

**BỘ CÔNG THƯƠNG  
TRƯỜNG ĐẠI HỌC ĐIỆN LỰC**



**NGUYỄN TUẤN ANH**

**NGHIÊN CỨU MÔ HÌNH HỌC MÁY HỖN HỢP DỰ BÁO  
CÔNG SUẤT PHÁT ĐIỆN MẶT TRỜI TRONG ĐIỀU KIỆN DỮ  
LIỆU THỰC TẾ TẠI VIỆT NAM**

**LUẬN ÁN TIẾN SĨ NGÀNH KỸ THUẬT NĂNG LƯỢNG  
MÃ SỐ: THÍ ĐIỂM**

**Hà Nội - Năm 2026**

**Công trình được thực hiện tại:  
Trường Đại học Điện lực**

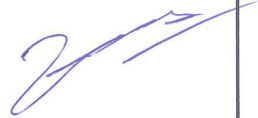
**Người hướng dẫn khoa học:**

**Người hướng dẫn 1**

**Người hướng dẫn 2**



**TS. Phạm Mạnh Hải**



**TS. Vũ Minh Pháp**

**Phản biện 1: PGS. TS. Lê Minh Thùy**

**Phản biện 2: PGS.TS. Nguyễn Thị Hoài Thu**

**Phản biện 3: PGS. TS. Nguyễn Tùng Linh**

Luận án được bảo vệ trước Hội đồng đánh giá Luận án Tiến sĩ cấp trường tại Trường Đại học Điện lực vào hồi 14h00 ngày 27 tháng 02 năm 2026.

Có thể tìm hiểu luận án tại:

1. Thư viện Trường Đại học Điện lực
2. Thư viện Quốc gia Việt Nam

## LỜI CAM ĐOAN

Nghiên cứu sinh đã tìm hiểu và nắm vững các quy định liên quan đến hành vi vi phạm tính trung thực trong học thuật. Nghiên cứu sinh cam kết bằng danh dự cá nhân rằng công trình này do chính nghiên cứu sinh thực hiện và tuân thủ đầy đủ các yêu cầu về liêm chính học thuật, không có bất kỳ hành vi vi phạm nào.

Hà Nội, Ngày 28... Tháng 09... năm 2026

**Tập thể hướng dẫn**

**Nghiên cứu sinh**

**Người hướng dẫn 1**

**Người hướng dẫn 2**

**TS. Phạm Mạnh Hải**

**TS. Vũ Minh Pháp**

**Nguyễn Tuấn Anh**

## LỜI CẢM ƠN

Để hoàn thành luận án tiến sĩ, cùng với những nỗ lực của bản thân, nghiên cứu sinh đã nhận được sự động viên, giúp đỡ của rất nhiều cá nhân và tập thể. Nghiên cứu sinh xin trân trọng cảm ơn Trường Đại học Điện lực nơi nghiên cứu sinh thực hiện đề tài đã tạo điều kiện thuận lợi về cơ sở vật chất, học thuật và môi trường nghiên cứu nghiêm túc.

Nghiên cứu sinh cũng trân trọng gửi lời cảm ơn đến hai thầy hướng dẫn là Tiến sĩ Phạm Mạnh Hải (Trường Đại học Điện lực) và Tiến sĩ Vũ Minh Pháp (Viện Khoa học Công nghệ Năng lượng và Môi trường, Viện Hàn lâm Khoa học và Công nghệ Việt Nam) đã tận tình chỉ bảo, định hướng và hỗ trợ nghiên cứu sinh trong suốt quá trình học tập, nghiên cứu và hoàn thành luận án.

Nghiên cứu sinh cũng xin gửi lời cảm ơn chân thành tới các thầy cô khoa Kỹ thuật điện và khoa Năng lượng mới đã luôn tận tình động viên và hỗ trợ Nghiên cứu sinh trong quá trình triển khai đề tài nghiên cứu.

Nghiên cứu sinh 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, nghiên cứu sinh 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 nghiên cứu sinh có được chỗ dựa vững chắc để hoàn thành luận án này.

Hà Nội, ngày 28 tháng 09 năm 2026

Nghiên cứu sinh

Nguyễn Tuấn Anh

## I. MỞ ĐẦU

### 1. Lý do chọn đề tài

Trong xu thế chuyển dịch năng lượng toàn cầu, Việt Nam đang đẩy mạnh phát triển năng lượng tái tạo nhằm đảm bảo an ninh năng lượng và giảm phát thải khí nhà kính. Trong các nguồn năng lượng tái tạo, điện mặt trời nổi lên như một nguồn năng lượng chủ đạo nhờ tiềm năng dồi dào, tính bền vững và khả năng mở rộng nhanh chóng. Tuy nhiên, sản lượng điện mặt trời lại phụ thuộc chặt chẽ vào điều kiện thời tiết (đặc biệt là bức xạ mặt trời), gây dao động mạnh về công suất phát và tạo ra thách thức lớn cho công tác điều độ vận hành hệ thống điện. Đáng chú ý, nhiều nhà máy điện mặt trời mới ở Việt Nam chưa có dữ liệu lịch sử dài hạn, dữ liệu thu thập thường ngắn, gián đoạn hoặc không đồng nhất, trong khi sự phân hóa rõ rệt theo mùa và vùng miền càng làm bài toán dự báo trở nên phức tạp. Trước những khó khăn đó, nhu cầu cấp thiết đặt ra là phải phát triển một mô hình dự báo công suất ngắn hạn có độ chính xác cao, ổn định và đáng tin cậy ngay cả khi dữ liệu lịch sử thiếu hụt, đồng thời thích ứng theo mùa vụ và khác biệt vùng khí hậu, và có khả năng mở rộng áp dụng cho nhiều nhà máy khác nhau.

### 2. Mục tiêu nghiên cứu

**Mục tiêu tổng quát:** Xây dựng một mô hình dự báo công suất phát điện mặt trời ngắn hạn hoạt động hiệu quả, ổn định ngay cả trong điều kiện chuỗi dữ liệu lịch sử không đầy đủ, đồng thời thích ứng với biến động thời tiết theo mùa và có thể mở rộng dự báo cho nhiều nhà máy tại các vùng khí hậu khác nhau.

#### **Mục tiêu cụ thể:**

Xây dựng mô hình dự báo bức xạ mặt trời. So sánh đánh giá hiệu suất ba mô hình LightGBM, LSTM, GRU trong dự báo bức xạ mặt trời.

So sánh đánh giá hiệu suất dự báo công suất phát điện mặt trời của các mô hình cây quyết định gồm XGBoost, LightGBM, Random Forest và các mô hình học sâu chuỗi thời gian gồm LightGBM, LSTM, GRU.

Xây dựng tổ hợp ba mô hình cây quyết định (XGBoost, LightGBM, Random Forest) để dự báo công suất.

Tối ưu trọng số theo mùa nhằm tăng độ chính xác.

Áp dụng trong kịch bản dự báo cho nhiều nhà máy bằng mô hình chọn lọc (Selector-Model) tự động thích nghi với từng nhà máy.

### 3. Đối tượng và phạm vi nghiên cứu

**Đối tượng nghiên cứu:** Công suất phát điện ngắn hạn tại các nhà máy điện mặt trời quy mô công nghiệp ở Việt Nam.

**Phạm vi về không gian:** Dữ liệu thực nghiệm từ các nhà máy tại Quảng Trị, Thanh Hóa và Đắk-Lắk và hai nhà máy tại Trung Quốc.

**Phạm vi thời gian:** Dự báo ngắn hạn trong khoảng 1–3 ngày.

**Đặc trưng và phạm vi dữ liệu:** Bao gồm bức xạ mặt trời, nhiệt độ không khí, nhiệt độ tấm pin, các yếu tố mùa vụ. Tình huống đặc biệt: thiếu chuỗi dữ liệu, dữ liệu ngắt quãng.

#### 4. Phương pháp nghiên cứu

**Lý thuyết:** Luận án kế thừa lý thuyết học máy, học sâu và phương pháp tổ hợp mô hình (Stacking Ensemble) trong xử lý chuỗi thời gian.

##### Phương pháp nghiên cứu:

Khảo sát lý thuyết và các mô hình hiện có.

Thu thập-xử lý dữ liệu từ ba nhà máy tại ba vùng khí hậu.

Xây dựng mô hình dự báo đơn lẻ và tổ hợp.

Đề xuất tối ưu trọng số theo mùa và mô hình chọn lọc tự động.

Đánh giá mô hình qua các chỉ số RMSE, NRMSE, NMAPE, MAPE, thời gian huấn luyện, khả năng mở rộng.

#### 5. Nguồn tài liệu tham khảo

Tài liệu được sử dụng trong nghiên cứu gồm:

Nguồn dữ liệu thực nghiệm từ ba nhà máy điện mặt trời ở Việt Nam và hai nhà máy tại Trung Quốc.

Các công trình nghiên cứu trong và ngoài nước về mô hình dự báo công suất năng lượng tái tạo, mô hình học máy và học sâu.

Tài liệu kỹ thuật, báo cáo và công bố khoa học về các mô hình dự báo và ứng dụng học máy

#### 6. Đóng góp của luận án

Luận án đã đưa ra những đóng góp khoa học và thực tiễn nổi bật trong lĩnh vực dự báo công suất phát điện mặt trời, tập trung giải quyết những hạn chế về dữ liệu và khả năng ứng dụng tại Việt Nam. Cụ thể bao gồm:

Đề xuất mô hình tổ hợp SE-XGB-LGBM-RF-OW, mô hình Stacking Ensemble kết hợp ba thuật toán cây quyết định (XGBoost, LightGBM, Random Forest) và cơ chế tối ưu trọng số theo mùa (Optimized Weights). Mô hình này giúp nâng cao độ chính xác và tính ổn định của dự báo trong điều kiện khí hậu thay đổi theo mùa.

Xây dựng mô hình kết hợp LightGBM-LSTM, trong đó LightGBM được sử dụng để xử lý và nội suy dữ liệu thiếu, tạo chuỗi dữ

liệu hoàn chỉnh làm đầu vào cho LSTM dự báo công suất. Cách tiếp cận này giúp khắc phục tình trạng thiếu dữ liệu thực tế, cải thiện đáng kể hiệu quả dự báo ngắn hạn.

Phát triển mô hình Selector-Model có khả năng tự động lựa chọn mô hình phù hợp cho từng nhà máy và từng thời điểm dự báo, đảm bảo khả năng mở rộng và tổng quát hóa mô hình cho nhiều nhà máy điện mặt trời ở các vùng khí hậu khác nhau mà không cần huấn luyện lại hoàn toàn.

## **7. Bố cục của luận án**

Ngoài phần mở đầu, kết luận, danh mục tài liệu, phụ lục, nội dung luận án có kết cấu 4 chương chính như sau:

**Chương 1:** Tổng quan nghiên cứu.

**Chương 2:** Các mô hình dự báo bức xạ mặt trời và công suất phát điện mặt trời trong ngắn hạn.

**Chương 3:** Phân tích đánh giá hiệu suất các mô hình dự báo bức xạ và công suất phát điện mặt trời.

**Chương 4:** Đề xuất mô hình dự báo cho một số nhà máy điện mặt trời tại Việt Nam.

## II. NỘI DUNG NGHIÊN CỨU

### CHƯƠNG 1. TỔNG QUAN NGHIÊN CỨU

Năng lượng điện giữ vai trò then chốt trong chiến lược phát triển bền vững, công nghiệp hóa và nâng cao chất lượng sống. Tuy nhiên, sự phụ thuộc vào nhiên liệu hóa thạch đã dẫn đến ô nhiễm môi trường và cạn kiệt tài nguyên, thúc đẩy xu hướng chuyển dịch sang năng lượng tái tạo. Trong đó, điện mặt trời nổi lên như một giải pháp khả thi nhờ tính thân thiện và khả năng ứng dụng rộng rãi. Tại Việt Nam, lĩnh vực này phát triển nhanh chóng, nhưng lại chịu ảnh hưởng lớn từ điều kiện thời tiết, gây biến động công suất và thách thức cho công tác điều độ hệ thống.

Trước yêu cầu vận hành ổn định và tuân thủ quy hoạch năng lượng quốc gia, bài toán dự báo công suất phát điện mặt trời ngắn hạn trở nên cấp thiết. Chương này sẽ tổng quan về bối cảnh phát triển điện mặt trời, thực trạng và nhu cầu dự báo công suất, cùng các phương pháp dự báo phổ biến hiện nay qua đó làm nền tảng cho việc đề xuất mô hình phù hợp với điều kiện Việt Nam trong các chương sau.

#### 1.1. Tiềm năng điện mặt trời tại Việt Nam

Việt Nam có lợi thế lớn về tài nguyên mặt trời. Bức xạ trung bình đạt 4–5 kWh/m<sup>2</sup>/ngày, cao nhất ~5,7 kWh/m<sup>2</sup>/ngày. Số giờ nắng trung bình: 1.600–2.600 giờ/năm. Phân bố bức xạ khác nhau theo vùng miền, đặc biệt cao ở các tỉnh Nam Trung Bộ và Tây Nguyên.

Triển vọng và mục tiêu tiêu đến 2030 theo quy hoạch điện VIII điều chỉnh năng lượng tái tạo sẽ chiếm 28%–36% tổng công suất lắp đặt. Trong đó, điện mặt trời được kỳ vọng là trụ cột chính để hiện thực hóa mục tiêu này.

#### 1.2. Thực trạng phát triển điện mặt trời tại Việt Nam

Tổng công suất lắp đặt năm 2023 là hơn 16.600 MW. Giai đoạn 2019–2020 chứng kiến sự bùng nổ nhờ cơ chế giá FIT ưu đãi. Tuy nhiên, sự phát triển nhanh khiến hệ thống truyền tải quá tải, và các chính sách hỗ trợ sau đó chưa được điều chỉnh kịp thời.

#### Các khó khăn thách thức trong dự báo công suất phát điện mặt trời

Thách thức Hạ tầng truyền tải quá tải, chính sách thiếu nhất quán, cạnh tranh đất đai với nông nghiệp và bảo tồn sinh thái. Ngoài ra, một số khu vực có tiềm năng nhưng chưa được đầu tư do thiếu kết nối hạ tầng.

#### Nhu cầu và hiện trạng dự báo công suất điện mặt trời

Biến động công suất do ảnh hưởng thời tiết đặt ra yêu cầu dự báo chính xác. Các quy định hiện hành (Quyết định 67/QĐ-ĐTĐL, 4608/ĐDQG-NLTT) yêu cầu dự báo theo khung 15 phút, D+1, D+2, tuần. Độ chính xác yêu cầu: NMAPE <15%.

### Các khung thời gian dự báo

Trong phạm vi tại Việt Nam, nghiên cứu sinh lựa chọn khoảng thời gian từ 1 đến 3 ngày là dự báo ngắn hạn dài ngày nhằm thống nhất với các ứng dụng kỹ thuật phổ biến tại Việt Nam và phù hợp với mục tiêu điều độ, vận hành ngắn hạn của nhà máy

### 1.3. Tổng quan các nghiên cứu về dự báo công suất phát điện mặt trời

#### - Các sai số dùng để đánh giá dự báo

Để đánh giá hiệu quả các mô hình dự báo công suất điện mặt trời, luận án sử dụng năm chỉ số sai số phổ biến.

**RMSE (Root Mean Square Error)** là sai số bình phương trung bình gốc giữa các giá trị dự báo và thực tế, phản ánh độ lệch chuẩn của sai số.

**NRMSE (Normalized RMSE)** là RMSE đã được chuẩn hóa theo tổng công suất lắp đặt (Capacity), giúp so sánh giữa các nhà máy có quy mô khác nhau.

**MAPE (Mean Absolute Percentage Error)** đo sai số tuyệt đối trung bình theo phần trăm, cho biết mức sai lệch tương đối giữa dự báo và thực tế.

**NMAPE (Normalized MAPE)** là sai số MAPE chuẩn hóa theo công suất định mức của nhà máy.

Cuối cùng, **MAE (Mean Absolute Error)** đo sai số tuyệt đối trung bình, thường được dùng cho dữ liệu bức xạ mặt trời ( $W/m^2$ ).

#### - Tổng quan các nghiên cứu

Trong giai đoạn 2018–2024, dự báo công suất điện mặt trời đã trở thành hướng nghiên cứu sôi động với nhiều công trình quốc tế và trong nước tập trung nâng cao độ chính xác và tính ứng dụng thực tiễn. Các hướng tiếp cận nổi bật có thể kể đến gồm: (i) cải tiến các kiến trúc học sâu (ví dụ LSTM, GRU và Transformer) nhằm mô tả tốt hơn các quan hệ phi tuyến phức tạp; (ii) ứng dụng hiệu quả các thuật toán học máy dựa trên cây quyết định (như XGBoost, LightGBM) vốn huấn luyện nhanh và xử lý tốt dữ liệu thiếu; và (iii) phát triển các mô hình kết hợp (hybrid/ensemble) để tận dụng ưu điểm của từng phương pháp.

### 1.4. Khoảng trống nghiên cứu và hướng tiếp cận của luận án

Mặc dù đạt nhiều tiến bộ, tại Việt Nam việc ứng dụng các mô hình hiện đại này còn hạn chế, đặc biệt trong bối cảnh nhiều nhà máy mới có chuỗi dữ liệu lịch sử ngắn hoặc gián đoạn. Khoảng trống nghiên cứu đặt ra là cần một mô hình dự báo linh hoạt, có khả năng thích ứng với dữ liệu đầu vào chưa hoàn hảo và biến động khí hậu địa phương. Ngoài ra, các nghiên cứu hiện tại cũng chưa có một mô hình nào có khả năng thích ứng linh hoạt và khả chuyên giữa các nhà máy nằm ở những vùng khí hậu khác nhau.

Luận án này tập trung giải quyết khoảng trống đó bằng cách đề xuất một cách tiếp cận mô hình tổ hợp thích ứng, nhằm đảm bảo độ chính xác dự báo ngay cả khi dữ liệu đầu vào bị thiếu hụt, đồng thời có thể mở rộng dự báo cho nhiều nhà máy tại các vùng khí hậu khác nhau.

### **1.5. Tiểu kết chương 1**

Việt Nam có tiềm năng lớn để phát triển điện mặt trời nhờ điều kiện khí hậu thuận lợi, song công tác dự báo công suất ngắn hạn vẫn đang đối mặt với nhiều thách thức do dữ liệu đo đạc còn hạn chế, chuỗi dữ liệu vận hành chưa đầy đủ và sự biến động mạnh của thời tiết theo mùa. Chương 1 đã trình bày bối cảnh phát triển điện mặt trời và nhấn mạnh tầm quan trọng của công tác dự báo công suất, đồng thời tổng quan các phương pháp dự báo đã được nghiên cứu và ứng dụng trong và ngoài nước. Kết quả tổng quan cho thấy, mặc dù các mô hình học máy hiện đại đặc biệt là các mô hình tổ hợp có tiềm năng nâng cao độ chính xác dự báo, nhưng vẫn thiếu nghiên cứu phù hợp với điều kiện dữ liệu thực tế của Việt Nam, nơi dữ liệu thường ngắn, gián đoạn và không đồng nhất.

Trên cơ sở đó, luận án xác định khoảng trống nghiên cứu nằm ở việc chưa có một mô hình dự báo hoạt động tốt trong điều kiện thiếu dữ liệu, vừa thích ứng linh hoạt theo mùa và vùng khí hậu, đồng thời có khả năng tổng quát hóa cho nhiều nhà máy khác nhau. Hướng tiếp cận của luận án là phát triển mô hình học máy tổ hợp (Stacking Ensemble) kết hợp học sâu và cây quyết định, cùng cơ chế tối ưu trọng số theo mùa và lựa chọn mô hình thích ứng (Selector-Model), nhằm nâng cao tính ổn định, độ chính xác và khả năng mở rộng dự báo cho nhiều nhà máy, phù hợp với điều kiện thực tế trong việc dự báo công suất phát điện mặt trời tại Việt Nam.

## CHƯƠNG 2: CÁC MÔ HÌNH DỰ BÁO BỨC XẠ MẶT TRỜI VÀ CÔNG SUẤT PHÁT ĐIỆN MẶT TRỜI TRONG NGẮN HẠN

Chương 2 trình bày cơ sở dữ liệu thực tế và các mô hình nền tảng trong dự báo bức xạ và công suất phát điện mặt trời ngắn hạn. Đây là nền tảng để phát triển và đánh giá các mô hình đề xuất trong các chương tiếp theo.

### 2.1. Bộ dữ liệu thực tế và tiền xử lý dữ liệu

Bộ dữ liệu chính được sử dụng trong nghiên cứu được thu thập từ nhà máy điện mặt trời Quảng Trị (công suất 49,5 MW), trong giai đoạn 01/01/2022 – 31/12/2022, với tần suất lấy mẫu 5 phút/lần, tổng cộng 105.120 điểm dữ liệu. Bộ dữ liệu gồm các đặc trưng đầu vào chính: bức xạ mặt trời (GHI), nhiệt độ môi trường, nhiệt độ tấm pin, và công suất phát thực tế. Phân tích thống kê cho thấy bức xạ mặt trời có tương quan mạnh nhất với công suất phát ( $r = 0,987$ ), tiếp theo là nhiệt độ tấm pin ( $r = 0,893$ ), phản ánh rõ mối quan hệ vật lý giữa các yếu tố đầu vào và đầu ra trong hệ thống phát điện mặt trời.

Để đảm bảo chất lượng dữ liệu, quy trình tiền xử lý được áp dụng gồm: (i) làm sạch dữ liệu bằng cách loại bỏ các giá trị bất thường (ví dụ như các điểm có bức xạ bằng 0 nhưng công suất khác 0...). Quy trình tiền xử lý được thực hiện sâu nhằm phân biệt biến động thời tiết và sai lệch vận hành. Luận án đã loại bỏ 265 điểm dữ liệu bất thường (0,25%) do lỗi cảm biến hoặc các nghịch lý vật lý (như bức xạ tăng nhưng công suất giảm mạnh do lệnh điều độ hệ thống hoặc bảo trì). Việc làm sạch này giúp mô hình tập trung học đúng quy luật vật lý tự nhiên, nâng cao độ tin cậy cho dự báo công suất khả dụng; (ii) chuẩn hóa dữ liệu về cùng thang [0,1] theo phương pháp Min–Max scaling để tăng tốc độ huấn luyện và tính ổn định của mô hình; và (iii) lưu giữ tham số chuẩn hóa để bảo toàn tính nhất quán khi áp dụng cho tập kiểm định và dữ liệu dự báo.

Bên cạnh bộ dữ liệu chính, luận án còn sử dụng ba bộ dữ liệu bổ sung từ các nhà máy tại Đắk-Lắk, Thanh Hóa, và Quảng Trị (2024) nhằm kiểm chứng khả năng tổng quát hóa và thích ứng đa vùng của mô hình đề xuất. Đồng thời, nghiên cứu còn mở rộng thu thập dữ liệu vận hành thực tế trong 02 năm từ hai nhà máy tại khu vực phía Bắc và Tây Bắc Trung Quốc nhằm phục vụ công tác kiểm chứng độc lập trên tập dữ liệu nước ngoài. Việc sử dụng dữ liệu quốc tế với đặc thù khí hậu lục địa khắc nghiệt này giúp khẳng định tính khách quan và khả năng thích nghi của các kiến trúc học máy đề xuất trước sự thay đổi của vùng

địa lý, quy mô công suất và đặc tính vận hành khác nhau. Các bộ dữ liệu này có cùng tần suất lấy mẫu và đặc trưng khí tượng giống với bộ dữ liệu chính, song thiếu dữ liệu nhiệt độ tâm pin một yếu tố quan trọng lý giải cho việc lựa chọn bộ dữ liệu Quảng Trị 2022 làm tập nền tảng cho phân tích và so sánh mô hình. Tất cả các bộ dữ liệu bổ sung đều được xử lý theo cùng quy trình làm sạch, chuẩn hóa và đồng bộ đặc trưng, nhằm bảo đảm tính nhất quán khi sử dụng trong các chương sau, đặc biệt ở phần đánh giá mô hình tổ hợp và Selector-Model.

## 2.2. Tổng quan về bức xạ mặt trời và các phương pháp dự báo

Bức xạ mặt trời là tổng năng lượng Mặt trời phát ra đến bề mặt Trái Đất trong một đơn vị diện tích và thời gian ( $W/m^2$ ), là đầu vào cơ bản của hệ thống điện mặt trời và giữ vai trò quyết định trong dự báo công suất phát. Cường độ bức xạ phụ thuộc vào vị trí địa lý, thời gian trong ngày và mùa trong năm, cùng các yếu tố khí quyển như mây, bụi, hơi nước và ozone. Việc hiểu rõ và mô hình hóa các yếu tố này là điều kiện tiên quyết để tăng độ chính xác trong dự báo công suất điện mặt trời.

Các mô hình tính toán bức xạ được phân loại theo mức độ phức tạp. Mô hình cơ bản xác định cường độ bức xạ từ hằng số mặt trời, góc thiên đỉnh và hệ số truyền qua khí quyển. Các mô hình vật lý – thực nghiệm như ASHRAE hay Ineichen–Perez sử dụng dữ liệu khí tượng hoặc ảnh vệ tinh để hiệu chỉnh phù hợp với điều kiện địa phương. Mức độ phức tạp và độ chính xác của mô hình phụ thuộc vào chất lượng dữ liệu đầu vào và năng lực tính toán của hệ thống.

Dự báo bức xạ là bước trung gian bắt buộc trong chuỗi dự báo công suất điện mặt trời, vì sai số trong dự báo bức xạ ảnh hưởng trực tiếp đến sai số công suất. Dự báo bức xạ chính xác giúp nâng cao hiệu quả điều độ, lập kế hoạch vận hành và tối ưu sử dụng pin lưu trữ.

Hiện nay, có bốn hướng tiếp cận chính trong dự báo bức xạ:

Dữ liệu vệ tinh, như Himawari (Nhật Bản) hay GOES (Mỹ), cung cấp ảnh mây và tham số khí quyển theo thời gian thực;

Vector chuyển động của đám mây (CMV), dự báo ngắn hạn bằng phân tích chuyển động của mây;

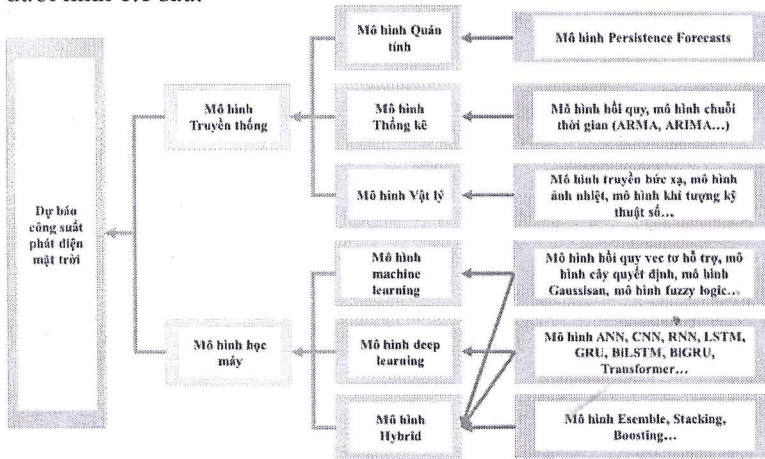
Mô hình dự báo thời tiết số (NWP), mô phỏng vật lý khí quyển ở quy mô trung hạn;

Học máy và học sâu (ML/DL), như ANN, SVM, LightGBM, LSTM, GRU có khả năng xử lý dữ liệu phi tuyến, tự học và cập nhật mô hình khi dữ liệu thay đổi.

Để thu thập và hiệu chỉnh dữ liệu phục vụ dự báo, các nhà máy điện mặt trời thường được trang bị pyranometer (đo bức xạ toàn phần), pyrliometer (đo bức xạ trực tiếp) và hệ thống SCADA ghi nhận công suất phát theo thời gian thực. Hệ thống quan trắc này đảm bảo nguồn dữ liệu chính xác và liên tục yếu tố cốt lõi cho việc huấn luyện và kiểm chứng mô hình dự báo.

### 2.3. Các mô hình dự báo công suất phát điện mặt trời

Các mô hình dự báo công suất phát điện mặt trời được thể hiện dưới hình 1.1 sau:



Hình 2.1. Các mô hình dự báo công suất phát điện mặt trời

Các mô hình dự báo công suất phát điện mặt trời được phân loại thành 2 nhóm chính:

#### - Nhóm Mô hình truyền thống

**Mô hình quán tính (Persistence Forecasts):** Giả định công suất phát tại thời điểm hiện tại sẽ được giữ nguyên trong tương lai gần. Mô hình đơn giản, dùng làm đường cơ sở để so sánh.

**Mô hình thống kê:** Bao gồm các phương pháp hồi quy tuyến tính và chuỗi thời gian như ARMA, ARIMA. Phù hợp khi dữ liệu có tính quy luật và ổn định.

**Mô hình vật lý:** Dựa trên mô phỏng truyền bức xạ mặt trời, ảnh nhiệt, hoặc dữ liệu khí tượng kỹ thuật số. Yêu cầu thông tin chi tiết về điều kiện khí quyển và môi trường địa phương.

#### Nhóm mô hình học máy

**Mô hình machine learning:** Áp dụng các thuật toán như SVM (hồi quy véc-tơ hỗ trợ), cây quyết định, mô hình Gaussian và fuzzy logic. Thích hợp với dữ liệu phi tuyến và có nhiễu.

**Mô hình deep learning:** Gồm các mạng nơ-ron nhiều tầng như ANN, CNN, RNN, LSTM, GRU, BiLSTM, BiGRU, và gần đây là Transformer. Có khả năng tự học đặc trưng dữ liệu, đặc biệt hiệu quả với chuỗi thời gian dài.

**Mô hình hybrid (kết hợp):** Kết hợp nhiều mô hình như học máy-học sâu, hoặc truyền thống-học máy. Các kỹ thuật như Ensemble, Stacking, Boosting được sử dụng để cải thiện độ chính xác và tính ổn định của dự báo.

### 2.3. Tiểu kết chương 2

Chương 2 đã giới thiệu bộ dữ liệu thực tế, quy trình tiền xử lý và các mô hình phổ biến trong dự báo bức xạ cũng như công suất phát điện mặt trời. Các mô hình được phân loại theo nhóm truyền thống, học máy và mô hình kết hợp. Phân tích lý thuyết cho thấy tiềm năng của mô hình học sâu và mô hình tổ hợp trong việc nâng cao độ chính xác, đặc biệt trong điều kiện dữ liệu thiếu hoặc biến động mạnh. Những nội dung này làm cơ sở cho việc xây dựng và kiểm chứng các mô hình cải tiến ở chương sau.

### CHƯƠNG 3: PHÂN TÍCH ĐÁNH GIÁ CÁC MÔ HÌNH DỰ BÁO BỨC XẠ VÀ CÔNG SUẤT PHÁT ĐIỆN MẶT TRỜI

Chương 3 trình bày quá trình xây dựng, đánh giá và so sánh hiệu quả của các mô hình dự báo bức xạ và công suất phát điện mặt trời trong bối cảnh dữ liệu quá khứ không đầy đủ một thách thức phổ biến tại nhiều nhà máy điện mặt trời ở Việt Nam. Các mô hình học máy đơn lẻ và mô hình kết hợp được thử nghiệm nhằm xác định hướng tiếp cận phù hợp nhất, làm nền tảng cho giải pháp tổ hợp trong chương sau.

#### 3.1. Xây dựng mô hình dự báo bức xạ mặt trời.

Dựa trên dữ liệu quan trắc thực tế tại nhà máy điện mặt trời, luận án tiến hành xây dựng và thử nghiệm các mô hình dự báo bức xạ ngắn hạn bằng phương pháp học máy. Các mô hình được lựa chọn gồm LightGBM, LSTM và GRU, đại diện cho hai hướng tiếp cận: học máy truyền thống và học sâu theo chuỗi thời gian.

Quy trình xây dựng mô hình gồm các bước: tiền xử lý dữ liệu, lựa chọn đặc trưng, huấn luyện mô hình và đánh giá hiệu suất bằng các chỉ số sai số như RMSE, MAE, MAPE. Kết quả dự báo được trình bày trong Bảng 3.1 dưới đây nhằm so sánh hiệu quả giữa các mô hình.

Bảng 3.1. Kết quả dự báo bức xạ mặt trời của ba mô hình LightGBM, LSTM, GRU với các tập dữ liệu huấn luyện và dự báo

Các chỉ số		Mô hình LightGBM	Mô hình LSTM	Mô hình GRU
Tập kiểm tra	Thời gian thực hiện (s)	0,534	456,571	397,197
	RMSE (W/m <sup>2</sup> )	54,817	59,164	59,331
	MAE (W/m <sup>2</sup> )	27,599	34,788	34,749
Tập dữ liệu thời tiết ngày 15-02-2021	Thời gian thực hiện (s)	0,121	1,454	1,762
	RMSE (W/m <sup>2</sup> )	60,099	58,837	57,452
	MAE (W/m <sup>2</sup> )	36,219	44,446	42,753
Tập dữ liệu thời tiết các ngày từ 01-03/03/2021	Thời gian thực hiện (s)	0,210	2,479	1,566
	RMSE (W/m <sup>2</sup> )	36,274	38,239	38,136
	MAE (W/m <sup>2</sup> )	21,048	27,478	26,973

Kết quả cho thấy LightGBM là mô hình có hiệu suất ổn định nhất với RMSE và MAE thấp nhất trên cả ba khoảng thời gian dự báo, đồng thời có tốc độ xử lý nhanh vượt trội. LSTM và GRU cho kết quả dự báo tương đương với LightGBM nhưng chậm hơn đáng kể. Như vậy, LightGBM được lựa chọn là mô hình dự báo bức xạ xét về độ ổn định và thời gian huấn luyện nhanh, 2 mô hình LSTM và GRU có thể được xem xét lựa chọn nếu dự báo theo chuỗi thời gian và có đủ tài nguyên tính toán.

### 3.2. Xây dựng mô hình dự báo công suất phát điện mặt trời

Các mô hình được chia thành hai nhóm chính: (1) nhóm chuỗi thời gian gồm LSTM, GRU và BiGRU; (2) nhóm cây quyết định gồm XGBoost, LightGBM và Random Forest. Việc phân nhóm giúp làm rõ ưu nhược điểm của từng hướng tiếp cận trong điều kiện dữ liệu khuyết.

### 3.3. Kết quả đánh giá

Luận án thực hiện so sánh hệ thống hai nhóm mô hình trong kịch bản đầy đủ và thiếu hụt ngẫu nhiên 75% dữ liệu huấn luyện để đánh giá giới hạn ổn định.

Bảng 3.2. So sánh kết quả của các mô hình trong kịch bản đầy đủ và thiếu 75% dữ liệu huấn luyện

Đầy đủ dữ liệu huấn luyện						
Các so sánh	Mô hình XGBoost	Mô hình LightGBM	Mô hình Random Forest	Mô hình LSTM	Mô hình GRU	Mô hình BiGRU
Thời gian hoàn thành (s)	2,12	0,607	16,58	594,512	609,591	562,68
RMSE (kW)	1397,277	1373,272	1431,003	1670,691	1717,464	1554,82
NRMSE (%)	2,823	2,774	2,888	3,387	3,469	3,141
NMAPE (%)	1,012	1,004	1,033	1,35	1,521	1,18

MAPE (%)	Không xác định	Không xác định	Không xác định	Không xác định	Không xác định	Không xác định
<b>Thiếu ngẫu nhiên 75% tập dữ liệu huấn luyện</b>						
<b>Các so sánh</b>	<b>Mô hình XGBoost</b>	<b>Mô hình LightGBM</b>	<b>Mô hình Random Forest</b>	<b>Mô hình LSTM</b>	<b>Mô hình GRU</b>	<b>Mô hình BiGRU</b>
Thời gian hoàn thành (s)	0,56	0,51	2,32	450,51	469,42	440,34
RMSE (kW)	2501,27	3047,88	2519,07	3339,81	3406,79	3196,59
NRMSE (%)	5,05	6,16	5,09	6,75	6,88	6,46
NMAPE (%)	2,12	2,61	1,52	2,99	2,64	2,74
MAPE (%)	Không xác định	Không xác định	Không xác định	Không xác định	Không xác định	Không xác định

Kết quả tại Bảng 3.2 cho thấy nhóm cây quyết định (XGBoost, LightGBM, RF) vượt trội hoàn toàn về tốc độ huấn luyện và độ chính xác so với nhóm học sâu. Đặc biệt, khi dữ liệu huấn luyện bị thiếu hụt nghiêm trọng, Random Forest duy trì tính ổn định tốt nhất với sai số NMAPE chỉ 1,52%.

Nhằm đánh giá tác động của sai số dữ liệu đầu vào đến kết quả dự báo công suất, luận án thực hiện so sánh hiệu năng của các mô hình đơn lẻ trong hai trường hợp sử dụng bức xạ thực tế và bức xạ dự báo tại Bảng 3.3.

Bảng 3.3. So sánh sai số dự báo công suất phát điện mặt trời giữa các mô hình đơn lẻ trong giai đoạn 01-03/03/2021 trong hai trường hợp sử dụng bức xạ dự báo từ mô hình LightGBM tại chương 2 và sử dụng bức xạ thực tế

Sử dụng bức xạ thực tế

Các mô hình	RMSE (kW)	NRMSE (%)	MAPE (%)	NMAPE (%)
XGB	1457,02	2,94	59,31	1,26
LGBM	1438,34	2,91	59,8	1,25
RF	1483,57	3,00	59,36	1,3
LSTM	9489,09	19,16	99,5	9,05
GRU	9489,06	19,16	99,5	9,05
BiGRU	1941,6	3,92	124,11	1,81
<b>Sử dụng bức xạ dự báo từ mô hình LightGBM</b>				
Các mô hình	RMSE (kW)	NRMSE (%)	MAPE (%)	NMAPE (%)
XGB	1591,75	3,21	144,23	1,87
LGBM	1534,36	3,1	145,63	1,82
RF	1591,03	3,21	151,3	1,88
LSTM	9489,13	19,17	99,6	9,06
GRU	9489,41	19,171	99,51	9,06
BiGRU	1636,78	3,31	178,8	1,97

Kết quả thực nghiệm cho thấy nhóm cây quyết định hoạt động ổn định, có sai số thấp và thời gian huấn luyện ngắn. Ngược lại, nhóm chuỗi thời gian gặp khó khăn khi chuỗi dữ liệu không liên tục, dẫn đến sai số tăng cao. Trong nhóm này, BiGRU tỏ ra hiệu quả hơn so với LSTM và GRU.

### 3.3. Mô hình kết hợp LightGBM-LSTM

Nghiên cứu sinh tiến hành thử nghiệm phương án kết hợp LightGBM để nội suy dữ liệu bị thiếu và LSTM để dự báo công suất. Kết quả dự báo của mô hình kết hợp được trình bày trong bảng 3.4 dưới đây:

Bảng 3.4. So sánh sai số dự báo giữa các mô hình đơn và mô hình kết hợp LGBM-LSTM cho các ngày 01-03/03/2021

Kết quả dự báo với thông số bức xạ thực tế					
Mô hình	Tỷ lệ dữ liệu quá khứ bị thiếu (%)	RMSE (kW)	NRMSE (%)	MAPE (%)	NMAPE (%)
LSTM (đầy đủ dữ liệu quá khứ)	0	1460,52	2,95	41,04	1,40

<b>XGBoost</b>	<b>0-100</b>	1510,83	3,10	64,47	1,30
<b>LightGBM</b>	<b>0-100</b>	1431,22	2,90	57,73	1,20
<b>Random Forest</b>	<b>0-100</b>	1493,81	3,01	65,02	1,30
<b>LSTM (không được lấp đầy dữ liệu quá khứ)</b>	<b>10</b>	2268,99	4,58	42,92	1,86
	<b>20</b>	3188,13	6,44	48,93	2,59
	<b>30</b>	3872,06	7,82	55,62	3,18
	<b>50</b>	5225,56	10,56	62,67	4,42
	<b>100</b>	8940,90	18,06	97,41	8,58
<b>LightGBM-LSTM</b>	<b>10</b>	1513,61	3,06	42,18	1,45
	<b>20</b>	1519,97	3,07	46,55	1,46
	<b>30</b>	1553,51	3,14	45,37	1,48
	<b>50</b>	1712,14	3,46	49,06	1,6
	<b>100</b>	1977,93	4,06	62,62	1,85
<b>Kết quả dự báo với thông số bức xạ dự báo từ mô hình LightGBM</b>					
<b>Mô hình</b>	<b>Tỷ lệ thiếu dữ liệu quá khứ (%)</b>	<b>RMSE (kW)</b>	<b>NRMSE (%)</b>	<b>MAPE (%)</b>	<b>NMAPE (%)</b>
<b>LSTM (đầy đủ dữ liệu quá khứ)</b>	<b>0</b>	1447,32	2,92	115,79	1,63
<b>XGBoost</b>	<b>0-100</b>	1562,32	3,16	138,98	1,80
<b>LightGBM</b>	<b>0-100</b>	1543,65	3,12	138,04	1,81
<b>Random Forest</b>	<b>0-100</b>	1535,78	3,10	163,65	1,85

<b>LSTM (không được lấp đầy dữ liệu quá khứ)</b>	<b>10</b>	2231,18	4,51	116,49	2,07
	<b>20</b>	3255,04	6,58	115,47	2,68
	<b>30</b>	3753,65	7,58	119,56	3,25
	<b>50</b>	5007,9	10,12	127,86	4,45
	<b>100</b>	8902,82	17,99	151,08	8,60
<b>LightGBM- LSTM</b>	<b>10</b>	1491,53	3,01	121,33	1,69
	<b>20</b>	1561,36	3,15	130,03	1,76
	<b>30</b>	1650,54	3,33	132,73	1,85
	<b>50</b>	1666,44	3,37	143,67	1,98
	<b>100</b>	2064,96	4,17	192,64	2,41

Dù kết quả được cải thiện so với LSTM đơn lẻ, mô hình LightGBM-LSTM này vẫn chưa vượt qua hiệu suất của các mô hình cây quyết định độc lập trong điều kiện dữ liệu khuyết.

### **3.4. Hướng chuyển sang Stacking Ensemble**

Từ các kết quả trên, luận án đề xuất hướng xây dựng mô hình tổ hợp dạng Stacking Ensemble. Cách tiếp cận này nhằm tận dụng thế mạnh riêng của từng mô hình, giảm thiểu sai số và tăng tính linh hoạt trong bối cảnh dữ liệu hạn chế.

### **3.5. Tiểu kết chương 3**

Chương 3 đã tiến hành đánh giá hệ thống các mô hình dự báo trong điều kiện thiếu dữ liệu quá khứ. Kết quả cho thấy các mô hình cây quyết định (LightGBM, XGBoost, Random Forest) có độ chính xác cao và thời gian xử lý nhanh, trong khi nhóm mô hình học sâu theo chuỗi thời gian (LSTM, GRU, BiGRU) bị suy giảm hiệu suất khi dữ liệu không liên tục. Mô hình kết hợp LightGBM-LSTM cải thiện sai số so với LSTM đơn lẻ nhưng vẫn chưa vượt trội hơn nhóm cây quyết định. Những phân tích này làm cơ sở cho việc đề xuất mô hình tổ hợp Stacking Ensemble trong chương kế tiếp, nhằm tận dụng ưu điểm của từng mô hình và tăng tính linh hoạt khi ứng phó với dữ liệu thiếu hụt.

## CHƯƠNG 4: ĐỀ XUẤT MÔ HÌNH DỰ BÁO CHO MỘT SỐ NHÀ MÁY ĐIỆN MẶT TRỜI TẠI VIỆT NAM

Chương 4 đề xuất mô hình tổ hợp Stacking Ensemble nhằm khắc phục hạn chế về dữ liệu không đầy đủ và nâng cao hiệu quả dự báo công suất phát điện mặt trời trong thực tiễn. Mô hình được thiết kế với cấu trúc linh hoạt, có khả năng tối ưu hóa theo mùa và thích ứng với nhiều điều kiện khí hậu vùng miền, qua đó mở rộng khả năng ứng dụng tại các nhà máy điện mặt trời ở Việt Nam và trên thế giới.

### 4.1. Mô hình SE-XGB-LGBM-RF Tổ hợp 3 mô hình cây quyết định.

Luận án đề xuất mô hình tổ hợp Stacking gồm ba thuật toán cây quyết định: XGBoost, LightGBM và Random Forest (SE-XGB-LGBM-RF). Mô hình này tận dụng ưu điểm chung của các mô hình gốc như tốc độ huấn luyện nhanh, khả năng xử lý dữ liệu thiếu và độ chính xác cao. Kết quả thực nghiệm cho thấy đây là mô hình có hiệu suất tốt nhất trong nghiên cứu với sai số RMSE và NMAPE thấp trên nhiều tập dữ liệu thử nghiệm.

#### +) So sánh với mô hình lai SE-XGB-LGBM-LSTM, SE-XGB-LGBM-GRU, SE-XGB-LGBM-BiGRU

Các mô hình tổ hợp lai giữa cây quyết định và mạng nơ-ron (SE-XGB-LGBM-LSTM, v.v.) cũng được xây dựng nhằm kiểm tra khả năng cải thiện độ chính xác. Kết quả so sánh các mô hình được thể hiện trong bảng 4.1 dưới đây:

Bảng 4.1: Kết quả so sánh dự báo công suất phát điện mặt trời của các mô hình SE-XGB-LGBM-RF, SE-XGB-LGBM-BiGRU, SE-XGB-LGBM-GRU, SE-XGB-LGBM-LSTM

Các tiêu chí đánh giá	Mô hình SE-XGB-LGBM-RF	Mô hình SE-XGB-LGBM-BiGRU	Mô hình SE-XGB-LGBM-GRU	Mô hình SE-XGB-LGBM-LSTM
Thời gian dự báo (s)	23,28	101,57	98,39	98,94
RMSE (kW)	1453,85	1497,67	3359,07	3359,08

<b>NRMSE (%)</b>	2,93	3,02	6,78	6,79
<b>MAPE (%)</b>	59,39	79,95	54,67	54,68
<b>NMAPE (%)</b>	1,26	1,35	3,07	3,08

Kết quả từ bảng 4.1 cho thấy do đặc trưng của các mô hình mạng nơ-ron phụ thuộc vào chuỗi liên tục, hiệu suất của các mô hình lai này không vượt qua mô hình SE-XGB-LGBM-RF khi dữ liệu bị khuyết. Do đó nghiên cứu sinh lựa chọn mô hình SE-XGB-LGBM-RF để tiến hành xây dựng trọng số tối ưu theo mùa.

#### 4.2. Tối ưu trọng số theo mùa

Nhằm tăng khả năng thích nghi theo điều kiện khí hậu vùng miền, mô hình SE-XGB-LGBM-RF được tối ưu trọng số bằng thuật toán L-BFGS-B cho từng mùa (xuân, hạ, thu, đông) được gọi là mô hình SE-XGB-LGBM-RF-OW. Kết quả dự báo của mô hình đã được tối ưu trọng số theo các ngày điển hình trong mỗi mùa được thể hiện dưới bảng 4.2 dưới đây. Kết quả cho thấy sai số NMAPE giảm rõ rệt, duy trì dưới 2% ở tất cả các mùa, chứng minh tính linh hoạt và hiệu quả của tối ưu hóa theo mùa.

Bảng 4.2. Kết quả dự báo trong các ngày điển hình của một mùa theo với trọng số tối ưu cho mô hình SE-XGB-LiGBM-RF-OW

Công suất		$P_{max}$ (kW)	$P_{min}$ (kW)	$P_{trung\ bình}$ (kW)	RMSE (kW)	NRMSE (%)	MAPE (%)	NMAPE (%)
Mùa xuân	Thực tế	41541,56	0	7928,31	1183,33	2,390	8,572	1,112
	Dự báo	40841,49	1,673					
	$\Delta P_{thực\ tế.}$ dự báo	700,07	1,673					
Mùa hạ	Thực tế	40287,79	0	9970,87	939,332	1,897	12,616	0,744
	Dự báo	39869,53	0					
	$\Delta P_{thực\ tế.}$ dự báo	418,26	0					
Thực tế		40710,07	0	6495,09	635,920	1,284	16,721	0,539

Mùa thu	Dự báo	39194,77	0					
	$\Delta P_{\text{thực tế}}$ dự báo	1515,3	0					
Mùa đông	Thực tế	39052,66	0	3780,31	975,106	1,969	17,797	0,874
	Dự báo	38945,49	0					
	$\Delta P_{\text{thực tế}}$ dự báo	107,17	0					

### 4.3. Dự báo cho nhiều nhà máy

Dữ liệu từ ba nhà máy ở ba vùng khí hậu (Quảng Trị – miền Trung, Đắk Lắk – Tây Nguyên, Thanh Hóa – Bắc Trung Bộ) được sử dụng để kiểm chứng khả năng mở rộng của mô hình SE-XGB-LGBM-RF. Hai kịch bản được xây dựng:

**Kịch bản 1 (Selector-model):** gộp dữ liệu từ cả ba nhà máy để huấn luyện một mô hình duy nhất, sau đó chọn đầu ra theo từng vùng;

**Kịch bản 2:** huấn luyện riêng biệt cho từng nhà máy với dữ liệu đặc thù từng vùng.

Kết quả dự báo cho ba nhà máy của mô hình SE-XGB-LGBM-RF trong hai kịch bản được thể hiện dưới bảng 4.3 sau đây:

Bảng 4.3. Kết quả dự báo từ mô hình Selector-Model và mô hình SE-XGB-LGBM-RF khi được huấn luyện riêng biệt cho từng nhà máy

Mô hình dự báo sử dụng Selector-Model				
Nhà máy	RMSE (kW)	NRMSE (%)	MAPE (%)	NMAPE (%)
Thanh Hóa	2718,84	9,06	85,5	4,94
Quảng Trị	2310,55	4,66	49,2	2,07
Đắk-Lắk	3938,48	7,87	49,5	4,22
Mô hình SE-XGB-LGBM-RF dự báo riêng cho 3 nhà máy				
Nhà máy	RMSE (kW)	NRMSE (%)	MAPE (%)	NMAPE (%)
Thanh Hóa	2773,94	9,26	78,75	4,97
Quảng Trị	1897,83	3,83	8,7	1,32
Đắk-Lắk	3309,02	6,68	40,67	3,26

Từ bảng 4.3 cho thấy mô hình huấn luyện riêng có sai số thấp hơn tại các nhà máy có điều kiện thời tiết biến động mạnh. Tuy nhiên, kịch bản Selector-model cho thấy khả năng khái quát tốt và thuận lợi hơn khi áp dụng cho mở rộng hệ thống dự báo công suất phát, quản lý vận hành cho các nhà máy điện mặt trời có tính tương đồng trong điều kiện khí hậu.

Luận án áp dụng cơ chế Học chuyển giao (Transfer Learning) để cập nhật bộ trọng số tối ưu của mô hình SE-XGB-LGBM-RF-OW và Selector-Model định kỳ 3-5 năm. Việc này giúp các mô hình thích nghi với sự xuống cấp tự nhiên của tấm pin (0,5%–1%/năm) và thay đổi cấu hình thiết bị, đảm bảo hiệu suất dự báo bền vững suốt vòng đời dự án.

#### 4.4. Đánh giá mô hình SE-XGB-LGBM-RF-OW và Selector-Model trên tập dữ liệu nước ngoài

Để kiểm chứng khả năng tổng quát hóa, mô hình được thử nghiệm trên dữ liệu vận hành của hai nhà máy tại Trung Quốc (nhà máy 1 có công suất lắp đặt 130 MW và nhà máy 2 có công suất 110 MW) có khí hậu lục địa khắc nghiệt và áp suất khí quyển thấp. Kết quả được thể hiện tại Bảng 4.4

Bảng 4.4. Kết quả dự báo từ mô hình Selector-Model và mô hình SE-XGB-LGBM-RF-OW cho hai nhà máy tại Trung Quốc

Nhà máy 1 (công suất 130MW)				
Mô hình	RMSE (MW)	NRMSE (%)	MAPE (%)	NMAPE (%)
Selector-Model	6,89	5,13	18,3	2,54
SE-XGB-LGBM-RF-OW	6,76	5,2	18,52	2,52
Nhà máy 2 (công suất 110MW)				
Mô hình	RMSE (MW)	NRMSE (%)	MAPE (%)	NMAPE (%)
Selector-Model	2,35	1,87	10,83	1,25
SE-XGB-LGBM-RF-OW	2,19	1,8	10,77	1,24

Bảng 4.4 cho thấy sai số NMAPE của hai mô hình dự báo tại cả hai nhà máy đều dưới 3%, qua đó cho thấy các kiến trúc đề xuất có khả năng thích nghi và tổng quát hóa tốt. Kết quả này cho thấy các mô hình có thể nắm bắt được các quy luật quang điện mang tính phổ quát mà không phụ thuộc vào đặc điểm địa lý cục bộ.

#### 4.5. Tiểu kết chương 4

Chương 4 đã xây dựng và thử nghiệm mô hình tổ hợp SE-XGB-LGBM-RF, cho thấy hiệu suất vượt trội so với các mô hình đơn lẻ và các mô hình học sâu trong điều kiện dữ liệu khuyết. Việc tối ưu trọng số theo mùa (SE-XGB-LGBM-RF-OW) giúp giảm đáng kể sai số dự báo, đặc biệt là chỉ số NMAPE luôn duy trì dưới 2%. Thử nghiệm tại ba nhà máy tại Việt Nam ở các vùng khí hậu khác nhau cũng khẳng định khả năng tổng quát hóa và mở rộng mô hình. Hai kịch bản ứng dụng huấn luyện riêng và Selector-model cung cấp lựa chọn linh hoạt tùy theo yêu cầu triển khai thực tế, làm cơ sở cho ứng dụng mô hình vào vận hành và điều độ điện mặt trời quy mô lớn. Đặc biệt, việc kiểm chứng trên hai nhà máy tại Trung Quốc với đặc thù khí hậu lục địa khắc nghiệt cho thấy sai số NMAPE đều duy trì dưới 3%. Kết quả này khẳng định tính ổn định và khả năng tự thích nghi của mô hình trước sự thay đổi của vùng địa lý, quy mô công suất và đặc tính vận hành quốc tế.

## KẾT LUẬN VÀ KIẾN NGHỊ

### 1. Kết luận chính

L luận án đã phát triển một hệ thống mô hình dự báo công suất phát điện mặt trời ngắn hạn có độ chính xác cao, thích ứng tốt với đặc điểm khí hậu biến động của Việt Nam. Những đóng góp nổi bật gồm:

- Đề xuất mô hình tổ hợp SE-XGB-LGBM-RF-OW, tối ưu trọng số theo mùa, giúp tăng độ chính xác và ổn định trong điều kiện khí hậu thay đổi.

- Xây dựng mô hình LightGBM-LSTM để khắc phục thiếu chuỗi dữ liệu quá khứ trong thời gian gần, cải thiện đáng kể hiệu suất so với mô hình LSTM khi không có chuỗi dữ liệu quá khứ đầy đủ.

- Phát triển mô hình Selector-Model, cho phép lựa chọn mô hình tối ưu theo vùng khí hậu và đặc điểm nhà máy, mở rộng khả năng ứng dụng thực tiễn.

### 2. Kiến nghị

Để mở rộng hiệu quả ứng dụng và tiếp tục phát triển hướng nghiên cứu này, tác giả đề xuất:

#### a) Ứng dụng thực tiễn:

- Triển khai mô hình SE-XGB-LGBM-RF-OW tại các nhà máy có thời tiết biến động hoặc thiếu dữ liệu lịch sử.

- Dùng Selector-Model cho các cụm nhà máy trong giai đoạn đầu, sau đó tùy chỉnh riêng biệt nếu cần độ chính xác cao hơn.

- Ứng dụng mô hình hybrid LightGBM-LSTM như một giải pháp kỹ thuật dự phòng nhằm lấp đầy các chuỗi dữ liệu công suất bị thiếu hụt trong ngắn hạn, giúp duy trì tính liên tục và ổn định của hệ thống dự báo ngay cả khi hạ tầng thu thập dữ liệu SCADA gặp sự cố gián đoạn cục bộ.

#### b) Về dữ liệu và hạ tầng:

- Đầu tư hệ thống thu thập dữ liệu tại chỗ, tích hợp dữ liệu vệ tinh và dự báo khí tượng số.

- Nâng cấp năng lực tính toán để triển khai mô hình học sâu trên diện rộng.

#### c) Hướng nghiên cứu tiếp theo:

- Mở rộng sang các nguồn năng lượng tái tạo khác và các hệ thống đa nguồn.

- Phát triển mô hình tổ hợp tự thích ứng, tích hợp các kỹ thuật hiện đại như Transformer, Attention, AutoML nhằm tăng độ chính xác và linh hoạt.

**DANH MỤC CÁC CÔNG TRÌNH KHOA HỌC CỦA TÁC GIẢ  
LIÊN QUAN ĐẾN LUẬN ÁN**

**Hội thảo quốc tế (Scopus): 03 công bố**

[1] Tuan-Anh Nguyen, Manh-Hai Pham, Vu Minh Phap, Quang-hiep Do, Ngoc-Trung Nguyen, Dang-Toan Nguyen, Thi Nga Nguyen., "Forecasting of solar power generation in Vietnam deploying a simple GRU model," Conf. Proc. - 2023 IEEE Asia Meet. Environ. Electr. Eng. EEE-AM 2023, pp. 1-6, 2023, doi: 10.1109/EEEAM58328.2023.10395560.

[2] Tuan-Anh Nguyen, Manh-Hai Pham, Minh-Phap Vu, and Trung-Kien Duong, "A recent invasion wave of deep learning in solar power forecasting techniques using ANN," IEEE Int. Futur. Energy Electron. Conf, 2021, doi: 10.1109/IFEEEC53238.2021.9661747.

[3] Ngoc-Trung Nguyen, Phap Vu Minh, Thi Anh Tho Vu, Tuan-Anh Nguyen, Manh-Hai Pham, Trung Nguyen Thanh, Dang Toan Nguyen, Thanh-Doanh Le, Son Tran-Thanh, "An Experimental Comparison of TCN, LSTM, and GRU for Short-term Solar Power Forecasting," IEEE-2025 12th International Conference on Power and Energy Systems Engineering (CPESE 2025). (đã được Acceptance Full PAPER).

**Bài báo quốc tế (Scopus): 01 công bố**

[1] Tuan Anh Nguyen, Manh Hai Pham, Minh Phap Vu, Ngoc Trung Nguyen, Dang Toan Nguyen, Thi Anh Tho Vu, Trong Tuan Tran, Anh Tuan Do, "A Study on Novel Solar Power Forecasting using XGB-LGBM-RF Hybrid Model and L-BFGS-B Optimization Algorithm", Engineering, Technology & Applied Science Research, vol. 15, no. 4, pp. 24516-24522, 2025.

**Tạp chí trong nước: 03 công bố**

[1] Nguyễn Tuấn Anh, Phạm Mạnh Hải, Vũ Minh Pháp, Nguyễn Đăng Toàn, 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.

[2] Nguyễn Tuấn Anh, Phạm Mạnh Hải, Nguyễn Đăng Toàn, Lê Thành Doanh, Nguyễn Ngọc Trung, Vũ Thị Anh Thơ, Vũ Minh Pháp, Trần Trọng Tuấn, "Đánh giá các mô hình dự báo bức xạ mặt trời: LightGBM, LSTM và GRU," J. Sci. Technol. - HaUI, vol. 60, no. 8, pp. 3-10, Aug. 2024, doi: 10.57001/huih5804.2024.256.

[3] Manh-Hai Pham, Nguyen Tuan Anh, Vu Minh Phap, Pham Văn Duy, Le Thanh Doanh, Vu Thi Anh Tho, Nguyen Dang Toan, “A Comparative Study on the Operational Effectiveness of Machine Learning Models in Solar Power Forecasting” ,” JST: Smart Systems and Devices Volume 35, Issue 2, May 2025, 054-061.



