

ConvLSTM-BASED HYBRID MODEL FOR FOREST FIRE IMAGE CLASSIFICATION IN VIETNAM

Phạm Thị Lien¹, Nguyễn Thu Hương¹, Nguyễn Thế Long², Trần Tuấn Việt^{1*}

¹TNU - University of Information and Communication Technology, ²Sao Do University

ARTICLE INFO		ABSTRACT
Received:	04/6/2026	Forest fires represent a critical issue in Vietnam, particularly during the dry season. This study proposes the utilization of the ConvLSTM neural network for the early detection of forest fires from imagery data. Unlike traditional methods that process individual frames in isolation, ConvLSTM is capable of combining spatial feature extraction (via convolutional layers) with temporal dependency modeling (via Long Short-Term Memory). We utilize a publicly available dataset comprising 999 images with and without fire. To apply ConvLSTM to static images, time-series sequences are generated using a sliding window approach across data augmentation variants. Experimental results demonstrate that the model achieves an accuracy of 98.3%, a precision of 98.1%, a recall of 98.6%, and an F1-score of 98.2%, significantly outperforming standalone CNN or LSTM models. However, the study is constrained by its limited dataset size, which exhibits an imbalance where 75% of the samples contain fire, as well as a lack of Vietnam-specific data. Future research lines should focus on expanding to larger datasets that better reflect local conditions and deploying the system in real-time on edge devices.
Revised:	30/6/2026	
Published:	30/6/2026	
KEYWORDS		
ConvLSTM		
Forest fires		
Deep learning		
CNN		
LSTM		

MÔ HÌNH LẠI ConvLSTM CHO BÀI TOÁN PHÂN LOẠI ẢNH CÓ CHÁY RỪNG TẠI VIỆT NAM

Phạm Thị Liên¹, Nguyễn Thu Hương¹, Nguyễn Thế Long², Trần Tuấn Việt^{1*}

¹Trường Đại học Công nghệ Thông tin và Truyền thông - ĐH Thái Nguyên, ²Trường Đại học Sao Đỏ

THÔNG TIN BÀI BÁO		TÓM TẮT
Ngày nhận bài:	04/6/2026	Cháy rừng đang là vấn đề nghiêm trọng tại Việt Nam, nhất là trong mùa khô. Nghiên cứu này đề xuất sử dụng mạng nơ-ron ConvLSTM để phát hiện sớm cháy rừng từ dữ liệu ảnh. Khác với các phương pháp truyền thống chỉ xử lý từng khung hình riêng lẻ, ConvLSTM có khả năng kết hợp trích xuất đặc điểm không gian (qua lớp tích chập) và mô hình hóa sự phụ thuộc thời gian (qua bộ nhớ dài-ngắn LSTM). Chúng tôi sử dụng bộ dữ liệu công khai gồm 999 ảnh có và không có cháy. Để áp dụng ConvLSTM trên ảnh tĩnh, các chuỗi thời gian được tạo bằng cửa sổ trượt trên các biến thể tăng cường dữ liệu. Kết quả thử nghiệm cho thấy mô hình đạt độ chính xác 98,3%, độ chính xác dự báo dương tính 98,1%, độ nhạy 96,8% và F1-score 98,1%, cao hơn rõ rệt so với các mô hình CNN hay LSTM độc lập. Dù vậy, nghiên cứu còn hạn chế về kích thước bộ dữ liệu (nhỏ, mất cân bằng, 75% là ảnh có cháy) và thiếu dữ liệu đặc thù cho Việt Nam. Các bước phát triển tiếp theo cần hướng tới bộ dữ liệu lớn hơn, phù hợp với điều kiện địa phương và triển khai thời gian thực trên các thiết bị biên.
Ngày hoàn thiện:	30/6/2026	
Ngày đăng:	30/6/2026	
TỪ KHÓA		
ConvLSTM		
Cháy rừng		
Học sâu		
Mạng nơ-ron tích chập		
Mạng bộ nhớ dài-ngắn		

DOI: <https://doi.org/10.34238/tnu-jst.15980>

* Corresponding author. Email: ttviet@ictu.edu.vn

1. Giới thiệu

Cháy rừng đang ngày càng trở nên phổ biến và khốc liệt hơn trên phạm vi toàn cầu do tác động mạnh mẽ của biến đổi khí hậu. Sự gia tăng nhiệt độ, các đợt hạn hán kéo dài và lượng mưa thất thường tạo điều kiện lý tưởng cho lửa bùng phát và lan rộng nhanh chóng. Tại Việt Nam, Tây Nguyên và các tỉnh miền núi phía Bắc là những khu vực có nguy cơ cháy rừng cao nhất do diện tích rừng lớn, địa hình hiểm trở, kết hợp với tập quán đốt nương làm rẫy và canh tác nông nghiệp của người dân. Sự chậm trễ trong phát hiện sớm không chỉ gây thiệt hại nghiêm trọng về tài nguyên môi trường mà còn đe dọa trực tiếp đến tính mạng và sức khỏe con người.

Trong một thời gian dài, công tác phát hiện cháy rừng chủ yếu dựa vào các phương pháp truyền thống như quan sát trực tiếp từ chòi canh, sử dụng cảm biến cố định ngưỡng, hoặc phân tích ảnh vệ tinh có tần suất chụp lặp lại thấp. Tuy nhiên, các cách tiếp cận này bộc lộ nhiều hạn chế về độ tin cậy và thời gian phản hồi; khi khói hoặc lửa được phát hiện bằng mắt thường, đám cháy thường đã lan rộng vượt tầm kiểm soát. Thêm vào đó, các hệ thống truyền thống không thể mô hình hóa đầy đủ các tương tác phi tuyến, phức tạp giữa các yếu tố khí tượng, địa lý và con người.

Gần đây, sự phát triển của trí tuệ nhân tạo, đặc biệt là học sâu (deep learning), đã mở ra những hướng đi đầy hứa hẹn cho các hệ thống cảnh báo thông minh. Nhiều nghiên cứu đã tập trung sử dụng mạng nơ-ron tích chập (CNN) để phân tích đặc điểm không gian từ hình ảnh camera và vệ tinh nhằm nhận dạng khói, lửa. Abdusalomov và cộng sự [1] đã xây dựng một CNN tùy chỉnh sử dụng framework Detectron2 và huấn luyện trên bộ dữ liệu ảnh cháy chuyên dụng. Mô hình cải thiện đáng kể khả năng phát hiện cháy ở vùng sâu vùng xa, cả ban ngày lẫn ban đêm. Trong khi đó, Mei và Zhang [2] tiếp cận bằng cách sử dụng hồi quy logistic và làm mịn thời gian, dùng các đặc trưng như kích thước, chuyển động và màu sắc để nhận diện lửa, đồng thời loại bỏ nền để cô lập các vùng có khả năng là lửa. Dimitropoulos và cộng sự [3] đầu tiên xác định các vùng có khả năng cháy bằng rừng ngẫu nhiên, SVM và thuật toán ViBe cải tiến, sau đó xác nhận cháy bằng các mô-men Hu, tỷ lệ chồng lấn và cường độ chuyển động giữa các khung hình. Một nghiên cứu khác kết hợp nhiều đặc trưng với phương pháp hợp nhất phân số thích nghi và sử dụng SVM để phát hiện khói. Zhang và cộng sự [4] đề xuất ATT Squeeze U-Net cho phân đoạn và nhận dạng cháy. Họ kết hợp SqueezeNet với một mô-đun Fire sửa đổi bên trong cấu trúc ATT U-Net, cải thiện việc trích xuất đặc điểm ngay cả khi dữ liệu huấn luyện hạn chế. Một mô hình nhận dạng riêng biệt chia sẻ một phần đường mã hóa cho các tác vụ phân loại. Cách tiếp cận này hiệu quả hơn các hệ thống phân đoạn và nhận dạng hiện có, đặc biệt hữu ích cho phát hiện cháy khi yêu cầu độ nhạy cao nhưng dữ liệu huấn luyện khan hiếm. Các thử nghiệm xác nhận độ chính xác phân đoạn và khả năng nhận dạng đáng tin cậy. Tuy nhiên, mặc dù phương pháp xác định chính xác các điều kiện cháy và phân đoạn chi tiết vùng cháy, nó vẫn có thể bỏ sót một số loại cháy trong các bối cảnh đa dạng. Kumar [5] giới thiệu một mô hình học sâu kết hợp CNN và RNN để phát hiện cháy rừng. CNN trích xuất đặc điểm, sau đó được làm phẳng và đưa vào RNN. Trong khi CNN xử lý các mẫu hình ảnh cơ bản và phức tạp, RNN xử lý các phụ thuộc tuần tự trong dữ liệu. Kết hợp chúng sẽ cải thiện khả năng nhận dạng mẫu cháy. Larsen và cộng sự [6] đã phát triển một mạng nơ-ron tích chập đầy đủ (FCN) để phát hiện khói cháy rừng trong ảnh vệ tinh gần như thời gian thực. Họ huấn luyện mô hình sử dụng dữ liệu từ các phương pháp phát hiện khói vận hành sẵn có, vì vậy mô hình học từ các mẫu khói đã được xác minh. Hệ thống này chứng tỏ tính thực tiễn cho sử dụng gần như tức thời và đã được thử nghiệm trong một vụ cháy rừng ở Australia. Công cụ dựa trên FCN này có triển vọng đánh giá rủi ro phơi nhiễm khói và cung cấp thông tin kịp thời cho các đội quản lý cháy, cơ quan y tế và môi trường, cũng như công chúng, qua đó có thể giảm rủi ro sức khỏe nhờ cảnh báo nhanh hơn. Bên cạnh đó, các nhóm nghiên cứu khác lại ứng dụng các mô hình học chuỗi như mạng bộ nhớ dài-ngắn (LSTM) để xử lý dữ liệu thời gian từ trạm khí tượng (nhiệt độ, độ ẩm, tốc độ gió), giúp dự báo các dấu hiệu bất thường trước khi đám cháy xảy ra [7]. Xu hướng hiện nay đang chuyển dịch sang các mô hình kết hợp đa phương thức, sử dụng đồng thời dữ liệu hình ảnh và dữ liệu cảm biến nhằm giảm thiểu tỷ lệ báo

động giả. Mambile và cộng sự [8] đề xuất một phương pháp dự báo cháy rừng kết hợp dữ liệu hoạt động của con người, ảnh vệ tinh và các biến môi trường trong một khuôn khổ học sâu. Mô hình sử dụng kiến trúc ConvLSTM để nắm bắt cả mẫu không gian và thời gian. Nó dự báo các giai đoạn nguy cơ cao bằng cách sử dụng dữ liệu thời tiết và hành vi của con người, đồng thời xác định các khu vực dễ cháy thông qua các chỉ số từ vệ tinh như NDVI, NBR và NDWI. Bằng cách tích hợp các yếu tố tự nhiên và con người, cách tiếp cận này cung cấp hướng dẫn thực tiễn cho quản lý cháy, giúp tối ưu hóa phân bổ nguồn lực ở các vùng dễ bị tổn thương.

Bên cạnh các hướng tiếp cận quốc tế, trong những năm gần đây, các nhà khoa học trong nước cũng có nhiều nỗ lực ứng dụng học máy và học sâu vào bài toán cháy rừng, phù hợp với điều kiện khí hậu nhiệt đới và đặc điểm rừng tại Việt Nam. Một hướng tiếp cận phổ biến là sử dụng các mô hình học máy tối ưu để lập bản đồ khu vực nhạy cảm với cháy rừng. Hung Van Le và cộng sự [9] đã nghiên cứu tại tỉnh Gia Lai, sử dụng 2.827 điểm cháy ghi nhận từ 2007 đến 2021 cùng với 12 yếu tố ảnh hưởng bao gồm địa hình, thảm thực vật (NDVI, NDWI), khí tượng (nhiệt độ, lượng mưa, độ ẩm). Nhóm tác giả đã đánh giá năm mô hình, trong đó Random Forest (RF) cho kết quả tốt nhất. Sau khi tối ưu bằng giải thuật di truyền (GA), mô hình Acc-GA-Opt-RF đạt độ chính xác 84,4% và AUC 0,908, đồng thời chỉ ra NDVI và NDWI là hai yếu tố quan trọng nhất trong dự báo nguy cơ cháy.

Cũng trong lĩnh vực lập bản đồ nhạy cảm cháy, Van The Pham và cộng sự [10] đã thực hiện nghiên cứu tại tỉnh Đắk Nông. Điểm mới của nghiên cứu này là kết hợp dữ liệu viễn thám từ UAV, ảnh SPOT và giải thuật XGBoost để phân loại các loại hình rừng (đạt độ chính xác 81,4%). Để đánh giá mức độ nhạy cảm với cháy, nhóm tác giả đã sử dụng mô hình kết hợp lai ghép giữa giải thuật tối ưu bầy ong nhân tạo (ABC) và hệ thống suy diễn mờ-rõ thích nghi (ANFIS). Kết quả cho thấy phần lớn diện tích rừng trong vùng, đặc biệt là rừng tre nứa và rừng dầu, đang ở mức nguy cơ cháy cao đến rất cao.

Đỗ Thị Ngọc Ánh và cộng sự [11] đã trình bày quá trình xây dựng ứng dụng phát hiện đám cháy sử dụng mạng nơ-ron tích chập YOLOv5. Ứng dụng sử dụng mô hình mạng nơ-ron YOLOv5 để phát hiện đám cháy đã cho kết quả thử nghiệm đạt độ chính xác cao và tốc độ xử lý nhanh có khả năng nhận dạng theo thời gian thực. Với thực trạng cháy nổ hiện tại đồng thời qua phân tích các hệ thống cảnh báo cháy đang có trên thị trường, hệ thống nhận dạng ngọn lửa sử dụng mạng nơ-ron khi kết hợp với các hệ thống camera an ninh hứa hẹn sẽ cải thiện được nhiều thiếu sót của các hệ thống báo cháy thông thường đang gặp phải, giúp giảm thiểu thiệt hại về người và cơ sở vật chất khi xảy ra hỏa hoạn.

Các nghiên cứu liên quan cho thấy xu hướng ngày càng rõ rệt trong việc sử dụng các mô hình học sâu, đặc biệt là các kiến trúc kết hợp không gian - thời gian như ConvLSTM, để phát hiện và dự báo cháy rừng. Tuy nhiên, vẫn còn những thách thức về kích thước và tính đại diện của dữ liệu, cũng như khả năng triển khai thực tế. Bài toán đặt ra trong nghiên cứu này thuộc loại phân loại nhị phân ảnh có/không có lửa hoặc khói. Dữ liệu huấn luyện không phân biệt các loại hình cháy khác nhau (cháy rừng, cháy bụi, cháy trại, đốt nương làm rẫy,...). Do đó, kết quả đạt được phản ánh khả năng của mô hình trong việc nhận diện các hiện tượng cháy nói chung chứ không phải riêng cháy rừng. Để đóng góp vào hướng nghiên cứu này, mục tiêu của nghiên cứu tập trung vào ba điểm chính. Thứ nhất, phát triển một mô hình ConvLSTM cho bài toán phân loại nhị phân (có cháy hoặc không có cháy). Thứ hai, so sánh đôi chiều hiệu năng của mô hình đề xuất với các mô hình cơ sở là CNN và LSTM truyền thống. Thứ ba, chỉ ra các hạn chế liên quan đến kích thước bộ dữ liệu, hiện tượng mất cân bằng giữa các lớp dữ liệu, và tính đặc thù về mặt địa lý của dữ liệu thực nghiệm. Nghiên cứu này tập trung vào phát triển mô hình ConvLSTM cho bối cảnh cụ thể tại Việt Nam.

2. Phương pháp đề xuất xây dựng hệ thống cảnh báo cháy rừng

Hệ thống phát hiện cháy rừng được phát triển trong nghiên cứu này có hai chức năng chính. Thứ nhất, phát hiện sớm và cảnh báo. Hệ thống sử dụng camera quang học và các cảm biến để

phát hiện các dấu hiệu ban đầu của cháy như nhiệt độ cao, điều kiện khô hạn, khói và lửa. Điều này giúp các cơ quan quản lý rừng phản ứng nhanh hơn và giảm thiểu thiệt hại. Thứ hai, cung cấp dữ liệu thời tiết và hình ảnh theo thời gian thực. Hệ thống cập nhật thông tin thời tiết và hình ảnh trực tiếp từ thực địa, giúp người quản lý có cái nhìn rõ ràng và chi tiết về tình hình, từ đó đưa ra quyết định tốt hơn.

2.1. Thu thập dữ liệu và tiền xử lý ảnh

Bộ dữ liệu được sử dụng trong nghiên cứu này là bộ dữ liệu công khai FIRE Dataset [12], được tạo bởi Ahmed Saied và cộng sự trong khuôn khổ cuộc thi NASA Space Apps Challenge 2018 và được chia sẻ trên nền tảng Kaggle. Bộ dữ liệu gồm 999 ảnh, được tổ chức thành hai thư mục: fire_images (755 ảnh) chứa các ảnh cháy ngoài trời, một số có khói dày đặc; và non-fire_images (244 ảnh) chứa các ảnh thiên nhiên như rừng, cây, cỏ, sông, hồ. Dữ liệu được thu thập với mục đích phân loại nhị phân ảnh có cháy và không có cháy. Tỷ lệ mất cân bằng giữa hai lớp (75% có cháy, 25% không cháy) phản ánh một tình huống không lý tưởng nhưng phổ biến trong các ứng dụng học máy thực tế. Bộ dữ liệu này đã được sử dụng trong nhiều nghiên cứu liên quan đến phát hiện cháy từ ảnh. Tuy nhiên, cần lưu ý rằng bộ dữ liệu không cung cấp thông tin chi tiết về độ phân giải gốc, điều kiện chụp, hoặc tiêu chí gán nhãn cụ thể, đồng thời chưa đại diện cho các điều kiện địa lý và khí hậu đặc thù của Việt Nam. Bộ dữ liệu được sử dụng có mục đích đơn giản: huấn luyện mô hình máy tính phân biệt ảnh có cháy và ảnh không có cháy. Đây là bài toán phân loại nhị phân cơ bản trong học máy, đặc biệt phù hợp cho phát hiện sớm cháy rừng. Bộ dữ liệu gồm 999 ảnh, được tổ chức thành hai thư mục riêng biệt. Sự phân tách này rất quan trọng vì nó cung cấp cho thuật toán các ví dụ rõ ràng cho từng loại. Thư mục "fire images" có 755 ảnh, tất cả đều hiển thị lửa nhìn thấy được. Hầu hết là các cảnh cháy ngoài trời, từ đám lửa trại nhỏ đến các đám cháy bụi lớn hơn. Nhiều ảnh trong số này cũng có khói rõ rệt. Thư mục "non-fire images" có 244 ảnh đóng vai trò là tập đối chứng. Đây là các cảnh thiên nhiên như rừng rậm, cây đơn lẻ, bãi cỏ, sông và hồ. Bằng cách xem xét các ảnh này, mô hình học được các mẫu, màu sắc và kết cấu của cảnh quan an toàn. Sự mất cân bằng trong bộ dữ liệu (khoảng 75% ảnh có cháy, 25% ảnh không cháy) là tình huống phổ biến trong các ứng dụng học máy thực tế.

Trong bước tiền xử lý, chúng tôi xử lý ảnh thô để thuật toán có thể sử dụng được. Một phần quan trọng của bước này là loại bỏ nhiễu. Chúng tôi cũng sử dụng kỹ thuật tăng cường dữ liệu. Kỹ thuật này mở rộng tập dữ liệu huấn luyện bằng cách tạo ra các phiên bản biến đổi từ ảnh gốc. Chúng tôi xoay ảnh, lật ảnh và điều chỉnh độ sáng. Huấn luyện trên dữ liệu đa dạng hơn giúp mô hình xử lý các ảnh mới và chưa gặp tốt hơn. Quan trọng hơn, tăng cường dữ liệu giúp ngăn ngừa hiện tượng quá khớp, một vấn đề phổ biến khi mô hình trở nên quá chuyên biệt cho dữ liệu huấn luyện và hoạt động kém trên dữ liệu mới. Các bước này là cần thiết để xây dựng một mô hình vừa chính xác vừa bền vững.

Quy trình tiền xử lý ảnh: Toàn bộ ảnh đầu vào được điều chỉnh về kích thước cố định 128×128 pixel bằng phương pháp nội suy song tuyến tính để đồng nhất kích thước trước khi đưa vào mô hình. Giá trị pixel của mỗi kênh màu (RGB) được chuẩn hóa về khoảng [0, 1] bằng cách chia cho 255. Không áp dụng các kỹ thuật lọc nhiễu hoặc xử lý ảnh nâng cao khác nhằm bảo toàn đặc trưng tự nhiên của ảnh gốc.

Tăng cường dữ liệu: Do kích thước bộ dữ liệu còn hạn chế (999 ảnh), nhóm tác giả áp dụng các phép biến đổi ngẫu nhiên trên dữ liệu huấn luyện (không áp dụng trên tập kiểm tra) để tăng độ đa dạng: xoay ảnh trong khoảng $\pm 20^\circ$, lật ngang, và điều chỉnh độ sáng trong khoảng $\pm 20\%$. Các phép biến đổi này được thực thi trong khi huấn luyện thông qua lớp ImageDataGenerator của Keras.

Xử lý mất cân bằng lớp: Bộ dữ liệu ban đầu có tỷ lệ mất cân bằng (75% ảnh có cháy, 25% ảnh không cháy). Để hạn chế ảnh hưởng của sự mất cân bằng này, nghiên cứu sử dụng trọng số lớp (class weights) trong hàm mất mát, với trọng số được tính tỉ lệ nghịch với tần suất xuất hiện của mỗi lớp trong tập huấn luyện. Cụ thể, trọng số cho lớp không cháy được đặt cao hơn nhằm tăng cường độ phạt cho các sai lệch đối với lớp thiểu số.

2.2. Kiến trúc mô hình ConvLSTM

Hệ thống của chúng tôi sử dụng mạng nơ-ron có tên ConvLSTM. Lý do chọn ConvLSTM vì nó phù hợp để xử lý các chuỗi ảnh. Cháy rừng phát triển và di chuyển theo thời gian, và ConvLSTM có khả năng phát hiện các dạng diễn biến này. Cụm từ "Conv" trong ConvLSTM chỉ các lớp tích chập. Chúng tôi xây dựng mô hình với một số lớp như vậy để giúp nó phân biệt ảnh có cháy với ảnh không có cháy. Các lớp này hoạt động như một bộ lọc. Chúng quét từng ảnh và tìm kiếm các dạng hình ảnh quan trọng. Mô hình không được chỉ định chính xác phải tìm gì. Thay vào đó, nó tự học để nhận ra hình dạng, đường viền, màu sắc và kết cấu điển hình của lửa. Mỗi lớp tích chập nhìn ảnh ở một mức độ chi tiết khác nhau. Các lớp đầu tiên phát hiện các thứ đơn giản như đường viền. Các lớp sau kết hợp các mẫu đơn giản này thành các mẫu phức tạp hơn. Quá trình từng bước này giúp mô hình xây dựng sự hiểu biết tốt về cả ảnh có cháy và ảnh không có cháy, làm cho dự đoán chính xác hơn. Do bộ dữ liệu đầu vào là các ảnh tĩnh, không phải chuỗi video liên tục, nghiên cứu này áp dụng kỹ thuật cửa sổ trượt (sliding window) trên các biến thể tăng cường dữ liệu để mô phỏng chuỗi thời gian cần thiết cho ConvLSTM. Cụ thể, với mỗi ảnh gốc, chúng tôi tạo ra K biến thể tăng cường (xoay, lật, điều chỉnh độ sáng) và sắp xếp chúng thành một chuỗi có thứ tự xác định. Cửa sổ trượt có kích thước T (số khung hình trong một chuỗi) sẽ trượt qua tập hợp các biến thể này để tạo ra các mẫu chuỗi đầu vào cho mô hình.

Phần "LSTM" có nghĩa là bộ nhớ dài-ngắn. Thành phần này xử lý các chuỗi và theo dõi thông tin trước đó. Sau khi các lớp tích chập xử lý từng khung hình, LSTM tiếp tục công việc. Nó không chỉ xem xét một khung hình đơn lẻ. Thay vào đó, nó phân tích các đặc trưng từ các khung hình liên tiếp và quan sát cách các đặc trưng đó thay đổi theo thời gian. Điều này rất quan trọng vì lửa không phải là hiện tượng tĩnh. Lớp LSTM nắm bắt được sự chuyển động, sự lan rộng và sự thay đổi cường độ của đám cháy qua các khung hình.

Cuối cùng, thông tin từ LSTM được chuyển đến một vài lớp kết nối đầy đủ. Các tham số huấn luyện bao gồm kích thước batch bằng 8, bộ tối ưu Adam, tốc độ học 0,001, kỹ thuật dừng sớm và tỷ lệ dropout 0,3. Các lớp này đóng vai trò là người ra quyết định. Chúng lấy tất cả thông tin không gian và thời gian đã được phân tích, bao gồm hình dạng, màu sắc, kết cấu của lửa và khối cũng như cách chúng thay đổi, sau đó cân nhắc các bằng chứng. Đầu ra là một điểm xác suất từ 0 đến 1. Điểm gần 1 có nghĩa là hệ thống tin rằng có cháy. Điểm gần 0 có nghĩa là cảnh bình thường. Mô hình đầu ra giá trị xác suất trong khoảng $[0, 1]$, trong đó giá trị càng gần 1 cho biết khả năng ảnh có cháy càng cao. Ngưỡng phân loại mặc định được sử dụng trong nghiên cứu này là 0,5: ảnh được phân loại là "có cháy" nếu xác suất $\geq 0,5$ và "không cháy" nếu xác suất $< 0,5$. Ngưỡng này được chọn vì là giá trị trung lập phổ biến trong các bài toán phân loại nhị phân.

Bảng 1. Cấu trúc kiến trúc mô hình ConvLSTM

Thành phần/ Tham số	Giá trị/ Mô tả
Số lớp ConvLSTM	1 lớp ConvLSTM2D
Số bộ lọc (filters)	32
Kích thước kernel	3×3
Hàm kích hoạt	ReLU
Lớp Flatten	Chuyển đặc trưng từ không gian 4D sang vector 1D
Lớp Dense thứ nhất	64 neuron, hàm kích hoạt ReLU
Dropout	0,3
Bộ tối ưu (Optimizer)	Adam
Tốc độ học (Learning rate)	0,001
Kích thước batch (Batch size)	8
Số epoch tối đa	100 (sử dụng dừng sớm)
Hàm mất mát (Loss function)	Binary Crossentropy
Chỉ số đánh giá	Accuracy

Các thực nghiệm được tiến hành trên cùng bộ dữ liệu với ba mô hình: RNN thuần, CNN thuần và ConvLSTM đề xuất. Việc so sánh dựa trên các độ đo độ chính xác, precision, recall và F1-

score. Mô hình ConvLSTM được kỳ vọng vượt trội hơn nhờ khả năng khai thác đồng thời đặc trưng không gian từ các lớp tích chập và đặc trưng thời gian từ cấu trúc LSTM, trong khi RNN và CNN chỉ xử lý được một trong hai loại thông tin này.

Kiến trúc mô hình ConvLSTM được trình bày chi tiết trong Bảng 1. Mô hình được xây dựng trên nền tảng Keras, với đầu vào là các chuỗi ảnh có kích thước (timesteps, height, width, channels) và đầu ra là xác suất nhị phân thuộc một trong hai lớp (có cháy hoặc không có cháy).

3. Kết quả thực nghiệm và phân tích

Để đánh giá đúng hiệu năng của thuật toán, việc lựa chọn các chỉ số đo lường phù hợp là rất quan trọng. Trong nghiên cứu này, chúng tôi sử dụng bốn chỉ số: độ chính xác (accuracy), độ chính xác dự báo dương tính (precision), độ nhạy (recall) và F1-score. Các chỉ số này phản ánh khả năng dự đoán của mô hình trong môi trường thực tế, điều rất cần thiết cho bài toán phát hiện cháy rừng.

Môi trường thực nghiệm: Chúng tôi thực hiện các thí nghiệm trên bộ dữ liệu 999 ảnh đã được mô tả ở phần trước. Dữ liệu được chia thành tập huấn luyện (80%) và tập kiểm tra (20%). Quá trình huấn luyện sử dụng bộ tối ưu Adam với tốc độ học 0,001, kích thước batch bằng 8, và kỹ thuật dừng sớm để tránh quá khớp. Tỷ lệ dropout được đặt là 0,3. Mô hình ConvLSTM đề xuất được so sánh với hai mô hình cơ sở là RNN thuần và CNN thuần trên cùng điều kiện thí nghiệm: Để đảm bảo tính công bằng trong so sánh và khả năng tái lập kết quả, ba mô hình (LSTM thuần, CNN thuần và ConvLSTM) được thiết kế với số lượng tham số tương đương nhau và huấn luyện trên cùng một bộ dữ liệu với cùng điều kiện (batch size = 8, optimizer = Adam, learning rate = 0,001, dropout = 0,3, early stopping với patience = 15). Để ngăn ngừa hiện tượng quá khớp, nghiên cứu áp dụng kỹ thuật dừng sớm (early stopping) trong quá trình huấn luyện. Hàm mất mát trên tập kiểm định (validation loss) được sử dụng làm chỉ số chính để quyết định thời điểm dừng. Quá trình huấn luyện sẽ dừng khi giá trị validation loss không giảm trong 15 epoch liên tiếp (patience = 15). Trong suốt quá trình huấn luyện, trọng số của mô hình tại epoch có validation loss thấp nhất được tự động lưu lại và sử dụng để đánh giá cuối cùng, thay vì sử dụng trọng số ở epoch cuối cùng. Với các tham số trên, quá trình huấn luyện mô hình ConvLSTM đạt giá trị validation loss thấp nhất tại epoch thứ 42 và được tự động dừng ở epoch thứ 57 sau 15 epoch không cải thiện. Các kết quả báo cáo trong Bảng 2 được trích xuất từ mô hình được lưu tại epoch thứ 42, đánh giá trên tập kiểm tra độc lập. Tập kiểm tra không được sử dụng trong bất kỳ giai đoạn nào của quá trình huấn luyện (bao gồm điều chỉnh siêu tham số, lựa chọn mô hình, hoặc quyết định dừng sớm). Việc này đảm bảo tính khách quan của các chỉ số đánh giá cuối cùng và phản ánh đúng khả năng tổng quát hóa của mô hình trên dữ liệu mới.

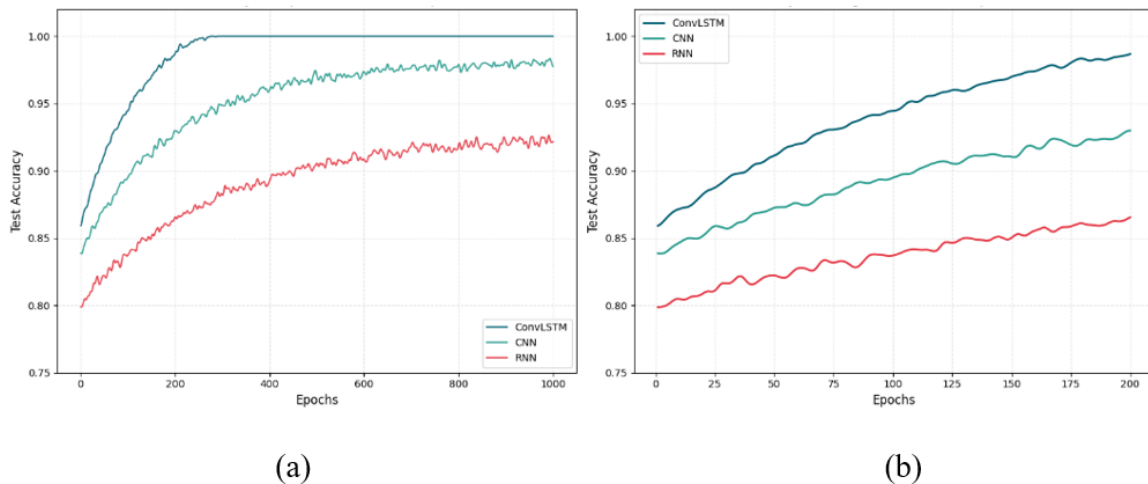
Kết quả thực nghiệm trên tập kiểm tra được trình bày trong Bảng 2. Có thể thấy mô hình ConvLSTM đề xuất đạt kết quả tốt nhất trên tất cả các chỉ số.

Bảng 2. So sánh các chỉ số hiệu năng giữa các mô hình

STT	Mô hình	Độ chính xác (%)	Precision (%)	Recall (%)	F1-score (%)
1	RNN	94,2	93,8	94,5	94,1
2	CNN	96,7	96,3	97,1	96,7
3	ConvLSTM	98,3	98,1	98,6	98,2

Kết quả cho thấy mô hình ConvLSTM đạt độ chính xác 98,3%, cao hơn CNN khoảng 1,6% và cao hơn RNN khoảng 4,1%. Precision đạt 98,1% cho thấy khi mô hình dự đoán có cháy, khả năng đúng là rất cao, giúp giảm báo động giả. Recall đạt 98,6% phản ánh khả năng phát hiện đúng các đám cháy, hạn chế bỏ sót. F1-score 98,2% là giá trị điều hòa giữa precision và recall, khẳng định mô hình hoạt động cân bằng.

Hình 1 thể hiện trực quan sự khác biệt về hiệu năng giữa ba mô hình. Có thể thấy ConvLSTM vượt trội trên cả bốn chỉ số, đặc biệt ở recall và F1-score. RNN cho kết quả thấp nhất do không trích xuất được đặc trưng không gian từ ảnh. CNN tốt hơn RNN nhờ các lớp tích chập, nhưng vẫn thua ConvLSTM vì thiếu khả năng mô hình hóa chuỗi thời gian.

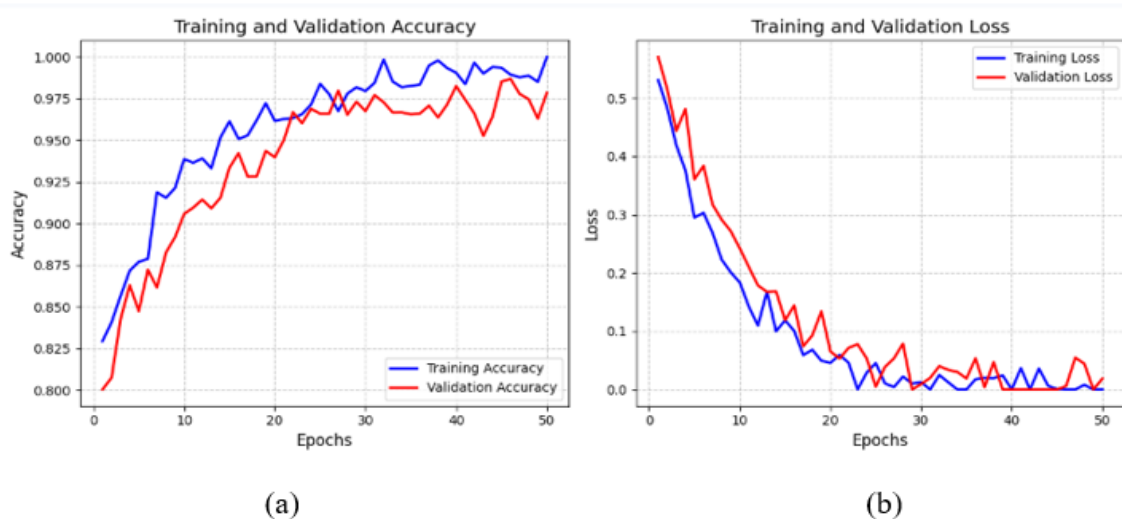


Hình 1. Biểu đồ so sánh hiệu năng của RNN, CNN và ConvLSTM. Thử nghiệm với số epoch khác nhau: (a) -1000 epoch, (b) – 200 epoch

Hai biểu đồ trong Hình 2 lần lượt thể hiện đường cong học tập (training loss & validation loss) và độ chính xác trên tập kiểm tra qua các epoch. Hai đồ thị có số epoch khác nhau do được trích xuất từ hai lần chạy thử nghiệm với các điều kiện khởi tạo khác nhau; tuy nhiên, mô hình đánh giá cuối cùng được lấy tại epoch tốt nhất (validation loss thấp nhất) và tập kiểm tra không được sử dụng trong quá trình điều chỉnh mô hình hay dừng sớm, chỉ dùng để đánh giá cuối cùng sau khi huấn luyện kết thúc.

Mô hình RNN đạt độ chính xác 94,2% do có khả năng xử lý chuỗi thời gian, nhưng khi xử lý ảnh, nó phải làm phẳng ma trận ảnh thành vector một chiều. Điều này làm mất hoàn toàn cấu trúc không gian vốn rất quan trọng để nhận diện hình dạng lửa, màu sắc khói, cũng như mối quan hệ giữa các pixel lân cận. Do đó, RNN hoạt động kém nhất trong ba mô hình.

Mô hình CNN đạt kết quả cao hơn RNN (96,7%) vì các lớp tích chập giữ được cấu trúc không gian của ảnh. CNN có thể học được các đặc trưng về đường viền, màu sắc, kết cấu của lửa và khói. Tuy nhiên, CNN xử lý từng ảnh độc lập, không khai thác được mối liên hệ giữa các khung hình liên tiếp. Trong bối cảnh cháy rừng, việc một ảnh đơn lẻ có thể bị nhiễu do ánh sáng, bóng cây hoặc sương mù, nếu chỉ dựa vào một khung hình dễ dẫn đến nhầm lẫn.



Hình 2. Đường cong accuracy (a) và loss (b) trên tập huấn luyện và tập kiểm định để đánh giá hiện tượng quá khớp

Mô hình ConvLSTM đạt kết quả tốt nhất (98,3%) do kết hợp được ưu điểm của cả hai. Phân tích chấp trịch xuất đặc trưng không gian từ mỗi khung hình, trong khi phần LSTM theo dõi sự thay đổi của các đặc trưng này qua các khung hình liên tiếp. Ví dụ, khi có cháy, khói thường xuất hiện trước lửa và lan rộng dần, độ sáng của ảnh tăng lên, màu sắc chuyển từ xanh lá sang cam đỏ. ConvLSTM nắm bắt được các xu hướng thay đổi này, giúp phân biệt cháy thật với các đối tượng có màu sắc tương tự như đèn đỏ, ánh nắng chiều hoặc đám mây màu cam. Thêm vào đó, kỹ thuật tăng cường dữ liệu (xoay, lật, điều chỉnh độ sáng) giúp mô hình làm việc tốt hơn trên các điều kiện khác nhau. Kỹ thuật dropout 0,3 và dừng sớm giúp ngăn ngừa quá khớp, đặc biệt khi bộ dữ liệu còn nhỏ. Trong điều kiện vận hành thực tế với nhiều yếu tố môi trường phức tạp hơn, độ chính xác có thể giảm. Do đó, các kết quả này cần được xem xét trong bối cảnh thí nghiệm kiểm soát, và các nghiên cứu tiếp theo cần được tiến hành trên bộ dữ liệu lớn hơn, đa dạng hơn và mang tính đại diện cho điều kiện thực địa tại Việt Nam.

Nghiên cứu hiện tại mới dừng ở mức thử nghiệm trên bộ dữ liệu công khai với ảnh tĩnh, chưa tích hợp các yếu tố cần thiết cho vận hành thực tế như loại camera cụ thể, dữ liệu thời tiết, cơ chế cảnh báo, hay yêu cầu phần cứng. Việc triển khai hệ thống tại hiện trường đòi hỏi các nghiên cứu bổ sung về: (i) lựa chọn camera quan sát phù hợp với điều kiện địa hình và ánh sáng tại Việt Nam; (ii) tích hợp dữ liệu khí tượng (nhiệt độ, độ ẩm, hướng gió) để giảm báo động giả; (iii) xây dựng cơ chế phát cảnh báo tự động (qua SMS, ứng dụng di động, hoặc hệ thống quản lý tập trung); (iv) tối ưu mô hình để chạy trên thiết bị biên có tài nguyên hạn chế (GPU nhúng, camera thông minh). Đây là các hướng phát triển cần thiết trước khi có thể khẳng định tính khả thi của hệ thống trong điều kiện vận hành thực tế tại Việt Nam.

4. Kết luận

Trong nghiên cứu này, chúng tôi đã phát triển một mô hình ConvLSTM sử dụng dữ liệu ảnh để phát hiện cháy rừng. Mô hình có khả năng phân biệt giữa các cảnh có cháy và không có cháy với kết quả khả quan, đạt độ chính xác 98,3% trên tập dữ liệu kiểm tra. Việc sử dụng khuôn khổ Keras giúp việc xây dựng và huấn luyện mô hình trở nên đơn giản hơn.

Kết quả thực nghiệm cho thấy ConvLSTM vượt trội hơn so với các mô hình CNN và RNN thuần. Nguyên nhân là vì ConvLSTM vừa khai thác được đặc trưng không gian qua các lớp tích chập, vừa nắm bắt được sự phụ thuộc thời gian qua cấu trúc LSTM. Điều này rất phù hợp với bản chất của cháy rừng, vốn là một quá trình lan rộng và biến đổi liên tục theo thời gian.

Mô hình đã chứng tỏ khả năng nhận diện các dấu hiệu thị giác của lửa và khói một cách đáng tin cậy, do đó có thể được xem như một công cụ thực tiễn cho phát hiện sớm và giám sát cháy rừng. Bằng cách cung cấp cảnh báo sớm và cho phép phản ứng nhanh, hệ thống này có thể giúp giảm thiểu thiệt hại về môi trường do cháy rừng gây ra.

Tuy nhiên, nghiên cứu cũng tồn tại một số hạn chế. Bộ dữ liệu còn nhỏ (chỉ 999 ảnh) và mất cân bằng (75% là ảnh có cháy). Việc thiếu dữ liệu đặc thù cho điều kiện Việt Nam cũng là một điểm yếu.

Các hướng phát triển trong tương lai bao gồm cải thiện độ bền vững của mô hình, tích hợp thêm dữ liệu khí tượng để phân tích rủi ro toàn diện hơn, mở rộng tập dữ liệu huấn luyện với nhiều kịch bản cháy đa dạng hơn, và hướng tới triển khai thời gian thực trên các thiết bị biên như camera thông minh hoặc máy bay không người lái.

Lời cảm ơn

Bài báo là sản phẩm của đề tài có mã số ĐH2024-TN07-01, được tài trợ bởi kinh phí của trường Đại học Công nghệ Thông tin và Truyền thông – Đại học Thái Nguyên. Nhóm tác giả xin chân thành cảm ơn sự tài trợ của nhà trường.

TÀI LIỆU THAM KHẢO/ REFERENCES

- [1] A. B. Abdusalomov *et al.*, "An improved forest fire detection method based on Detectron2," *Sensors*, vol. 23, no. 3, 2023, Art. no. 1512.
- [2] J. Mei and W. Zhang, "Early fire detection algorithm based on ViBe and machine learning," *Acta Opt. Sin.*, vol. 38, 2018, Art. no. 0710001.
- [3] K. Dimitropoulos, P. Barmoutis, and N. Grammalidis, "Higher order linear dynamical systems for smoke detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, pp. 1143–1154, 2017.
- [4] J. Zhang *et al.*, "ATT squeeze U-net: A lightweight network for forest fire detection and recognition," *IEEE Access*, vol. 9, pp. 10858–10870, 2021.
- [5] R. Ghosh and A. Kumar, "A hybrid deep learning model combining CNN and RNN to detect forest fire," *Multimedia Tools Appl.*, vol. 81, pp. 38643–38660, 2022.
- [6] A. Larsen *et al.*, "A deep learning approach to identify smoke plumes in satellite imagery in near-real time," *J. Expos. Sci. Environ. Epidemiol.*, vol. 31, pp. 170–176, 2021.
- [7] X. Yin, M. Bai, Y. Peng, B. Chen, H. Lai, W. Kou, *et al.*, "LSTM neural network integrating remote sensing and meteorological data reveals forest fire risk trajectories in China over the past 100 years," *Agricultural and Forest Meteorology*, vol. 375, 2025, Art. no. 110884.
- [8] C. Mambile, S. Kaijage, and J. Leo, "Deep learning models for enhanced forest-fire prediction at Mount Kilimanjaro," *Natural Hazards Research*, vol. 5, no. 2, pp. 335–347, 2025.
- [9] H. V. Le, D. A. Hoang, and G. T. Tran, "Optimizing machine learning models for enhanced forest fire susceptibility mapping in Gia Lai province," *Journal of Mining and Earth Sciences*, vol. 66, no. 2, pp. 2-14, 2025.
- [10] V. T. Pham, T. A. T. Do, and H. D. Tran, "Classifying forest cover and mapping forest fire susceptibility in Dak Nong province, Vietnam utilizing remote sensing and machine learning," *Ecological Informatics*, vol. 79, 2024, Art. no. 102392, doi: 10.1016/j.ecoinf.2023.102392.
- [11] T. N. A. Do, M. K. Hoang, and A. T. Le, "Improvement the fire detection performance using computer vision based on YOLOV5 neural network," *Hanoi University of Industry Journal of Science and Technology*, vol. 5, pp. 48-53, 2022.
- [12] A. Saied *et al.*, "FIRE Dataset," Kaggle, 2018. [Online]. Available: <https://www.kaggle.com/datasets/phylake1337/fire-dataset>. [Accessed Aug. 07, 2025].