

**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

---



**Nguyễn Mạnh Hiển**

**NGHIÊN CỨU VÀ ÁP DỤNG MÔ HÌNH LSTM  
CHO BÀI TOÁN DỰ BÁO CÔNG SUẤT PHÁT ĐIỆN  
NHÀ MÁY ĐIỆN GIÓ**

**ĐỀ ÁN TỐT NGHIỆP THẠC SĨ KỸ THUẬT  
(Theo định hướng ứng dụng)**

**HÀ NỘI – 2025**

**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**



**Nguyễn Mạnh Hiển**

**NGHIÊN CỨU VÀ ÁP DỤNG MÔ HÌNH LSTM  
CHO BÀI TOÁN DỰ BÁO CÔNG SUẤT PHÁT ĐIỆN  
NHÀ MÁY ĐIỆN GIÓ**

**CHUYÊN NGÀNH : HỆ THỐNG THÔNG TIN**

**MÃ SỐ: 8.48.01.04**

**ĐỀ ÁN TỐT NGHIỆP THẠC SĨ KỸ THUẬT**

**(Theo định hướng ứng dụng)**

**NGƯỜI HƯỚNG DẪN KHOA HỌC**

**TS. NGUYỄN NGỌC ĐIỆP**

**HÀ NỘI – 2025**

## LỜI CAM ĐOAN

Tôi cam đoan rằng đề án tốt nghiệp thạc sĩ: “**Nghiên cứu và áp dụng mô hình LSTM cho bài toán dự báo công suất phát điện nhà máy điện gió**” là công trình nghiên cứu của chính tôi.

Các số liệu, kết quả nêu trong đề án là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

Không có sản phẩm/nghiên cứu nào của người khác được sử dụng trong đề án này mà không được trích dẫn theo đúng quy định.

Hà Nội, ngày 29 tháng 07 năm 2025

Học viên thực hiện đề án



**Nguyễn Mạnh Hiền**

## LỜI CẢM ƠN

Trong suốt quá trình học tập và nghiên cứu thực hiện đề án tốt nghiệp thạc sĩ, ngoài nỗ lực của bản thân, tôi đã nhận được sự hướng dẫn nhiệt tình quý báu của quý Thầy Cô, cùng với sự động viên và ủng hộ của gia đình, bạn bè và đồng nghiệp. Với lòng kính trọng và biết ơn sâu sắc, tôi xin gửi lời cảm ơn chân thành tới:

Thầy **TS. Nguyễn Ngọc Diệp**, người thầy kính mến đã hết lòng giúp đỡ, hướng dẫn, động viên, tạo điều kiện cho tôi trong suốt quá trình thực hiện và hoàn thành đề án tốt nghiệp thạc sĩ.

Ban Giám Đốc, Phòng đào tạo sau đại học và quý Thầy Cô đã tạo mọi điều kiện thuận lợi giúp tôi hoàn thành đề án.

Tôi xin chân thành cảm ơn gia đình, bạn bè, đồng nghiệp trong cơ quan đã động viên, hỗ trợ tôi trong lúc khó khăn để tôi có thể học tập và hoàn thành đề án. Mặc dù đã có nhiều cố gắng, nỗ lực, nhưng do thời gian và kinh nghiệm nghiên cứu khoa học còn hạn chế nên không thể tránh khỏi những thiếu sót. Tôi rất mong nhận được sự góp ý của quý Thầy Cô cùng bạn bè đồng nghiệp để kiến thức của tôi ngày một hoàn thiện hơn.

Xin chân thành cảm ơn!

Hà Nội, ngày      tháng      năm 2025

Học viên thực hiện đề án



**Nguyễn Mạnh Hiền**

## MỤC LỤC

|                                                                                                 |            |
|-------------------------------------------------------------------------------------------------|------------|
| <b>LỜI CAM ĐOAN .....</b>                                                                       | <b>i</b>   |
| <b>LỜI CẢM ƠN .....</b>                                                                         | <b>ii</b>  |
| <b>MỤC LỤC.....</b>                                                                             | <b>iii</b> |
| <b>DANH SÁCH HÌNH VẼ.....</b>                                                                   | <b>v</b>   |
| <b>DANH SÁCH BẢNG BIỂU.....</b>                                                                 | <b>vi</b>  |
| <b>DANH MỤC CÁC THẬT NGỮ, CHỮ VIẾT TẮT .....</b>                                                | <b>vii</b> |
| <b>MỞ ĐẦU .....</b>                                                                             | <b>1</b>   |
| <b>CHƯƠNG 1: TỔNG QUAN VỀ BÀI TOÁN DỰ BÁO CÔNG SUẤT PHÁT ĐIỆN NHÀ MÁY ĐIỆN GIÓ.....</b>         | <b>4</b>   |
| 1.1 Tổng quan về năng lượng gió và thực trạng dự báo công suất phát điện nhà máy điện gió ..... | 4          |
| 1.1.1 Tổng quan về năng lượng gió.....                                                          | 4          |
| 1.1.2 Thực trạng dự báo công suất phát điện nhà máy điện gió.....                               | 5          |
| 1.2 Giới thiệu về bài toán dự báo công suất phát điện nhà máy điện gió.....                     | 6          |
| 1.2.1 Mục tiêu của bài toán dự báo công suất phát điện gió .....                                | 6          |
| 1.2.2 Đặc điểm và độ phức tạp của bài toán.....                                                 | 7          |
| 1.2.3 Ý nghĩa thực tiễn của bài toán.....                                                       | 7          |
| 1.3 Các phương pháp tiếp cận hiện tại và hạn chế .....                                          | 8          |
| 1.3.1 Các phương pháp vật lý (physical methods) .....                                           | 8          |
| 1.3.2. Các phương pháp thống kê truyền thống (các mô hình tuyến tính).....                      | 8          |
| 1.3.3 Các phương pháp học máy (Machine Learning).....                                           | 9          |
| 1.3.4 Các phương pháp lai(hybrid prediction methods).....                                       | 10         |
| 1.3.5 Hạn chế và thách thức .....                                                               | 11         |
| 1.4 Một số lý thuyết có liên quan .....                                                         | 12         |
| 1.4.1 Khái niệm chuỗi thời gian và đặc điểm của dữ liệu đo công suất phát điện .....            | 12         |
| 1.4.2 Các tiêu chí đánh giá mô hình dự báo:.....                                                | 12         |
| 1.4.3 Các kỹ thuật tiền xử lý dữ liệu .....                                                     | 14         |
| 1.5 Kết luận chương.....                                                                        | 15         |

|                                                           |           |
|-----------------------------------------------------------|-----------|
| <b>CHƯƠNG 2: PHƯƠNG PHÁP ĐỀ XUẤT .....</b>                | <b>16</b> |
| 2.1. Cơ sở lựa chọn phương pháp .....                     | 16        |
| 2.2 Thiết kế mô hình dự báo công suất gió bằng LSTM ..... | 18        |
| 2.2.1 Mô hình hóa bài toán dự báo công suất.....          | 18        |
| 2.2.2 Cấu trúc mạng LSTM được đề xuất .....               | 19        |
| 2.2.3 Quy trình huấn luyện mô hình.....                   | 21        |
| 2.2.4 Đánh giá ưu điểm mô hình LSTM .....                 | 26        |
| 2.3 Kết luận chương.....                                  | 26        |
| <b>CHƯƠNG 3: THỬ NGHIỆM VÀ ĐÁNH GIÁ.....</b>              | <b>27</b> |
| 3.1 Thu thập và tiền xử lý dữ liệu.....                   | 27        |
| 3.1.1 Nguồn dữ liệu.....                                  | 27        |
| 3.1.2 Tiền xử lý dữ liệu .....                            | 30        |
| 3.2 Thử nghiệm và đánh giá kết quả.....                   | 31        |
| 3.2.1 Cấu hình thí nghiệm .....                           | 31        |
| 3.2.2 Kết quả thử nghiệm .....                            | 32        |
| 3.3 Kết luận chương.....                                  | 40        |
| <b>KẾT LUẬN .....</b>                                     | <b>41</b> |
| <b>DANH MỤC CÁC TÀI LIỆU THAM KHẢO .....</b>              | <b>44</b> |

## DANH SÁCH HÌNH VẼ

|                                                                                   |    |
|-----------------------------------------------------------------------------------|----|
| Hình 1.1 Dự báo về ngành điện gió đến năm 2030.....                               | 5  |
| Hình 1.2 Hai loại mô hình dự báo kết hợp.....                                     | 11 |
| Hình 2.1 Mô hình mỗi khối (cell) trong LSTM .....                                 | 16 |
| Hình 2.2 Mô hình hóa đầu vào - đầu ra trong bài toán dự báo công suất phát điện . | 19 |
| Hình 2.3 Kiến trúc mô hình LSTM 1 lớp ẩn.....                                     | 20 |
| Hình 2.4 Kiến trúc mô hình LSTM 2 lớp ẩn.....                                     | 21 |
| Hình 3.1 Biểu đồ phân phối Công suất phát điện .....                              | 28 |
| Hình 3.2 Biểu đồ tương quan tốc độ gió – Công suất.....                           | 28 |
| Hình 3.3 Biểu đồ Công suất trung bình theo năm.....                               | 29 |
| Hình 3.4 Biểu đồ Công suất trung bình theo tháng.....                             | 29 |
| Hình 3.5 Kết quả dự báo 4h tiếp theo .....                                        | 34 |
| Hình 3.6 Kết quả dự báo 6h tiếp theo .....                                        | 35 |
| Hình 3.7 Kết quả dự báo 10h tiếp theo .....                                       | 36 |
| Hình 3.8 Kết quả dự báo 4h tiếp theo GRU .....                                    | 38 |
| Hình 3.9 Kết quả dự báo 4h tiếp theo MLP .....                                    | 39 |
| Hình 3.10 Kết quả dự báo 6h tiếp theo với GRU.....                                | 39 |
| Hình 3.11 Kết quả dự báo 6h tiếp theo với MLP .....                               | 40 |

## DANH SÁCH BẢNG BIỂU

|                                                                               |    |
|-------------------------------------------------------------------------------|----|
| Bảng 1.1 Phân loại dự báo công suất phát điện dựa trên khoảng thời gian ..... | 6  |
| Bảng 3.1 Dữ liệu sử dụng trong nghiên cứu .....                               | 27 |
| Bảng 3.2 Các tham số mô hình theo kích thước cửa sổ trượt .....               | 31 |
| Bảng 3.3: Đánh giá dự báo theo kích thước cửa sổ trượt .....                  | 32 |
| Bảng 3.4: Đánh giá dự báo theo số lần lặp .....                               | 32 |
| Bảng 3.5: Đánh giá dự báo theo số lần lặp .....                               | 33 |
| Bảng 3.7 Kết quả dự báo 4h tiếp theo .....                                    | 34 |
| Bảng 3.8 Kết quả dự báo 6h tiếp theo .....                                    | 34 |
| Bảng 3.9 Kết quả dự báo 10h tiếp theo .....                                   | 35 |
| Bảng 3.10 Các tham số của mô hình GRU .....                                   | 37 |
| Bảng 3.11 Kết quả dự báo 4h tiếp theo GRU .....                               | 37 |
| Bảng 3.12 Các tham số của mô hình MLP .....                                   | 38 |
| Bảng 3.13 Kết quả dự báo 4h tiếp theo MLP .....                               | 38 |
| Bảng 3.14 So sánh kết quả dự báo các mô hình(4h tiếp theo) .....              | 39 |

## DANH MỤC CÁC THẬT NGỮ, CHỮ VIẾT TẮT

|                |                                                   |
|----------------|---------------------------------------------------|
| LSTM           | Long Short-term Memory                            |
| CNN            | Convolutional Neural Network                      |
| RNN            | Recurrent Neural Network                          |
| ARIMA          | AutoRegressive Integrated Moving Average          |
| SARIMA         | Seasonal AutoRegressive Integrated Moving Average |
| SVM            | Support Vector Machine                            |
| SVR            | Support Vector Regression                         |
| XGBoost        | Extreme Gradient Boosting                         |
| RF             | Random Forest                                     |
| MAE            | Mean Absolute Error                               |
| RMSE           | Root Mean Square Error                            |
| MAPE           | Mean Absolute Percentage Error                    |
| R <sup>2</sup> | Coefficient of Determination                      |
| IRENA          | International Renewable Energy Agency             |
| IQR            | Interquartile Range                               |
| NWP            | Numerical Weather Prediction                      |
| MLP            | Multilayer Perceptron                             |
| GRU            | Gated Recurrent Unit                              |

## MỞ ĐẦU

### 1. Tính cấp thiết của đề tài:

Trong những năm gần đây, điện gió đã trở thành một trong những nguồn năng lượng tái tạo phát triển nhanh nhất tại Việt Nam. Tính đến tháng 7/2023, tổng công suất lắp đặt điện gió tại Việt Nam đã đạt 1.000 MW, chiếm khoảng 4% tổng công suất điện lắp đặt của cả nước.

Trong thời gian tới, điện gió tiếp tục được kỳ vọng là một trong những nguồn năng lượng tái tạo chủ lực của Việt Nam. Chính phủ Việt Nam đã đặt mục tiêu đến năm 2030, tổng công suất lắp đặt điện gió đạt 12.000 MW [1].

Bên cạnh các lợi ích như bù đắp năng lượng thiếu hụt, tận dụng năng lượng tại chỗ, giảm thiểu khí nhà kính, việc các nguồn điện gió tham gia vào lưới điện với tỉ trọng công suất cao cũng tạo ra một số thách thức đối với vận hành ổn định hệ thống điện quốc gia.

Cụ thể, sự phát triển mạnh mẽ của năng lượng điện gió gây ảnh hưởng đến việc điều độ, huy động các nhà máy điện khác và làm tăng nhu cầu dự phòng để đảm bảo tính ổn định hệ thống điện.

Dự báo công suất điện gió là một trong các giải pháp cấp thiết cho các thách thức trên, đóng vai trò quan trọng trong công tác lập kế hoạch, quản lý và vận hành hệ thống điện. Với mục đích đưa những nghiên cứu ứng dụng công nghệ dự báo vào bài toán thực tiễn, tôi xin chọn đề tài nghiên cứu “**Nghiên cứu và áp dụng mô hình LSTM cho bài toán dự báo công suất phát điện nhà máy điện gió**”. Tôi hy vọng những kết quả đạt được của đề tài này có thể là một cơ sở quan trọng trong việc xây dựng các hệ thống dự báo công suất phát điện cũng như mở rộng hơn là hệ thống quản lý vận hành nhà máy điện gió ở nước ta.

### 2. Tổng quan về vấn đề nghiên cứu:

Do những yêu cầu về quản lý, vận hành hệ thống điện cũng như nhằm khai thác, sử dụng hiệu quả nguồn tài nguyên này, một trong những công việc quan trọng là dự báo công suất phát nguồn điện gió. Dựa vào đặc điểm riêng của mỗi nhà máy

điện gió, có thể nghiên cứu, áp dụng những phương pháp dự báo công suất phát điện gió phù hợp.

Có thể phân loại dự báo công suất điện gió theo nhiều cách khác nhau. Xét theo thang thời gian, dự báo được chia theo 04 loại: Dự báo cực ngắn hạn – từ vài phút đến 01 giờ (ultra-short-term forecasting), Dự báo ngắn hạn – từ 01 giờ đến vài giờ (Short-term); Dự báo trung hạn – từ vài giờ đến 01 tuần (Medium-term); Dự báo dài hạn – từ 01 tuần đến 01 năm trở lên (Long-term) [2].

Về phương pháp dự báo, có nhiều mô hình dự báo được sử dụng. Trên thế giới, phương pháp sử dụng kỹ thuật học máy (machine learning) và các mô hình lai (hybrid methods) để dự báo công suất điện gió được sử dụng phổ biến. Các phương pháp học sâu (deep learning) sử dụng thư viện TensorFlow như: Phương pháp kết hợp TensorFlow và phân tích các thành phần chính PCA [3]; Phương pháp kết hợp mạng nơron liên kết (CNN) và mạng nơron chức năng cơ sở hướng tâm (RBFNN) [4].

Ở Việt Nam, tác giả Lê Hà Phan và các cộng sự đã nghiên cứu sử dụng thuật toán ANFIS kết hợp phương pháp phân nhóm (Clustering) để dự báo công suất gió trước một ngày [5]. Nhóm tác giả Đinh Thành Việt và cộng sự nghiên cứu và ứng dụng học sâu trong dự báo công suất phát nguồn điện gió [6]...

Mỗi phương pháp dự báo công suất phát điện gió đều có những ưu, nhược điểm riêng, phù hợp với từng đặc điểm của mỗi nhà máy.

Với những yêu cầu trong công tác vận hành hệ thống điện, đề án đề xuất một phương pháp dự báo điện gió thích hợp với bài toán dự báo trước 30 phút, 01 giờ, 24 giờ,... ứng dụng cho việc lập kế hoạch điều độ, đưa ra các quyết định vận hành hợp lý và đảm bảo ổn định hoạt động thị trường điện.

### **3. Mục đích nghiên cứu:**

Nghiên cứu các phương pháp dự báo công suất phát điện gió bằng mô hình LSTM và áp dụng dự báo công suất phát điện của các nhà máy điện gió.

Mục tiêu cụ thể bao gồm:

- Nắm được mô hình LSTM và các thuật toán tối ưu trong dự báo
- Áp dụng và đánh giá hiệu quả một số mô hình LSTM trong dự báo công suất phát điện của nhà máy điện gió

#### **4. Đối tượng và phạm vi nghiên cứu:**

##### ***4.1. Đối tượng nghiên cứu***

- Năng lượng điện gió và các phương pháp dự báo công suất phát điện của nhà máy điện gió

##### ***4.2. Phạm vi nghiên cứu***

- Các phương pháp dự báo tốc độ gió và công suất phát cho nhà máy điện gió

#### **5. Phương pháp nghiên cứu:**

##### ***4.3. Phương pháp nghiên cứu lý thuyết***

- Đọc và tìm hiểu về mạng LSTM và một số mô hình học máy học sâu
- Nghiên cứu các mô hình dự báo
- Nghiên cứu các kỹ thuật xử lý số liệu

##### ***4.4. Phương pháp thực nghiệm***

- Thu thập bộ dữ liệu huấn luyện từ dữ liệu về gió tại nhà máy điện gió
- Thử nghiệm mô hình đề xuất, phân tích và đánh giá kết quả đạt được

# CHƯƠNG 1: TỔNG QUAN VỀ BÀI TOÁN DỰ BÁO CÔNG SUẤT PHÁT ĐIỆN NHÀ MÁY ĐIỆN GIÓ

## 1.1 Tổng quan về năng lượng gió và thực trạng dự báo công suất phát điện nhà máy điện gió

### 1.1.1 Tổng quan về năng lượng gió

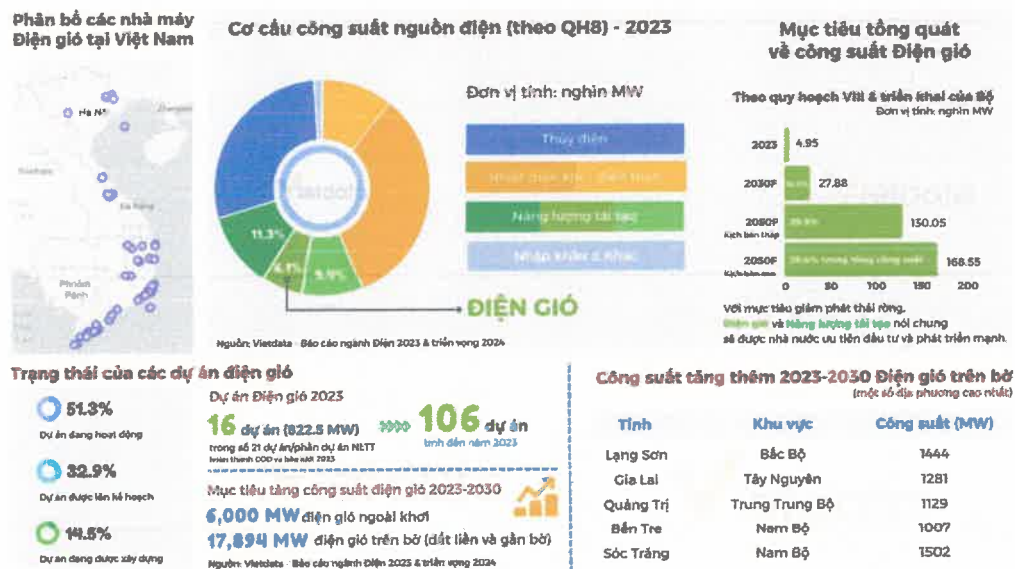
Năng lượng gió là một trong những nguồn năng lượng tái tạo quan trọng, được khai thác từ động năng của không khí chuyển động trong khí quyển. Khác với các nguồn năng lượng hóa thạch như than đá, dầu mỏ hay khí đốt, năng lượng gió không gây phát thải khí nhà kính và không tiêu tốn tài nguyên hữu hạn. Do vậy, phát triển năng lượng gió được xem là một trong những giải pháp chiến lược nhằm hướng đến phát triển bền vững và giảm thiểu biến đổi khí hậu toàn cầu.

Nguyên lý hoạt động của hệ thống điện gió dựa trên việc các tua-bin gió (wind turbines) chuyển đổi động năng của gió thành cơ năng, sau đó được máy phát điện chuyển hóa thành điện năng. Điện năng này có thể được hòa vào lưới điện quốc gia hoặc sử dụng tại chỗ. Với đặc điểm thân thiện với môi trường, hiệu quả khai thác ngày càng tăng và chi phí lắp đặt giảm dần theo thời gian, điện gió đang ngày càng được nhiều quốc gia ưu tiên phát triển.

Theo báo cáo của Cơ quan Năng lượng Tái tạo Quốc tế (IRENA), tổng công suất điện gió toàn cầu đã đạt khoảng 1.133GW vào cuối năm 2024 [7], tăng mạnh so với những năm trước. Nhiều quốc gia như Trung Quốc, Mỹ, Đức và Tây Ban Nha đã khai thác hiệu quả tiềm năng gió để đảm bảo an ninh năng lượng và giảm phụ thuộc vào nhiên liệu hóa thạch.

Việt Nam được đánh giá là một quốc gia có tiềm năng rất lớn về năng lượng gió, đặc biệt là tại các khu vực ven biển miền Trung và miền Nam. Với tổng công suất kỹ thuật lên đến 1.068 Gigawatt (GW) tại vùng đặc quyền kinh tế (EEZ) tính ở độ cao 100 mét; và 57,8 GW tại vùng ven bờ đến 6 hải lý, Việt Nam đứng trước cơ hội lớn để trở thành trung tâm năng lượng tái tạo khu vực, đồng thời thực hiện các cam kết quốc tế về biến đổi khí hậu[8].

Trong những năm gần đây, điện gió đã trở thành một trong những lĩnh vực được ưu tiên phát triển tại Việt Nam. Tính đến cuối năm 2023, tổng công suất điện gió lắp đặt tại Việt Nam đạt khoảng 4,948.5 MW[9], với hàng chục dự án đang vận hành và hàng trăm dự án đang trong quá trình khảo sát và phê duyệt. Tuy nhiên, phần lớn các nhà máy điện gió vẫn đang gặp nhiều thách thức trong quá trình vận hành, đặc biệt là vấn đề biến động công suất phát điện[16].



Hình 1.1 Dự báo về ngành điện gió đến năm 2030

### 1.1.2 Thực trạng dự báo công suất phát điện nhà máy điện gió

Một trong những đặc điểm lớn nhất của năng lượng gió là tính không ổn định và khó dự đoán. Cường độ gió có thể thay đổi nhanh chóng theo thời gian, phụ thuộc vào nhiều yếu tố như địa hình, thời tiết, mùa vụ, áp suất khí quyển,... Điều này khiến công suất phát điện từ các tuabin gió thay đổi liên tục, dẫn đến khó khăn trong việc điều độ nguồn điện, dự báo nguồn cung và đảm bảo an toàn cho hệ thống lưới điện[16].

Hiện nay, các nhà máy điện gió tại Việt Nam vẫn chủ yếu sử dụng các phương pháp dự báo truyền thống hoặc dựa vào công cụ dự báo thời tiết từ các cơ quan khí tượng, hoặc phụ thuộc vào các tính năng dự báo được tích hợp sẵn trong hệ thống điều hành mua từ đối tác nước ngoài. Những phương pháp này thường chưa đủ chính xác hoặc không phù hợp để đáp ứng nhu cầu vận hành theo thời gian thực, đặc biệt là trong môi trường biến động mạnh như năng lượng gió[18].

## 1.2 Giới thiệu về bài toán dự báo công suất phát điện nhà máy điện gió

Dự báo công suất phát điện là một trong những nhiệm vụ quan trọng trong vận hành hệ thống năng lượng tái tạo, đặc biệt là điện gió. Do đặc tính biến động, không ổn định và phụ thuộc vào thời tiết của nguồn năng lượng gió, công suất phát ra từ các tuabin gió không thể giữ ở mức cố định như các nhà máy nhiệt điện hay thủy điện truyền thống. Điều này đặt ra yêu cầu cấp thiết trong việc xây dựng các mô hình dự báo nhằm dự đoán trước lượng điện năng có thể phát trong tương lai gần, phục vụ điều độ hệ thống điện và vận hành thị trường điện.

### 1.2.1 Mục tiêu của bài toán dự báo công suất phát điện gió

Mục tiêu chính của bài toán dự báo công suất điện gió là ước lượng chính xác công suất đầu ra của hệ thống điện gió trong một khoảng thời gian xác định trong tương lai, dựa vào các dữ liệu quan trắc và vận hành trong quá khứ.

Tùy vào mục đích sử dụng, khi xét theo miền thời gian, dự báo công suất phát điện có thể được chia làm bốn loại chính[2]:

- Dự báo cực ngắn hạn (ultra-short-term forecasting): từ vài phút đến 1 giờ, ứng dụng cho bài toán điều hành lưới điện thời gian thực.
- Dự báo ngắn hạn (short-term forecasting): từ một giờ đến vài giờ, thường được dùng cho mục đích điều độ và ổn định hệ thống.
- Dự báo trung hạn (medium-term forecasting): từ vài giờ đến vài ngày, phục vụ cho lập kế hoạch vận hành và dự phòng.
- Dự báo dài hạn (long-term forecasting): từ vài tuần đến vài tháng, sử dụng cho quy hoạch phát triển và đầu tư.

**Bảng 1.1 Phân loại dự báo công suất phát điện dựa trên khoảng thời gian**

| Phân loại    | Khoảng thời gian   | Ứng dụng                                      |
|--------------|--------------------|-----------------------------------------------|
| Cực ngắn hạn | Vài phút đến 1h    | • Bài toán điều hành lưới điện thời gian thực |
| Ngắn hạn     | 1h đến vài giờ     | • Điều độ và ổn định hệ thống                 |
| Trung hạn    | Vài giờ đến 1 tuần | • Lập kế hoạch vận hành và dự phòng           |
| Dài hạn      | 1 tuần đến 1 năm   | • Quy hoạch phát triển và đầu tư              |

Trong khuôn khổ của đề án này, bài toán tập trung vào dự báo công suất ngắn hạn (short-term forecasting), trong đó đầu vào là chuỗi thời gian lịch sử các thông số như: tốc độ gió, hướng gió, áp suất không khí, nhiệt độ, và công suất phát điện thực tế tại từng thời điểm trong quá khứ; đầu ra là công suất phát điện tại thời điểm trong tương lai.

### ***1.2.2 Đặc điểm và độ phức tạp của bài toán***

Bài toán dự báo công suất phát điện gió là một bài toán dự báo chuỗi thời gian phi tuyến, nhiều biến đầu vào, với các đặc điểm sau:

- **Đặc tính phi tuyến và ngẫu nhiên:** Mối quan hệ giữa tốc độ gió và công suất phát điện không tuyến tính và chịu ảnh hưởng bởi nhiều yếu tố môi trường khác như nhiệt độ, áp suất, hướng gió[19].
- **Độ trễ thời gian và tính liên tục:** Công suất phát tại một thời điểm không chỉ phụ thuộc vào các giá trị đầu vào hiện tại mà còn phụ thuộc vào các trạng thái trước đó, đòi hỏi mô hình có khả năng học được mối quan hệ theo thời gian[20].
- **Nhu cầu xử lý và phân tích dữ liệu lớn:** Các dữ liệu đo đạc từ cảm biến tại nhà máy có thể chứa nhiều, sai số hoặc mất mát dữ liệu, làm tăng thách thức cho mô hình dự báo[20].

Chính vì vậy, các phương pháp học sâu như mạng nơ-ron hồi tiếp (RNN) và đặc biệt là Long Short-Term Memory (LSTM) vốn được thiết kế để xử lý chuỗi thời gian có tính phụ thuộc lâu dài và đã chứng minh hiệu quả trong nhiều bài toán dự báo phức tạp, sẽ có vai trò nổi bật trong bài toán dự báo này.

### ***1.2.3 Ý nghĩa thực tiễn của bài toán***

Việc dự báo chính xác công suất phát điện từ nhà máy điện gió mang lại nhiều lợi ích thiết thực:

- **Hỗ trợ điều độ hệ thống điện:** Giúp trung tâm điều độ lưới điện có kế hoạch phân bổ tải, tránh quá tải hoặc thiếu hụt nguồn.
- **Tối ưu hóa vận hành nhà máy:** Giảm thiểu chi phí khởi động, ngừng máy, và bảo trì thiết bị.
- **Tăng hiệu quả tham gia thị trường điện:** Đối với các đơn vị phát điện tham gia thị trường cạnh tranh, dự báo tốt giúp xây dựng chiến lược giao dịch điện năng

hiệu quả.

- Hỗ trợ quy hoạch năng lượng: Các kết quả dự báo dài hạn có thể phục vụ cho các cơ quan quản lý nhà nước trong việc xây dựng chiến lược phát triển năng lượng quốc gia[18].

### **1.3 Các phương pháp tiếp cận hiện tại và hạn chế**

Trong những năm gần đây, nhiều phương pháp dự báo công suất phát điện từ năng lượng gió đã được nghiên cứu và ứng dụng. Các phương pháp này có thể được phân thành bốn nhóm chính: phương pháp vật lý (physical methods), phương pháp thống kê (statistical methods), phương pháp học máy (machine learning methods) và phương pháp lai (hybrid prediction methods)[21]. Mỗi phương pháp có ưu nhược điểm khác nhau, phụ thuộc vào đặc tính dữ liệu, mục tiêu dự báo và yêu cầu thực tiễn.

#### ***1.3.1 Các phương pháp vật lý (physical methods)***

Được coi là phương pháp dự báo lâu đời nhất. Phương pháp này dựa trên việc sử dụng các mô hình dự báo thời tiết số (Numerical Weather Prediction - NWP) để mô phỏng các điều kiện khí tượng trong tương lai.

Sau khi có dự báo về các biến khí tượng, các thông tin này được sử dụng để tính toán công suất phát điện của các tua-bin gió thông qua các mô hình vật lý mô phỏng quá trình chuyển đổi từ năng lượng gió thành điện năng. Điều này bao gồm việc sử dụng các đặc tính kỹ thuật của tua-bin, như đường cong công suất, và các yếu tố địa hình, như độ nhám bề mặt và địa hình xung quanh.

Tuy nhiên, việc cập nhật dữ liệu NWP không thường xuyên và ở thời gian thực khiến mô hình vật lý gặp vấn đề về tính chính xác, điều này khiến chúng không phù hợp với các dự báo ngắn hạn, trung hạn và dài hạn[21].

#### ***1.3.2. Các phương pháp thống kê truyền thống (các mô hình tuyến tính)***

Các phương pháp thống kê truyền thống dự báo chủ yếu dựa trên dữ liệu lịch sử phát điện mà không dựa trên thông tin khí tượng liên quan.

- Phương pháp bền vững (Persistence method): Dự báo công suất hoặc tốc độ gió trong tương lai bằng chính giá trị hiện tại. Phương pháp này đơn giản và đôi khi hiệu quả cho dự báo rất ngắn hạn, nhưng độ chính xác giảm nhanh khi thời gian dự báo kéo dài[22].

- ARIMA (Autoregressive Integrated Moving Average):

Đây là mô hình dự báo chuỗi thời gian phổ biến, hoạt động hiệu quả với các chuỗi có tính tuyến tính cao. ARIMA sử dụng sự kết hợp giữa tự hồi quy (AR), trung bình trượt (MA) và phép lấy sai phân (I) để dự báo giá trị tương lai[23].

Hạn chế của nhóm các phương pháp này:

- Giả định dữ liệu phải tuyến tính và ổn định theo thời gian – điều không phù hợp với bản chất phi tuyến, biến động và không đều của dữ liệu thực tế từ điện gió – như đã đề cập bên trên.
- Khả năng xử lý dữ liệu quy mô lớn hoặc có tính phi tuyến cao rất hạn chế.
- Không tận dụng được mối quan hệ phụ thuộc dài hạn trong chuỗi thời gian, dẫn đến kết quả dự báo kém chính xác khi chuỗi phức tạp.

### ***1.3.3 Các phương pháp học máy (Machine Learning)***

Để giải quyết các giới hạn của các mô hình tuyến tính, các phương pháp học máy đã được áp dụng tương đối thành công trong bài toán này. Ứng dụng học máy trong dự báo năng lượng gió có thể được phân loại thành các mô hình học máy truyền thống và các mô hình học sâu.

Các phương pháp học máy truyền thống gồm:

- RF(Random Forest):

Mô hình rừng ngẫu nhiên (RF) xây dựng nhiều cây quyết định độc lập và kết hợp kết quả để tăng độ chính xác. RF có khả năng xử lý dữ liệu phi tuyến, chống overfitting khá tốt[24]. Tuy nhiên, do không khai thác được thông tin tuần tự theo thời gian, RF thường kém hiệu quả hơn các mô hình chuỗi thời gian(như LSTM) trong dự báo công suất phát điện gió.

- SVR(Support Vector Regression):

Một biến thể của Support Vector Machine (SVM), được dùng cho bài toán hồi quy. Phù hợp với dữ liệu có độ nhiễu thấp và số chiều không quá cao[25]. SVR không khai thác tốt quan hệ theo chuỗi thời gian, nên hiệu quả có thể giảm trong các bài toán dự báo công suất điện gió nhiều bước và có tính dao động cao.

Các phương pháp học sâu có thể phân làm hai loại dựa theo hướng tiếp cận.

Các mô hình không gian như:

- Mạng nơ-ron tích chập CNN (Convolutional neural networks)[26]

Mạng nơ-ron tích chập (CNN) được thiết kế chủ yếu cho xử lý dữ liệu dạng lưới như ảnh, nhưng đã được mở rộng hiệu quả sang các bài toán chuỗi thời gian bằng cách trích xuất đặc trưng cục bộ từ cửa sổ trượt. Tuy nhiên, do thiếu cơ chế ghi nhớ dài hạn, CNN thường không phù hợp với các bài toán dự báo công suất phát điện gió có tính phụ thuộc thời gian rõ rệt.

- DBN(Deep belief networks)[27]

DBN là một mô hình học sâu bao gồm các lớp Restricted Boltzmann Machines xếp chồng, có khả năng học biểu diễn đặc trưng ẩn hiệu quả từ dữ liệu đầu vào phức tạp. Tuy nhiên, DBN khó huấn luyện ổn định, thiếu khả năng mô hình hóa quan hệ theo thời gian một cách rõ rệt, nên ít được ưu tiên trong các bài toán dự báo công suất phát điện gió.

Các mô hình thời gian tập trung xử lý dữ liệu tuần tự, như:

- Mạng nơ-ron hồi quy RNN(Requirement Neural Network)

RNN là mô hình chuyên xử lý dữ liệu chuỗi, với khả năng ghi nhớ ngắn hạn nhờ kết nối hồi tiếp theo thời gian, giúp khai thác được tính liên tục của dữ liệu như trong dự báo công suất phát điện gió. RNN xử lý tốt trên các dữ liệu chuỗi, tuy nhiên, RNN thường gặp vấn đề về gradient biến mất hoặc bùng nổ khi xử lý chuỗi dài, khiến hiệu suất suy giảm so với các biến thể cải tiến như LSTM hay GRU trong thực tế[28].

- Mạng bộ nhớ dài ngắn hạn LSTM(Long short-term memory)

LSTM là một biến thể của RNN, được thiết kế đặc biệt để:

- Học các phụ thuộc thời gian dài hạn
- Tránh các vấn đề vanishing gradient trong RNN truyền thống
- Dự báo chuỗi thời gian hiệu quả (ví dụ : tốc độ gió, công suất phát điện) [29]

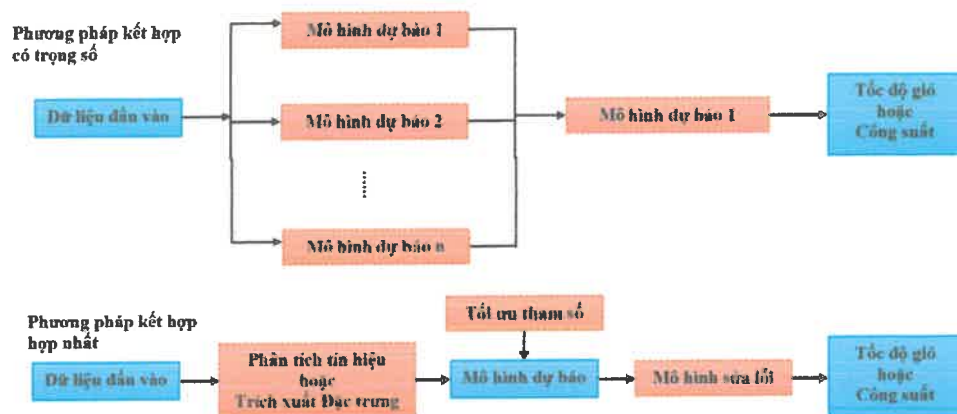
LSTM đặc biệt phù hợp cho bài toán dự báo công suất điện gió nhờ khả năng ghi nhớ và khai thác các quan hệ phụ thuộc theo thời gian, giúp mô hình hóa tốt đặc tính dao động và phi tuyến của dữ liệu gió. So với các mô hình truyền thống, LSTM thường cho độ chính xác cao hơn trong các bài toán dự báo nhiều bước.

#### ***1.3.4 Các phương pháp lai(hybrid prediction methods)***

Các phương pháp dự đoán kết hợp, tích hợp các điểm mạnh của nhiều mô hình

khác nhau và tối ưu hóa sự kết hợp của chúng, đóng vai trò quan trọng trong việc nâng cao độ chính xác của dự báo trong dự báo năng lượng gió. Điều này đặc biệt quan trọng do tính phức tạp của dự báo năng lượng gió do nhiều yếu tố ảnh hưởng, đặc biệt là trong điều kiện thời tiết khắc nghiệt, khi một mô hình duy nhất tỏ ra không đủ và dễ mắc lỗi.

Một số mô hình dự báo kết hợp được áp dụng ở [30,31]. Hình bên dưới mô tả hai loại mô hình dự báo kết hợp: phương pháp kết hợp có trọng số và phương pháp kết hợp hợp nhất:



Hình 1.2 Hai loại mô hình dự báo kết hợp

### 1.3.5 Hạn chế và thách thức

Dù có nhiều hướng tiếp cận, bài toán dự báo công suất phát điện gió vẫn còn nhiều hạn chế và thách thức:

- Tính không ổn định cao: Các mô hình dễ bị sai lệch khi điều kiện thời tiết thay đổi đột ngột.
- Phụ thuộc dữ liệu lịch sử chất lượng cao: Thiếu dữ liệu hoặc dữ liệu nhiễu ảnh hưởng lớn đến kết quả.
- Khó chọn tham số và kiến trúc tối ưu: Các mô hình deep learning có nhiều siêu tham số, dễ dẫn đến quá khớp (overfitting) nếu không xử lý cẩn thận.
- Thiếu đánh giá toàn diện: Nhiều nghiên cứu mới dừng ở mức kiểm nghiệm trên bộ dữ liệu nhỏ, chưa được thử nghiệm thực tế trên quy mô lớn hoặc dữ liệu nhà máy cụ thể tại Việt Nam [7].

## 1.4 Một số lý thuyết có liên quan

Trong quá trình nghiên cứu và xây dựng mô hình dự báo công suất phát điện gió sử dụng LSTM, đề án nghiên cứu một số lý thuyết cơ bản sau:

### 1.4.1 Khái niệm chuỗi thời gian và đặc điểm của dữ liệu đo công suất phát điện

Chuỗi thời gian (time series) là tập hợp các giá trị đo đạc theo thứ tự thời gian. Trong bối cảnh nhà máy điện gió, chuỗi dữ liệu thường bao gồm: tốc độ gió, hướng gió, nhiệt độ, áp suất không khí và công suất phát điện, được thu thập theo chu kỳ cố định (thường là theo giờ hoặc theo phút)[53].

Đặc điểm của chuỗi thời gian công suất phát điện gió gồm:

- Tính không ổn định (non-stationarity): dữ liệu có xu hướng thay đổi theo thời gian do ảnh hưởng của điều kiện thời tiết.
- Tính phi tuyến (non-linearity): mối quan hệ giữa tốc độ gió và công suất không hoàn toàn tuyến tính.
- Chu kỳ và mùa vụ (seasonality): biểu hiện rõ rệt theo thời điểm trong ngày hoặc mùa trong năm.
- Độ phân giải cao: dữ liệu thu thập liên tục với tần suất cao đòi hỏi mô hình có khả năng lưu trữ và học từ chuỗi dài.

### 1.4.2 Các tiêu chí đánh giá mô hình dự báo:

Trong lĩnh vực dự báo năng lượng điện gió, việc đánh giá chất lượng của mô hình dự báo được thực hiện thông qua một số chỉ số phổ biến[21]:

- MAE (Mean Absolute Error):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1.1)$$

Chỉ số MAE – Giá trị tuyệt đối trung bình – thể hiện độ sai lệch trung bình tuyệt đối giữa giá trị thực và giá trị dự báo. MAE càng nhỏ thể hiện mô hình dự báo càng chính xác. MAE thể hiện trực tiếp độ lệch trung bình mà mô hình dự báo mắc phải – dễ hiểu, dễ so sánh và minh họa trực quan. MAE là chỉ số cốt lõi trong đánh giá mô hình dự báo điện gió, giúp hiểu được độ sai lệch trung bình một cách trực tiếp và dễ giải thích.

- MSE (Mean Square Error):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1.2)$$

Là chỉ số đo sai số bình phương trung bình giữa giá trị thực tế và giá trị dự báo. Nó là trung bình cộng của bình phương hiệu số giữa dự báo và thực tế. MSE đo độ sai lệch trung bình, nhưng nhấn mạnh vào các sai số lớn do việc bình phương. Giá trị MSE càng nhỏ, mô hình càng chính xác. Trong bài toán dự báo điện gió, MSE thường được dùng làm hàm mất mát (loss function) trong quá trình huấn luyện mô hình vì dễ tính đạo hàm và ổn định trong tối ưu hóa.

- RMSE (Root Mean Square Error):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1.3)$$

Là chỉ số đo độ lệch chuẩn (sai số) giữa giá trị thực tế và giá trị dự báo, bằng cách bình phương sai số, tính trung bình, rồi căn bậc hai. RMSE đo lường độ lớn của sai số, ưu tiên tăng nặng hơn với các sai số lớn (vì có bình phương). Cũng giống như chỉ số MAE, giá trị RMSE càng nhỏ, mô hình càng chính xác. Vì RMSE phạt mạnh sai số lớn, nếu giá trị cao bất thường có thể phán đoán mô hình có vấn đề với các biến động bất thường của gió. RMSE thường được chọn làm hàm mất mát (loss function) trong mô hình hồi quy (như LSTM), mô hình cố gắng giảm RMSE qua các epoch.

- MAPE (Mean Absolute Percentage Error):

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (1.4)$$

Là chỉ số đo sai số tuyệt đối trung bình theo tỷ lệ phần trăm (%) giữa giá trị dự báo và giá trị thực tế. MAPE cho biết trung bình mô hình sai bao nhiêu phần trăm so với giá trị thực tế. Dễ hiểu, dễ so sánh giữa các mô hình và các tập dữ liệu khác nhau vì không phụ thuộc đơn vị. MAPE cho phép so sánh độ chính xác giữa các mô hình khác nhau trên cùng dữ liệu hoặc khác nhau về đơn vị. MAPE thường được dùng làm tiêu chí chính trong các báo cáo kỹ thuật hoặc học thuật. Trong bài toán dự báo công suất phát điện gió, MAPE là chỉ số đánh giá mô hình phổ biến nhất trong thực tiễn và nghiên cứu, nhờ tính trực quan và dễ hiểu.

Trong các công thức, các khái niệm:

- $y_i$ : Giá trị thực tế
- $\hat{y}_i$ : Giá trị dự đoán
- $\bar{y}_i$ : giá trị trung bình của tập dữ liệu thực tế
- $n$ : Số lượng mẫu

Mỗi chỉ số phản ánh một khía cạnh khác nhau của độ chính xác, và có thể được sử dụng kết hợp để có cái nhìn toàn diện hơn[22].

### 1.4.3 Các kỹ thuật tiền xử lý dữ liệu

Trước khi đưa dữ liệu vào mô hình dự báo, đặc biệt là các mô hình học máy hoặc học sâu như LSTM, cần thực hiện nhiều bước tiền xử lý nhằm chuẩn hóa, làm sạch và tối ưu hóa hiệu quả học của mô hình. Một số kỹ thuật tiền xử lý phổ biến:

#### a. Xử lý thiếu dữ liệu (Missing Values)

Trong thực tế, dữ liệu đo đạc từ các trạm khí tượng hoặc nhà máy điện gió thường xảy ra mất mát do lỗi cảm biến hoặc truyền dữ liệu. Một số cách xử lý gồm:

- Nội suy tuyến tính (linear interpolation) giữa các điểm lân cận.
- Lấp đầy bằng trung bình hoặc giá trị gần nhất (forward/backward fill).
- Loại bỏ bản ghi nếu tỷ lệ thiếu dữ liệu quá lớn.

#### b. Phát hiện và xử lý giá trị ngoại lai (Outlier Detection)

Giá trị bất thường có thể do lỗi đo hoặc các hiện tượng cực đoan trong tự nhiên.

Các phương pháp phát hiện bao gồm:

- Kiểm tra ngưỡng (z-score hoặc IQR).
- Phát hiện bằng mô hình học không giám sát như Isolation Forest.

#### c. Chuẩn hóa dữ liệu (Normalization / Standardization)

Dữ liệu đầu vào với thang đo khác nhau (vd: tốc độ gió m/s, áp suất hPa, công suất MW) cần được đưa về cùng phạm vi để mô hình học hiệu quả:

- Min-Max Scaling: đưa dữ liệu về khoảng  $[0, 1]$ .
- Standard Scaling: chuẩn hóa theo phân phối chuẩn (trung bình 0, độ lệch chuẩn 1).

#### d. Tạo đặc trưng thời gian (Time Features)

Việc thêm các đặc trưng thời gian giúp mô hình học được tính mùa vụ và xu hướng:

- Giờ trong ngày, ngày trong tuần, tháng trong năm (biến rời rạc hoặc chuyển đổi sin/cos để phản ánh chu kỳ).
- Lag features: giá trị tại các thời điểm trước đó (vd: công suất tại  $t-1$ ,  $t-2$ ,...).
- Rolling statistics: trung bình trượt, độ lệch chuẩn trượt trong khoảng thời gian nhất định.

*e. Tách tập huấn luyện, kiểm tra theo thời gian*

Do tính tuần tự của chuỗi thời gian, việc tách dữ liệu không thể ngẫu nhiên mà cần theo trình tự:

- Train-test split phải đảm bảo không rò rỉ thông tin tương lai vào quá khứ.
- Trong nhiều trường hợp, có thể sử dụng Cross-validation theo khối thời gian

(TimeSeriesSplit)

## 1.5 Kết luận chương

Chương 1 đã trình bày tổng quan về năng lượng gió, tầm quan trọng của việc dự báo công suất phát điện trong các nhà máy điện gió, cũng như các phương pháp tiếp cận hiện tại trong lĩnh vực này. Bên cạnh đó, chương cũng đã giới thiệu khái quát về bài toán dự báo công suất phát điện từ dữ liệu khí tượng và công suất lịch sử, đồng thời nêu bật các thách thức đặc thù như tính phi tuyến, sự biến động mạnh và sự phụ thuộc theo thời gian của dữ liệu.

Qua việc khảo sát các phương pháp truyền thống và hiện đại, có thể nhận thấy rằng các mô hình học máy, đặc biệt là mô hình mạng nơ-ron hồi tiếp (RNN) và LSTM, ngày càng được ứng dụng rộng rãi trong bài toán dự báo chuỗi thời gian, nhờ khả năng học các mối quan hệ phức tạp và dài hạn trong dữ liệu.

Bên cạnh đó, các khái niệm chuỗi thời gian cùng với các chỉ số đánh giá hiệu suất dự báo đã được trình bày nhằm làm cơ sở cho việc xây dựng mô hình trong các chương tiếp theo.

Trong chương 2, đề án sẽ trình bày chi tiết phương pháp đề xuất sử dụng mô hình LSTM để dự báo công suất phát điện nhà máy điện gió, bao gồm kiến trúc mô hình, chiến lược huấn luyện và các yếu tố kỹ thuật liên quan.

## CHƯƠNG 2: PHƯƠNG PHÁP ĐỀ XUẤT

### 2.1. Cơ sở lựa chọn phương pháp

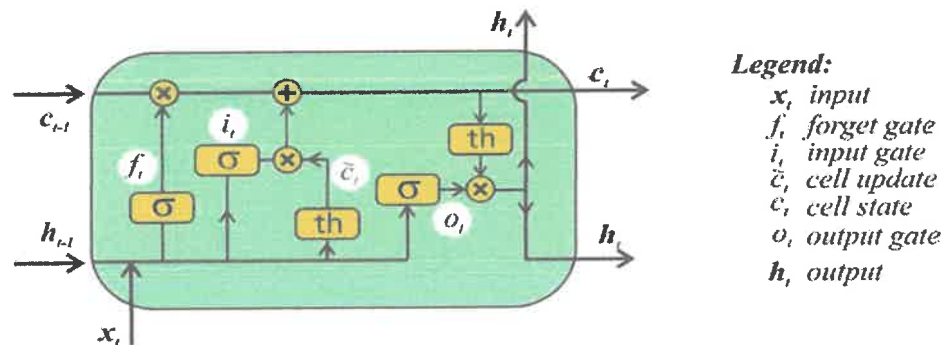
Dựa trên các phân tích tổng quan và đánh giá các phương pháp trong Chương 1, đề án đề xuất sử dụng LSTM (Long Short-Term Memory) để giải quyết bài toán dự báo công suất phát điện nhà máy điện gió. Lý do mô hình LSTM được đề xuất vì nó có khả năng vượt trội trong việc học và lưu giữ thông tin lâu dài, phù hợp với bài toán chuỗi thời gian. Các nghiên cứu của Kong và cộng sự ở [54] về dự báo nhu cầu sử dụng điện ngắn hạn của gia đình, nghiên cứu của Marino và cộng sự về Dự báo tải điện của tòa nhà ở [55], hay của Wang và cộng sự ở [37]..đã chứng minh điều này.

Mô hình LSTM được đề xuất bởi Hochreiter và Schmidhuber [10], là một cải tiến quan trọng của mạng nơ-ron hồi tiếp(RNN), đặc biệt hiệu quả trong việc học các phụ thuộc dài hạn trong chuỗi thời gian.

LSTM được thiết kế với cơ chế bộ nhớ có kiểm soát, cho phép nó vượt qua hiện tượng mất mát gradient – vốn là rào cản chính khiến RNN khó học được các chuỗi có độ dài lớn.

Cấu trúc của LSTM bao gồm ba cổng chính: cổng quên (forget gate), cổng vào (input gate) và cổng ra (output gate). Các cổng này được điều khiển bởi các hàm kích hoạt sigmoid( $\sigma$ ) và hàm tanh( $\tanh$ ) để điều tiết luồng thông tin trong quá trình huấn luyện, giúp mô hình quyết định nên lưu trữ, cập nhật hoặc loại bỏ thông tin nào tại mỗi bước thời gian.

Ngoài ra, LSTM còn có trạng thái ô nhớ (cell state  $C_t$ ), là nơi lưu giữ thông tin xuyên suốt chuỗi thời gian. Nhờ đó, LSTM có khả năng duy trì trạng thái nhớ ổn định và phản ánh các quan hệ phức tạp trong chuỗi dữ liệu phi tuyến.



Hình 2.1 Mô hình mỗi khối (cell) trong LSTM

Giả sử tại bước thời gian  $t$ , đầu vào là  $x_t$ , đầu ra trước đó là  $h_{t-1}$ , và trạng thái nhớ trước là  $C_{t-1}$ , các bước xử lý trong LSTM được mô tả như sau:

- Cổng quên (Forget Gate):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2.1)$$

$f_t \in [0,1]$  điều khiển mức độ thông tin  $C_t$  được giữ lại.

- Cổng vào (Input Gate):

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2.2)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (2.3)$$

$i_t$  xác định phần nào của  $\tilde{C}_t$  được thêm vào trạng thái nhớ.

- Cập nhật trạng thái bộ nhớ:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (2.4)$$

- Cổng đầu ra (Output Gate):

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (2.5)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (2.6)$$

$h_t$  là đầu ra hiện tại và sẽ được truyền cho bước kế tiếp.

Trong đó:

- $x_t$ : đầu vào tại thời điểm  $t$
- $h_{t-1}$ : đầu ra của LSTM ở thời điểm trước đó
- $C_t$ : trạng thái bộ nhớ tại thời điểm  $t$
- $\sigma$ : hàm sigmoid
- $\tanh$ : hàm tanh

Phương pháp dự báo sử dụng LSTM đề xuất bao gồm các bước chính:

1. Thu thập và tiền xử lý dữ liệu: đồng bộ thời gian, chuẩn hóa, loại bỏ dữ liệu lỗi.

2. Trích xuất các đặc trưng: tạo cửa sổ thời gian (sliding window) làm đầu vào cho mô hình.

3. Xây dựng mô hình LSTM: với đầu vào là chuỗi các đặc trưng khí tượng (tốc độ gió, hướng gió, nhiệt độ, áp suất, v.v.) và công suất thực tế trong quá khứ; đầu ra là công suất phát điện dự kiến tại thời điểm trong tương lai.

4. Huấn luyện và tối ưu hóa mô hình: điều chỉnh siêu tham số (hyperparameters) như số lớp, số nơ-ron, kích thước cửa sổ trượt, learning rate,...

5. Đánh giá hiệu suất dự báo: sử dụng các chỉ số MAE, MSE, RMSE, MAPE,  $R^2$  và so sánh với mô hình khác.

## 2.2 Thiết kế mô hình dự báo công suất phát điện gió bằng LSTM

### 2.2.1 Mô hình hóa bài toán dự báo công suất

Trong bài toán dự báo công suất phát điện, mô hình đầu vào bao gồm chuỗi thời gian của các đặc trưng khí tượng (nhiệt độ, áp suất không khí, tốc độ gió, hướng gió) và công suất phát điện thực tế tại các thời điểm trước đó. Việc đưa công suất lịch sử vào mô hình giúp tận dụng tính tự tương quan của chuỗi, từ đó cải thiện độ chính xác của dự báo. Đầu ra của mô hình là giá trị công suất phát điện tại bước thời gian tiếp theo, cụ thể:

- Tham số đầu vào:
  - Nhiệt độ
  - Áp suất
  - Hướng gió
  - Tốc độ gió
  - Công suất phát điện tại các thời điểm trước đó
- Đầu ra cần dự báo:
  - Công suất phát điện tại thời điểm tương lai

Dữ liệu đầu ra là Công suất phát điện dự báo tại thời điểm trong tương lai.

Có thể mô tả bài toán dự báo công suất phát điện qua công thức toán học như bên dưới:

$$Y_{t+h} = f(x_t, x_{t-1}, x_{t-2}, \dots, x_{t-w}) \quad (2.7)$$

Trong đó:

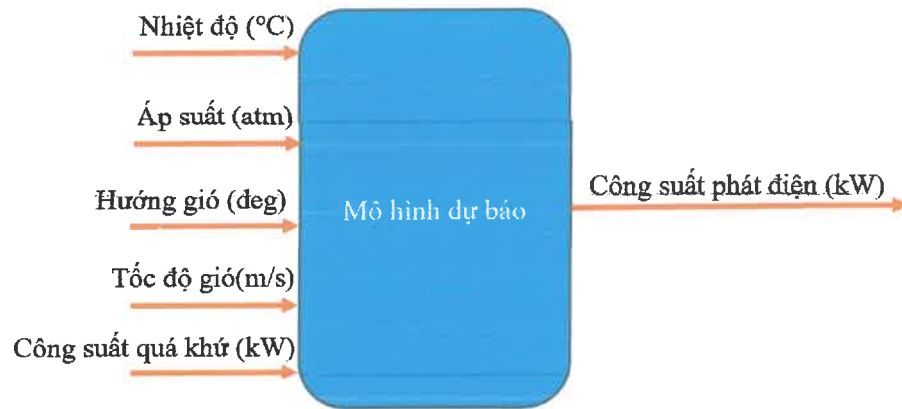
$Y_{t+h}$ : Công suất cần dự báo tại thời điểm  $t+h$

$x_t$ : Các đặc trưng khí tượng và công suất phát điện thực tế tại thời điểm  $t$

$w$ : Số bước lookback quá khứ

$f()$ : Hàm thể hiện mô hình dự báo

Mình họa bài toán như hình vẽ:



Hình 2.2 Mô hình hóa đầu vào - đầu ra trong bài toán dự báo công suất phát điện

### 2.2.2 Cấu trúc mạng LSTM được đề xuất

Trong nghiên cứu này, mô hình LSTM được thiết kế để dự báo công suất phát điện tại thời điểm  $t+h$  dựa trên chuỗi dữ liệu thời tiết và công suất ở các thời điểm trước đó. Trong đó,  $h$  là số giờ dự báo tiếp theo.

Kiến trúc mạng LSTM được thiết kế dựa trên đặc điểm dữ liệu chuỗi thời gian có quan hệ phụ thuộc dài hạn và phi tuyến - đặc trưng điển hình của dữ liệu công suất phát điện gió [21].

Cấu trúc tổng thể của mô hình bao gồm các thành phần chính như sau:

- Đầu vào: là chuỗi thời gian của các đặc trưng khí tượng và kỹ thuật có ảnh hưởng đến công suất phát điện, bao gồm: tốc độ gió, hướng gió, áp suất không khí, nhiệt độ, và công suất phát trong quá khứ. Các đặc trưng này được chuẩn hóa và chia thành các chuỗi con có độ dài cố định theo phương pháp cửa sổ trượt (sliding window).

Mỗi chuỗi đầu vào là một chuỗi có độ dài cố định  $w$  (window side), với  $d$  đặc trưng tại mỗi bước thời gian, tạo thành một tensor ba chiều có kích thước ( $samples, w, d$ ). Việc sử dụng nhiều đặc trưng đầu vào giúp mô hình khai thác tốt hơn các mối tương quan khí tượng và tăng độ chính xác của dự báo [31].

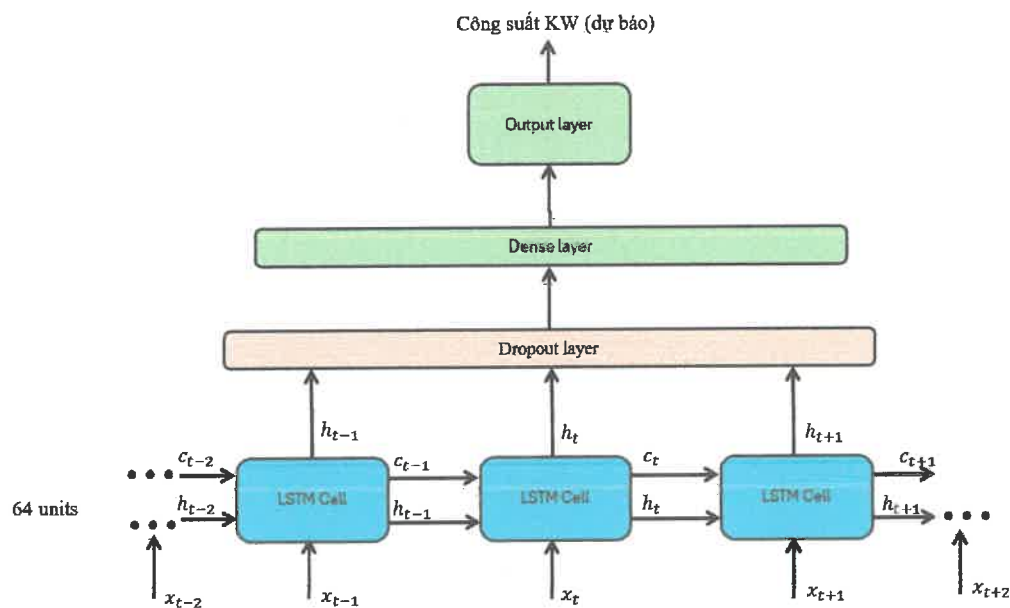
Lớp LSTM: Một hoặc nhiều lớp LSTM được sử dụng để trích xuất đặc trưng tuần tự từ chuỗi dữ liệu. Các lớp này giúp mô hình học được cả quan hệ ngắn hạn và dài hạn, nhờ vào cơ chế cổng ghi nhớ (memory gates) đặc trưng của LSTM[10]. Số

lượng nơ-ron (units) trong mỗi lớp là một siêu tham số. Để án thử nghiệm mô hình LSTM 1 lớp ẩn với 64 đơn vị và mô hình LSTM 2 lớp ẩn, với lớp 1 có 128 đơn vị, lớp 2 có 64 đơn vị, sau thử nghiệm sẽ đánh giá tính chính lựa chọn mô hình phù hợp. Biểu đồ 2 kiến trúc mô hình được trình bày ở Hình 2.3 và Hình 2.4 bên dưới.

- Lớp Dropout: được chèn giữa các lớp LSTM và ở cuối trước lớp Dense đầu ra nhằm giảm thiểu hiện tượng overfitting. Dropout hoạt động bằng cách vô hiệu hóa ngẫu nhiên một phần các nơ-ron trong quá trình huấn luyện, giúp tăng tính mô hình[36, 54].

- Lớp Dense đầu ra: trường hợp dự báo 1 bước kế tiếp, lớp Dense đầu ra là một lớp kết nối đầy đủ với một nơ-ron duy nhất, thực hiện nhiệm vụ dự báo công suất phát điện tại bước thời gian tiếp theo. Với mục đích nghiên cứu dự báo ngắn hạn, và nhằm giảm phụ thuộc giữa các bước, để án thử nghiệm dự báo nhiều bước trong 1 lần, lớp Dense bao gồm nhiều nơ-ron đầu ra – tương ứng với số bước cần dự báo.

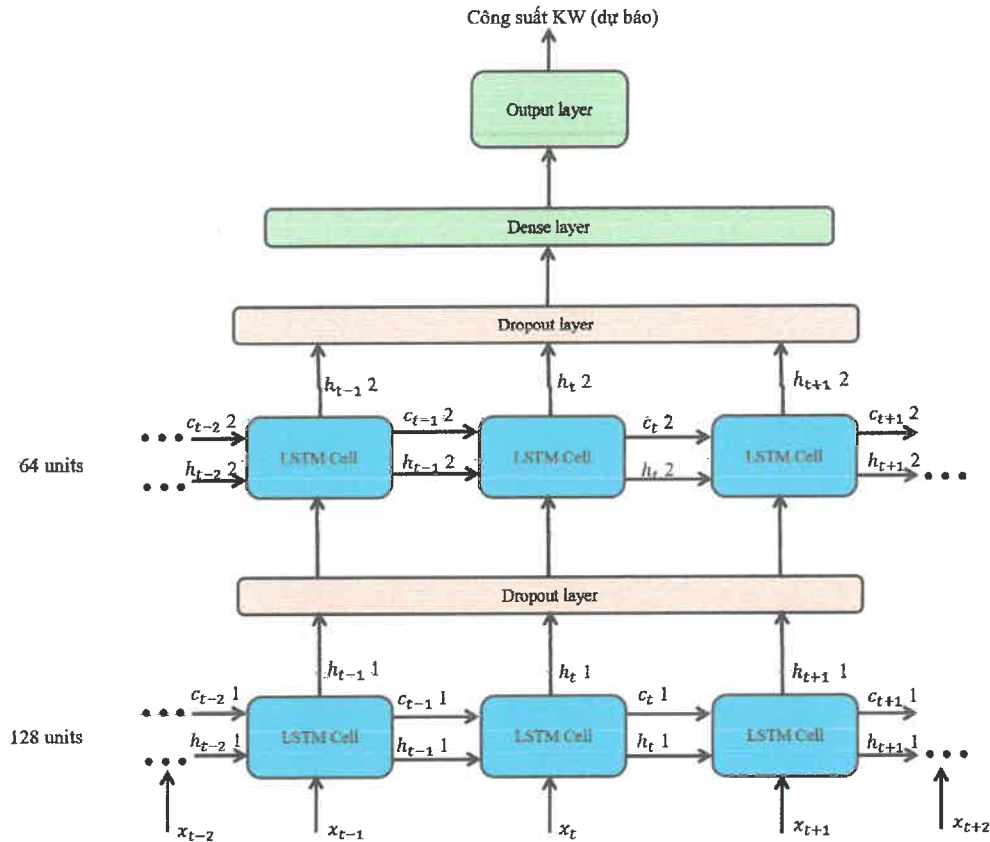
Kiến trúc mô hình LSTM 1 lớp ẩn, với số lượng nơ-ron của lớp là 64 đơn vị (units). 1 lớp Dropout để tránh overfitting trước khi output ra kết quả. 1 lớp Dense đầu ra.



Trong đó:  $x_t = [\text{Nhiệt độ}, \text{Áp suất}, \text{Hướng gió}, \text{Tốc độ gió}, \text{Công suất}]$  tại thời điểm  $t$

**Hình 2.3 Kiến trúc mô hình LSTM 1 lớp ẩn**

Kiến trúc mô hình LSTM 2 lớp ẩn, với số lượng nơ-ron của lớp 1 là 128 đơn vị (units). Lớp Dropout xen giữa 2 lớp ẩn để tránh Overfitting. Lớp ẩn thứ 2 với 64 đơn vị. 1 lớp Dropout để tránh overfitting trước khi output ra kết quả. 1 lớp Dense đầu ra.



Trong đó:  $x_t = [\text{Nhiệt độ, Áp suất, Hướng gió, Tốc độ gió, Công suất}]$  tại thời điểm  $t$

**Hình 2.4 Kiến trúc mô hình LSTM 2 lớp ẩn**

Đề án tập trung nghiên cứu xây dựng mô hình dự báo đa bước trực tiếp, với đầu ra là  $h$  bước tiếp theo được dự báo trong một lần. Với khoảng dự báo mức ngắn hạn đến vài giờ, dự báo đa bước trực tiếp có thể tránh được các sai số lan truyền qua từng bước so với dùng đa bước đệ quy. Ngoài ra, với việc bộ dữ liệu có đầy đủ các đặc trưng đầu vào và có thể tách các đặc trưng để đưa vào mô hình, dự báo đa bước trực tiếp có thể học được xu hướng giữa các bước.

### 2.2.3 Quy trình huấn luyện mô hình

#### a. Tiền xử lý dữ liệu:

Trong các bài toán học sâu trên chuỗi thời gian, đặc biệt với dữ liệu thực tế từ nhà máy điện gió, tiền xử lý dữ liệu là bước bắt buộc nhằm đảm bảo độ sạch, tính nhất quán và khả năng huấn luyện ổn định của mô hình. Dữ liệu thực tế thường gặp vấn đề về thiếu giá trị, nhiễu, và sai lệch thang đo giữa các đặc trưng.

- Xử lý dữ liệu thiếu và nhiễu:

Dữ liệu đầu vào được kiểm tra và loại bỏ các bản ghi có giá trị bị thiếu (missing values) hoặc không hợp lệ.

- Chuẩn hóa dữ liệu:

Dữ liệu sau khi được làm sạch sẽ được chuẩn hóa để đưa các đặc trưng về cùng thang đo, giúp tăng hiệu quả học của mô hình. Trong nghiên cứu này, phương pháp Min-Max Scaling được sử dụng để chuẩn hóa dữ liệu về khoảng  $[0, 1]$ , theo công thức:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2.8)$$

Việc chuẩn hóa là cần thiết khi sử dụng mạng LSTM, vốn thường hoạt động với các hàm kích hoạt như sigmoid hoặc tanh, có hiệu suất tốt hơn khi đầu vào nằm trong một khoảng nhỏ [40]. Đặc biệt, dữ liệu được chuẩn hóa riêng biệt trên tập huấn luyện để tránh hiện tượng rò rỉ dữ liệu (data leakage) [41].

Ngoài ra, chuẩn hóa theo Min-Max giúp đảm bảo mô hình học được mối quan hệ tuyến tính hóa và dễ dàng đảo ngược sau khi dự báo, từ đó thuận tiện cho việc đánh giá bằng các chỉ số thực tế như MAE, RMSE [43].

- Tạo chuỗi đầu vào theo cửa sổ trượt:

Dữ liệu chuỗi thời gian được chuyển đổi thành các mẫu huấn luyện bằng cách áp dụng kỹ thuật cửa sổ trượt (sliding window) - phương pháp phổ biến để chuyển đổi chuỗi đơn chiều thành tập các chuỗi con liên tục có độ dài cố định. Kỹ thuật này giúp mô hình học được mối quan hệ theo thời gian giữa các quan sát liên tiếp [35][42].

Trong bài toán dự báo công suất phát điện gió, độ dài cửa sổ trượt thường được chọn là 24, 48, 72..bước thời gian (tương ứng với các bước lấy dữ liệu các khoảng 24h, 48h, 72h..trong quá khứ). Mục tiêu là giúp mô hình nhận diện được các chu kỳ

ngắn hạn và biến động theo ngày của công suất phát điện - một đặc điểm đặc trưng trong dữ liệu tua-bin gió.

*b. Chia tập dữ liệu:*

Với tập dữ liệu quy mô ở mức trung bình, có các khuyến nghị chia tập dữ liệu theo tỷ lệ 80 - 20, tương ứng với 80% dữ liệu dùng để huấn luyện, 20% dữ liệu dùng để đánh giá.

Nhiều nghiên cứu trong lĩnh vực dự báo năng lượng và học sâu đã sử dụng các tỷ lệ chia tập này để so sánh kết quả giữa các mô hình [37, 41].

Trong dự báo chuỗi thời gian, việc chia tập dữ liệu không được thực hiện ngẫu nhiên, mà phải chia theo trục thời gian để giữ nguyên cấu trúc tuần tự. Tức là, toàn bộ các mẫu ở phần đầu chuỗi thời gian được dùng để huấn luyện, trong khi phần cuối chuỗi (dữ liệu tương lai) được dùng để đánh giá mô hình. Điều này giúp mô hình phản ánh đúng bối cảnh ứng dụng thực tế, nơi dự báo luôn dựa trên dữ liệu lịch sử.

Các framework như Keras, Scikit-learn cũng mặc định hoặc gợi ý các tỷ lệ này như một quy chuẩn để thực hành tốt.

*c. Xây dựng mô hình:*

Trong nghiên cứu này, mô hình được xây dựng dựa trên kiến trúc mạng Long Short-Term Memory (LSTM) - một biến thể của mạng nơ-ron hồi tiếp (RNN), được thiết kế để xử lý và học các mối quan hệ phụ thuộc theo thời gian dài hạn nhờ vào cơ chế cổng ghi nhớ (memory gates)[10].

Để án thử nghiệm triển khai mô hình LSTM với các tham số khác nhau để đưa ra các kết quả đánh giá. Nghiên cứu của Bolboacă và cộng sự[56] chỉ ra rằng, với tập dữ liệu trung bình (~52,560 dòng) và số lượng đặc trưng không nhiều, việc lựa chọn số lớp LSTM trong kiến trúc là 1 lớp hoặc 2 lớp với số nơ-ron tương ứng là 128, 64 sẽ ảnh hưởng không đáng kể đến kết quả dự báo. Vì vậy, đề án đề xuất sử dụng mô hình 2 lớp LSTM, lớp 1 có 128 units, lớp 2 có 64 units. Với kiến trúc này, mô hình có thể học được sâu hơn, trong khi có thể sử dụng các tham số khác để tránh việc overfitting.

Lớp Dropout: Giữa hai lớp LSTM hoặc trước khi đẩy ra lớp Dense, một lớp Dropout với tỷ lệ dropout = 0.2 được chèn vào để tránh hiện tượng overfitting[48].

Kỹ thuật này giúp mô hình trở nên ổn định hơn, bằng cách loại bỏ ngẫu nhiên một tỷ lệ nhất định các nơ-ron trong quá trình huấn luyện.

Lớp đầu ra (Dense layer): Một lớp Dense đầu ra, dùng hàm kích hoạt tuyến tính để tổng hợp đầu ra là công suất phát điện cần dự báo:

$$y = W \cdot x + b \quad (2.9)$$

Việc sử dụng hàm tuyến tính là phù hợp trong các bài toán hồi quy, nơi đầu ra là một đại lượng thực liên tục và không bị giới hạn về miền giá trị [40].

Mô hình được hiện thực bằng thư viện Keras trên nền TensorFlow, cho phép tùy chỉnh linh hoạt và mở rộng với các thành phần như Dropout hoặc thêm nhiều đầu ra nếu bài toán là dự báo đa bước.

#### *d. Huấn luyện mô hình:*

Giai đoạn huấn luyện là bước quan trọng để tối ưu hóa các tham số của mạng LSTM nhằm giảm sai số dự báo giữa giá trị thực tế và giá trị mô hình ước lượng. Trong đề án này, quá trình huấn luyện mô hình được thiết lập theo các cấu hình phổ biến trong lĩnh vực học sâu ứng dụng cho chuỗi thời gian.

- Hàm mất mát: MSE (Mean Squared Error)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y - \hat{y}_i)^2 \quad (2.10)$$

MSE có đặc tính làm nổi bật sai số lớn, nhờ đó hỗ trợ việc học chính xác hơn trong các bước dự báo sai lệch nghiêm trọng. Đây cũng là hàm mất mát mặc định trong nhiều framework học sâu như Keras khi xử lý bài toán hồi quy [45].

- Thuật toán tối ưu: Adam Optimizer

Để tối ưu hóa hàm mất mát, thuật toán Adam (Adaptive Moment Estimation) được sử dụng. Adam kết hợp ưu điểm của hai thuật toán SGD có momentum và RMSprop, cho phép điều chỉnh tốc độ học (learning rate) một cách tự động cho từng tham số dựa trên ước lượng trung bình động của gradient và bình phương gradient [46]. Adam đặc biệt hiệu quả và ổn định với dữ liệu phi tuyến và nhiễu như chuỗi công suất phát điện gió.

- Số epoch:

Mô hình thử nghiệm với epochs thay đổi để tìm ra giá trị cân bằng giúp mô hình học đủ sâu để giảm sai số mà vẫn tránh được tình trạng quá khớp (overfitting). Việc xác định số epoch được thực hiện dựa trên thực nghiệm, với việc theo dõi giá trị mất mát qua từng epoch.

- Kích thước lô huấn luyện (Batch size): 64

Batch size được đặt là 64, giúp tối ưu giữa hiệu suất tính toán và khả năng hội tụ. Theo nghiên cứu của Masters và Luschi (2018), batch size trung bình (như 32 hoặc 64) thường cho kết quả tốt hơn batch nhỏ hoặc rất lớn trong các bài toán chuỗi thời gian [47]. Ở mô hình bài toán dự báo công suất phát, với tập dữ liệu đầu vào ở mức trung bình, số lượng đặc trưng và dữ liệu không quá phức tạp, batch size được chọn ở mức 64.

Việc lựa chọn cấu hình huấn luyện nhằm tìm ra mô hình học được hiệu quả, tránh overfitting và ổn định trên tập kiểm thử.

#### *e. Đánh giá và kiểm thử:*

Sau khi hoàn tất quá trình huấn luyện, mô hình được đánh giá trên tập kiểm thử nhằm đo lường hiệu suất dự báo và khả năng khái quát hóa. Trong nghiên cứu này, hiệu năng mô hình được đánh giá bằng các chỉ số phổ biến trong hồi quy và dự báo chuỗi thời gian, bao gồm MAE, RMSE, MAPE. Các chỉ số này đã được trình bày chi tiết trong Chương 1, Mục IV.2 – Các tiêu chí đánh giá mô hình dự báo. Các chỉ số này cung cấp một cái nhìn nhiều chiều về độ chính xác, mức độ sai lệch và khả năng giải thích của mô hình.

Các chỉ số đánh giá mô hình:

- MAE (Mean Absolute Error):
- RMSE (Root Mean Square Error):
- MAPE (Mean Absolute Percentage Error)

Mỗi chỉ số phản ánh một khía cạnh khác nhau của độ chính xác, và có thể được sử dụng kết hợp để có cái nhìn toàn diện hơn[21].

Ngoài các chỉ số định lượng, mô hình còn được đánh giá thông qua trực quan hóa kết quả dự báo so với dữ liệu thực tế. Các biểu đồ như:

- Đường biểu diễn công suất dự báo với thực tế

- Biểu đồ sai số tại từng bước dự báo
- Đồ thị phân bố phần trăm lỗi (MAPE theo bước)

giúp minh họa rõ hơn hiệu năng và xu hướng sai số của mô hình theo thời gian, từ đó hỗ trợ phân tích và điều chỉnh mô hình hiệu quả hơn.

#### **2.2.4 Đánh giá ưu điểm mô hình LSTM**

**Khả năng học dài hạn:** Với cơ chế cổng (gates) và trạng thái bộ nhớ (cell state), LSTM có thể học và lưu giữ các mẫu biến động theo thời gian dài — điều mà các mô hình RNN truyền thống hay ARIMA khó đạt được [10].

**Khả năng khử nhiễu:** Khi kết hợp với các kỹ thuật tiền xử lý như chuẩn hóa và chọn đặc trưng, LSTM vẫn giữ được độ chính xác cao trong điều kiện dữ liệu thực có nhiễu hoặc mất cân bằng tần suất [41]. Điều này đặc biệt phù hợp với dữ liệu thời tiết, vốn có sự biến động theo mùa và ngẫu nhiên.

**Mở rộng linh hoạt:** Kiến trúc LSTM cho phép dễ dàng mở rộng về cả không gian đặc trưng đầu vào (bổ sung thêm các biến khí tượng như độ ẩm, bức xạ mặt trời, nhiệt độ đất, v.v.) và chiều thời gian đầu ra. Trong đề án này, mô hình được huấn luyện theo cấu trúc dự báo đa bước (multi-step) để dự đoán công suất phát điện cho nhiều bước thời gian tiếp theo (từ  $t+1$  đến  $t+h$ ). Điều này chứng minh khả năng linh hoạt và thích nghi của LSTM đối với các yêu cầu thực tiễn trong dự báo năng lượng gió ngắn hạn, trung hạn [42].

### **2.3 Kết luận chương**

Trong chương này, luận văn đã trình bày phương pháp đề xuất để dự báo công suất phát điện gió dựa trên mạng nơ-ron hồi tiếp LSTM. Phương pháp này tận dụng khả năng ghi nhớ dài hạn và mô hình hóa chuỗi thời gian của LSTM, hứa hẹn mang lại kết quả chính xác và ổn định hơn so với các phương pháp truyền thống. Kiến trúc mô hình, đầu vào, chiến lược huấn luyện và các yếu tố kỹ thuật cũng đã được mô tả chi tiết. Các bước triển khai mô hình sẽ được tiến hành và đánh giá thực nghiệm trong Chương 3 tiếp theo.

## CHƯƠNG 3: THỬ NGHIỆM VÀ ĐÁNH GIÁ

### 3.1 Thu thập và tiền xử lý dữ liệu

#### 3.1.1 Nguồn dữ liệu

Dữ liệu sử dụng trong nghiên cứu được thu thập từ dữ liệu được thu thập từ dữ liệu từ NREL (Mỹ)[17]. Tuy nhiên, tại thời điểm nghiên cứu, NREL chỉ còn cung cấp dữ liệu khí tượng được đo đạc trong quá khứ còn dữ liệu công suất phát thực tế đã được loại bỏ (deprecated), thay vào đó NREL khuyến nghị nhà nghiên cứu chuyển đổi công suất phát dựa trên thông số của tuabin. Vì vậy, đề án đã tham khảo nguồn dữ liệu của NREL thông qua nghiên cứu của Shashwat Arghode ở đề tài CSE-523, “*Time series Prediction for Power Generation from Wind*”[14]. Dữ liệu bao gồm:

- Thông số thời gian: Thời gian
- Thông số khí tượng: nhiệt độ (°C), áp suất (atm), tốc độ gió (m/s), hướng gió (deg)
- Thông số vận hành: công suất phát thực tế (kW) theo từng thời điểm.

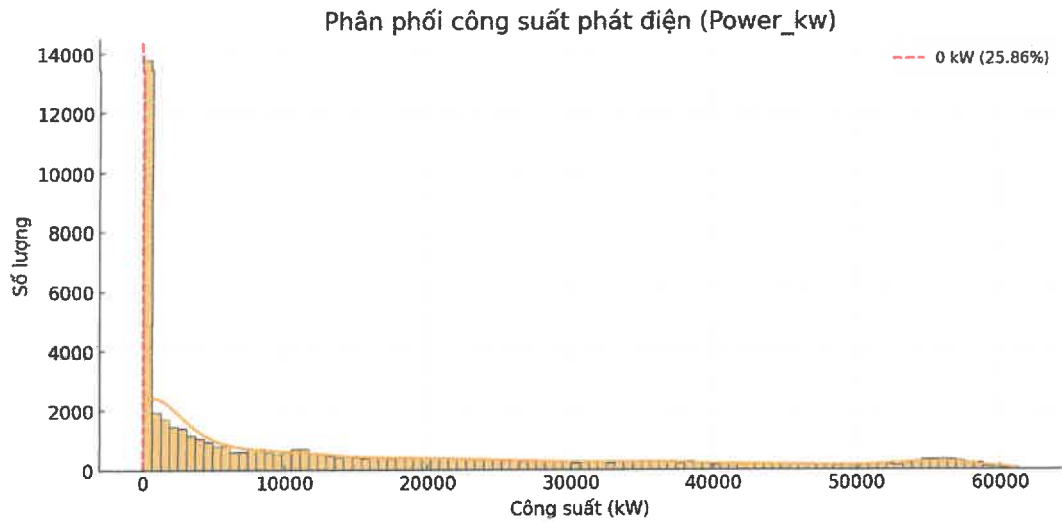
Tần suất lấy mẫu là theo giờ, với tổng số bản ghi khoảng 52.560 dòng dữ liệu.

Dưới đây là một số thống kê liên quan tập dữ liệu.

**Bảng 3.1 Dữ liệu sử dụng trong nghiên cứu**

| DateTime            | Air temperature   (°C) | Pressure   (atm) | Wind speed   (m/s) | Wind direction   (deg) | Power generated by system   (kW) |
|---------------------|------------------------|------------------|--------------------|------------------------|----------------------------------|
| 2010-04-02 08:00:00 | 9.954                  | 0.979708         | 8.227              | 188                    | 26038.4                          |
| 2010-04-02 09:00:00 | 9.646                  | 0.97986          | 8.449              | 186                    | 28059.7                          |
| 2010-04-02 10:00:00 | 9.045                  | 0.980181         | 8.02               | 184                    | 24121.9                          |
| 2010-04-02 11:00:00 | 8.726                  | 0.980244         | 7.77               | 183                    | 21935.4                          |
| 2010-04-02 12:00:00 | 11.296                 | 0.980513         | 6.808              | 180                    | 14067.7                          |
| 2010-04-02 13:00:00 | 14.104                 | 0.980804         | 4.228              | 162                    | 1857.95                          |
| 2010-04-02 14:00:00 | 17.038                 | 0.980873         | 3.807              | 155                    | 903.492                          |
| 2010-04-02 15:00:00 | 19.638                 | 0.980888         | 4.112              | 166                    | 1565.82                          |

Biểu đồ phân phối Công suất phát điện:

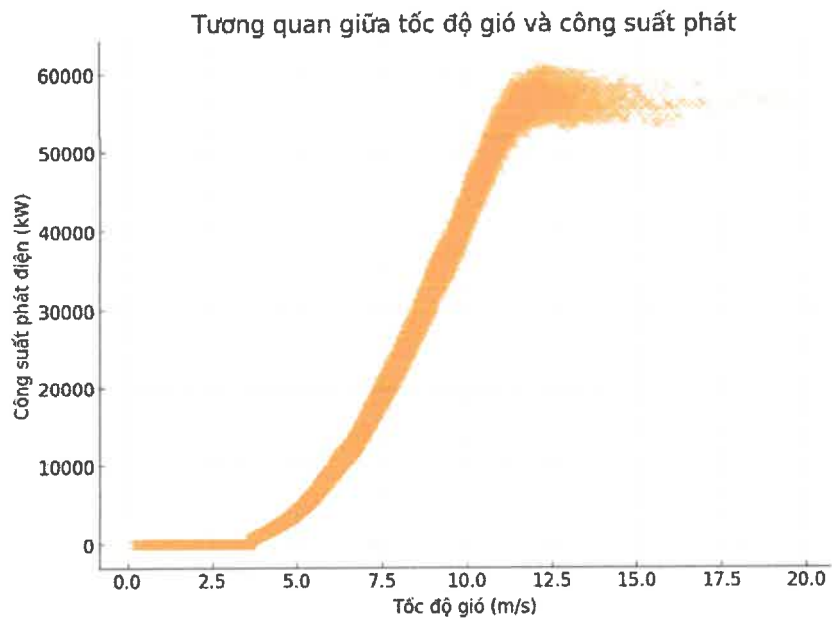


**Hình 3.1 Biểu đồ phân phối Công suất phát điện**

Dựa vào biểu đồ Phân phối công suất có thể đánh giá:

- Dải phân phối phát điện rộng, và lệch phải, với đỉnh ở gần mức công suất phát = 0 kW.
- Khoảng ~25% số liệu công suất phát = 0 kW, phản ánh thời điểm dừng phát điện do gió yếu hoặc bảo trì, bảo dưỡng thiết bị.

Biểu đồ tương quan Tốc độ gió (m/s) – Công suất (kW)

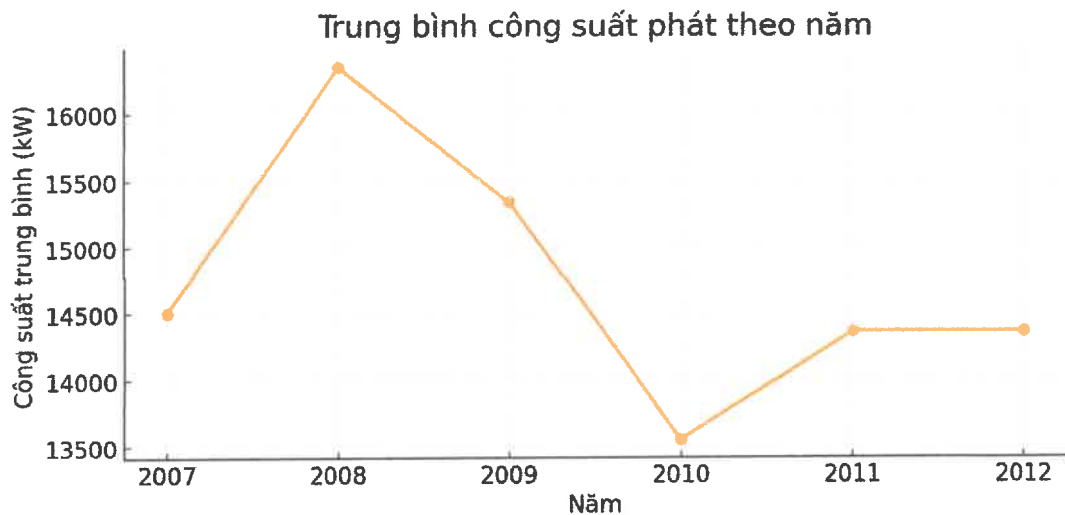


**Hình 3.2 Biểu đồ tương quan tốc độ gió – Công suất**

Dựa vào biểu đồ Tương quan tốc độ gió – Công suất phát có thể thấy:

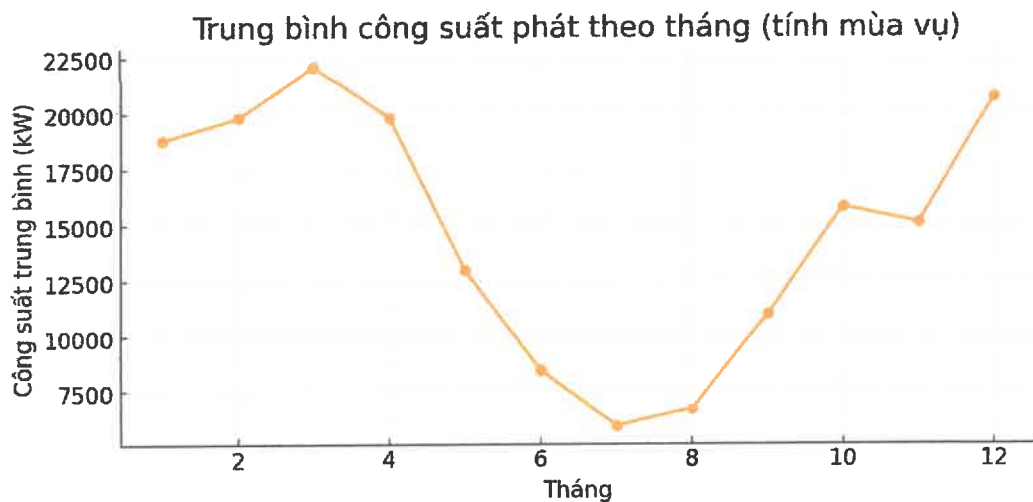
- Ở mức tốc độ gió thấp ( $< 3\text{m/s}$ ) thì công suất phát  $\sim 0\text{ kW}$ .
- Ở mức tốc độ gió trung bình ( $4\text{-}10\text{m/s}$ ): công suất tăng theo đồ thị parabol.
- Ở mức tốc độ gió cao ( $> 12\text{m/s}$ ): Công suất phát điện ổn định, có thể công suất phát đạt định mức của tuabin.

Thống kê trung bình công suất phát điện theo năm:



**Hình 3.3 Biểu đồ Công suất trung bình theo năm**

Thống kê trung bình công suất phát điện theo mùa vụ (tháng):



**Hình 3.4 Biểu đồ Công suất trung bình theo tháng**

Thống kê Công suất phát trung bình theo năm và tháng có thể thấy:

- Công suất trung bình mỗi năm có dao động nhưng không có xu hướng tăng/giảm đột ngột.

- Có dấu hiệu mùa vụ rõ rệt: các tháng 1–3 và 10–12 thường có công suất trung bình cao hơn, có thể do gió mùa Đông mạnh hơn.

- Các tháng giữa năm (tháng 5–9) có xu hướng giảm.

Đánh giá chung về bộ dữ liệu:

- Dữ liệu có biến động theo năm, theo mùa vụ và theo thời gian trong ngày.

Điều này thể hiện đặc điểm phi tuyến của công suất phát điện.

- Dữ liệu công suất phát tại nhiều điểm có sự tăng/giảm đột ngột, thể hiện sự không ổn định liên quan yếu tố khí tượng.

- Bộ dữ liệu có nhiều dòng dữ liệu công suất phát điện  $\sim 0$  kW, cho thấy có nhiều thời điểm tuabin không hoạt động có thể do nguyên nhân tốc độ gió thấp hoặc thiết bị đang trong quá trình bảo hành bảo dưỡng. Dữ liệu rõ ràng có tính thực tiễn cao, phù hợp với nghiên cứu, thực nghiệm.

### **3.1.2 Tiền xử lý dữ liệu**

Tập dữ liệu gồm 52.560 bản ghi, không bao gồm các bản ghi bị trống (missing values). Việc đảm bảo dữ liệu đầy đủ giúp tránh phát sinh lỗi trong quá trình huấn luyện và đánh giá mô hình.

Để đảm bảo chất lượng dữ liệu đầu vào cho mô hình, các bước tiền xử lý được thực hiện gồm:

- Xử lý dữ liệu thiếu (missing values): không có dòng trống trong dữ liệu.

- Loại bỏ dữ liệu có thể gây nhiễu: các dòng có công suất phát điện bằng 0 - chiếm khoảng 25% dữ liệu - có thể gây nhiễu cho mô hình dự báo công suất trong điều kiện hoạt động bình thường.

- Chuẩn hóa dữ liệu về khoảng  $[0,1]$  bằng kỹ thuật Min-Max Scaling để tăng hiệu quả huấn luyện mạng LSTM.

- Chuyển đổi chuỗi thời gian thành dạng mẫu học bằng kỹ thuật sliding window.

## 3.2 Thử nghiệm và đánh giá kết quả

### 3.2.1 Cấu hình thử nghiệm

Chương trình sử dụng thư viện mã nguồn mở Keras[32] - đây là 1 API chạy trên nền tảng TensorFlow[33]. Keras hỗ trợ mạnh mẽ các mạng nơ-ron như: RNN, LSTM, GRU. Chương trình được viết bằng ngôn ngữ lập trình Python, trên nền tảng thư viện Google Colaboratory[34] và được format dưới định dạng jupyter notebook(.ipynb), cho phép xử lý trực quan, hiển thị kết quả nhanh chóng. Phương thức lập trình đơn giản, cấu trúc rõ ràng. Chương trình chạy trên hạ tầng của Google Colab, cấu hình T4 GPU.

Các thực nghiệm sẽ tiến hành với các thông số cấu hình khác nhau của mô hình để tìm ra tham số tốt nhất cho mô hình. Mục đích thử nghiệm các mô hình khác nhau để huấn luyện mạng nơ-ron, so sánh giá trị MAPE của các mô hình và từ đó chọn ra mô hình tốt nhất để áp dụng để dự báo công suất phát điện gió.

Trong bài toán dự báo công suất phát điện gió kích thước cửa sổ trượt – window size – hay là số bước thời gian(quá khứ) mà mô hình sẽ đưa vào đầu vào để dự báo đầu ra có ý nghĩa quan trọng, ảnh hưởng đến mức độ học và khả năng dự báo xu hướng của chuỗi thời gian. Thử nghiệm các kích thước của sổ trượt – lookback khác nhau để tìm ra kích thước cửa sổ trượt phù hợp với mô hình.

**Bảng 3.2 Các tham số mô hình theo kích thước cửa sổ trượt**

| <b>Tham số</b>            | <b>Giá trị</b>                      |
|---------------------------|-------------------------------------|
| Số bước thời gian đầu vào | <b>24, 48, 72, 96</b>               |
| Số lớp LSTM               | 2                                   |
| Số đơn vị LSTM            | Lớp 1: 128 units<br>Lớp 2: 64 units |
| Batch size                | 64                                  |
| Số epoch huấn luyện       | 10                                  |
| Hàm tối ưu                | Adam                                |
| Số giờ dự báo tiếp theo   | 4h                                  |

Xét trên số lượng lần lặp trên tập dữ liệu huấn luyện (epoch), đề án thử nghiệm số lần lặp khác nhau để tìm ra mức lặp phù hợp với huấn luyện mô hình. Epoch là một trong những tham số quan trọng của mô hình dự báo, đặc biệt với LSTM, RNN, CNN, nó ảnh hưởng đến việc học của mô hình. Nếu epoch quá nhỏ, có thể mô hình chưa học đủ vì duyệt qua dữ liệu quá ít. Nếu epoch quá lớn, mô hình có thể học quá chi tiết, đánh mất đi khả năng tổng quát hóa. Epoch ở mức hợp lý sẽ đảm bảo mô hình học vừa đủ và ổn định.

Đề án thử nghiệm và đánh giá thông qua chỉ số MAPE để lựa chọn số epoch phù hợp với mô hình.

**Bảng 3.3: Các tham số mô hình theo số lần lặp**

| Tham số                   | Giá trị                             |
|---------------------------|-------------------------------------|
| Số bước thời gian đầu vào | 48                                  |
| Số lớp LSTM               | 2                                   |
| Số đơn vị LSTM            | Lớp 1: 128 units<br>Lớp 2: 64 units |
| Batch size                | 64                                  |
| Số epoch huấn luyện       | <b>10, 20, 30, 40, 50</b>           |
| Hàm tối ưu                | Adam                                |
| Số giờ dự báo tiếp theo   | 4h                                  |

### 3.2.2 Kết quả thử nghiệm

Sau khi huấn luyện, mô hình LSTM đạt kết quả dự báo (4h tiếp theo) trên tập kiểm thử như bảng bên dưới. Chỉ số MAPE được dùng để đánh giá tính chính xác của mô hình.

**Bảng 3.4: Đánh giá dự báo theo kích thước cửa sổ trượt**

| STT | Kích thước cửa sổ trượt | MAPE% trung bình |
|-----|-------------------------|------------------|
| 1   | 24                      | 131.44%          |
| 2   | <b>48</b>               | <b>4.38%</b>     |
| 3   | 72                      | 6.25%            |
| 4   | 96                      | 19.37%           |

Có thể thấy, với kích thước cửa sổ trượt thấp (24h), không đủ thông tin để dự đoán xu hướng, đặc biệt trong chuỗi nhiều dao động. Kích thước cửa sổ trượt 48h cho thấy tính chính xác cao nhất, vừa đủ để nắm được xu hướng của chuỗi, nhưng không quá dài để gây nhiễu (như 72h hay 96h). Với kích thước cửa sổ trượt 48h có thể tương ứng với ~ 2 ngày dữ liệu trong quá khứ, có thể phản ánh chu kỳ gió ngắn hạn. Từ thực nghiệm, đề án nhận thấy kích thước cửa sổ trượt (window size) 48h là điểm cân bằng tối ưu cho mô hình dự báo.

**Bảng 3.5: Đánh giá dự báo theo số lần lặp**

| STT | Epochs | MAPE% trung bình |
|-----|--------|------------------|
| 1   | 10     | 4.38%            |
| 2   | 20     | <b>4.06%</b>     |
| 3   | 30     | 15.88%           |
| 4   | 40     | 18.89%           |
| 5   | 50     | 20.87%           |

Bảng kết quả thử nghiệm với số epoch khác nhau cho thấy số epoch ở mức 20 là phù hợp nhất, với MAPE trung bình 4.06%. Trong khi mô hình có xu hướng giảm loss trên tập huấn luyện, khả năng tổng quát hóa bị giảm dần khi tăng số epoch.

Từ 2 thử nghiệm trên, đề án nhận định sử dụng kích thước cửa sổ trượt (window size) 48h và số lần lặp trên tập huấn luyện (epoch) 20 lần sẽ phù hợp với mô hình bài toán dự báo đang xây dựng.

**Bảng 3.6 Các tham số chung của mô hình thử nghiệm**

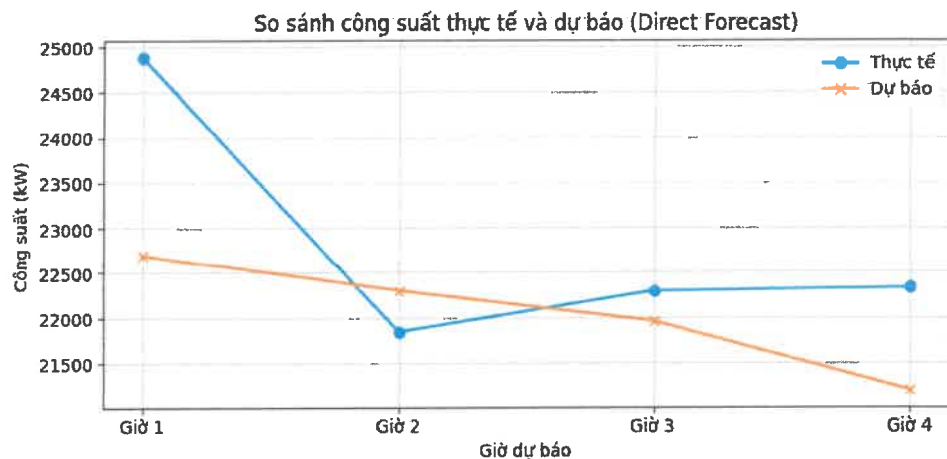
| Tham số                   | Giá trị                             |
|---------------------------|-------------------------------------|
| Số bước thời gian đầu vào | 48                                  |
| Số lớp LSTM               | 2                                   |
| Số đơn vị LSTM            | Lớp 1: 128 units<br>Lớp 2: 64 units |
| Batch size                | 64                                  |
| Số epoch huấn luyện       | 20                                  |
| Hàm tối ưu                | Adam                                |

Dưới đây là một số kết quả thử nghiệm dựa trên các tham số của mô hình dựa trên bảng tham số mô hình chung.

Với số bước là 4, tương đương với dự báo 4h tiếp theo, kết quả đánh giá mô hình thu được như bên dưới:

**Bảng 3.7 Kết quả dự báo 4h tiếp theo**

| Giờ   | Thực tế (kW) | Dự báo (kW) | MAE      | MAPE (%) |
|-------|--------------|-------------|----------|----------|
| Giờ 1 | 24,888.20    | 23,327.10   | 1,561.10 | 6.27%    |
| Giờ 2 | 21,839.80    | 22,607.21   | 767.41   | 3.51%    |
| Giờ 3 | 22,291.30    | 21,855.58   | 435.72   | 1.95%    |
| Giờ 4 | 22,325.40    | 21,317.34   | 1,008.06 | 4.52%    |

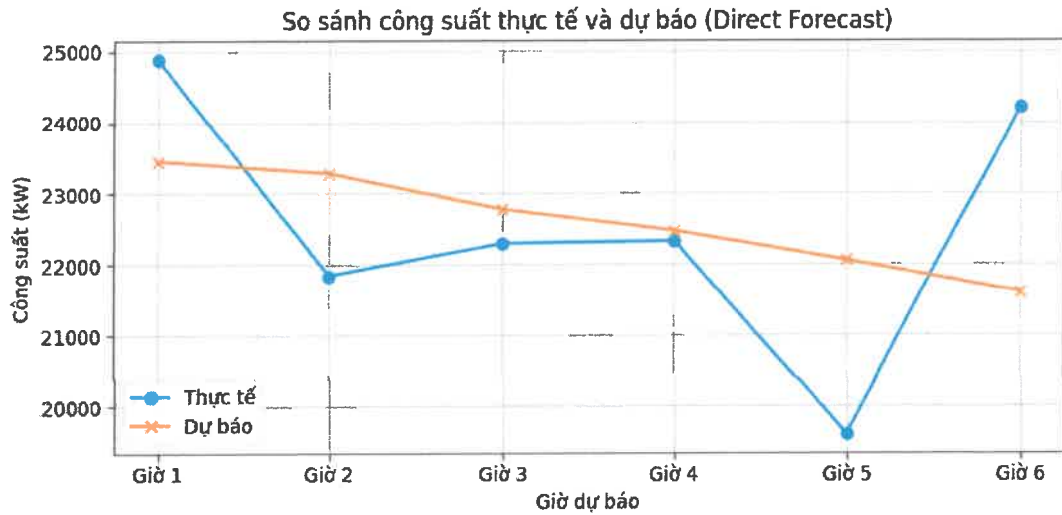


**Hình 3.5 Kết quả dự báo 4h tiếp theo**

Với số bước là 6, tương đương với dự báo 6h tiếp theo, kết quả đánh giá mô hình thu được như bên dưới:

**Bảng 3.8 Kết quả dự báo 6h tiếp theo**

| Giờ   | Thực tế (kW) | Dự báo (kW) | MAE      | MAPE (%) |
|-------|--------------|-------------|----------|----------|
| Giờ 1 | 24,888.20    | 23,206.10   | 1,682.10 | 6.76%    |
| Giờ 2 | 21,839.80    | 22,263.70   | 423.9    | 1.94%    |
| Giờ 3 | 22,291.30    | 21,413.36   | 877.94   | 3.94%    |
| Giờ 4 | 22,325.40    | 20,476.85   | 1,848.55 | 8.28%    |
| Giờ 5 | 19,605.80    | 20,263.07   | 657.27   | 3.35%    |
| Giờ 6 | 24,217.60    | 19,996.53   | 4,221.07 | 17.43%   |

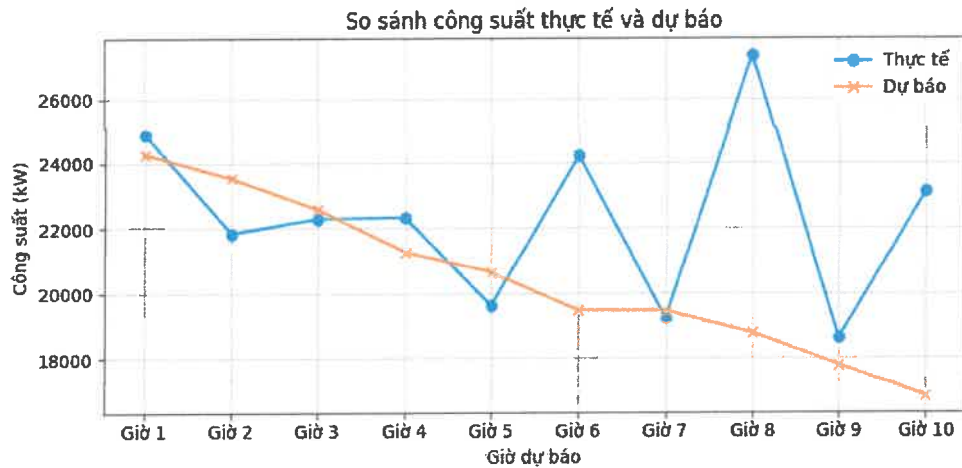


**Hình 3.6 Kết quả dự báo 6h tiếp theo**

Đề án cũng thử nghiệm với số bước xa hơn (so với bài toán dự báo ngắn hạn 1h đến vài giờ). Kết quả dự báo với số bước là 10h tiếp theo như bên dưới:

**Bảng 3.9 Kết quả dự báo 10h tiếp theo**

| Giờ    | Thực tế (kW) | Dự báo (kW) | MAE (kW) | MAPE (%) |
|--------|--------------|-------------|----------|----------|
| Giờ 1  | 24,888.20    | 24,269.43   | 618.77   | 2.49%    |
| Giờ 2  | 21,839.80    | 23,521.58   | 1,681.78 | 7.7%     |
| Giờ 3  | 22,291.30    | 22,569.67   | 278.37   | 1.25%    |
| Giờ 4  | 22,325.40    | 21,268.24   | 1,057.16 | 4.74%    |
| Giờ 5  | 19,605.80    | 20,645.80   | 1,040.00 | 5.3%     |
| Giờ 6  | 24,217.60    | 19,453.58   | 4,764.02 | 19.67%   |
| Giờ 7  | 19,216.60    | 19,436.80   | 220.2    | 1.15%    |
| Giờ 8  | 27,333.50    | 18,762.22   | 8,571.28 | 31.36%   |
| Giờ 9  | 18,609.60    | 17,787.54   | 822.06   | 4.42%    |
| Giờ 10 | 23,106.20    | 16,826.18   | 6,280.02 | 27.18%   |



**Hình 3.7 Kết quả dự báo 10h tiếp theo**

Từ những kết quả trên cho thấy mô hình LSTM có khả năng dự báo tương đối chính xác công suất phát điện trong ngắn hạn (4 giờ) – với MAPE trung bình  $\sim 4\%$ . Tuy nhiên, ở mức xa hơn (6h), mô hình có xu hướng lệch thực tế nhiều hơn, chưa duy trì được ổn định như ở 4 bước đầu, với MAPE trung bình ở mức 6.95%. Ở các bước xa hơn (10h), mô hình đang có xu hướng lệch rõ rệt hơn so với thực tế, MAPE trung bình ở mức 10.52%. Mặc dù với các mô hình dự báo phi tuyến, các kết quả dự báo cho chỉ số MAPE trung bình  $< 10\%$  được nhận định là khả quan, nhưng rõ ràng, sự biến động đột ngột của các yếu tố khí tượng có ảnh hưởng lớn đến việc dự báo chính xác kết quả, đặc biệt ở các bước xa. Mặc dù kết quả dự báo là tương đối chính xác, tuy nhiên, có sự khác biệt về độ lệch MAPE ở các bước khác nhau và không hẳn độ lệch này tăng (không chính xác) ở các bước xa hơn. Có thể thấy, ở các giờ gần vẫn có độ lệch MAPE cao, trong khi ở một số giờ xa hơn, độ lệch MAPE thấp hơn. Điều này có thể giải thích bởi 2 yếu tố, đầu tiên là việc thử nghiệm mô hình thực hiện trên một tập dữ liệu duy nhất, thứ hai là do sự biến thiên đột ngột về thời tiết ở các giờ khác nhau dẫn đến mô hình chưa cho kết quả dự báo chính xác.

Có thể nhận thấy, nếu chỉ dựa vào LSTM thì kết quả dự báo cho kết quả đáng tin cậy ở các bước gần (4-6h tiếp theo). Tuy nhiên, ở các bước xa, để có thể đạt mức độ chính xác cao hơn, cần có những thử nghiệm kết hợp LSTM với các mô hình khác.

### 3.2.3 So sánh với mô hình khác

Đề án tiến hành so sánh đánh giá dự báo với mô hình tuyến tính MLP(Multilayer Perceptron) và mô hình GRU(Gated Recurrent Unit) - ở bước dự báo 4h tiếp theo. Trong đó, GRU là mô hình học sâu tương đương với LSTM, còn MLP là mạng nơ-ron đơn giản gồm các lớp fully-connected. Mục đích để có thể thấy được sự vượt trội của LSTM trong kết quả thử nghiệm.

MLP(Multilayer Perceptron) là mạng nơ-ron truyền thẳng gồm nhiều lớp ẩn, có khả năng học các mối quan hệ phi tuyến trong dữ liệu[51].

Tuy nhiên, do không có cơ chế ghi nhớ thời gian, MLP cần các đặc trưng đầu vào dạng cửa sổ thời gian (sliding window) để xử lý chuỗi.

GRU(Gated Recurrent Unit) là biến thể đơn giản hơn của LSTM, thuộc nhóm mạng nơ-ron hồi tiếp (RNN), có khả năng ghi nhớ trạng thái theo thời gian.

Nhờ cấu trúc gọn nhẹ với ít tham số hơn, GRU thường huấn luyện nhanh hơn LSTM và vẫn duy trì được hiệu quả tương đối cao trên chuỗi thời gian[52].

#### a. Mô hình GRU:

Thông số mô hình GRU tương tự như LSTM theo bảng bên dưới:

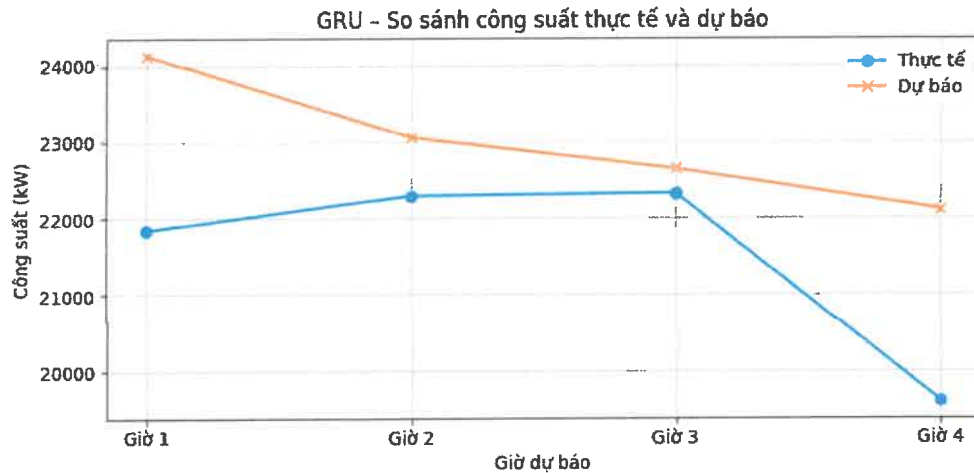
**Bảng 3.10 Các tham số của mô hình GRU**

| Tham số                   | Giá trị |
|---------------------------|---------|
| Số bước thời gian đầu vào | 48      |
| Batch size                | 64      |
| Số epoch huấn luyện       | 20      |
| Hàm tối ưu                | Adam    |
| Số bước dự báo            | 4h      |

Kết quả dự báo của mô hình GRU:

**Bảng 3.11 Kết quả dự báo 4h tiếp theo GRU**

| Giờ   | Thực tế (kW) | Dự báo (kW) | MAE      | MAPE (%) |
|-------|--------------|-------------|----------|----------|
| Giờ 1 | 21,839.80    | 24,133.64   | 2,293.84 | 10.50%   |
| Giờ 2 | 22,291.30    | 23,051.21   | 759.91   | 3.41%    |
| Giờ 3 | 22,325.40    | 22,639.62   | 314.22   | 1.41%    |
| Giờ 4 | 19,605.80    | 22,112.12   | 2,506.32 | 12.78%   |



**Hình 3.8 Kết quả dự báo 4h tiếp theo GRU**

*b. Mô hình MLP:*

Thông số mô hình MLP theo bảng bên dưới:

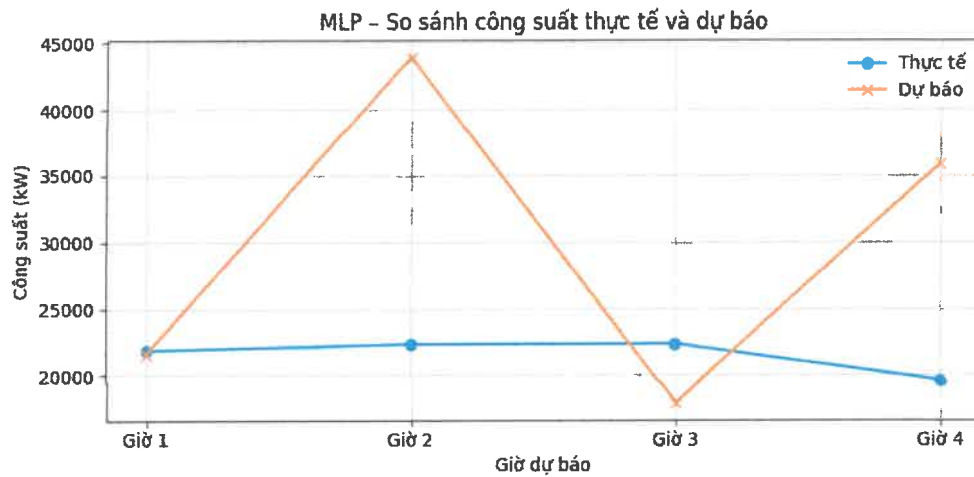
**Bảng 3.12 Các tham số của mô hình MLP**

| Tham số                   | Giá trị   |
|---------------------------|-----------|
| Số bước thời gian đầu vào | 48        |
| Số layer ẩn               | (128, 64) |
| Activation                | Relu      |
| Hàm tối ưu                | Adam      |
| Số bước dự báo            | 4h        |

Kết quả dự báo của mô hình MLP:

**Bảng 3.13 Kết quả dự báo 4h tiếp theo MLP**

| Giờ   | Thực tế (kW) | Dự báo (kW) | MAE       | MAPE (%) |
|-------|--------------|-------------|-----------|----------|
| Giờ 1 | 21,839.80    | 21,524.80   | 315       | 1.44%    |
| Giờ 2 | 22,291.30    | 43,939.75   | 21,648.45 | 97.12%   |
| Giờ 3 | 22,325.40    | 17,862.34   | 4,463.06  | 19.99%   |
| Giờ 4 | 19,605.80    | 35,870.89   | 16,265.09 | 82.96%   |



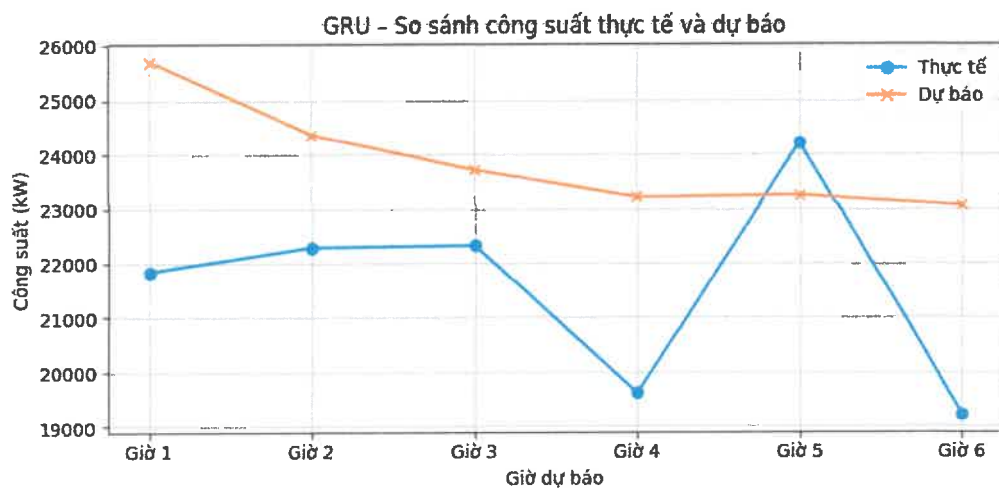
**Hình 3.9 Kết quả dự báo 4h tiếp theo MLP**

Từ kết quả thử nghiệm các mô hình, bảng thống kê bên dưới chỉ ra hiệu suất và độ chính xác cao của LSTM so với MLP và GRU.

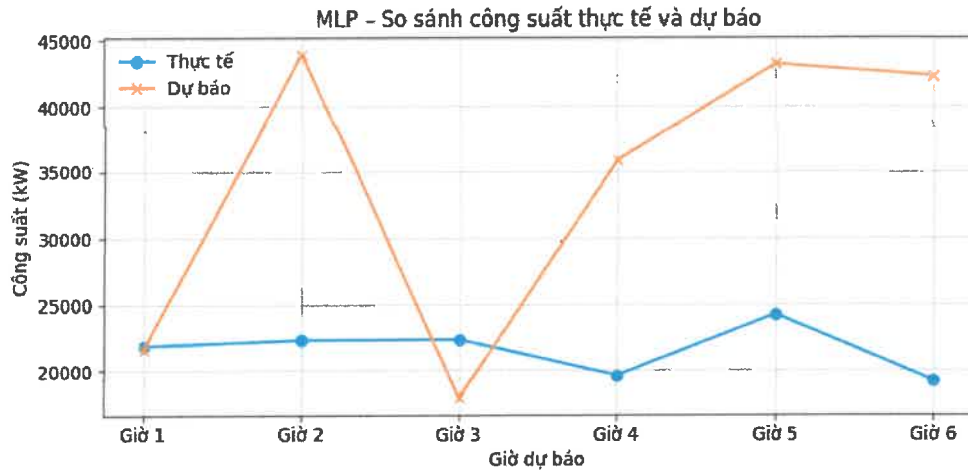
**Bảng 3.14 So sánh kết quả dự báo các mô hình(4h tiếp theo)**

| STT | Mô hình | MAPE (%) trung bình |
|-----|---------|---------------------|
| 1   | LSTM    | 4.06%               |
| 2   | MLP     | 50.38%              |
| 3   | GRU     | 7.03%               |

Ở số bước xa hơn, kết quả dự báo của 2 mô hình GRU và MLP có sự sai số rõ rệt.



**Hình 3.10 Kết quả dự báo 6h tiếp theo với GRU**



**Hình 3.11 Kết quả dự báo 6h tiếp theo với MLP**

Một số nhận xét dựa trên kết quả thử nghiệm 3 mô hình được tổng hợp như bên dưới:

- LSTM cho thấy khả năng học và dự báo ổn định xuyên suốt 4 bước dự báo. MAPE trung bình duy trì dưới 10%, với mức dao động nhẹ từ bước 1 đến bước 4. Đây là mô hình thể hiện tốt nhất toàn dải.
- GRU có hiệu suất gần tương đương với LSTM, có tốc độ nhanh hơn. Sai số dự báo theo từng giờ thấp và ổn định, phù hợp với các bài toán cần mô hình nhẹ hơn LSTM.
- MLP tuy dự báo rất chính xác ở bước đầu ( $MAPE \approx 1.44\%$ ), nhưng nhanh chóng mất độ chính xác ở các bước sau, do lỗi lan truyền tích lũy. MAPE của các bước sau tăng lên đến hơn 100%, khiến mô hình không phù hợp với bài toán multi-step.

### 3.3 Kết luận chương

Chương 3 đã trình bày chi tiết quá trình thử nghiệm mô hình LSTM trong bài toán dự báo công suất phát điện từ nhà máy điện gió, bao gồm xử lý dữ liệu đầu vào, cấu hình mô hình, huấn luyện và đánh giá hiệu quả. Mô hình LSTM cho kết quả tốt hơn so với các mô hình tương đương như MLP, GRU, với sai số tương đối thấp và khả năng bắt được các đặc trưng phức tạp của chuỗi thời gian khí tượng.

## KẾT LUẬN

### 1. Tóm tắt nội dung nghiên cứu

Trong đề án này, tác giả đã nghiên cứu và áp dụng mô hình mạng nơ-ron hồi tiếp LSTM (Long Short-Term Memory) cho bài toán dự báo công suất phát điện của nhà máy điện gió – một bài toán có ý nghĩa thực tiễn cao trong lĩnh vực năng lượng tái tạo và quản lý lưới điện.

Nội dung đề án được triển khai theo ba chương chính:

- **Chương 1** trình bày tổng quan về năng lượng gió, thực trạng dự báo công suất phát điện và các phương pháp đang được nghiên cứu và ứng dụng. Từ đó, xác định được tiềm năng và tính khả thi của mô hình LSTM trong bài toán này.
- **Chương 2** đề xuất phương pháp sử dụng mạng LSTM để mô hình hóa chuỗi thời gian dữ liệu khí tượng và công suất phát điện. Kiến trúc mô hình, các bước xử lý và chiến lược huấn luyện được mô tả chi tiết.
- **Chương 3** thực hiện huấn luyện và thử nghiệm mô hình trên dữ liệu thực tế. Mô hình LSTM cho kết quả chính xác cao với sai số dự báo thấp hơn so với các mô hình khác: MLP, GRU.

### 2. Kết luận

Qua quá trình nghiên cứu và thực nghiệm, đề án rút ra các kết luận sau:

- Mô hình LSTM có khả năng học và khai thác tốt các mối quan hệ phi tuyến và phụ thuộc thời gian dài hạn trong dữ liệu, phù hợp với bản chất biến động của công suất phát điện từ gió.
- Hiệu quả của mô hình LSTM vượt trội so với các mô hình truyền thống, đặc biệt là khi dữ liệu được tiền xử lý hợp lý và các đặc trưng khí tượng được lựa chọn phù hợp.
- Việc dự báo chính xác công suất phát điện giúp nâng cao khả năng lập kế hoạch vận hành, tối ưu hóa hiệu quả khai thác nguồn năng lượng tái tạo, và hỗ trợ điều độ lưới điện.

### 3. Hạn chế và hướng phát triển

Trong quá trình nghiên cứu và thực nghiệm mô hình LSTM cho bài toán dự báo công suất phát điện của nhà máy điện gió, bên cạnh những kết quả đạt được, đề án vẫn còn tồn tại một số hạn chế nhất định. Việc nhận diện và đánh giá các hạn chế này là cơ sở quan trọng để định hướng cho các bước phát triển tiếp theo, nhằm nâng cao độ chính xác, độ tin cậy và khả năng ứng dụng thực tế của mô hình. Bên cạnh đó, các hướng phát triển cũng được đề xuất nhằm khắc phục những điểm chưa hoàn thiện, đồng thời có thể khai thác tốt hơn tiềm năng của các mô hình học sâu trong lĩnh vực dự báo năng lượng tái tạo nói chung và dự báo công suất phát điện gió nói riêng.

### **1. Hạn chế**

Mặc dù mô hình LSTM đã cho thấy khả năng trong việc dự báo công suất phát điện nhà máy điện gió, tuy nhiên trong quá trình nghiên cứu và thực nghiệm vẫn còn tồn tại những hạn chế:

- Chưa tích hợp các yếu tố bên ngoài như lịch bảo trì, sự cố kỹ thuật hay điều kiện địa hình.
- Chưa kết hợp với các mô hình dự báo khí tượng theo thời gian thực.
- Mô hình hiện tại mới chỉ dự báo ngắn hạn (trong vài giờ tiếp theo), chưa mở rộng cho bài toán dự báo trung/dài hạn.
- Việc chọn lựa tham số và tối ưu mô hình hiện đang thực hiện thủ công, chưa áp dụng các phương pháp tự động như Grid Search hay Bayesian Optimization.

### **2. Hướng phát triển**

Từ phân tích về các hạn chế, có thể thấy đề án có nhiều tiềm năng để hoàn thiện và mở rộng hướng ứng dụng. Trong hướng phát triển tiếp theo, việc tích hợp thêm các yếu tố đầu vào như dữ liệu khí tượng theo thời gian thực, dữ liệu vận hành thực tế, điều kiện kỹ thuật của thiết bị, hay thông tin về bảo trì sẽ giúp mô hình phản ánh tốt hơn các yếu tố tác động đến công suất phát điện.

- Tối ưu hóa mô hình LSTM với các thuật toán khác: Firework Algorithm[15]...
- Kết hợp LSTM với các mô hình học sâu khác như CNN, Attention để cải thiện độ chính xác.

- Áp dụng các kỹ thuật ensemble learning hoặc hybrid modeling để tăng tính ổn định.
- Xây dựng hệ thống triển khai thực tế (real-time forecasting system) cho các nhà máy điện gió.
- Mở rộng nghiên cứu sang dự báo nhu cầu phụ tải hoặc kết hợp dự báo nguồn & phụ tải phục vụ điều độ thông minh.

## DANH MỤC CÁC TÀI LIỆU THAM KHẢO

- [1] Intracom, 9 September 2023, “*TIỀM NĂNG PHÁT TRIỂN NĂNG LƯỢNG GIÓ Ở VIỆT NAM*”. [Online]. Available: <https://intracom.com.vn/nang-luong-gio-o-viet-nam/>
- [2] Wen-Yeau Chang, “*A Literature Review of Wind Forecasting Methods*”, *Journal of Power and Energy Engineering*, 2, 161–168, 2014.
- [3] M. Khan, Tianqi Liu, F. Ullah, “*A New Hybrid Approach to Forecast Wind Power for Large Scale Wind Turbine Data Using Deep Learning with TensorFlow Framework and Principal Component Analysis*”, vol 12, *Energies* 2019.
- [4] Ying-Yi Hong, C. L. Paulo, P. Rioflorido, “*A hybrid deep learningbased neural network for 24-h ahead wind power forecasting*”, *ScienceDirect*, Pages 530-539, Volume 250, 15 September 2019.
- [5] Lê Hà Phan, PECC2, “*Điện gió và trí tuệ nhân tạo: Đã đến lúc máy móc có thể đưa ra dự báo chính xác?*”. [Online]. Available: <https://pecc2.com/en/dien-gio-va-tri-tue-nhan-cao-da-den-luc-may-moc-co-the-dua-ra-du-bao-chinh-xac-.html>
- [6] Đinh Thành Việt, Võ Văn Phương, Dương Minh Quân, Nguyễn Đình Ngọc Hải, Chu Văn Long. “*Nghiên cứu ứng dụng học sâu Trong dự báo công suất phát nguồn điện Gió*”. *Tạp Chí Khoa học Và Công nghệ - Đại học Đà Nẵng*, vol 19, số p.h 3, Tháng Ba 2021, tr 6-11, <https://jst-ud.vn/jst-ud/article/view/7374>.
- [7] EVN, 27 March 2025, “*Công suất năng lượng tái tạo toàn cầu tăng 585GW năm 2024*”. [Online]. Available: [Công suất năng lượng tái tạo toàn cầu tăng 585GW năm 2024](#).
- [8] NETZERO, 23 April 2025, “<https://moitruong.net.vn/viet-nam-phat-hien-tiem-nang-dien-gio-ngoai-khoi-hon-1-000-gw-nho-ve-tinh-sieu-hien-dai-82325.html>”.
- [Online]. Available: [Việt Nam phát hiện tiềm năng điện gió ngoài khơi hơn 1.000 GW nhờ vệ tinh siêu hiện đại](#)

- [9] VietData, 02 May 2024, “Điện Gió chiếm gần 80% số dự án năng lượng tái tạo hòa lưới 2023”, [Online]. Available: [Điện Gió chiếm gần 80% số dự án năng lượng tái tạo hòa lưới 2023](#).
- [10] Hochreiter, S., & Schmidhuber, J. (1997). “Long short-term memory”. *Neural Computation* (1997) 9 (8): 1735–1780. Doi: <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [11] Zhiheng Huang, Wei Xu, Kai Yu, 9 Aug 2015, “*Bidirectional LSTM-CRF Models for Sequence Tagging*”. Doi: <https://doi.org/10.48550/arXiv.1508.01991>.
- [12] Md Fazle Rabby, Yazhou Tu, Md Imran Hossen, Insup Le, Anthony S Maida, Xiali Hei, 18 Jan 2021, “*Stacked LSTM Based Deep Recurrent Neural Network with Kalman Smoothing for Blood Glucose Prediction*”. Doi: <https://doi.org/10.48550/arXiv.2101.06850>.
- [13] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, Wang-chun Woo, 19 Sep 2015, “*Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting*”. Doi: <https://doi.org/10.48550/arXiv.1506.04214>.
- [14] Shashwat Arghode, 19 May 2018, “*Wind-Energy-Prediction-using-LSTM*”. [Online]. Available: <https://github.com/ShashwatArghode/Wind-Energy-Prediction-using-LSTM>.
- [15] Bilin Shao, Dan Song, Genqing Bian, Yu Zhao, 28 September 2021, “*Wind Speed Forecast Based on the LSTM Neural Network Optimized by the Firework Algorithm*”. Doi: <https://doi.org/10.1155/2021/4874757>.
- [16] EVN, “*Phát triển năng lượng tái tạo: Thách thức trong điều hành hệ thống điện Việt Nam*” Tập đoàn Điện lực Việt Nam, Hà Nội, 2018. [Online] Available: [Phát triển năng lượng tái tạo: Thách thức trong điều hành hệ thống điện Việt Nam](#)
- [17] NREL, Wind Research. Online. Available: [Wind Data and Tools | Wind Research | NREL](#).
- [18] Tạp chí Công Thương, 16 August 2023, “*Hiện trạng hệ thống dự báo công suất phát nhà máy điện gió - mặt trời ở Việt Nam và xu hướng trong thời gian tới*”.

[Online]. Available: Hiện trạng hệ thống dự báo công suất phát nhà máy điện gió - mặt trời ở Việt Nam và xu hướng trong thời gian tới.

[19] M. H. El-Ahmar, A. -H. M. El-Sayed and A. M. Hemeida, "Evaluation of factors affecting wind turbine output power," *2017 Nineteenth International Middle East Power Systems Conference (MEPCON)*, Cairo, Egypt, 2017, pp. 1471-1476, doi: 10.1109/MEPCON.2017.8301377.

[20] Wang, D., Xu, M., Guangming, Z. *et al.* Enhancing wind power forecasting accuracy through LSTM with adaptive wind speed calibration (C-LSTM). *Sci Rep* 15, 5352 (2025). <https://doi.org/10.1038/s41598-025-89398-y>.

[21] Li, F., Wang, H., Wang, D., Liu, D., & Sun, K. (2025). "A Review of Wind Power Prediction Methods Based on Multi-Time Scales". *Energies*, 18(7), 1713. <https://doi.org/10.3390/en18071713> .

[22] Xin Zhao, Shuangxin Wang, Tao Li, "Review of Evaluation Criteria and Main Methods of Wind Power Forecasting", *Energy Procedia*, Volume 12, 2011, Pages 761-769, ISSN 1876-6102, <https://doi.org/10.1016/j.egypro.2011.10.102>

[23] Ergin Erdem, Jing Shi, "ARMA based approaches for forecasting the tuple of wind speed and direction", *Applied Energy*, Volume 88, Issue 4, 2011, Pages 1405-1414, ISSN 0306-2619, <https://doi.org/10.1016/j.apenergy.2010.10.031>.

[24] A. Lahouar, J. Ben Hadj Slama, "Hour-ahead wind power forecast based on random forests," *Renewable Energy*, Volume 109, 2017, Pages 529-541, ISSN 0960-1481, <https://doi.org/10.1016/j.renene.2017.03.064>.

[25] Qinghua Hu, Shiguang Zhang, Zongxia Xie, Jusheng Mi, Jie Wan, "Noise model based v-support vector regression with its application to short-term wind speed forecasting", *Neural Networks*, Volume 57, 2014, Pages 1-11, ISSN 0893-6080, <https://doi.org/10.1016/j.neunet.2014.05.003>.

[26] Huang, C.-J., & Kuo, P.-H. (2018). "A Short-Term Wind Speed Forecasting Model by Using Artificial Neural Networks with Stochastic Optimization for Renewable Energy Systems". *Energies*, 11(10), 2777. <https://doi.org/10.3390/en1102777> .

[27] Rui Yang, Hui Liu, Nikolaos Nikitas, Zhu Duan, Yanfei Li, Ye Li, "Short-term

wind speed forecasting using deep reinforcement learning with improved multiple error correction approach”, *Energy*, Volume 239, Part B, 2022, 122128, ISSN 0360-5442, <https://doi.org/10.1016/j.energy.2021.122128> .

[28] Zaccheus O. Olofe, “A 5-day wind speed & power forecasts using a layer recurrent neural network (LRNN)”, *Sustainable Energy Technologies and Assessments*, Volume 6, 2014, Pages 1-24, ISSN 2213-1388, <https://doi.org/10.1016/j.seta.2013.12.001>.

[29] Bo Gu, Tianren Zhang, Hang Meng, Jinhua Zhang, “Short-term forecasting and uncertainty analysis of wind power based on long short-term memory, cloud model and non-parametric kernel density estimation”, *Renewable Energy*, Volume 164, 2021, Pages 687-708, ISSN 0960-1481, <https://doi.org/10.1016/j.renene.2020.09.087>.

[30] Jianzhou Wang, Jianming Hu, Kailiang Ma, Yixin Zhang, “A self-adaptive hybrid approach for wind speed forecasting”, *Renewable Energy*, Volume 78, 2015, Pages 374-385, ISSN 0960-1481, <https://doi.org/10.1016/j.renene.2014.12.074> .

[31] Lu, P., Ye, L., Sun, B., Zhang, C., Zhao, Y., & Teng, J. (2018). “A New Hybrid Prediction Method of Ultra-Short-Term Wind Power Forecasting Based on EEMD-PE and LSSVM Optimized by the GSA”. *Energies*, 11(4), 697. <https://doi.org/10.3390/en11040697> .

[32] F. Chollet, *Keras*, GitHub repository, 2015. [Online]. Available: <https://github.com/keras-team/keras>

[33] M. Abadi *et al.*, “TensorFlow: Large-scale machine learning on heterogeneous distributed systems,” *arXiv preprint arXiv:1603.04467*, 2016. [Online]. Available: <https://arxiv.org/abs/1603.04467>

[34] J. Z. Wang, N. T. Chau, and B. H. W. Zhang, “Google Colaboratory,” Google Research, [Online]. Available: <https://colab.research.google.com> , Accessed on: Jun. 3, 2025.

[35] H. A. Rezaie and H. A. Jalab, “Time series forecasting using LSTM networks: A sliding window approach,” *2020 9th International Conference on Computer and Communication Engineering (ICCCCE)*, Kuala Lumpur, Malaysia, 2020, pp. 75–80.

[Online]. Available: <https://ieeexplore.ieee.org/document/9275402>.

- [36] N. Srivastava et al., "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
- [37] Wang, D., Cui, X., & Niu, D. (2022). "Wind Power Forecasting Based on LSTM Improved by EMD-PCA-RF". *Sustainability*, 14(12), 7307. <https://doi.org/10.3390/su14127307>.
- [38] Won Hee Chung, Yeong Hyeon Gu, Seong Joon Yoo, "District heater load forecasting based on machine learning and parallel CNN-LSTM attention", *Energy*, Volume 246, 2022, 123350, ISSN 0360-5442, <https://doi.org/10.1016/j.energy.2022.123350>.
- [39] Gama, João & Aguilar-Ruiz, Jesús & Klinkenberg, Ralf. (2008). "Knowledge discovery from data streams". *Intell. Data Anal.* 12. 251-252. 10.3233/IDA-2008-12301.
- [40] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning", MIT Press, 2016.
- [41] S. Brownlee, "Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python", Machine Learning Mastery, 2018.
- [42] Wang, J., Jiang, W., Li, Z., & Lu, Y. (2021). "A New Multi-Scale Sliding Window LSTM Framework (MSSW-LSTM): A Case Study for GNSS Time-Series Prediction". *Remote Sensing*, 13(16), 3328. <https://doi.org/10.3390/rs13163328>.
- [43] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, Springer, 2013.
- [44] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A Search Space Odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, 2017.
- [45] F. Chollet, *Deep Learning with Python*, 2nd ed., Manning Publications, 2021.
- [46] Kingma, Diederik & Ba, Jimmy. (2014). "Adam: A Method for Stochastic Optimization". International Conference on Learning Representations.
- [47] Masters, Dominic & Luschi, Carlo. (2018). "Revisiting Small Batch Training for Deep Neural Networks". 10.48550/arXiv.1804.07612.
- [48] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R.

- (2014). "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". *Journal of Machine Learning Research*, 15(56), 1929–1958.
- [49] Casolaro, A., Capone, V., Iannuzzo, G., & Camastra, F. (2023). "Deep Learning for Time Series Forecasting: Advances and Open Problems". *Information*, 14(11), 598. <https://doi.org/10.3390/info14110598> .
- [50] M. -L. Jiang, J. -L. Zhu, J. Chen, M. -W. Wu and J. -Q. Wu, "Research on Experimental Teaching of Communication Principles Based on Visible Light Communication," *2021 16th International Conference on Computer Science & Education (ICCSE)*, Lancaster, United Kingdom, 2021, pp. 1019-1023, doi: <https://doi.org/10.1109/ICCSE51940.2021.9569693>.
- [51] H. Ramchoun, M. A. J. Idrissi, Y. Ghanou, và M. Ettaouil, "Multilayer Perceptron: Architecture Optimization and Training," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 1, pp. 26–30, 2016, doi: 10.9781/ijimai.2016.415.
- [52] Cho, Kyunghyun; van Merriënboer, Bart; Bahdanau, Dzmitry; Bougares, Fethi; Schwenk, Holger; Bengio, Yoshua (2014). "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation". arXiv:1406.1078.
- [53] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, "Time Series Analysis: Forecasting and Control", 4th ed., Wiley, 2008.
- [54] Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2017). "Short-term residential load forecasting based on LSTM recurrent neural network". *IEEE Transactions on Smart Grid*, 10(1), 841–851. <https://doi.org/10.1109/TSG.2017.2753802>.
- [55] Marino, D. L., Amarasinghe, K., & Manic, M. (2016). "Building energy load forecasting using Deep Neural Networks". *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*. <https://doi.org/10.1109/IECON.2016.7793243>.
- [56] Bolboacă, R., & Haller, P. (2023). "Performance Analysis of Long Short-Term Memory Predictive Neural Networks on Time Series Data". *Mathematics*, 11(6), 1432. <https://doi.org/10.3390/math11061432>.

## **BẢN CAM ĐOAN**

Tôi cam đoan đã thực hiện việc kiểm tra mức độ tương đồng nội dung luận văn qua phần mềm <https://app.kiemtratailieu.vn> một cách trung thực và đạt kết quả mức độ tương đồng 15% toàn bộ nội dung luận văn. Bản luận văn kiểm tra qua phần mềm là bản cứng luận văn đã nộp để bảo vệ trước hội đồng. Nếu sai tôi xin chịu các hình thức kỷ luật theo quy định hiện hành của Học viện.

Hà Nội, ngày 31 tháng 07 năm 2025

**HỌC VIÊN CAO HỌC**



**Nguyễn Mạnh Hiển**



## BÁO CÁO KIỂM TRA TRÙNG LẬP

### Thông tin tài liệu

Tên tài liệu: Dean\_Thacsi\_NguyenManhHien  
Tác giả: Mạnh Hiến Nguyễn  
Điểm trùng lặp: 15  
Thời gian tải lên: 09:32 01/08/2025  
Thời gian sinh báo cáo: 09:33 01/08/2025  
Các trang kiểm tra: 60/60 trang



### Kết quả kiểm tra trùng lặp

15%

Có 15% nội dung trùng  
lặp

85%

Có 85% nội  
dung không  
trùng lặp

0%

Có 0% nội dung  
người dùng loại  
trừ

0%

Có 0% nội dung  
hệ thống bỏ qua

### Nguồn trùng lặp tiêu biểu

[tailieu.vn](http://tailieu.vn) [123docz.net](http://123docz.net) [arxiv.org](http://arxiv.org)

Học viên

Nguyễn Mạnh Hiến

Giảng viên hướng dẫn

TS. Nguyễn Ngọc Diệp

**BÁO CÁO GIẢI TRÌNH**  
**SỬA CHỮA, HOÀN THIỆN ĐỀ ÁN TỐT NGHIỆP**

Họ và tên học viên: Nguyễn Mạnh Hiền

Chuyên ngành: HTTT

Khóa: 2022 đợt 2

Tên đề tài: Nghiên cứu và áp dụng mô hình LSTM cho bài toán dự báo công suất phát điện nhà máy điện gió

Người hướng dẫn khoa học: TS. Nguyễn Ngọc Diệp

Ngày bảo vệ: 19/07/2025

Các nội dung học viên đã sửa chữa, bổ sung trong đề án tốt nghiệp theo ý kiến đóng góp của Hội đồng chấm đề án tốt nghiệp:

| TT | Ý kiến hội đồng                                                           | Sửa chữa của học viên                                                                                                                                                                                                                                                                                 |
|----|---------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | Rà soát trình bày                                                         | Tiếp thu góp ý của Hội đồng, học viên đã rà soát, chỉnh sửa các lỗi soạn thảo, trong đó có sửa sai công thức 1.4.                                                                                                                                                                                     |
| 2  | Câu hỏi học viên đã thực nghiệm bao nhiêu lần?                            | Sau quá trình thử nghiệm để tìm được các tham số mô hình phù hợp: số epoch, kích thước cửa sổ trượt, học viên đã thử nghiệm 10 lần trên tập dữ liệu gốc với tỉ lệ training và testing là 80:20 để lấy kết quả dự báo của mô hình.                                                                     |
| 3  | Giải thích kết quả thực nghiệm có sự biến đổi ở các khung giờ khác nhau?  | Kết quả thực nghiệm ở các khung giờ khác nhau có sự sai khác thể giải thích được là do sự biến thiên đột ngột của thời tiết, ở đây là tốc độ gió tăng hoặc giảm đột ngột, dẫn đến thực tế công suất phát cũng tăng giảm đột ngột mà mô hình chưa theo kịp. Học viên đã bổ sung trình bày ở mục 3.2.2. |
| 4  | Tại sao chọn LSTM thay vì chọn các phương pháp học sâu tiên tiến hiện nay | Qua khảo sát, học viên tìm hiểu LSTM được sử dụng nhiều ở các bài toán dự báo phi tuyến, nhất là các dự báo năng lượng điện gió. Học viên đã giải trình trong phần 2.1.                                                                                                                               |
| 5  | Giải thích lý do lựa chọn so sánh với GRU và MLP?                         | GRU và MLP thường được sử dụng trong các bài toán dự báo năng lượng gió và thường được sử dụng làm cơ sở để so sánh, đánh giá hiệu quả các                                                                                                                                                            |

|   |                                        |                                                                                        |
|---|----------------------------------------|----------------------------------------------------------------------------------------|
|   |                                        | mô hình trong nghiên cứu ở tài liệu [21], [22].                                        |
| 6 | Tổ chức nội dung trình bày gắn kết hơn | Tiếp thu góp ý của Hội đồng, học viên đã rà soát tổ chức nội dung trình bày của đề án. |

Hà Nội, ngày 04 tháng 08 năm 2025

**Ký xác nhận của**

**CHỦ TỊCH HỘI ĐỒNG  
CHẤM ĐỀ ÁN**



**GS.TS. Từ Minh Phương**

**THƯ KÝ HỘI  
ĐỒNG**



**TS. Đỗ Thị Liên**

**NGƯỜI HƯỚNG  
DẪN KHOA HỌC**



**TS. Nguyễn Ngọc Điệp**

**HỌC VIÊN**



**Nguyễn Mạnh Hiên**

**BIÊN BẢN  
HỌP HỘI ĐỒNG CHẤM ĐỀ ÁN TỐT NGHIỆP THẠC SĨ**

Căn cứ quyết định số Quyết định số 1098/QĐ-HV ngày 26 tháng 06 năm 2025 của Giám đốc Học viện Công nghệ Bưu chính Viễn thông về việc thành lập Hội đồng chấm đề án tốt nghiệp thạc sĩ. Hội đồng đã họp vào hồi...9...giờ...55...phút, ngày 19 tháng 07 năm 2025 tại Học viện Công nghệ Bưu chính Viễn thông để chấm đề án tốt nghiệp thạc sĩ cho:

Học viên: **Nguyễn Mạnh Hiên**

Tên đề án tốt nghiệp: **Nghiên cứu và áp dụng mô hình LSTM cho bài toán dự báo công suất phát điện nhà máy điện gió**

Chuyên ngành: **Hệ thống thông tin**

Mã số: **8480104**

Các thành viên của Hội đồng chấm đề án tốt nghiệp có mặt: ...../ 05

| TT | HỌ VÀ TÊN                        | TRÁCH NHIỆM TRONG HĐ | GHI CHÚ |
|----|----------------------------------|----------------------|---------|
| 1  | <b>GS.TS. Từ Minh Phương</b>     | Chủ tịch             |         |
| 2  | <b>TS. Đỗ Thị Liên</b>           | Thư ký               |         |
| 3  | <b>PGS. TS. Nguyễn Mạnh Hùng</b> | Phản biện 1          |         |
| 4  | <b>TS. Nguyễn Văn Vinh</b>       | Phản biện 2          |         |
| 5  | <b>PGS.TS. Trần Đăng Hưng</b>    | Ủy viên              |         |

**Các nội dung thực hiện:**

- Chủ tịch Hội đồng điều khiển buổi họp. Công bố quyết định của Giám đốc Học viện Công nghệ Bưu chính Viễn thông về việc thành lập Hội đồng chấm đề án tốt nghiệp thạc sĩ.
- Người hướng dẫn khoa học hoặc thư ký đọc lý lịch khoa học và các điều kiện bảo vệ đề án tốt nghiệp của học viên. (có bản lý lịch khoa học và kết quả các môn học cao học của học viên kèm theo).
- Học viên trình bày tóm tắt đề án tốt nghiệp.
- Phản biện 1 đọc nhận xét (có văn bản kèm theo)
- Phản biện 2 đọc nhận xét (có văn bản kèm theo)

**6. Các câu hỏi của thành viên Hội đồng:**

- Thầy nghiên cứu đề án kết quả thực nghiệm bao nhiêu lần?
- Giải thích kết quả thực nghiệm lần đầu 3' các lượng giá khác nhau.
- Tại sao chọn mô hình 'LSTM' thay vì các phương pháp học sâu hiện hành nay?
- Tại sao chủ so sánh phương pháp LSTM với phương pháp CNN, MLP,

**7. Trả lời của học viên:**

- Học viên giải trình và làm rõ trong báo cáo

8. Thư ký đọc nhận xét về quá trình thực hiện đề án tốt nghiệp của học viên (có văn bản kèm theo).

9. Hội đồng họp riêng:

- Bầu Ban kiểm phiếu:

1. Trưởng Ban kiểm phiếu:

Trần Đăng Hưng

2. Ủy viên Ban kiểm phiếu:

Đỗ Thị Liên

3. Ủy viên Ban kiểm phiếu:

Nguyễn Mạnh Hùng

- Hội đồng chấm đề án tốt nghiệp bằng bỏ phiếu kín.

- Ban kiểm phiếu làm việc:

- Trưởng Ban kiểm phiếu báo cáo kết quả kiểm phiếu (có Biên bản họp Ban kiểm phiếu kèm theo)

- Điểm trung bình của đề án tốt nghiệp: 8,1

Kết luận:

1. Các nội dung cần chỉnh sửa, hoàn thiện sau bảo vệ đề án tốt nghiệp:

- Bỏ coat trình bày

- Làm rõ những nhận xét theo góp ý của các thành viên trong hội đồng

- Tổ chức nội dung trình bày gắn kết hơn

2. Đề nghị Học viện công nhận (hoặc không) và cấp bằng (hoặc không) thạc sĩ cho học viên:

Đề nghị học viện công nhận và cấp bằng thạc sĩ cho học viên

3. Đề án tốt nghiệp có thể phát triển thành đề tài nghiên cứu cho

NCS. Không

Buổi làm việc kết thúc vào... 10h40... cùng ngày.

Chủ tịch



GS.TS. Từ Minh Phương

Thư ký



TS. Đỗ Thị Liên

**CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM**  
**Độc lập – Tự do – Hạnh phúc**  
-----o0o-----

**BẢN NHẬN XÉT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ**

*(Dùng cho người phản biện)*

Tên đề án: *Nghiên cứu và áp dụng mô hình LSTM cho bài toán dự báo công suất phát điện nhà máy điện gió*

Chuyên ngành: Hệ thống thông tin

Mã chuyên ngành: 8.48.01.04

Tên học viên: *Nguyễn Mạnh Hiên*

Họ và tên người phản biện: *Nguyễn Mạnh Hùng*

Học hàm, học vị: PGS.TS.

Nơi công tác: Khoa Công nghệ thông tin 1 - Học viện Công nghệ Bưu chính viễn thông.

Số điện thoại liên hệ: .....

**NỘI DUNG NHẬN XÉT**

**I/ Cơ sở khoa học và thực tiễn, tính cấp thiết của đề tài:**

Bài toán dự báo công suất điện gió là một trong các bài toán có ý nghĩa thực tiễn hiện nay.

**II/ Nội dung của đề án, các kết quả đã đạt được:**

Đề án bao gồm 3 chương:

Chương 1 trình bày về bài toán dự đoán công suất điện gió.

Chương 2 trình bày phương pháp áp dụng mô hình LSTM vào bài toán dự đoán công suất điện gió.

Chương 3 trình bày thử nghiệm và đánh giá mô hình

Nhìn chung, đề án đáp ứng được các yêu cầu cơ bản của một đề án thạc sĩ kỹ thuật.

Tuy nhiên, đề án vẫn còn một số vấn đề:

- Công thức 1.4 trang 13 còn thiếu giá trị tuyệt đối.
- Phần kết quả thực nghiệm mới thấy dự đoán trong vòng 4.6.10h tiếp theo, chưa thấy kết quả đánh giá dự đoán theo mùa vụ hay theo năm.
- Phần so sánh với các phương pháp khác không trình bày chung kết quả dự đoán của tất cả phương pháp được so sánh vào cùng biểu đồ cho dễ đối chiếu.

**III/ Những vấn đề cần giải thích thêm:**

- Các thực nghiệm dự đoán cho kết quả ở bảng 3.7 được thực hiện bao nhiêu lần?
- So sánh kết quả bảng 3.7 và 3.11, tác giả có thể giải thích lí do tại sao GRU cho kết quả tốt hơn ở giờ 2 và giờ 3 nhưng lại kém hơn ở giờ 1 không?

**IV/ Kết luận:**

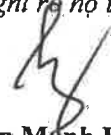
Đề án đáp ứng được yêu cầu của một đề án thạc sĩ kỹ thuật.

Đồng ý cho phép học viên được bảo vệ đề án tốt nghiệp.

Hà Nội, ngày 15 tháng 07 năm 2025

**Người nhận xét**

(Ký ghi rõ họ tên)



**Nguyễn Mạnh Hùng**

**CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM**

**Độc lập – Tự do – Hạnh phúc**

**BẢN NHẬN XÉT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ**  
**(Dùng cho người phản biện)**

Tên đề tài đề án tốt nghiệp: **NGHIÊN CỨU VÀ ỨNG DỤNG MÔ HÌNH LSTM CHO BÀI TOÁN DỰ BÁO CÔNG SUẤT PHÁT ĐIỆN NHÀ MÁY ĐIỆN GIÓ**

Chuyên ngành: **Hệ thống thông tin**

Mã chuyên ngành: .....

Họ và tên học viên: **Nguyễn Mạnh Hiền**

Họ và tên người nhận xét: **Nguyễn Văn Vinh**

Học hàm, học vị: **TS**

Chuyên ngành: **KHMT**

Cơ quan công tác: **Trường ĐH Công Nghệ, ĐHQG Hà Nội**

Số điện thoại:.....E-mail:.....

**NỘI DUNG NHẬN XÉT**

**I/ Cơ sở khoa học và thực tiễn, tính cấp thiết của đề tài:**

Năng lượng sạch ngày càng quan trọng, đặc biệt năng lượng dựa vào gió. Việc dự đoán chính xác phát điện của nhà máy phát điện gió là quan trọng trong việc quản lý và điều phối nguồn điện. Vì vậy đề tài dự đoán công suất phát điện nhà máy điện gió là có ý nghĩa về mặt khoa học cũng như thực tiễn.

**II/ Nội dung của đề án tốt nghiệp, các kết quả đã đạt được:**

- Nghiên cứu tổng quan về nhà máy phát điện gió
- Nghiên cứu về mô hình LSTM và ứng dụng phương pháp học máy này cho bài toán dự đoán công suất nhà máy phát điện gió.
- Thử nghiệm và đánh giá với bộ dữ liệu NREL (Mỹ) cho kết quả khả quan.

**III/ Những vấn đề cần giải thích thêm:**

- Dữ liệu nghiên cứu khá cũ (2010) nên như hiện nay có thể có nhiều yếu tố khác ảnh hưởng đến khả năng dự đoán công suất điện gió này. Có thể tham khảo và thử nghiệm dữ liệu mới hơn.
- Bổ xung các nghiên cứu liên quan đến bài toán này trong và ngoài nước.
- Đề án cũng nên giải thích rõ hơn tại sao lại lựa chọn phương pháp dựa vào LSTM trong khi hiện nay có thể dựa vào mô hình Transformer hoặc Large language models có thể dự đoán tốt hơn?

- Chương 3 sang chương mới cũng không break trang. Chương 3 nhưng các mục con lại bắt đầu từ 4.x.
- Tài liệu tham khảo sơ sài và không có trích dẫn trong báo cáo.

**III/ Những vấn đề cần giải thích thêm:**

- CSDL có phân tán theo các vi dịch vụ không?

**IV/ Kết luận:**

Đề án đáp ứng được yêu cầu của một đề án thạc sĩ kỹ thuật.

Đồng ý cho phép học viên được bảo vệ đề án tốt nghiệp.

Hà Nội, ngày 15 tháng 07 năm 2025

**Người nhận xét**  
(Ký ghi rõ họ tên)



**Nguyễn Mạnh Hùng**

the 1990s, the number of people in the UK who are employed in the public sector has increased by 1.5 million, from 2.5 million in 1980 to 4 million in 1995 (Department of Health 1996).

There is a growing emphasis on the need to improve the quality of care in the public sector, and to ensure that the public sector is able to meet the needs of the population. This has led to a number of initiatives, including the introduction of the Health Care Act 1999, which aims to improve the quality of care in the public sector, and to ensure that the public sector is able to meet the needs of the population. The Health Care Act 1999 also aims to improve the quality of care in the private sector, and to ensure that the private sector is able to meet the needs of the population.

The Health Care Act 1999 also aims to improve the quality of care in the public sector, and to ensure that the public sector is able to meet the needs of the population. This has led to a number of initiatives, including the introduction of the Health Care Act 1999, which aims to improve the quality of care in the public sector, and to ensure that the public sector is able to meet the needs of the population.

The Health Care Act 1999 also aims to improve the quality of care in the public sector, and to ensure that the public sector is able to meet the needs of the population. This has led to a number of initiatives, including the introduction of the Health Care Act 1999, which aims to improve the quality of care in the public sector, and to ensure that the public sector is able to meet the needs of the population.

The Health Care Act 1999 also aims to improve the quality of care in the public sector, and to ensure that the public sector is able to meet the needs of the population. This has led to a number of initiatives, including the introduction of the Health Care Act 1999, which aims to improve the quality of care in the public sector, and to ensure that the public sector is able to meet the needs of the population.

The Health Care Act 1999 also aims to improve the quality of care in the public sector, and to ensure that the public sector is able to meet the needs of the population. This has led to a number of initiatives, including the introduction of the Health Care Act 1999, which aims to improve the quality of care in the public sector, and to ensure that the public sector is able to meet the needs of the population.

The Health Care Act 1999 also aims to improve the quality of care in the public sector, and to ensure that the public sector is able to meet the needs of the population. This has led to a number of initiatives, including the introduction of the Health Care Act 1999, which aims to improve the quality of care in the public sector, and to ensure that the public sector is able to meet the needs of the population.

The Health Care Act 1999 also aims to improve the quality of care in the public sector, and to ensure that the public sector is able to meet the needs of the population. This has led to a number of initiatives, including the introduction of the Health Care Act 1999, which aims to improve the quality of care in the public sector, and to ensure that the public sector is able to meet the needs of the population.

The Health Care Act 1999 also aims to improve the quality of care in the public sector, and to ensure that the public sector is able to meet the needs of the population. This has led to a number of initiatives, including the introduction of the Health Care Act 1999, which aims to improve the quality of care in the public sector, and to ensure that the public sector is able to meet the needs of the population.