

Bài báo khoa học

# Xây dựng bản đồ phân vùng nhạy cảm ngập lụt sử dụng mô hình Light Gradient Boosting Machine

Nguyễn Thanh Tuấn<sup>1</sup>, Vũ Cao Đạt<sup>1\*</sup>, Nguyễn Đức Đảm<sup>1</sup>, Phạm Thái Bình<sup>1</sup>

<sup>1</sup> Trường Đại học Công nghệ GTVT, 54 Triều Khúc, Thanh Xuân, Hà Nội, Việt Nam;  
datvc@utt.edu.vn; damnd@utt.edu.vn; tuannt94@utt.edu.vn; binhpt@utt.edu.vn

\*Tác giả liên hệ: datvc@utt.edu.vn; Tel.: +84-384026586

Ban Biên tập nhận bài: 15/12/2022; Ngày phản biện xong: 22/1/2023; Ngày đăng bài: 25/1/2023

**Tóm tắt:** Mục tiêu chính của nghiên cứu này là xây dựng bản đồ phân vùng nhạy cảm ngập lụt khu vực huyện Quảng Trạch, tỉnh Quảng Bình (Việt Nam) sử dụng mô hình *Light Gradient Boosting Machine LGBM* – một trong những mô hình trí tuệ nhân tạo hiệu quả trong đánh giá và dự báo thiên tai. Cơ sở dữ liệu bao gồm 173 vị trí ngập lụt trong quá khứ và 07 tham số thành phần (mưa, địa mạo, độ bao phủ mặt đất, độ cao, hình dáng bề mặt địa hình, góc mái dốc, và hướng mái dốc) đã được thu thập để xây dựng dữ liệu đào tạo (70%) và dữ liệu kiểm tra (30%) dùng cho xây dựng và kiểm chứng mô hình. Độ chính xác của mô hình được đánh giá thông qua nhiều thông số thống kê định lượng bao gồm diện tích dưới đường cong ROC (AUC). Kết quả nghiên cứu thể hiện rằng LGBM có độ chính xác cao trong dự báo và phân vùng nhạy cảm ngập lụt tại khu vực nghiên cứu (AUC = 0,96 cho dữ liệu đào tạo và AUC = 0,88 cho dữ liệu kiểm tra). Bản đồ phân vùng nhạy cảm ngập lụt xây dựng từ mô hình có độ chính xác cao có thể được dùng trong việc nâng cao hiệu quả quy hoạch sử dụng đất và đưa ra các giải pháp phù hợp để giảm thiểu tác hại gây ra bởi ngập lụt tại khu vực nghiên cứu.

**Từ khóa:** Ngập lụt; Light Gradient Boosting Machine; Quảng Trạch; Việt Nam.

## 1. Giới thiệu

Việt Nam chịu ảnh hưởng nghiêm trọng từ biến đổi khí hậu gây ra các hiện tượng thời tiết cực đoan như mưa lớn kéo dài làm xuất hiện nhiều loại thiên tai nguy hiểm trong đó có ngập lụt [1]. Ngập lụt có thể gây ra nhiều tác hại như phá hủy tài sản, gây ô nhiễm môi trường, gây ra cái chết của nhiều người. Dữ liệu thống kê của Ban Chỉ đạo Quốc gia về phòng chống thiên tai cho thấy trong 20 năm trở lại đây ở Việt Nam các loại thiên tai như bão, lũ, lở đất... đã gây thiệt hại tài sản trên 6,4 tỷ USD, khiến hơn 13 nghìn người thiệt mạng, khoảng hơn 70% dân số và 60% diện tích đất bị ảnh hưởng từ thiên tai [1]. Vì vậy, nghiên cứu và xây dựng các công cụ đánh giá và dự báo ngập lụt là một trong những nhiệm vụ quan trọng nhằm giảm thiểu các tác động xấu gây ra bởi ngập lụt.

Xây dựng bản đồ phân vùng nhạy cảm ngập lụt là một trong những công việc hữu ích trong việc nâng cao hiệu quả quy hoạch sử dụng đất nhằm giảm thiểu tác động tiêu cực gây ra bởi ngập lụt [2]. Bản đồ này thể hiện sức phân cấp các khu vực có mức độ nhạy cảm (xác suất xảy ra ngập lụt) từ cao đến thấp. Có nhiều nghiên cứu trong những năm gần đây nhằm phân vùng nhạy cảm ngập lụt ở nhiều khu vực trên thế giới như Trung Quốc [3], Nhật Bản [4], Đài Loan [5], Iran [6]. Ở Việt Nam, một số nghiên cứu đã được thực hiện để phân vùng nhạy cảm ngập lụt [7–11]. Hầu hết các nghiên cứu này sử dụng cách tiếp cận dựa trên các

mô hình trí tuệ nhân tạo trong phân tích và dự báo nguy cơ ngập lụt tại các khu vực được lựa chọn để nghiên cứu. Kết quả từ các nghiên cứu này cho thấy các mô hình trí tuệ nhân tạo có hiệu quả và là các công cụ tiềm năng trong việc đánh giá và phân vùng nhạy cảm ngập lụt.

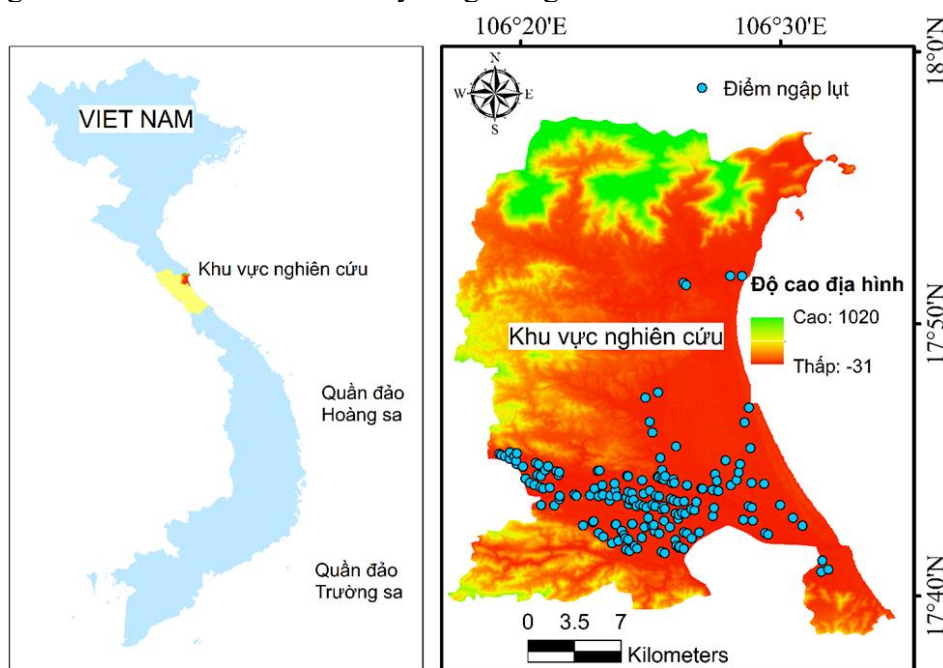
Quảng Bình là tỉnh duyên hải miền Trung của Việt Nam có đặc điểm về khí hậu và địa hình rất phức tạp, nơi đây cũng là địa phương chịu ảnh hưởng nặng nề của hầu hết các loại thiên tai hàng năm trong đó có ngập lụt. Vì vậy, mục tiêu chính của nghiên cứu này là nhằm xây dựng bản đồ phân vùng nhạy cảm ngập lụt tại huyện Quảng Trạch, tỉnh Quảng Bình sử dụng một trong những kỹ thuật trí tuệ nhân tạo hiệu quả có tên là *Light Gradient Boosting Machine (LightGBM)* phục vụ cho việc nâng cao hiệu quả của công tác lập quy hoạch sử dụng đất, giảm thiểu tác hại gây ra bởi ngập lụt tại khu vực này. Việc kiểm chứng độ chính xác của mô hình LGBM được thực hiện bằng việc sử dụng nhiều kỹ thuật đánh giá trong đó có kỹ thuật đường cong ROC. Các công cụ như ArcGIS và nguồn mã mô hình mở trong phần mềm Python được sử dụng để xây dựng cơ sở dữ liệu và mô hình hóa bài toán.

## 2. Dữ liệu và phương pháp nghiên cứu

### 2.1. Đặc điểm của khu vực nghiên cứu

Huyện Quảng Trạch có diện tích 448 km<sup>2</sup>. Với địa thế trải dài từ 106°15' đến 106°59' kinh độ đông và 17°42' đến 17°59' vĩ độ bắc, địa hình của khu vực đa dạng, có cả vùng đồi núi và trung du chiếm đến 65% diện tích của huyện (Hình 1). Huyện Quảng Trạch nằm ở phía bắc tỉnh Quảng Bình, Phía bắc giáp thị xã Kỳ Anh và huyện Kỳ Anh, tỉnh Hà Tĩnh. Phía đông giáp Biên Đông. Phía tây giáp huyện Tuyên Hóa. Phía nam giáp thị xã Ba Đồn. Địa chất của huyện Quảng Trạch được tạo nên bởi một số tầng đá, đặc biệt là đá trầm tích hóa học, các loại đất sét, đá đỏ và đất đỏ.

Huyện thuộc khu vực nhiệt đới gió mùa khô nên có sự phân hóa sâu sắc của địa hình và chịu ảnh hưởng khí hậu chuyển tiếp giữa miền Bắc và Nam. Khí hậu có 2 mùa rõ rệt là mùa mưa và mùa khô, mùa mưa kéo dài từ tháng 9 tới tháng 3 năm sau. Độ ẩm khá cao ở mức 83–84%. Lượng mưa trung bình năm đạt ngưỡng 2.000–2.3000 mm/năm, lượng mưa lớn nhất vào tháng 9, 10, 11. Mức nhiệt trung bình năm tầm 25°C tới 26°C, có xu hướng tăng dần theo hướng Bắc vào Nam và chiều từ Tây sang Đông.



**Hình 1.** Vị trí địa lý của khu vực huyện Quảng Trạch, tỉnh Quảng Bình và các vị trí ngập lụt trong quá khứ.

## 2.2. Giới thiệu về mô hình LGBM

LGBM là một mô hình học máy – trí tuệ nhân tạo, được thiết kế để tăng tốc độ huấn luyện và giảm chi phí bộ nhớ bằng cách sử dụng các thuật toán tối ưu hóa để tối ưu hóa các quyết định trên cây quyết định. LGBM cũng cung cấp các tính năng tối ưu hóa cho các biến liên tục và rời rạc, giúp nâng lực dự báo mô hình tốt hơn phù hợp với các tập dữ liệu phức tạp. LGBM có thể được sử dụng để giải quyết các bài toán phân loại và hồi quy. LGBM đã được sử dụng thành công trong nhiều bài toán khác nhau như phát hiện gian lận nhấp chuột trực tuyến [12], dự báo rủi ro tài chính doanh nghiệp [13], nguy cơ an toàn thực phẩm [14]. Trong nghiên cứu này, LGBM được sử dụng để dự báo và đánh giá phân vùng ngập cảm ngập lụt; là bài toán phân loại nhị phân; trong đó có 02 biến đầu ra “ngập lụt” được mã hóa là “1” và “không ngập lụt” được mã hóa là “0”.

## 2.3. Các kỹ thuật đánh giá độ chính xác của mô hình

Các kỹ thuật đánh giá độ chính xác của mô hình như đường cong ROC, các chỉ số thống kê như độ chính xác (ACC), kappa, độ nhạy (SST), độ đặc hiệu (SPF), giá trị dự đoán dương (PPV), giá trị dự đoán âm (NPV), sai số tuyệt đối trung bình (MAE), căn của sai số toàn phương trung bình gốc (RMSE) được sử dụng để đánh giá độ chính xác của mô hình LGBM. Kỹ thuật đường cong ROC là một công cụ để đánh giá hiệu năng của một mô hình phân loại. Đường cong ROC đại diện cho sự tương quan giữa tỷ lệ False Positive (FP) và True Positive (TP) trong mô hình phân loại. Trong đó, FR là số lượng của các mẫu được phân loại sai trong số các mẫu thực tế là “âm”, còn tỷ lệ TP là số lượng của các mẫu được phân loại đúng trong số các mẫu thực tế là “dương”. Diện tích dưới đường cong ROC (AUC) được sử dụng để đánh giá định lượng hiệu năng của mô hình phân loại. Giá trị AUC càng lớn, mô hình phân loại càng tốt.  $AUC = 1$  cho thấy mô hình phân loại hoàn toàn chính xác, còn  $AUC = 0,5$  cho thấy mô hình phân loại không chính xác. Các giá trị ACC, SST, SPF, PPV, NPV, MAE, RMSE, and kappa được xác định theo các công thức được trình bày chi tiết trong nghiên cứu [15]. Nhìn chung, giá trị ACC, SST, SPF, PPV, NPV và kappa cao thể hiện năng lực dự báo của mô hình là tốt và ngược lại. Giá trị MAE và RMSE càng thấp thể hiện năng lực dự báo của mô hình càng tốt.

## 2.4. Dữ liệu nghiên cứu

### 2.4.1. Hiện trạng ngập lụt

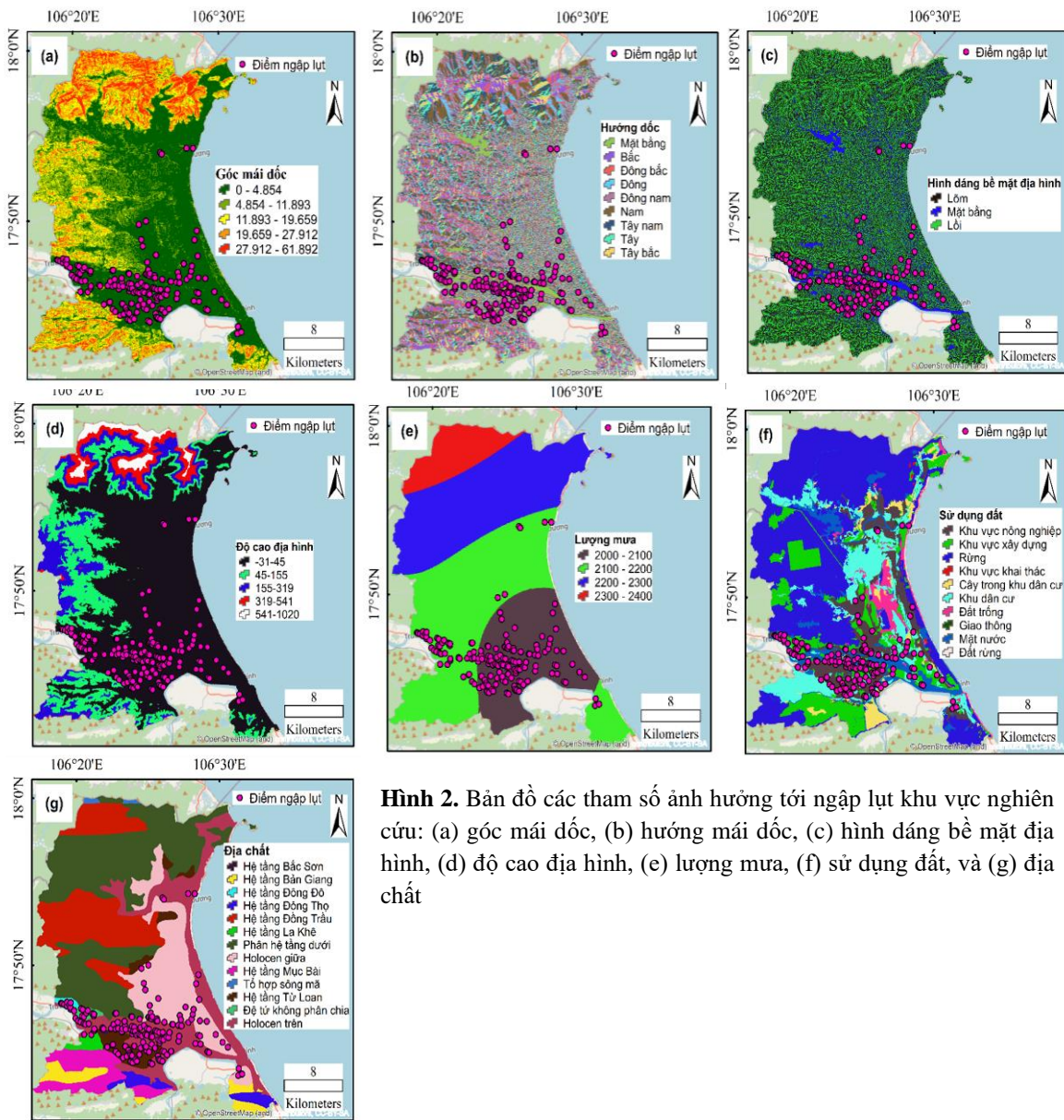
Hiện trạng ngập lụt bao gồm các vị trí ngập lụt trong quá khứ là dữ liệu quan trọng trong đánh giá và dự báo phân vùng ngập cảm ngập lụt [16]. Trong nghiên cứu này, có tổng cộng 173 vị trí ngập lụt đã được thu thập từ Trung tâm dự báo khí tượng Quảng Bình năm 2007, 2010, 2016 để xây dựng bản đồ hiện trạng ngập lụt. Bài toán dự báo phân vùng ngập cảm ngập lụt là bài toán phân loại nhị phân, vì vậy, 173 vị trí không ngập lụt cũng được nhận diện và sử dụng như là biến phân loại cùng với các vị trí ngập lụt. Dữ liệu ngập lụt và không ngập lụt được kết hợp và chia ngẫu nhiên để xây dựng dữ liệu đào tạo (70%) và dữ liệu kiểm chứng (30%) dùng cho xây dựng mô hình và kiểm chứng mô hình.

### 2.4.2. Các tham số ảnh hưởng tới ngập lụt

Căn cứ vào phân tích bản chất của quá trình ngập lụt trong quá khứ tại khu vực nghiên cứu và các nghiên cứu có liên quan [7–9], trong nghiên cứu này, chúng tôi lựa chọn 07 tham số điển hình ảnh hưởng tới quá trình ngập lụt tại khu vực nghiên cứu bao gồm: góc mái dốc, hướng mai dốc, hình dáng bề mặt địa hình, độ cao địa hình, mưa, độ bao phủ mặt đất, và địa chất. Bản đồ các tham số địa hình như góc mái dốc, hướng mai dốc, hình dáng bề mặt địa hình, độ cao địa hình được trích xuất từ mô hình số độ cao DEM (30m) thu thập từ nguồn dữ liệu mở của Hội địa chất công trình Mỹ (<https://earthexplorer.usgs.gov>), bản đồ độ bao phủ



mặt đất được xây dựng từ Bộ Tài nguyên và Môi trường, bản đồ mưa xây dựng từ dữ liệu mưa trung bình năm thu thập từ mười trạm đo ở tỉnh Quảng Bình trong khoảng thời gian từ 30 năm từ năm 1986 đến 2016, bản đồ địa chất được thu thập và xây dựng từ sở tài nguyên môi trường tỉnh Quảng Bình. Các bản đồ tham số thành phần được chuyển đổi sang định dạng raster và sample với dữ liệu hiện trạng để xây dựng cơ sở dữ liệu dùng cho mô hình dự báo phân vùng ngập cảm ngập lụt (Hình 2).



**Hình 2.** Bản đồ các tham số ảnh hưởng tới ngập lụt khu vực nghiên cứu: (a) góc mái dốc, (b) hướng mái dốc, (c) hình dáng bề mặt địa hình, (d) độ cao địa hình, (e) lượng mưa, (f) sử dụng đất, và (g) địa chất

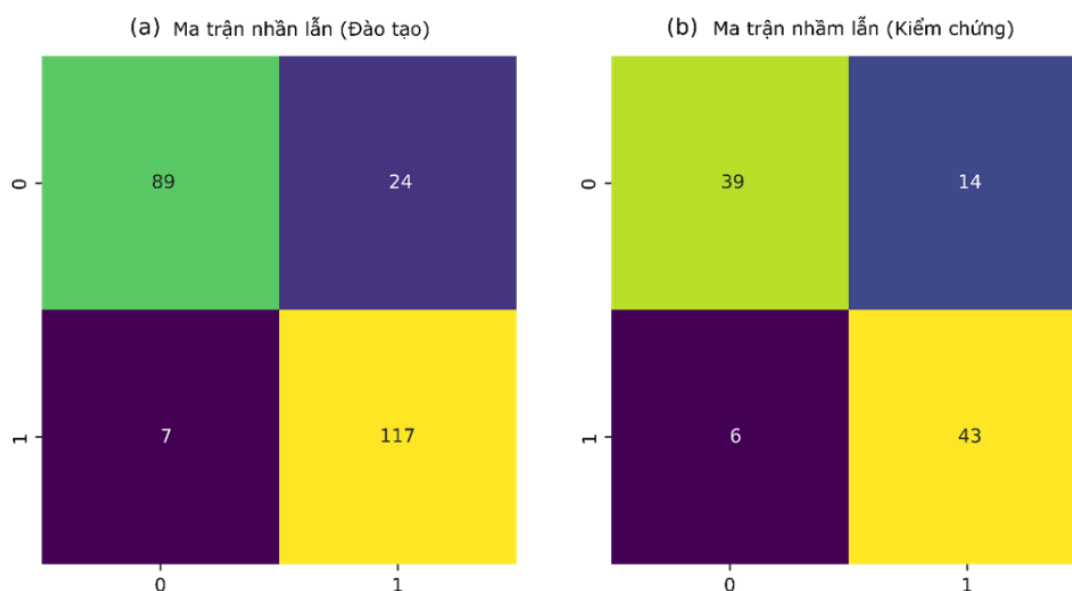
### 3. Kết quả và thảo luận

#### 3.1. Kiểm chứng độ chính xác của mô hình dự báo

Kết quả kiểm chứng độ chính xác của mô hình LGBM được thể hiện trên Hình 3, Bảng 1, Hình 4 và Hình 5. Hình 3 thể hiện các giá trị của ma trận lỗi cho tập dữ liệu đào tạo và kiểm chứng. Đối với tập dữ liệu đào tạo, có 93 dữ liệu không ngập lụt “0” được dự báo đúng và 20 dữ liệu dự báo sai trong khi đó có 121 dữ liệu ngập lụt “1” được dự báo đúng và 03 dữ liệu dự báo sai. Với dữ liệu kiểm chứng, có 41 dữ liệu không ngập lụt “0” được dự báo đúng và 12 dữ liệu dự đoán sai trong khi đó có 42 dữ liệu ngập lụt “1” dự báo đúng và 07 dữ liệu dự đoán sai. Sử dụng dữ liệu của ma trận lỗi, các giá trị PPV, NPV, SST, SPF, ACC được

xác định và thể hiện trong Bảng 1. Kết quả kiểm chứng mô hình trên tập dữ liệu đào tạo cho thấy các giá trị PPV, NPV, SST, SPF, ACC lần lượt là 82,301%; 97,581%; 96,875%; 85,516%; 90,295%; trong khi đó các giá trị này là 77,358%; 87,5%; 87,234%; 77,778%; 82,187% cho tập dữ liệu kiểm chứng. Giá trị chỉ số kappa của mô hình LGBM sử dụng dữ liệu đào tạo và dữ liệu kiểm chứng lần lượt là 0,804 và 0,628. Hình 4 thể hiện giá trị lỗi của mô hình và Bảng 1 thể hiện các giá trị lỗi RMSE và MAE của mô hình. Giá trị RMSE của mô hình LGBM sử dụng dữ liệu đào tạo và dữ liệu kiểm chứng lần lượt là 0,312 và 0,432. Giá trị MAE của mô hình LGBM sử dụng dữ liệu đào tạo và dữ liệu kiểm chứng lần lượt là 0,097 và 0,168. Hình 5 thể hiện kết quả phân tích đường cong ROC và các giá trị diện tích dưới đường cong (AUC) của mô hình LGBM. Kết quả phân tích thể hiện rằng giá trị AUC của mô hình LGBM là 0,96 và 0,88 lần lượt cho tập dữ liệu đào tạo và dữ liệu kiểm chứng.

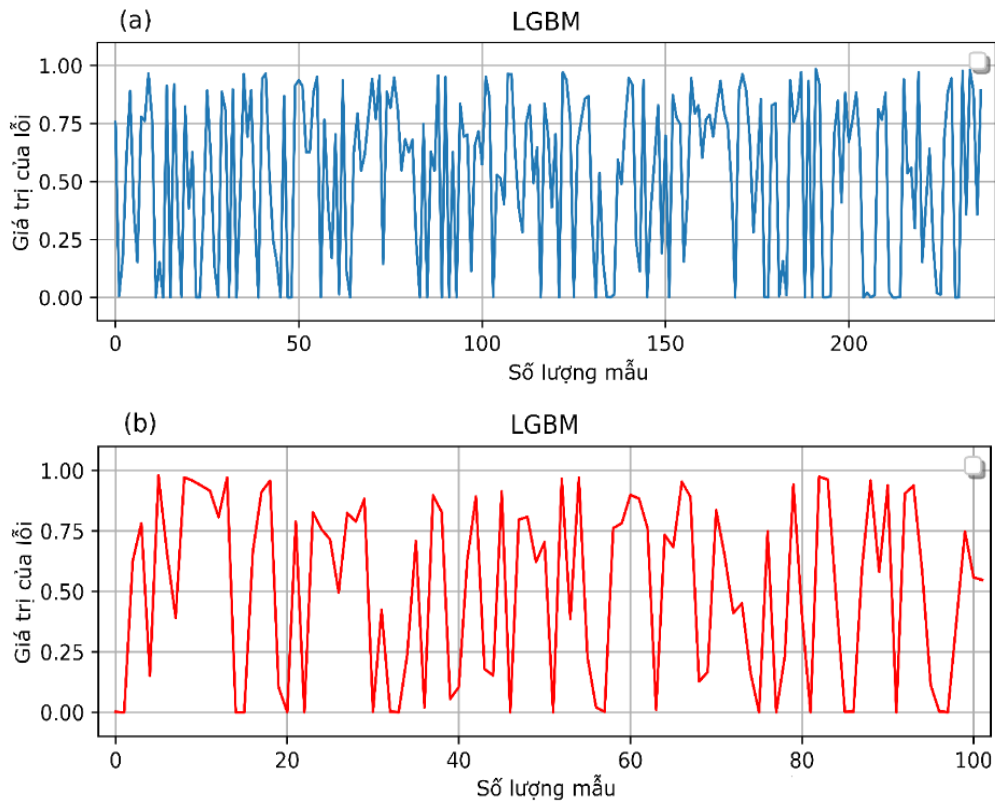
Nhìn chung, kết quả kiểm chứng độ chính xác của mô hình LGBM cho thấy mô hình LGBM có độ chính xác cao trong phân tích và mô hình hóa độ nhạy cảm ngập lụt tại khu vực nghiên cứu. Kết quả này là phù hợp do mô hình LGBM là một trong những mô hình trí tuệ nhân tạo có năng lực dự báo và phân tích tốt với nhiều ưu điểm như sau [17, 18] (i) LGBM sử dụng thuật toán tối ưu để tối ưu tốc độ học máy, giúp cho việc huấn luyện và dự đoán nhanh trên một tập dữ liệu lớn, (ii) LGBM có hiệu suất học tốt với độ chính xác cao so với các mô hình khác như XGBoost hoặc Gradient Boosting Decision Tree (GBDT), (iii) LGBM có thể học từ một lượng lớn dữ liệu và đặc trưng, giúp cho việc dự đoán chính xác hơn, (iv) LGBM có thể học từ cả dữ liệu liên tục và dữ liệu cố định, giúp cho việc phân tích dữ liệu đa kiểu. Kết quả so sánh độ chính xác của mô hình LGBM trong nghiên cứu này và các nghiên cứu dự báo ngập lụt ở khu vực khác cho thấy mô hình LGBM trong nghiên cứu này tương đương với mô hình LGBM được sử dụng trong nghiên cứu khác như Aydin, Iban [19] (AUC = 0,8978).



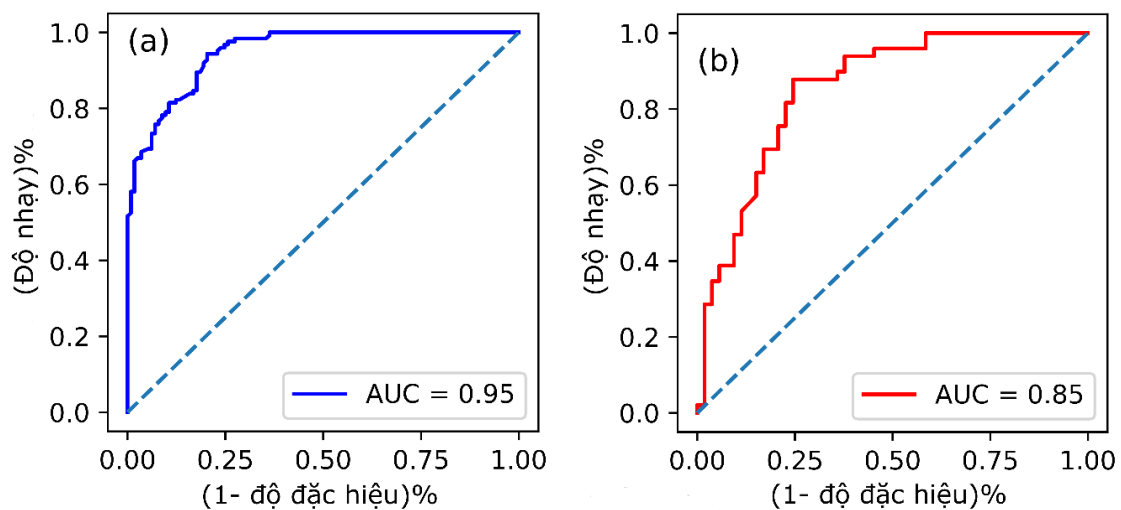
**Hình 3.** Ma trận lỗi của mô hình LGBM.

**Bảng 1.** Kết quả phân tích độ chính xác của mô hình LGBM.

Các chỉ số đánh giá	PPV (%)	NPV (%)	SST (%)	SPF (%)	ACC (%)	kappa	RMSE	MAE
Dữ liệu đào tạo	82,301	97,581	96,875	85,816	90,295	0,804	0,312	0,097
Dữ liệu kiểm chứng	77,358	87,500	87,234	77,778	82,178	0,628	0,432	0,168



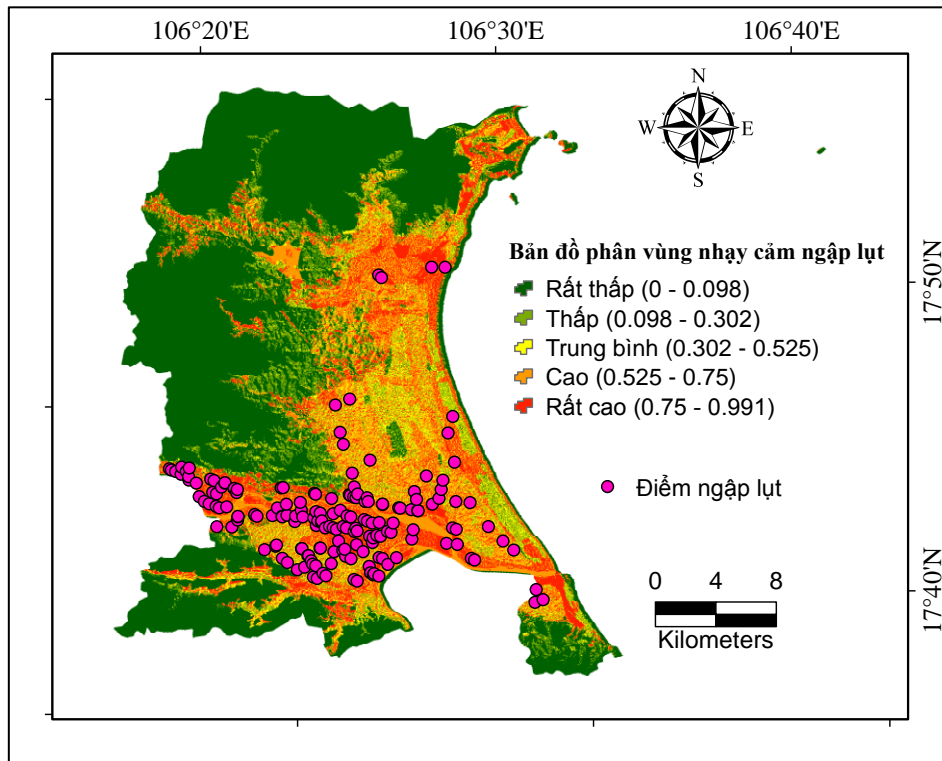
**Hình 4.** Kết quả phân tích lỗi của mô hình LGBM sử dụng: (a) Dữ liệu đào tạo; (b) Dữ liệu kiểm chứng.



**Hình 5.** Kết quả phân tích đường cong ROC sử dụng mô hình LGBM.

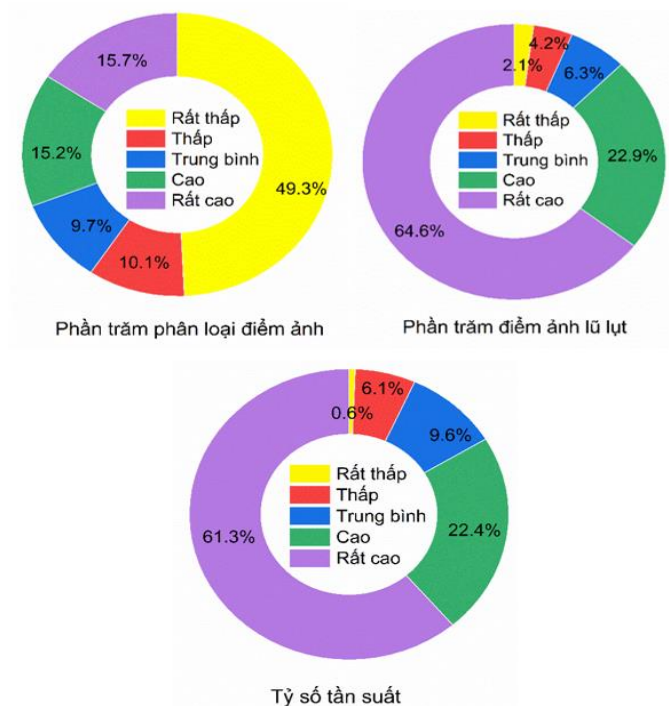
### 3.2. Xây dựng bản đồ nhạy cảm ngập lụt tại Quảng Trạch, Quảng Bình

Bản đồ nhạy cảm ngập lụt tại Quảng Trạch, Quảng Bình được xây dựng sử dụng kết quả đào tạo từ mô hình LGBM và được thể hiện trên Hình 6. Giá trị xác suất xảy ra ngập lụt cho toàn bộ các điểm ảnh trên bản đồ được tính toán và xác định thông qua việc kiểm chứng mô hình LGBM trên toàn bộ dữ liệu của khu vực nghiên cứu. Các giá trị này có giá trị giao động trong khoảng từ 0 đến 1, giá trị gần 0 thể hiện xác suất thấp xảy ra ngập lụt và giá trị 1 thể hiện xác suất xảy ra ngập lụt. Các giá trị này được phân chia thành 5 khoảng giá trị sử dụng phương pháp phân loại điểm ngẫu nhiên được tích hợp trong công cụ ArcGIS tương ứng với nó là 5 lớp nhạy cảm ngập lụt được hiển thị trên bản đồ bao gồm: rất thấp, thấp, trung bình, cao và rất cao (Hình 6).



**Hình 6.** Bản đồ phân vùng nhạy cảm ngập lụt khu vực Huyện Quảng Trạch sử dụng mô hình LGBM.

Hình 7 thể hiện kết quả kiểm chứng độ chính xác của bản đồ phân vùng nhạy cảm ngập lụt sử dụng phân tích tỷ số tần suất (FR) là tỷ số giữa số lần xuất hiện của các điểm ngập lụt trong quá khứ so với số điểm ảnh của các lớp nhạy cảm tương ứng. Kết quả cho thấy rằng, các lớp nhạy cảm cao và rất cao có tỷ số tần suất cao nhất thể hiện rằng tần suất xuất hiện ngập lụt ở các lớp cao và rất cao là lớn nhất. Kết quả này chứng tỏ bản đồ phân vùng nhạy cảm ngập lụt được xây dựng có độ chính xác tốt và có thể được sử dụng trong việc lập quy hoạch sử dụng đất và quản lý hiệu quả thiên tai ngập lụt tại khu vực nghiên cứu.



**Hình 7.** Kết quả đánh giá độ chính xác của bản đồ phân vùng nhạy cảm ngập lụt.



#### 4. Kết luận

Nghiên cứu xây dựng bản đồ nhạy cảm ngập lụt tại khu vực Huyện Quảng Trạch, tỉnh Quảng Bình đã được thực hiện trong nghiên cứu này sử dụng mô hình trí tuệ nhân tạo LGBM. Có tổng cộng 173 vị trí ngập lụt trong quá khứ và 07 tham số nguyên nhân (góc mái dốc, hướng mái dốc, độ cong địa hình, độ cao địa hình, mưa, địa mạo, sử dụng đất) được thu thập để xây dựng cơ sở dữ liệu không gian cho mô hình phân tích. Các kỹ thuật đánh giá phổ biến như đường cong ROC, các chỉ số đánh giá độ chính xác đã được lựa chọn để đánh giá độ chính xác của mô hình phân loại LGBM.

Kết quả phân tích và đánh giá thể hiện rằng mô hình LGBM có độ chính xác cao trong dự báo và phân vùng nhạy cảm ngập lụt tại khu vực nghiên cứu ( $AUC = 0,96$  cho dữ liệu đào tạo và  $AUC = 0,88$  cho dữ liệu kiểm chứng). Vì vậy, có thể kết luận rằng mô hình LGBM là công cụ hữu ích và tiềm năng trong đánh giá và phân vùng nhạy cảm ngập lụt và có thể áp dụng để phân tích và đánh giá cho các khu vực khác chịu ảnh hưởng bởi ngập lụt khi xem xét các yếu tố và đặc điểm địa môi trường cụ thể cho từng khu vực nghiên cứu. Hạn chế của nghiên cứu này là chưa xem xét tới các yếu tố liên quan đến mạng lưới dòng chảy ảnh hưởng tới khả năng xảy ra ngập lụt, và nguồn gốc phát sinh lũ tại khu vực nghiên cứu.

**Đóng góp của tác giả:** Xây dựng ý tưởng nghiên cứu: P.T.B., V.C.Đ., N.Đ.Đ., N.T.T.; Xử lý số liệu: N.Đ.Đ., N.T.T., V.C.Đ.; Chạy mô hình: V.C.Đ., N.Đ.Đ.; Viết bản thảo bài báo: P.T.B., V.C.Đ., N.Đ.Đ., N.T.T.; Chỉnh sửa bài báo: P.T.B., V.C.Đ., N.T.T., N.Đ.Đ.

**Lời cảm ơn:** Nghiên cứu này được tài trợ bởi Quỹ Phát triển khoa học và công nghệ Quốc gia (NAFOSTED) trong đề tài mã số 105.08–2019.03.

**Lời cam đoan:** Tập thể tác giả cam đoan bài báo này là công trình nghiên cứu của tập thể tác giả, chưa được công bố ở đâu, không được sao chép từ những nghiên cứu trước đây; không có sự tranh chấp lợi ích trong nhóm tác giả.

#### Tài liệu tham khảo

1. Tùng, T. Thiên tai bất thường, gây thiệt hại lớn. Bộ tài nguyên và môi trường, 2022.
2. Roy, P.; Pal, S.C.; Chakraborty, R.; Chowdhuri, I.; Malik, S.; Das, B. Threats of climate and land use change on future flood susceptibility. *J. Cleaner Prod.* **2020**, *272*, 122757.
3. Zhao, G.; Pang, B.; Xu, Z.; Yue, J.; Tu, T. Mapping flood susceptibility in mountainous areas on a national scale in China. *Sci. Total Environ.* **2018**, *615*, 1133–1142.
4. Chen, H.; Ito, Y.; Sawamukai, M.; Tokunaga, T. Flood hazard assessment in the Kujukuri Plain of Chiba Prefecture, Japan, based on GIS and multicriteria decision analysis. *Nat. Hazards* **2015**, *78*, 105–120.
5. Khoirunisa, N.; Ku, C.Y.; Liu, C.Y. A GIS-based artificial neural network model for flood susceptibility assessment. *Int. J. Environ. Res. Public Health* **2021**, *18*(3), 1072.
6. Rahmati, O.; Pourghasemi, H.R.; Zeinivand, H. Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. *Geocarto. Int.* **2016**, *31*(1), 42–70.
7. Pham, B.T.; Luu, C.; Dao, D.V.; Phong, T.V.; Nguyen, H.D.; Le, H.V.; von Meding J.; Prakash, I. Flood risk assessment using deep learning integrated with multi-criteria decision analysis. *Knowledge-Based Syst.* **2012**, *219*, 106899.
8. Pham, B.T.; Luu, C.; Phong, T.V.; Nguyen, H.D.; Le, H.V.; Tran, T.Q.; Ta, H.T.; Prakash, I. Flood risk assessment using hybrid artificial intelligence models integrated with multi-criteria decision analysis in Quang Nam Province, Vietnam. *J. Hydrol.* **2021**, *592*, 125815.
9. Pham, B.T.; Luu, C.; Phong, T.V.; Trinh, P.T.; Shirzadi, A.; Renoud, S.; Asadi, S.; Le, H.V.; von Meding, J.; Clague, J.J. Can deep learning algorithms outperform



- benchmark machine learning algorithms in flood susceptibility modeling? *J. Hydrol.* **2021**, 592, 125615.
10. Luu, C.; Pham, B.T.; Phong, T.V.; Costache, R.; Nguyen, H.D.; Amiri, M.; Bui, Q.D.; Nguyen, L.T.; Le, H.L.; Prakash, I. GIS-based ensemble computational models for flood susceptibility prediction in the Quang Binh Province, Vietnam. *J. Hydrol.* **2021**, 599, 126500.
  11. Nguyen, H.D.; Thanh, B.Q.; Nguyen, Q.H.; Nguyen, T.G.; Pham, L.T.; Nguyen, X.L.; Vu, P.L.; Nguyen, T.H.T.; Nguyen, A.T.; Petrisor, A.I. A novel hybrid approach to flood susceptibility assessment based on machine learning and land use change. Case study: a river watershed in Vietnam. *Hydrol. Sci. J.* **2022**, 67(7), 1065–1083.
  12. Minastireanu, E.A.; Mesnita, G. Light gbm machine learning algorithm to online click fraud detection. *J. Inform. Assur. Cybersecur* **2019**, 263928.
  13. Wang, D.N.; Li, L.; Zhao, D. Corporate finance risk prediction based on LightGBM. *Inf. Sci.* **2022**, 602, 259–268.
  14. Jing, W.; Qian, B.; Yannian, L. Study on food safety risk based on LightGBM model: a review. *Food Sci. Technol.* **2022**, 42, e42021.
  15. Đức, Đ.N.; Thanh, T.N.; Văn, P.T.; Thái, B.P. Phát triển mô hình học máy cây quyết định và cây quyết định xen kẽ thành lập bản đồ dự báo không gian sạt lở đất tại huyện Mường Nhé, tỉnh Điện Biên, Việt Nam. *Tap chí điện tử Khoa học và Công nghệ Giao thông* **2022**, 2(1), 36–56.
  16. Mahmoud, S.H.; Gan, T.Y. Multi-criteria approach to develop flood susceptibility maps in arid regions of Middle East. *J. Cleaner Prod.* **2018**, 196, 216–229.
  17. Weng, T.; Liu, W.; Xiao, J. Supply chain sales forecasting based on lightGBM and LSTM combination model. *Ind. Manage. Data Syst.* **2020**, 120(2), 265–279.
  18. Zhang, Y.; Zhu, C.; Wang, Q. LightGBM-based model for metro passenger volume forecasting. *IET Intel. Transport Syst.* **2020**, 14(13), 1815–1823.
  19. Aydin, H.E.; Iban, M.C. Predicting and analyzing flood susceptibility using boosting-based ensemble machine learning algorithms with SHapley Additive exPlanations. *Nat. Hazards* **2022**, 1–35.

## Flood susceptibility mapping using Light Gradient Boosting Machine

Nguyen Thanh Tuan<sup>1</sup>, Vu Cao Dat<sup>1\*</sup>, Dam Duc Nguyen<sup>1</sup>, Binh Thai Pham<sup>1</sup>

<sup>1</sup> University of Transport and Technology; datvc@utt.edu.vn; binhpt@utt.edu.vn; damnd@utt.edu.vn; tuannt94@utt.edu.vn

**Abstract:** The main objective of this study is to construct flood susceptibility map at Quang Trach district, Quang Binh province (Vietnam) using artificial intelligence based model namely Light Gradient Boosting Machine (LGBM). Database including 173 past flood locations and a set of seven flood conditioning factors (rainfall, geomorphology, land cover, elevation, curvature, slope and aspect) were collected and built training dataset (70%) and validating dataset (30%) for constructing and validating the model. Accuracy of the model was validated using various quantitative statistical indexes including area under the ROC curve (AUC). Results presented that LGBM has a good performance in predicting and assessing flood susceptibility at the study area (AUC = 0.96 for training dataset and AUC = 0.88 for validating dataset). Flood susceptibility map constructed from the modeling with good performance can be used for effective land use planning and giving suitable measures in reducing the harmful impacts of floods at the study area.

**Keywords:** Flood; Light Gradient Boosting Machine; Quang Trach; Vietnam.