

Mẫu B24. BCTT-BNN

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

BÁO CÁO TÓM TẮT KẾT QUẢ  
ĐỀ TÀI TIỀM NĂNG CẤP BỘ

NGHIÊN CỨU ÚNG 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

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

**BÁO CÁO TÓM TẮT KẾT QUẢ  
ĐỀ TÀI TIỀM NĂNG CẤP BỘ**

**NGHIÊN CỨU ÚNG 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

I. Đặt vấn đề .....	1
1. Tính cấp thiết.....	1
2. Mục tiêu của đề tài .....	2
3. Cách tiếp cận .....	2
a. Thu thập Dữ liệu.....	2
b. Chuẩn bị Dữ liệu .....	2
c. Mô hình hóa và Thuật toán .....	2
d. Đầu ra - Phân vùng lũ quét.....	2
II. Vật liệu, nội dung và phương pháp nghiên cứu .....	3
1. Vật liệu nghiên cứu .....	3
2. Nội dung nghiên cứu.....	3
3. Phương pháp nghiên cứu.....	4
a. Phương pháp thu thập dữ liệu .....	4
b. Phương pháp nghiên cứu GIS .....	4
c. Phương pháp học máy .....	4
d. Phương pháp học sâu .....	5
e. Phương pháp tối ưu hóa mô hình .....	5
III. Kết quả và thảo luận.....	5
1. Xây dựng bộ dữ liệu phân phục vụ phân vùng lũ quét .....	5
a. Xây dựng dữ liệu thảm phủ từ ảnh viễn thám.....	5
b. Xây dựng bộ cơ sở dữ liệu không gian khác phục vụ phân vùng lũ quét....	11
2. Xây dựng mô hình trí tuệ nhân tạo phân vùng lũ quét cho khu vực nghiên cứu .....	14
a. Chuẩn hóa dữ liệu đầu vào .....	14
b. Chuẩn bị nhãn dữ liệu dự đoán .....	16
c. Xây dựng mô hình học máy .....	18
d. Xây dựng mô hình học sâu .....	19
e. Kết quả xây dựng mô hình .....	20
3. Xây dựng bản đồ phân vùng lũ quét theo kịch bản mưa.....	23
a. Xây dựng kịch bản mưa .....	23
b. Xây dựng bản đồ phân vùng nguy cơ bằng mô hình CNN.....	25
IV. Kết luận và kiến nghị.....	27
Kết luận .....	27
Những hạn chế còn tồn tại .....	27
Kiến nghị .....	28

## I. Đặt vấn đề

### 1. Tính cấp thiết

Trong bối cảnh biến đổi khí hậu ngày càng phức tạp, lũ quét đã trở thành một trong những thách thức lớn đối với an ninh con người và phát triển bền vững tại Việt Nam, đặc biệt ở các khu vực miền núi như huyện Mù Cang Chải, tỉnh Yên Bái. 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” được xây dựng nhằm đáp ứng nhu cầu phát triển các giải pháp khoa học công nghệ tiên tiến, tận dụng hướng tiếp cận mới trong phân tích rủi ro thiên tai, vốn đang thu hút sự quan tâm của nhiều nhà khoa học trên thế giới.

Lũ quét, với đặc trưng xảy ra đột ngột và sức tàn phá mạnh mẽ, đã gây ra những tổn thất nghiêm trọng tại nhiều địa phương ở Việt Nam. Các sự kiện lũ quét tại Hòa Bình, Yên Bái, Sơn La năm 2017 và Lai Châu, Hà Giang năm 2018 ghi nhận hàng chục trường hợp tử vong, hàng trăm hộ dân mất nơi cư trú, và thiệt hại kinh tế ước tính lên đến hàng nghìn tỷ đồng. Những tổn thất này phản ánh mức độ nghiêm trọng của lũ quét, đặc biệt trong bối cảnh biến đổi khí hậu làm gia tăng tần suất và cường độ của các hiện tượng thời tiết cực đoan như mưa lớn bất thường. Thực trạng này đòi hỏi các giải pháp phân tích rủi ro thiên tai có khả năng cung cấp thông tin chi tiết, đáng tin cậy, và phù hợp với đặc điểm địa phương, từ đó hỗ trợ công tác phòng chống và giảm thiểu thiệt hại.

Hướng tiếp cận mới trong phân vùng rủi ro lũ quét, dựa trên sự kết hợp giữa trí tuệ nhân tạo và dữ liệu địa không gian, đang được nhiều nhà khoa học trên thế giới quan tâm và phát triển. Các thuật toán học máy tiên tiến, như Random Forest, Light Gradient Boosting Machine (LGBM), và mạng nơ-ron tích chập (CNN), đã chứng minh hiệu quả trong việc xử lý các tập dữ liệu lớn và phức tạp, cho phép phân tích các mối quan hệ phi tuyến giữa các yếu tố gây lũ quét, bao gồm độ dốc địa hình, lượng mưa, thảm thực vật, và đặc tính đất.

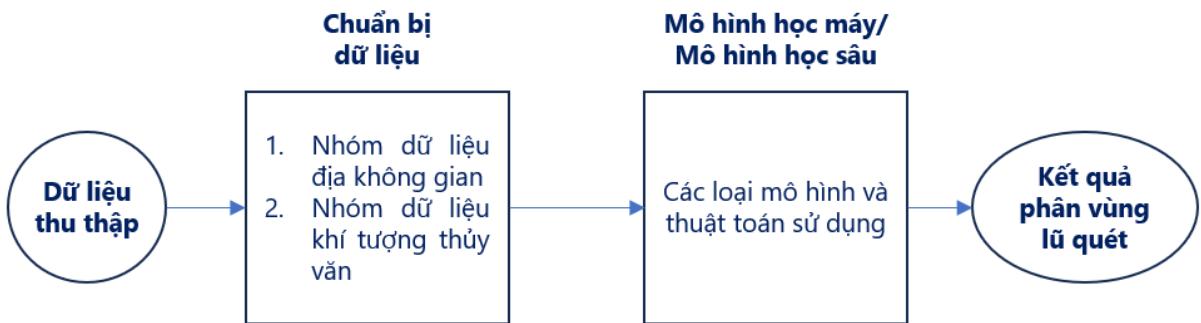
Trí tuệ nhân tạo (AI) và dữ liệu địa không gian (GeoAI) đang cách mạng hóa cách cách tiếp cận và quản lý rủi ro thiên tai, trong đó có lũ quét. Trí tuệ nhân tạo với khả năng học hỏi từ dữ liệu, nhận diện các mẫu phức tạp và đưa ra dự đoán chính xác, đã trở thành công cụ đắc lực trong việc phân tích dữ liệu phi tuyến và đa biến liên quan đến lũ quét. Các thuật toán học máy có thể xử lý hiệu quả lượng lớn dữ liệu từ nhiều nguồn khác nhau, bao gồm hình ảnh vệ tinh, dữ liệu khí tượng và thông tin môi trường, để xác định các yếu tố ảnh hưởng đến lũ quét và dự báo các khu vực có nguy cơ cao.

Cách tiếp cận này cho phép tích hợp nhiều yếu tố đầu vào, từ địa hình, thảm phủ đến khí tượng, để tạo ra các mô hình dự báo có độ tin cậy cao và khả năng áp dụng linh hoạt ở các quy mô địa lý khác nhau, đặc biệt không chỉ sử dụng dữ liệu đầu vào một cách độc lập, nghiên cứu này sử dụng các dữ liệu thủy văn lưu vực làm yếu tố đầu vào nhằm mô tả/diễn đạt nguyên nhân sinh lũ quét. Trên cơ sở đó, nghiên cứu này sẽ đánh giá khả năng ứng dụng công nghệ trí tuệ nhân tạo và dữ liệu địa không gian để nâng cao độ tin cậy trong phân vùng lũ quét.

## 2. Mục tiêu của đề tài

Đánh giá khả năng ứng dụng công nghệ trí tuệ nhân tạo và dữ liệu địa không gian để nâng cao độ tin cậy trong phân vùng lũ quét quy mô cấp huyện.

## 3. Cách tiếp cận



Hình I - 1. Sơ đồ tiếp cận trong nghiên cứu

### a. Thu thập Dữ liệu

Giai đoạn đầu tiên tập trung vào việc thu thập dữ liệu thô từ nhiều nguồn khác nhau. Đây là bước quan trọng nhất vì chất lượng của toàn bộ quá trình phân tích phụ thuộc vào độ chính xác và tính đầy đủ của dữ liệu đầu vào. Dữ liệu có thể được thu thập từ các nguồn như hình ảnh vệ tinh, dữ liệu GPS, khảo sát thực địa, hoặc các cơ sở dữ liệu địa lý có sẵn.

### b. Chuẩn bị Dữ liệu

Sau khi thu thập, dữ liệu thô cần được xử lý qua hai bước chính:

**Bước 1: Nhóm dữ liệu địa không gian** - Quá trình này bao gồm việc tổ chức và phân loại dữ liệu theo các tiêu chí địa lý như vị trí, tọa độ, hoặc các đặc điểm không gian cụ thể. Mục đích là tạo ra các nhóm dữ liệu có tính chất tương đồng về mặt không gian.

**Bước 2: Nhóm dữ liệu khí tượng thủy văn** - Đây là bước chuyên biệt hóa, tập trung vào việc xử lý các dữ liệu liên quan đến điều kiện khí hậu, thời tiết và tài nguyên nước. Các thông số về lượng mưa thời đoạn được tổ chức và chuẩn hóa để đưa vào mô hình.

### c. Mô hình hóa và Thuật toán

Trong giai đoạn này, các loại mô hình và thuật toán sử dụng được áp dụng lên dữ liệu đã được chuẩn bị. Điều này có thể bao gồm các phương pháp machine learning, deep learning kết hợp các thuật toán chuyên biệt cho phân tích không gian. Việc lựa chọn mô hình phù hợp phụ thuộc vào mục tiêu cụ thể của nghiên cứu và tính chất của dữ liệu.

### d. Đầu ra - Phân vùng lũ quét

Kết quả cuối cùng của quy trình là việc tạo ra bản đồ phân vùng lũ quét. Đây là sản phẩm có giá trị thực tiễn cao, giúp xác định các khu vực có nguy cơ cao, trung bình và thấp về lũ quét.

## **II. Vật liệu, nội dung và phương pháp nghiên cứu**

### **1. Vật liệu nghiên cứu**

**Đối tượng nghiên cứu:** Đối tượng nghiên cứu là lũ quét do mưa lớn, không xét đến các yếu tố sạt lở/nghẽn dòng.

**Địa điểm nghiên cứu:** Huyện Mù Cang Chải, tỉnh Yên Bai

**Thời gian nghiên cứu:** 01/2023-06/2025

### **2. Nội dung nghiên cứu**

**Nội dung 1: Thu thập, điều tra, và đánh giá tình hình thiên tai lũ quét tại khu vực nghiên cứu (huyện Mù Cang Chải, tỉnh Yên Bai).**

- Công việc 1.1: Phân tích, lựa chọn và đánh giá hiện trạng khu vực nghiên cứu.
- Công việc 1.2: Điều tra thực địa kết hợp số liệu lịch sử để xây dựng bộ cơ sở dữ liệu phục vụ xây dựng mô hình phân vùng rủi ro lũ quét. (Tổ điều tra gồm 5 người thực hiện trong vòng 10 ngày).
- Công việc 1.3: Lập báo cáo và đánh giá kết quả điều tra thực địa tại khu vực nghiên cứu.

**Nội dung 2: Nghiên cứu và xây dựng bộ dữ liệu phục vụ mô hình phân vùng rủi ro lũ quét từ dữ liệu địa không gian.**

- Công việc 2.1: Nghiên cứu về cấu trúc các loại dữ liệu viễn thám được sử dụng trong mô hình phân rủi ro vùng lũ quét.
- Công việc 2.2: Nghiên cứu xây dựng dữ liệu thảm phủ, ứng dụng trí tuệ nhân tạo, dữ liệu viễn thám, và số liệu thực địa.
- Công việc 2.3: Nghiên cứu, thu thập và phân tích để xây dựng các bộ dữ liệu khác phục vụ xây dựng mô hình phân vùng rủi ro lũ quét.

**Nội dung 3: Đánh giá các phương pháp tiềm năng ứng dụng trí tuệ nhân tạo trong phân vùng rủi ro lũ quét.**

- Công việc 3.1: Phân tích tổng hợp các phương pháp tiềm năng ứng dụng trí tuệ nhân tạo trong phân vùng rủi ro lũ quét.
- Công việc 3.2: Nghiên cứu và đánh giá nhằm đề xuất phương pháp tối ưu ứng dụng trí tuệ nhân tạo cho phân vùng rủi ro lũ quét ở khu vực nghiên cứu.

**Nội dung 4: Đề xuất quy trình xây dựng phương pháp tối ưu ứng dụng trí tuệ nhân tạo và dữ liệu địa không gian để phân vùng rủi ro lũ quét cho khu vực nghiên cứu (huyện Mù Cang Chải, tỉnh Yên Bai).**

- Công việc 4.1: Đánh giá tính hiệu quả của phương pháp lựa chọn ứng dụng phân vùng rủi ro lũ quét cho vùng nghiên cứu.
- Công việc 4.2: Xây dựng phương pháp tối ưu ứng dụng trí tuệ nhân tạo và dữ liệu viễn thám để phân vùng rủi ro lũ quét cho khu vực nghiên cứu.
- Công việc 4.3: Nghiên cứu phương pháp tích hợp kết quả bản đồ phân vùng rủi ro lũ quét và bản đồ hành chính.

## **Nội dung 5: Báo cáo tổng kết, đánh giá nghiệm thu kết quả thực hiện đề tài**

### **3. Phương pháp nghiên cứu**

Nghiên cứu ứng dụng trí tuệ nhân tạo (AI) kết hợp dữ liệu địa không gian để phân vùng lũ quét tại khu vực thí điểm Mù Cang Chải, sử dụng các phương pháp học máy và học sâu, cùng với phân tích GIS để đánh giá nguy cơ lũ quét. Sơ đồ tiếp cận bao gồm bốn khía cạnh chính: thu thập dữ liệu, nghiên cứu GIS, học máy và học sâu.

#### **a. Phương pháp thu thập dữ liệu**

Dữ liệu được thu thập từ năm nhóm chính: (1) địa hình, (2) sử dụng đất – thảm phủ, (3) loại đất, (4) dữ liệu vệ tinh quan trắc bề mặt, và (5) dữ liệu mưa. Dữ liệu địa hình sử dụng DEM từ ảnh ALOS (độ phân giải 12,5m, năm 2006-2011) để tạo các sản phẩm như độ dốc, độ cong địa hình, diện tích lưu vực, và chỉ số ám địa hình (TWI). Dữ liệu đất bao gồm tốc độ thẩm bình quân và nhóm đất thủy văn (A, B, C, D) từ USDA và ORNL DAAC, phục vụ tính toán chỉ số CN (Curve Number). Dữ liệu vệ tinh (Sentinel, Landsat, Modis) cung cấp các chỉ số như NDVI, MNDVI, và CN. Dữ liệu mưa bao gồm lượng mưa giờ, 3 giờ, 6 giờ, và 24 giờ lớn nhất, thu thập từ trạm quan trắc (2021-2024) và vệ tinh. Dữ liệu lũ quét lịch sử được thu thập từ báo cáo thực địa và tài liệu, dù còn thiếu sót do các sự kiện ở khu vực hẻo lánh không được ghi nhận đầy đủ.

#### **b. Phương pháp nghiên cứu GIS**

Phương pháp GIS được sử dụng để xử lý dữ liệu địa không gian, tạo bản đồ các đặc trưng liên quan đến lũ quét. Từ DEM, các đặc trưng địa hình như độ dốc, độ dốc lòng dẫn, chiều dài dòng chảy, diện tích lưu vực, khoảng cách và độ cao tương đối đến sông/suối, và TWI được tính toán bằng các công cụ trong ArcGIS hoặc QGIS (Slope, Flow Length, Watershed, Euclidean Distance). Dữ liệu vệ tinh cung cấp chỉ số NDVI và CN, trong khi dữ liệu đất được phân tích để xác định tốc độ thẩm và nhóm đất thủy văn. Các đặc trưng được chuẩn hóa để đảm bảo tính tương thích và tăng độ chính xác cho mô hình AI.

#### **c. Phương pháp học máy**

Nghiên cứu sử dụng bốn mô hình học máy để phân loại nguy cơ lũ quét: Rừng ngẫu nhiên (Random Forest - RF), Máy hỗ trợ vectơ (Support Vector Machine - SVM), Hồi quy Logistic (Logistic Regression - LR), và LightGBM (LGBM). Các đặc trưng đầu vào được chia thành bốn nhóm:

- **Địa hình:** độ dốc, TWI, độ cong địa hình.
- **Thủy văn:** khoảng cách/chênh lệch độ cao đến sông/suối, độ dốc lòng dẫn, diện tích lưu vực.
- **Thực phủ:** NDVI, CN.
- **Khí tượng:** lượng mưa giờ, 3 giờ, 6 giờ, và 24 giờ lớn nhất.

Dữ liệu được làm sạch để loại bỏ nhiễu, chuẩn hóa bằng phương pháp Standardization (Z-score Normalization) do dữ liệu như lượng mưa và TWI có phân bố lệch. RF và

LGBM xử lý tốt dữ liệu định tính mà không cần mã hóa phức tạp, trong khi SVM và LR yêu cầu chuẩn hóa chặt chẽ. RF dự đoán dựa trên kết hợp nhiều cây quyết định, giảm sai số thông qua lấy mẫu ngẫu nhiên. SVM tạo ranh giới phân chia trong không gian dữ liệu, phù hợp với dữ liệu chuẩn hóa tốt. LR gán trọng số cho đặc trưng, đơn giản nhưng kém hiệu quả với mối quan hệ phi tuyến. LGBM xây dựng cây quyết định tuần tự, tối ưu với dữ liệu lớn và mối quan hệ phi tuyến.

#### **d. Phương pháp học sâu**

Hai mô hình học sâu được sử dụng: Mạng nơ-ron sâu (DNN) và Mạng nơ-ron tích chập (CNN). DNN xử lý đồng thời nhiều đặc trưng, tự động học mối quan hệ phức tạp. CNN phân tích dữ liệu không gian, xem các đặc trưng như “bản đồ” để tìm mẫu lũ quét. Các mô hình này sử dụng dữ liệu tương tự học máy nhưng linh hoạt hơn, có thể xử lý cả đặc trưng không trực tiếp liên quan (như cao độ).

#### **e. Phương pháp tối ưu hóa mô hình**

- **Tinh chỉnh siêu tham số:** Tìm giá trị tối ưu cho tỷ lệ học, số lớp ẩn, số cây (n\_estimators), độ sâu cây (max\_depth), và batch size.
- **Tăng cường dữ liệu:** Tạo vùng đệm quanh dòng chảy chính và gán nhãn guy cơ (4: cao, 3: trung bình, 2: thấp, 1: rất thấp, 0: không lũ) dựa trên độ cao tương đối và lượng mưa.
- **Tối ưu hóa mạng nơ-ron:** Điều chỉnh cấu trúc mạng (số lớp, kích thước lớp) và thuật toán huấn luyện (Adam, RMSprop) để tăng hiệu suất và giảm quá khứ.

### **III. Kết quả và thảo luận**

#### **1. Xây dựng bộ dữ liệu phân phục vụ phân vùng lũ quét**

##### **a. Xây dựng dữ liệu thảm phủ từ ảnh viễn thám**

Dữ liệu thảm phủ là dữ liệu quan trọng mô tả bề mặt và có tác động đến dòng chảy trong thủy văn. Một số phân loại thảm phủ chính thường được sử dụng như của Esri và Jaxa do các đơn vị này liên tục phân loại thảm phủ hàng năm.

Hệ thống phân loại của Esri bao gồm 9 lớp chính như Mặt nước, Rừng cây, Thực vật ngập nước, Cây trồng, Khu vực xây dựng, Đất trống, Tuyết/Băng, Mây che phủ, và Đồng cỏ/Thảo nguyên. Hệ thống này có xu hướng tập trung vào đặc điểm tự nhiên và nhân tạo, với sự nhấn mạnh vào các khu vực ngập nước và thực vật (như rừng ngập mặn, ruộng lúa), phù hợp với phân tích đa dạng sinh học và quản lý tài nguyên ở các khu vực nhiệt đới như Việt Nam.

Hệ thống của JAXA bao gồm 12 lớp như Nước, Khu vực đô thị, Ruộng lúa, Cây trồng khác, Cỏ/cây bụi, Cây gỗ/vườn cây, Đất trống, Rừng thường xanh, Rừng rụng lá, Rừng trảng, Rừng ngập mặn, và Nuôi trồng thủy sản. Hệ thống này chi tiết hơn với sự phân biệt rõ ràng giữa các loại rừng (thường xanh, rụng lá, trồng) và bổ sung lớp Nuôi trồng thủy sản, phản ánh sự quan tâm đến các hoạt động kinh tế ven biển.

Mặc dù các hệ thống phân loại này là hiện hữu, việc phân loại lại thảm phủ cho một khu vực cụ thể là cần thiết khi mục tiêu chính là tối ưu hóa phân tích thủy văn, đặc biệt trong bối cảnh quản lý tài nguyên nước và dự đoán dòng chảy. Thảm phủ bề mặt ảnh hưởng trực tiếp đến các quá trình thủy văn như thấm nước, dòng chảy bề mặt, và bốc hơi, do đó một hệ thống phân loại không phù hợp hoặc thiếu chi tiết có thể dẫn đến sai lệch trong các mô hình thủy văn. Tại một khu vực cụ thể, các đặc điểm tự nhiên và nhân tạo như rừng, đồng cỏ, đất canh tác, hoặc khu vực xây dựng có thể thay đổi theo thời gian do tác động của con người hoặc biến đổi khí hậu, khiến các phân loại có sẵn như của Esri hoặc JAXA không còn phản ánh chính xác thực tế địa phương.

Mục tiêu của việc phân loại lại thảm phủ để phù hợp với thủy văn là xây dựng một cơ sở dữ liệu chi tiết, phản ánh chính xác điều kiện thực địa, từ đó cải thiện độ tin cậy của các mô hình tính toán dòng chảy như phương pháp hệ số đường cong (curve number) của NRCS. Một phân loại mới có thể tập trung vào các yếu tố như mật độ thực vật, mức độ thấm nước của đất, và cách quản lý đất (chẳng hạn canh tác theo đường đồng mức hay chăn thả), vốn ảnh hưởng lớn đến lượng nước thấm vào lòng đất hoặc chảy tràn ra sông ngòi. Ngoài ra, việc này cho phép đánh giá tác động của các thay đổi sử dụng đất, như chuyển từ rừng sang khu vực đô thị hóa, lên tài nguyên nước, đặc biệt trong mùa mưa hoặc lũ lụt. Tại khu vực cụ thể, phân loại lại cũng giúp xác định các vùng dễ bị ngập úng hoặc khô hạn, từ đó hỗ trợ lập kế hoạch thủy lợi và bảo vệ môi trường một cách hiệu quả hơn. Vì vậy, quá trình này không chỉ là cập nhật dữ liệu mà còn là bước quan trọng để đảm bảo các quyết định quản lý dựa trên cơ sở khoa học, phù hợp với nhu cầu thực tiễn của khu vực nghiên cứu.

Nghiên cứu này tập trung vào lũ quét, điều này đồng nghĩa với việc đi tìm những đặc điểm về sử dụng đất có liên quan đến yếu tố dòng chảy. Nghiên cứu lũ quét đòi hỏi một cách tiếp cận chi tiết về thảm phủ để đánh giá chính xác tiềm năng dòng chảy và ngập úng, đặc biệt trong các khu vực dễ bị tổn thương. Hệ thống phân loại thảm phủ của NRCS, được trình bày trong Chương 8 và Chương 9 của National Engineering Handbook (NEH) [92], cung cấp một khuôn khổ khoa học để phân tích các yếu tố thủy văn như thấm nước, dòng chảy bề mặt, và khả năng giữ nước. Theo NRCS, thảm phủ được chia thành các lớp sử dụng đất và xử lý đất (land use and treatment classes), bao gồm đất canh tác, đồng cỏ, rừng, và khu vực đô thị, với các điều kiện thủy văn được đánh giá qua các cấp độ như tốt, trung bình, và kém. Đối với lũ quét, việc áp dụng phân loại này giúp xác định các khu vực có hệ số đường cong runoff (CN) cao, vốn phản ánh khả năng dòng chảy nhanh và mạnh, đặc biệt trên đất trống, đất canh tác không che phủ, hoặc khu vực đô thị hóa với nhiều bề mặt không thấm nước.

Phân loại theo NRCS cũng nhấn mạnh vai trò của các biện pháp xử lý đất như trồng cây theo đường đồng mức (contouring) hoặc làm bậc thang (terracing), vốn làm chậm dòng chảy và tăng khả năng thấm, từ đó giảm rủi ro lũ quét. Trong khu vực nghiên cứu, việc kết hợp dữ liệu thực địa về mật độ thực vật, lượng rác hữu cơ, và mức độ chăn thả với bảng CN (như Bảng 9-1)[92] cho phép xây dựng mô hình thủy văn chính xác hơn.

Đặc biệt, đối với các khu vực rừng hoặc đồng cỏ, NRCS phân loại theo điều kiện thủy văn (poor, fair, good) dựa trên tỷ lệ che phủ và quản lý, giúp đánh giá tác động của quản lý đất đến dòng chảy cực đoan. Ngoài ra, việc áp dụng phương trình tính CN tổng hợp (equation 9-1 hoặc 9-2) từ Chương 9 [92] cho các khu vực đô thị hoặc hỗn hợp sử dụng đất giúp điều chỉnh rủi ro lũ quét khi có sự thay đổi trong tỷ lệ bờ mặt không thấm. Vì vậy, việc sử dụng hệ thống NRCS không chỉ hỗ trợ phân tích hiện trạng mà còn định hướng chiến lược quản lý thảm phủ để giảm thiểu tác động của lũ quét, đặc biệt trong các khu vực có địa hình dốc hoặc mưa lớn như ở Việt Nam.

Nghiên cứu này thể hiện một nỗ lực hệ thống hóa các đặc điểm tự nhiên và nhân tạo của khu vực Mù Cang Chải nhằm phục vụ mục tiêu thủy văn, đặc biệt trong quản lý dòng chảy và lũ quét. Chi tiết được thể hiện trong bảng sau đây:

Bảng III - 1. Nhóm phân loại thảm phủ sử dụng trong nghiên cứu

Lớp	Mô tả	Điều kiện thủy văn	Ảnh hưởng đến dòng chảy
1	Mặt nước (Hồ, sông, đầm lầy)	Bè mặt nước	Tạo dòng chảy trực tiếp, không thấm thấu
2	Rừng rậm (Rừng dày, điều kiện Good)	Good	Tăng thấm thấu, giảm dòng chảy bờ mặt
3	Đất trống (Fallow hoặc đất không có thực vật)	Poor	Tăng dòng chảy bờ mặt, giảm thấm thấu
4	Khu xây dựng (Bờ mặt không thấm nước)	Impervious	Tạo dòng chảy nhanh, không thấm thấu
5	Rừng thưa (Rừng thưa, điều kiện Fair/Poor)	Fair/Poor	Thấm thấu trung bình, dòng chảy trung bình
6	Ruộng bậc thang (Đất canh tác với hệ thống terracing)	Contoured/terraced	Giảm dòng chảy nhờ lưu trữ nước trong ruộng
7	Đất trồng lúa (Đất canh tác ngập nước)	Wet meadow	Giữ nước lâu dài, dòng chảy chậm

### Chuẩn bị dữ liệu

Nghiên cứu lựa chọn 16 chỉ số đầu vào (NDVI, EVI, GNDVI, GRVI, NDWI1, NDWI2, GSI, NDBI, NDWI, BSI, Slope, Red, Green, Blue, VV, VH) để phân loại thảm phủ theo 7 nhóm (Mặt nước, Rừng rậm, Đất trống, Khu xây dựng, Rừng thưa, Ruộng bậc thang, Đất trồng lúa) nhằm tận dụng tối đa thông tin từ dữ liệu đa phổ và radar để phân biệt các đặc trưng quang phổ, cấu trúc, và địa hình của từng loại thảm phủ.

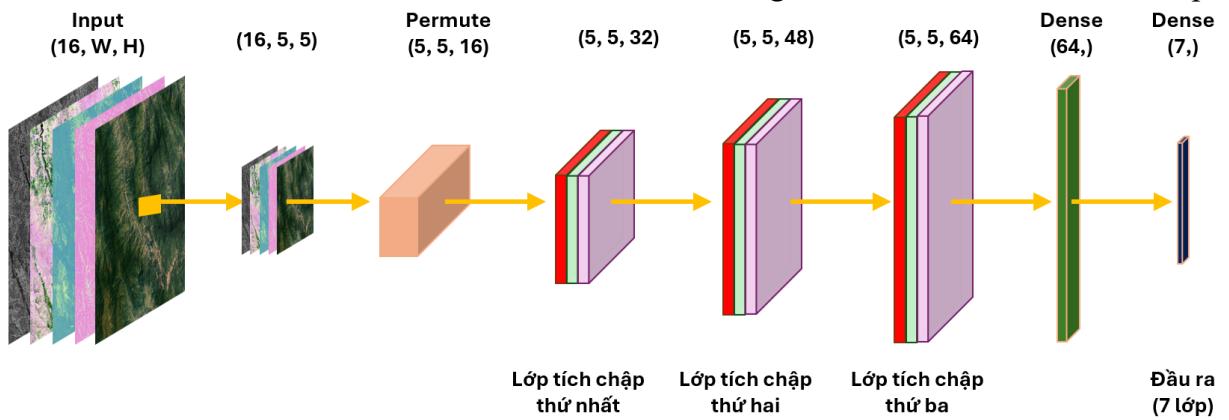
Bảng III - 2. Mô tả các tham số đầu vào trong phân loại thảm phủ

T T	Tham số	Mô tả	Ý nghĩa	Lớp tác động chính
1	NDVI	Chỉ số chênh lệch thực vật chuẩn hóa (NIR - Red)/(NIR + Red)	Đánh giá mật độ và sức khỏe thực vật dựa trên phản xạ hồng ngoại gần (NIR) và đỏ (Red), rất hữu ích trong việc phân biệt Rừng rậm và Rừng thưa dựa trên sự khác biệt về độ che phủ và sinh khối.	Rừng rậm (2), Rừng thưa (5)
2	EVI	Chỉ số thực vật cải tiến ( $2.5^*(NIR - Red)/(NIR + 6^*Red - 7.5^*Blue + 1)$ )	Cải thiện NDVI, giảm nhiễu khí quyển và đất nền, đặc biệt hiệu quả trong việc xác định các khu vực có thực vật dày đặc so với khu vực thưa thớt.	Rừng rậm (2), Rừng thưa (5)

T T	Tham số	Mô tả	Ý nghĩa	Lớp tác động chính
3	GNDV I	Chỉ số chênh lệch thực vật xanh (NIR - Green)/(NIR + Green)	Đánh giá sức khỏe thực vật qua phản xạ kênh xanh, hỗ trợ phân loại Đất trồng lúa nhờ khả năng phát hiện sự thay đổi quang phổ trong môi trường ngập nước.	Rừng rậm (2), Đất trồng lúa (7)
4	GRVI	Chỉ số thực vật xanh - đỏ (Green - Red)/(Green + Red)	Phân biệt thực vật qua sự chênh lệch giữa kênh xanh và đỏ, hỗ trợ phát hiện Đất trồng nhờ phản xạ đặc trưng của đất không che phủ.	Rừng thưa (5), Đất trồng (3)
5	NDWI 1	Chỉ số nước (Green - NIR)/(Green + NIR)	Phát hiện nước qua sự hấp thụ mạnh ở dài NIR, giúp nhận diện Mặt nước và Đất trồng lúa nhờ đặc tính quang phổ của nước bề mặt.	Mặt nước (1), Đất trồng lúa (7)
6	NDWI 2	Chỉ số nước cải tiến (Green - SWIR)/(Green + SWIR)	Tăng cường phát hiện nước qua dài hòng ngoại ngắn (SWIR), giảm nhiễu từ thực vật, hỗ trợ phân biệt giữa mặt nước tự nhiên và đất ngập nước canh tác.	Mặt nước (1), Đất trồng lúa (7)
7	GSI	Chỉ số xanh-SWIR (Green - SWIR)/(Green + SWIR)	Đánh giá độ ẩm đất và nước trên bề mặt, hỗ trợ phân loại Đất trồng lúa và Đất trồng nhờ khả năng phát hiện sự thay đổi độ ẩm.	Đất trồng lúa (7), Đất trồng (3)
8	NDBI	Chỉ số xây dựng (SWIR - NIR)/(SWIR + NIR)	Phát hiện khu vực xây dựng qua phản xạ SWIR cao, nổi bật trong việc nhận diện Khu xây dựng nhờ đặc trưng bề mặt không thấm nước.	Khu xây dựng (4)
9	NDWI	Chỉ số nước tổng quát (NIR - SWIR)/(NIR + SWIR)	Nhận diện nước và vùng ngập qua sự chênh lệch NIR-SWIR, ứng dụng rộng rãi trong phân tích thủy văn cho Mặt nước và Đất trồng lúa.	Mặt nước (1), Đất trồng lúa (7)
10	BSI	Chỉ số đất tràn ((SWIR + Red) - (NIR + Blue))/((SWIR + Red) + (NIR + Blue))	Phát hiện đất tràn qua phản xạ đặc trưng của đất ở dài SWIR, hỗ trợ phân loại Đất trồng nhờ khả năng nhận diện đất không che phủ.	Đất trồng (3)
11	Slope	Độ dốc địa hình (tính bằng độ hoặc phần trăm)	Đánh giá độ nghiêng của bề mặt, ảnh hưởng trực tiếp đến dòng chảy và khả năng thấm nước, quan trọng trong phân loại Ruộng bậc thang và Rừng rậm.	Ruộng bậc thang (6), Rừng rậm (2)
12	Red	Dải đỏ từ dữ liệu đa phổ	Phản xạ ánh sáng đỏ, hỗ trợ phân biệt bề mặt như Đất trồng và Khu xây dựng, cung cấp cơ sở cho các chỉ số tổng hợp khác.	Đất trồng (3), Khu xây dựng (4)
13	Green	Dải xanh từ dữ liệu đa phổ	Phản xạ ánh sáng xanh, hỗ trợ phát hiện nước và thực vật, đặc biệt hiệu quả trong nhận diện Mặt nước và Đất trồng lúa.	Mặt nước (1), Đất trồng lúa (7)
14	Blue	Dải xanh lam từ dữ liệu đa phổ	Phản xạ ánh sáng xanh lam, hỗ trợ phân biệt đất và nước, giảm nhiễu khí quyển trong phân loại Đất trồng và Mặt nước.	Đất trồng (3), Mặt nước (1)
15	VV	Độ phân cực thẳng đứng (Vertical-Vertical) từ radar Sentinel-1	Phản ánh độ nhám bề mặt và cấu trúc thực vật, hiệu quả trong phân loại Ruộng bậc thang nhờ tín hiệu tán xạ đặc trưng.	Ruộng bậc thang (6), Đất trồng (3)
16	VH	Độ phân cực chéo (Vertical-Horizontal) từ radar Sentinel-1	Phân tích cấu trúc thực vật và bề mặt phức tạp, nhạy với sinh khối và hỗ trợ phân loại Rừng rậm và Ruộng bậc thang.	Rừng rậm (2), Ruộng bậc thang (6)

## 2. Mô hình CNN

Mô hình CNN là mô hình học sâu được thiết kế để tận dụng mối quan hệ không gian giữa các pixel lân cận, một yếu tố quan trọng trong phân loại LULC do sự khác biệt về cấu trúc bê mặt giữa các lớp.



Hình III - 1. Cấu trúc mạng CNN

Tập dữ liệu bao gồm 100,000 mẫu (trên tổng số 122,393 mẫu dựa trên thuật toán không ché số lượng mẫu tối đa) được chia thành tập huấn luyện (80%) và tập xác thực (20%) bằng phương pháp phân chia ngẫu nhiên có phân tầng để đảm bảo phân bố nhẵn đồng đều. Toàn bộ mẫu được lấy ở dạng vùng và chuyển về dạng điểm thông qua công cụ rasterize trong QGIS. Mô hình được huấn luyện trong 50 epoch với kích thước batch là 32. Các callback được sử dụng bao gồm:

**EarlyStopping:** Dừng huấn luyện nếu hàm mất mát xác thực không cải thiện sau 5 epoch, đồng thời khôi phục trọng số tốt nhất.

**ReduceLROnPlateau:** Giảm tốc độ học xuống 50% nếu hàm mất mát xác thực không cải thiện sau 3 epoch, với tốc độ học tối thiểu là 1e-6.

Bảng III - 3. Phân bổ nhãn trong tập dữ liệu huấn luyện

Lớp	Tên	Số mẫu	Tỷ lệ
1	Mặt nước	4,929	4.93%
2	Rừng rậm	24,223	24.22%
3	Đất trống	17,784	17.78%
4	Khu xây dựng	25,505	25.51%
5	Rừng thưa	4,986	4.99%
6	Ruộng bậc thang	17,012	17.01%
7	Đất trồng lúa	5,561	5.56%

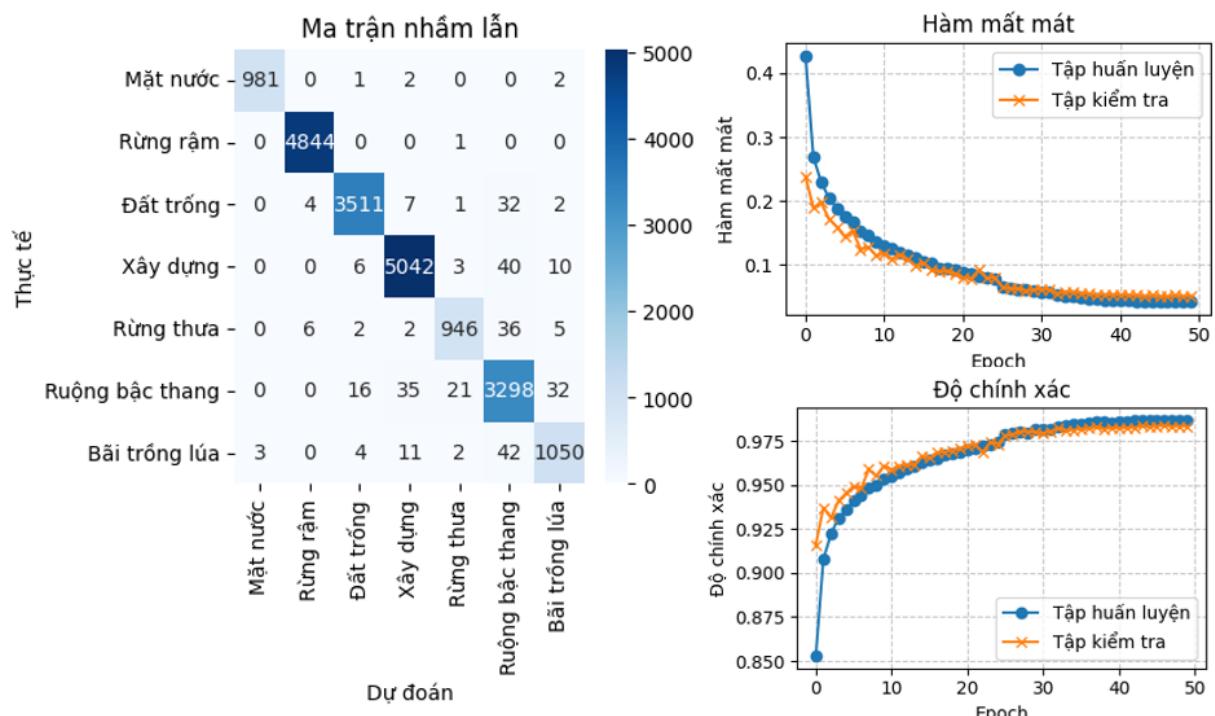
Tập dữ liệu huấn luyện thể hiện sự mất cân bằng đáng kể giữa các lớp, với các lớp như "Khu xây dựng" (25.51%) và "Rừng rậm" (24.22%) chiếm tỷ lệ lớn, trong khi các lớp như "Mặt nước" (4.93%) và "Rừng thưa" (4.99%) có số mẫu ít hơn. Tuy nhiên, mô hình CNN vẫn đạt độ chính xác cao (98.32%) nhờ vào các chiến lược thiết kế hiệu quả. Thứ nhất, việc sử dụng phân chia ngẫu nhiên có phân tầng (stratified random

splitting) đảm bảo rằng tỷ lệ các lớp được duy trì trong cả tập huấn luyện và xác thực, giúp mô hình học được các đặc trưng của các lớp thiểu số. Thứ hai, các lớp tích chập giãn nở và kernel lớn cho phép mô hình nắm bắt các mẫu không gian phức tạp, đặc biệt là ở các lớp hiếm như "Mặt nước", vốn có đặc trưng quang học rõ rệt (NDWI, NDWI1). Ngoài ra, hàm mất mát sparse categorical crossentropy và kỹ thuật dropout (0.2) giúp giảm thiểu tình trạng quá khớp đối với các lớp chiếm ưu thế, đảm bảo mô hình tổng quát hóa tốt trên tất cả các lớp. Những yếu tố này đã giảm thiểu tác động của dữ liệu mất cân bằng, dẫn đến hiệu suất phân loại đồng đều, như được thể hiện trong ma trận nhầm lẫn.

### 3. Kết quả xây dựng bản đồ thảm phủ

#### a. Hiệu suất mô hình

Mô hình CNN đạt độ chính xác huấn luyện là 98.68% và độ chính xác xác thực là 98.32% sau 50 epoch. Quá trình hội tụ được thể hiện trong Hình 3, cho thấy hàm mất mát giảm đều và ổn định, đặc biệt sau khi tốc độ học được điều chỉnh tự động. Ma trận nhầm lẫn (Hình 4) cho thấy mô hình phân loại chính xác hầu hết các lớp, với một số nhầm lẫn nhỏ giữa rừng rậm (lớp 2) và rừng thưa (lớp 5) do sự tương đồng về mật độ thực vật. Độ chính xác và độ nhạy của từng lớp được trình bày trong Bảng III - 4.



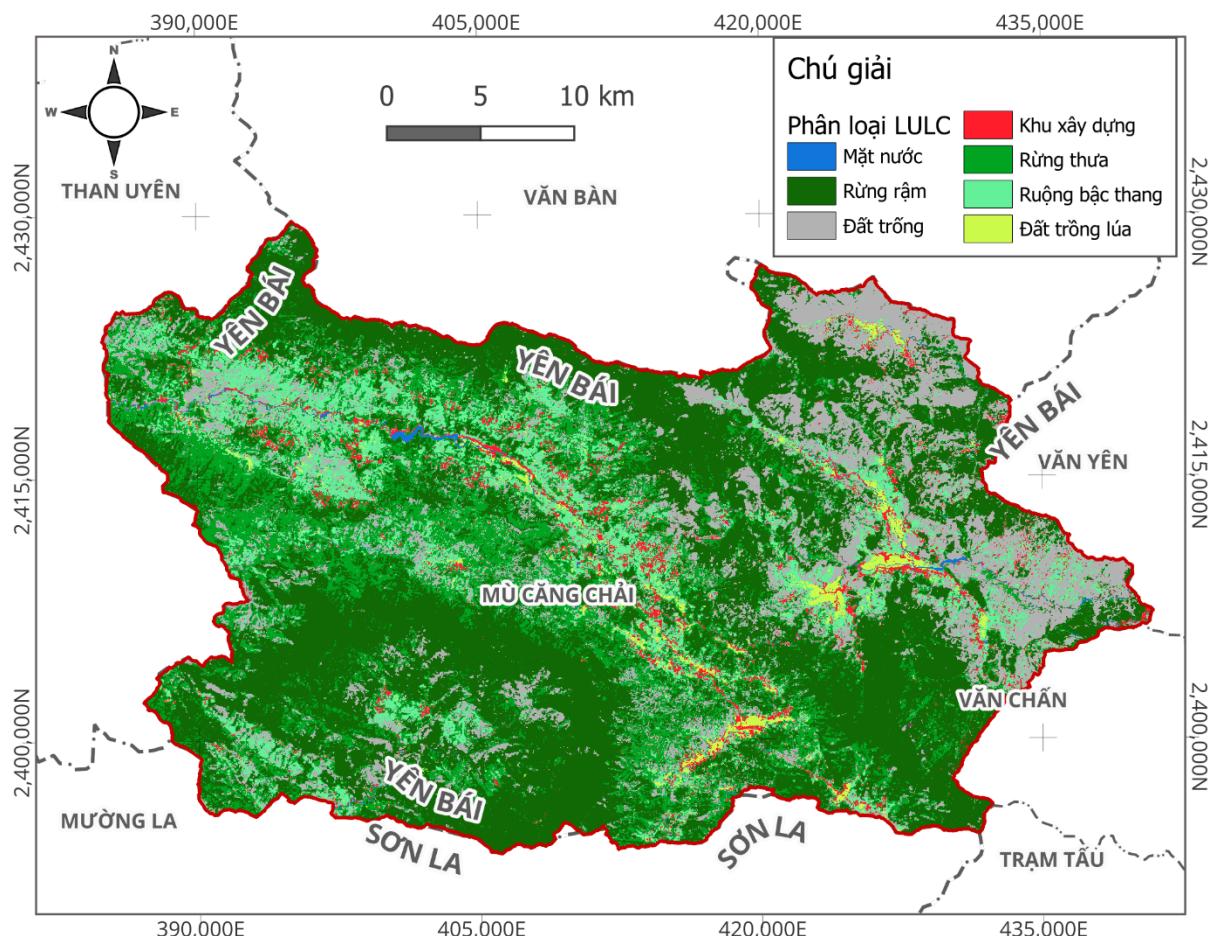
Hình III - 2. Ma trận nhầm lẫn của mô hình CNN trên tập kiểm tra và đồ thị hàm mất mát, độ chính xác trong quá trình huấn luyện, kiểm tra mô hình

Bảng III - 4. Độ chính xác và độ nhạy của từng lớp phân loại thảm phủ

Lớp	Tên	Độ chính xác (%)	Độ nhạy (%)	Phương pháp nghiên cứu	ESRI Land Cover	JAXA ALOS
1	Mặt nước	99.7	99.49	Có	Có	Có
2	Rừng rậm	99.79	99.98	Có (Good)	Có (Trees)	Có (chưa phân loại)

Lớp	Tên	Độ chính xác (%)	Độ nhạy (%)	Phương pháp nghiên cứu	ESRI Land Cover	JAXA ALOS
3	Đất trống	99.18	98.71	Có (Poor)	Có (Bare ground)	Có (Barren)
4	Khu xây dựng	98.88	98.84	Có (Impervious)	Có (Built-up)	Có (Urban)
5	Rừng thưa	97.13	94.88	Có (Fair/Poor)	Không	Có (chưa phân loại)
6	Ruộng bậc thang	95.65	96.94	Có (Contoured)	Không	Không
7	Đất trồng lúa	95.37	94.42	Có (Wet meadow)	Không	Có (rice)

b. Bản đồ thẩm phu



Hình III - 3. Bản đồ LULC được tạo ra từ mô hình CNN

Bằng cách sử dụng 16 tham số đầu vào từ Sentinel-1, Sentinel-2 và ALOS, mô hình CNN đã phân loại thành công bảy lớp LULC, mỗi lớp có mối liên hệ rõ ràng với các quá trình thủy văn như thẩm thấu và dòng chảy. So với các sản phẩm LULC hiện có như ESRI Land Cover và JAXA ALOS, phương pháp này cung cấp các lớp phân loại chuyên biệt hơn, đặc biệt là ruộng bậc thang, đất trồng lúa và chất lượng rừng ( thông qua CSDL rừng), vốn quan trọng trong mô phỏng dòng chảy ở các khu vực nông nghiệp miền núi phía Bắc Việt Nam.

b. Xây dựng bộ cơ sở dữ liệu không gian khác phục vụ phân vùng lũ quét

Trong bối cảnh dự đoán nguy cơ lũ quét, việc lựa chọn và hiểu rõ các đặc trưng dữ liệu là yếu tố cốt lõi để xây dựng các mô hình học máy hiệu quả. Các đặc trưng được sử dụng trong nghiên cứu này thuộc bốn nhóm chính: địa hình, thủy văn, thực phủ và khí tượng. Mỗi nhóm đặc trưng phản ánh một khía cạnh cụ thể của môi trường tự nhiên và đóng vai trò quan trọng trong việc xác định mức độ nguy cơ lũ quét. Trong nội dung này, nhóm nghiên cứu sẽ phân tích chi tiết các đặc trưng trong đó nhấn mạnh ý nghĩa vật lý, cách thu thập, và vai trò của chúng trong việc mô hình hóa nguy cơ lũ quét.

Bảng III - 5. Phân tích lựa chọn các yếu tố dữ liệu đầu vào

TT	Đặc trưng	Ý nghĩa	Phân tích lý do lựa chọn	Các yếu tố liên quan và lý do loại bỏ
1	Cao độ so với sông suối	Chênh lệch độ cao giữa điểm địa hình với sông, suối gần nhất theo hướng dòng chảy	Các điểm có chênh lệch cao độ so với sông, suối thấp có nguy cơ lũ quét cao hơn	Không có
2	Khoảng cách đến sông suối	Khoảng cách giữa điểm địa hình với sông, suối gần nhất theo hướng dòng chảy	Các điểm có khoảng cách gần sông, suối có nguy cơ lũ quét cao hơn	Không có
3	Độ dốc	Độ dốc địa hình (độ dốc cục bộ điểm)	Có tác động đến lũ quét: độ dốc lớn làm cho nước di chuyển nhanh hơn.	Không có
4	Độ dốc lòng dẫn	Độ dốc của lòng dẫn (theo hướng dòng chảy)	Độ dốc lòng dẫn lớn → năng lượng dòng chảy lớn hơn trong cùng điều kiện về lưu lượng. Tác động đến thời gian tập trung dòng chảy và các yếu tố thủy văn khác	Không có
5	Chiều dài dòng chảy	Khoảng cách từ điểm xa nhất của lưu vực đến điểm tính toán	Ảnh hưởng đến thời gian tập trung dòng chảy, chiều dài ngắn thì thời gian tập trung nhanh hơn	Không có
6	Diện tích lưu vực	[Nguyên tắc lưu vực] – Diện tích từng lưu vực. Mỗi điểm được coi là cửa ra một lưu vực	Diện tích lớn → lưu lượng lớn hơn trong cùng lượng mưa, là tham số ảnh hưởng trực tiếp đến kết quả tính toán thủy văn	Không có
7	Chỉ số âm địa hình	Phản ánh khả năng tích tụ nước tại một điểm.	TWI cao mô tả các khu vực thấp, trũng (chân đồi, lòng dẫn...) → nguy cơ ngập lụt và lũ quét cao hơn.	Không có

TT	Đặc trưng	Ý nghĩa	Phân tích lý do lựa chọn	Các yếu tố liên quan và lý do loại bỏ
8	Chỉ số sức mạnh dòng chảy	Mô tả năng lượng tiềm năng của dòng chảy bề mặt trong đánh giá xói mòn và vận chuyển trầm tích	SPI cao → năng lượng lớn → khả năng xói mòn mạnh → tăng nguy cơ lũ quét	Không có
9	Chỉ số vị trí địa hình	Chênh lệch cao độ của một điểm so với bình quân các điểm xung quanh	TWI thấp ( $<0$ ) là khu vực tích tụ, bị tập trung nước. TWI cao ( $>0$ ) là khu vực thoát nước	Không có
10	Chỉ số NDVI	[Nguyên tắc lưu vực] Mô tả mật độ và sức khỏe của thực vật bình quân trong một lưu vực.	NDVI cao → thảm thực vật dày → giữ nước tốt → giảm nguy cơ lũ quét	Yếu tố liên quan trực tiếp là thảm phủ và sử dụng đất. Hai yếu tố này là định tính đối với lũ quét, do đó không lựa chọn trong nghiên cứu này. (Thay bằng chỉ số NDVI, CN)
11	Chỉ số CN	[Nguyên tắc lưu vực] Phản ánh khả năng thấm nước, trữ nước, sử dụng đất và loại đất	CN cao → trữ nước kém và tạo dòng chảy mặt lớn → tăng nguy cơ lũ quét	Lý do: chưa có các nghiên cứu rõ ràng về đất trồng lúa và hoa màu thì loại nào có tác động bất lợi đến lũ quét...
12	Tốc độ thấm bình quân	[Nguyên tắc lưu vực] Mô tả tốc độ thấm của nước mưa xuống bề mặt	Tốc độ thấm càng lớn → giảm dòng chảy mặt → giảm nguy cơ lũ quét	Yếu tố liên quan trực tiếp là loại đất (phân loại theo nhóm), các loại đất có tác động đến lũ quét khác nhau nhưng chưa có chỉ tiêu định lượng rõ ràng, do đó loại đất không được lựa chọn trong nghiên cứu này (thay bằng tốc độ thấm bình quân của đất).
13	Cao độ địa hình	Giá trị cao độ tại một vị trí	Ảnh hưởng trực tiếp hoặc gián tiếp đến lũ quét. Cao độ địa hình thấp có nguy cơ bị lũ quét hơn.	Không có
14	Cao độ bình quân lưu vực	[Nguyên tắc lưu vực] Giá trị cao độ bình quân lưu vực	Lưu vực có cao độ bình quân cao là các lưu vực ở vùng đồi, núi. Có ảnh hưởng trực tiếp hoặc gián tiếp đến lũ quét.	Không có
15	Độ cong địa hình	Mô tả bề mặt của địa hình theo hướng dốc	Nếu âm (lõm) → tích tụ dòng chảy (chân dốc,	Không có

TT	Đặc trưng	Ý nghĩa	Phân tích lý do lựa chọn	Các yếu tố liên quan và lý do loại bỏ
	(theo hướng dốc)		thung lũng) → tăng nguy cơ lũ quét	
16	Độ cong địa hình (phuong ngang)	Mô tả bề mặt của địa hình theo hướng ngang	Nếu âm (lõm) → hội tụ dòng chảy → tăng nguy cơ lũ quét	Không có
17	Lượng mưa giờ lớn nhất	Lượng mưa giờ lớn nhất trong một thời đoạn	Lượng mưa thời đoạn ngắn lớn có khả năng sinh lũ quét lớn. Lũ quét trong tự nhiên được hình thành do mưa lớn trong thời đoạn ngắn.	
18	Lượng mưa 3 giờ lớn nhất	Lượng mưa 3 giờ lớn nhất trong một thời đoạn		Không có
19	Lượng mưa 6 giờ lớn nhất	Lượng mưa 6 giờ lớn nhất trong một thời đoạn	Thời gian thường dưới 6 giờ.	
20	Lượng mưa 24 giờ lớn nhất	Lượng mưa 24 giờ lớn nhất trong một thời đoạn		Yếu tố liên quan trực tiếp là độ ẩm kỳ trước (bằng các bản đồ như SMAP...). Nghiên cứu này sử dụng lượng mưa 24 giờ lớn nhất thay thế cho độ ẩm kỳ trước (bởi các bản đồ vệ tinh) do chu kỳ ảnh không liên tục.

## 2. Xây dựng mô hình trí tuệ nhân tạo phân vùng lũ quét cho khu vực nghiên cứu

### a. Chuẩn hóa dữ liệu đầu vào

Quá trình xây dựng mô hình bắt đầu với việc thu thập và tiền xử lý dữ liệu không gian từ các tệp raster, bao gồm các đặc trưng địa hình như độ cao, khoảng cách đến dòng chảy, độ dốc, chỉ số độ ẩm địa hình (TWI), chỉ số sức mạnh dòng chảy (SPI), và các đặc trưng khí tượng như lượng mưa tối đa trong các khung thời gian khác nhau (3 giờ, 6 giờ, 24 giờ). Dữ liệu không gian được lưu trữ dưới định dạng GeoTIFF, đòi hỏi các kỹ thuật xử lý đặc biệt để đảm bảo tính nhất quán và khả năng sử dụng trong học máy.

Một nguyên tắc nhỏ là các biến đầu vào nên có giá trị nhỏ, có thể nằm trong khoảng  $0 \div 1$  hoặc được chuẩn hóa với giá trị trung bình là 0 và độ lệch chuẩn là 1. Tuy nhiên, nếu các giá trị của biến nhỏ (gần với 0 và 1) và phân phối dữ liệu bị hạn chế (độ lệch chuẩn lân cận 1) thì có thể không cần chia tỷ lệ dữ liệu. Điều này sẽ giúp mô hình đào tạo nhanh hơn và giảm khả năng mắc kẹt trong các tối ưu cục bộ [93].

Do vậy, toàn bộ số liệu đầu vào được chuẩn hóa theo nguyên tắc này và được thể hiện trong bảng sau:

Bảng III - 6. Chuẩn hóa các dữ liệu đầu vào cho mô hình

TT	Đặc trưng	Ký hiệu	Đơn vị	Phương pháp	Chuẩn hóa	Khoảng giá trị	Công thức Bước 1
1	Cao độ so với sông suối	eleStream	m	Cục bộ	Robust Scaling + MinMax	[0, 1]	$X' = \frac{X - \text{median}}{\text{IQR}}$
2	Khoảng cách đến sông suối	disStream	m	Cục bộ	Log + MinMax	[0, 1]	$X' = \log(X + 1)$
3	Độ dốc	wSlope	độ	Lưu vực	MinMax	[0, 1]	$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$
4	Độ dốc lòng dẫn	stream Slope	m/m	Cục bộ	MinMax	[0, 1]	$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$
5	Chiều dài dòng chảy	flowLength	m	Cục bộ	Log + MinMax	[0, 1]	$X' = \log(X + 1)$
6	Diện tích lưu vực	area	$\text{m}^2$	Lưu vực	Log + MinMax	[0, 1]	$X' = \log(X + 1)$
7	Chỉ số âm địa hình	twi		Cục bộ	MinMax	[0, 1]	$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$
8	Chỉ số sức mạnh dòng chảy	spi		Cục bộ	Log + MinMax	[0, 1]	$X' = \log(X + 1)$
9	Chỉ số vị trí địa hình	tpi		Cục bộ	Z-score + MinMax	[0, 1]	$X' = \frac{X - \mu}{\sigma}$
10	Chỉ số NDVI	wNdvi		Lưu vực	MinMax	[0, 1]	$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$
11	Chỉ số CN	wCN		Lưu vực	MinMax	[0, 1]	$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$
12	Tốc độ thám bình quân	wInfiRate	mm/hour	Lưu vực	Robust Scaling + MinMax	[0, 1]	$X' = \frac{X - \text{median}}{\text{IQR}}$
13	Cao độ địa hình	dem	m	Cục bộ	Robust Scaling + MinMax	[0, 1]	$X' = \frac{X - \text{median}}{\text{IQR}}$
14	Cao độ bình quân lưu vực	eleWatershed	m	Lưu vực	Robust Scaling + MinMax	[0, 1]	$X' = \frac{X - \text{median}}{\text{IQR}}$
15	Độ cong địa hình (theo hướng dốc)	profCurvature		Cục bộ	Z-score + MinMax	[0, 1]	$X' = \frac{X - \mu}{\sigma}$
16	Độ cong địa hình (phuong ngang)	planCurvature		Cục bộ	Z-score + MinMax	[0, 1]	$X' = \frac{X - \mu}{\sigma}$
17	Lượng mưa giờ lớn nhất	max_p_recip	mm	Lưu vực	Log & “÷10”	[0, 1]	$X' = \log(X + 1)$

TT	Đặc trưng	Ký hiệu	Đơn vị	Phương pháp	Chuẩn hóa	Khoảng giá trị	Công thức Bước 1
18	Lượng mưa 3 giờ lớn nhất	max_3 h_prec_ip	mm	Lưu vực	Log & “÷10”	[0, 1]	$X' = \log (X+1)$
19	Lượng mưa 6 giờ lớn nhất	max_6 h_prec_ip	mm	Lưu vực	Log & “÷10”	[0, 1]	$X' = \log (X+1)$
20	Lượng mưa 24 giờ lớn nhất	max_2 4h_prec_ip	mm	Lưu vực	Log & “÷10”	[0, 1]	$X' = \log (X+1)$

Nếu phương pháp chuẩn hóa chỉ có 1 bước, thì thực hiện theo công thức ghi trong cột cuối, nếu có 2 bước, thì bước thứ hai là theo phương pháp MinMax (tham khảo chỉ số NDVI hoặc CN)

### b. Chuẩn bị nhãn dữ liệu dự đoán

Dữ liệu dự đoán là dữ liệu nhãn, được gán các giá trị từ 0 đến 4. Các giá trị này được đánh giá định lượng theo số (từ 0 đến 4) dựa trên đánh giá của nhóm nghiên cứu thực địa tại khu vực huyện Mù Cang Chải cho các suối chính ở một số xã điển hình cho trận lũ năm 2023.

Bảng III - 7. Nhãn mức độ lũ và ý nghĩa

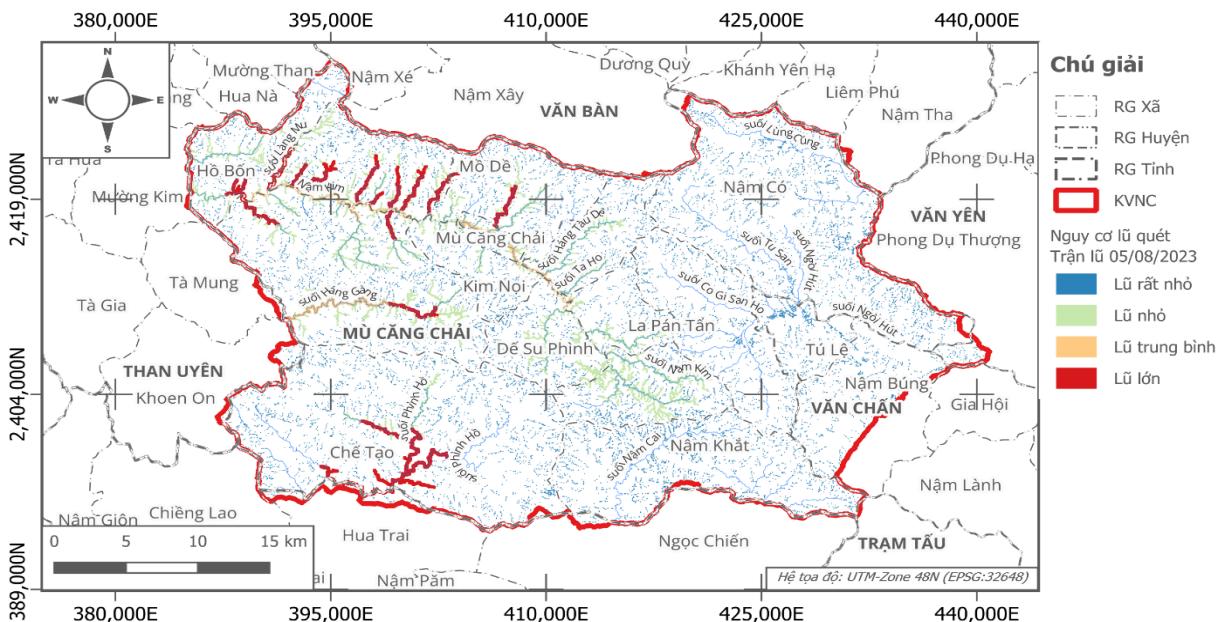
TT	Nhãn	Giá trị nhãn	Ý nghĩa
0	Không có lũ	0	Các điểm thuộc mái dốc của núi, đỉnh núi, nơi không có tập trung dòng chảy hoặc có tập trung dòng chảy không đáng kể.
1	Lũ rất nhỏ	1	Dòng chảy trên suối không gây nguy hiểm đến các đối tượng, là dòng chảy phổ biến xuất hiện trên khu vực.
2	Lũ nhỏ	2	Dòng chảy trên suối là dòng chảy nhanh, hình thành do mưa nhưng không gây nguy hiểm đến các đối tượng.
3	Lũ trung bình	3	Dòng chảy trên sông suối là dòng chảy xiết, nằm trong lòng dẫn và an toàn để có thể đi qua các công trình cầu treo, không cuốn trôi các vật liệu lớn gây nguy hiểm cho cộng đồng sinh sống quanh khu vực
4	Lũ lớn	4	Dòng chảy trên sông suối là dòng chảy xiết, có cuốn trôi các vật liệu lớn hoặc nhỏ trong lòng dẫn, có thể tác động đến các công trình như cầu qua sông và các đối tượng cộng đồng nhà dân sinh sống xung quanh khu vực.

Dựa trên các tiêu chí này, nhóm nghiên cứu đã tiến hành thu thập thông tin tại các xã được đánh giá là có ảnh hưởng bởi lũ bao gồm các xã Khang Mao, Hồ Bón, Mò Dè, Lao Chải, Ché Tạo và Nậm Có. Chi tiết như sau:

Bảng III - 8. Nhãn phân loại đánh giá theo tình hình mưa lũ thực tế tại địa phương

TT	Địa chỉ	Suối	Đánh giá đợt lũ 8/2023	Nhãn phân loại
1	Xã Hò Bón	Háng Nhù, Thống Gầu Bua, Làng Mu	Hầu hết các suối thuộc khu vực xã Hò Bón có dòng chảy mạnh, xiết. Người dân rất cảnh giác trong đợt mưa lũ đầu tháng 8/2023. Dòng chảy trên các suối này cuốn trôi nhiều vật liệu có đường kính lên tới 1m, phô biến là vài chục cm. Các khu vực nhánh suối đổ ra hướng suối Nậm Kim được đánh giá có nguy cơ cao, trong khi các nhánh suối đổ về phía suối Hàng Đè Chu ghi nhận dòng chảy lũ bình thường (có lũ nhưng ít nguy hiểm)	Các nhánh suối chính như Hàng Nhù, Thống Gầu Bua, Làng Mu và lân cận suối được gán nhãn 4, các suối còn lại gán nhãn 3.
2	Xã Khao Mang	Suối Hàng B La Ha; Hàng B La Đê; Hàng Tàu Đê; Páo Sơ Dao; Tủa Má Pán; Giàng Xua; Hàng Trán	Trong các nhánh suối này, khu vực Hàng B La Ha và Hàng B La Đê ghi nhận lũ lớn tại khu vực đổ vào suối chính Nậm Kim; trong khi các nhánh suối khác có lũ nhỏ hơn. Riêng suối Hàng Trán trong trận lũ này không có lũ lớn, không được coi là lũ quét.	Một số suối nhánh Hàng B La Ha và Hàng B La Đê gán nhãn 4, suối còn lại gán nhãn 3.
3	Xã Lao Chải	Suối Hàng Đè Sủa, Hàng Gàng, Lao Chải	Riêng suối Hàng Đè Sủa là suối được ghi nhận có lũ quét rất lớn và đổ trực tiếp vào suối Nậm Kim đợt mưa lũ này. Suối này cuốn các vật liệu lên tới 2-3m. Suối Hàng Gàng ở khu vực cuối Bản Hàng Gàng cũng xảy ra lũ rất lớn.	2 nhánh suối Hàng Đè Sủa và Hàng Gàng được gán nhãn 4. Khu vực thượng nguồn bản Hàng Gàng và các suối khác đổ vào gán nhãn 3.
4	Xã Mồ Dề	Suối Nà Hàng	Suối Nà Hàng thuộc xã Mồ Dề là suối có lũ rất lớn trong đợt mưa lũ 8/2023. Các nhánh suối khác có ghi nhận lũ nhỏ hơn.	Suối Nà Hàng được gán nhãn 4, trong khi các nhánh suối khác đổ vào Nậm Kim gán nhãn 3.
5	Xã Chẽ Tạo	Suối Nậm Khắt, Nậm Khốt, Phình Hồ	Các nhánh suối đổ vào suối Phình Hồ đợt mưa này đều ghi nhận lũ lớn, đặc biệt là tại các bản Chẽ Tạo, Phú Vá, Tà Sung. Khu vực bản Nà Hàng không ghi nhận lũ lớn.	Các nhánh suối Nậm Khắt, Nậm Khốt và Phình Hồ được gán nhãn 4, các nhánh suối đổ vào được gán nhãn 3.
6	Xã Nậm Có		Các suối trên khu vực xã Nậm Có không ghi nhận lũ lớn.	Các nhánh suối trên khu vực này được gán nhãn 2.

Ngoài ra theo mô tả, khu vực bản Mí Háng Táu (thuộc xã Púng Luông) cũng có ghi nhận lũ rất lớn, gây ảnh hưởng đến sinh hoạt của người dân. Các khu vực xã khác và các nhánh suối khác không ghi nhận lũ lớn. Nhìn chung, phần lớn các nhánh suối khu vực hạ du suối Nậm Kim đều ghi nhận lũ lên, trong khi đó, phía suối chính Ngòi Hút (đô về huyện Văn Yên) không có ghi nhận lũ lớn.



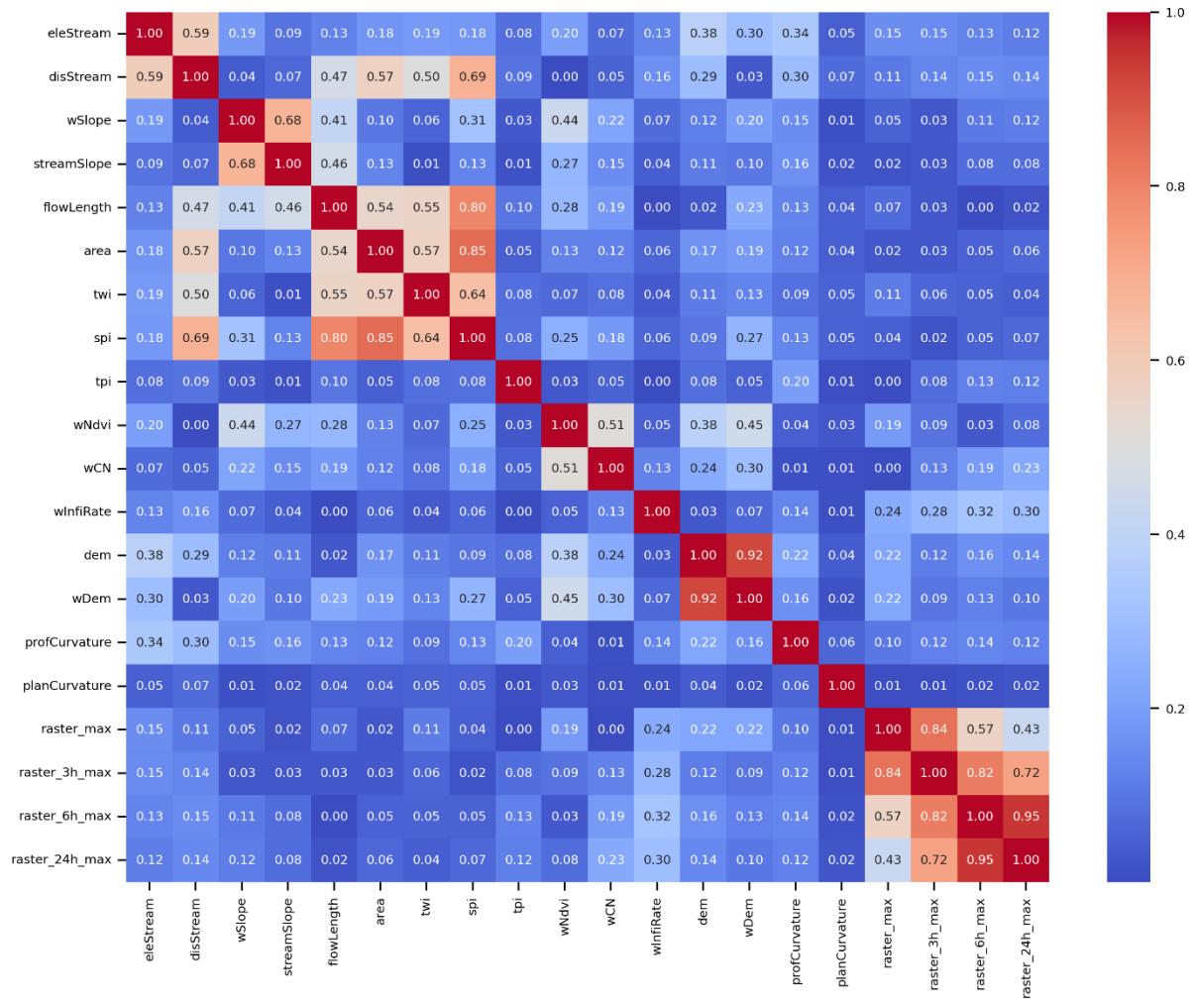
Hình III - 4. Kết quả điều tra các nhánh sông bị lũ quét trong trận lũ 05/08/2023 tại huyện Mù Cang Chải và phân loại nguy cơ lũ quét dựa trên đánh giá.

Ngoài các khu vực mô tả phía trên (chủ yếu ghi nhận lũ lớn), các khu vực còn lại được đánh nhãn 1 nếu nằm trên các sườn núi có độ dốc lớn (mái dốc núi), đỉnh núi và các khu vực ruộng bậc thang. Nhãn 2 được đánh cho các nhánh suối nhỏ và rất nhỏ ở thượng nguồn các khu vực không ghi nhận lũ lớn.

Mặc dù có dữ liệu mưa của trạm Mù Cang Chải theo giờ tại các trận lũ khác trước năm 2021, tuy nhiên, dữ liệu mưa tại một trạm không đủ đại diện cho một khu vực nhỏ bé, do đó, khó có thể đánh giá được chính xác lượng mưa sinh lũ của các trận lũ trước năm 2021. Từ năm 2021 trở đi, mật độ quan trắc mưa có thể được coi là tương đối tốt, do đó, nghiên cứu sử dụng dữ liệu năm 2023 (cho trận lũ xảy ra tại Hồ Bón) làm cơ sở để xác định các điểm phân loại nguy cơ như đã trình bày phía trên.

### c. Xây dựng mô hình học máy

Lựa chọn đặc trưng là một bước quan trọng để giảm độ phức tạp của mô hình và cải thiện hiệu suất. Trong nghiên cứu này, hai kỹ thuật chính được sử dụng để giảm số lượng đặc trưng: loại bỏ các đặc trưng có tương quan cao và lựa chọn đặc trưng dựa trên điểm số thống kê. Các tương quan được tính toán để xác định các đặc trưng có mức tương quan tuyệt đối lớn hơn 0,85 sẽ bị loại bỏ nhằm làm giảm đa cộng tuyến.



Hình III - 5. Ma trận tương quan các đặc trưng

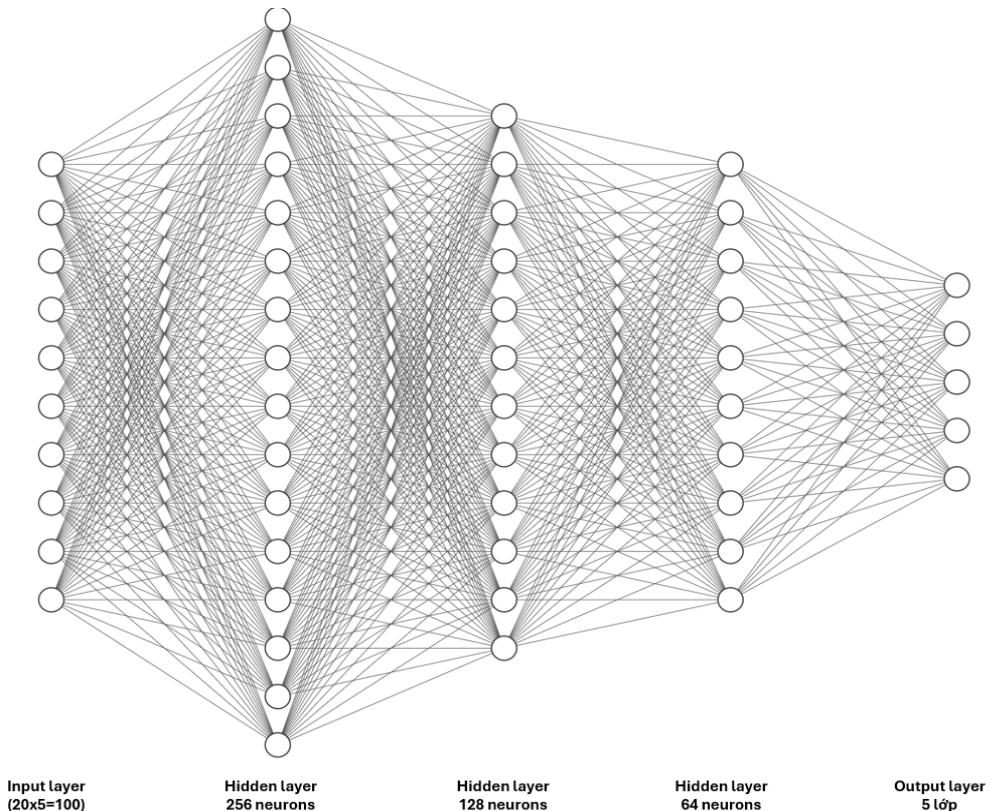
Theo ma trận tương quan các đặc trưng của dữ liệu đưa vào mô hình học máy, cao độ bình quân lưu vực (wDem) có mức độ tương quan lớn với cao độ cửa ra (0,92), bên cạnh đó, lượng mưa lớn nhất 24 giờ cũng có tương quan rất lớn với lượng mưa lớn nhất 6 giờ. Do lũ quét thường xảy ra trong khoảng 6 giờ và thời gian tập trung dòng chảy của các lưu vực nhỏ sinh lũ quét cũng trong khoảng này, nhóm nghiên cứu loại bỏ 2/20 đặc trưng là cao độ bình quân lưu vực (wDem) và lượng mưa 24 giờ lớn nhất. Như vậy, 18/20 đặc trưng sẽ được đưa vào đánh giá và xây dựng mô hình học máy.

#### d. Xây dựng mô hình học sâu

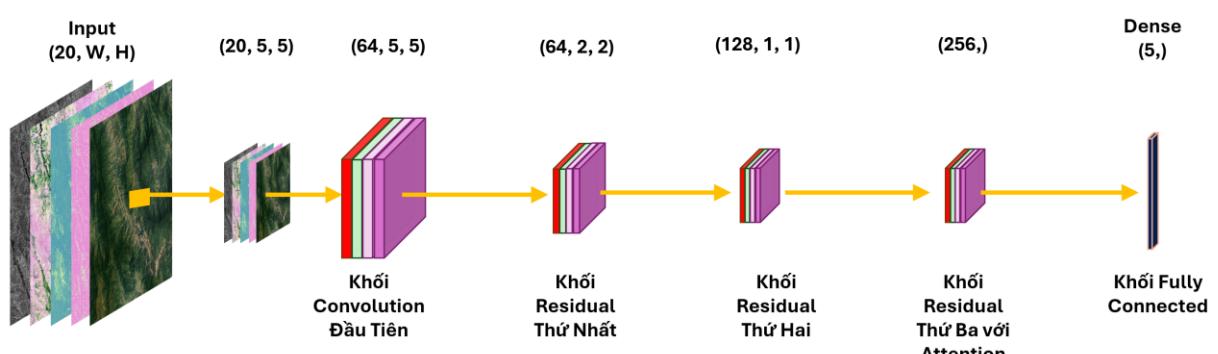
Không giống mô hình học máy, mô hình học sâu có thể tự lựa chọn các đặc trưng theo thuật toán để đưa vào dự đoán, do đó, không cần phải loại bỏ các đặc trưng khác khi sử dụng mô hình học sâu. Các yếu tố không gian xung quanh được lựa chọn bằng việc thử dần các tham số. Nghiên cứu lựa chọn phương pháp thử dần cho các vùng lân cận từ 3x3 đến 11x11, kết quả lựa chọn được mô hình CNN được lấy vùng lân cận 5x5 và mô hình DNN được lấy đặc trưng lân cận 7x7 cho ra sự dự đoán tốt nhất (chỉ số Accuracy tốt nhất).

Tất cả 20 tham số được giữ lại mà không loại bỏ bất kỳ tham số nào. Lý do là CNN được thiết kế để xử lý dữ liệu không gian và có khả năng tự động trích xuất các mẫu

không gian phức tạp từ các vùng lân cận. Việc giữ nguyên toàn bộ tham số đảm bảo rằng mô hình có thể khai thác mọi thông tin không gian có sẵn, từ các mẫu đơn giản như độ dốc địa hình đến các mẫu phức tạp hơn như sự kết hợp giữa độ ẩm và lượng mưa. Các giá trị dữ liệu cũng đã được chuẩn hóa tương tự mô hình học máy, do đó, mô hình này mang tính kế thừa dữ liệu từ các mô hình học máy đã được xây dựng từ trước.



Hình III - 6. Cấu trúc thiết kế mô hình DNN



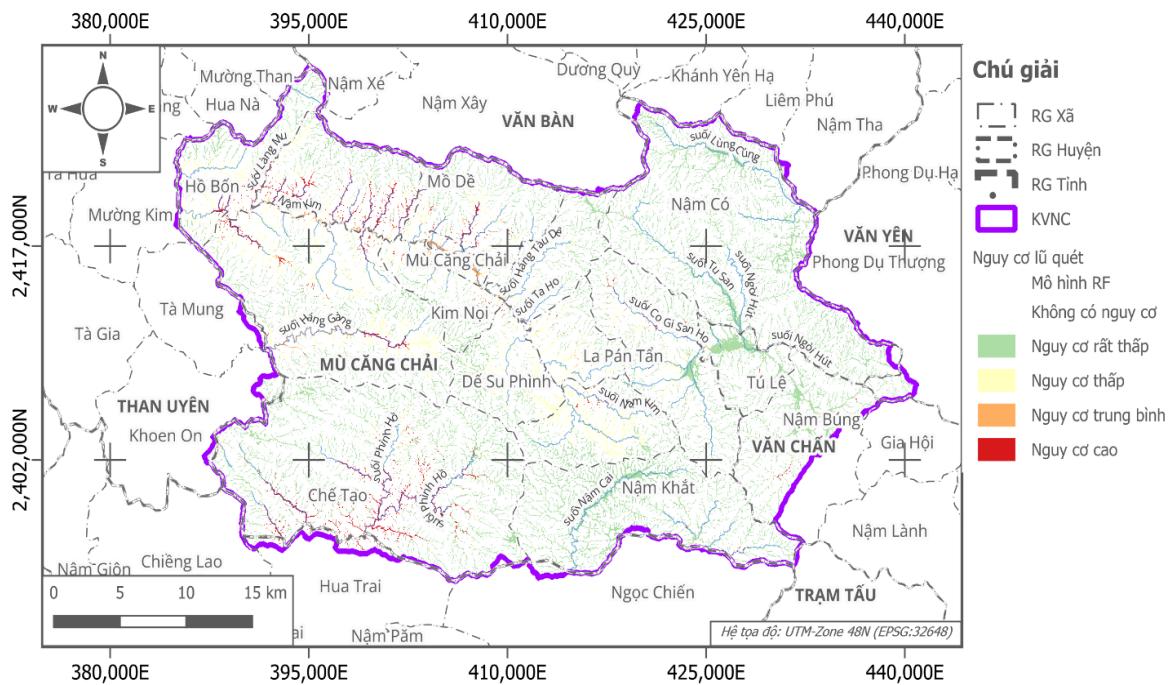
Hình III - 7. Cấu trúc các khối thuật toán trong mô hình CNN

#### e. Kết quả xây dựng mô hình

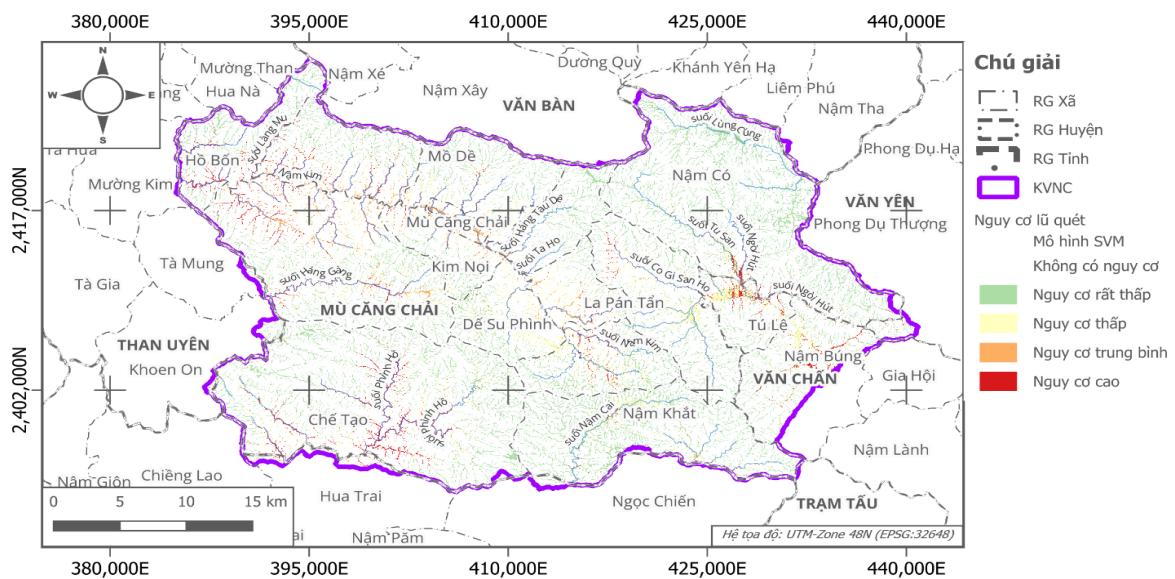
Bảng III - 9. Kết quả tổng hợp đánh giá và khuyến nghị lựa chọn mô hình phân vùng lũ quét

Mô hình	Accuracy	Precision	Recall	F1 Score	Thời gian	Tỷ lệ Hiệu suất/Thời gian	Xếp hạng
LGBM	95.24%	95.23%	95.24%	95.22%	10 p	Xuất sắc	1

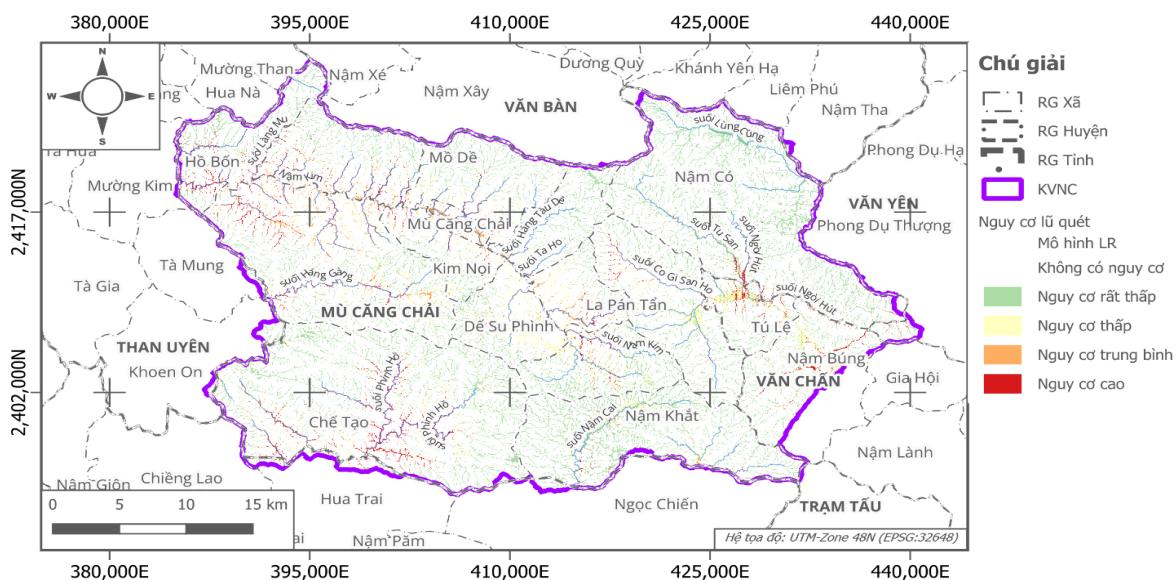
Random Forest	93.95%	93.92%	93.95%	93.91%	10 p	Rất tốt	2
Ensemble	93.26%	93.22%	93.26%	93.22%	8 p	Tốt	3
CNN	90.00%	89.98%	90.00%	89.95%	4h 40p	Trung bình	4
DNN	89.92%	89.77%	89.92%	89.77%	5h 40p	Trung bình	5
SVM	69.47%	68.91%	69.47%	69.08%	6h 5p	Rất kém	6
Logistic Regression	68.97%	68.39%	68.97%	68.57%	8 p	Kém	7



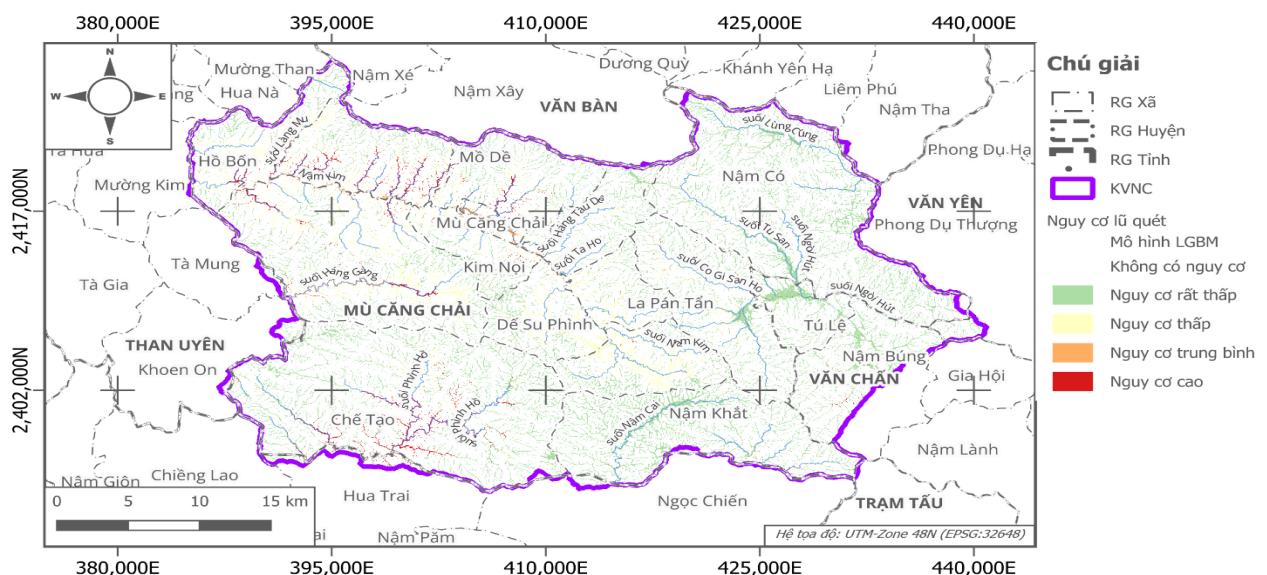
Hình III - 8. Kết quả xác định nguy cơ lũ quét bằng mô hình RF



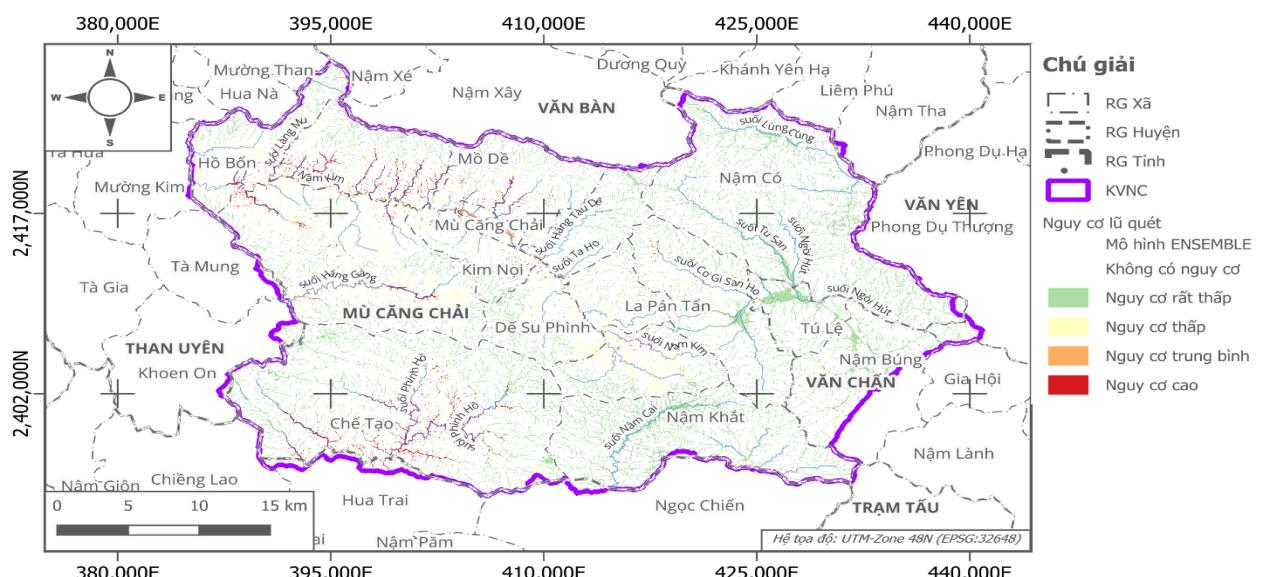
Hình III - 9. Kết quả xác định nguy cơ lũ quét bằng mô hình SVM



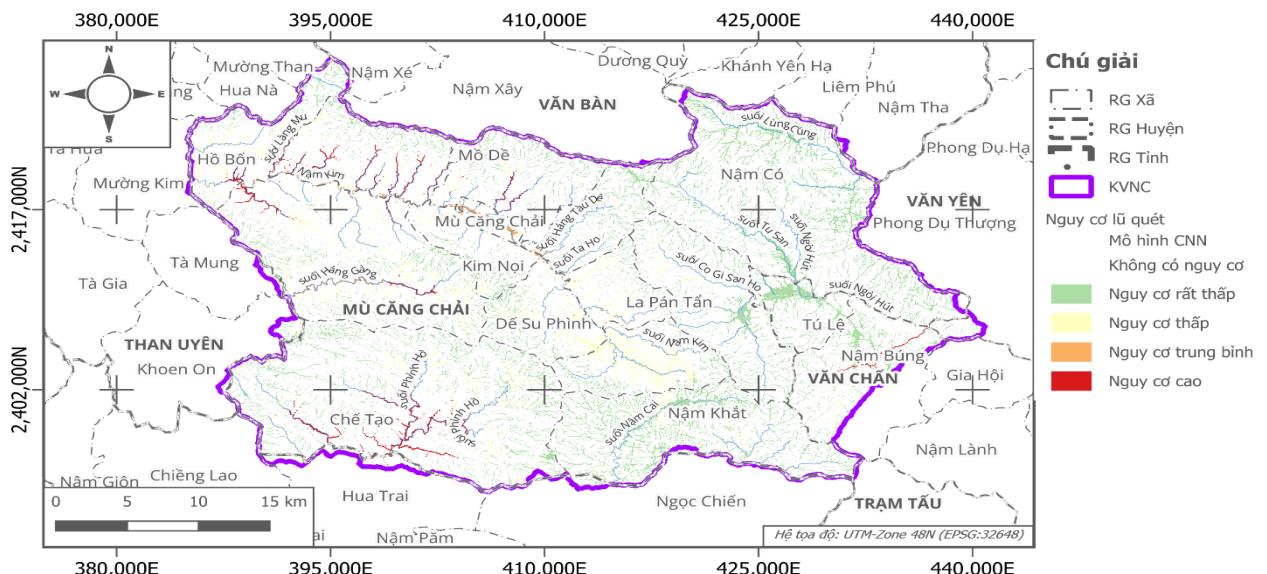
Hình III - 10. Kết quả xác định nguy cơ lũ quét bằng mô hình LR



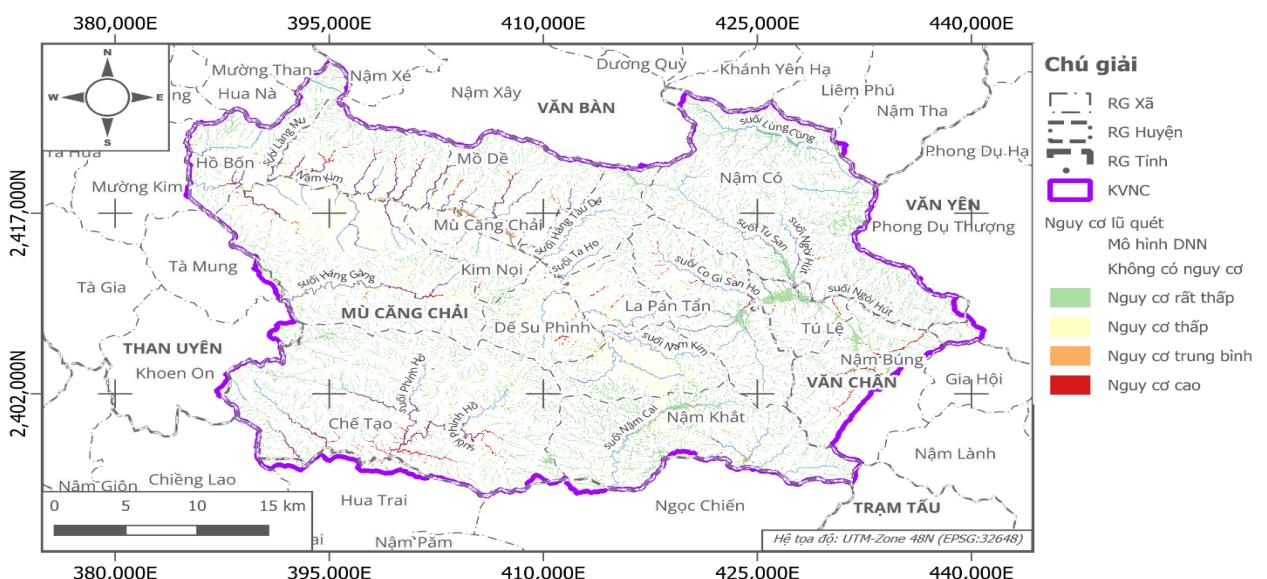
Hình III - 11. Kết quả xác định nguy cơ lũ quét bằng mô hình LGBM



Hình III - 12. Kết quả xác định nguy cơ lũ quét bằng mô hình ENSEMBLE



Hình III - 13. Kết quả xác định nguy cơ lũ quét bằng mô hình CNN

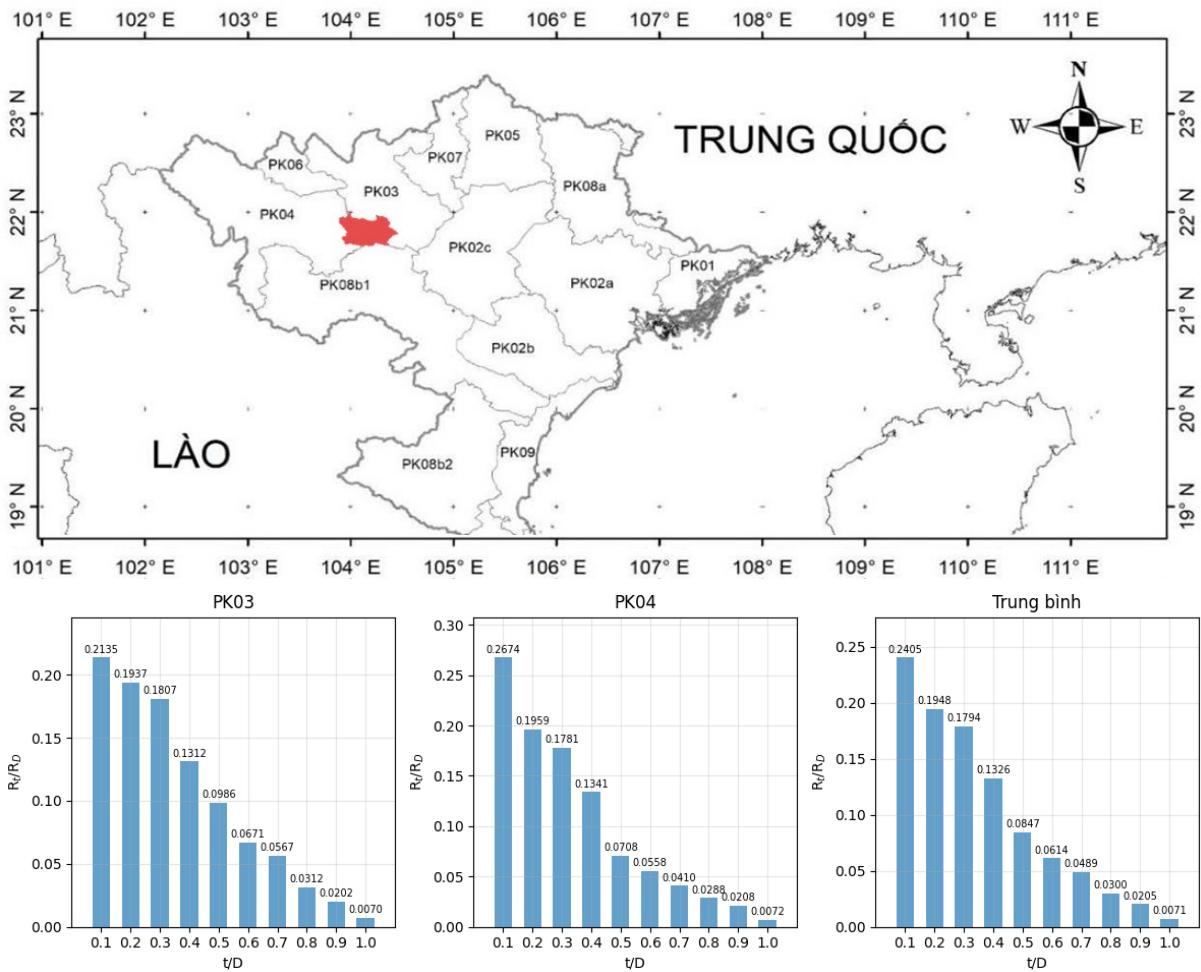


Hình III - 14. Kết quả xác định nguy cơ lũ quét bằng mô hình DNN

### 3. Xây dựng bản đồ phân vùng lũ quét theo kịch bản mưa

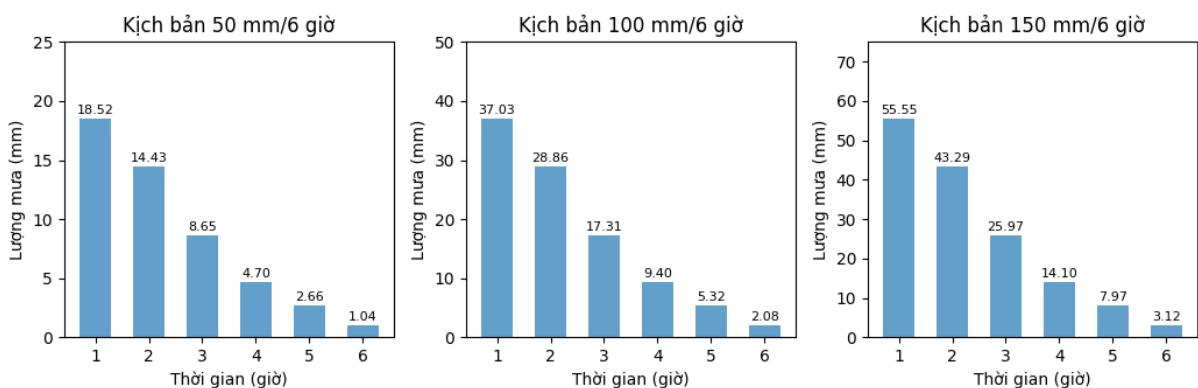
#### a. Xây dựng kịch bản mưa

Kết quả nghiên cứu phía trên là kết quả phân vùng lũ quét cho khu vực Mù Cang Chải trận lũ 8/2023. Trên cơ sở đó, nghiên cứu tiếp tục triển khai xây dựng bản đồ phân vùng lũ quét cho các kịch bản mưa giả định. Do có rất nhiều kịch bản mưa, nghiên cứu này sử dụng kịch bản mưa bát lợi nhất theo TCVN 13615:2022 làm cơ sở xác định biểu đồ phân bố tương ứng với lượng mưa thời đoạn 6 giờ cho các dự báo với tổng lượng mưa lần lượt là 50mm; 100mm; và 150mm. Mô hình CNN được áp dụng.



Hình III - 15. Phân vùng mưa rào theo TCVN 13615:2022 và khu vực nghiên cứu

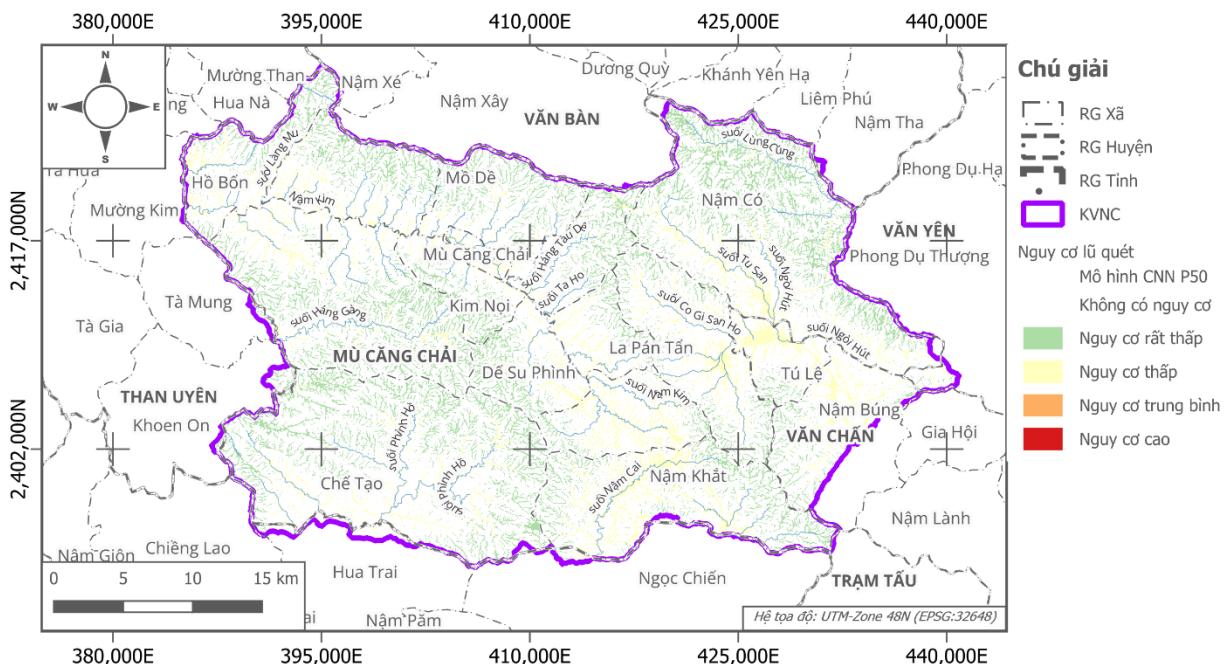
Khu vực nghiên cứu nằm trong vùng PK04 và PK03. Do đó, nghiên cứu này sẽ lấy bình quân phân vùng mưa rào bát lợi nhất của cả hai vùng làm cơ sở để xây dựng biểu đồ phân bố mưa cho các kịch bản mưa. Các kịch bản thể hiện như sau:



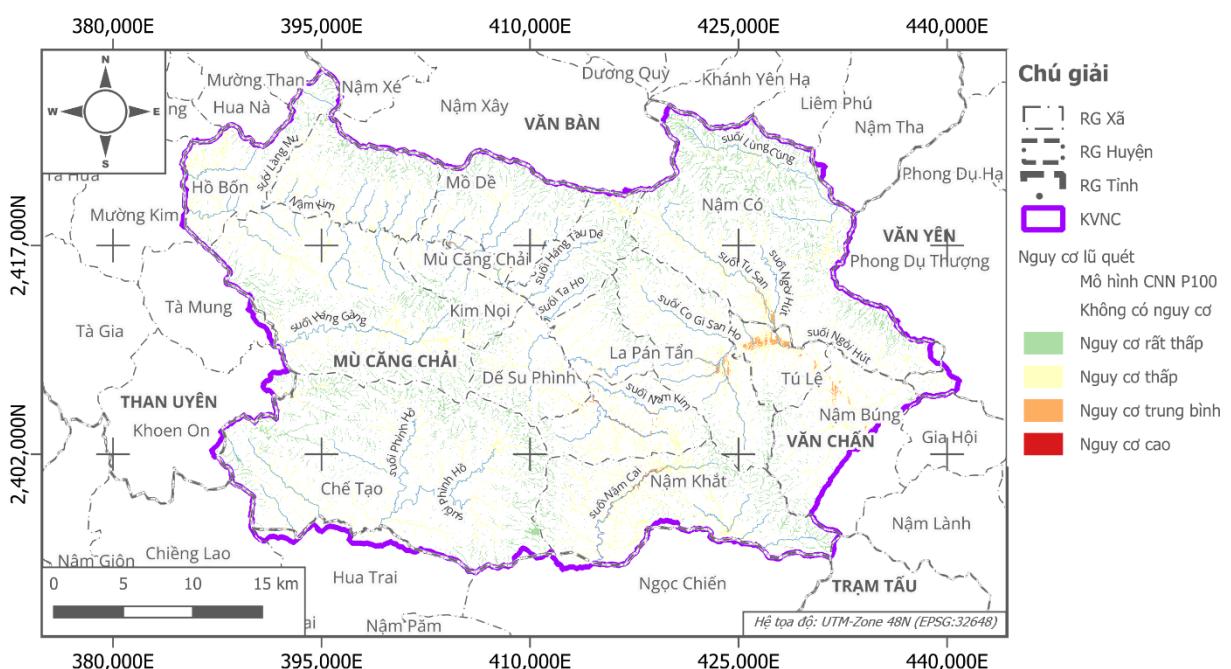
Hình III - 16. Phân bố mưa tương ứng với mô hình mưa bát lợi nhất cho KVNC

Tham số	Kịch bản mưa trong 6 giờ (mm)		
	50 mm	100 mm	150 mm
1 giờ max	18.52	37.03	55.55
3 giờ max	41.6	83.2	124.81
6 giờ max	50	100	150
24 giờ max	102.39	204.79	307.18

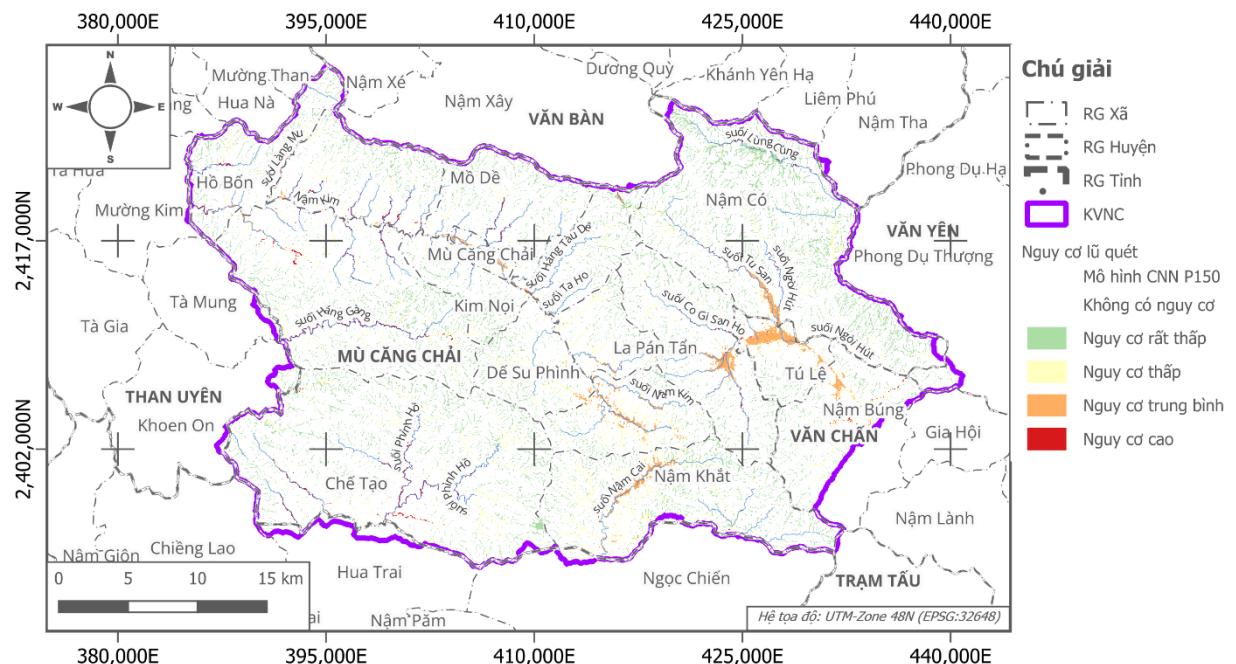
## b. Xây dựng bản đồ phân vùng nguy cơ bằng mô hình CNN



Hình III - 17. Kết quả xây dựng bản đồ phân vùng nguy cơ kịch bản mưa 50mm/6 giờ



Hình III - 18. Kết quả xây dựng bản đồ phân vùng nguy cơ kịch bản mưa 100mm/6 giờ



Hình III - 19. Kết quả xây dựng bản đồ phân vùng nguy cơ kịch bản mưa 150mm/6 giờ

## **IV. Kết luận và kiến nghị**

### **Kết luận**

Kết quả của nghiên cứu đã đạt được những thành tựu trong việc ứng dụng trí tuệ nhân tạo cho lĩnh vực phòng chống thiên tai, thể hiện qua việc xây dựng thành công hệ thống phân vùng lũ quét với độ các chỉ số đánh giá tốt. Mô hình LGBM (Light Gradient Boosting Machine) đã khẳng định vị thế dẫn đầu với độ chính xác lên tới 95.24%, kèm theo các chỉ số Precision, Recall và F1-Score đều duy trì ở mức cao trên 95%. Điều đáng chú ý là thời gian huấn luyện mô hình chỉ trong vòng 10 phút, thể hiện tính hiệu quả vượt trội trong việc xử lý dữ liệu địa hình và khí tượng phức tạp, mở ra tiềm năng triển khai ứng dụng thực tế với chi phí tính toán hợp lý.

Sự thành công của các mô hình trí tuệ nhân tạo còn thể hiện qua việc phát triển hệ thống phân loại thảm phủ sử dụng mô hình CNN, đạt độ chính xác huấn luyện 98.68% và độ chính xác thực 98.32% sau 50 epoch trong việc phân loại 7 lớp thảm phủ khác nhau. Kết quả này không chỉ chứng minh khả năng ứng dụng mạnh mẽ của học sâu trong xử lý dữ liệu viễn thám mà còn tạo nền tảng quan trọng cho việc đánh giá rủi ro lũ quét, vì thảm phủ là một trong những yếu tố then chốt ảnh hưởng đến khả năng thám nước và dòng chảy bờ mặt. Sự chênh lệch nhỏ giữa độ chính xác huấn luyện và xác thực (chỉ 0.36%) cho thấy mô hình có khả năng tổng quát hóa tốt, tránh được hiện tượng overfitting thường gặp trong các mô hình học sâu, đặc biệt các mô hình học sâu rất phù hợp sử dụng trong phân loại hình ảnh.

Việc tích hợp thành công 20 tham số đầu vào đa dạng, từ các đặc trưng địa hình như độ cao so với sông suối (eleStream), độ dốc lưu vực (wSlope), chỉ số độ ẩm địa hình (TWI) đến các yếu tố khí tượng như lượng mưa tối đa trong các khoảng thời gian khác nhau, đã tạo nên một hệ thống đánh giá toàn diện và khoa học. Đặc biệt, việc sử dụng kết hợp cả các tham số điểm và tham số trung bình lưu vực thể hiện các nguyên tắc cơ bản về tính chất đa tỷ lệ không gian của hiện tượng lũ quét dưới vai trò thủy văn học, từ đó nâng cao độ chính xác của mô hình dự báo.

Một điểm đáng chú ý khác là các mô hình phi tuyến sẽ có sự phù hợp tốt hơn các mô hình tuyến tính trong bài toán phân vùng lũ quét. Mô hình Random Forest với độ chính xác 93.95% đã chứng minh khả năng xử lý tốt các biến đầu vào có tính phi tuyến cao, trong khi các mô hình ensemble đạt 93.26% cho thấy tiềm năng kết hợp sức mạnh của nhiều thuật toán. Điều này không chỉ tạo ra sự linh hoạt trong lựa chọn mô hình phù hợp với từng điều kiện cụ thể mà còn mở ra hướng nghiên cứu tối ưu hóa kết hợp các phương pháp để đạt hiệu quả cao nhất. Tuy nhiên, các mô hình tuyến tính như hồi quy logistic hay máy hỗ trợ vectơ lại chưa thể hiện tốt với những tương quan phức tạp và phi tuyến.

### **Những hạn chế còn tồn tại**

Mặc dù đạt được những kết quả khả quan, nghiên cứu vẫn còn những hạn chế đáng chú ý. Sự chênh lệch lớn về hiệu suất giữa các mô hình cho thấy tính không ổn định trong việc lựa chọn thuật toán phù hợp. Trong khi LGBM đạt 95.24% thì Logistic

Regression chỉ đạt 68.97%, điều này phản ánh sự phức tạp trong việc lựa chọn mô hình tối ưu cho từng điều kiện địa lý và khí hậu cụ thể.

Thời gian huấn luyện của các mô hình học sâu như CNN và DNN tương đối dài (4-5 giờ) nhưng hiệu suất lại không vượt trội so với các mô hình truyền thống, đặt ra câu hỏi về hiệu quả kinh tế trong việc triển khai thực tế. Sự khác biệt về hiệu suất này có thể được giải thích qua bản chất khác nhau giữa bài toán phân vùng lũ quét và phân loại thảm phủ. Trong khi phân loại thảm phủ dựa trên các đặc trưng không gian liên tục và có tính chất tương quan cao giữa các pixel lân cận - điều mà CNN rất giỏi trong việc trích xuất thông qua các bộ lọc tích chập, thì phân vùng lũ quét lại phụ thuộc vào sự tương tác phức tạp giữa nhiều yếu tố địa hình, thủy văn và khí tượng mà không nhất thiết có tính liên tục không gian mạnh.

Hơn nữa, dữ liệu đầu vào cho phân vùng lũ quét thường là các giá trị số đơn lẻ (như độ cao, độ dốc, lượng mưa) thay vì dữ liệu hình ảnh có cấu trúc không gian rõ ràng như trong phân loại thảm phủ. CNN và DNN đòi hỏi một lượng lớn dữ liệu để học được các mẫu hình phức tạp, trong khi các mô hình truyền thống như LGBM và Random Forest lại có khả năng xử lý tốt hơn các dữ liệu dạng bảng với số lượng mẫu hạn chế và có thể nắm bắt được mối quan hệ phi tuyến giữa các biến một cách hiệu quả hơn.

Tuy nhiên, về mức độ phù hợp trong phân loại, các mô hình CNN và DNN lại tỏ ra phù hợp hơn dù độ chính xác hay hiệu suất kém hơn các mô hình cây quyết định. Do đó, trong quá trình kiểm chứng, đánh giá cho các trận lũ tiếp theo, cần có sự so sánh và kiểm chứng độc lập nhằm khẳng định độ tin cậy của mô hình phân vùng lũ quét.

Một hạn chế quan trọng khác là việc đánh giá chỉ dựa trên các chỉ số thống kê mà chưa xem xét đến tính khả thi trong ứng dụng thực tế. Độ chính xác cao trên tập dữ liệu thử nghiệm không hoàn toàn đảm bảo hiệu quả khi triển khai trong môi trường thực tế với những biến động không lường trước về khí hậu và địa hình. Ngoài ra, việc đánh giá mức độ lũ chỉ là sự đánh giá chủ quan và thiếu phân định rõ ràng (như thế nào là lớn, thế nào là trung bình...), điều này gây khó khăn thực sự trong phân vùng lũ quét nói chung và phân loại lũ nói riêng. Tuy nhiên, sự nhầm lẫn giữa các cấp độ phân vùng liên tục (giữa các cấp độ và các cấp độ lân cận) là có thể chấp nhận được do thiếu ranh giới rõ ràng.

## Kiến nghị

Bất cập lớn nhất trong việc sử dụng mô hình trí tuệ nhân tạo trong phân vùng/dự báo/cảnh báo nguy cơ lũ quét là dữ liệu, đặc biệt là nhãn dữ liệu. Các trận lũ quét xảy ra thường không phổ biến, do đó có rất ít mẫu được lấy cho việc huấn luyện mô hình trí tuệ nhân tạo. Ở Việt Nam, thông thường các trận lũ có thiệt hại về người hoặc cuốn trôi nhà cửa mới được xem là lũ quét, do đó các trận lũ quét không gây thiệt hại về người và tài sản, hoặc xảy ra ở các khu vực hẻo lánh không được ghi nhận. Điều này làm hạn chế rất lớn đối với việc xác định lũ quét cho mô hình. Đặc biệt, lũ quét tự nhiên sinh ra bởi mưa lớn, trong khi đó, lượng mưa phân bố ở khu vực miền núi là rất cục bộ, điều này

dẫn đến tình trạng lượng mưa quan trắc không đủ độ tin cậy hoặc phản ánh được chính xác nguyên nhân gây lũ quét, đặc biệt là khoảng trước năm 2021, khi số lượng các trạm quan trắc là không đáng kể (có những trạm cách xa khu vực xảy ra lũ quét hàng vài chục km). Do đó, cần tiến hành điều tra hàng năm về các trận lũ quét với các thông số đầy đủ bao gồm: (1) Phạm vi xảy ra và mức độ lũ quét: Đó không phải là một điểm, đó là một đoạn suối, hoặc một khu vực. Cần định lượng được phạm vi này dựa trên điều tra tại địa phương bằng tọa độ thể hiện trên GIS; (2) Lượng mưa sinh lũ quét: Đánh giá được lượng mưa giờ, lượng mưa tích lũy của đợt lũ quét nhằm phản ánh đúng nguyên nhân sinh lũ.

Ngoài ra, các nghiên cứu về lũ quét sử dụng dữ liệu địa không gian và mô hình trí tuệ nhân tạo trước đây thường sử dụng giá trị nội tại một điểm (độ dốc nội tại, chỉ số thực vật nội tại...), điều này chưa phản ánh đúng về sự hình thành lũ quét (mà phù hợp hơn đối với loại hình sạt lở đất hoặc dòng chảy sinh ra bởi sạt lở). Do đó, cần tiếp cận dựa trên nguyên tắc lưu vực đối với loại hình thiên tai lũ quét nhằm diễn tả quá trình vật lý về sự hình thành lũ. Đây cũng là những hạn chế đã tồn tại nhiều năm trong nghiên cứu lũ quét cần được nghiên cứu chuyên sâu hơn.

Qua nghiên cứu này, nhóm nghiên cứu đề xuất hướng nghiên cứu tiếp theo: **Nghiên cứu xây dựng hệ thống cảnh báo lũ quét ngắn hạn độ phân giải cao cho các lưu vực vừa và nhỏ: tích hợp radar-AI-thủy văn.**

Hướng nghiên cứu này hướng đến một hệ thống cảnh báo hoàn chỉnh cho một khu vực cụ thể, dựa vào dữ liệu radar hiện tại để dự báo lượng mưa trong khoảng 3÷6 giờ tiếp theo (sử dụng mô hình trí tuệ nhân tạo), làm đầu vào cho mô hình thủy văn (có khả năng dự báo đến từng vị trí – pixel) nhằm đánh giá lũ, mức độ lũ phục vụ cảnh báo nguy cơ lũ quét. Hệ thống này có khả năng chạy 24/7 theo thời gian thực nhằm đáp ứng nhu cầu phòng, chống thiên tai với thời gian cảnh báo sớm từ 3÷6 giờ dựa vào thời đoạn mưa dự báo và thời gian tập trung dòng chảy của mỗi lưu vực. Để thực hiện được điều này, mỗi pixel trong khu vực nghiên cứu được xem như cửa ra của một lưu vực con. Từ đó, lưu vực thượng nguồn của mỗi pixel sẽ được xác định và mô hình thủy văn dự báo lũ quét sẽ được xây dựng cho từng lưu vực này. Hệ thống này có tính khả thi cao nhờ kế thừa và phát huy các điểm mạnh của trí tuệ nhân tạo và kết hợp với mô hình thủy văn truyền thống (mô phỏng vật lý), từ đó đưa ra những cảnh báo phù hợp có độ chi tiết cao.