

**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**

LUẬN ÁN TIẾN SỸ KỸ THUẬT NĂNG LƯỢNG

HÀ NỘI – 2024

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

LUẬN ÁN TIẾN SĨ KỸ THUẬT NĂNG LƯỢNG

Mã số : Thí điểm

Xác nhận của Học viện
Khoa học và Công nghệ

Người hướng dẫn 1

Người hướng dẫn 2

TS. Nguyễn Quang Ninh

TS. Đoàn Văn Bình

HÀ NỘI – 2024

LỜI CAM ĐOAN

Tôi xin cam đoan đề tài nghiên cứu trong luận án này là công trình nghiên cứu của tôi dựa trên những tài liệu, số liệu do chính tôi tự tìm hiểu và nghiên cứu. Chính vì vậy, các kết quả nghiên cứu đảm bảo trung thực và khách quan nhất. Trong quá trình nghiên cứu, tác giả công bố một số kết quả trên các tạp chí khoa học của ngành và của lĩnh vực năng lượng. Kết quả nghiên cứu của luận án này chưa từng được công bố trong bất kỳ công trình nghiên cứu nào khác ngoài các công trình nghiên cứu của tác giả.

Tác giả luận án

Bùi Duy Linh

LỜI CẢM ƠN

Tác giả xin bày tỏ lời cảm ơn và lòng kính trọng đối với các thầy hướng dẫn: Tiến sĩ Đoàn Văn Bình, Tiến sĩ Nguyễn Quang Ninh - Viện Khoa học công nghệ Năng lượng và Môi trường, Viện Hàn lâm Khoa học và Công nghệ Việt Nam bởi những chỉ dẫn quý báu về định hướng nghiên cứu và phương pháp luận để luận án này được hoàn thành.

Tác giả cũng xin chân thành cảm ơn Ban lãnh đạo Học viện Khoa học và Công nghệ, Viện Khoa học công nghệ Năng lượng và Môi trường - Viện Hàn lâm Khoa học và Công nghệ Việt Nam đã tạo điều kiện về cơ sở vật chất và thời gian để tác giả hoàn thành luận án.

Tác giả xin trân trọng cảm ơn các nhà khoa học, các đồng nghiệp đã phản biện, đóng góp các ý kiến quý báu để xây dựng và trao đổi các vấn đề lý thuyết cũng như thực tiễn giúp hoàn thiện luận án.

Đặc biệt, tác giả xin được bày tỏ lời cảm ơn sâu sắc nhất đến gia đình đã luôn chia sẻ những khó khăn, luôn hỗ trợ, động viên tinh thần giúp tác giả có được chỗ dựa vững chắc để hoàn thành luận án này.

MỤC LỤC

MỤC LỤC	iii
DANH MỤC CÁC TỪ VIẾT TẮT	vi
DANH MỤC BẢNG	vii
DANH MỤC HÌNH.....	ix
MỞ ĐẦU	1
1. Lý do chọn đề tài.....	1
2. Mục tiêu nghiên cứu	5
3. Phương pháp nghiên cứu	5
4. Phạm vi nghiên cứu.....	6
5. Cơ sở khoa học và thực tiễn của đề tài	6
6. Những đóng góp mới của luận án.....	6
7. Cấu trúc luận án	7
CHƯƠNG 1. TỔNG QUAN NGHIÊN CỨU	8
1.1. Mở đầu	8
1.2. Công suất phát đầu ra của nhà máy ĐMT và các yếu tố ảnh hưởng	8
1.2.1. Công suất phát đầu ra của nhà máy ĐMT	8
1.2.2. Các yếu tố ảnh hưởng đến công suất phát nhà máy ĐMT.....	9
1.3. Phân loại các khung dự báo công suất ĐMT	16
1.4. Đánh giá chất lượng mô hình dự báo.....	17
1.4.1. Đánh giá sai số thông qua các chỉ số	17
1.4.2. Đánh giá sai số thông qua biểu đồ phân bố sai số	18
1.5. Các phương pháp dự báo ngắn hạn công suất nhà máy ĐMT.....	19
1.5.1. Phương pháp Vật lý	21
1.5.2. Phương pháp Thống kê theo chuỗi thời gian.....	22
1.5.3. Phương pháp Học máy.....	24
1.5.4. Phương pháp Kết hợp	29
1.6. Tình hình nghiên cứu	30

1.7. Kết luận	35
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	37
2.1. Mở đầu	37
2.2. Thu thập số liệu.....	37
2.2.1. Số liệu khí tượng thực tế.....	37
2.2.2. Số liệu công suất phát	38
2.2.3. Bộ dữ liệu phục vụ xây dựng mô hình	38
2.3. Môi trường thí nghiệm.....	39
2.4. Tiền xử lý số liệu	40
2.4.1. Trích xuất các đặc trưng quan trọng từ bộ dữ liệu.....	41
2.4.2. Làm sạch dữ liệu.....	42
2.4.3. Phân chia dữ liệu huấn luyện và kiểm tra.....	45
2.5. Xây dựng mô hình LTSM và các mô hình so sánh	45
2.5.1. Mô hình quán tính.....	45
2.5.2. Mô hình ARIMA	45
2.5.3. Mô hình MLP.....	47
2.5.4. Xây dựng mô hình LSTM.....	47
2.6. Huấn luyện mô hình LSTM	49
2.7. So sánh kết quả dự báo từ các mô hình	50
2.8. Huấn luyện mô hình sử dụng tập kiểm chứng và kỹ thuật dừng sớm	52
2.9. So sánh mô hình nhiều đầu vào với mô hình một đầu vào	53
2.10. Dự báo 01 bước tới.....	54
2.11. Dự báo nhiều bước tới	55
2.12. Đánh giá và kết luận	58
CHƯƠNG 3. GIẢI PHÁP CẢI TIẾN HIỆU QUẢ MÔ HÌNH VÀ XÂY DỰNG QUY TRÌNH, CÔNG CỤ DỰ BÁO.....	61
3.1. Kỹ thuật tiền xử lý số liệu với hệ số P/GHI kết hợp với phân cụm GHI.....	61
3.1.1. Đặt vấn đề	61
3.1.2. Phương án đề xuất	62

3.1.3.	Thí nghiệm chứng minh hiệu quả.....	64
3.2.	Kỹ thuật huấn luyện sử dụng số liệu khí tượng dự báo trong quá trình huấn luyện và dự báo.....	67
3.2.1.	Đặt vấn đề.....	67
3.2.2.	Phương án đề xuất.....	68
3.2.3.	Thí nghiệm chứng minh hiệu quả.....	68
3.3.	Kỹ thuật sử dụng dữ liệu bức xạ trời trong thay thế cho các đầu vào chỉ số thời gian 71	
3.3.1.	Đặt vấn đề.....	71
3.3.2.	Phương án đề xuất.....	72
3.3.3.	Thí nghiệm kiểm chứng hiệu quả.....	73
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.....	76
3.4.1.	Bước 1 - Xử lý dữ liệu đầu vào.....	77
3.4.2.	Bước 2 - Huấn luyện mô hình.....	79
3.4.3.	Bước 3 – Dự báo.....	83
3.4.4.	Công cụ phần mềm dự báo công suất ĐMT.....	84
	KẾT LUẬN VÀ HƯỚNG NGHIÊN CỨU TIẾP THEO.....	87
1.	KẾT LUẬN.....	87
2.	HƯỚNG NGHIÊN CỨU TIẾP THEO.....	88
	DANH MỤC CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ.....	90
	TÀI LIỆU THAM KHẢO.....	91
	PHỤ LỤC 1: Giao diện công cụ dự báo.....	101
	PHỤ LỤC 2: Kết quả chi tiết các chỉ số sai số dự báo trên tập dữ liệu kiểm tra của từng bước trong mô hình dự báo nhiều bước đồng thời.....	103

DANH MỤC CÁC TỪ VIẾT TẮT

Tiếng Việt

STT	Từ viết tắt	Giải thích
1	ĐMT	Điện mặt trời
2	NLTT	Năng lượng tái tạo

Tiếng Anh

STT	Từ viết tắt	Giải thích
1	AC	Alternating Current: Điện xoay chiều
2	ADF	Augmented Dickey-Fuller: Kiểm định Dickey-Fuller tăng cường
3	AIC	Akaike Information Criteria: Tiêu chuẩn kiểm định thông tin Akaike
4	ANN	Artificial Neural Network: Mạng nơ-ron nhân tạo
5	APE	Absolute Percentage Error: Sai số tuyệt đối phần trăm
6	ARIMA	Autoregressive Integrated Moving Average: Tự hồi quy tích hợp trung bình trượt
7	BPTT	Back Propagation Through Time: Thuật toán lan truyền ngược qua thời gian
8	DC	Direct Current: Điện một chiều
9	GHI	Global Horizontal Irradiance: Bức xạ toàn phần theo phương ngang
10	IQR	Interquartile Range: Khoảng tứ phân vị
11	LSTM	Long Short Term Memory: Mạng Nơ-ron hồi quy Bộ Nhớ Dài – Ngắn
12	MAE	Mean Absolute Error: Sai số tuyệt đối trung bình
13	MLP	Multilayer Perceptron: Mạng nơ-ron nhiều lớp
14	MAPE	Mean Absolute Percentage Error: Sai số tuyệt đối phần trăm trung bình
15	MPPT	Maximum Power Point Tracker: Bộ theo dõi điểm công suất cực đại
16	MSE	Mean Square Error: Sai số trung bình bình phương
17	nRMSE	normalize Root Mean Square Error: Sai số trung bình bình phương gốc chuẩn hóa
18	NWP	Numerical Weather Prediction: Mô hình dự báo thời tiết số
19	RMSE	Root Mean Square Error: Sai số trung bình bình phương gốc
20	RNN	Recurrent Neural Networks: Mạng nơ-ron hồi quy
21	STC	Standard Test Condition: Điều kiện thí nghiệm tiêu chuẩn

DANH MỤC BẢNG

Bảng 1.1. Ví dụ về hệ số ảnh hưởng của nhiệt độ đến thông số vận hành tấm pin.....	13
Bảng 1.2. Phân loại các khung dự báo.....	16
Bảng 1.3. Tổng hợp phân loại các phương pháp dự báo công suất ĐMT	20
Bảng 1.4: Tổng hợp một số các mô hình dự báo ĐMT sử dụng LSTM đã công bố từ 2019-2023	35
Bảng 2.1. Dữ liệu lịch sử của nhà máy	38
Bảng 2.2. Kết quả kiểm định hệ số tương quan Pearson	42
Bảng 2.3. Kết quả kiểm định ADF	46
Bảng 2.4. Kết quả kiểm định AIC	47
Bảng 2.5. Hệ số tương quan Pearson của các thông số khí tượng và công suất tại các bước trễ khác nhau.....	48
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.....	51
Bảng 2.7. Kết quả huấn luyện tối ưu mô hình	52
Bảng 2.8. So sánh mô hình sử dụng đầu vào đa biến và đơn biến	53
Bảng 2.9. Kết quả sai số của 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	55
Bảng 2.10. Bảng dữ liệu rút gọn kết quả chi tiết các chỉ số sai số dự báo trên tập dữ liệu kiểm tra của các bước trong mô hình dự báo nhiều bước đồng thời.....	56
Bảng 2.11. So sánh các mô hình.....	58
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	67
Bảng 3.2. Kết quả dự báo của mô hình LSTM-4L100N trên tập dữ liệu kiểm tra.....	67
Bảng 3.3. Tổng hợp so sánh sai số của các mô hình với các đầu vào khác nhau.....	69
Bảng 3.4. Các nhân thời gian thường được dùng trong dự báo công suất ĐMT	72
Bảng 3.5. So sánh cấu hình thiết lập thực nghiệm với các mô hình.....	75
Bảng 3.6. So sánh kết quả dự báo trên tập dữ liệu kiểm tra	75

DANH MỤC HÌNH

Hình MĐ.1.1. Công suất lắp đặt ĐMT tại Việt Nam	1
Hình MĐ.1.2. Cơ cấu nguồn ngày vận hành thông thường và ngày nghỉ lễ	2
Hình MĐ.1.3. Công suất phát của một nhà máy ĐMT biến động trong ngày	3
Hình MĐ.1.4. Mức độ biến động của ĐMT trong vận hành thực tế	4
Hình 1.1. Đường cong I-V và đặc tính công suất của tấm pin quang điện.....	8
Hình 1.2. Các thành phần bức xạ mà tấm pin mặt trời nhận được	10
Hình 1.3: Đặc tính I-V của tấm pin MSP290AS-36.EU	11
Hình 1.4. Quan hệ giữa GHI và công suất phát đầu ra của một nhà máy ĐMT	12
Hình 1.5. Ảnh hưởng của nhiệt độ lên điện áp và dòng điện của tấm pin.....	13
Hình 1.6. Nguyên lý thay đổi góc nghiêng theo mùa để đạt sản lượng tối đa	14
Hình 1.7. Giới hạn công suất do nghẽn mạch truyền tải	15
Hình 1.8. Sự thay đổi của sai số dự báo theo khung thời gian dự báo	17
Hình 1.9. Biểu đồ phân bố sai số phần trăm của hai phương pháp dự báo	19
Hình 1.10. Các nguồn dữ liệu khí tượng	21
Hình 1.11. Mô phỏng cấu trúc của một mạng nơ-ron nhân tạo.....	25
Hình 1.12. Các khối trong chuỗi khối của mạng nơ-ron hồi quy	27
Hình 1.13. Cấu trúc của một chuỗi khối LSTM	28
Hình 1.14: Tỷ lệ sử dụng các phương pháp trong các nghiên cứu về dự báo công suất ĐMT	31
Hình 2.1. Dữ liệu vận hành quá khứ của nhà máy	39
Hình 2.2. Ngày có công suất phát đầu ra ổn định.....	41
Hình 2.3. Ngày có công suất phát đầu ra biến động	41
Hình 2.4. Phương pháp IQR xác định điểm ngoại lai	43
Hình 2.5. Phân phối của công suất đầu ra trước khi loại bỏ các điểm ngoại lai.....	44
Hình 2.6. Phân phối của công suất đầu ra sau khi loại bỏ các điểm ngoại lai.....	44
Hình 2.7. Quá trình xử lý dữ liệu công suất phát trong khoảng thời gian 10 ngày	44
Hình 2.8. Phân chia dữ liệu Huấn luyện - Kiểm tra	45
Hình 2.9. Đồ thị ACF trên tập dữ liệu công suất phát đầu ra	46
Hình 2.10. Sơ đồ thí nghiệm xây dựng mô hình dự báo.....	49
Hình 2.11. So sánh thời gian huấn luyện của các mô hình.....	50
Hình 2.12. Giá trị hàm mất mát (MAE) trong quá trình huấn luyện	50
Hình 2.13. Phân bố sai số tương đối trên tập dữ liệu kiểm tra của các mô hình	52
Hình 2.14. So sánh sai số mô hình sử dụng đầu vào đa biến và mô hình sử dụng đầu vào đơn biến	54

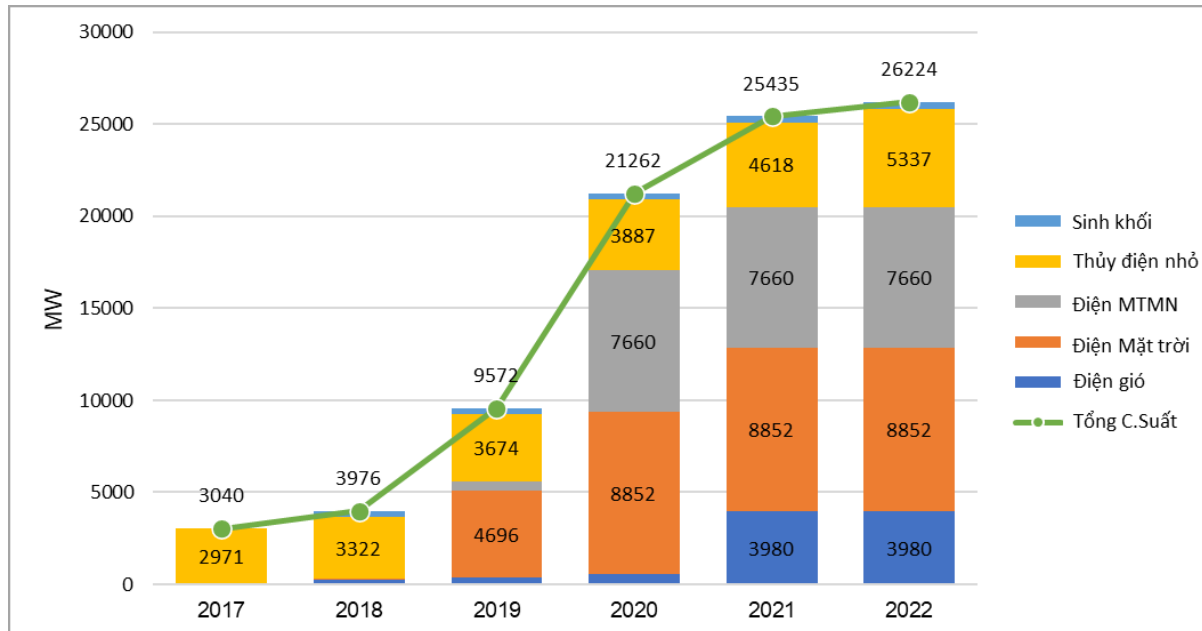
Hình 2.15. Phân bố sai số của dự báo công suất cho chu kỳ t sử dụng các số liệu khí tượng quá khứ đầu vào (1) $t-4$ đến t và (2) từ $t-4$ đến $t-1$	55
Hình 2.16. Kết quả chỉ số MAPE (%) trên tập dữ liệu kiểm tra của từng bước dự báo trong mô hình dự báo nhiều bước đồng thời (multistep).....	56
Hình 2.17. Kết quả chỉ số RMSE (MW) trên tập dữ liệu kiểm tra của từng bước dự báo trong mô hình dự báo nhiều bước đồng thời (multistep).....	56
Hình 2.18. Hệ số tương quan Pearson của các yếu tố khí tượng đến công suất đầu ra theo các bước trễ khác nhau	57
Hình 3.1. Dữ liệu trung bình giờ theo từng tháng của nhà máy ĐMT	61
Hình 3.2. Các bước trong kỹ thuật xử lý dữ liệu ĐMT sử dụng hệ số P/GHI.....	64
Hình 3.3. Dữ liệu vận hành nhà máy ĐMT	64
Hình 3.4. Dữ liệu quá khứ trong 1 tuần của nhà máy ĐMT	65
Hình 3.5. Biểu diễn dữ liệu trước và sau xử lý.....	66
Hình 3.6. Phân phối sai số của mô hình LSTM-4L100N	67
Hình 3.7. Mô tả sự khác biệt của cấu trúc dữ liệu đầu vào của mô hình đề xuất.....	69
Hình 3.8. Phân bố sai số tương đối với mô hình 07 và 09 đặc trưng	71
Hình 3.9. So sánh phân bố sai số tương đối trên bộ dữ liệu khí tượng dự báo của theo ngày trong 5/2020	71
Hình 3.12. Bức xạ GHI thực đo và Bức xạ GHI trời trong tính toán (W/m^2).....	74
Hình 3.16. Quy trình xử lý bài toán dự báo công suất nhà máy ĐMT	77
Hình 3.17. Sơ đồ quá trình huấn luyện mô hình.....	80
Hình 3.18. Sơ đồ chức năng chính của Phần mềm	84

MỞ ĐẦU

1. Lý do chọn đề tài

Sử dụng điện của Việt Nam đã tăng rất nhanh trong những năm gần đây, từ 99 tỷ kWh năm 2010 tăng lên 271 tỷ kWh năm 2022, đạt tốc độ tăng bình quân xấp xỉ gần 10%/ năm. Cùng thời gian đó, công suất đặt của hệ thống điện đã tăng từ 21,5 GW năm 2010 lên khoảng 80 GW vào cuối năm 2022 [1]. Dự kiến sản lượng điện của Việt Nam tiếp tục tăng cao, đến năm 2030 là từ 567 tỷ kWh và 2050 là 1.224 -1.378 tỷ kWh. Cùng với đó công suất lắp đặt năm dự kiến đạt khoảng 150 GW vào 2030 và khoảng 490-573 GW vào năm 2050. Trong đó, các nguồn năng lượng tái tạo (NLTT) trong tổng cung năng lượng đạt 30,9%-39,2% vào năm 2030 và 67,5-71,5% vào năm 2050. Riêng đối với điện mặt trời (ĐMT), công suất lắp đặt toàn hệ thống sẽ đạt dự kiến khoảng 20 GW vào năm 2030 và 168-189GW vào năm 2050 [2].

Mặc dù công nghệ ĐMT đã có từ rất lâu trên thế giới nhưng tại Việt Nam do những chính sách khuyến khích phát triển của nhà nước mà ĐMT mới chỉ thực sự bùng nổ từ năm 2019. Theo số liệu từ [3], đến cuối năm 2020, công suất lắp đặt của ĐMT quy mô trang trại là 8.852 MW và của ĐMT mái nhà là 7.660 MW như Hình MĐ.1.1.



Hình MĐ.1.1. Công suất lắp đặt ĐMT tại Việt Nam

(Nguồn: “Vietnam, a System in Rapid Transition” - ESIG Spring Technical Workshop [3])

Cùng với sự phát triển mạnh trong thời gian gần đây của loại hình công nghệ NLTT như ĐMT, điện gió, tỷ trọng của loại hình này trong vận hành thời gian thực cũng đang ngày một tăng. ĐMT với tính chất là một nguồn năng lượng tái tạo đã mang đến nhiều lợi ích to lớn như giảm thiểu phát thải CO₂, đóng góp sản lượng đáng kể cho hệ

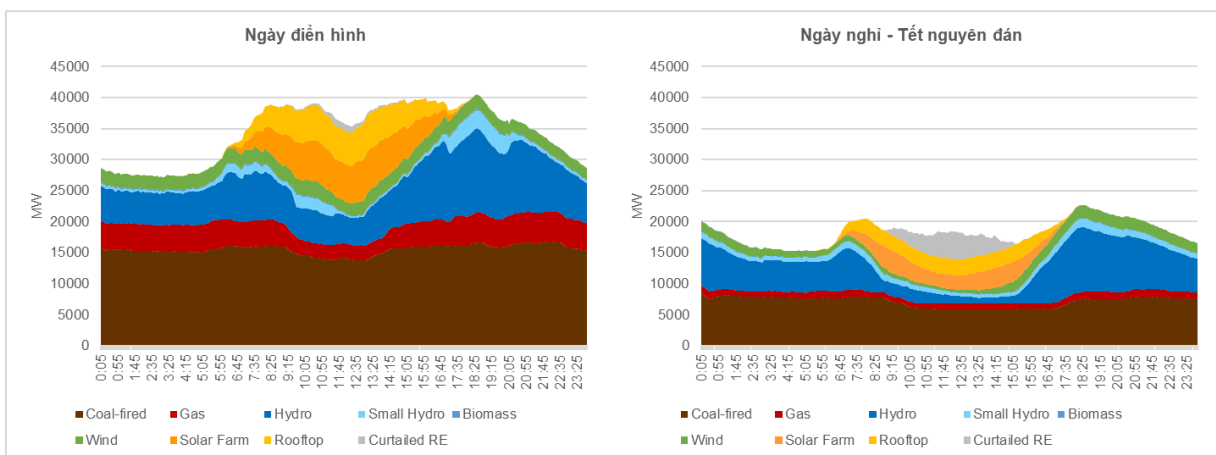
thống trong bối cảnh nguồn cung năng lượng từ nhiên liệu hóa thạch đang ngày càng cạn kiệt, hỗ trợ cân bằng công suất tại một số khu vực trước đây chỉ thuần nhận điện từ lưới điện, hỗ trợ công suất cho hệ thống trong những khung giờ cao điểm tiêu thụ điện.

Tuy nhiên, các nhà máy ĐMT có đặc tính vận hành rất bất định, công suất phát không ổn định và phụ thuộc lớn vào các yếu tố thời tiết, đặc biệt là các thông số như bức xạ, nhiệt độ... và các thông số này có thể thay đổi mạnh trong thời gian rất ngắn. Việc thay đổi đột ngột công suất phát của các nhà máy ĐMT khi loại hình này đang dần chiếm tỷ trọng lớn trong cơ cấu nguồn phát nếu không được dự báo trước để kiểm soát và điều độ, sẽ là tác nhân quan trọng gây lên mất ổn định hệ thống điện, có khả năng làm rã lưới, gây mất điện trên diện rộng, mất an ninh cung cấp điện và thiệt hại nghiêm trọng về kinh tế và xã hội. 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 [4], 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.

Sự gia tăng tỷ trọng nguồn ĐMT trong hệ thống điện Việt Nam đã mang đến nhiều thách thức mới đặc biệt là áp lực đảm bảo vận hành an toàn, ổn định hệ thống trong thời gian thực, do sự biến động mạnh mẽ của công suất ĐMT phụ thuộc vào yếu tố thời tiết.

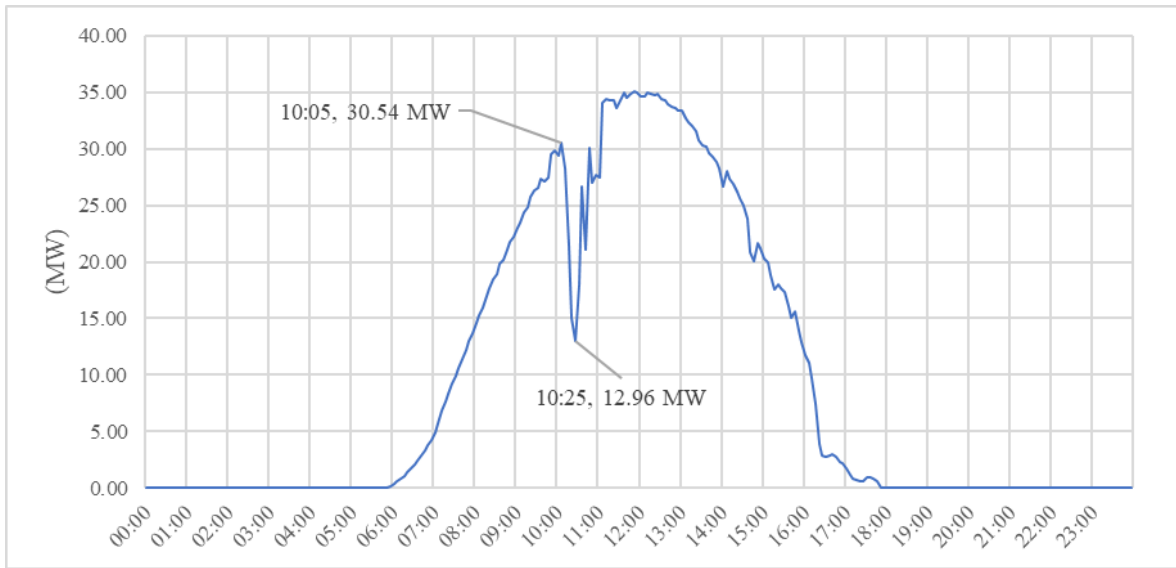


Hình 1.2. Cơ cấu nguồn ngày vận hành thông thường và ngày nghỉ lễ

(Nguồn: “Vietnam, a System in Rapid Transition” - ESIG Spring Technical Workshop [3])

Hiện nay tỷ trọng nguồn 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% như Hình MĐ.1.2 [3]. Tỷ trọng ngày càng tăng của ĐMT đã ảnh hưởng không nhỏ đến khả năng vận hành an toàn ổn định của hệ thống điện.

Ở quy mô từng nhà máy, mức độ thay đổi công suất cũng ở mức rất lớn. Ví dụ như một nhà máy ĐMT có công suất 35 MW thì chỉ trong khoảng thời gian rất ngắn 20 phút từ 10h05 đến 10h25, công suất phát của nhà máy đã thay đổi giảm khoảng 17,58 MW (ứng với 50% công suất đặt của nhà máy) như Hình MĐ.1.3.



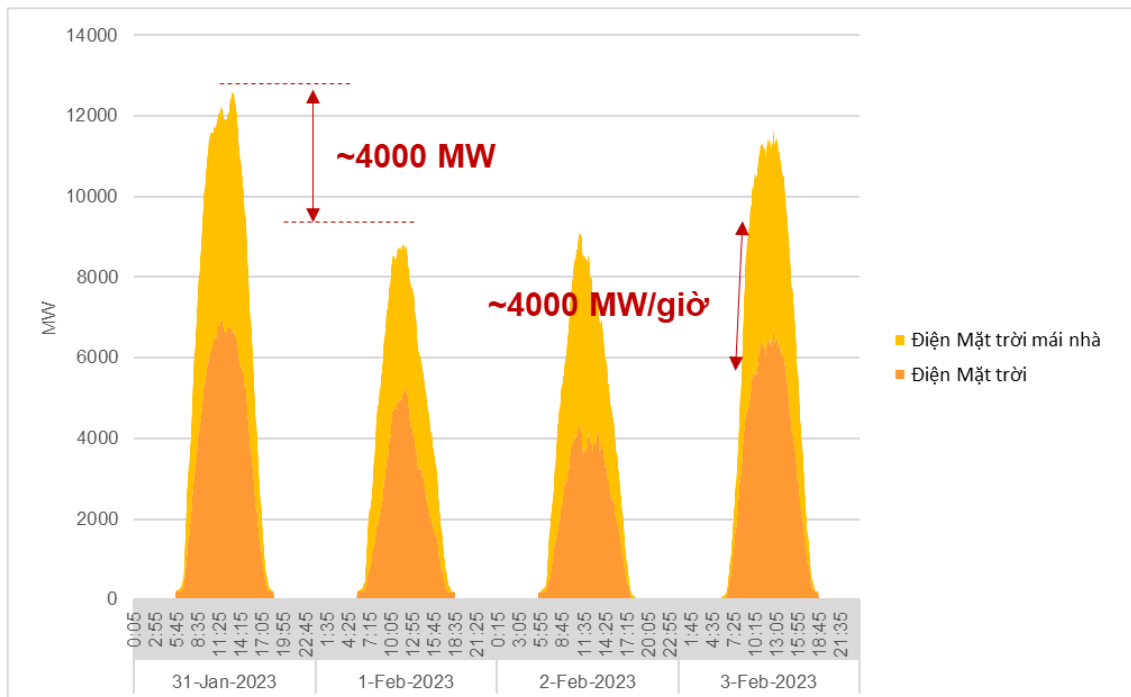
Hình MĐ.1.3. Công suất phát của một nhà máy ĐMT biến động trong ngày

Trên quy mô toàn hệ thống, mức độ thay đổi công suất của ĐMT theo thời gian trong ngày cũng ở mức rất lớn, có thể lên đến khoảng 4000 MW trong 01 giờ. Giữa các ngày liền kề nhau, mức độ thay đổi công suất có thể phát cũng lên đến khoảng 4000 MW như có thể thấy trên Hình MĐ.1.4.

Do đó, để đảm bảo an toàn cho công tác vận hành thì dự báo với độ chính xác cao các nguồn này là một yêu cầu hết sức cấp thiết. Từ đó có thể thấy việc kiểm soát và dự báo trước công suất phát trong ngắn hạn của mỗi nhà máy ĐMT có ý nghĩa rất lớn đối với vận hành cả hệ thống điện.

Dự báo công suất ĐMT là quá trình dự đoán giá trị công suất phát trong tương lai dựa trên các yếu tố như điều kiện thời tiết, vị trí địa lý và dữ liệu lịch sử. Dự báo công suất ĐMT ngắn hạn có vai trò quan trọng trong quản lý và điều khiển hoạt động của nhà máy ĐMT. Ngoài ra nó còn cung cấp thông tin dự kiến về công suất, giúp các đơn vị điều độ, quản lý vận hành lưới điện lập kế hoạch vận hành cho hệ thống điện. Những ý nghĩa có thể kể đến bao gồm:

- *Tối ưu hóa vận hành nguồn điện và lưới điện:* Dự báo công suất của nhà máy điện mặt trời giúp đơn vị vận hành nắm rõ về lượng phát điện mặt trời, giúp điều độ nguồn điện hợp lý, kinh tế từ các nguồn khác nhau để đảm bảo cân bằng nguồn-tải và duy trì ổn định cho toàn bộ hệ thống. Ngoài ra việc dự báo chính xác còn giúp các cơ quan điều độ truyền tải và phân phối đưa ra các quyết định về đóng cắt, vận hành lưới điện một cách hợp lý, an toàn và tối ưu cho hệ thống.
- *Tối ưu hóa lưu trữ năng lượng:* Dự báo chính xác về công suất phát từ nhà máy ĐMT giúp đưa ra quyết định về việc lưu trữ năng lượng dư thừa (sử dụng pin lưu trữ hoặc hệ thống lưu trữ năng lượng khác...) một cách hợp lý để phát vào các khung thời gian hệ thống cần công suất như khi có các sụt giảm, biến động lớn mà các nguồn khác không đủ hoặc chưa kịp đáp ứng.



Hình MD.1.4. Mức độ biến động của ĐMT trong vận hành thực tế

(Nguồn: “Vietnam, a System in Rapid Transition” - ESIG Spring Technical Workshop [3])

Qua những phân tích trên có thể thấy được rằng vai trò và nhu cầu dự báo ngắn hạn công suất phát của các nhà máy ĐMT đang là một đòi hỏi cấp thiết đặt ra cho không chỉ ở quy mô hệ thống mà còn ở từng nhà máy ĐMT quy mô công nghiệp ở Việt Nam.

Việc nghiên cứu về các phương pháp dự báo công suất phát của điện mặt trời đã và đang là một chủ đề được giới nghiên cứu quan tâm. Bài toán dự báo ngắn hạn công suất phát nhà máy ĐMT đã bắt đầu được nghiên cứu từ khoảng một thập kỷ trở lại đây và các phương pháp dự báo công suất phát của ĐMT đã trải qua một quá trình liên tục của sự tiến bộ và cải tiến. Bắt đầu từ các mô hình dự báo ứng dụng theo các lý thuyết dự báo chuỗi thời gian của thống kê cổ điển đến nay các kỹ thuật dự báo đã có sự phát

triển mạnh mẽ với nhiều hướng nghiên cứu mới đặc biệt là ứng dụng mạng nơ-ron nhân tạo trong giải quyết bài toán. Với sự tiến bộ của công nghệ trong thời gian qua, nhiều mô hình học máy, học sâu hiện đại đã ra đời và đã chứng minh hiệu quả trong giải quyết các vấn đề thực tiễn như xử lý ngôn ngữ tự nhiên, xử lý hình ảnh, hệ thống khuyến nghị... Việc nghiên cứu các công nghệ này 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 bao gồm không chỉ công tác nghiên cứu khai phá dữ liệu quá khứ để lựa chọn các đặc trưng 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.
- Giải pháp cải tiến hiệu quả xây dựng mô hình dự báo
- Quy trình và công cụ phần mềm xử lý bài toán dự báo ngắn hạn công suất nhà máy ĐMT sử dụng mạng nơ-ron hồi quy bộ nhớ dài ngắn.

3. Phương pháp nghiên cứu

Luận án sử dụng các phương pháp nghiên cứu sau:

- *Thu thập thông tin thông qua nghiên cứu tài liệu*

Thu thập, thống kê, phân tích từ các tài liệu công bố khoa học trong và ngoài nước cũng như các tài liệu vận hành thực tế từ các đơn vị để đánh giá hiện trạng, các phương pháp và phân tích ưu điểm, nhược điểm các công nghệ dự báo công suất phát ĐMT.

- *Thực nghiệm*

Thực nghiệm xây dựng mô hình dự báo công suất phát nhà máy ĐMT sử dụng mạng Nơ-ron hồi quy thông qua các khâu công việc bao gồm: thu thập số liệu, xử lý dữ liệu, lập trình xây dựng mô hình, huấn luyện mô hình. Trong quá trình thực nghiệm, tác giả cũng tiến hành đề xuất và thử nghiệm các giải pháp nhằm cải tiến hiệu quả của việc xây dựng mô hình dự báo. Đồng thời trong quá trình thực nghiệm này cũng sẽ tiến hành xây dựng, hoàn thiện phần mềm công cụ dự báo ngắn hạn công suất phát nhà máy ĐMT sử dụng mạng nơ-ron hồi quy.

- *Phân tích và tổng hợp*

Sử dụng các chỉ tiêu tiêu chuẩn về sai số dự báo để tính toán kiểm tra chất lượng mô hình dự báo đã xây dựng so sánh với số liệu thực đo thu thập trong thực tế và có sự

so sánh đối chiếu chất lượng mô hình dự báo với các mô hình khác để rút ra kết luận về tính hiệu quả và khả thi của mô hình.

4. Phạm vi nghiên cứu

Phạm vi nghiên cứu của luận án như sau:

- *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 máy 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. Cơ sở khoa học của đề tài là sự phát triển và ứng dụng các phương pháp học sâu và trí tuệ nhân tạo trong việc xây dựng mô hình dự báo công suất phát của ĐMT. Các mạng nơ-ron nhân tạo như Recurrent Neural Network (RNN) và Long Short-Term Memory (LSTM) đã được chứng minh là hiệu quả trong việc xử lý và dự báo dữ liệu chuỗi, bao gồm cả dự báo công suất phát ĐMT. Các nghiên cứu trong và ngoài nước 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. Công suất phát của các nhà máy ĐMT có tính biến động cao và phụ thuộc mạnh vào các yếu tố khí tượng như bức xạ mặt trời, nhiệt độ.... 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 với độ chính xác tốt hơn so với các mô hình dự báo truyền thống và có chất lượng tốt so với các mô hình dựa trên LSTM đã được công bố trong nước và quốc tế trong 5 năm trở lại đây.

- Đề xuất các giải pháp cải tiến hiệu quả trong xây dựng mô hình dự báo ngắn hạn công suất nhà máy ĐMT sử dụng mạng nơ-ron hồi quy, góp phần nâng cao chất lượng dự báo thông qua các giải pháp gồm:
 - o Tiềm xử lý số liệu với hệ số P/GHI kết hợp với phân cụm GHI
 - o Kỹ thuật sử dụng số liệu dự báo trong quá trình huấn luyện
 - o Kỹ thuật sử dụng dữ liệu bức xạ trời trong thay thế cho các chỉ dấu thời gian
- Xây dựng được quy trình hoàn chỉnh và công cụ phần mềm dự báo ngắn hạn công suất nhà máy ĐMT sử dụng mạng nơ-ron hồi quy, tạo điều kiện thuận lợi cho người sử dụng 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 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

Để có góc nhìn toàn diện về vấn đề dự báo công suất phát nhà máy ĐMT, chương này tập trung làm rõ các nội dung sau:

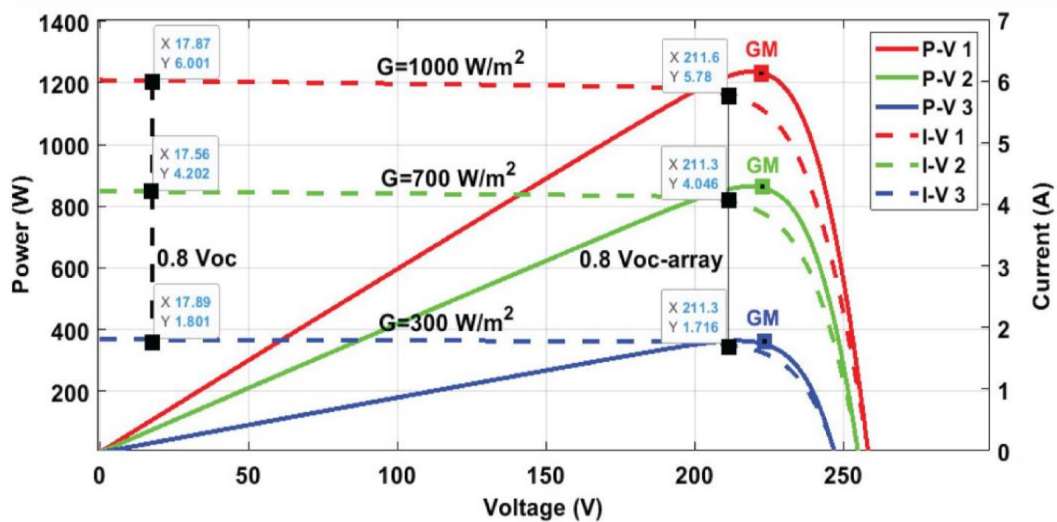
- Trình bày về công suất phát đầu ra của nhà máy ĐMT và các yếu tố ảnh hưởng đến công suất phát đầu ra này.
- Phân loại các khung thời gian của việc dự báo công suất phát đầu ra của nhà máy ĐMT từ đó đi sâu vào ý nghĩa của dự báo cho khung ngắn hạn
- Nghiên cứu tổng quan về các phương pháp dự báo ngắn hạn công suất phát nhà máy ĐMT
- Tình hình nghiên cứu bài toán dự báo ngắn hạn công suất phát nhà máy ĐMT quốc tế và trong nước
- Đánh giá và rút ra kết luận về hướng nghiên cứu.

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

1.2.1. Công suất phát đầu ra của nhà máy ĐMT

Công suất điện P được đo bằng Watt [W] của mô hình điện một chiều mô phỏng vận hành của tế bào quang điện, là tích của điện áp V được đo bằng Volt [V] và dòng điện I được đo bằng Ampere [A]:

$$P = V \times I \quad (1)$$



Hình 1.1. Đường cong I-V và đặc tính công suất của tấm pin quang điện

(Nguồn: S. Sarwar et al., 2022 [5])

Điện áp của tế bào quang điện được giữ ổn định dưới điều kiện ánh sáng mặt trời bình thường. Tuy nhiên, dòng điện của tế bào rất nhạy cảm với cường độ ánh sáng mặt trời. Cường độ ánh sáng mặt trời cao tạo ra dòng điện mạnh. Diện tích bề mặt của tế bào

quang điện cũng ảnh hưởng đến đầu ra dòng điện. Tế bào có diện tích bề mặt lớn sẽ tạo ra dòng điện nhiều hơn so với tế bào có diện tích bề mặt nhỏ. Vì vậy, cường độ ánh sáng chiếu vào tế bào và kích thước của tế bào quang điện là hai yếu tố chính ảnh hưởng đến đầu ra của một tế bào quang điện.

Đường cong I-V trên Hình 1.1 mô tả đặc tính công suất của tấm pin quang điện ở các mức bức xạ khác nhau. Đường cong công suất của tấm pin quang điện được thể hiện bằng đường cong màu xanh lam. Đường cong này biểu diễn công suất đầu ra theo các mức các điện áp. Trong điều kiện hoạt động bình thường, đường cong thay đổi theo cả trục dòng điện và trục điện áp. Điểm công suất tối đa (MPP – Maximum Power Point), điểm mà tấm pin quang điện đạt được công suất tối đa, xảy ra ở "điểm đầu gối" của đường cong công suất, nơi tích của điện áp và dòng điện là lớn nhất. Công suất cực đại P_{MPP} (W) được xác định bằng công thức:

$$P_{MPP} = V_{MPP} \times I_{MPP} \quad (2)$$

Tại bất kỳ thời điểm nào, một tấm pin điện đang hoạt động với một dòng điện và điện áp cụ thể nằm dọc theo đường cong I-V của nó (đường cong màu đỏ trong hình). Đường cong này cho thấy dòng điện I được tạo ra theo sự biến đổi của các mức điện áp. I_{SC} đại diện cho dòng điện ngắn mạch (SC – Short Circuit), tức là giá trị mà dòng điện đạt tới giá trị tối đa và điện áp bằng không. V_{OC} đại diện cho điện áp mở mạch (OC – Open Circuit), tức là giá trị mà điện áp đạt tối đa và dòng điện bằng không.

1.2.2. Các yếu tố ảnh hưởng đến công suất phát nhà máy ĐMT

1.2.2.1. Khí tượng

a. Bức xạ mặt trời

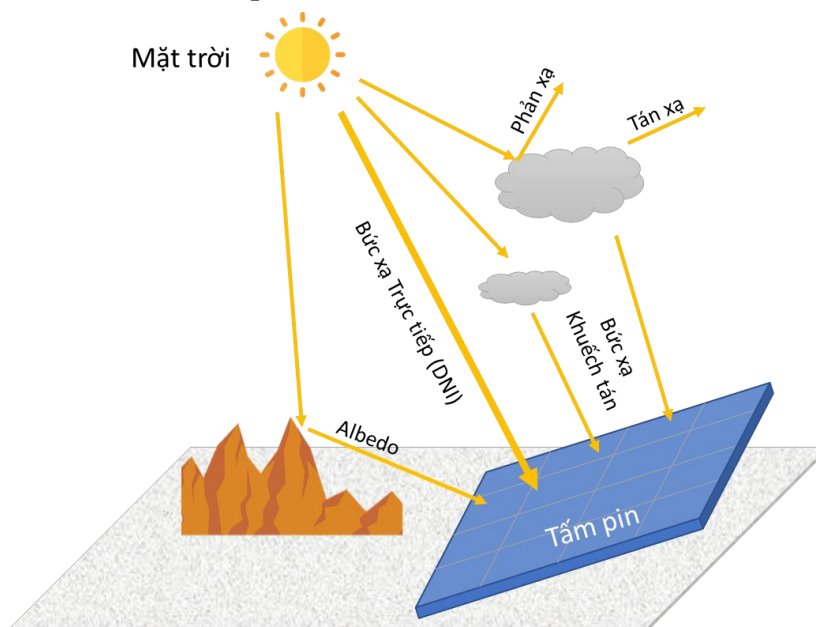
Khi bức xạ mặt trời đi qua khí quyển trái đất, nó sẽ bị ảnh hưởng bởi tương tác với các thành phần có mặt ở đó. Mây, những giọt nước và bụi sẽ gây ra hiện tượng phản xạ. Trong khi đó, những thành phần khác như ozone, oxy, carbon dioxide và hơi nước sẽ hấp thụ đáng kể bức xạ ở một số dải tần số cụ thể. Kết quả của tất cả các quá trình này là phân tách của bức xạ mặt trời chiếu vào vật nhận ở bề mặt trái đất, mà cụ thể trong trường hợp đang nghiên cứu là các tấm pin mặt trời, thành các thành phần khác nhau rõ rệt.

Bức xạ trực tiếp (Direct hoặc Beam Radiation), hình thành từ các tia sáng không bị phản xạ hoặc tán xạ mà đi thành đường thẳng trực tiếp từ mặt trời đến bề mặt tấm pin. Bức xạ này được gọi là DNI (Direct Normal Irradiance).

Bức xạ khuếch tán (Diffuse Radiation) mà bề mặt tấm pin nhận được do các nguồn sáng gián tiếp đến từ bầu trời mà không phải là Bức xạ trực tiếp.

Bức xạ Albedo là bức xạ phản xạ từ mặt đất.

Tổng hợp bức xạ mà bề mặt tấm pin nhận được gồm DNI, Diffuse Radiation và Albedo hình thành nên một giá trị được gọi là Bức xạ toàn phần theo phương ngang (*GHI – Global Horizontal Irradiance*) có đơn vị là W/m^2 . Hình 1.2 mô tả chi tiết các thành phần bức xạ mà một tấm pin mặt trời nhận được.



Hình 1.2. Các thành phần bức xạ mà tấm pin mặt trời nhận được

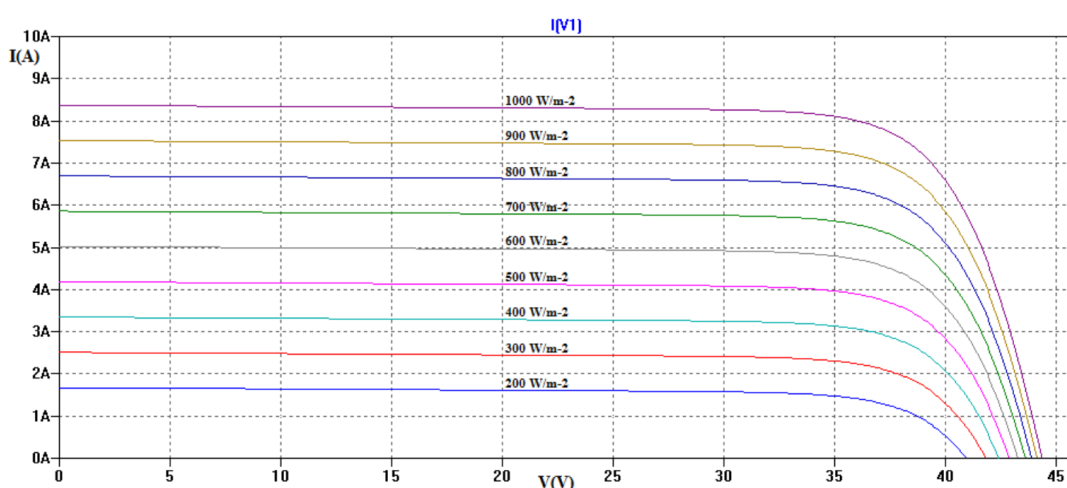
Bức xạ trời trong là lượng bức xạ mặt trời tại bề mặt Trái Đất trong điều kiện không bị che khuất, không có mây. Nó đại diện cho bức xạ mặt trời tối đa có thể nhận được tại một vị trí cụ thể trên Trái Đất. Các giá trị tính toán bức xạ trời trong cung cấp thông tin quan trọng để đánh giá tiềm năng sinh năng lượng điện mặt trời tối đa của một nhà máy điện mặt trời tại bất kỳ vị trí địa lý nào trong suốt cả năm. Việc tính toán bức xạ trời trong tại một vị trí lắp đặt nhà máy điện mặt trời liên quan đến việc sử dụng các mô hình và phương pháp khác nhau [6]. Nghiên cứu [7] trình bày đánh giá so sánh các mô hình khác nhau.

Một số mô hình đáng chú ý bao gồm:

- Mô hình Ineichen là một mô hình thường được sử dụng, ước tính bức xạ trời trong dựa trên các thông số khí quyển như nồng độ hơi nước, nồng độ ozone và độ sâu quang phổ của các hạt bụi. Mô hình này xem xét vị trí của mặt trời, vị trí địa lý và thời điểm trong năm để tính toán bức xạ trời trong [8], [9].
- Mô hình Haurwitz là một phương pháp khác được sử dụng để tính toán bức xạ trời trong. Nó sử dụng góc zenith mặt trời, vĩ độ và thời điểm trong năm để ước tính bức xạ. Mô hình này xem xét vị trí của mặt trời liên quan đến vị trí và tính toán theo cong quỹ đạo của Trái Đất. Theo báo cáo [10] mô hình Haurwitz có hiệu suất tốt nhất trong số các mô hình chỉ yêu cầu góc zenith.

- Ngoài ra, mô hình Solis đơn giản cũng được sử dụng để ước tính bức xạ trời trong. Mô hình này tính đến góc zenith mặt trời và vĩ độ của vị trí để tính toán bức xạ trời trong. Độ chính xác của mô hình được công bố là 15, 20 và 18 W/m^2 đối với các thành phần bức xạ GHI và DNI [11].
- Mô hình Bird Clear Sky: Mô hình này kết hợp mô hình bức xạ mặt trời Bird với điều kiện trời quang để ước tính GHI trong trời quang. Nó xem xét các biến số như nồng độ hơi nước trong khí quyển, nồng độ ozone và độ sâu phổ của các hạt bụi trong không khí [6].

Theo báo cáo phân tích so sánh về các mô hình bức xạ trong trời trong, mô hình Ineichen/Perez cho thấy hiệu suất tốt nhất với số lượng dữ liệu đầu vào tối thiểu [10]. Do đó, luận án này sẽ sử dụng mô hình Ineichen/Perez để tính toán bức xạ trời trong tại vị trí của nhà máy.



Hình 1.3: Đặc tính I-V của tấm pin MSP290AS-36.EU

(Nguồn: Tài liệu kỹ thuật tấm pin do hãng công bố)

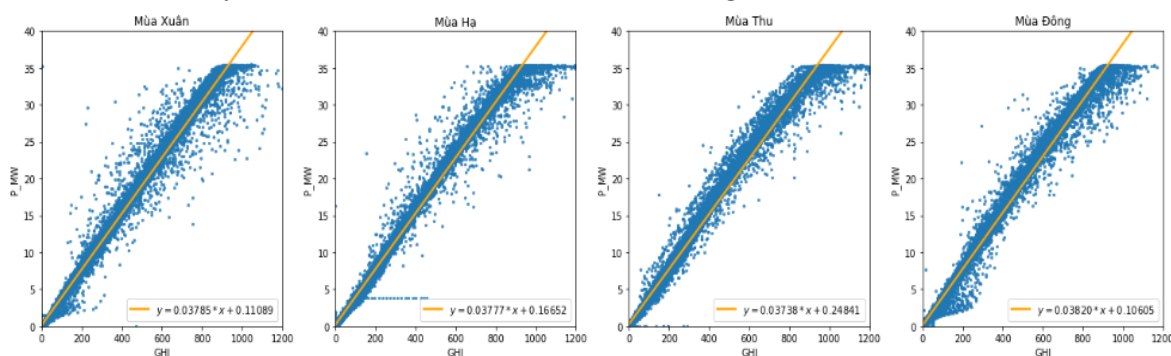
Công suất đầu ra của một tế bào hoặc tấm pin quang điện phụ thuộc trực tiếp vào bức xạ mặt trời chiếu trực tiếp vào bề mặt của nó [12]. Hình 1.3 mô tả đặc tính I-V của tấm pin MSP290AS-36.EU với nhiều mức bức xạ khác nhau từ 200W/m^2 đến 1000W/m^2 ở 25°C (nhiệt độ tiêu chuẩn phòng thí nghiệm).

Ngay khi ánh sáng mặt trời chiếu sáng vào bề mặt của tế bào/tấm pin, điện áp tăng lên một giá trị gần với V_{OC} . Bất kể sự thay đổi của bức xạ mặt trời, chẳng hạn như bị che khuất bởi đám mây, điện áp chỉ dao động một chút dưới khoảng giá trị V_{OC} .

Trong khi đó, khi bức xạ mặt trời bề mặt tấm pin nhận được tăng, dòng điện I tăng do sự gia tăng của hiệu ứng quang điện. Dòng điện, tuy nhiên, sẽ tăng tỷ lệ thuận với độ chiếu sáng, chỉ đạt đến dòng điện định mức của tế bào/tấm pin trong điều kiện độ chiếu sáng mạnh, chẳng hạn như 1000W/m^2 . Khi có vật thể che khuất như đám mây, động vật bay, cành cây... sẽ tạo ra bóng đổ trên tấm pin, và gây ra sụt giảm dòng điện.

Vì dòng điện tỷ lệ thuận trực tiếp với công suất đầu ra, vì vậy bóng đổ có ảnh hưởng đáng kể đến công suất đầu ra.

Theo nghiên cứu [13], chỉ số GHI có mức độ tương quan rất cao đối với công suất phát đầu ra của nhà máy ĐMT, hệ số tương quan lên đến khoảng 98%. Do đó, bức xạ mặt trời có ảnh hưởng rất mạnh đến công suất phát đầu ra của nhà máy ĐMT. Hình 1.4 thể hiện mối quan hệ giữa các yếu tố GHI (W/m^2) với công suất phát đầu ra P (MW) của một nhà máy ĐMT vào các mùa khác nhau trong năm.



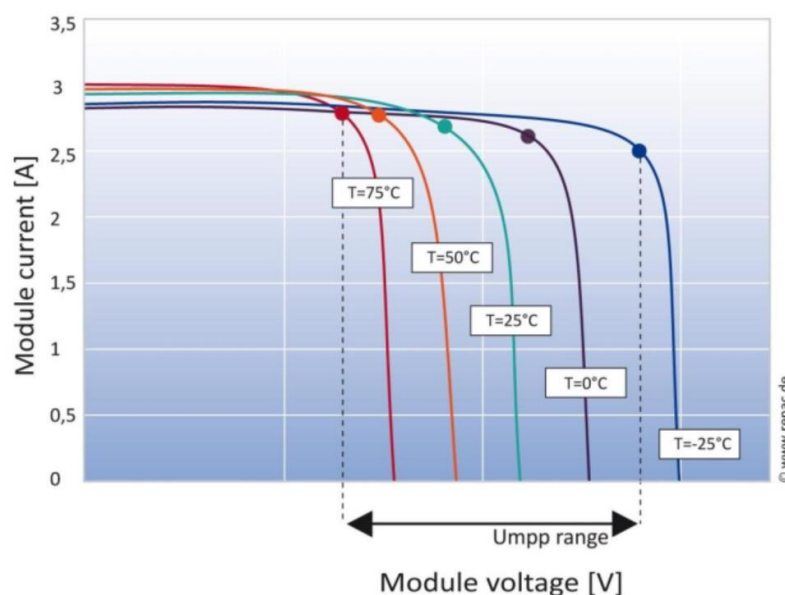
Hình 1.4. Quan hệ giữa GHI và công suất phát đầu ra của một nhà máy ĐMT

b. Nhiệt độ

Khi tấm pin nóng lên, công suất đầu ra của nó giảm đi. Điều này là do điện áp mạch mở V_{OC} giảm đáng kể khi nhiệt độ nội bộ của tế bào quang điện tăng lên. Dòng điện ngắn mạch I_{sc} tăng nhẹ với sự tăng nhiệt độ, nhưng không đủ để bù đắp cho sự giảm V_{OC} lớn. Hình 1.5 mô tả mối quan hệ giữa nhiệt độ tấm pin với các thông số điện áp và dòng điện của tấm pin [14].

Dưới điều kiện bình thường, tấm pin mặt trời sẽ luôn hoạt động ở nhiệt độ cao hơn nhiệt độ môi trường. Khi tấm pin nóng lên, công suất đầu ra giảm xuống do điện áp hở mạch giảm. Để đánh giá độ suy giảm của điện áp, dòng điện hay công suất phát đầu ra theo độ tăng nhiệt độ, người ta sử dụng hệ số nhiệt độ. Các datasheet của các tấm pin đều được nhà sản xuất cung cấp hệ số nhiệt độ (T_C) cho điện áp mạch mở V_{OC} trong điều kiện thí nghiệm tiêu chuẩn (STC – Standard Test Condition), dòng điện mạch ngắn I_{sc} trong STC và giá trị công suất tối đa (MPP – Maximum Power Point) trong STC.

Hệ số ảnh hưởng của nhiệt độ đến điện áp là hệ số được sử dụng phổ biến nhất. Inverter (và các thiết bị khác như điều khiển sạc) có thể bị hư hỏng do điện áp của tấm pin/dây tấm pin quá mức so với điện áp đầu vào định mức của các inverter (và các thiết bị khác). Ngược lại, điện áp quá thấp có thể gây giảm hiệu suất hoạt động của hệ thống. Bảng 1.1 trình bày ví dụ về hệ số ảnh hưởng của nhiệt độ đến các thông số vận hành tấm pin. Ở đây mỗi $^{\circ}C$ tăng lên hay giảm đi được hiểu là chênh lệch so với nhiệt độ tại điều kiện thí nghiệm tiêu chuẩn ($25^{\circ}C$).



Hình 1.5. Ảnh hưởng của nhiệt độ lên điện áp và dòng điện của tấm pin

(Nguồn: RENAC [14])

Bảng 1.1. Ví dụ về hệ số ảnh hưởng của nhiệt độ đến thông số vận hành tấm pin

Loại	Dạng đơn vị	Ví dụ
Hệ số nhiệt độ cho điện áp	%/°C	-0,36%/°C
	V/°C	-0,156 V/°C hoặc -156 mV/°C
Hệ số nhiệt độ cho dòng điện	%/°C	+0,045%/°C
	A/°C	+0,0029 A/°C hoặc +2,9 mA/°C
Hệ số nhiệt độ cho công suất	%/°C	-0,42%/°C
	W/°C	+0,957 W/°C

c. Tốc độ gió, độ ẩm

Tốc độ gió hay độ ẩm cũng có mối liên hệ tương đối đối với công suất phát đầu ra của nhà máy ĐMT [15], [16], [17].

Các tác giả trong [13] đã tiến hành phân tích hệ số tương quan (Pearson test) của tốc độ gió và độ ẩm với công suất đầu ra của nhà máy ĐMT. Kết quả cho thấy hệ số tương quan của gió là khoảng 0,19-0,38 và hệ số tương quan của độ ẩm là từ -0,49 đến -0,76. Như vậy có nghĩa là tốc độ gió có tương quan thuận tuy nhiên tương đối yếu đến công suất phát đầu ra. Điều này có thể lý giải bởi gió sẽ tương quan gián tiếp thông qua việc làm mát bề mặt tấm pin giúp tăng hiệu suất tấm pin.

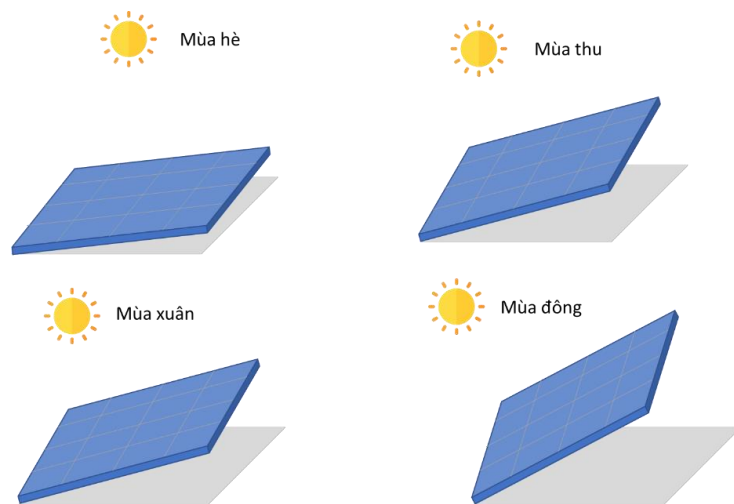
Đối với độ ẩm, mặc dù tương quan nghịch tương đối mạnh nhưng vấn đề tương tự như đã phân tích đối với nhiệt độ đó là việc tùy vào thời điểm khác nhau trong ngày thì độ ẩm cũng biến thiên. Vào thời điểm giữa trưa, khi bức xạ cao nhất thì thường độ

âm giảm xuống thấp nhất trong khi điều này ngược lại vào ban đêm. Từ đó dẫn đến tương quan nghịch của độ ẩm không khí.

1.2.2.2. Góc lắp đặt tấm pin

Góc nghiêng của tấm pin mặt trời ảnh hưởng đáng kể đến công suất phát đầu ra của hệ thống điện mặt trời. Khi tấm pin mặt trời được lắp đặt dưới góc nghiêng tối ưu, các tia sáng mặt trời sẽ chiếu thẳng vào bề mặt của tấm pin, làm cho lượng ánh sáng được hấp thụ bởi tấm pin đạt đến mức cao nhất. Khi ánh sáng chiếu thẳng vào tấm pin, hiệu suất chuyển đổi năng lượng từ ánh sáng thành điện năng của tấm pin cũng đạt đến mức tối đa. Do đó, công suất phát đầu ra của hệ thống điện mặt trời sẽ tăng lên. Tuy nhiên, khi tấm pin mặt trời được lắp đặt dưới góc nghiêng không tối ưu, ví dụ như khi tấm pin được lắp đặt quá nằm ngang hoặc quá thẳng đứng, lượng ánh sáng chiếu vào tấm pin sẽ giảm đi, gây ra sự giảm công suất phát đầu ra của hệ thống điện mặt trời.

Do đó, việc lắp đặt tấm pin mặt trời dưới góc nghiêng tối ưu là rất quan trọng để đảm bảo rằng hệ thống điện mặt trời hoạt động ở hiệu suất tối đa. Góc nghiêng tối ưu của tấm pin mặt trời phụ thuộc vào vị trí địa lý, mùa trong năm và địa hình xung quanh, và có thể được tính toán bằng các công cụ và phần mềm tính toán hiệu suất của hệ thống điện mặt trời. Điều chỉnh hướng và độ nghiêng của tấm pin ảnh hưởng đáng kể đến lượng bức xạ mặt phẳng nhận được và do đó là lượng năng lượng mà tấm pin sản xuất ra.



Hình 1.6. Nguyên lý thay đổi góc nghiêng theo mùa để đạt sản lượng tối đa

Ở bán cầu bắc, tấm pin nên được hướng về phía nam; ở bán cầu nam, nó nên được hướng về phía bắc. Điều này đảm bảo mức độ bức xạ tối đa lên tấm pin trong suốt năm [18]. Một ví dụ về nguyên lý thay đổi góc nghiêng theo mùa để đạt được sản lượng tối đa được trình bày như Hình 1.6.

1.2.2.3. Phụ tải và chế độ điều khiển

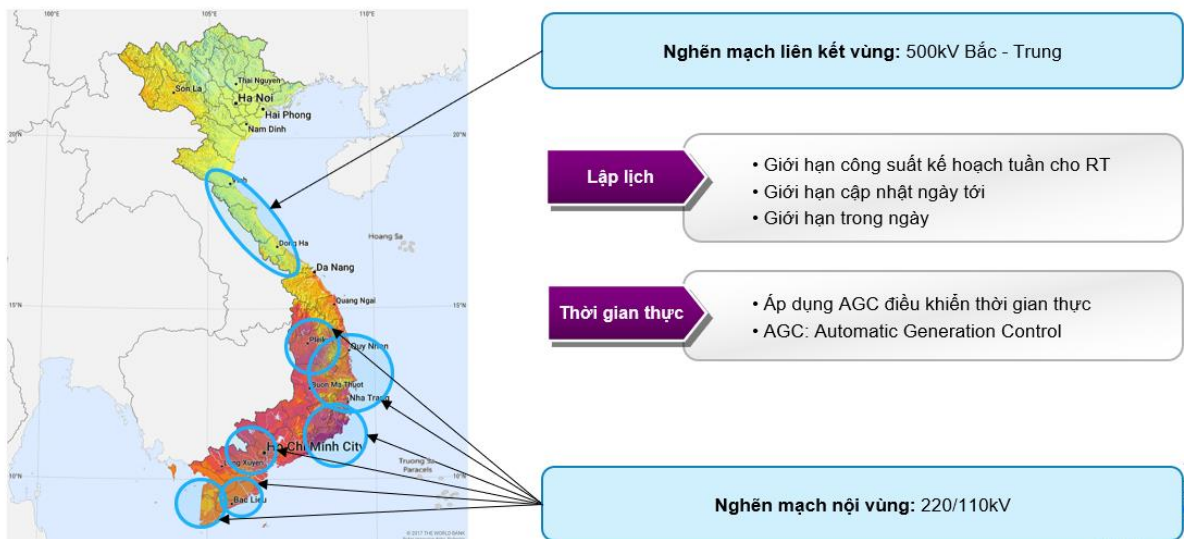
Một phụ tải tải điện (hay chính là điện trở trong mô hình DC) được kết nối với một tấm pin sẽ xác định điểm làm việc trên đường cong đặc tính I-V của tấm pin. Phụ tải điện có thể là bóng đèn, bơm, lưới điện hoặc ắc quy hoặc chính là toàn bộ hệ thống điện mà nhà máy điện kết nối. Điện áp qua tải và cường độ dòng điện qua tải sẽ được xác định, cho một bộ giá trị I-V cố định, bởi phương trình:

$$I = \frac{V}{R} \quad (3)$$

trong đó I là dòng điện tính bằng Ampe, V là điện áp tính bằng Volt và R là trở kháng tải tính bằng Ohm.

Trong chế độ làm việc nối lưới hiện nay, trở kháng kết nối với các hệ thống điện mặt trời thường xuyên biến động, tùy thuộc vào điều kiện của lưới. Khi điều khiển giữ V, điện áp đầu ra của hệ thống được giữ ở mức cố định, nhưng lượng điện năng được sản xuất có thể thay đổi do ảnh hưởng của ánh sáng mặt trời và các yếu tố khác. Trong khi đó, khi điều khiển giữ Q, lượng điện năng được sản xuất được giữ ở mức cố định, nhưng điện áp đầu ra có thể thay đổi tùy thuộc vào điều kiện ánh sáng và khác nhau giữa các bộ phận của hệ thống.

1.2.2.4. Giới hạn lưới điện đấu nối



Hình 1.7. Giới hạn công suất do nghẽn mạch truyền tải

(Nguồn: “Vietnam, a System in Rapid Transition” - ESIG Spring Technical Workshop [3])

Một vấn đề đáng chú ý trong vận hành là vấn đề nghẽn mạch, hiện tại do sự phân bố không cân đối của phụ tải và nguồn năng lượng tái tạo, trong các thời điểm đặc biệt là khung giờ giữa trưa, hệ thống gặp phải vấn đề nghẽn mạch đối với một số nhóm nhà máy đấu nối vào các đường dây 220/110kV có khả năng truyền tải bị hạn chế. Ngoài ra

khi năng lượng tái tạo phát quá cao thì việc nghẽn mạch do liên kết 500kV từ miền Trung ra miền Bắc cũng diễn ra.

Để xử lý vấn đề này, đơn vị điều độ đã sử dụng hệ thống AGC (Automatic Generation Control – Hệ thống tự động điều khiển công suất) để tự động điều chỉnh công suất các nhà máy trong các nhóm nghẽn mạch an toàn cho đường dây và đảm bảo công bằng trong huy động giữa các nhà máy. Điều này dẫn đến khả năng nhà máy được điều chỉnh phát thấp hơn so với khả năng thực phát của nhà máy theo bức xạ và các điều kiện thời gian thực.

1.3. Phân loại các khung dự báo công suất ĐMT

Khung dự báo là khoảng cách về thời gian giữa thời điểm đưa ra bản tin dự báo và thời điểm mà bản tin dự báo đó có cung cấp thông tin. Việc phân chia khung thời gian dự báo cũng chưa có sự thống nhất hoàn toàn, theo các nghiên cứu gần đây, có thể chia ra thành 3 khung: dài hạn, trung hạn, ngắn hạn hoặc một số nghiên cứu lại bổ sung thêm một khung nữa đó là khung rất ngắn hạn. Tổng hợp cụ thể từ các nghiên cứu [19], [20], [21], [22], [23] về việc phân loại khung thời gian dự báo và ứng dụng cụ thể của từng khung có thể chia ra như sau:

Bảng 1.2. Phân loại các khung dự báo

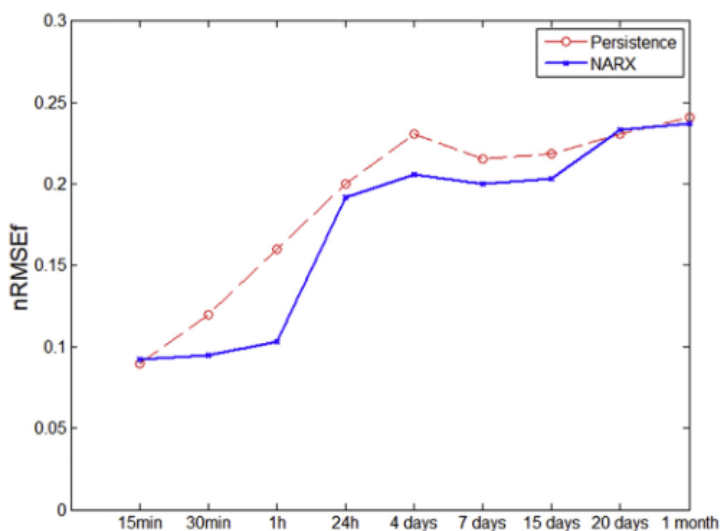
Loại dự báo	Khung dự báo	Ứng dụng
Dự báo dài hạn ([19], [24])	Từ 1 năm đến 10 năm	Đầu tư xây dựng, phát triển, nâng cấp lưới điện, nguồn điện.
Dự báo trung hạn ([13], [19], [23])	Từ 1 tuần hoặc 1 tháng đến 1 năm	Sắp xếp lịch sửa chữa định kỳ cho nhà máy và cho lưới điện, đánh giá xác định cấu hình vận hành cơ bản cho lưới điện.
Dự báo ngắn hạn ([13], [24])	Từ 30 phút hoặc 60 phút đến 6 tiếng hoặc 01 tuần	Vận hành kinh tế hệ thống điện, cân bằng nguồn tải, chào giá thị trường điện ngày tới, sắp xếp lịch sửa chữa đột xuất, quản lý hoạt động nhà máy điện...
Dự báo rất ngắn hạn ([13], [23])	Từ 1 phút – 30 phút hoặc 60 phút	Điều khiển thời gian thực của hoạt động hệ thống điện và của các hệ thống lưu trữ, đảm bảo dự phòng điều tần và dự phòng quay, chào giá trong thị trường điện trong ngày...

Sai số dự báo của các khung thời gian dự báo khác nhau cũng rất khác nhau:

- Đối với khung rất ngắn hạn, sai số đạt được cho 15 phút tới tại nghiên cứu [25] là nRMSE 9% cho các phương pháp thí nghiệm.
- Đối với khung ngắn hạn, một số nghiên cứu đã công bố sai số và nhìn chung kết quả kém hơn so với khung rất ngắn hạn. Chen và cộng sự [26] đã thực hiện nghiên cứu

mô hình dự báo cho các khung từ 6 tiếng đến 48 tiếng tới cho các kiểu hình thời tiết khác nhau và kết quả thu được sai số MAPE là 9,45% cho trời nắng, 9,88% cho trời nhiều mây, 38,12% cho trời mưa. Đối với dự báo cho 24 tiếng tới, nghiên cứu [25] đã đạt được một mô hình sai số có nRMSE là 19% tốt hơn một chút so với giá trị tiêu chuẩn theo phương pháp quán tính là 20%.

- Đối với khung trung hạn, sai số đạt được tăng dần theo thời gian. Tại nghiên cứu [25] (Hình 1.8) đã chỉ ra khi lên đến khung khoảng 1 tháng sai số sẽ lên đến gần 25%.



Hình 1.8. Sự thay đổi của sai số dự báo theo khung thời gian dự báo

(Nguồn: A. G. R. Vaz và cộng sự, 2016 [25])

- Đối với khung dài hạn chưa có thông tin công bố nào về độ chính xác hay sai số đạt được của các mô hình dự báo.

1.4. Đánh giá chất lượng mô hình dự báo

1.4.1. Đánh giá sai số thông qua các chỉ số

Rất nhiều nghiên cứu đã được thực hiện để phát triển các mô hình dự báo công suất điện mặt trời. Tuy nhiên, việc đánh giá hiệu quả hoạt động của các phương pháp dự báo khác nhau còn tương đối phức tạp, do các nghiên cứu sử dụng các thước đo khác nhau làm tiêu chí riêng. Đánh giá, đo lường sai số đóng vai trò rất quan trọng để đánh giá độ chính xác của mô hình dự báo, từ nhiều nghiên cứu khác nhau [27], [28], [29], [30], [31] có thể thấy các nhóm tiêu chí thường được sử dụng bao gồm:

- MAE (Mean Absolute Error) hay Sai số tuyệt đối trung bình
- PE (Percentage Error) hay Sai số tương đối
- APE (Percentage Error) hay Sai số tuyệt đối phần trăm
- MAPE (Mean Absolute Percentage Error) hay Sai số tuyệt đối phần trăm trung bình
- MSE (Mean Square Error) hay Sai số trung bình bình phương

- RMSE (Root Mean Square Error) hay Sai số trung bình bình phương gốc
- nRMSE (normalize Root Mean Square Error) hay Sai số trung bình bình phương gốc chuẩn hóa.

Theo [13], [22], [24], [31], công thức sử dụng cho các tiêu chí đánh giá sai số mô hình này như sau:

$$MAE = \frac{1}{N} \sum_{k=1}^N |P_{db} - P_{td}| \quad (4)$$

$$PE = \frac{P_{db} - P_{td}}{P_{dm}} \times 100\% \quad (5)$$

$$APE = \frac{|P_{db} - P_{td}|}{P_{dm}} \times 100\% \quad (6)$$

$$MAPE = \frac{1}{N} \sum_{k=1}^N \frac{|P_{db} - P_{td}|}{P_{dm}} \times 100\% \quad (7)$$

$$MSE = \frac{1}{N} \sum_{k=1}^N (P_{db} - P_{td})^2 \quad (8)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^N (P_{db} - P_{td})^2} \quad (9)$$

$$nRMSE = \frac{RMSE}{P_{dm}} \quad (10)$$

Trong đó: P_{td} , P_{db} and P_{dm} là giá trị thực đo, giá trị trung bình và công suất đặt của nhà máy ĐMT; N là số điểm mẫu trong bộ dữ liệu để thực hiện đánh giá chỉ số.

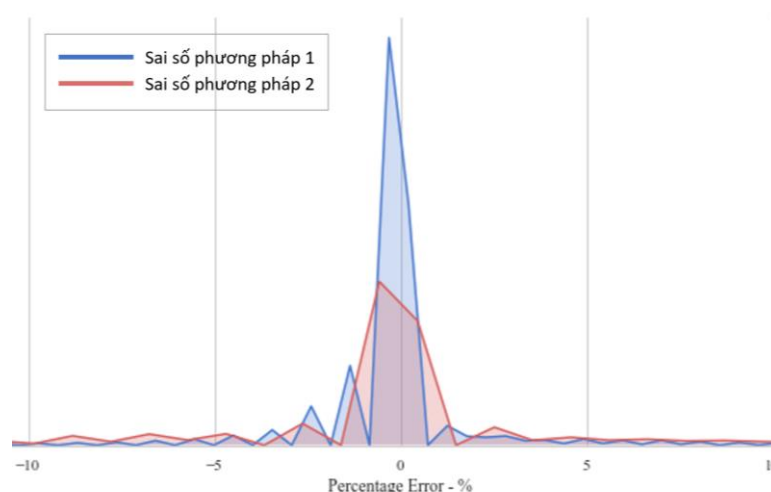
Về mặt ý nghĩa, MAE là trung bình trị tuyệt đối của tất cả các sai số giữa công suất thực tế và dự báo trên toàn bộ mẫu kiểm tra, MAPE là MAE được chuẩn hóa bởi công suất định mức của nhà máy. Hai chỉ số này mang thông tin đo lường độ chính xác trung bình của các dự báo mà không xem xét chiều của sai số.

Trong khi đó MSE là trung bình bình phương của tất cả các sai số giữa công suất thực tế và dự báo trên toàn bộ mẫu kiểm tra, RMSE là căn bậc hai của MSE, nRMSE là giá trị RMSE được chuẩn hóa bởi công suất định mức của nhà máy. Các chỉ số này đo lường độ chính xác trung bình của các dự báo mà không xem xét chiều sai số và tăng trọng số đối với các sai số lớn.

1.4.2. Đánh giá sai số 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 các sai số đã xảy ra trên tập dữ liệu đánh giá và có thể được hiểu như xác suất mà một giá trị sai số nhất định sẽ xảy ra. Ngoài ra, các lỗi lớn nhất

cũng có thể được phát hiện thông qua biểu đồ. Tuy nhiên, đây chỉ là dữ liệu thống kê từ quá khứ và một giá trị sai số khác không nằm trong phân bố này hoàn toàn có thể xảy ra trong tương lai. Vì biểu đồ phân bố sai số là hình vẽ mà phải là một con số cụ thể nên khả năng hiểu và phân tích dữ liệu từ đây rất quan trọng để có thể rút ra được những nhận xét đánh giá. Hình 1.9 dưới đây là một ví dụ về biểu đồ phân bố sai số phần trăm của hai phương pháp dự báo. Để nhận thấy rằng phương pháp 1 cho kết quả tốt hơn so với phương pháp 2 do sai số tập trung quanh mức 0% của phương pháp cao hơn hẳn so với phương pháp 2. Đồng thời ở những mức sai số lớn, xác suất xảy ra của phương pháp 1 cũng thấp hơn so với phương pháp 2.



Hình 1.9. Biểu đồ phân bố sai số phần trăm của hai phương pháp dự báo

1.5. Các phương pháp dự báo ngắn hạn công suất nhà máy ĐMT

Có một số phương pháp để phân loại các mô hình dự báo. Theo [13], có thể phân loại theo các tiếp cận là trực tiếp và gián tiếp. Trong cách tiếp cận gián tiếp thì bằng cách phương pháp khác nhau người ta sẽ xác định hoặc dự báo ra các yếu tố ảnh hưởng đến công suất phát (thông số tấm pin, bức xạ, nhiệt độ..) sau đó sẽ sử dụng các công thức quan hệ để tính toán ra công suất phát. Trong các tiếp cận trực tiếp, các yếu tố ảnh hưởng được thu thập sau đó người ta sẽ xây dựng mô hình để dự báo trực tiếp công suất phát từ các yếu tố này. Nghiên cứu [32] bằng phương pháp trực tiếp đã cho thấy kết quả đạt được tốt hơn phương pháp gián tiếp.

Để phân loại các phương pháp dự báo công suất phát của các nhà máy ĐMT còn có thể dựa trên phương pháp dự báo. Như trình bày ở Bảng 1.3, có 4 nhóm chính của kỹ thuật dự báo: phương pháp vật lý, phương pháp thống kê chuỗi thời gian, phương pháp học máy và phương pháp kết hợp (phương pháp lai hoặc hỗn hợp). Mô hình dự báo theo Quán tính (Persistence model) có thể coi là một nhánh trong phương pháp thống kê. Phương pháp ANN (Artificial Neural Network – Mạng nơ-ron nhân tạo) trước

đây có thể coi là một nhánh của phương pháp thống kê. Tuy nhiên gần đây không chỉ có ANN mà rất nhiều tác giả đã đào sâu hướng nghiên cứu ứng dụng Học máy (Machine learning) hay Học sâu (Deep learning) vào bài toán dự báo công suất ĐMT. Do đó, xu hướng mới gần đây coi phương pháp sử dụng Học máy là một nhánh độc lập với nhánh sử dụng thống kê cổ điển.

Bảng 1.3. Tổng hợp phân loại các phương pháp dự báo công suất ĐMT

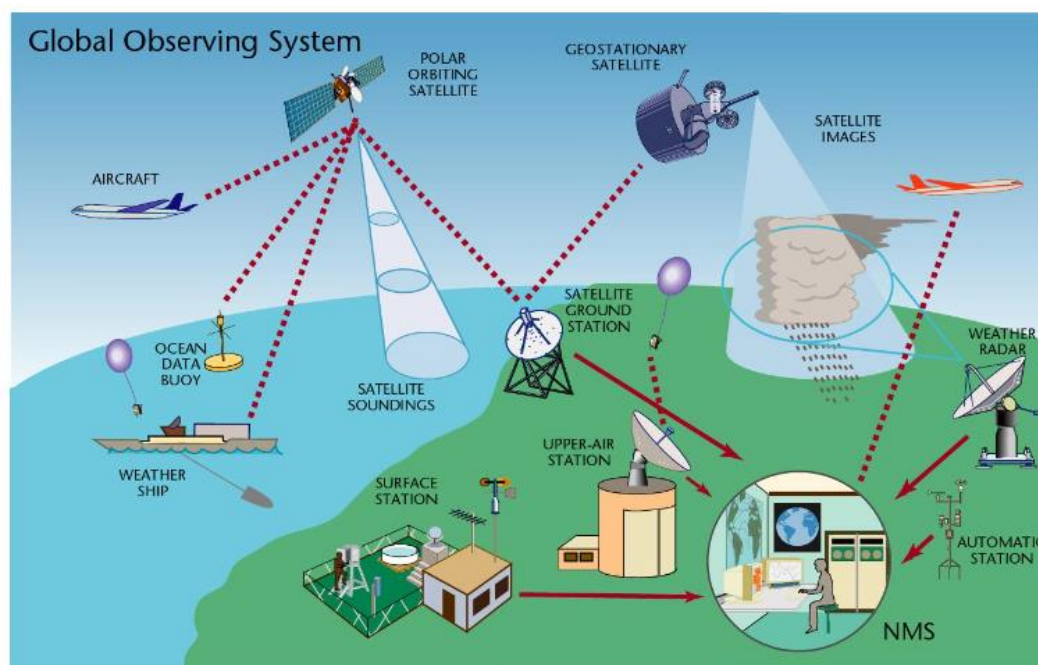
Phương pháp	Cách tiếp cận	Mô tả	Các kỹ thuật chính
Vật lý	Gián tiếp	Sử dụng các mô hình vật lý, khí tượng để tính toán ra các thông số như Bức xạ, nhiệt độ... sau đó sử dụng công thức quan hệ để tính ra công suất phát	- Dự báo Thời tiết Số - Ảnh mây
Thống kê	Trực tiếp hoặc Gián tiếp	Sử dụng mô hình thống kê để xác định ra công suất (trực tiếp) hoặc xác định ra các thông số khác có thể tính toán ra công suất (gián tiếp)	- Quán tính - Mô hình thống kê chuỗi thời gian
Học máy	Trực tiếp hoặc Gián tiếp	Sử dụng mô hình học máy mà chủ yếu là các mạng nơ-ron để xác định ra công suất (trực tiếp) hoặc xác định ra các thông số khác có thể tính toán ra công suất (gián tiếp)	- Mạng nơ-ron đa lớp - Mạng nơ-ron hồi quy
Kết hợp	Gián tiếp	Sử dụng các mô hình kết hợp kết quả của nhiều mô hình khác nhau để tính toán công suất.	- Kỹ thuật hợp tác - Kỹ thuật cạnh tranh

Như vậy có thể thấy đến nay đã và đang có rất nhiều phương pháp đã và đang được áp dụng cho dự báo công suất ngắn hạn ĐMT. Tuy nhiên những kỹ thuật mới đang không ngừng được nghiên cứu và cải tiến giúp nâng cao hiệu quả dự báo. Do đó, để hiểu rõ bản chất, ưu nhược điểm của từng phương pháp và có sự so sánh để tìm ra phương pháp hiệu quả nhất từ đó đề xuất ra những hướng đi áp dụng mới chính là một trong những định hướng quan trọng của luận án này.

1.5.1. Phương pháp Vật lý

Phương pháp vật lý chủ yếu dựa trên Dự báo Thời tiết Số (NWP) khi sử dụng các phương trình động lực học và nhiệt động học để ước tính trạng thái của khí tại một thời điểm trong tương lai từ trạng thái của khí quyển tại một thời điểm nhất định.

Những dữ liệu đầu vào được thu thập từ các trạm thời tiết, thiết bị đo, cũng như từ các tàu biển và máy bay. Dữ liệu này được phân tích cùng với hình ảnh vệ tinh để có được tình trạng hiện tại của khí quyển. Việc xử lý các bộ dữ liệu lớn và thực hiện các tính toán phức tạp liên quan đến các mô hình toán học NWP với độ phân giải tốt đủ để kết quả có ích yêu cầu sử dụng một số máy tính siêu mạnh nhất trên thế giới. Do đó, hầu hết các mô hình NWP đã được phát triển và triển khai tại các Trung tâm Khí tượng Quốc gia.



Hình 1.10. Các nguồn dữ liệu khí tượng

(Nguồn: Sarkar [33])

Mô hình toàn cầu sẽ bao phủ toàn bộ trái đất. Hiện tại có khoảng 9 mô hình toàn cầu với khả năng độ dài dự báo khác nhau:

+ Nhóm chất lượng cao nhất: ECMWF – Châu Âu (10 ngày), GFS – Mỹ (15 ngày), UKMO – Anh (6 ngày).

+ Nhóm còn lại: ICON – Đức (7 ngày), ARPEGE – Pháp (5 ngày), GEM – Canada (10 ngày), ACCES-G – Úc (10 ngày), GDAPS/UM – Nhật (12 ngày), CMA – Trung Quốc (10 ngày).

Dữ liệu khí tượng được cung cấp từ mô hình toàn cầu này thường là dữ liệu khí tượng thô, muốn biến thành dữ liệu cho từng site phải qua xử lý. Việc cung cấp miễn

phí chủ yếu chỉ cho các đơn vị quốc gia thành viên, có thỏa thuận hoặc cho mục đích nghiên cứu. Ngoài ra các tổ chức này cũng cung cấp dữ liệu dự báo khí tượng thương mại cho các tổ chức, công ty thương mại khai thác và sử dụng. Dữ liệu thường được bán theo gói và theo khu vực: ví dụ ECMWF bán theo từng gói dữ liệu có giá dao động từ 50.000 Euro tới vài trăm nghìn Euro/năm.

Do bài toán dự báo ngắn hạn NLTT nói chung và bài toán dự báo ngắn hạn ĐMT nói riêng chỉ là một lĩnh vực ứng dụng chuyên biệt từ mô hình NWP nên từ mô hình NWP cần có các bước xử lý tiếp theo để có thể thu được kết quả mong muốn. Theo đó, các mô hình sử dụng các phương trình vật lý hoặc các công cụ thống kê hay Mạng nơ-ron nhân tạo (Artificial Neural Networks - ANNs) để tính toán công suất phát hay sản lượng phát từ thời tiết sẽ được nghiên cứu phát triển và sử dụng.

Đối với phương pháp NWP, ban đầu, các viện nghiên cứu và phát triển khí tượng thủy văn đã khởi xướng và phát triển các công cụ dự báo ngắn hạn về sức mạnh gió. Hai tổ chức nghiên cứu trong lĩnh vực này là Phòng thí nghiệm Quốc gia Risø (Risø National Laboratory) và Đại học Kỹ thuật Đan Mạch (DTU - Technical University of Denmark) từ những năm 1990. Kể từ đó, các dự án tương tự đã được thực hiện bởi nhiều viện nghiên cứu và công ty tư nhân trên toàn thế giới. Trong những năm gần đây, các viện nghiên cứu và công ty này đã tiếp tục phát triển các hệ thống dự báo cho năng lượng mặt trời. Giải pháp mà họ sử dụng đó là tiếp nhận/mua lại số dự báo từ mô hình NWP sau đó kết hợp với dữ liệu tự thu thập, quan trắc để xử lý đưa ra mô hình dự báo riêng, phù hợp với nhu cầu của người sử dụng cuối cùng. Theo công bố từ các nhà cung cấp dịch vụ dự báo thương mại có uy tín trên thế giới như Solargis, Solcast, Meteorologica..., họ đang thường sử dụng mô hình ECMWF, ICON cho dự báo ngày tới và ECMWF, GFS dự báo 10-14 ngày ngày tới. Đối với dự báo trong ngày tới, phương pháp sử dụng là kết hợp giữa các mô hình dự báo trên với ảnh vệ tinh cho dự báo cập nhật thời gian thực (nowcasting).

1.5.2. Phương pháp Thống kê theo chuỗi thời gian

1.5.2.1. Mô hình Quán tính (Persistence)

Trong dự báo ĐMT ngắn hạn và rất ngắn hạn mô hình dự báo theo quán tính được sử dụng khá phổ biến do việc tính toán đơn giản, độ trễ thời gian thấp và độ chính xác chấp nhận được [17]. Bản chất của mô hình Quán tính dựa trên ý tưởng về sự lặp lại một cách đơn giản. Phương pháp này giả định rằng giá trị của bước thời gian ($t + i$) sẽ bằng với bước thời gian trước đó (t). Về mặt toán học, công thức để đưa ra dự đoán được thể hiện như sau:

$$y_{t+i} = y_t \quad (11)$$

Trong đó i là một bước thời gian tương lai tùy ý, y_t và y_{t+i} là giá trị đã biết và giá trị dự báo tại bước thời gian t và $t+i$.

Trong phương pháp thống kê, đối với phương pháp quán tính, giả định "hôm nay bằng ngày mai" được áp dụng trong phương pháp này, nghĩa là điều kiện khí hậu (ví dụ như bức xạ mặt trời) sẽ giữ nguyên trong ngày mai so với ngày hôm trước [34].

1.5.2.2. Mô hình tự hồi quy (AR - Autoregressive)

Mô hình tự hồi quy được sử dụng để dự đoán giá trị của một biến số trong tương lai dựa trên giá trị của biến số đó trong quá khứ và hiện tại. Trong mô hình tự hồi quy, biến y được dự báo bằng cách sử dụng hàm tuyến tính của các giá trị của y trong quá khứ [35]. Công thức của mô hình tự hồi quy bậc p hay còn được ký hiệu là AR(p), trong đó p là số lượng giá trị trước đó được sử dụng để dự đoán giá trị hiện tại, như sau:

$$y_t = c + \varphi_1 y_{(t-1)} + \varphi_2 y_{(t-2)} + \dots + \varphi_p y_{(t-p)} + \varepsilon_t \quad (12)$$

$$y_t = c + \sum_{i=1}^p \varphi_i y_{(t-i)} + \varepsilon_t \quad (13)$$

Trong đó y_t là giá trị của biến số trong thời điểm hiện tại, c là hệ số chặn (intercept), φ_i là các hệ số tự hồi quy (autoregressive coefficients) và ε_t là các thành phần lỗi ngẫu nhiên.

1.5.2.3. Mô hình trung bình động (MA - Moving Average)

Mô hình trung bình động (Moving Average) là một trong những mô hình thống kê dùng để dự báo xu hướng của dữ liệu chuỗi thời gian. Mô hình này dựa trên giả định rằng giá trị hiện tại của chuỗi thời gian phụ thuộc vào các giá trị trước đó của chuỗi. Mô hình trung bình động được xây dựng bằng cách tính toán trung bình cộng của các giá trị trong một khoảng thời gian cố định trước đó [35]. Công thức của mô hình trung bình trượt bậc q hay MA(q) như sau:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{(t-1)} + \theta_2 \varepsilon_{(t-2)} + \dots + \theta_p \varepsilon_{(t-p)} \quad (14)$$

$$y_t = c + \varepsilon_t + \sum_{i=1}^p \theta_i \varepsilon_{(t-i)} \quad (15)$$

Trong đó, y_t là giá trị của chuỗi thời gian tại thời điểm t , c là một hằng số, ε_t là một biến ngẫu nhiên với kỳ vọng bằng 0 và phương sai ổn định, và $\theta_1, \theta_2, \dots, \theta_q$ là các hệ số trọng số.

1.5.2.4. Mô hình ARMA

Khi kết hợp hai mô hình AR và MA lại ta có mô hình ARMA. Công thức của mô hình ARMA như sau:

$$y_{(t)} = \sum_{i=1}^p \alpha_i y_{(t-i)} + \sum_{j=1}^q \beta_j e_{(t-j)} \quad (16)$$

Trong công thức trên bao gồm hai phần, phần tự động hồi quy (AR) và phần đường trung bình (MA). Trong phương trình trên, $y_{(t)}$ là giá trị dự báo tại thời điểm t .

1.5.2.5. Mô hình ARIMA

Các tác giả trong [35] đã mang đến một mô hình dự báo mới có thể được sử dụng để giải quyết các vấn đề trong lĩnh vực thời tiết học, tài chính và công nghệ kỹ thuật xử lý lượng lớn dữ liệu được quan sát trong quá khứ. Công cụ này dựa theo phương pháp luận ARIMA (Autoregressive Integrated Moving Average - Tự hồi quy tích hợp trung bình trượt) và được biết rộng rãi dưới cái tên phương pháp luận Box-Jenkins (BJ).

Các mô hình chuỗi thời gian như công suất của nhà máy ĐMT đã trình bày được dựa vào giả thiết là các chuỗi thời gian nghiên cứu có tính dừng tức là giá trị trung bình và phương sai của chuỗi thời gian có tính dừng yếu là hằng số và đồng phương sai của nó không đổi theo thời gian. Tuy nhiên khi nhiều chuỗi thời gian không có tính dừng, tức là chúng sẽ có tính kết hợp (I - Integrated). Nếu một chuỗi thời gian là kết hợp bậc nhất [có nghĩa là nó có dạng I(1)], thì các sai phân bậc một của nó là I(0), tức là, có tính dừng. Tương tự, nếu một chuỗi thời gian là I(2), sai phân bậc hai của nó là I(0) sẽ có tính dừng. Nói chung, nếu một chuỗi thời gian là I(d), sau khi tính sai phân d lần ta có một chuỗi I(0) có tính dừng.

Do vậy, nếu phải tính sai phân một chuỗi thời gian d lần để làm cho nó có tính dừng và sau đó áp dụng mô hình ARMA(p, q), ta nói rằng chuỗi thời gian ban đầu là ARIMA(p, d, q), tức là nó là một chuỗi thời gian trung bình trượt kết hợp tự hồi quy, với p biểu thị số các số hạng tự hồi quy, d biểu thị số lần chuỗi thời gian phải được tính sai phân cho tới khi có tính dừng, và q là số các số hạng trung bình trượt.

Các mô hình thống kê chuỗi thời gian cho kết quả tốt hơn trong dự báo ngắn hạn so với các mô hình dự báo thời tiết số (NWP) [17]. Để đáp ứng nhu cầu dự báo công suất ĐMT, nghiên cứu [36] cho thấy rằng ARIMA là phương pháp chuỗi thời gian được sử dụng phổ biến nhất, chủ yếu vì nó là lựa chọn phổ biến cho các phương pháp dùng để tham chiếu.

1.5.3. Phương pháp Học máy

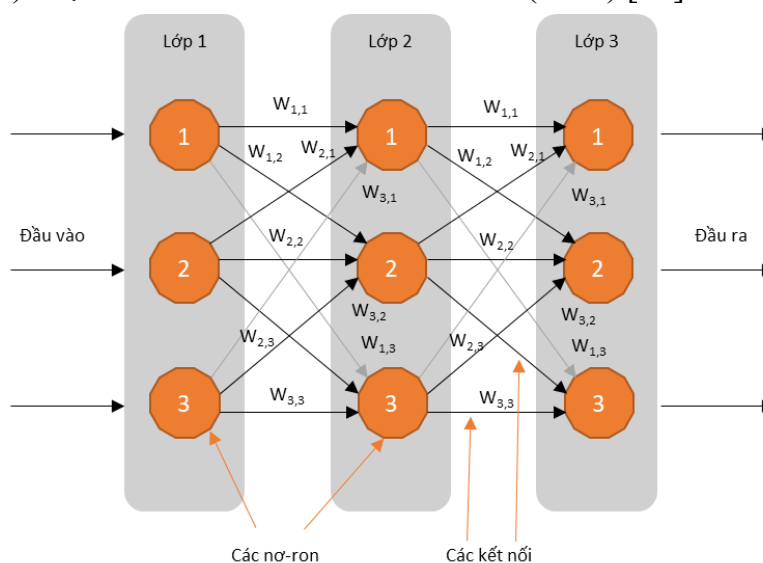
1.5.3.1. Học máy

Học máy (ML - Machine Learning) là một tập con của trí tuệ nhân tạo. Nó là một lĩnh vực trong khoa học máy tính, có khả năng tự học hỏi dựa trên dữ liệu được đưa vào mà không cần phải được lập trình cụ thể. Thay vì việc mã hóa các quy trình phần mềm bằng tay với một bộ chỉ thị cụ thể để thực hiện một nhiệm vụ cụ thể, máy được "đào

tạo" bằng cách sử dụng lượng lớn dữ liệu và các thuật toán cung cấp cho nó khả năng học hỏi và thực hiện tác vụ một cách hiệu quả. Đối với phương pháp học máy, đã có nhiều nghiên cứu sử dụng các dạng thức khác nhau của mạng nơ-ron để thực hiện dự báo ĐMT, có thể kể đến như [24], [37], [38], [39], [40], [41].

1.5.3.2. Mạng nơ-ron nhân tạo

Với ý tưởng từ mạng lưới tế bào thần kinh sinh học, mạng nơ-ron nhân tạo được cấu thành từ các nơ-ron nhân tạo với nhiều lớp, mỗi lớp kết nối với tất cả các lớp trước hoặc sau nó. Một mạng nơ-ron với nhiều hơn hai lớp (layer) còn được gọi là Multilayer Perceptrons (MLPs) hoặc Feedforward Neural Network (FNN) [42].



Hình 1.11. Mô phỏng cấu trúc của một mạng nơ-ron nhân tạo

Hình 1.11 thể hiện một mô hình mạng ba lớp, mỗi lớp có ba nơ-ron nhân tạo (hay còn gọi là các nút mạng), mỗi nút được kết nối với mọi nút khác trong lớp trước và lớp tiếp theo. Vấn đề đặt ra là cần xác định những thông số cần điều chỉnh trong quá trình huấn luyện mạng. Đầu tiên đó là điều chỉnh độ mạnh của kết nối giữa các nút hay chính là các trọng số $w_{i,j}$ (weight). Trong một nút, chúng ta có thể điều chỉnh tổng của các đầu vào, hoặc chúng ta có thể điều chỉnh hình dạng của hàm kích hoạt, nhưng điều đó phức tạp hơn so với việc chỉnh sửa độ mạnh của kết nối giữa các nút. Trọng số thấp sẽ giảm tín hiệu và trọng số cao sẽ tăng cường tín hiệu. Nghĩa là khi mạng học để cải thiện đầu ra của mạng bằng cách tinh chỉnh trọng số liên kết bên trong mạng, một số trọng số trở thành không hoặc gần bằng không. Trọng số bằng không hoặc gần bằng không có nghĩa là những liên kết đó không đóng góp gì cho mạng vì các tín hiệu không truyền qua được. Trọng số bằng không có nghĩa là các tín hiệu được nhân với số không, dẫn đến kết quả bằng không, vì vậy liên kết đó được coi như bị đứt.

Các mạng nơ-ron nhân tạo có thể có một hoặc nhiều nút mạng ở lớp cuối cùng tạo thành một tầng đầu ra (output layer), hoặc có thể thêm các lớp trung gian giữa lớp

đầu vào (input layer) và lớp đầu ra, được gọi là lớp ẩn (hidden layer). Các mạng nơ-ron thường có nhiều lớp ẩn và các lớp có thể có các hàm kích hoạt khác nhau.

Mạng nơ-ron là một lớp các thuật toán học máy phổ biến nhất hiện nay. Mạng nơ-ron được lấy cảm hứng từ hiểu biết của chúng ta về sinh học của bộ não - tất cả các kết nối giữa các nơ-ron đó. Tuy nhiên, khác với bộ não sinh học, trong đó bất kỳ nơ-ron nào cũng có thể kết nối với bất kỳ nơ-ron khác trong một khoảng cách vật lý nhất định, các mạng nơ-ron nhân tạo này được cấu trúc thành các lớp, kết nối và hướng dẫn dữ liệu riêng biệt. Mạng nơ-ron chính là nền tảng cho các mô hình Học sâu (Deep Learning – DL) phức tạp hiện nay

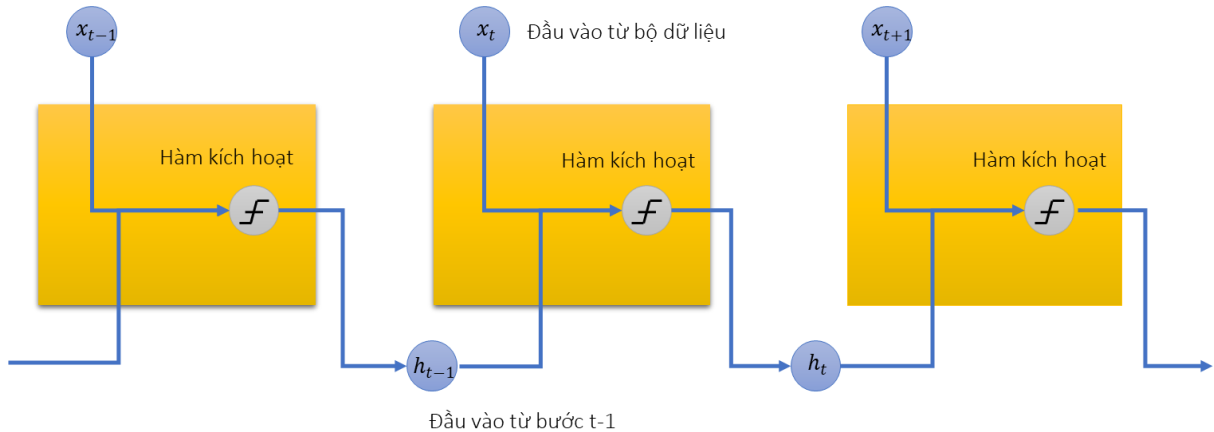
a. Mạng Nơ-ron hồi quy

Mạng nơ-ron nhân tạo (ANN - Artificial Neural Networks) đã được sử dụng thành công trong các ứng dụng dự đoán và vượt qua các mô hình thống kê cổ điển như Auto Regressive Integrated Moving Average (ARIMA) [43]. Tính phi tuyến và khả năng xử lý chuỗi thời gian mà không cần biết trước cách chuỗi đó được tạo ra là những đặc điểm nổi bật của các mạng ANN. Kiến trúc Feedforward (tức là một cấu trúc mà thông tin chảy theo một hướng duy nhất) là nền tảng của hầu hết các ứng dụng của ANN [44]. Tuy nhiên, mạng nơ-ron Feedforward (FNN) chủ yếu được sử dụng để xử lý các dữ liệu có tính chất ổn định mà không được thiết kế để xử lý các dữ liệu có tính chất biến động. Vì vậy, một kiến trúc ANN có khả năng xử lý các dữ liệu có tính chất biến động là cần thiết để đạt được mô hình hóa hiệu quả khi làm việc với các chuỗi thời gian như của nhà máy ĐMT.

Mạng nơ-ron hồi quy (RNNs - Recurrent Neural Networks) là một loại mạng nơ-ron được thiết kế với các kết nối phản hồi cho phép thông tin đi cả theo hướng ngược lại. Những kết nối này đóng vai trò là bộ nhớ trong mạng nơ-ron. Việc thiết kế thêm bộ nhớ trong mạng này cho phép RNN ghi nhớ trạng thái trước đó của nó trong quá trình xử lý, từ đó đưa ra khả năng xử lý các dữ liệu liên tục thay đổi. RNN đã được ứng dụng từ đầu thập niên 1990 trong bài toán nhận dạng chữ chỉ [45], sau đó là bài toán xử lý ngôn ngữ tự nhiên [46] và các ứng dụng chuỗi thời gian [47] trong những năm gần đây. Các RNNs thường được áp dụng cho bài toán dự báo là mạng Elman và mạng Jordan, thường được gọi là Mạng Nơ-ron Hồi Quy Đơn Giản (SRNN - Simple Recurrent Neural Networks). Khi kết hợp giữa mạng Elman và mạng Jordan ta thu được mô hình mạng với tên gọi là Mạng Nơ-ron Hồi Quy Đa Lớp (MRNN - Multi-Recurrent Neural Networks) [48].

Rất nhiều kỹ thuật đã được đề xuất để giải quyết bài toán dự báo dữ liệu dạng chuỗi thời gian trong nhiều lĩnh vực khác nhau. Trong số đó, mạng nơ-ron hồi quy như MRNN đã chứng minh được hiệu quả và được ứng dụng rất rộng rãi trong các vấn đề

liên quan đến dự báo chuỗi thời gian [48], [49]. Trong một khối RNN tiêu chuẩn của mạng nơ-ron hồi quy, cách mà thông tin được truyền đi được mô tả theo Hình 1.12. Ở mỗi khối điển hình trong chuỗi khối của mạng RNN (block), trạng thái đầu ra của bước trước trước ký hiệu là h_{t-1} và dữ liệu đầu vào của bước này được ký hiệu là x_t , hàm mô tả tín hiệu đầu ra được ký hiệu là h_t . Qua mô tả sơ đồ có thể thấy việc kết nối các thông tin phía trước để dự đoán cho hiện tại chính là đặc điểm điểm nổi bật của mạng nơ-ron hồi quy.



Hình 1.12. Các khối trong chuỗi khối của mạng nơ-ron hồi quy

Thuật toán lan truyền ngược qua thời gian (Back Propagation Through Time - BPTT) được sử dụng để huấn luyện bởi mạng RNN. Phép nhân ma trận và việc cập nhật các ô nhớ là các thao tác cơ bản mà BPTT cần thực hiện tại mỗi bước lặp. Do đó, quá trình tính toán cần rất nhiều phép nhân lặp lại của ma trận trọng số cũng như sử dụng nhiều lần đạo hàm của hàm kích hoạt (activation function). Khi đó, có một hiện tượng không mong muốn xảy ra đó là vấn đề tiêu biến đạo hàm hay “vanishing gradient” [49]. Việc huấn luyện mạng có thể sẽ không còn hiệu quả khi các đạo hàm ngày càng nhỏ hơn do thực hiện các phép nhân ma trận lặp đi lặp lại. Đây chính là điểm hạn chế của các mạng RNN thông thường và sẽ được khắc phục, cải tiến trong những mô hình mạng hiện đại hơn.

b. Mạng bộ nhớ dài ngắn

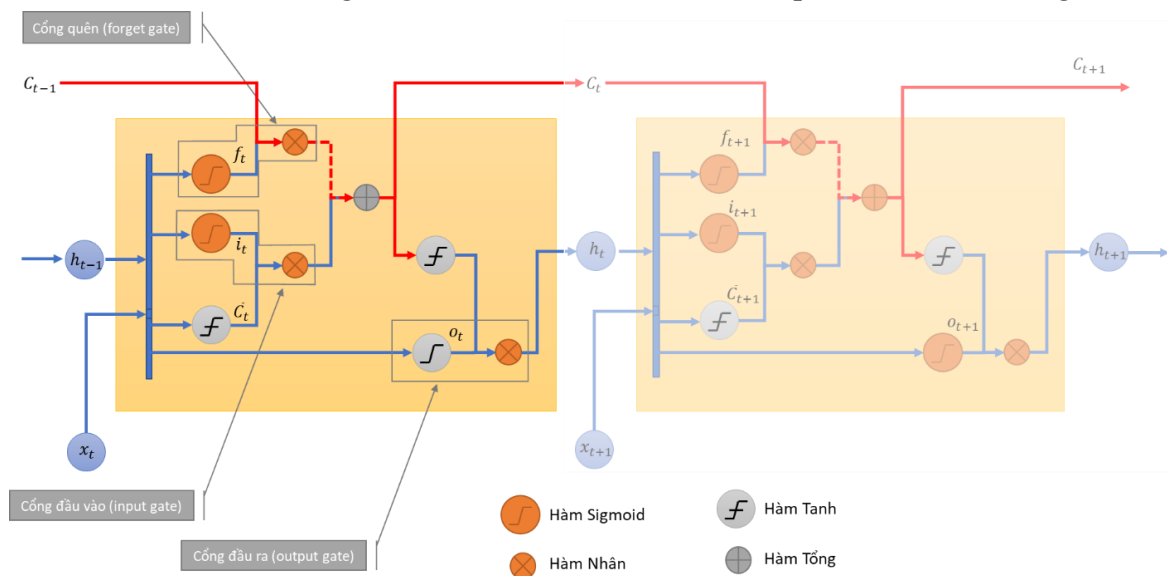
LSTM (Long Short Term Memory) là một loại RNN đặc biệt có thể học được cả sự phụ thuộc ngắn hạn cũng như dài hạn [49]. Mạng LSTM được thiết kế để giải quyết vấn đề tiêu biến đạo hàm (vanishing gradient) trong các mạng RNN truyền thống. Vấn đề này xảy ra khi giá trị đạo hàm của hàm mất mát được tính toán cập nhật liên tục qua các lớp trong mạng ngày càng trở nên rất nhỏ, khiến cho quá trình huấn luyện mạng học trở nên khó khăn và đôi khi không thể huấn luyện tiếp được.

LSTM cũng có cấu trúc giống như chuỗi RNN ở trên nhưng mô-đun lặp lại chứa trong khối LSTM phức tạp hơn. Các khối LSTM có các ô nhớ (ô trạng thái) được kết

nối qua các lớp kế tiếp nhau. Chúng cũng chứa các lớp tương tác khác nhau kiểm soát chọn lọc luồng thông tin trong các khối. Hình 1.13 mô tả cấu trúc của một chuỗi khối LSTM tiêu chuẩn.

Mạng LSTM được thiết kế để giải quyết vấn đề tiêu biến đạo hàm bằng cách sử dụng các cổng (gates) để điều chỉnh lượng thông tin được truyền qua mạng. Mỗi khối LSTM bao gồm một ô nhớ (memory cell) và các cổng. Các cổng này cho phép đơn vị LSTM quyết định lượng thông tin nào nên được giữ lại trong bộ nhớ, lượng thông tin nào nên bị loại bỏ, và lượng thông tin nào nên được truyền tiếp qua các đơn vị khác trong mạng. Cổng quên (forget gate) sẽ quyết định thông tin nào sẽ được loại bỏ khỏi ô nhớ trạng thái để quên lịch sử không liên quan vì một số thông tin đó có thể không quan trọng. Cổng đầu vào (input gate) quyết định phần nào của thông tin đầu vào mới có liên quan và lưu trữ vào ô nhớ trạng thái của nó. Cổng đầu ra (output gate) quyết định đầu ra của khối từ trạng thông tin đầu vào hiện tại và trạng thái của bộ nhớ. Mỗi cổng có cấu trúc gồm một tầng mạng sigmoid và một phép nhân nhằm mục đích sàng lọc thông tin đi qua cổng đó.

- Đầu vào cho tầng mạng sigmoid của các cổng đều gồm trạng thái nội bộ đầu ra từ bước trước đó h_{t-1} và giá trị đầu vào ở bước hiện tại x_t .
- Mỗi Đầu ra của tầng mạng Sigmoid của các cổng khác nhau sẽ thực hiện phép nhân với một thông tin khác nhau để thu được kết quả đầu ra cuối cùng, cụ thể:



Hình 1.13. Cấu trúc của một chuỗi khối LSTM

- Cổng Quên (forget gate): đầu ra của hàm Sigmoid là f_t sẽ được nhân với giá trị ô nhớ của bước trước C_{t-1} .
- Cổng Đầu vào (input gate): đầu ra của hàm Sigmoid là i_t sẽ được nhân với giá trị \bar{C}_t (là một hàm Tanh với đầu vào là h_{t-1} và x_t)

- Giá trị của ô nhớ C_t sẽ được cập nhật bằng cách cộng giá trị đầu ra của cổng Quên và cổng Đầu vào.
- Cổng Đầu ra (output gate): đầu ra của hàm Sigmoid là o_t sẽ được nhân với giá trị của ô nhớ (sau khi đã được xử lý qua hàm Tanh). Kết quả cuối cùng thu được h_t chính là trạng thái đầu ra của bước hiện tại.

Các hàm biểu diễn luồng thông tin của khối LSTM cụ thể như sau:

$$f_t = \text{sigmoid}(W_f[x_t, h_{t-1}] + b_f) \quad (17)$$

$$i_t = \text{sigmoid}(W_i[x_t, h_{t-1}] + b_i) \quad (18)$$

$$\bar{C}_t = \text{tanh}(W_c[x_t, h_{t-1}] + b_c) \quad (19)$$

$$C_t = (i_t \otimes \bar{C}_t) \oplus (f_t \otimes C_{t-1}) \quad (20)$$

$$o_t = \text{sigmoid}(W_o[x_t, h_{t-1}] + b_o) \quad (21)$$

$$h_t = o_t \otimes \text{tanh}(C_t) \quad (22)$$

Đặc điểm nổi bật của LSTM là ô nhớ trạng thái bên trong C_t cho phép dòng vận chuyển đạo hàm không bị gián đoạn theo thời gian. Việc duy trì kết nối liên tục với ô nhớ qua các khối LSTM làm giảm bớt loại trừ sự suy giảm mạnh của đạo hàm sau các phép nhân ma trận (vanishing gradient). Tổng kết lại có thể thấy, các vi phân của hàm mất mát không biến mất nhanh chóng khi chúng được lan truyền ngược theo thời gian do trạng thái ô nhớ tích lũy các hoạt động theo thời gian và các vi phân được tính theo các hàm cộng [50].

1.5.4. Phương pháp Kết hợp

Phương pháp kết hợp (Ensemble) là một phương pháp tổ hợp nhiều mô hình dự báo khác nhau để tạo ra một dự đoán chính xác hơn bằng cách kết hợp thông tin từ các mô hình riêng lẻ [51]. Có 2 hướng kết hợp có thể được thực hiện đó là Hợp tác (Cooperative) và Cạnh tranh (Competitive).

1.5.4.1. Kỹ thuật Hợp tác

Kỹ thuật này chia nhiệm vụ dự báo thành một số nhiệm vụ phụ và giải quyết từng nhiệm vụ phụ một cách riêng biệt. Kết quả dự báo tổng thể được thu được bằng cách tổng hợp các giá trị dự báo từ tất cả các bộ dự đoán. Có hai dạng chính của Kỹ thuật Hợp tác gồm tiền xử lý và hậu xử lý.

Tiền xử lý là quá trình chia tập dữ liệu đầu vào thành nhiều tập con và mỗi tập con được mô hình hóa và dự đoán bởi một bộ dự đoán (predictor). Thông thường, bộ dự đoán là giống nhau cho tất cả các tập con. Dự đoán cuối cùng là tổng của tất cả các đầu ra của các bộ dự đoán. Một số nghiên cứu gần đây đã triển khai dự báo điện mặt trời, điện gió theo hướng này như [52], [53].

Hậu xử lý là quá trình dự báo chuỗi thời gian liên tiếp bằng hai hoặc nhiều bộ dự đoán khác nhau do một tập dữ liệu chuỗi thời gian có thể có nhiều đặc điểm và mỗi đặc điểm phù hợp với một phương pháp cụ thể. Có nhiều mô hình dự báo kết hợp sử dụng

kỹ thuật hợp tác dựa trên hậu xử lý như ARIMA-GARCH , ARIMA-ANN, ARIMA-SVM và SVR-SVC [54], [55], [56].

1.5.4.2. Kỹ thuật cạnh tranh

Kỹ thuật này huấn luyện ra các bộ dự đoán khác nhau một cách riêng biệt với các tập dữ liệu khác nhau hoặc cùng một tập dữ liệu nhưng với các tham số khác nhau, sau đó dự đoán được thu được bằng cách lấy trung bình (hoặc các phép tương đương khác) của các quyết định từ tất cả các bộ dự đoán cá nhân (bộ dự đoán gốc).

Các mô hình dự báo riêng lẻ trong dự báo kết hợp có thể bao gồm các phương pháp như Random Forest, Gradient Boosting, Support Vector Regression, Neural Networks và nhiều phương pháp khác. Mỗi mô hình này có những ưu điểm và hạn chế riêng, và việc kết hợp chúng trong một bộ dự đoán có thể tận dụng được những lợi ích của từng mô hình.

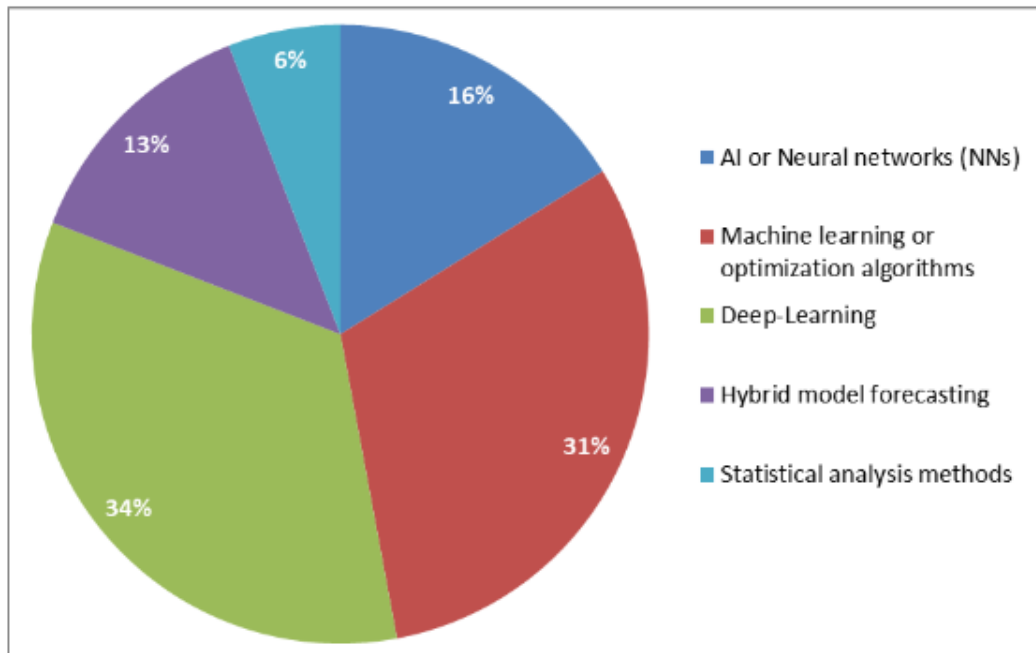
Sau khi các mô hình dự báo riêng lẻ được huấn luyện, quá trình kết hợp dự đoán được thực hiện. Có nhiều phương pháp để kết hợp các dự đoán, bao gồm trung bình, trọng số và phiếu bầu. Trong trường hợp trung bình, các dự đoán từ các mô hình riêng lẻ được lấy trung bình để tạo ra một dự đoán cuối cùng. Trong trường hợp trọng số, các mô hình được đánh trọng số khác nhau dựa trên hiệu suất của chúng trong quá khứ và dự đoán cuối cùng được tính dựa trên trọng số này. Trong trường hợp phiếu bầu, mỗi mô hình đưa ra một phiếu bầu cho mỗi dự đoán và dự đoán cuối cùng được xác định bằng cách chọn dự đoán có số phiếu cao nhất [51].

Phương pháp kết hợp cho dự báo công suất ĐMT được một số nghiên cứu cho thấy có khả năng cung cấp dự báo tốt hơn so với mô hình dự báo đơn [13], [22], [57]. Tuy nhiên, các mô hình kết hợp cũng có một số nhược điểm. Hiệu suất kém của mô hình đơn trong kiến trúc kết hợp có thể dẫn đến lỗi dự đoán cao hơn. Để khắc phục điều này, cần thiết phải thiết kế một mô hình mà mỗi bộ dự đoán không ảnh hưởng đến chất lượng của nhau. Đã có những hướng nghiên cứu giải quyết điều này như trong [58], nhóm nghiên cứu đã rút ra kết luận có thể cải thiện đầu ra dự đoán tổng thể bằng cách thực hiện dự đoán dựa trên nhiều bộ dự đoán và kết hợp đầu ra của mỗi bộ dự đoán theo cách mà các tác giả đề xuất.

1.6. Tình hình nghiên cứu

Phương pháp dự báo công suất ĐMT sử dụng học máy đã được nhiều nghiên cứu công bố là cho độ chính xác hiệu quả hơn so với các phương pháp thống kê do có thể xử lý được mối quan hệ phi tuyến phức tạp giữa đầu vào và đầu ra [59], [60]. Theo nghiên cứu [19], [61], mạng nơ-ron nhân tạo là một phương pháp học máy phổ biến để dự báo ngắn hạn công suất phát điện mặt trời. Điều này bởi vì các mạng nơ-ron có thể

thích nghi với các biến động mạnh trong mối quan hệ đầu vào-đầu ra do điều kiện môi trường thay đổi [19].



Hình 1.14: Tỷ lệ sử dụng các phương pháp trong các nghiên cứu về dự báo công suất ĐMT

Nguồn: Tsai (2023) [62]

Trong một nghiên cứu đáng chú ý năm 2023 của Tsai và cộng sự [62], 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 các nhánh ứng dụng của phương pháp học máy, mô hình mạng học sâu phổ biến nhất được sử dụng là mô hình mạng nơ-ron hồi quy bộ nhớ ngắn dài hay Long - Short Term Memory (LSTM), đã được chứng minh là hiệu quả trong việc dự báo công suất ĐMT trong ngắn hạn [63].

Đối với các nghiên cứu công bố bởi các tác giả quốc tế trên các tạp chí/hội nghị uy tín, số lượng các nghiên cứu liên quan rất lớn và đa dạng, đặc biệt từ 2019 trở lại đây, một số nghiên cứu đáng chú ý như:

Trong nghiên cứu của Wang và cộng sự năm 2019 [64] đã đề xuất mô hình dự báo CNN-LSTM phát triển từ mô hình LSTM cơ bản để tiến hành thực nghiệm dự báo cho hệ thống ĐMT có công suất 23,4 kWp (tương đương khoảng 19,5 kWac). Các tác giả kết luận rằng khi có đủ dữ liệu lớn, LSTM vượt trội hơn so với các phương pháp truyền thống khác. Kết quả thu được là mô hình LSTM 2 lớp kết hợp với 1 lớp mạng CNN (mô hình LSTM-CNN) để dự báo cho bước hiện tại sử dụng dữ liệu khí tượng hiện tại kết hợp với các dữ liệu khí tượng quá khứ và cho sai số đạt được là MAE 0,221

kW và RMSE là 0,621 (tính toán trong cả ngày 24 giờ). Khi chuẩn hóa một cách tương đối để so sánh trong luận án cho khung thời gian có nắng từ 5h-18h (13 giờ), giá trị chuẩn hóa MAPE là 2,092%, nRMSE là 5,879%.

Trong nghiên cứu của Ospina và cộng sự năm 2019 [65] mạng LSTM được sử dụng để dự báo công suất cho một hệ thống điện mặt trời có công suất 12,6 MW ở Florida, Mỹ. Các nhà nghiên cứu đã sử dụng bộ dữ liệu có độ dài 2 năm với độ phân giải 30 phút gồm 35,089 mẫu dữ liệu để huấn luyện và kiểm tra mô hình. Mô hình thu được dự báo cho 1 bước 30 phút tiếp theo có sai số MAPE là 3,555% và nRMSE là 6,600% (tính toán cho cả ngày 24 giờ). Khi chuẩn hóa một cách tương đối để so sánh trong luận án cho khung thời gian có nắng từ 5h-18h (13 giờ), giá trị chuẩn hóa MAPE là 6,563%, nRMSE là 12,184%.

Trong nghiên cứu của Zhou H và cộng sự năm 2019 [66], nhóm tác giả đã phát triển mô hình mạng LSTM để dự báo công suất cho hệ thống ĐMT có công suất 20 kW tại tỉnh Zhejiang, Trung Quốc với bộ dữ liệu thu thập trong 4 năm từ 2014 đến 2018 với độ phân giải dữ liệu 7,5 phút. Kết quả là mô hình thu được có sai số đã quy đổi trong các giờ ban ngày theo các bước dự báo như sau: dự báo 1 bước MAPE 4,0%, nRMSE 6,950 %; dự báo 2 bước MAPE 5,0%, nRMSE 8,0%; dự báo 4 bước MAPE 6,1%, nRMSE 9,05%; dự báo 8 bước MAPE 7,35%, nRMSE 10,45% (trong đó MAPE được quy đổi từ kết quả MAE trong nghiên cứu).

Trong nghiên cứu của Wen và cộng sự công bố năm 2019 [67], nhóm tác giả đã xây dựng mô hình dự báo công suất hệ thống ĐMT tại Trung Quốc có công suất 106,6 kW với bộ dữ liệu từ 1/1/2018 đến 1/2/2018. Với mô hình đề xuất dựa trên LSTM, nhóm tác giả đã đạt được sai số MAE là 4,369 kW và RMSE là 7,536 kW (tính trong 24h). Khi chuẩn hóa một cách tương đối để so sánh trong luận án cho khung thời gian có nắng từ 5h-18h (13 giờ), giá trị chuẩn hóa MAPE là 7,566% và nRMSE là 13,051%.

Trong nghiên cứu của Harrou và cộng sự năm 2020 [68], các tác giả đã thực hiện nghiên cứu trên bộ dữ liệu độ phân giải 15 phút của một nhà máy ĐMT có công suất 9 MWp với thời gian là 1 năm từ 1/2018 đến tháng 12/2018 với mô hình LSTM 1 lớp ẩn với 200 nút ẩn trên lớp ẩn. Kết quả thu được của mô hình sau khi huấn luyện 90% dữ liệu và thử nghiệm 10% dữ liệu cho sai số MAPE là 8,93%.

Trong nghiên cứu của Zhou N và cộng sự công bố năm 2020 [69], các tác giả sử dụng bộ dữ liệu của một nhà máy ĐMT 5,83 kWp (khoảng 4,85 kWAC) đặt tại Úc có độ dài 1 năm trong 2017 gồm dữ liệu công suất và khí tượng. Kết quả thu được mô hình LSTM dự báo cho 1 ngày tới cho sai số tốt nhất với MAE là 0,3063 kW, RMSE là 0,5283 kW tính cho cả ngày 24 giờ. Khi chuẩn hóa một cách tương đối để so sánh cho

khung có nắng từ 5-18h (13 giờ), giá trị chuẩn hóa MAPE là 11,639%, nRMSE là 20,075%.

Trong nghiên cứu của Zhang và cộng sự công bố năm 2020 [70], các tác giả sử dụng bộ dữ liệu của một nhà máy ĐMT 6,41 kW đặt tại California, Hoa Kỳ có độ dài từ 1/7/2015-31/12/2016. Kết quả thu được mô hình LSTM cho sai số tốt nhất với MAPE là 4,56% và nRMSE 8,39% (tính cho cả ngày 24 giờ). Sau khi chuẩn hóa tương đối để so sánh cho khung có nắng từ 5-18h (13 giờ), giá trị chuẩn hóa MAPE là 8,418%, nRMSE là 15,489%.

Trong nghiên cứu của Park và cộng sự năm 2021 [71], nhóm tác giả đã phát triển mô hình dự báo công suất dựa trên mô hình LSTM với 1 lớp ẩn và 3 lớp ẩn cho bộ dữ liệu của một hệ thống điện mặt trời với quy mô khoảng hơn 500 kWp trong thời gian 7 tháng đầu năm 2019. Kết quả thu được cho thấy mô hình 3 lớp ẩn có sai số nRMSE tốt nhất là 13,2%.

Trong nghiên cứu của Li và cộng sự năm 2021 [72], các tác giả đã nghiên cứu bộ dữ liệu của một hệ thống ĐMT với công suất cực đại khoảng 40 kW (theo hình vẽ trong bài báo nghiên cứu) từ 1/2018-12/2018 với các dữ liệu công suất và khí tượng như nhiệt độ, độ ẩm, tốc độ gió... Với các mô hình thử nghiệm khác nhau, mô hình cho kết quả tốt nhất là mô hình LSTM có RMSE là 2,560 kW ứng với nRMSE (quy đổi theo phương pháp của luận án) là 6,4% trong các giờ ban ngày.

Trong nghiên cứu của Zhou H và cộng sự công bố năm 2021 [73], các tác giả đã thực hiện nghiên cứu cho hệ thống ĐMT 20 kW tại thành phố Shaoxing, miền đông Trung Quốc. Bộ dữ liệu trong vòng 4 năm từ 2014 đến 2018 gồm công suất và các thông số khí tượng với độ phân giải 7,5 phút. Mô hình LSTM mà nhóm tác giả đề xuất được cải tiến nâng cấp so với mô hình đã công bố tại nghiên cứu [66] có sai số dự báo tốt nhất trong các mô hình so sánh với giá trị chuẩn hóa MAPE đạt được là 3,5% và nRMSE là 6,5% cho dự báo 1 bước tới, MAPE 4,25% và nRMSE 7,0% cho dự báo 2 bước tới, MAPE 6,9% và nRMSE 10,2% cho dự báo 4 bước tới.

Trong nghiên cứu của Liu và cộng sự công bố năm 2021 [74], nhóm tác giả đã sử dụng dữ liệu có độ phân giải 15 phút trong khoảng 09 tháng từ hệ thống ĐMT 18,78 kW tại Chiang Rai (Thái Lan) để thực nghiệm xây dựng mô hình dự báo công suất phát. Tác giả xây dựng mô hình. Sau quá trình huấn luyện mô hình dự báo cho tốt nhất dựa trên LSTM thu được có RMSE là 0,497 kW (tính cho cả ngày 24h). Sau khi chuẩn hóa tương đối để so sánh cho khung có nắng từ 5-18h (13 giờ), giá trị chuẩn hóa nRMSE là 4,886%.

Trong nghiên cứu của Pombo và cộng sự năm 2022 [75], các tác giả đã thực hiện thực nghiệm xây dựng một số mô hình dự báo khác nhau cho bộ dữ liệu của một hệ

thống ĐMT 10 kW đặt trong khuôn viên trường Đại học Kỹ thuật Đan Mạch (DTU - Danish Technical University). Bộ dữ liệu độ phân giải 05 phút, dài 15 tháng từ tháng 1/2019 đến tháng 12/2020 dữ liệu gồm dữ liệu công suất và dữ liệu đo các thông số khí tượng như bức xạ, nhiệt độ.... Kết quả cho thấy sai số MAPE tốt nhất đạt được cho dự báo 60 bước tới (5 tiếng) là khoảng 14,990%.

Trong nghiên cứu của Suresh và cộng sự năm 2022 [76], nhóm tác giả đã phát triển mô hình dự báo công suất dựa trên mô hình LSTM cho bộ dữ liệu độ phân giải 60 phút của một hệ thống ĐMT với quy mô khoảng hơn 317 kWp (khoảng 265 kWac) tại Ba Lan. Kết quả thu được cho thấy mô hình dự báo 1 chu kỳ tới có sai số RMSE là 15,59 kW và MAE là 8,36 kW. Sau khi chuẩn hóa tương đối để so sánh cho khung có nắng từ 5-18h (13 giờ), giá trị chuẩn hóa MAPE là 4,8% và nRMSE là 9,046%.

Đối với các nghiên cứu công bố bởi các tác giả trong nước trên các tạp chí/hội nghị uy tín, số lượng bài nghiên cứu còn chưa nhiều, mới bắt đầu xuất hiện trong thời gian gần đây, có thể kể đến như:

Trong nghiên cứu của nhóm tác giả Nguyễn Đức Tuyên và cộng sự công bố năm 2020 [77], các tác giả đã xây dựng mô hình dự báo trên cơ sở số liệu thu thập từ hệ thống ĐMT tại Kentucky Hoa Kỳ trong năm 2006, công suất đặt của hệ thống ước tính khoảng 78 MW, độ phân giải dữ liệu là 60 phút. Kết quả thu được mô hình dự báo 1 ngày tới dựa trên LSTM có sai số tốt nhất với RMSE là 4,03 MW và MAE là 2,92 MW, tương ứng với giá trị chuẩn hóa nRMSE là 5,167% và MAPE là 3,743%. Lưu ý rằng, dữ liệu công suất dùng để so sánh với giá trị dự báo là giá trị cực đại lý thuyết được nhóm tác giả tính toán từ các giá trị khí tượng.

Trong nghiên cứu của nhóm tác giả Nguyễn Thị Hoài Thu và cộng sự công bố năm 2022 [78], các tác giả đã thực nghiệm với bộ số liệu trong vòng 6 tháng từ 1/2022 đến 6/2022 của các hệ thống ĐMT mái nhà tại khu vực Bảo Hà, Lâm Đồng, của Việt Nam. Kết quả đạt được là mô hình đề xuất của nhóm tác giả dựa trên mạng LSTM có RMSE đạt 2,641 MW và MAE đạt 1,896 MW, tính cho cả ngày 24 giờ. Mặc dù nhóm tác giả không công bố chính xác giá trị công suất đặt của bộ số liệu tuy nhiên qua biểu đồ hình vẽ có thể xác định tương đối giá trị này khoảng 60 MW. Sau khi chuẩn hóa tương đối để so sánh cho khung có nắng từ 5-18h (13 giờ), giá trị chuẩn hóa MAPE là 4,8% và nRMSE là 9,046%. Như vậy giá trị ước tính nRMSE là 8,127% và MAPE là 5,834%. Lưu ý rằng, đây là nghiên cứu dành cho dự báo ĐMT mái nhà trên một khu vực địa lý rộng và dữ liệu công suất tổng sẽ có tính chất của hiệu ứng “smoothing effect”.

Bảng 1.4 dưới đây trình bày tổng hợp các kết quả nghiên cứu của các mô hình dự báo công suất ĐMT có sử dụng LSTM và bộ dữ liệu quá khứ từ các hệ thống

DMT/nhà máy ĐMT, đã được công bố trong thời gian gần đây từ 2019 đến nay từ các tác giả quốc tế và trong nước.

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

TT	Mô hình	Năm	Quy mô nhà máy	Độ phân giải	Độ dài dự báo	Số bước dự báo	Khung dự báo	MAPE (%)	nRMSE (%)
1	Wang và cộng sự	2019	23,4 kWp (19,5 kW)	5 phút	t	1	5 phút	2,092	5,879
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+288	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 đến t+4	4	30 phút	6,900	10,200
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

Lưu ý, trong tổng hợp chỉ xét đến các nghiên cứu dành cho dự báo của các hệ thống riêng tại vị trí xác định (dự báo điểm) không xét đến các nghiên cứu dành cho ĐMT mái nhà của một khu vực do không cùng đối tượng nghiên cứu của luận án.

1.7. Kết luận

Đối với mỗi phương pháp, có những lợi điểm và hạn chế riêng. Đối với phương pháp vật lý, do tính chất phức tạp của bài toán vật lý khí tượng toàn cầu hoặc khu vực mà phương pháp này thường chỉ được triển khai ở các cơ quan khí tượng lớn trên thế

giới. Phương pháp thống kê chuỗi thời gian gặp khó khăn trong bài toán dự báo công suất ĐMT khi đối mặt với các biến động ngẫu nhiên và không thể mô phỏng được các mô hình phi tuyến. Như đã phân tích về các yếu tố ảnh hưởng đến công suất phát của nhà máy ĐMT hầu hết đều mang tính chất phi tuyến từ các yếu tố khí tượng đến các yếu tố lắp đặt, điều khiển, vận hành hệ thống do đó việc ứng dụng dự báo ĐMT sẽ có những hạn chế nhất định. Phương pháp kết hợp các phương pháp trên để tận dụng ưu điểm của từng phương pháp riêng biệt. Tuy nhiên, việc kết hợp các phương pháp yêu cầu quá trình phân tích thử nghiệm kỹ lưỡng để chọn mô hình đơn và phương pháp kết hợp do sai số từng thành phần sẽ ảnh hưởng rất lớn để việc tổng hợp sai số.

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 mạng 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

Nội dung này tập trung vào việc xây dựng một mô hình dự báo công suất ngắn hạn của nhà máy ĐMT sử dụng Mạng nơ-ron hồi quy bộ nhớ dài ngắn (LSTM) trên các bộ dữ liệu thực tế. Mục tiêu của nghiên cứu này là xác định tính khả thi và hiệu quả của mô hình LSTM trong việc dự báo công suất ĐMT trong ngắn hạn.

Mạng LSTM là một biến thể của mạng nơ-ron hồi quy, được thiết kế đặc biệt để xử lý và mô hình hóa dữ liệu chuỗi thời gian. LSTM có khả năng lưu trữ thông tin lâu dài trong bộ nhớ dài hạn và điều chỉnh luồng thông tin thông qua các cổng. Điều này giúp nó khắc phục được nhược điểm của mạng nơ-ron hồi quy truyền thống và cải thiện khả năng dự báo.

Nghiên cứu sẽ thực hiện việc xây dựng mô hình LSTM dự báo công suất ĐMT bằng cách sử dụng các dữ liệu quá khứ công suất nhà máy và các yếu tố tác động như bức xạ, nhiệt độ, độ ẩm.... Bằng cách phân tích các yếu tố liên quan, luận án sẽ tiến hành thử nghiệm trên một số cấu hình LSTM khác nhau. Sau khi huấn luyện, số liệu dự báo trên tập dữ liệu kiểm tra của các mô hình LSTM sẽ được đánh giá so sánh dự báo với dữ liệu thực tế. Các chỉ số đánh giá sai số được sử dụng để đánh giá hiệu quả của mô hình so với các mô hình thông dụng khác.

Ngoài ra, trong phần này cũng tiến hành so sánh hiệu quả sai số đạt được của mô hình trên hai chỉ số thông dụng nhất được nhiều nghiên cứu khác lựa chọn là MAPE và nRMSE với mô hình dựa trên LSTM của các nghiên cứu trong và ngoài nước đã công bố trong khoảng từ năm 2019 trở lại đây để có sự đối chiếu hiệu quả đạt được.

2.2. Thu thập số liệu

2.2.1. Số liệu khí tượng thực tế

Trong các công trình lắp đặt vừa và lớn (>1 MW), sai số đo đạc có tác động đáng kể đến việc đánh giá hiệu quả vận hành của nhà máy. Sai số vài phần trăm đối với một nhà máy có công suất đặt lên đến hàng trăm MW dẫn đến là có sự khác biệt đáng kể trong số liệu dự báo công suất đầu ra nhà máy và trực tiếp ảnh hưởng đến vận hành của nhà máy. Các phép đo GHI rất quan trọng để xác định hiệu suất làm việc của nhà máy, đặc biệt trong bối cảnh nhiều nhà máy đang bị giới hạn công suất như hiện nay. Đối với các dự án quy mô lớn, ở khu vực nơi điều kiện khí tượng có thể thay đổi do sự khác biệt về khí hậu. Trong những trường hợp này, các trạm theo dõi bức xạ và trạm thời tiết có thể cần được đặt ở nhiều vị trí. Việc sử dụng hai hoặc nhiều vị trí theo dõi bức xạ đảm

bảo tính dự phòng của các thiết bị đo để khi một số thiết bị cần được thay thế hoặc hỏng hóc, quá trình thu thập dữ liệu sẽ vẫn diễn ra liên tục. Ngoài ra, các thông số khí tượng khác liên quan đến nhà máy, chẳng hạn như nhiệt độ không khí, độ ẩm, lượng mưa, tốc độ và hướng gió cần được theo dõi bởi một trạm thời tiết chuyên dụng.

2.2.2. Số liệu công suất phát

Các số liệu công suất phát là các số liệu làm việc của nhà máy thông thường được ghi nhận qua hệ thống SCADA/DCS của nhà máy. Các tín hiệu công suất phát này thường có độ phân giải rất cao (1 giây – 2 giây) một tín hiệu và sẽ được truyền về phòng điều khiển của nhà máy.

2.2.3. Bộ dữ liệu phục vụ xây dựng mô hình

Bộ dữ liệu này được thu thập từ một nhà máy điện mặt trời tại khu vực Nam Bộ của Việt Nam. Khu vực này rất thuận lợi cho việc phát triển năng lượng mặt trời nhờ vào số giờ nắng trung bình hàng năm tương đối cao, từ 2.200-2.500 giờ/năm. Nhà máy có công suất đặt 48 MWAC ứng với sản lượng điện trung bình hàng năm là khoảng 104,8 GWh.

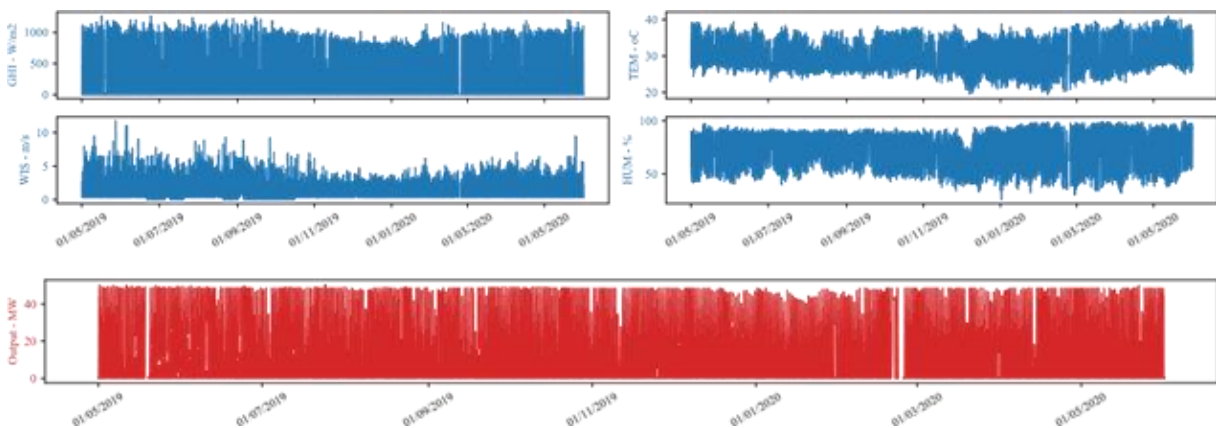
Nhà máy điện có 2 trạm đo các thông số khí tượng như bức xạ mặt trời, nhiệt độ môi trường, tốc độ gió và độ ẩm. Sau khi được thu thập từ các cảm biến, dữ liệu này cùng với công suất đầu ra của nhà máy điện được gửi về trung tâm điều độ thông qua hệ thống SCADA. Đối với mục đích vận hành, dữ liệu này đã được xử lý thành định dạng dữ liệu với độ phân giải 05 phút. Điều này có nghĩa là mỗi điểm dữ liệu có một chỉ dấu thời gian đại diện cho trung bình của tất cả các điểm dữ liệu SCADA trong khoảng thời gian 5 phút kể từ thời điểm của nhãn chỉ dấu. Ví dụ, dữ liệu nhiệt độ ngày 1/5/2019 0:00 là 28.5°C, có nghĩa là trung bình các giá trị ghi nhận bởi hệ thống SCADA từ 1/5/2019 0:00 đến 1/5/2019 0:05 là 28.5°C. Bảng 2.1 mô tả bộ dữ liệu lịch sử của nhà máy từ tháng 5 năm 2019 đến tháng 6 năm 2020 (13 tháng). Ý nghĩa của mỗi cột dữ liệu được mô tả như sau:

- GHI: (W/m^2) bức xạ toàn phần là tổng lượng bức xạ ngắn của ánh sáng nhận được từ trên bề mặt nằm ngang so với mặt đất.
- Nhiệt độ (TEM): ($^{\circ}C$) Nhiệt độ môi trường tại vị trí nhà máy ĐMT.
- Tốc độ gió (WIS): (m/s) Tốc độ gió tại vị trí nhà máy ĐMT.
- Độ ẩm (HUM): (%) Độ ẩm tương đối tại vị trí nhà máy ĐMT.
- Công suất (Output): (MW) Công suất đầu ra của nhà máy ĐMT.

Bảng 2.1. Dữ liệu lịch sử của nhà máy

STT	Thời gian	GHI (W/m^2)	Nhiệt độ ($^{\circ}C$)	Tốc độ gió (m/s)	Độ ẩm (%)	Công suất (MW)
1	5/1/2019 0:00	0	28,5	1,85	82,34	0

STT	Thời gian	GHI (W/m ²)	Nhiệt độ (°C)	Tốc độ gió (m/s)	Độ ẩm (%)	Công suất (MW)
2	5/1/2019 0:05	0	28,39	1,85	82,54	0
3	5/1/2019 0:10	0	28,39	1,85	82,75	0
4	5/1/2019 0:15	0	28,39	1,09	83,08	0
5	5/1/2019 0:20	0	28,39	1,85	83,08	0
...
114333	5/31/2020 23:40	0	27,42	0,33	94,46	0
114334	5/31/2020 23:45	0	27,42	0,33	94,57	0
114335	5/31/2020 23:50	0	27,5	0,33	94,69	0
114336	5/31/2020 23:55	0	27,52	0,33	94,6	0



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

2.3. Môi trường thí nghiệm

Các thử nghiệm được thực hiện với môi trường chi tiết như sau:

- Cấu hình của máy tính là CPU Intel Core i7-6700 3,4GHz, RAM 20GB, 64bit-Windows 10.
- Các phần mềm sử dụng:
 - Python phiên bản 3.7.5: Python là một ngôn ngữ lập trình mạnh mẽ và linh hoạt, được sử dụng rộng rãi trong lĩnh vực học máy. Điểm mạnh của Python nằm ở cú pháp đơn giản, dễ đọc và dễ hiểu, giúp người dùng nhanh chóng triển khai các thuật toán và mô hình học máy một cách dễ dàng. Bên cạnh đó, Python cũng hỗ trợ nhiều thư viện mạnh mẽ và đa dạng để phục vụ các nhu cầu phát triển học máy.
 - Visual Studio Code: là một môi trường phát triển tích hợp (IDE) mã nguồn mở và miễn phí, được phát triển bởi Microsoft. Nó là một trong những công cụ phổ biến nhất được sử dụng cho phát triển phần mềm và có nhiều ứng dụng trong lĩnh vực học máy đặc biệt cho ngôn ngữ lập trình python.

- TensorFlow 2.0: TensorFlow là một thư viện mã nguồn mở được phát triển bởi Google. Với TensorFlow, người dùng có thể xây dựng và huấn luyện các mạng nơ-ron. Thư viện TensorFlow cung cấp các lớp và hàm tối ưu hóa cho việc xử lý dữ liệu, trích xuất đặc trưng, huấn luyện mô hình và đánh giá hiệu suất.
- Scikit-learn (với hai thư viện chính là Pandas và NumPy): Pandas và NumPy là hai thư viện quan trọng trong ngôn ngữ lập trình Python, được sử dụng rộng rãi trong các ứng dụng học máy và học sâu. Cả hai thư viện đều cung cấp các cấu trúc dữ liệu và công cụ mạnh mẽ để xử lý và phân tích dữ liệu, giúp người dùng thực hiện các thao tác tính toán và biến đổi dữ liệu dễ dàng.
- Matplotlib và Seaborn: Matplotlib là một thư viện trực quan hóa dữ liệu mạnh mẽ trong Python. Nó cho phép người dùng tạo ra các biểu đồ như đường cong, cột, scatter plot và histogram để trực quan hóa dữ liệu. Matplotlib cung cấp nhiều tùy chọn tùy chỉnh cho việc điều chỉnh màu sắc, kiểu đồ thị, nhãn và tiêu đề, giúp người dùng tạo ra biểu đồ chuyên nghiệp. Matplotlib cũng tích hợp tốt với các thư viện khác như NumPy và Pandas, giúp người dùng trực quan hóa dữ liệu dễ dàng. Seaborn là một thư viện trực quan hóa dữ liệu được xây dựng trên Matplotlib, nhằm cung cấp một giao diện cao cấp và tùy chỉnh hơn cho việc trực quan hóa dữ liệu. Seaborn cung cấp các biểu đồ và kiểu trực quan hóa cao cấp như heatmaps, violin plots, pair plots và joint plots, giúp người dùng khám phá và hiểu rõ hơn về mối quan hệ giữa các biến trong dữ liệu. Seaborn cũng cung cấp các chủ đề màu sắc hỗ trợ và các chức năng tùy chỉnh khác để tạo ra các biểu đồ có ngoại hình đẹp và chuyên nghiệp.
- PVLIB (Photovoltaic Library): là một thư viện phần mềm mã nguồn mở được phát triển cho Python, nhằm hỗ trợ phân tích và mô phỏng hệ thống điện mặt trời. PVLIB cung cấp các công cụ và chức năng để tính toán liên quan đến điện mặt trời, giúp nghiên cứu và ứng dụng trong lĩnh vực học máy và bài toán dự báo điện mặt trời.

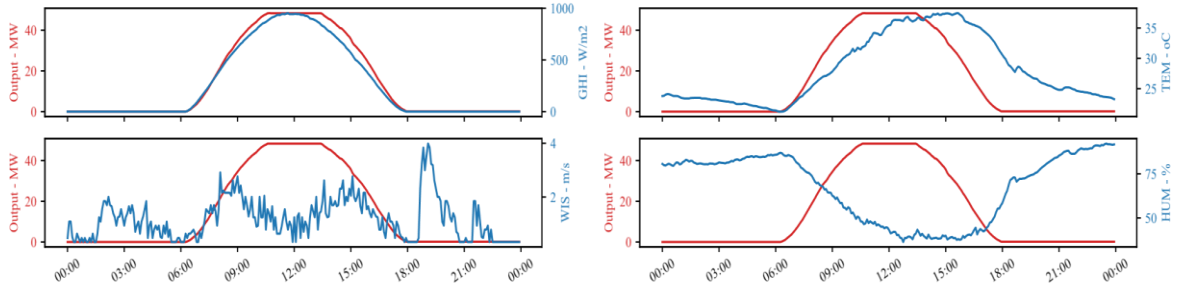
2.4. Tiền xử lý số liệu

Sau khi thu thập được các dữ liệu thô từ các nhà máy ĐMT phục vụ công việc xây dựng mô hình dự báo, giai đoạn tiếp theo là tiền xử lý số liệu. Giai đoạn tiền xử lý dữ liệu được đánh giá là một trong những khâu rất quan trọng trong quá trình khai thác dữ liệu để xây dựng mô hình dự báo.

2.4.1. Trích xuất các đặc trưng quan trọng từ bộ dữ liệu

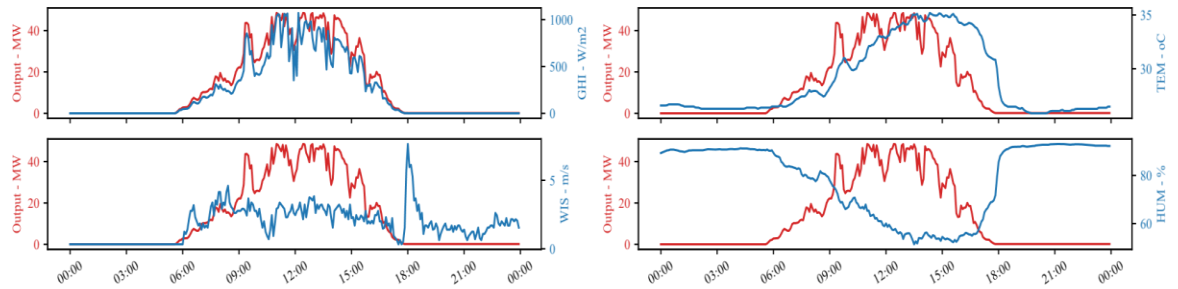
Trong bài toán dự báo công suất điện mặt trời, dữ liệu được trích xuất để xây dựng mô hình bao gồm công suất đầu ra thực tế và dữ liệu khí tượng thực tế. Ngoài ra, dữ liệu tính liên quan đến nhà máy điện mặt trời, như công suất lắp đặt và vị trí địa lý, cũng được sử dụng trong việc tính toán các biến đầu vào. Tuy nhiên việc lựa chọn dữ liệu cần các kỹ thuật phân tích để xác định các dữ liệu phù hợp.

Các kỹ thuật phân tích dữ liệu sẽ được sử dụng để nghiên cứu về mối tương quan giữa dữ liệu đầu vào và dữ liệu đầu ra. Khi nghiên cứu bộ dữ liệu, ta nhận thấy rằng công suất đầu ra của nhà máy bị ảnh hưởng mạnh bởi các yếu tố thời tiết.



Hình 2.2. Ngày có công suất phát đầu ra ổn định

Hình 2.2 và Hình 2.3 cho thấy mối tương quan giữa các yếu tố thời tiết khác nhau (màu xanh) và công suất đầu ra (màu đỏ) trong ngày điều kiện phát ổn định và không ổn định.



Hình 2.3. Ngày có công suất phát đầu ra biến động

Để kiểm tra mối quan hệ thống kê hoặc tương quan giữa chuỗi dữ liệu thời tiết và công suất đầu ra, hệ số tương quan Pearson (r) với công thức tổng quát như sau được sử dụng:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (20)$$

Trong đó:

- n : Là số mẫu
- x_i, y_i : Là giá trị của x và y ứng với chỉ số i
- \bar{x} : Là giá trị trung bình của x

\bar{y} : Là giá trị trung bình của y

Hệ số tương quan Pearson có giá trị dao động trong khoảng từ -1 đến +1.

- $r = 0$: Hai biến không có tương quan tuyến tính.
- $r = 1$; $r = -1$: Hai biến có tương quan tuyến tính tuyệt đối.
- $r < 0$: Hệ số tương quan âm.
- $r > 0$: Hệ số tương quan dương.

Bảng 2.2 cho thấy kết quả tính toán hệ số tương quan Pearson giữa công suất đầu ra và các yếu tố thời tiết tại cùng thời điểm. Với kết quả này, có thể nhận thấy rằng đối với công suất đầu ra của nhà máy ĐMT và GHI có một mối tương quan dương mạnh ($r = 0,982249$), nhiệt độ và độ ẩm tương quan mạnh mặc dù độ ẩm có tương quan âm. Tốc độ gió có mối tương quan dương nhưng ở mức thấp ($r = 0,379356$).

Bảng 2.2. Kết quả kiểm định hệ số tương quan Pearson

Thông số	GHI (Bức xạ)	TEM (Nhiệt độ)	WIS (Tốc độ gió)	HUM (Độ ẩm)
Hệ số Pearson (r)	0,982249	0,750528	0,379356	-0,762010

2.4.2. Làm sạch dữ liệu

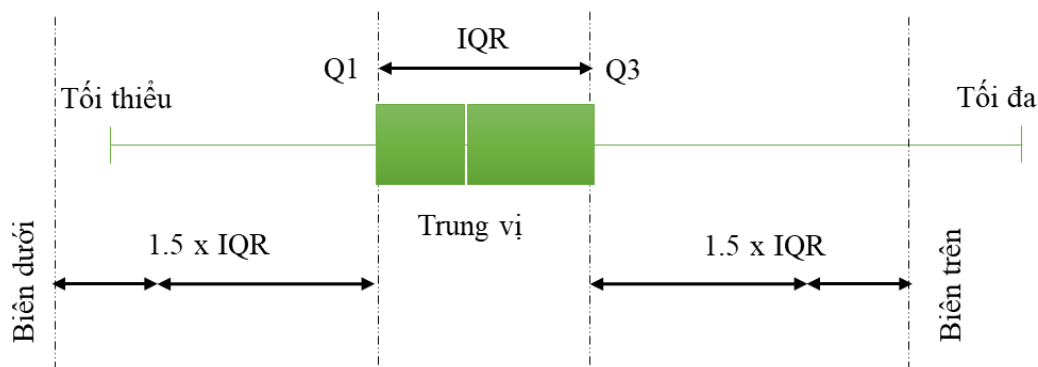
Các tập dữ liệu quá khứ phục vụ dự báo ĐMT thường có các yếu tố gián đoạn tạm thời (thiếu) hoặc những tín hiệu sụt giảm gây ra bởi biến động thời tiết hoặc theo mùa, và các sự cố lưới điện. Cụ thể, trong điều kiện ánh sáng yếu xuất hiện vào buổi sáng sớm hoặc chiều muộn, có thể xảy ra hiện tượng số liệu bức xạ mặt trời âm và giá trị công suất đi kèm bị thiếu. Nguyên nhân là do cảm biến bức xạ mặt trời hoạt động kém tin cậy và inverter chưa hoạt động. Một trường hợp khác là việc dữ liệu về bức xạ mặt trời, nhiệt độ và công suất đầu ra bị thiếu trong giữa ngày có thể bị lỗi cảm biến bức xạ mặt trời và nhiệt độ, và sự cố của inverter hoặc lưới điện. Đây là các giá trị ngoại lai không có xu hướng, bị ảnh hưởng bởi các sự kiện ngẫu nhiên và có ảnh hưởng đáng kể đến dự báo. Hơn nữa, dữ liệu đôi khi có thể bị sai hoặc thiếu do lỗi cảm biến hoặc đường truyền tín hiệu. Do đó, việc tiền xử lý dữ liệu đầu vào sai lỗi bằng cách phân tách, nội suy hoặc điều chỉnh theo mùa (tức là làm sạch dữ liệu và thay đổi cấu trúc) là rất quan trọng [79].

Theo các nghiên cứu gần đây, kỹ thuật khoảng tứ phân vị (Interquartile Range - IQR) thường 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 [80]. Đây là kỹ thuật thường được ứng dụng nhiều trong các bài toán liên quan đến lĩnh vực năng lượng [17],

[81], [82], [83], [84], [85]. Phạm vi biến thiên của dữ liệu trong bộ dữ liệu có thể được mô tả như trong Hình 2.4, trong đó:

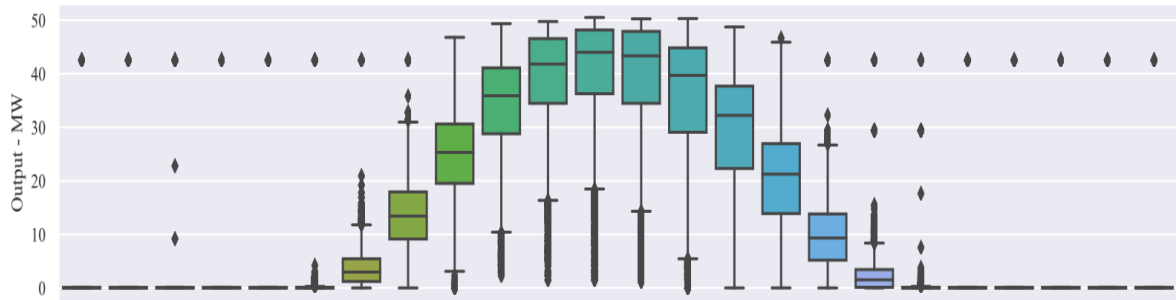
- Trung vị (Median) là điểm trung tâm của tập dữ liệu chia tập dữ liệu thành hai nửa, dưới và trên;
- Q1: là tứ phân đầu tiên của tập dữ liệu, có nghĩa là 25% điểm dữ liệu nằm giữa Q1 và giá trị tối thiểu;
- Q3: là tứ phân thứ ba của tập dữ liệu, có nghĩa là 25% điểm dữ liệu nằm giữa Q3 và giá trị tối đa;
- Khoảng tứ phân (IQR) là sự chênh lệch giữa tứ phân đầu tiên (Q1) và thứ ba (Q3) được định nghĩa theo công thức: $IQR = Q3 - Q1$
- Biên dưới (Lower Bound) có giá trị là $(Q1 - 1.5 * IQR)$
- Biên trên (Upper Bound) có giá trị là $(Q3 + 1.5 * IQR)$
- **Điểm ngoại lai** là bất kỳ điểm dữ liệu nào nằm ngoài phạm vi được định nghĩa bởi Biên dưới và Biên trên.

Sau đó, các điểm dữ liệu lỗi được thay thế hoặc hiệu chỉnh bằng các phương pháp sao cho tập dữ liệu kết quả có đủ kích thước để phục vụ việc huấn luyện. Điều này đặc biệt có ý nghĩa đối với việc xây dựng mô hình cho các nhà máy điện mặt trời mới phải bắt đầu hoạt động khi dữ liệu quá khứ là ngắn và không ổn định.

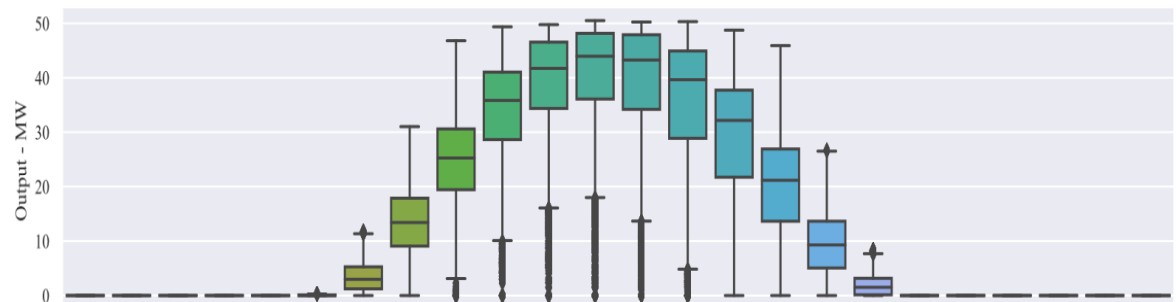


Hình 2.4. Phương pháp IQR xác định điểm ngoại lai

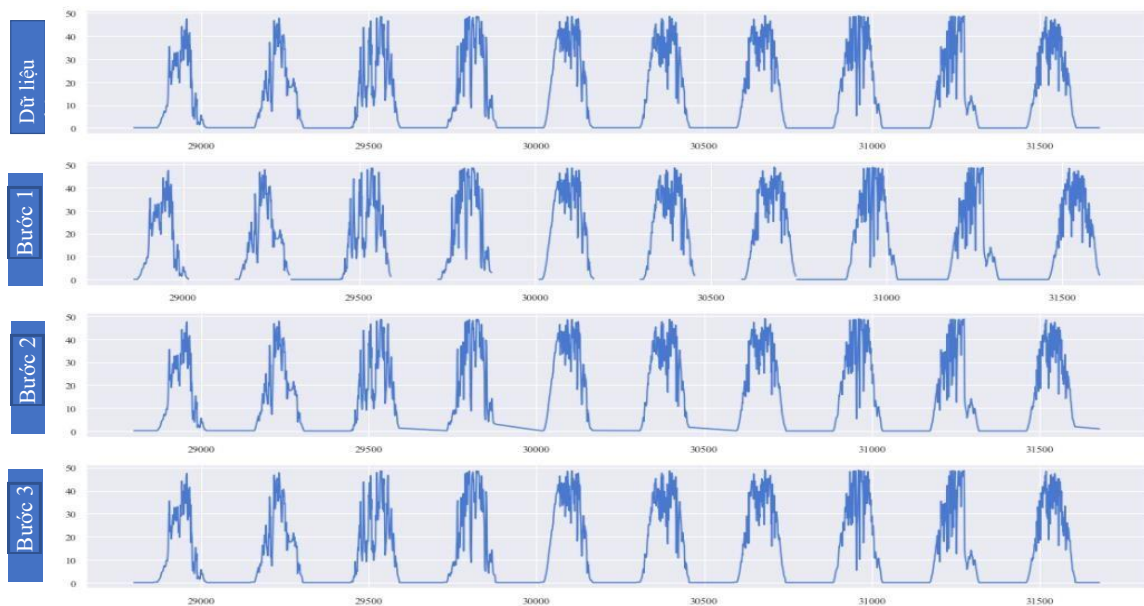
Ứng dụng cụ thể cho bộ dữ liệu đang xem xét, trong mỗi giờ trong một ngày, dữ liệu như công suất, có một phạm vi biến thiên nhất định. Phạm vi biến thiên cho mỗi giờ có thể được mô tả như trong Hình 2.5, sau quá trình áp dụng thuật toán IQR, các điểm ngoại lai được loại bỏ và dữ liệu được biểu diễn như Hình 2.6.



Hình 2.5. Phân phối của công suất đầu ra trước khi loại bỏ các điểm ngoại lai.



Hình 2.6. Phân phối của công suất đầu ra sau khi loại bỏ các điểm ngoại lai.



Hình 2.7. Quá trình xử lý dữ liệu công suất phát trong khoảng thời gian 10 ngày

Hình 2.7 biểu diễn quá trình xử lý dữ liệu hoàn chỉnh dựa trên việc áp dụng IQR trong một khoảng thời gian 10 ngày.

- Trong Bước 1, dữ liệu gốc được xử lý và lọc bằng IQR như đã mô tả ở trên.
- 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.

Sau quá trình xử lý, thu được bộ dữ liệu với tổng số 114.336 mẫu dữ liệu.

2.4.3. Phân chia dữ liệu huấn luyện và kiểm tra

Sau khi làm sạch dữ liệu, tập huấn luyện và tập kiểm tra cho thí nghiệm sẽ được xác định. Dữ liệu đầu vào là dữ liệu chỉ dấu thời gian theo Lịch (Ngày trong năm, Giờ trong ngày, Phút trong giờ) và dữ liệu khí tượng (Bức xạ mặt trời - GHI, Nhiệt độ - TEM, Tốc độ gió - WIS, Độ ẩm - HUM). Dữ liệu đầu ra là các điểm dữ liệu công suất đầu ra (Output).

Với 114.336 mẫu dữ liệu, tập dữ liệu được chia thành 2 phần: tập huấn luyện và tập kiểm tra. Phần đầu được sử dụng để huấn luyện mô hình và phần còn lại được coi là những mẫu dữ liệu chưa biết để đo lường độ chính xác của mô hình. Độ dài của tập huấn luyện là 365 ngày để phản ánh quy luật trong cả năm. Độ dài của tập kiểm tra là khoảng 1 tháng. Hình 2.8 minh họa cách chia dữ liệu thành tập huấn luyện và tập kiểm tra từ tập dữ liệu gốc.



Hình 2.8. Phân chia dữ liệu Huấn luyện - Kiểm tra

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

2.5.1. Mô hình quán tính

Đối với bài toán dự báo công suất điện mặt trời, tính lặp lại của các ngày là rõ ràng. Do đó, với 288 chu kỳ 05 phút ứng với 01 ngày, công thức cho phương pháp dự báo này như sau:

$$y(t + 288) = y(t)$$

2.5.2. Mô hình ARIMA

Để đáp ứng nhu cầu dự báo công suất ĐMT, nghiên cứu [36] cho thấy rằng ARIMA là phương pháp chuỗi thời gian được sử dụng phổ biến nhất, chủ yếu vì nó là lựa chọn phổ biến cho các phương pháp dùng để tham chiếu.

Để xác định hệ số của mô hình ARIMA(p,d,q), có thể sử dụng phương pháp ACF và PACF như sau:

Bước 1: Kiểm định Augmented Dickey-Fuller (ADF). Bảng 2.3. Kết quả kiểm định ADF cho thấy rằng chuỗi dữ liệu gốc của công suất đầu ra là dừng và không cần thực hiện việc chuyển đổi chuỗi thời gian ban đầu bằng việc lấy vi phân. Như vậy, giá trị d của mô hình ARIMA được thiết lập là 0.

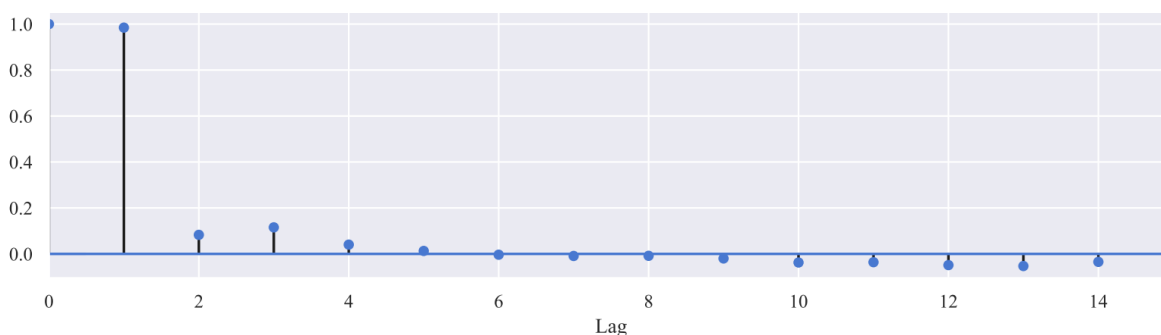
Bảng 2.3. Kết quả kiểm định ADF

Mục	Giá trị
ADF Statistic	-40,26710939
n_lags	0
p-value	0
Critical Values: 1%	-3,4304120801
Critical Values: 5%	-2,8615674384
Critical Values: 10%	-2,5667846045

Bước 2: Vẽ đồ thị hàm tự tương quan (ACF) của chuỗi dữ liệu đã được chuyển đổi sai phân.

Bước 3: Dựa trên đồ thị ACF, có thể xác định giá trị của p là số lượng lớn nhất của lag trong ACF mà giá trị tương quan có ý nghĩa thống kê (tức là vượt qua đường chân không).

Hình 2.9 cho thấy đồ thị ACF trên tập dữ liệu. Khoảng trễ từ 1 đến 5 có tương quan lớn nhất với giá trị hiện tại. Điều đó có nghĩa là hệ số p của mô hình ARIMA thay đổi trong khoảng [1, 2, 3, 4, 5].



Hình 2.9. Đồ thị ACF trên tập dữ liệu công suất phát đầu ra

Bậc q của mô hình ARIMA cũng được thiết lập là 0 theo nghiên cứu [86] đã được công bố.

Để xác định chính xác giá trị của p, AIC (Akaike Information Criteria) [87] được kiểm tra khi p thay đổi trong khoảng trên [88].

Công thức của AIC được biểu diễn như sau:

$$AIC = 2p - 2 \ln(L)$$

Trong đó:

- p : tham số p của mô hình ARIMA
- L : giá trị tối đa của hàm hợp lý của mô hình.

Mô hình có điểm AIC thấp hơn được kỳ vọng có sự cân đối tốt hơn giữa khả năng phù hợp với tập dữ liệu và khả năng tránh việc phù hợp quá mức với tập dữ liệu.

Bảng 2.4. Kết quả kiểm định AIC

ARIMA(p,d,q)	AIC
(1,0,0)	509122,9
(2,0,0)	508405,2
(3,0,0)	506841,4
(4,0,0)	506680,8
(5,0,0)	506681,1

Kết quả được hiển thị trong Bảng 2.4. Kết quả kiểm định AIC. Dựa trên kết quả này, giá trị phù hợp của p là 4 vì nó có AIC thấp nhất.

2.5.3. Mô hình MLP

Từ nghiên cứu tổng hợp [19], phương pháp phổ biến để dự báo công suất nhà máy ĐMT trong ngắn hạn là mạng nơ-ron nhân tạo (ANN). ANN là một mạng lưới gồm các nơ-ron nhân tạo được kết nối trong các lớp khác nhau: lớp đầu vào, lớp ẩn và lớp đầu ra. Một mạng nơ-ron có hơn hai lớp còn được gọi là mạng nơ-ron nhiều lớp hay mạng Perceptron nhiều lớp (ký hiệu là MLP). Luận án sử dụng mô hình MLP từ gói Scikit-learn [89] để triển khai các thí nghiệm. Các thông số mặc định của mô hình này là như sau:

- Số lớp mạng biến thiên từ 1 đến 4.
- Số nơ-ron trong mỗi lớp ẩn: 100.
- Bộ solver cho quá trình huấn luyện: ADAM, là thuật toán tối ưu hóa dựa trên gradient ngẫu nhiên do Diederik P. Kingma và Jimmy Ba đề xuất. Theo [90], bộ solver 'ADAM' hoạt động tốt trên các tập dữ liệu có kích thước lớn (với hàng nghìn mẫu huấn luyện trở lên) về cả thời gian huấn luyện và kết quả.
- Hàm kích hoạt: 'relu'. ReLU đã được chứng minh giúp huấn luyện mạng nơ-ron nhiều lớp và mạng sâu (có nhiều lớp ẩn) nhanh hơn so với hàm tanh [91].

Vì mục đích so sánh, các mô hình với số lượng lớp ẩn từ 1 đến 4 đã được kiểm tra. Mỗi đầu vào là một mảng phẳng (mảng 1D) gồm 7 phần tử tương ứng với 7 đặc trưng trên đã nêu.

2.5.4. Xây dựng mô hình LSTM

Tương tự như MLP, việc chọn các đặc trưng đầu vào, tối ưu hóa cấu trúc mạng và siêu tham số cũng rất quan trọng trong mô hình LSTM. Dựa trên phân tích hệ số

tương quan trong Bảng 2.2, dữ liệu thời tiết bao gồm GHI, TEM, WIS, HUM sẽ được sử dụng trong quá trình huấn luyện như các đầu vào. Ngoài ra, các yếu tố chỉ dấu thời gian cũng được bao gồm trong quá trình huấn luyện: ngày trong năm, giờ trong ngày và phút trong giờ.

Mỗi đầu vào để huấn luyện hoặc kiểm tra mạng LSTM là một ma trận 2 chiều với số cột c là số lượng đặc trưng được chọn như đã nêu (7 đặc trưng). Số hàng r trong ma trận là số lượng điểm dữ liệu trước đó có thể ảnh hưởng đến công suất dự báo. Mặc dù khả năng học các phụ thuộc xa của LSTM là tương đối mạnh, nhưng việc lựa chọn giá trị r quá lớn sẽ ảnh hưởng đến thời gian và quy mô tính toán.

Luận án sẽ sử dụng tham số p của mô hình ARIMA ở trên để lựa chọn cài đặt phù hợp cho ma trận đầu vào. Với $p = 4$ có nghĩa là ngoài giá trị khí tượng tại điểm dự báo, khoảng thời gian có khả năng ảnh hưởng đến công suất hiện tại là 20 phút trước đó. Do đó, mỗi đầu vào sẽ là một ma trận có kích thước 5 hàng x 7 cột, trong đó 5 hàng thể hiện cho 5 điểm dữ liệu $t, t-1, t-2, t-3, t-4$ để tiến hành dự báo công suất cho bước t .

Để kiểm tra mức độ ảnh hưởng của các dữ liệu với các độ trễ khác nhau đến dữ liệu công suất phát, luận án tính toán hệ số tương quan Pearson theo các độ trễ. Kết quả như bảng sau. Từ đó có thể thấy rằng ở các độ trễ đến $t-4$ với các bước dữ liệu 05 phút, mức độ tương quan chưa thay đổi nhiều và vẫn có ảnh hưởng tương đối lớn đến dữ liệu công suất phát đầu ra.

Bảng 2.5. Hệ số tương quan Pearson của các thông số khí tượng và công suất tại các bước trễ khác nhau

Bước trễ	GHI (Bức xạ)	TEM (Nhiệt độ)	WIS (Tốc độ gió)	HUM (Độ ẩm)
t	0,982249	0,750528	0,379356	-0,762010
t-1	0,968379	0,764664	0,386954	-0,77615
t-2	0,955569	0,776043	0,393618	-0,78766
t-3	0,946792	0,785243	0,39906	-0,79742
t-4	0,938187	0,793383	0,404118	-0,80600

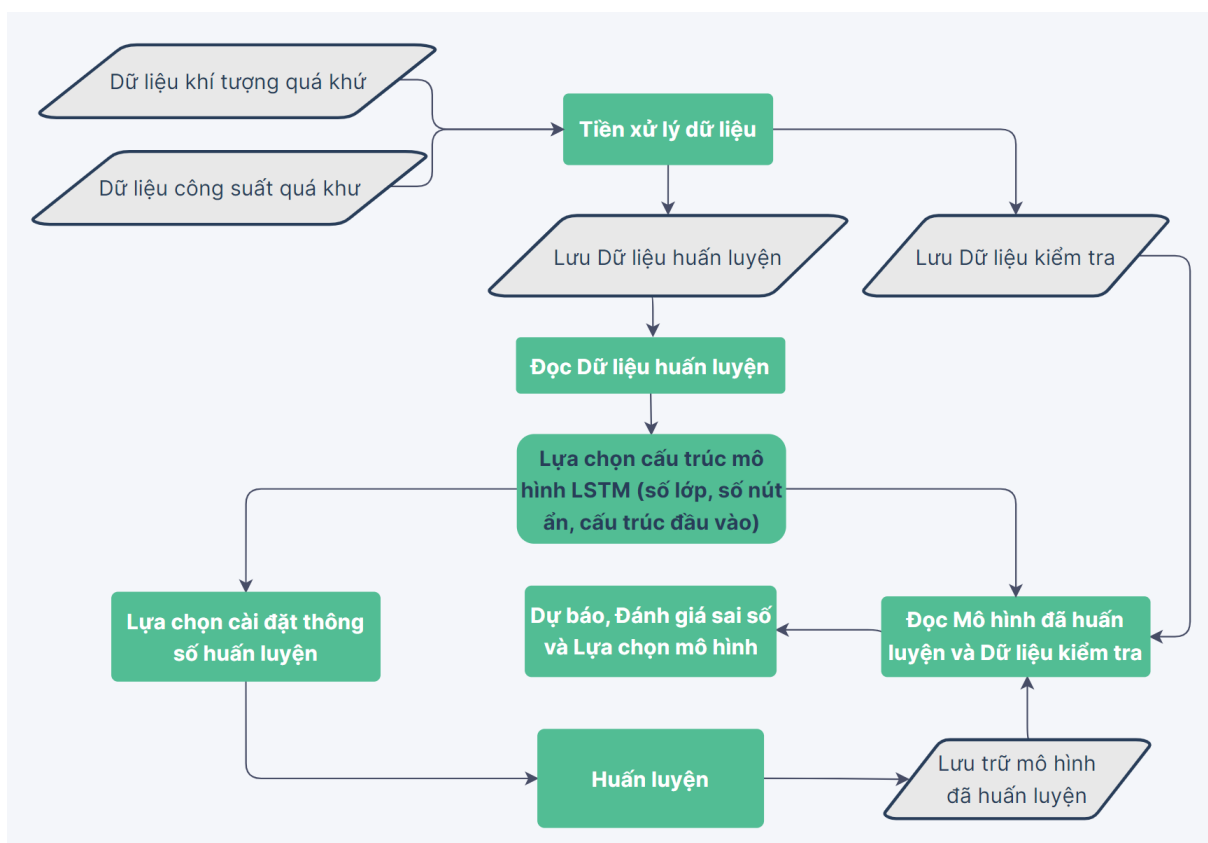
Để lựa chọn các tham số tối ưu cho mô hình, cấu hình của mạng LSTM và thiết lập cho giai đoạn huấn luyện được lựa chọn như sau:

- 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.
- Bộ solver cho quá trình huấn luyện: ADAM, là thuật toán tối ưu hóa dựa trên gradient ngẫu nhiên do Diederik P. Kingma và Jimmy Ba đề xuất. Theo [90], bộ solver 'ADAM' hoạt động tốt trên các tập dữ liệu có kích thước lớn (với hàng nghìn mẫu huấn luyện trở lên) về cả thời gian huấn luyện và kết quả.

- Hàm kích hoạt: 'relu'. ReLU đã được chứng minh giúp huấn luyện mạng nơ-ron nhiều lớp và mạng sâu (có nhiều lớp ẩn) nhanh hơn so với hàm tanh [91].
- Các thiết lập huấn luyện khác bao gồm
 - o Hàm mất mát: Hàm sai số trung bình tuyệt đối (MAE),
 - o Số epoch huấn luyện: 50 epoch.

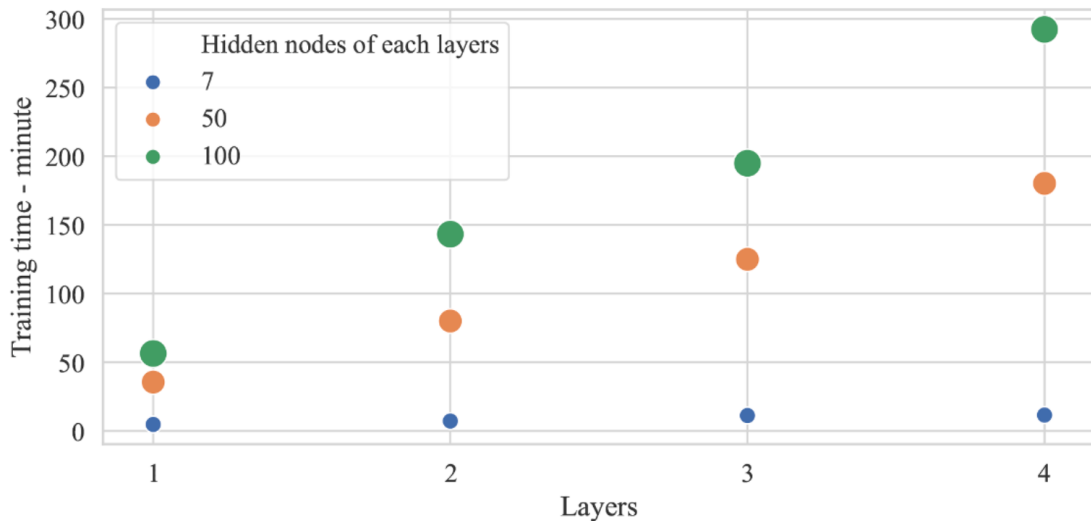
2.6. Huấn luyện mô hình LSTM

Hình 2.10 mô tả trình tự thực hiện thí nghiệm nghiên cứu đối với việc xây dựng mô hình dự báo sử dụng mạng LSTM.

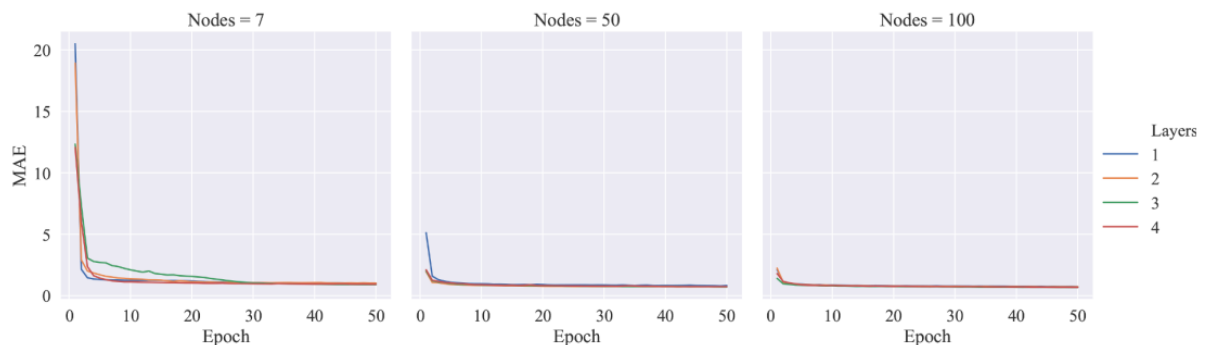


Hình 2.10. Sơ đồ thí nghiệm xây dựng mô hình dự báo

Hình 2.11 và Hình 2.12 mô tả thời gian huấn luyện và giá trị hàm mất mát trong quá trình huấn luyện các mô hình LSTM có cấu trúc khác nhau. Thời gian huấn luyện tỉ lệ thuận trực tiếp với độ phức tạp của mạng. Kết quả huấn luyện cho thấy khi mỗi lớp ẩn có một số lượng nút ẩn nhỏ, thời gian huấn luyện vẫn tương đối ngắn (dưới 15 phút) khi tăng số lớp ẩn từ 1 đến 4 lớp. Tuy nhiên, khi tăng số nút ẩn trên mỗi lớp lên thành 50 hoặc 100 nút, thời gian huấn luyện tăng đáng kể. Đối với mạng phức tạp nhất với 4 lớp ẩn và 100 nút ẩn trên mỗi lớp, thời gian huấn luyện lên đến gần 300 phút (5 giờ). Sau khi mô hình đã được huấn luyện tốt, trong thời gian thực nó có thể được sử dụng để dự đoán mà không cần huấn luyện lại, và thời gian dự báo chỉ khoảng 20 giây.



Hình 2.11. So sánh thời gian huấn luyện của các mô hình



Hình 2.12. Giá trị hàm mất mát (MAE) trong quá trình huấn luyện

Chỉ trong trường hợp dự báo giảm đáng kể trong quá trình đánh giá dài hạn (tuần, tháng), mới cần huấn luyện lại mô hình trên toàn bộ tập dữ liệu lớn. Qua hình mô tả giá trị hàm mất mát trong quá trình huấn luyện, với mạng 50 nút hoặc 100 nút, giá trị của hàm mất mát đã ổn định sau khoảng 10 epoch huấn luyện. Tại 50 epoch, sự cải thiện trong giá trị hàm mất mát giữa các mô hình khác nhau là không đáng kể.

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

Sau khi huấn luyện các mô hình trên tập dữ liệu huấn luyện, tập dữ liệu kiểm tra chứa dữ liệu thời tiết thực tế trong tháng 5 năm 2020 thu thập từ khu vực nhà máy được sử dụng để kiểm tra chất lượng của các mô hình. Bảng 2.6 mô tả kết quả của việc áp dụng các mô hình huấn luyện khác nhau để dự đoán trên tập dữ liệu kiểm tra.

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

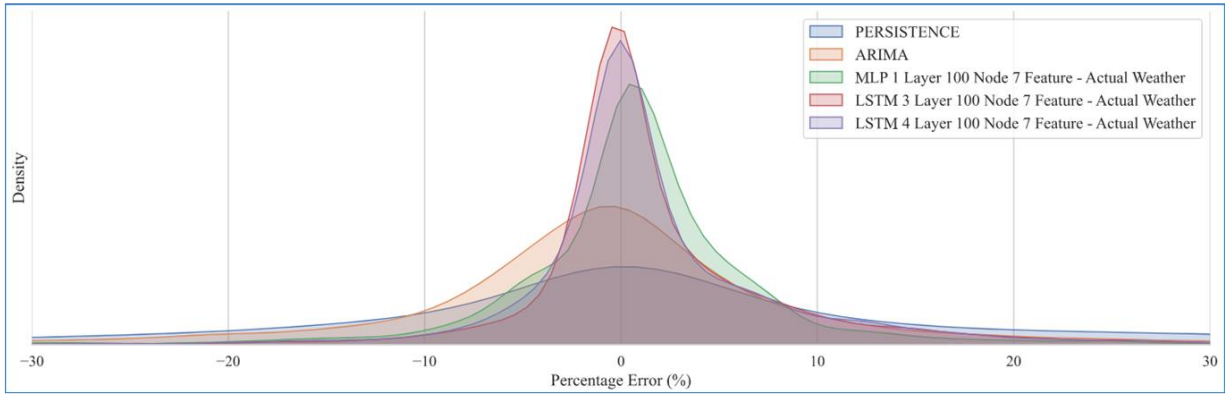
Kết quả kiểm tra trên tập dữ liệu kiểm tra cho thấy mô hình quán tính cho kết quả lỗi tương đối lớn; RMSE là 11,325 MW và MAE là 7,401 MW, tương ứng với MAPE là 15,418%.

Sử dụng mô hình ARIMA cải thiện đáng kể kết quả trên tất cả các độ đo lỗi. Việc sử dụng ARIMA giảm RMSE một cách đáng kể còn 5,410 MW và MAPE 7,329%.

Khi áp dụng mô hình MLP 1L-100N với một lớp ẩn như đã mô tả ở trên, RMSE cải thiện đáng kể giảm còn 3,470 MW và MAPE cải thiện giảm còn 4,346%. Tăng số lượng lớp ẩn không cải thiện kết quả lỗi cho các mô hình MLP.

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).

Hình 2.13 mô tả phân phối của chỉ số sai số tương đối (PE) của các điểm dữ liệu trên tập dữ liệu kiểm tra với mạng LSTM-4L100N cũng như các mô hình khác để so sánh.



Hình 2.13. Phân bố sai số tương đối trên tập dữ liệu kiểm tra của các mô hình

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

Trong quá trình huấn luyện một mô hình máy học, bộ dữ liệu kiểm chứng (validation dataset) có thể được sử dụng để đánh giá hiệu suất của mô hình sau mỗi vòng lặp huấn luyện. Bộ dữ liệu kiểm chứng tăng cường khả năng tổng quát hóa của mô hình trên dữ liệu mới, mà nó chưa từng thấy trong quá trình huấn luyện, qua đó làm giảm đi tình trạng quá khớp “over-fit” của mô hình. Khi huấn luyện, dữ liệu huấn luyện được sử dụng để điều chỉnh các trọng số và siêu tham số của mô hình. Sau mỗi vòng lặp huấn luyện, mô hình được đánh giá trên bộ dữ liệu kiểm chứng để đo lường hiệu suất. Điều này giúp theo dõi sự tiến bộ của mô hình và đánh giá các điểm yếu cần cải thiện.

Với bộ dữ liệu sử dụng đã trình bày ở trên, trong lần huấn luyện này, luận án sử dụng tỷ lệ kiểm chứng 10% (trích 10% dữ liệu trong tập huấn luyện để làm kiểm chứng trong quá trình huấn luyện). Đồng thời, huấn luyện sử dụng thêm kỹ thuật dừng sớm (early stopping) với cài đặt patient là 10 bước lặp để giảm thiểu thời gian huấn luyện. Cụ thể, nếu thấy xu hướng chất lượng mô hình qua từng bước lặp trên tập kiểm chứng đã đạt mức tốt nhất, nếu trong 10 bước lặp tiếp theo mà không tiếp tục cải thiện thì việc huấn luyện sẽ dừng lại. Mô hình được chọn là mô hình có sai giá trị hàm mất mát trên tập kiểm chứng là nhỏ nhất.

Bảng 2.7. 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
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 không có kiểm chứng.

- RMSE giảm từ 3,082 MW xuống còn 2,166 MW, cải thiện khoảng 29,7% so với huấn luyện không có kiểm chứng.

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

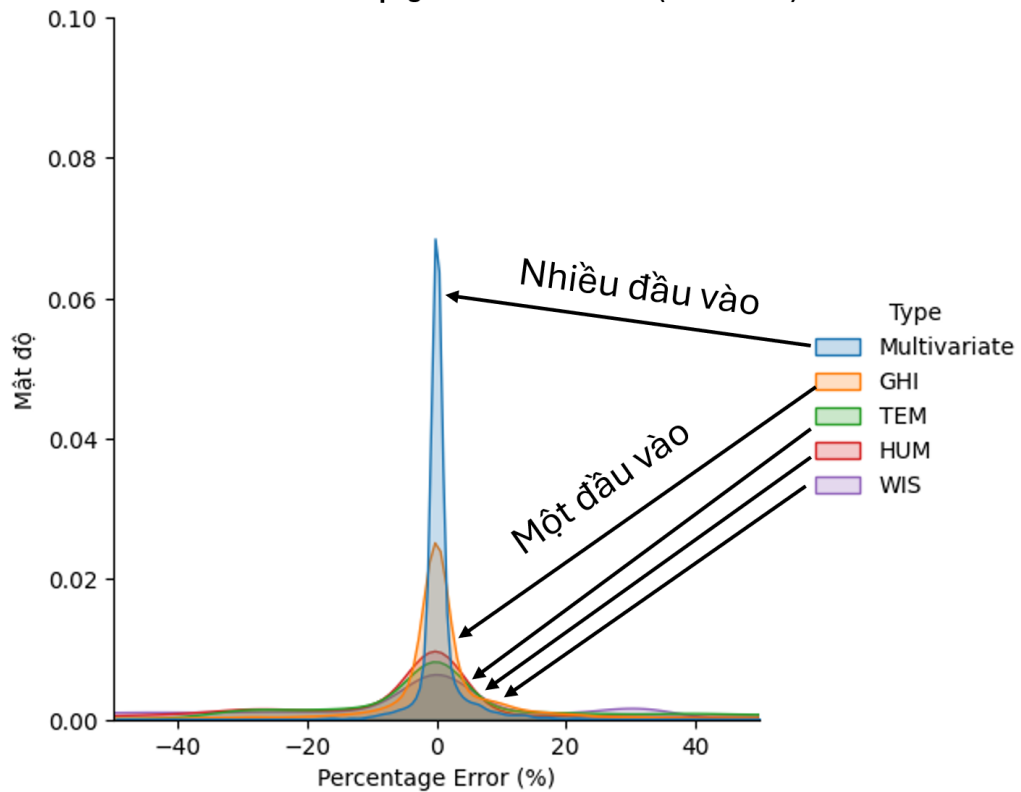
Mô hình đã được xây dựng tại Chương 2 này là mô hình đa biến (multivariate) khi sử dụng kết hợp nhiều đầu vào khác nhau để dự báo công suất. Để có sự đánh giá đầy đủ hơn về mô hình, luận án cũng tiến hành thực nghiệm và so sánh kết quả từ mô hình này so với mô hình sử dụng đầu vào đơn biến (univariate). Các mô hình đầu vào đơn biến được tiến hành thí nghiệm gồm mô hình chỉ có đầu vào bức xạ (GHI), nhiệt độ (TEM), độ ẩm (HUM) và tốc độ gió (WIS). Nghiên cứu tiếp tục sử dụng mô hình LSTM 4 lớp, 100 nút ẩn trên mỗi lớp đã thu được trước đó cùng cấu trúc độ dài đầu vào như Mục 2.5.4 để huấn luyện và dự báo các bộ đầu vào đơn biến này. Bảng 2.8 thể hiện kết quả đánh giá các chỉ số sai số nhận được khi sử dụng mô hình đầu vào đơn biến so với kết quả đã đạt được trong mô hình sử dụng đầu vào đa biến trước đó. Trong các mô hình đơn biến, mô hình sử dụng chỉ một đầu vào bức xạ (GHI) cho kết quả tốt nhất với RMSE đạt 7,568 MW và MAPE đạt 10,192 MW. Các mô hình sử dụng dữ liệu đầu vào đơn biến nhiệt độ, độ ẩm, tốc độ gió thể hiện kết quả kém hơn so với mô hình sử dụng đầu vào bức xạ tương đối lớn với MAPE thu được lần lượt là 22,281%, 28,106%, 35,585%.

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

So với kết quả đạt được từ mô hình sử dụng kết hợp nhiều đầu vào đã huấn luyện ở các phần trước, sai số của mô hình sử dụng đầu vào đơn biến là lớn hơn. Kết quả này cũng được phản ánh thông qua biểu đồ phân phối sai số tỷ lệ (PE – percentage error) của các mô hình trên cùng tập dữ liệu kiểm tra. Mức độ tập trung sai số nhỏ của các mô hình đơn biến quanh mốc 0% khá thấp trong khi xuất hiện nhiều sai số lớn khiến biểu đồ có xu hướng bè ra hai bên mốc 0% như hình Hình 2.14.

So sánh sai số trên tập dữ liệu kiểm tra của mô hình sử dụng đầu vào đa biến (multivariate) và mô hình sử dụng đầu vào đơn biến (univariate)



Hình 2.14. So sánh sai số mô hình sử dụng đầu vào đa biến và mô hình sử dụng đầu vào đơn biến

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.

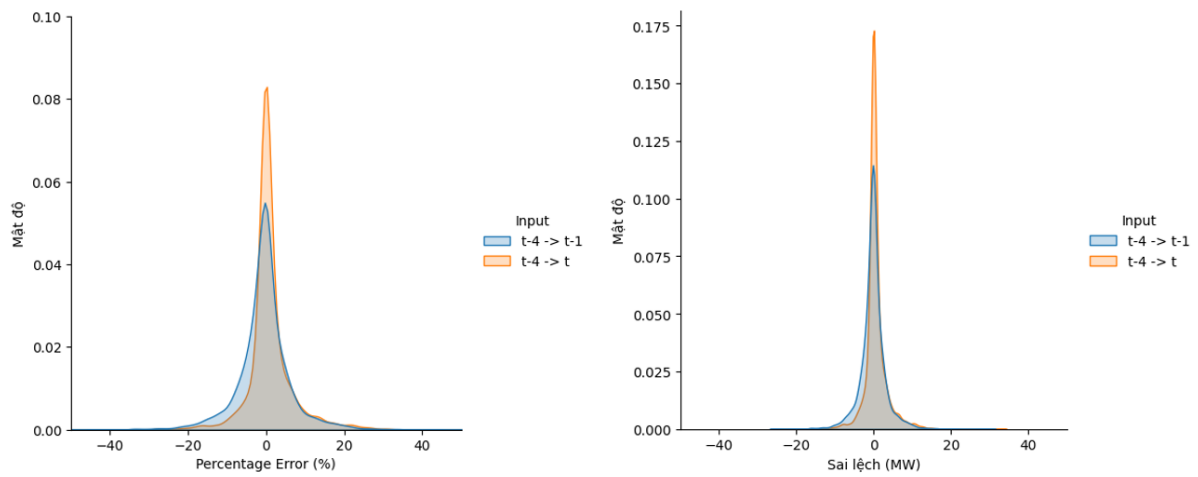
Các thiết lập cụ thể của mô hình như sau:

- Đầu vào: ma trận 4 hàng 7 cột. Trong đó 7 cột tương ứng với 7 đầu vào khí tượng và thời gian, 4 hàng gồm các dữ liệu tại thời điểm $t-1$, $t-2$, $t-3$, $t-4$.
- Đầu ra: Công suất tại thời điểm t
- Mô hình: LSTM 4 lớp, mỗi lớp 100 nút ẩn
- Các thiết lập huấn luyện tương tự như tại Mục 2.5.4.

Các kết quả được trình bày trong Bảng 2.9 và Hình 2.15.

Bảng 2.9. Kết quả sai số của 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
			MW	%	MW^2	MW	%
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



Hình 2.15. Phân bố sai số của 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 $t-1$

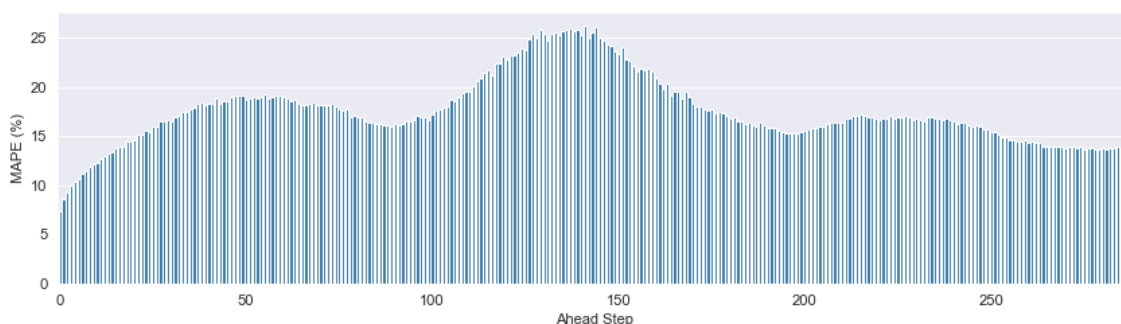
Kết quả cho thấy việc dự báo cho 1 bước tới từ dữ liệu quá khứ có sai số tương đối lớn với MAPE là 6,549% và nRMSE là 10,140%.

Hiện tượng sai số tăng lên nhanh như trên cho thấy khó khăn của việc dự báo đối với các nhà máy ĐMT quy mô công suất lớn. Như đã trình bày ở các phần trước, ở các nhà máy quy mô công suất lớn, thực tế chỉ có từ 1-2 trạm đo các thông số khí tượng cho cả nhà máy ĐMT quy mô 48 MW với diện tích gần 50 hecta. Do đó thông số khí tượng đo được chỉ mang tính trung bình, đặc biệt là bức xạ. Một số đám mây nhỏ ngang qua cũng có thể ảnh hưởng mạnh đến bức xạ đo được tại vị trí trạm khí tượng nhưng lại ít ảnh hưởng đến công suất toàn bộ. Hoặc ngược lại một đám mây lớn đi qua ảnh hưởng mạnh đến công suất toàn bộ nhưng tại vị trí trạm khí tượng thì lại ghi nhận mức sụt giảm GHI thấp.

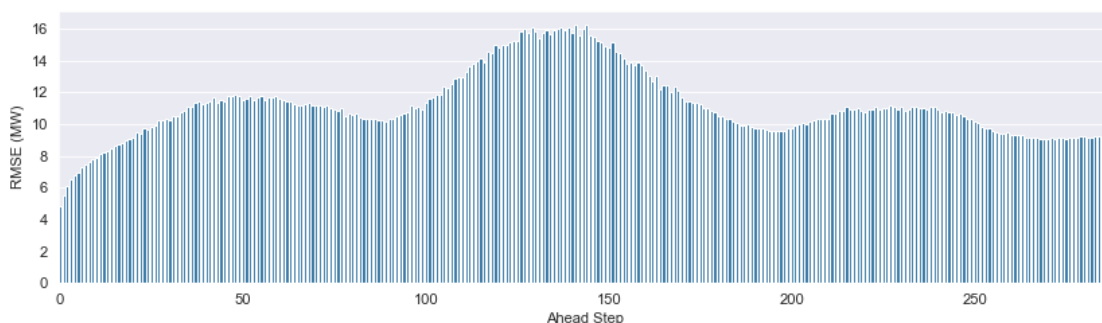
2.11. Dự báo nhiều bước tiếp theo

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. Kết quả huấn luyện và kiểm tra được trình bày như Hình 2.16 và Hình 2.17 dưới đây.



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)



Hình 2.17. Kết quả chỉ số RMSE (MW) trên tập dữ liệu kiểm tra của từng bước dự báo trong mô hình dự báo nhiều bước đồng thời (multistep)

Kết quả chi tiết của từng bước được trình bày trong bảng số liệu tại Phụ lục 2 của luận án. Dưới đây là Bảng 2.10 trình bày dữ liệu rút gọn kết quả chỉ số sai số dự báo trên tập dữ liệu kiểm tra của các bước trong mô hình dự báo nhiều bước đồng thời.

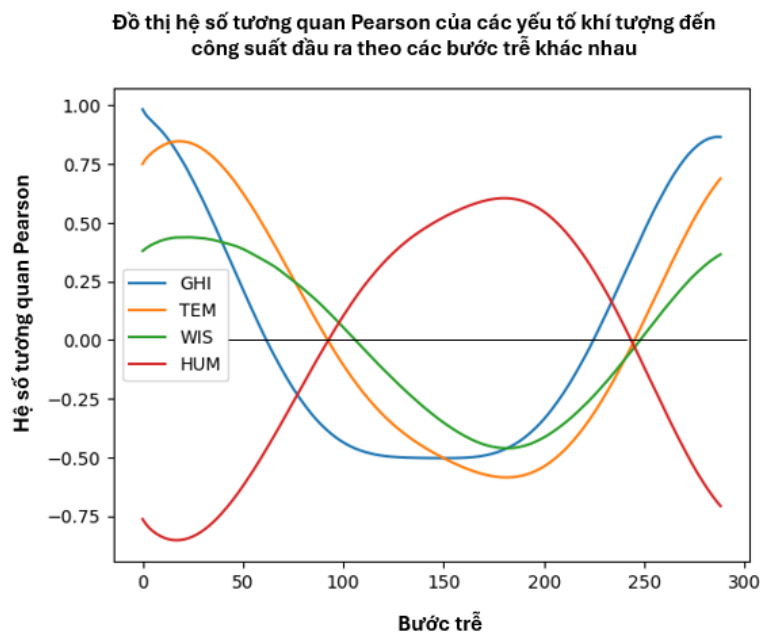
Bảng 2.10. Bảng dữ liệu rút gọn kết quả chi tiết các chỉ số sai số dự báo trên tập dữ liệu kiểm tra của các bước trong mô hình dự báo nhiều bước đồng thời

Bước dự báo	MSE	RMSE	MAE	MAPE	Bước dự báo	MSE	RMSE	MAE	MAPE
	MW ²	MW	MW	%		MW ²	MW	MW	%
1	23,079	4,804	3,57	7,438	145	263,48	16,232	12,549	26,145
12	65,929	8,12	6,082	12,67	156	189,56	13,768	10,364	21,591
24	94,156	9,703	7,476	15,574	168	145,07	12,045	9,069	18,894
36	122,59	11,072	8,558	17,83	180	114,87	10,718	8,201	17,085
48	139,02	11,791	9,153	19,069	192	94,272	9,709	7,593	15,82
60	137,68	11,734	9,182	19,13	204	101,34	10,067	7,628	15,892
72	124,15	11,142	8,744	18,216	216	122,52	11,069	8,283	17,257
84	107,16	10,352	7,842	16,337	228	124,32	11,15	8,185	17,053
96	116,23	10,781	7,985	16,636	240	122,55	11,07	8,016	16,7

Bước	MSE	RMSE	MAE	MAPE	Bước	MSE	RMSE	MAE	MAPE
dự báo	MW ²	MW	MW	%	dự báo	MW ²	MW	MW	%
108	157,22	12,539	9,112	18,984	252	101,23	10,061	7,393	15,402
120	223,84	14,961	11,095	23,114	264	86,324	9,291	6,859	14,289
132	238,12	15,431	11,85	24,688	276	82,762	9,097	6,583	13,715
144	255,68	15,99	12,285	25,593	288	84,835	9,211	6,673	13,902

Từ các hình vẽ và bảng số liệu trên có thể nhận thấy sai số dự báo đạt được khi dự báo ở bước 288 (01 ngày tới) là 9,211 MW với RMSE và 13,902% với MAPE. Như vậy có thể thấy mô hình mới hoạt động tốt trên dự báo đơn bước và còn có sai số lớn trong mô hình dự báo nhiều bước, cụ thể là cho mốc 01 ngày tới. Khi huấn luyện cho dự báo đa bước, kết quả dự báo cho 1 bước tới cũng có sai số tăng tương đối mạnh, RMSE tăng từ 2,166 MW trong mô hình dự báo đơn bước lên thành 4,804 MW, MAPE tăng từ 2,942% trong mô hình dự báo đơn bước lên thành 7,438%. Nguyên nhân là do khi này quá trình huấn luyện phải xem xét cải thiện độ chính xác trên toàn bộ đầu ra 288 điểm dữ liệu trong mỗi lần huấn luyện thay vì chỉ 1 điểm dữ liệu.

Đồ thị sai số của từng bước dự báo trong dự báo nhiều bước còn cho thấy hiện tượng sai số cao nhất ở khoảng 140-150 bước 05 phút tới ứng với độ trễ khoảng 12 tiếng. Để tìm hiểu nguyên nhân gây ra hiện tượng này, luận án tiến hành phân tích hệ số tương quan của các yếu tố khí tượng đến công suất đầu ra theo các bước trễ khác nhau với phạm vi từ 1 đến 288 bước trễ 05 phút. Kết quả thu được như trong Hình 2.18. Ở độ trễ khoảng 12 tiếng, hệ số tương quan của cả bức xạ, nhiệt độ, độ ẩm và tốc độ gió đều đảo chiều rất mạnh so với các độ trễ nhỏ, đây có thể là nguyên nhân chính dẫn đến hiện tượng sai số của khoảng độ trễ này tăng mạnh.



Hình 2.18. Hệ số tương quan Pearson của các yếu tố khí tượng đến công suất đầu ra theo các bước trễ khác nhau

2.12. Đánh giá và kết luận

Đề 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 được công bố

Mô hình	Năm	Quy mô nhà máy	Độ phân giải	Độ dài dự báo	Số bước dự báo	Khung dự báo	MAPE (%)	nRMSE (%)
Dự báo cho chu kỳ đang có dữ liệu khí tượng tương ứng								
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
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. Từ việc so sánh kết quả đạt được có thể rút ra một số kết luận sau:

- Đối với khung dự báo cho chu kỳ hiện tại:

Có thể nhận thấy sai số đạt được của mô hình tương đối tốt so với nhóm các nghiên cứu trên dữ liệu có cùng độ phân giải 05 phút với MAPE đạt được đều nằm trong khoảng từ 2-3% và RMSE đạt được trong khoảng từ 4,5%-6,0%. Kết quả này cũng tốt

hơn khi so sánh kết quả nghiên cứu từ các bộ dữ liệu có độ phân giải lớn hơn từ 30 phút đến 60 phút kết quả cho thấy sai số đạt được của các bộ dữ liệu phân giải lớn có xu hướng lớn hơn với MAPE dao động trong khoảng từ 4,0-7,5% và nRMSE dao động trong khoảng từ 6,5% đến 13,2%. Khi xét trên quy mô, một số nghiên cứu có quy mô nhà máy tương đương với quy mô của luận án nhưng ở điều kiện độ phân giải dữ liệu lớn hơn cũng cho kết quả sai số cao hơn.

- Đối với dự báo cho 01 chu kỳ tiếp theo:

Ở mô hình để tiến hành dự báo cho 1 bước tiếp theo. So với mô hình công bố của Ospina và cộng sự (nghiên cứu nhà máy quy mô 12,6 MW, độ phân giải dữ liệu 30 phút), mô hình Luận án đạt được có MAPE tương đương khoảng 6,5% và nRMSE tốt hơn khoảng 2% (10,140% so với 12,184%). Sự khác biệt có thể lý giải đến từ sự khác nhau về độ phân giải dữ liệu. Cả mô hình luận án và mô hình của Ospina và cộng sự đều có sai số MAPE tốt hơn so với mô hình của Harrou và cộng sự nghiên cứu trên bộ dữ liệu nhà máy 9 MWp độ phân giải 15 phút.

Tuy nhiên, khi so sánh với các mô hình nghiên cứu trên các bộ dữ liệu độ phân giải 7,5 phút nhưng quy mô nghiên cứu trên hệ thống chỉ có 20 kW của Zhou H và cộng sự công bố năm 2019 và 2021 thì sự khác biệt là tương đối lớn. Sai số của nghiên cứu của Zhou H đạt được là MAPE từ 3,5%-4% và nRMSE từ 6,5%-6,95%. Điều này cho thấy sự khác biệt trong quy mô công suất dự báo ảnh hưởng đến chất lượng dự báo. Như đã trình bày ở các phần trước, ở các nhà máy quy mô công suất lớn, thực tế chỉ có từ 1-2 trạm đo các thông số khí tượng cho cả nhà máy ĐMT quy mô vài chục MW với diện tích gần vài chục hecta. Do đó thông số khí tượng đo được chỉ mang tính trung bình, đặc biệt là bức xạ. Một số đám mây nhỏ ngang qua cũng có thể ảnh hưởng mạnh đến bức xạ đo được tại vị trí trạm khí tượng nhưng lại ít ảnh hưởng đến công suất toàn bộ. Hoặc ngược lại một đám mây lớn đi qua ảnh hưởng mạnh đến công suất toàn bộ nhưng tại vị trí trạm khí tượng thì lại ghi nhận mức sụt giảm GHI thấp.

- Đối với dự báo cho khung 01 ngày tới:

Việc thí nghiệm ứng dụng mô hình thu được để dự báo đa bước (multistep), có thể thấy kết quả mô hình chưa thực sự hoạt động tốt để phục vụ dự báo ngày tới với sai số dự báo ở bước 1 ngày tới chỉ đạt sai số MAPE khoảng 13,902% và nRMSE là 9,211 MW. So với nghiên cứu của Zhou N và cộng sự trên bộ dữ liệu nhà máy 5,83 kWp với cùng độ phân giải dữ liệu có thể thấy mô hình luận án cho kết quả MAPE cao hơn khoảng 2,3% nhưng nRMSE thấp hơn khoảng 1%. Tuy nhiên so với nghiên cứu của Zhang và cộng sự trên bộ dữ liệu nhà máy 6,41 kWp với độ phân giải dữ liệu 15 phút thì mô hình luận án cho kết quả MAPE cao hơn khoảng 5,5% và nRMSE cao hơn khoảng 3,7%, sự khác biệt có thể đến từ quy mô hệ thống ĐMT mà luận án nghiên cứu là gần

48 MW, lớn hơn rất nhiều so với hệ thống 6,41 kWp mà Zhang nghiên cứu. So với các kết quả của luận án đạt được hoặc một số kết quả đã công bố khác với cùng khung 1 ngày tới, kết quả của nhóm nghiên cứu Nguyễn Đức Tuyên và công sự đạt được có sai số tốt hơn rất nhiều với MAPE khoảng 3,7% và nRMSE khoảng 5,4%. Tuy nhiên do dữ liệu công suất dùng để so sánh với giá trị dự báo là giá trị cực đại lý thuyết được nhóm tác giả tính toán từ các giá trị khí tượng quá khứ nên sẽ có ảnh hưởng khi so sánh với kết quả từ các mô hình khác.

Như vậy, luận án đã xây dựng được mô hình dự báo công suất phát nhà máy ĐMT và thử nghiệm trên các khung dự báo cũng như so sánh với một số nghiên cứu công bố liên quan trong thời gian từ năm 2019 đến nay. Kết quả đạt được cho thấy mô hình thu được của luận án có chất lượng tương đối tốt, đặc biệt là ở khung dự báo cho chu kỳ hiện tại với sai số đạt được ở nhóm cao nhất với MAPE 2,942% và nRMSE đạt được là 4,513%. Nội dung trong phần tiếp theo sẽ nghiên cứu các giải pháp nhằm nâng cao chất lượng dự báo từ mô hình LSTM cơ bản đã xây dựng được trong phần này.

CHƯƠNG 3.

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

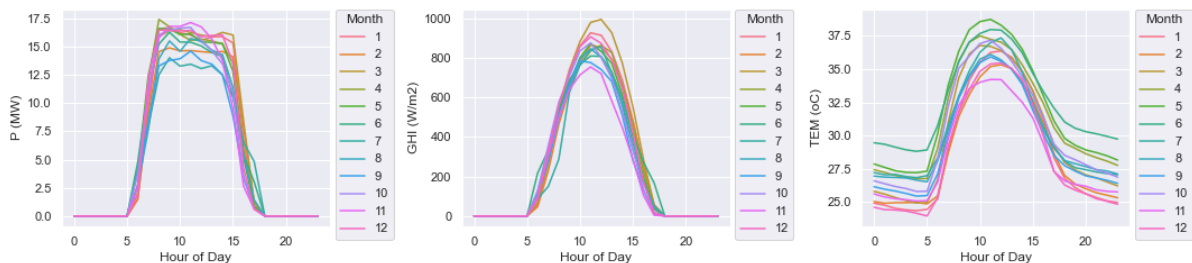
3.1. Kỹ thuật tiền xử lý số liệu với hệ số P/GHI kết hợp với phân cụm GHI

3.1.1. Đặt vấn đề

Việc xử lý dữ liệu đầu vào là bước bắt buộc đối với tất cả các phương pháp huấn luyện mô hình LSTM nhằm mục đích loại bỏ các dữ liệu nhiễu, xấu thường gặp trong bộ dữ liệu của một nhà máy điện mặt trời.

Trong điều kiện nhà máy điện được phát toàn bộ công suất thì công suất phát phụ thuộc vào khả năng của hệ thống điện mặt trời, tức là phụ thuộc vào GHI (bức xạ toàn phần, tính bằng đơn vị W/m^2) và một số điều kiện thời tiết khác. Tuy nhiên, khi nhà máy điện mặt trời bị giới hạn công suất hay sự cố thì xảy ra hiện tượng tập dữ liệu thu được có nhiều điểm dữ liệu không đáng tin cậy như có giá trị P (công suất phát, MW) khi GHI bằng 0, hoặc tỷ lệ giữa GHI và P là bất thường, GHI lớn nhưng công suất phát P lại nhỏ, không tương xứng.

Ví dụ, dữ liệu thu thập từ một nhà máy bị giới hạn công suất được phân tích để làm rõ về vấn đề này. Trước tiên, dữ liệu trung bình hàng giờ trong một ngày trong các tháng của năm được phân tích như trong Hình 3.1. Các dữ liệu bao gồm công suất đầu ra, Bức xạ GHI và nhiệt độ môi trường. Dữ liệu cho thấy lượng bức xạ mà khu vực nhà máy nhận được có hình dạng chuông chuẩn, nhưng công suất đầu ra của nhà máy có hình dạng khá đặc biệt. Nhà máy có công suất thiết kế là 39 MWp nhưng giá trị trung bình đầu ra từ 9:00 đến 15:00 chỉ khoảng 17 MW, khá đồng đều nhưng ở mức thấp so với công suất thiết kế.



Hình 3.1. Dữ liệu trung bình giờ theo từng tháng của nhà máy DMT

3.1.2. Phương án đề xuất

Phương pháp khoảng tứ phân vị (IQR) trong xử lý dữ liệu đã được áp dụng bởi các nhà nghiên cứu, tuy nhiên trong lĩnh vực dự báo công suất của nhà máy ĐMT, phương pháp này chưa được phổ biến. Trong nghiên cứu [92], tác giả sử dụng phương pháp IQR để xử lý dữ liệu đầu vào dựa trên mối quan hệ giữa GHI và P. Các tác giả tính toán hiệu suất điện (Effpv) cho tất cả các điểm dữ liệu, sau đó xem xét các điểm dữ liệu có Effpv nằm ngoài phạm vi Giới hạn Dưới - Giới hạn Trên là ngoại lai. Phương pháp này chỉ cho kết quả tốt khi nhà máy hoạt động bình thường và sai sót dữ liệu chỉ xuất phát từ các nguyên nhân thông thường như lỗi hệ thống, đo lường không chính xác, lỗi truyền tải hoặc lưu trữ dữ liệu, v.v. Tuy nhiên, trong trường hợp nhà máy ĐMT hoạt động dưới điều kiện giới hạn công suất, phương pháp sử dụng thuần túy IQR để xử lý dữ liệu cũng không hiệu quả, khả năng phát hiện ngoại lai của bộ lọc không đủ mạnh.

Trong nghiên cứu [93] liên quan đến dự báo bức xạ GHI, trong quá trình xử lý dữ liệu ngoại lai, tác giả sử dụng phương pháp IQR nhưng chỉ áp dụng cho các giá trị GHI trong một khung thời gian cụ thể (ví dụ, từ 7:00-8:00). Thực tế là biến động GHI trong một khoảng thời gian 01 giờ rất lớn. Và việc chỉ sử dụng IQR cho tập dữ liệu về bức xạ có biến động theo từng giờ với một phạm vi biến động lớn sẽ làm cho bộ lọc các giá trị ngoại lai không hiệu quả, chỉ có thể lọc ra các giá trị bất thường quá cao hoặc âm. Ví dụ, trong dữ liệu GHI được trình bày, tác giả đã chỉ ra, từ 11:00-14:00, phạm vi GHI đáng tin cậy dao động từ 0 đến khoảng 1000 (W/m^2).

Để khắc phục nhược điểm này, trong nghiên cứu đã được công bố năm 2022, nghiên cứu sinh và nhóm nghiên cứu của mình đã đề xuất một 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, loại bỏ các giá trị ngoại lai trong các khoảng hoạt động có mức độ GHI khác nhau, cụ thể như sau:

Bước 1: Tính toán đặc trưng cho mỗi điểm dữ liệu theo tỷ lệ P/GHI. Giá trị này phản ánh lượng công suất đầu ra tương ứng với lượng GHI mà vị trí nhà máy nhận được. Trong dữ liệu dạng chuỗi thời gian, mỗi điểm dữ liệu là một tập hợp các thông số công suất, bức xạ và nhiệt độ môi trường đo tại cùng một khoảng thời gian.

Bước 2: Chia tập dữ liệu thành các nhóm nhỏ theo GHI của từng điểm dữ liệu. Mỗi nhóm có khoảng cách từ giá trị thấp nhất đến cao nhất là 50 đơn vị W/m^2 . Việc chia

nhỏ các khoảng GHI là nhằm mục đích giúp tăng cường độ chính xác của các bộ lọc ngoại lai theo phương pháp IQR.

Các điểm dữ liệu sẽ được sắp xếp theo thứ tự thông số GHI tăng dần từ thấp đến cao, sau đó phân nhóm với khoảng cách 50 đơn vị mỗi nhóm.

Công suất phát đầu ra là một đại lượng phụ thuộc mạnh vào bức xạ. Tuy nhiên công suất này còn bị chịu tác động bởi nhiều yếu tố khác như nhiệt độ môi trường, khả dụng thiết bị.... Hơn nữa, trong một số điều kiện như sự cố hay bị giới hạn công suất do lưới thường xảy ra tại các khung giờ có bức xạ cao, tỷ lệ P/GHI cũng sẽ có sự biến động mạnh so với các khung giờ khác. Nếu không thực hiện việc chi nhỏ khoảng GHI, bộ lọc IQR sẽ rất khó phát huy tác dụng loại được điểm ngoại lai, lỗi thực sự ứng với từng điều kiện khác nhau của dữ liệu.

Bước 3: Sử dụng phương pháp IQR cho các giá trị P/GHI trong mỗi nhóm nhỏ để loại bỏ các giá trị ngoại lai trong nhóm.

Quy tắc IQR được hiểu là, mỗi tập dữ liệu sẽ được sắp xếp theo thứ tự từ thấp đến cao, sau đó xác định các điểm tứ phân vị như sau:

+ Q1: là giá trị mà 25% các điểm dữ liệu sẽ nhỏ hơn giá trị này.

+ Q3: là giá trị mà 75% các điểm dữ liệu sẽ nhỏ hơn giá trị này.

+ $IQR = Q3 - Q1$

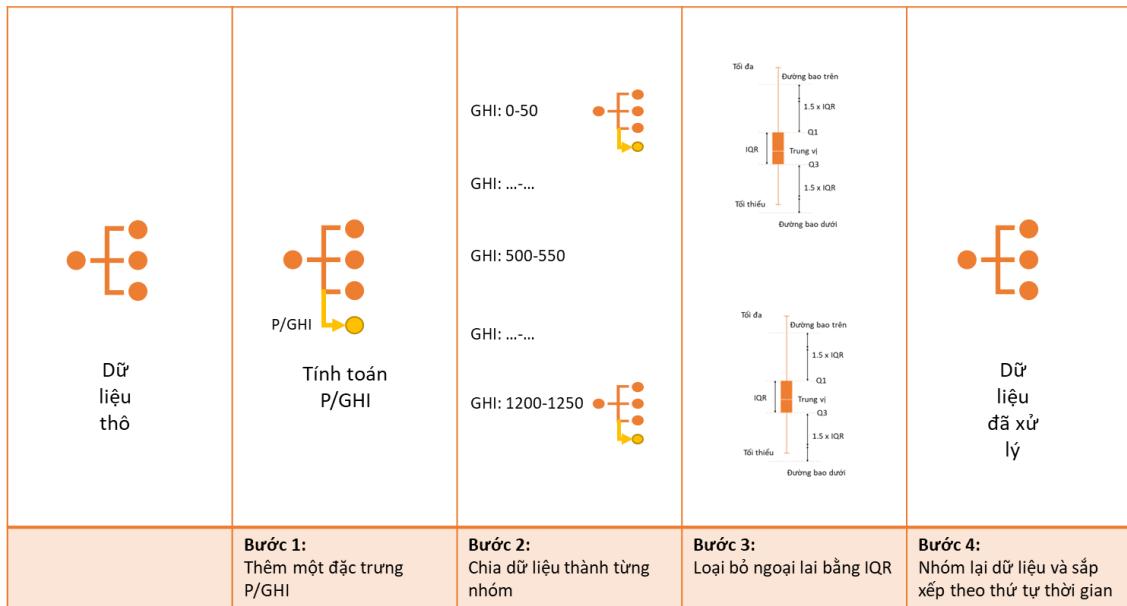
+ Upper bound (Giới hạn trên) = $Q3 + 1,5 \times IQR$

+ Lower bound (Giới hạn dưới) = $Q1 - 1,5 \times IQR$

Một giá trị được coi là tin cậy khi nằm trong khoảng từ Giới hạn dưới đến Giới hạn trên. Những điểm ngoài khoảng này được coi là các điểm ngoại lai.

Bước 4: sắp xếp lại theo thứ tự thời gian và phân dữ liệu thành các nhóm.

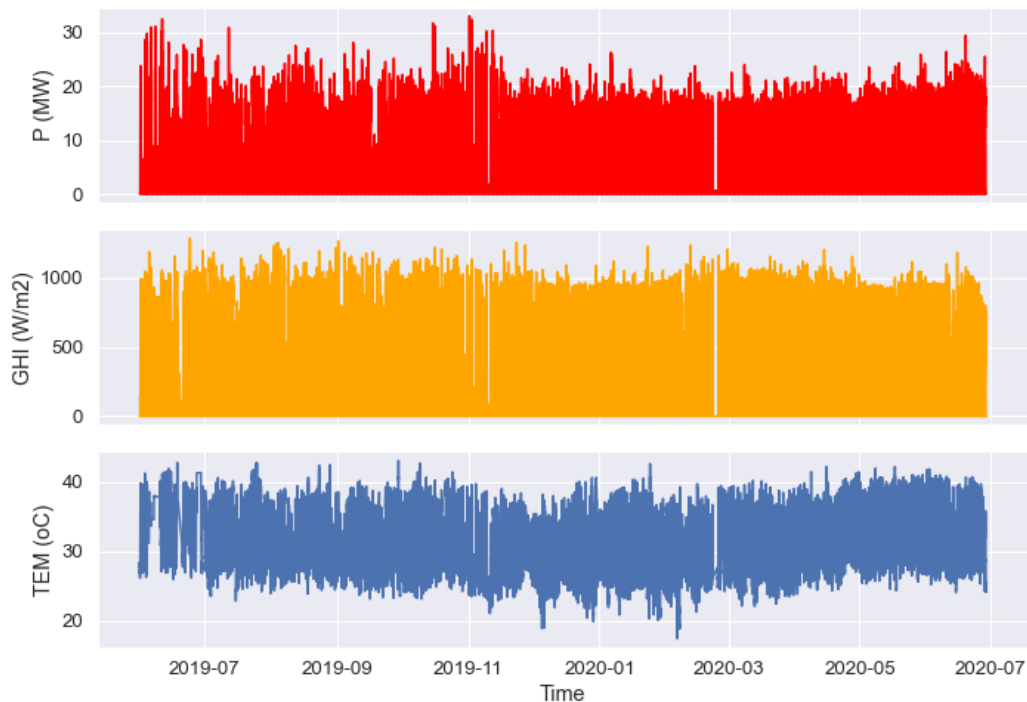
Do ở các bước trước, dữ liệu được sắp xếp theo nhóm bức xạ (GHI) mà không quan tâm đến tiêu chí thời gian nên tại bước này, trước khi đưa chuỗi dữ liệu thời gian vào huấn luyện cần thực hiện sắp xếp lại theo thứ tự thời gian. Tổng hợp mô tả lại kỹ thuật này được trình bày trong Hình 3.2.



Hình 3.2. Các bước trong kỹ thuật xử lý dữ liệu ĐMT sử dụng hệ số P/GHI

3.1.3. Thí nghiệm chứng minh hiệu quả

3.1.3.1. Dữ liệu



Hình 3.3. Dữ liệu vận hành nhà máy ĐMT

Bộ dữ liệu được thu thập từ nhà máy điện mặt trời 39MWp đặt tại Việt Nam, dữ liệu từ tháng 6 năm 2019 đến hết tháng 6 năm 2020. Dữ liệu thu thập thông qua hệ thống đo đếm của nhà máy bao gồm công suất đầu ra của nhà máy, P [MW], bức xạ mặt trời GHI tính bằng [W/m²] và nhiệt độ môi trường TEM tính bằng [°C]. Dữ liệu có độ phân

giải 05 phút, mỗi điểm của tập dữ liệu là giá trị trung bình được đo trong 05 phút cho mỗi thông số. Tập dữ liệu đầy đủ được hiển thị trong Hình 3.3.

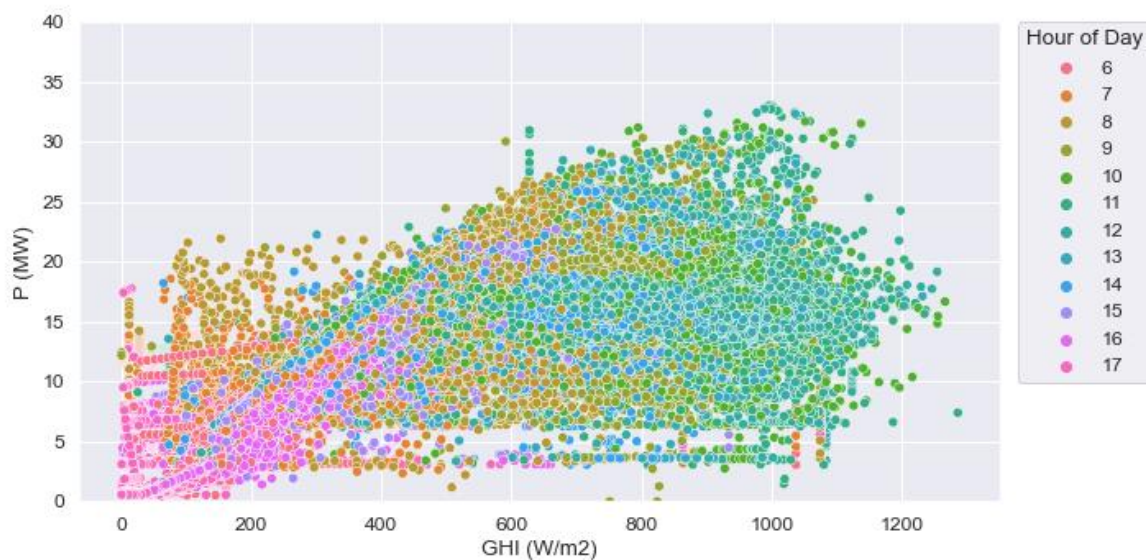
Hình 3.4 mô tả dữ liệu đo đạc trong 1 tuần của nhà máy. Có thể thấy rằng trong những giờ trưa, mặc dù Bức xạ GHI tăng lên, công suất phát đầu ra tại nhiều thời điểm có xu hướng giảm. Nguyên nhân chính đến từ việc nhà máy được kết nối vào một khu vực lưới gồm nhiều nhà máy điện mặt trời và gió khác, mà tổng công suất của những nhà máy này ở trạng thái công suất tối đa lớn hơn công suất vận hành cho phép của đường truyền tải điện, vì vậy các nhà máy trong khu vực thường phải giảm công suất trong các giờ giữa trưa. Yếu tố này là khách quan, do đó nhà đầu tư của nhà máy gặp khó khăn lớn trong việc dự báo công suất của nhà máy. Vấn đề là cần có một thuật toán có khả năng dự đoán công suất đầu ra của nhà máy với độ chính xác cao trong điều kiện bị giới hạn công suất.



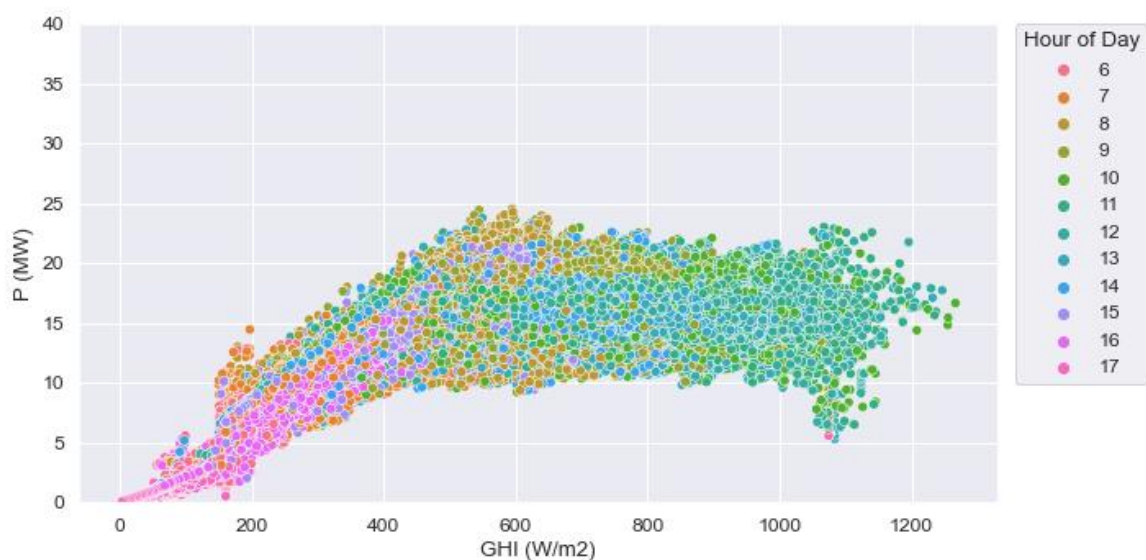
Hình 3.4. Dữ liệu quá khứ trong 1 tuần của nhà máy ĐMT

3.1.3.2. Tiền xử lý dữ liệu theo phương án đề xuất

Biểu diễn về hình ảnh dữ liệu trước xử lý và kết quả sau khi loại bỏ các giá trị ngoại lai được hiển thị trong Hình 3.5 dưới đây.



a. Dữ liệu trước xử lý



b. Dữ liệu sau xử lý

Hình 3.5. Biểu diễn dữ liệu trước và sau xử lý

3.1.3.3. Kết quả và phân tích

Để thấy rõ hiệu quả của việc tiền xử lý dữ liệu theo phương án đề xuất, luận án đã tiến hành sử dụng mô hình LSTM-4L100N (đã thu được từ Chương 2) để huấn luyện và dự báo cho hai bộ dữ liệu trước và sau xử lý. 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%.

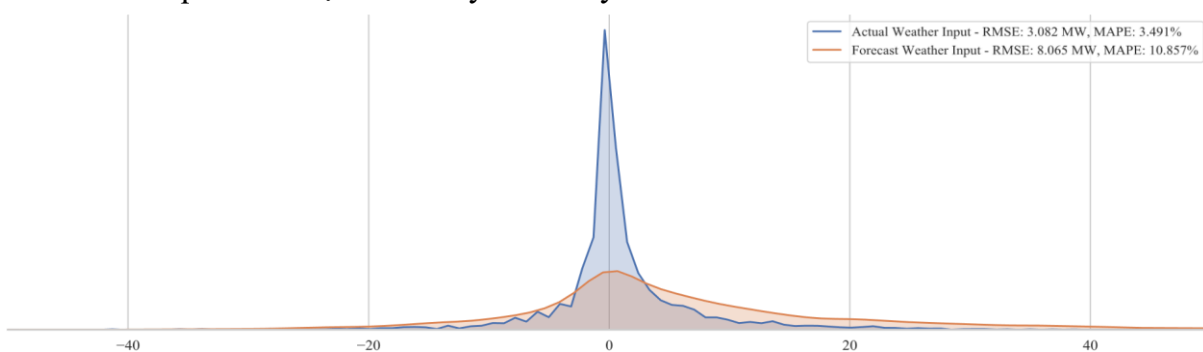
Bảng 3.1. Kết quả sai số dự báo trên tập dữ liệu kiểm tra với bộ dữ liệu huấn luyện trước và sau khi xử lý với kỹ thuật P/GHI kết hợp phân cụm GHI

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

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

3.2.1. Đặt vấn đề

Từ kết quả thu được ở Chương 2, mô hình LSTM-4L100N với MAPE tốt nhất sẽ được sử dụng để dự báo công suất đầu ra của nhà máy ĐMT với dữ liệu đầu vào cho việc huấn luyện được lấy từ dữ liệu thời tiết lịch sử thực tế thu thập tại hiện trường nhà máy. Trong phần này, mô hình được chọn sẽ được sử dụng để dự báo công suất đầu ra của nhà máy ĐMT với dữ liệu đầu vào là dữ liệu dự báo thời tiết được cung cấp bởi bên thứ ba. Kết quả thu được trình bày dưới đây.



Hình 3.6. Phân phối sai số của mô hình LSTM-4L100N

Hình 3.6 thể hiện một so sánh về phân phối sai số của mô hình LSTM-4L100N trên bộ dữ liệu khí tượng thực tế và bộ dữ liệu dự báo khí tượng được cung cấp bởi nhà cung cấp thương mại.

Bảng 3.2. Kết quả dự báo của mô hình LSTM-4L100N trên tập dữ liệu kiểm tra

Kết quả dự báo của mô hình đã huấn luyện trên bộ dữ liệu khí tượng	MSE	RMSE	MAE	MAPE
	MW ²	MW	MW	%
Thực tế	9,499	3,082	1,676	3,491
Dự báo	65,041	8,065	5,211	10,857

Từ kết quả Bảng 3.2 có thể thấy, mặc dù mô hình hoạt động rất 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). Như vậy, vấn đề đặt ra là cần giải pháp để cải thiện chất lượng dự báo khi sử dụng số liệu đầu vào khí tượng dự báo.

3.2.2. Phương án đề xuất

Từ kết quả thu được ở trên, 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 và đến từ chất lượng của nhà cung cấp dự báo. Tuy nhiên, để có được giải pháp khả thi giúp áp dụng mô hình tính toán công suất đầu ra trong thực tế với kết quả tốt, nghiên cứu sinh và nhóm nghiên cứu đã đề xuất và công bố một giải pháp với ý tưởng chính là mô hình học cũng từ dữ liệu dự báo khí tượng trong quá khứ, để với cùng nguồn dữ liệu dự báo, trong tương lai mô hình có khả năng cải thiện độ chính xác của mình.

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.

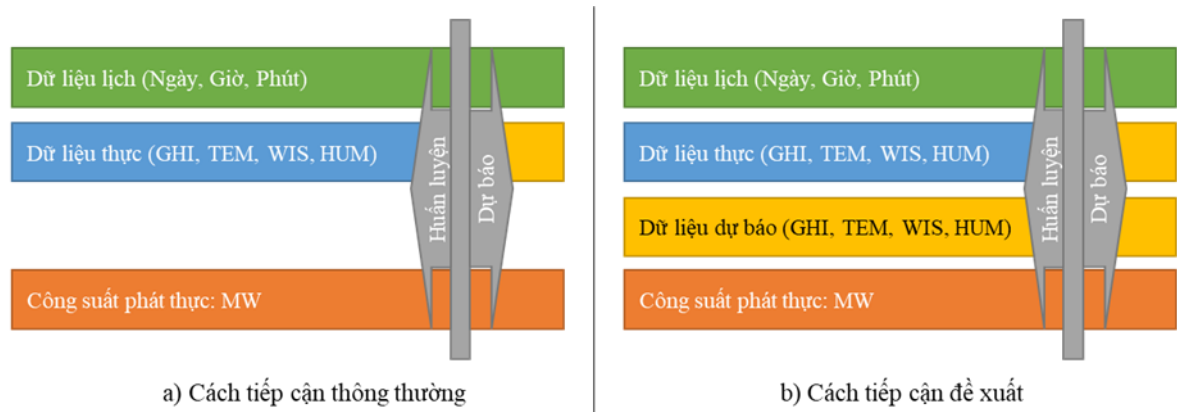
3.2.3. Thí nghiệm chứng minh hiệu quả

3.2.3.1. Dữ liệu và tiền xử lý dữ liệu

Nghiên cứu sử dụng bộ dữ liệu và tiền xử lý dữ liệu tương tự bộ dữ liệu đã trình bày trong Chương 2.

3.2.3.2. Thiết lập thí nghiệm

Hình 3.7 mô tả sự khác biệt của cấu trúc dữ liệu đầu vào của các mô hình đề xuất. Để chọn đầu vào thích hợp cho quá trình huấn luyện tiếp theo, dựa trên hệ số tương quan tính toán ở trên, các đặc trưng GHI, TEM, HUM, WIS được thêm vào lần lượt và đánh giá sai số trên bộ dữ liệu kiểm tra.



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

Trong thí nghiệm này, có các mô hình LSMT-4L100N với cấu trúc đầu vào bổ sung mới sẽ được huấn luyện để kiểm tra các chỉ số sai số bao gồm:

- Mô hình 8 đặc trưng đầu vào (1 đặc trưng dự báo mới: GHI)
- Mô hình 9 đặc trưng đầu vào (2 đặc trưng dự báo mới: GHI, TEM)
- Mô hình 10 đặc trưng đầu vào (3 đặc trưng dự báo mới: GHI, TEM, HUM)
- Mô hình 11 đặc trưng đầu vào (4 đặc trưng dự báo mới: GHI, TEM, HUM, WIS)

3.2.3.3. Kết quả và phân tích

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 + GHI, TEM, HUM, WIS dự báo)	65,041	8,065	5,211	10,857
7I+1F	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

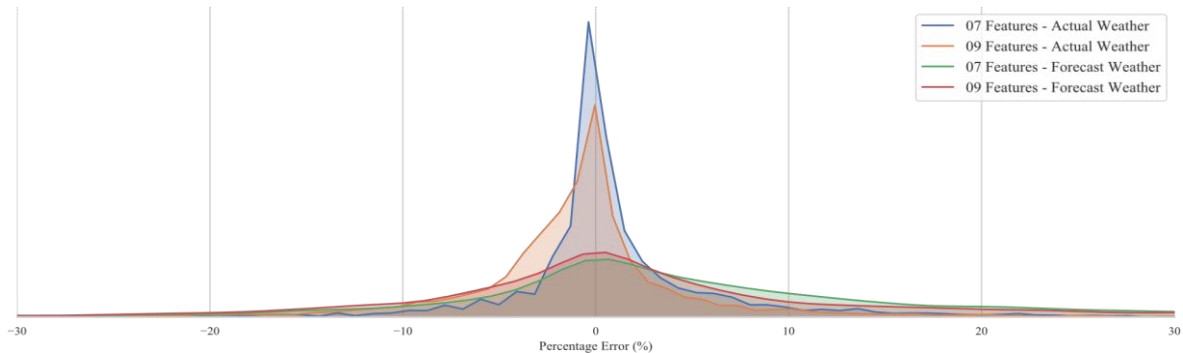
Bảng 3.3 cho thấy sự so sánh chi tiết về kết quả sai số trên mô hình LSTM-4L100N với các cấu hình đầu vào và kịch bản kiểm tra khác nhau khi dự báo công suất tại thời điểm t , sử dụng đầu vào khí tượng tại $t, t-1, t-2, t-3, t-4$.

Như đã thực nghiệm ở Chương 2, mô hình LSTM 4 lớp 100 nút ẩn (mô hình 7I) cho sai số RMSE là 3,082 MW và MAPE là 3,491% khi có 7 đầu vào (3 đầu vào thời gian, 4 đầu vào khí tượng thực tế thu thập trong quá khứ). Các mô hình 7I+1F, 7I + 2F, 7I+3F, 7I+4F là các mô hình được thêm các đầu vào là số liệu khí tượng dự báo trong quá khứ để cùng huấn luyện. Từ bảng số liệu tại mục (a) của Bảng 3.3 có thể nhận thấy, khi thêm các đầu vào là số liệu khí tượng dự báo, sai số mô hình đạt được trên tập dữ liệu kiểm tra tăng lên so với mô hình chỉ gồm các đầu vào khí tượng quá khứ thực đo.

Tuy nhiên khi ứng dụng mô hình dự báo này cùng với các số liệu khí tượng dự báo, các đầu vào khí tượng thực tế sẽ cần được thay thế bởi số liệu khí tượng dự báo. Bảng số liệu tại mục (b) của Bảng 3.3 trình bày 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). 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

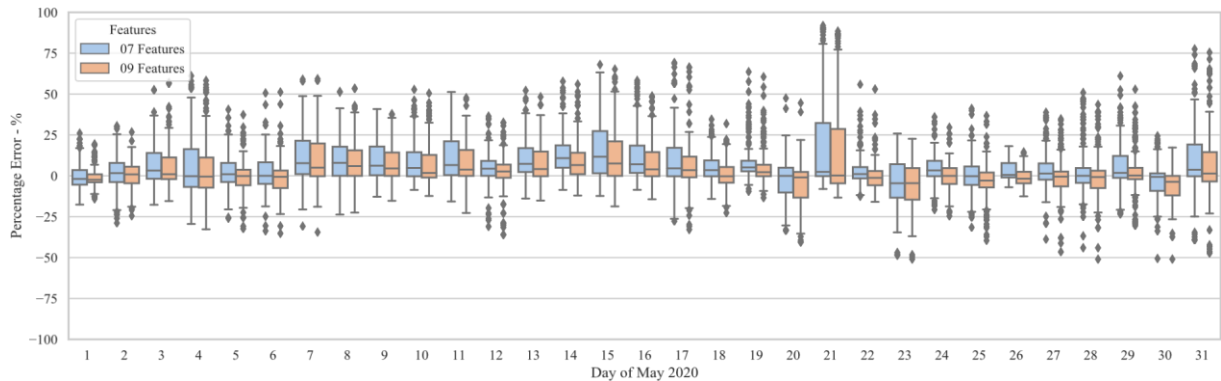
hình 7I+2F với MAPE là 9,881% đạt được mức độ cải thiện đáng kể nhất khoảng 0,976% so với mô hình 7I, mức cải thiện là khoảng 9%.

Phân phối sai số tương đối (PE) trong Hình 3.8 thể hiện rõ kết quả này.



Hình 3.8. Phân bố sai số tương đối với mô hình 07 và 09 đặc trưng

Hình 3.9 minh họa sai số dự báo công suất hàng ngày trong tháng 5 năm 2020 của mô hình LSTM 09 đặc trưng.



Hình 3.9. So sánh phân bố sai số tương đối trên bộ dữ liệu khí tượng dự báo của theo ngày trong 5/2020

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

3.3.1. Đặt vấn đề

Rất nhiều công trình nghiên cứu về dự báo ngắn hạn công suất phát của nhà máy ĐMT đã sử dụng chỉ dấu về mặt thời gian như một dữ liệu đầu vào quan trọng trong mô hình mà họ xây dựng [17], [19], [22], [57], [94], [95], [96]. Thống kê các yếu tố chỉ dấu thường được các tác giả lựa chọn sử dụng được trình bày trong Bảng 3.4. Các nhãn thời gian thường được dùng trong dự báo công suất ĐMT. Do đặc tính của ĐMT có quy luật tuần hoàn theo năm (do trái đất quay quanh mặt trời) và theo ngày (trái đất tự quay quanh trục) nên các yếu tố như ngày trong năm giờ trong ngày thường được lựa chọn. Trong khi đó nhãn thời gian chỉ tháng trong năm hoặc tuần trong năm hầu như không được lựa chọn sử dụng.

Bảng 3.4. Các nhãn thời gian thường được dùng trong dự báo công suất ĐMT

Chỉ dấu thời gian	Ý nghĩa	Ví dụ
Ngày trong năm	Phản ánh thứ tự các ngày trong một năm tính từ 1 đến 365	Ngày 1/1 thường gán nhãn là 1, ngày 31/12 thường được gán nhãn là 365
Giờ trong ngày	Xác định giá trị từ 1 đến 24 giờ trong ngày	Có thể đánh nhãn từ 0 đến 23 hoặc 1 đến 24
Phút trong giờ	Xác định giá trị từ 1 đến 60. Tùy vào độ phân giải dữ liệu có thể đánh chính xác theo nhãn phút hoặc đánh theo kiểu từng khoảng	Đối với dữ liệu độ phân giải 15 phút có thể đánh nhãn: 0, 15, 30, 45 hoặc đánh nhãn 0, 1, 2, 3

Tuy nhiên, khi sử dụng các chỉ dấu thời gian trong dự báo, thường tạo ra các tín hiệu đầu vào có dạng răng cưa, trong khi công suất đầu ra của nhà máy ĐMT thường có dạng đường cong hình chuông. Vấn đề đặt ra là liệu có một đầu vào khác có thể được sử dụng một cách hiệu quả để thay thế các chỉ dấu thời gian như ngày, giờ và phút vẫn hay thường được sử dụng phổ biến hay không. Luận án đề xuất một phương pháp sử dụng bức xạ trời trong như một thay thế cho các chỉ số thông thường như ngày, giờ và phút, nhằm nâng cao hiệu quả của mô hình dự báo và giảm sai số dự báo.

3.3.2. Phương án đề xuất

Việc sử dụng bức xạ trời trong vào trong dự báo ĐMT cũng đã được một số các tác giả nghiên cứu. Trong [97], các tác giả đã thử nghiệm các mô hình trời trong khác nhau để làm đầu vào cho mô hình dự báo bức xạ mặt trời sử dụng phương pháp quán tính với kết luận là mô hình tính toán bức xạ trời trong tốt hơn không làm cải thiện chất lượng dự báo. Tuy nhiên khẳng định này chỉ mới được thử nghiệm với mô hình dự báo sử dụng phương pháp quán tính là phương pháp dự báo đơn giản nhất với chất lượng kém hơn nhiều so với các phương pháp dự báo hiện tại sử dụng các công nghệ hiện đại như mạng nơ-ron nhân tạo. Trong [98], các tác giả sử dụng một mô hình mạng nơ-ron để dự báo các thông số đầu vào cho mô hình BIRD (một mô hình tính bức xạ trời trong, sau đó sử dụng các công thức quy đổi để tính toán ra công suất phát). Kết luận rút ra được là mô hình có sai số đạt được tương đối tốt tuy nhiên không có sự so sánh với các mô hình khác. Trong [99], các tác giả đề xuất phương pháp tính toán mới cho chỉ số trời trong của hệ thống điện mặt trời và thông qua đó để tính toán dự báo cho các hệ thống điện mặt trời ở khu vực lân cận. Trong [100], các tác giả nghiên cứu mô hình tích hợp các mô đun aerosol và bức xạ thời gian thực sau đó đồng hóa dữ liệu AOD (Aerosol Optical Depth – Độ dày quang học sol khí) từ Himawari 8 sử dụng 3DVAR để tối ưu

quá trình dự báo bức xạ trời trong của các khu vực tại Trung Quốc. Trong [101], các tác giả đề xuất sử dụng một mô hình tính bức xạ trời trong mới sau đó sử dụng mô hình này để tính toán bức xạ trời trong của nhà máy và dùng nó làm đầu vào cho mô hình dự báo công suất theo phương pháp thống kê chuỗi thời gian. Kết quả là sai số dự báo ngắn hạn cải thiện khoảng 3%. Trong công bố đáng chú ý liên quan đến chủ đề này gần đây năm 2022 [102], nghiên cứu dự báo công suất ĐMT bằng các mô hình RF, SVR, CNN, LSTM, Hybrid với các cấu hình tập dữ liệu huấn luyện gồm: (1) Dữ liệu gốc (GHI, nhiệt độ, độ ẩm, tốc độ gió...), (2) Dữ liệu mở rộng gồm dữ liệu gốc và dữ liệu tính toán từ dữ liệu gốc, (3) Dữ liệu nguyên bản, bổ sung dữ liệu tính toán mở rộng và bổ sung chỉ dấu thời gian là giờ trong ngày. Kết luận thu được là bộ dữ liệu thứ 3 cho kết quả tốt nhất cho thấy vai trò của việc bổ sung thêm chỉ dấu thời gian giờ trong ngày trong thí nghiệm làm tăng hiệu quả dự báo.

Như vậy, bức xạ trời trong là một trong những đầu vào tiềm năng trong việc cải thiện chất lượng các mô hình dự báo, tuy nhiên các tác giả mới chỉ nghiên cứu sử dụng bức xạ trời trong trong các mô hình dự báo tương đối đơn giản chủ yếu là các mô hình thống kê chuỗi thời gian với một số ghi nhận về cải tiến chất lượng khoảng 3%.

Trong luận án này, ý tưởng được sử dụng là thay bộ dữ liệu chỉ dấu thời gian (Ngày trong năm, Giờ trong ngày, phút trong giờ) bằng một đầu vào duy nhất khác mà có hiệu quả hơn. Đầu vào được đề xuất là Bức xạ trời trong. Bức xạ trời trong 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). Thực tế kết quả thí nghiệm sẽ chứng minh nhận định trên.

3.3.3. Thí nghiệm kiểm chứng hiệu quả

3.3.3.1. Dữ liệu

Sử dụng bộ dữ liệu tại Chương 2.

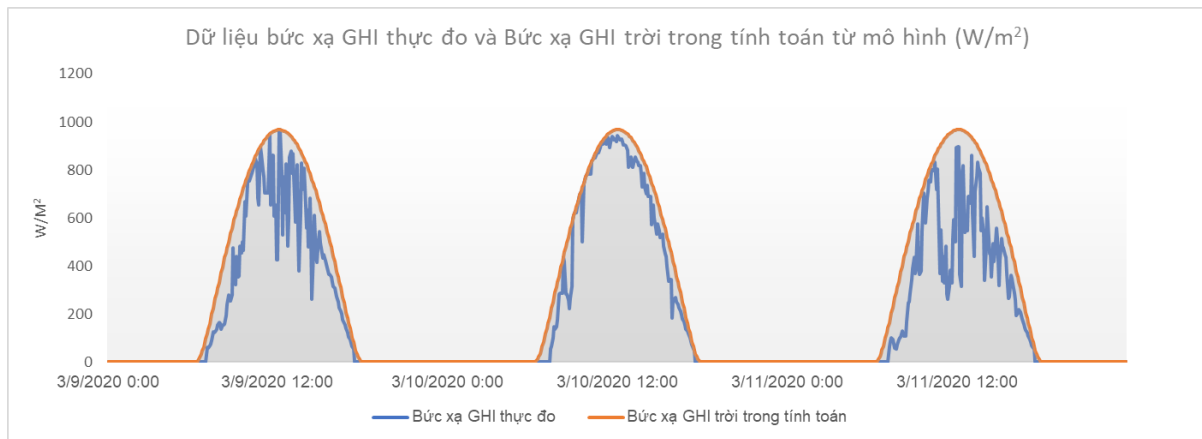
3.3.3.2. Tiền xử lý dữ liệu

Trong luận án này, như đã được đề cập trong phần đặt vấn đề, sẽ tìm cách thay dữ liệu chỉ dấu thời gian thường được sử dụng trong hầu hết các nghiên cứu dự báo công suất ĐMT bằng một đặc trưng dữ liệu mới đó là bức xạ trời trong. Do đó, thay vì sử dụng đặc trưng chỉ dấu thời gian thông thường (ngày trong năm, giờ trong ngày, phút trong giờ), luận án đề xuất sử dụng bức xạ trời trong tại vị trí địa lý của nhà máy như một đầu vào để dự báo.

Theo báo cáo phân tích so sánh về các mô hình bức xạ trong trời trong, mô hình Ineichen/Perez cho thấy hiệu quả tốt nhất với số lượng dữ liệu đầu vào tối thiểu [10]. Do đó, luận án này sẽ sử dụng mô hình Ineichen/Perez để tính toán bức xạ trời trong tại vị trí của nhà máy.

Công cụ được sử dụng để tính toán bức xạ mặt trời là PVLIB, một thư viện phần mềm mã nguồn mở được sử dụng trong lĩnh vực năng lượng mặt trời. Nó cung cấp các công cụ và chức năng để tính toán và mô phỏng hiệu suất của hệ thống năng lượng mặt trời, bao gồm các thông số như sản lượng điện, bức xạ mặt trời, góc mặt trời và các yếu tố khác liên quan đến hoạt động của hệ thống điện mặt trời [103].

Một ví dụ cụ thể về việc tính toán GHI trời trong tại vị trí nhà máy điện mặt trời so với GHI đo được sử dụng các thiết bị của nhà máy được trình bày trong Hình 3.10.



Hình 3.10. Bức xạ GHI thực đo và Bức xạ GHI trời trong tính toán (W/m^2)

Bộ dữ liệu đã thu thập sau đó được chia thành hai phần riêng biệt:

- Tập huấn luyện: Dữ liệu từ tháng 5 năm 2019 đến tháng 4 năm 2020 được sử dụng để huấn luyện các mô hình dự báo.
- Tập kiểm tra: Dữ liệu tháng 5 năm 2020 được sử dụng làm đầu vào để tính toán công suất dự báo đầu ra và so sánh với công suất thực đo, từ đó đánh giá hiệu quả của mô hình dự báo.

3.3.3.3. Thiết lập thí nghiệm

Trong thí nghiệm tại Mục 3.1, luận án đã thực hiện thí nghiệm trên dữ liệu của một nhà máy ĐMT quy mô công nghiệp và đề xuất một phương pháp sử dụng mạng LSTM cùng các phương pháp để chọn các tham số cho mô hình. Luận án này sẽ tái tạo các điều kiện thực nghiệm tương tự nhưng sau đó sử dụng các đầu vào thay thế mới để đánh giá hiệu quả.

Bảng 3.5. So sánh cấu hình thiết lập thực nghiệm với các mô hình

Thông tin	Mô hình sử dụng chỉ dấu thời gian	Mô hình sử dụng Bức xạ trời trong
Độ phân giải dữ liệu	05 phút	05 phút
Dữ liệu đầu vào	+ Dữ liệu khí tượng (GHI, TEM, WIS, HUM) + Dữ liệu chỉ dấu thời gian (Ngày trong năm, Giờ trong ngày, Phút trong giờ)	+ Dữ liệu khí tượng (GHI, TEM, WIS, HUM) + Dữ liệu Bức xạ trời trong tính toán tại khu vực vị trí nhà máy từ mô hình Ineichen/Perez bằng công cụ PVLIB
Dữ liệu đầu ra	Công suất phát nhà máy	Công suất phát nhà máy
Cấu hình mạng LSTM	Tương tự	Tương tự
Huấn luyện	Huấn luyện sử dụng hàm mất mát MAE, hàm kích hoạt ReLU, bộ tối ưu hóa Adam và 50 epochs, sử dụng kỹ thuật Early Stopping.	Huấn luyện sử dụng hàm mất mát MAE, hàm kích hoạt ReLU, bộ tối ưu hóa Adam và 50 epochs, sử dụng kỹ thuật Early Stopping.
Hàm mất mát	MAE	MAE
Huấn luyện	Validation 10%	Validation 10%

3.3.3.4. Kết quả và phân tích

Kết quả dự báo trên tập dữ liệu kiểm tra được so sánh với công suất phát thực tế để đánh giá sai số, kết quả trình bày trong Bảng 3.6.

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

Mô hình	Khung dự báo	MAE	MAPE	MSE	RMSE	nRMSE
	Ngày tới	5,909	12,310	80,966	8,988	18,745

Từ bảng kết quả trên có thể thấy:

- Đối với dự báo chu kỳ hiện tại, Mô hình sử dụng bức xạ trời trong có giá trị MAPE là 2,728%, thấp hơn 0,214% so với sai số MAPE của Mô hình sử dụng chỉ dấu thời gian. Mức giảm nRMSE cũng là khoảng 0,165%. Sai số có cải thiện giảm nhẹ.
- Đối với dự báo 1 chu kỳ tiếp theo, Mô hình sử dụng bức xạ trời trong có giá trị MAPE là 4,388%, thấp hơn 2,161% so với sai số MAPE của Mô hình sử dụng chỉ dấu thời gian. Mức giảm nRMSE cũng là khoảng 1,381%. Sai số cải thiện giảm đáng kể.
- Đối với dự báo 1 ngày tới, Mô hình sử dụng bức xạ trời trong có giá trị MAPE là 4,388%, thấp hơn 1,592% so với sai số MAPE của Mô hình sử dụng chỉ dấu thời gian. Mức giảm nRMSE cũng là khoảng 0,223%. Sai số MAPE cải thiện đáng kể, sai số nRMSE cải thiện giảm nhẹ.

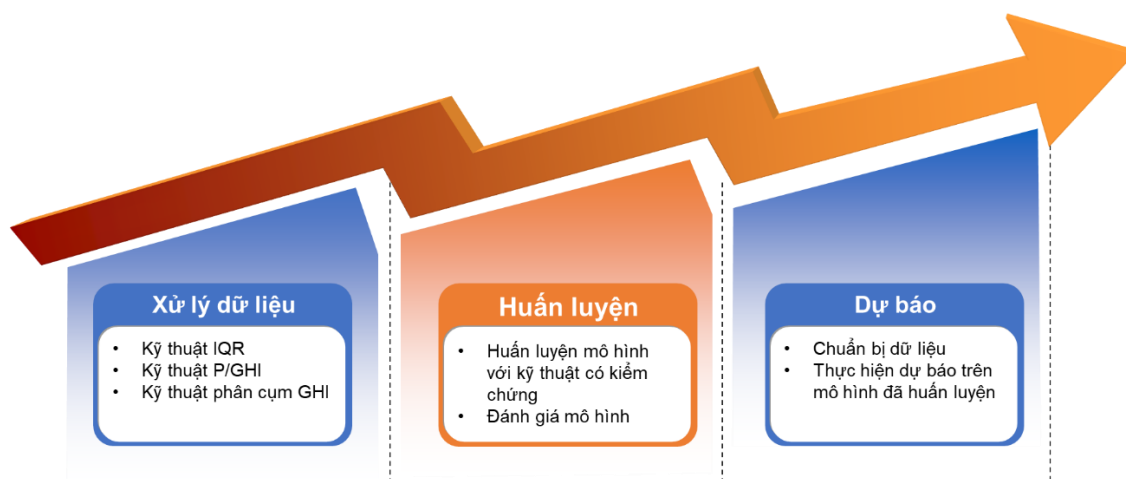
3.4. Quy trình và công cụ xử lý bài toán dự báo ngắn hạn công suất ĐMT sử dụng mạng Nơ-ron hồi quy

Nội dung này trình bày cụ thể quy trình và công cụ xử lý cụ thể một bài toán dự báo ngắn hạn công suất nhà máy ĐMT sử dụng mạng nơ-ron hồi quy mà luận án đã tổng kết được qua quá trình nghiên cứu thực nghiệm.

Quy trình này giúp xác định các bước cần thiết để xây dựng và đào tạo mô hình dự báo. Quy trình cung cấp hướng dẫn chi tiết từ việc chuẩn bị dữ liệu đến huấn luyện mô hình, kiểm định, dự báo và cải thiện chất lượng mô hình. Nó giúp đảm bảo tính nhất quán và hiệu quả của quá trình dự báo.

Về mặt công cụ, hiện nay có nhiều công cụ có khả năng ứng dụng để giải quyết bài toán tuy nhiên, mỗi công cụ sẽ có ưu, nhược điểm riêng. Việc đề xuất lựa chọn các giải pháp phù hợp và xây dựng thành một công cụ phần mềm tích hợp sẽ là khuyến nghị hữu ích cho các nhà nghiên cứu, chuyên gia và người dùng trong lĩnh vực dự báo công suất ĐMT, cung cấp một công cụ tiện lợi và hiệu quả để huấn luyện, dự báo và đánh giá hiệu quả mô hình, phục vụ công tác quản lý vận hành.

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:



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

3.4.1. Bước 1 - Xử lý dữ liệu đầu vào

3.4.1.1. Mô tả

Dữ liệu đầu vào phục vụ việc huấn luyện mô hình LSTM bao gồm:

- Dữ liệu công suất phát trung bình của nhà máy ĐMT (đơn vị MW), thể hiện công suất phát đầu ra đo được của nhà máy ĐMT trung bình trong một khoảng thời gian nhất định (ví dụ: trung bình trong 05 phút, trung bình trong 15 phút...).

- Dữ liệu bức xạ toàn phần (GHI) đo được tại vị trí bề mặt tấm pin của hệ thống (đơn vị W/m^2), thể hiện chỉ số bức xạ mặt trời toàn phần tại vị trí bề mặt tấm pin trung bình trong một khoảng thời gian nhất định (ví dụ: trung bình trong 05 phút, trung bình trong 15 phút...)

- Dữ liệu nhiệt độ môi trường đo được tại vị trí bề mặt tấm pin của nhà máy (đơn vị $^{\circ}C$) trong một khoảng thời gian nhất định (ví dụ: trung bình trong 05 phút, trung bình trong 15 phút...).

Việc xử lý dữ liệu đầu vào là bước bắt buộc đối với tất cả các phương pháp huấn luyện mô hình LSTM nhằm mục đích loại bỏ các dữ liệu nhiễu, xấu thường gặp trong bộ dữ liệu của một nhà máy ĐMT.

Trong điều kiện nhà máy điện được phát toàn bộ công suất thì công suất phát phụ thuộc vào khả năng của hệ thống điện mặt trời, tức là phụ thuộc vào GHI (bức xạ toàn phần, tính bằng đơn vị W/m^2) và một số điều kiện thời tiết khác. Tuy nhiên, khi nhà máy điện mặt trời bị giới hạn công suất hay sự cố thì xảy ra hiện tượng tập dữ liệu thu được có nhiều điểm dữ liệu không đáng tin cậy như có giá trị P (công suất phát, MW) khi GHI bằng 0, hoặc tỷ lệ giữa GHI và P là bất thường, GHI lớn nhưng công suất phát P lại nhỏ, không tương xứng.

Việc xử lý số liệu đầu vào được thực hiện bằng kỹ thuật phân chia bức xạ GHI và hệ số tỷ lệ P/GHI giữa công suất phát P và bức xạ GHI.

Sau khi dữ liệu được loại bỏ các thành phần xấu, nhiễu, tức là các dữ liệu thuộc điểm ngoại lai, nó sẽ được xử lý tiếp bằng cách chia thành bộ dữ liệu huấn luyện và bộ dữ liệu kiểm tra mô hình. Cuối cùng là việc biến đổi bộ dữ liệu huấn luyện, bộ dữ liệu kiểm tra và bộ dữ liệu kiểm chứng thành dạng đầu vào tiêu chuẩn của mô hình LSTM.

3.4.1.2. Quy trình Xử lý dữ liệu

Quy trình xử lý dữ liệu được thiết kế để tiền xử lý dữ liệu đầu vào trước khi huấn luyện mô hình LSTM. Dưới đây là mô tả thiết kế quy trình xử lý dữ liệu:

- ❖ Bước 1: Đọc dữ liệu đầu vào
 - ✓ Mô tả: Quy trình bắt đầu bằng việc đọc dữ liệu từ file CSV đầu vào.
 - ✓ Hoạt động:
 - Mở file CSV và đọc nội dung dữ liệu vào bộ nhớ.
 - Lưu trữ dữ liệu vào biến hoặc cấu trúc dữ liệu phù hợp để tiếp tục xử lý.
- ❖ Bước 2: Phân tích dữ liệu
 - ✓ Mô tả: Bước này thực hiện phân tích dữ liệu đầu vào để hiểu cấu trúc và tính chất của nó.
 - ✓ Hoạt động:
 - Kiểm tra và xử lý dữ liệu thiếu, bị lặp lại hoặc nhiễu.
 - Tiến hành nghiên cứu dữ liệu, bao gồm việc tính toán các thống kê mô tả, tạo đồ thị và biểu đồ để trực quan hóa dữ liệu.
- ❖ Bước 3: Làm sạch dữ liệu
 - ✓ Mô tả: Bước này nhằm loại bỏ các nhiễu và dữ liệu không hợp lệ để chuẩn bị cho quá trình huấn luyện mô hình.
 - ✓ Hoạt động:
 - Loại bỏ các dòng dữ liệu bị thiếu hoặc không hợp lệ.
 - Thực hiện điều chỉnh và chuẩn hóa dữ liệu nếu cần thiết.
 - Xử lý và loại bỏ nhiễu, điểm ngoại lai trong dữ liệu.
- ❖ Bước 4: Biến đổi dữ liệu
 - ✓ Mô tả: Bước này thực hiện các biến đổi và chuyển đổi dữ liệu để phù hợp với yêu cầu của mô hình LSTM.
 - ✓ Hoạt động:
 - Áp dụng kỹ thuật trích xuất đặc trưng (feature extraction) nếu cần thiết.
 - Chuyển đổi dữ liệu thành định dạng phù hợp với mô hình LSTM, chẳng hạn như chuỗi thời gian.

❖ **Bước 5: Chuẩn bị dữ liệu huấn luyện**

- ✓ Mô tả: Tạo ra các tập dữ liệu huấn luyện và kiểm tra từ dữ liệu đã được xử lý.
- ✓ Hoạt động:
 - Chia dữ liệu thành tập huấn luyện và tập kiểm tra theo tỷ lệ xác định.
 - Chuẩn bị dữ liệu thành định dạng phù hợp cho huấn luyện mô hình LSTM.

3.4.2. Bước 2 - Huấn luyện mô hình

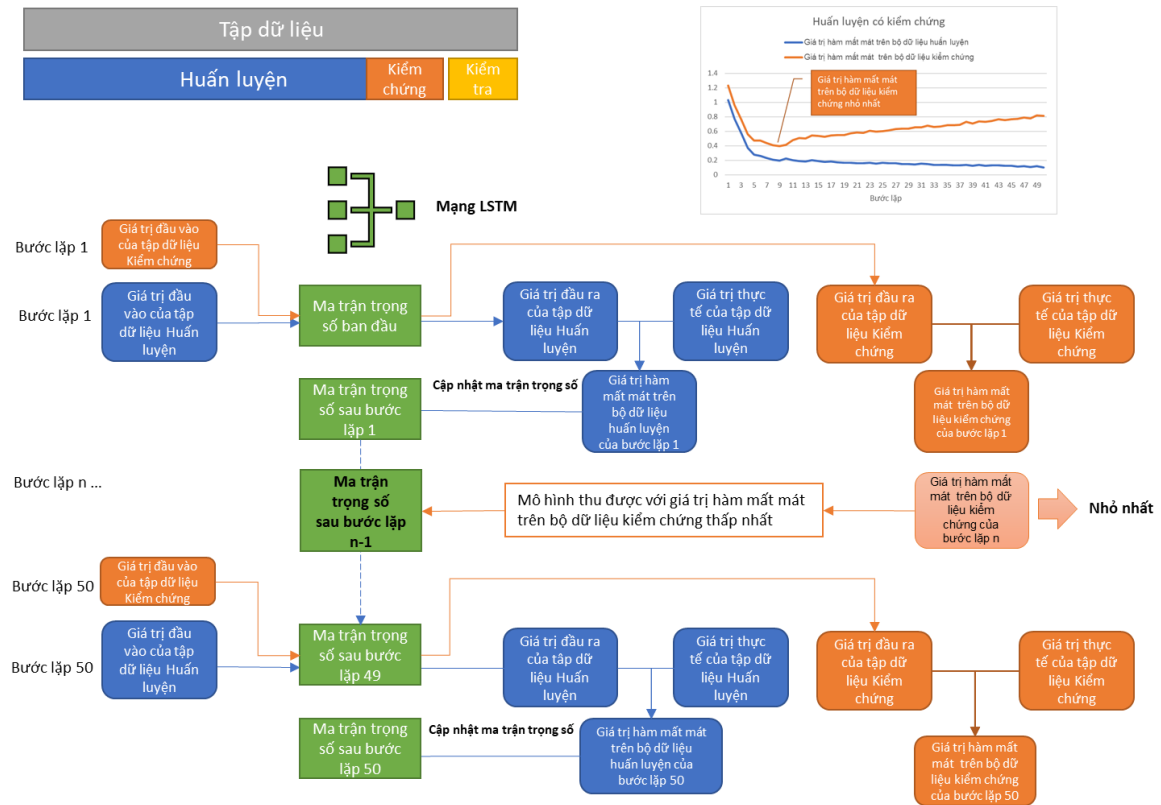
3.4.2.1. Mô tả

Thông thường, việc huấn luyện mô hình LSTM sẽ được thực hiện bằng cách sử dụng toàn bộ dữ liệu huấn luyện để tìm ra ma trận trọng số. Sau đó, sử dụng ma trận trọng số tìm được để tính toán dựa trên bộ dữ liệu kiểm tra để kiểm tra xem kết quả dự báo của mô hình có chấp nhận được không.

Tuy nhiên, với cách làm như trong tình trạng kỹ thuật nêu trên sẽ xuất hiện hiện tượng quá khớp (overfit), tức là hiện tượng xảy ra khi ma trận trọng số thu được sau khi huấn luyện mô hình bằng bộ dữ liệu huấn luyện thể hiện rất tốt trên tập dữ liệu huấn luyện nhưng khi áp dụng ma trận trọng số đó vào mô hình LSTM để chạy trên bộ dữ liệu kiểm tra thì kết quả lại không như mong muốn.

Việc tách một phần bộ dữ liệu huấn luyện để tạo thành bộ dữ liệu kiểm chứng như đã trình bày ở bên trên giúp cải thiện tình trạng này. Sau đó, sử dụng kết hợp bộ dữ liệu huấn luyện và bộ dữ liệu kiểm chứng phục vụ việc huấn luyện mô hình LSTM để tìm ra mô hình trọng số tối ưu, như Hình 3.12. Sơ đồ quá trình huấn luyện mô hình.

Quá trình huấn luyện sẽ được thực hiện theo các bước lặp (epoch). Các dữ liệu phục vụ huấn luyện trong bước lặp được nạp vào mô hình LSTM với ma trận trọng số mặc định để tìm ra giá trị sai số huấn luyện (hay chính là hàm mất mát MAE), sau đó ma trận trọng số được điều chỉnh tương ứng; ngay sau đó, bộ dữ liệu kiểm chứng cũng được nạp vào mô hình LSTM với ma trận trọng số mặc định và tìm ra giá trị sai số kiểm chứng (hay chính là hàm mất mát MAE) mà không chờ đến khi toàn bộ dữ liệu huấn luyện được sử dụng hết.



Hình 3.12. Sơ đồ quá trình huấn luyện mô hình

Việc huấn luyện mô hình LSTM theo các bước lặp sẽ được thực hiện cụ thể như sau:

Bước lặp (Epoch) đầu tiên- tương ứng với Epoch 1 trên hình: ứng với ma trận hệ số ban đầu (ma trận hệ số mặc định) của Mô hình LSTM, các dữ liệu đầu vào (dữ liệu về thời tiết, bức xạ GHI, nhiệt độ môi trường, độ ẩm...) của bộ dữ liệu huấn luyện được sử dụng để dự báo công suất phát của nhà máy điện mặt trời. Kết quả công suất phát dự báo huấn luyện được so sánh với giá trị công suất phát thực tế của nhà máy để có thể tính ra các giá trị sai số huấn luyện của bước lặp thứ nhất theo các phương pháp tính sai số. Căn cứ trên việc đánh giá các giá trị sai số này, mô hình cập nhật, thay đổi ma trận trọng số mới của mô hình LSTM.

Đồng thời, ma trận hệ số ban đầu cũng được sử dụng cho mô hình LSTM để tính toán bộ dữ liệu kiểm chứng. Kết quả dự báo công suất phát kiểm chứng của mô hình lúc này được so sánh với giá trị công suất phát thực tế của nhà máy để có thể tính ra giá trị sai số kiểm chứng của bước lặp đầu tiên theo các phương pháp tính sai số. Giá trị sai số kiểm chứng này được lưu lại trong bộ nhớ.

Bước lặp tiếp theo - tương ứng với epoch n trong hình: ứng với ma trận trọng số đã được cập nhật sau bước lặp liền trước của Mô hình LSTM, các dữ liệu đầu vào (dữ

liệu về thời tiết, bức xạ GHI, nhiệt độ, độ ẩm...) của bộ dữ liệu huấn luyện tiếp tục được sử dụng để dự báo công suất phát của nhà máy điện mặt trời. Kết quả công suất phát dự báo huấn luyện (Output train) được so sánh với giá trị công suất phát thực tế của nhà máy để có thể tính ra các giá trị sai số huấn luyện của tiếp theo theo các phương pháp tính sai số. Căn cứ trên việc đánh giá các giá trị sai số này, mô hình tiếp tục cập nhật, thay đổi ma trận hệ số mới của mô hình LSTM.

Đồng thời, ma trận trọng số đã được cập nhật sau bước lặp liền trước của mô hình LSTM cũng được sử dụng để tính toán bộ dữ liệu kiểm chứng. Kết quả dự báo công suất phát kiểm chứng của mô hình lúc này được so sánh với giá trị công suất phát thực tế của nhà máy để có thể tính ra giá trị sai số kiểm chứng của bước lặp tiếp theo theo các phương pháp tính sai số và giá trị sai số kiểm chứng này được lưu lại trong bộ nhớ.

Việc huấn luyện mô hình LSTM được thực hiện với số lượng bước lặp định trước và thu được hai đại lượng là giá trị sai số huấn luyện và giá trị sai số kiểm chứng. Kết thúc quá trình huấn luyện, tức là sau khi toàn bộ dữ liệu huấn luyện trong các bước lặp với số lượng định trước và bộ dữ liệu kiểm chứng tương ứng được nạp vào mô hình LSTM, căn cứ vào đồ thị hàm giá trị sai số huấn luyện và hàm giá trị sai số kiểm chứng, có thể rút ra được xu hướng chất lượng của mô hình qua từng bước lặp. Và ma trận trọng số nào cho ra kết quả giá trị sai số kiểm chứng nhỏ nhất được coi là ma trận trọng số cần tìm kiếm.

Trong quá trình huấn luyện, bộ dữ liệu kiểm chứng (validation dataset) được sử dụng để đánh giá hiệu suất của mô hình sau mỗi vòng lặp huấn luyện. Bộ dữ liệu kiểm chứng tăng cường khả năng tổng quát hóa của mô hình trên dữ liệu mới, mà nó chưa từng thấy trong quá trình huấn luyện. Khi huấn luyện, dữ liệu huấn luyện được sử dụng để điều chỉnh các trọng số và siêu tham số của mô hình. Sau mỗi vòng lặp huấn luyện, mô hình được đánh giá trên bộ dữ liệu kiểm chứng để đo lường hiệu suất. Điều này giúp theo dõi sự tiến bộ của mô hình và đánh giá các điểm yếu cần cải thiện.

Việc huấn luyện mô hình LSTM được thực hiện với điều kiện dừng sớm (early stopping) để giảm thiểu thời gian huấn luyện. Cụ thể, nếu thấy xu hướng chất lượng mô hình qua từng bước lặp trên tập validation đã đạt mức tốt nhất, không tiếp tục cải thiện thì việc huấn luyện sẽ dừng lại và ma trận trọng số nào cho ra kết quả giá trị sai số kiểm chứng nhỏ nhất được coi là ma trận trọng số cần tìm kiếm.

3.4.2.2. *Đánh giá mô hình*

Dữ liệu về thời tiết, bức xạ GHI, nhiệt độ, độ ẩm... của bộ dữ liệu kiểm tra được nạp vào Mô hình LSTM với ma trận trọng số thu được để dự báo công suất phát của nhà máy điện mặt trời.

Kết quả công suất phát dự báo được so sánh với giá trị công suất phát thực tế của nhà máy điện mặt trời có trong bộ dữ liệu kiểm tra để đánh giá sai số.

Trong trường hợp kết quả dự báo có độ chính xác thấp hơn kết quả dự báo bằng các phương pháp dự báo khác thì việc huấn luyện được thực hiện lại từ Bước 1.

3.4.2.3. *Quy trình Huấn luyện mô hình*

Quy trình huấn luyện mô hình được mô tả như dưới đây:

- ❖ Bước 1: Chuẩn bị dữ liệu huấn luyện
 - ✓ Mô tả: Bắt đầu bằng việc chuẩn bị dữ liệu huấn luyện cho mô hình LSTM.
 - ✓ Hoạt động:
 - Sử dụng dữ liệu đã được xử lý và biến đổi từ quy trình xử lý dữ liệu.
 - Chia dữ liệu thành các nhóm để huấn luyện mô hình một cách hiệu quả.
- ❖ Bước 2: Xây dựng mô hình LSTM
 - ✓ Mô tả: Bước này tạo ra mô hình LSTM để huấn luyện trên dữ liệu.
 - ✓ Hoạt động:
 - Xác định kiến trúc mô hình LSTM, bao gồm số lớp LSTM, số đơn vị (units), số lượng lớp ẩn, và các tham số khác.
 - Xác định hàm mất mát (loss function) và thuật toán tối ưu hóa để huấn luyện mô hình.
- ❖ Bước 3: Huấn luyện mô hình
 - ✓ Mô tả: Bước này thực hiện quá trình huấn luyện mô hình LSTM trên dữ liệu huấn luyện.
 - ✓ Hoạt động:
 - Sử dụng thuật toán tối ưu hóa và hàm mất mát đã được xác định để điều chỉnh các trọng số trong mô hình LSTM.
 - Lặp lại quá trình huấn luyện trên các batch dữ liệu cho đến khi đạt được điều kiện dừng. Điều kiện dừng là khi giá trị hàm mất mát trên tập dữ liệu kiểm tra (validation dataset) không cải thiện sau một số bước lặp xác định (được ấn định trước với thông lệ khoảng 10-20 epoch).
- ❖ Bước 4: Đánh giá mô hình
 - ✓ Mô tả: Đánh giá hiệu suất của mô hình huấn luyện trên dữ liệu kiểm tra.
 - ✓ Hoạt động:

- Sử dụng dữ liệu kiểm tra để đánh giá mô hình, tính toán các độ đo đánh giá như độ chính xác, RMSE, MAE, MAPE, và các độ đo khác.
 - Vẽ biểu đồ hoặc đồ thị để trực quan hóa kết quả đánh giá mô hình.
- ❖ **Bước 5: Lưu trữ mô hình**
- ✓ **Mô tả:** Quy trình kết thúc bằng việc lưu trữ mô hình đã được huấn luyện để sử dụng sau này.
 - ✓ **Hoạt động:**
 - Lưu trữ các thông số và trọng số của mô hình LSTM vào một tệp.
 - Ghi lại các thông tin quan trọng như kiến trúc mô hình, tham số huấn luyện, và độ đo đánh giá để sử dụng và tham khảo trong tương lai.

3.4.3. *Bước 3 – Dự báo*

3.4.3.1. *Mô tả*

Bước này sử dụng mô hình LSTM đã được huấn luyện để thực hiện dự báo công suất phát trong tương lai. Các công việc cần thực hiện như sau:

- ✓ **Chuẩn bị dữ liệu dự báo:** Trước khi thực hiện dự báo, cần chuẩn bị dữ liệu đầu vào cho mô hình. Điều này bao gồm việc chuẩn hóa, mã hóa và xử lý dữ liệu dự báo để đảm bảo rằng chúng tương thích với định dạng dữ liệu mà mô hình đã được huấn luyện.
- ✓ **Áp dụng mô hình huấn luyện:** Sau khi chuẩn bị dữ liệu dự báo, mô hình LSTM đã được huấn luyện được áp dụng để thực hiện dự báo công suất phát. Mô hình sẽ nhận dữ liệu đầu vào và tạo ra dự báo dựa trên các mẫu và quy luật mà nó đã học được từ quá trình huấn luyện.

3.4.3.2. *Quy trình Dự báo*

Quy trình dự báo được trình bày như dưới đây:

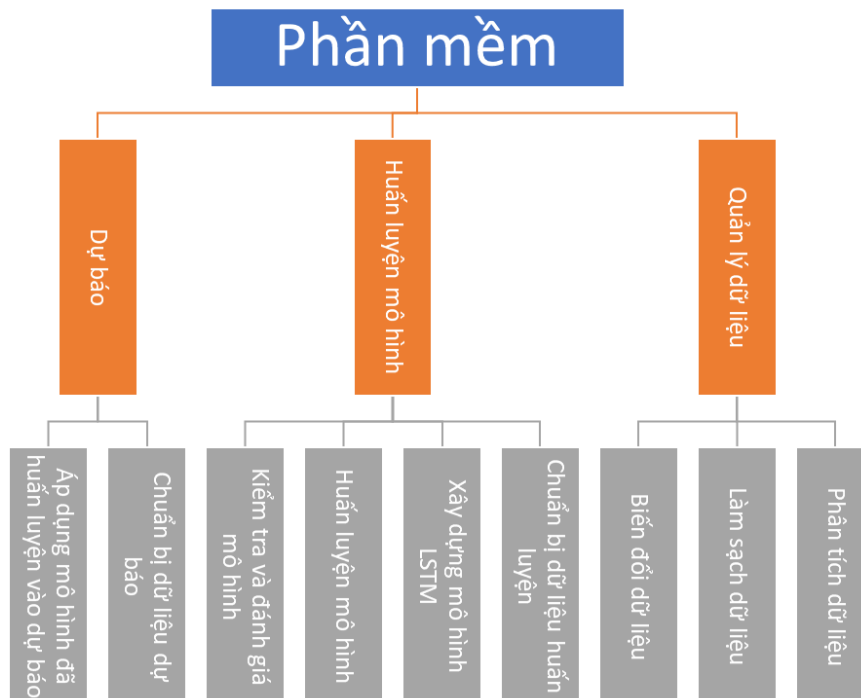
- ❖ **Bước 1: Chuẩn bị dữ liệu dự báo**
- ✓ **Mô tả:** Bắt đầu bằng việc chuẩn bị dữ liệu đầu vào cho quá trình dự báo.
 - ✓ **Hoạt động:**
 - Sử dụng dữ liệu mới nhất để tạo dữ liệu đầu vào cho dự báo.
 - Áp dụng các quy tắc và quy trình xử lý dữ liệu đã được áp dụng trong quy trình xử lý dữ liệu để chuẩn bị dữ liệu dự báo.
- ❖ **Bước 2: Tải mô hình huấn luyện**
- ✓ **Mô tả:** Bước này bao gồm tải mô hình đã được huấn luyện từ quy trình huấn luyện mô hình.
 - ✓ **Hoạt động:**
 - Truy cập vào tệp hoặc cơ sở dữ liệu lưu trữ mô hình huấn luyện.
 - Tải các thông số và trọng số của mô hình để sử dụng cho quá trình dự báo.

- ❖ Bước 3: Dự báo công suất
 - ✓ Mô tả: Bước này thực hiện dự báo công suất phát nhà máy điện mặt trời sử dụng mô hình đã được huấn luyện.
 - ✓ Hoạt động:
 - Áp dụng mô hình LSTM đã được tải để dự báo công suất phát.
 - Sử dụng dữ liệu đầu vào chuẩn bị từ bước trước để thực hiện dự báo.
- ❖ Bước 4: Xuất kết quả dự báo
 - ✓ Mô tả: Quá trình kết thúc bằng việc xuất kết quả dự báo công suất.
 - ✓ Hoạt động:
 - Hiện thị kết quả dự báo công suất trên giao diện người dùng hoặc lưu trữ vào tệp hoặc cơ sở dữ liệu.
 - Cung cấp thông tin về dự báo công suất như thời gian dự báo, giá trị dự báo, và các độ đo đánh giá.

3.4.4. Công cụ phần mềm dự báo công suất ĐMT

Sơ đồ chức năng chính của phần mềm được trình bày trong Hình 3.13, công cụ phần mềm có 3 khối chức năng chính bao gồm:

- Quản lý dữ liệu: gồm các chức năng phân tích dữ liệu, làm sạch dữ liệu và biến đổi dữ liệu
- Huấn luyện mô hình: gồm các chức năng huấn luyện và kiểm tra mô hình
- Dự báo: thực hiện dự báo từ mô hình đã huấn luyện



Hình 3.13. Sơ đồ chức năng chính của Phần mềm

3.4.4.1. Chức năng Quản lý dữ liệu

Phần chức năng "Quản lý dữ liệu" bao gồm các công việc phân tích dữ liệu, làm sạch dữ liệu và biến đổi dữ liệu, nhằm đảm bảo dữ liệu đầu vào cho mô hình huấn luyện và dự báo là chính xác và hợp lệ. Cụ thể, phần chức năng "Quản lý dữ liệu" có các chức năng con sau:

- Phân tích dữ liệu: Trước khi xử lý dữ liệu, cần phân tích và hiểu rõ về dữ liệu đầu vào. Quá trình phân tích dữ liệu bao gồm việc khám phá cấu trúc và đặc điểm của dữ liệu, xác định các thuộc tính quan trọng, kiểm tra tính đầy đủ và chính xác của dữ liệu, cũng như phát hiện và xử lý các giá trị thiếu, ngoại lai, hoặc các vấn đề khác trong dữ liệu.
- Làm sạch dữ liệu: Dữ liệu thô thường chứa nhiễu và các giá trị không hợp lệ. Quá trình làm sạch dữ liệu bao gồm loại bỏ các giá trị trùng lặp, điền giá trị thiếu, loại bỏ nhiễu và xử lý các giá trị ngoại lai. Điều này giúp làm sạch dữ liệu và cung cấp dữ liệu chính xác và đáng tin cậy cho việc huấn luyện và dự báo. Đặc biệt trong phần mềm này tích hợp kỹ thuật xử lý dữ liệu căn cứ theo P/GHI với việc phân khoảng các cụm dữ liệu GHI.
- Biến đổi dữ liệu: Dữ liệu ban đầu cần được biến đổi hoặc chuẩn hóa để phù hợp với yêu cầu của mô hình. Các biến đổi dữ liệu có thể bao gồm chuyển đổi đơn vị đo lường, chuẩn hóa dữ liệu về cùng một khoảng giá trị, mã hóa biến đầu vào thành dạng số, hoặc áp dụng các phép biến đổi khác để tạo ra dữ liệu có ý nghĩa và hữu ích cho mô hình huấn luyện và dự báo. Đặc biệt, trong phần mềm này còn tích hợp chức năng tính toán bức xạ trời trong để làm đầu vào mới tăng cường hiệu quả của mô hình dự báo.

3.4.4.2. Chức năng Huấn luyện mô hình

Phần chức năng "Huấn luyện mô hình" bao gồm các công việc huấn luyện mô hình và kiểm tra độ chính xác của mô hình, nhằm đảm bảo rằng mô hình có khả năng dự báo chính xác và tin cậy. Cụ thể, phần chức năng "Huấn luyện mô hình" có các chức năng sau:

- Chuẩn bị dữ liệu huấn luyện: Trước khi huấn luyện mô hình, cần chuẩn bị dữ liệu huấn luyện phù hợp, đúng form mẫu.
- Xây dựng mô hình LSTM: Trong phần chức năng này, mô hình LSTM được xây dựng với các lớp LSTM và số lượng nút ẩn trên mỗi lớp khác nhau.
- Huấn luyện mô hình: Sau khi xây dựng mô hình, quá trình huấn luyện được thực hiện bằng cách đưa dữ liệu huấn luyện vào mô hình và cập nhật các trọng số mô hình dựa trên độ sai số giữa dự đoán và giá trị thực tế. Quá trình huấn luyện được lặp lại qua nhiều epoch để mô hình học và cải thiện dự đoán theo thời gian.

- Kiểm tra và đánh giá mô hình: Sau khi huấn luyện, mô hình được kiểm tra bằng cách đưa dữ liệu kiểm tra vào mô hình và đánh giá hiệu suất của mô hình bằng các độ đo như sai số trung bình, độ chính xác, hệ số xác định và các độ đo khác tùy thuộc vào yêu cầu của bài toán.

3.4.4.3. Chức năng Dự báo

Chức năng "Dự báo" sử dụng mô hình LSTM đã được huấn luyện để thực hiện dự báo công suất phát trong tương lai gồm:

- Chuẩn bị dữ liệu dự báo: Trước khi thực hiện dự báo, cần chuẩn bị dữ liệu đầu vào cho mô hình. Bước này gồm việc chuẩn hóa, mã hóa và xử lý dữ liệu dự báo để đảm bảo rằng chúng tương thích với định dạng dữ liệu mà mô hình đã được huấn luyện.
- Áp dụng mô hình huấn luyện: Sau khi chuẩn bị dữ liệu dự báo, mô hình LSTM đã được huấn luyện được áp dụng để thực hiện dự báo công suất phát. Mô hình sẽ nhận dữ liệu đầu vào và tạo ra dự báo theo đúng form mẫu yêu cầu.

KẾT LUẬN VÀ HƯỚNG NGHIÊN CỨU TIẾP THEO

1. KẾT LUẬN

Luận án “**Nghiên cứu mô hình dự báo ngắn hạn công suất phát của nhà máy điện mặt trời sử dụng mạng nơ-ron hồi quy**” đã hoàn thành toàn bộ nội dung nghiên cứu đề ra 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 (real-time):

Với cách tiếp cận hiện tại, mô hình đã được huấn luyện tốt có thể được sử dụng trong thực tế với số liệu đầu vào liên tục được cập nhật mà không cần phải huấn luyện lại. Trong thực tế, sự phát triển của các kỹ thuật học máy hiện đại cho phép mô hình có khả năng huấn luyện nâng cao độ chính xác chỉ với một số lượng tập mẫu dữ liệu bổ sung nhỏ mà không cần huấn luyện lại trên toàn bộ tập dữ liệu với thời gian tính toán lớn. Hai hướng có thể kể đến là:

- Online Learning (Huấn luyện trực tuyến): là phương pháp học máy cho phép mô hình được cập nhật mỗi khi có dữ liệu mới. Mô hình được huấn luyện ban đầu bằng tập dữ liệu huấn luyện, và sau đó nó có thể tiếp tục học từ dữ liệu mới một cách trực tuyến.
- Real-Time Learning (Huấn luyện thời gian thực): là một khía cạnh của Online learning, nhưng nó đặc trưng hơn bởi việc cập nhật mô hình ngay khi dữ liệu mới xuất hiện. Mô hình thường được huấn luyện với dữ liệu ban đầu và sau đó liên tục điều chỉnh mô hình bằng cách học từ dữ liệu mới ngay khi nó xuất hiện.

Các kỹ thuật nâng cao chất lượng mô hình:

Hướng phát triển tiếp theo của luận án là nghiên cứu các mô hình học sâu mới dựa trên LSTM cũng như phương pháp như tối ưu hóa thuật toán, tinh chỉnh siêu tham số, tăng cường dữ liệu huấn luyện và kiểm soát quá khớp mới nhất để cải thiện chất lượng mô hình. Ngoài ra, với các kỹ thuật đã đề xuất trong luận án, hướng nghiên cứu tiếp theo sẽ bao gồm việc thí nghiệm đánh giá giải pháp tích hợp các kỹ thuật đã đưa ra sẽ đạt được hiệu quả tổng hợp như thế nào.

Hoàn thiện quy trình và công cụ:

Mục tiêu tiếp theo của nghiên cứu là triển khai mô hình dự báo vào thực tế nâng cao tính ứng dụng. Đồng thời, nghiên cứu các hướng mở rộng như web hóa để công khai và chia sẻ giải pháp dự báo công suất ĐMT với cộng đồng người dùng, hướng tới hoàn thiện hơn nữa sản phẩm.

Nghiên cứu mô hình ứng dụng ảnh bầu trời để dự báo sụt giảm ngắn hạn

Mô hình dự báo kết quả thu được của Luận án được xây dựng sử dụng phương pháp mạng Nơ-ron LSTM và huấn luyện mạng trên bộ dữ liệu thực tế. Về mặt lý thuyết, khi lượng dữ liệu đủ lớn thì mô hình có khả năng nhận diện được các yếu tố này. Tuy nhiên, trong điều kiện thực tế nghiên cứu và ứng dụng ở Việt Nam, các đầu vào có thể thu thập được chủ yếu là bức xạ, nhiệt độ, độ ẩm, tốc độ gió tại vị trí nhà máy. Những yếu tố khí tượng này chỉ mô phỏng được một phần sự biến động của môi trường xung quanh. Với tính chất độ phân giải dữ liệu nhỏ (5 phút), các sụt giảm trong phạm vi ngắn một vài chu kỳ tiếp theo có thể được mô hình nhận ra, tuy nhiên mô hình không thể dự báo được các sụt giảm ở phạm vi dài như trong một vài tiếng tới. Hướng nghiên cứu để xử lý phạm vi này có thể là sử dụng thêm ảnh chụp bầu trời (sky-image) tại vị trí nhà máy. Hướng triển khai này cần có đầu tư trang bị thêm thiết bị nên hiện chưa nằm trong phạm vi nghiên cứu đã đặt ra của đề tài và sẽ nằm trong các hướng nghiên cứu tiếp theo của đề tài

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 ĐỂ 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*).

TÀI LIỆU THAM KHẢO

- [1] EVNNLDC, “Báo cáo vận hành hệ thống điện và thị trường điện Việt Nam 2022,” 2022.
- [2] Thủ Tướng Chính Phủ, “Quyết định Phê duyệt Quy hoạch phát triển điện lực quốc gia thời kỳ 2021-2030, tầm nhìn đến năm 2050 (Quy hoạch điện VIII).” 2023.
- [3] Linh Duy Bui and Quynh Pham, “Vietnam, a System in Rapid Transition - ESIG Spring Technical Workshop 2023.” Accessed: May 21, 2023. [Online]. Available: <https://www.esig.energy/event/2023-spring-technical-workshop/>
- [4] Ahmed G. Abo-Khalil, “Impacts of Wind Farms on Power System Stability,” in *Modeling and Control Aspects of Wind Power Systems*, InTech, 2013. doi: 10.5772/55090.
- [5] S. Sarwar *et al.*, “A Novel Hybrid MPPT Technique to Maximize Power Harvesting from PV System under Partial and Complex Partial Shading,” *Applied Sciences (Switzerland)*, vol. 12, no. 2, Jan. 2022, doi: 10.3390/app12020587.
- [6] F. Antonanzas-Torres, R. Urraca, J. Polo, O. Perpiñán-Lamigueiro, and R. Escobar, “Clear sky solar irradiance models: A review of seventy models,” *Renewable and Sustainable Energy Reviews*, vol. 107. Elsevier Ltd, pp. 374–387, Jun. 01, 2019. doi: 10.1016/j.rser.2019.02.032.
- [7] P. Ineichen, “Validation of models that estimate the clear sky global and beam solar irradiance,” *Solar Energy*, vol. 132, pp. 332–344, Jul. 2016, doi: 10.1016/j.solener.2016.03.017.
- [8] R. Perez *et al.*, “A new operational model for satellite-derived irradiances: description and validation,” *Solar Energy*, vol. 73, no. 5, pp. 307–317, Nov. 2002, doi: 10.1016/S0038-092X(02)00122-6.
- [9] P. Ineichen and R. Perez, “A NEW AIRMASS INDEPENDENT FORMULATION FOR THE LINKE TURBIDITY COEFFICIENT,” 2002. [Online]. Available: www.elsevier.com/locate/solener
- [10] M. J. Reno, C. W. Hansen, and J. S. Stein, “Global Horizontal Irradiance Clear Sky Models: Implementation and Analysis,” 2012. [Online]. Available: <http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online>
- [11] P. Ineichen, “A broadband simplified version of the Solis clear sky model,” *Solar Energy*, vol. 82, no. 8, pp. 758–762, 2008, doi: 10.1016/j.solener.2008.02.009.

- [12] J. Cubas, S. Pindado, and C. de Manuel, “Explicit Expressions for Solar Panel Equivalent Circuit Parameters Based on Analytical Formulation and the Lambert W-Function,” MDPI AG, Mar. 2014, p. c013. doi: 10.3390/ece-1-c013.
- [13] U. K. Das *et al.*, “Forecasting of photovoltaic power generation and model optimization: A review,” *Renewable and Sustainable Energy Reviews*, vol. 81. Elsevier Ltd, pp. 912–928, Jan. 01, 2018. doi: 10.1016/j.rser.2017.08.017.
- [14] RENAC, “PV Application,” 2017.
- [15] S. S. Reddy, “Optimal scheduling of thermal-wind-solar power system with storage,” *Renew Energy*, vol. 101, pp. 1357–1368, Feb. 2017, doi: 10.1016/j.renene.2016.10.022.
- [16] S. S. Reddy and P. R. Bijwe, “Real time economic dispatch considering renewable energy resources,” *Renew Energy*, vol. 83, pp. 1215–1226, Nov. 2015, doi: 10.1016/j.renene.2015.06.011.
- [17] R. Ahmed, V. Sreeram, Y. Mishra, and M. D. Arif, “A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization,” *Renewable and Sustainable Energy Reviews*, vol. 124. Elsevier Ltd, May 01, 2020. doi: 10.1016/j.rser.2020.109792.
- [18] Antonio Luque and Steven Hegedus, *Handbook of Photovoltaic Science and Engineering*. Chichester, UK: John Wiley & Sons, Ltd, 2003. doi: 10.1002/0470014008.
- [19] M. Q. Raza, M. Nadarajah, and C. Ekanayake, “On recent advances in PV output power forecast,” *Solar Energy*, vol. 136, Oct. 2016, doi: 10.1016/j.solener.2016.06.073.
- [20] M. Q. Raza and A. Khosravi, “A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings,” *Renewable and Sustainable Energy Reviews*, vol. 50. Elsevier Ltd, pp. 1352–1372, Jun. 18, 2015. doi: 10.1016/j.rser.2015.04.065.
- [21] A. Nespoli *et al.*, “Day-ahead photovoltaic forecasting: A comparison of the most effective techniques,” *Energies (Basel)*, vol. 12, no. 9, 2019, doi: 10.3390/en12091621.
- [22] S. Sobri, S. Koochi-Kamali, and N. A. Rahim, “Solar photovoltaic generation forecasting methods: A review,” *Energy Conversion and Management*, vol. 156. Elsevier Ltd, pp. 459–497, Jan. 15, 2018. doi: 10.1016/j.enconman.2017.11.019.
- [23] Y. Ren, P. N. Suganthan, and N. Srikanth, “Ensemble methods for wind and solar power forecasting - A state-of-the-art review,” *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 82–91, Oct. 2015, doi: 10.1016/j.rser.2015.04.081.

- [24] M. K. Behera, I. Majumder, and N. Nayak, "Solar photovoltaic power forecasting using optimized modified extreme learning machine technique," *Engineering Science and Technology, an International Journal*, vol. 21, no. 3, pp. 428–438, Jun. 2018, doi: 10.1016/j.jestch.2018.04.013.
- [25] A. G. R. Vaz, B. Elsinga, W. G. J. H. M. van Sark, and M. C. Brito, "An artificial neural network to assess the impact of neighbouring photovoltaic systems in power forecasting in Utrecht, the Netherlands," *Renew Energy*, vol. 85, pp. 631–641, Jan. 2016, doi: 10.1016/j.renene.2015.06.061.
- [26] C. Chen, S. Duan, T. Cai, and B. Liu, "Online 24-h solar power forecasting based on weather type classification using artificial neural network," *Solar Energy*, vol. 85, no. 11, pp. 2856–2870, Nov. 2011, doi: 10.1016/j.solener.2011.08.027.
- [27] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Conversion and Management*, vol. 198. Elsevier Ltd, Oct. 15, 2019. doi: 10.1016/j.enconman.2019.111799.
- [28] RENAC, "Short term prediction of wind and solar power generation," 2017.
- [29] Q. Hu, R. Zhang, and Y. Zhou, "Transfer learning for short-term wind speed prediction with deep neural networks," *Renewable Energy*, vol. 85. pp. 83–95, 2016. doi: 10.1016/j.renene.2015.06.034.
- [30] A. S. Qureshi, A. Khan, A. Zameer, and A. Usman, "Wind power prediction using deep neural network based meta regression and transfer learning," *Applied Soft Computing Journal*, vol. 58, pp. 742–755, Sep. 2017, doi: 10.1016/j.asoc.2017.05.031.
- [31] C. Möhrlen, W. Dk, J. Zack, J. Messner, A. Analytics, and J. Browell, "IEA Wind Task 36 - Recommended Practice on Renewable Energy Forecast Solution Selection," 2019. Accessed: May 02, 2023. [Online]. Available: <https://www.ieawindforecasting.dk/publications/recommendedpractice>
- [32] M. Kudo, A. Takeuchi, Y. Nozaki, H. Endo, and S. Jiro, "Forecasting electric power generation in a photovoltaic power system for an energy network," *Electrical Engineering in Japan (English translation of Denki Gakkai Ronbunshi)*, vol. 167, no. 4, pp. 16–23, Jun. 2009, doi: 10.1002/eej.20755.
- [33] A. Sarkar, "Global and Regional Ensemble Prediction Systems of NCMRWF." Accessed: May 07, 2023. [Online]. Available: www.ncmrwf.gov.in
- [34] S. Dutta *et al.*, "Load and Renewable Energy Forecasting for a Microgrid using Persistence Technique," in *Energy Procedia*, Elsevier Ltd, 2017, pp. 617–622. doi: 10.1016/j.egypro.2017.12.736.

- [35] George E. P. Box and Gwilym M. Jenkins, *Time series analysis, forecasting and control*, Rev. ed. Holden-Day series in time series analysis and digital processing, 1976.
- [36] D. Yang, J. Kleissl, C. A. Gueymard, H. T. C. Pedro, and C. F. M. Coimbra, “History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining,” *Solar Energy*, vol. 168, pp. 60–101, Jul. 2018, doi: 10.1016/j.solener.2017.11.023.
- [37] C. Feng, J. Zhang, W. Zhang, and B.-M. Hodge, “Convolutional neural networks for intra-hour solar forecasting based on sky image sequences,” *Appl Energy*, vol. 310, p. 118438, Mar. 2022, doi: 10.1016/j.apenergy.2021.118438.
- [38] J. Liu, X. Huang, Q. Li, Z. Chen, G. Liu, and Y. Tai, “Hourly stepwise forecasting for solar irradiance using integrated hybrid models CNN-LSTM-MLP combined with error correction and VMD,” *Energy Convers Manag*, vol. 280, p. 116804, Mar. 2023, doi: 10.1016/j.enconman.2023.116804.
- [39] A. Mellit and A. M. Pavan, “A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy,” *Solar Energy*, vol. 84, no. 5, pp. 807–821, May 2010, doi: 10.1016/j.solener.2010.02.006.
- [40] A. Mellit, S. Sağlam, and S. A. Kalogirou, “Artificial neural network-based model for estimating the produced power of a photovoltaic module,” *Renew Energy*, vol. 60, pp. 71–78, Dec. 2013, doi: 10.1016/j.renene.2013.04.011.
- [41] H. T. Yang, C. M. Huang, Y. C. Huang, and Y. S. Pai, “A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output,” *IEEE Trans Sustain Energy*, vol. 5, no. 3, pp. 917–926, 2014, doi: 10.1109/TSTE.2014.2313600.
- [42] Vũ Hữu Tiệp, *Machine Learning cơ bản*. Nhà xuất bản Khoa học và Kỹ thuật, 2018. [Online]. Available: www.facebook.com/
- [43] G. P. Zhang, B. E. Patuwo, and M. Y. Hu, “A simulation study of artificial neural networks for nonlinear time-series forecasting,” 2001.
- [44] G. Zhang, B. Eddy Patuwo, and M. Y. Hu, “Forecasting with artificial neural networks: The state of the art,” *Int J Forecast*, vol. 14, no. 1, pp. 35–62, Mar. 1998, doi: 10.1016/S0169-2070(97)00044-7.
- [45] K. Murakami and H. Taguchi, “Gesture Recognition using Recurrent Neural Networks,” in *SIGCHI Conference on Human Factors in Computing Systems*, 1991.

- [46] S. Lawrence, C. L. Giles, and S. Fong, “Natural Language Grammatical Inference with Recurrent Neural Networks,” *IEEE Trans Knowl Data Eng*, vol. 12, no. 1, 2000.
- [47] M. Qi and G. P. Zhang, “Trend Time Series Modeling and Forecasting with Neural Networks,” *IEEE Trans Neural Netw*, vol. 19, no. 5, 2008.
- [48] Georg Dorffner, “Neural Networks for Time Series Processing,” *Neural Network World*, vol. 6, no. 447–468, 1996.
- [49] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput*, vol. 9, no. 8, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
- [50] M. Abdel-Nasser and K. Mahmoud, “Accurate photovoltaic power forecasting models using deep LSTM-RNN,” *Neural Comput Appl*, vol. 31, no. 7, pp. 2727–2740, Jul. 2019, doi: 10.1007/s00521-017-3225-z.
- [51] D. Opitz and R. Maclin, “Popular Ensemble Methods: An Empirical Study,” 1999.
- [52] H. Liu, H. Q. Tian, C. Chen, and Y. fei Li, “A hybrid statistical method to predict wind speed and wind power,” *Renew Energy*, vol. 35, no. 8, pp. 1857–1861, Aug. 2010, doi: 10.1016/j.renene.2009.12.011.
- [53] J. C. Cao and S. H. Cao, “Study of forecasting solar irradiance using neural networks with preprocessing sample data by wavelet analysis,” *Energy*, vol. 31, no. 15, pp. 3435–3445, 2006, doi: 10.1016/j.energy.2006.04.001.
- [54] GUO-RUI J, PU HAN, and YONG-JIE ZHAI, “Wind speed forecasting based on Support Vector Machine with forecasting error estimation,” in *Sixth International Conference on Machine Learning and Cybernetics*, IEEE Xplore, 2007.
- [55] Alicja Lojowska, Dorota Kurowicka, Georgios Papaefthymiou, and Lou van der Sluis, “Advantages of ARMA-GARCH Wind Speed Time Series Modeling,” in *EEE internationalconference on probabilistic methods applied to power systems (PMAPS2010)*, IEEE, 2010.
- [56] J. Shi, J. Guo, and S. Zheng, “Evaluation of hybrid forecasting approaches for wind speed and power generation time series,” *Renewable and Sustainable Energy Reviews*, vol. 16, no. 5, pp. 3471–3480, Jun. 2012. doi: 10.1016/j.rser.2012.02.044.
- [57] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres, “Review of photovoltaic power forecasting,” *Solar Energy*, vol. 136. Elsevier Ltd, pp. 78–111, Oct. 15, 2016. doi: 10.1016/j.solener.2016.06.069.

- [58] H. Li *et al.*, “Ensemble learning for overall power conversion efficiency of the all-organic dye-sensitized solar cells,” *IEEE Access*, vol. 6, pp. 34118–34126, Jun. 2018, doi: 10.1109/ACCESS.2018.2850048.
- [59] J. Gaboitaolelwe, A. M. Zungeru, A. Yahya, C. K. Lebekwe, D. N. Vinod, and A. O. Salau, “Machine Learning Based Solar Photovoltaic Power Forecasting: A Review and Comparison,” *IEEE Access*, vol. 11. Institute of Electrical and Electronics Engineers Inc., pp. 40820–40845, 2023. doi: 10.1109/ACCESS.2023.3270041.
- [60] Y. K. Wu, C. L. Huang, Q. T. Phan, and Y. Y. Li, “Completed Review of Various Solar Power Forecasting Techniques Considering Different Viewpoints,” *Energies*, vol. 15, no. 9. MDPI, May 01, 2022. doi: 10.3390/en15093320.
- [61] C. Brester, V. Kallio-Myers, A. V. Lindfors, M. Kolehmainen, and H. Niska, “Evaluating neural network models in site-specific solar PV forecasting using numerical weather prediction data and weather observations,” *Renew Energy*, May 2023, doi: 10.1016/j.renene.2023.02.130.
- [62] W. C. Tsai, C. S. Tu, C. M. Hong, and W. M. Lin, “A Review of State-of-the-Art and Short-Term Forecasting Models for Solar PV Power Generation,” *Energies*, vol. 16, no. 14. Multidisciplinary Digital Publishing Institute (MDPI), Jul. 01, 2023. doi: 10.3390/en16145436.
- [63] D. Lee and K. Kim, “Recurrent Neural Network-Based Hourly Prediction of Photovoltaic Power Output Using Meteorological Information,” *Energies (Basel)*, vol. 12, no. 2, p. 215, Jan. 2019, doi: 10.3390/en12020215.
- [64] K. Wang, X. Qi, and H. Liu, “Photovoltaic power forecasting based LSTM-Convolutional Network,” *Energy*, vol. 189, p. 116225, Dec. 2019, doi: 10.1016/j.energy.2019.116225.
- [65] J. Ospina, A. Newaz, and M. O. Faruque, “Forecasting of PV plant output using hybrid wavelet-based LSTM-DNN structure model,” *IET Renewable Power Generation*, vol. 13, no. 7, pp. 1087–1095, May 2019, doi: 10.1049/iet-rpg.2018.5779.
- [66] H. Zhou, Y. Zhang, L. Yang, Q. Liu, K. Yan, and Y. Du, “Short-Term photovoltaic power forecasting based on long short term memory neural network and attention mechanism,” *IEEE Access*, vol. 7, pp. 78063–78074, 2019, doi: 10.1109/ACCESS.2019.2923006.
- [67] L. Wen, K. Zhou, S. Yang, and X. Lu, “Optimal load dispatch of community microgrid with deep learning based solar power and load forecasting,” *Energy*, vol. 171, pp. 1053–1065, Mar. 2019, doi: 10.1016/j.energy.2019.01.075.

- [68] F. Harrou, F. Kadri, and Y. Sun, “Forecasting of Photovoltaic Solar Power Production Using LSTM Approach,” in *Advanced Statistical Modeling, Forecasting, and Fault Detection in Renewable Energy Systems*, IntechOpen, 2020. doi: 10.5772/intechopen.91248.
- [69] N. R. Zhou, Y. Zhou, L. H. Gong, and M. L. Jiang, “Accurate prediction of photovoltaic power output based on long short-term memory network,” *IET Optoelectronics*, vol. 14, no. 6, pp. 399–405, Dec. 2020, doi: 10.1049/iet-opt.2020.0021.
- [70] Y. Zhang, C. Qin, A. K. Srivastava, C. Jin, and R. K. Sharma, “Data-Driven Day-Ahead PV Estimation Using Autoencoder-LSTM and Persistence Model,” *IEEE Trans Ind Appl*, vol. 56, no. 6, pp. 7185–7192, Nov. 2020, doi: 10.1109/TIA.2020.3025742.
- [71] M. K. Park, J. M. Lee, W. H. Kang, J. M. Choi, and K. H. Lee, “Predictive model for PV power generation using RNN (LSTM),” *Journal of Mechanical Science and Technology*, vol. 35, no. 2, pp. 795–803, Feb. 2021, doi: 10.1007/s12206-021-0140-0.
- [72] Y. Li, F. Ye, Z. Liu, Z. Wang, and Y. Mao, “A Short-Term Photovoltaic Power Generation Forecast Method Based on LSTM,” *Math Probl Eng*, vol. 2021, 2021, doi: 10.1155/2021/6613123.
- [73] H. Zhou, Q. Liu, K. Yan, and Y. Du, “Deep Learning Enhanced Solar Energy Forecasting with AI-Driven IoT,” *Wirel Commun Mob Comput*, vol. 2021, 2021, doi: 10.1155/2021/9249387.
- [74] C. H. Liu, J. C. Gu, and M. T. Yang, “A Simplified LSTM Neural Networks for One Day-Ahead Solar Power Forecasting,” *IEEE Access*, vol. 9, pp. 17174–17195, 2021, doi: 10.1109/ACCESS.2021.3053638.
- [75] D. V. Pombo, P. Bacher, C. Ziras, H. W. Bindner, S. V. Spataru, and P. E. Sørensen, “Benchmarking physics-informed machine learning-based short term PV-power forecasting tools,” *Energy Reports*, vol. 8, pp. 6512–6520, Nov. 2022, doi: 10.1016/j.egy.2022.05.006.
- [76] V. Suresh, F. Aksan, P. Janik, T. Sikorski, and B. S. Revathi, “Probabilistic LSTM-Autoencoder Based Hour-Ahead Solar Power Forecasting Model for Intra-Day Electricity Market Participation: A Polish Case Study,” *IEEE Access*, vol. 10, pp. 110628–110638, 2022, doi: 10.1109/ACCESS.2022.3215080.
- [77] N. Duc Tuyen, V. Xuan, S. Huu, and L. V. Thinh, “Photovoltaic Power Generation Forecasting Utilizing Long Short Term Memory,” *Measurement, control and automation*, vol. 1, no. 2, 2020.

- [78] N. T. Hoai Thu, P. Quoc Bao, and N. V. Nhat Nam, “Multiple Step Ahead Forecasting of Rooftop Solar Power Based on a Novel Hybrid Model of CEEMDAN - Bidirectional LSTM Network with Structure Optimized by PSO Method,” in *2022 11th International Conference on Control, Automation and Information Sciences (ICCAIS)*, IEEE, Nov. 2022, pp. 522–528. doi: 10.1109/ICCAIS56082.2022.9990323.
- [79] Z. Yang and J. Wang, “A hybrid forecasting approach applied in wind speed forecasting based on a data processing strategy and an optimized artificial intelligence algorithm,” *Energy*, vol. 160, pp. 87–100, Oct. 2018, doi: 10.1016/j.energy.2018.07.005.
- [80] N. C. Schwertman, M. A. Owens, and R. Adnan, “A simple more general boxplot method for identifying outliers,” *Comput Stat Data Anal*, vol. 47, no. 1, pp. 165–174, Aug. 2004, doi: 10.1016/j.csda.2003.10.012.
- [81] F. Xiao and C. Fan, “Data mining in building automation system for improving building operational performance,” *Energy Build*, vol. 75, pp. 109–118, Jun. 2014, doi: 10.1016/j.enbuild.2014.02.005.
- [82] Z. (Jerry) Yu, F. Haghghat, B. C. M. Fung, and L. Zhou, “A novel methodology for knowledge discovery through mining associations between building operational data,” *Energy Build*, vol. 47, pp. 430–440, Apr. 2012, doi: 10.1016/j.enbuild.2011.12.018.
- [83] 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.
- [84] 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.
- [85] 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.

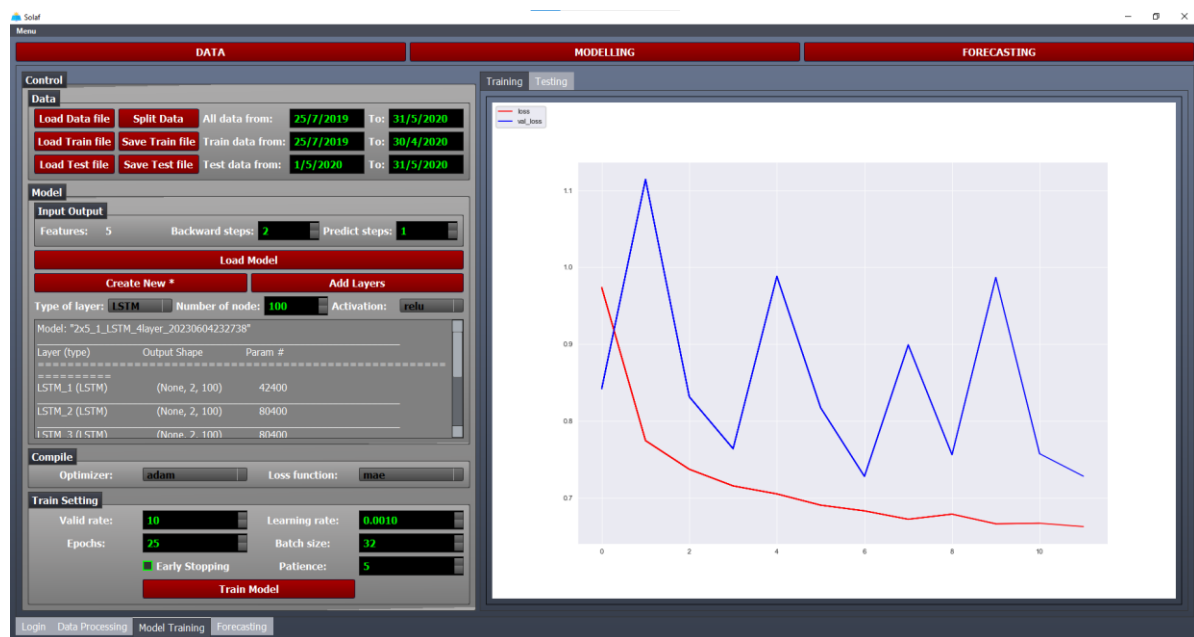
- [86] G. Reikard, “Predicting solar radiation at high resolutions: A comparison of time series forecasts,” *Solar Energy*, vol. 83, no. 3, pp. 342–349, Mar. 2009, doi: 10.1016/j.solener.2008.08.007.
- [87] H. Akaike, “TIME SERIES ANALYSIS AND CONTROL THROUGH PARAMETRIC MODELS,” in *Applied Time Series Analysis I*, Elsevier, 1978, pp. 1–23. doi: 10.1016/B978-0-12-257250-0.50006-4.
- [88] Z. Dong, D. Yang, T. Reindl, and W. M. Walsh, “Short-term solar irradiance forecasting using exponential smoothing state space model,” *Energy*, vol. 55, pp. 1104–1113, Jun. 2013, doi: 10.1016/j.energy.2013.04.027.
- [89] David Cournapeau, Matthieu Brucher, Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, and Vincent Michel, “scikit-learn.” Accessed: May 30, 2023. [Online]. Available: <https://scikit-learn.org/stable/index.html>
- [90] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” Dec. 2014.
- [91] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
- [92] F. Wang, Z. Zhen, B. Wang, and Z. Mi, “Comparative study on KNN and SVM based weather classification models for day ahead short term solar PV power forecasting,” *Applied Sciences (Switzerland)*, vol. 8, no. 1, Dec. 2017, doi: 10.3390/app8010028.
- [93] C. Feng and J. Zhang, “Hourly-Similarity Based Solar Forecasting Using Multi-Model Machine Learning Blending,” Mar. 2018, [Online]. Available: <http://arxiv.org/abs/1803.03623>
- [94] F. Barbieri, S. Rajakaruna, and A. Ghosh, “Very short-term photovoltaic power forecasting with cloud modeling: A review,” *Renewable and Sustainable Energy Reviews*, vol. 75. Elsevier Ltd, pp. 242–263, 2017. doi: 10.1016/j.rser.2016.10.068.
- [95] C. Voyant *et al.*, “Machine learning methods for solar radiation forecasting: A review,” *Renewable Energy*, vol. 105. Elsevier Ltd, pp. 569–582, 2017. doi: 10.1016/j.renene.2016.12.095.
- [96] S. Yin, J. Wang, Z. Li, and X. Fang, “State-of-the-art short-term electricity market operation with solar generation: A review,” *Renewable and Sustainable Energy Reviews*, vol. 138. Elsevier Ltd, Mar. 01, 2021. doi: 10.1016/j.rser.2020.110647.
- [97] D. Yang, “Choice of clear-sky model in solar forecasting,” *Journal of Renewable and Sustainable Energy*, vol. 12, no. 2, Mar. 2020, doi: 10.1063/5.0003495.

- [98] J. Li, “Short-term Photovoltaic Power Prediction Based on Moderate-resolution Imaging Spectroradiometer Clear Sky Data,” in *Proceedings - 2020 Chinese Automation Congress, CAC 2020*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020, pp. 930–934. doi: 10.1109/CAC51589.2020.9326823.
- [99] N. A. Engerer and F. P. Mills, “KPV: A clear-sky index for photovoltaics,” *Solar Energy*, vol. 105, pp. 679–693, 2014, doi: 10.1016/j.solener.2014.04.019.
- [100] S. Wang, T. Dai, C. Li, Y. Cheng, G. Huang, and G. Shi, “Improving Clear-Sky Solar Power Prediction over China by Assimilating Himawari-8 Aerosol Optical Depth with WRF-Chem-Solar,” *Remote Sens (Basel)*, vol. 14, no. 19, Oct. 2022, doi: 10.3390/rs14194990.
- [101] Y. Ma, X. Zhang, S. Mei, Z. Zhen, R. Gao, and Z. Zhou, “Ultra-short-term solar power forecasting based on a modified clear sky model,” in *39th Chinese Control Conference*, 2020.
- [102] D. V. Pombo, P. Bacher, C. Ziras, H. W. Bindner, S. V. Spataru, and P. E. Sørensen, “Benchmarking physics-informed machine learning-based short term PV-power forecasting tools,” *Energy Reports*, vol. 8, pp. 6512–6520, Nov. 2022, doi: 10.1016/j.egyr.2022.05.006.
- [103] W. F. Holmgren, C. W. Hansen, and M. A. Mikofski, “pvlib python: a python package for modeling solar energy systems,” *J Open Source Softw*, vol. 3, no. 29, p. 884, Sep. 2018, doi: 10.21105/joss.00884.

PHỤ LỤC 1: Giao diện công cụ dự báo



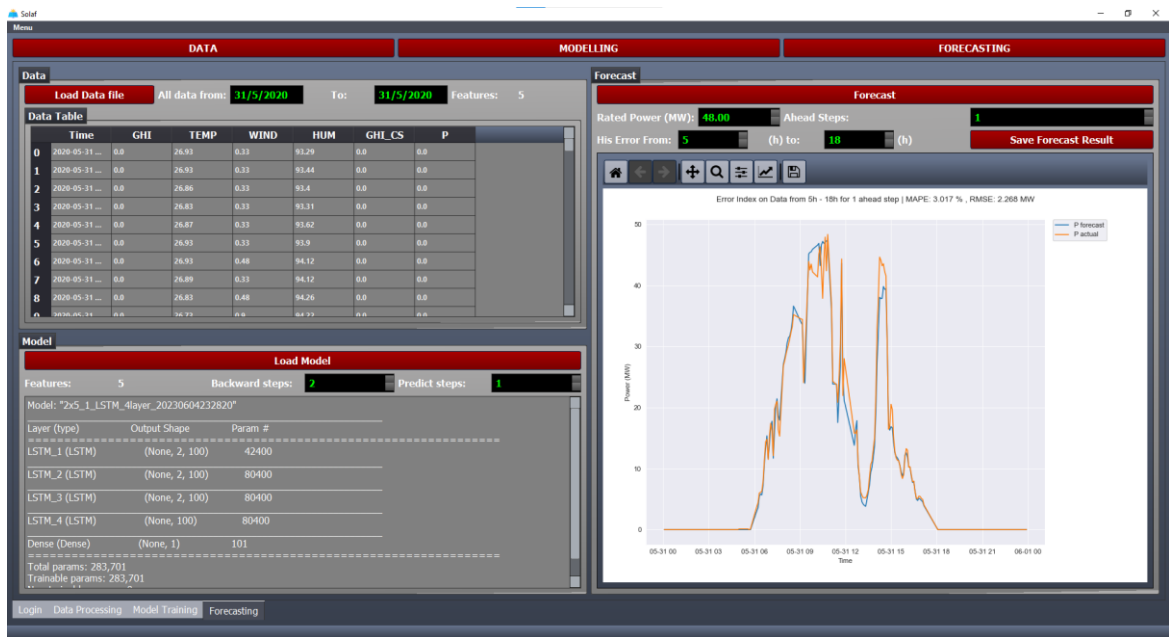
Hình PL 1. Giao diện chức năng Xử lý dữ liệu



Hình PL 2. Giao diện chức năng Huấn luyện mô hình



Hình PL 3. Giao diện chức năng Kiểm tra mô hình sau huấn luyện



Hình PL 4. Giao diện chức năng Dự báo

PHỤ LỤC 2: Kết quả chi tiết các chỉ số sai số dự báo trên tập dữ liệu kiểm tra của từng bước trong mô hình dự báo nhiều bước đồng thời

Bước	MSE	RMS E	MAE	MAP E	Bước	MSE	RMS E	MAE	MAP E
dự báo	MW ²	MW	MW	%	dự báo	MW ²	MW	MW	%
1	23,07 9	4,804	3,57	7,438	145	263,4 8	16,232	12,54 9	26,145
2	30,65 9	5,537	4,106	8,555	146	243,0 7	15,591	12,01 1	25,023
3	37,01	6,084	4,482	9,337	147	239,0 9	15,463	11,88 3	24,757
4	42,28 4	6,503	4,813	10,027	148	232,2 1	15,238	11,66 7	24,306
5	46,04 4	6,786	5,001	10,418	149	230,3 3	15,177	11,59 5	24,156
6	48,89 6	6,993	5,147	10,722	150	223,0 9	14,936	11,35 8	23,663
7	52,97 8	7,279	5,386	11,22	151	219,4 3	14,813	11,20 2	23,338
8	55,69	7,463	5,516	11,492	152	230,2 7	15,175	11,51 2	23,984
9	58,35 1	7,639	5,677	11,827	153	211,2 5	14,534	10,97 5	22,864
10	61,36	7,833	5,846	12,18	154	208,8	14,45	10,89 4	22,696
11	62,76 1	7,922	5,906	12,304	155	199,6	14,128	10,63 2	22,15
12	65,92 9	8,12	6,082	12,67	156	189,5 6	13,768	10,36 4	21,591
13	68,07 1	8,25	6,206	12,929	157	192,6 1	13,879	10,48 9	21,852
14	69,30 7	8,325	6,336	13,2	158	189,4 3	13,763	10,39 8	21,662
15	71,55 6	8,459	6,442	13,42	159	191,8 9	13,852	10,46 1	21,794
16	74,00 3	8,602	6,595	13,74	160	188,1 5	13,717	10,34 2	21,546
17	76,6	8,752	6,704	13,966	161	178,5 7	13,363	10,05 6	20,95
18	77,13 7	8,783	6,72	13,999	162	170,7 2	13,066	9,798	20,413
19	81,25 2	9,014	6,926	14,43	163	161,4 1	12,705	9,531	19,856
20	81,95 9	9,053	6,94	14,458	164	170,3 2	13,051	9,783	20,382
21	83,45 9	9,136	6,997	14,576	165	149,2 3	12,216	9,203	19,173
22	89,77 2	9,475	7,269	15,143	166	154,3 9	12,425	9,36	19,499
23	89,20 4	9,445	7,279	15,164	167	154,1 3	12,415	9,358	19,495
24	94,15 6	9,703	7,476	15,574	168	145,0 7	12,045	9,069	18,894

Bước	MSE	RMS E	MAE	MAP E	Bước	MSE	RMS E	MAE	MAP E
dự báo	MW ²	MW	MW	%	dự báo	MW ²	MW	MW	%
25	92,66 2	9,626	7,444	15,509	169	152,5 2	12,35	9,351	19,482
26	97,37 3	9,868	7,642	15,92	170	145,9 9	12,082	9,108	18,975
27	98,46 5	9,923	7,697	16,036	171	136,3 5	11,677	8,774	18,278
28	104,5 7	10,226	7,931	16,523	172	130,5	11,424	8,655	18,031
29	105,1 1	10,252	7,925	16,511	173	131,2 2	11,455	8,669	18,059
30	106,7 9	10,334	8,001	16,669	174	128,3 1	11,328	8,503	17,716
31	104,8 5	10,239	7,951	16,564	175	128,0 3	11,315	8,463	17,631
32	110,1 1	10,493	8,125	16,927	176	127,1 8	11,277	8,532	17,775
33	111,2 1	10,546	8,187	17,056	177	121,3 7	11,017	8,313	17,318
34	115,9 4	10,767	8,363	17,423	178	121,5 5	11,025	8,402	17,504
35	117,6 8	10,848	8,409	17,519	179	118,3 2	10,877	8,335	17,364
36	122,5 9	11,072	8,558	17,83	180	114,8 7	10,718	8,201	17,085
37	122,8 3	11,083	8,563	17,839	181	109,8 4	10,48	8,067	16,805
38	129,8 1	11,393	8,808	18,351	182	109,5 2	10,465	8,12	16,917
39	130,4	11,419	8,842	18,42	183	106,5 9	10,324	7,948	16,558
40	127,8 1	11,305	8,739	18,206	184	106,4 3	10,316	7,953	16,569
41	129,3 4	11,373	8,808	18,35	185	103,8 6	10,191	7,818	16,287
42	130,3 9	11,419	8,817	18,37	186	101,6 9	10,084	7,847	16,348
43	137,5	11,726	9,047	18,847	187	98,53 4	9,926	7,729	16,103
44	128,2 4	11,324	8,775	18,281	188	97,77	9,888	7,682	16,005
45	133,5 9	11,558	8,948	18,642	189	99,60 3	9,98	7,839	16,331
46	131,4 8	11,467	8,895	18,532	190	97,12 2	9,855	7,728	16,1
47	139,2 6	11,801	9,142	19,046	191	95,31 6	9,763	7,619	15,873
48	139,0 2	11,791	9,153	19,069	192	94,27 2	9,709	7,593	15,82
49	140,3 6	11,847	9,2	19,167	193	94,22 5	9,707	7,586	15,805

Bước	MSE	RMS E	MAE	MAP E	Bước	MSE	RMS E	MAE	MAP E
dự báo	MW ²	MW	MW	%	dự báo	MW ²	MW	MW	%
50	139,64	11,817	9,19	19,145	194	92,956	9,641	7,497	15,619
51	133,6	11,559	9,016	18,784	195	92,389	9,612	7,444	15,509
52	135,28	11,631	9,041	18,836	196	91,749	9,579	7,378	15,372
53	137,96	11,746	9,142	19,045	197	91,323	9,556	7,361	15,336
54	133,23	11,543	9,018	18,788	198	92,447	9,615	7,365	15,345
55	136,7	11,692	9,142	19,047	199	92,405	9,613	7,315	15,239
56	139,39	11,806	9,217	19,202	200	94,325	9,712	7,42	15,459
57	133,35	11,548	9,06	18,874	201	95,435	9,769	7,458	15,537
58	136,05	11,664	9,132	19,026	202	97,899	9,894	7,529	15,685
59	136,94	11,702	9,168	19,099	203	100,38	10,019	7,625	15,885
60	137,68	11,734	9,182	19,13	204	101,34	10,067	7,628	15,892
61	135,01	11,619	9,111	18,982	205	100,25	10,013	7,644	15,925
62	133,2	11,541	9,044	18,841	206	103,16	10,157	7,701	16,044
63	130,16	11,409	8,941	18,628	207	104,77	10,236	7,792	16,234
64	131,54	11,469	8,996	18,742	208	107,2	10,354	7,862	16,38
65	126,22	11,235	8,795	18,324	209	107,26	10,357	7,879	16,415
66	125,89	11,22	8,728	18,184	210	107,52	10,369	7,875	16,405
67	125,58	11,206	8,708	18,142	211	106,98	10,343	7,856	16,367
68	126,72	11,257	8,758	18,245	212	113,09	10,634	8,042	16,755
69	128,54	11,337	8,845	18,428	213	113,62	10,659	8,084	16,842
70	124,15	11,142	8,692	18,107	214	116,84	10,809	8,194	17,07
71	125,14	11,187	8,739	18,206	215	118,42	10,882	8,193	17,069
72	124,15	11,142	8,744	18,216	216	122,52	11,069	8,283	17,257
73	123,72	11,123	8,724	18,175	217	120,14	10,961	8,198	17,078
74	125,54	11,205	8,771	18,272	218	119,25	10,92	8,159	16,997

Bước	MSE	RMS E	MAE	MAP E	Bước	MSE	RMS E	MAE	MAP E
dự báo	MW ²	MW	MW	%	dự báo	MW ²	MW	MW	%
75	121,6 5	11,03	8,641	18,002	219	121,5 5	11,025	8,133	16,943
76	119,7 5	10,943	8,498	17,704	220	118,0 3	10,864	8,066	16,804
77	118,5	10,886	8,459	17,623	221	116,0 3	10,772	7,986	16,637
78	120,8 8	10,995	8,536	17,782	222	120,3 5	10,97	8,07	16,813
79	111,0 9	10,54	8,109	16,894	223	120,3 2	10,969	8,094	16,863
80	114,6 1	10,706	8,221	17,127	224	123,7 7	11,125	8,192	17,066
81	112,1 9	10,592	8,153	16,985	225	118,7 1	10,895	8,061	16,794
82	113,0 6	10,633	8,135	16,949	226	120,5 8	10,981	8,109	16,894
83	109,2	10,45	7,951	16,564	227	121,4 7	11,021	8,138	16,954
84	107,1 6	10,352	7,842	16,337	228	124,3 2	11,15	8,185	17,053
85	107,2 2	10,355	7,849	16,351	229	123,8 2	11,127	8,13	16,938
86	106,3 1	10,311	7,8	16,25	230	120,0 1	10,955	8,003	16,673
87	106,9 4	10,341	7,811	16,273	231	122,5 3	11,069	8,035	16,74
88	104,6 4	10,229	7,714	16,071	232	118,4 9	10,885	7,978	16,621
89	105,8 3	10,287	7,763	16,174	233	119,8 4	10,947	7,962	16,587
90	103,6 4	10,18	7,675	15,989	234	123,6 9	11,121	8,107	16,889
91	107,6 6	10,376	7,81	16,271	235	123,5 3	11,114	8,146	16,97
92	106,6 2	10,325	7,761	16,168	236	120,6 2	10,983	8,048	16,766
93	109,5 3	10,466	7,831	16,315	237	121,3 5	11,016	8,051	16,773
94	111,6 5	10,566	7,945	16,552	238	119,2 7	10,921	7,983	16,632
95	113,7 9	10,667	7,955	16,572	239	122,6 2	11,073	8,047	16,766
96	116,2 3	10,781	7,985	16,636	240	122,5 5	11,07	8,016	16,7
97	124,2 4	11,146	8,215	17,114	241	120,0 4	10,956	7,952	16,567
98	121,0 2	11,001	8,145	16,969	242	116,0 9	10,774	7,825	16,302
99	123,4 2	11,109	8,141	16,961	243	117,5 9	10,844	7,847	16,348

Bước	MSE	RMS E	MAE	MAP E	Bước	MSE	RMS E	MAE	MAP E
dự báo	MW ²	MW	MW	%	dự báo	MW ²	MW	MW	%
100	120,2 3	10,965	8,021	16,71	244	116,6 7	10,801	7,856	16,366
101	128,9 6	11,356	8,25	17,188	245	115,5 9	10,751	7,768	16,182
102	134,6 3	11,603	8,447	17,599	246	111,6 8	10,568	7,68	16
103	137,1 9	11,713	8,513	17,735	247	113,5 6	10,657	7,749	16,144
104	139,8 3	11,825	8,607	17,932	248	110,8	10,526	7,67	15,98
105	140,5 5	11,855	8,667	18,057	249	107,2 6	10,357	7,572	15,774
106	152,4 6	12,347	9,01	18,772	250	107,1 2	10,35	7,564	15,758
107	151,2 6	12,299	8,948	18,642	251	102,7 3	10,136	7,433	15,486
108	157,2 2	12,539	9,112	18,984	252	101,2 3	10,061	7,393	15,402
109	165,3 8	12,86	9,342	19,463	253	97,33 8	9,866	7,253	15,11
110	168,4 6	12,979	9,391	19,565	254	94,83 3	9,738	7,152	14,901
111	168,2 2	12,97	9,365	19,511	255	94,76 8	9,735	7,172	14,942
112	177,8 2	13,335	9,621	20,045	256	91,58 8	9,57	7,021	14,628
113	185,4 9	13,619	9,892	20,608	257	90,36	9,506	7,018	14,62
114	190,4 2	13,799	10,03 8	20,913	258	89,16 5	9,443	6,939	14,456
115	195,3 7	13,978	10,26 9	21,393	259	88,95 1	9,431	6,928	14,433
116	200,9 7	14,176	10,44 5	21,761	260	90,38 7	9,507	7,038	14,663
117	192,0 6	13,858	10,17 8	21,204	261	87,44 1	9,351	6,904	14,383
118	211,3 9	14,539	10,75 8	22,412	262	87,32	9,345	6,94	14,457
119	209,8 4	14,486	10,72 8	22,349	263	86,16	9,282	6,88	14,334
120	223,8 4	14,961	11,09 5	23,114	264	86,32 4	9,291	6,859	14,289
121	219,8	14,826	10,97 9	22,873	265	84,07 4	9,169	6,718	13,996
122	224,9 7	14,999	11,15 2	23,232	266	83,43 2	9,134	6,7	13,958
123	223,9 1	14,963	11,16 1	23,252	267	83,26	9,125	6,696	13,95
124	228,8 1	15,127	11,28 9	23,52	268	83,07 3	9,114	6,697	13,953

Bước	MSE	RMS E	MAE	MAP E	Bước	MSE	RMS E	MAE	MAP E
dự báo	MW ²	MW	MW	%	dự báo	MW ²	MW	MW	%
125	233,1 6	15,27	11,45 6	23,866	269	82,45 3	9,08	6,674	13,905
126	231,2 6	15,207	11,41 7	23,784	270	82,79 1	9,099	6,684	13,924
127	249,6 2	15,799	11,95 7	24,91	271	82,84 8	9,102	6,649	13,853
128	255,5 1	15,985	12,19 5	25,407	272	83,15 1	9,119	6,681	13,92
129	248,6 9	15,77	11,99 6	24,993	273	82,58 8	9,088	6,672	13,899
130	259,1 1	16,097	12,39 6	25,825	274	83,16 3	9,119	6,651	13,857
131	250,8 5	15,838	12,20 1	25,419	275	83,56 3	9,141	6,67	13,896
132	238,1 2	15,431	11,85	24,688	276	82,76 2	9,097	6,583	13,715
133	249,5	15,795	12,16 9	25,351	277	84,05 1	9,168	6,641	13,836
134	252,3	15,884	12,28 9	25,603	278	84,16 1	9,174	6,611	13,772
135	244,8 1	15,647	12,1	25,208	279	83,69 3	9,148	6,583	13,715
136	252,8 4	15,901	12,31 5	25,655	280	84,65 8	9,201	6,588	13,726
137	256,2 6	16,008	12,42 1	25,876	281	85,01 3	9,22	6,627	13,807
138	259,4 9	16,109	12,43 2	25,901	282	84,09 5	9,17	6,572	13,691
139	253	15,906	12,32	25,668	283	84,13 7	9,173	6,6	13,751
140	259,4 8	16,108	12,41 8	25,871	284	84,61 8	9,199	6,649	13,851
141	247,7 6	15,74	12,15 9	25,331	285	85,27 9	9,235	6,66	13,874
142	265,0 7	16,281	12,60 3	26,257	286	85,06 3	9,223	6,636	13,826
143	243,2 1	15,595	12,02	25,042	287	86,04 2	9,276	6,77	14,103
144	255,6 8	15,99	12,28 5	25,593	288	84,83 5	9,211	6,673	13,902