

Đánh giá ảnh hưởng của các đặc trưng khí tượng đến độ chính xác của mô hình dự báo công suất điện mặt trời

Assessment of the Influence of Meteorological features on the Accuracy of a Solar power forecasting model

> THS HOÀNG THUYỀN

Khoa KTHT&MTĐT, Trường Đại học Kiến trúc Hà Nội

Email: thuyen.htd.hau@gmail.com

TÓM TẮT

Dự báo công suất điện mặt trời đóng vai trò then chốt trong việc đảm bảo tính ổn định của điều độ hệ thống điện và tối ưu hóa tích hợp năng lượng tái tạo. Nghiên cứu này thực hiện đánh giá định lượng ảnh hưởng của các đặc trưng khí tượng đầu vào đến độ chính xác của mô hình dự báo sử dụng mạng nơ-ron hồi tiếp Long Short-Term Memory (LSTM). Dựa trên tập dữ liệu vận hành thực tế của nhà máy điện mặt trời công suất 50 MW, hai kịch bản thực nghiệm đã được thiết lập để so sánh giữa bộ 3 đặc trưng cơ sở (Bức xạ mặt trời, Nhiệt độ môi trường, Nhiệt độ tấm pin) và bộ 7 đặc trưng mở rộng (bổ sung độ ẩm, áp suất, tốc độ gió và độ che phủ mây). Kết quả thực nghiệm cho thấy mô hình sử dụng bộ đặc trưng tối giản đạt mức sai số NMAPE thấp (dưới 4%), trong khi việc bổ sung các biến khí tượng nâng cao chỉ mang lại sự cải thiện không đáng kể, dao động từ 0,05% đến 0,11%. Điều này chứng minh rằng các biến số bức xạ và nhiệt độ đã mang đủ thông tin cốt lõi để mô hình học được các quy luật phi tuyến của hệ thống. Nghiên cứu khẳng định chiến lược tối giản đặc trưng là giải pháp tối ưu về mặt kỹ thuật - kinh tế, đặc biệt phù hợp cho các hệ thống điện mặt trời tích hợp công trình đô thị hoặc áp mái, nơi việc cắt giảm chi phí hạ tầng cảm biến mà vẫn đảm bảo độ tin cậy dự báo là mục tiêu ưu tiên hàng đầu.

Từ khóa: Dự báo công suất phát điện mặt trời; đặc trưng thời tiết; LSTM.

ABSTRACT

Solar power forecasting plays a pivotal role in ensuring the stability of power system dispatching and optimizing the integration of renewable energy. This study conducts a quantitative assessment of the influence of input meteorological features on the accuracy of a forecasting model using Long Short-Term Memory (LSTM) networks. Based on real operational data from a 50 MW solar power plant, two experimental scenarios were established to compare a baseline 3-feature set (Solar Irradiance, Ambient Temperature, Module Temperature) with an expanded 7-feature set (including humidity, atmospheric pressure, wind speed, and cloud cover). Experimental results show that the model using the minimalist feature set achieves a low NMAPE (below 4%), while the integration of additional meteorological variables provides only marginal improvement, ranging from 0.05% to 0.11%. This demonstrates that solar radiation and temperature variables provide sufficient core information for the model to learn the nonlinear patterns of the system. The study confirms that a minimalist feature strategy is a techno-economically optimal solution, particularly suitable for urban-integrated or rooftop solar systems where reducing sensor infrastructure costs while maintaining forecasting reliability is a top priority.

Keywords: Solar power forecasting; meteorological features; LSTM

1. ĐẶT VẤN ĐỀ

Sự gia tăng nhanh chóng của các nguồn năng lượng tái tạo, đặc biệt là điện mặt trời, đã đặt ra những thách thức lớn đối với công tác điều độ và vận hành hệ thống điện do tính biến động và phụ thuộc chặt chẽ vào điều kiện thời tiết [1]. Để đảm bảo tính ổn định của lưới điện và tối ưu hóa việc tích hợp năng lượng sạch, các phương pháp

dự báo công suất chính xác trong ngắn hạn và trung hạn trở thành yêu cầu cấp thiết [2].

Trong những năm gần đây, các mô hình học sâu, đặc biệt là mạng nơ-ron hồi tiếp Long Short-Term Memory (LSTM) đã chứng minh được ưu thế vượt trội trong việc xử lý các dữ liệu chuỗi thời gian có tính phụ thuộc dài hạn. Khả năng ghi nhớ các quy luật thời

tiết từ quá khứ giúp LSTM bám sát được các biến động công suất thực tế tốt hơn so với các phương pháp thống kê truyền thống [3].

Tuy nhiên, một vấn đề quan trọng thường bị bỏ qua trong các nghiên cứu dự báo là bài toán lựa chọn đặc trưng (Feature Selection) gắn liền với chi phí đầu tư hạ tầng quan trắc. Hầu hết các mô hình hiện nay có xu hướng tích hợp tối đa các biến khí tượng như áp suất, độ ẩm, tốc độ và hướng gió để đạt được độ chính xác cao nhất. Mặc dù điều này phù hợp với các nhà máy điện mặt trời quy mô lớn (Utility-scale), nhưng lại gây ra rào cản tài chính và kỹ thuật đáng kể đối với các hệ thống điện mặt trời áp mái hoặc tích hợp trong công trình đô thị (BIPV). Việc lắp đặt và bảo trì các trạm quan trắc khí tượng đầy đủ cho từng tòa nhà không chỉ làm tăng chi phí đầu tư ban đầu mà còn gây khó khăn trong công tác vận hành hệ thống quản lý năng lượng thông minh.

Nghiên cứu này tập trung phân tích mối tương quan giữa việc tối ưu hóa chi phí hạ tầng thiết bị và duy trì độ chính xác dự báo, qua đó đánh giá khả năng đáp ứng ngưỡng sai số cho phép của bộ đặc trưng khí tượng tối giản. Tác động định lượng của các thành phần khí tượng đầu vào đến hiệu năng mô hình dự báo Long Short-Term Memory (LSTM) được khảo sát chi tiết thông qua hai kịch bản thực nghiệm.

Kết quả nghiên cứu không chỉ đóng góp về mặt học thuật trong việc xác định độ nhạy của mô hình LSTM mà còn cung cấp cơ sở tham chiếu quan trọng cho việc triển khai các giải pháp dự báo chi phí thấp, phù hợp với xu thế phát triển đô thị thông minh và công trình xanh hiện nay. Nghiên cứu được chia ra làm 4 phần bao gồm: Phần 1 là giới thiệu tổng quan về nghiên cứu, phần 2 là phương pháp nghiên cứu, phần 3 trình bày các kết quả và cuối cùng là kết luận.

2. PHƯƠNG PHÁP NGHIÊN CỨU

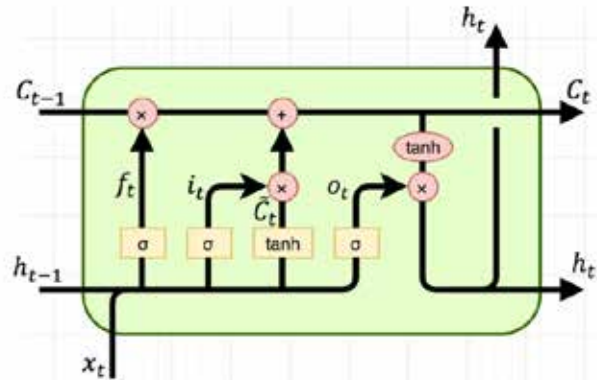
2.1. Dữ liệu nghiên cứu

Dữ liệu huấn luyện được thu thập từ một nhà máy điện mặt trời có công suất lắp đặt 50 MW tại tỉnh Đắk Lắk trong khoảng thời gian từ ngày 01/01/2022 đến ngày 31/12/2022. Với chu kỳ lấy mẫu 5 phút, bộ dữ liệu bao gồm 105,120 điểm đo, phản ánh chi tiết sự biến thiên của công suất phát điện dưới tác động của các yếu tố khí tượng. Nhằm đánh giá khả năng tổng quát hóa của mô hình, nghiên cứu thiết lập hai tập dữ liệu dự báo độc lập: Tập dự báo 1 (dữ liệu ngày 30/06/2023) đại diện cho kịch bản dự báo ngắn hạn trong ngày và Tập dự báo 2 (từ ngày 04/06 đến 06/06/2023) phục vụ đánh giá tính ổn định trong dự báo trung hạn. Việc tách biệt hoàn toàn các tập dữ liệu này đảm bảo tính khách quan và loại bỏ hiện tượng rò rỉ dữ liệu trong quá trình thực nghiệm.

Trước khi đưa vào mô hình, dữ liệu thô được chuẩn hóa về khoảng [0], [1] bằng phương pháp Min-Max Scaler. Quá trình này giúp loại bỏ sự khác biệt về đơn vị đo lường giữa các đặc trưng khí tượng, đồng thời tăng tốc độ hội tụ của thuật toán tối ưu. Các tham số chuẩn hóa được xác định từ tập huấn luyện và giữ cố định khi áp dụng cho các tập dự báo để đảm bảo tính đồng nhất về phân phối dữ liệu.

2.2. Mô hình LSTM

Mạng LSTM được lựa chọn nhờ cơ chế cổng (gate mechanism) đặc thù, cho phép mô hình kiểm soát luồng thông tin và lưu trữ các phụ thuộc dài hạn trong chuỗi thời gian. Trong cấu trúc này, cổng quên (forget gate) quyết định thông tin nào từ quá khứ sẽ bị loại bỏ, cổng đầu vào (input gate) cập nhật trạng thái ô hiện tại, và cổng đầu ra (output gate) trích xuất thông tin cần thiết cho bước thời gian tiếp theo [4]. Khả năng này đặc biệt quan trọng đối với dữ liệu điện mặt trời, vốn mang tính chu kỳ và phụ thuộc vào các xu hướng thời tiết từ các bước thời gian trước đó.



Hình 1. Mô hình LSTM [5]

Mô hình LSTM trong nghiên cứu này được thiết kế với cấu trúc tối ưu gồm một lớp LSTM 64 nơ-ron, kết hợp cùng lớp ẩn Dense 16 nơ-ron sử dụng hàm kích hoạt Tanh để tăng tính phi tuyến. Lớp đầu ra thực hiện dự báo giá trị công suất đơn lẻ. Các tham số huấn luyện được thiết lập cụ thể như sau:

- Hàm tối ưu: Adam (learning rate = 0,001)
- Look-back window: 12 bước (60 phút)
- Batch size: 32
- Epoch: 100

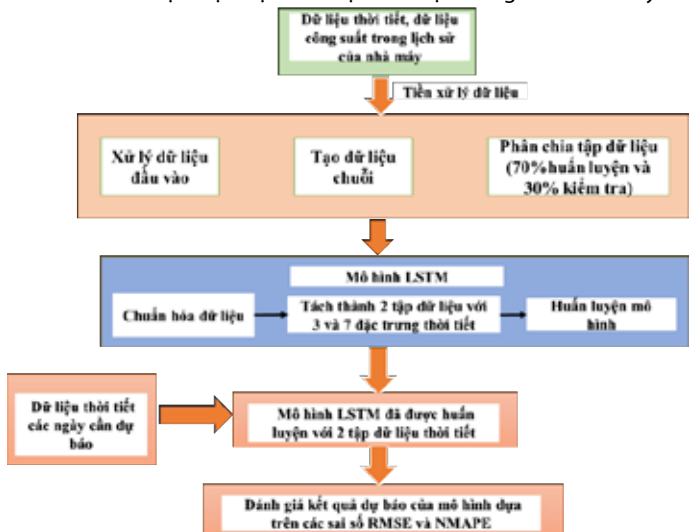
Dữ liệu của tập huấn luyện được chia theo tỷ lệ 70% huấn luyện và 30% kiểm tra.

2.3. Kịch bản thử nghiệm

Nghiên cứu tiến hành so sánh hiệu năng thông qua hai kịch bản dựa trên số lượng đặc trưng đầu vào, nhằm xác định mức độ phụ thuộc của độ chính xác dự báo vào sự phức tạp của hệ thống quan trắc:

- Kịch bản 1 (Mô hình tối giản): Sử dụng bộ 3 đặc trưng cơ sở gồm bức xạ mặt trời (GHI), nhiệt độ môi trường và nhiệt độ tấm pin. Kịch bản này nhằm kiểm chứng xem mô hình LSTM với số lượng đặc trưng đầu vào tối thiểu có đáp ứng yêu cầu dự báo hay không.
- Kịch bản 2 (Mô hình mở rộng): Bổ sung thêm 4 đặc trưng khí tượng gồm độ ẩm không khí, tốc độ gió, độ che phủ mây và áp suất khí quyển. Việc thiết lập kịch bản này nhằm kiểm chứng liệu sự gia tăng các đặc trưng dữ liệu có mang lại sự cải thiện đáng kể về mặt kỹ thuật cho mô hình dự báo hay không.

Các bước thực hiện dự báo được thể hiện trong sơ đồ sau đây:



Hình 2. Các bước thực hiện dự báo

2.4. Các sai số được sử dụng

Hiệu năng của mô hình được định lượng thông qua hai chỉ số thống kê phổ biến:

Sai số trung bình bình phương (RMSE): Chỉ số RMSE đo lường độ lệch giữa giá trị dự báo và giá trị thực tế, có xu hướng nhạy cảm với các sai số lớn. Giá trị RMSE càng thấp cho thấy mô hình có độ chính xác càng cao trong việc bám sát các biến động mạnh của công suất. Công thức tính RMSE được thể hiện sau đây [6].

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

Trong đó: \hat{y}_i là công suất dự báo (kW), y_i là công suất thực tế (kW), n là số lượng điểm dữ liệu.

Sai số tuyệt đối phần trăm trung bình đã được chuẩn hóa (NMAPE): Đối với hệ thống điện mặt trời, việc sử dụng NMAPE là cực kỳ quan trọng do công suất thường về mức 0 vào ban đêm. NMAPE thực hiện chuẩn hóa sai số dựa trên công suất định mức của nhà máy là 50 MW (tương đương 50000 kW) thay vì giá trị thực tế tại thời điểm đo, giúp tránh lỗi chia cho phần tử bằng 0 và cung cấp cái nhìn khách quan về hiệu suất mô hình trên toàn dải công suất. Công thức tính NMAPE được thể hiện như sau [7]:

$$NMAPE = \frac{100}{n} \sum_{i=1}^n \frac{|P_i^{db} - P_i^{tt}|}{P_{dm}} \quad (2)$$

Trong đó: P_i^{db} giá trị công suất phát của dự báo thứ i (kW), P_i^{tt} là giá trị công suất trong thực tế của tín hiệu thứ i (kW), P_{dm} là công suất định mức của nhà máy (kW), n là số tín hiệu dự báo được đánh giá trong khung thời gian của loại dự báo.

3. KẾT QUẢ

Hiệu suất dự báo của mô hình mạng LSTM đối với các kịch bản thực nghiệm được tổng hợp chi tiết tại Bảng 1.

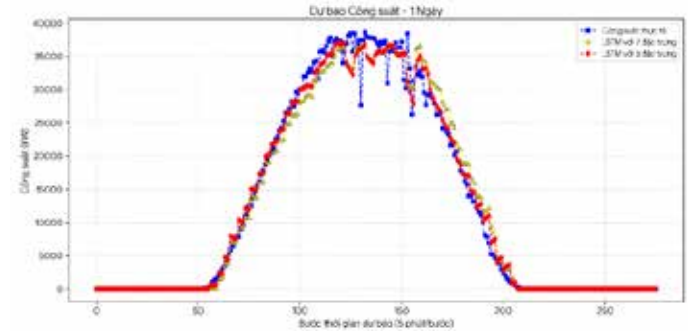
Bảng 1. Kết quả dự báo của mô hình LSTM cho 2 kịch bản

Kịch bản	Mô hình	RMSE (kW)	NMAPE (%)
Kịch bản 1 (3 đặc trưng khí tượng)	LSTM cho dự báo 1 ngày	2465,12	2,94
	LSTM cho dự báo 3 ngày	4219,23	3,92
Kịch bản 2 (7 đặc trưng khí tượng)	LSTM cho dự báo 1 ngày	2399,86	3,05
	LSTM cho dự báo 3 ngày	4137,45	3,87

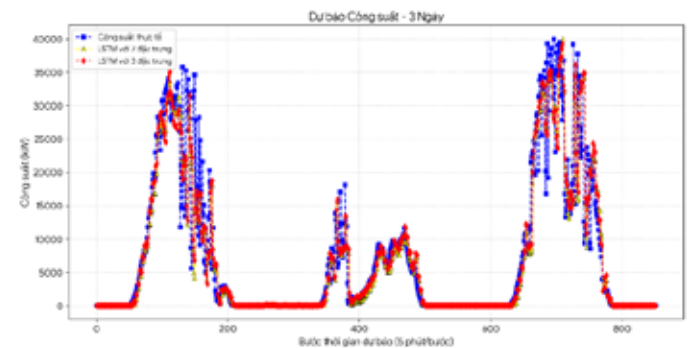
Bảng 1 cho thấy sự chênh lệch về chỉ số NMAPE giữa hai kịch bản là rất nhỏ. Với dự báo cho 1 ngày thì độ chênh lệch của 2 kịch bản là 0,11% và với dự báo cho 3 ngày thì độ chênh lệch là 0,05%. Với công suất định mức của nhà máy là 50 MW, mức chênh lệch 0,05%-0,11% tương đương sai khác công suất trung bình chỉ khoảng 25-55 kW. So với biên độ dao động công suất hàng ngày lên tới hàng chục MW, mức sai lệch này là không đáng kể và hầu như không tạo ra khác biệt về mặt vận hành hệ thống hoặc lập kế hoạch điều độ. Đáng chú ý, tất cả các kết quả dự báo đều nằm trong ngưỡng sai số vận hành cho phép của Công ty TNHH MTV Vận hành hệ thống điện và Thị trường điện quốc gia (NSMO) với yêu cầu NMAPE dưới 15% [8]. Ngoài ra, kết quả huấn luyện lặp lại nhiều lần cho thấy độ lệch chuẩn của NMAPE dao động trong khoảng từ ±0,06% đến ±0,08%, tương đương với mức chênh lệch giữa hai kịch bản. Điều này cho thấy sự khác biệt về độ chính xác giữa bộ dữ liệu

3 đặc trưng và 7 đặc trưng không tạo ra sự khác biệt có ý nghĩa trong bối cảnh vận hành thực tế.

Việc so sánh công suất thực tế và kết quả dự báo của mô hình LSTM trong hai cấu hình đặc trưng được minh họa chi tiết qua Hình 3 và Hình 4.



Hình 3. So sánh công suất dự báo của mô hình LSTM theo hai cấu hình đặc trưng và công suất thực tế trong dự báo ngắn hạn (1 ngày)



Hình 4. So sánh công suất dự báo của mô hình LSTM theo hai cấu hình đặc trưng và công suất thực tế trong dự báo trung hạn (3 ngày)

Đồ thị tại Hình 3 và Hình 4 cho thấy cả hai cấu hình đặc trưng mô hình dự báo đều có khả năng tái tạo tốt hình dạng đường cong công suất thực tế, đặc biệt là trong giai đoạn đạt đỉnh. Tại các bước thời gian có biến động thời tiết nhanh (mây che cục bộ), mô hình LSTM sử dụng 3 đặc trưng vẫn duy trì được xu hướng bám sát đường công suất thực tế không hề thua kém so với mô hình đầy đủ 7 đặc trưng. Điều này chứng minh rằng các biến số bức xạ và nhiệt độ đã mang đủ thông tin cốt lõi để mô hình học được các quan hệ phi tuyến phức tạp. Việc mở rộng thêm các biến khí tượng bổ trợ như áp suất, độ ẩm và các biến bổ trợ khác không mang lại cải thiện đáng kể trong quá trình dự báo.

❖ Khả năng ứng dụng thực tiễn và mở rộng cho hệ thống điện mặt trời đô thị

Mặc dù các dữ liệu thực nghiệm trong nghiên cứu này được trích xuất từ nhà máy điện mặt trời quy mô lớn (50 MW), kết quả này mang lại những giá trị tham chiếu quan trọng cho các hệ thống quy mô nhỏ hơn.

Các đặc tính vật lý chi phối quá trình biến đổi quang điện như cường độ bức xạ và tác động của nhiệt độ lên tế bào quang điện là đồng nhất trên mọi quy mô. Do đó, mô hình dự báo với bộ đặc trưng tối giản có thể dễ dàng chuyển giao cho các hệ thống điện mặt trời tích hợp công trình đô thị hoặc điện mặt trời mái nhà tại các khu chung cư, nhà máy.

Trong bối cảnh phát triển đô thị thông minh, việc ưu tiên sử dụng bộ 3 đặc trưng (Bức xạ, Nhiệt độ môi trường, Nhiệt độ tấm pin) mang lại nhiều lợi ích thiết thực:

1. Tối ưu hóa chi phí đầu tư: Giảm bớt nhu cầu lắp đặt các cảm biến chuyên dụng đắt tiền như máy đo tốc độ gió, cảm biến áp suất khí quyển, camera toàn cảnh bầu trời.

2. Đơn giản hóa hạ tầng dữ liệu: Giảm khối lượng truyền tải và lưu trữ dữ liệu, phù hợp với các hệ thống điều khiển tại tòa nhà.

3. Hỗ trợ quản lý năng lượng thông minh: Cung cấp đầu vào tin cậy cho hệ thống quản lý năng lượng tòa nhà để điều phối phụ tải, tối ưu hóa việc sạc/xả hệ thống lưu trữ và lập kế hoạch mua bán điện theo biểu giá thời gian một cách hiệu quả nhất.

Để minh họa cụ thể cho tiềm năng ứng dụng này, Hình 5 mô tả cấu trúc điển hình của một hệ thống điện mặt trời mái nhà tích hợp trên các công trình đô thị. Đây chính là bối cảnh mà giải pháp tối giản hóa bộ đặc trưng khí tượng phát huy tối đa ưu thế về mặt kinh tế và kỹ thuật.



Hình 5. Hệ thống điện mặt trời mái nhà tích hợp trên các công trình đô thị

Việc áp dụng mô hình LSTM dựa trên bộ 3 đặc trưng cốt lõi (bức xạ, nhiệt độ môi trường, nhiệt độ tấm pin) giúp các đơn vị quản lý tòa nhà giảm bớt gánh nặng về hạ tầng quan trắc mà vẫn đạt được độ chính xác dự báo cần thiết. Điều này tạo tiền đề cho việc tích hợp sâu rộng năng lượng tái tạo vào lưới điện phân phối tại các khu vực dân cư đông đúc.

4. KẾT LUẬN

Nghiên cứu đã thực hiện đánh giá định lượng về ảnh hưởng của các đặc trưng khí tượng đến độ chính xác của mô hình dự báo công suất điện mặt trời dựa trên mạng LSTM. Thông qua việc phân tích dữ liệu thực tế và thử nghiệm trên các kịch bản dự báo ngắn hạn và trung hạn, một số kết luận quan trọng được rút ra như sau:

Thứ nhất, mô hình LSTM thể hiện sự ổn định và chính xác cao trong cả hai kịch bản dự báo. Khi sử dụng bộ 3 đặc trưng cơ sở (bức xạ mặt trời, nhiệt độ môi trường và nhiệt độ tấm pin), sai số NMAPE vẫn đạt mức thấp dưới 4%, hoàn toàn đáp ứng các tiêu chuẩn vận hành nghiêm ngặt của hệ thống điện.

Thứ hai, kết quả thực nghiệm khẳng định rằng việc bổ sung các đặc trưng nâng cao như độ ẩm, áp suất khí quyển, tốc độ gió và độ che phủ của đám mây chỉ mang lại sự cải thiện sai số không đáng kể (dao động từ 0,05% đến 0,11%). Điều này chỉ ra rằng các biến số cơ sở đã mang đủ thông tin đặc trưng để mô hình học được quy luật phát điện của hệ thống, trong khi các đặc trưng bổ trợ có thể làm nhiễu thông tin đối với các mô hình học sâu như LSTM.

Thứ ba, nghiên cứu đề xuất một chiến lược tiếp cận tối giản trong việc xây dựng hạ tầng dữ liệu dự báo. Việc tập trung vào bộ 3 đặc trưng cốt lõi giúp giảm thiểu đáng kể chi phí đầu tư cảm biến và đơn giản hóa quy trình vận hành hệ thống giám sát. Đây là cơ sở quan trọng để triển khai rộng rãi các mô hình dự báo cho hệ thống điện mặt trời mái nhà và tích hợp công trình đô thị, nơi yếu tố chi phí và sự tinh gọn của thiết bị luôn được ưu tiên hàng đầu.

Trong tương lai, nghiên cứu có thể được mở rộng bằng cách khảo sát ảnh hưởng của các yếu tố đặc thù trong đô thị như bóng đổ từ các công trình lân cận hoặc tác động của ô nhiễm không khí đến độ chính xác dự báo, nhằm hoàn thiện hơn nữa khả năng ứng dụng của mô hình trong thực tế quản lý năng lượng thông minh.

TÀI LIỆU THAM KHẢO

- [1] N. D. Tuyen, V. Xuan, S. Huu, and L. V. Thinh, "Photovoltaic Power Generation Forecasting Utilizing Long Short Term Memory," *Meas. Control Autom.*, vol. 1, no. April, pp. 2-7, 2021.
- [2] Nguyễn Tuấn Anh, Phạm Mạnh Hải, Vũ Minh Pháp, Nguyễn Đăng Toàn và Vũ Thị Anh Thơ, "Dự báo công suất nguồn phát điện mặt trời bằng mô hình BiGRU," *Tạp chí Khoa học và Công nghệ Năng lượng - Trường Đại học Điện Lực*, vol. 34, pp. 1-14, 2024.
- [3] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735-1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
- [4] 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 a case study in Vietnam," *Electr. Power Syst. Res.*, vol. 213, 2022, doi: 10.1016/j.epr.2022.108706.
- [5] N. N. V. Nhat, D. N. Huu, and T. T. H. Nguyen, "Short-term multi-step forecasting of rooftop solar power generation using a combined data decomposition and deep learning model of EEMD-GRU," *J. Renew. Sustain. Energy*, vol. 16, no. 1, 2024, doi: 10.1063/5.0176951.
- [6] E. Nziyumba, R. Hu, Y. Xu, J. Niyogisubizo, Z. Li, and P. C. Nshimyumukiza, "Solar Radiation Forecasting Using Ensemble-Based Hybrid LGBM-GB-MLP Model: A Novel Stacked Generalization Method," *SSRN Electron. J.*, no. October, 2022, doi: 10.2139/ssrn.4076358.
- [7] Bộ Công Thương, "67 QĐ/quy trình dự báo công suất, điện năng phát của các nguồn điện năng lượng tái tạo," pp. 1-22, 2021.
- [8] EVNNLDC, "4608/ĐĐQG-NLTT, yêu cầu độ chính xác đối với dự báo công suất phát các NMD NLTT năm 2024," pp. 1-10, 2023.