

# PHÁT HIỆN SỰ CỐ TRONG HỆ THỐNG ĐIỆN MẶT TRỜI DỰA TRÊN HỌC MÁY

## PV SYSTEM FAULT RECOGNITION BASED ON MACHINE LEARNING

Nguyễn Quốc Minh\*, Nguyễn Tiến Thành

*Trường Điện - Điện tử, Đại học Bách Khoa Hà Nội<sup>1</sup>*

\*Tác giả liên hệ: minh.nguyenquoc@hust.edu.vn  
(Nhận bài: 21/01/2022; Chấp nhận đăng: 30/4/2022)

**Tóm tắt** - Việc xác định được sự cố và vị trí xảy ra sự cố trong hệ thống điện mặt trời đóng vai trò quan trọng trong việc đảm bảo an toàn vận hành, độ tin cậy và hiệu suất sử dụng tối đa của hệ thống điện mặt trời. Các sự cố thường đa dạng và xuất hiện tại nhiều vị trí khác nhau trên hệ thống đặt ra thách thức lớn cho người giám sát và vận hành. Trong nghiên cứu này, nhóm tác giả đề xuất sử dụng phương pháp học máy, cụ thể là mô hình thuật toán học tập theo nhóm để tự động hóa việc phát hiện các sự cố trong hệ thống điện mặt trời. Mô hình được huấn luyện và kiểm thử trên bộ dữ liệu gồm hơn 2 triệu các trạng thái sự cố khác nhau. Kết quả độ chính xác của thuật toán đạt được là 98,83% cho thấy mô hình đề xuất có thể phát hiện được các sự cố trong hệ thống điện mặt trời với độ chính xác cao.

**Từ khóa** - Hệ thống điện mặt trời; học máy; sự cố; học tập theo nhóm

### 1. Đặt vấn đề

Hiện nay, khi tiềm năng các nguồn năng lượng thủy điện đã khai thác gần hết, các nguồn nhiệt điện gây ô nhiễm môi trường, thì việc sử dụng các nguồn năng lượng mới và tái tạo như điện mặt trời, điện gió, điện sinh khối thay thế cho các dạng năng lượng truyền thống đang là xu hướng tất yếu. Theo số liệu của Tập đoàn điện lực quốc gia Việt Nam (EVN), tính đến thời điểm đầu tháng 5/2021, tổng công suất điện mặt trời đã lên tới 19900 MW (trong đó, 10317 MW điện mặt trời trang trại, 9583 MW điện mặt trời áp mái) tăng hơn 20 lần so với quy hoạch [1]. Với sự gia tăng nhanh chóng của điện mặt trời hiện nay, các dự án, các nhà máy điện mặt trời được xây dựng và đưa vào sử dụng càng nhiều. Trong quá trình đi vào vận hành, các nhà máy điện mặt trời không thể tránh khỏi việc gặp các hư hỏng, sự cố, ảnh hưởng tới sự làm việc bình thường của hệ thống điện. Các dạng hư hỏng, sự cố rất đa dạng và có thể xảy ra ở các vị trí khác nhau trên hệ thống điện mặt trời như sự cố các bộ điều tần Inverter đấu nối ở phía DC cũng như phía AC, lỗi các cảm biến, tia lửa điện hư hỏng cáp kết nối tấm pin năng lượng mặt trời hay các thiết bị lưu điện, sự cố sụt áp, hở mạch trong lưới điện, các vết nứt vỡ hoặc biến dạng trên bề mặt các tấm pin năng lượng mặt trời [2]. Các lỗi này làm giảm chất lượng và tuổi thọ công trình cũng như hiệu quả đầu tư.

Một vài sự cố điển hình có thể kể đến như sự cố vào ngày 23/9/2020, tại nhà máy sản xuất của Công ty cổ phần Điện Gia Lai, đã xảy ra vụ cháy 60 tấm pin mặt trời với nguyên nhân vụ cháy được xác định là do các mối nối của các dây dẫn từ các tấm pin đến bộ phận xử lý Inverter chưa đảm bảo, dẫn đến thoát nhiệt gây phóng điện và cháy; Vào ngày 9/9/2019, một đám cháy bất ngờ xuất hiện

**Abstract** - Identifying faults in PV systems plays an important role in ensuring safety, reliability and maximum efficiency of the solar energy system. Faults in PV systems are often difficult to detect due to the complexity of the systems, posing big challenges for supervisors and operators. In this research, we propose a novel model based on machine learning method, specifically the Ensemble Learning model, to detect the faults in the PV systems automatically. The proposed model is trained and tested on a dataset of 2 million PV system fault states. The accuracy result of the algorithm is 98,83%, showing that the proposed model can detect the faults in the solar power system with high accuracy.

**Key words** - PV system; machine learning; fault; Ensemble Learning

tại nhà máy điện mặt trời lớn nhất Nhật Bản nằm tại thành phố Ichihara tỉnh Chiba, công trình năng lượng mặt trời này với diện tích khoảng 18 hec-ta với quy mô 50 nghìn tấm pin năng lượng mặt trời... Việc phát hiện nhanh và chính xác các sự cố xảy ra tại hệ thống điện mặt trời là một thách thức lớn đối với các kỹ sư giám sát và vận hành hệ thống. Đã có một số nghiên cứu về vấn đề phát hiện sự cố trên hệ thống điện mặt trời như phát hiện sự cố ngắn mạch dựa trên thuật toán bám điểm công suất cực đại [3], phát hiện sự cố chạm đất 1 pha dựa trên thiết bị đo sóng phản xạ miền thời gian [4], phát hiện sự cố trên tấm pin mặt trời dựa trên phân tích tần số [5]. Các phương pháp phát hiện sự cố nêu trên mặc dù có độ chính xác cao nhưng chỉ có thể phát hiện được một dạng sự cố tại một vị trí nhất định, trong khi các sự cố trên hệ thống điện mặt trời là khá đa dạng và có thể xuất hiện tại nhiều vị trí khác nhau.

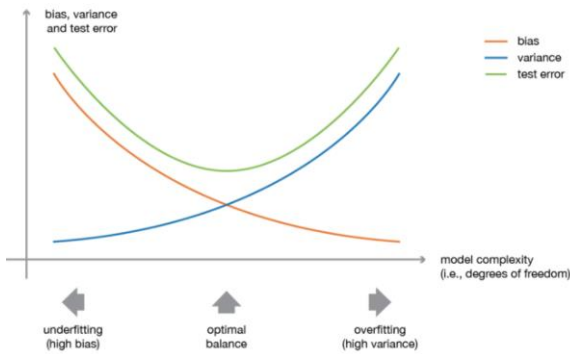
Ngày nay, với sự phát triển của trí tuệ nhân tạo nói chung và học máy nói riêng đã làm cho các bài toán phát hiện và phân loại trở nên ngày càng chính xác, thậm chí một số tác vụ có thể vượt qua con người. Một số nghiên cứu gần đây đã tập trung vào việc ứng dụng các phương pháp học máy để phát hiện và phân loại sự cố trong hệ thống điện [6, 7]. Ưu điểm của các thuật toán học máy là có thể phát hiện và phân loại được các sự cố đa dạng với tốc độ nhanh và cho độ chính xác cao. Bên cạnh đó, các thuật toán học máy không cần thông tin từ mô hình vật lý của hệ thống, các đặc trưng này sẽ được trích xuất tự động trong quá trình huấn luyện. Vì vậy, trong nghiên cứu này, nhóm tác giả đề xuất sử dụng mô hình thuật toán học máy là học tập theo nhóm (Ensemble Learning) để phát hiện các sự cố trong hệ thống điện mặt trời.

<sup>1</sup> School of Electrical and Electronic Engineering, Hanoi University of Science & Technology (Nguyen Quoc Minh, Nguyen Tien Thanh)

## 2. Mô hình thuật toán

Ensemble Learning (Học tập theo nhóm) là một nhánh của lĩnh vực học máy. Ý tưởng của việc học tập theo nhóm có liên quan chặt chẽ với ý tưởng về “sự khôn ngoan của đám đông”, theo đó thì các quyết định, lựa chọn hoặc ước tính độc lập khác nhau được tổng hợp thành lựa chọn cuối cùng với độ chính xác thu được tốt hơn độ chính xác của các lựa chọn đơn lẻ. Trong học máy, điều đó có nghĩa là kết hợp các dự đoán từ nhiều mô hình được đào tạo trên tập dữ liệu huấn luyện. Thay vì cố gắng xây dựng một mô hình tốt nhất, chúng ta sẽ đi xây dựng một họ các mô hình yếu hơn, tuy nhiên khi kết hợp các mô hình lại, sẽ thu được một mô hình vượt trội.

Khi gặp một các bài toán phân loại (classification) hay hồi quy (regression) thì việc chọn ra một mô hình đủ tốt luôn là vấn đề mấu chốt. Khác với Deep Learning (học sâu), việc tìm ra mô hình tốt là việc cố gắng thay đổi số lớp hay thay đổi cấu trúc mạng, đối với các thuật toán học máy thì việc lựa chọn mô hình là việc tối ưu tham số, quan sát các đặc điểm về số chiều của không gian dữ liệu, hiểu hết sự đa dạng của các đặc trưng, lựa chọn cẩn thận cách kết hợp các đặc trưng đó, đặt ra các giả thiết về phân phối dữ liệu. Tất cả các bước xây dựng mô hình bản chất là xoay quanh giải quyết vấn đề giữa độ lệch (bias) và phương sai (variance), (bias-variance trade off). Hiểu đơn giản về bias-variance trade off tức là chúng ta mong muốn huấn luyện mô hình có độ lệch thấp và phương sai thấp. Tuy nhiên, độ lệch và phương sai thường có xu hướng nghịch đảo với nhau, và chỉ có thể lựa chọn tăng cái này và chấp nhận giảm cái kia [8]. Một điểm lưu ý là nếu chỉ sử dụng một mô hình duy nhất để huấn luyện thì vấn đề “bias-variance trade off” là điều không thể tránh khỏi.



Hình 1. Mối quan hệ giữa bias và variance

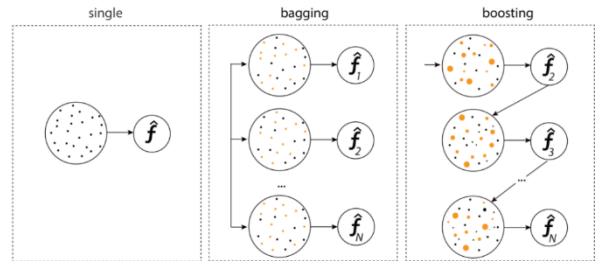
Để giải quyết được vấn đề “bias-variance trade off”, một hướng giải quyết được đề xuất là sử dụng mô hình kết hợp các lựa chọn có độ chính xác thấp để đưa ra được lựa chọn có độ chính xác tốt hơn. Hiểu một cách đơn giản về phương pháp này là nếu một mô hình không thể tự giải quyết được, hãy để nhiều mô hình cùng nhau giải quyết. Tuy nhiên, nhiều mô hình ở đây có thể là cùng loại nhưng áp dụng trên những phần dữ liệu khác nhau (kì vọng là độc lập với nhau) hoặc những mô hình hoàn toàn khác loại được kết hợp lại.

Mỗi kiểu kết hợp các mô hình được áp dụng tùy theo mục đích nhất định. Dựa vào tính chất này, các thuật toán Ensemble Learning được chia thành 3 nhóm chính, bao gồm:

- Phương pháp đóng bao (Bagging Ensembles): Xây dựng một lượng lớn các mô hình, thường là cùng loại, trên các mẫu khác nhau được lấy ngẫu nhiên từ tập dữ liệu huấn luyện, đảm bảo mỗi mô hình là một mẫu dữ liệu không trùng lặp. Những mô hình này được huấn luyện độc lập và song song với nhau nhưng đầu ra của chúng sẽ được trung bình cộng để cho ra kết quả cuối cùng. Một số thuật toán điển hình như: Decision Tree, Random Subspaces, Random Forest, Extra Trees.

- Phương pháp tăng cường (Boosting Ensembles): Xây dựng một lượng lớn các mô hình một cách tuần tự, mỗi mô hình sau sẽ học cách sửa lỗi của những mô hình trước tạo thành một chuỗi các mô hình mà mô hình sau sẽ tốt hơn mô hình trước bởi trọng số được cập nhật qua mỗi lần xây dựng, cụ thể ở đây là trọng số của những dữ liệu dự đoán đúng sẽ không đổi, còn trọng số của những dữ liệu dự đoán sai sẽ được tăng thêm. Kết quả của mô hình cuối cùng trong chuỗi này sẽ được lấy làm kết quả trả về. Một số thuật toán điển hình như: AdaBoost, Gradient Boosting, XGBoost, LightGBM, CatBoost.

- Phương pháp xếp chồng (Stacking Ensembles): Kỹ thuật này đề cập tới việc xây dựng một số mô hình và một mô hình giám sát (meta model), huấn luyện các mô hình này độc lập, sau đó mô hình giám sát sẽ học cách kết hợp kết quả dự đoán của một số mô hình trên một cách tốt nhất. Một số kỹ thuật điển hình như: Voting, Weighted Averrange, Blending, Stacking, Super Learner.



Hình 2. Sơ đồ khối minh họa các nhóm chính

Trong nghiên cứu này, nhóm tác giả đề xuất sử dụng các thuật toán thuộc nhóm phương pháp tăng cường (Boosting Ensembles) để phát hiện các sự cố trong hệ thống điện mặt trời. Nhóm tăng cường đề cập đến một lớp các thuật toán học máy kết hợp thường được sử dụng phổ biến cho các vấn đề dự đoán có cấu trúc, chẳng hạn như phân loại và hồi quy trên dữ liệu dạng bảng, rất phù hợp với bộ dữ liệu mà nhóm nghiên cứu đang sử dụng.

### 2.1. Thuật toán AdaBoost

Adaptive Boosting, gọi tắt là AdaBoost, là mô hình xây dựng thuật toán nhằm giải quyết bài toán tối ưu sau:

$$\min_{C_n = 1:N, w_n = 1:N} L \left( y, \sum_{n=1}^N C_n w_n \right) \quad (1)$$

Trong đó, L là giá trị loss function, y là nhãn (label),  $C_n$  là trọng số của weak learner thứ n (confidence score),  $w_n$  là weak learner thứ n.

AdaBoost tiến hành huấn luyện các mô hình mới dựa trên việc đánh lại trọng số cho các điểm dữ liệu hiện tại, nhằm giúp các mô hình mới có thể tập trung hơn vào các

mẫu dữ liệu đang bị học sai, từ đó làm giảm giá trị loss của mô hình [9]. Cụ thể, các bước triển khai như sau:

- Khởi tạo trọng số ban đầu là bằng nhau (bằng  $1/N$ ) cho mỗi điểm dữ liệu.
- Tại vòng lặp thứ  $i$ :
  - + Huấn luyện mô hình với trọng số  $w_i$  (weak learner) mới thêm vào;
  - + Tính toán giá trị mất mát (loss), từ đó tính toán ra giá trị trọng số  $C_i$  của mô hình vừa huấn luyện;
  - + Cập nhật mô hình chính  $W = W + C_i.w_i$ ;
  - + Cuối cùng, đánh lại trọng số cho các điểm dữ liệu (các điểm dữ liệu bị đoán sai được tăng trọng số, các điểm dữ liệu được dự đoán đúng thì giảm trọng số).
- Sau đó lặp lại với vòng lặp thêm mô hình tiếp theo  $i+1$ .

## 2.2. Thuật toán Gradient Boosting

Gradient Boosting là một dạng tổng quát hóa của AdaBoost. Thay vì cố gắng tìm tất cả các giá trị  $C_n, w_n$  để tìm nghiệm tối ưu toàn cục như AdaBoost - một công việc tốn nhiều thời gian và tài nguyên, thì Gradient Boosting đi tìm các giá trị nghiệm cục bộ sau khi thêm mỗi một mô hình mới vào chuỗi mô hình với mong muốn dẫn đi đến nghiệm toàn cục.

$$\min_{C_n, w_n} L(y, W_{n-1} + C_n w_n) \quad (2)$$

$$\text{với } W_{n-1} = \sum_{n=1}^{N-1} C_n w_n$$

Nếu chúng ta coi chuỗi các mô hình là một hàm số  $W$ , thì mỗi hàm học có thể coi là một tham số  $w$ . Để cực tiểu hóa hàm mất mát  $L(y, W)$  chúng ta áp dụng phương pháp gradient descent:

$$W_n = W_{n-1} - \eta \frac{\partial}{\partial W} L(W_{n-1}) \quad (3)$$

Đến đây, ta có thể thấy mối quan hệ liên quan sau:

$$C_n w_n \approx -\eta \frac{\partial}{\partial W} L(W_{n-1})$$

với,  $w_n$  là mô hình được thêm vào tiếp theo. Khi đó, mô hình mới cần học để khớp với giá trị  $-\eta \frac{\partial}{\partial W} L(W_{n-1})$ . Giá trị

$-\eta \frac{\partial}{\partial W} L(W_{n-1})$  còn có một tên gọi khác là “pseudo-residuals” [10]

Tóm lại, có thể tóm tắt quá trình thuật toán như sau:

- Huấn luyện mô hình mới thêm vào để làm phù hợp giá trị của “pseudo-residuals” đã có.
- Tính toán giá trị trọng số  $C_i$  của mô hình vừa huấn luyện.
- Cập nhật mô hình chính  $W = W + C_i.w_i$ .
- Cuối cùng, tính toán giá trị “pseudo-residuals” để làm nhân cho mô hình tiếp theo
- Lặp lại với vòng lặp  $i+1$ .

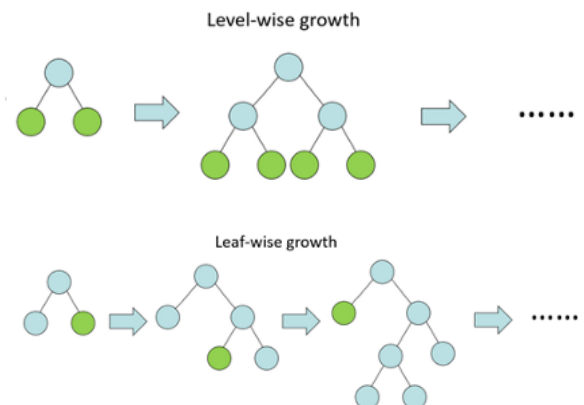
Có thể thấy, phương pháp cập nhật trọng số của điểm dữ liệu của AdaBoost cũng là một trong các trường hợp của Gradient Boosting. Do đó, Gradient Boosting bao quát được nhiều trường hợp hơn. Về mặt lý thuyết, phương pháp Gradient Boosting hoạt động tốt hơn AdaBoost, tuy nhiên nó dễ gặp các sự cố quá mức (overfitting) và thời gian chạy lâu hơn để tăng độ dốc.

## 2.3. Thuật toán XGBoost

XGBoost (Extreme Gradient Boosting) là một giải thuật được xây dựng dựa trên nền tảng của Gradient Boosting. Tuy nhiên, kèm theo đó là những cải tiến to lớn về mặt tối ưu thuật toán, về sự kết hợp giữa sức mạnh phần mềm và phần cứng, giúp đạt được những kết quả vượt trội cả thời gian huấn luyện cũng như bộ nhớ sử dụng. Thứ nhất, nó cải thiện việc quá mức (overfitting) bằng cách sử dụng chính quy hóa. Thứ hai, nó cải thiện tốc độ chạy bằng cách tối ưu hóa việc sắp xếp bằng việc chạy song song [11] XGBoost có thể được sử dụng để quyết định tất cả các vấn đề từ hồi quy (regression), phân loại (classification), ranking và giải quyết các vấn đề do người dùng tự định nghĩa.

## 2.4. Thuật toán LightGBM

Light Gradient Boosted Machine, gọi tắt là LightGBM, là một thư viện tăng cường độ dốc mã nguồn mở xuất phát từ thuật toán học tập dựa trên cây quyết định (Decision Tree). Nó áp dụng thuật toán dựa trên biểu đồ chia tách biến liên tục thành các nhóm khác nhau thay vì sắp xếp chúng riêng lẻ. Nó sử dụng phương pháp phát triển cây theo lá thay vì phương pháp tăng trưởng cây theo cấp (được sử dụng bởi hầu hết các phương pháp dựa trên cây quyết định khác), điều này làm nên điểm khác biệt của LightGBM. Kích thước dữ liệu đang tăng lên từng ngày và các thuật toán khoa học dữ liệu truyền thống đang trở nên khó khăn để đưa ra kết quả nhanh hơn. LightGBM đã khắc phục được vấn đề này, thuật toán đã cải thiện thời gian chạy rất nhiều bằng cách làm cho khối lượng công việc tính toán trở nên ‘nhẹ nhàng’ nhưng vẫn có thể duy trì mức hiệu suất mô hình bằng hoặc hơn so với các thuật toán khác [12].



**Hình 3.** Kỹ thuật phát triển cây

Một vài điểm chính tạo nên tốc độ và độ chính xác đáng kinh ngạc của mô hình có thể kể đến như:

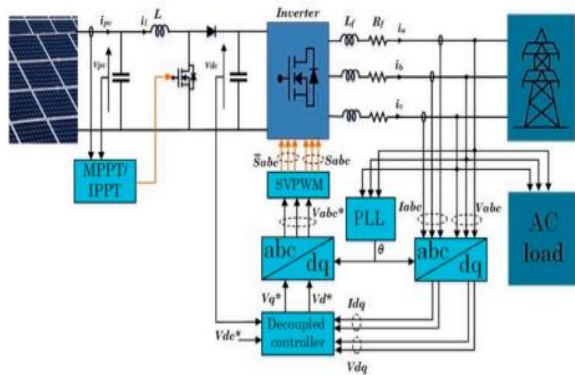
- LightGBM sử dụng "histogram-based algorithms"

thay thế cho "pre-sort-based algorithms" thường được dùng trong các giải thuật tăng cường để tìm kiếm điểm phân chia (split point) trong quá trình xây dựng cây (tree). Cải tiến này giúp LightGBM tăng tốc độ huấn luyện, đồng thời làm giảm bộ nhớ cần sử dụng. Ngoài ra, điểm tối ưu của LightGBM là ở 2 thuật toán: GOSS (Gradient Based One Side Sampling) và EFB (Exclusive Feature Bundling) giúp tăng tốc đáng kể trong quá trình tính toán [12].

- LightGBM phát triển cây dựa trên phương pháp leaf-wise, trong khi hầu hết các giải thuật tăng cường khác, kể cả XGBoost, dựa trên phương pháp level-wise. Phương pháp leaf-wise lựa chọn nút để phát triển cây dựa trên tối ưu toàn bộ cây, trong khi phương pháp level-wise tối ưu trên nhánh đang xét, do đó, với số nút nhỏ, các cây xây dựng từ leaf-wise thường có kết quả tốt hơn nhiều so với level-wise.

### 3. Bộ dữ liệu dùng trong nghiên cứu

Trong nghiên cứu này, nhóm tác giả sử dụng bộ dữ liệu được tạo ra từ phòng thí nghiệm PV microgrid system [13], mô hình được miêu tả ở Hình 4. Hệ thống PV nối lưới bao gồm bộ nguồn mô phỏng các tấm pin năng lượng mặt trời được tạo ra qua lập trình Chroma 62150H-1000s. Bộ nguồn mô phỏng này có độ ổn định cao và có thiết kế đáp ứng được các thay đổi trong các điều kiện thời tiết thay đổi khác nhau giống như tấm PV thực tế như mức độ bức xạ, nhiệt độ và hiệu ứng đám mây. Nguồn xoay chiều lập trình bởi Chroma 61511 Programmable AC Source được sử dụng trong mô phỏng lưới, đáp ứng đa dạng các hiện tượng trên đường dây. Các loại mạch điều khiển được mô hình hóa bằng phần mềm Matlab Simulink. Thuật toán điều khiển được thực hiện trên bo mạch Dspace 1104, và cũng được sử dụng cho việc thu thập dữ liệu. Kỹ thuật điều khiển định hướng điện áp thì được sử dụng kết hợp với điều chế độ rộng xung không gian để điều khiển công suất tác dụng và công suất phản kháng dựa trên tín hiệu từ lưới. Điện áp ra đồng bộ với điện áp lưới qua vòng khóa pha. Mô hình cũng sử dụng tải xoay chiều dùng cho mục đích bảo vệ hệ thống trong khi đưa đầu vào là lỗi xuất hiện trong thực tế.



**Hình 4.** Mô hình hệ thống pin mặt trời nối lưới mô phỏng trong phòng thí nghiệm [13]

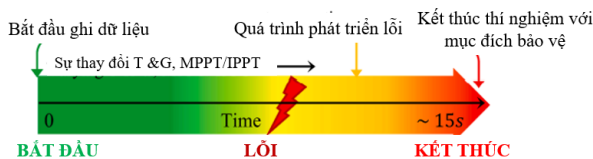
Hệ thống điều khiển ở đây dựa trên kỹ thuật tối ưu hóa phân tử bầy đàn (Particle Swarm Optimization) để đảm bảo công suất trong cả hai chế độ điểm công suất tối đa MPPT (Maximum Power Point Tracking) khi  $P_{available} \leq P_{limit}$  và

chế độ điểm công suất trung gian IPPT (Intermediate Power Point Tracking) khi  $P_{available} > P_{limit}$  hệ thống này được dùng để tạo và thu thập dữ liệu lỗi thực tế theo thời gian thực.

**Bảng 1.** Các sự cố trong hệ thống điện mặt trời

Lỗi	Loại lỗi	Mô tả
F0	Không có lỗi	
F1	Lỗi trên bộ biến tần	Hỏng hoàn toàn một trong sáu IGBTs
F2	Lỗi bộ cảm biến phản hồi	Cảm biến một pha lỗi 20%
F3	Sự cố trên lưới điện	Hiện tượng sụt áp trong lưới điện
F4	Lỗi không đồng bộ trên các tấm PV	10 ÷ 20% các tấm PV bị bóng râm che khuất
F5	Lỗi không đồng bộ trên các tấm PV	15% hở mạch trên các tấm PV
F6	Lỗi bộ điều khiển MPPT/IPPT	Hệ số khuếch đại của bộ điều khiển PI trong điều khiển chế độ MPPT/IPPT của boost converter bị giảm 20%
F7	Lỗi bộ điều khiển Boost converter	Tham số thời gian của bộ điều khiển PI trong điều khiển chế độ MPPT/IPPT của boost converter bị tăng thêm 20%

Dữ liệu được thu thập từ các phép đo cảm biến và được trích xuất các thành phần bởi PMU (Phase Measurement Unit) vi mô. Để không làm mất tính tổng quát thì các PMU vi mô được sử dụng với độ phân giải micro giây với độ chính xác đến mili giây. Có tất cả là 8 kịch bản bao gồm cả lỗi và không lỗi được đưa ra, ký hiệu lần lượt là F0, F1, F2, F3, F4, F5, F6, F7, các kịch bản được thể hiện như trong Bảng 1. Các lỗi này thuộc nhiều dạng và tại các vị trí khác nhau trong hệ thống để đảm bảo phân tích và nhận dạng được đầy đủ các sự cố ảnh hưởng trực tiếp đến hệ thống. Các thí nghiệm được thực hiện một cách độc lập và liên tiếp trong khoảng thời gian từ 10 đến 15 giây trong đó sự cố được đưa vào khoảng giây thứ 7 đến giây thứ 8, như được mô tả trên Hình 5. Thời gian lấy mẫu để thu thập dữ liệu trong khoảng thời gian  $T_s = 100 \mu s$ .

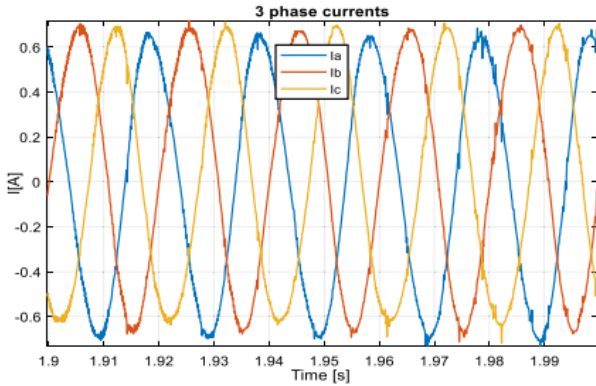


**Hình 5.** Quá trình xảy ra lỗi trong hệ thống PV

Các lỗi trên bộ biến tần hay sự cố trên lưới điện như F1 và F3 xảy ra ở phía lưới của hệ thống PV, chúng gây ảnh hưởng đến phía AC, tuy các thông số của lưới thay đổi rất nhỏ được minh họa như trong Hình 6 nhưng hậu quả gây nên với lưới rất nghiêm trọng, nên những lỗi này cần được phát hiện trong giai đoạn đầu và xử lý ngay lập tức. Các lỗi không đồng bộ trên các tấm pin PV như lỗi F4, F5, gây ra những biến đổi lớn các cảm biến ở phía DC, những lỗi này có mức độ nguy hiểm thấp, gây ảnh hưởng đến đến hệ thống không nhiều nhưng gây ra tổn thất năng lượng khá lớn trong hệ thống. Thuật toán còn phát hiện ra lỗi bộ cảm biến phản hồi (F2), lỗi trong các bộ điều khiển tích phân tỷ lệ PI



trong điều khiển chế độ MPPT/IPPT ở phía DC (F6, F7). Trong đó, lỗi bộ điều khiển F7 cho biết hệ số khuếch đại của bộ điều khiển PI trong điều khiển chế độ MPPT/ IPPT của bộ Boost Converter bị giảm xuống 20% trong khi F6 tăng thêm 20%, việc này dẫn đến giảm hiệu suất của việc theo dõi quỹ đạo MPPT/ IPPT. Những lỗi trên rất phổ biến trong thực tế, tác động của các sự cố đến hệ thống PV, mô tả về phân lý thuyết và đường đặc tính U-I được trình bày

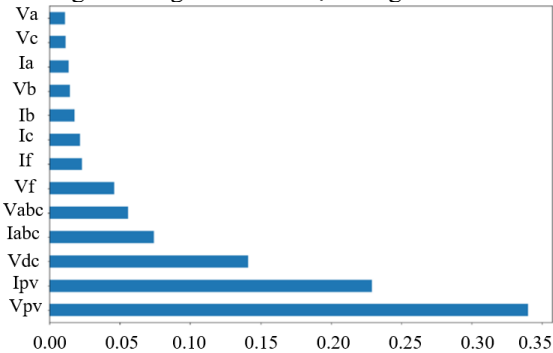


Hình 6. Sự dao động của các thông số lưới khi gặp sự cố F1, F3 [13]

Mỗi sự cố trên Bảng 1, được mô tả bởi các thông số liên quan trực tiếp đến lỗi là:

$$\{I_{PV}, V_{PV}, V_{DC}, I_a, I_b, I_c, V_a, V_b, V_c, I_{abc}, f_I, V_{abc}, f_V\}$$

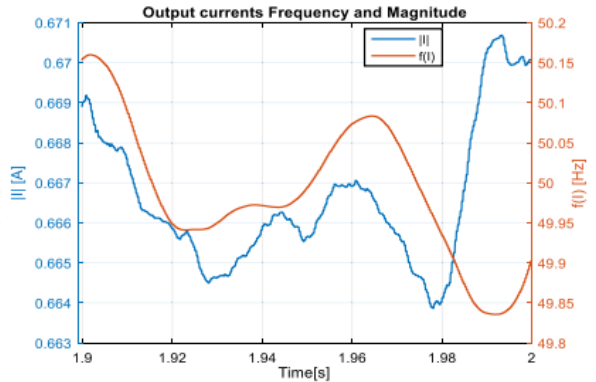
Trong đó,  $I_{PV}$  là dòng điện các tấm pin PV;  $V_{PV}$  là điện áp các tấm pin PV;  $V_{DC}$  là điện áp phía một chiều;  $I_a, I_b, I_c$  lần lượt là dòng điện pha A, B, C phía xoay chiều;  $V_a, V_b, V_c$  là điện áp pha A, B, C phía xoay chiều;  $I_{abc}, V_{abc}$  là cường độ dòng điện, điện áp thứ tự thuận;  $f_I, f_V$  là tần số dòng điện và điện áp. Các thông số này cũng chính là các đặc trưng mà nhóm nghiên cứu sử dụng làm đầu vào trong mô hình các thuật toán để phân loại các sự cố trong hệ thống điện mặt trời. Vì vậy, đầu vào của các thuật toán là một mảng có kích thước 1x13 bao gồm thông số của các đặc trưng nêu trên.



Hình 7. Mức độ ảnh hưởng của các đặc trưng

Trong quá trình huấn luyện mô hình, nhóm nghiên cứu trích xuất ra được mức độ ảnh hưởng của các đặc trưng như trên Hình 7. Trục hoành là các chỉ số đã được chuẩn hóa sao cho chỉ số ảnh hưởng của tất cả các đặc trưng bằng 1. Trục tung là các đặc trưng được sắp xếp giảm dần từ trên xuống. Từ đó có cái nhìn tổng quan hơn về tầm quan trọng hay đóng góp tương đối của mỗi đặc trưng đến bài toán phân loại các sự cố trong hệ thống điện mặt trời.

và đánh giá chi tiết trong [14]. Ngoài ra, các lỗi trong hệ thống PV xảy ra do sự hao mòn các tấm pin, suy thoái các tế bào quang điện hay hư hỏng các mối nối, những lỗi xuống cấp này không quá nghiêm trọng và chúng được tránh thông qua quá trình bảo trì, kiểm tra định kỳ, hơn thế nữa việc phát hiện chúng đòi hỏi cần có dữ liệu dài hạn với khoảng thời gian lấy mẫu lớn vì thế các lỗi liên quan đến vấn đề trên không được xét trong nghiên cứu này.



#### 4. Kết quả

Với bộ dữ liệu đầu vào rất lớn với hơn 2 triệu các trạng thái lỗi khác nhau không thể tránh khỏi các điểm dữ liệu bị lỗi, nhiễu và bị thiếu, trong quá trình tiền xử lý dữ liệu, nhóm nghiên cứu đã tiến hành loại bỏ các điểm này ra khỏi bộ Data. Ngoài ra, mỗi trạng thái có các đặc trưng  $I_{PV}, V_{PV}, V_{DC}, I_a, I_b, I_c, V_a, V_b, V_c, I_{abc}, f_I, V_{abc}, f_V$  có giá trị biến thiên tương đối khác nhau. Điều này ảnh hưởng rất lớn tới các sai số trong việc xác định các trọng số của mô hình thuật toán, do đó nhóm nghiên cứu đã chuẩn hóa các đặc trưng theo phương pháp min-max scaling. Đây là phương pháp cho phép biến một đại lượng  $X$  có giá trị biến thiên bất kỳ thành đại lượng  $X'$  có giá trị biến thiên từ 0 đến 1.

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (4)$$

Trong đó,  $X$  là giá trị ban đầu;  $X'$  là giá trị sau khi chuẩn hóa;  $X_{\max}, X_{\min}$  là giá trị lớn nhất và nhỏ nhất được tính trên toàn bộ dữ liệu ở cùng một thành phần.

Sau khi đã có bộ dữ liệu hoàn chỉnh, dữ liệu được chia thành hai phần là tập huấn luyện và tập kiểm thử với tỷ lệ 70-30. Bước tiếp theo là xây dựng mô hình, nhóm nghiên cứu đã sử dụng kỹ thuật GridSearch nhằm thay đổi và tìm ra hyperparameter tối ưu cho mô hình, từ đó kết hợp các tham số và cải thiện hiệu suất tổng quát hóa của một mô hình. Bước cuối cùng là đánh giá độ chính xác của thuật toán. Các bước được tóm tắt trong sơ đồ Hình 8.

Sau nhiều lần xây dựng các thuật toán, tối ưu các tham số bên trong mô hình, độ chính xác lớn nhất của mô hình phát hiện các sự cố đạt được là 98,83 %.

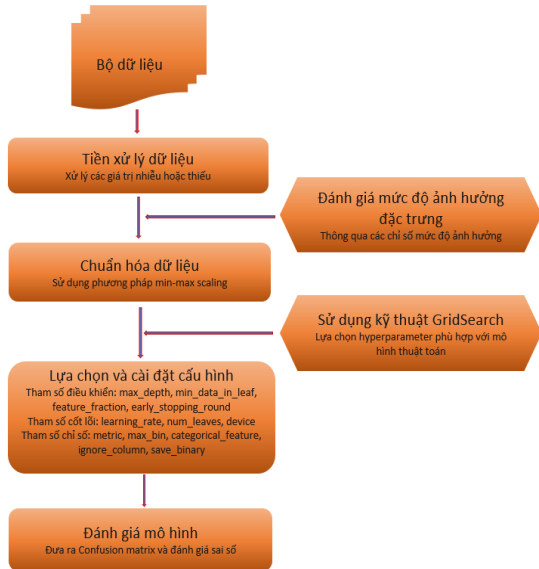
Để đánh giá trực quan độ chính xác của mô hình, chúng tôi sử dụng ma trận hợp nhất (Confusion matrix). Hình 9 thể hiện ma trận hợp nhất của thuật toán LightGBM, ma trận có kích thước 8x8 do có 8 kịch bản lỗi cần phát hiện bao gồm F0, F1, F2, F3, F4, F5, F6, F7. Trục hoành thể hiện các trạng

thái dự đoán của mô hình, còn trực tiếp thể hiện các trạng thái thực tế. Với định nghĩa như vậy thì có thể thấy, các phần tử trên đường chéo chính của ma trận thể hiện số trạng thái dự đoán giống với thực tế, còn các phần tử nằm ngoài đường chéo chính thể hiện số dự đoán khác với thực tế. Ví dụ như phần tử C(4,3) có giá trị bằng 90, tức là có 90 trạng thái trong thực tế là lỗi F5 nhưng mô hình dự đoán nhầm thành trạng thái lỗi F3 hay phần tử C(4,5) có giá trị là 50712 tức trạng thái trong thực tế là lỗi F3 và mô hình dự đoán chính xác là lỗi F3. Căn cứ vào ma trận hợp nhất ta có thể tính được độ chính xác của các thuật toán theo công thức:

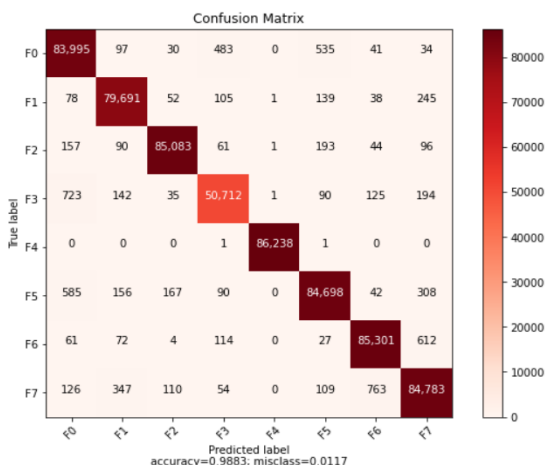
$$acc = \frac{Số\ dự\ đoán\ đúng}{Tổng\ số\ dự\ đoán} \quad (4)$$

Tương tự, ma trận hợp nhất của các thuật toán AdaBoost, Gradient Boosting, XGBoost được thể hiện lần lượt ở Hình 10-12.

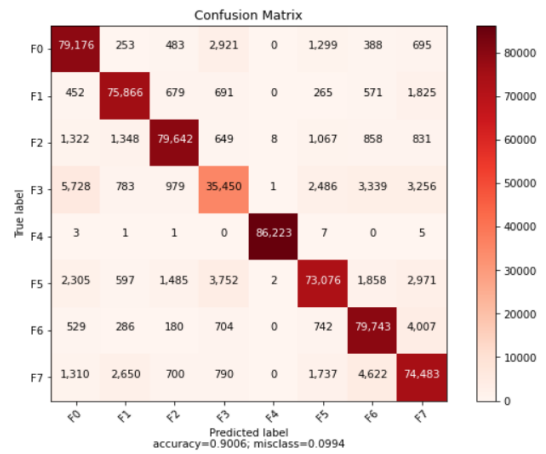
Nhóm nghiên cứu cũng thống kê lại được thời gian huấn luyện và nhận dạng của các thuật toán. Kết quả cho thấy mô hình thuật toán LightGBM đã cải thiện đáng kể tốc độ tính toán so với các thuật toán khác trong nhóm Ensemble Learning khi được huấn luyện trên bộ dữ liệu rất lớn, với hơn 2 triệu các trạng thái khác nhau.



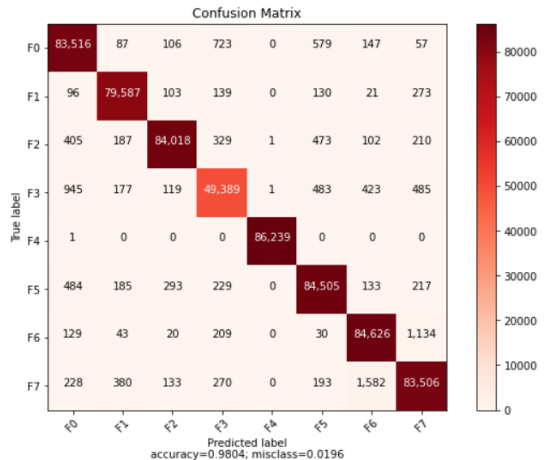
Hình 8. Sơ đồ khối của mô hình phát hiện các sự cố trong hệ thống điện mặt trời dựa trên học máy



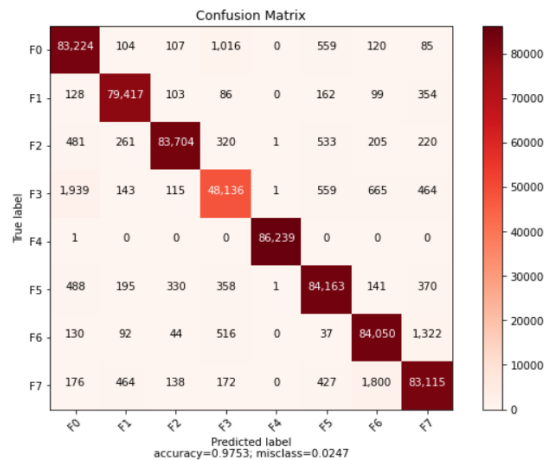
Hình 9. Ma trận hợp nhất của thuật toán LightGBM



Hình 10. Ma trận hợp nhất của thuật toán AdaBoost



Hình 11. Ma trận hợp nhất của thuật toán XGBoost



Hình 12. Ma trận hợp nhất của thuật toán G.Boosting

Bảng 2. Độ chính xác (%) của từng loại lỗi

Lỗi	AdaBoost	Gradient Boosting	Light GBM	XGBoost
F0	92,91	97,66	98,57	98,01
F1	94,42	98,84	99,18	99,05
F2	92,90	97,64	99,25	98,01
F3	68,14	92,63	97,48	94,94
F4	99,98	99,99	100	99,98
F5	84,88	97,81	98,43	98,21
F6	92,52	97,52	98,97	98,18
F7	86,32	96,32	98,25	96,77

**Bảng 3.** So sánh thời gian huấn luyện, nhận dạng và độ chính xác của thuật toán

	Ada Boost	Gradient Boosting	Light GBM	XGBoost
Thời gian huấn luyện (s)	6238	4226,1	321,7	5038,3
Thời gian nhận dạng (s)	0,0233	0,0082	0,0035	0,0104
Độ chính xác (%)	90,06	97,53	98,83	98,04

## 5. Kết luận

Nghiên cứu này đề xuất một cách tiếp cận để phát hiện tự động các sự cố trong hệ thống điện mặt trời dựa trên thuật toán học máy. Kết quả nghiên cứu cho thấy, mô hình học máy có thể phát hiện các sự cố rất đa dạng trên hệ thống điện mặt trời với độ chính xác lên tới 98,83%. Trong các nghiên cứu tiếp theo, nhóm tác giả sẽ tiếp tục cải thiện mô hình để nâng cao hơn nữa tốc độ tính toán của thuật toán trên các bộ dữ liệu có kích thước lớn.

## TÀI LIỆU THAM KHẢO

- [1] Tập đoàn Điện lực Việt Nam “Báo cáo thường niên”, 2021.
- [2] K. AbdulMawjood, S. S. Refaat and W. G. Morsi, "Detection and prediction of faults in photovoltaic arrays: A review", *2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2018)*, 2018, pp. 1-8.
- [3] D. S. Pillai and N. Rajasekar, "An MPPT-Based Sensorless Line-Line and Line-Ground Fault Detection Technique for PV Systems", in *IEEE Transactions on Power Electronics*, vol. 34, no. 9, Sept. 2019, pp. 8646-8659.
- [4] S. Roy, M. K. Alam, F. Khan, J. Johnson and J. Flicker, "An Irradiance-Independent, Robust Ground-Fault Detection Scheme for PV Arrays Based on Spread Spectrum Time-Domain Reflectometry (SSTDR)", in *IEEE Transactions on Power Electronics*, vol. 33, no. 8, Aug. 2018, pp. 7046-7057.
- [5] P. Sevilla-Camacho, M. Zuñiga-Reyes, J. Robles-Ocampo, R. Castillo-Palamera, J. Muñoz and J. Rodríguez-Reséndiz, "A Novel Fault Detection and Location Method for PV Arrays Based on Frequency Analysis", in *IEEE Access*, vol. 7, 2019, pp. 72050-72061.
- [6] H. A. Tokel, R. A. Halaseh, G. Alirezai and R. Mathar, "A New Approach for Machine Learning-based Fault Detection and Classification in Power Systems", *2018 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2018, pp. 1-5.
- [7] K. Moloi and A. O. Akumu, "Power Distribution Fault Diagnostic Method Based on Machine Learning Technique", *2019 IEEE PES/IAS PowerAfrica*, 2019, pp. 238-242.
- [8] Belkin, M., Hsu, D., Ma, S., Mandal, S. Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proc. Natl. Acad. Sci.* 32, 116 (2019), 15849—15854
- [9] Wu, P., Zhao, H. Some Analysis and Research of the AdaBoost Algorithm. In: Chen, R. (eds) *Intelligent Computing and Information Science. ICICIS 2011. Communications in Computer and Information Science*, vol 134. Springer, Berlin, Heidelberg, 2011.
- [10] Haihao Lu, Sai Praneeth Karimireddy, Natalia Ponomareva, Vahab Mirrokni, "Accelerating Gradient Boosting Machines", *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, PMLR 108:516-526, 2020.
- [11] Jason Brownlee, "Gradient Boosted Trees with XGBoost and Scikit-learn", *edition v1.15*, 2021.
- [12] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, Tie-Yan Liu, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree", *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 2017.
- [13] A.Guichi, A.Talha, E.M.Berkouk, S.Mekhilef and S.Gassab, "A New Method for Intermediate Power Point Tracking for PV Generator under Partially Shaded Conditions in Hybrid System", *Solar Energy*, vol. 170, Aug. 2018, pp. 974-987.
- [14] Pillai DS, Rajasekar N. A comprehensive review on protection challenges and fault diagnosis in PV systems. *Renew. Sust. Energy Rev.* 2018;91:18-40.