

# ĐÁNH GIÁ NGUY CƠ CHÁY RỪNG KHU VỰC TỈNH PHÚ YÊN TỪ DỮ LIỆU VIỄN THÁM VÀ GIS TRÊN CƠ SỞ KỸ THUẬT HỌC MÁY RANDOM FOREST

TRẦN XUÂN TRƯỜNG<sup>(1)</sup>, LÊ THANH NGHỊ<sup>(1)</sup>  
NGUYỄN NHƯ HÙNG<sup>(2)</sup>, LÊ HÙNG CHIẾN<sup>(3)</sup>

<sup>(1)</sup>Trường Đại học Mở - Địa chất

<sup>(2)</sup>Trường Đại học kỹ thuật Lê Quý Đôn

<sup>(3)</sup>Trường Đại học Lâm Nghiệp

## **Tóm tắt:**

Cháy rừng là một trong những nguyên nhân gây tổn thất to lớn về kinh tế và môi trường sinh thái, làm suy giảm diện tích và chất lượng rừng. Do ảnh hưởng của biến đổi khí hậu và các hoạt động của con người, xu hướng và tần suất cháy rừng tăng lên và đạt đến mức đáng báo động ở nhiều khu vực ở Việt Nam. Nhiều mô hình dự báo nguy cơ cháy rừng bằng dữ liệu viễn thám và GIS đã được đề xuất và thử nghiệm cho các khu vực với điều kiện tự nhiên, kinh tế - xã hội khác nhau. Bài báo này trình bày kết quả xây dựng bản đồ dự báo nguy cơ cháy rừng khu vực tỉnh Phú Yên từ dữ liệu viễn thám và GIS. Trong nghiên cứu, 10 lớp thông tin đầu vào chiết xuất từ dữ liệu viễn thám và GIS, bao gồm: chỉ số thực vật NDVI, chỉ số nước NDWI, lớp phủ/sử dụng đất, khoảng cách đến đường giao thông, độ ẩm tương đối, nhiệt độ bề mặt, lượng mưa trung bình tháng, hướng sườn, độ dốc và độ cao được sử dụng để thành lập bản đồ phân vùng nguy cơ cháy rừng trên cơ sở kỹ thuật Random Forest. Kết quả nhận được trong nghiên cứu là giải pháp hiệu quả trong dự báo sớm nguy cơ cháy rừng cũng như giảm thiểu thiệt hại do cháy rừng gây ra.

*Từ khóa:* dự báo nguy cơ cháy rừng, viễn thám, GIS, mô hình hóa, Random Forest, tỉnh Phú Yên

## **1. Mở đầu**

Do ảnh hưởng của biến đổi khí hậu, biểu hiện bởi lượng mưa ít hơn và gia tăng nhiệt độ, mùa khô kéo dài hơn và sự can thiệp của các hoạt động của con người, xu hướng và tần suất cháy rừng tăng lên và đạt đến mức đáng báo động ở nhiều khu vực trên thế giới, trong đó có Việt Nam. Việc dự báo nguy cơ cháy rừng là một vấn đề có tính cấp thiết, cung cấp nguồn thông tin kịp thời phục vụ công tác bảo

vệ và phát triển tài nguyên rừng, hỗ trợ chính quyền địa phương trong quản lý và quy hoạch rừng, phân bổ nguồn lực, xử lý tình huống khẩn cấp và cảnh báo sớm cháy rừng.

Trên thế giới, các phương pháp thống kê dựa trên các chỉ số khí tượng, khí hậu đã được sử dụng cho nghiên cứu cháy rừng do tính chất ngẫu nhiên cố hữu của hiện tượng cháy rừng. Mặc dù đã đạt được những kết quả quan trọng, các phương pháp này đã bỏ qua nhiều yếu tố

đầu vào quan trọng của mô hình dự báo nguy cơ cháy rừng như các chỉ số thảm thực vật, khoảng cách đường giao thông, khoảng cách tới khu dân cư... Ngoài ra, với các bài toán có khối lượng dữ liệu lớn, nhiều đầu vào như mô hình dự báo nguy cơ cháy rừng, độ chính xác của các mô hình thống kê vẫn còn hạn chế. Những hạn chế này có thể được khắc phục khi áp dụng các kỹ thuật học máy (machine learning) trong xây dựng các mô hình dự báo nguy cơ cháy rừng từ dữ liệu viễn thám và GIS do các thuật toán này hiệu quả hơn với dữ liệu lớn, có nhiều đầu vào.

Các nghiên cứu trên thế giới đã kết hợp các kỹ thuật học máy như Random Forest (RF), Support Vector Machine (SVM), Classification and Regression Tree (CART), mạng neural nhân tạo (ANN) để nâng cao độ chính xác của các mô hình dự báo nguy cơ cháy rừng (Barmoutis et al., 2020; Vasilakos et al., 2009; Oliveira et al., 2012; Dieu Tien Bui et al., 2016). Ngoài ra, các kỹ thuật hồi quy đa biến (multiple regression) (Oliveira et al., 2012), hồi quy logistic (Pourghasemi, 2015), hồi quy trọng số địa lý (geographically weighted regression - GWR) (Fernandez et al., 2012), hồi quy tuyến tính tổng quát hóa GLMs (Ruano et al., 2022), kỹ thuật khai phá dữ liệu (data mining) (Arpaci et al., 2014) cũng được sử dụng để xây dựng các mô hình dự báo nguy cơ cháy rừng trên cơ sở đánh giá mối quan hệ giữa các yếu tố tự nhiên, xã hội và khả năng xảy ra cháy. Số lượng các biến đầu vào trong các mô hình hồi quy này là rất lớn, góp phần nâng cao độ chính xác cho kết quả dự báo nguy cơ cháy rừng. Tại Việt Nam, một số nghiên cứu gần đây như (Đặng Ngô Bảo Toàn, 2021; Dieu Tien Bui et al., 2017;

Hoang V.T. et al., 2020) đã sử dụng các kỹ thuật học máy dựa trên dữ liệu viễn thám và GIS để dự báo nguy cơ cháy rừng cho một số khu vực cụ thể. Mặc dù vậy, do mỗi địa phương có điều kiện tự nhiên, kinh tế - xã hội khác nhau, cần thiết phải có sự nghiên cứu, đánh giá và lựa chọn mô hình phù hợp với điều kiện cụ thể từng khu vực.

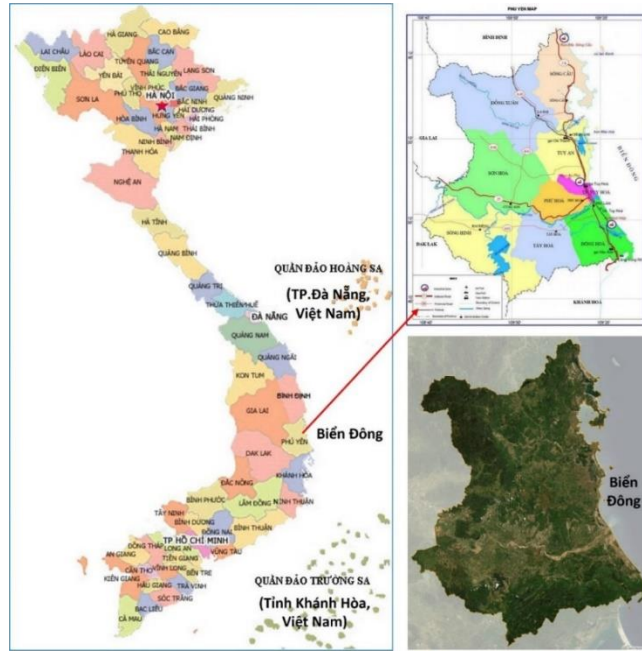
Bài báo này trình bày kết quả xây dựng bản đồ dự báo nguy cơ cháy rừng khu vực tỉnh Phú Yên từ dữ liệu viễn thám và GIS. 10 lớp thông tin đầu vào đại diện cho đặc điểm địa hình, lớp phủ, khí hậu và điều kiện kinh tế - xã hội cùng bộ dữ liệu các điểm cháy trong quá khứ được sử dụng kết hợp các thuật toán học máy nhằm đánh giá nguy cơ cháy rừng khu vực thực nghiệm.

## **2. Dữ liệu và phương pháp nghiên cứu**

### **2.1. Dữ liệu và khu vực nghiên cứu**

#### *a) Khu vực nghiên cứu*

Khu vực nghiên cứu được lựa chọn là địa bàn tỉnh Phú Yên (Hình 1). Phú Yên là một tỉnh duyên hải miền Trung, có diện tích tự nhiên đạt 5.045 km<sup>2</sup>, dân số năm 2022 là 876.619 người. Phú Yên thuộc vùng khí hậu nhiệt đới gió mùa, chịu ảnh hưởng của khí hậu đại dương. Thời tiết có 2 mùa rõ rệt: mùa khô từ tháng 1 đến tháng 8, mùa mưa từ tháng 9 đến tháng 12, tập trung từ 70 - 80% lượng mưa cả năm. Theo số liệu thống kê, diện tích đất rừng ở Phú Yên đạt gần 182.000 hecta, bao gồm 3 kiểu rừng chính: rừng kín lá rộng thường xanh, rừng rụng lá (rừng khộp) và rừng trồng. Thời gian gần đây, do ảnh hưởng của biến đổi khí hậu và các hoạt động nhân sinh, cháy rừng ở Phú Yên diễn ra phức tạp, với hàng chục vụ cháy mỗi năm.



Hình 1: Vị trí địa lý tỉnh Phú Yên

### b) Dữ liệu sử dụng

Dữ liệu viễn thám sử dụng trong nghiên cứu là ảnh vệ tinh Landsat 8 chụp ngày 28/02/2022 nhằm xác định các chỉ số thực vật và chỉ số nước. Ảnh được thu thập vào mùa khô, có chất lượng tốt và không bị ảnh hưởng bởi điều kiện thời tiết.

Mô hình số độ cao ALOS với độ phân giải không gian 30 m cung cấp bởi Cơ quan Hàng không vũ trụ Nhật Bản (JAXA) được sử dụng để chiết xuất thông tin các yếu tố địa hình như độ dốc, độ cao, hướng sườn. Dữ liệu về lớp phủ/sử dụng đất được thu thập từ sản phẩm bản đồ LULC độ phân giải 30 m năm 2020 do JAXA cung cấp.

Dữ liệu mạng lưới đường giao thông được thu thập từ dự án Open Street Map và bản đồ địa hình quốc gia tỉ lệ 1:50.000 để xây dựng lớp dữ liệu khoảng cách tới đường giao thông.

Dữ liệu về khí hậu và thời tiết, bao gồm: nhiệt độ bề mặt, độ ẩm tương đối, lượng mưa trung bình tháng được thu thập từ dự án POWER, thuộc Cơ quan Hàng không và Vũ

trụ Quốc gia Hoa Kỳ (NASA), truy cập tại địa chỉ <https://power.larc.nasa.gov/data-access-viewer/>.

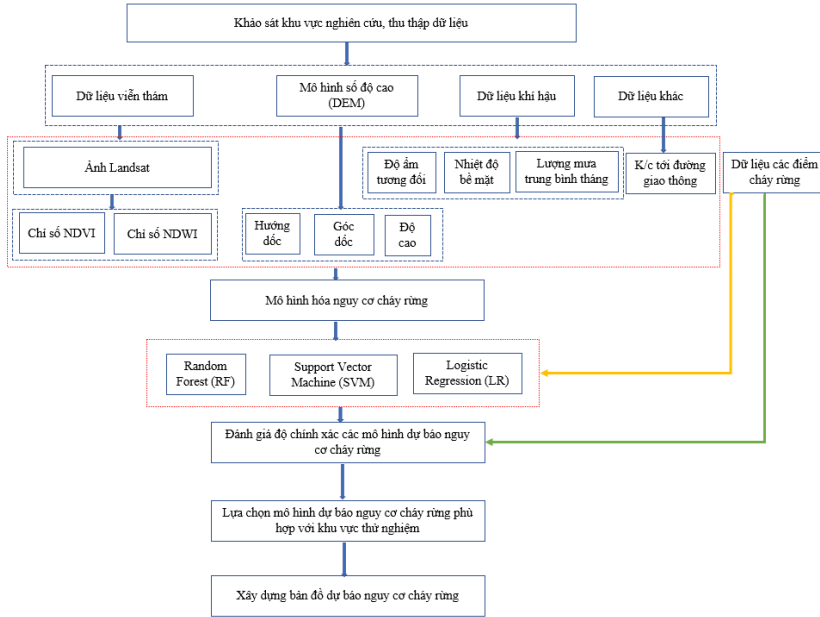
Bộ dữ liệu điểm cháy bao gồm 306 điểm được thu thập trong giai đoạn 2019 - 2023 dựa trên hồ sơ cháy của Cục Kiểm lâm (Bộ NN&PTNT); dữ liệu hỏa hoạn từ hệ thống FIRMS (NASA); thông tin các vụ cháy rừng thu thập từ phương tiện truyền thông và các điểm cháy rừng từ kiểm tra thực địa bằng máy GPS cầm tay.

### 2.2. Phương pháp nghiên cứu

Từ phân tích đặc điểm tự nhiên, kinh tế - xã hội khu vực nghiên cứu và dữ liệu thu thập được, trong bài báo tiến hành lựa chọn và xây dựng 10 lớp thông tin đầu vào cho mô hình dự báo nguy cơ cháy rừng, bao gồm: (1) Chỉ số thực vật NDVI; (2) Chỉ số nước NDWI; (3) Sử dụng đất/Lớp phủ (LULC); (4) Khoảng cách tới đường giao thông; (5) Độ ẩm tương đối; (6) Nhiệt độ bề mặt; (7) Lượng mưa trung bình tháng; (8) Độ dốc; (9) Độ cao và (10) Hướng sườn.

Các thuật toán học máy, bao gồm RF, SVM và hồi quy Logistic (LR) được lựa chọn thử nghiệm để xây dựng mô hình dự báo nguy cơ cháy rừng khu vực nghiên cứu, sau đó đánh giá độ chính xác dựa trên bộ dữ liệu kiểm tra và lựa chọn thuật toán có độ chính xác cao

nhất. Từ bộ dữ liệu điểm cháy thu thập được, 70% số lượng điểm được sử dụng để huấn luyện mô hình, trong khi 30% số lượng điểm cháy còn lại dùng để kiểm tra độ chính xác mô hình. Sơ đồ quy trình nghiên cứu được trình bày trên hình 2.

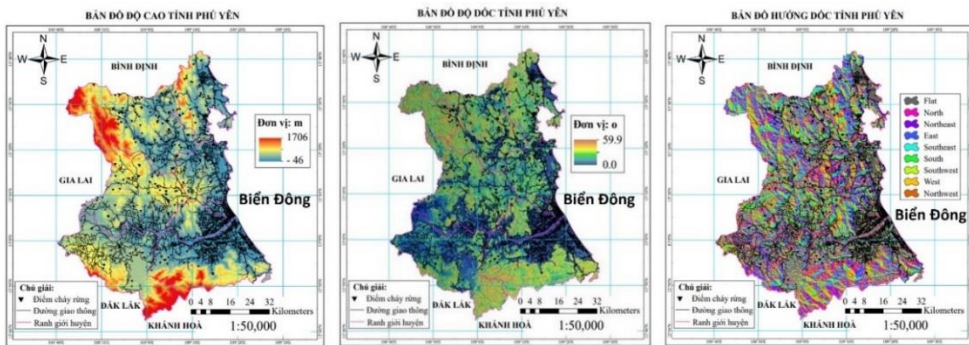


Hình 2: Sơ đồ quy trình công nghệ dự báo nguy cơ cháy rừng từ dữ liệu viễn thám và GIS

### 3. Kết quả nghiên cứu và thảo luận

Trên hình 3 trình bày kết quả xây dựng các lớp thông tin về địa hình, bao gồm độ cao,

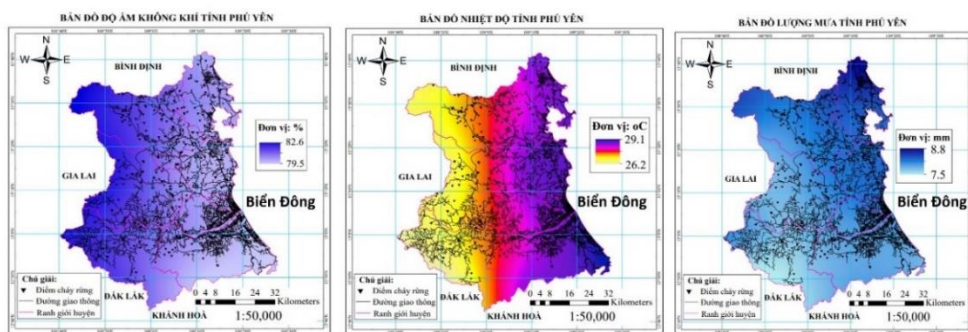
độ dốc và hướng sườn khu vực tỉnh Phú Yên từ mô hình số độ cao ALOS với độ phân giải không gian 30m, cung cấp bởi Cơ quan hàng không vũ trụ Nhật Bản (JAXA).



Hình 3: Các lớp dữ liệu về độ cao, độ dốc, hướng sườn khu vực tỉnh Phú Yên

Từ bộ dữ liệu về khí hậu thu thập từ dự án POWER của Cơ quan Hàng không vũ trụ Hoa Kỳ (NASA), trong nghiên cứu tiến hành xây dựng các lớp dữ liệu đầu vào của mô hình liên

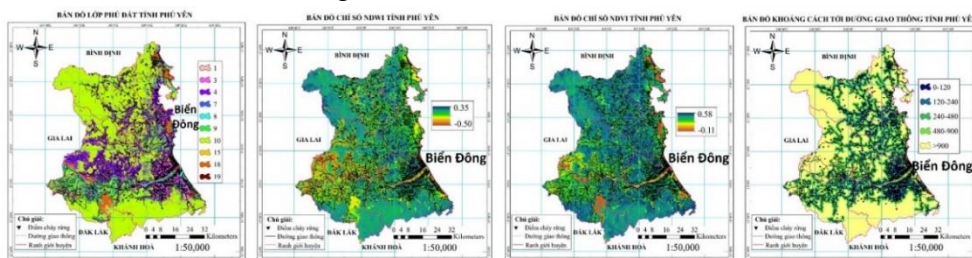
quan đến yếu tố khí hậu như độ ẩm không khí tương đối, nhiệt độ bề mặt và lượng mưa trung bình tháng (hình 4).



Hình 4: Các lớp dữ liệu về độ ẩm không khí, nhiệt độ bề mặt và lượng mưa trung bình tháng khu vực tỉnh Phú Yên

Các lớp thông tin đầu vào của mô hình dự báo nguy cơ cháy rừng về sử dụng đất/lớp phủ (LULC), lớp phủ thực vật (chỉ số NDVI), lớp phủ nước mặt (NDWI) và khoảng cách tới

đường giao thông được trình bày trên hình 5. Tất cả 10 lớp dữ liệu đầu vào của mô hình đều có độ phân giải không gian 30 m.



Hình 5: Các lớp dữ liệu về LULC, chỉ số NDWI, NDVI và khoảng cách tới đường giao thông khu vực tỉnh Phú Yên

Từ bộ dữ liệu đầu vào này và 306 điểm cháy rừng khu vực tỉnh Phú Yên trong quá khứ, trong nghiên cứu tiến hành thử nghiệm với 03 thuật toán học máy thông dụng: RF, SVM và LR, sau đó đánh giá hiệu suất của các mô hình để lựa chọn mô hình phù hợp với điều kiện tự nhiên, kinh tế - xã hội khu vực thực nghiệm. 214 điểm cháy (tương đương 70% bộ dữ liệu điểm cháy) được sử dụng để huấn luyện các mô hình học máy, trong khi đó 92 điểm (tương đương 30% bộ dữ liệu điểm cháy) được sử dụng để đánh giá độ chính xác các mô hình. Các tham số đối với các thuật toán được thử nghiệm với nhiều giá trị khác nhau để lựa chọn tham số có độ chính xác cao nhất.

Đối với bộ dữ liệu huấn luyện, trong nghiên cứu tiến hành đánh giá qua các chỉ số thống kê đo lường độ chính xác bao gồm True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Độ nhạy cảm (Sens), Độ đặc hiệu (Spec), Độ chính xác tổng thể (Acc), F-score, chỉ số Kappa và chỉ số AUC. Dựa trên số liệu thu được ở bảng 1, cho thấy mô hình học máy RF có hiệu suất tốt hơn so với các thuật toán khác bao gồm SVM và LR. Các giá trị TP và TN thu được qua mô hình RF lần lượt là 161 và 189 điểm dữ liệu, cao hơn so với các thuật toán SVM. Với thuật toán LR, giá trị TP tuy có cao hơn so với RF (165 so với 161), tuy nhiên giá trị TN lại nhỏ hơn khá nhiều

(158 so với 189). Tương tự như vậy, các giá trị FP và FN khi sử dụng thuật toán RF là nhỏ hơn so với SVM, lần lượt là 53 và 25 điểm dữ liệu, khi so sánh với 53 và 62 điểm dữ liệu khi sử dụng SVM. Trong khi đó, với thuật toán LR, giá trị FP đạt thấp nhất (49), tuy nhiên FN

lại đạt cao hơn rất nhiều so với RF (56 so với 25). Với các thông số còn lại, thuật toán RF đều thể hiện hiệu suất cao hơn khi so với SVM và LR, đặc biệt là giá trị hệ số Kappa (0.636) và AUC (0.896) khi so với SVM (0.463 và 0.846) và LR (0.509 và 0.851).

*Bảng 1: Hiệu suất của các mô hình RF, SVM và LR khi sử dụng tập dữ liệu huấn luyện*

Thuật toán	Chỉ số thống kê đo lường độ chính xác											
	TP	TN	FP	FN	PPV	NPV	Sens	Spec	Acc	F-score	Kappa	AUC
RF	161	189	53	25	75.2	88.3	86.6	78.1	81.8	0.805	0.636	<b>0.896</b>
SVM	161	152	53	62	75.2	71.0	72.2	74.1	73.1	0.737	0.463	<b>0.846</b>
LR	165	158	49	56	77.1	73.8	74.7	76.3	75.5	0.759	0.509	<b>0.851</b>

Với bộ dữ liệu kiểm tra, thuật toán RF cũng thể hiện hiệu quả cao hơn khi so với thuật toán SVM và RF, thể hiện ở phần lớn các thông số đánh giá độ chính xác (bảng 2). Giá trị các thông số TN, NPV, độ nhạy cảm

(Sens), độ đặc hiệu (Spec), độ chính xác tổng thể (Acc), F-score, hệ số Kappa và diện tích dưới đường cong (AUC) khi sử dụng thuật toán RF đều đạt cao nhất.

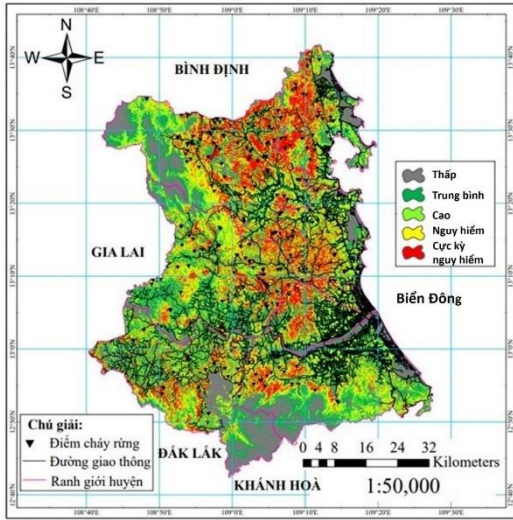
*Bảng 2: Hiệu suất của các mô hình RF, SVM và LR khi sử dụng bộ dữ liệu kiểm tra*

Thuật toán	Chỉ số thống kê đo lường độ chính xác											
	TP	TN	FP	FN	PPV	NPV	Sens	Spec	Acc	F-score	Kappa	AUC
RF	72	74	20	18	78.3	80.4	80.0	78.7	79.3	0.791	0.587	<b>0.925</b>
SVM	76	64	16	28	82.6	69.6	73.1	80.0	76.1	0.776	0.522	<b>0.874</b>
LR	72	65	20	27	78.3	70.7	72.7	76.5	74.5	0.754	0.489	<b>0.856</b>

Từ kết quả phân tích các thông số thống kê trong đánh giá hiệu suất của các mô hình dự báo nguy cơ cháy rừng, trong nghiên cứu lựa chọn thuật toán RF để phân loại và xây dựng bản đồ nguy cơ cháy rừng khu vực tỉnh Phú Yên. Bản đồ cảnh báo cháy rừng tỉnh Phú Yên thành lập từ mô hình học máy RF được trình bày trong hình 6, trong đó nguy cơ cháy rừng khu vực tỉnh Phú Yên được chia thành 5 mức độ: thấp, trung bình, cao, nguy hiểm và cực kỳ nguy hiểm.

Phân tích kết quả đạt được cho thấy, các khu vực có nguy cơ cháy rừng «nguy hiểm» và «cực kỳ nguy hiểm» tập trung chủ yếu ở địa bàn các huyện Đồng Xuân, Sông Cầu, Tuy

An và một phần các huyện Tây Hòa, Sông Hinh. Các khu vực có nguy cơ cháy rừng «thấp» và «trung bình» phân bố chủ yếu ở phía nam các huyện Tây Hòa và Sông Hinh cùng phần phía tây khu vực tỉnh Phú Yên. So sánh với vị trí các điểm cháy xảy ra trong quá khứ (ký hiệu bởi biểu tượng hình tam giác màu đen trên hình 6) cho thấy, phần lớn các điểm cháy được phân bố ở các khu vực có nguy cơ cháy rừng «nguy hiểm» và «cực kỳ nguy hiểm», đặc biệt ở khu vực huyện Đồng Xuân và Sông Cầu và phần trung tâm của tỉnh thuộc địa bàn các huyện Sơn Hòa, Tuy An.



Hình 6: Bản đồ dự báo nguy cơ cháy rừng khu vực tỉnh Phú Yên xây dựng từ dữ liệu viễn thám và GIS trên cơ sở thuật toán RF

#### 4. Kết luận

Trong nghiên cứu này, 10 lớp dữ liệu đầu vào đại diện cho đặc điểm lớp phủ, địa hình, khí hậu và hoạt động của con người được sử dụng để xây dựng bản đồ dự báo nguy cơ cháy rừng khu vực tỉnh Phú Yên trên cơ sở các kỹ thuật học máy. Các lớp thông tin đầu vào này được chiết xuất từ dữ liệu viễn thám và GIS, kết hợp bộ dữ liệu các điểm cháy trong quá khứ phục vụ quá trình huấn luyện của các mô hình. 03 mô hình học máy phổ biến, bao gồm RF, SVM và LR được thử nghiệm để lựa chọn mô hình phù hợp nhất với đặc điểm khu vực nghiên cứu. Kết quả nhận được cho thấy, thuật toán Random Forest cho phép dự báo nguy cơ cháy rừng khu vực tỉnh Phú Yên với độ chính xác cao nhất, thể hiện qua so sánh các thông số thống kê khi đánh giá hiệu suất của bộ dữ liệu huấn luyện và bộ dữ liệu kiểm tra.

Kết quả nhận được trong nghiên cứu có thể sử dụng hiệu quả phục vụ công tác dự báo sớm nguy cơ cháy rừng, cung cấp thông tin giúp các nhà quản lý trong theo dõi, ứng phó và giảm thiểu thiệt hại do cháy rừng gây ra. ○

#### Tài liệu tham khảo

- [1]. Đặng Ngô Bảo Toàn (2021). Nghiên cứu nguy cơ và cảnh báo cháy rừng tỉnh Sơn La trên cơ sở ứng dụng công nghệ địa thông tin, *Luận án Tiến sĩ Địa lý*, Đại học Quốc gia Hà Nội.
- [2]. Arpacı A., Eastaugh C., Vacik H. (2013). Selecting the best performing fire weather indices for Austrian ecoregions, *Theoretical and Applied Climatology*, 114, 393 - 406.
- [3]. Barmpoutis P., Papaioannou P., Dimitropoulos K., Grammalidis N. (2020). A Review on Early Forest Fire Detection Systems Using Optical Remote Sensing, *Sensors*, 20(22), 6442; <https://doi.org/10.3390/s20226442>
- [4]. Dieu Tien Bui, Kim Thoa Thi Le, Van Cam Nguyen, Hoang Duc Le, Revhaug I. (2016). Tropical Forest Fire Susceptibility Mapping at the Cat Ba National Park Area, Hai Phong City, Vietnam, Using GIS-Based Kernel Logistic Regression, *Remote Sensing*, 8, 347, doi:10.3390/rs8040347.
- [5]. Dieu Tien Bui, Quang Thanh Bui, Quoc Phi Nguyen, Pradhan B., Nampak H., Phan Trong Trinh (2017). A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area, *Agricultural and Forest Meteorology*, 233, 32 - 44.
- [6]. Fernandez J., Chuvieco E., Koutsias N. (2012). Modelling long-term fire occurrence factors in Spain by accounting for local variations with geographically weighted regression, *Natural Hazards Earth System Sciences*, 12, 1-17.

- [7]. Hoang V.T., Chou T., Fang Y., Nguyen N.T., Nguyen Q.H., Pham X.C., Dang N.B.T., Nguyen X.L., Meadows M. (2020). Mapping forest fire risk and development of early warning system for NW Vietnam using AHP and MCA/GIS methods, *Applied Sciences*, 10(12), 4348.
- [8]. Oliveira S., Oehler F., Ayanz J., Camia A., Pereira J. (2012). Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest, *Forest Ecology and Management*, Vol. 275, 117 – 129.
- [9]. Pourghasemi H. (2015). GIS-based forest fire susceptibility mapping in Iran: A comparison between evidential belief function and binary logistic regression models, *Scandinavian Journal of Forest Research*, 40 pp., DOI: 10.1080/02827581.2015.1052750.
- [10]. Ruano A., Jolly W., Freeborn P., Nieva D., Vega N., Herrera C., Rodrigues M. (2022). Spatial Predictions of Human and Natural-Caused Wildfire Likelihood across Montana (USA), *Remote Sensing*, 13(8), 1200.
- [11]. Vasilakos C., Kalabokidis K., Hatzopoulos J., Matsinos T. (2009). Identifying wildland fire ignition factors through sensitivity analysis of a neural network, *Natural Hazards*, 50, 125 - 143.
- [12]. <https://phuyen.gov.vn/> ○

## Summary

### **Assessment of forest fire risk in Phu Yen province from remote sensing and GIS data based on Random Forest algorithm**

*Tran Xuan Truong, Le Thanh Nghi, Hanoi University of Mining and Geology*

*Nguyen Nhu Hung, Le Quy Don Technical University*

*Le Hung Chien, Vietnam National University of Forestry*

Forest fires are one of the causes of great economic and ecological losses, reducing the area and quality of forests. Due to the effects of climate change and human activities, the trend and frequency of forest fires have increased and reached alarming levels in many areas in Vietnam. Many models of forest fire risk prediction by remote sensing and GIS data have been proposed and tested for regions with different natural and socio-economic conditions. This paper presents the results of building a model to predict forest fire risk in Phu Yen province using remote sensing and GIS data. In the study, 10 input layers were extracted from remote sensing and GIS data, including the NDVI index, NDWI index, land cover/land use, distance to roads, relative humidity, land surface temperature, average monthly rainfall, aspect, slope and elevation are used to build a forest fire risk prediction map based on the Random Forest algorithm. The results obtained in the study are an effective solution in early forecasting the risk of forest fires as well as minimising damage caused by forest fires. ○

Keywords: forest fire risk prediction, remote sensing, GIS, modeling, Random Forest, Phu Yen province.