

**BỘ GIÁO DỤC  
VÀ ĐÀO TẠO**

**VIỆN HÀN LÂM KHOA HỌC  
VÀ CÔNG NGHỆ VIỆT NAM**

**HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ**



**Hoàng Anh Đức**

**MỘT SỐ KỸ THUẬT PHÂN LOẠI DỮ LIỆU VÀ ỨNG DỤNG  
TRONG THÀNH LẬP BẢN ĐỒ NGUY CƠ CHÁY RỪNG, SẠT  
LỎ ĐẤT**

**TÓM TẮT LUẬN ÁN TIẾN SĨ HỆ THỐNG THÔNG TIN**

**Mã số: 9 48 01 04**

**Hà Nội - 2025**

**Công trình được hoàn thành tại: Học viện Khoa học và Công nghệ,  
Viện Hàn lâm Khoa học và Công nghệ Việt Nam**

Người hướng dẫn khoa học:

Người hướng dẫn 1: PGS.TS. Đặng Văn Đức, Viện Công nghệ thông tin

Người hướng dẫn 2: PGS. TS. Lê Văn Hưng, Đại học Thủy Lợi

Phản biện 1:

Phản biện 2:

Phản biện 3:

Luận án được bảo vệ trước Hội đồng đánh giá luận án tiến sĩ cấp Học viện họp tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam vào hồi giờ , ngày tháng năm 2025.

Có thể tìm hiểu luận án tại:

1. Thư viện Học viện Khoa học và Công nghệ
2. Thư viện Quốc gia Việt Nam

**DANH MỤC CÁC BÀI BÁO ĐÃ XUẤT BẢN  
LIÊN QUAN ĐẾN LUẬN ÁN**

1. **CT1.** Lê Văn Hưng, Nguyễn Thị Thanh, Đặng Hữu Nghị, Hoàng Anh Đức, "Xây dựng mô hình học sâu đánh giá nguy cơ cháy rừng tại Lâm Đồng", Nghiên cứu cơ bản trong lĩnh vực Khoa học trái đất và Môi trường (CAREES), 2019, doi: 10.15625/vap.2019.000162.
2. **CT2.** Hung Van Le, **Duc Anh Hoang**, Chuyen Trung Tran, Phi Quoc Nguyen, Van Hai Thi Tran, Nhat Duc Hoang, Mahdis Amiri, Thao Phuong Thi Ngo, Ha Viet Nhu, Thong Van Hoang, Dieu Tien Bui, "A new approach of deep neural computing for spatial prediction of wildfire danger at tropical climate areas," Ecological Informatics, vol. 63, pp. 101300, 2021. DOI: 10.1016/j.ecoinf.2021.101300.
3. **CT3. Duc Anh Hoang**, Hung Van Le, Dong Van Pham, Hoa Viet Pham, and Dieu Tien Bui, "Hybrid BBO-DE Optimized SPAARCTree Ensemble for Landslide Susceptibility Mapping," Remote Sensing, vol. 15, no. 8, p. 2187, Apr. 2023, doi: 10.3390/rs15082187.

## GIỚI THIỆU

### 1. Tính cấp thiết của đề tài nghiên cứu

Cháy rừng và sạt lở đất là hai loại thiên tai phổ biến và nguy hiểm, gây tổn thất lớn về kinh tế, xã hội và môi trường tại Việt Nam và nhiều nơi trên thế giới. Biến đổi khí hậu làm tăng tần suất và cường độ của các hiện tượng này, tạo ra thách thức lớn cho công tác quản lý và phòng chống. Cháy rừng, mặc dù có vai trò sinh thái nhất định, có thể gây hậu quả nghiêm trọng khi xảy ra ở quy mô lớn, đặc biệt ở vùng nhiệt đới và gần khu dân cư. Sạt lở đất, một hiểm họa địa chất đáng kể, ảnh hưởng đến hàng triệu người, gây thiệt hại về sinh mạng và tài sản hàng năm. Xu hướng gia tăng mưa lớn và bão, nhất là ở vùng núi các nước đang phát triển, dự báo sẽ làm tăng số vụ lở đất trong tương lai. Trong bối cảnh đó, việc ứng dụng các kỹ thuật phân loại dữ liệu để thành lập bản đồ nguy cơ cháy rừng và sạt lở đất trở nên cấp thiết, giúp cung cấp thông tin quan trọng cho công tác dự báo, cảnh báo và quản lý rủi ro thiên tai.

Việt Nam, với vị trí địa lý nằm trong khu vực Đông Nam Á và địa hình đa dạng, được đánh giá là một trong những quốc gia dễ bị ảnh hưởng bởi thiên tai nhất trên thế giới. Địa hình đồi núi và khí hậu nhiệt đới gió mùa tạo điều kiện thuận lợi cho sự xuất hiện của cả cháy rừng và lở đất. Tình trạng này càng trở nên nghiêm trọng hơn trong bối cảnh biến đổi khí hậu toàn cầu, khi mà các hiện tượng thời tiết cực đoan ngày càng trở nên phổ biến và khó dự đoán. Điều này đặt ra thách thức lớn cho công tác quy hoạch sử dụng đất, bố trí dân cư và các hoạt động canh tác.

Trong bối cảnh đó, việc dự đoán và thành lập bản đồ phân vùng nguy cơ thiên tai đóng vai trò then chốt trong công tác quản lý và giảm thiểu rủi ro. Các phương pháp học máy kết hợp với hệ thống tin địa lý (GIS) đang mở ra những hướng tiếp cận mới, mang lại nhiều ưu điểm vượt trội so với các phương pháp truyền thống. Các nghiên cứu trước đây tại Việt Nam về dự báo cháy rừng và lở đất thường sử dụng các phương pháp thống kê truyền thống, chưa khai thác được tiềm năng của các phương pháp học máy tiên

tiên. Việc áp dụng học máy và GIS trong dự báo thiên tai giúp xác định chính xác hơn các khu vực có nguy cơ cao, từ đó hỗ trợ việc lập kế hoạch phòng ngừa, quản lý rừng hiệu quả và quy hoạch sử dụng đất an toàn.

Với những lý do trên, luận án này nhằm giải quyết nhu cầu thực tế, có ý nghĩa khoa học và thực tiễn. Mục tiêu chính là chứng minh tính hiệu quả của việc sử dụng dữ liệu viễn thám và GIS kết hợp với các kỹ thuật học máy tiên tiến trong xây dựng các mô hình dự đoán nguy cơ xảy ra thiên tai cũng như xây dựng bản đồ phân vùng nguy cơ. Kết quả của nghiên cứu này sẽ góp phần quan trọng vào việc cải thiện công tác quản lý và phòng chống thiên tai tại Việt Nam, đồng thời cung cấp một phương pháp tiếp cận mới có thể áp dụng cho các khu vực khác có điều kiện tương tự.

## **2. Mục tiêu nghiên cứu**

Mục tiêu chính: xây dựng các mô hình học máy hiệu quả để dự đoán và thành lập bản đồ phân vùng nguy cơ thiên tai.

Mục tiêu cụ thể: thu thập và xử lý dữ liệu, phát triển và huấn luyện mô hình học máy, đánh giá hiệu suất của mô hình và thành lập bản đồ phân vùng nguy cơ xảy ra thiên tai.

## **3. Nội dung nghiên cứu**

Để đạt được các mục tiêu của luận án này, luận án sẽ tiến hành các nội dung:

- Tìm hiểu các lý thuyết cơ bản về học máy, các phương pháp phân loại dữ liệu và các vấn đề liên quan đến xây dựng bản đồ phân vùng nguy cơ thiên tai, tình hình nghiên cứu về vấn đề xây dựng
- Xây dựng quy trình thu thập, tiền xử lý dữ liệu các vùng nghiên cứu
- Thu thập, xử lý và tiền xử lý dữ liệu cho các khu vực nghiên cứu, bao gồm dữ liệu về cháy rừng của tỉnh Gia Lai, Việt Nam, và dữ liệu về lở đất của huyện Than Uyên, tỉnh Lai Châu, Việt Nam.
- Tạo và huấn luyện mô hình có thể được sử dụng cho hai bộ dữ liệu.
- Đánh giá các tiêu chí, thử nghiệm và đánh giá các mô hình đề xuất.
- Biểu diễn kết quả ra bản đồ phân vùng nguy cơ cháy rừng và sạt lở đất.

## **4. Đóng góp của luận án**

- 1) Đề xuất một số kỹ thuật tổng hợp, xử lý, phân loại dữ liệu phục vụ phân vùng nguy cơ cháy rừng và sạt lở đất.
- 2) Sử dụng các kết quả phân loại kể trên để thành lập các bản đồ phân vùng nguy cơ cháy rừng ở tỉnh Gia Lai và sạt lở đất ở huyện Than Uyên, tỉnh Lai Châu.

## **5. Cấu trúc luận án**

Luận án bao gồm phần mở đầu, 3 chương chính và kết luận.

Chương 1: Tổng quan về lĩnh vực nghiên cứu

Chương 2: Mô hình xây dựng bản đồ phân vùng nguy cơ cháy rừng ở Gia Lai sử dụng Deep-NC

Chương 3: Mô hình thành lập bản đồ phân vùng nguy cơ sạt lở đất sử dụng cây tổng hợp BBO-DE-StreeEns

Cuối cùng là phần kết luận và đề xuất hướng nghiên cứu trong tương lai

## **CHƯƠNG 1 TỔNG QUAN VỀ LĨNH VỰC NGHIÊN CỨU**

### **1.1 Lý thuyết cơ bản**

#### **1.1.1 Học sâu**

Học sâu là một nhánh của học máy, lấy cảm hứng từ cấu trúc não người. Nó sử dụng mạng nơ-ron nhân tạo nhiều lớp để phân tích dữ liệu phức tạp, là nền tảng cho nhiều ứng dụng AI hiện đại. Các thành phần chính của học sâu bao gồm: mạng nơ-ron nhân tạo, trọng số và độ lệch (bias), hàm kích hoạt, lan truyền ngược, bỏ qua (dropout), CNNs, RNNs.

Ưu điểm: hiệu quả ở nhiều lĩnh vực như xử lý ngôn ngữ, nhận dạng hình ảnh. Hạn chế: cần nhiều dữ liệu và tài nguyên tính toán. Một mô hình học sâu điển hình gồm lớp đầu vào, các lớp ẩn và lớp đầu ra. Độ sâu của mô hình được xác định bởi số lớp ẩn. Có thể áp dụng kiến trúc này để lập bản đồ phân vùng nguy cơ cháy rừng. Các lớp ẩn xử lý dữ liệu đầu vào, lớp đầu ra tính xác suất của các lớp “cháy” hay “không cháy”.

#### **1.1.2 Học kết hợp**

Học kết hợp là phương pháp học máy kết hợp nhiều mô hình cơ sở để tạo ra một mô hình dự đoán tổng hợp chính xác và tối ưu hơn. Nguyên lý cơ bản là tạo ra "mô hình mạnh" từ nhóm "mô hình yếu".

Các phương pháp chính trong học kết hợp bao gồm:

- Bagging: Huấn luyện mô hình trên các tập con dữ liệu, kết hợp kết quả.
- Boosting: Huấn luyện mô hình liên tiếp, mô hình sau học từ mô hình trước.
- Stacking: Huấn luyện nhiều mô hình khác nhau, sau đó huấn luyện mô hình cuối cùng trên kết quả của chúng.
- Voting: Áp dụng nhiều mô hình song song, kết quả là đa số hoặc trung bình.
- Blending: Kết hợp mô hình cơ bản bằng trọng số tối ưu.

Ưu điểm chính của học kết hợp: Tăng độ chính xác nhờ kết hợp nhiều mô hình. Giảm overfitting nhờ trung bình hóa các độ lệch và phương sai, cũng như tăng sự ổn định của mô hình.

Tuy nhiên, phương pháp này cũng có nhược điểm như tốn tài nguyên tính toán và không phù hợp với mọi bài toán.

## **1.2 Một số mô hình học máy**

### **1.2.1 Máy vector hỗ trợ (SVM)**

SVM là một thuật toán học máy được sử dụng cho cả bài toán phân loại và hồi quy. Nó tìm ra một ranh giới tuyến tính hoặc phi tuyến tính để phân chia các điểm dữ liệu vào các nhóm khác nhau. SVM cũng có khả năng xử lý các dữ liệu không tuyến tính thông qua việc sử dụng các hàm nhân.

### **1.2.2 Máy vector liên quan (RVM)**

Tương tự như SVM, RVM cũng là một thuật toán học máy sử dụng cho phân loại và hồi quy. Tuy nhiên, RVM sử dụng một số lượng nhỏ các vector quan trọng để xác định mô hình, giúp giảm chi phí tính toán so với SVM.

### **1.2.3 Rừng ngẫu nhiên (RF)**

RF là một thuật toán học máy dựa trên việc xây dựng một tập hợp các cây quyết định. Mỗi cây được xây dựng dựa trên một tập dữ liệu con và sử dụng phương pháp "rừng ngẫu nhiên" để đưa ra dự đoán. RF có khả năng xử lý các dữ liệu không tuyến tính và cũng có tính khả diễn giải cao.

### 1.2.4 Hồi quy logistic (LR)

LR là một thuật toán học máy sử dụng cho bài toán phân loại. Nó sử dụng hàm logistic để ước lượng xác suất của việc thuộc vào một lớp nhất định. LR thường được sử dụng khi cần thông tin về xác suất

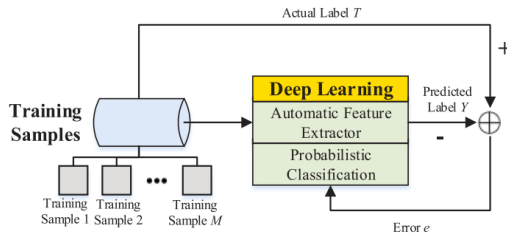
### 1.2.5 Mạng nơ-ron perceptron đa lớp (MLP)

MLP là một kiến trúc mạng nơ-ron nhân tạo với nhiều lớp ẩn. Nó được sử dụng cho các bài toán phân loại và hồi quy. MLP tổ hợp các nút nơ-ron và trọng số giữa chúng để học một mô hình tương quan giữa đầu vào và đầu ra.

### 1.2.6 Bộ phân loại giảm điểm và thuộc tính chia (SPAARC)

SPAARC Là thuật toán cây quyết định mở rộng từ CART. SPAARC tích hợp hai kỹ thuật: Lấy mẫu thuộc tính node và lấy mẫu điểm phân chia để tăng tốc độ huấn luyện mà không ảnh hưởng nhiều đến độ chính xác.

## 1.3 Các thuật toán tối ưu



Hình 1-7 Quá trình chung của việc thích ứng trọng số mạng.

Một trong những lợi thế đáng kể của phương pháp Học máy nâng cao này là khả năng suy luận các đại diện cấp cao từ các đặc trưng ban đầu. Tuy nhiên, hiệu quả của việc trích xuất đặc trưng cấp cao trong các mạng nơ-ron sâu phụ thuộc nhiều vào các thuật toán tối ưu được sử dụng để tinh chỉnh trọng số mạng, như minh họa trong Hình 1-7. Trong ngữ cảnh này, luận án đánh giá một số thuật toán tối ưu phổ biến, bao gồm SGD, RMSProp, Adam, Adadelta, BBO và Tiến hóa vi phân (DE).

### 1.3.1 Stochastic Gradient Descent (SGD)

SGD là thuật toán tối ưu thường được sử dụng trong học máy. Nó cập nhật các tham số mô hình bằng cách sử dụng một ví dụ huấn luyện hoặc một tập con nhỏ tại mỗi lần lặp.



### 1.3.2 RMSProp

RMSProp là một thuật toán tối ưu hóa gradient thích ứng được Geoffrey Hinton đề xuất năm 2012, nhằm cải thiện hiệu suất của Gradient Descent trong huấn luyện mạng nơ-ron sâu. Nó tự động điều chỉnh tốc độ học cho từng tham số dựa trên trung bình bình phương của gradient gần đây, giúp hội tụ nhanh hơn và xử lý tốt các gradient thưa thớt trong các bài toán học sâu phức tạp.

### 1.3.3 Thuật toán Ước lượng moment thích ứng (Adam)

Adam là thuật toán tối ưu tốc độ học thích ứng, thường được sử dụng để huấn luyện mạng nơ-ron sâu, được đề xuất bởi Kingma và Ba năm 2015. Adam kết hợp ưu điểm của 2 thuật toán phổ biến khác: AdaGrad và RMSProp.

### 1.3.4 Thuật toán tối ưu Adadelata

Thuật toán tối ưu Adadelata được đề xuất bởi Zeiler năm 2012 giải quyết vấn đề về tốc độ học giảm dần theo thời gian. Nó đạt được điều này bằng cách giới hạn kích thước của số tích lũy gradient quá khứ. Đặc biệt, phương pháp Adadelata loại bỏ nhu cầu thiết lập tốc độ học mặc định.

### 1.3.5 Tối ưu hóa dựa trên Địa sinh học (BBO)

BBO là một thuật toán tối ưu hóa meta-heuristic lấy cảm hứng từ quá trình di cư của các loài sinh vật giữa các đảo sinh học. BBO mô phỏng sự phân bố và di chuyển của các loài dựa trên các yếu tố như chỉ số thích nghi của môi trường sống (HSI) và chỉ số phù hợp của loài (SIV), từ đó tìm ra giải pháp tối ưu cho các bài toán tối ưu hóa phức tạp trong nhiều lĩnh vực khác nhau.

### 1.3.6 Tiến hóa vi phân

DE là một thuật toán tối ưu hóa tiến hóa mạnh mẽ được phát triển bởi Storn và Price vào năm 1997, sử dụng các phép toán đột biến, lai ghép và chọn lọc để tìm kiếm giải pháp tối ưu trong không gian tìm kiếm đa chiều. DE nổi bật với khả năng xử lý hiệu quả các bài toán tối ưu hóa phi tuyến và đa mục tiêu, đồng thời dễ dàng triển khai với ít tham số điều chỉnh, làm cho nó trở thành một công cụ phổ biến trong nhiều lĩnh vực ứng dụng.

## 1.4 Cấu trúc nghiên cứu

### 1.4.1 Quy trình của nghiên cứu

Quy trình thành lập bản đồ phân vùng nguy cơ xảy ra thiên tai gồm 11 bước:

Bước 1: Thu thập dữ liệu

- Thu thập dữ liệu từ các cơ quan quản lý rừng, phòng chống thiên tai địa phương về các vụ cháy rừng và sạt lở đất trong quá khứ.
- Thu thập dữ liệu từ ảnh vệ tinh, mô hình số độ cao (DEM), bản đồ địa chất, dữ liệu khí tượng thủy văn.
- Thực hiện khảo sát thực địa để thu thập thông tin bổ sung về thảm thực vật, cấu trúc đất.

Bước 2: Tiền xử lý dữ liệu

- Sử dụng phần mềm GIS để xử lý dữ liệu không gian.
- Sử dụng ArcPy, mã nguồn Python để điền các giá trị còn thiếu.
- Chuẩn hóa tất cả các biến số về cùng một thang đo.

Bước 3: Phân tích và lựa chọn các yếu tố ảnh hưởng

- Thực hiện phân tích riêng cho cháy rừng và sạt lở để lựa chọn các yếu tố ảnh hưởng, bao gồm 12 yếu tố ảnh hưởng đến khả năng cháy rừng và 10 yếu tố ảnh hưởng đến khả năng sạt lở đất.
- Xác định các yếu tố quan trọng nhất cho mỗi loại thiên tai, sử dụng thuật toán Average Impurity Decrease (AID) để đánh giá tầm quan trọng của các yếu tố ảnh hưởng đến khả năng cháy rừng và thuật toán wrapper kết hợp với phương pháp cross-validation với năm lần chia dữ liệu để đánh giá vai trò của 10 yếu tố ảnh hưởng.

Bước 4: Xây dựng CSDL GIS

- Tạo geodatabase cho từng loại thiên tai ở hai vùng nghiên cứu.
- Tạo các lớp bản đồ riêng biệt cho từng yếu tố ảnh hưởng

Bước 5: Xây dựng và huấn luyện mô hình

- Thiết kế mô hình cho cháy rừng (Deep-NC) và sạt lở đất (BBO-DE StreeEns)
- Chia dữ liệu thành tập huấn luyện và tập kiểm tra cho mỗi mô hình
- Huấn luyện các mô hình.

Bước 6: Đánh giá và tinh chỉnh mô hình

- Đánh giá hiệu suất của cả hai mô hình
- Tinh chỉnh từng mô hình riêng biệt để đạt hiệu suất tốt nhất
- So sánh hiệu suất của hai mô hình với các mô hình tham chiếu

Bước 7: Áp dụng mô hình để dự đoán

- Sử dụng mô hình cháy rừng trên Geodatabase của tỉnh Gia Lai để dự đoán nguy cơ cháy rừng
- Sử dụng mô hình sạt lở đất trên Geodatabase của huyện Than Uyên để dự đoán nguy cơ sạt lở đất
- Lưu kết quả dự đoán của cả hai mô hình

Bước 8: Phân loại mức độ nguy cơ

- Xác định ngưỡng phân loại riêng cho cháy rừng và sạt lở đất
- Phân loại kết quả dự đoán thành các mức độ nguy cơ cho mỗi loại thiên tai, sử dụng phương pháp phân lớp ngắt tự nhiên (Natural Break)

Bước 9: Tạo bản đồ phân vùng nguy cơ

- Tạo bản đồ nguy cơ cháy rừng
- Tạo bản đồ nguy cơ sạt lở đất

Bước 10: Kiểm chứng và đánh giá bản đồ:

- So sánh kết quả dự đoán với dữ liệu thực tế cho cả hai loại thiên tai
- Thực hiện kiểm chứng thực địa cho nguy cơ cháy rừng tại tỉnh Gia Lai và sạt lở đất cho huyện Than Uyên, Lai Châu.

Bước 11: Hoàn thiện và trình bày bản đồ:

- Tạo bản đồ phân vùng nguy cơ cháy rừng tại tỉnh Gia Lai và bản đồ phân vùng nguy cơ sạt lở đất tại huyện Than Uyên, Lai Châu.
- Thêm chú thích, tỷ lệ, và các thông tin cần thiết khác cho bản đồ.

#### **1.4.2 Công tác chuẩn bị dữ liệu**

Các bước chuẩn bị dữ liệu cho nghiên cứu, bao gồm:

##### **1.4.2.1 Thu thập dữ liệu đối với bộ dữ liệu cháy rừng của tỉnh Gia Lai**

- Dữ liệu cháy rừng: 2530 vị trí cháy từ 2007-2016 từ CSDL của Bộ Nông nghiệp và Phát triển Nông thôn (MARD) tại <http://www.kiemlam.org.vn>

- Dữ liệu địa hình: Bản đồ địa hình tỷ lệ 1:50.000 từ Bộ Tài nguyên và Môi trường (MONRE)
  - Ảnh vệ tinh: Landsat-8 OLI năm 2016, thu thập từ EarthExplorer (<http://earthexplorer.usgs.gov>)
  - Dữ liệu khí hậu (2007-2014): Nhiệt độ, tốc độ gió, độ ẩm tương đối, lượng mưa. Nguồn thu thập từ Trung tâm Quốc gia về Thông tin Môi trường (NCEI), Website: <https://www.ncdc.noaa.gov/>
  - Dữ liệu sử dụng đất: Bản đồ sử dụng đất tỉnh Gia Lai tỷ lệ 1:50.000 bao gồm 11 hạng mục sử dụng đất, nguồn từ Dự án kiểm kê đất đai quốc gia năm 2013 của Tổng cục Quản lý Đất đai
- Nghiên cứu đã kết hợp các nguồn dữ liệu đa dạng này để xây dựng CSDL GIS và phân tích nguy cơ cháy rừng cho tỉnh Gia Lai.

#### **1.4.2.2 Thu thập dữ liệu sạt lở đất của huyện Than Uyên**

- Dữ liệu sạt lở đất: 970 điểm sạt lở toàn tỉnh Lai Châu, trong đó 114 điểm tại huyện Than Uyên được sử dụng. Nguồn thu thập từ Dự án quốc gia do Viện Khoa học và Kỹ thuật Địa chất Việt Nam thực hiện từ 2012
- Bản đồ địa hình tỷ lệ 1:50.000 của tỉnh Lai Châu độ phân giải 20x20m cho huyện Than Uyên
- Bản đồ địa chất và vết đứt gãy lấy từ Viện Địa chất Khoáng sản Việt Nam
- Bản đồ thổ nhưỡng lấy từ Viện Khoa học Nông nghiệp Việt Nam (VAAS)
- Bản đồ đường giao thông và phân bố sông suối lấy nguồn từ Bộ Tài nguyên và Môi trường (MONRE)

#### **1.4.2.3 Tiền xử lý dữ liệu**

Tiền xử lý dữ liệu cháy rừng tại Gia Lai bao gồm các bước chính:

- Chuyển đổi dữ liệu phi cấu trúc, bao gồm các bước: Thu thập từ nhiều nguồn, số hóa tài liệu giấy, tổ chức thành file CSV, tích hợp vào hệ thống GIS, kết nối với bản đồ số, mã hóa tình trạng cháy rừng (dạng 0/1), xuất ra dạng shapefile
- Xử lý dữ liệu thiếu: Sử dụng Python và SimpleImputer() trong Scilearn được dùng để thay thế giá trị NULL

- Tạo đặc trưng: Sử dụng ArcGIS tạo các lớp độ dốc, hướng dốc, độ cao, độ cong bề mặt từ DEM; Tính các chỉ số NDVI, NDWI, NDMI từ ảnh vệ tinh.
  - Đáp ứng yêu cầu mô hình: Mã hóa dữ liệu phân loại thành số
  - Chuẩn hóa dữ liệu: Sử dụng scaler từ -1 đến 1 cho tất cả đặc trưng
  - Cân bằng dữ liệu: Chọn ngẫu nhiên 2530 điểm không cháy rừng
  - Chia dữ liệu: 70% (3542 mẫu) cho huấn luyện, 30% (1518 mẫu) cho kiểm tra
- Tiền xử lý dữ liệu sạt lở đất tại huyện Than Uyên, Lai Châu bao gồm các bước:
- Sử dụng công cụ Clip trong ArcGIS Pro để quy các dữ liệu về diện tích trong ranh giới huyện Than Uyên.
  - Xây dựng mô hình số độ cao DEM bằng Topo to Raster.
  - Tạo bản đồ các yếu tố ảnh hưởng (độ cao, độ dốc, độ cong, độ nghiêng, độ chênh cao) từ DEM, sử dụng các công cụ Spatial Analysis của ArcGIS Pro.
  - Sử dụng bản đồ Thổ nhưỡng Quốc gia để tạo lớp đặc trưng về Loại đất và Địa chất.
  - Tạo lớp đặc trưng khoảng cách đến đứt gãy, đường, sông bằng công cụ Distance.
  - Xác định 114 vị trí sạt lở tại Than Uyên.
  - Cân bằng dữ liệu bằng cách lấy mẫu ngẫu nhiên 114 điểm không sạt lở.
  - Chia dữ liệu 228 điểm thành tập huấn luyện (70%) và tập kiểm tra (30%) bằng `train_test_split()`.
  - Phân lớp và chuẩn hóa dữ liệu bằng Reclassify và Raster Calculator.

## 1.5 Đánh giá mô hình

### 1.5.1 Các phương pháp đánh giá

- Phương pháp kiểm định Hold-Back: Chia dữ liệu thành tập huấn luyện và tập kiểm định để huấn luyện và đánh giá mô hình. Còn gọi là phân chia Train/Test hoặc Train/Validate.
- Kiểm tra chéo (Cross-Validation): Chia dữ liệu thành k phân đoạn và huấn luyện mô hình trên k-1 phân đoạn, sau đó đánh giá trên phân đoạn còn lại. Lặp lại k lần để đảm bảo mỗi phân đoạn được dùng 1 lần để đánh giá.

- Kiểm tra chéo phân tầng (Stratified Cross-Validation): Tương tự như kiểm tra chéo nhưng đảm bảo tỷ lệ lớp là như nhau trong mỗi phân đoạn dữ liệu.
- Phương pháp Bootstrapping: Sinh ra nhiều tập huấn luyện từ tập dữ liệu gốc bằng cách lấy mẫu ngẫu nhiên có hoàn lại.

### 1.5.2 Các chỉ số đánh giá

Đối với bài toán phân loại nhị phân, NCS sử dụng các chỉ số đánh giá sau:  
Ma trận nhầm lẫn (Confusion Matrix): Biểu diễn ma trận của các dự đoán và kết quả thực tế, với các giá trị TP, TN, FP, FN.

Độ chính xác (Accuracy): Tỷ lệ các dự đoán chính xác trên tổng số mẫu.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}};$$

Độ chính xác dương (Precision - PPV): Tỷ lệ TP/tổng các dự đoán dương.

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}};$$

Giá trị dự đoán âm (NPV): Tỷ lệ TN trong tổng số các dự đoán âm.

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}};$$

Độ nhạy (Sensitivity): Tỷ lệ TP trong tổng số các mẫu dương thực sự.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}};$$

Độ đặc hiệu (Specificity): Tỷ lệ TN trong tổng số các dự đoán âm.

$$\text{Spe} = \frac{\text{TN}}{\text{TN} + \text{FP}};$$

F1-score: Trung bình hài hòa của Precision và Recall.

$$\text{Fscore} = \frac{2 \times \text{PPV} \times \text{Sen}}{\text{PPV} + \text{Sen}} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}};$$

Đường cong ROC (Receiver Operating Characteristic) và AUC (Area Under the Curve): Các thước đo đo lường và số liệu về hiệu suất mô hình với các ngưỡng phân loại khác nhau.

Hệ số Kappa: Thước đo thống kê mức độ đồng thuận giữa hai đánh giá viên cho một tập các mục tiêu.

$$\text{Kappa} = \frac{2 \times (\text{TP} \times \text{TN} - \text{FN} \times \text{FP})}{(\text{TP} + \text{FP}) \times (\text{FP} + \text{TN}) + (\text{TP} + \text{FN}) \times (\text{FN} + \text{TN})}$$

MSE và RMSE: Các chỉ số đánh giá thường dùng trong các bài toán hồi quy.



- Địa hình: độ dốc, hướng dốc, độ cao, độ cong.
- Môi trường: mục đích sử dụng đất, chỉ số NDVI, NDWI, NDMI.
- Khí hậu: nhiệt độ, tốc độ gió, độ ẩm, lượng mưa.

Các yếu tố này được biểu diễn dưới dạng các lớp bản đồ số.

#### **2.1.4 Xây dựng Geodatabase về cháy rừng tại tỉnh Gia Lai**

CSDL GIS được xây dựng bao gồm 12 yếu tố ảnh hưởng và bản đồ phân bố cháy rừng. Tổng cộng 5060 bản ghi được tiền xử lý để đưa vào mô hình. Quá trình xử lý và tiền xử lý dữ liệu được nhắc đến ở mục 1.6 của luận án.

### **2.2 Mô hình Deep-NC**

#### **2.2.1 Lựa chọn mô hình**

Các mô hình SVM, RVM, Random Forest và Deep-NC đã được tác giả đánh giá bằng việc chạy thử trên cùng bộ dữ liệu bằng phương pháp Trial and Error. Kết quả cho thấy Deep-NC có hiệu suất tốt nhất nên đã được lựa chọn cho nghiên cứu này.

#### **2.2.2 Đánh giá tầm quan trọng của các yếu tố**

Phương pháp Average Impurity Decrease (AID) được sử dụng để đánh giá tầm quan trọng trong dự đoán của các yếu tố ảnh hưởng đến khả năng cháy rừng. Kết quả cho thấy NDVI, NDWI và NDMI là quan trọng nhất.

#### **2.2.3 Hàm mục tiêu huấn luyện mô hình Deep-NC**

Việc huấn luyện mô hình Deep-NC liên quan đến quá trình tìm kiếm và cập nhật trọng số của mô hình để giảm thiểu sự khác biệt giữa các điểm có nguy cơ cháy rừng và thực tế. Một hàm mục tiêu cần thiết để định lượng sự khác biệt này. Trong nghiên cứu này, tác giả luận án chọn Mean Squared Error (MSE) làm hàm mục tiêu, được mô tả như sau:

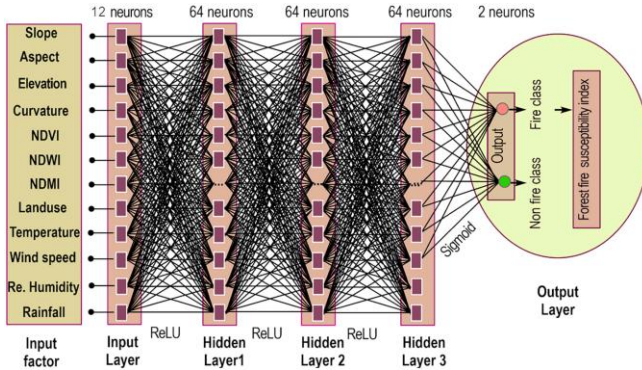
$$L(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Trong đó, 'N' đại diện cho số lượng mẫu dữ liệu, trong khi 'y<sub>i</sub>' và 'y<sub>hat</sub><sub>i</sub>' lần lượt là giá trị thực tế và tính toán cho mẫu 'i'.

#### **2.2.4 Kiến trúc mô hình Deep-NC**

Kiến trúc mô hình Deep-NC (Xem Hình 2-5) gồm 5 lớp:





Hình 2-5: Cấu trúc mô hình Deep-NC được sử dụng trong nghiên cứu Lớp đầu vào (với 12 nơ-ron) tương ứng với 12 yếu tố ảnh hưởng đến cháy rừng, mỗi yếu tố được biểu diễn bởi một nơ-ron riêng biệt, các lớp ẩn (3 lớp, mỗi lớp có 64 nơ-ron): Các lớp ẩn này học các đại diện phức tạp và phi tuyến từ các yếu tố đầu vào. Mỗi nơ-ron trong các lớp ẩn sử dụng hàm kích hoạt ReLU (Rectified Linear Unit). Hàm ReLU là một hàm kích hoạt phi tuyến phổ biến trong các mô hình học sâu, với công thức:

$$f(x) = \max(0, x)$$

ReLU sẽ trả về  $x$  nếu  $x > 0$  và trả về 0 nếu  $x \leq 0$ . ReLU thường được sử dụng trong các lớp ẩn vì nó giúp giảm hiện tượng biến mất gradient trong quá trình lan truyền ngược.

Lớp đầu ra (2 nơ-ron): Lớp này bao gồm 2 nơ-ron; mỗi nơ-ron đại diện cho một lớp được phân vùng: "Lớp không cháy" và "Lớp cháy". Hàm kích hoạt Sigmoid được sử dụng trong lớp này để chuyển đổi đầu ra của mô hình thành xác suất.

$$f(x) = 1/(1+e^{(-x)})$$

Để tinh chỉnh trọng số của mô hình Deep-NC, tác giả sử dụng các thuật toán tối ưu như SGD, RMSProp, Adam và Adadelta. Mô hình Deep-NC giúp xác định ranh giới quyết định tách bản đồ khu vực nghiên cứu thành hai loại riêng biệt – "không cháy rừng" và "cháy rừng". Sau đó, kết quả tính toán từ mô hình Deep-NC có thể được chuyển thành định dạng raster để phân tích sâu hơn bằng phần mềm ArcGIS.

### 2.3 Đánh giá mô hình

Mô hình Deep-NC gồm 5 lớp với 206 neuron, cho hiệu suất cao trong dự đoán cháy rừng. Mô hình sử dụng thuật toán Adam để tối ưu hóa 9.294 trọng số. Khi chạy mô hình trên bộ dữ liệu kiểm tra, kết quả AUC của Deep-NC được so sánh với mô hình SVM (AUC = 0,786), mô hình RVM (AUC = 0,793), và mô hình Random Forest (AUC = 0,790). Mô hình Deep-NC đề xuất, với giá trị AUC 0,894, cho thấy hiệu quả vượt trội.

Bảng 2-1 Ma trận nhầm lẫn của các mô hình RVM, SVM, RandomForest và Deep-NC với Bộ dữ liệu huấn luyện và Bộ dữ liệu kiểm tra

Chỉ số	RVM		SVM		Random Forest		Deep-NC	
	Bộ dữ liệu huấn luyện	Bộ dữ liệu kiểm tra	Bộ dữ liệu huấn luyện	Bộ dữ liệu kiểm tra	Bộ dữ liệu huấn luyện	Bộ dữ liệu kiểm tra	Bộ dữ liệu huấn luyện	Bộ dữ liệu kiểm tra
TP	1760	344	1822	359	1748	340	1742	673
TN	1494	270	1372	252	1432	259	1637	564
FP	358	68	296	53	370	72	29	86
FN	624	142	746	160	686	153	134	195

Bảng 2-4 Hiệu suất của các mô hình RVM, SVM, RandomForest và Deep-NC với Bộ dữ liệu huấn luyện và Bộ dữ liệu kiểm tra

Chỉ số	RVM		SVM		Random Forest		Deep-NC	
	Bộ dữ liệu huấn luyện	Bộ dữ liệu xác thực	Bộ dữ liệu huấn luyện	Bộ dữ liệu xác thực	Bộ dữ liệu huấn luyện	Bộ dữ liệu xác thực	Bộ dữ liệu huấn luyện	Bộ dữ liệu xác thực
PPV(%)	83.10	83.50	82.53	82.52	86.02	87.14	98.36	88.7
NPV(%)	70.54	65.53	67.61	62.86	64.78	61.17	92.43	74.3
Sens(%)	73.83	70.78	71.82	68.97	70.95	69.17	92.86	77.5
Spe(%)	80.67	79.88	79.47	78.25	82.25	82.62	98.26	86.8
Acc(%)	76.82	74.51	75.07	72.69	75.40	74.15	95.40	81.5
AUC	0.842	0.793	0.830	0.790	0.813	0.786	0.983	0.894
Kappa	0.536	0.490	0.501	0.454	0.508	0.483	0.908	0.630

## 2.4 Đánh giá mô hình Deep-NC với các thuật toán tối ưu hoá khác nhau

Bảng 2-7 Ma trận nhầm lẫn của mô hình Deep-NC sử dụng 4 thuật toán tối ưu Adam, SGD, RMSprop và AdaDelta trên 10 trường hợp lấy mẫu ngẫu nhiên của tập kiểm tra

Chỉ số	Chỉ số với thuật toán Adam				Chỉ số với thuật toán SGD				Chỉ số với thuật toán RMSprop				Chỉ số với thuật toán AdaDelta			
	Min	Max	Mean	STD	Min	Max	Mean	STD	Min	Max	Mean	STD	Min	Max	Mean	STD
TP	570	662	623	35.6	612	719	669	26.9	493	698	588	66.1	540	729	635	54.3
TN	537	661	605	31.9	450	546	491	32.6	419	602	528	57.5	405	565	499	47.8
FP	97	189	137	35.6	40	147	90.2	26.9	61	266	171	66.1	30	219	124	54.3
FN	98	222	154	31.9	213	309	269	31.5	157	340	231	57.5	194	354	260	47.8

Bảng 2-7 trình bày Ma trận nhầm lẫn của Deep-NC với 4 thuật toán tối ưu trên 10 mẫu ngẫu nhiên. False Negative (FN) - số trường hợp mô hình dự đoán sai là không cháy rừng - là chỉ số quan trọng, thể hiện khả năng bỏ sót các trường hợp cháy rừng. AUC được sử dụng làm chỉ số chính để so sánh hiệu suất dự đoán của mô hình Deep-NC với các thuật toán tối ưu khác nhau. Vì vậy, mô hình Deep-NC tối ưu bằng Adam vượt trội hơn các mô hình khác trong việc dự đoán cháy rừng, với giá trị AUC trung bình cao nhất là 0.893.

Bảng 2-8 Hiệu suất phân lớp cho mô hình Deep-NC sử dụng 4 thuật toán tối ưu Adam, SGD, RMSprop và AdaDelta trên mười trường hợp lấy mẫu ngẫu nhiên của Test set

Chỉ số	Chỉ số với thuật toán Adam				Chỉ số với thuật toán SGD				Chỉ số với thuật toán RMSprop				Chỉ số của 10 mẫu thử với thuật toán AdaDelta			
	Min	Max	Mean	STD	Min	Max	Mean	STD	Min	Max	Mean	STD	Min	Max	Mean	STD
PPV(%)	75.1	87.2	82	4.69	80.6	94.7	88.1	3.55	65	92	77.5	8.7	71.2	96.1	83.7	7.15
NPV(%)	70.8	87.1	79.7	4.21	59.3	71.9	64.6	4.19	55.2	79.3	69.5	7.6	53.4	74.4	65.8	6.3
Sens(%)	72	86.8	80.2	3.79	69.1	74.2	71.4	1.82	67.2	75.9	72.2	2.6	67.3	73.6	71.2	1.99
Spe(%)	74	86.5	81.7	4.32	78.8	91.8	84.8	3.26	69.4	87.3	76.6	6	72.1	93.1	81.1	6.02
Acc(%)	72.9	85.8	80.8	3.74	74.1	78.3	76.4	1.14	72.1	74.5	73.5	0.9	72.8	76.8	74.7	0.96
AUC	0.82	0.93	0.89	0.03	0.81	0.86	0.84	0.01	0.81	0.84	0.82	0.01	0.82	0.85	0.83	0.01

## 2.5 Thành lập bản đồ phân vùng nguy cơ cháy rừng



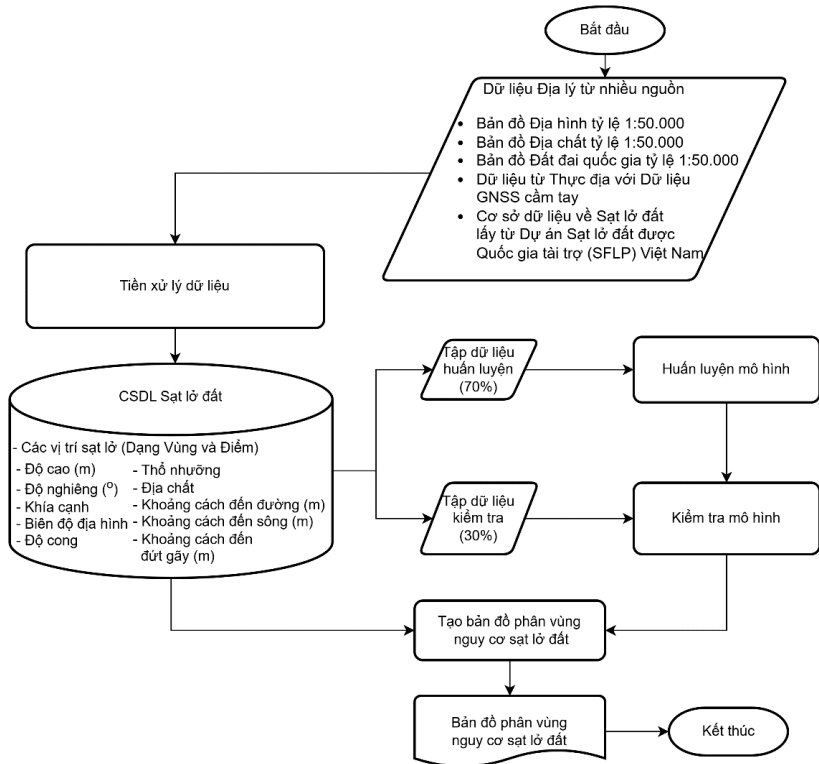


sạt lở đất ở huyện Than Uyên, Việt Nam, thay thế cho Deep-NC không hiệu quả với bộ dữ liệu nhỏ (228 bản ghi). Mô hình mới kết hợp:

- Subbagging và Không gian phụ ngẫu nhiên để tạo bộ dữ liệu phụ.
- Thuật toán SPAARC để xây dựng cây quyết định, giảm 70% thời gian xây dựng so với Random Forest.
- Tối ưu hóa siêu tham số bằng kết hợp BBO và DE (DE/BBO).

Phương pháp này nhằm cải thiện hiệu suất, tốc độ xử lý và độ chính xác so với thuật toán Rừng ngẫu nhiên truyền thống, đồng thời khắc phục hạn chế của Deep-NC với bộ dữ liệu nhỏ.

### 3.2.2 Quy trình thành lập bản đồ nhạy cảm nguy cơ sạt lở đất bằng mô hình BBO-DE-StreeEns



Hình 3-2 Quy trình xây dựng mô hình đánh giá nguy cơ sạt lở đất

Để xây dựng mô hình đánh giá nguy cơ sạt lở đất, ta thực hiện các bước trong Hình 3-6 như sau:

- Thu thập dữ liệu 10 yếu tố ảnh hưởng: độ cao, độ dốc, độ cong, hướng dốc, chênh cao, thổ nhưỡng, địa chất, khoảng cách đến đứt gãy, đường và sông (Xem phần 1.6.2.2 trong báo cáo toàn văn)

- Xây dựng CSDL sạt lở:

- + Trích xuất giá trị 10 yếu tố từ bản đồ raster

- + Chuyển đổi dữ liệu rời rạc thành giá trị số cho các yếu tố tương ứng

Sau các bước thu thập bộ dữ liệu, tiền xử lý dữ liệu, CSDL sạt lở đất được hình thành với 10 yếu tố ảnh hưởng và vị trí các vụ sạt lở đất, được tạo thành một tập gồm 228 điểm (bao gồm 114 điểm sạt lở và 114 điểm không sạt lở). Dữ liệu này được chia thành 2 tập dữ liệu huấn luyện (70%) và thẩm định (30%). Ở bước này sẽ tiến hành các bước để xây dựng mô hình BBO-DE-StreeEns đề xuất (Hình 3-7). Trong đó, các bước như sau:

Bước 1: Chia tập dữ liệu thành 2 tập huấn luyện (70%) và kiểm tra (30%)

Bước 2: Khởi tạo các siêu tham số ban đầu của thuật toán BBO và DE. Các siêu tham số này bao gồm tỷ lệ đột biến, tỷ lệ chéo, kích thước quần thể, số lần lặp, giá trị ràng buộc các tham số tối ưu hóa.

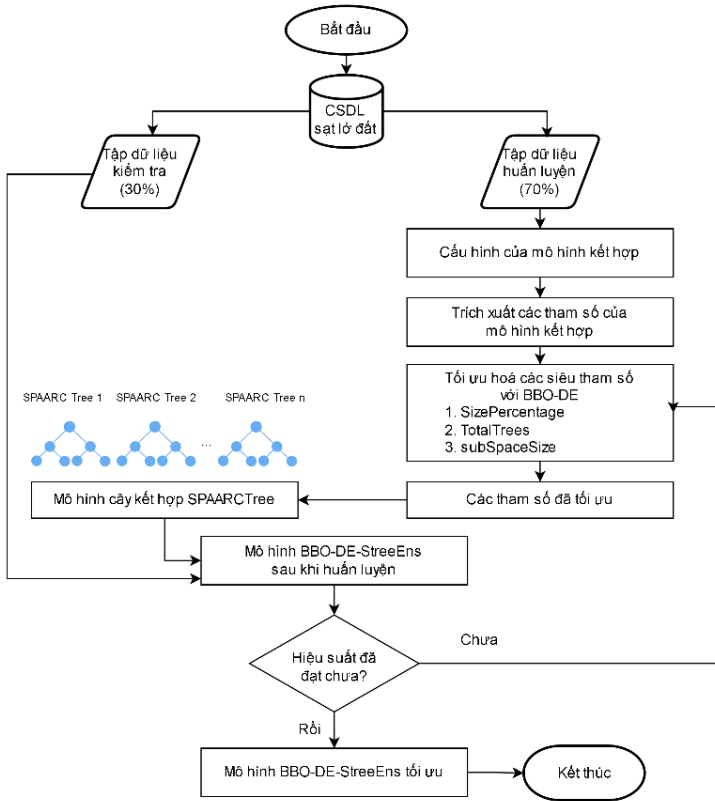
Bước 3: Tối ưu hóa 3 siêu tham số của mô hình gồm TotalTrees, SizePercentage và SubSpaceSize bằng thuật toán lai BBO-DE. Các siêu tham số này được sử dụng để tính toán hàm mục tiêu MAE trên tập huấn luyện.

Bước 4: Đánh giá mô hình trên tập kiểm tra. Nếu kết quả đạt sẽ dừng việc tối ưu hóa. Ngược lại, điều chỉnh các siêu tham số.

Bước 5: Lặp lại Bước 3, 4 cho đến khi tìm được bộ siêu tham số tối ưu.

Bước 6: Mô hình ở bước này là tối ưu, các siêu tham số có thể được sử dụng trong mô hình để tạo ra bản đồ nguy cơ sạt lở đất ở bước tiếp theo.

Các thông số trên được trích xuất vào mô hình tổ hợp, sau đó BBO-DE sẽ tối ưu hóa ba siêu tham số: SizePercentage, TotalTrees và subSpaceSize. Trong nghiên cứu này, chúng là: SizePercentage=0,9 (90%), TotalTrees = 30 và subSpaceSize = 0,5 (50%).



Hình 3-7 Kiến trúc của mô hình đề xuất BBO-DE StreeEns

### 3.2.3 Tác động của các yếu tố ảnh hưởng đến sạt lở đất

Đánh giá đóng góp của 10 yếu tố ảnh hưởng đến trượt lở trong mô hình BBO-DE-StreeEns bằng phương pháp wrapper và 5 lần cross-validation. Kết quả: độ dốc ảnh hưởng lớn nhất (0,299), sau đó là khoảng cách đến đường (0,224) và độ cao (0,142). Các yếu tố còn lại ít ảnh hưởng hơn (0,026-0,084).

### 3.2.4 Hàm chi phí và các siêu tham số

Để đạt hiệu năng tốt nhất, mô hình BBO-DE-StreeEns phụ thuộc vào việc chọn 3 siêu tham số: TotalTrees, SizePercentage và SubSpaceSize.

Trong nghiên cứu, tối ưu hóa các siêu tham số này bằng kỹ thuật lai BBO-DE, sử dụng hàm mục tiêu là Mean Absolute Error (MAE).



$$MAE = \frac{1}{n} \sum_{i=1}^n |LS_i - \widehat{LS}_i|$$

Trong đó:  $LS_i$ : giá trị nhạy cảm trượt lở dự đoán cho mẫu thứ  $i$ .  
 $\widehat{LS}_i$ : giá trị thực tế tương ứng.  
 $n$ : tổng số mẫu trong tập huấn luyện.

### 3.2.5 Kết quả đánh giá

Bảng 3-5 Ma trận nhầm lẫn của mô hình BBO-DE-StrreeEns được đề xuất cùng các mô hình tham chiếu trên tập dữ liệu kiểm tra.

Mô hình	TP	TN	FN	FP
BBO-DE-STreeEns	28	31	6	3
Lregr	24	28	10	6
MLPNeuNet	26	18	8	16
SPAARC	28	29	6	5
SVM	24	28	10	6

Bảng 3-6 Các chỉ số đánh giá hiệu suất của mô hình BBO-DE-STreeEns được đề xuất và các mô hình tham chiếu trên tập dữ liệu kiểm tra

Mô hình	Các chỉ số hiệu suất							
	PPV (%)	NPV (%)	Sen (%)	Spe (%)	Acc (%)	Fscore	Kappa	AUC
BBO-DE-STreeEns	90.3	83.8	82.4	91.2	86.8	0.862	0.735	0.940
Lregr	80.0	73.7	70.6	82.4	76.5	0.750	0.529	0.853
MLPNeuNet	61.9	69.2	76.5	52.9	64.7	0.684	0.294	0.748
SPAARC	84.8	82.9	82.4	85.3	83.8	0.836	0.676	0.915
SVM	80.0	73.7	70.6	82.4	76.5	0.750	0.529	0.767

5 mô hình (BBO-DE-StrreeEns, LRegr, MLPNeuNet, SVM, SPAARC) được huấn luyện bằng kiểm định chéo 10 lần trên 228 điểm (114 trượt lở, 114 không trượt lở). Trên tập huấn luyện, mô hình đề xuất (với siêu tham số tối ưu: TotalTrees=30, SizePercentage=0.9, subSpaceSize=0.5) và SPAARC cho kết quả tốt nhất về các chỉ số (AUC, Kappa, F-score, Accuracy).

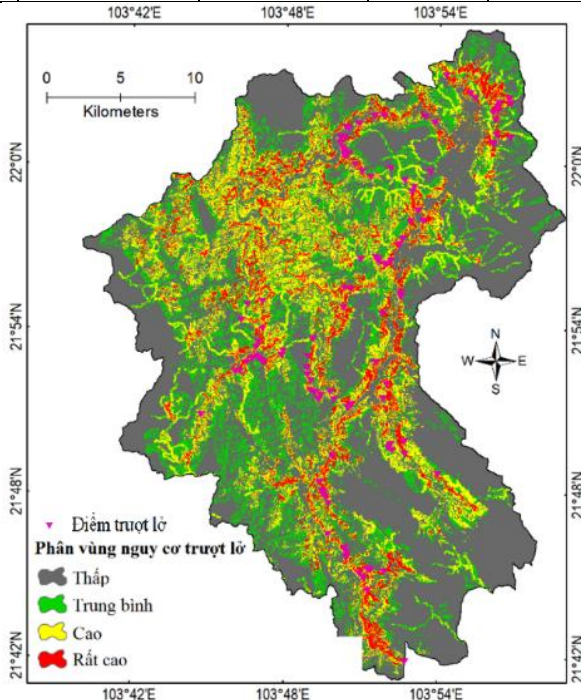
Trên tập kiểm tra, BBO-DE-StrreeEns và SPAARC vẫn duy trì khả năng dự đoán tốt nhất. Các mô hình LRegr, SVM, MLPNeuNet cho kết quả kém hơn

trên cả tập huấn luyện và kiểm định. Như vậy, mô hình BBO-DE-StreeEns đề xuất là tốt nhất trong các mô hình được khảo sát.

### 3.3.6 Tạo bản đồ phân vùng nguy cơ sạt lở đất cho huyện Than Uyên

Bảng 3-8 Các chỉ số của 4 lớp cấp độ nguy cơ sạt lở đất của mô hình đề xuất

STT	Cấp nguy cơ	Khoảng giá trị chỉ số nguy cơ	Tỷ lệ diện tích(%)	Diện tích (km <sup>2</sup> )	Số điểm trượt lở	Tỷ lệ điểm trượt lở (%)
1	Thấp	0,062–0,508	50,00	394,5	12	10,53
2	TB	0,508–0,606	20,00	157,8	7	6,14
3	Cao	0,606–0,737	20,00	157,8	23	20,17
4	Rất cao	0,737–0,910	10,00	78,9	72	63,16
<b>Tổng cộng</b>			<b>100</b>	<b>789</b>	<b>114</b>	<b>100</b>



Hình 3-9 Bản đồ nhạy cảm nguy cơ sạt lở đất huyện Than Uyên sử dụng mô hình BBO-DE-StreeEns

Mô hình BBO-DE-StreeEns được chọn để tính chỉ số nhạy cảm trượt lở (0,062-0,910) cho từng pixel ở Than Uyên. Bản đồ nhạy cảm trượt lở được

tạo bằng cách phân chỉ số thành 4 cấp sử dụng phương pháp Natural Break: rất cao, cao, trung bình và thấp. Ngưỡng các cấp độ xác định qua phân tích đồ thị tỷ lệ % diện tích trượt lở/bản đồ nhạy cảm. Phân bố: rất cao 50% (0,737-0,910), cao 20% (0,674-0,737), trung bình 20% (0,502-0,674), thấp 10% (0,062-0,502). Cấp rất cao và cao chiếm 83,33% điểm trượt lở. Bản đồ cuối cùng dựa trên các ngưỡng và phân cấp này.

### **KẾT LUẬN**

Luận án đã trình bày quá trình thu thập và xử lý dữ liệu từ tỉnh Gia Lai và huyện Than Uyên, tỉnh Lai Châu, đồng thời đề xuất và chứng minh hiệu quả của hai mô hình quan trọng. Mô hình Deep-NC được áp dụng trong dự báo nguy cơ cháy rừng, vượt trội hơn các mô hình RVMs, SVMs, và RF, với thuật toán tối ưu Adam cho hiệu suất tốt nhất. Đối với bản đồ khả năng sạt lở đất, mô hình BBO-DE-StreeEns, kết hợp Cây SPAARC và tối ưu hóa hybrid BBO-DE, đã thể hiện hiệu suất vượt trội, xác định độ dốc và khoảng cách đến đường là yếu tố quan trọng nhất. Cả hai mô hình đều chứng minh hiệu quả trong việc phân lớp nguy cơ cháy rừng và sạt lở ở Việt Nam, với BBO-DE-StreeEns còn nổi bật về tốc độ đào tạo nhanh và độ chính xác cao.

### **HƯỚNG NGHIÊN CỨU TRONG TƯƠNG LAI**

Hướng nghiên cứu trong tương lai bao gồm việc tối ưu hóa cấu trúc mô hình Deep-NC và thuật toán huấn luyện, bổ sung dữ liệu như lượng mưa vào mô hình BBO-DE-StreeEns, và kiểm định khả năng áp dụng rộng rãi của các mô hình trên các khu vực và kịch bản khác nhau. Ngoài ra, nghiên cứu cũng hướng đến việc tích hợp Deep-NC và BBO-DE-StreeEns thông qua học kết hợp hoặc học đa nhiệm, phát triển Học trực tuyến để cập nhật kết quả theo thời gian thực, và mở rộng nghiên cứu sang các tỉnh khác. Trong quá trình này, cần xem xét đặc thù riêng về địa hình, khí hậu, thảm thực vật và mối quan hệ giữa các yếu tố ảnh hưởng của từng địa phương, đồng thời so sánh và điều chỉnh mô hình dựa trên sự tương đồng hoặc khác biệt giữa các khu vực.