

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



Nguyễn Mạnh Hiền

**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

TÓM TẮT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

HÀ NỘI – 2025

Đề án tốt nghiệp được hoàn thành tại:
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: **Tiến sĩ Nguyễn Ngọc Điệp**

Phản biện 1:

Phản biện 2:

Đề án tốt nghiệp sẽ được bảo vệ trước Hội đồng chấm đề án tốt nghiệp thạc sĩ tại Học viện
Công nghệ Bưu chính Viễn thông

Vào lúc: giờ ngày tháng năm

Có thể tìm hiểu đề án tốt nghiệp tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông.

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:

- Hiểu 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

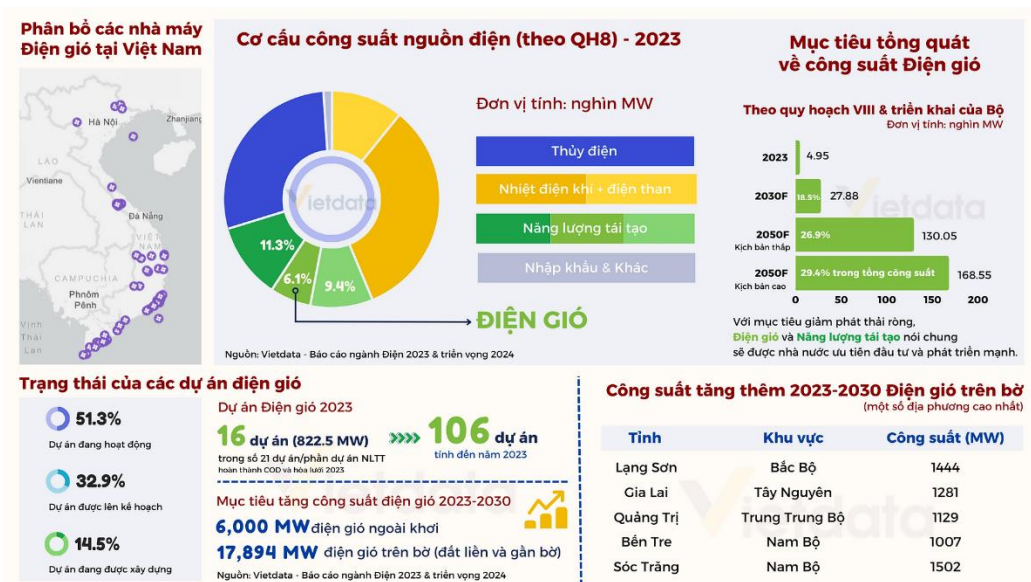
- Thu thập bộ dữ liệu huấn luyện từ dữ liệu về gió tại nhà máy điện gió
- Thử nghiệm mô hình đề xuất, phân tích và đánh giá kết quả đạt được

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.



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.

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.

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

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:

- Dự báo cực ngắn hạn (ultra-short-term forecasting): từ vài phút đến 1 giờ.
- Dự báo ngắn hạn (short-term forecasting): từ một giờ đến vài giờ.
- Dự báo trung hạn (medium-term forecasting): từ vài giờ đến vài ngày.
- Dự báo dài hạn (long-term forecasting): từ vài tuần đến vài tháng.

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.
- Độ trễ thời gian và tính liên tục.
- Nhu cầu xử lý và phân tích dữ liệu lớn.

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.
- Tối ưu hóa vận hành nhà máy.

- Tăng hiệu quả tham gia thị trường điện.
- Hỗ trợ quy hoạch năng lượng.

1.3 Các phương pháp tiếp cận hiện tại và hạn chế

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.

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.

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)
- ARIMA (Autoregressive Integrated Moving Average)

1.3.3 Các phương pháp học máy (Machine Learning)

Các phương pháp học máy truyền thống gồm:

- RF(Random Forest)
- SVR(Support Vector Regression)

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

- Mạng nơ-ron tích chập CNN (Convolutional neural networks)
- DBN(Deep belief networks)
- Mạng nơ-ron hồi quy RNN(Requirement Neural Network)
- Mạng bộ nhớ dài ngắn hạn LSTM(Long short-term memory)

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.

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.
- Phụ thuộc dữ liệu lịch sử chất lượng cao.
- Khó chọn tham số và kiến trúc tối ưu.
- Thiếu đánh giá toàn diện.

1.4 Một số lý thuyết có liên quan

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).

Đặc điểm của chuỗi thời gian công suất gió gồm:

- Tính không ổn định (non-stationarity).
- Tính phi tuyến (non-linearity).
- Chu kỳ và mùa vụ (seasonality).
- Độ phân giải cao.

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 như:

- MAE (Mean Absolute Error):

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

- MSE (Mean Square Error):

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

- RMSE (Root Mean Square Error):

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

- MAPE (Mean Absolute Percentage Error):

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i}$$

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.

1.4.3 Các kỹ thuật tiền xử lý dữ liệu

- Xử lý thiếu dữ liệu (*Missing Values*)
- Phát hiện và xử lý giá trị ngoại lai (*Outlier Detection*)
- Chuẩn hóa dữ liệu (*Normalization / Standardization*)
- Tạo đặc trưng thời gian (*Time Features*)
- Tách tập huấn luyện, kiểm tra theo thời gian

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.

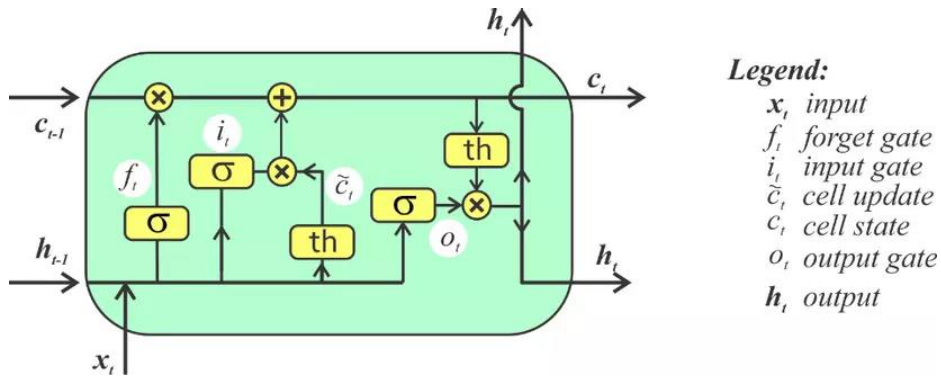
Trong chương 2, đề án sẽ trình bày chi tiết phương pháp đề xuất sử dụng mô hình LSTM để dự báo công suất phát điện nhà máy điện gió, bao gồm kiến trúc mô hình, chiến lược huấn luyện và các yếu tố kỹ thuật liên quan.

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.

Mô hình LSTM được đề xuất bởi Hochreiter và Schmidhuber, 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.



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):

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

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

- Cổng vào (Input Gate):

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

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

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

- Cập nhật trạng thái bộ nhớ:

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

- Cổng đầu ra (Output Gate):

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

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

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
- σ : hàm sigmoid
- $\tanh$: hàm tanh

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.

- 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:

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

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.

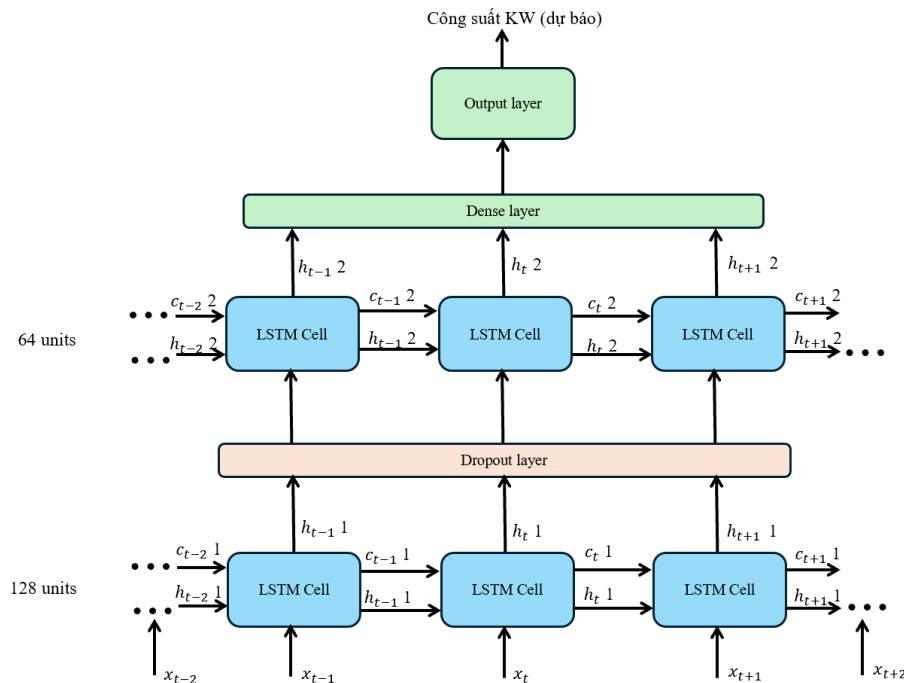
Cấu trúc tổng thể của mô hình bao gồm các thành phần chính như sau:

- 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.

- Lớp LSTM: Một hoặc nhiều lớp LSTM được sử dụng để trích xuất đặc trưng tuần tự từ chuỗi dữ liệu. Số lượng nơ-ron (units) trong mỗi lớp là một siêu tham số, được tinh chỉnh thông qua quá trình thử nghiệm.

- Lớp Dropout: được chèn giữa các lớp LSTM nhằm giảm thiểu hiện tượng overfitting.

- Lớp Dense đầu ra: trường hợp mở rộng thành dự báo nhiều bước, lớp này có thể mở rộng thành nhiều nơ-ron 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.

2.2.3 Quy trình huấn luyện mô hình

a. Tiền xử lý dữ liệu:

- Xử lý dữ liệu thiếu và nhiễu:

- Chuẩn hóa dữ liệu:

Trong nghiên cứu này, phương pháp Min-Max Scaling được sử dụng để chuẩn hóa dữ liệu về khoảng $[0, 1]$, theo công thức:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

- Tạo chuỗi đầu vào theo cửa sổ trượt:

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).

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.

c. Xây dựng mô hình:

Đề án thử nghiệm triển khai mô hình LSTM với các tham số khác nhau để đưa ra các kết quả đánh giá. Các tham số thử nghiệm mô hình bao gồm:

- Số lớp ẩn (layer)
- Số đơn vị nơ-ron trên 1 lớp ẩn (units)

Ngoài ra, dựa trên kết quả của các nghiên cứu, một tham số khác trong mô hình được sử dụng dựa trên đánh giá của các nghiên cứu này.

Lớp Dropout: Dropout = 0.2.

Lớp đầu ra (Dense layer): một lớp Dense đầu ra, dùng hàm kích hoạt tuyến tính:

$$y = W \cdot x + b$$

để tổng hợp đầu ra là công suất phát điện cần dự báo.

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 mở rộng sang bài toán multi-step.

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.

- Hàm mất mát: MSE (Mean Squared Error)

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

- Thuật toán tối ưu: Adam Optimizer
- Số epoch:
- Kích thước lô huấn luyện (Batch size): 64

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.

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.

2.2.4 Đánh giá ưu điểm mô hình LSTM

Khả năng học dài hạn.

Khả năng khử nhiễu.

Mở rộng linh hoạt.

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

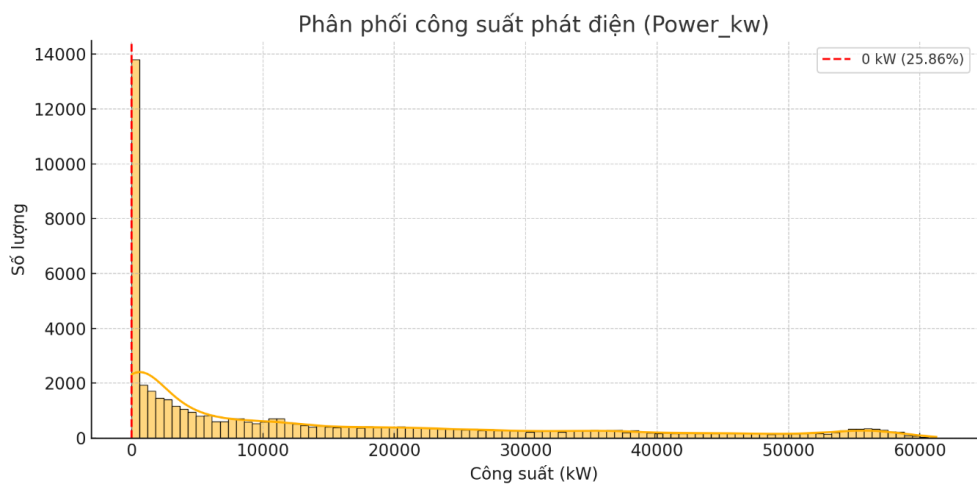
Dữ liệu sử dụng trong nghiên cứu được thu thập từ dữ liệu được thu thập từ dữ liệu từ NREL (Mỹ). Dữ liệu bao gồm:

- 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.

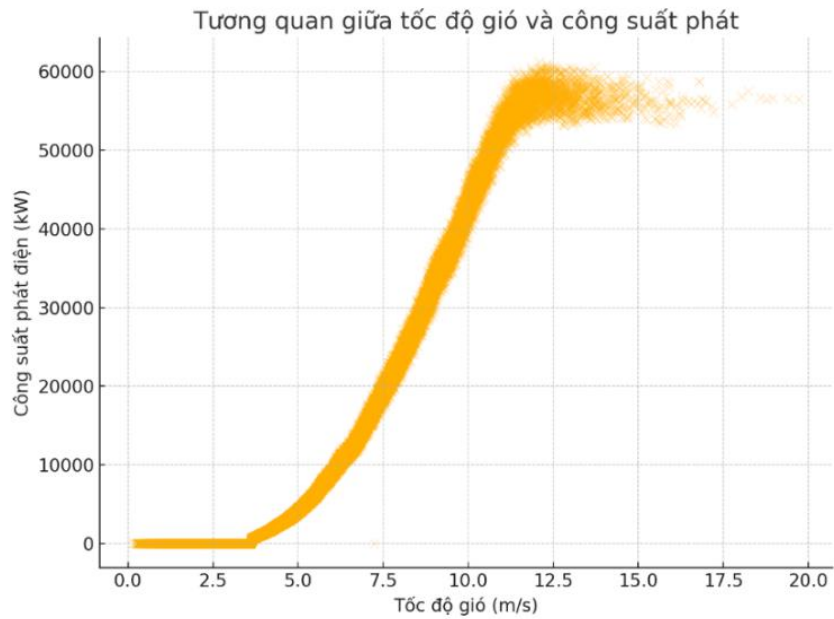
Tần suất lấy mẫu là theo giờ, với tổng số bản ghi khoảng 52.560 dòng dữ liệu.

Bảng 3.1 Dữ liệu sử dụng trong nghiên cứu

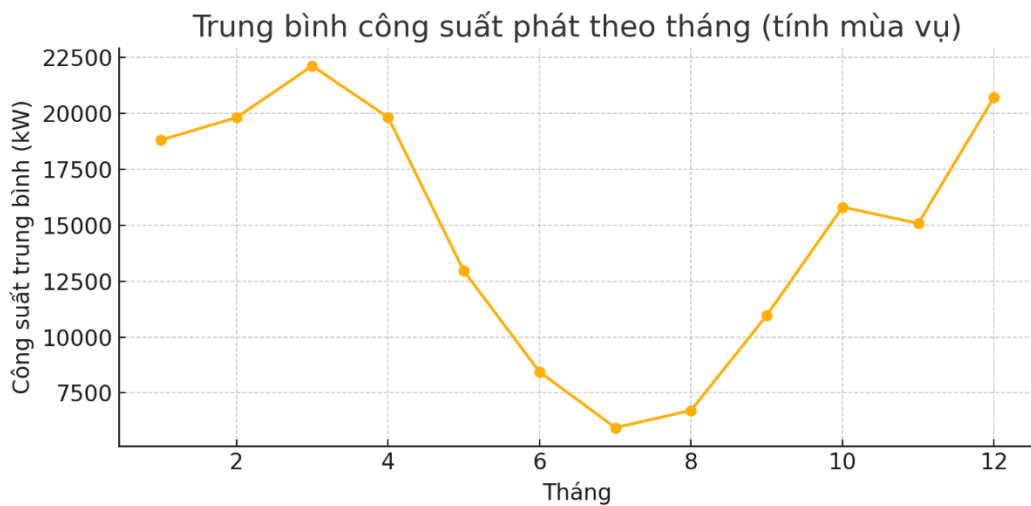
DateTime	Air temperature (°C)	Pressure (atm)	Wind speed (m/s)	Wind direction (deg)	Power generated by system (kW)
2010-04-02 08:00:00	9.954	0.979708	8.227	188	26038.4
2010-04-02 09:00:00	9.646	0.97986	8.449	186	28059.7
2010-04-02 10:00:00	9.045	0.980181	8.02	184	24121.9
2010-04-02 11:00:00	8.726	0.980244	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



Hình 3.1 Biểu đồ phân phối Công suất phát điện



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



Hình 3.4 Biểu đồ Công suất trung bình theo tháng

3.1.2 Tiền xử lý dữ liệu

- Xử lý dữ liệu thiếu (missing values): không có dòng trống trong dữ liệu.
- Loại bỏ dữ liệu có thể gây nhiễu
- Chuẩn hóa dữ liệu về khoảng $[0,1]$ bằng kỹ thuật Min-Max Scaling.
- 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 thí nghiệm

Chương trình sử dụng thư viện mã nguồn mở Keras – 1 API cao cấp chạy trên nền tảng TensorFlow. 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.

3.2.2 Kết quả thử nghiệm

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 3.3: Đánh giá dự báo theo kích thước cửa sổ trượt

STT	Kích thước cửa sổ trượt	MAPE% trung bình
1	24	131.44%
2	48	4.38%
3	72	6.25%
4	96	19.37%

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

Bảng 3.5: Đánh giá dự báo theo số lần lặp

STT	Epochs	MAPE% trung bình
1	10	4.38%
2	20	4.06%
3	30	15.88%
4	40	18.89%
5	50	20.87%

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.

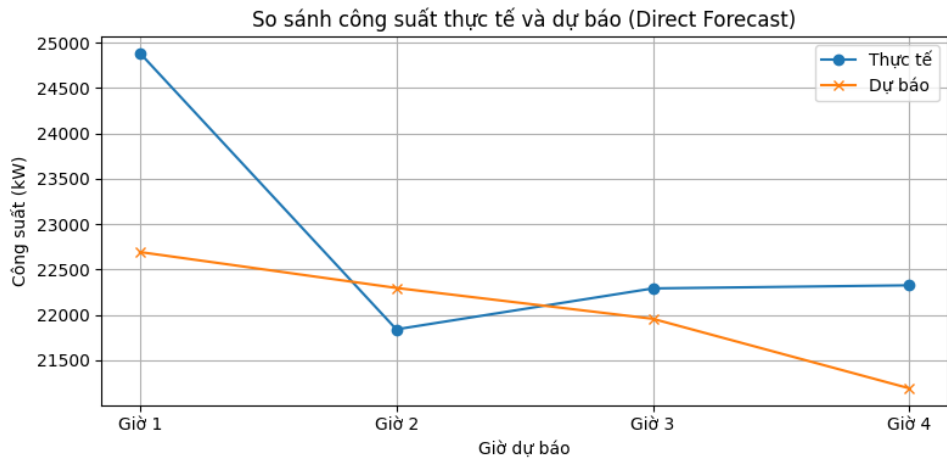
Dưới đây là một số kết quả thử nghiệm dựa trên các tham số của mô hình dựa trên bảng tham số mô hình chung.

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 Lớp 2: 64 units
Batch size	64
Số epoch huấn luyện	20
Hàm tối ưu	Adam

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%



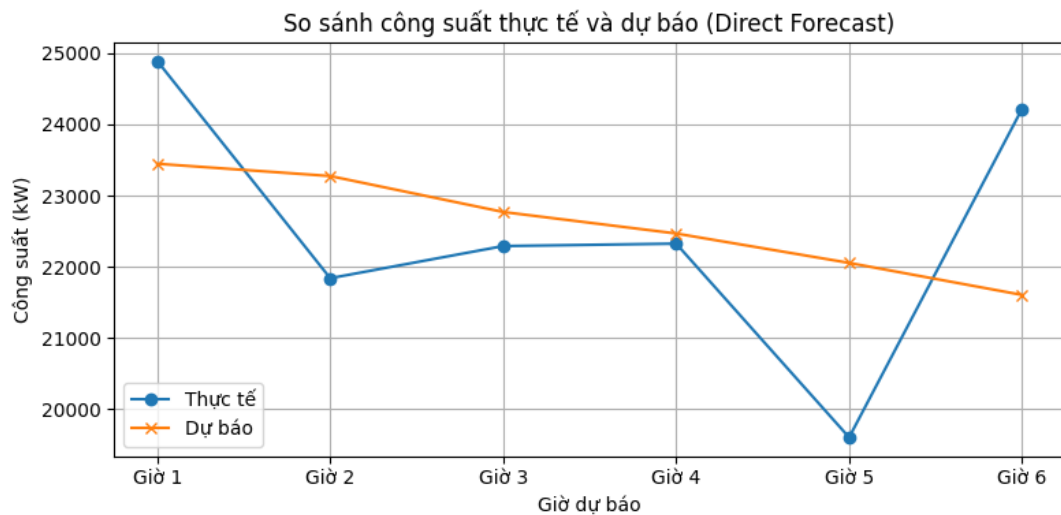
Hình 3.5 Kết quả dự báo 4h tiếp theo

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

Kết quả 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%.

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.

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

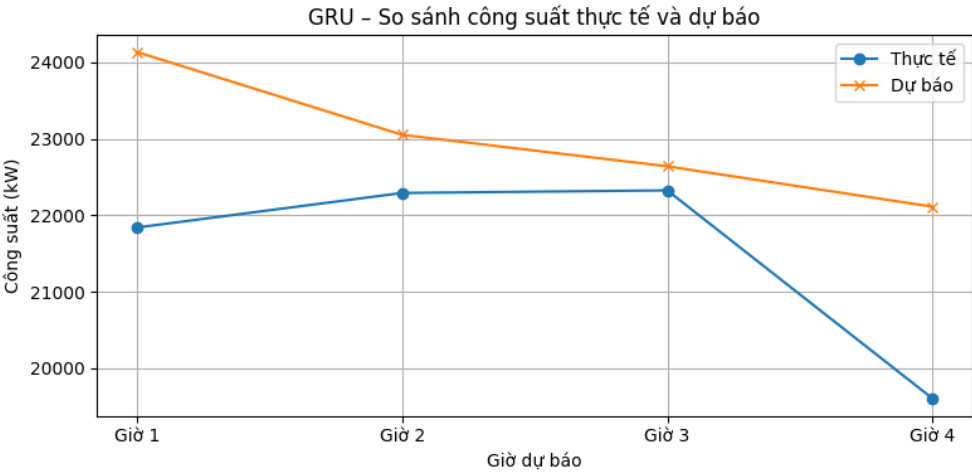
Tham số	Giá trị
Số bước thời gian đầu vào	48
Batch size	64
Số epoch huấn luyện	20
Hàm tối ưu	Adam

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:

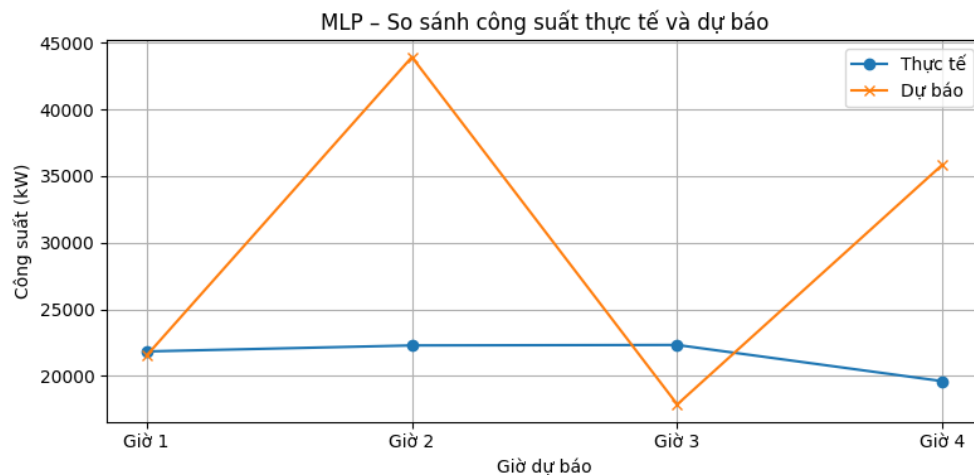
Bảng 3.12 Các tham số của mô hình MLP

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%

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.