

**BỘ GIÁO DỤC
VÀ ĐÀO TẠO**

**VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM**

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ



BÙI DUY LINH

**NGHIÊN CỨU MÔ HÌNH DỰ BÁO NGẮN HẠN
CÔNG SUẤT PHÁT CỦA NHÀ MÁY ĐIỆN MẶT TRỜI
SỬ DỤNG MẠNG NƠ-RON HỒI QUY**

TÓM TẮT LUẬN ÁN TIẾN SỸ KỸ THUẬT NĂNG LƯỢNG

HÀ NỘI – 2024

**Công trình được hoàn thành tại: Học viện Khoa học & Công nghệ
- Viện Hàn lâm Khoa học và Công nghệ Việt Nam.**

Người hướng dẫn khoa học 1: TS. Nguyễn Quang Ninh

Người hướng dẫn khoa học 2: TS. Đoàn Văn Bình

Phản biện 1:

Phản biện 2

Phản biện 3:

Luận án được bảo vệ trước Hội đồng chấm luận án tiến sĩ, Học viện Khoa học và Công nghệ - Viện Hàn lâm Khoa học và Công nghệ Việt Nam vào hồi giờ ngày tháng năm 20..

Có thể tìm hiểu luận án tại thư viện:

- Thư viện Quốc gia Việt Nam
- Thư viện Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

MỞ ĐẦU

1. Tính cấp thiết của luận án

- Hiện nay tỷ trọng nguồn điện mặt trời (ĐMT) so với công suất phát toàn hệ thống Việt Nam trong ngày cao nhất thường ở mức khoảng 30%-35%, tuy nhiên trong những ngày phụ tải thấp như các kỳ nghỉ lễ dài ngày, tỷ trọng này có thể lên tới gần 50%. Do tính chất biến động mạnh theo các điều kiện thời tiết, đây là loại hình nguồn khó dự báo và mang đến nhiều thách thức trong vận hành, cụ thể đó là: Hệ thống phải duy trì lượng dự phòng công suất lớn để bảo đảm đáp ứng phụ tải khi công suất phát của các nguồn NLTT thay đổi mạnh, liên tục với biên độ lớn, bất định; Do các nguồn này sử dụng chủ yếu các thiết bị điện tử công suất, hầu như không đóng góp quán tính cho hệ thống nên khi tỷ trọng vận hành lớn sẽ dẫn đến độ dự trữ ổn định của hệ thống điện bị suy giảm và nguy cơ mất ổn định tần số và điện áp trên lưới điện sẽ tăng lên.

- Do đó, để đảm bảo an toàn, tối ưu cho công tác vận hành thì dự báo với độ chính xác cao công suất phát các nguồn này là một yêu cầu hết sức cấp thiết. Việc nghiên cứu các công nghệ để phát triển các mô hình dự báo ngắn hạn công suất các nhà máy ĐMT đặc biệt là các mô hình mạng nơ-ron hồi quy là một hướng đi tiềm năng, có tính ứng dụng cao trong bối cảnh xu hướng phát triển mạnh mẽ của các nguồn ĐMT như hiện nay. Việc ứng dụng không chỉ bao gồm công tác xử lý dữ liệu và huấn luyện kiểm thử để xây dựng mô hình tốt nhất, mà còn bao gồm cả việc tổng kết đưa ra được một quy trình triển khai khả thi, có thể áp dụng được trong thực tế tại các nhà máy ĐMT có quy mô công nghiệp.

2. Mục tiêu nghiên cứu

- Xây dựng mô hình dự báo ngắn hạn công suất nhà máy ĐMT với

mạng nơ-ron hồi quy bộ nhớ dài ngắn (LSTM).

- Đề xuất các giải pháp cải tiến hiệu quả xây dựng mô hình dự báo.
- Xây dựng quy trình và công cụ xử lý bài toán dự báo ngắn hạn công suất nhà máy ĐMT sử dụng mạng LSTM.

3. Phương pháp nghiên cứu

Các phương pháp nghiên cứu sử dụng trong luận án gồm:

- Thu thập thông tin về cơ sở khoa học và cơ sở thực tiễn thông qua nghiên cứu tài liệu
- Thực nghiệm
- Phân tích và tổng hợp

4. Phạm vi nghiên cứu

- *Khung dự báo*: Luận án tập trung nghiên cứu các phương pháp dự báo phù hợp cũng như tiến hành thực nghiệm đối với dự báo ngắn hạn.

- *Đối tượng nghiên cứu*: Luận án tập trung nghiên cứu mô hình dự báo ngắn hạn công suất phát cho các nhà ĐMT dạng trang trại quy mô công nghiệp.

- *Hướng tiếp cận bài toán dự báo*: Luận án tập trung vào hướng tiếp cận dự báo trực tiếp công suất phát đầu ra của nhà máy ĐMT từ các yếu tố khí tượng và các yếu tố ảnh hưởng khác được lựa chọn trong quá trình phân tích mô hình.

5. Cơ sở khoa học và thực tiễn của đề tài

- Đề tài dựa trên cơ sở khoa học của lĩnh vực năng lượng mặt trời, mạng nơ-ron hồi quy và các phương pháp dự báo công suất. Các nghiên cứu đã chỉ ra rằng việc áp dụng các thuật toán học sâu này có thể cải thiện đáng kể độ chính xác của dự báo công suất ĐMT.

- Cơ sở thực tiễn của đề tài là nhu cầu ngày càng tăng về dự báo công suất phát của các nhà máy ĐMT trong hệ thống điện Việt Nam nói riêng và thế giới nói chung. Sự phát triển nhanh chóng của ĐMT

trong những năm gần đây đã tạo ra những thách thức trong việc kiểm soát và điều hành hệ thống điện, đặc biệt trong dự báo công suất phát của ĐMT trong ngắn hạn.

6. Những đóng góp mới của luận án

- Xây dựng mô hình dự báo ngắn hạn công suất phát của nhà máy ĐMT sử dụng mạng LSTM, thử nghiệm thành công mô hình cho các nhà máy ĐMT tại Việt Nam.

- Đề xuất các giải pháp cải tiến hiệu quả trong xây dựng mô hình dự báo gồm: (1) Kỹ thuật tiền xử lý số liệu với hệ số P/GHI kết hợp với phân cụm GHI; (2) Kỹ thuật huấn luyện sử dụng số liệu khí tượng dự báo; (3) Kỹ thuật sử dụng dữ liệu bức xạ trời trong thay thế các chỉ dấu thời gian. Đã thực hiện công bố chính thức các kết quả nghiên cứu liên quan trên các tạp chí quốc tế (02 bài báo SCI – Q1, 02 bài báo Scopus – Q3, Q4).

- Xây dựng được quy trình hoàn chỉnh và công cụ dự báo ngắn hạn công suất nhà máy ĐMT sử dụng mạng nơ-ron hồi quy, tạo điều kiện thuận lợi cho người dùng có nhu cầu cho việc triển khai và sử dụng công nghệ dự báo trong thực tế cho các nhà máy ĐMT.

7. Cấu trúc của luận án

Luận án được xây dựng với cấu trúc như sau:

- Mở đầu
- Chương 1: Tổng quan nghiên cứu
- Chương 2: Xây dựng mô hình dự báo ngắn hạn công suất nhà máy ĐMT với mạng nơ-ron hồi quy bộ nhớ dài ngắn
- Chương 3: Giải pháp cải tiến hiệu quả mô hình và xây dựng quy

trình, công cụ dự báo

- Kết luận và hướng nghiên cứu tiếp theo
- Danh mục các công trình đã công bố
- Tài liệu tham khảo
- Phụ lục

CHƯƠNG 1: TỔNG QUAN NGHIÊN CỨU

1.1. Mở đầu

Các nội dung cần nghiên cứu để có góc nhìn toàn diện về vấn đề dự báo công suất phát nhà máy ĐMT.

1.2. Công suất phát đầu ra của nhà máy ĐMT và các yếu tố ảnh hưởng

Các yếu tố ảnh hưởng đến công suất phát đầu ra của một nhà máy điện mặt trời trong thực tế gồm:

- Các yếu tố khí tượng (Bức xạ mặt trời, nhiệt độ, tốc độ gió, độ ẩm...)
- Góc lắp đặt tấm pin
- Phụ tải và chế độ điều khiển
- Giới hạn lưới điện đầu nối

1.3. Phân loại các khung dự báo công suất ĐMT

- Dự báo dài hạn
- Dự báo trung hạn
- Dự báo ngắn hạn

1.4. Phương pháp đánh giá chất lượng mô hình dự báo

1.4.1. Đánh giá thông qua các chỉ số

Từ các nghiên cứu khác nhau tổng hợp các nhóm tiêu chí thường được sử dụng bao gồm: MAE, PE, APE, MAPE, MSE, RMSE, nRMSE.

1.4.2. *Đánh giá thông qua biểu đồ phân bố sai số*

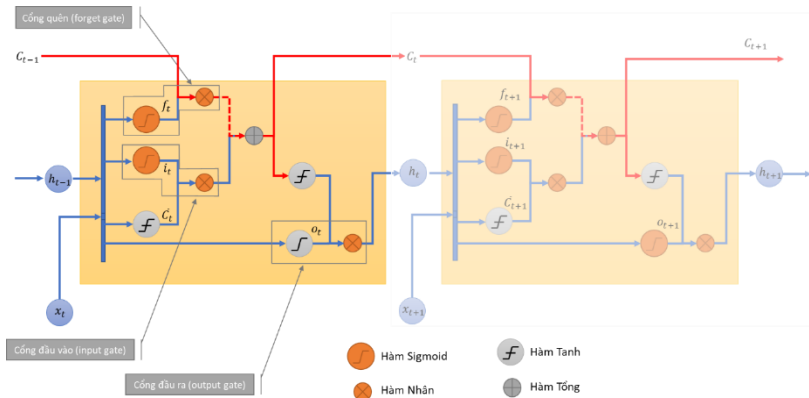
Biểu đồ phân bố sai số (hay còn gọi là phổ sai số) cung cấp một cái nhìn tổng quan cũng như trực quan về xác suất mà một giá trị sai số nhất định sẽ xảy ra.

1.5. Các phương pháp dự báo ngắn hạn công suất NMDMT

Trình bày nghiên cứu về các phương pháp: Vật lý, Thống kê, Học máy, Kết hợp; Các hướng tiếp cận: Gián tiếp, trực tiếp.

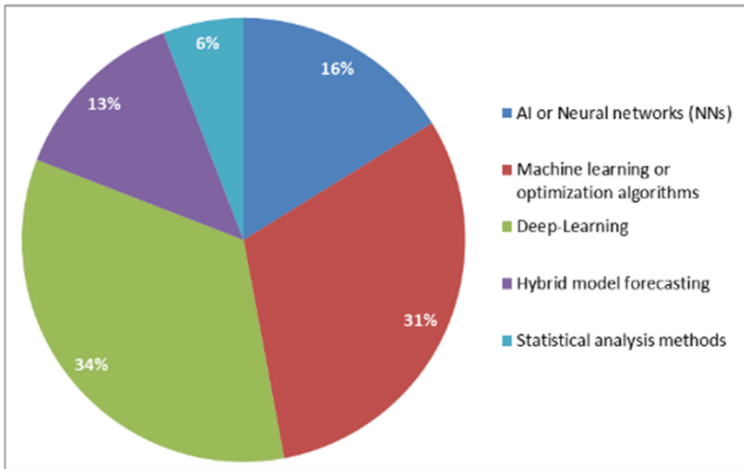
Mạng bộ nhớ ngắn dài LSTM (Long Short Term Memory)

LSTM (Long Short Term Memory) là một loại mạng nơ-ron hồi quy đặc biệt có thể học được cả sự phụ thuộc ngắn hạn cũng như dài hạn. Mạng LSTM được thiết kế để giải quyết vấn đề tiêu biến đạo hàm (vanishing gradient) trong các mạng nơ-ron hồi quy truyền thống.



Hình 1.13. Cấu trúc của một chuỗi khối LSTM

1.6. Tình hình nghiên cứu



Hình 1.14: Tỷ lệ sử dụng các phương pháp trong các nghiên cứu về dự báo công suất ĐMT. Nguồn: Tsai (2023)

Trong nghiên cứu 2023 của Tsai và cộng sự công bố năm 2023, các tác giả tổng hợp 70 nghiên cứu đã công bố trên các tạp chí uy tín gồm từ năm 2020 đến năm 2023 về chủ đề dự báo công suất ĐMT. Qua biểu đồ thể hiện trong Hình 1.14 có thể thấy kết luận rút ra là các phương pháp dự báo dựa trên học sâu đang chiếm ưu thế chủ đạo với 34% các nghiên cứu sử dụng phương pháp này. Trong đó, mô hình mạng học sâu phổ biến nhất được sử dụng là mô hình Long - Short Term Memory (LSTM), đã được chứng minh là hiệu quả trong việc dự báo công suất ĐMT trong ngắn hạn.

Bảng 1.4: Tổng hợp một số các mô hình dự báo ĐMT sử dụng LSTM đã công bố từ 2019

T T	Mô hình	Năm	CS NM	Độ phân giải	Độ dài dự báo	Số bước dự báo	Khu ng dự báo	MAPE (%)	nRMS E (%)
1	Wang và cộng sự	2019	23,4 kWp (19,5 kW)	5 phút	t	1	5 phút	2,092	5,879

T T	Mô hình	Năm	CS NM	Độ phân giải	Độ dài dự báo	Số bước dự báo	Khu ng dự báo	MAPE (%)	nRMS E (%)
2	Ospina và cộng sự	2019	12,6 MW	30 phút	t+1	1	30 phút	6,563	12,184
3	Zhou H và cộng sự	2019	20 kW	7,5 phút	t+1	1	7,5 phút	4,000	6,950
				7,5 phút	Từ t+1 đến t+2	2	15 phút	5,000	8,000
				7,5 phút	Từ t+1 đến t+4	4	30 phút	6,100	9,050
				7,5 phút	Từ t+1 đến t+8	8	60 phút	7,350	10,450
4	Wen và cộng sự	2019	106,6 kWp	60 phút	t	1	60 phút	7,566	13,051
5	Harrou và cộng sự	2020	9 MWp	15 phút	t+1	1	15 phút	8,930	
6	Zhou N và cộng sự	2020	5,83 kWp (4,85 kW)	05 phút	Từ t+1 đến t+28	288	1 ngày	11,639	20,075
7	Zhang và cộng sự	2020	6,41 kW	15 phút	Từ t+1 đến t+96	96	1 ngày	8,418	15,489
8	Park và cộng sự	2021	500 kWp	60 phút	t	1	60 phút		13,200
9	Li và cộng sự	2021	40 kW	60 phút	t	1	60 phút		6,400
10	Zhou H và cộng sự	2021	20 kW	7,5 phút	t+1	1	7,5 phút	3,500	6,500
			20 kW	7,5 phút	Từ t+1 đến t+2	2	15 phút	4,250	7,000
			20 kW	7,5 phút	Từ t+1	4	30 phút	6,900	10,200

T T	Mô hình	Năm	CS NM	Độ phân giải	Độ dài dự báo	Số bước dự báo	Khu ng dự báo	MAPE (%)	nRMS E (%)
					đến t+4				
11	Liu và cộng sự	2021	18,78 kW	05 phút	t	1	05 phút		4,886
12	Pombo và cộng sự	2022	10 kW	05 phút	Từ t+1 đến t+60	60	5 giờ		14,990
13	Suresh và cộng sự	2022	317 kWp (265 kW)	60 phút	Từ t+1 đến t+24	24	1 ngày	4,800	9,046
14	Nguyễn Đức Tuyên và cộng sự	2020	78 MW	60 phút	Từ t+1 đến t+24	24	1 ngày	3,743	5,402

1.7. Kết luận

Các nghiên cứu gần đây đã chứng minh rằng mạng Nơ ron hồi quy, đặc biệt là mô hình LSTM, là một công cụ hiệu quả trong dự báo công suất điện mặt trời trong ngắn hạn. Mặc dù số lượng các nghiên cứu đã công bố trên thế giới trong 5 năm trở lại đây là tương đối nhiều và đa dạng, tuy nhiên các công bố từ các tác giả trong nước còn đang hạn chế. Một phần nguyên nhân bởi ĐMT mới chỉ thực sự bùng nổ phát triển tại Việt Nam từ năm 2019 đến nay dẫn đến việc thiếu các bộ số liệu để phục vụ triển khai các nghiên cứu. Ngoài ra, qua thống kê các nghiên cứu tại Mục 1.6 đã cho thấy các nghiên cứu mới chỉ tập trung vào dự báo công suất cho các hệ thống điện mặt trời quy mô nhỏ và trung bình từ vài chục kW đến vài MW, trong khi mới có ít nghiên cứu về dự báo cho các nhà máy ĐMT quy mô lớn. Điều này cũng sẽ gây khó khăn cho việc triển khai các kết quả ứng dụng trong thực tế bởi tính chất bộ dữ liệu của các nhà máy ĐMT quy mô lớn có tính chất

rất khác khi các bộ dữ liệu khí tượng thu thập được là bộ mang tính đại diện khu vực rất rộng lớn của nhà máy thay vì phản ánh tương đối chính xác như đối với các hệ quy mô nhỏ. Nguyên nhân là bởi thông thường theo quy định, đối với nhà máy ĐMT quy mô công nghiệp thì trung bình từ cứ 30 MW - 50 MW công suất đặt mới có từ 1 đến 2 trạm đo khí tượng.

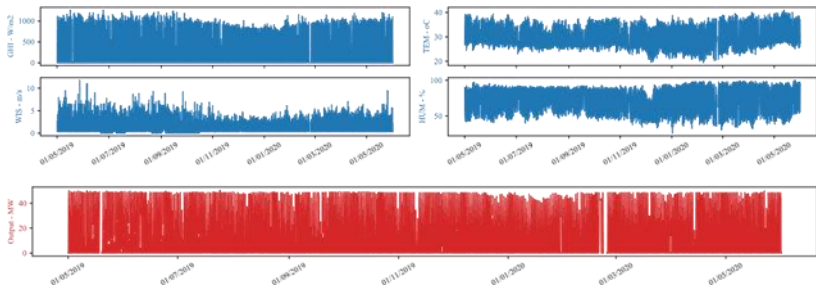
Do đó, luận án này sẽ đi sâu vào hướng nghiên cứu sử dụng mạng nơ-ron hồi quy LSTM để xây dựng mô hình dự báo ngắn hạn công suất phát đầu ra của nhà máy ĐMT quy mô công nghiệp ở Việt Nam. Không chỉ đơn thuần ứng dụng phương pháp LSTM, luận án còn đề xuất các giải pháp nhằm tăng hiệu quả của việc ứng dụng LSTM trong dự báo ngắn hạn công suất phát nhà máy ĐMT. Ngoài ra, để tạo điều kiện cho các nhà nghiên cứu tiếp theo phát triển các kỹ thuật liên quan, luận án phân tích sâu kỹ thuật được đề xuất dựa trên số liệu thống kê được mô tả chi tiết và được sử dụng làm cơ sở để xác định các tham số của mô hình.

CHƯƠNG 2.

XÂY DỰNG MÔ HÌNH DỰ BÁO NGẮN HẠN CÔNG SUẤT NHÀ MÁY ĐMT VỚI MẠNG NƠ-RON HỒI QUY BỘ NHỚ DÀI NGẮN

2.1. Mở đầu

2.2. Thu thập dữ liệu



Hình 2.1. Dữ liệu vận hành quá khứ của nhà máy

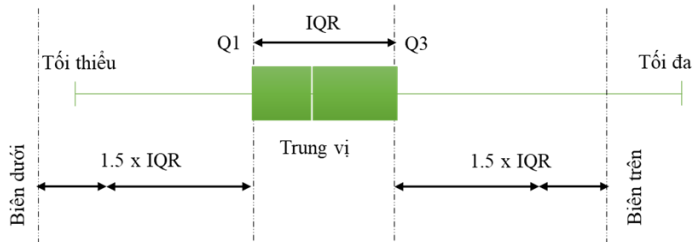
Bộ dữ liệu để thực nghiệm được thu thập từ một nhà máy điện mặt trời tại khu vực phía Nam của Việt Nam từ 5/2019 đến 6/2020.

2.3. Môi trường thí nghiệm

Trình bày về môi trường và công cụ sử dụng thí nghiệm

2.4. Tiền xử lý số liệu

Kỹ thuật khoảng tứ phân vị (Interquartile Range - IQR) được sử dụng để thực hiện việc phát hiện các điểm ngoại lai để loại bỏ tránh gây nhiễu trong quá trình huấn luyện và kiểm tra mô hình.



Hình 2.4. Phương pháp IQR xác định điểm ngoại lai

Quá trình xử lý dữ liệu dựa trên việc áp dụng IQR:

- Trong Bước 1, dữ liệu gốc được xử lý và lọc bằng IQR.
- Trong Bước 2, các điểm dữ liệu bị thiếu được tính toán bằng cách nội suy từ các điểm dữ liệu xung quanh. Tuy nhiên, điều này có thể dẫn đến dữ liệu không chính xác, vì phạm vi dữ liệu bị thiếu là lớn.
- Trong Bước 3, dữ liệu sau Bước 2 tiếp tục được lọc bằng IQR. Các điểm dữ liệu hợp lý sẽ được giữ lại trong khi những điểm ngoại lai sẽ được thay thế bằng trung vị (Q2) của tập giá trị trong khoảng thời gian đó.

2.5. Xây dựng mô hình LSTM và các mô hình so sánh

Các mô hình được lựa chọn để so sánh gồm:

- Mô hình quán tính
- Mô hình ARIMA (4,0,0)
- Các Mô hình MLP (1-4 lớp; 7, 50, 100 nút ẩn trên mỗi lớp)

Để lựa chọn các tham số tối ưu cho mô hình, cấu hình của mạng LSTM và thiết lập cho giai đoạn huấn luyện được lựa chọn như sau:

- Mỗi đầu vào là một ma trận có kích thước 5 hàng x 7 cột
- Số lớp mạng thay đổi từ 1 đến 4.
- Số nút ẩn trên mỗi lớp mạng thay đổi từ 7 (tương ứng với 7 nút đầu vào) đến 50 và 100 nút.

2.6. Huấn luyện mô hình LSTM

Tiến hành huấn luyện các mô hình LSTM theo các cấu hình đã nêu.

2.7. So sánh kết quả dự báo từ các mô hình

Bảng 2.6. Tổng hợp các chỉ số đánh giá kết quả sai số trên tập dữ liệu kiểm tra cho các mô hình khác nhau

Mô hình	Lớp ẩn (L) - số nút (N)	MSE	RMSE	nRMSE	MAE	MAPE
		MW ²	MW	%	MW	%
	Quán tính	128,259	11,325	23,594	7,401	15,418
	ARIMA	29,268	5,410	11,271	3,518	7,329

Mô hình	Lớp ẩn (L) - số nút (N)	MSE MW^2	RMSE MW	nRMSE %	MAE MW	MAPE %
MLP	1L-100N	12,038	3,470	7,229	2,086	4,346
MLP	2L-100N	12,063	3,473	7,235	2,165	4,511
MLP	3L-100N	15,276	3,908	8,142	2,339	4,873
MLP	4L-100N	17,641	4,200	8,750	2,546	5,304
LSTM	1L-7N	12,951	3,599	7,498	2,094	4,362
LSTM	1L-50N	10,700	3,271	6,815	1,901	3,960
LSTM	1L-100N	12,126	3,482	7,254	2,037	4,244
LSTM	2L-7N	12,778	3,575	7,448	2,153	4,485
LSTM	2L-50N	10,425	3,229	6,727	1,831	3,815
LSTM	2L-100N	10,176	3,190	6,646	1,766	3,679
LSTM	3L-7N	10,880	3,299	6,873	1,928	4,017
LSTM	3L-50N	10,414	3,227	6,723	1,811	3,772
LSTM	3L-100N	9,223	3,037	6,327	1,694	3,529
LSTM	4L-7N	12,104	3,479	7,248	2,110	4,395
LSTM	4L-50N	10,085	3,176	6,617	1,803	3,757
LSTM	4L-100N	9,499	3,082	6,421	1,676	3,491

Các mạng LSTM cho thấy ưu thế về độ chính xác trên tập kiểm tra so với các phương pháp truyền thống khác. Mạng LSTM đơn giản nhất với cấu trúc một lớp ẩn và 7 nút ẩn có kết quả khá tương đồng với mạng MLP tốt nhất. Bằng cách tăng độ sâu của mạng LSTM (độ phức tạp), kết quả thu được MAPE tốt nhất (3,491%) từ mạng 4 lớp với 100 nút (LSTM-4L100N).

2.8. Huấn luyện mô hình sử dụng sử dụng với tập kiểm chứng và kỹ thuật dừng sớm

Bảng 2.6. Kết quả huấn luyện tối ưu mô hình

Chỉ số	MAE	MAPE	MSE	RMSE	nRMSE
Đơn vị	MW	%	MW^2	MW	%
Huấn luyện cơ bản	1,676	3,491	9,499	3,082	6,421

Chỉ số	MAE	MAPE	MSE	RMSE	nRMSE
Huấn luyện có kiểm chứng 10% và có dừng sớm	1,412	2,942	4,692	2,166	4,513

Nhận xét: Mô hình thu được từ huấn luyện có kiểm chứng và dừng sớm cho kết quả sai số dự báo có:

- MAPE giảm từ 3,491% xuống còn 2,942%, cải thiện khoảng 16,0% so với huấn luyện cơ bản.

- RMSE giảm từ 3,082% xuống còn 2,166%, cải thiện khoảng 29,7% so với huấn luyện cơ bản.

2.9. So sánh mô hình nhiều đầu vào với mô hình một đầu vào

Bảng 2.8.. So sánh mô hình sử dụng đầu vào đa biến và đơn biến

Đầu vào		MSE	RMSE	nRMSE	MAE	MAPE
		MW ²	MW	%	MW	%
Đa biến		4,692	2,166	4,513	1,412	2,942
Đơn biến	Bức xạ - GHI	57,268	7,568	15,767	4,892	10,192
	Nhiệt độ - TEM	192,344	13,869	28,894	10,695	22,281
	Độ ẩm - HUM	299,621	17,31	36,063	13,491	28,106
	Tốc độ gió - WIS	453,961	21,306	44,388	17,081	35,585

2.10. Dự báo 01 bước tiếp theo

Trong các phần trước, luận án đã nghiên cứu mô hình dự báo bước t sử dụng các đầu vào tại t, t-1, t-2, t-3, t-4. Mô hình thu được sẽ được ứng dụng khi có số liệu khí tượng dự báo để dự báo công suất

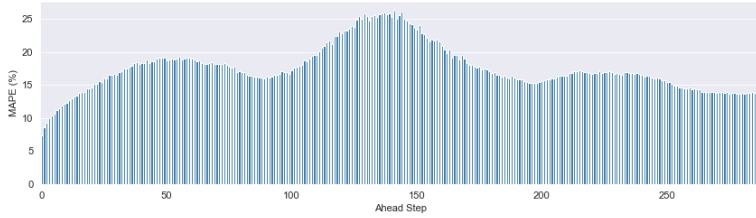
trong các chu kỳ tương lai. Trong phần này, luận án tiến hành thử nghiệm việc dự báo 01 bước tiếp theo chỉ sử dụng dữ liệu quá khứ. Mô hình này áp dụng trong trường hợp khi triển khai không sử dụng dữ liệu khí tượng dự báo.

Bảng 2.9. Kết quả sai số dự báo công suất cho chu kỳ t sử dụng các số liệu khí tượng quá khứ đầu vào (1) $t-4$ đến t và (2) từ $t-4$ đến $t-1$

STT	Đầu vào khí tượng	Đầu ra công suất	MAE	MAPE	MSE	RMSE	nRMSE
			<i>MW</i>	%	<i>MW²</i>	<i>MW</i>	%
1	t-4, t-3, t-2, t-1, t	t	1,412	2,942	4,692	2,166	4,513
2	t-4, t-3, t-2, t-1	t	3,143	6,549	23,664	4,865	10,140

2.11. Dự báo nhiều bước tiếp theo

Nghiên cứu thử nghiệm việc dự báo nhiều bước đầu ra cho 01 ngày tới (24 giờ tới). Do bộ dữ liệu có độ phân giải 05 phút nên số bước đầu ra tiếp theo cần dự báo là 288 bước. Ở mỗi bước huấn luyện, giá trị hàm mất mát MAE được tính toán từ toàn bộ đồng thời 288 giá trị dự báo tiếp theo thay vì chỉ 1 giá trị tiếp theo.



Hình 2.16. Kết quả chỉ số MAPE (%) trên tập dữ liệu kiểm tra của từng bước dự báo trong mô hình dự báo nhiều bước đồng thời (multistep)

2.12. Đánh giá và kết luận

Để có sự đánh giá luận án tiến hành so sánh kết quả mô hình đạt được với một số kết quả từ các công bố trong và ngoài nước liên quan đến dự báo công suất ĐMT sử dụng mạng LSTM từ năm 2019 đến nay đã trình bày trong Mục 1.6 về tổng kết tình hình nghiên cứu.

Bảng 2.11. So sánh kết quả nghiên cứu với các mô hình đã công bố

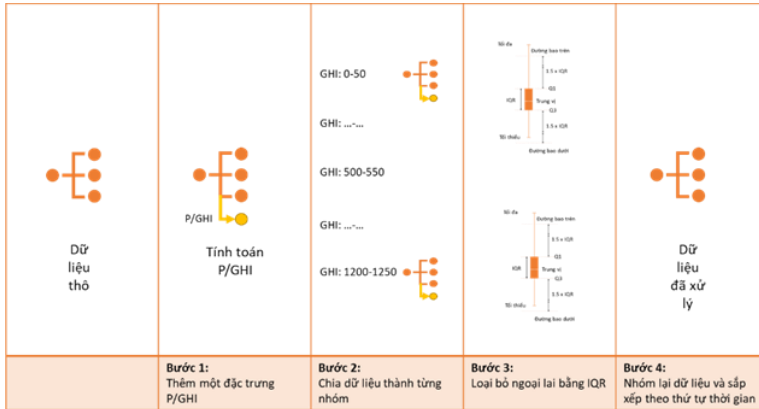
Mô hình	Năm	Quy mô nhà máy	Độ phân giải	Độ dài dự báo	Số bước dự báo	Khu ng dự báo	MAPE (%)	nRMSE (%)
Dự báo cho chu kỳ hiện tại (có dữ liệu khí tượng chu kỳ hiện tại)								
Mô hình Luận án		48 MW	05 phút	t	1	05 phút	2,942	4,513
Liu và cộng sự	2021	18,78 kW	05 phút	t	1	05 phút		4,886
Wang và cộng sự	2019	23,4 kWp	05 phút	t	1	05 phút	2,092	5,879
Li và cộng sự	2021	40 kW	60 phút	t	1	60 phút		6,400
Wen và cộng sự	2019	106,6 kWp	60 phút	t	1	60 phút	7,566	13,051
Suresh và cộng sự	2022	317 kWp	60 phút	t	1	60 phút	4,800	9,046

Mô hình	Năm	Quy mô nhà máy	Độ phân giải	Độ dài dự báo	Số bước dự báo	Khu ng dự báo	MAPE (%)	nRMSE (%)
Park và cộng sự	2021	500 kWp	60 phút	t	1	60 phút		13,200
Dự báo cho 01 chu kỳ tiếp theo (chỉ có dữ liệu khí tượng quá khứ)								
Mô hình Luận án		48 MW	05 phút	t+1	1	05 phút	6,549	10,140
Zhou H và cộng sự	2019	20 kW	7,5 phút	t+1	1	7,5 phút	4,000	6,950
Zhou H và cộng sự	2021	20 kW	7,5 phút	t+1	1	7,5 phút	3,500	6,500
Harrou và cộng sự	2020	9 MWp	15 phút	t+1	1	15 phút	8,930	
Ospina và cộng sự	2019	12,6 MW	30 phút	t+1	1	30 phút	6,563	12,184
Dự báo cho 01 ngày tiếp theo (chỉ có dữ liệu khí tượng quá khứ)								
Mô hình Luận án		48 MW	05 phút	Từ t+1 đến t+288	288	1 ngày	13,902	19,190
Zhou N và cộng sự	2020	5,83 kWp	05 phút	Từ t+1 đến t+288	288	1 ngày	11,639	20,075
Zhang và cộng sự	2020	6,41 kW	15 phút	Từ t+1 đến t+96	96	1 ngày	8,418	15,489
Nguyễn Đức Tuyên và cộng sự	2020	78 MW	60 phút	Từ t+1 đến t+24	24	1 ngày	3,743	5,402

Bảng 2.11 trình bày so sánh kết quả tính toán từ mô hình trong luận án với một số mô hình đã công bố theo Sai số trung bình bình phương gốc chuẩn hóa (nRMSE) và Sai số tuyệt đối phần trăm trung bình (MAPE) theo các khung thời gian dự báo.

CHƯƠNG 3. GIẢI PHÁP CẢI TIẾN HIỆU QUẢ MÔ HÌNH VÀ XÂY DỰNG QUY TRÌNH, CÔNG CỤ DỰ BÁO

3.1. Kỹ thuật tiền xử lý số liệu với hệ số P/GHI kết hợp với phân cụm GHI



Hình 3.2. Mô tả các bước trong kỹ thuật xử lý dữ liệu đề xuất

Để khắc phục khó khăn trong xử lý dữ liệu ngoại lai trong bộ dữ liệu nhà máy ĐMT, nghiên cứu đã công bố năm 2022 đề xuất giải pháp kết hợp phương pháp IQR với hệ số P/GHI và chia khoảng GHI thành các đoạn hợp lý để xử lý dữ liệu đầu vào với các bước như mô tả Hình 3.2.

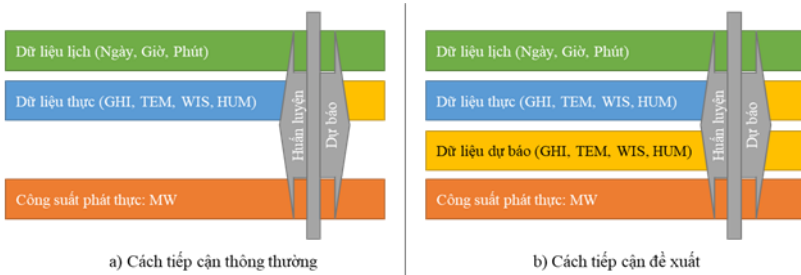
Bảng 3.1. Kết quả sai số dự báo trên tập dữ liệu kiểm tra với bộ dữ liệu huấn luyện trước và sau khi xử lý với kỹ thuật P/GHI kết hợp phân cụm GHI

Chỉ số	MAE	MAPE	MSE	RMSE
Bộ dữ liệu	MW	%	MW ²	MW
Trước xử lý	2,501	6,413	10,354	3,218
Sau xử lý với kỹ thuật P/GHI kết hợp phân cụm GHI	1,602	4,109	4,706	2,169

Bảng 3.1 trình bày kết quả sai số dự báo trên tập dữ liệu kiểm tra với bộ dữ liệu huấn luyện trước và sau khi xử lý với kỹ thuật P/GHI kết hợp phân cụm GHI. Qua đó có thể thấy, áp dụng kỹ thuật xử lý đề xuất giúp giảm MAPE từ 6,413% xuống còn 4,109 % với mức cải thiện 36,2%.

3.2. Kỹ thuật huấn luyện sử dụng số liệu khí tượng dự báo trong quá trình huấn luyện và dự báo

Mặc dù mô hình thu được tại Chương 2 hoạt động tốt với bộ dữ liệu thời tiết thực tế, khi sử dụng bộ dữ liệu dự báo thời tiết, sai số tăng đáng kể với RMSE là 8,065 MW (cao gấp 2,62 lần) và MAPE là 10,857% (cao gấp 3,11 lần). Từ đó, có thể nhận thấy sự gia tăng đáng kể trong sai số khi sử dụng dữ liệu dự báo khí tượng, mà nguyên nhân chủ yếu là do sai số dự báo của dữ liệu khí tượng. Sai số này là không thể tránh khỏi đến từ chất lượng của nhà cung cấp dự báo.



Hình 3.7. Sự khác biệt của cấu trúc dữ liệu đầu vào của mô hình đề xuất

Ý tưởng chính của giải pháp là nếu thuật toán dự báo dữ liệu thời tiết là ổn định, thì sai số dự báo thời tiết cũng có một quy luật ẩn mà mô hình LSTM có thể phát hiện ra. Quá trình huấn luyện dựa chỉ trên dữ liệu vận hành thực tế sẽ không thể học được quy luật này. Tuy nhiên, việc thêm dữ liệu dự báo khí tượng trong quá khứ trong quá trình huấn luyện mô hình, có thể dẫn đến sai số khi kiểm tra dữ liệu

thực tế không tốt như trước, nhưng có thể cải thiện kết quả công suất dự báo đầu ra trong thực tế.

Bảng 3.3. Tổng hợp so sánh sai số của các mô hình với các đầu vào khác nhau

(a) Sai số trên tập dữ liệu kiểm tra

Ký hiệu mô hình	Số đầu vào	MSE	RMSE	MAE	MAPE
		MW ²	MW	MW	%
7I	7 (3 đầu vào thời gian + 4 đầu vào khí tượng quá khứ)	9,499	3,082	1,676	3,491
7I+1F	8 (3 đầu vào thời gian + 4 đầu vào khí tượng quá khứ + GHI dự báo trong quá khứ)	12,482	3,533	1,983	4,132
7I+2F	9 (3 đầu vào thời gian + 4 đầu vào khí tượng quá khứ + GHI, TEM dự báo trong quá khứ)	10,829	3,291	1,909	3,978
7I+3F	10 (3 đầu vào thời gian + 4 đầu vào khí tượng quá khứ + GHI, TEM, HUM dự báo trong quá khứ)	11,490	3,390	1,887	3,932
7I+4F	11 (3 đầu vào thời gian + 4 đầu vào khí tượng quá khứ + GHI, TEM, HUM, WIS dự báo trong quá khứ)	12,093	3,477	1,946	4,054

(b) Sai số trên tập dữ liệu kiểm tra khi ứng dụng dự báo trong thực tế (thay thế đầu vào khí tượng thực tế bằng số liệu khí tượng dự báo)

Ký hiệu mô hình	Số đầu vào	MSE	RMSE	MAE	MAPE
		MW ²	MW	MW	%
7I	7 (3 đầu vào thời gian)	65,041	8,065	5,211	10,857

7I+1F	+ GHI, TEM, HUM, WIS dự báo) 8 (3 đầu vào thời gian + GHI, TEM, HUM, WIS dự báo + GHI dự báo)	67,119	8,193	5,182	10,796
7I+2F	9 (3 đầu vào thời gian + GHI, TEM, HUM, WIS dự báo + GHI, TEM dự báo)	56,348	7,507	4,743	9,881
7I+3F	10 (3 đầu vào thời gian + GHI, TEM, HUM, WIS dự báo + GHI, TEM, HUM dự báo)	61,027	7,812	4,921	10,253
7I+4F	11 (3 đầu vào thời gian + GHI, TEM, HUM, WIS dự báo + GHI, TEM, HUM, WIS dự báo)	66,762	8,171	5,184	10,801

Kết quả kiểm tra sai số phản ánh hiệu quả của mô hình đề xuất. Mô hình 7I lúc này cho sai số tăng lên đáng kể với MAPE là 10,857 % và RMSE là 8,065%. Các mô hình 8 đầu vào (7I+1F), 9 đầu vào (7I+2F), 10 đầu vào (7I+3F), 11 đầu vào (7I+4F), bổ sung thêm lần lượt GHI, TEM, HUM, WIS dự báo trong quá khứ đều ghi nhận có sự cải thiện về MAPE so với mô hình 7I. Trong số đó, sai số của mô hình 7I+2F với MAPE là 9,881% đạt được mức độ cải thiện đáng kể nhất khoảng 0,976% so với mô hình 7I, mức cải thiện là khoảng 9%.

3.3. Kỹ thuật sử dụng dữ liệu bức xạ trời trong thay thế cho các đầu vào chỉ số thời gian

Ý tưởng được sử dụng là thay bộ dữ liệu chỉ dấu thời gian (Ngày trong năm, Giờ trong ngày, phút trong giờ) bằng bức xạ trời trong. Bức xạ trời trong là một trong những đầu vào tiềm năng trong việc cải thiện

chất lượng các mô hình dự báo. Bức xạ này vốn không được tính toán từ các giá trị vận hành của nhà máy mà nó đến từ bản chất vị trí đặt nhà máy và thời gian trong năm xét đến. Bức xạ trời trong có thể tính toán trước cho bất kỳ thời điểm nào trong năm vì có mô hình tính toán sẵn có với chỉ cần số liệu vị trí tọa độ của nhà máy. Giá trị này vừa biến động theo ngày trong năm, vừa thay đổi theo thời gian trong ngày (giờ, phút) nên là một lựa chọn tiềm năng để thay thế cho cả ba giá trị (Ngày trong năm, Giờ trong ngày, phút trong giờ).

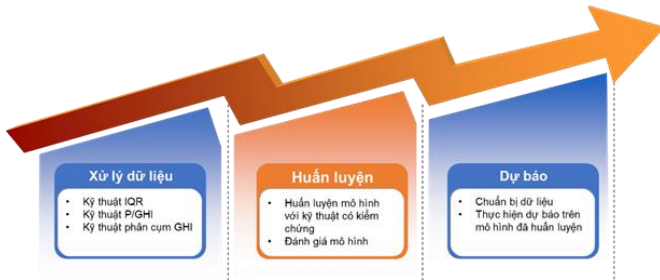
Bảng 3.6. So sánh kết quả dự báo trên tập dữ liệu kiểm tra

Mô hình	Khung dự báo	MAE	MAPE	MSE	RMSE	nRMSE
		<i>MW</i>	%	<i>MW²</i>	<i>MW</i>	%
Mô hình sử dụng chỉ dấu thời gian	Chu kỳ hiện tại	1,412	2,942	4,692	2,166	4,513
	1 chu kỳ tiếp theo	3,143	6,549	23,664	4,865	10,140
	Ngày tới	6,673	13,902	84,835	9,211	19,190
Mô hình sử dụng bức xạ trời trong	Chu kỳ hiện tại	1,309	2,728	4,002	2,001	4,168
	1 chu kỳ tiếp theo	2,106	4,388	12,139	3,484	7,258
	Ngày tới	5,909	12,310	80,966	8,988	18,745

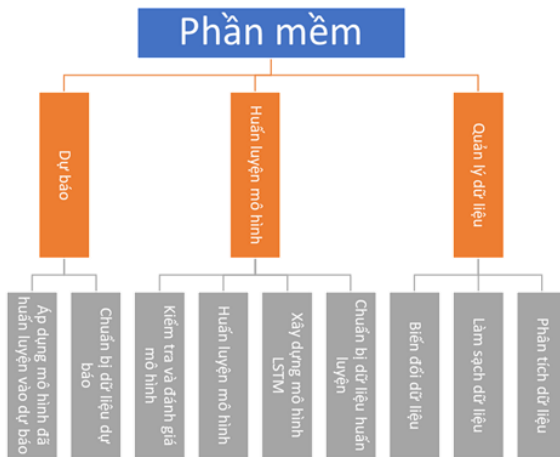
3.4. Quy trình và công cụ xử lý bài toán dự báo ngắn hạn công suất ĐMT sử dụng mạng nơ-ron hồi quy

Tổng hợp từ quá trình nghiên cứu và thực hiện các thí nghiệm của mình, luận án đưa ra một quy trình hoàn chỉnh để xử lý bài toán dự báo công suất ngắn hạn của một nhà máy ĐMT sử dụng mạng Nơ-ron hồi quy LSTM như sau:

- Bước 1: Xử lý dữ liệu đầu vào
- Bước 2: Huấn luyện mô hình
- Bước 3: Dự báo



Hình 3.11. Quy trình xử lý bài toán dự báo công suất nhà máy ĐMT



Hình 3.13. Sơ đồ chức năng chính của công cụ Phần mềm

Sơ đồ chức năng chính của công cụ phần mềm được trình bày trong Hình 3.13, công cụ phần mềm có 3 khối chức năng chính bao gồm: (1) Quản lý dữ liệu (2) Huấn luyện mô hình (3) Dự báo.

KẾT LUẬN VÀ HƯỚNG NGHIÊN CỨU TIẾP THEO

1. Kết luận

Qua quá trình nghiên cứu và thực nghiệm, đề tài “**Nghiên cứu mô hình dự báo ngắn hạn công suất phát của nhà máy điện mặt trời sử dụng mạng nơ-ron hồi quy**” nghiên cứu sinh đã hoàn thành

toàn bộ nội dung nghiên cứu của đề tài và đã đạt được các kết quả quan trọng cụ thể như sau:

- Đã thu thập thông tin và tổng hợp hệ thống hóa lại các kiến thức liên quan đến lĩnh vực dự báo công suất phát của ĐMT trên toàn diện các khía cạnh gồm: dữ liệu đầu vào, khung dự báo, tiêu chí đánh giá chất lượng, các phương pháp dự báo. Qua việc tổng hợp, phân tích, đánh giá các nghiên cứu trong và ngoài nước đã công bố liên quan đến dự báo công suất nhà máy ĐMT sử dụng mạng LSTM trong 5 năm trở lại đây đã củng cố lựa chọn về hướng đi sâu nghiên cứu mô hình dự báo ngắn hạn công suất phát của nhà máy ĐMT sử dụng mạng nơ-ron hồi quy LSTM.

- Đã thực nghiệm thành công xây dựng mô hình dự báo ngắn hạn công suất phát của nhà máy điện mặt trời sử dụng mạng nơ-ron hồi quy với biến thể nâng cao là mạng bộ nhớ dài ngắn (LSTM). Thông qua thực nghiệm trên số liệu của nhà máy ĐMT, luận án đã chứng minh hiệu quả của mô hình LSTM 4 lớp với 100 nút ẩn trên mỗi lớp cho các chỉ số sai số MAPE đạt 3,491%, tốt hơn đáng kể so với kết quả thực nghiệm từ các phương pháp so sánh phổ biến khác (MLP 4,346%, ARIMA 7,329%, Quán tính 15,418%). Mô hình của luận án sau khi được áp dụng huấn luyện tối ưu (với bộ dữ liệu kiểm chứng 10% và kỹ thuật dừng sớm) giúp giảm sai số MAPE còn 2,942%, sai số nRMSE còn 4,513% cho khung chu kỳ hiện tại, đạt kết quả tương đương với các mô hình tốt nhất được công bố trong 05 năm trở lại đây. Đối với khung dự báo 01 chu kỳ tiếp theo, sai số MAPE đạt 6,549% và nRMSE đạt 10,140%, tương đương với một số mô hình đã công bố với các điều kiện độ phân giải dữ liệu và quy mô nhà máy gần giống với dữ liệu của Luận án. Đối với khung dự báo ngày tới, sai số đạt được của mô hình luận án còn khá cao, tuy nhiên khi so sánh với các mô hình đã công bố khác, sự khác biệt là không lớn và nguyên nhân

chính dẫn đến sự chênh lệch chất lượng đến từ sự khác nhau về quy mô của đối tượng nhà máy ĐMT của mỗi nghiên cứu.

- Đã tiến hành nghiên cứu đề xuất các giải pháp cải tiến và thực nghiệm chứng minh hiệu quả trong việc xây dựng mô hình dự báo bao gồm: (1) Kỹ thuật tiền xử lý số liệu với hệ số P/GHI kết hợp với phân cụm GHI giúp giảm sai số MAPE từ 6,413% xuống còn 4,109% với mức cải thiện khoảng 36,2% (2) Kỹ thuật huấn luyện sử dụng số liệu khí tượng dự báo giúp giảm sai số MAPE trên bộ dữ liệu dự báo từ 10,857% còn 9,881%, mức cải thiện khoảng 9,0% (3) Kỹ thuật sử dụng dữ liệu bức xạ trời trong thay thế các chỉ dấu thời gian giúp giảm MAPE từ 0,214% đến 2,161% và nRMSE từ 0,165% đến 1,381% cho các khung dự báo.

- Đã xây dựng được quy trình hoàn chỉnh và công cụ hỗ trợ với giao diện thân thiện, thao tác thuận tiện xử lý bài toán dự báo ngắn hạn công suất nhà máy ĐMT sử dụng mạng nơ-ron hồi quy.

- Đã thực hiện công bố chính thức các kết quả nghiên cứu liên quan đến đề tài trên các tạp chí quốc tế (02 bài báo SCI – Q1, 02 bài báo Scopus – Q3, Q4).

2. Hướng nghiên cứu tiếp theo

- Hệ thống huấn luyện online và huấn luyện thời gian thực (realtime)
- Các kỹ thuật nâng cao chất lượng mô hình
- Hoàn thiện quy trình và công cụ
- Nghiên cứu mô hình ứng dụng ảnh bầu trời để dự báo sụt giảm ngắn hạn

DANH MỤC CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ

Tạp chí quốc tế: 04 công bố

- [1] **L. D. Bui**, N. Q. Nguyen, B. Van Doan, and E. R. Sanseverino, “Forecasting energy output of a solar power plant in curtailment condition based on LSTM using P/GHI coefficient and validation in training process, a case study in Vietnam,” *Electric Power Systems Research*, vol. 213, p. 108706, Dec. 2022, doi: 10.1016/J.EPSR.2022.108706 (**IF = 3.818, 2022, SCI - Q1**).
- [2] N. Q. Nguyen, **L. Duy Bui**, B. Van Doan, E. R. Sanseverino, D. Di Cara, and Q. D. Nguyen, “A new method for forecasting energy output of a large-scale solar power plant based on long short-term memory networks, a case study in Vietnam,” *Electric Power Systems Research*, vol. 199, p. 107427, 2021, doi: 10.1016/j.epsr.2021.107427 (**IF = 3.818, 2022, SCI - Q1**).
- [3] **D. L. Bui**, Q. N. Nguyen, V. B. Doan, T. K. Pham, and D. D. Le, “Evaluating an Effectiveness of a Solar Power Plant Output Forecasting Model Based on LSTM Method Using Validation in Different Seasons of a Year in Vietnam,” *GMSARN International Journal*, vol. 18, pp. 114–122, 2024 (**Scopus - Q4**).
- [4] N. Quang, **L. Duy**, **B. Van**, and Q. Dinh, “Applying Artificial Intelligence in Forecasting the Output of Industrial Solar Power Plant in Vietnam,” *EAI Endorsed Transactions on Energy Web*, p. 169166, Jul. 2021, doi: 10.4108/eai.29-3-2021.169166 (**Scopus – Q3**).

Sáng chế: 01

Nghiên cứu sinh đã tham gia nghiên cứu thực hiện sáng chế
**“PHƯƠNG PHÁP HUẤN LUYỆN MÔ HÌNH LSTM DỰ
BÁO CÔNG SUẤT PHÁT HỆ THỐNG ĐIỆN MẶT TRỜI”.**

Hiện sáng chế đã được nộp đơn lên Cục sở hữu trí tuệ và đã được quyết
định về việc chấp nhận đơn hợp lệ (*văn bản số 14874w/QĐ-SHTT*
ngày 31/03/2023 của Cục trưởng Cục Sở hữu trí tuệ).