều kiện địa hình, đất đai tương tự, nhưng cần kiểm chứng thêm bằng dữ liệu thực địa tại khu vực mới trước khi áp dụng.

3.2. Các yếu tố ảnh hưởng đến sạt lở đất tại xã Tú Lệ, tỉnh Lào Cai

Đánh giá các yếu tố ảnh hưởng đến cấp độ sạt lở đất tại xã Tú Lệ tỉnh Lào Cai (đây là xã được sạt nhập từ xã Tú Lệ và xã Cao Phạ, tỉnh Yên Bái cũ) phân tích CHAID tương tự trên 30 OTC với Parent Node = 6 và Child Node = 2. Kết quả phân tích được thể hiện ở cây quyết định như Hình 3.



Risk		
Method	Estimate	Std. Error
Resubstitution	.133	.062
Cross-Validation	.533	.091

Growing Method: CHAID
Dependent Variable: capsatlo

Classification				
Observed	Predicted			Percent Correct
	1	2	3	
1	15	0	0	100.0%
2	0	0	4	0.0%
3	0	0	11	100.0%
Overall Percentage	50.0%	0.0%	50.0%	86.7%

Growing Method: CHAID
Dependent Variable: capsatlo

Hình 3. Cây quyết định CHAID xác định các yếu tố ảnh hưởng đến cấp độ sạt lở đất tại xã Tú Lệ, Lào Cai

Kết quả cho thấy mật độ tăng cây cao (*matdocaycao*) là yếu tố ảnh hưởng mạnh nhất đến cấp độ sạt lở đất với giá trị Chi-square = 12,857 ($p = 0,008$). Cụ thể, khi mật độ cây cao > 500 cây/ha thì 100% các ô tiêu chuẩn thuộc cấp sạt lở thấp (cấp 1). Ngược lại, khi mật độ

cây cao ≤ 500 cây/ha, thì trên 50% số OTC ở cấp sạt lở cao (cấp 3). Trong cùng 1 nhánh mật độ cây cao ≤ 500 cây/ha, độ che phủ cây bụi thảm tươi (*ChephuThamtuoil*) là biến có ảnh hưởng rõ rệt đến sạt lở đất với giá trị Adj. P-value = 0,014, Chi-square = 21, df = 2 với ý

nghĩa thống kê ($p < 0.05$). Trong nhánh này khi độ che phủ thảm tươi ở mức thấp (0; 25; 30; 35; 40%), thì 73,3% các OTC thuộc cấp sạt lở cao (cấp 3); khi che phủ thảm tươi ở mức cao hơn 45% thì không có bất cứ ô nào sạt lở ở cấp 2 và 3, mà tất cả các ô đều thuộc cấp sạt lở 1 (cấp thấp). Như vậy có thể thấy, khi mật độ cây gỗ thấp (≤ 500 cây/ha), thảm tươi trở thành yếu tố quyết định đến cấp độ sạt lở. Điều này có nghĩa là khi mật độ cây gỗ thấp, nhưng nếu duy trì độ che phủ cây bụi thảm tươi trên 40% thì vẫn đảm bảo an toàn để sạt lở không xảy ra hoặc xảy ra cấp thấp. Mô hình đạt độ chính xác phân loại tổng thể 86,7% trên dữ liệu 30 OTC, trong đó dự đoán đúng hoàn toàn cấp 1 và cấp 3, nhưng đối với cấp 2 dự đoán còn chưa khả thi. Điều này có thể giải thích do số lượng ô tiêu chuẩn thuộc cấp sạt lở trung bình (cấp 2) chỉ có 4 mẫu, chiếm tỷ lệ rất nhỏ so với cấp 1 (15 mẫu) và cấp 3 (11 mẫu). Với cỡ mẫu hạn chế như vậy, thuật toán CHAID gặp khó khăn trong việc xác định ngưỡng phân tách ổn định cho nhóm này.

Tuy nhiên, mô hình vẫn đạt độ chính xác 100% đối với hai cấp cực đoan (cấp 1 và cấp 3) – đây là các cấp có ý nghĩa thực tiễn quan trọng nhất đối với công tác cảnh báo và phòng ngừa sạt lở đất. Sai số theo kiểm chứng chéo 10-fold của mô hình tại xã Tú Lệ là khá cao (0,533) do quy mô mẫu nhỏ ($n = 30$). Do đó, độ tin cậy của mô hình chỉ ở mức trung bình. Kết quả nghiên cứu này khá phù hợp với các kết quả nghiên cứu khác trên thế giới. Theo Phillips Chris và cộng sự (2021) thực vật, đặc biệt là cây gỗ, đóng vai trò quan trọng trong việc giảm tần suất và quy mô sạt lở đất tăng nông trên các sườn dốc cao phủ lớp đất phong hóa thông qua cơ chế lực kháng của rễ [1]. Nhóm tác giả cũng chỉ ra rằng ở quy mô ngắn hạn 10-30 năm, che phủ thực vật giúp ổn định sườn dốc không chỉ nhờ rễ mà còn qua tác động đến cơ chế thủy văn (cơ chế nước giữ lại trên tán, bốc thoát hơi...). Vì vậy, khi cây gỗ bị loại bỏ do khai thác, cháy rừng hoặc sâu bệnh hại sẽ xuất hiện tổn thương lỗ trống liên quan đến sạt lở đất kéo dài khoảng 3–20 năm, khiến tỷ lệ sạt lở tăng mạnh (2–10 lần) trước khi rễ của rừng mới phát triển. Các kết quả đánh giá về vai trò của cây gỗ cũng được khẳng định rất rõ ràng trong nghiên cứu lựa chọn các yếu tố để dự báo sạt lở đất bằng các mô hình học máy của Chen và cộng sự (2023) [8]. Kết quả nghiên cứu này chứng minh rằng độ che phủ cây bụi, thảm tươi có tác dụng

bảo vệ đất và quan trọng; ngay cả khi mật độ cây cao không đạt, việc duy trì che phủ cây bụi thảm tươi $\geq 45\%$ vẫn có thể đưa toàn bộ các OTC về mức sạt lở thấp nhất (cấp 1).

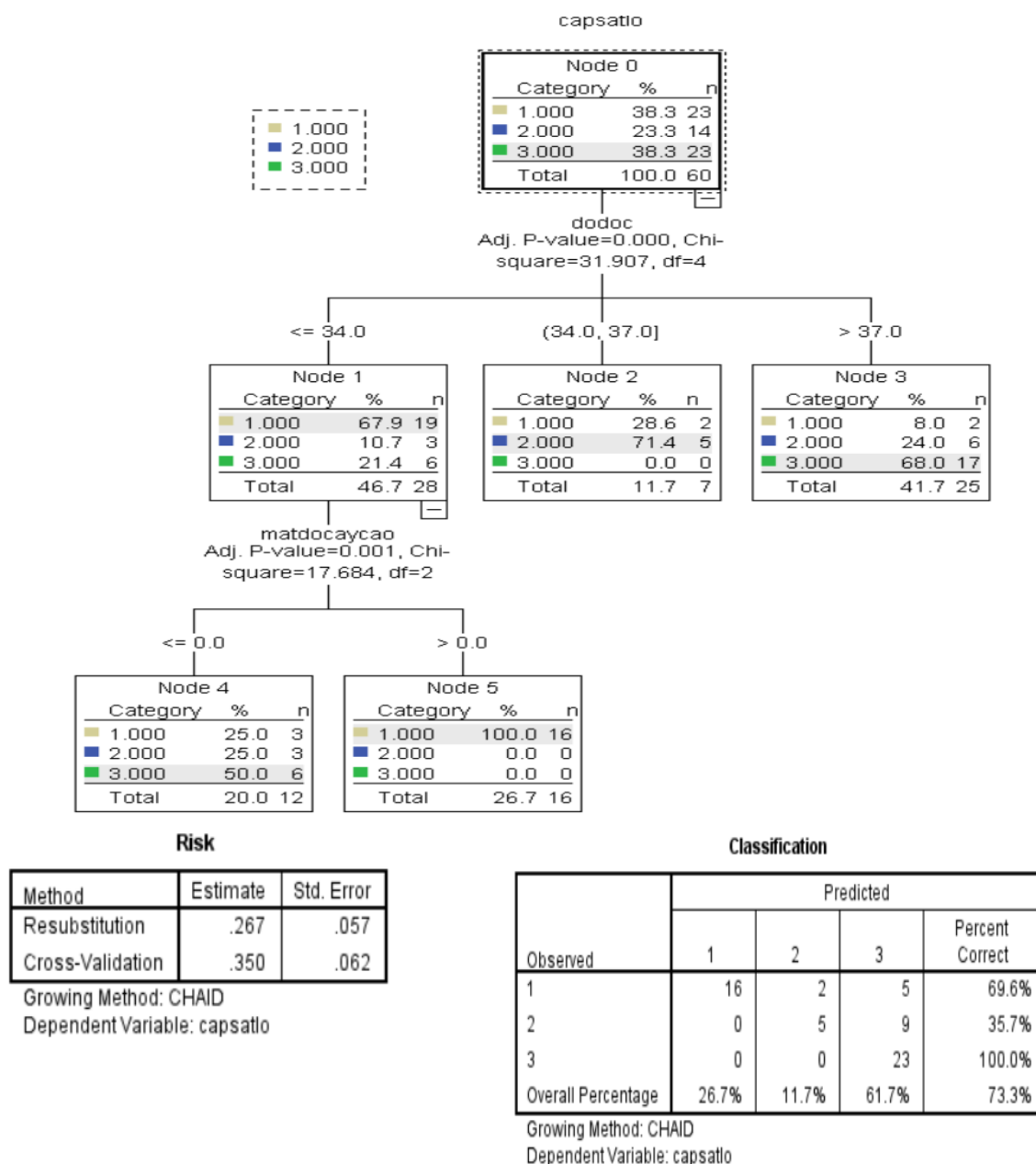
3.3. Kết quả phân tích tổng hợp xác định các yếu tố ảnh hưởng đến sạt lở đất tại khu vực nghiên cứu

Phân tích được thực hiện trên tổng hợp 60 OTC (gộp xã Trung Chải và xã Tú Lệ) với phương pháp CHAID, thiết lập Parent Node = 6, Child Node = 2. Kết quả cho thấy mô hình chỉ giữ lại 2 biến độc lập có ý nghĩa là độ dốc và mật độ cây cao (Hình 4).

Kết quả nghiên cứu cho thấy độ dốc là yếu tố ảnh hưởng mạnh nhất đến cấp độ sạt lở đất tăng nông, với ngưỡng nguy cơ thấp khi độ dốc $\leq 34^\circ$, trung bình ở khoảng độ dốc $34^\circ\text{--}37^\circ$ và cao nhất khi độ dốc $> 37^\circ$. Kết quả này hoàn toàn phù hợp với nghiên cứu của Çellek và cộng sự (2020), nhóm đã xác định khu vực có độ dốc cao là vùng có nguy cơ sạt lở cao nhất do lực cắt tăng mạnh dưới tác động của trọng lực [3]. Kết quả phân tích CHAID trên cũng cho thấy rõ trong cùng nhóm độ dốc thấp $\leq 34^\circ$, biến mật độ cây cao (*matdocaycao*) tiếp tục phân tách rõ ràng và có ý nghĩa thống kê cao (Adj. P-value = 0,001). Cụ thể, các ô có mật độ cây > 0 đạt 100% ở cấp sạt lở 1 (rủi ro thấp), không có bất kỳ ô nào rơi vào cấp 2 hay cấp 3. Ngược lại, các ô nơi có độ dốc $\leq 34^\circ$ mà không có cây gỗ (*matdocaycao* = 0) có 25% OTC ở cấp 1, 25% OTC ở cấp 2 và có tới 50% OTC ở cấp 3 (rủi ro cao). Kết quả này khá tương đồng với kết quả nghiên cứu của Murgia và cộng sự (2022), Asada và cộng sự (2023) và Lorella Montrasio và cộng sự (2023) về cơ chế tăng cường sức kháng trượt nhờ hệ rễ cây gỗ trong nghiên cứu về sạt lở đất [2, 9, 10]. Kết quả này cũng góp phần định lượng thêm các nhận định về yếu tố ảnh hưởng đến sạt lở đất đã được trình bày trong nghiên cứu gần đây của nhóm tác giả Kiều Thị Dương và cộng sự (2025) [11]. Ngoài ra, kết quả nghiên cứu này tương đồng với kết luận của Viet Du và cộng sự (2023) khi sử dụng deep learning đánh giá tác động của thay đổi lớp phủ đất đến sạt lở đất tại lưu vực sông Trà Khúc ở Việt Nam về ảnh hưởng của sự chuyển đổi từ rừng sang đất trồng hoặc nông nghiệp làm tăng mạnh nguy cơ sạt lở [7]. Có thể thấy vai trò của thảm thực vật, là một thành phần quan trọng đóng vai trò then chốt trong việc thay đổi đặc tính đất và ảnh hưởng đến độ ổn định tổng thể

của sườn dốc... Nghiên cứu này đã chỉ rõ vai trò của thảm thực vật bao gồm cả cây gỗ và cây bụi thảm tươi như một lá chắn tự nhiên giúp giảm sạt lở đất, tuy nhiên cây gỗ có tác dụng chủ yếu bởi tăng lực gia cố của rễ cây đâm sâu xuống đất, trong khi cây bụi thảm tươi lại có tác dụng bởi khả năng giảm động năng của nước mưa và điều tiết khả năng thấm nước của đất. Mặc dù

một số biến thực vật hoặc biến liên quan đến đất có tương quan với nhau (ví dụ mật độ cây cao và độ tàn che; độ ẩm và độ pH của đất), mô hình CHAID vẫn chỉ giữ lại độ dốc và mật độ cây cao là hai biến có ý nghĩa thống kê cao nhất, cho thấy thuật toán đã loại bỏ hiệu quả các biến dư thừa.



Hình 4. Cây quyết định CHAID tổng hợp xác định các yếu tố ảnh hưởng đến cấp độ sạt lở đất tại 2 xã Tú Lệ và Trung Chải, Lào Cai

Kết quả CHAID này khẳng định mạnh mẽ vai trò của rừng trong việc giảm thiểu sạt lở đất tầng nông. Ngay cả khi độ dốc thấp ($\leq 34^\circ$), việc mất rừng sẽ làm tăng gấp nhiều lần nguy cơ sạt lở (từ 0% ở cấp 3 lên đến 50% cấp 3). Đây là bằng chứng thực nghiệm khoa học để lý giải ý

nghĩa và vai trò của các giải pháp bảo vệ và phục hồi rừng đối với bảo vệ đất, chống sạt lở tại các khu vực đồi núi của Việt Nam.

Mặc dù mô hình phân tích trên toàn bộ 60 ô tiêu chuẩn cho thấy ngưỡng phân tách độ dốc ở 34° , nhóm tác giả chọn sử dụng giá trị 33° làm

ngưỡng đại diện trong phần Tóm tắt và Kết luận. Lý do là ngưỡng này mang tính phù hợp với mục tiêu tăng cường an toàn trong công tác cảnh báo sớm sạt lở đất. Sự chênh lệch 1° nằm trong phạm vi sai số đo đạc thực địa và không làm thay đổi ý nghĩa thực tiễn của các quy luật đã phát hiện được.

4. KẾT LUẬN

Nghiên cứu ứng dụng thuật toán cây quyết định CHAID để xác định các yếu tố ảnh hưởng và ngưỡng giá trị của chúng đến cấp độ sạt lở đất tầng nông tại xã Trung Chải và xã Tú Lệ, tỉnh Lào Cai. Kết quả cho thấy, độ dốc là yếu tố có ảnh hưởng mạnh mẽ nhất và ổn định nhất ở cả hai khu vực nghiên cứu. Tại xã Trung Chải, các yếu tố tiếp theo có ý nghĩa là bề dày tầng đất và chiều cao dưới cành của tầng cây cao, trong khi tại xã Tú Lệ, mật độ cây cao và che phủ cây bụi, thảm tươi đóng vai trò quan trọng hơn.

Phân tích trên toàn bộ 60 OTC cho thấy mô hình CHAID xác định được các ngưỡng giá trị có ý nghĩa thực tiễn đối với cấp độ sạt lở, đồng thời khẳng định vai trò của thảm thực vật. Đặc biệt ngay cả ở những khu vực có độ dốc thấp, sự hiện diện của tầng cây gỗ, cây bụi và thảm tươi vẫn đóng vai trò quan trọng trong việc giảm thiểu nguy cơ sạt lở thông qua cơ chế gia cố hệ rễ và giảm động năng của nước mưa. Kết quả nghiên cứu nhấn mạnh rằng việc bảo vệ và phục hồi rừng là giải pháp hiệu quả để giảm thiểu sạt lở đất tầng nông tại khu vực đồi núi của Việt Nam.

Việc ứng dụng thuật toán CHAID cho phép xác định thứ tự ảnh hưởng và ngưỡng giá trị của các yếu tố một cách định lượng và trực quan. Kết quả có tiềm năng vận dụng trong công tác cảnh báo, quy hoạch phòng ngừa sạt lở. Tuy nhiên, do quy mô mẫu còn hạn chế, sai số còn cao, các kết quả này cần được kiểm chứng thêm trên quy mô rộng hơn, bộ dữ liệu lớn hơn để nâng cao độ tin cậy và khả năng khái quát hóa trước khi áp dụng rộng rãi. Bên cạnh đó, sự chênh lệch tỷ lệ không gian giữa bản đồ đất, lưới ERA5-Land với quy mô OTC 100 m² có thể dẫn đến sai lệch khi gán giá trị các biến thứ cấp cho từng OTC. Đồng thời việc chưa thực hiện kiểm tra độ tin cậy giữa các điều tra viên đối với biến phụ thuộc (cấp độ sạt lở) cũng là hạn chế cần được khắc phục trong các nghiên cứu tiếp theo.

TÀI LIỆU THAM KHẢO

- [1]. Phillips C., Hales T., Smith H. & Basher L. (2021). Shallow landslides and vegetation at the catchment scale: A perspective. *Ecological Engineering*. 173: 106436. DOI: 10.1016/j.ecoleng.2021.106436.
- [2]. Lorella Montrasio, Michele Placido Antonio Gatto & Chiara Miodini (2023). The role of plants in the prevention of soil-slip: the G-SLIP model and its application on territorial scale through G-XSLIP platform. *Landslides*. 20: 1149-1165. DOI: 10.1007/s10346-023-02031-9.
- [3]. Çellek S. (2020). Effect of the slope angle and its classification on landslide. *Natural Hazards and Earth System Sciences*. 20(12): 3251-3265. DOI: 10.5194/nhess-2020-87.
- [4]. Law Y. K., Lee C. K., Chan A. H., Mak N. P., Hau B. C. & Wu J. (2024). Unveiling the role of forests in landslide occurrence, recurrence and recovery. *Journal of Applied Ecology*. 61(9): 2033-2046. DOI: 10.1111/1365-2664.14741.
- [5]. Li Y. & Duan W. (2023). Decoding Vegetation's Role in Landslide Susceptibility Mapping: An Integrated Review of Techniques and Future Directions. *Biogeotechnics*. 2(1): 100056. DOI: 10.1016/j.bgtech.2023.100056.
- [6]. Nagarajan R., Roy A., Vinod Kumar R., Mukherjee A. & Khire M. V. (2000). Landslide hazard susceptibility mapping based on terrain and climatic factors for tropical monsoon regions. *Bulletin of Engineering Geology and the Environment*. 58(4): 275-287. DOI: 10.1007/s100649900032.
- [7]. Quan Vu Viet Du, Huu Duy Nguyen, Viet Thanh Pham, Cao Huan Nguyen, Quoc Huy Nguyen, Quang Thanh Bui, Thanh Thuy Doan, Anh Tuan Tran & Alexandru-Ionut Petrisor (2023). Deep learning to assess the effects of land use/land cover and climate change on landslide susceptibility in the Tra Khuc river basin of Vietnam. *Geocarto International*. 38(1): 2172218. DOI: 10.1080/10106049.2023.2172218.
- [8]. Chen C. & Fan L. (2023). Selection of contributing factors for predicting landslide susceptibility using machine learning and deep learning models. *Stochastic Environmental Research and Risk Assessment*. 39: 4169-4194. DOI: 10.1007/s00477-023-02556-4.
- [9]. Asada H. & Minagawa T. (2023). Impact of Vegetation Differences on Shallow Landslides: A Case Study in Aso, Japan. *Water*. 15(18): 3193. DOI: 10.3390/w15183193.
- [10]. Murgia I., Giadrossich F., Mao Z., Cohen D., Capra G. F. & Schwarz M. (2022). Modeling shallow landslides and root reinforcement: A review. *Ecological Engineering*. 181: 106671. DOI: 10.1016/j.ecoleng.2022.106671.
- [11]. Kieu Thi Duong, Bui Xuan Dung, Tran Thi Dang Thuy, Nguyen Thi Thanh An, Ngo Tai Anh & Nguyen Duc Phuong (2025). Status of landslides and influencing factors in Bao Yen district, Lao Cai province and Quan Hoa district, Thanh Hoa province: A community-based survey. *Journal of Forestry Science and Technology*. 10(2): 61-71. DOI: 10.55250/Jo.vnuf.10.2.2025.061-071.