

**BỘ NÔNG NGHIỆP VÀ MÔI TRƯỜNG
VIỆN KHOA HỌC THỦY LỢI VIỆT NAM**

**BÁO CÁO SẢN PHẨM 3
BÁO CÁO ĐÁNH GIÁ CÁC PHƯƠNG PHÁP
TIỀM NĂNG ỨNG DỤNG TRÍ TUỆ NHÂN TẠO
TRONG PHÂN VÙNG LŨ QUÉT**

**ĐỀ TÀI: NGHIÊN CỨU ỨNG DỤNG TRÍ TUỆ NHÂN TẠO
VÀ DỮ LIỆU ĐỊA KHÔNG GIAN ĐỂ PHÂN VÙNG LŨ
QUÉT QUY MÔ CẤP HUYỆN**

Cơ quan chủ quản: Bộ Nông nghiệp và Môi trường
Tổ chức chủ trì: Viện Khoa học Thủy lợi Việt Nam
Chủ nhiệm: Lê Văn Thìn
Thời gian thực hiện: 01/2023÷06/2025

HÀ NỘI - 2025

MỤC LỤC

CHƯƠNG 1. TỔNG QUAN NGHIÊN CỨU XÁC ĐỊNH LŨ QUÉT VÀ ỨNG DỤNG TRÍ TUỆ NHÂN TẠO TRONG PHÂN VÙNG LŨ QUÉT	1
1.1. Sơ lược về trí tuệ nhân tạo và ứng dụng trí tuệ nhân tạo trong nghiên cứu lũ quét	1
1.1.1 Trí tuệ nhân tạo	1
1.1.2 Ứng dụng trí tuệ nhân tạo trong nghiên cứu lũ quét	2
1.2. Các nghiên cứu về phân vùng lũ quét	3
1.2.1 Các nghiên cứu trên thế giới	3
1.2.2 Các nghiên cứu ở Việt Nam	20
1.3. Các sản phẩm chính từ kết quả phân vùng lũ quét sử dụng trí tuệ nhân tạo và dữ liệu địa không gian	25
1.3.1 Bản đồ phát hiện lũ quét	25
1.3.2 Bản đồ nhạy cảm với lũ quét	26
1.3.3 Bản đồ nguy cơ lũ quét	27
1.4. Đánh giá các phương pháp xác định lũ quét và dữ liệu sử dụng	29
1.4.1 Các phương pháp phổ biến xác định lũ quét	29
1.4.2 Dữ liệu sử dụng	32
1.4.3 Sự khác biệt dữ liệu giữa học máy và học sâu	33
CHƯƠNG 2. PHƯƠNG PHÁP PHÂN VÙNG LŨ QUÉT ỨNG DỤNG TRÍ TUỆ NHÂN TẠO VÀ DỮ LIỆU ĐỊA KHÔNG GIAN	35
2.1. Phương pháp thu thập dữ liệu	35
2.2. Phương pháp nghiên cứu GIS	35
2.3. Phương pháp học máy	35
2.4. Phương pháp học sâu	40
2.5. Một số kỹ thuật tối ưu hóa mô hình trí tuệ nhân tạo	42
CHƯƠNG 3. ĐỀ XUẤT PHƯƠNG PHÁP TỐI ƯU ỨNG DỤNG TRÍ TUỆ NHÂN TẠO CHO PHÂN VÙNG RỦI RO LŨ QUÉT Ở KHU VỰC MÙ CANG CHẢI	43
3.1. Lựa chọn khu vực nghiên cứu	43
3.2. Phân tích, đề xuất phương pháp tối ưu ứng dụng trí tuệ nhân tạo cho phân vùng lũ quét	44

3.2.1 Chi tiết các thuật toán và lựa chọn tham số tối ưu cho mô hình	45
3.2.2 Tổng hợp các thuật toán và lựa chọn tham số tối ưu cho mô hình.....	54
3.3. Phân tích, lựa chọn mô hình trí tuệ nhân tạo trong phân vùng lũ quét.....	56
3.3.1 Đánh giá các mô hình trí tuệ nhân tạo và đề xuất lựa chọn mô hình. .	56
3.3.2 Đánh giá tổng hợp và lựa chọn mô hình xác định lũ quét	59
KẾT LUẬN.....	61
TÀI LIỆU THAM KHẢO	63

CHƯƠNG 1. TỔNG QUAN NGHIÊN CỨU XÁC ĐỊNH LŨ QUÉT VÀ ỨNG DỤNG TRÍ TUỆ NHÂN TẠO TRONG PHÂN VÙNG LŨ QUÉT

1.1. Sơ lược về trí tuệ nhân tạo và ứng dụng trí tuệ nhân tạo trong nghiên cứu lũ quét

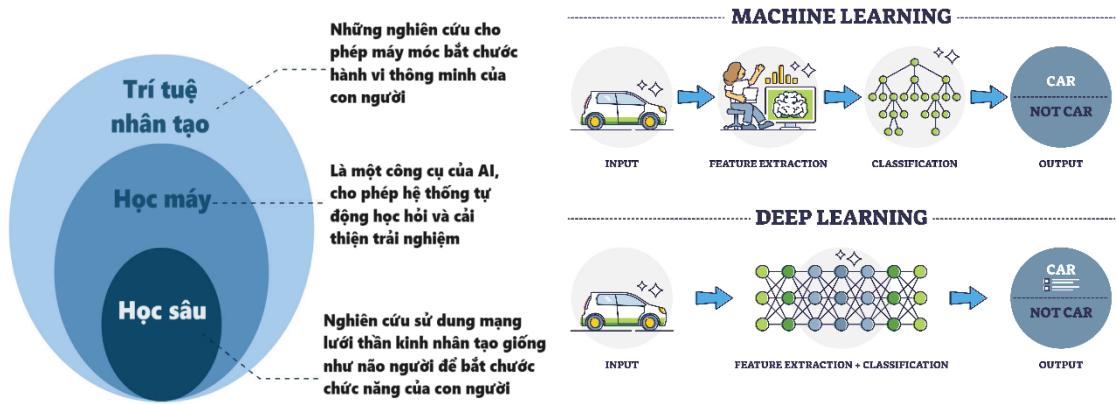
1.1.1 Trí tuệ nhân tạo

Trí tuệ nhân tạo (AI) là lĩnh vực khoa học máy tính chuyên giải quyết các vấn đề nhận thức thường liên quan đến trí tuệ con người, chẳng hạn như học tập, sáng tạo và nhận diện hình ảnh. Các tổ chức hiện đại thu thập vô số dữ liệu từ nhiều nguồn khác nhau như cảm biến thông minh, nội dung do con người tạo, công cụ giám sát và nhật ký hệ thống. Mục tiêu của AI là tạo ra các hệ thống tự học có thể tìm ra ý nghĩa của dữ liệu. Sau đó, AI áp dụng kiến thức thu được để giải quyết các vấn đề mới theo cách giống như con người.

Có thể hiểu một cách đơn giản, trí tuệ nhân tạo chỉ đơn giản là quá trình học tập, từ đó đưa ra phán đoán với những thông tin tương tự những thông tin đã được học tập. Ví dụ như để mô hình trí tuệ nhân tạo có thể đoán được giá nhà tại một khu vực, chỉ với điều kiện dữ liệu đầu vào bao gồm: (1) vị trí; (2) tên chủ đầu tư; (3) vị trí số tầng; (4) diện tích; và (5) tiện ích. Mô hình cần có đủ dữ liệu (có thể lên tới hàng ngàn dữ liệu) các căn hộ tương tự để học, sau đó dự đoán giá nhà cho một nhà bất kỳ với đủ 5 yếu tố trên. Việc này hoàn toàn tương tự như việc cung cấp tư vấn dịch vụ bất động sản của một nhân viên sale bất động sản. Các nhân viên càng lâu năm, dữ liệu được học càng nhiều thì việc dự đoán giá nhà mới càng tiệm cận giá thực tế (tương tự như mô hình trí tuệ nhân tạo, với dữ liệu học càng lớn, mô hình được huấn luyện càng tốt hơn).

Đây là một ví dụ điển hình để có thể hiểu được cách hoạt động của trí tuệ nhân tạo. Trên thực tế, khi càng có nhiều dữ liệu, giá nhà càng có sự biến động mạnh làm cho ngay cả những người làm việc lâu năm cũng khó có thể đưa ra giá tiệm cận do độ phức tạp của dữ liệu. Điều này dẫn đến các thuật ngữ như Overfitting hay Underfitting... Do đó, một lượng dữ liệu vừa đủ và “đáng tin cậy” sẽ tốt hơn rất nhiều việc đưa quá nhiều các “dữ liệu bẩn” vào để đào tạo mô hình làm suy giảm chất lượng mô hình trí tuệ nhân tạo xây dựng. Người xây dựng mô hình trí tuệ nhân tạo cần hiểu rõ bản chất của các yếu tố phụ thuộc dẫn đến kết quả dự đoán thay vì “tra tấn dữ liệu” để xây dựng mô hình.

Trí tuệ nhân tạo bao gồm các thuật toán học máy (machine learning) và học sâu (deep learning) đều được hoạt động theo nguyên tắc sử dụng dữ liệu đầu vào → thông qua thuật toán → dữ liệu đầu ra. Học máy cũng tương tự như “mô hình hộp trắng” và học sâu có những nét tương tự như “mô hình hộp đen”. Do đó về bản chất, học máy có thể được giải thích thông qua các thuật toán một cách tương đối rõ ràng (có thể hiện sự logic bên trong và có thể kiểm tra được), còn học sâu thì mang tính trừu tượng hơn và có những tham số ẩn.



Hình 1. Sự khác biệt giữa học máy và học sâu

Phương pháp học máy đòi hỏi người sử dụng có những hiểu biết nhất định và định hướng thuật toán (như dự đoán dạng hàm số, kiểu phân bố dữ liệu), học máy sử dụng nhiều loại thuật toán khác nhau, trong khi đó học sâu lại tập trung vào mạng nơ-ron nhân tạo. Để sử dụng được mô hình học máy, người dùng cần trích xuất những đặc trưng theo hiểu biết (ví dụ như độ dốc từ địa hình...) và đưa vào mô hình, nhưng học sâu có thể tự động học những đặc trưng phức tạp từ dữ liệu đầu vào. Do đó, học sâu cũng yêu cầu nhiều dữ liệu hơn và tài nguyên tính toán mạnh mẽ hơn. Học máy cũng yêu cầu dạng dữ liệu đầu vào là vector (dạng số/matrice...) đã được chuẩn hóa/xử lý, tuy nhiên học sâu có thể nhận dữ liệu thô với sự đa dạng hơn (âm thanh, hình ảnh, số liệu...)

1.1.2 *Ứng dụng trí tuệ nhân tạo trong nghiên cứu lũ quét*

AI có tiềm năng to lớn trong nhiều lĩnh vực, bao gồm y tế, tài chính, giáo dục, và môi trường. Một trong những ứng dụng đáng chú ý của AI là trong nghiên cứu và dự đoán lũ quét. Lũ quét là lũ xảy ra bất ngờ trên sườn dốc và trên các sông suối nhỏ miền núi, dòng chảy xiết, thường kèm theo bùn đá, lũ lên nhanh, xuống nhanh, có sức tàn phá lớn [1]. Đây là một loại hình thiên tai khó dự báo và gây nhiều thiệt hại cho khu vực miền núi nước ta.

Nói như vậy có nghĩa rằng lũ quét là một dạng lũ, do đó, nó là kết quả phản ứng thủy văn của một lưu vực (được xác định từ mưa → dòng chảy). Điều này dẫn đến quá trình xây dựng mô hình trí tuệ nhân tạo để dự đoán lũ quét tại một vị trí thì vị trí đó cần phải có các thông số đại diện cho lưu vực mà cửa ra chính là điểm dự đoán. Nếu đưa các thông số chỉ đại diện cho các điểm (thông số nội tại), việc dự đoán lũ quét có thể gây ra sai lệch rất lớn (ví dụ một điểm trên mái nhà có các yếu tố bất lợi như độ dốc lớn, không có thảm phủ...). Bên cạnh đó, các thông số được đưa vào cũng cần phải là các thông số ảnh hưởng trực tiếp đến quá trình hình thành dòng chảy và mức độ sẵn có của dữ liệu.

Bên cạnh ứng dụng dự đoán (tính nhạy cảm/nguy cơ) lũ quét, mô hình trí tuệ nhân tạo cũng được sử dụng để phát hiện lũ quét (nhận dạng lũ) từ ảnh viễn thám. Về bản

chất, đây là quá trình phát hiện “nước” trên bề mặt giữa hai thời kỳ trước và sau lũ để so sánh và phát hiện khu vực bị ngập lũ. Ứng dụng này ban đầu sử dụng cho phát hiện lũ và ngập lũ, sau đó triển khai linh hoạt để ứng dụng trong phát hiện lũ quét.

Nghiên cứu này tin rằng với sự kết hợp của các tham số địa không gian (từ ảnh vệ tinh và quan sát bề mặt) và lượng mưa, mô hình trí tuệ nhân tạo có thể phát hiện ra được nguy cơ lũ quét một cách đáng tin cậy.

1.2. Các nghiên cứu về phân vùng lũ quét

Phân vùng lũ quét đã được nhiều nhà nghiên cứu thực hiện nhằm tìm ra các khu vực chịu tác động bởi lũ quét, về cơ bản, phân vùng lũ quét được chia thành hai loại: phân vùng theo điểm và phân vùng theo lưu vực.

Phân vùng lũ quét theo điểm là lập một bản đồ không gian các vị trí chịu tác động bởi lũ quét dựa trên đặc điểm không gian của điểm đó. Thông thường, quá trình phân vùng theo điểm được xây dựng bằng phương pháp chồng chập đa nhân tố như: Phương pháp phân tích thứ bậc (AHP) [2, 3]; phương pháp tỷ số tần suất (FR) [4]; phương pháp trọng số dẫn chứng (WOE) [5]; phương pháp học máy (ML) [6]; phương pháp học sâu (DL) [7]...

Phân vùng lũ quét theo lưu vực thường được tiếp cận dưới góc độ thủy văn, là quá trình xác định các yếu tố đặc trưng của dòng chảy lũ (từ mô phỏng mưa – dòng chảy) như lưu lượng đỉnh lũ, vận tốc dòng chảy, thời gian tập trung dòng chảy... để xác định lũ quét tại cửa ra của một lưu vực. Khi dòng chảy thuộc lưu vực có khả năng là lũ quét (vượt ngưỡng quy định nào đó) thì lưu vực đó được cảnh báo có nguy cơ lũ quét. Một trong những phương pháp điển hình của loại hình phân vùng lũ quét theo lưu vực là phương pháp FFG (Flash flood guidance) trong hệ thống FFGS (Flash flood guidance system) [8].

1.2.1 Các nghiên cứu trên thế giới

Mặc dù các nghiên cứu về lũ quét có từ rất sớm, các nghiên cứu về phân vùng lũ quét lại khởi đầu muộn hơn. Trước năm 2000, hầu hết các nghiên cứu về lũ quét tập trung vào đánh giá các trận lũ quét [9, 10, 11] và tính toán thủy văn cho các vị trí có khả năng xảy ra lũ quét [12, 13]. Đến năm 1992, Sweeney viết hướng dẫn hiện đại hóa lũ quét lưu vực (Modernized Areal Flash Flood Guidance) và đăng tải trên bản ghi nhớ hướng dẫn kỹ thuật NWS HYDRO 44 [14]. Trong đó mô tả đầy đủ các khái niệm cũng như hướng dẫn các bước tính toán xác định ngưỡng mưa sinh lũ. Khái niệm FFG (Flash Flood Guidance) cũng được hình thành (mặc dù lưu lượng tràn bờ đã được Emmett nghiên cứu trước đó gần 20 năm [15]), mô tả về một lượng mưa bình quân đủ để sinh ra lũ lụt trên các sông suối nhỏ. Cơ sở của FFG là tính toán các giá trị ngưỡng mưa hiệu quả trong một khoảng thời gian cần thiết để gây ra lũ, lượng mưa hiệu quả này được

hiểu là lượng mưa dư sau tổn thất do thẩm thấu, giam giữ và bốc hơi trên lưu vực để hình thành dòng chảy tràn trên bề mặt. Sau đó, các ngưỡng FFG được Trung tâm dự báo xác định cụ thể cho từng khu vực.

Một số phân tích khác sử dụng bản đồ lượng mưa tàn suất để dự báo khả năng xuất hiện lũ quét [16] với giả thiết lượng mưa hàng ngày lớn hơn giá trị xác định. Phương pháp này giúp đưa ra khu vực có tiềm năng lũ quét hoàn toàn dựa vào lượng mưa, tuy nhiên chưa xét đến các yếu tố bề mặt. Ngoài ra, sử dụng phân bố mưa có thể gây ra cảnh báo không do vị trí xuất hiện lũ quét có thể nằm ở hạ lưu khu vực mưa lớn – nơi không có lượng mưa phản ánh đúng thực trạng lũ quét.

Năm 1997, Sharada trình bày một phương pháp xác định ảnh hưởng của nguy cơ lũ quét tới tuyến đường sắt tại quận Warangal, Ấn Độ [17] bằng việc kết hợp nhiều yếu tố địa không gian bao gồm (1) Mặt nước; (2) đặc điểm bờ kè; (3) số lượng hồ trong khu vực; (4) mật độ sông/suối; (5) đặc điểm thoát nước; (6) khoảng cách tới đường sắt; (7) sử dụng đất; (8) độ lún; và (9) độ dốc giữa suối và đường sắt. Phương pháp này gần giống với các nghiên cứu chống chập đà nhân tố hiện nay và đưa ra được bản đồ tác động được phân cấp nguy cơ (từ rất thấp đến rất cao) bằng việc sử dụng công cụ GIS.

Từ năm 2000 trở lại đây, các nghiên cứu phân vùng lũ quét đã trở nên dần phổ biến dưới sự hỗ trợ của các công cụ GIS và máy tính hiệu năng cao và đi về hai hướng: (1) phân vùng lũ quét dựa trên nguyên lý thủy văn và các yếu tố liên quan đến thủy văn (bao gồm mưa, kết quả tính toán thủy văn, thủy lực) và (2) phân vùng lũ quét dựa trên các tham số địa không gian. Các nghiên cứu dựa trên dữ liệu địa không gian được phân tích, đánh giá trên hàng loạt các yếu tố được raster hóa cho khu vực nghiên cứu, từ đó phân vùng lũ quét từ nhiều tham số. Trong nghiên cứu này, tác giả phân loại nghiên cứu phân vùng lũ quét theo dữ liệu địa không gian làm 2 loại: (1) sử dụng phương pháp trọng số và (2) sử dụng phương pháp trí tuệ nhân tạo.

1.2.1.1 Phân vùng lũ quét dựa trên lượng mưa

Mưa lớn là một trong những nguyên nhân chính trong quá trình hình thành lũ quét tự nhiên, do vậy, yếu tố lượng mưa được xét đến trong các nghiên cứu về lũ nói chung và lũ quét nói riêng là một hướng đi đúng đắn. Mặc dù vậy, dự báo được lượng mưa một cách chính xác không phải là một vấn đề dễ dàng trong các nghiên cứu.

Borga và cộng sự đã có một rà soát các nghiên cứu ở châu Âu nói riêng và công bố quốc tế nói chung đến năm 2014 về hệ thống cảnh báo lũ quét và lũ bùn đá về các khía cạnh: kỹ thuật sử dụng, khoảng trống nghiên cứu và đề xuất các hướng nghiên cứu tiếp theo [18]. Trong đó, đề xuất tiếp theo bao gồm: xây dựng hệ thống giám sát và dự báo và xác định ngưỡng mưa cho lũ quét. Ghomash và cộng sự đã nghiên cứu tác động của diễn biến lượng mưa đến quá trình hình thành lũ quét cho các khu vực miền núi bán

khô cằn ở lưu vực Kan (Iran) [19]. Các hình thái mưa khác nhau cho ra kết quả khác nhau về đỉnh lũ, đường quá trình lũ và các yếu tố thủy văn.

Alfieri và cộng sự đã trình bày hướng dẫn dự báo lũ quét dựa trên ngưỡng về lượng mưa trong cuốn sổ tay dự báo khí tượng thủy văn vào năm 2015 [20]. Alfieri cho biết, chỉ sử dụng dữ liệu mưa có thể ước tính được thiệt hại do lũ lụt gây ra. Một trong những phương pháp phổ biến là phương pháp chỉ dẫn lũ quét (FFG) trong hệ thống chỉ dẫn lũ quét (FFGS) đang được sử dụng rộng rãi trên toàn thế giới.

Một nghiên cứu khác của Alfieri năm 2011 đã đề xuất hệ thống cảnh báo sớm lũ quét dựa trên các sản phẩm dự báo khí tượng như dự báo thời tiết số trị (NWP), kết hợp NWP-Radar và truyền sóng radar cho các khu vực miền núi ở Catalonia (thuộc Tây Ban Nha) [21]. Nhược điểm của phương pháp này là bỏ qua các điều kiện ban đầu của bề mặt và hiệu chỉnh của lượng mưa.

Phương pháp phổ biến nhất là xây dựng ngưỡng mưa dựa trên quan hệ cường độ - thời gian (I-D) được Caine đề xuất năm 1980 [22] với công thức $I = \alpha D^\beta$ (công thức kinh nghiệm). Trong đó, α và β là các hệ số và D là thời gian mưa (theo giờ), I là cường độ mưa. Trên cơ sở thu thập các trận mưa trong quá khứ, Caine đã tìm ra được ngưỡng mưa $I = 14.82D^{-0.39}$ (với thời đoạn mưa nhỏ hơn 500 giờ) là ngưỡng mưa có khả năng gây ra lũ bùn đá và sạt lở đất.

Năm 2007, tiếp tục nghiên cứu dựa theo Caine, Guzzetti và cộng sự [23] đã tìm ra được ngưỡng mưa $I = 2.28D^{-0.2}$ cho các trận mưa có thời đoạn dưới 48 giờ và $I = 0.48D^{-0.11}$ cho các trận mưa từ 48 giờ đến dưới 1000 giờ. Nghiên cứu này sử dụng cơ sở dữ liệu toàn cầu từ 2.626 trận mưa và 19.800 trạm quan trắc mưa trên toàn thế giới. Con số này thấp hơn khá đáng kể so với nghiên cứu của Caine và làm tăng ngưỡng an toàn về mưa.

Bezak và cộng sự đã xây dựng đặc điểm các đợt mưa cực lớn gây ra sạt lở đất và lũ quét ở Slovenia trong 25 năm [24]. Kết quả phân tích ở các trạm cho thấy công thức của Caine phù hợp nhất trong việc xây dựng ngưỡng mưa cảnh báo.

Turkington và cộng sự đã đưa ra ngưỡng cảnh báo về lượng mưa cho lũ bùn đá và lũ quét ở khu vực miền núi phía Nam nước Pháp (thung lũng Ubaye) là lượng mưa lớn hơn 20mm/ngày hoặc 22mm/4 ngày [25]. Mặc dù trong nghiên cứu có nhiều ngưỡng phân loại khác nhau độc lập (cho các yếu tố khí tượng khác), đây là một ngưỡng mưa quá an toàn để đưa ra cảnh báo và có khả năng cảnh báo không cao trong khi đáp ứng đầy đủ về tiêu chí phân loại cho các trận mưa sinh lũ ở khu vực nghiên cứu.

Ramos Filho và cộng sự đã trình bày một phương pháp cải tiến trong việc sử dụng ngưỡng cường độ mưa cực đại để đánh giá và cảnh báo lũ quét ở bang São Paulo (Brazil) [26] bằng mô hình số mũ có mối quan hệ với cường độ mưa và chỉ số lượng mưa kỳ trước (API) bằng công thức $I = ae^{b \cdot API} + c$. Trong đó: a , b , c là các hằng số cần xác định. Nghiên cứu cũng chỉ ra giá trị của a nằm trong khoảng từ cường độ mưa tối thiểu đến

ba lần cường độ mưa tối đa (của các trận lũ quét đã xảy ra), giá trị b thay đổi trong khoảng từ $-1.00 \div -0.01$ và giá trị c nằm trong khoảng từ cường độ mưa tối thiểu đến cường độ mưa trung bình (của các trận lũ quét đã xảy ra). Kết quả nghiên cứu cho thấy xác suất phát hiện các trận lũ quét trên ngưỡng trên tăng trung bình 14%.

Zhai và cộng sự đã nghiên cứu ngưỡng mưa sinh lũ quét tại ba khu vực miền núi ở miền Nam Trung Quốc (Zhong, Balisi và Yu) dựa trên mô hình thủy văn lũ quét Trung Quốc [27]. Quá trình xác định bao gồm mực nước và lưu lượng sinh lũ quét tại các khu vực, từ đó tính ngược lượng mưa sinh lũ quét. Kết quả nghiên cứu chỉ ra: (1) tại Zhong, lưu lượng sinh lũ quét là $356,2\text{m}^3/\text{s}$ và ngưỡng mưa dao động từ $93 \div 344\text{mm/trận}$; (2) tại Balisi, lưu lượng sinh lũ quét là $544,69\text{m}^3/\text{s}$ và ngưỡng mưa dao động từ $77 \div 246\text{mm/trận}$; (3) tại Yu, lưu lượng sinh lũ quét là $335,2\text{m}^3/\text{s}$ và ngưỡng mưa dao động từ $111 \div 420\text{mm/trận}$.

Yuan và cộng sự cho ra kết quả về ngưỡng mưa tương ứng với điều kiện độ ẩm kỳ trước (ASMC) và sự phân bố mưa [28]. Lượng mưa được tính toán cho các độ ẩm kỳ trước là 0.2 (khô); 0.5 (trung bình); 0.8 (ướt) và ngưỡng mưa tương ứng từ $95 \div 140\text{mm}$; $79 \div 125\text{mm}$; và $73 \div 117\text{mm/6 giờ}$ cho lưu vực Xinxian, huyện Tín Dương, tỉnh Hà Nam, Trung Quốc.

1.2.1.2 Phương pháp phân vùng lũ quét dựa vào thủy văn, thủy lực

Saber và cộng sự đã sử dụng lượng mưa vệ tinh GSMP để lập mô hình dự báo lũ quét cho lưu vực sông Karpuz (Thổ Nhĩ Kỳ) [29], mô hình thủy văn Hydro-BEAM được sử dụng để xác định dòng chảy dựa trên dữ liệu GSMP đã được hiệu chỉnh. Mặc dù đã có sự hiệu chỉnh độ lệch nhưng không có sự đảm bảo rằng dữ liệu mưa vệ tinh có đủ tin cậy để dự báo dòng chảy lũ. Thậm chí, Westerberg và cộng sự [30] trong một nghiên cứu khác đã chỉ ra rằng việc hiệu chỉnh lại lượng mưa vệ tinh không phải lúc nào cũng là một phương pháp đáng tin cậy vì các lỗi vệ tinh có thể dẫn đến những suy luận sai lệch trong quá trình hiệu chỉnh mô hình thủy văn cho các trận lũ tiếp theo.

Silvestro và cộng sự đã sử dụng mô hình thủy văn bán phân bố DriFt kết hợp với lượng mưa thực đo và lượng mưa dự báo trước 2 giờ để dự báo nguy cơ lũ quét cho lưu vực sông Entella ở phía đông vùng Liguria, Ý [31]. Số liệu mưa dự báo được lấy từ các nguồn ECMWF; COSMO-LAMI; BOLAM; MOLOCH. Kết quả tính toán cho thấy, sai lệch % đỉnh lũ dao động từ $-15 \div 43\%$ với khoảng thời gian dự báo trước từ $1:40\text{p} \div 4:40\text{p}$ cho 7 trận lũ quét đã xảy ra trong quá khứ. Nghiên cứu không đánh giá khả năng ứng dụng trong tương lai và có đưa ra một số khuyến nghị khi áp dụng mô hình.

Việc sử dụng các tham số thủy lực để tính toán nguy cơ lũ quét (mực nước, vận tốc...) đã được áp dụng, tuy nhiên về bản chất là các mô hình toán một chiều trên sông. Để có thể nói là mô phỏng lũ quét (với đặc điểm diễn biến nhanh và không chậm biến đổi đều) là chưa phù hợp, tuy nhiên, hầu hết các mô hình này sử dụng chỉ với mục đích

xác định dòng chảy lũ trên sông suối, khi đạt ngưỡng về mực nước, vận tốc...sẽ được xem xét cảnh báo nguy cơ lũ nói chung và lũ quét nói riêng.

Ngưỡng sinh dòng chảy được sử dụng để đánh giá dòng chảy trên sông khi vượt qua ngưỡng cần cảnh báo. Đây là kết quả của việc sử dụng mô hình thủy lực để tính toán các yếu tố về mực nước, vận tốc... từ mưa và các mô hình thủy văn. Các kết quả tính toán sẽ được so sánh trực tiếp với giá trị ngưỡng dòng chảy. Ngưỡng dòng chảy có thể là mực nước lớn nhất đã xảy ra tương ứng với trận lũ quét trong quá khứ, cũng có thể là một giá trị tương ứng với tần suất khi có quan trắc trong nhiều năm về mực nước (phương pháp thống kê) [32].

Trong hệ thống FFG, việc xác định lưu lượng tràn bờ được sử dụng bằng phương pháp thủy lực mặt cắt, từ đó đưa ra ngưỡng tràn bờ tại các vị trí cửa ra của lưu vực tính toán. Một số các nghiên cứu khác có thể sử dụng ngưỡng tràn bờ bằng công thức kinh nghiệm [33].

1.2.1.3 Phân vùng lũ quét theo trọng số dữ liệu địa không gian

Phần lớn các kết quả nghiên cứu về lũ quét được công bố có liên quan đến đánh giá tính nhạy cảm với lũ quét. Mặc dù là phương pháp đơn giản, các nghiên cứu này lại có lượng dữ liệu lớn và đa dạng. Các loại dữ liệu được sử dụng thông thường bao gồm: (1) Dữ liệu liên quan đến yếu tố địa hình như: độ dốc, độ cong, cao độ...; (2) Dữ liệu liên quan đến yếu tố thủy văn như: khoảng cách đến sông suối, mật độ sông suối, chiều dài dòng chảy...; (3) Dữ liệu liên quan đến thảm phủ bề mặt như: sử dụng đất, chỉ số NDVI...; (4) dữ liệu liên quan đến loại đất như: nhóm đất thủy văn, nhóm đất, kết cấu đất...

Để kết hợp được rất nhiều các yếu tố, các nhà nghiên cứu sử dụng phương pháp chồng chập đa nhân tố để đưa ra bản đồ cuối cùng về nguy cơ. Mỗi loại dữ liệu được xác định cho một hệ số tác động đến kết quả hình thành lũ quét. Các phương pháp sau đây là phổ biến trong việc đánh giá tính nhạy cảm với lũ quét:

1. Phương pháp phân tích thứ bậc

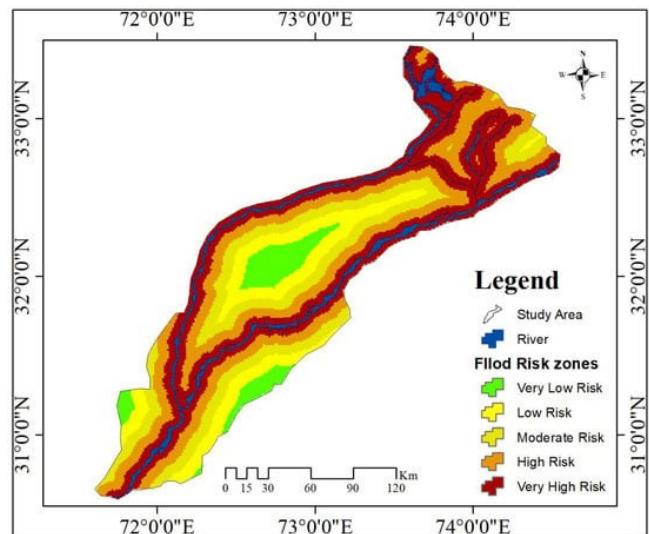
Phương pháp phân tích thứ bậc (AHP) được biết đến là một phương pháp đưa ra phương án quyết định dựa trên đa tiêu chí. Phương pháp này được phát triển bởi Thomas L. Saaty năm 1977 [34]. Bài báo này có số lượt trích dẫn lên tới hơn 5.000 lượt theo thống kê của Crossref (tính đến tháng 9/2023).

Nội dung của phương pháp AHP bao gồm 5 bước: (1) Xác định các yếu tố liên quan và phương án quyết định; (2) Xây dựng ma trận tương quan để đánh giá mức độ quan trọng của các yếu tố liên quan. Trong ma trận này, mỗi ô thể hiện mức độ quan trọng của yếu tố dòng (row) đối với yếu tố cột (column); (3) Xác định trọng số của các yếu tố thể hiện mức độ quan trọng của các yếu tố đó. Trọng số của các yếu tố được tính

toán từ ma trận tương quan; (4) Xây dựng ma trận quyết định để đánh giá mức độ ưu tiên của các phương án quyết định đối với các yếu tố liên quan. Trong ma trận này, mỗi ô thể hiện mức độ ưu tiên của phương án dòng (row) đối với yếu tố cột (column); và (5) Tính toán kết quả là thứ tự ưu tiên của các phương án quyết định. Thứ tự ưu tiên này được tính toán từ ma trận quyết định và trọng số của các yếu tố.

Với ưu điểm là một phương pháp tổng quát, có thể áp dụng cho nhiều loại vấn đề khác nhau, phương pháp này đã được áp dụng trong việc xác định nguy cơ lũ quét với các yếu tố đã được trình bày ở trên. Do sử dụng dữ liệu lớn với phần là dữ liệu về không gian, phương pháp này đang dần phổ biến trong những năm gần đây.

Aquil Tariq và cộng sự đã sử dụng phương pháp AHP trong việc xác định nguy cơ lũ quét cho 2 lưu vực sông Jhelum và Chenab ở Ấn Độ [35] dựa trên 8 yếu tố bao gồm (1) độ dốc; (2) độ cao; (3) mật độ sông suối; (4) Lượng mưa; (5) thảm phủ; (6) khoảng cách đến sông suối; (7) địa chất; và (8) loại đất. Kết quả xây dựng bản đồ và phân tích trọng số được thể hiện ở hình bên.



Bảng 1. Bản đồ ma trận và trọng số quyết định theo AHP

Các yếu tố	SL	E	DD	R	LULC	DS	G	So	SFWV
Độ dốc (SL)	1.00	0.20	5.00	4.00	2.00	1.00	4.00	0.30	0.0152
Độ cao (E)	1.00	0.23	0.90	5.00	1.00	4.00	2.00	0.30	0.0254
Mật độ sông suối (DD)	2.00	5.00	4.00	0.20	1.00	5.00	3.00	1.00	0.2569
Lượng mưa (R)	0.25	0.12	1.00	0.23	0.50	1.00	1.00	0.30	0.0452
Thảm phủ (LULC)	0.10	0.12	0.17	0.45	0.50	0.13	0.25	1.00	0.1400
KC đến sông suối (DS)	1.00	1.00	3.00	5.00	3.00	5.00	4.00	1.00	0.2592
Địa chất (G)	0.34	0.23	0.25	1.00	0.25	0.18	2.00	0.35	0.1190
Loại đất (So)	0.25	4.00	0.25	1.00	2.00	4.00	3.00	1.00	0.1800

Yếu tố lượng mưa trong nghiên cứu này sử dụng là sự chênh lệch lượng mưa so với lượng mưa trung bình nhiều năm tại các trạm (cụ thể trong nghiên cứu này là lượng mưa trung bình từ năm 2010 đến 2015 của 5 trạm quan trắc trong khu vực nghiên cứu).

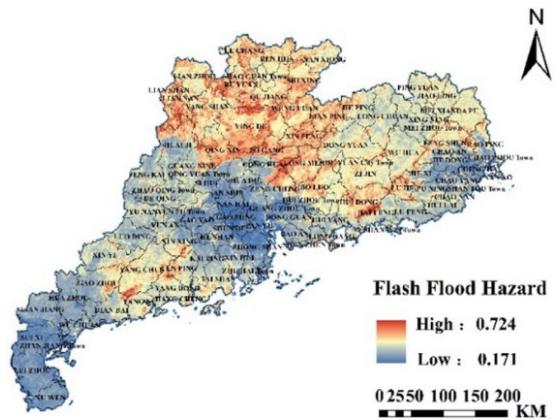
Giá trị phương sai của lượng mưa sẽ được tính toán theo các trạm và được nội suy thành raster mưa dựa trên thuật toán trọng số nghịch đảo IDW.

Kairong Lin và cộng sự cũng sử dụng phương pháp AHP để xác định nguy cơ lũ quét cho tỉnh Quảng Đông ở Trung Quốc [36], Lin chỉ sử dụng 4 yếu tố nguy cơ đầu vào bao gồm (1) cao độ địa hình; (2) độ dốc; (3) lượng mưa; và (4) mật độ sông suối. Kết quả được thể hiện như hình bên. Độ dốc và lượng mưa được xét với trọng số như nhau (khoảng 34%), trong khi đó, cao độ địa hình là gần 20% và mật độ sông suối chiếm gần 12%.

Một biến thể khác của AHP là FAHP (Fuzzy Analytic Hierarchy Process) [3, 37] khi không phân loại các yếu tố thành 9 ngưỡng (từ 1 đến 9) như phương pháp gốc mà thay vì đó là quy đổi về phạm vi từ 0 đến 1. Các tác giả cho rằng FAHP có lợi thế hơn vì nó có thể đáp ứng được sự không chắc chắn trong dữ liệu và phân tích và phù hợp để xác định khu vực dễ bị lũ lụt ở các thành phố, đặc biệt là do sự trùng khớp với các khu vực bị tàn phá nhiều nhất được xác định bằng cách phát hiện sự thay đổi giữa hai hình ảnh Sentinel-2.

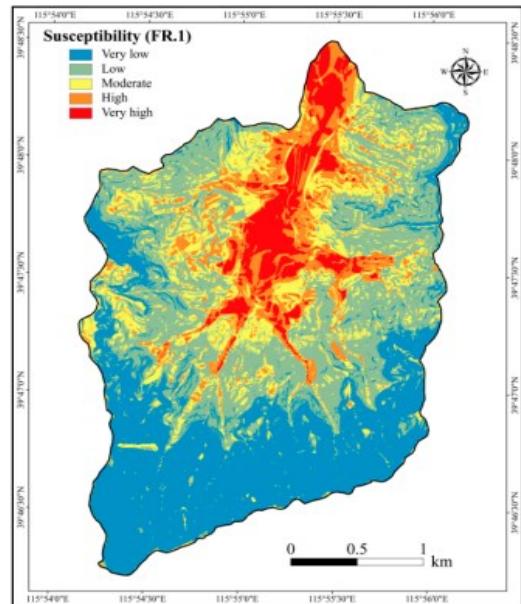
2. Phương pháp tỷ lệ tần suất

Phương pháp tỷ lệ tần suất - Frequency Ratio (FR) là một trong những phương pháp phổ biến đo lường tần suất xuất hiện của lũ quét trong một khu vực nhất định. Các khu vực có chỉ số FR cao hơn thể hiện nguy cơ lũ quét cao hơn. Chỉ số FR là thương của tỷ lệ xuất hiện trên tỷ lệ diện tích. Tương tự như AHP, phương pháp này sử dụng các yếu tố đa dạng về địa không gian.



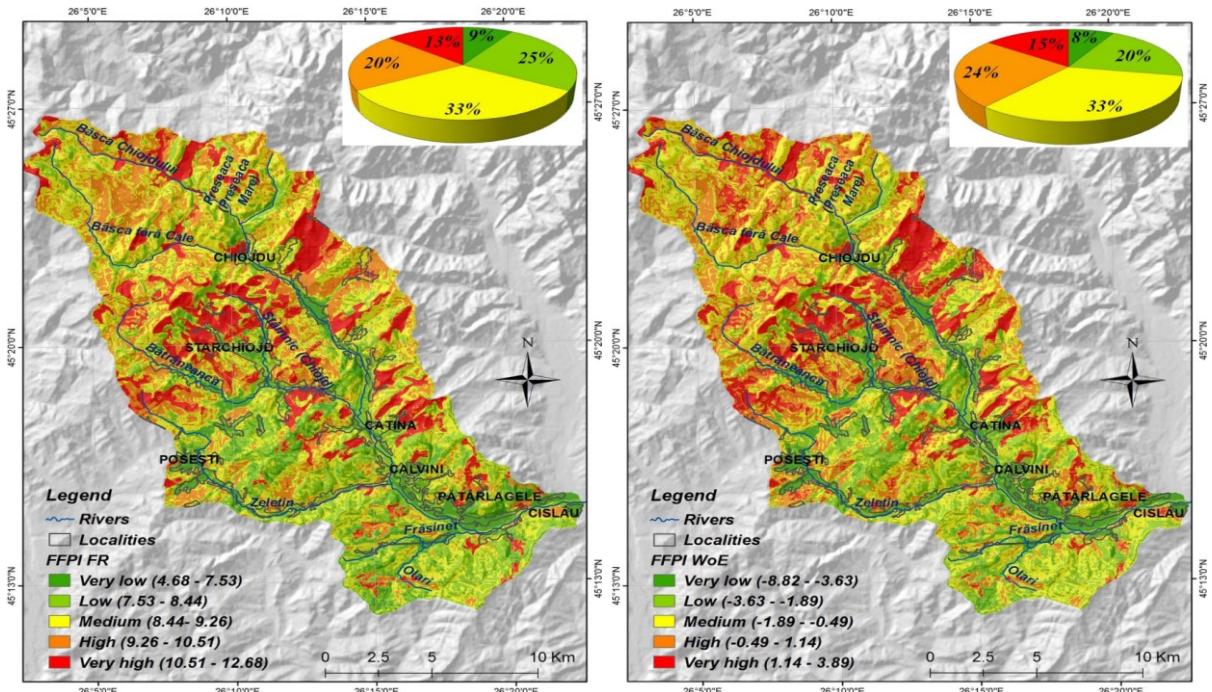
Hình 2. Bản đồ nguy cơ lũ quét tỉnh Quảng Đông – Trung Quốc

Chen Cao đã sử dụng phương pháp FR để xây dựng bản đồ nguy cơ lũ quét cho khu vực Xiqu Gully (thuộc Bắc Kinh – Trung Quốc) dựa trên 85 điểm đã xảy ra lũ quét [4] và 10 yếu tố bao gồm: độ cao; độ dốc; độ cong; sử dụng đất; địa chất; kết cấu đất; khu vực có nguy cơ sụt lún; chỉ số năng lượng dòng chảy; chỉ số ẩm địa hình và mưa lớn ngắn hạn. Kết quả xây dựng bản đồ cho thấy các vị trí có nguy cơ rất cao về lũ quét nằm ở khu vực trũng thấp của địa hình.



Hình 3. Bản đồ nguy cơ lũ quét khu vực Xiqu Gully

Ở Romania, Romulus Costache đã xây dựng bản đồ tiềm năng lũ quét bằng phương pháp FR kết hợp với phương pháp trọng số dẫn chứng (WOE) với 9 yếu tố cho lưu vực sông Bâscă Chiojdului có diện tích khoảng 340km² [38].



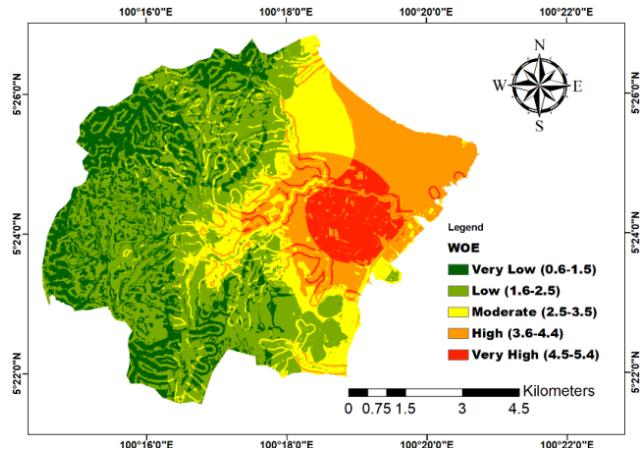
Hình 4. Kết quả xây dựng bản đồ tiềm năng lũ quét lưu vực Bâscă Chiojdului bằng phương pháp FR (bên trái) và WOE (bên phải)

3. Phương pháp trọng số dẫn chứng

Phương pháp trọng số dẫn chứng - Weights Of Evidence (WOE) được sử dụng để đo lường mức độ liên quan của một biến độc lập với một biến phụ thuộc. WOE được

tính bằng cách sử dụng hàm logistic để chuyển đổi các giá trị của biến độc lập thành các giá trị số. $WOE = \ln(p/(1 - p))$, trong đó p là xác suất biến phụ thuộc có giá trị là 1 và $(1-p)$ là xác suất biến phụ thuộc có giá trị là 0. Để tính toán WOE, cần chia dữ liệu thành 2 nhóm, một nhóm có biến phụ thuộc có giá trị là 1 và một nhóm có biến phụ thuộc có giá trị là 0. Sau đó, tính xác suất của biến phụ thuộc có giá trị là 1 cho từng nhóm. Cuối cùng, sử dụng công thức trên để tính WOE cho từng giá trị của biến độc lập.

A Saleh đã sử dụng phương pháp WOE để xây dựng bản đồ nhạy cảm với lũ quét cho lưu vực Sungai Pinang ở Malaysia dựa trên 6 yếu tố bao gồm: độ cao, độ dốc, lượng mưa, độ che phủ đất, khoảng cách từ sông và thạch học [5]. Kết quả xây dựng bản đồ cho độ tin cậy lên tới 0,839.



Hình 5. Bản đồ nhạy cảm lũ quét lưu vực Sungai Pinang

1.2.1.4 Phân vùng lũ quét ứng dụng trí tuệ nhân tạo

Trí tuệ nhân tạo được ứng dụng rộng rãi trong phân vùng lũ quét những năm gần đây. Nghiên cứu này phân tích hơn 300 nghiên cứu về lũ quét trong những năm gần đây về vấn đề xây dựng các bản đồ nhạy cảm và bản đồ nguy cơ lũ quét trên cơ sở dữ liệu của CrossRef. Bảng sau đây tổng hợp một số mô hình trí tuệ nhân tạo (bao gồm học máy và học sâu) thường được sử dụng trong phân vùng lũ quét.

Bảng 2. Danh sách các mô hình phổ biến được sử dụng trong phân vùng lũ quét

TT	Tên	Đặc điểm
I	Mô hình học máy (machine learning)	
1	Hồi quy tuyến tính (Linear Regression)	Đặc điểm: Mô hình hóa mối quan hệ tuyến tính giữa biến đầu vào và đầu ra. Ưu điểm: Đơn giản, dễ hiểu, nhanh chóng. Nhược điểm: Chỉ phù hợp với quan hệ tuyến tính, nhạy cảm với các điểm đột biến. Ứng dụng trong lũ quét: Dự đoán mực nước lũ dựa trên lượng mưa.
2	Hồi quy logistic (Logistic Regression)	Đặc điểm: Dùng cho phân loại nhị phân, ước tính xác suất của một sự kiện. Ưu điểm: Hiệu quả với dữ liệu phân loại, dễ giải thích. Nhược điểm: Giả định quan hệ tuyến tính giữa biến đầu vào và log-odds của đầu ra.

TT	Tên	Đặc điểm
		Ứng dụng trong lũ quét: Dự đoán khả năng xảy ra lũ quét (có/không).
3	Cây quyết định (Decision Trees)	<p>Đặc điểm: Mô hình dựa trên quy tắc quyết định dạng cây.</p> <p>Ưu điểm: Dễ hiểu và diễn giải, xử lý tốt dữ liệu categorical và số.</p> <p>Nhược điểm: Có thể overfitting, không ổn định với thay đổi nhỏ trong dữ liệu.</p> <p>Ứng dụng trong lũ quét: Phân loại mức độ nguy hiểm của lũ quét.</p>
4	Máy véc-tơ hỗ trợ (Support Vector Machines - SVM)	<p>Đặc điểm: Tìm siêu phẳng tối ưu để phân loại dữ liệu.</p> <p>Ưu điểm: Hiệu quả trong không gian nhiều chiều, linh hoạt với nhiều kernel.</p> <p>Nhược điểm: Khó chọn kernel phù hợp, kém hiệu quả với dữ liệu lớn.</p> <p>Ứng dụng trong lũ quét: Phân loại các khu vực có nguy cơ lũ quét.</p>
5	Mô hình phân lớp Naive Bayes	<p>Đặc điểm: Dựa trên định lý Bayes với giả định độc lập giữa các đặc trưng.</p> <p>Ưu điểm: Đơn giản, nhanh, hiệu quả với dữ liệu văn bản và phân loại.</p> <p>Nhược điểm: Giả định độc lập giữa các đặc trưng thường không đúng trong thực tế.</p> <p>Ứng dụng trong lũ quét: Phân loại các loại cảnh báo lũ quét dựa trên nhiều yếu tố.</p>
6	K láng giềng gần nhất (K-Nearest Neighbors - KNN)	<p>Đặc điểm: Phân loại dựa trên khoảng cách với K điểm dữ liệu gần nhất.</p> <p>Ưu điểm: Đơn giản, hiệu quả với dữ liệu có ranh giới quyết định phức tạp.</p> <p>Nhược điểm: Chậm với dữ liệu lớn, nhạy cảm với tỷ lệ các đặc trưng.</p> <p>Ứng dụng trong lũ quét: Dự đoán mức độ lũ quét dựa trên các sự kiện tương tự trong quá khứ.</p>
7	Phân cụm K-means (K-Means Clustering)	<p>Đặc điểm: Phân nhóm dữ liệu thành K cụm dựa trên khoảng cách.</p> <p>Ưu điểm: Đơn giản, hiệu quả với dữ liệu lớn, dễ hiểu.</p> <p>Nhược điểm: Cần xác định số cụm trước, nhạy cảm với outliers và điểm khởi tạo.</p> <p>Ứng dụng trong lũ quét: Phân nhóm các khu vực có đặc điểm lũ quét tương tự.</p>
II <i>Mô hình học sâu (deep learning)</i>		
1	Mạng nơ-ron tích chập (Convolutional Neural	<p>Đặc điểm: Được thiết kế để xử lý dữ liệu có cấu trúc lưới, đặc biệt là hình ảnh.</p> <p>Cấu trúc: Gồm các lớp tích chập, lớp gộp (pooling) và lớp kết nối đầy đủ.</p> <p>Ưu điểm:</p>

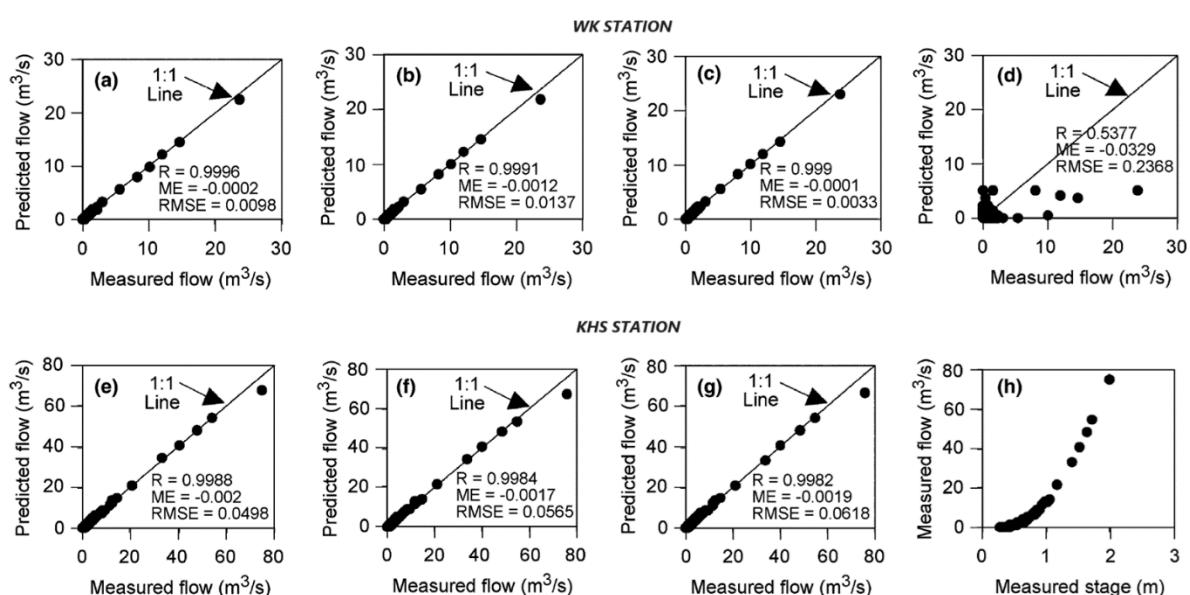
TT	Tên	Đặc điểm
	Networks CNN)	<p>- Hiệu quả trong việc trích xuất đặc trưng từ dữ liệu không gian.</p> <p>- Giảm số lượng tham số cần học so với mạng nơ-ron thông thường.</p> <p>- Có khả năng học tự động các đặc trưng phức tạp.</p> <p>Nhược điểm:</p> <ul style="list-style-type: none"> - Cần lượng dữ liệu lớn để huấn luyện. - Tính toán nặng, đòi hỏi phần cứng mạnh. <p>Ứng dụng trong nghiên cứu lũ quét: Phân tích ảnh vệ tinh hoặc radar để đánh giá nguy cơ lũ quét.</p>
2	Mạng nơ-ron truy hồi (Recurrent Neural Networks RNN)	<p>Đặc điểm: Được thiết kế để xử lý dữ liệu chuỗi và dữ liệu thời gian.</p> <p>Cấu trúc: Có các kết nối truy hồi, cho phép thông tin được truyền từ bước thời gian trước đến bước hiện tại.</p> <p>Ưu điểm:</p> <ul style="list-style-type: none"> - Hiệu quả trong việc xử lý dữ liệu chuỗi thời gian. - Có khả năng học các phụ thuộc thời gian. <p>Nhược điểm:</p> <ul style="list-style-type: none"> - Khó huấn luyện do vấn đề gradient biến mất/bùng nổ. - Khó xử lý các phụ thuộc dài hạn. <p>Ứng dụng trong nghiên cứu lũ quét: Dự báo lũ quét dựa trên dữ liệu thời gian thực về lượng mưa, mực nước sông.</p>
3	Bộ nhớ dài-ngắn hạn (Long Short-Term Memory)	<p>Đặc điểm: Một loại đặc biệt của RNN, được thiết kế để khắc phục vấn đề của RNN thông thường.</p> <p>Cấu trúc: Sử dụng các cổng (gates) để kiểm soát luồng thông tin, bao gồm cổng quên, cổng đầu vào và cổng đầu ra.</p> <p>Ưu điểm:</p> <ul style="list-style-type: none"> - Có khả năng học các phụ thuộc dài hạn tốt hơn RNN thông thường. - Giải quyết vấn đề gradient biến mất. - Hiệu quả trong việc xử lý chuỗi dài. <p>Nhược điểm:</p> <ul style="list-style-type: none"> - Phức tạp hơn RNN, có nhiều tham số hơn cần học. - Vẫn có thể gặp khó khăn với chuỗi rất dài. <p>Ứng dụng trong nghiên cứu lũ quét: Dự báo lũ quét dài hạn, phân tích xu hướng lũ quét theo thời gian.</p>
4	Mạng đối nghịch tạo sinh (Generative adversarial networks)	<p>Đặc điểm: Bao gồm hai mạng neural cạnh tranh nhau: mạng sinh (generator) và mạng phân biệt (discriminator).</p> <p>Cấu trúc: Mạng sinh tạo ra dữ liệu giả, mạng phân biệt cố gắng phân biệt dữ liệu thật và giả.</p> <p>Ưu điểm:</p> <ul style="list-style-type: none"> - Có khả năng tạo ra dữ liệu mới rất giống với dữ liệu thật. - Học được các phân phối phức tạp của dữ liệu. <p>Nhược điểm:</p> <ul style="list-style-type: none"> - Khó huấn luyện và điều chỉnh.

TT	Tên	Đặc điểm
		<ul style="list-style-type: none"> Có thể không ổn định trong quá trình huấn luyện. Ứng dụng trong nghiên cứu lũ quét: Tạo ra các kịch bản lũ quét giả định để đánh giá rủi ro, tăng cường dữ liệu cho các mô hình dự báo.

Bản chất của mô hình trí tuệ nhân tạo là quá trình học hỏi những đặc trưng của các dữ liệu đầu vào để dự đoán kết quả đầu ra từ các dữ liệu được đào tạo. Mặc dù các mô hình tuyến tính và phi tuyến đã được áp dụng trong mô hình mưa – dòng chảy từ rất sớm, các nghiên cứu trước đây thường đưa ra các phát hiện dưới dạng công thức thay vì “học” [39]. Năm 1995, Hsu đã sử dụng mô hình mạng thần kinh nhân tạo (ANN) để mô phỏng mưa – dòng chảy ở Mississippi [40], đây là một trong những nghiên cứu điển hình ứng dụng trí tuệ nhân tạo trong tính toán thủy văn.

Năm 2006, Sahoo đã ứng dụng mô hình ANN để mô phỏng lũ quét cho lưu vực suối ở Oahu, Hawaii [41] thông qua quá trình mô phỏng mưa dòng chảy. ANN được kết hợp với thuật toán lan truyền ngược và gọi là thuật toán BPNN với 2 lớp ẩn. Đầu vào của mô hình bao gồm (1) lượng mưa tại thời điểm t ; (2) bốc hơi tại thời điểm t ; (3) mực nước tại thời điểm t ; và (4) mực nước tại thời điểm $t - 1$. Sahoo đã mô phỏng với các trường hợp bao gồm: (a) với tất cả dữ liệu đầu vào; (b) tất cả dữ liệu ngoại trừ bốc hơi; (c) chỉ sử dụng mực nước; (d) tất cả ngoại trừ mực nước. Kết quả mô phỏng cho 2 trạm WK và KHS cho thấy, hệ số tương quan R trong tất cả các trường hợp đều lớn hơn 0,99.

Mặc dù đạt với độ tương quan rất cao, tác giả chưa trình bày quá trình dự đoán cho một thời đoạn khác, do vậy chưa thể đánh giá được khả năng dự đoán dòng chảy từ lượng mưa (do chuỗi dữ liệu trong tương lai chưa có mực nước). Khởi đầu này cho thấy tiềm năng của dự báo dòng chảy bằng phương pháp ANN trong tính toán thủy văn.



Hình 6. Kết quả mô phỏng mưa – dòng chảy bằng mô hình BPNN của Sahoo

Năm 2009, Janál đã sử dụng mô hình tập mờ (Fuzzy) để dự báo tình trạng khẩn cấp lưu vực sông trong tình huống lũ quét cho lưu vực Morva và Odra (cộng hòa Séc) trên MATLAB [42]. Với các tham số đầu vào bao gồm: (1) cường độ mưa; (2) thời lượng mưa; (3) chỉ số mưa tích lũy; (4) diện tích lưu vực; (5) hệ số hình dạng lưu vực; (6) độ dốc lưu vực; và (7) độ che phủ lưu vực. Đầu ra của mô hình là giá trị dòng chảy cực đại (lưu lượng). Các đại lượng đầu vào được chuẩn hóa về từ 0÷1, kết quả đầu ra được phân thành 4 cấp độ lũ lụt, trong đó cấp 3 và cấp 4 được xem là thảm khốc.

Toukourou đã sử dụng phương pháp học thống kê để dự báo dòng chảy lũ dựa vào lượng mưa và mực nước, áp dụng cho lưu vực Gardon (Pháp) [43]. Mô hình ANN được sử dụng với một lớp ẩn. Kết quả cho thấy hệ số NAS dự đoán lưu lượng lũ giảm dần theo thời đoạn dự báo. Hệ số lớn nhất đạt 0,98 cho thời đoạn dự báo sau 30 phút, sau đó giảm dần đến 0,84 cho thời đoạn dự báo sau 4 giờ và thấp nhất là 0,58 cho thời đoạn dự báo lớn hơn 5 giờ.

Bảng 3. Kết quả dự báo theo thời đoạn của mô hình ANN cho lưu vực sông Gardon

Forecasting horizon (f)	0.5 h	1 h	2 h	3 h	4 h	5 h	mean
N_C	2	2	5	3	3	3	3
Persistency criterion	0.45	0.65	0.32	0.28	0.23	0.59	0.42
R^2 (Nash-Sutcliffe criterion)	0.98	0.93	0.87	0.93	0.84	0.58	0.85
Estimated/Observed peak values	0.90	0.84	0.73	0.82	0.79	0.60	0.78

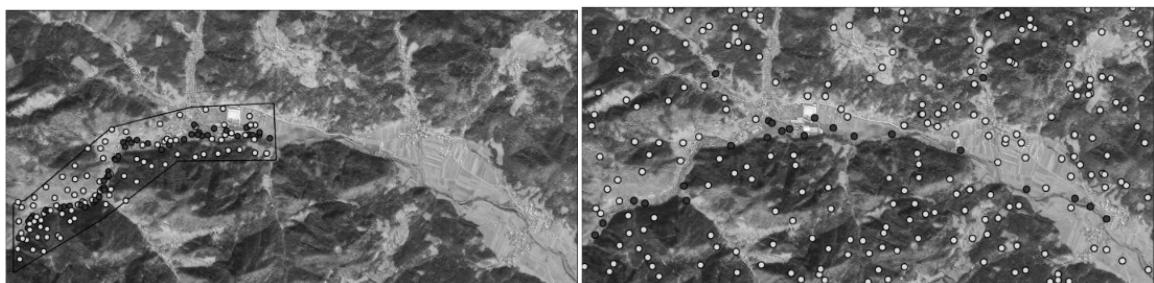
Số lượng lớp ẩn và tham số đầu vào sẽ quyết định các kết quả đầu ra của mô hình dự báo. Các tham số này rất khó điều chỉnh/quyết định để có thể lựa chọn được mô hình hiệu quả. Thông thường, đối với dữ liệu đầu vào, các tham số cần đạt một số tiêu chí bao gồm:

- Tính tương quan với các tham số khác: các tham số độc lập sẽ cho ra kết quả tốt hơn các tham số phụ thuộc. Ví dụ nếu cùng đưa chỉ số tập trung dòng chảy và diện tích lưu vực làm thông số đầu vào, mô hình sẽ hoạt động kém hiệu quả hơn (do tham số diện tích phụ thuộc vào tham số tập trung dòng chảy).
- Tác động đến kết quả đầu ra: các yếu tố đầu vào phải là các yếu tố có tác động đến kết quả đầu ra về mặt lý thuyết. Mô hình có các yếu tố đầu vào càng tác động mạnh đến kết quả đầu ra thì sẽ hoạt động hiệu quả hơn các mô hình đưa các tham số đầu vào ít có liên quan đến kết quả đầu ra.

Khi đã thỏa mãn các điều kiện trên, một mô hình hiệu quả cần phải xác định các tham số hiệu quả (through qua các thuật toán lựa chọn/số lớp ẩn....). Việc này khó xác định và cần một bước thử dần/tối ưu dựa trên các thuật toán đánh giá kết quả đầu ra. Một mô hình có kết quả đầu ra tốt sẽ có bộ tham số tốt.

Các tiếp cận trên hướng đến việc dự báo dòng chảy lũ cho lưu vực sông, tương tự như cách hoạt động của mô hình thủy văn/thủy lực. Các yếu tố tác động đến quá trình hình thành dòng chảy được chú trọng làm đầu vào cho mô hình dự báo. Có thể nói, đây là một hướng đi đúng đắn và phù hợp thực tế do lũ quét là một dạng lũ nên kết quả phản ứng thủy văn của một lưu vực sẽ giúp xác định được dòng chảy lũ quét. Tuy nhiên, kết quả này không thể phân vùng lũ quét mà chỉ có thể xác định dòng chảy tại một vị trí cụ thể và phân loại ngưỡng nhằm cảnh báo lũ/lũ quét.

Năm 2013, Lamovec đã sử dụng mô hình học máy để phát hiện vùng ngập lũ cho trận lũ quét tại sông Selška Sora (Slovenia) xảy ra vào tháng 9 năm 2007 [44]. Hàng loạt các mô hình được sử dụng bao gồm mô hình Bayes (NavieBayes; BayesNet), mô hình cây quyết định (J48; Random Tree; Random Forest); mô hình Rules (JRip) và một số thuật toán bao gồm AdaBoostM1, LogitBoost, Bagging dựa trên bộ dữ liệu ảnh SPOT (2,5m), DTM ở độ phân giải 12,5m, mạng sông. Các tham số đầu vào bao gồm: (1) cao độ; (2) độ dốc; (3) hướng; (4) NDVI; (5) NDBI; (6) NBI cho 145 điểm thuộc mô phạm vi nhỏ được trích xuất phục vụ đào tạo cho các thuộc tính có độ phân giải cao nhất và 255 điểm thuộc phạm vi lớn hơn cho độ phân giải thấp hơn.

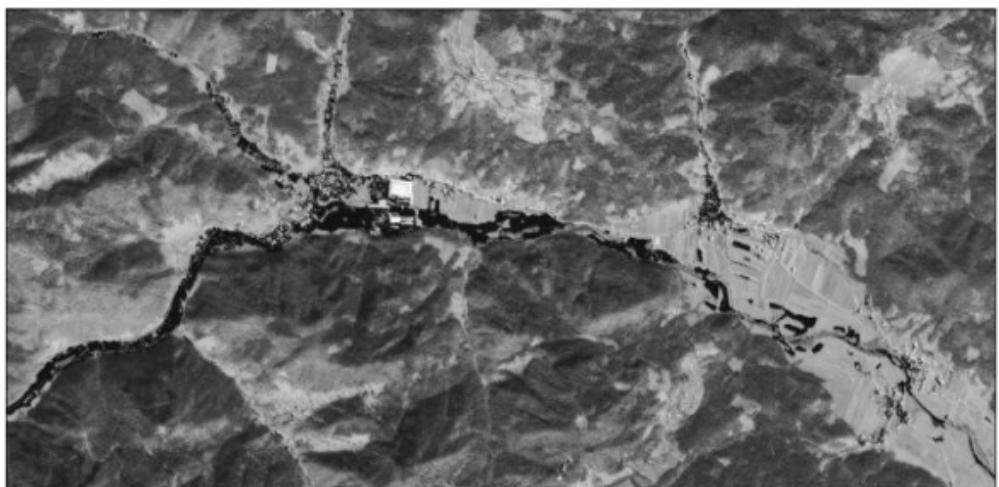


Hình 7. Các điểm trích xuất cho dữ liệu đào tạo mô hình

Toàn bộ các điểm (400 điểm) được sử dụng cho đào tạo mô hình (trainning), trong khi đó, 145 điểm ở độ phân giải cao sẽ được sử dụng trích xuất thành 145 điểm ở độ phân giải thấp phục vụ kiểm tra (testing). Kết quả về độ chính xác dự đoán được thể hiện ở bảng sau:

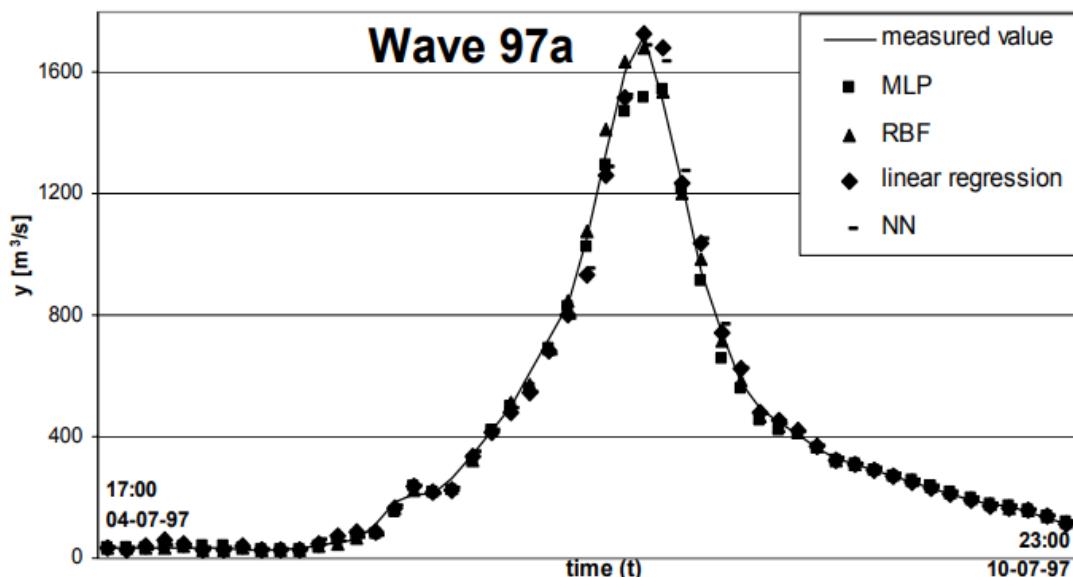
Bảng 4. Độ chính xác của phát hiện ngập lũ theo các thuật toán

Method:	Correctly classified instances	
	Training	Testing
J48	88%	95%
JRip	86%	85%
Bagging J48 (10 trees)	90%	92%
Random Forest C45 (10 trees)	89%	92%



Hình 8. Kết quả phát hiện ngập lũ theo thuật toán J48

Piotrowski đã trình bày một nghiên cứu đánh giá tính hiệu quả của các kỹ thuật học máy khác nhau, đặc biệt là mạng lưới thần kinh (Neural Network) trong việc dự báo các sự kiện mưa – dòng chảy lũ nhằm so sánh hiệu quả của các mô hình khác nhau trong việc dự báo lũ quét, từ đó nâng cao khả năng dự báo lũ [45]. Nghiên cứu này sử dụng các mô hình bao gồm Multi-Layer Perceptron (MLP), Mạng chức năng cơ sở xuyên tâm (RBF), Phương pháp tiếp cận láng giềng gần nhất (NN), Hồi quy tuyến tính cho khu vực sông Nysa Kłodzka với dữ liệu bao gồm lượng mưa và lưu lượng (mỗi 3 giờ) trong thời gian từ năm 1965 đến năm 2000. Các trận lũ được phân tích phải thỏa mãn độ trễ tối đa 24 giờ. Kết quả nghiên cứu cho thấy, mạng RBF vượt trội đáng kể so với các kỹ thuật dự báo khác, bên cạnh đó, kết quả dự báo cho thời đoạn 3 giờ có độ tin cậy cao.

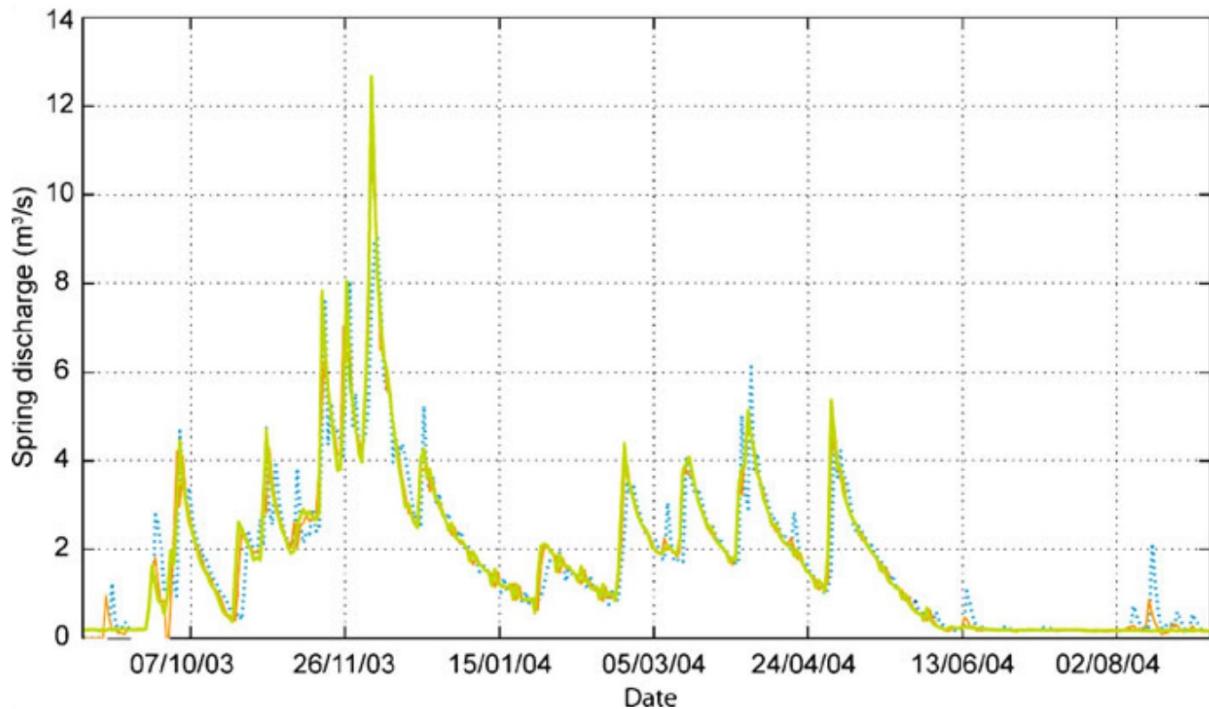


Hình 9. Kết quả dự báo thời đoạn 3 giờ của các mô hình trong nghiên cứu của Piotrowski

Kong A Siou đã phát triển mô hình dự báo lũ quét ở lưu vực Lez (miền nam nước Pháp) sử dụng mạng lưới thần kinh nhân tạo nhằm mục đích nâng cao sự hiểu biết về

đặc tính thủy động lực của lưu vực Lez và cải thiện độ chính xác của dự báo lũ [46]. Tác giả đã sử dụng mạng thần kinh nhân tạo (ANN) để dự báo dòng chảy ra dựa trên dữ liệu lượng mưa và lưu lượng lịch sử. Quá trình đào tạo bao gồm các kỹ thuật như dừng sớm và xác thực chéo để ngăn chặn việc trang bị quá mức và đảm bảo khả năng khai thác hóa của mô hình. Tập dữ liệu bao gồm dữ liệu lịch sử của 16 năm, bao gồm dữ liệu lượng mưa từ ba trạm quan trắc và trạm đo lưu lượng nằm trên suối Lez. Dữ liệu được sử dụng để huấn luyện mạng lưới thần kinh, tập trung vào việc lựa chọn các tham số đầu vào tối ưu thông qua phân tích tương quan chéo và kiểm soát độ phức tạp của mô hình. Kết quả chỉ ra rằng mô hình mạng thần kinh đã mô phỏng và dự báo thành công các dòng chảy ra với tiêu chí Nash xấp xỉ 0,95 cho mô phỏng và 0,84 cho dự báo trong hai ngày. Mô hình này thể hiện hiệu suất thỏa đáng trong việc đồng bộ hóa dòng chảy định với dữ liệu quan sát được, cho thấy độ tin cậy của nó trong việc dự báo lũ ngắn hạn.

Forecasting horizon f	0	1 day	2 days	3 days	Selected architecture
Valflaunès	0.95	0.90	0.84	0.73	8 inputs, 5 hidden neurons
Prades	0.94	0.90	0.84	0.73	9 inputs, 3 hidden neurons
St Martin	0.94	0.90	0.87	0.75	9 inputs, 3 hidden neurons

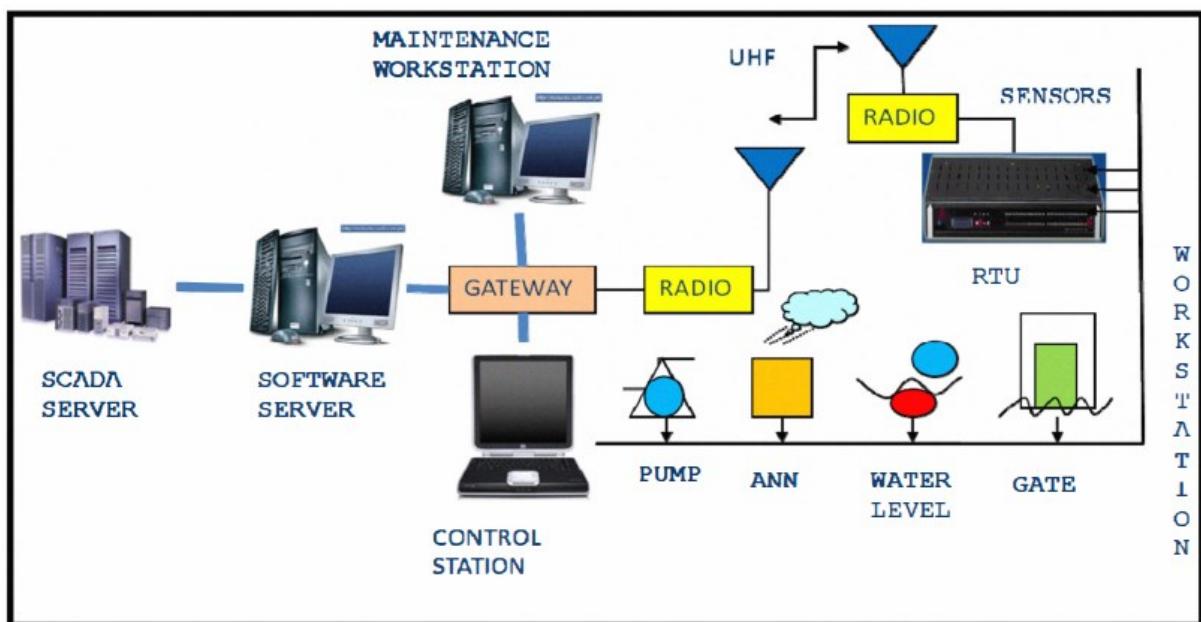


Hình 10. Kết quả mô phỏng dự báo lưu lượng dựa trên lượng mưa (màu xanh nõn chuối: thực đo; màu cam: mô phỏng; màu xanh chấm: dự báo trước 2 ngày)

Năm 2011, Artigue đã nghiên cứu phát triển mô hình dự báo lũ quét hiệu quả bằng cách sử dụng mạng thần kinh, được thiết kế đặc biệt để vận hành mà không dựa vào dự báo lượng mưa trong tương lai hoặc dữ liệu lưu lượng quan sát được trước đó. Nghiên cứu này nhằm mục đích tăng cường hệ thống cảnh báo sớm lũ quét ở các lưu vực không có trạm đo [47]. Khu vực áp dụng là lưu vực Gardon de Mialet, một tiểu lưu vực của

vùng Anduze ở miền Nam nước Pháp. Khu vực này có đặc điểm là độ dốc lớn, đất đá mỏng và dễ bị ảnh hưởng bởi lượng mưa lớn cục bộ, khiến nơi đây trở thành địa điểm quan trọng trong việc dự báo lũ quét. Nghiên cứu sử dụng mô hình mạng nơ-ron perceptron đa lớp (MLP). Mô hình này được chọn vì các đặc tính gần đúng và phân tích phổ quát, cho phép mô hình hóa một cách hiệu quả mối quan hệ phức tạp giữa lượng mưa và lưu lượng trong lưu vực. Kiến trúc bao gồm cả các thành phần tuyến tính và phi tuyến tính để nắm bắt động lực của phản ứng thủy văn.

Nghiên cứu sử dụng dữ liệu từ ba máy đo mưa đặt tại Barre-des-Cévennes, Mialet và Saint-Roman-de-Tousque, cùng với các phép đo lưu lượng từ máy đo tại Mialet. Bộ dữ liệu bao gồm các bản ghi lượng mưa và lưu lượng, được xử lý để chọn các biến có liên quan cho việc huấn luyện mạng lưới thần kinh. Kết quả chỉ ra rằng mô hình mạng lưới thần kinh cung cấp chất lượng dự báo tuyệt vời, đạt các giá trị tiêu chí Nash trên 0,8, được coi là chuẩn mực cho hiệu suất tốt trong lĩnh vực thủy văn. Đáng chú ý, mô hình này duy trì độ chính xác dự đoán ngay cả khi phạm vi dự báo tăng lên, thể hiện tính chắc chắn trong dự đoán của nó. Tuy nhiên, thời điểm dự đoán lưu lượng đỉnh điểm kém khả quan hơn, cho thấy còn nhiều cơ hội để cải thiện.



Hình 11. Hệ thống cảnh báo lũ quét khu vực Selangor, Malaysia

Izyan đã sử dụng mô hình ANN kết hợp với hệ thống SCADA bằng cách sử dụng dữ liệu khí tượng hàng ngày được thu thập từ ba trạm chính ở Selangor trong những năm 2007 đến 2010. Các thông số đầu vào bao gồm nhiệt độ bầu ướt, độ ẩm tương đối, tốc độ gió, độ mây và áp suất không khí, trong khi biến mục tiêu là lượng mưa đo bằng máy đo mưa tại các trạm đã chọn để dự báo nguy cơ lũ quét cho khu vực Selangor, Malaysia [48]. Kết quả chỉ ra rằng hiệu suất dự báo hàng ngày của mô hình không cao như mong đợi nhưng nó vẫn cung cấp những dự đoán hữu ích cho các ứng dụng

thực tế trong dự báo lượng mưa và quản lý lũ lụt. Mô hình ANN đã chứng tỏ tính hiệu quả trong việc xử lý dữ liệu thời tiết ồn ào và không ổn định, đặc trưng trong các tình huống như vậy. Việc tích hợp các kết quả dự báo lượng mưa vào hệ thống SCADA được coi là một bước quan trọng hướng tới cải thiện các chiến lược phòng chống lũ lụt.

Boukharouba năm 2013 đã sử dụng một phương pháp liên quan đến việc phân cụm các sự kiện lũ lụt thành các nhóm riêng biệt dựa trên đặc điểm của chúng, cho phép tạo các mô hình Hồi quy vectơ hỗ trợ (SVR) cụ thể cho từng cụm [49]. Cách tiếp cận này trái ngược với các phương pháp truyền thống sử dụng một mô hình toàn cầu duy nhất. Nghiên cứu sử dụng cơ sở dữ liệu toàn diện về lượng mưa và dữ liệu mực nước được thu thập từ năm 1993 đến năm 2008, bao gồm 23 trận lũ lớn, với thời gian lấy mẫu là 30 phút.

Việc dự báo lũ quét dựa trên lượng mưa của mô hình trí tuệ nhân tạo bẩn chất vẫn hướng về dự báo mưa – dòng chảy và đây là một cách tiếp cận đúng bởi mưa là yếu tố tự nhiên chính gây ra lũ quét (các trường hợp sự cố/nghẽn dòng là các yếu tố bất định). Mặc dù vậy, nhiều nghiên cứu hướng đến mục tiêu xác định các khu vực dễ ảnh hưởng bởi lũ quét dựa trên đặc điểm địa hình, địa mạo (susceptibility). Cách tiếp cận này có thể chỉ ra những khu vực dễ bị ảnh hưởng bởi lũ quét, nhưng không chỉ ra được khi nào thì sẽ xảy ra lũ quét. Tuy nhiên, kết quả nghiên cứu của các dạng nghiên cứu này lại có ý nghĩa lớn đối với quy hoạch sử dụng đất hoặc bố trí dân cư/tái định cư.

1.2.2 Các nghiên cứu ở Việt Nam

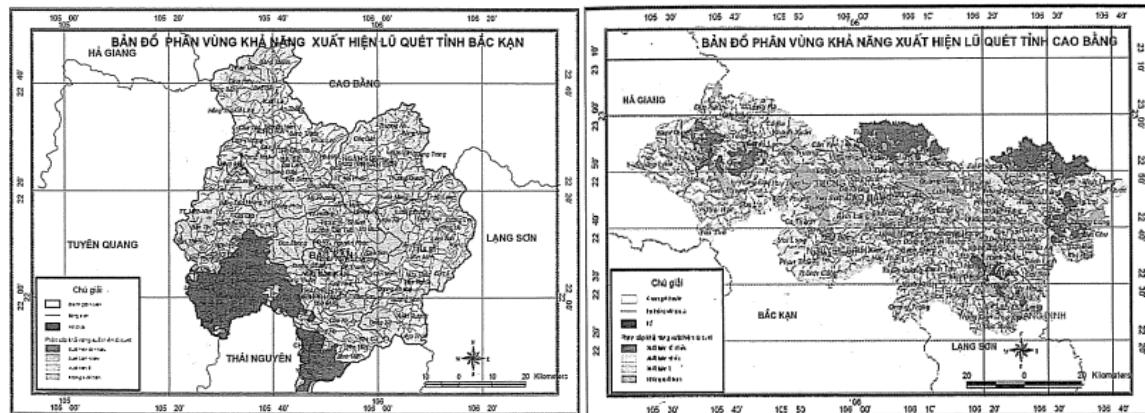
1.2.2.1 Phân vùng lùi quét theo số dữ liệu địa không gian

Năm 1995, Cao Đăng Dư đã phân vùng khả năng xuất hiện lũ quét cho vùng Tây Bắc bằng phương pháp phân tích nhân tố [50], các nhân tố được đưa vào đánh giá bao gồm: (1) lượng mưa một ngày lớn nhất ứng với tần suất 1% và 5%; (2) độ dốc lưu vực; (3) độ dốc lòng sông; (4) lớp phủ; (5) loại đất; và (6) mô đun dòng chảy đỉnh lũ tần suất 1% và 5%. Bản đồ được xây dựng ở tỷ lệ 1:500.000 và kết quả được phân thành 4 cấp độ: (1) khả năng xuất hiện cao; (2) khả năng xuất hiện khá; (3) khả năng xuất hiện trung bình; (4) ít có khả năng xuất hiện.

Phạm Thị Hương Lan và Vũ Minh Cát đã xây dựng bản đồ tiềm năng lũ quét phục vụ công tác cảnh báo lũ quét vùng núi Đông Bắc Việt Nam [51] sử dụng 4 yếu tố bao gồm: (1) độ dốc bề mặt; (2) lượng mưa một ngày lớn nhất; (3) thảm thực vật; (4) khả năng thấm của đất. Mỗi yếu tố được phân thành 20 cấp, có 4.845 tổ hợp được xem xét, trong đó 97 tổ hợp cho thấy có



khả năng xuất hiện lũ quét cao và phù hợp với điều kiện tự nhiên ở Việt Nam. Kết quả cho thấy tác giả đã xác định được ngưỡng mưa một ngày cho từng trạm bao gồm: Ở Thái Nguyên – trạm Kỳ Phú 220mm; Phú Bình 167mm; Cao Bằng – trạm Trùng Khánh 286,3mm, Hà Quảng 212mm; Bắc Kạn – trạm Bắc Kạn 205,7mm, Thác Riềng 106,3mm, Chợ Mới 143mm.



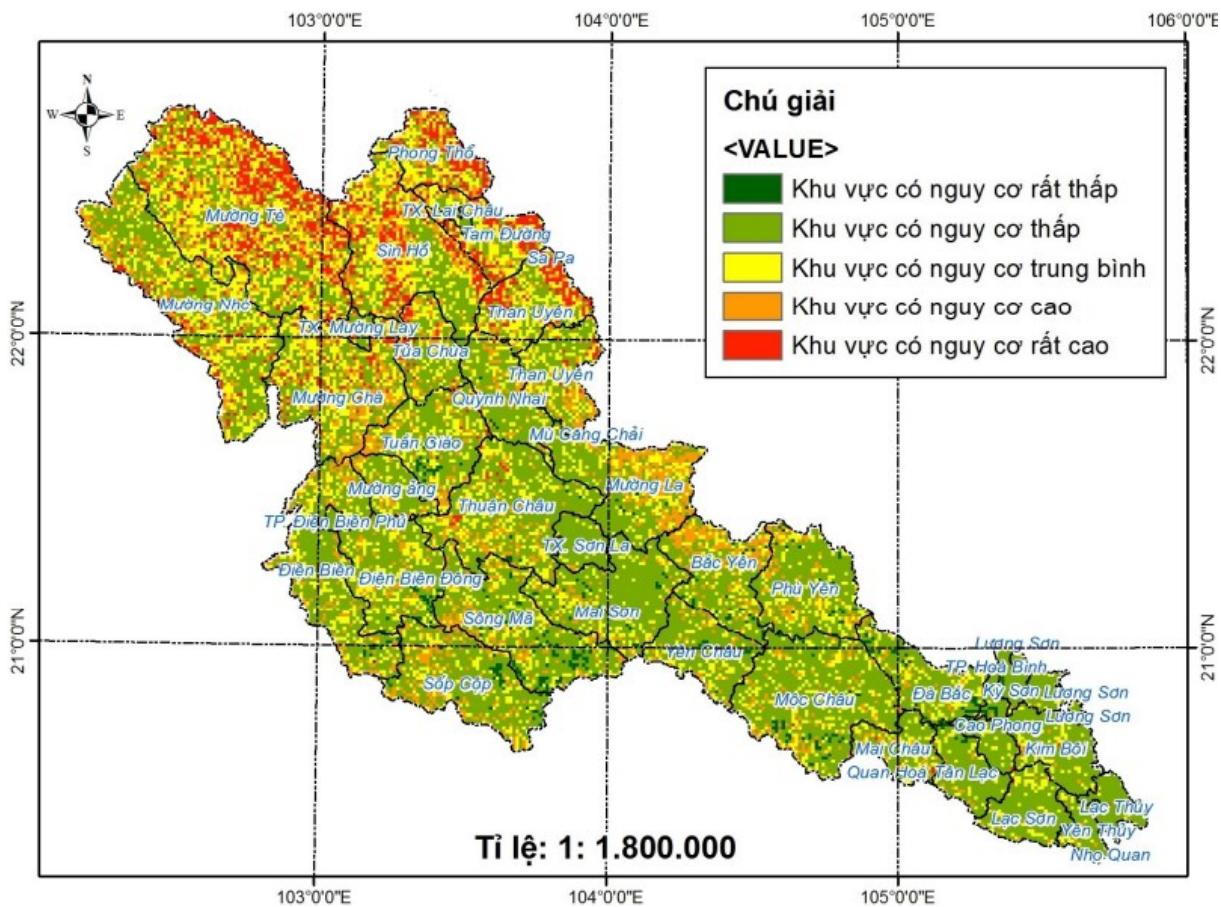
Hình 12. Bản đồ phân vùng khả năng xuất hiện lũ quét các tỉnh Đông Bắc

Lã Thanh Hà (2009) đã xây dựng bản đồ phân vùng nguy cơ lũ quét cho tỉnh Yên Bái [52] dựa trên phương pháp trọng số thử dàn với số liệu kiểm chứng là 58 trận lũ quét đã xảy ra trên địa bàn tỉnh. Nghiên cứu sử dụng các số liệu đầu vào bao gồm: (1) lượng mưa ngày lớn nhất ứng với tần suất 50%; (2) nguy cơ xói mòn đất; (3) độ dốc lưu vực; và (4) khả năng phòng hộ của rừng. Kết quả phân vùng được xây dựng trên bản đồ tỷ

Dương Thị Lợi [53] đã sử dụng phương pháp phân tích thứ bậc (AHP) để phân tích đa chỉ tiêu các yếu tố bao gồm: (1) độ dốc; (2) thành phần cơ giới của đất; (3) sử dụng đất; (4) lượng mưa một ngày lớn nhất. Kết quả ma trận trọng số thể hiện như sau:

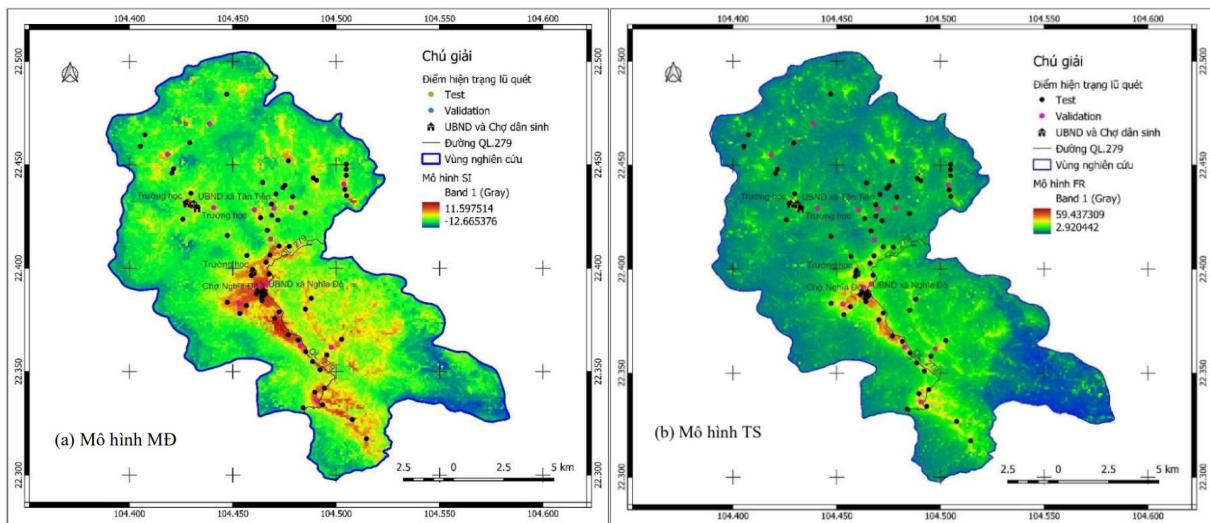
Bảng 5. Bảng đánh giá trọng số theo phương pháp AHP

	Độ dốc	LM	HTSDD	TPCGĐ	Tổng
Độ dốc	1,00	3,00	5,00	7,00	16,00
LM	0,33	1,00	3,00	5,00	9,33
HTSDD	0,2	0,33	1,00	3,00	4,53
TPCGĐ	0,14	0,20	0,33	1,00	1,68
Total	2	5	9	16	



Hình 13. Bản đồ nguy cơ lũ quét khu vực Tây Bắc

Đào Minh Đức và cộng sự đã xây dựng bản đồ nguy cơ lũ quét cho suối Nghĩa Đô, huyện Bảo Yên, tỉnh Lào Cai bằng việc sử dụng kết hợp kết quả chòng chập đa nhân tố và lượng mưa ngày để đưa ra bản đồ theo thời gian [54].



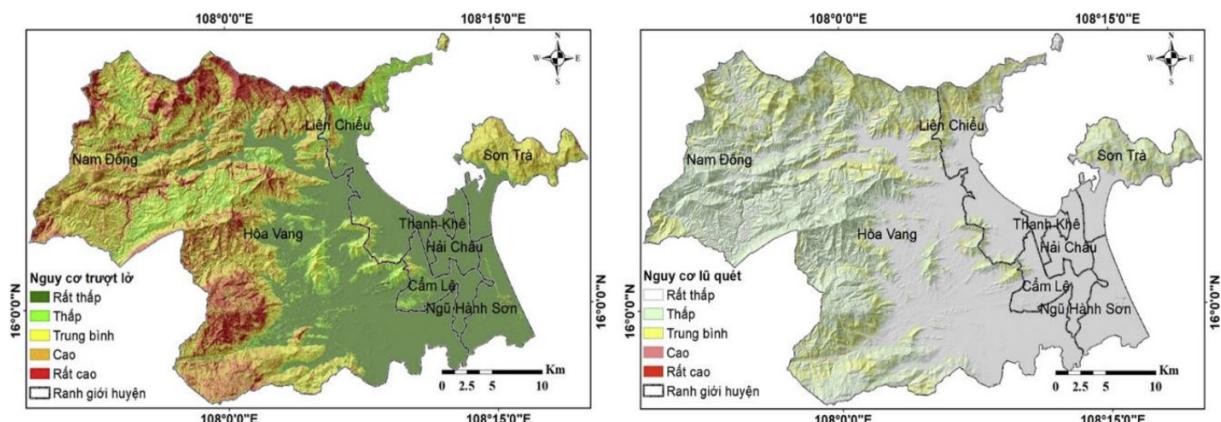
Hình 14. Kết quả xây dựng bản đồ nguy cơ lũ quét theo 2 phương pháp chòng chập đa nhân tố cho suối Nghĩa Đô

Có 9 yếu tố đưa vào đánh giá chòng chập đa nhân tố bằng phương pháp tỷ lệ tần suất và phương pháp thống kê mật độ bao gồm: (1) cao độ; (2) độ dốc; (3) độ cong; (4)

phân cắt sâu; (5) phân cắt ngang; (6) năng lượng dòng chảy; (7) âm địa hình; (8) NDVI; và (9) thạch học. Kết quả xây dựng bản đồ nguy cơ ở phía trên một lần nữa lại được kết hợp với bản đồ lượng mưa trong 24 giờ để tạo ra bản đồ nguy cơ lũ quét theo thời gian. Trong đó, cả hai loại bản đồ đều được phân thành 5 cấp độ nguy cơ và kết quả cuối cùng được phân thành 7 cấp nguy cơ lũ quét

		<10mm	10-25mm	25-45mm	45-70mm	>70mm
		1	2	3	4	5
Tiềm năng hình thành lũ quét trên lưu vực suối Nghĩa Đô	1	I	I	II	II	II
	2	I	III	III	III	III
3	II	III	IV	IV	IV	IV
4	III	IV	V	V	V	V
5	III	IV	V	VI	VI	VII

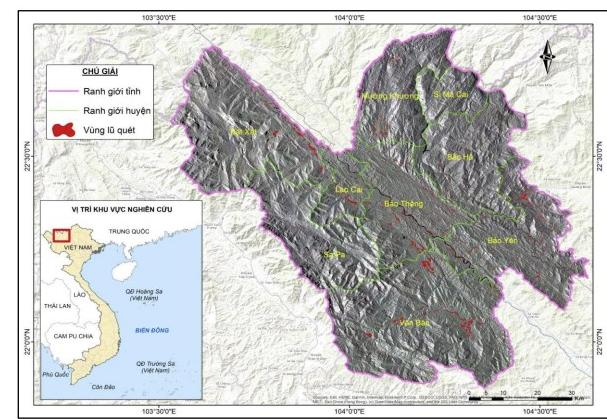
Nguyễn Thị Huyền và cộng sự [55] đã khoanh định các khu vực nhạy cảm về trượt lở, lũ quét cho khu vực thành phố Đà Nẵng bằng phương pháp đánh giá đa tiêu chí SMCE với 9 yếu tố. Mỗi yếu tố được chia thành 5 cấp độ, sau đó kết hợp lại với bộ tiêu chí thành phần để xây dựng bản đồ. Kết quả đã xây dựng được 2 bản đồ cho trượt lở và lũ quét tờ tỷ lệ 1:50.000 cho khu vực thành phố Đà Nẵng.



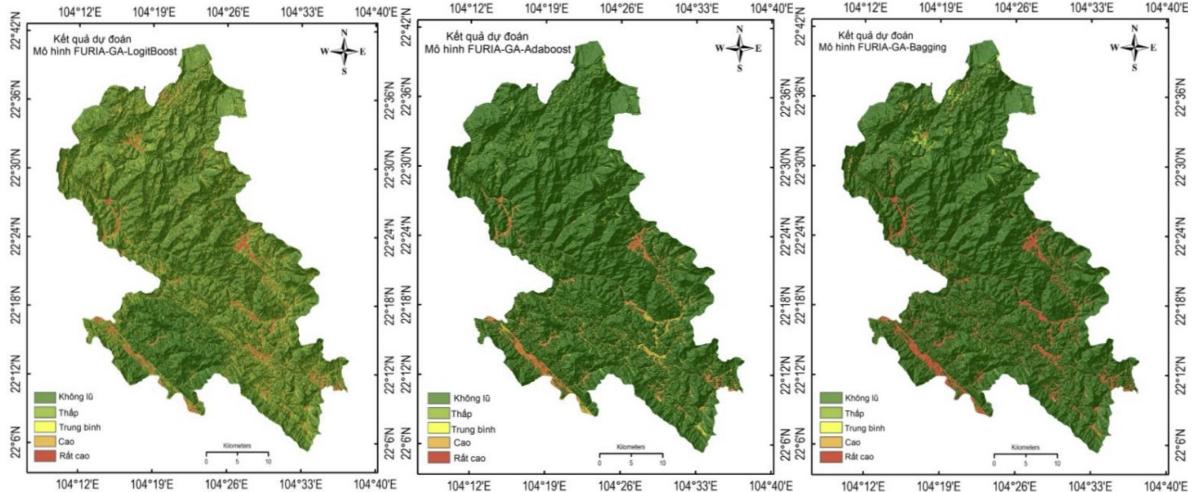
Hình 15. Bản đồ nguy cơ trượt lở và lũ quét cho khu vực thành phố Đà Nẵng

1.2.2.2 Phân vùng lũ quét ứng dụng trí tuệ nhân tạo

Trí tuệ nhân tạo và dữ liệu địa không gian trong những năm gần đây cũng được các nhà khoa học ở Việt Nam quan tâm và ứng dụng trong nghiên cứu về lũ quét. Ngô Thị Phương Thảo [56] đã sử dụng ảnh Sentinel-1A để phát hiện lũ quét và xây dựng bản đồ hiện trạng lũ quét Lào Cai năm 2017. Dữ liệu lựa chọn sử dụng là băng tần VV cho các thời điểm trước và sau khi xảy ra lũ. Ngoài ra, trong một nghiên cứu khác [57], tác giả cũng đã xây dựng mô hình FA-LM-ANN dựa trên nền tảng thuật toán ANN để tự động cập nhật và tối ưu các trọng số của mô hình dự báo lũ quét, đồng thời sử dụng mô hình tổ hợp học máy với



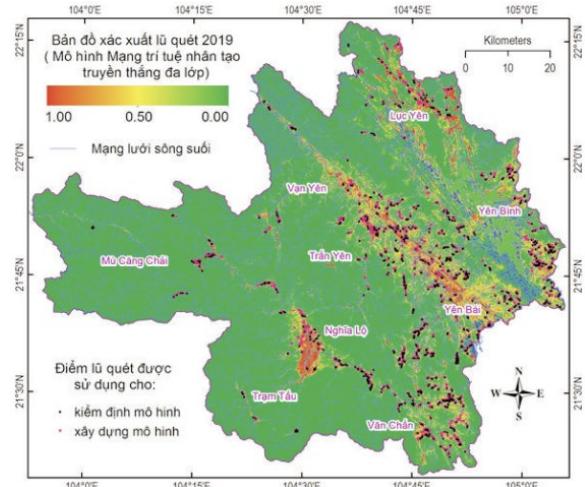
các thuật toán di truyền GA, luật mờ FURIA và cây quyết định DT để dự đoán nguy cơ lũ quét.



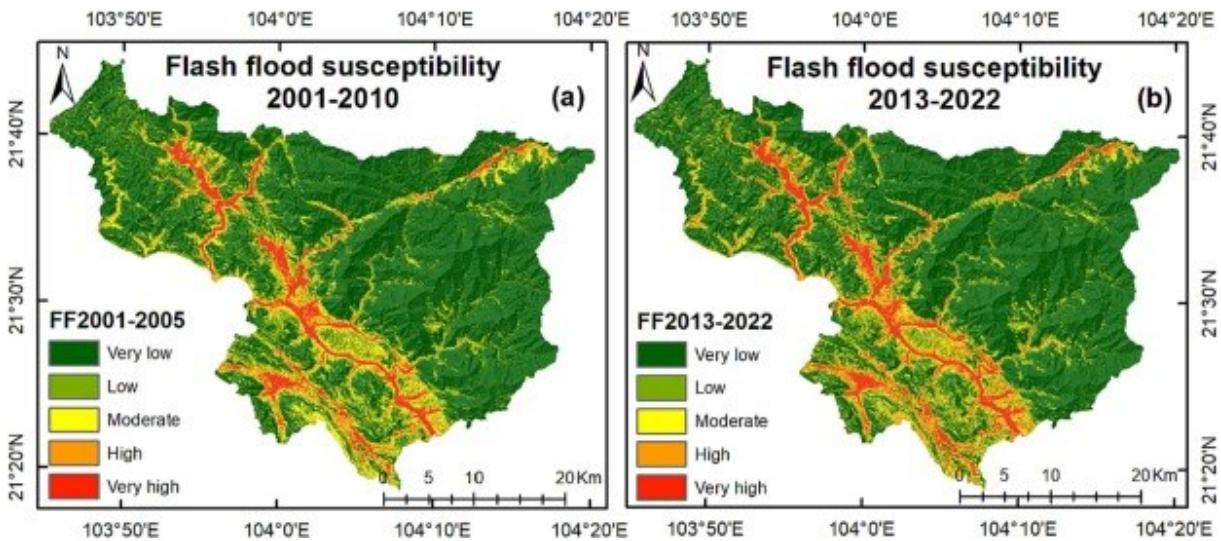
Hình 16. Bản đồ phân vùng nguy cơ lũ quét bởi các mô hình học máy khác nhau [57]

Hà Thị Hằng [58] đã phát triển mô hình học máy để dự báo lũ quét và sạt lở đất trên các tuyến đường bộ ở khu vực miền núi với dữ liệu thu thập là 235 điểm sạt lở và 88 khu vực lũ quét trên quốc lộ 6 (địa phận tỉnh Hòa Bình, Việt Nam). Các mô hình được phát triển bao gồm hình học máy kết hợp nâng cao Decorate-SYS, Rotation Forest-SYS, và Vote-SYS với bộ phân lớp cơ bản là thuật toán Systematically Developed Forest of Multiple Decision Trees (SYS).

Nguyễn Việt Nghĩa [59] đã ứng dụng mạng nơ-ron nhân tạo đa lớp trong thành lập mô hình phân vùng lũ quét khu vực miền núi Tây Bắc và xây dựng bản đồ cho tỉnh Yên Bái. Mô hình sử dụng 7 yếu tố chính bao gồm độ dốc, độ cao, địa chất, lượng mưa ngày lớn nhất, thảm phủ, mật độ dòng chảy và loại đất để làm đầu vào. Kết quả đầu ra là bản đồ xác suất lũ quét cho tỉnh Yên Bái năm 2019.



Hoàng Đức Vinh [60] đã đánh giá định lượng tác động của con người đến khả năng xảy ra lũ quét ở vùng núi Việt Nam bằng việc sử dụng 6 mô hình Multi-layer Perceptron, Gaussian Naïve Bayes, Support Vector Machines, K-Nearest Neighbors, XGBoost, and Random Forest để xây dựng bản đồ nhạy cảm với lũ quét cho khu vực huyện Mường La, tỉnh Sơn La. Kết quả lựa chọn mô hình RF với độ tin cậy cao nhất để lập bản đồ. Kết quả dự đoán được phân loại thành 5 cấp độ dựa trên phương pháp Natural Break.



Hình 17. Bản đồ nhạy cảm với lũ quét cho 2 giai đoạn tại huyện Mường La

1.3. Các sản phẩm chính từ kết quả phân vùng lũ quét sử dụng trí tuệ nhân tạo và dữ liệu địa không gian.

Sự kết hợp giữa trí tuệ nhân tạo (AI), bao gồm học máy (ML) và học sâu (DL), cùng với dữ liệu địa không gian đã tạo ra những bước tiến đáng kể trong việc phân vùng và quản lý lũ quét. Các sản phẩm chính từ quá trình này bao gồm bản đồ phát hiện lũ quét, bản đồ nhạy cảm với lũ quét (susceptibility maps), và bản đồ nguy cơ lũ quét (hazard maps). Những sản phẩm này không chỉ hỗ trợ trong việc dự báo và cảnh báo sớm mà còn đóng vai trò quan trọng trong việc lập kế hoạch giảm thiểu rủi ro và quản lý tài nguyên nước.

1.3.1 Bản đồ phát hiện lũ quét

Bản đồ phát hiện lũ quét là sản phẩm trực tiếp từ các mô hình AI nhằm xác định và phân loại các khu vực đang hoặc đã trải qua lũ quét dựa trên dữ liệu thời gian thực hoặc lịch sử. Đây là công cụ quan trọng trong việc phát hiện nhanh các sự kiện lũ quét, đặc biệt ở những khu vực có nguy cơ cao và thiêu trạm quan trắc (ungauged basins). Các nghiên cứu như “Flash Flood Detection and Alert System Using Machine Learning” (2024) [61] đã sử dụng MobileNet CNN để phân loại hình ảnh vệ tinh, đạt độ chính xác lên đến 90%, từ đó tạo ra bản đồ phát hiện lũ quét với khả năng cảnh báo kịp thời qua tin nhắn hoặc mạng xã hội.

1. Quy trình phát triển:

Quá trình xây dựng bản đồ phát hiện lũ quét thường bắt đầu bằng việc thu thập dữ liệu địa không gian, bao gồm hình ảnh vệ tinh (Sentinel-1, Landsat), dữ liệu mưa từ các sản phẩm như IMERG hoặc PERSIANN, và thông tin thủy văn như mực nước hoặc lưu lượng. Các mô hình DL như Convolutional Neural Network (CNN) hoặc U-Net được sử dụng để phân đoạn (segmentation) các khu vực ngập lụt từ hình ảnh vệ tinh, trong khi các mô hình ML như Random Forest (RF) hoặc XGBoost được áp dụng để phân loại

các sự kiện lũ dựa trên dữ liệu đa nguồn. Ví dụ, nghiên cứu “A modern method for building damage evaluation using deep learning approach - Case study: Flash flooding in Derna, Libya” (2024) [62] đã khai thác dữ liệu SAR từ Sentinel-1 để xác định phạm vi ngập lụt, kết hợp với U-Net để trích xuất các khu vực bị ảnh hưởng, từ đó tạo bản đồ chi tiết về mức độ thiệt hại.

2. Ứng dụng:

Bản đồ phát hiện lũ quét có giá trị lớn trong các hệ thống cảnh báo sớm (Early Warning Systems - EWS). Chúng cung cấp thông tin thời gian thực về vị trí và phạm vi lũ quét, giúp các nhà quản lý triển khai các biện pháp ứng phó khẩn cấp như sơ tán dân cư hoặc bảo vệ cơ sở hạ tầng. Ví dụ, trong nghiên cứu “Nowcasting for urban flash floods in Africa” (2021) [63], mô hình TrajGRU kết hợp với dữ liệu MSG-SEVIRI đã tạo ra bản đồ dự báo ngắn hạn (0-2 giờ) cho Ghana, hỗ trợ cảnh báo lũ quét đô thị với độ phân giải 3 km.

3. Ưu điểm và hạn chế:

Ưu điểm của bản đồ phát hiện lũ quét là khả năng xử lý dữ liệu lớn và cung cấp kết quả nhanh chóng, đặc biệt khi kết hợp AI với dữ liệu vệ tinh. Tuy nhiên, hạn chế nằm ở độ chính xác phụ thuộc vào chất lượng dữ liệu đầu vào (ví dụ: độ phân giải của hình ảnh vệ tinh) và khả năng phân biệt lũ quét với các loại ngập lụt khác. Ngoài ra, các mô hình DL như CNN yêu cầu dữ liệu huấn luyện phong phú, điều này có thể là thách thức ở các khu vực thiếu dữ liệu lịch sử.

1.3.2 Bản đồ nhạy cảm với lũ quét

Bản đồ nhạy cảm với lũ quét (flash flood susceptibility maps) là sản phẩm dự đoán các khu vực có khả năng xảy ra lũ quét dựa trên các yếu tố điều kiện địa lý, thủy văn và khí tượng. Đây là công cụ quan trọng trong việc lập kế hoạch dài hạn và quản lý rủi ro, giúp xác định các vùng dễ bị tổn thương để ưu tiên các biện pháp giảm thiểu. Các nghiên cứu như “Flash-Flood Susceptibility Modeling Using New Approaches of Hybrid and Ensemble Tree-Based Machine Learning Algorithms” (2020) [64] đã sử dụng RF và Extremely Randomized Trees (ERT) để lập bản đồ độ nhạy lũ quét ở lưu vực Kalvan, Iran, đạt AUC lên đến 0.82.

1. Quy trình phát triển:

Để xây dựng bản đồ này, các nhà nghiên cứu thu thập dữ liệu địa không gian như độ dốc, độ cao, khoảng cách đến sông, lượng mưa, và sử dụng đất từ các nguồn như DEM (Digital Elevation Model), Landsat, hoặc dữ liệu khí tượng vệ tinh. Các yếu tố này được đưa vào các mô hình ML/DL như RF, SVM, hoặc DNN để tính toán chỉ số độ nhạy lũ quét (Flash-Flood Potential Index - FFPI). Nghiên cứu “Flash-Flood Potential Mapping Using Deep Learning, Alternating Decision Trees and Data Provided by

Remote Sensing Sensors” (2021) [65] đã sử dụng DNN-WOE với AUC 0.96, kết hợp dữ liệu từ Google Earth và các yếu tố địa hình để lập bản đồ ở lưu vực Bâscă Chiojdului, Romania, cho thấy 59.38% khu vực có độ nhạy cao.

2. Ứng dụng:

Bản đồ nhạy cảm với lũ quét được sử dụng để xác định các khu vực cần ưu tiên trong quy hoạch đô thị, xây dựng hệ thống thoát nước, hoặc bảo vệ đất nông nghiệp. Ví dụ, nghiên cứu “Flash flood susceptibility mapping using stacking ensemble machine learning models” (2022) [66] đã tạo bản đồ cho một khu vực ở Hy Lạp, hỗ trợ các nhà quản lý trong việc giảm thiểu thiệt hại kinh tế và môi trường. Chúng cũng là cơ sở để đánh giá tác động của biến đổi khí hậu lên tần suất lũ quét.

3. Ưu điểm và hạn chế:

Ưu điểm của bản đồ này là khả năng tích hợp nhiều yếu tố điều kiện để đưa ra dự đoán không gian chính xác, đặc biệt khi sử dụng các mô hình ensemble như RF hoặc XGBoost. Tuy nhiên, hạn chế bao gồm việc phụ thuộc vào dữ liệu lịch sử lũ quét để huấn luyện mô hình, điều này có thể không khả thi ở các khu vực chưa từng ghi nhận sự kiện lũ. Ngoài ra, việc lựa chọn các yếu tố đầu vào đôi khi thiếu cơ sở lý thuyết rõ ràng, dẫn đến sự không nhất quán giữa các nghiên cứu.

1.3.3 Bản đồ nguy cơ lũ quét

Bản đồ nguy cơ lũ quét (flash flood hazard maps) là sản phẩm phân tích không gian nhằm xác định các khu vực có khả năng xảy ra lũ quét trong những điều kiện cụ thể, chẳng hạn như lượng mưa cực đại, đặc điểm địa hình, hoặc đặc tính thủy văn. Khác với bản đồ rủi ro, bản đồ nguy cơ tập trung hoàn toàn vào xác suất và phạm vi xảy ra của hiện tượng lũ quét, không xem xét đến các tác động tiềm tàng đối với con người, cơ sở hạ tầng, hay các yếu tố kinh tế-xã hội. Đây là công cụ nền tảng trong việc dự báo và lập kế hoạch giảm thiểu lũ quét, cung cấp thông tin quan trọng để các nhà quản lý hiểu rõ những khu vực nào dễ bị ảnh hưởng bởi sự kiện lũ trong các kịch bản thời tiết nhất định. Các nghiên cứu như “Google Earth Engine and Machine Learning for Flash Flood Exposure Mapping—Case Study: Tetouan, Morocco” (2024) [67] đã sử dụng Random Forest (RF) với độ chính xác 96% để lập bản đồ nguy cơ, tập trung vào các yếu tố tự nhiên như lượng mưa và độ dốc.

1. Quy trình phát triển:

Quá trình xây dựng bản đồ nguy cơ lũ quét bắt đầu bằng việc thu thập dữ liệu địa không gian, bao gồm các yếu tố như độ cao (elevation), độ dốc (slope), khoảng cách đến sông (distance to river), lượng mưa (rainfall), và chỉ số độ ẩm địa hình (Topographic Wetness Index - TWI). Dữ liệu này thường được lấy từ Digital Elevation Model (DEM), hình ảnh vệ tinh (Sentinel-1, Landsat), hoặc sản phẩm mưa vệ tinh như IMERG. Các

mô hình trí tuệ nhân tạo (AI) như RF, Support Vector Machine (SVM), hoặc Deep Neural Network (DNN) được áp dụng để phân tích mối quan hệ giữa các yếu tố đầu vào và khả năng xảy ra lũ quét. Ví dụ, nghiên cứu “Flash-Flood Potential Mapping Using Deep Learning, Alternating Decision Trees and Data Provided by Remote Sensing Sensors” (2021) [65] đã sử dụng DNN kết hợp với Weights of Evidence (WOE) để tạo bản đồ nguy cơ cho lưu vực Bâscă Chiojdului, Romania, đạt AUC 0.96, xác định 59.38% khu vực có nguy cơ cao dựa trên các điều kiện địa hình và lượng mưa.

Các mô hình AI thường được huấn luyện với dữ liệu lịch sử lũ quét để dự đoán xác suất xảy ra trong các điều kiện cụ thể, chẳng hạn như mưa lớn kéo dài 6 giờ với cường độ 50 mm/giờ. Một số nghiên cứu khác, như “Flash flood susceptibility mapping using stacking ensemble machine learning models” (2022) [66], đã sử dụng mô hình ensemble (RF kết hợp XGBoost) để lập bản đồ nguy cơ với độ chính xác cao, tập trung vào các yếu tố tự nhiên mà không đưa vào dữ liệu về dân số hay cơ sở hạ tầng.

2. Ứng dụng:

Bản đồ nguy cơ lũ quét đóng vai trò quan trọng trong việc dự báo và lập kế hoạch phòng chống thiên tai ở cấp độ khu vực. Chúng giúp xác định các khu vực cần theo dõi sát sao trong mùa mưa hoặc khi có dự báo thời tiết cực đoan, từ đó hỗ trợ triển khai các biện pháp phòng ngừa như xây dựng hệ thống thoát nước hoặc cảnh báo sớm. Ví dụ, nghiên cứu “Flash Flood Detection and Susceptibility Mapping in the Monsoon Period” (2022) [68] đã tạo bản đồ nguy cơ cho một khu vực ở Ấn Độ, cung cấp thông tin về các vùng có nguy cơ cao để điều chỉnh chiến lược quản lý nước trong mùa gió mùa. Bản đồ này cũng là nền tảng để phát triển các bản đồ rủi ro khi kết hợp với dữ liệu về tác động và thiệt hại.

3. Ưu điểm và hạn chế:

Ưu điểm của bản đồ nguy cơ lũ quét là khả năng cung cấp thông tin không gian chính xác về các khu vực tiềm ẩn lũ quét dựa trên dữ liệu tự nhiên, với sự hỗ trợ của các mô hình AI như RF hoặc DNN mang lại độ tin cậy cao (AUC thường >0.90). Chúng đơn giản hơn bản đồ rủi ro vì không yêu cầu dữ liệu phức tạp về kinh tế-xã hội, do đó dễ triển khai ở các khu vực thiếu thông tin chi tiết. Tuy nhiên, hạn chế nằm ở chỗ bản đồ này không phản ánh đầy đủ mức độ nghiêm trọng của lũ quét đối với cộng đồng hoặc cơ sở hạ tầng, mà chỉ dừng lại ở việc chỉ ra khả năng xảy ra. Ngoài ra, độ chính xác phụ thuộc vào chất lượng dữ liệu đầu vào (ví dụ: độ phân giải của DEM) và khả năng mô phỏng các điều kiện thời tiết cụ thể, vốn có thể thay đổi do biến đổi khí hậu.

Bản đồ nguy cơ lũ quét là sản phẩm quan trọng trong việc xác định các khu vực có khả năng xảy ra lũ quét dưới các điều kiện cụ thể, dựa trên dữ liệu địa không gian và các mô hình AI như RF, SVM, hoặc DNN. Với vai trò tập trung vào yếu tố tự nhiên, nó cung cấp nền tảng cho dự báo và phòng ngừa lũ quét, đồng thời là bước đệm để phát

triển bản đồ rủi ro khi cần đánh giá tác động rộng hơn. Mặc dù đơn giản và hiệu quả, bản đồ nguy cơ cần được cập nhật thường xuyên để phản ánh các thay đổi trong điều kiện khí hậu và địa hình, đảm bảo tính ứng dụng trong quản lý thiên tai dài hạn.

1.4. Đánh giá các phương pháp xác định lũ quét và dữ liệu sử dụng

Để đánh giá bao quát hơn về các phương pháp/mô hình được sử dụng trong lũ quét và dữ liệu sử dụng, nghiên cứu đã phân tích, đánh giá gần 500 nghiên cứu trên toàn thế giới dựa trên cơ sở dữ liệu CrossRef với các từ khóa bao gồm: flash flood; susceptibility; hazard; machine learning; deep learning. Kết quả xác định như sau:

1.4.1 Các phương pháp phổ biến xác định lũ quét

Các mô hình thành ba nhóm: Học máy truyền thống (ML), Học sâu (DL), và Khác (Other) (bao gồm các phương pháp thống kê hoặc lai ghép không thuộc ML/DL thuận tự). Dưới đây là bảng tổng hợp các mô hình được sử dụng, mức độ phổ biến, và hiệu suất:

Bảng 6. Tổng hợp các mô hình được sử dụng để xác định lũ quét

Loại mô hình	Mô hình cụ thể	Số lần xuất hiện	Hiệu suất nổi bật	Nhận xét
ML	Random Forest (RF)	68	AUC: 0.86–0.98, Accuracy: 81–96%	Phổ biến nhất, hiệu quả cao trong lập bản đồ độ nhạy lũ và dự báo.
	Support Vector Machine (SVM)	46	AUC: 0.75–0.96, Accuracy: ~90%	Hiệu quả trong phân loại, nhưng cần tối ưu hóa tham số.
	Logistic Regression (LR)	38	AUC: 0.72–0.93, Accuracy: ~75–90%	Đơn giản, phù hợp cho dữ liệu tuyến tính, hiệu suất trung bình.
	Decision Tree (DT)	24	AUC: ~0.93	Đã hiểu, nhưng dễ bị overfitting nếu không được tối ưu.
	Classification and Regression Trees (CART)	20	AUC: ~0.91	Phù hợp cho dữ liệu phi tuyến, hiệu suất khá.
	XGBoost	18	AUC: 0.85–0.98	Hiệu suất cao, đặc biệt trong các mô hình ensemble.
	K-Nearest Neighbors (kNN)	16	AUC: 0.75–0.91	Đơn giản, hiệu quả với dữ liệu nhỏ, nhạy cảm với nhiễu.
	Naïve Bayes (NB)	14	AUC: 0.76–0.83	Phù hợp cho phân loại, hiệu suất trung bình.
	Alternating Decision Tree (ADT)	10	AUC: ~0.95	Ít phổ biến, nhưng hiệu quả trong một số nghiên cứu.
	Boosted Regression Tree (BRT)	10	AUC: ~0.82	Hiệu quả trong độ nhạy lũ, nhưng ít được sử dụng hơn RF.

Loại mô hình	Mô hình cụ thể	Số lần xuất hiện	Hiệu suất nổi bật	Nhận xét
	LightGBM	6	KGE: >0.9	Hiệu suất cao trong dự báo dòng chảy ngắn hạn.
	Extreme Learning Machine (ELM)	5	AUC: ~0.90	Nhanh, nhưng ít phổ biến trong lũ quét.
	Gradient Boosting Machine (GBM)	4	AUC: ~0.90	Hiệu quả, nhưng ít được sử dụng hơn XGBoost.
	GLMnet	3	AUC: ~0.94	Ít phổ biến, dùng trong hồi quy tuyến tính tổng quát.
	TreeBag	3	AUC: ~0.94	Ít được sử dụng, hiệu suất khá.
	Functional Tree (FT)	2	AUC: ~0.95	Ít phổ biến, hiệu quả trong một số trường hợp cụ thể.
	Kernel Logistic Regression (KLR)	2	AUC: ~0.95	Biến thể của LR, ít được sử dụng.
	Quadratic Discriminant Analysis (QDA)	2	AUC: ~0.95	Ít phổ biến, dùng trong phân loại.
DL	Deep Neural Network (DNN)	28	AUC: 0.92–0.96, Accuracy: >90%	Phổ biến trong các nghiên cứu gần đây, hiệu suất cao với dữ liệu phức tạp.
	Long Short-Term Memory (LSTM)	22	NSE: 0.80–0.87, AUC: ~0.90	Hiệu quả trong dự báo chuỗi thời gian (time-series forecasting).
	Convolutional Neural Network (CNN)	16	AUC: ~0.935, Dice: ~79.75%	Hiệu quả trong phân tích ảnh vệ tinh và lập bản đồ lũ.
	Deep Belief Network (DBN)	6	AUC: ~0.90	Ít phổ biến, nhưng hiệu quả với dữ liệu phức tạp.
	U-Net	5	Dice: ~79.75%	Chủ yếu dùng cho phân đoạn ảnh vệ tinh (segmentation).
	Deep 1D-CNN	4	AUC: 0.969, Accuracy: 90.2%	Hiệu suất rất cao trong dự báo không gian.
	PSO-BP Neural Network	3	MAE: 2.51%, RMSE: 3.74%	Kết hợp tối ưu hóa, hiệu quả trong dự báo lưu lượng đỉnh.
	MobileNet CNN	2	Accuracy: ~90%	Dùng cho phân loại hình ảnh lũ, hiệu quả cao.
	TrajGRU	1	-	Dùng trong dự báo ngắn hạn (nowcasting), ít phổ biến.
Other	Frequency Ratio (FR)	32	AUC: ~0.72–0.93	Phương pháp thống kê, thường kết hợp với ML/DL để tăng hiệu suất.
	Analytical Hierarchy Process (AHP)	26	AUC: ~0.80–0.845	Phương pháp ra quyết định đa tiêu chí, thường kết hợp với ML.
	Weights of Evidence (WOE)	12	AUC: ~0.96	Thống kê, hiệu quả khi kết hợp với DL (như DLNN-WOE).
	Fuzzy Logic	10	AUC: ~0.90	Xử lý dữ liệu không chắc chắn, thường kết hợp với ML/DL.

Loại mô hình	Mô hình cụ thể	Số lần xuất hiện	Hiệu suất nổi bật	Nhận xét
	Monte Carlo	5	-	Dùng trong mô phỏng xác suất, ít phổ biến trong lũ quét.
	Reinforcement Learning	3	-	Mới xuất hiện, tiềm năng trong cảnh báo thời gian thực.
	Bayesian Belief Network (BBN)	3	AUC: ~0.90	Ít phổ biến, dùng trong mô hình xác suất.
	Hydrological Models (HEC-HMS, GR4H)	15	NSE: 0.56–0.94	Mô hình thủy văn truyền thống, thường kết hợp với ML/DL.
	Statistical Methods (khác)	10	-	Các phương pháp thống kê chung, ít được chỉ rõ.

Các mô hình học máy phổ biến nhất: Random Forest (RF) là mô hình được sử dụng nhiều nhất (68 lần), nhờ tính linh hoạt, khả năng xử lý dữ liệu phi tuyến, và hiệu suất cao (AUC thường >0.90). SVM (46 lần) và Logistic Regression (38 lần) cũng rất phổ biến, đặc biệt trong các nghiên cứu về lập bản đồ độ nhạy lũ (susceptibility mapping). XGBoost (18 lần) và LightGBM (6 lần) đang nổi lên như các lựa chọn thay thế mạnh mẽ cho RF, với hiệu suất cạnh tranh.

Các mô hình học sâu phổ biến nhất: Deep Neural Network (DNN) (28 lần) và Long Short-Term Memory (LSTM) (22 lần) là các mô hình DL được sử dụng nhiều nhất, đặc biệt trong các nghiên cứu từ 2020 trở đi. Convolutional Neural Network (CNN) (16 lần) và Deep 1D-CNN (4 lần) cho thấy hiệu quả vượt trội trong xử lý dữ liệu hình ảnh vệ tinh và dự báo không gian (AUC lên đến 0.969).

Các mô hình/phương pháp khác: Frequency Ratio (FR) (32 lần) và Analytical Hierarchy Process (AHP) (26 lần) là các phương pháp thống kê phổ biến, thường được dùng để xác định trọng số của các yếu tố đầu vào hoặc kết hợp với ML/DL. Hydrological Models như HEC-HMS và GR4H (15 lần) vẫn được sử dụng, đặc biệt trong các nghiên cứu kết hợp với ML/DL để cải thiện dự báo.

Hiệu suất của các phương pháp/mô hình: Các mô hình ensemble (như RF, XGBoost, DNN-FR) thường đạt AUC >0.90 , vượt trội so với các mô hình đơn lẻ như LR hoặc NB. Trong DL, Deep 1D-CNN ghi nhận AUC cao nhất (0.969), nhưng chỉ xuất hiện trong một số nghiên cứu cụ thể. LSTM cho thấy hiệu suất tốt trong dự báo chuỗi thời gian (NSE: 0.80–0.87).

Xu hướng: Các mô hình DL (LSTM, CNN) và ensemble (RF, XGBoost) ngày càng được ưa chuộng, đặc biệt trong các khu vực thiếu dữ liệu (ungauged basins) hoặc cần xử lý dữ liệu phức tạp (hình ảnh vệ tinh, chuỗi thời gian). RF, SVM, LR, DNN, và LSTM là những mô hình chủ đạo, với RF dẫn đầu về số lần xuất hiện và hiệu suất ổn định.

1.4.2 Dữ liệu sử dụng

Các yếu tố đầu vào (input factors) được sử dụng trong các nghiên cứu rất đa dạng, bao gồm các đặc trưng địa lý, thủy văn, khí tượng, và nhân tạo. Dưới đây là danh sách tổng hợp các yếu tố phổ biến:

Bảng 7. Tổng hợp các dữ liệu phổ biến sử dụng trong nghiên cứu lũ quét

Dữ liệu	Mô tả	Loại dữ liệu
Slope (Độ dốc)	Góc nghiêng của bề mặt địa hình, ảnh hưởng đến tốc độ dòng chảy.	Địa hình (Topographic)
Elevation (Độ cao)	Độ cao so với mực nước biển, ảnh hưởng đến phân bố mưa và dòng chảy.	Địa hình
Aspect (Hướng dốc)	Hướng của độ dốc, ảnh hưởng đến tiếp xúc ánh sáng và độ ẩm đất.	Địa hình
Distance to River (Khoảng cách đến sông)	Khoảng cách từ điểm nghiên cứu đến sông gần nhất, liên quan đến nguy cơ lũ.	Thủy văn (Hydrological)
Topographic Wetness Index (TWI)	Chỉ số độ ẩm địa hình, biểu thị khả năng tích tụ nước.	Thủy văn
Stream Power Index (SPI)	Chỉ số sức mạnh dòng chảy, liên quan đến năng lượng dòng chảy.	Thủy văn
Topographic Position Index (TPI)	Chỉ số vị trí địa hình, xác định vị trí tương đối (đỉnh, sườn, thung lũng).	Địa hình
Rainfall (Lượng mưa)	Lượng mưa tích lũy hoặc cường độ mưa, yếu tố chính gây lũ quét.	Khí tượng (Meteorological)
Land Use/Land Cover (LULC)	Loại sử dụng đất (nông nghiệp, đô thị, rừng), ảnh hưởng đến khả năng thấm nước.	Nhân tạo (Anthropogenic)
Soil Type (Loại đất)	Đặc tính đất (đất sét, cát, v.v.), ảnh hưởng đến khả năng thấm và dòng chảy.	Địa chất (Geological)
Lithology (Thạch học)	Lớp bề mặt, ảnh hưởng đến khả năng thấm nước và xói mòn.	Địa chất
Stream Density (Mật độ sông)	Mật độ mạng lưới sông trong khu vực, liên quan đến thoát nước.	Thủy văn
Curvature (Độ cong)	Độ cong của địa hình (lồi, lõm), ảnh hưởng đến dòng chảy và tích tụ nước.	Địa hình
Profile Curvature	Độ cong theo hướng dốc, ảnh hưởng đến tốc độ dòng chảy.	Địa hình
Plan Curvature	Độ cong ngang, ảnh hưởng đến hướng dòng chảy.	Địa hình
Distance to Road (Khoảng cách đến đường)	Khoảng cách đến đường, liên quan đến tác động nhân tạo và khả năng tiếp cận.	Nhân tạo
Population Density (Mật độ dân số)	Mật độ dân cư, ảnh hưởng đến mức độ tổn thương và thiệt hại.	Nhân tạo
River Length (Chiều dài sông)	Tổng chiều dài sông trong lưu vực, liên quan đến dòng chảy.	Thủy văn
Basin Area (Diện tích lưu vực)	Diện tích lưu vực, ảnh hưởng đến lượng nước tập trung.	Thủy văn
Soil Moisture (Độ ẩm đất)	Độ ẩm của đất, ảnh hưởng đến khả năng thấm nước và dòng chảy bề mặt.	Thủy văn

Dữ liệu	Mô tả	Loại dữ liệu
Gully Density (Mật độ rãnh xói mòn)	Mật độ các rãnh xói mòn, liên quan đến nguy cơ lũ quét.	Địa hình
Normalized Difference Built-up Index (NDBI)	Chỉ số xây dựng, biểu thị mức độ đô thị hóa.	Nhân tạo
Temperature (Nhiệt độ)	Nhiệt độ môi trường, ảnh hưởng đến bốc hơi và độ ẩm đất.	Khí tượng
Vegetation Cover (Độ phủ thực vật)	Mức độ che phủ thực vật, ảnh hưởng đến xói mòn và thâm nước.	Nhân tạo
Precipitation Estimates (Uớc lượng mưa)	Dữ liệu mưa từ vệ tinh (IMERG, PERSIANN), dùng khi thiếu trạm đo.	Khí tượng
Discharge (Lưu lượng)	Lưu lượng nước trong sông, yếu tố trực tiếp liên quan đến lũ.	Thủy văn
Water Level (Mực nước)	Mực nước sông hoặc hồ, biểu thị tình trạng thủy văn.	Thủy văn

Các yếu tố độ dốc (slope), độ cao (elevation), khoảng cách đến sông (distance to river), lượng mưa (rainfall) và sử dụng đất (LULC) là những yếu tố được sử dụng phổ biến nhất, xuất hiện trong hầu hết các nghiên cứu, do chúng có tác động trực tiếp đến dòng chảy bề mặt và nguy cơ lũ quét.

Dữ liệu khí tượng (như lượng mưa) và địa hình (như độ dốc, độ cao) đóng vai trò quan trọng nhất trong việc dự báo lũ quét. Sử dụng đất (LULC) và Loại đất (Soil Type) ngày càng được chú trọng, đặc biệt trong các nghiên cứu về độ nhạy lũ, vì chúng ảnh hưởng đến khả năng thâm nước và xói mòn.

Dữ liệu từ vệ tinh (IMERG, PERSIANN, Sentinel) ngày càng được sử dụng phổ biến để bổ sung cho các khu vực thiếu trạm đo (ungauged basins). Một số nghiên cứu gần đây đã bắt đầu sử dụng các yếu tố nhân tạo như mật độ dân số và khoảng cách đến đường để đánh giá mức độ tổn thương (vulnerability).

1.4.3 Sự khác biệt dữ liệu giữa học máy và học sâu

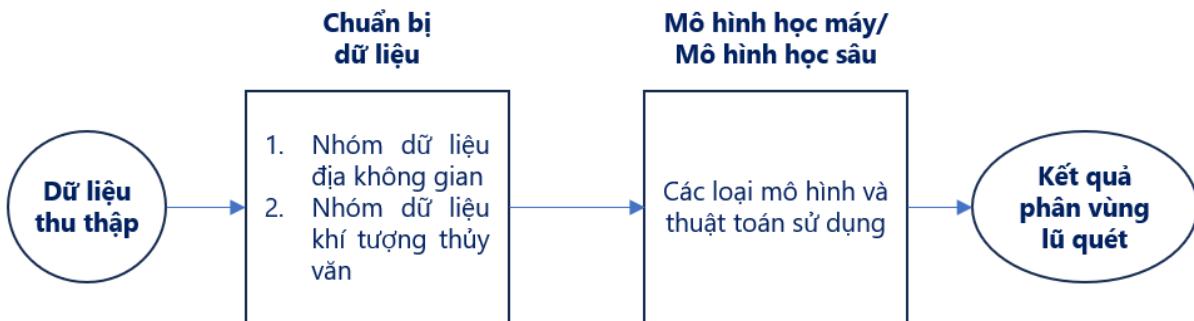
Dựa trên các tài liệu tìm hiểu, mô hình học máy và học sâu trong AI có sự khác biệt đáng kể. Chi tiết thể hiện như sau:

Bảng 8. Sự khác biệt về dữ liệu giữa hai mô hình trí tuệ nhân tạo

Tiêu chí	Học máy (Machine Learning)	Học sâu (Deep Learning)
Loại dữ liệu đầu vào	<ul style="list-style-type: none"> - Dữ liệu có cấu trúc: bảng số liệu (lượng mưa, mực nước, độ dốc địa hình, lưu lượng sông). - Đặc trưng được trích xuất thủ công (ví dụ: tổng mưa theo giờ, độ ẩm đất). 	<ul style="list-style-type: none"> - Dữ liệu thô hoặc ít xử lý: hình ảnh vệ tinh, radar thời tiết, video dòng chảy, bản đồ địa hình. - Tự động trích xuất đặc trưng từ dữ liệu như hình ảnh hoặc chuỗi thời gian.
Yêu cầu chuẩn hóa dữ liệu	<ul style="list-style-type: none"> - Bắt buộc chuẩn hóa dữ liệu số (lượng mưa, mực nước) về cùng thang đo (ví dụ: [0, 1]). - Phương pháp: Min-Max Scaling, Z-score. 	<ul style="list-style-type: none"> - Chuẩn hóa ít nghiêm ngặt hơn, nhưng cần cho hình ảnh (pixel về [0, 1]) hoặc chuỗi thời gian (chuẩn hóa biên độ). - Dữ liệu radar/hình ảnh thường tự động chuẩn hóa trong mô hình.

Tiêu chí	Học máy (Machine Learning)	Học sâu (Deep Learning)
	- Dữ liệu địa hình cần đồng nhất định dạng.	
Phạm vi giá trị	- Dữ liệu số trong phạm vi chuẩn hóa (ví dụ: lượng mưa 0-100 mm, mực nước 0-10 m). - Outliers (mưa cực lớn) cần xử lý để tránh sai lệch mô hình.	- Hình ảnh: pixel [0, 255] hoặc [0, 1]. - Chuỗi thời gian (mưa, mực nước): chuẩn hóa về [-1, 1] hoặc [0, 1]. - Outliers ít ảnh hưởng nhờ mạng sâu.
Cần gán giá trị (xử lý dữ liệu thiếu)	- Cần gán giá trị cho dữ liệu thiếu (ví dụ: lượng mưa thiếu tại trạm đo). - Phương pháp: trung bình, trung vị, hoặc mô hình dự đoán (KNN, Random Forest).	- Ít cần gán giá trị thủ công. - Mô hình sâu có thể xử lý dữ liệu thiếu qua các tầng mạng hoặc sử dụng dữ liệu thô (hình ảnh, radar) để bù đắp.
Dữ liệu có phải là số không?	- Bắt buộc là số. - Dữ liệu phi số (loại đất, vùng địa lý) cần mã hóa (One-Hot Encoding, Label Encoding).	- Không bắt buộc là số. - Hình ảnh vệ tinh, radar, hoặc video dòng chảy được xử lý trực tiếp và chuyển thành số trong quá trình huấn luyện.
Xử lý dữ liệu tạp nham (noise)	- Nhạy cảm với nhiễu (ví dụ: dữ liệu mưa sai lệch từ cảm biến). - Cần làm sạch dữ liệu (loại bỏ outliers, lọc nhiễu) trước khi huấn luyện.	- Chịu nhiễu tốt hơn nhờ kiến trúc mạng sâu. - Có thể học từ dữ liệu radar/hình ảnh có nhiễu, nhưng nhiễu quá lớn vẫn cần làm sạch cơ bản.
Kích thước dữ liệu	- Hoạt động tốt với dữ liệu nhỏ/trung bình (ví dụ: vài nghìn bản ghi từ trạm đo mưa, mực nước). - Hiệu quả phụ thuộc vào chất lượng đặc trưng thủ công.	- Yêu cầu dữ liệu lớn (hàng chục nghìn hình ảnh vệ tinh, chuỗi thời gian dài). - Dữ liệu nhỏ dễ gây overfitting, đặc biệt với mạng nơ-ron sâu.
Loại bài toán chính	- Dự báo lũ quét (hồi quy: dự đoán mực nước, phân loại: nguy cơ lũ cao/thấp). - Phản phẩm chính: dự báo mưa, phân tích rủi ro lũ. - Phân tích cụm (clustering) vùng nguy cơ lũ.	- Dự báo lũ quét phức tạp (dự đoán thời gian, vị trí lũ). - Nhận diện vùng ngập từ hình ảnh vệ tinh. - Phân tích chuỗi thời gian mưa/mực nước. - Mô phỏng dòng chảy lũ bằng mạng nơ-ron.
Công cụ xử lý dữ liệu	- Thư viện: Scikit-learn, Pandas, NumPy. - Công cụ trích xuất đặc trưng: thống kê (tổng mưa, trung bình mực nước), GIS. - Làm sạch: lọc nhiễu, quy tắc kinh doanh.	- Thư viện: TensorFlow, PyTorch, Keras. - Công cụ tiền xử lý: OpenCV (hình ảnh vệ tinh), xarray (dữ liệu radar), Hugging Face (NLP cho báo cáo lũ). - Tiền xử lý tập trung vào định dạng dữ liệu thô.
Ví dụ ứng dụng dữ liệu	- Dự báo nguy cơ lũ quét dựa trên dữ liệu bảng (lượng mưa, mực nước, độ dốc). - Phân loại vùng nguy cơ lũ (cao/thấp) dựa trên đặc trưng địa hình, đất đai.	- Dự báo lũ quét từ hình ảnh vệ tinh và radar thời tiết. - Phân tích video dòng chảy để phát hiện lũ quét. - Dự đoán thời gian lũ dựa trên chuỗi thời gian mưa/mực nước.

CHƯƠNG 2. PHƯƠNG PHÁP PHÂN VÙNG LŨ QUÉT ỨNG DỤNG TRÍ TUỆ NHÂN TẠO VÀ DỮ LIỆU ĐỊA KHÔNG GIAN



Hình 18. Sơ đồ tiếp cận trong nghiên cứu

Hình 18 thể hiện sơ đồ tiếp cận trong việc ứng dụng trí tuệ nhân tạo và dữ liệu địa không gian để phân vùng lũ quét. Bốn khối công việc được thể hiện bao gồm:

2.1. Phương pháp thu thập dữ liệu

Phương pháp thu thập dữ liệu giúp nghiên cứu có đầy đủ bộ dữ liệu để phục vụ đánh giá và phân vùng lũ quét cho khu vực thí điểm (Mù Cang Chải), bên cạnh đó, công tác thu thập dữ liệu nghiên cứu trước đây sẽ góp phần làm tăng cường nhận thức và nâng cao hiểu biết về lũ quét, trí tuệ nhân tạo trong nghiên cứu lũ quét và cách thức triển khai thực hiện. Các dữ liệu thu thập dữ liệu được thể hiện chi tiết trong mục **Error! Reference source not found.**, và công tác chuẩn bị dữ liệu được thể hiện chi tiết trong mục **Error! Reference source not found..**

2.2. Phương pháp nghiên cứu GIS

Phương pháp nghiên cứu GIS là một tập hợp các phương pháp diễn toán các dữ liệu địa không gian nhằm xây dựng các bản đồ có nghĩa trong một lĩnh vực hay một phạm vi nào đó. Ví dụ đơn giản nhất là dữ liệu DEM địa hình (có thể được thu thập bằng khảo sát hoặc từ ảnh vệ tinh) là đầu vào để xác định các dữ liệu có liên quan bao gồm độ dốc, độ cong địa hình hoặc địa mạo...

Trong nghiên cứu về lũ quét, các loại dữ liệu địa không gian cơ bản bao gồm: (1) dữ liệu địa hình; (2) các dữ liệu ảnh vệ tinh; và (3) dữ liệu khí tượng thủy văn.

2.3. Phương pháp học máy

1. Lựa chọn mô hình, dữ liệu và phương pháp chung để chuẩn hóa dữ liệu

Nghiên cứu sử dụng bốn mô hình học máy để đánh giá nguy cơ lũ quét dưới dạng bài toán phân loại: Rừng ngẫu nhiên (Random Forest - RF), Máy hỗ trợ vectơ (Support Vector Machine - SVM), Hồi quy Logistic (Logistic Regression - LR) và LightGBM - LGBM. Các mô hình này được chọn dựa trên khả năng xử lý dữ liệu phức tạp và phù hợp với đầu ra là một lớp về phân loại lũ.

Các nhóm dữ liệu đầu vào bao gồm 4 nhóm chính:

- Đặc trưng địa hình: bao gồm bản đồ mô hình số độ cao và các sản phẩm từ mô hình số độ cao như độ dốc, chỉ số ẩm địa hình, độ cong địa hình...
- Đặc trưng thủy văn: bao gồm khoảng cách/chênh lệch độ cao so với sông/suối, độ dốc lòng dẫn, chiều dài dòng chảy, diện tích lưu vực...
- Đặc trưng thực phủ: bao gồm các chỉ số liên quan đến bề mặt như NDVI (liên quan đến thảm phủ); chỉ số CN (liên quan đến thảm phủ và đặc trưng địa chất).
- Đặc trưng khí tượng: bao gồm các dữ liệu đặc trưng mưa như lượng mưa giờ lớn nhất, tổng lượng mưa 3, 6 giờ lớn nhất...

Chuẩn hóa đặc trưng: Thay vì sử dụng các đặc trưng định tính như loại đất hoặc loại hình sử dụng đất, nghiên cứu chọn các chỉ số định lượng như tốc độ thấm, NDVI, và CN để đại diện cho đất và thảm phủ. Điều này giúp giảm thiểu việc mã hóa phức tạp và tăng tính chính xác.

Phương pháp chuẩn hóa: Với các đặc trưng định tính (nếu có), các phương pháp như trọng số dẫn chứng (WOE) hoặc tỷ lệ tần suất (FR) có thể được sử dụng. Tuy nhiên, các phương pháp này yêu cầu dữ liệu lớn và đa dạng để đảm bảo độ tin cậy.

Làm sạch dữ liệu: Dữ liệu nhiễu (như các điểm không đại diện cho lũ quét) được loại bỏ trong quá trình tiền xử lý, đảm bảo chất lượng đầu vào.

Nghiên cứu xem lũ quét như phản ứng thủy văn của một lưu vực, tập trung vào các điểm ở cửa ra lưu vực thay vì các vị trí không đặc trưng (như mái nhà, đỉnh núi). Điều này phản ánh thực tế rằng lũ quét thường tích tụ ở thượng nguồn và gây hậu quả ở hạ du, giúp mô hình dự đoán chính xác hơn.

Không giống SVM và LR, RF và LGBM có khả năng xử lý dữ liệu định tính (nếu cần) mà không yêu cầu mã hóa phức tạp. Các đặc trưng định tính có thể được mã hóa bằng One-Hot Encoding hoặc mã hóa ngẫu nhiên. RF hoạt động dựa trên việc phân chia dữ liệu theo các đặc trưng tối ưu, tận dụng ý tưởng từ thống kê (như xác suất mẫu) nhưng được nâng cấp thông qua kỹ thuật lấy mẫu ngẫu nhiên và kết hợp nhiều cây quyết định.

Bằng việc kết hợp kiến thức thủy văn và học máy, nghiên cứu xây dựng một mô hình hồi quy đánh giá nguy cơ lũ quét phù hợp với các khu vực miền núi. Các đặc trưng được chọn đảm bảo tính khoa học, trong khi quá trình xử lý dữ liệu tăng cường độ tin cậy của mô hình.

2. Các mô hình sử dụng và nguyên tắc xác định lũ quét

a. Rừng ngẫu nhiên (Random Forest - RF)

Rừng ngẫu nhiên là một mô hình kết hợp nhiều cây quyết định, mỗi cây đưa ra một dự đoán riêng, sau đó lấy trung bình để có kết quả cuối cùng. Mỗi cây quyết định hoạt động như một chuỗi câu hỏi dựa trên các đặc trưng dữ liệu.

Ví dụ: Một cây quyết định có thể hỏi:

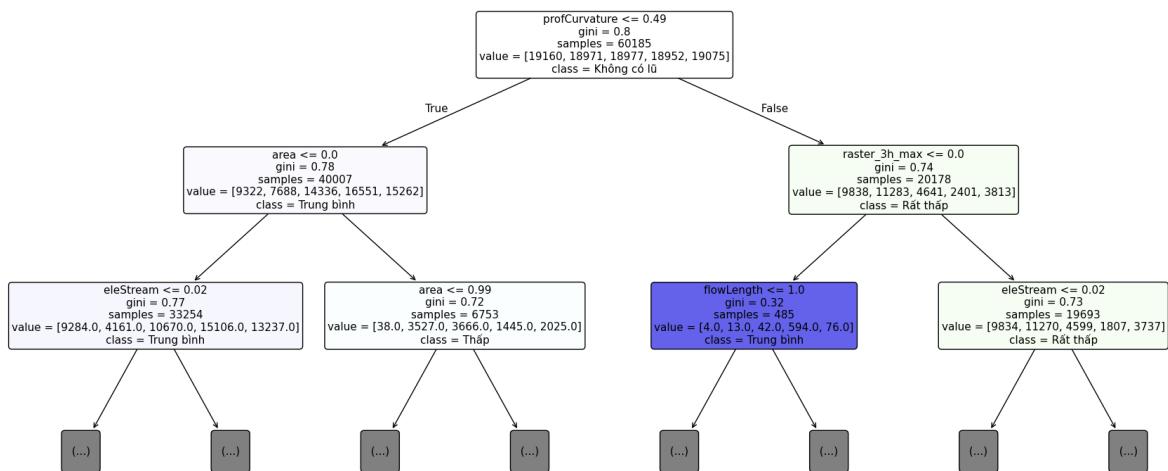
- Lượng mưa ngày lớn nhất có vượt quá 100 mm không?
- Nếu có, độ dốc bình quân lưu vực có lớn hơn 20 độ không?
- Nếu cả hai đều đúng, dự đoán nguy cơ lũ quét là 80/100.

Rừng ngẫu nhiên tạo ra hàng trăm cây như vậy, mỗi cây sử dụng một phần dữ liệu ngẫu nhiên (như lượng mưa, NDVI, hoặc tốc độ thẩm thấu), rồi kết hợp kết quả để tăng độ chính xác.

- **Nguyên tắc dự đoán nguy cơ lũ quét**

- Phân tích nhiều đặc trưng: Mô hình xem xét đồng thời toàn bộ đặc trưng (lượng mưa, độ dốc, NDVI, v.v.) và tự động xác định đặc trưng nào quan trọng nhất (ví dụ: lượng mưa lớn thường làm tăng nguy cơ lũ).
- Kết hợp dự đoán: Bằng cách lấy trung bình từ nhiều cây quyết định, mô hình giảm thiểu sai sót và đưa ra dự đoán đáng tin cậy hơn.
- Đầu ra số: Kết quả là một giá trị từ 0 đến 100, biểu thị mức độ nguy cơ lũ quét tại cửa ra của lưu vực.

Rừng ngẫu nhiên rất hiệu quả khi xử lý dữ liệu phức tạp, như các yếu tố thủy văn (mưa, độ dốc, thẩm phủ). Nó có thể phát hiện các mối quan hệ không đơn giản, ví dụ: nguy cơ lũ tăng mạnh khi lượng mưa lớn kết hợp với tốc độ thẩm thấu thấp. Tuy nhiên, mô hình này khó giải thích chi tiết lý do đưa ra một con số cụ thể, vì nó dựa trên nhiều cây quyết định.



Hình 19. Hình mô tả sự phân loại cho 5 lớp (cây có độ sâu là 2)

Ví dụ trong một cây được thể hiện như hình trên (với dữ liệu đã được chuẩn hóa về khoảng từ -1 đến 1), tại nút gốc (trên cùng), cây này sẽ xem xét độ cong theo phương dốc của địa hình và so sánh với ngưỡng 0,49. Khi đó, số lượng mẫu phân loại cho 5 lớp

được thể hiện trong value (có hơn 19 nghìn mẫu ở lớp 0 – không có lũ; gần 19 nghìn mẫu ở lớp 1 – nguy cơ rất thấp; gần 19 nghìn mẫu ở lớp 2 – nguy cơ thấp....). Số lượng mẫu ở lớp 0 – không có lũ là lớn nhất, do đó tại nút này sẽ phân loại là không có lũ. Tuy nhiên, phân loại này bị phiến diện do các lớp khác cũng có số lượng tương tự, không quá chênh lệch. Do đó, tiếp tục xét đến tầng thứ hai, tại tầng này, nếu độ cong địa hình theo phương dốc nhỏ hơn 0,49, mô hình sẽ xét tiếp diện tích lưu vực và nếu độ cong lớn hơn 0,49, mô hình sẽ xét tổng lượng mưa 3 giờ lớn nhất. Cứ như vậy cho đến khi kết thúc. Một mô hình thường có rất nhiều cây (nên gọi là rừng), số lượng cây trong mô hình được thể hiện qua tham số n_n_estimators, trong khi đó, số tầng (hay còn gọi là độ sâu cây) được thể hiện qua tham số max_depth. Các cây khác nhau có thuật toán phân loại khác nhau. Mỗi bộ dữ liệu đầu vào sẽ được từng cây đánh giá, sau đó cho ra kết quả dưới dạng bỏ phiếu. Tỷ lệ các cây bỏ phiếu cho một lớp nào đó càng cao thì dữ liệu tương ứng sẽ được phân vào lớp đó.

b. Máy hỗ trợ vectơ (Support Vector Machine - SVM)

Máy hỗ trợ vectơ tìm cách phân biệt các mức độ nguy cơ lũ quét bằng cách tạo ra một ranh giới trong không gian dữ liệu. Không gian này bao gồm các đặc trưng như lượng mưa, độ dốc, và chỉ số CN.

Ví dụ: SVM có thể xác định rằng các lưu vực với lượng mưa ngày lớn hơn 120 mm và độ dốc trên 25 độ có nguy cơ lũ cao (ví dụ: 85/100), trong khi các lưu vực có mưa dưới 50 mm và độ dốc thấp có nguy cơ thấp (ví dụ: 20/100). Mô hình sẽ tìm ranh giới tối ưu để phân chia các mức nguy cơ khác nhau và dự đoán một giá trị số.

- **Nguyên tắc dự đoán nguy cơ lũ quét**
- Xác định ranh giới: SVM tạo ra một ranh giới (hay mặt phẳng) trong không gian dữ liệu, sao cho các lưu vực có nguy cơ lũ cao và thấp được tách biệt rõ ràng nhất.
- Dự đoán mức độ nguy cơ: Dựa trên vị trí của một lưu vực (được xác định bởi các đặc trưng), SVM tính toán một giá trị nguy cơ từ 0 đến 100.
- Xử lý mẫu phức tạp: SVM có thể nhận diện các mối quan hệ phức tạp giữa các đặc trưng (như lượng mưa và NDVI) bằng cách sử dụng kỹ thuật biến đổi dữ liệu.

SVM có sự phù hợp với bài toán lũ quét vì nó có thể xử lý nhiều đặc trưng cùng lúc và tìm ra các mẫu nguy cơ dựa trên dữ liệu thủy văn. Mô hình này đặc biệt hiệu quả khi dữ liệu được chuẩn hóa tốt (ví dụ: lượng mưa và độ dốc được đưa về cùng thang đo). Tuy nhiên, SVM có thể chậm khi xử lý lượng dữ liệu lớn và yêu cầu điều chỉnh tham số cẩn thận, bên cạnh đó vấn đề phi tuyến trong dự đoán lũ quét có thể sẽ làm khó cho mô hình này mặc dù mô hình có thể đưa ra các mặt phẳng cong trong ranh giới phân chia.

c. Hồi quy logistic (Logistic Regression - LR)

Hồi quy Logistic là một mô hình đơn giản, sử dụng một công thức để kết hợp các đặc trưng và dự đoán nguy cơ lũ quét. Mỗi đặc trưng (như lượng mưa, độ dốc, NDVI) được gán một trọng số, thể hiện mức độ ảnh hưởng của nó.

Ví dụ: Nếu lượng mưa ngày lớn hơn 100 mm có trọng số +4, độ dốc trên 20 độ có trọng số +2, và chỉ số NDVI thấp có trọng số +1, mô hình sẽ tính tổng các trọng số này và đưa ra một con số nguy cơ (ví dụ: 70/100).

- **Nguyên tắc dự đoán nguy cơ lũ quét**
- Kết hợp đặc trưng: Mô hình gán trọng số cho từng đặc trưng (ví dụ: mưa lớn có trọng số cao hơn diện tích lưu vực) và tính toán tổng để dự đoán nguy cơ.
- Dự đoán số liệu: Kết quả là một giá trị từ 0 đến 100, được điều chỉnh để dễ hiểu, thể hiện mức độ nguy cơ lũ quét.
- Giả định tuyến tính: Mô hình cho rằng các đặc trưng ảnh hưởng đến nguy cơ lũ theo cách đơn giản (ví dụ: mưa càng lớn, nguy cơ càng cao).

Hồi quy Logistic dễ triển khai, nhanh, và dễ giải thích, giúp xác định đặc trưng nào (như lượng mưa hay độ dốc) ảnh hưởng mạnh nhất đến lũ quét. Tuy nhiên, mô hình này có thể không hiệu quả nếu các đặc trưng có mối quan hệ phức tạp, ví dụ: nguy cơ lũ chỉ tăng khi lượng mưa lớn kết hợp với tốc độ thẩm đất thấp, đặc biệt là các quan hệ phi tuyến của các dữ liệu.

d. LightGBM (Light Gradient Boosting Machine - LGBM)

LightGBM (Light Gradient Boosting Machine) là một mô hình học máy tiên tiến dựa trên gradient boosting, được thiết kế để xử lý dữ liệu lớn và phức tạp với tốc độ cao và hiệu quả. Mô hình này xây dựng nhiều cây quyết định liên tiếp, trong đó mỗi cây cố gắng sửa lỗi của các cây trước đó, từ đó cải thiện dự đoán nguy cơ lũ quét.

Ví dụ LightGBM có thể phân tích các đặc trưng như lượng mưa ngày lớn nhất, độ dốc lưu vực, chỉ số NDVI, và tốc độ thẩm đất. Nếu lượng mưa vượt quá 100 mm và độ dốc lớn hơn 20 độ, mô hình có thể dự đoán nguy cơ lũ quét cao (ví dụ: 85/100). Mỗi cây trong LightGBM học từ sai số của cây trước, giúp dự đoán chính xác hơn.

- **Nguyên tắc dự đoán nguy cơ lũ quét**
 - Xây dựng tuần tự: LightGBM tạo ra các cây quyết định tuần tự, trong đó mỗi cây tập trung vào việc sửa lỗi dự đoán của các cây trước đó bằng cách sử dụng gradient của hàm mất mát.
 - Tối ưu hóa hiệu quả: Mô hình sử dụng kỹ thuật "histogram-based" để nhóm dữ liệu thành các bin, giảm thời gian tính toán và tăng tốc độ xử lý, đặc biệt với dữ liệu lớn.

- Dự đoán số liệu: Kết quả là một giá trị từ 0 đến 100, biểu thị mức độ nguy cơ lũ quét, được tính dựa trên sự kết hợp của các cây trong mô hình.
- Xử lý đặc trưng quan trọng: LightGBM tự động xác định và ưu tiên các đặc trưng quan trọng (ví dụ: lượng mưa lớn hoặc độ dốc cao) để cải thiện độ chính xác.

LightGBM đặc biệt hiệu quả trong bài toán dự đoán nguy cơ lũ quét nhờ khả năng xử lý nhanh dữ liệu thủy văn phức tạp và phát hiện các mối quan hệ phi tuyến giữa các đặc trưng, chẳng hạn như sự kết hợp giữa lượng mưa cao và tốc độ thẩm thấu đất thấp. Mô hình này cũng hỗ trợ dữ liệu lớn và có thể được điều chỉnh thông qua các tham số như số lượng cây (`n_estimators`), độ sâu cây (`max_depth`), và tỷ lệ học (`learning_rate`). Tuy nhiên, LightGBM có thể yêu cầu điều chỉnh tham số cẩn thận để tránh hiện tượng quá khớp (overfitting) và đôi khi khó giải thích chi tiết lý do đưa ra dự đoán cụ thể.

2.4. Phương pháp học sâu

1. Lựa chọn mô hình, dữ liệu và phương pháp chuẩn hóa dữ liệu

Đối với phương pháp học sâu, nghiên cứu sử dụng 3 mô hình bao gồm Mạng nơ-ron sâu (Deep Neural Network - DNN), Mạng nơ-ron tích chập (Convolutional Neural Network - CNN), và Mạng nơ-ron hồi tiếp dài ngắn hạn (Long Short-Term Memory - LSTM) để đánh giá lũ quét cho khu vực nghiên cứu. Khác với các mô hình học máy, các dữ liệu đầu vào của mô hình học sâu đa dạng và linh hoạt hơn. Các yếu tố dữ liệu như cao độ (không thể hiện xu hướng lũ quét) cũng có thể được sử dụng để làm đầu vào cho mô hình.

2. Các mô hình sử dụng và nguyên tắc xác định lũ quét

Nghiên cứu sử dụng ba mô hình học sâu để dự đoán nguy cơ lũ quét dưới dạng bài toán phân loại: DNN, CNN và LSTM. Mỗi mô hình dựa vào các tham số đầu vào sẽ dự đoán nhãn phân loại lũ ở đầu ra dựa trên dữ liệu đào tạo và kiểm tra. Các mô hình tận dụng các đặc trưng định lượng, bao gồm lượng mưa, độ dốc lưu vực, chỉ số NDVI, và các yếu tố thủy văn khác, để phân tích và dự đoán. Dưới đây là mô tả chi tiết về từng mô hình, cách chúng hoạt động, và lý do chúng phù hợp với bài toán lũ quét.

a. Mạng nơ-ron tích chập (Convolutional Neural Network - CNN)

Mạng nơ-ron tích chập được thiết kế để phân tích dữ liệu có cấu trúc không gian, như hình ảnh, nhưng cũng có thể áp dụng cho dữ liệu số nếu được tổ chức phù hợp. Trong bài toán lũ quét, CNN xem các đặc trưng như lượng mưa, độ dốc, và NDVI như một "bản đồ" của lưu vực, tìm ra các mẫu liên quan đến nguy cơ lũ.

Ví dụ: Nếu tổ chức dữ liệu thành một bảng, với các cột là lượng mưa, độ dốc, và NDVI tại các điểm trong lưu vực, CNN sẽ phân tích bảng này để tìm các khu vực có nguy cơ cao (như nơi mưa lớn và độ dốc cao xuất hiện cùng nhau).

- **Nguyên tắc dự đoán nguy cơ lũ quét**
- Phân tích cục bộ: CNN xem xét các đặc trưng (như lượng mưa, độ dốc, chỉ số CN) trong các nhóm nhỏ, ví dụ: kiểm tra xem mưa lớn có đi kèm độ dốc cao ở một khu vực cụ thể không.
- Tìm mẫu phức tạp: Qua các lớp tích chập, CNN nhận diện các mẫu lớn hơn, như sự kết hợp của lượng mưa giờ lớn nhất và tốc độ thẩm thấu thấp, để đánh giá nguy cơ lũ.
- Dự đoán số liệu: Kết quả là một giá trị từ 0 đến 100, thể hiện nguy cơ lũ quét, ví dụ: 90/100 cho lưu vực có nhiều khu vực với mưa lớn và địa hình bất lợi.

CNN phù hợp khi dữ liệu lũ quét được tổ chức thành các cấu trúc không gian, ví dụ: các giá trị lượng mưa và NDVI tại nhiều điểm trong lưu vực. Mô hình tìm ra các mẫu cục bộ, như khu vực có mưa lớn gần sông suối, giúp dự đoán chính xác hơn. Tuy nhiên, CNN cần dữ liệu lớn và định dạng phù hợp, đồng thời tồn thời gian huấn luyện hơn so với các mô hình đơn giản.

b. Mạng nơ-ron sâu (Deep Neural Network - DNN)

Mạng nơ-ron sâu giống như một hệ thống xử lý thông tin, lấy dữ liệu đầu vào (như lượng mưa, độ dốc) và biến đổi chúng qua nhiều lớp (layers) để đưa ra dự đoán. Mỗi lớp giống như một bước xử lý, tìm ra các mối quan hệ giữa các đặc trưng.

Ví dụ: Khi nhận dữ liệu về lượng mưa ngày (150 mm), độ dốc lưu vực (25 độ), và chỉ số NDVI (0.3), DNN phân tích các con số này qua các lớp để xác định mức độ nguy cơ lũ. Nó tự động học cách kết hợp các đặc trưng, chẳng hạn nhận ra rằng mưa lớn và độ dốc cao thường dẫn đến nguy cơ lũ cao.

- **Nguyên tắc dự đoán nguy cơ lũ quét**
- Xử lý đồng thời các đặc trưng: DNN nhận tất cả các đặc trưng (lượng mưa, độ dốc, NDVI, v.v.) và biến đổi chúng qua nhiều lớp. Mỗi lớp tìm ra các mẫu phức tạp, ví dụ: lượng mưa lớn kết hợp với tốc độ thẩm thấu thấp làm tăng nguy cơ lũ.
- Tự động học đặc trưng: Không cần con người chọn yếu tố nào quan trọng, DNN tự tìm ra cách các đặc trưng ảnh hưởng đến lũ quét qua quá trình huấn luyện.
- Dự đoán số liệu: Kết quả là một giá trị từ 0 đến 100, thể hiện nguy cơ lũ quét tại cửa ra của lưu vực, ví dụ: 85/100 cho khu vực có mưa lớn và địa hình dốc.

DNN phù hợp với bài toán lũ quét vì nó có thể xử lý nhiều đặc trưng cùng lúc và tìm ra các mối quan hệ phức tạp mà các mô hình đơn giản có thể bỏ sót. Ví dụ, nó có thể nhận ra rằng nguy cơ lũ tăng không chỉ do mưa lớn mà còn do sự kết hợp với chỉ số CN cao và độ cao tương đối thấp. Tuy nhiên, DNN cần nhiều dữ liệu để học hiệu quả và có thể khó giải thích chi tiết tại sao đưa ra một con số cụ thể.

c. Mạng nơ-ron hồi tiếp dài ngắn hạn (Long Short-Term Memory - LSTM)

Mạng LSTM là một loại mạng nơ-ron chuyên xử lý dữ liệu theo thời gian, như các chuỗi số thay đổi qua từng giờ hoặc ngày. Trong bài toán lũ quét, LSTM phân tích các đặc trưng như lượng mưa theo thời gian để dự đoán nguy cơ lũ.

Ví dụ: Nếu lượng mưa giờ lớn nhất tăng từ 10 mm lên 50 mm trong 3 giờ, đồng thời chỉ số NDVI thấp (thảm phủ kém), LSTM sẽ xem xét xu hướng này để dự đoán nguy cơ lũ quét tăng cao.

- **Nguyên tắc dự đoán nguy cơ lũ quét**
- Phân tích theo thời gian: LSTM xem xét các đặc trưng như lượng mưa giờ hoặc ngày theo thứ tự thời gian, nhận ra xu hướng dần đến lũ quét, ví dụ: mưa lớn kéo dài làm đất bão hòa.
- Ghi nhớ dài hạn: Mô hình "nhớ" các sự kiện trước đó, như lượng mưa tích lũy trong 24 giờ, để đánh giá tác động của chúng đến nguy cơ lũ hiện tại.
- Dự đoán số liệu: Kết quả là một giá trị từ 0 đến 100, thể hiện nguy cơ lũ quét, ví dụ: 80/100 nếu mưa lớn kéo dài kết hợp với độ dốc lưu vực cao.

LSTM rất hiệu quả khi nguy cơ lũ quét phụ thuộc vào dữ liệu thời gian, như lượng mưa tăng dần hoặc sự thay đổi của tốc độ thẩm thấu qua các trận mưa. Mô hình có thể dự đoán chính xác hơn bằng cách xem xét lịch sử dữ liệu, ví dụ: một lưu vực nhận mưa lớn liên tục trong 24 giờ có nguy cơ lũ cao hơn so với mưa lớn trong 1 giờ. Tuy nhiên, LSTM phức tạp và yêu cầu dữ liệu thời gian chi tiết, cùng với thời gian huấn luyện dài.

2.5. Một số kỹ thuật tối ưu hóa mô hình trí tuệ nhân tạo

Về cơ bản, kỹ thuật tối ưu nhằm tìm kiếm các tham số/thuật toán để giảm thiểu các chi phí đào tạo mô hình và tăng cường chất lượng dự đoán.

2.5.1.1 Tinh chỉnh siêu tham số (*Hyperparameter tuning*)

Siêu tham số là các tham số không được học trong quá trình huấn luyện, như tỷ lệ học (learning rate), số lượng lớp ẩn (hidden layer size), số lượng epochs, và batch size. Tinh chỉnh siêu tham số là quá trình tìm kiếm và chọn giá trị tốt nhất cho các siêu tham số này để cải thiện hiệu suất mô hình.

2.5.1.2 Tăng cường dữ liệu (*Data augmentation*)

Tăng cường dữ liệu là quá trình tạo ra thêm dữ liệu huấn luyện bằng cách áp dụng các biến đổi nhỏ lên dữ liệu hiện có, như xoay, thu phóng, cắt tia, lật ngang, hay thay đổi ánh sáng. Việc tăng cường dữ liệu giúp mô hình học được các biến thể khác nhau của dữ liệu và giảm hiện tượng quá khớp (overfitting).

2.5.1.3 Tối ưu hóa mạng nơ-ron (*Network optimization*)

Cải thiện cấu trúc mạng nơ-ron bằng cách thay đổi số lượng lớp, kích thước lớp, số lượng tham số, và các thay đổi khác để cải thiện hiệu suất. Có thể sử dụng các kiến trúc mạng nơ-ron sâu (deep neural networks) như Convolutional Neural Networks (CNN) hay Recurrent Neural Networks (RNN) để nâng cao hiệu suất mô hình.

2.5.1.4 Tối ưu hóa thuật toán huấn luyện (Training algorithm optimization)

Cải tiến thuật toán huấn luyện bằng cách sử dụng các phương pháp tối ưu hóa như Adam, RMSprop, hoặc Stochastic Gradient Descent (SGD) với các biến thể như Momentum hay Nesterov. Cũng có thể sử dụng kỹ thuật regularization như L1 regularization hoặc L2 regularization để giảm hiện tượng quá khớp.

CHƯƠNG 3. ĐỀ XUẤT PHƯƠNG PHÁP TỐI ƯU ỦNG DỤNG TRÍ TUỆ NHÂN TẠO CHO PHÂN VÙNG RỦI RO LŨ QUÉT Ở KHU VỰC MÙ CANG CHẢI

3.1. Lựa chọn khu vực nghiên cứu

Mù Cang Chải có lịch sử chịu thiệt hại nặng nề từ các trận lũ quét. Một ví dụ điển hình là trận lũ quét ngày 5 tháng 8 năm 2023 tại xã Hồ Bón [80], gây thiệt hại nghiêm trọng. Trận lũ này không chỉ làm rõ tính bất ngờ và sức tàn phá của lũ quét mà còn nhấn mạnh sự cần thiết của các công cụ dự báo và quản lý rủi ro.



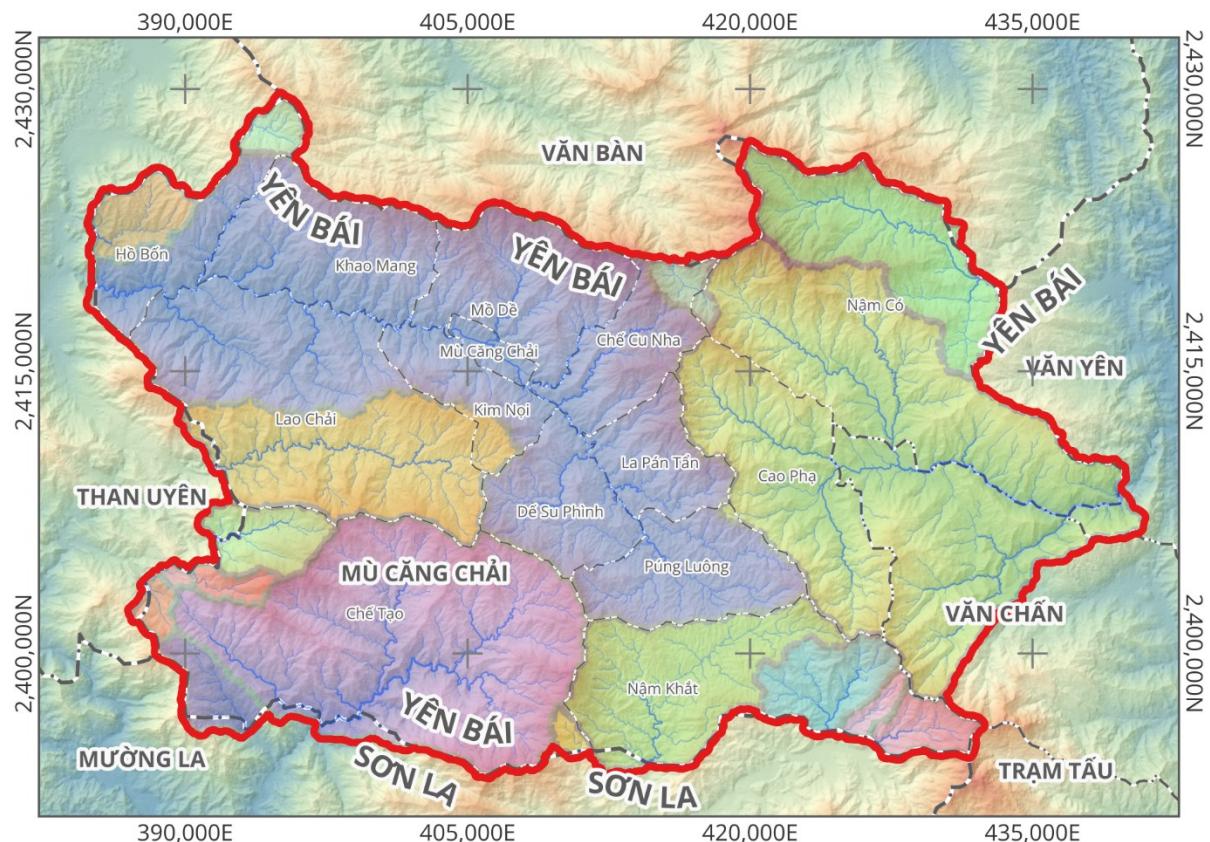
Hình 20. Trận lũ xảy ra tại Hồ Bón năm 2023

Các sự kiện lũ quét trước đó, như trận lũ năm 2017 tại trung tâm thị trấn Mù Cang Chải, cũng gây thiệt hại lớn về người và tài sản [74, 75, 76]. Những sự kiện này cho thấy Mù Cang Chải là một trong những khu vực có nguy cơ lũ quét cao nhất ở miền Bắc Việt Nam, với các yếu tố như độ dốc lớn, mật độ dân cư cao ở các khu vực ven sông suối và thiếu hệ thống cảnh báo sớm.



Hình 21. Lũ quét tại suối Háng Chú năm 2017 [91]

Mù Cang Chải được chọn vì nó đại diện cho các khu vực miền núi Việt Nam, nơi mà các yếu tố như địa hình dốc, mưa lớn và thay đổi sử dụng đất tạo ra nguy cơ lũ quét cao. Việc nghiên cứu khu vực này cho phép các nhà khoa học phát triển các mô hình có thể áp dụng rộng rãi, cung cấp thông tin chi tiết về cách giảm thiểu rủi ro lũ quét trong bối cảnh biến đổi khí hậu và đô thị hóa. Phân tích địa hình thủy văn cho khu vực Mù Cang Chải cho thấy khu vực nghiên cứu là thượng nguồn của các lưu vực sông như lưu vực Nậm Kim; lưu vực ngòi Hút; lưu vực Nậm Tha; và lưu vực Nậm Chang.



Hình 22. Huyện Mù Cang Chải là thượng nguồn của nhiều lưu vực, không chịu sự tác động lũ quét từ bên ngoài mà chỉ xảy ra ở nội tại huyện

3.2. Phân tích, đề xuất phương pháp tối ưu ứng dụng trí tuệ nhân tạo cho phân vùng lũ quét.

3.2.1 Chi tiết các thuật toán và lựa chọn tham số tối ưu cho mô hình

1. Rừng ngẫu nhiên

Rừng ngẫu nhiên là một thuật toán ensemble dựa trên tập hợp nhiều cây quyết định. Mỗi cây được huấn luyện trên một tập con dữ liệu (bootstrap sample) và một tập con ngẫu nhiên của đặc trưng. Kết quả dự đoán được tổng hợp thông qua bỏ phiếu đa số trong bài toán phân loại. Cụ thể:

- **Bootstrap aggregating (bagging):** Tạo nhiều tập dữ liệu con bằng cách lấy mẫu có hoán lại từ tập dữ liệu gốc.
- **Random feature selection:** Tại mỗi node của cây quyết định, chỉ một tập con đặc trưng được chọn ngẫu nhiên để tìm ngưỡng phân chia tối ưu.
- **Voting:** Kết quả cuối cùng là trung bình (regression) hoặc đa số (classification) của dự đoán từ tất cả các cây.

Phương pháp tối ưu hóa RF:

Các tham số tối ưu:

- n_estimators: Số lượng cây quyết định. Giá trị lớn (100-1000) cải thiện độ chính xác nhưng tăng chi phí tính toán.
- max_depth: Độ sâu tối đa của cây. Giới hạn độ sâu giảm overfitting nhưng có thể làm giảm khả năng học.
- min_samples_split và min_samples_leaf: Số mẫu tối thiểu để chia node hoặc tại lá, giúp kiểm soát kích thước cây.
- max_features: Số đặc trưng được chọn ngẫu nhiên tại mỗi node. Giá trị nhỏ (ví dụ: $\text{sqrt}(n_features)$) tăng tính đa dạng giữa các cây.

Xử lý mất cân bằng lớp: Sử dụng tham số class_weight (ví dụ: balanced) để tăng trọng số cho các lớp thiểu số, đặc biệt quan trọng trong bài toán lũ quét nơi các lớp nguy cơ cao thường hiếm.

Bổ sung đặc trưng không gian: Tính toán thủ công các đặc trưng như giá trị trung bình, phương sai, hoặc chỉ số texture trong vùng lân cận (3×3 hoặc 5×5 pixel) để cải thiện khả năng học mẫu không gian.

Song song hóa: Sử dụng tham số n_jobs để tận dụng đa luồng, giảm thời gian huấn luyện.

a. Cơ chế hoạt động chi tiết:

RF xây dựng nhiều cây quyết định (mặc định 100 cây), mỗi cây học trên một tập con dữ liệu (bootstrap) và tập con đặc trưng (mặc định $\text{sqrt}(n_features)$).

Trong bài toán phân loại 5 lớp, mỗi cây dự đoán một lớp, và kết quả cuối cùng là lớp nhận được đa số phiếu từ tất cả các cây.

Tham số random_state=42 đảm bảo tính tái lập (reproducibility) của mô hình.

n_jobs=threads tận dụng đa luồng để tăng tốc huấn luyện.

class_weight='balanced' tự động điều chỉnh trọng số lớp dựa trên tần suất, tăng trọng số cho các lớp hiếm (như nguy cơ rất cao) để xử lý mất cân bằng.

b. Lý do chọn tham số:

random_state=42: Đảm bảo kết quả huấn luyện và đánh giá nhất quán, hỗ trợ so sánh với các mô hình khác. Đây là giá trị phổ biến trong nghiên cứu để tái lập.

n_jobs=threads: Tận dụng toàn bộ tài nguyên CPU, giảm thời gian huấn luyện, đặc biệt với dữ liệu lớn (hàng triệu pixel).

class_weight='balanced': Dữ liệu lũ quét thường mất cân bằng (lớp nguy cơ cao chiếm tỷ lệ nhỏ). Tham số này tăng trọng số cho các lớp hiếm, cải thiện khả năng dự đoán các khu vực nguy cơ cao – yếu tố quan trọng trong quản lý thiên tai.

Không tuning thêm các tham số khác (như n_estimators, max_depth): Cấu hình mặc định của RF (n_estimators=100, max_features='sqrt') đã đủ mạnh cho nhiều bài toán thực tế. Việc không điều chỉnh thêm có thể do:

- Ưu tiên tốc độ triển khai ban đầu.
- Dữ liệu 20 đặc trưng không quá phức tạp, phù hợp với cấu hình mặc định.
- Kế hoạch tuning sau khi đánh giá hiệu suất sơ bộ.

c. Các cách tối ưu RF và lý do chọn cách hiện tại:

Tuning hyperparameters: Có thể sử dụng grid search hoặc random search để tìm n_estimators, max_depth, min_samples_split. Tuy nhiên, cách này tốn thời gian với dữ liệu lớn, nên cấu hình mặc định được chọn để đánh giá ban đầu.

Feature engineering: Bổ sung đặc trưng không gian (ví dụ: trung bình vùng lân cận 5×5) cải thiện khả năng học mẫu không gian, nhưng chưa được áp dụng trong cấu hình hiện tại, có thể do ưu tiên đơn giản hóa pipeline dữ liệu.

SMOTE hoặc oversampling: Xử lý mất cân bằng lớp bằng cách tạo mẫu tổng hợp cho lớp thiểu số, nhưng class_weight='balanced' là cách đơn giản và hiệu quả hơn, phù hợp với dữ liệu lớn.

Pruning cây: Giới hạn độ sâu cây để giảm overfitting, nhưng RF đã có cơ chế bagging chống overfitting, nên không cần thiết trong cấu hình ban đầu.

2. Máy hỗ trợ vectơ (SVM)

SVM tìm một siêu phẳng tối ưu trong không gian đặc trưng để phân tách các lớp, tối đa hóa khoảng cách (margin) đến các điểm dữ liệu gần nhất (support vectors). Với

dữ liệu phi tuyến, SVM sử dụng kernel trick (ví dụ: kernel RBF) để ánh xạ dữ liệu vào không gian chiều cao hơn, nơi các lớp có thể được phân tách tuyến tính. Trong bài toán phân loại 5 lớp, SVM sử dụng chiến lược one-vs-rest để xử lý đa lớp.

a. Phương pháp tối ưu hóa SVM:

- Các tham số tối ưu:
 - C: Tham số điều chỉnh độ phạt lỗi phân loại. Giá trị nhỏ ưu tiên margin lớn (mô hình đơn giản), giá trị lớn ưu tiên phân loại đúng (có thể dẫn đến overfitting).
 - gamma: Điều chỉnh độ cong của kernel (RBF). Giá trị nhỏ tạo ranh giới mượt, giá trị lớn làm mô hình nhạy hơn với từng điểm dữ liệu.
 - kernel: Lựa chọn kernel phù hợp (RBF, polynomial, linear). Kernel RBF thường được chọn cho dữ liệu phi tuyến.
- Xử lý mất cân bằng lớp: Sử dụng class_weight='balanced' để tăng trọng số cho các lớp thiểu số, cải thiện khả năng dự đoán lớp nguy cơ cao.
- Chuẩn hóa dữ liệu: SVM nhạy với thang đo của đặc trưng, nên cần chuẩn hóa (scaling) dữ liệu đầu vào (ví dụ: MinMaxScaler hoặc StandardScaler).
- Song song hóa: Sử dụng n_jobs để tăng tốc huấn luyện, đặc biệt với chiến lược one-vs-rest.

b. Cơ chế hoạt động chi tiết:

Trong câu hình SVC(random_state=42, probability=True, class_weight='balanced'):

- SVM xây dựng một siêu phẳng phân tách 5 lớp nguy cơ lũ quét, sử dụng kernel RBF (mặc định) để xử lý dữ liệu phi tuyến.
- probability=True cho phép tính xác suất dự đoán, cần thiết để so sánh với các mô hình khác (ví dụ: trong ensemble).
- class_weight='balanced' tự động điều chỉnh trọng số lớp dựa trên tần suất, tăng trọng số cho các lớp hiếm.
- random_state=42 đảm bảo tính tái lập của mô hình.
- Mỗi pixel được biểu diễn dưới dạng vector 20 đặc trưng, không bao gồm thông tin không gian trừ khi được bổ sung thủ công.

c. Lý do chọn tham số:

- random_state=42: Đảm bảo tính tái lập, hỗ trợ so sánh với các mô hình khác.
- probability=True: Cần thiết để tính xác suất đầu ra, đặc biệt khi tích hợp với mô hình ensemble hoặc đánh giá bằng AUC-ROC.

- `class_weight='balanced'`: Xử lý mất cân bằng lớp, đảm bảo mô hình tập trung vào các lớp nguy cơ cao.
- Không tuning C và gamma: Cấu hình mặc định ($C=1.0$, $\gamma='scale'$) phù hợp cho đánh giá ban đầu, tránh chi phí tính toán cao của grid search. Điều này có thể do:
 - Ưu tiên tốc độ triển khai.
 - Dữ liệu 20 đặc trưng không quá phức tạp, phù hợp với kernel RBF mặc định.
 - Kế hoạch tuning sau khi đánh giá sơ bộ.

d. Các cách tối ưu SVM và lý do chọn cách hiện tại:

- Tuning hyperparameters: Grid search hoặc random search cho C và gamma có thể cải thiện độ chính xác, nhưng tốn thời gian với dữ liệu lớn, nên cấu hình mặc định được chọn ban đầu.
- Feature engineering: Bổ sung đặc trưng không gian (ví dụ: trung bình vùng lân cận) cải thiện khả năng học mẫu không gian, nhưng chưa áp dụng để đơn giản hóa pipeline.
- SMOTE hoặc undersampling: Tạo mẫu tổng hợp hoặc giảm mẫu lớp đa số, nhưng `class_weight='balanced'` là cách hiệu quả và đơn giản hơn.
- Kernel khác: Có thể thử kernel polynomial hoặc linear, nhưng RBF thường phù hợp hơn với dữ liệu phi tuyến như lũ quét.

3. Hồi quy logistic (Logistic Regression - LR)

Hồi quy logistic là một mô hình thống kê sử dụng hàm sigmoid để dự đoán xác suất thuộc về một lớp. Trong bài toán đa lớp, LR sử dụng chiến lược one-vs-rest hoặc softmax regression để phân loại 5 lớp nguy cơ lũ quét.

a. Phương pháp tối ưu hóa LR:

- Tuning hyperparameters:
 - C: Nghịch đảo của hệ số regularization. Giá trị nhỏ tăng regularization, giảm overfitting; giá trị lớn ưu tiên phân loại đúng.
 - solver: Lựa chọn thuật toán tối ưu hóa (ví dụ: `lbfgs`, `liblinear`).
 - `max_iter`: Số vòng lặp tối đa để đảm bảo hội tụ.
- Xử lý mất cân bằng lớp: Sử dụng `class_weight='balanced'` để tăng trọng số cho lớp thiểu số.
- Chuẩn hóa dữ liệu: LR nhạy với thang đo đặc trưng, cần chuẩn hóa dữ liệu.
- Song song hóa: Sử dụng `n_jobs` để tăng tốc với chiến lược one-vs-rest.

b. Cơ chế hoạt động chi tiết:

Trong cấu hình LogisticRegression(random_state=42, max_iter=1000, class_weight='balanced', n_jobs=threads):

- LR sử dụng softmax regression để dự đoán xác suất cho 5 lớp nguy cơ.
- random_state=42 đảm bảo tính tái lập.
- max_iter=1000 đảm bảo mô hình hội tụ, đặc biệt với dữ liệu lớn.
- class_weight='balanced' tăng trọng số cho lớp thiểu số.
- n_jobs=threads tận dụng đa luồng để tăng tốc huấn luyện.

c. Lý do chọn tham số:

- random_state=42: Đảm bảo tính tái lập.
- max_iter=1000: Đảm bảo hội tụ với dữ liệu lớn và đa lớp.
- class_weight='balanced': Xử lý mất cân bằng lớp, cải thiện dự đoán cho lớp nguy cơ cao.
- n_jobs=threads: Tăng tốc huấn luyện.
- Không tuning C hoặc solver: Cấu hình mặc định ($C=1.0$, solver='lbfgs') phù hợp cho đánh giá ban đầu, do:
 - LR là mô hình baseline, không cần tối ưu phức tạp.
 - Dữ liệu 20 đặc trưng không quá phức tạp.

d. Các cách tối ưu LR và lý do chọn cách hiện tại:

- Tuning hyperparameters: Grid search cho C và solver có thể cải thiện hiệu suất, nhưng không cần thiết cho baseline.
- Feature engineering: Bổ sung đặc trưng không gian, nhưng chưa áp dụng để đơn giản hóa pipeline.
- SMOTE: Có thể áp dụng, nhưng class_weight='balanced' đủ hiệu quả.

4. Light Gradient Boosting Machine (LGBM)

LGBM là thuật toán gradient boosting sử dụng kỹ thuật leaf-wise tree growth và histogram-based splitting để tăng tốc độ và giảm bộ nhớ. Mỗi cây được xây dựng tuần tự, tập trung sửa lỗi của các cây trước.

a. Phương pháp tối ưu hóa LGBM:

- Tuning hyperparameters:
 - n_estimators: Số cây quyết định.
 - learning_rate: Tốc độ học, ảnh hưởng đến tốc độ hội tụ.
 - max_depth và num_leaves: Kiểm soát độ phức tạp cây.

- o min_child_samples: Số mẫu tối thiểu tại lá.
- Xử lý mất cân bằng lớp: Sử dụng class_weight='balanced' hoặc is_unbalance=True.
- Feature engineering: Bổ sung đặc trưng không gian.
- Song song hóa: Sử dụng n_jobs để tăng tốc.

b. Cơ chế hoạt động chi tiết:

Trong cấu hình LGBMClassifier(random_state=42, n_jobs=threads, class_weight='balanced'):

- LGBM xây dựng các cây quyết định tuần tự, sử dụng histogram để chia dữ liệu nhanh hơn.
- random_state=42 đảm bảo tái lập.
- n_jobs=threads tăng tốc huấn luyện.
- class_weight='balanced' xử lý mất cân bằng lớp.

c. Lý do chọn tham số:

- random_state=42: Đảm bảo tái lập.
- n_jobs=threads: Tăng tốc huấn luyện.
- class_weight='balanced': Xử lý mất cân bằng lớp.
- Không tuning thêm: Cấu hình mặc định (n_estimators=100, learning_rate=0.1) phù hợp cho đánh giá ban đầu, do:
 - o LGBM có hiệu suất tốt ngay cả với cấu hình mặc định.
 - o Ưu tiên tốc độ triển khai.

d. Các cách tối ưu LGBM và lý do chọn cách hiện tại:

- Tuning hyperparameters: Grid search cho n_estimators, learning_rate, nhưng chưa áp dụng để đánh giá nhanh.
- Feature engineering: Bổ sung đặc trưng không gian, nhưng chưa áp dụng để đơn giản hóa.
- SMOTE: class_weight='balanced' là cách đơn giản và hiệu quả.

5. Mô hình kết hợp (Ensemble)

Ensemble sử dụng VotingClassifier kết hợp RF và LGBM với cơ chế soft voting, lấy trung bình xác suất dự đoán từ hai mô hình.

a. Phương pháp tối ưu hóa Ensemble:

- Tuning trọng số mô hình: Điều chỉnh trọng số của RF và LGBM.
- Tuning mô hình con: Tối ưu RF và LGBM riêng lẻ.

- Song song hóa: Sử dụng n_jobs.

b. Cơ chế hoạt động chi tiết:

Trong cấu hình VotingClassifier(estimators=[('rf', ...), ('lgbm', ...)], voting='soft', n_jobs=threads):

- Soft voting lấy trung bình xác suất từ RF và LGBM.
- random_state=42 đảm bảo tái lập.
- n_jobs=threads tăng tốc.

c. Lý do chọn tham số:

- voting='soft': Tận dụng xác suất để cải thiện độ chính xác.
- n_jobs=threads: Tăng tốc.
- Không tuning trọng số: Trọng số mặc định (1:1) đủ hiệu quả cho đánh giá ban đầu.

d. Các cách tối ưu Ensemble và lý do chọn cách hiện tại:

- Tuning trọng số: Có thể điều chỉnh, nhưng chưa áp dụng để đơn giản hóa.
- Tối ưu mô hình con: Tuning RF và LGBM riêng lẻ, nhưng chưa cần thiết.

6. Mạng nơ-ron sâu (Deep Neural Network - DNN)

DNN gồm 3 lớp ẩn (256, 128, 64 nơ-ron), sử dụng BatchNormalization, LeakyReLU, Dropout, và L2 regularization.

Đầu ra là softmax cho 5 lớp.

```
Sequential([
    Input(shape=(input_dim,)),
    Dense(256, kernel_regularizer=l2(0.001)),
    BatchNormalization(),
    LeakyReLU(alpha=0.1),
    Dropout(0.3),
    Dense(128, kernel_regularizer=l2(0.001)),
    BatchNormalization(),
    LeakyReLU(alpha=0.1),
    Dropout(0.3),
    Dense(64, kernel_regularizer=l2(0.001)),
    BatchNormalization(),
    LeakyReLU(alpha=0.1),
    Dropout(0.2),
    Dense(5, activation='softmax', dtype='float32')
])
```

DNN là mạng nơ-ron với nhiều lớp ẩn, học các mối quan hệ phi tuyến phức tạp từ dữ liệu dạng vector.

a. Phương pháp tối ưu hóa DNN:

- Tuning kiến trúc: Số lớp, số nơ-ron mỗi lớp.
- Regularization: Sử dụng Dropout, L2 regularization, BatchNormalization.
- Tối ưu tham số: Learning rate, batch size.

b. Lý do chọn tham số:

- Kiến trúc 256-128-64: Cân bằng giữa độ phức tạp và khả năng học phi tuyến.
- L2 regularization (0.001): Giảm overfitting.
- Dropout (0.3, 0.2): Ngăn chặn học quá mức trên dữ liệu huấn luyện.
- LeakyReLU (alpha=0.1): Xử lý tốt gradient vanishing so với ReLU.
- BatchNormalization: Ôn định huấn luyện, tăng tốc hội tụ.

c. Các cách tối ưu DNN và lý do chọn cách hiện tại:

- Tuning kiến trúc: Có thể thử số lớp/nơ-ron khác, nhưng cấu hình hiện tại đủ tốt.
- Tuning learning rate: Chưa áp dụng, ưu tiên cấu hình mặc định.
- Data augmentation: Không áp dụng do dữ liệu raster tĩnh.

7. Mạng nơ-ron tích chập (Convolutional Neural Network - CNN)

CNN sử dụng các lớp convolution để học đặc trưng không gian từ dữ liệu raster, phù hợp với bài toán lũ quét.

CNN sử dụng các lớp Conv2D, residual blocks, attention block, SpatialDropout, và GlobalAveragePooling. Đầu ra là softmax cho 5 lớp.

```
def build_cnn(self, input_shape: tuple[int, int, int]) -> Model:  
    inputs = Input(shape=input_shape)  
    x = Conv2D(64, (3, 3), padding='same', kernel_regularizer=l2(0.005))(inputs)  
    x = BatchNormalization()(x)  
    x = LeakyReLU(alpha=0.1)(x)  
    x = SpatialDropout2D(0.3)(x)  
    # Residual blocks, MaxPooling, Attention block, GlobalAveragePooling  
    outputs = Dense(5, activation='softmax', kernel_regularizer=l2(0.005))(x)  
    model = Model(inputs, outputs)  
    return model
```

a. Phương pháp tối ưu hóa CNN:

- Tuning kiến trúc: Số lớp convolution, số bộ lọc, kích thước kernel.
- Regularization: SpatialDropout, L2 regularization, BatchNormalization.
- Tuning hyperparameters: Learning rate, batch size, kích thước patch.

b. Lý do chọn tham số:

- Conv2D (64, 128, 256 bộ lọc): Tăng dần số bộ lọc để học đặc trưng phức tạp.
- L2 regularization (0.005): Giảm overfitting.
- SpatialDropout (0.3): Phù hợp với dữ liệu raster, giảm overfitting không gian.
- LeakyReLU (alpha=0.1): Xử lý gradient vanishing.
- Residual blocks và attention block: Cải thiện khả năng học mẫu không gian phức tạp.

c. Các cách tối ưu CNN và lý do chọn cách hiện tại:

- Tuning kiến trúc: Có thể thử số lớp/bộ lọc khác, nhưng cấu hình hiện tại cân bằng hiệu suất và chi phí.
- Tuning patch size: Patch 15×15 hoặc 21×21 là lựa chọn hợp lý.
- Data augmentation: Không áp dụng do dữ liệu tĩnh.

8. Mạng nơ-ron nhớ dài ngắn hạn (Long Short – Term Memory – LSTM)

LSTM sử dụng các cổng điều khiển (input, forget, output gates) để học các phụ thuộc thời gian dài hạn từ dữ liệu chuỗi, được thiết kế cho bài toán dự báo có yếu tố thời gian.

LSTM sử dụng các lớp LSTM bidirectional, Dense layers, Dropout, và BatchNormalization. Đầu ra là softmax cho 5 lớp rủi ro lũ quét.

a. Phương pháp tối ưu hóa LSTM:

- Tuning kiến trúc: Số lớp LSTM, số units, sử dụng bidirectional hoặc stacked LSTM.
- Regularization: Dropout, recurrent dropout, L2 regularization cho Dense layers.
- Tuning hyperparameters: Learning rate, batch size, sequence length, timesteps.

b. Hạn chế và lý do không phù hợp:

- Yêu cầu dữ liệu chuỗi thời gian: LSTM cần dữ liệu có tính tuần tự theo thời gian, trong khi dữ liệu nghiên cứu chủ yếu là các đặc trưng không gian tĩnh (địa hình, thảm phủ, loại đất).
- Thiếu temporal dependencies: Các yếu tố như độ dốc, độ cao, CN không thay đổi theo thời gian, không tạo ra các phụ thuộc thời gian mà LSTM có thể học.
- Chi phí tính toán cao: LSTM yêu cầu nhiều tài nguyên tính toán hơn nhưng không mang lại lợi ích với dữ liệu tĩnh.

c. Khuyến nghị loại bỏ LSTM trong phân vùng lũ quét:

- Bản chất bài toán: Phân vùng rủi ro lũ quét dựa trên đặc trưng không gian tĩnh không phù hợp với kiến trúc LSTM.
- Hiệu quả kém: Việc ép buộc dữ liệu tĩnh vào mô hình chuỗi thời gian dẫn đến hiệu suất kém hơn các mô hình không gian như CNN.
- Tính khoa học: Lựa chọn mô hình phù hợp với bản chất dữ liệu là nguyên tắc cơ bản trong machine learning.

3.2.2 Tổng hợp các thuật toán và lựa chọn tham số tối ưu cho mô hình

Bảng 9. Bảng tổng hợp thuật toán và tham số tối ưu

Mô hình	Tham số có thể tối ưu/Thuật toán áp dụng	Ý nghĩa tham số và thuật toán	Lựa chọn trong nghiên cứu và giải thích
Random Forest (RF)	<ul style="list-style-type: none"> n_estimators (số cây) max_depth (độ sâu cây) min_samples_split (số mẫu tối thiểu để chia) min_samples_leaf (số mẫu tối thiểu tại lá) max_features (số đặc trưng xét khi chia) class_weight (cân bằng lớp) random_state (seed ngẫu nhiên) n_jobs (song song hóa) 	<ul style="list-style-type: none"> n_estimators: Càng lớn càng ổn định nhưng tăng thời gian tính toán max_depth: Kiểm soát overfitting, quá lớn dễ overfit min_samples_split/leaf: Regularization, ngăn cây quá chi tiết max_features: Tăng tính đa dạng giữa các cây class_weight='balanced': Tự động cân bằng trọng số theo tỷ lệ nghịch đảo tần suất lớp 	<ul style="list-style-type: none"> random_state=42: Đảm bảo tái lập kết quả class_weight='balanced': Xử lý mất cân bằng dữ liệu lũ quét (vùng nguy hiểm ít hơn vùng an toàn) n_jobs=threads: Tối ưu tốc độ với xử lý song song Các tham số khác sử dụng mặc định để tránh overfitting trên dữ liệu không gian
Support Vector Machine (SVM)	<ul style="list-style-type: none"> C (tham số regularization) kernel (hàm kernel) gamma (tham số kernel RBF) degree (bậc polynomial) class_weight (cân bằng lớp) probability (ước lượng xác suất) random_state (seed) 	<ul style="list-style-type: none"> C: Kiểm soát trade-off giữa margin và classification error. C lớn → ít regularization kernel: RBF cho phi tuyến, linear cho tuyến tính gamma: Ảnh hưởng của từng điểm training. Lớn → overfitting probability=True: Cho phép dự đoán xác suất thay vì chỉ nhãn 	<ul style="list-style-type: none"> probability=True: Cần thiết để có xác suất dự đoán cho ensemble class_weight='balanced': Xử lý mất cân bằng lớp random_state=42: Đảm bảo reproducibility Kernel và C sử dụng mặc định (RBF, C=1.0) phù hợp cho dữ liệu phi tuyến của bài toán địa không gian
Logistic Regression (LR)	<ul style="list-style-type: none"> C (nghịch đảo regularization) penalty (loại regularization) solver (thuật toán tối ưu) max_iter (số vòng lặp tối đa) class_weight (cân bằng lớp) n_jobs (song song hóa) 	<ul style="list-style-type: none"> C: Nghịch đảo strength của regularization. C lớn → ít regularization penalty: L1 (Lasso), L2 (Ridge), ElasticNet solver: 'liblinear' (nhỏ), 'lbfgs' (trung bình), 'saga' (lớn) max_iter: Ngừng sớm để tránh vô hạn 	<ul style="list-style-type: none"> max_iter=1000: Đảm bảo hội tụ với dữ liệu lớn (nhiều pixel raster) class_weight='balanced': Xử lý mất cân bằng n_jobs=threads: Tăng tốc với multiprocessing Các tham số khác mặc định phù hợp cho bài toán phân loại đa lớp
LightGBM (LGBM)	<ul style="list-style-type: none"> num_leaves (số lá mỗi cây) learning_rate (tốc độ học) n_estimators (số boosting rounds) max_depth (độ sâu tối đa) min_child_samples (số mẫu tối thiểu tại lá) subsample (tỷ lệ mẫu con) colsample_bytree (tỷ lệ đặc trưng) 	<ul style="list-style-type: none"> num_leaves: Kiểm soát độ phức tạp mô hình, quá lớn dễ overfit learning_rate: Tốc độ học, nhỏ cần nhiều estimators subsample: Phòng overfitting bằng random sampling reg_alpha/lambda: L1 sparse features, L2 smooth weights Leaf-wise growth: Khác level-wise, tối ưu loss nhanh hơn 	<ul style="list-style-type: none"> class_weight='balanced': Xử lý mất cân bằng dữ liệu lũ quét random_state=42: Reproducibility n_jobs=threads: Tận dụng đa luồng Các tham số khác mặc định: tối ưu cho dữ liệu lớn, cân bằng tốc độ và accuracy Leaf-wise growth phù hợp với dữ liệu raster phức tạp

Mô hình	Tham số có thể tối ưu/Thuật toán áp dụng	Ý nghĩa tham số và thuật toán	Lựa chọn trong nghiên cứu và giải thích
	<ul style="list-style-type: none"> reg_alpha/reg_lambda (L1/L2 regularization) class_weight (cân bằng lớp) 		
Ensemble (Voting)	<ul style="list-style-type: none"> estimators (danh sách mô hình) voting (hard/soft voting) weights (trọng số các mô hình) n_jobs (song song hóa) 	<ul style="list-style-type: none"> voting='soft': Sử dụng xác suất dự đoán thay vì vote cứng weights: Cho phép ưu tiên mô hình tốt hơn Ensemble: Giảm variance, tăng stability và performance 	<ul style="list-style-type: none"> RF + LGBM: Kết hợp tree-based methods, hỗ trợ nhau (RF ổn định, LGBM tối ưu) voting='soft': Tận dụng confidence của từng mô hình n_jobs=threads: Song song hóa toàn bộ ensemble Không sử dụng weights để tránh bias
CNN	<p>Kiến trúc:</p> <ul style="list-style-type: none"> Số lớp Convolution Kích thước filter (kernel_size) Số filter (channels) Padding strategy Pooling type và size <p>Regularization:</p> <ul style="list-style-type: none"> L2 regularization Batch Normalization Dropout/SpatialDropout2D LeakyReLU activation <p>Attention & Residual:</p> <ul style="list-style-type: none"> Residual blocks Attention mechanism Global Average Pooling 	<ul style="list-style-type: none"> Conv2D(64, (3,3)): 64 filters 3x3 trích xuất spatial features BatchNorm: Ổn định training, tăng tốc hội tụ LeakyReLU($\alpha=0.1$): Tránh dying ReLU problem SpatialDropout2D: Dropout cả feature map thay vì từng neuron Residual blocks: Giải quyết vanishing gradient Attention: Tập trung vào vùng quan trọng L2(0.005): Regularization tránh overfitting 	<ul style="list-style-type: none"> Progressive channels ($64 \rightarrow 128 \rightarrow 256$): Học từ low-level đến high-level features Residual blocks: Cho phép mạng sâu hơn mà không bị degradation Attention mechanism: Tự động focus vào vùng có nguy cơ lũ cao SpatialDropout2D(0.3): Phù hợp với dữ liệu spatial, tránh co-adaptation GlobalAveragePooling: Giảm parameters, tránh overfitting hơn Flatten L2 regularization: Cân bằng với spatial complexity
DNN	<p>Kiến trúc:</p> <ul style="list-style-type: none"> Số lớp ẩn Số neurons mỗi lớp Activation functions <p>Regularization:</p> <ul style="list-style-type: none"> L2 regularization Batch Normalization Dropout rate Learning rate <p>Optimization:</p> <ul style="list-style-type: none"> Optimizer (Adam, SGD) Loss function Batch size 	<ul style="list-style-type: none"> Dense layers ($256 \rightarrow 128 \rightarrow 64$): Giảm dần để học hierarchical representations BatchNormalization: Tăng tốc training và ổn định LeakyReLU: Tránh dying neurons Dropout: Ngẫu nhiên loại bỏ neurons để tránh overfitting L2 regularization: Penalty trên weights 	<ul style="list-style-type: none"> Pyramid architecture ($256 \rightarrow 128 \rightarrow 64$): Funnel information từ 20 features đến 5 classes Progressive dropout ($0.3 \rightarrow 0.3 \rightarrow 0.2$): Giảm dần khi gần output Batch Normalization: Ổn định với dữ liệu đa scale (elevation, rainfall, etc.) L2(0.001): Nhẹ hơn CNN do ít parameters hơn LeakyReLU: Tránh gradient vanishing với dữ liệu địa lý

Bảng 10. Tổng hợp các thuật toán tối ưu chung

Thuật toán	Áp dụng cho	Ý nghĩa	Lựa chọn trong nghiên cứu
Class Weight Balancing	Tất cả mô hình	Xử lý mất cân bằng dữ liệu bằng trọng số nghịch đảo tần suất	class_weight='balanced': Tự động tính toán trọng số phù hợp với tỷ lệ các lớp nguy cơ lũ
Cross-Validation	Grid Search, Model Selection	Đánh giá độ tổng quát hóa trên dữ liệu chưa thấy	Stratified K-Fold để đảm bảo tỷ lệ các lớp được giữ nguyên
Ensemble Methods	Meta-learning	Kết hợp nhiều mô hình để tăng performance và giảm variance	Voting Classifier với RF+LGBM: kết hợp stability và efficiency
Regularization	Deep Learning	Ngăn overfitting với penalty terms	L1/L2 regularization kết hợp với Dropout và BatchNorm
Attention Mechanism	CNN	Tập trung vào các vùng quan trọng trong spatial data	Self-attention để tự động identify các khu vực có nguy cơ lũ cao

3.3. Phân tích, lựa chọn mô hình trí tuệ nhân tạo trong phân vùng lũ quét

Dựa trên các phân tích tại chương 2, có thể thấy bốn mô hình học máy gồm rừng ngẫu nhiên (RF), máy hỗ trợ vectơ (SVM), hồi quy logistic (LR) và LightGBM (LGBM) cùng với hai mô hình học sâu là mạng nơ-ron sâu (DNN), mạng nơ-ron tích chập (CNN) và mạng nơ-ron hồi tiếp dài ngắn hạn (LSTM) là các phương pháp nổi bật hiện nay trong lĩnh vực dự đoán nguy cơ lũ quét. Tuy nhiên, không phải mô hình nào cũng phù hợp cho mọi loại dữ liệu hoặc mọi mục đích sử dụng, đặc biệt là trong bối cảnh dữ liệu địa không gian có tính chất phức tạp và không đồng nhất.

Trong nghiên cứu này, bộ dữ liệu bao gồm 20 raster đặc trưng, mỗi raster đại diện cho một yếu tố ảnh hưởng đến nguy cơ lũ quét như địa hình, thảm phủ, lớp phủ đất, lượng mưa, chỉ số CN, v.v. Dữ liệu đầu ra (label) là bản đồ nguy cơ lũ quét dạng raster, trong đó mỗi pixel mang một nhãn (ví dụ: nguy cơ cao/thấp hoặc các cấp nguy cơ). Một điểm quan trọng là bản chất không gian của dữ liệu: mỗi điểm dữ liệu (pixel) không chỉ phụ thuộc vào đặc trưng của chính nó mà còn bị ảnh hưởng bởi khu vực xung quanh.

Do đó, việc lựa chọn mô hình cần xem xét kỹ khả năng xử lý dữ liệu raster, khả năng học các đặc trưng không gian, và cả tính toán hiệu quả với dữ liệu lớn.

3.3.1 Đánh giá các mô hình trí tuệ nhân tạo và đề xuất lựa chọn mô hình.

1. Nhóm mô hình học máy

a. Rừng ngẫu nhiên (Random Forest – RF)

RF là một thuật toán học máy phổ biến với ưu điểm:

- Có khả năng xử lý dữ liệu có nhiều đặc trưng mà không cần giả định phân phối.
- Khả năng chống overfitting tốt nhờ trung bình hóa nhiều cây quyết định.
- Dữ liệu đầu vào: RF yêu cầu dữ liệu dạng bảng (tabular), tức là mỗi pixel được biểu diễn dưới dạng một vector đặc trưng gồm 20 giá trị từ 20 raster tương ứng.

Hạn chế: RF không tự động nắm bắt thông tin không gian, tức là không xem xét mối liên hệ giữa pixel và các pixel lân cận. Để bổ sung yếu tố không gian, cần thủ công mở rộng đặc trưng (ví dụ: tính toán giá trị trung bình hoặc phương sai của từng raster trong vùng 3×3 hoặc 5×5 quanh điểm trung tâm).

b. Máy hỗ trợ vectơ (Support Vector Machine – SVM)

SVM là thuật toán mạnh về phân loại nhị phân:

- Đặc biệt hiệu quả khi dữ liệu không tuyến tính (nhờ kernel trick).
- Ôn định với dữ liệu nhỏ và trung bình.
- Dữ liệu đầu vào: Giống RF, SVM nhận đầu vào là vector đặc trưng tại từng pixel.

Hạn chế: SVM cũng không tự khai thác đặc trưng không gian. Ngoài ra, với tập dữ liệu lớn (nhiều pixel), chi phí tính toán của SVM có thể rất cao.

c. Hồi quy logistic (Logistic Regression – LR)

LR là mô hình thống kê cổ điển:

- Dễ triển khai, dễ diễn giải.
- Thích hợp cho những bài toán phân loại tuyến tính.
- Dữ liệu đầu vào: Mỗi pixel là một vector đặc trưng gồm 20 đặc trưng.

Hạn chế: Khả năng học phi tuyến kém hơn RF và SVM, và hoàn toàn không nắm bắt được ngữ cảnh không gian trừ khi có đặc trưng bổ sung.

d. Light Gradient Boosting Machine (LGBM)

LGBM là thuật toán gradient boosting hiện đại với nhiều cải tiến vượt trội:

- Tốc độ huấn luyện nhanh nhờ kỹ thuật leaf-wise tree growth thay vì level-wise như các thuật toán boosting truyền thống.
- Hiệu quả xử lý dữ liệu lớn với yêu cầu bộ nhớ thấp nhờ tối ưu hóa lưu trữ histogram.
- Khả năng học mồi quan hệ phi tuyến phức tạp giữa các đặc trưng thông qua việc kết hợp tuần tự nhiều weak learner.
- Tích hợp sẵn các kỹ thuật regularization để giảm overfitting như dropout cho boosting và constraining tree complexity.
- Dữ liệu đầu vào: LGBM xử lý dữ liệu dạng bảng, với mỗi pixel được biểu diễn bởi vector 20 đặc trưng từ các raster tương ứng.

Hạn chế: Tương tự RF và SVM, LGBM không tự động khai thác thông tin không gian giữa các pixel lân cận. Mô hình chỉ xem xét từng pixel như một điểm dữ liệu độc lập, không nắm bắt được cấu trúc không gian của ảnh viễn thám. Để tận dụng ngữ cảnh

không gian, cần bổ sung thủ công các đặc trưng texture hoặc spatial features như local variance, spatial autocorrelation trong cửa sổ lân cận.

2. Nhóm mô hình học sâu

a. Mạng nơ-ron sâu (Deep Neural Network – DNN)

DNN là phiên bản mở rộng của mạng perceptron đa lớp:

- Có khả năng học phi tuyến mạnh mẽ.
- Linh hoạt và dễ tùy biến về kiến trúc.
- Dữ liệu đầu vào: DNN nhận vector đặc trưng giống như RF/SVM, có thể là 20 đặc trưng gốc hoặc mở rộng (ví dụ 20 đặc trưng + 20 đặc trưng từ vùng lân cận).
- Tự động học các mối quan hệ phi tuyến phức tạp.

Hạn chế: Dù mạnh về học phi tuyến, DNN không tự nắm bắt spatial pattern nếu dữ liệu chỉ là vector.

b. Mạng nơ-ron tích chập (Convolutional Neural Network – CNN)

CNN nổi bật nhờ khả năng xử lý dữ liệu dạng ảnh hoặc raster:

- Có khả năng học đặc trưng không gian (spatial pattern) thông qua các lớp convolution.
- Phù hợp tự nhiên cho bài toán bản đồ và dữ liệu raster.
- Dữ liệu đầu vào: Một patch raster (ví dụ: 15×15 pixel) với 20 kênh tương ứng với 20 raster đặc trưng. Pixel trung tâm của patch là điểm cần dự đoán.
- Học được mối quan hệ không gian giữa pixel và khu vực xung quanh.
- Tận dụng tốt cấu trúc dữ liệu raster mà không cần thủ công mở rộng đặc trưng.

Lưu ý: Kích thước patch cần hợp lý để vừa đảm bảo có đủ thông tin không gian, vừa tối ưu tài nguyên tính toán.

c. Mạng nơ-ron hồi tiếp dài ngắn hạn (Long Short-Term Memory – LSTM)

LSTM là một kiến trúc nối bật cho chuỗi thời gian:

- Có khả năng học các phụ thuộc dài hạn trong dữ liệu tuần tự.
- Rất mạnh nếu dữ liệu mang tính chất thời gian.
- Dữ liệu đầu vào: Chuỗi đặc trưng theo thời gian tại từng pixel (ví dụ: mưa tích lũy theo giờ/ngày).

Hạn chế: Trong bài toán hiện tại, dữ liệu chính là các raster tĩnh (không có chuỗi thời gian rõ ràng), do đó LSTM không thực sự phù hợp nếu chỉ có dữ liệu tĩnh. Tuy nhiên, nếu mở rộng thêm dữ liệu khí tượng theo thời gian, LSTM hoặc hybrid CNN-LSTM có thể phát huy sức mạnh.

3.3.2 Đánh giá tổng hợp và lựa chọn mô hình xác định lũ quét

Bảng 11. Đặc điểm tổng hợp các mô hình trí tuệ nhân tạo trong nghiên cứu lũ quét

Mô hình	Định dạng dữ liệu	Khả năng học phi tuyến	Khả năng khai thác không gian	Độ phức tạp tính toán	Phù hợp bài toán hiện tại
RF	Vector	Tốt	Không (cần thủ công)	Trung bình	Tốt
LGB M	Vector	Rất tốt	Không (cần thủ công)	Thấp	Rất tốt
SVM	Vector	Trung bình	Không (cần thủ công)	Cao	Khá
LR	Vector	Yếu	Không	Thấp	Trung bình
DNN	Vector	Rất tốt	Không (cần thủ công)	Cao	Tốt
CNN	Ảnh raster (patch)	Rất tốt	Tự động học các đặc trưng không gian	Cao	Rất tốt
LSTM	Chuỗi thời gian	Rất tốt	Không	Cao	Hạn chế (trừ khi dữ liệu tuân theo chuỗi thời gian liên tục)

Trong quá trình lựa chọn mô hình dự đoán nguy cơ lũ quét, việc cân nhắc giữa các phương pháp học máy và học sâu phải dựa trên cả đặc điểm dữ liệu và năng lực của từng mô hình. Bộ dữ liệu sử dụng trong nghiên cứu bao gồm 20 raster đặc trưng, đại diện cho các yếu tố địa lý, khí tượng và thủy văn khác nhau. Đây là dữ liệu dạng không gian (spatial data), mà mỗi pixel trong bản đồ nguy cơ lũ quét không chỉ chịu tác động của các yếu tố tại chính nó mà còn phụ thuộc rất lớn vào điều kiện của khu vực xung quanh. Chính vì thế, việc lựa chọn mô hình cần xem xét khả năng khai thác thông tin không gian một cách hiệu quả.

Mạng nơ-ron tích chập (CNN) nổi bật là một lựa chọn phù hợp nhờ kiến trúc được thiết kế chuyên biệt để xử lý dữ liệu ảnh hoặc raster. Khác với các mô hình học máy truyền thống vốn chỉ xem xét từng điểm dữ liệu như một vector đặc trưng rời rạc, CNN có khả năng tiếp nhận đầu vào dạng patch raster (ví dụ: một vùng 15×15 pixel với 20 kênh đặc trưng), trong đó pixel trung tâm là đối tượng dự đoán. Nhờ cơ chế convolution, CNN tự động học được mối quan hệ không gian giữa pixel và vùng lân cận, cho phép mô hình nhận diện các pattern phức tạp như đường rãnh, sườn dốc hay khu vực tập trung nước mưa – những yếu tố quan trọng trong nguy cơ lũ quét. Đây chính là lợi thế vượt trội mà các mô hình như Logistic Regression, SVM hay thậm chí DNN không thể tự động làm được nếu chỉ sử dụng vector đặc trưng đơn lẻ. Việc xây dựng dữ liệu đầu vào dưới dạng patch cũng giúp CNN linh hoạt mở rộng quy mô mà không làm mất đi tính liên kết không gian.

Tuy nhiên, CNN yêu cầu tài nguyên tính toán lớn và thời gian huấn luyện dài hơn so với các phương pháp truyền thống. Trong bối cảnh đó, Rừng ngẫu nhiên (RF) trở

thành một lựa chọn thay thế hiệu quả, nhất là khi cần một mô hình dễ triển khai và đánh giá nhanh. RF xử lý tốt dữ liệu có nhiều đặc trưng và không yêu cầu giả định phức tạp về phân phối dữ liệu. Dù không có khả năng học thông tin không gian một cách trực tiếp như CNN, RF cho phép người nghiên cứu mở rộng bộ đặc trưng bằng cách tích hợp thêm các thông kê không gian thủ công, ví dụ như giá trị trung bình hoặc phương sai trong vùng lân cận 5×5 hoặc 7×7 pixel quanh điểm trung tâm. Cách tiếp cận này tuy không tối ưu bằng CNN nhưng lại là một giải pháp trung gian đáng tin cậy, giúp tận dụng sức mạnh của RF mà vẫn bổ sung yếu tố không gian ở mức độ nhất định.

Mặc dù vậy, khi triển khai thực tế trên tập dữ liệu raster gồm 20 lớp đặc trưng không gian với số lượng ô raster rất lớn, mô hình RF bộc lộ một hạn chế rõ rệt: thời gian dự đoán chậm do phải thực hiện voting trên hàng trăm cây quyết định, mỗi cây lại duyệt qua toàn bộ tập dữ liệu.

Trước thực tế đó, Light Gradient Boosting Machine (LightGBM) mới nổi lên trong những năm gần đây như một giải pháp thay thế tối ưu hơn (do đó chưa có nhiều nghiên cứu về lũ quét sử dụng). LightGBM về bản chất vẫn thuộc nhóm các mô hình cây quyết định – tương tự RF – nhưng cải tiến bằng kỹ thuật gradient boosting. Điều này giúp LightGBM khắc phục hai vấn đề chính mà RF gặp phải: (1) tối ưu hóa chất lượng dự đoán nhờ việc xây dựng các cây liên tiếp để liên tục sửa lỗi dự đoán, và (2) tốc độ huấn luyện cũng như dự đoán nhanh hơn nhờ thuật toán leaf-wise và kỹ thuật chia nhỏ theo histogram, tối ưu cho dữ liệu lớn.

Ngoài ra, LightGBM vẫn duy trì được các ưu điểm cốt lõi vốn là lý do lựa chọn RF: khả năng xử lý tốt dữ liệu không chuẩn hóa, đánh giá tầm quan trọng đặc trưng rõ ràng, và khả năng phân tích cấu trúc cây quyết định để hỗ trợ diễn giải mô hình. Điểm mạnh này giúp LightGBM không chỉ đáp ứng được yêu cầu về hiệu quả tính toán mà còn giữ nguyên tính minh bạch và dễ giải thích trong ứng dụng khoa học và quản lý tài nguyên thiên nhiên.

Từ các phân tích trên có thể kết luận rằng, trong điều kiện lý tưởng về tài nguyên tính toán và mục tiêu tối ưu hóa chất lượng mô hình, CNN là lựa chọn ưu tiên nhờ khả năng khai thác triệt để cấu trúc không gian của dữ liệu raster. Trong khi đó, LightGBM đóng vai trò như một mô hình nền tảng, vừa đơn giản triển khai vừa cho kết quả khá tốt khi biết cách mở rộng đặc trưng hợp lý. Việc kết hợp cả hai mô hình này trong nghiên cứu không chỉ giúp so sánh hiệu quả giữa học sâu và học máy truyền thống mà còn tăng độ tin cậy và tính khách quan của kết quả cuối cùng.

Trong nghiên cứu này, nhóm nghiên cứu sẽ phát triển tất cả các mô hình để có sự đánh giá tốt hơn bao gồm 6 mô hình bao gồm: RF, SVM, LR, LGBM, CNN và DNN. Riêng mô hình LSTM được đánh giá là chưa phù hợp trong nghiên cứu này đối với dữ liệu không phải là dạng chuỗi theo thời gian.

KẾT LUẬN

Qua quá trình phân tích và đánh giá toàn diện các phương pháp trí tuệ nhân tạo, nghiên cứu đã đưa ra những kết luận quan trọng về việc ứng dụng các mô hình AI trong bài toán phân vùng nguy cơ lũ quét. Việc lựa chọn và tối ưu hóa mô hình không chỉ phụ thuộc vào đặc tính kỹ thuật của từng thuật toán mà còn cần xem xét kỹ lưỡng bản chất của dữ liệu không gian và yêu cầu thực tiễn của bài toán quản lý thiên tai.

Nhóm mô hình học máy truyền thống đã thể hiện những ưu điểm rõ rệt trong việc xử lý dữ liệu đặc trưng đa dạng và khả năng diễn giải kết quả. Trong số đó, Random Forest (RF) khẳng định được vị thế như một baseline model đáng tin cậy nhờ khả năng chống overfitting tốt và tính ổn định cao. Tuy nhiên, hạn chế về tốc độ xử lý đã được khắc phục hiệu quả bởi LightGBM (LGBM), mô hình nổi bật với kiến trúc leaf-wise growth và các kỹ thuật tối ưu hóa tiên tiến. LGBM không chỉ vượt trội về tốc độ huấn luyện và dự đoán mà còn duy trì được chất lượng dự đoán cao, đặc biệt phù hợp với bài toán có khối lượng dữ liệu lớn như phân tích raster đa lớp.

Support Vector Machine (SVM) và Logistic Regression (LR) đóng vai trò quan trọng trong việc cung cấp các baseline method, giúp đánh giá mức độ cải thiện của các mô hình phức tạp hơn. Mặc dù SVM có khả năng xử lý dữ liệu phi tuyến tốt thông qua kernel trick, nhưng chi phí tính toán cao và khó khăn trong việc xử lý dữ liệu quy mô lớn là những hạn chế đáng kể. LR, với tính đơn giản và khả năng diễn giải tuyệt vời, phù hợp cho các ứng dụng cần tính minh bạch cao trong quá trình ra quyết định.

Nhóm mô hình học sâu mang lại những đột phá quan trọng, đặc biệt trong việc khai thác cấu trúc không gian của dữ liệu. Mạng nơ-ron tích chập (CNN) đã chứng minh được khả năng vượt trội trong việc học các đặc trưng không gian phức tạp thông qua cơ chế convolution. Việc tích hợp các kỹ thuật tiên tiến như residual connections, attention mechanism, và spatial dropout sẽ giúp CNN không chỉ học được các pattern địa lý phức tạp mà còn tự động xác định được các vùng có ý nghĩa quan trọng trong việc đánh giá nguy cơ lũ quét. Cấu trúc encoder-decoder (U-Net style) ($64 \rightarrow 128 \rightarrow 256 \rightarrow 128 \rightarrow 64 \rightarrow 5$) cho phép mô hình học được các đặc trưng từ mức độ cơ bản đến phức tạp, sau đó lại tự động giảm dần và phục hồi các thông tin không gian để tạo ra đầu ra với 5 kênh phân loại, phù hợp với tính chất đa tầng của các yếu tố địa lý tác động đến lũ quét.

Deep Neural Network (DNN), với kiến trúc fully-connected layers, thể hiện khả năng học các mối quan hệ phi tuyến phức tạp giữa các đặc trưng. Mặc dù không có khả năng khai thác thông tin không gian tự động như CNN, DNN vẫn đạt được hiệu quả tốt nhờ khả năng học các interaction pattern giữa 20 đặc trưng đầu vào. Việc áp dụng pyramid architecture ($100 \rightarrow 256 \rightarrow 128 \rightarrow 64$) và progressive dropout strategy sẽ giúp mô hình cân bằng được giữa khả năng học và tránh overfitting.

Trong quá trình phân tích, Long Short-Term Memory (LSTM) đã được loại bỏ khỏi danh sách các mô hình triển khai do bản chất không phù hợp với đặc tính của bài toán phân vùng lũ quét hiện tại. LSTM được thiết kế đặc biệt để xử lý dữ liệu chuỗi thời gian (sequential data) với khả năng học các phụ thuộc dài hạn (long-term dependencies). Tuy nhiên, bộ dữ liệu trong nghiên cứu này chủ yếu bao gồm các raster đặc trưng tĩnh như địa hình, thảm phủ, loại đất, và các chỉ số CN - những yếu tố không thay đổi theo thời gian hoặc thay đổi rất chậm.

Mặc dù LSTM có thể phát huy sức mạnh nếu được tích hợp với dữ liệu khí tượng theo chuỗi thời gian (như lượng mưa theo giờ, mực nước sông theo ngày), nhưng trong bối cảnh nghiên cứu hiện tại với dữ liệu chủ yếu là static spatial features, việc ép buộc sử dụng LSTM sẽ không mang lại hiệu quả tốt mà có thể dẫn đến hiệu quả kém hơn các mô hình được thiết kế phù hợp với dữ liệu không gian tĩnh. Do đó, nhóm nghiên cứu quyết định loại bỏ LSTM khỏi danh sách mô hình nghiên cứu phân vùng nguy cơ lũ quét trong nhiệm vụ này.

TÀI LIỆU THAM KHẢO

- [1] *Quyết định số 18/2021/QĐ-TTg ngày 22 tháng 4 năm 2021 quy định về dự báo, cảnh báo, truyền tin thiên tai và cấp độ rủi ro thiên tai*, 2021. Địa chỉ: vanban.chinhphu.vn/default.aspx?pageid=27160&docid=203152.
- [2] Romdani, R. P, Tamamadin, M, Susandi, A, Pratama, A and Wijaya, A. R, "Development of Flash Flood Hazard Map in Bima City (NTB) using Analytical Hierarchy Process," *IOP Conference Series: Earth and Environmental Science*, vol. 166, p. 012035, 2018. DOI:10.1088/1755-1315/166/1/012035.
- [3] S. Talha, M. Maanan, H. Atika and H. Rhinane, "Prediction of flash flood susceptibility using Fuzzy Analytical Hierarchy Process (FAHP) algorithms and gis: A study case of Guelmim region in southwestern of Morocco," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vols. XLII-4/W19, pp. 407-414, 2019. DOI:10.5194/isprs-archives-xlii-4-w19-407-2019.
- [4] Chen Cao, Peihua Xu, Yihong Wang and et al., "Flash Flood Hazard Susceptibility Mapping Using Frequency Ratio and Statistical Index Methods in Coalmine Subsidence Areas," *Sustainability*, vol. 8, no. 9, p. 948, 2016. DOI:10.3390/su8090948.
- [5] A Saleh, N Sabtu and M R Bunmi, "Flash Flood Susceptibility Mapping in Sungai Pinang catchment using Weight of Evidence," *IOP Conference Series: Earth and Environmental Science*, vol. 1091, no. 1, p. 012017, 2022. DOI:10.26226/m.61a82ed54a84e7b4701d8bdb.
- [6] Costache, Romulus, Pham, Quoc Bao, Sharifi, Ehsan and et al., "Flash-Flood Susceptibility Assessment Using Multi-Criteria Decision Making and Machine Learning Supported by Remote Sensing and GIS Techniques," *Remote Sensing*, vol. 12, p. 106, 2019. DOI:10.3390/rs12010106.
- [7] Zhao, Gang, Liu, Ronghua, Yang, Mingxiang, Tu, Tongbi, Ma, Meihong, Hong, Yang and Wang, Xiekang, "Large-scale flash flood warning in China using deep learning," *Journal of Hydrology*, vol. 604, p. 127222, 2022. DOI:10.1016/j.jhydrol.2021.127222.
- [8] Carpenter, T.M., Sperfslage, J.A., Georgakakos, K.P., Sweeney, T. and Fread, D.L., "National threshold runoff estimation utilizing GIS in support of operational

- flash flood warning systems," *Journal of Hydrology*, vol. 224, no. 1-2, p. 21–44, 1999. DOI:10.1016/S0022-1694(99)00115-8.
- [9] Forest Service, "CAUSES OF FLASHY FLOODS AND MUD FLOWS IN UTAH," *Monthly Weather Review*, vol. 59, p. 122–122, 1931. DOI:10.1175/1520-0493(1931)59<122a:coffam>2.0.co;2.
- [10] Kuz'min, K. K., "The catastrophic flash flood of 1973 and the medeo dam," *Hydrotechnical Construction*, vol. 8, no. 3, p. 203–206, 1974. DOI:10.1007/bf02403378.
- [11] Hales, John E., "The Kansas City Flash Flood of 12 September 1977," *Bulletin of the American Meteorological Society*, vol. 59, no. 6, p. 706–710, 1978. DOI:10.1175/1520-0477-59.6.706.
- [12] NOAA, National Weather Service forecasting handbook (No. 1 - 1979), University of Michigan Library (January 1, 1979), 1979. Địa chỉ: books.google.com.vn/books?id=8jRRAAAAMAAJ&printsec=frontcover&hl=vi&#v=onepage&q&f=false.
- [13] Georgakakos, Konstantine P., "A generalized stochastic hydrometeorological model for flood and flash-flood forecasting: 1. Formulation," *Water Resources Research*, vol. 22, no. 13, p. 2083–2095, 1986. DOI:10.1029/WR022i013p02096.
- [14] Sweeney, Timothy L., "Modernized Areal Flash Flood Guidance," NOAA Technical Memorandum NWS HYDRO 44, 1992. Địa chỉ: <blob:https://repository.library.noaa.gov/17df97fc-73c5-44db-8f38-5064333c9f3d>.
- [15] William W. Emmett, "The Channels and Waters of the Upper Salmon River Area, Idaho," United States Government Printing Office, Washington, 1975. DOI:10.3133/pp870a.
- [16] Chang, Tiao J. and Sun, Hong Y., "Study of Potential Flash Floods by Kriging Method," *Journal of Hydrologic Engineering*, vol. 2, no. 3, p. 104–108, 1997. DOI:10.1061/(asce)1084-0699(1997)2:3(104).
- [17] Sharada, D., Devi, D. Kaveri, Prasad, S. and Kumar, Seelan Santhosh, "Modelling flash flood hazard to a railway line: A GIS approach," *Geocarto International*, vol. 12, no. 3, p. 77–82, 1997. DOI:10.1080/10106049709354600.

- [18] Marco Borga, Markus Stoffel, Lorenzo Marchi, Francesco Marra and Matthias Jakob, "Hydrogeomorphic response to extreme rainfall in headwater systems: Flash floods and debris flows," *Journal of Hydrology*, vol. 518, pp. 194-205, 2014. DOI:10.1016/j.jhydrol.2014.05.022.
- [19] Shahin Khosh Bin Ghomash, Daniel Bachmann, Daniel Caviedes-Voulli`eme and Christoph Hinz, "Impact of Rainfall Movement on Flash Flood Response: A Synthetic Study of a Semi-Arid Mountainous Catchment," *Water*, vol. 14, no. 12, p. 1844, 2022. DOI:10.3390/w14121844.
- [20] Lorenzo Alfieri, Marc Berenguer, Valentin Knechtl and et al., "Handbook of Hydrometeorological Ensemble Forecasting," in *Flash Flood Forecasting Based on Rainfall Thresholds*, Berlin, Springer Berlin Heidelberg, 2015, pp. 1-38. DOI:10.1007/978-3-642-40457-3_49-1.
- [21] L. Alfieri, D. Velasco and J. Thielen, "Flash flood detection through a multi-stage probabilistic warning system for heavy precipitation events," *Advances in Geosciences*, vol. 29, pp. 69-75, 2011. DOI:10.5194/adgeo-29-69-2011.
- [22] Nel Caine, "The Rainfall Intensity: Duration Control of Shallow Landslides and Debris Flows," *Geografiska Annaler. Series A, Physical Geography*, vol. 62, no. 1/2, p. 23, 1980. DOI:10.2307/520449.
- [23] Fausto Guzzetti, Silvia Peruccacci, Mauro Rossi and Colin P. Stark, "The rainfall intensity-duration control of shallow landslides and debris flows: an update," *Landslides*, vol. 5, no. 1, pp. 3-17, 2007. DOI:10.1007/s10346-007-0112-1.
- [24] Nejc Bezak, Mojca vSraj and Matjaz Mikovs, "Copula-based IDF curves and empirical rainfall thresholds for flash floods and rainfall-induced landslides," *Journal of Hydrology*, vol. 541, pp. 272--284, 2016. DOI:10.1016/j.jhydrol.2016.02.058.
- [25] T. Turkington, J. Ettema, C. J. van Westen and K. Breinl, "Empirical atmospheric thresholds for debris flows and flash floods in the southern French Alps," *Natural Hazards and Earth System Sciences*, vol. 14, no. 6, pp. 1517-1530, 2014. DOI:10.5194/nhess-14-1517-2014.
- [26] Geraldo Moura Ramos Filho, Victor Hugo Rabelo Coelho, Emerson da Silva Freitas and et al., "An improved rainfall-threshold approach for robust prediction and warning of flood and flash flood hazards," *Natural Hazards*, vol. 105, no. 3, pp. 2409-2429, 2020. DOI:10.1007/s11069-020-04405-x.

- [27] Xiaoyan Zhai, Liang Guo, Ronghua Liu and Yongyong Zhang, "Rainfall threshold determination for flash flood warning in mountainous catchments with consideration of antecedent soil moisture and rainfall pattern," *Natural Hazards*, vol. 94, no. 2, pp. 605-625, 2018. DOI:10.1007/s11069-018-3404-y.
- [28] Wenlin Yuan, Xinyu Tu, Chengguo Su and et al., "Research on the Critical Rainfall of Flash Floods in Small Watersheds Based on the Design of Characteristic Rainfall Patterns," *Water Resources Management*, vol. 35, no. 10, pp. 3297-3319, 2021. DOI:10.1007/s11269-021-02893-5.
- [29] Mohamed Saber and Koray Yilmaz, "Evaluation and Bias Correction of Satellite-Based Rainfall Estimates for Modelling Flash Floods over the Mediterranean region: Application to Karpuz River Basin, Turkey," *Water*, vol. 10, no. 5, p. 657, 2018. DOI:10.3390/w10050657.
- [30] I. K. Westerberg, J.-L. Guerrero, P. M. Younger and et al., "Calibration of hydrological models using flow-duration curves," *Hydrology and Earth System Sciences*, vol. 15, no. 7, pp. 2205-2227, 2011. DOI:10.5194/hess-15-2205-2011.
- [31] F. Silvestro, N. Rebora, G. Cummings and L. Ferraris, "Experiences of dealing with flash floods using an ensemble hydrological nowcasting chain: implications of communication, accessibility and distribution of the results," *Journal of Flood Risk Management*, vol. 10, no. 4, pp. 446-462, 2015. DOI:10.1111/jfr3.12161.
- [32] Seann Reed, John Schaake and Ziya Zhang, "A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations," *Journal of Hydrology*, vol. 337, no. 3-4, pp. 402-420, 2007. DOI:10.1016/j.jhydrol.2007.02.015.
- [33] Jonathan D Phillips, "Geomorphic impacts of flash flooding in a forested headwater basin," *Journal of Hydrology*, vol. 269, no. 3-4, pp. 236-250, 2002. DOI:10.1016/s0022-1694(02)00280-9.
- [34] Thomas L Saaty, "A scaling method for priorities in hierarchical structures," *Journal of Mathematical Psychology*, vol. 15, no. 3, pp. 234-281, 1977. DOI:10.1016/0022-2496(77)90033-5.
- [35] Aqil Tariq, Jianguo Yan, Bushra Ghaffar and et al., "Flash Flood Susceptibility Assessment and Zonation by Integrating Analytic Hierarchy Process and Frequency Ratio Model with Diverse Spatial Data," *Water*, vol. 14, no. 19, p. 3069, 2022. DOI:10.3390/w14193069.

- [36] Kairong Lin, Haiyan Chen, Chong-Yu Xu and et al., "Assessment of flash flood risk based on improved analytic hierarchy process method and integrated maximum likelihood clustering algorithm," *Journal of Hydrology*, vol. 584, p. 124696, 2020. DOI:10.1016/j.jhydrol.2020.124696.
- [37] Mohammed Sadek and Xuxiang Li, "Low-Cost Solution for Assessment of Urban Flash Flood Impacts Using Sentinel-2 Satellite Images and Fuzzy Analytic Hierarchy Process: A Case Study of Ras Ghareb City, Egypt," *Advances in Civil Engineering*, vol. 2019, pp. 1-15, 2019. DOI:10.1155/2019/2561215.
- [38] Romulus Costache and Liliana Zaharia, "Flash-flood potential assessment and mapping by integrating the weights-of-evidence and frequency ratio statistical methods in GIS environment - case study: Basca Chiojdului River catchment (Romania)," *Journal of Earth System Science*, vol. 126, no. 4, pp. 1-19, 2017. DOI:10.1007/s12040-017-0828-9.
- [39] Amoroch, J. and Brandstetter, A., "Determination of Nonlinear Functional Response Functions in Rainfall-Runoff Processes," *Water Resources Research*, vol. 7, no. 5, pp. 1087-1101, 1971. DOI:10.1029/wr007i005p01087.
- [40] Hsu, Kuo-lin, Gupta, Hoshin Vijai and Sorooshian, Soroosh, "Artificial Neural Network Modeling of the Rainfall-Runoff Process," *Water Resources Research*, vol. 31, no. 10, pp. 2517-2530, 1995. DOI:10.1029/95wr01955.
- [41] Sahoo, G.B., Ray, C. and De Carlo, E.H., "Use of neural network to predict flash flood and attendant water qualities of a mountainous stream on Oahu, Hawaii," *Journal of Hydrology*, vol. 327, no. 3-4, pp. 525-538, 2006. DOI:10.1016/j.jhydrol.2005.11.059.
- [42] Janál, Petr and Starý, Miloš, "Fuzzy model use for prediction of the state of emergency of river basin in the case of flash flood," *Journal of Hydrology and Hydromechanics*, vol. 57, no. 3, 2009. DOI:10.2478/v10098-009-0013-1.
- [43] Toukourou, Mohamed Samir, Johannet, Anne and Dreyfus, Gérard, "Flash Flood Forecasting by Statistical Learning in the Absence of Rainfall Forecast: A Case Study," in *Communications in Computer and Information Science*, Springer Berlin Heidelberg, 2009, pp. 98-107. DOI:10.1007/978-3-642-03969-0_10.
- [44] Lamovec, Peter, Veljanovski, Tatjana, Mikoš, Matjaž and Oštir, Krištof, "Detecting flooded areas with machine learning techniques: case study of the

Selška Sora river flash flood in September 2007," *Journal of Applied Remote Sensing*, vol. 7, no. 1, p. 073564, 2013. DOI:10.1117/1.jrs.7.073564.

- [45] Piotrowski, A., Napiórkowski, J. J. and Rowiński, P.M., "Flash-flood forecasting by means of neural networks and nearest neighbour approach – a comparative study," *Nonlinear Processes in Geophysics*, vol. 13, no. 4, pp. 443-448, 2006. DOI:10.5194/npg-13-443-2006.
- [46] Kong A Siou, L., Johannet, A., Pistre, S. and Borrell, V., "Flash Floods Forecasting in a Karstic Basin Using Neural Networks: the Case of the Lez Basin (South of France)," in *Environmental Earth Sciences*, Springer Berlin Heidelberg, 2010, pp. 215-221. DOI:10.1007/978-3-642-12486-0_33.
- [47] Artigue, G., Johannet, A., Borrell, V. and Pistre, S., "Flash floods forecasting without rainfalls forecasts by recurrent neural networks. Case study on the Mialet basin (Southern France)," in *2011 Third World Congress on Nature and Biologically Inspired Computing*, 2011. DOI:10.1109/nabic.2011.6089612.
- [48] Izyan 'Izzati Abdul Rahman and Nik Mohd Asrol Alias, "Rainfall forecasting using an artificial neural network model to prevent flash floods," in *8th International Conference on High-capacity Optical Networks and Emerging Technologies*, 2011. DOI:10.1109/honet.2011.6149841.
- [49] Boukharouba, Khaled, Roussel, Pierre, Dreyfus, Gerard and Johannet, Anne, "Flash flood forecasting using Support Vector Regression: An event clustering based approach," in *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2013. DOI:10.1109/mlsp.2013.6661958.
- [50] Cao Đăng Dư and Lương Tuấn Anh, "Phân vùng khả năng xuất hiện lũ quét," *Tạp chí Khí tượng thủy văn*, p. 1÷8, 1995. Địa chỉ: tapchikttv.vn/data/article/2126/2.pdf.
- [51] Phạm Thị Hương Lan and Vũ Minh Cát, "Một số kết quả nghiên cứu xây dựng bản đồ tiềm năng lũ quét phục vụ công tác cảnh báo lũ quét vùng núi đồng bắc Việt Nam," *Tạp chí Khí tượng Thủy văn*, vol. 556, no. 2, pp. 11-16, 2008. Địa chỉ: http://tapchikttv.vn/data/article/1401/thang%202%202008_02.pdf.
- [52] Lã Thanh Hà, "Nghiên cứu xây dựng bản đồ phân vùng nguy cơ lũ quét phục vụ công tác phòng tránh lũ quét cho tỉnh Yên Bái," *Tạp chí Khí tượng Thủy văn*, vol. 578, no. 2, pp. 11-15, 2009. Địa chỉ: <http://tapchikttv.vn/data/article/362/B%C3%A0i%203.pdf>.

- [53] Dương Thị Lợi and Đặng Phương Lan, "Ứng dụng mô hình đa chi tiêu nhằm đánh giá nguy cơ lũ quét trong bối cảnh biến đổi khí hậu toàn cầu. Trường hợp nghiên cứu cụ thể: miền núi Tây Bắc - Việt Nam," *Tạp chí Khí tượng thủy văn*, vol. 721, p. 31÷45, 2021. Địa chỉ: <http://tapchikttv.vn/data/article/1388/4.pdf>.
- [54] Minh Đức, Đào, Cao Minh, Vũ, Hải Yến, Hoàng, Quang Anh, Phạm and Kinh Bắc, Đặng, "Đánh giá nguy cơ hình thành lũ quét trên suối Nghĩa Đô, huyện Bảo Yên, tỉnh Lào Cai bằng phương pháp phân tích thống kê," *Vietnam Journal of Hydrometeorology*, vol. EME4, no. 1, pp. 341-354, 2022. DOI:10.36335/vnjhm.2022(eme4).341-354.
- [55] Thị Huyền, Nguyễn, Quốc Khánh, Nguyễn, Huy Dương, Nguyễn, Hoàng Ninh, Nguyễn and Đức Hà, Nguyễn, "Kết quả khoanh định các khu vực nhạy cảm về trượt lở, lũ quét khu vực Thành phố Đà Nẵng," *Vietnam Journal of Hydrometeorology*, vol. 1, no. 745, pp. 21-33, 2023. DOI:10.36335/vnjhm.2023(745).21-33.
- [56] Thị Phương Thảo, Ngô, Hùng Long, Ngô, Anh Tuấn, Trần and Minh Hằng, Lê, "Sử dụng ảnh Sentinel-1A đa thời gian để phát hiện lũ quét, thử nghiệm tại tỉnh Lào Cai," *Journal of Hydro-meteorology*, vol. 8, no. 764, pp. 29-37, 2024. DOI:10.36335/vnjhm.2024(764).29-37.
- [57] Ngô Thị Phương Thảo, "Luận án: Nghiên cứu phát triển mô hình trí tuệ nhân tạo trong phân vùng nguy cơ lũ quét ở Việt Nam," Trường đại học Mỏ - Địa chất, 2024. Địa chỉ: <https://humg.edu.vn/content/tintuc/Lists/News/Attachments/9000/Luan%20an%20tien%20si-Ngo%20Thi%20Phuong%20Thao.pdf>.
- [58] Ha, Hang, Bui, Quynh Duy, Khuc, Thanh Dong, Tran, Dinh Trong, Pham, Binh Thai, Mai, Sy Hung, Nguyen, Lam Phuong and Luu, Chinh, "A machine learning approach in spatial predicting of landslides and flash flood susceptible zones for a road network," *Modeling Earth Systems and Environment*, vol. 8, no. 4, p. 4341÷4357, 2022. DOI:10.1007/s40808-022-01384-9.
- [59] Nguyễn Viết Nghĩa and Nguyễn Cao Cường, "Ứng dụng mạng nơ-ron nhân tạo đa lớp trong thành lập mô hình phân vùng lũ quét khu vực miền núi Tây Bắc, thực nghiệm tại tỉnh Yên Bái," *Tạp chí khoa học Đo đạc và Bản đồ*, vol. 44, no. 6, p. 56÷64, 2020. Địa chỉ: <https://jgac.vn/index.php/journal/article/view/304/291>.

- [60] Hoang, Duc-Vinh and Liou, Yuei-An, "Assessing the influence of human activities on flash flood susceptibility in mountainous regions of Vietnam," *Ecological Indicators*, vol. 158, no. , p. 111417, 2024. DOI:10.1016/j.ecolind.2023.111417.
- [61] Lal, Aleena B, M, Megha, P, Muhammed Sufiyan M, Thekkel, David Joseph, V, Aswathy M and Meckamalil, Rotney Roy, "Flash Flood Detection and Alert System Using Machine Learning," *International Journal for Research in Applied Science and Engineering Technology*, vol. 12, no. 6, p. 45÷51, 2024. DOI:10.22214/ijraset.2024.62971.
- [62] Sellami, El Mehdi and Rhinane, Hassan, "A modern method for building damage evaluation using deep learning approach - Case study: Flash flooding in Derna, Libya," *E3S Web of Conferences*, vol. 502, no. , p. 03010, 2024. DOI:10.1051/e3sconf/202450203010.
- [63] Lugt, Dorien, van Hoek, Mattijn, Meirink, Jan Fokke and van der Kooij, Eva, "Nowcasting for urban flash floods in Africa: a machine-learning and satellite-observation based model," , vol. , no. , p. , 2021. DOI:10.5194/egusphere-egu21-16002.
- [64] Band, Shahab S., Janizadeh, Saeid, Chandra Pal, Subodh, Saha, Asish, Chakrabortty, Rabin, Melesse, Assefa M. and Mosavi, Amirhosein, "Flash Flood Susceptibility Modeling Using New Approaches of Hybrid and Ensemble Tree-Based Machine Learning Algorithms," *Remote Sensing*, vol. 12, no. 21, p. 3568, 2020. DOI:10.3390/rs12213568.
- [65] Costache, Romulus, Arabameri, Alireza, Blaschke, Thomas, Pham, Quoc Bao, Pham, Binh Thai, Pandey, Manish, Arora, Aman, Linh, Nguyen Thi Thuy and Costache, Iulia, "Flash-Flood Potential Mapping Using Deep Learning, Alternating Decision Trees and Data Provided by Remote Sensing Sensors," *Sensors*, vol. 21, no. 1, p. 280, 2021. DOI:10.3390/s21010280.
- [66] Ilia, Ioanna, Tsangaratos, Paraskevas, Tzampoglou, Ploutarchos, Chen, Wei and Hong, Haoyuan, "Flash flood susceptibility mapping using stacking ensemble machine learning models," *Geocarto International*, vol. 37, no. 27, p. 15010÷15036, 2022. DOI:10.1080/10106049.2022.2093990.
- [67] SELLAMI, EL Mehdi and Rhinane, Hassan, "Google Earth Engine and Machine Learning for Flash Flood Exposure Mappingâ€”Case Study: Tetouan, Morocco," *Geosciences*, vol. 14, no. 6, p. 152, 2024. DOI:10.3390/geosciences14060152.

- [68] Razavi-Termeh, Seyed Vahid, Seo, MyoungBae, Sadeghi-Niaraki, Abolghasem and Choi, Soo-Mi, "Flash flood detection and susceptibility mapping in the Monsoon period by integration of optical and radar satellite imagery using an improvement of a sequential ensemble algorithm," *Weather and Climate Extremes*, vol. 41, no. , p. 100595, 2023. DOI:10.1016/j.wace.2023.100595.
- [69] MDE, "Method for Designing Infiltration Structures," [Trực tuyến]. Địa chỉ: https://mde.maryland.gov/programs/water/StormwaterManagementProgram/Documents/www.mde.state.md.us/assets/document/sedimentstormwater/Appnd_D13.pdf. [Truy cập 15/6/2023].
- [70] NRCS, "Soil Infiltration," [Trực tuyến]. Địa chỉ: https://web.archive.org/web/20240301064123/https://cropwatch.unl.edu/documents/USDA_NRCS_infiltration_guide6-4-14.pdf. [Truy cập 11/6/2023].
- [71] Rahmati, Mehdi, Weihermüller, Lutz, Vanderborght, Jan, Pachepsky, Yakov A., Mao, Lili, Sadeghi, Seyed Hamidreza, Moosavi, Niloofar, Kheirfam, Hossein, Montzka, Carsten, Van Looy, Kris, Toth, Brigitta, Hazbavi, Zeinab, Al Yamani, Wafa, Albalasmeh, Ammar A., Alghzawi, Ma'in Z., Angulo-Jaramillo, Rafael, Antonino, Antônio Celso Dantas, Arampatzis, George, Armindo, Robson André, Asadi, Hossein, Bamutaze, Yazidhi, Batlle-Aguilar, Jordi, Béchet, Béatrice, Becker, Fabian, Blöschl, Günter, Bohne, Klaus, Braud, Isabelle, Castellano, Clara, Cerdà, Artemi, Chalhoub, Maha, Cichota, Rogerio, Císlarová, Milena, Clothier, Brent, Coquet, Yves, Cornelis, Wim, Corradini, Corrado, Coutinho, Artur Paiva, de Oliveira, Muriel Bastista, de Macedo, José Ronaldo, Durães, Matheus Fonseca, Emami, Hojat, Eskandari, Iraj, Farajnia, Asghar, Flammini, Alessia, Fodor, Nándor, Gharaibeh, Mamoun, Ghavimipanah, Mohamad Hossein, Ghezzehei, Teamrat A., Giertz, Simone, Hatzigiannakis, Evangelos G., Horn, Rainer, Jiménez, Juan José, Jacques, Diederik, Keesstra, Saskia Deborah, Kelishadi, Hamid, Kiani-Harchegani, Mahboobeh, Kouselou, Mehdi, Kumar Jha, Madan, Lassabatere, Laurent, Li, Xiaoyan, Liebig, Mark A., Lichner, Lubomír, López, María Victoria, Machiwal, Deepesh, Mallants, Dirk, Mallmann, Micael Stolben, de Oliveira Marques, Jean Dalmo, Marshall, Miles R., Mertens, Jan, Meunier, Félicien, Mohammadi, Mohammad Hossein, Mohanty, Binayak P., Pulido-Moncada, Mansonia, Montenegro, Suzana, Morbidelli, Renato, Moret-Fernández, David, Moosavi, Ali Akbar, Mosaddeghi, Mohammad Reza, Mousavi, Seyed Bahman, Mozaffari, Hasan, Nabiollahi, Kamal, Neyshabouri, Mohammad Reza, Ottoni, Marta Vasconcelos, Ottoni Filho, Theophilo Benedicto, Pahlavan-Rad, Mohammad Reza, Panagopoulos, Andreas, Peth, Stephan, Peyneau, Pierre-

Emmanuel, Picciafuoco, Tommaso, Poesen, Jean, Pulido, Manuel, Reinert, Dalvan José, Reinsch, Sabine, Rezaei, Meisam, Roberts, Francis Parry, Robinson, David, Rodrigo-Comino, Jesús, Rotunno Filho, Otto Corrêa, Saito, Tadaomi, Suganuma, Hideki, Saltalippi, Carla, Sádor, Renáta, Schütt, Brigitta, Seeger, Manuel, Sepehrnia, Nasrollah, Sharifi Moghaddam, Ehsan, Shukla, Manoj, Shutaro, Shiraki, Sorando, Ricardo, Stanley, Ajayi Asishana, Strauss, Peter, Su, Zhongbo, Taghizadeh-Mehrjardi, Ruhollah, Taguas, Encarnación, Teixeira, Wenceslau Geraldes, Vaezi, Ali Reza, Vafakhah, Mehdi, Vogel, Tomas, Vogeler, Iris, Votrubova, Jana, Werner, Steffen, Winarski, Thierry, Yilmaz, Deniz, Young, Michael H., Zacharias, Steffen, Zeng, Yijian, Zhao, Ying, Zhao, Hong and Vereecken, Harry, "Development and analysis of the Soil Water Infiltration Global database," *Earth System Science Data*, vol. 10, no. 3, pp. 1237-1263, 2018. DOI:10.5194/essd-10-1237-2018.

- [72] C. ROSS, L. PRIHODKO, J. ANCHANG, S. KUMAR, W. JI and N. HANAN, *Global Hydrologic Soil Groups (HYSOGs250m) for Curve Number-Based Runoff Modeling*, ORNL Distributed Active Archive Center, 2018. DOI:10.3334/ORNLDAA/1566. Địa chỉ: https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1566.
- [73] Thùy Thanh, "Yên Bai: Chủ động ứng phó với mưa lớn diện rộng từ chiều tối ngày 14 đến 16/10," Báo Yên Bai, 13/10/2020. [Trực tuyến]. Địa chỉ: <https://baoyenbai.com.vn/PrintPreview/198871/>. [Truy cập 15/6/2023].
- [74] Theo Vietnamplus, "Yên Bai: Bảy người mất tích do lũ ống, lũ quét ở Mù Cang Chải," [Trực tuyến]. Địa chỉ: <https://daidoanket.vn/yen-bai-bay-nguo-mat-tich-do-lu-ong-lu-quet-o-mu-cang-chai-10078324.html>. [Truy cập 17/4/2025].
- [75] Vũ Bá Thao and Bùi Xuân Việt, "Phân tích ngưỡng mưa phát sinh một số trận lũ quét, lũ bùn đá thuộc các tỉnh Lai Châu, Điện Biên, Yên Bai, Sơn La," *Tạp chí Khí tượng thủy văn*, vol. 749, pp. 96-110, 2023. DOI:10.36335/VNJHM.2023(749).96-110.
- [76] JICA, "Khảo sát thu thập dữ liệu về các giải pháp phòng chống lũ quét và sạt lở đất tại khu vực miền núi phía Bắc của Việt Nam," Cơ quan hợp tác quốc tế Nhật Bản (JICA), Hà Nội, 2021. Địa chỉ: https://openjicareport.jica.go.jp/pdf/12357885_01.pdf.
- [77] Thanh Thùy, "Lũ quét ở Mù Cang Chải: Ước thiệt hại khoảng 150 tỷ đồng," [Trực tuyến]. Địa chỉ:

https://baoyenbai.com.vn/12/151798/Luquet_o_Mu_Cang_Chai_Uoc_thiet_hai_khoang_150_ty_dong.htm. [Truy cập 17/4/2025].

- [78] Đình Sơn, "Mưa lũ làm 3 người chết và 11 nạn nhân mất tích ở Yên Bai," 2018. [Trực tuyến]. Địa chỉ: <https://znews.vn/mua-lu-lam-3-nguoi-chet-va-11-nan-nhan-mat-tich-o-yen-bai-post861935.html>. [Truy cập 17/4/2025].
- [79] Mạnh Cường, "Thiệt hại do mưa, lũ gây ra trên địa bàn huyện Mù Cang Chải ước khoảng 920 triệu đồng," Trang Thông tin điện tử Mù Cang Chải, 22/7/2019. [Trực tuyến]. Địa chỉ: <https://mucangchai.yenbai.gov.vn/news/tin-moi/?UserKey=Thiet-hai-do-mua-lu-gay-ra-tren-dia-ban-huyen-Mu-Cang-Chai-uoc-khoang-920-trieu-dong&PageIndex=12>. [Truy cập 17/4/2025].
- [80] Báo Nông Nghiệp VÀ Môi Trường and Thanh Ngà - Trần Nam, "Mưa lũ ở Mù Cang Chải, Yên Bai: Nhiều thôn bản của xã Hò Bón vẫn chưa thể tiếp cận," 2023. [Trực tuyến]. Địa chỉ: <https://nongnghiepmoitruong.vn/mua-lu-o-mu-cang-chai-yen-bai-nhieu-thon-ban-cua-xa-ho-bon-van-chua-the-tiep-can-i717002.html>. [Truy cập 17/4/2025].
- [81] "Yên Bai: Hiệu quả từ mô hình phòng chống, giảm nhẹ thiên tai tại trường THPT Mù Cang Chải," 2011. [Trực tuyến]. Địa chỉ: <https://baochinhphu.vn/yen-bai-hieuqua-tu-mo-hinh-phong-chong-giam-nhe-thien-tai-tai-truong-thpt-mu-cang-chai-102104536.htm>. [Truy cập 17/4/2025].
- [82] "Ứng dụng khoa học công nghệ trong phòng, chống thiên tai trên địa bàn tỉnh Yên Bai còn nhiều hạn chế," [Trực tuyến]. Địa chỉ: <https://phongchongthientai.mard.gov.vn/Pages/Ung-dung-khoa-hoc-cong-nghe-trong-phong-chong-thie-5821976939.aspx>. [Truy cập 17/4/2025].
- [83] He, Fei, Liu, Suxia, Mo, Xingguo and Wang, Zhonggen, "Interpretable flash flood susceptibility mapping in Yarlung Tsangpo River Basin using H2O Auto-ML," *Scientific Reports*, vol. 15, no. 1, p. , 2025. DOI:10.1038/s41598-024-84655-y.
- [84] Alarifi, Saad S., Abdelkareem, Mohamed, Abdalla, Fathy and Alotaibi, Mislat, "Flash Flood Hazard Mapping Using Remote Sensing and GIS Techniques in Southwestern Saudi Arabia," *Sustainability*, vol. 14, no. 21, p. 14145, 2022. DOI:10.3390/su142114145.
- [85] Petr Sercl, Martin Pecha, Petr Novak, Hana Kyznarova, Ondrej Ledvinka, Vojtech Svoboda and Jan Danhelka, "Flash Flood Indicator," Czech Hydrometeorological

Institute, Prague, 2023. Địa chỉ:
<https://www.chmi.cz/files/portal/docs/reditel/SIS/nakladatelstvi/assets/ffi.pdf>.

- [86] Han J., Kamber M. and Pei J., Data Mining: Concepts and Techniques, Elsevier, 2012. DOI:10.1016/c2009-0-61819-5.
- [87] T. Hastie, R. Tibshirani and J. Friedman, "Model Assessment and Selection," in *The Elements of Statistical Learning*, Springer New York, 2008, pp. 219-259. DOI:10.1007/978-0-387-84858-7_7.
- [88] Pedregosa, Fabian, "Scikit-learn: Machine Learning in Python," 2012.
- [89] Box G. E. P. and Cox D. R., "An Analysis of Transformations," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 26, no. 2, pp. 211-252, 1964. Địa chỉ: <https://www.ime.usp.br/~abe/lista/pdfQWaCMboK68.pdf>.
- [90] In-Kwon Yeo and Richard A. Johnson, "A New Family of Power Transformations to Improve Normality or Symmetry," *Biometrika*, vol. 87, no. 4, pp. 954-959, 2000. Địa chỉ: <https://www.jstor.org/stable/2673623>.
- [91] Bá Thao, Vũ, "Phương pháp xác định khu vực rủi ro lũ bùn đá dựa vào bản đồ địa hình," *Vietnam Journal of Hydrometeorology*, vol. 713, no. 5, pp. 37-46, 2020. DOI:10.36335/vnjhm.2020(713).37-46.
- [92] Christopher M. Bishop, Neural Networks for Pattern Recognition, Birmingham: Oxford, 1995.
- [93] I. D. Moore, R. B. Grayson and A. R. Ladson, "Digital terrain modelling: A review of hydrological, geomorphological, and biological applications," *Hydrological Processes*, vol. 5, no. 1, pp. 3-30, jan/1991. DOI:10.1002/hyp.3360050103.
- [94] NRCS, "Part 630 - Hydrology," 3/2020. [Trực tuyến]. Địa chỉ: <https://directives.nrcc.usda.gov/sites/default/files2/1712930634/Part%20630%20-%20Hydrology.pdf>. [Truy cập 11/12/2023].