PHÁT HIỆN TRẠNG THÁI HỆ THỐNG ĐIỆN BỊ TẤN CÔNG AN NINH MẠNG DỰA TRÊN MÁY HỌC

POWER SYSTEMS CYBER-ATTACK DETECTION BASED ON MACHINE LEARNING

Nguyễn Quốc Minh[1*], Nguyễn Trần Minh Trang[1], Nguyễn Tiến Thành[1], Đàm Tá Hải[2]

[1]Trường Đại học Bách khoa Hà Nội
[2]Công ty TNHH MTV Thí nghiệm điện miền Bắc

*Tác giả liên hệ: minh.nguyenquoc@hust.edu.vn

(Nhận bài: 21/6/2021; Chấp nhận đăng: 09/8/2021)

Tóm tắt - Trong nghiên cứu này, nhóm tác giả đề xuất sử dụng các thuật toán máy học (machine learning) để phát hiện trạng thái hệ thống điện bị tấn công an ninh mạng. Bộ dữ liệu sử dụng được lấy từ phòng thí nghiệm Oak Ridge National Laboratory của Hoa Kỳ. Bộ dữ liệu bao gồm 128 các đặc trưng thu được từ các Phasor Measurement Unit (PMU), là các giá trị biên độ, góc pha của điện áp và dòng điện, tần số, tổng trở, và các trạng thái của hệ thống điều khiển bảo vệ. Bộ dữ liệu được đưa vào lớp trích chọn đặc trưng, nhằm loại bớt các đặc trưng không ảnh hưởng hoặc ít ảnh hưởng đến bài toán nhận dạng, sau đó được đưa vào lớp nhận dạng để phát hiện các trạng thái bị tấn công an ninh mạng. Kết quả cho thấy, các thuật toán machine learing có thể nhận dạng được trạng thái hệ thống điện bị tấn công an ninh mạng với độ chính xác đạt được là 92,39%.

Từ khóa - Hệ thống điện; an ninh mạng; trích xuất đặc trưng; phân loại; máy học

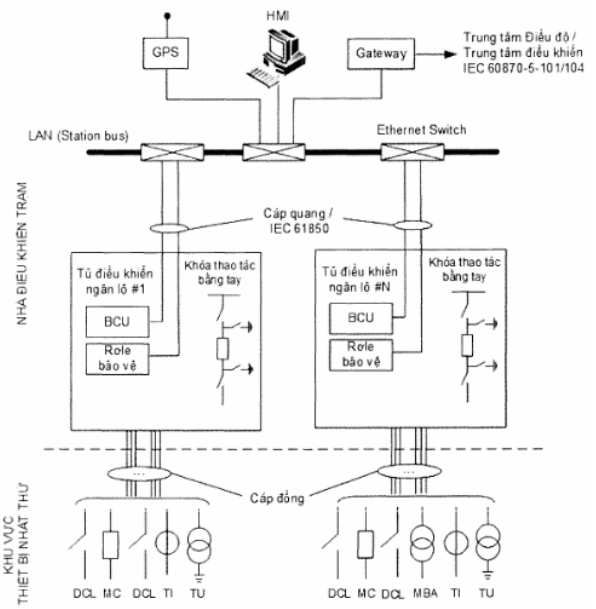
Abstract - In this research, the authors propose a novel method to detect 34567890- based on machine learning. We use the data from the Oak Ridge National Laboratory, USA. The data consist of 128 features from Phasor Measurement Unit (PMU) including phase and magnitude of the voltage and current, frequency, impedance and status from control panel. The data are first fed into feature extraction layer to detect and eliminate the unaffected features. The data are then split into training and testing sets. We use several machine learning algorithms to train the power system cyber-attack detection model such as random forest, support vector machine, K-nearest neighbor and neural network. The results show that, the cyber-attack can be detected with the accuracy of 92.39% by proposed method.

Key words - Power system; cyber security; feature extraction; classification; machine learning

1. Đặt vấn đề

Ngày nay, mức độ tự động hóa trong hệ thống điện ngày càng cao, đặc biệt với sự phát triển của lưới điện thông minh, tích hợp các nguồn điện phân tán. Ở Việt Nam, với công nghệ trạm không người trực, việc thu thập, giám sát, điều khiển, vận hành các trạm biến áp được thực hiện bởi các trung tâm điều khiển xa thông qua hệ thống máy tính human machine interface (HMI), các giao thức truyền thông như Modbus, IEC60870-5-101/103/104, DNP3 và IEC61850 (Hình 1). Việc ứng dụng công nghệ thông tin có vai trò lớn trong việc hiện đại hóa vận hành trạm biến áp, tiết giảm nhân lực và tiết kiệm chi phí vận hành hệ thống điện. Tuy nhiên, việc ngày càng phụ thuộc vào công nghệ thông tin trong điều khiển, vận hành hệ thống điện cũng khiến cho vấn đề an ninh và bảo mật trong hệ thống điện trở nên cấp thiết.

Ngày càng có nhiều các vụ tấn công an ninh mạng vào hệ thống điện trên thế giới được ghi nhận. Điển hình nhất có thể kể đến là vụ tấn công an ninh mạng vào hệ thống điện của Ukraine năm 2015. Vào ngày 23/12/2015, hacker đã thâm nhập vào hệ thống thông tin của ba công ty phân phối điện ở Ukraine. Một trong ba công ty bị ảnh hưởng nặng nhất với 30 trạm biến áp (7 trạm 110kV và 23 trạm 35kV) bị cắt điện trong thời gian từ 1 đến 6 giờ, khoảng 230 nghìn người bị ảnh hưởng bởi mất điện. Theo kết quả điều tra sau đó, vụ tấn công được thực hiện từ máy tính có địa chỉ IP từ Nga. Hacker đã gửi các mã độc đến các công ty điện lực này qua email, sau đó truy cập vào quyền điều khiển hệ thống SCADA, gửi lệnh cắt điện đến các trạm



Hình 1. *Quy định 176/EVN về cấu hình SCADA trạm 500/220/110kV*

biến áp, xóa các file dữ liệu trên hệ thống máy tính, tấn công từ chối dịch vụ chăm sóc khách hàng để ngăn chặn thông tin phản ánh mất điện về tổng đài. Đây là vụ tấn công an ninh mạng vào hệ thống điện thành công đầu tiên được ghi nhận. Các vụ tấn công an ninh mạng tương tự vào hệ thống điện các nước Mỹ, Nga, Iran ... cũng đã được ghi

[1] Hanoi University of Science and Technology (Nguyen Quoc Minh, Nguyen Tran Minh Trang, Nguyen Tien Thanh)
[2] Northern Electrical Testing one member Company Limited (Dam Ta Hai)

nhận với mức độ ảnh hưởng và thiệt hại khác nhau. Ở Việt Nam, mặc dù chưa ghi nhận vụ tấn công an ninh mạng vào hệ thống điện nào, tuy nhiên với sự phát triển lưới điện hiện đại, ứng dụng công nghệ thông tin ngày càng mạnh mẽ vào các khâu giám sát, điều khiển và vận hành hệ thống điện thì nguy cơ này ngày càng trở nên hiện hữu. Cách thức các vụ tấn công an ninh mạng thường sử dụng đó là truy cập vào quyền điều khiển hệ thống SCADA/EMS, gửi lệnh đóng cắt các thiết bị để gây mất điện, thay đổi cài đặt của hệ thống rơ le bảo vệ, xóa hoặc thay đổi các thông số vận hành như dòng điện, điện áp, công suất đo được khiến hệ thống bảo vệ hiểu nhầm là có sự cố. Việc phân loại được các trạng thái hệ thống điện bị sự cố một cách tự nhiên và trạng thái hệ thống điện bị tấn công an ninh mạng căn cứ vào các thông số đo được trong trường hợp này là rất khó khăn, ngay cả đối với các kỹ sư vận hành lâu năm cũng khó có thể phát hiện được.

Trong những năm gần đây, sự xuất hiện của lưới điện thông minh đã góp phần thúc đẩy các nghiên cứu về kỹ thuật phát hiện các hành vi xâm nhập và tấn công vào hệ thống điện. Một trong phương pháp phát hiện xâm nhập là tập trung vào các thiết bị điện tử thông minh (IED) trong lưới điện. Nghiên cứu của Chee-Wooi Ten [1] đã phát triển một phương pháp phát hiện sự xâm nhập dựa trên lịch sử bản ghi sự kiện của các thiết bị thông minh này. Phương pháp của Chee-Wooi Ten có hạn chế, đó là chỉ có thể phát hiện sự xâm nhập vào 01 thiết bị điện tử thông minh trong một thời điểm. Một phương pháp khác được đề xuất bởi Chen [2], nhằm phát hiện xâm nhập vào hệ thống điện của các hộ dân và tòa nhà thông minh. Trong phương pháp này, Chen đề xuất mô hình hàm thuần nhất để phát hiện xâm nhập căn cứ vào 03 yếu tố: Mức độ bảo mật của các thiết bị, lịch sử sử dụng điện và giá điện. Mô hình này có thể phát hiện được sự xâm nhập vào nhiều thiết bị điện tử thông minh cùng một lúc. Một hướng nghiên cứu khác là tập trung vào phân tích, đánh giá luồng dữ liệu thông tin trao đổi trong hệ thống điện thông qua các giao thức công nghiệp như IEC61850, Modbus/TCP. Nghiên cứu của Hadeli [3] đã đề xuất một phương pháp phát hiện xâm nhập dựa trên phân tích các mẫu dữ liệu tạo ra bởi các thiết bị truyền qua các giao thức công nghiệp. Phương pháp của Hadeli tỏ ra hiệu quả trong việc phát hiện sự xâm nhập thông qua các mẫu dữ liệu bất thường truyền qua mạng; Tuy nhiên, phương pháp này không phát hiện được việc can thiệp trực tiếp vào hệ thống điều khiển, truyền đi lệnh đóng cắt đến các thiết bị đóng cắt gây mất điện diện rộng. Một số nghiên cứu đã cho thấy, các thuật toán máy học có khả năng ứng dụng mạnh mẽ trong các vấn đề của hệ thống điện như bài toán dự báo phụ tải [4-5], dự báo bức xạ/ công suất phát của điện mặt trời [6], nhận dạng và định vị sự cố [7-8] … Trong nghiên cứu này, nhóm tác giả đề xuất sử dụng các thuật toán máy học nhằm phát hiện trạng thái hệ thống điện bị tấn công an ninh mạng.

2. Mô hình các thuật toán máy học

Trong phần này, nhóm tác giả sẽ giới thiệu mô hình một số thuật toán máy học ứng dụng trong lớp các bài toán nhận dạng trạng thái, và từ đó áp dụng vào bài toán nhận dạng trạng thái hệ thống điện bị tấn công an ninh mạng.

2.1. Thuật toán random forest

Thuật toán random forest là một trong những thuật toán máy học phổ biến, có khả năng ứng dụng trong các lớp bài toán hồi quy và phân loại [9]. Tư tưởng của thuật toán là sẽ tạo ra một khu rừng với nhiều cây quyết định (decision tree). Nói chung, nếu càng có nhiều cây quyết định thì các dự đoán sẽ càng chắc chắn, và do đó độ chính xác của mô hình càng cao. Mỗi một cây quyết định trong mô hình sẽ có các node. Các node thể hiện câu hỏi là node hình chữ nhật, còn các node thể hiện kết quả là các node hình tròn. Các câu hỏi trong mô hình cây quyết định là các câu hỏi dưới dạng nhị phân (đúng hoặc sai). Khi bộ dữ liệu huấn luyện được đưa vào thì các cây quyết định sẽ đưa ra kết quả phân loại, dựa trên các bộ câu hỏi nhị phân. Kết quả phân loại của thuật toán random forest sẽ dựa trên số lượng phiếu bầu (vote) lớn nhất từ các cây quyết định này. Thuật toán random forest có ưu điểm là có khả năng phân loại với độ chính xác cao ngay cả với các bộ dữ liệu bị thiếu, có khả năng tính toán với dữ liệu đầu vào lớn, đa chiều.

2.2. Thuật toán support vector machine (SVM)

SVM là một thuật toán có thể sử dụng cho cả bài toán phân loại và hồi quy, tuy nhiên chủ yếu được sử dụng cho bài toán phân loại [9]. Trong thuật toán này, dữ liệu được biểu diễn dưới dạng các điểm trong không gian n chiều (với n là số các đặc trưng của dữ liệu). Thuật toán này sẽ tìm đường ranh giới (hyperlane) để phân chia các điểm dữ liệu thành 2 hay nhiều loại sao cho khoảng cách từ các điểm dữ liệu tới đường ranh giới là xa nhất có thể.

2.3. Thuật toán K-nearest neighbor (KNN)

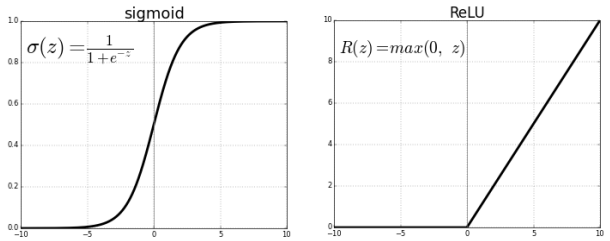
KNN là một trong những thuật toán máy học đơn giản nhất. Khi huấn luyện mô hình, thuật toán này không học từ dữ liệu, mọi tính toán sẽ được thực hiện khi nó cần dự đoán kết quả của dữ liệu mới [9]. KNN có thể áp dụng trong cả bài toán phân loại và hồi quy. Trong bài toán phân loại, một điểm dữ liệu mới sẽ được phân loại trực tiếp từ K điểm dữ liệu gần nhất trong tập dữ liệu huấn luyện.

2.4. Thuật toán XGBoost

XGBoost, viết tắt của từ eXtreme Gradient Boosting là một thuật toán máy học dựa trên cây quyết định, sử dụng phương pháp độ dốc tăng cường [9]. Đây là thuật toán mới được phát triển từ năm 2016 tại Đại học Washington, Hoa Kỳ và có khả năng ứng dụng để giải quyết các bài toán hồi quy, phân loại, xếp hạng và dự đoán. Thuật toán này có ưu điểm là có tốc độ tính toán rất nhanh với các bộ dữ liệu lớn, đa chiều.

2.5. Mạng nơ ron nhân tạo

Khi nói đến dữ liệu dạng bảng có cấu trúc thì các thuật toán máy học dựa trên cây quyết định thường sẽ cho kết quả tốt. Tuy nhiên, đối với các dạng dữ liệu phi cấu trúc như hình ảnh, giọng nói, văn bản thì mạng nơ ron nhân tạo lại có xu hướng làm việc tốt hơn. Mạng nơ ron nhân tạo mô phỏng hoạt động của bộ não con người. Cấu trúc này bao gồm lớp dữ liệu đầu vào (input layer), các lớp ẩn (hidden layer) và lớp kết quả đầu ra (output layer). Các nơ ron của một lớp liên kết với các nơ ron của lớp liền kề thông qua các hàm kích hoạt (activation function) có trọng số. Các hàm kích hoạt này là các hàm phi tuyến, đặc trưng cho mối quan hệ phức tạp của dữ liệu.



Hình 2. *Hàm kích hoạt sigmoid (trái) và ReLU (phải)*

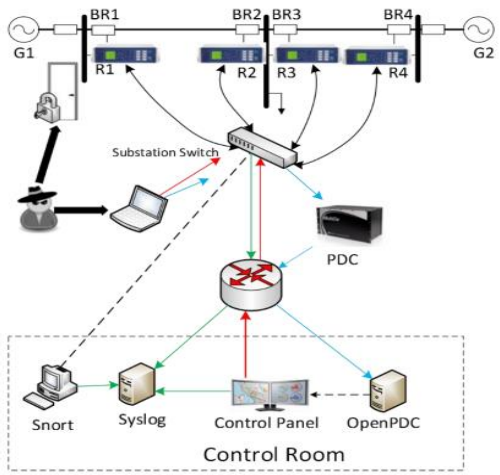
Có hai hàm kích hoạt được sử dụng phổ biến trong mạng nơ ron nhân tạo là hàm sigmoid (PT. 1) và hàm reLU (PT. 2).

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (1)$$

$$R(x) = max(0, x) \quad (2)$$

Hàm sigmoid là hàm kích hoạt phi tuyến được sử dụng phổ biến nhất trong mạng nơ ron nhân tạo. Hàm này nhận giá trị đầu vào bất kỳ và cho giá trị đầu ra biến thiên trong khoảng (0-1). Chính vì đặc điểm này nên hàm sigmoid thường được dùng để biến một giá trị thực thành xác suất. Với một giá trị đầu vào âm lớn thì hàm sigmoid sẽ tiến dần tới 0, và ngược lại với giá trị đầu vào dương lớn thì hàm sigmoid sẽ tiến dần tới 1, nếu đầu vào bằng 0 thì hàm sigmoid sẽ có giá trị bằng 0,5. Chính vì vậy, phương trình giá trị đầu vào bằng 0 thường được coi là đường biên để phân loại đầu ra theo dạng nhị phân. Nhược điểm của hàm sigmoid là nhanh chóng bão hòa đến giá trị 0 hoặc 1 khi trị tuyệt đối của giá trị đầu vào lớn, điều này dẫn tới đạo hàm bị triệt tiêu, khiến cho tốc độ tính toán bị suy giảm đáng kể. Một số nghiên cứu gần đây đã chỉ ra rằng, việc sử dụng hàm kích hoạt ReLU có thể khắc phục được vấn đề triệt tiêu đạo hàm, tốc độ tính toán cũng được cải thiện do đạo hàm của hàm ReLU là hằng số khi giá trị đầu vào dương.

3. Bộ dữ liệu dùng trong nghiên cứu



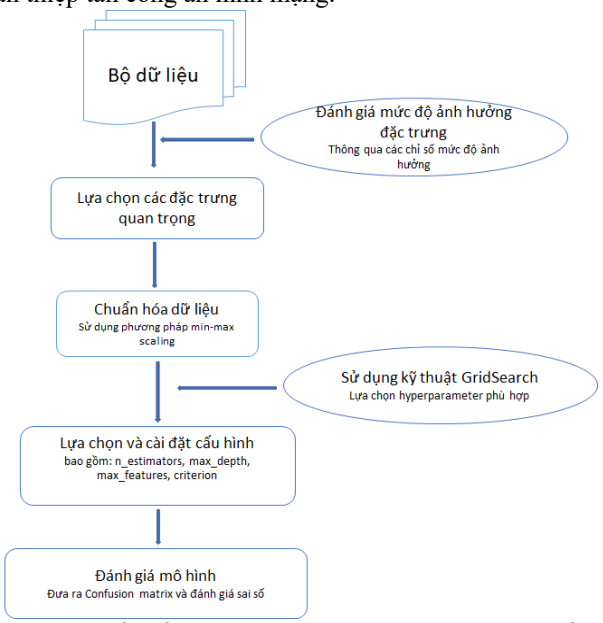
Hình 3. *Mô hình thí nghiệm của ORNL*

Trong nghiên cứu này, nhóm tác giả sử dụng bộ dữ liệu được tạo ra từ phòng thí nghiệm Oak Rigdge National Laboratory, Hoa Kỳ [10]. Mô hình này được thể hiện ở Hình 3. Mô hình này bao gồm một hệ thống điện đơn giản với 2 máy phát G1, G2 nối vào 2 đường dây. Các máy cắt BR1, BR2, BR3, BR4 được đặt ở 2 đầu mỗi đường dây. Bốn thiết bị PMU R1, R2, R3, R4 được đặt ở vị trí các thanh cái để đo các giá trị biên độ, góc pha của dòng điện và điện áp các pha; Dòng điện và điện áp các thành phần thứ tự thuận, nghịch, không; Tần số; tốc độ biến thiên tần số; Tổng trở. Mỗi PMU đo được 29 thông số, như vậy 4 PMU sẽ đo được 116 thông số, ngoài ra có thêm 12 thông số từ các bộ điều khiển, trạng thái của rơ le nên tổng số các thông số đầu vào là 128 (Hình 4). Đây cũng chính là 128 đặc trưng được sử dụng trong mô hình nhận dạng. Các thiết bị này được nối trực tiếp tới hệ thống điều khiển trung tâm. Hệ thống này sẽ tạo ra 5 kịch bản: 1) Hệ thống làm việc bình thường; 2) Sự cố ngắn mạch trên đường dây; 3) Thay đổi cài đặt của rơ le; 4) Gửi lệnh đóng cắt tới rơ le; 5) Chèn dữ liệu, thay đổi các thông số U, I.

Feature	Description
PA1:VH – PA3:VH	Phase A - C Voltage Phase Angle
PM1: V – PM3: V	Phase A - C Voltage Phase Magnitude
PA4:IH – PA6:IH	Phase A - C Current Phase Angle
PM4: I – PM6: I	Phase A - C Current Phase Magnitude
PA7:VH – PA9:VH	Pos. – Neg. – Zero Voltage Phase Angle
PM7: V – PM9: V	Pos. – Neg. – Zero Voltage Phase Magnitude
PA10:VH - PA12:VH	Pos. – Neg. – Zero Current Phase Angle
PM10: V - PM12: V	Pos. – Neg. – Zero Current Phase Magnitude
F	Frequency for relays
DF	Frequency Delta (dF/dt) for relays
PA:Z	Appearance Impedance for relays
PA:ZH	Appearance Impedance Angle for relays
S	Status Flag for relays

Hình 4. *Các thông số PMU đo được*

Trong số 5 kịch bản này thì 2 kịch bản ban đầu là các chế độ làm việc của hệ thống điện không có sự can thiệp tấn công an ninh mạng, còn 3 kịch bản sau là các kịch bản có sự can thiệp tấn công an ninh mạng. Để thuận lợi cho việc xây dựng mô hình thì nhóm tác giả rút gọn 5 kịch bản trên thành 3 kịch bản: 1) Hệ thống điện làm việc bình thường; 2) Có sự cố ngắn mạch trên đường dây; 3) Có sự can thiệp tấn công an ninh mạng.



Hình 5. *Sơ đồ khối của mô hình nhận dạng trạng thái hệ thống điện bị tấn công an ninh mạng dựa trên machine learning*

Hình 5 thể hiện sơ đồ khối của mô hình nhận dạng trạng thái hệ thống điện bị tấn công an ninh mạng dựa trên machine learning. Đầu tiên, mô hình sẽ đánh giá mức độ ảnh hưởng của các đặc trưng, các đặc trưng có mức độ ảnh hưởng đáng kể sẽ được dữ lại. Bước tiếp theo, dữ liệu sẽ được chuẩn hóa theo phương pháp min-max scaling để đưa

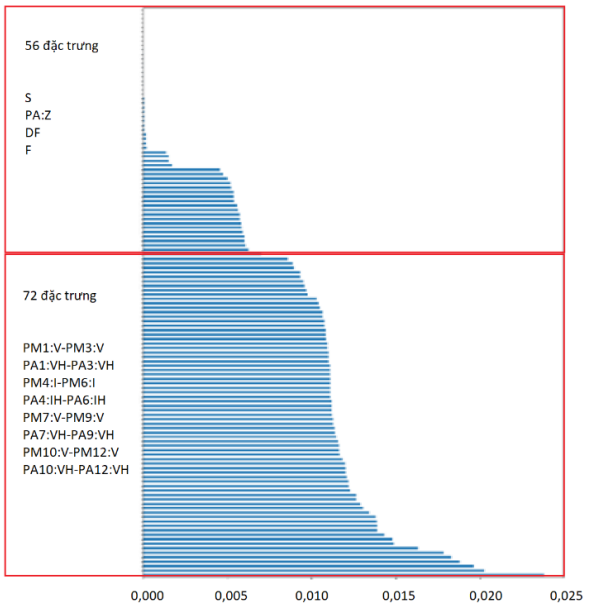
các giá trị biến thiên trong các khoảng khác nhau của các đặc trưng về cùng một khoảng biến thiên. Sau khi chuẩn hóa dữ liệu, nhóm nghiên cứu có sử dụng kỹ thuật Grid Search, là một phương pháp hiệu quả để tìm được bộ tham số tốt nhất trong học có giám sát và cải thiện hiệu suất tổng quát hóa của một mô hình. Sau khi tìm được bộ tham số tối ưu cho mô hình thì bước cuối cùng là đánh giá độ chính xác của mô hình.

4. Kết quả

	A	B	C	D	E	F	G
1	id	R1-PA1:VH	R1-PM1:V	R1-PA2:VH	R1-PM2:V	R1-PA3:VH	R1-PM3:V
2	1	70.39932	127673.1	-49.5723	127648	-169.5783	127723.2
3	2	73.6881	130280.7	-46.3007	130255.6	-166.2781	130355.9
4	3	73.73394	130305.8	-46.2549	130280.7	-166.2322	130381
5	4	74.08344	130581.6	-45.8996	130556.5	-165.8827	130656.8
6	5	74.55327	131083.1	-45.4241	131058	-165.4244	131158.3
7	6	74.54754	131058	-45.4413	131032.9	-165.4416	131133.2
8	7	74.53608	131007.8	-45.4585	131007.8	-165.453	131083.1
9	8	74.5017	130982.8	-45.4757	130957.7	-165.4759	131058
10	9	74.44441	130932.6	-45.533	130932.6	-165.5218	131007.8
11	10	74.28398	130857.4	-45.7163	130857.4	-165.7051	130932.6
12	11	74.08917	130832.3	-45.8939	130832.3	-165.8942	130932.6
13	12	73.88864	130832.3	-46.0887	130807.2	-166.0947	130907.5
14	13	73.70529	130807.2	-46.2893	130807.2	-166.2724	130907.5
15	14	73.68237	130832.3	-46.3122	130807.2	-166.2895	130907.5
16	15	73.65373	130832.3	-46.3408	130807.2	-166.3411	130907.5
17	16	73.41308	130832.3	-46.5815	130782.2	-166.5703	130907.5
18	17	73.20682	130832.3	-46.782	130807.2	-166.7823	130907.5
19	18	73.20109	130832.3	-46.7935	130782.2	-166.7823	130907.5
20	19	73.09223	130832.3	-46.8737	130807.2	-166.8854	130882.5
21	20	73.08077	130832.3	-46.8909	130807.2	-166.9083	130907.5

Hình 6. *Dạng dữ liệu đo được từ PMU*

Hình 6 thể hiện cấu trúc bảng dữ liệu đầu vào. Bảng dữ liệu này bao gồm 128 cột (tính từ cột B) thể hiện 128 đặc trưng đo được từ các PMU, và 74490 hàng đại diện cho số lượng các trạng thái được tạo ra từ hệ thống điều khiển trung tâm. Đây là một bộ dữ liệu tương đối lớn với rất nhiều đặc trưng nên cần thiết phải có các bước tiền xử lý nhằm giảm thời gian tính toán của mô hình. Trước hết có thể nhận thấy, các đặc trưng như dòng điện, điện áp, tần số, tổng trở có giải biến thiên tương đối khác nhau. Điều này có thể dẫn đến các sai số trong việc xác định các trọng số của mô hình, do đó nhóm tác giả đã chuẩn hóa các đặc trưng theo phương pháp min-max scaling (PT. 3).



Hình 7. *Mức độ ảnh hưởng của 128 đặc trưng đến mô hình nhận dạng trạng thái*

Đây là phương pháp đơn giản cho phép biến một đại lượng X có giải biến thiên bất kỳ thành đại lượng X' có giải biến thiên từ 0 đến 1.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

Sau khi đã chuẩn hóa các đặc trưng, bước tiếp theo nhóm tác giả sử dụng mô hình random forest để đánh giá sơ bộ ảnh hưởng của các đặc trưng đến bài toán nhận dạng trạng thái, kết quả được thể hiện ở Hình 12. Trong hình này, trục tung là các đặc trưng, còn trục hoành thể hiện mức độ ảnh hưởng của từng đặc trưng đến bài toán nhận dạng trạng thái, sao cho tổng của chúng bằng 1. Căn cứ vào Hình 12 có thể thấy, các đặc trưng có mức độ ảnh hưởng khác nhau đến mô hình nhận dạng trạng thái. Có 72 đặc trưng có ảnh hưởng lớn, đó là các đặc trưng về biên độ và góc pha của dòng điện, điện áp các pha; dòng điện, điện áp các thành phần thứ tự thuận, nghịch không. Các đặc trưng này sẽ được dữ lại để huấn luyện mô hình. 56 đặc trưng còn lại là các đặc trưng về trạng thái on/ off của rơ le (S), tổng trở rơ le đo được (PA:Z), tốc độ biến thiên của tần số (DF), độ lớn của tần số (F) ít ảnh hưởng đến mô hình nhận dạng trạng thái nên sẽ bị loại bỏ. Việc loại bỏ các đặc trưng ít hoặc không ảnh hưởng đến bài toán nhận dạng trạng thái là một thủ thuật phổ biến nhằm làm tăng tốc độ tính toán của thuật toán mà vẫn đảm bảo được độ chính xác.

Sau khi đã loại bỏ các đặc trưng không quan trọng, một vấn đề nữa cần giải quyết là bộ số liệu nhóm tác giả sử dụng có sự mất cân bằng lớn về tỉ lệ các trạng thái. Trong số 03 trạng thái thì trạng hệ thống điện bị tấn công chiếm đa số (> 70%) trong bộ dữ liệu mà nhóm nghiên cứu có sử dụng. Việc mất cân bằng dữ liệu lớn như vậy sẽ làm cho việc nhận dạng kém chính xác trên nhóm các trạng thái thiểu số là trạng thái bình thường và trạng thái có ngắn mạch trên đường dây. Trong nghiên cứu này, nhóm tác giả đề xuất sử dụng phương pháp SMOTE (Synthetic Minority Oversampling Technique) nhằm giải quyết vấn đề mất cân bằng của dữ liệu. Kỹ thuật này tạo ra dữ liệu tổng hợp cho các trạng thái dữ liệu thiểu số thiểu số, tiến hành bằng cách nối các điểm của lớp thiểu số với các đoạn thẳng và sau đó đặt các điểm nhân tạo trên các đoạn thẳng này.Về cơ bản, thuật toán SMOTE hoạt động theo 4 bước đơn giản:

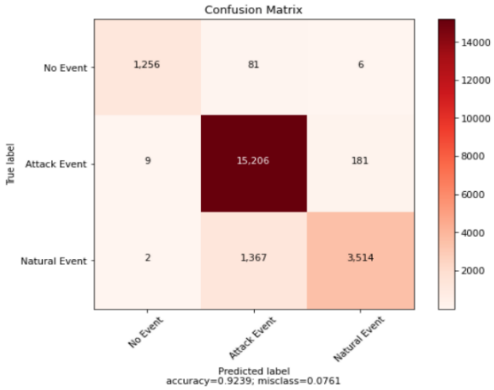
1. Chọn một vectơ đầu vào của lớp thiểu số.
2. Tìm k lân cận gần nhất của vectơ đó.
3. Chọn một trong những vùng lân cận này và đặt một điểm tổng hợp ở bất kỳ đâu trên đường thẳng nối với điểm đang xem xét và điểm lân cận đã chọn của nó.
4. Lặp lại các bước cho đến khi dữ liệu được cân bằng.

Để đánh giá trực quan độ chính xác của mô hình nhận dạng, nhóm tác giả sử dụng ma trận hợp nhất (confusion matrix). Hình 8 thể hiện ma trận hợp nhất của thuật toán random forest. Ma trận này có kích thước là 3×3 do có 3 trạng thái cần phân loại: no event là trạng thái hệ thống điện làm việc bình thường, natural event là trạng thái sự cố ngắn mạch trên đường dây (không có sự can thiệp của tấn công an ninh mạng) và attack event là trạng thái có sự can thiệp của tấn công an ninh mạng. Trục hoành thể hiện trạng thái dự đoán của mô hình, còn trục tung thể hiện trạng thái thực tế của mô hình.

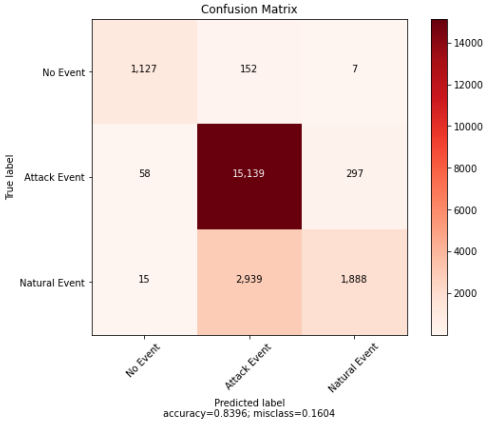
Với định nghĩa như vậy thì có thể thấy, các phần tử trên đường chéo chính của ma trận thể hiện số trạng thái có dự đoán giống với thực tế, còn các phần tử nằm ngoài đường chéo chính thể hiện số dự đoán khác với thực tế. Ví dụ, phần tử C(3,2) của ma trận có trị số bằng 1367, nghĩa là có 1367 trạng thái thực tế là chế độ sự cố ngắn mạch nhưng mô hình dự đoán nhầm thành trạng thái bị tấn công an ninh mạng. Căn cứ vào ma trận hợp nhất ta có thể tính được độ chính xác của mô hình theo công thức:

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

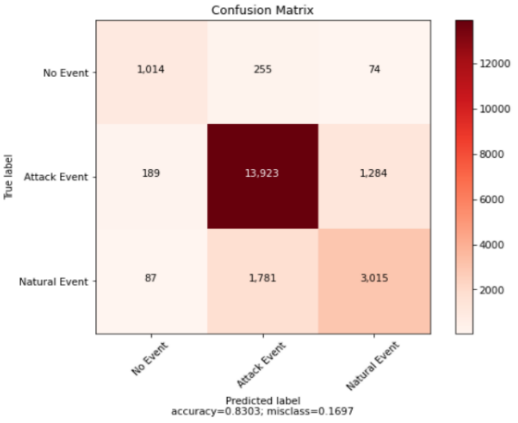
Ta có thể thấy, độ chính xác của thuật toán random forest trong trường hợp này là 0,9239. Tương tự, ma trận hợp nhất của các thuật toán XGBoost, KNN, SVM và ANN được thể hiện ở Hình 9-12.



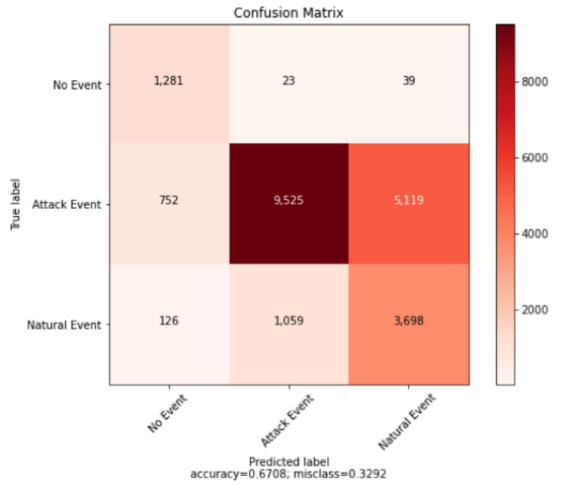
Hình 8. *Ma trận hợp nhất của thuật toán random forest*



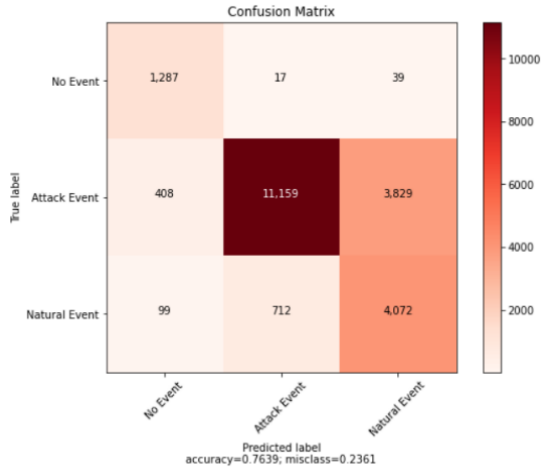
Hình 9. *Ma trận hợp nhất của thuật toán XGBoost*



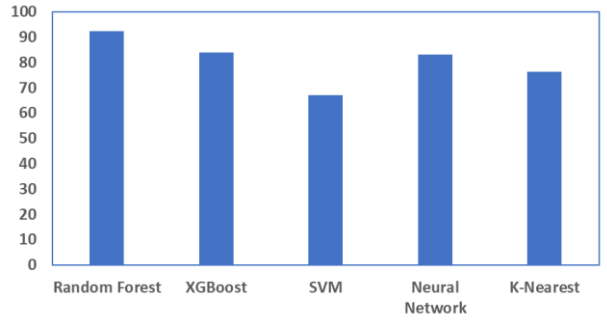
Hình 10. *Ma trận hợp nhất của thuật toán KNN*



Hình 11. *Ma trận hợp nhất của thuật toán SVM*



Hình 12. *Ma trận hợp nhất của thuật toán ANN*



Hình 13. *So sánh độ chính xác của các thuật toán*

Độ chính xác của các thuật toán theo phần trăm được thể hiện ở Hình 13. Có thể thấy, thuật toán random forest đạt được độ chính xác cao nhất là 92,39%. Các thuật toán máy học khác như XGBoost, SVM, KNN chỉ đạt được độ chính xác trong khoảng 67%-83%. Thuật toán random forest đã thể hiện được ưu điểm rõ rệt trong các bộ dữ liệu lớn có cấu trúc dạng bảng, do cơ chế sử dụng nhiều cây quyết định nên có khả năng phân loại tốt trong các trường hợp mà dữ liệu đầu vào có sự chênh lệch lớn giữa số lượng các trạng thái. Bên cạnh độ chính xác thì thời gian huấn luyện và nhận dạng cũng là một yếu tố quan trọng để đánh giá hiệu quả của thuật toán. Bảng 1 thể hiện thời gian huấn luyện và tính toán của các thuật toán được sử

dụng trong nghiên cứu. Có thể thấy, mạng nơ ron nhân tạo không thể hiện được ưu thế trong các dạng dữ liệu cấu trúc khi có thời gian tính toán lớn (1742 giây) và chỉ đạt được độ chính xác là 77,58%.

Bảng 1. *So sánh thời gian huấn luyện và nhận dạng của các thuật toán*

	Random Forest	XGBoost	SVM	KNN	Neural Network
Thời gian huấn luyện (s)	384	59	450	12	1742
Thời gian nhận dạng (s)	0,0611	0,00461	0,0160	0,0040	0,0432

5. Kết luận

Trong nghiên cứu này, nhóm tác giả đề xuất sử dụng các thuật toán máy học nhằm phát hiện trạng thái hệ thống điện bị tấn công an ninh mạng. Với đặc điểm dữ liệu đo được từ hệ thống điều khiển xa là loại dữ liệu có cấu trúc thì các thuật toán máy học tỏ ra có ưu điểm, đặc biệt thuật toán random forest đạt được độ chính xác 92,39%. Việc phát hiện được các trạng thái hệ thống điện có sự can thiệp tấn công an ninh mạng đóng vai trò quan trọng, giúp cho người và hệ thống điều khiển đưa ra các quyết định chính xác và kịp thời nhằm ngăn chặn và giảm thiểu nguy cơ từ không gian mạng đến sự vận hành an toàn của hệ thống điện. Trong các nghiên cứu tiếp theo, nhóm tác giả sẽ tập trung vào việc cải thiện các mô hình hiện có nhằm nâng cao độ chính xác của bài toán nhận dạng trạng thái.

TÀI LIỆU THAM KHẢO

[1] Chee-Wooi Ten, Junho Hong and Chen-Ching Liu, "Anomaly Detection for Cybersecurity of the Substations", *IEEE Transactions on Smart Grids*,vol. 2, no. 4, pp.865,873, Dec. 2011

[2] Y. Chen and B. Lou, "S2a: Secure smart household appliances", in *ACM Conference in Data Application Security Privacy*, San Antonio, TX, USA, pp. 217-228, Feb. 2012.

[3] Hadeli, H.; Schierholz, R.; Braendle, M. and Tuduce, C., "Leveraging determinism in industrial control systems for advanced anomaly detection and reliable security configuration", *Emerging Technologies & Factory Automation (ETFA)*, pp.1-8, 22-25, Sept. 2009.

[4] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo and Y. Xu, "Short-Term Residential Load Forecasting Based on Resident Behaviour Learning", in *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 1087-1088, Jan. 2018.

[5] C. Huang and P. Kuo, "Multiple-Input Deep Convolutional Neural Network Model for Short-Term Photovoltaic Power Forecasting", in *IEEE Access*, vol. 7, pp. 74822-74834, 2019.

[6] B. P. Mukhoty, V. Maurya and S. K. Shukla, "Sequence to sequence deep learning models for solar irradiation forecasting", *IEEE Milan PowerTech*, pp. 1-6, 2019.

[7] K. Moloi and A. O. Akumu, "Power distribution fault diagnostic method based on machine learning technique", 2019 IEEE PES/IAS PowerAfrica, pp. 238-242, 2019.

[8] T. Goswami and U. B. Roy, "Predictive Model for Classification of Power System Faults using Machine Learning", in *IEEE Region 10 Conference (TENCON)*, pp. 1881-1885, 2019.

[9] Aurélien Géron, "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow", *O'Reilly*, 2nd edition, 2019, ISBN: 978-1-492-03264-9.

[10] S. Pan, T. Morris and U. Adhikari, "Developing a Hybrid Intrusion Detection System Using Data Mining for Power Systems", in *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 3104-3113, Nov. 2015.