

APPLICATION OF MACHINE LEARNING MODELS FOR FAULT CLASSIFICATION AND LOCATION IN POWER TRANSMISSION LINES

Nguyen Quoc Minh*, Tran Van Tien, Bui Thi Phuong Thao, Nguyen Minh Hoang, Pham Thi Kim Huong

School of Electrical and Electronic Engineering - Hanoi University of Science and Technology

ARTICLE INFO	ABSTRACT
Received: 10/01/2025	This paper presents a method for fault classification and localization on transmission lines using machine learning models. The training data for these models were derived from simulations of the IEEE 9-bus power system in MATLAB Simulink software, with faults generated under various conditions. The study employed machine learning models such as Decision Tree, Logistic Regression, XGBoost and Artificial Neural Networks for fault classification, and recurrent neural network such as Convolutional Neural Network, Long Short-Term Memory, Gated Recurrent Unit, as well as hybrid models like CNN-LSTM and CNN-GRU for fault localization. By utilizing these machine learning models, the research focused on evaluating the accuracy of fault classification and localization on transmission lines with the goal of enhancing the stability and reliability of the power system while reducing fault recovery time. The results demonstrate the effectiveness of the machine learning models, with ANN achieving a fault classification accuracy of up to 99.974%, while CNN-GRU can localize faults with a mean absolute error of less than 0.029 km.
Revised: 27/02/2025	
Published: 27/02/2025	
KEYWORDS	
Fault analysis	
Fault localization	
Fault classification	
Power transmission line	
Machine learning	

NGHIÊN CỨU ỨNG DỤNG CÁC MÔ HÌNH HỌC MÁY NHẪM PHÂN LOẠI VÀ ĐỊNH VỊ SỰ CỐ TRÊN ĐƯỜNG DÂY TRUYỀN TẢI

Nguyễn Quốc Minh*, Trần Văn Tiến, Bùi Thị Phương Thảo, Nguyễn Minh Hoàng, Phạm Thị Kim Hương

Trường Điện – Điện tử, Đại học Bách khoa Hà Nội

THÔNG TIN BÀI BÁO	TÓM TẮT
Ngày nhận bài: 10/01/2025	Nghiên cứu này đề xuất phương pháp phân loại và định vị sự cố trên đường dây truyền tải dựa trên các mô hình học máy. Dữ liệu huấn luyện cho các mô hình được tạo ra từ việc mô phỏng lưới điện IEEE 9 nút trong phần mềm Matlab Simulink với các sự cố tạo ra trong nhiều điều kiện khác nhau. Các thuật toán học máy như Decision Tree, Logistic Regression, XGBoost, Support Vector Machine, Artificial Neural Network được sử dụng để phân loại sự cố, trong khi các mô hình mạng nơ ron hồi quy như Convolutional Neural Network, Long-Short Term Memory, Gated Recurrent Unit, mô hình kết hợp CNN-LSTM và CNN-GRU được sử dụng để định vị sự cố. Bằng cách sử dụng các mô hình học máy, nghiên cứu tập trung xem xét độ chính xác của việc phân loại và định vị sự cố trên đường dây truyền tải với mục tiêu nâng cao độ tin cậy và ổn định của hệ thống điện đồng thời góp phần giảm thời gian khắc phục sự cố. Kết quả cho thấy hiệu quả của các mô hình học máy, trong đó ANN đạt độ chính xác phân loại sự cố là 99,974%, còn CNN-GRU đạt sai số trung bình trong định vị sự cố là 0,029 km.
Ngày hoàn thiện: 27/02/2025	
Ngày đăng: 27/02/2025	
TỪ KHÓA	
Phân tích sự cố	
Định vị sự cố	
Phân loại sự cố	
Đường dây truyền tải	
Học máy	

DOI: <https://doi.org/10.34238/tnu-jst.11857>

* Corresponding author. Email: minh.nguyenquoc@hust.edu.vn

1. Giới thiệu

Sự phát triển mạnh mẽ của đô thị hóa và công nghiệp hóa đã dẫn đến nhu cầu sử dụng điện ngày càng tăng. Để đảm bảo cung cấp điện an toàn, liên tục và bền vững, hệ thống điện phải không ngừng mở rộng và vận hành với hiệu suất cao. Đường dây truyền tải điện đóng vai trò then chốt trong việc truyền tải điện năng từ các nhà máy điện đến các khu vực tiêu thụ, đảm bảo hệ thống điện vận hành hiệu quả và liên tục [1]. Song song với đó, việc áp dụng công nghệ hiện đại vào quản lý và vận hành hệ thống truyền tải điện là yếu tố không thể thiếu. Điều này giúp nâng cao khả năng phát hiện và định vị sự cố, tăng cường tính linh hoạt và khả năng ứng phó nhanh chóng của toàn hệ thống [2].

Hệ thống truyền tải đã và đang phải đối mặt với nhiều thách thức, đặc biệt là sự cố có thể xảy ra trên đường dây, làm gián đoạn cung cấp điện và gây ra những thiệt hại kinh tế nghiêm trọng [3]. Việc phát hiện, phân loại và định vị sự cố trên đường dây truyền tải đóng vai trò quan trọng trong việc đảm bảo độ tin cậy và ổn định của hệ thống điện. Trong những năm qua, các phương pháp khác nhau đã được đề xuất để giải quyết các thách thức này, từ các phương pháp truyền thống đến các cách tiếp cận hiện đại sử dụng học máy. Các phương pháp truyền thống để phân tích sự cố trên đường dây truyền tải bao gồm kỹ thuật dựa trên trở kháng [4], phương pháp sóng truyền và biến đổi wavelet [5]. Kỹ thuật dựa trên trở kháng tính toán khoảng cách đến điểm sự cố bằng cách sử dụng dữ liệu điện áp và dòng điện, nhưng độ chính xác của chúng bị ảnh hưởng bởi các yếu tố như điện trở sự cố và điều kiện hệ thống. Phương pháp sóng truyền dựa vào sự lan truyền của tín hiệu chuyển tiếp do sự cố gây ra, mang lại độ chính xác cao nhưng đòi hỏi các hệ thống đo lường tốc độ cao và đồng bộ [6]. Kỹ thuật biến đổi wavelet cung cấp khả năng phân tích tín hiệu sự cố ở nhiều độ phân giải, nhưng đi kèm với các phép tính phức tạp [7].

Với sự nổi bật của các phương pháp học máy với các tham số và siêu tham số tối ưu, những thách thức nêu trên có thể được giải quyết hiệu quả và chính xác hơn [8]. Phương pháp học máy dựa trên ý tưởng các mô hình có thể tìm được những đặc trưng quan trọng từ dữ liệu, trích xuất các quan hệ giữa các biến đầu vào và cải thiện độ chính xác qua từng bước huấn luyện [9]. Với sự hiện đại hóa được áp dụng trong vận hành và bảo vệ thống điện, các rơ-le hiện đại có thể tác động chính xác và lưu trữ những dữ liệu quan trọng khi xảy ra sự cố. Những dữ liệu này là cơ sở để việc áp dụng học máy vào phân tích sự cố trở nên thực tiễn.

Trong nghiên cứu này, nhóm tác giả áp dụng các mô hình học máy để giải quyết hai bài toán: phân loại và định vị sự cố trên đường dây truyền tải. Gần 60.000 kịch bản sự cố được tạo ra để phục vụ bài toán phân loại, và 240.000 kịch bản sự cố được xây dựng cho bài toán định vị. Các kịch bản này được mô phỏng trên phần mềm Matlab Simulink, bao gồm đầy đủ các yếu tố ảnh hưởng như loại sự cố, vị trí sự cố, điện trở sự cố, và mức độ mang tải tại thời điểm xảy ra sự cố.

2. Phương pháp nghiên cứu

2.1. Các mô hình học máy

Học máy, một công nghệ mang tính cách mạng, giúp máy tính cải thiện hiệu suất thông qua việc học từ dữ liệu mà không cần lập trình cụ thể. Dựa trên các thuật toán phức tạp, nó cho phép hệ thống nhận dạng mẫu, đưa ra dự đoán và tối ưu hóa các quyết định. Những thuật toán này thuộc ba loại: học có giám sát, học không giám sát và học tăng cường. Học có giám sát xây dựng các hàm từ dữ liệu huấn luyện, dự đoán kết quả dựa trên các cặp đầu vào và đầu ra đã biết [10]. Trong nghiên cứu này, nhóm nghiên cứu sử dụng các mô hình học máy có giám sát như Logistic Regression, Decision Tree, Support Vector Machine, XG Boost, ANN cho phân loại sự cố và các mô hình CNN, LSTM, GRU, CNN-LSTM, CNN-GRU. Cách tiếp cận này cung cấp góc nhìn mới nhằm tối ưu hóa các phương pháp phân loại và định vị sự cố, minh chứng cho tính linh hoạt và tiềm năng của học máy trong các ứng dụng thực tế. Dưới đây là các mô hình học máy được sử dụng trong nghiên cứu này.

Extreme Gradient Boosting: XGBoost là một thuật toán học máy tiên tiến dựa trên phương pháp gradient boosting, rất hiệu quả cho các bài toán phân loại. XGBoost xây dựng một tập hợp

các mô hình (thường là cây quyết định) theo cách tuần tự. Mỗi cây mới được huấn luyện để sửa chữa các lỗi của tập hợp mô hình trước đó. Mô hình dự đoán bằng cách kết hợp đầu ra của nhiều cây khác nhau. XGBoost sử dụng một hàm mục tiêu bao gồm cả hàm mất mát để đo lường hiệu suất của mô hình và một điều khoản điều chỉnh để ngăn ngừa hiện tượng overfitting.

Decision Tree: Mô hình DT hoạt động bằng cách phân chia dữ liệu thành các tập con dựa trên thuộc tính quan trọng nhất ở mỗi bước, tạo ra một hình dạng cây với các quyết định và các kết quả có thể xảy ra.

Logistic Regression: LR là một phương pháp phân loại thống kê lý tưởng cho dữ liệu có mối quan hệ tuyến tính giữa các biến đầu vào và xác suất đầu ra. Cơ chế hoạt động của LR như sau. LR sẽ sử dụng một phương trình tuyến tính như sau:

$$z = w^T x + b \quad (1)$$

Trong đó, z là giá trị đầu ra của phần tuyến tính; w^T là vector chuyển vị của vector w với w là vector trọng số ứng với mức độ quan trọng của biến đầu ra trong quá trình dự đoán; x là vector biến đầu vào (hay còn gọi là đặc trưng) dùng để dự đoán đầu ra; b là hệ số điều chỉnh giúp đường tuyến tính dịch chuyển linh hoạt để phù hợp hơn với dữ liệu.

Khi giải bài toán phân loại nhị phân, giá trị z được chuyển đổi thành xác suất thông qua hàm sigmoid. Công thức của hàm sigmoid là:

$$P(y = 1|x) = \frac{1}{1 + e^{-z}} \quad (2)$$

Trong đó, $P(y = 1|x)$ là xác suất có điều kiện mà biến mục tiêu y thuộc về lớp dương khi biết vector đầu vào x ; e là cơ sở của logarit tự nhiên.

Hàm sigmoid (2) biến đổi z với giá trị có thể biến thiên trong khoảng từ $-\infty$ đến $+\infty$ thành một giá trị trong khoảng $(0, 1)$, phù hợp để diễn giải dưới dạng xác suất.

Support Vector Machine: SVM tìm ra các siêu phẳng tối ưu trong không gian đa chiều, xuất sắc trong việc phân loại các điểm dữ liệu. Chúng có khả năng xử lý tốt các nhiệm vụ với khoảng cách lớp rõ ràng. Mặc dù linh hoạt, SVM yêu cầu phải điều chỉnh tham số cẩn thận, đặc biệt là đối với các tập dữ liệu lớn, do độ phức tạp tính toán của chúng [10].

Artificial Neural Network: Cấu trúc ANN bao gồm 3 lớp với số lượng nơ-ron được tối ưu hóa cho mỗi lớp như sau.

Lớp đầu vào: Nhận dữ liệu từ môi trường bên ngoài và mỗi nơ-ron của lớp đầu vào đại diện cho mỗi đặc trưng của dữ liệu.

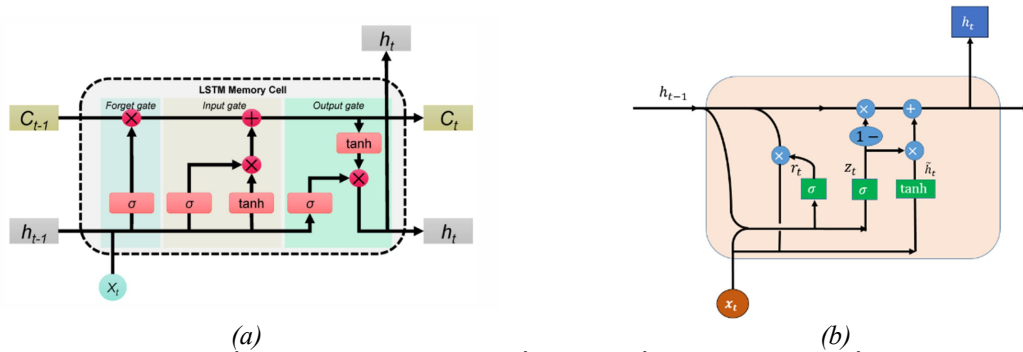
Lớp ẩn: Thực hiện quá trình xử lý thông tin và tính toán thông qua các trọng số và độ lệch để giảm thiểu lỗi giữa đầu ra dự đoán và đầu ra thực tế. Mỗi nơ-ron tạo ra một tín hiệu từ lớp trước thông qua hàm quy tuyến tính đa biến và hàm kích hoạt.

Lớp đầu ra: Bao gồm các nơ-ron dự đoán đầu ra của mô hình [10].

Convolutional Neural Network: CNN là mạng nơ-ron chuyên về các nhiệm vụ liên quan đến dữ liệu dạng lưới, chẳng hạn như hình ảnh và video. Các lớp tích chập của chúng cho phép nắm bắt các mẫu không gian, khiến chúng trở nên cần thiết trong các ứng dụng nhận dạng hình ảnh và hiểu hình ảnh. Nhưng với những ưu điểm đó CNN lại không hiệu quả trong việc xử lý các dữ liệu tuần tự vì không có cơ chế ghi nhớ thông tin theo thời gian. [11].

Long-Short Term Memory: Mạng LSTM được thiết kế để giải quyết vấn đề phụ thuộc xa với đặc tính mặc định là ghi nhớ thông tin trong suốt thời gian dài. Mạng LSTM bao gồm nhiều tế bào LSTM liên kết với nhau. Cấu trúc cụ thể của mỗi tế bào được biểu diễn như trong (Hình 1a) [12].

Gated Recurrent Unit: GRU là một biến thể của mạng nơ-ron hồi tiếp được thiết kế nhằm khắc phục các hạn chế của RNN truyền thống như là giữ lại thông tin quan trọng, đơn giản hóa mô hình so với LSTM giúp giảm tốc độ tính toán và giảm yêu cầu về bộ nhớ, đặc biệt là vấn đề tiêu biến gradient (Hình 1b) [13].



Hình 1. Cấu trúc bên trong của các tế bào: (a) tế bào LSTM và (b) tế bào GRU

Kết hợp CNN-LSTM: Mô hình là sự kết hợp giữa mạng nơ-ron tích chập (CNN) và bộ nhớ dài ngắn hạn (LSTM), được thiết kế để tận dụng thế mạnh của cả hai mô hình [14]. Là một phương pháp tận dụng ưu điểm của CNN và dùng LSTM để khắc phục những nhược điểm của CNN, do đó CNN-LSTM thường được sử dụng cho các bài toán liên quan đến dữ liệu có cả đặc trưng không gian (hình ảnh, dữ liệu cảm biến) và thời gian (chuỗi thời gian). Bên cạnh các ưu điểm thì CNN-LSTM cũng có những nhược điểm như: Phức tạp hơn so với từng mô hình đơn lẻ, yêu cầu nhiều tài nguyên tính toán hơn, việc tối ưu hóa có thể khó khăn hơn do phải cân bằng giữa các thành phần CNN và LSTM.

Kết hợp CNN-GRU: CNN-GRU là một kiến trúc mạnh mẽ và linh hoạt, kết hợp khả năng trích xuất đặc trưng không gian của CNN và khả năng xử lý thông tin tuần tự của GRU. Với những ưu điểm tương tự mô hình CNN-LSTM nhưng CNN-GRU lại mang lại tốc độ nhanh hơn và tiêu tốn ít tài nguyên hơn do GRU đơn giản hơn LSTM. Dù ít phức tạp hơn nhưng đôi với một số bài toán nhất định thì cấu trúc CNN-GRU có thể mang lại kết quả tốt hơn CNN-LSTM. Điều này có thể phụ thuộc vào đặc điểm của bài toán, đặc điểm của dữ liệu và kích thước của bộ dữ liệu [15].

3. Kết quả và bàn luận

3.1. Phương pháp đánh giá các mô hình học máy

Với phân loại sự cố, tính hiệu quả của mô hình được đánh giá thông qua độ chính xác phân loại, ngoài ra thời gian huấn luyện của các mô hình cũng được đề cập.

Độ chính xác của các thuật toán phân loại được tính toán như sau:

$$\text{độ chính xác} = \frac{\text{số mẫu phân loại đúng}}{\text{tổng số mẫu}} \quad (3)$$

Trong nhiệm vụ định vị sự cố, đây là bài toán hồi quy với biến đầu ra là giá trị liên tục của vị trí sự cố (km). Để đánh giá được tính hiệu quả của các mô hình các chỉ số: Sai số toàn phương trung bình (MSE), sai số tuyệt đối trung bình (MAE), sai số phần trăm tuyệt đối trung bình (MAPE) được sử dụng.

MSE đo lường sự khác biệt trung bình giữa giá trị dự đoán ($\hat{y}_i$) và giá trị thực tế (y_i) (4):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

MAE tính sai số trung bình dưới dạng trị tuyệt đối của sự khác biệt giữa giá trị thực tế (y_i) và giá trị dự đoán ($\hat{y}_i$) (5).

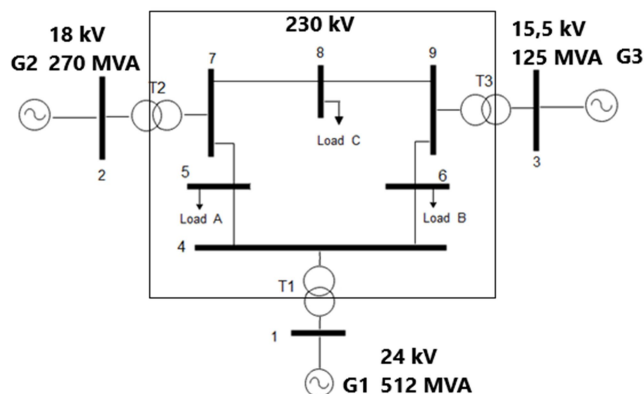
$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

MAPE tính sai số theo phần trăm, giúp chuẩn hóa giá trị sai số (6).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (6)$$

3.2. Phân loại sự cố

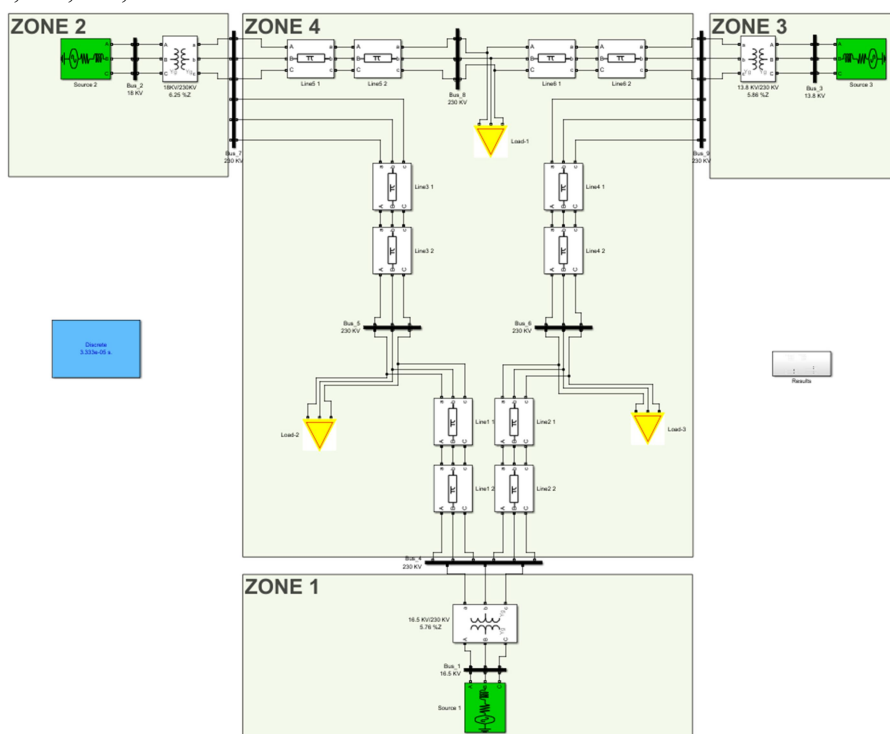
Dữ liệu sử dụng để huấn luyện và kiểm thử các mô hình phân loại sự cố được mô phỏng với các điều kiện khác nhau bao gồm cả trường hợp có sự cố và không có sự cố trên lưới điện IEEE 9 nút gồm 3 máy phát điện, 3 phụ tải và 3 máy biến áp (Hình 2) trong phần mềm Matlab Simulink.



Hình 2. Sơ đồ lưới điện IEEE 9 nút

Với trường hợp không có sự cố, 27.000 kịch bản được mô phỏng với 30 khả năng cho mỗi phụ tải, 30 khả năng đó gồm 3 nhóm 45% đến 55%, 65% đến 75%, 95% đến 105% giá trị định mức với bước thay đổi là 1%.

Với trường hợp có sự cố ngắn mạch, các sự cố được mô phỏng lần lượt tại giữa 12 vị trí trong lưới điện IEEE 9 nút bao gồm: trước và sau các máy biến áp T1, T2, T3, giữa các đường dây 7-8, 8-9, 7-5, 9-6, 5-4, 6-4.



Hình 3. Lưới IEEE 9 nút trong phần mềm Matlab Simulink

Mỗi vị trí sự cố đó được mô phỏng với 10 loại sự cố khác nhau gồm: Pha A chạm đất (AG), pha B chạm đất (BG), pha C chạm đất (CG), pha A chạm pha B (AB), pha B chạm pha C (BC), pha A chạm pha C (AC), pha A chạm pha B chạm đất (ABG), pha B chạm pha C chạm đất

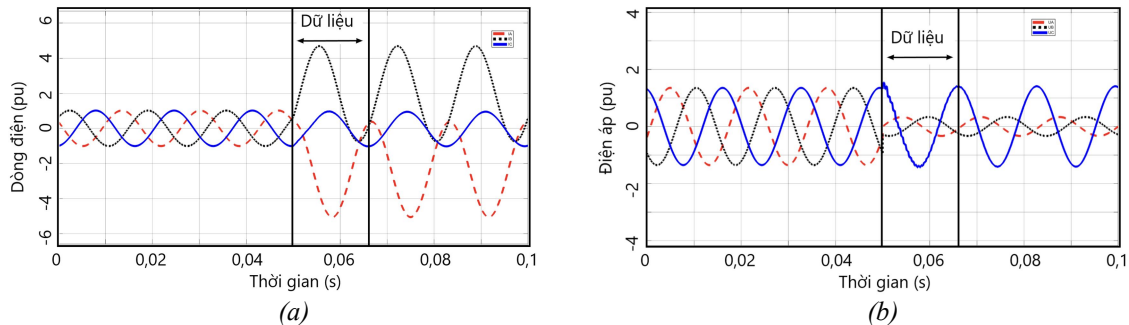
(BCG), pha A chạm pha C chạm đất (ACG), ba pha chạm nhau (ABC) và 10 giá trị kháng sự cố khác nhau là các giá trị điện trở tại vị trí sự cố. Tạo ra 100 khả năng khác nhau tại mỗi vị trí sự cố. Các phụ tải cũng được mô phỏng nhưng với 3 khả năng là 50%, 75% và 100%. Từ đó tạo ra 32.400 kịch bản cho trường hợp có sự cố.

Tổng 59.400 kịch bản có và không có sự cố, được gán với 17 nhãn khác nhau (Bảng 1) biểu thị khu vực xảy ra sự cố (Hình 3), loại sự cố và 1 nhãn không có sự cố.

Bảng 1. Các nhãn cho dữ liệu phân loại sự cố

Loại sự cố	Khu vực 1	Khu vực 2	Khu vực 3	Khu vực 4
1 pha chạm đất	Zone 1 LG	Zone 2 LG	Zone 3 LG	Zone 4 LG
2 pha	Zone 1 LL	Zone 2 LL	Zone 3 LL	Zone 4 LL
2 pha chạm đất	Zone 1 LLG	Zone 2 LLG	Zone 3 LLG	Zone 4 LLG
3 pha	Zone 1 LLL	Zone 2 LLL	Zone 3 LLL	Zone 4 LLL
Không có sự cố	Normal			

Các đặc trưng đầu vào của các mô hình sẽ là dòng điện và điện áp ở đầu cực của 3 máy phát trong lưới điện. Với mỗi đặc trưng của dòng điện và điện áp, dữ liệu lấy vào sẽ là 20 giá trị trong một chu kỳ đầu tiên sau khi xảy ra sự cố với tần số lấy mẫu là 1.200 Hz (Hình 4).



Hình 4. Dòng điện và điện áp đầu cực MF 1 khi bị sự cố 2 pha chạm đất sau MBA1: (a) dòng điện và (b) điện áp

Dữ liệu sau đó được chia ngẫu nhiên thành 2 phần với 80% dùng để huấn luyện các mô hình và 20% dùng để kiểm thử. Các mô hình Logistic Regression, Decision Tree, Support Vector Machine, XG Boost, ANN được sử dụng để phân loại sự cố. Sau khi huấn luyện và kiểm thử, độ chính xác của các mô hình được trình bày trong (Bảng 2).

Bảng 2. Độ chính xác của các mô hình phân loại sự cố

Mô hình	Logistic Regression	Decision Tree	Support Vector Machine	XG Boost	ANN
Độ chính xác (%)	94,023	99,579	96,953	99,941	99,974
Thời gian huấn luyện (s)	16,8	24,7	72,5	134,7	106,8

Có thể thấy độ chính xác của các mô hình đều >90%, đặc biệt các mô hình như Decision Tree, XG Boost, ANN có độ chính xác >99%. Nhưng thời gian huấn luyện của các mô hình này cũng lâu hơn so với các mô hình khác. Độ chính xác của mô hình ANN có độ chính xác cao nhất với 99,974% và thời gian huấn luyện là 106,8s.

3.3. Định vị sự cố

Với định vị sự cố, bộ dữ liệu được tạo ra bằng cách mô phỏng sự cố trên đường dây từ nút 4 đến nút 5 với chiều dài 89,93 km. Vị trí xảy ra sự cố sẽ là từng kilomet trên đường dây dài 89,93 km. Và tại mỗi vị trí sự cố cũng được mô phỏng tương tự như nhiệm vụ phân loại sự cố với 10 loại sự cố gồm AG, BG, CG, AB, BC, AC, ABG, BCG, ACG, ABC và 10 giá trị điện trở kháng sự cố khác nhau. Các phụ tải cũng được mô phỏng tương tự với 3 kịch bản cho mỗi phụ tải. Từ đó tạo ra bộ dữ liệu với 240.300 kịch bản để huấn luyện và kiểm thử các mô hình.

Dữ liệu đầu vào sẽ là điện áp và dòng điện tại nút số 4 và số 5, cách lấy dữ liệu cũng tương tự mô hình phân loại với tần số lấy mẫu là 1.200Hz (Hình 4). Đầu ra của bộ dữ liệu là số kilomet tính từ nút 5 đến vị trí xảy ra sự cố trên đường dây 4-5. Bộ dữ liệu được chia thành 2 phần với 70% huấn luyện và 30% kiểm thử. Các mô hình CNN, LSTM, GRU, kết hợp CNN-LSTM, kết hợp CNN-GRU được sử dụng để định vị sự cố.

Các thông số của mô hình được chọn bằng cách huấn luyện mô hình với từng tổ hợp tham số có thể với giá trị của các tham số nằm trong khoảng giới hạn và so sánh các kết quả của chỉ số để chọn ra bộ thông số tốt nhất cho mô hình. Các mô hình trên sử dụng hàm mất mát MSE, tùy chọn tối ưu hóa Nadam với tốc độ học khởi tạo ban đầu bằng 0,01 và giảm dần theo giá trị của hàm mất mát tính trong tập dữ liệu kiểm thử. Nadam là kết hợp các ưu điểm của Adam trong việc điều chỉnh tốc độ học dựa trên moment của gradient và bổ sung thêm thành phần Nesterov momentum để dự đoán trước hướng của gradient, giúp việc cập nhật tham số trở nên hiệu quả hơn.

Sau khi huấn luyện và kiểm thử các thông số MSE, MAE, MAPE của các mô hình được trình bày trong (Bảng 3).

Bảng 3. Các chỉ số MAE, MSE, MAPE của các mô hình định vị sự cố

Mô hình	CNN	LSTM	GRU	CNN-LSTM	CNN-GRU
MAE (km)	0,0479	0,0354	0,0356	0,0335	0,0290
MSE	0,0179	0,0131	0,0134	0,0079	0,0068
MAPE (%)	0,2591	0,2099	0,2067	0,1181	0,0932

Các mô hình đều có các chỉ số đánh giá tốt, mô hình kết hợp CNN-GRU có chỉ số tốt nhất với MAE = 0,029 km, MSE = 0,0068 và MAPE = 0,093%. Mô hình LSTM và GRU có độ chính xác gần như tương đương nhau, điều này có thể bắt nguồn từ việc mô hình GRU được phát triển dựa trên mô hình LSTM. Chỉ số của CNN-LSTM cũng cho thấy mô hình này chính xác hơn 3 mô hình đơn lẻ CNN, LSTM, GRU.

4. Kết luận

Bài báo này đã trình bày một hướng tiếp cận mới trong việc phân loại và định vị sự cố trên hệ thống truyền tải thông qua các mô hình học máy và học sâu. Mô phỏng hệ thống lưới điện IEEE 9 nút trong phần mềm Matlab Simulink với 299,700 kịch bản xảy ra đã cung cấp bộ dữ liệu có kích thước lên tới 3GB phục vụ cho việc huấn luyện và kiểm tra các mô hình học máy. Kết quả chỉ ra các mô hình có khả năng phân loại và định vị với độ chính xác cao, mô hình ANN có thể phân loại với độ chính xác 99,974%, mô hình CNN-GRU định vị sự cố với MAE = 0,029 km và MAPE = 0,093% với thiết lập sự cố trên đường dây dài 89,93 km.

Trong các nghiên cứu tiếp theo, dữ liệu sẽ được mô phỏng sát với thực tế hơn, bao gồm các kịch bản khác nhau và tín hiệu nhiễu xuất hiện trong dòng điện và điện áp để kiểm chứng khả năng áp dụng của các mô hình học máy trong điều kiện thực tế. Đồng thời, mô phỏng sẽ được mở rộng với các lưới điện lớn hơn, giúp tăng tính chính xác và khả năng hoạt động của các mô hình trong môi trường phức tạp. Các mô hình học máy sẽ được cải tiến nhằm nâng cao độ chính xác và giảm thời gian huấn luyện. Những cải tiến này sẽ góp phần nâng cao hiệu quả phát hiện sự cố, giảm thiểu thời gian khắc phục và củng cố tính ổn định cho hệ thống điện.

TÀI LIỆU THAM KHẢO/ REFERENCES

- [1] N. D. Tleis, "1 - Introduction to power system faults," in *Power Systems Modelling and Fault Analysis*, in *Newnes Power Engineering Series*. Oxford: Newnes, 2008, pp. 1-27, doi: 10.1016/B978-075068074-5.50005-9.
- [2] M. S. Uddin *et al.*, "On the protection of power system: Transmission line fault analysis based on an optimal machine learning approach," *Energy Rep.*, vol. 8, pp. 10168–10182, Nov. 2022, doi: 10.1016/j.egyr.2022.07.163.

-
- [3] M. Bhatnagar and A. Yadav, "Fault Detection and Classification in Transmission Line Using Fuzzy Inference System," in *2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, Dec. 2020, pp. 1-6, doi: 10.1109/ICRAIE51050.2020.9358386.
- [4] L. Hulka, U. Klapper, M. Putter, and W. Wurzer, "Measurement of line impedance and mutual coupling of parallel lines to improve the protection system," in *CIREN 2009 - 20th International Conference and Exhibition on Electricity Distribution - Part 1*, Jun. 2009, pp. 1-4, doi: 10.1049/cp.2009.0800.
- [5] T.-C. Lin, B. Simachew, and M.-Y. Cho, "A Novel Single-Ended Fault Location Algorithm for Digital Distance Relays Based on A New FPGA Design," in *2023 IEEE Power & Energy Society General Meeting (PESGM)*, Jul. 2023, pp. 1–5, doi: 10.1109/PESGM52003.2023.10253255.
- [6] W. Chen, D. Wang, D. Cheng, F. Qiao, X. Liu, and M. Hou, "Novel travelling wave fault location principle based on frequency modification algorithm," *Int. J. Electr. Power Energy Syst.*, vol. 141, Oct. 2022, Art. no. 108155, doi: 10.1016/j.ijepes.2022.108155.
- [7] I. A. França, C. W. Vieira, D. C. Ramos, L. H. Sathler, and E. G. Carrano, "A machine learning-based approach for comprehensive fault diagnosis in transmission lines," *Comput. Electr. Eng.*, vol. 101, Jul. 2022, Art. no. 108107, doi: 10.1016/j.compeleceng.2022.108107.
- [8] S. Banerjee, P. S. Bhowmik, and A. K. Bohre, "Detection and Location of Fault in Microgrid using Discrete Wavelet Transform based Technique," in *2022 IEEE 6th International Conference on Condition Assessment Techniques in Electrical Systems (CATCON)*, Dec. 2022, pp. 417-421, doi: 10.1109/CATCON56237.2022.10077697.
- [9] S. Vieira, W. H. Lopez Pinaya, and A. Mechelli, "Chapter 1 - Introduction to machine learning," in *Machine Learning*, A. Mechelli and S. Vieira, Eds., Academic Press, 2020, pp. 1-20, doi: 10.1016/B978-0-12-815739-8.00001-8.
- [10] A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Incorporated, 2019.
- [11] W. H. L. Pinaya, S. Vieira, R. Garcia-Dias, and A. Mechelli, "Chapter 10 - Convolutional neural networks," in *Machine Learning*, A. Mechelli and S. Vieira, Eds., Academic Press, 2020, pp. 173-191, doi: 10.1016/B978-0-12-815739-8.00010-9.
- [12] C. H. Pham, Q. M. Nguyen, D. T. Nguyen, and T. Q. A. Tao, "Short-Term Electricity Load Forecasting Based on Temporal Fusion Transformer Model," *IEEE Access*, vol. 10, pp. 106296-106304, 2022, doi: 10.1109/ACCESS.2022.3211941.
- [13] K. Cho *et al.*, "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, A. Moschitti, B. Pang, and W. Daelemans, Eds., Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1724-1734, doi: 10.3115/v1/D14-1179.
- [14] M. Abumohsen, A. Y. Owda, M. Owda, and A. Abumihsan, "Hybrid machine learning model combining of CNN-LSTM-RF for time series forecasting of Solar Power Generation," *E-Prime - Adv. Electr. Eng. Electron. Energy*, vol. 9, Sep. 2024, Art. no. 100636, doi: 10.1016/j.prime.2024.100636.
- [15] M.-C. Chiu, H.-W. Hsu, K.-S. Chen, and C.-Y. Wen, "A hybrid CNN-GRU based probabilistic model for load forecasting from individual household to commercial building," *Energy Rep.*, vol. 9, pp. 94-105, Oct. 2023, doi: 10.1016/j.egyr.2023.05.090.