

ỨNG DỤNG CNN-LSTM ĐỂ PHÂN TÍCH NỒNG ĐỘ $PM_{2.5}$ SỬ DỤNG TRẠM QUAN TRẮC TẠI KHU CÔNG NGHIỆP HÒA XÁ - NAM ĐỊNH

CNN-LSTM APPLICATION TO ANALYZE $PM_{2.5}$ CONCENTRATION USING MONITORING STATION AT HOA XA INDUSTRIAL PARK - NAM DINH

Nguyễn Thiên Tân^{1,*},
Bùi Văn Hậu¹, Đỗ Quang Hiệp²

DOI: <https://doi.org/10.57001/huinh5804.2025.283>

TÓM TẮT

Ô nhiễm không khí là một vấn đề toàn cầu quan trọng ảnh hưởng đến sức khỏe con người, hệ sinh thái và môi trường. Sự phức tạp của ô nhiễm không khí phát sinh từ các nguồn, loại và các tương tác phức tạp trong khí quyển dẫn đến các tác động môi trường và sức khỏe khác nhau. Nhiều phương pháp khác nhau đã được phát triển để dự đoán chất lượng không khí, chủ yếu được phân loại thành các phương pháp thống kê và học máy. Các phương pháp này tận dụng dữ liệu lịch sử và các thông số môi trường để dự báo các chỉ số chất lượng không khí và nồng độ chất ô nhiễm. Việc tích hợp các kỹ thuật này đã cho thấy hứa hẹn trong việc cải thiện độ chính xác dự đoán và cung cấp các cảnh báo kịp thời để quản lý chất lượng không khí. Trong nghiên cứu này, chúng tôi đề xuất sử dụng mô hình kết hợp CNN-LSTM (Convolutional Neural Network Long Short-Term Memory Networks) để phân tích, dự báo chất lượng không khí tại khu công nghiệp Hòa Xá, tỉnh Nam Định. Kết quả chứng minh rằng mô hình CNN-LSTM có những phân tích, dự báo tốt và có thể được sử dụng để đánh giá chất lượng không khí trong các vùng lớn hơn.

Từ khóa: Giám sát và dự báo ô nhiễm không khí chi tiết, mô hình CNN-LSTM, $PM_{2.5}$, khu công nghiệp Hòa Xá - Nam Định.

ABSTRACT

Air pollution is a major global problem affecting human health, ecosystems, and the environment. The complexity of air pollution arises from its sources, types, and complex atmospheric interactions that lead to various environmental and health impacts. Different methods have been developed for air quality prediction, which are mainly classified into statistical and machine learning methods. These methods utilize historical data and environmental parameters to forecast air quality indices and pollutant concentrations. The integration of these techniques has shown promise in improving prediction accuracy and providing timely warnings for air quality management. In this study, we propose to use a CNN-LSTM (Convolutional Neural Network Long Short-Term Memory Networks) hybrid model to analyze and forecast air quality in Hoa Xa Industrial Park, Nam Dinh Province. The results demonstrate that the CNN-LSTM model has good analysis and prediction and can be used to assess air quality in larger regions.

Keywords: Detailed air pollution monitoring and forecasting, CNN-LSTM, $PM_{2.5}$, Hoa Xa Industrial Park - Nam Dinh.

¹Khoa Điện tử và Kỹ thuật máy tính, Trường Đại học Kinh tế Kỹ thuật Công nghiệp

²Khoa Điện - Tự động hóa, Trường Đại học Kinh tế Kỹ thuật Công nghiệp

*Email: nttan@uneti.edu.vn

Ngày nhận bài: 21/4/2025

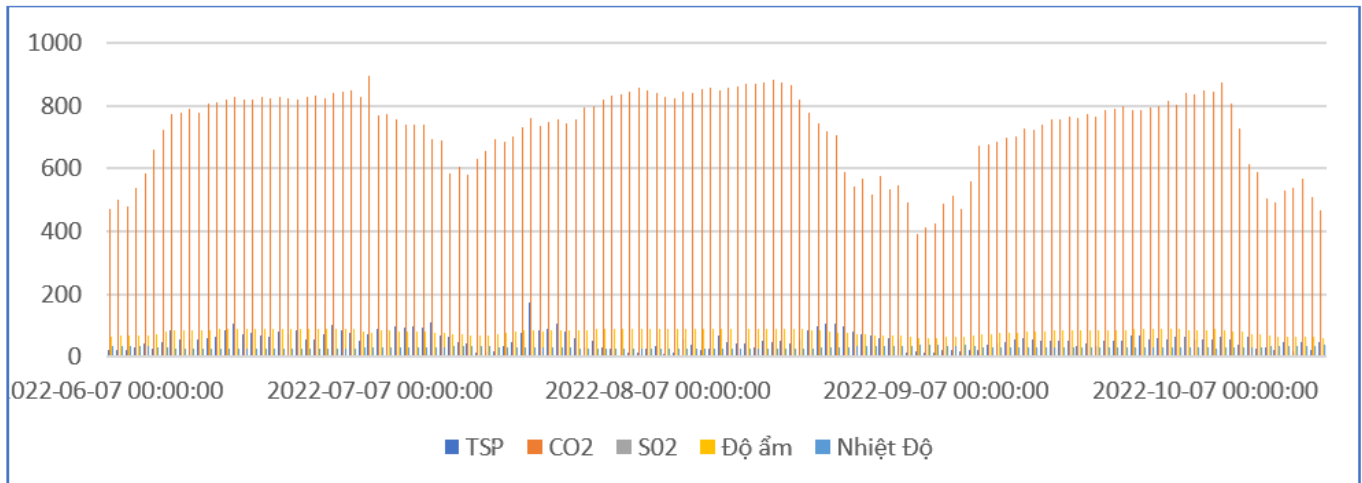
Ngày nhận bài sửa sau phản biện: 02/6/2025

Ngày chấp nhận đăng: 28/7/2025

1. GIỚI THIỆU

Biến đổi khí hậu và sự biến thiên khí hậu được dự báo sẽ gây ra những vấn đề nghiêm trọng đối với hệ sinh thái (ví dụ, làm tăng ô nhiễm không khí trong tương lai) [1, 2].

Quá trình công nghiệp hóa và việc sử dụng ồ ạt nhiên liệu hóa thạch đã làm gia tăng lượng khí nhà kính, dẫn đến gia tăng tác động của biến đổi khí hậu [3, 4]. Điều này đã gây ra ảnh hưởng đáng kể đến môi trường sống tại nhiều khu



Hình 1. Bộ dữ liệu quan trắc AQM-Air-UNETI

vực trên thế giới [5]. Những tác động này khác nhau tùy theo khu vực, loại hình ảnh hưởng và mức độ quan trọng của chúng. Các yếu tố khí hậu đã tác động trực tiếp và gián tiếp đến cả con người và môi trường sống của họ trong các giai đoạn khác nhau [6].

Chất lượng không khí được đánh giá dựa trên sáu thông số chính: tổng bụi lơ lửng (TSP), bụi mịn (PM_{10} , $PM_{2.5}$), carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), ozone (O_3) và chì (Pb) [7]. Trong đó, $PM_{2.5}$ đóng vai trò quan trọng nhất do tác động nghiêm trọng đến sức khỏe con người [8, 9]. $PM_{2.5}$ bao gồm các hạt vật chất có đường kính khí động học $\leq 2,5\mu m$, có thể di chuyển xa khỏi nguồn phát thải và bị ảnh hưởng bởi các yếu tố lý-hóa, khí tượng và địa hình. Bụi $PM_{2.5}$ có thể hình thành từ vật chất sơ cấp hoặc thông qua quá trình chuyển hóa khí thành hạt, tạo ra sol khí thứ cấp chứa carbon, sulfur, nitrogen và kim loại nặng. Do kích thước siêu nhỏ, $PM_{2.5}$ dễ dàng xâm nhập sâu vào phổi, đi vào hệ tuần hoàn và gây ảnh hưởng tiêu cực đến hệ hô hấp, tim mạch [10]. Theo Tổ chức Y tế Thế giới (WHO), ô nhiễm không khí gây ra khoảng 4,2 triệu ca tử vong mỗi năm [11]. Các nghiên cứu về $PM_{2.5}$ tập trung vào nguồn phát thải, đặc tính vật lý và tác động sinh học của chúng [2, 3]. Đặc biệt, kích thước hạt càng nhỏ thì mức độ ảnh hưởng càng cao do khả năng xâm nhập sâu hơn vào hệ hô hấp [11]. Nồng độ bụi mịn PM và CO_2 được coi là yếu tố khí hậu quan trọng nhất, ảnh hưởng đến sự tăng trưởng, phát triển và năng suất cây trồng [7]. Ngoài ra, hệ thống nhà ở cũng được xây dựng để thích ứng với các yếu tố khí hậu [8, 9]. Tại Việt Nam, các nghiên cứu về ô nhiễm không khí chủ yếu tập trung vào Hà Nội và TP. Hồ Chí Minh, hai khu vực có mức độ ô nhiễm cao nhất cả nước [12]. Vì vậy, việc giám sát và dự báo $PM_{2.5}$ là cần thiết để

đề xuất các giải pháp kiểm soát ô nhiễm và bảo vệ sức khỏe cộng đồng

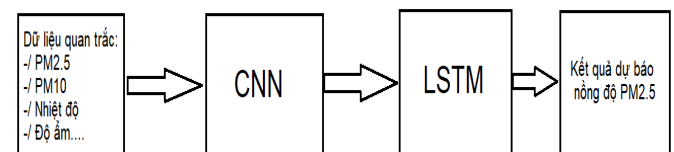
2. KHU VỰC NGHIÊN CỨU VÀ BỘ DỮ LIỆU

Khu công nghiệp Hòa Xá là trung tâm công nghiệp trọng điểm của tỉnh Nam Định, tập trung nhiều nhà máy sản xuất, chế biến với mật độ cao [13]. Hoạt động công nghiệp và giao thông là nguồn phát thải chính, làm tăng nồng độ $PM_{2.5}$. Với kích thước nhỏ hơn $2,5\mu m$, $PM_{2.5}$ có thể dễ dàng xâm nhập vào hệ hô hấp, gây ra các bệnh lý nghiêm trọng.

Bộ dữ liệu lịch sử về dữ liệu nồng độ bụi $PM_{2.5}$, $PM_{1.0}$, CO_2 , NO , nhiệt độ, độ ẩm được thu thập trong 06 tháng (6/2022 - 11/2022 với 720 dữ liệu quan trắc hàng ngày) đã được sử dụng để xây dựng và đánh giá mô hình [14].

3. PHƯƠNG PHÁP ĐỀ XUẤT

3.1. Phương pháp đề xuất



Hình 2. Sơ đồ khối của hệ thống dự báo

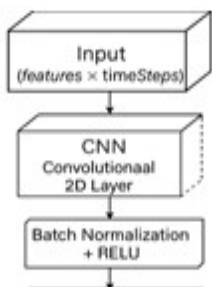
Nghiên cứu này đề xuất ước tính ô nhiễm không khí chi tiết trên toàn KCN Hòa Xá và dự báo tại trạm bằng CNN-LSTM cho các tác vụ phân loại dữ liệu quan trắc qua trạm AQM-Air-Uneti bằng cách kết hợp mạng nơ-ron tích chập 2 chiều (CNN) với lớp bộ nhớ dài hạn ngắn (LSTM). CNN xử lý dữ liệu chuỗi bằng cách áp dụng các bộ lọc tích chập trượt vào đầu vào. CNN có thể học các đặc điểm từ cả chiều không gian và chiều thời gian. Mạng LSTM xử lý dữ liệu chuỗi bằng cách lặp qua các bước thời gian và học các mối phụ thuộc dài hạn giữa các bước thời gian. Mạng

CNN-LSTM sử dụng các lớp tích chập và LSTM để học từ dữ liệu đào tạo[16]. Sơ đồ khối của hệ thống được giới thiệu trên hình 2.

Cấu trúc CNN-LSTM là một kiến trúc học sâu lai kết hợp Mạng nơ-ron Convolutional (CNN) và mạng Bộ nhớ ngắn hạn dài (LSTM) để tận dụng thế mạnh tương ứng của chúng trong việc xử lý dữ liệu không gian và thời gian. Cấu trúc này đặc biệt hiệu quả trong các tác vụ yêu cầu cả trích xuất tính năng từ dữ liệu thô và phân tích dữ liệu tuần tự, chẳng hạn như nhận dạng hoạt động, phân loại hình ảnh và xử lý tín hiệu. Thành phần CNN thành thạo trong việc nắm bắt các phân cấp không gian trong dữ liệu, trong khi thành phần LSTM vượt trội trong việc mô hình hóa các phụ thuộc thời gian. Sự kết hợp này cho phép phân tích toàn diện các bộ dữ liệu phức tạp, làm cho CNN-LSTM trở thành một công cụ linh hoạt trong các lĩnh vực khác nhau. CNN được sử dụng để trích xuất các đặc điểm không gian từ dữ liệu đầu vào, chẳng hạn như hình ảnh hoặc kết quả đọc cảm biến. Chúng đặc biệt hiệu quả trong việc xác định các mẫu và cấu trúc trong dữ liệu do khả năng trích xuất tính năng phân cấp của chúng. LSTM được sử dụng để nắm bắt các phụ thuộc thời gian trong dữ liệu. Chúng được thiết kế để xử lý dữ liệu tuần tự và có khả năng học các phụ thuộc lâu dài, điều này rất quan trọng đối với các nhiệm vụ liên quan đến dữ liệu chuỗi thời gian.

3.2. Kiến trúc CNN

Kiến trúc CNN-LSTM bao gồm ba phần chính: (1) tầng đầu vào nhận chuỗi dữ liệu nồng độ bụi PM_{2.5} và các yếu tố thời tiết như nhiệt độ, độ ẩm, lượng mưa, tốc độ gió; (2) tầng CNN giúp trích xuất đặc trưng cục bộ trong chuỗi; (3) tầng LSTM học các phụ thuộc thời gian dài; cuối cùng là tầng Fully Connected để tạo ra đầu ra là nồng độ PM_{2.5} dự báo cho nhiều bước thời gian trong tương lai.



Hình 3. Sơ đồ khối của hệ thống

Bảng 1. Giải thích chi tiết từng thành phần

Thành phần	Vai trò
Input Layer	Nhận dữ liệu đầu vào dạng [timeSteps × 1 × numFeatures]

Convolution2D Layer	Trích xuất đặc trưng cục bộ theo thời gian trong từng cửa sổ (ví dụ 3 ngày liên tiếp)
BatchNorm + ReLU	Ổn định phân phối và tạo phi tuyến tính
Flatten	Chuyển feature map 2D thành vector 1D để đưa vào LSTM
LSTM Layer	Học mối quan hệ tuần tự (temporal patterns), ví dụ xu hướng nhiệt độ
FullyConnected Layer	Dự báo n bước tiếp theo (multi-step)
Regression Layer	So sánh đầu ra liên tục với nhãn thực tế

Hệ thống mạng nơ-ron CNN-LSTM cho dự báo thời tiết được thiết kế dựa trên một chuỗi các lớp (layers), mỗi lớp thực hiện một chức năng cụ thể, lần lượt xử lý dữ liệu theo chiều sâu. Các lớp được sắp xếp theo thứ tự sau:

a) Lớp đầu vào (Input Layer)

Đầu vào của mạng là một tensor 3 chiều có kích thước [features × timeSteps × 1], trong đó:

- features: số đặc trưng đầu vào (ví dụ: 4 đặc trưng thời tiết)
- timeSteps: số bước thời gian trong một chuỗi (ví dụ: 10 ngày trước)
- 1: chiều không gian thứ ba để tương thích với CNN

Ví dụ, nếu ta có 4 đặc trưng và chọn cửa sổ thời gian là 10, mỗi mẫu đầu vào sẽ có kích thước [4 × 10], được định hình lại để phù hợp với kiến trúc CNN.

b) Lớp tích chập 2D (2D Convolutional Layer)

Lớp tích chập 2D là trung tâm của CNN, có nhiệm vụ trích xuất các đặc trưng cục bộ (local patterns) từ chuỗi thời gian. Tích chập được thực hiện theo chiều thời gian, giúp mạng học được các xu hướng ngắn hạn trong dữ liệu.

- Bộ lọc sử dụng có kích thước [3 × 1], nghĩa là tích chập theo chiều thời gian (3 bước liên tiếp).
- Số lượng bộ lọc: 32, tức là sau lớp này ta sẽ thu được 32 map đặc trưng.
- Padding 'same': đảm bảo đầu ra giữ nguyên chiều dài thời gian.

Kết quả của lớp này là một tensor kích thước [timeSteps × 1 × 32].

c) Lớp chuẩn hóa và phi tuyến (Batch Normalization + ReLU)

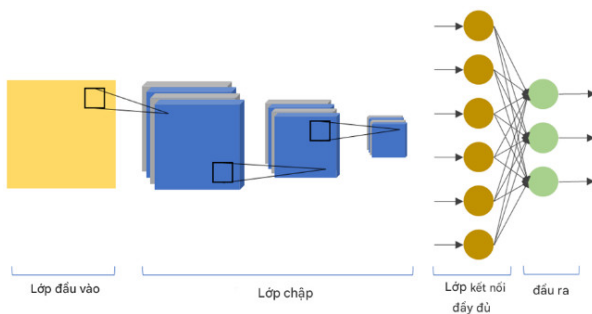
- **Batch Normalization:** ổn định phân phối dữ liệu qua từng mini-batch, giúp tăng tốc độ huấn luyện và giảm hiện tượng overfitting.

• **ReLU (Rectified Linear Unit):** kích hoạt phi tuyến, giữ lại các giá trị dương, loại bỏ các giá trị âm, giúp mạng học được các đặc trưng phi tuyến hiệu quả hơn.

d) Lớp làm phẳng (Flatten Layer)

Sau khi trích xuất đặc trưng không gian từ CNN, kết quả cần được "làm phẳng" thành vector 1 chiều để phù hợp với đầu vào của LSTM. Lớp flatten chuyển tensor $[timeSteps \times 1 \times numFilters]$ thành $[timeSteps \times numFilters]$.

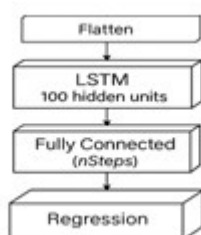
Ví dụ: $[10 \times 1 \times 32] \rightarrow [10 \times 32]$



Hình 4. Sơ đồ kiến trúc mạng CNN

CNN đặc biệt nổi tiếng với khả năng xử lý dữ liệu với cấu trúc liên kết giống lưới, chẳng hạn như hình ảnh, và đã được áp dụng thành công trong các lĩnh vực khác nhau, từ nhận dạng hình ảnh đến xử lý ngôn ngữ tự nhiên và quản lý mạng xe cộ. Ưu điểm của CNN bắt nguồn từ kiến trúc độc đáo của chúng, bao gồm các lớp chập cho phép trích xuất tính năng tự động và chia sẻ trọng lượng, dẫn đến các quy trình học tập hiệu quả và hiệu quả.

3.3. Kiến trúc LSTM



Hình 5. Sơ đồ kiến trúc mạng LSTM

LSTM là mạng hồi tiếp đặc biệt, có khả năng ghi nhớ các trạng thái lâu dài trong chuỗi, rất phù hợp với dữ liệu tuần tự như chuỗi thời gian.

• Số nút ẩn (hidden units): 100

• OutputMode: 'last' – chỉ lấy trạng thái cuối cùng để làm đầu ra, dùng cho bài toán nhiều bước dự báo (multi-step forecasting).

LSTM xử lý dữ liệu theo trật tự thời gian và học được sự phụ thuộc giữa các đặc trưng tại các thời điểm khác nhau.

Lớp kết nối đầy đủ (Fully Connected Layer)

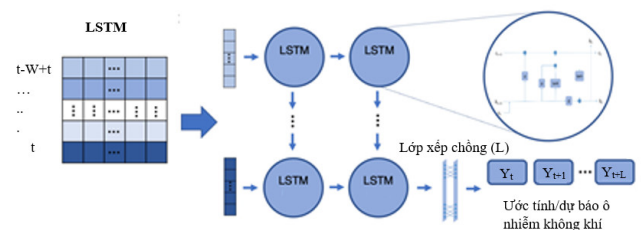
Lớp Fully Connected đóng vai trò ánh xạ từ không gian đặc trưng (vector đầu ra của LSTM) sang không gian mục tiêu – trong trường hợp này là nSteps bước dự báo nhiệt độ tiếp theo.

• Số node: bằng số bước thời gian muốn dự báo (nSteps), ví dụ 5 bước.

• Kết quả là một vector $[1 \times nSteps]$.

Lớp hồi quy (Regression Layer)

Lớp cuối cùng là lớp hồi quy tuyến tính, dùng để tính toán sai số giữa đầu ra mạng và giá trị thực tế. Hàm mất mát thường dùng là Mean Squared Error (MSE), đo lường mức độ khác biệt giữa dự đoán và giá trị thật.



Hình 6. Cấu trúc của một khung LSTM

Một khối bộ nhớ LSTM được định nghĩa trong Phương trình (1). Trạng thái ẩn cuối cùng của thành phần LSTM được sử dụng để ước tính và dự báo ô nhiễm không khí, thông qua một lớp kết nối đầy đủ (fully connected layer). Đối với ước tính ô nhiễm không khí chi tiết ở cấp độ toàn KCN Hòa Xá, một đầu ra được tạo ra cho một ma trận đầu vào, đại diện cho mức độ ô nhiễm không khí trong một lưới tại giờ hiện tại. Đối với dự báo ô nhiễm không khí tại trạm giám sát, nhiều đầu ra được tạo ra cho một ma trận đầu vào, đại diện cho mức độ ô nhiễm không khí theo giờ tại trạm giám sát trong các giờ tiếp theo.

$$f_t = \sigma(W_f x_t + U_f h_{t-1} W_f + \text{bias}_f) \quad (a)$$

$$g_t = \tanh(W_g x_t + h_{t-1} W_g + \text{bias}_g) \quad (b)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} W_i + \text{bias}_i) \quad (c) \quad (1)$$

$$c_t = f_t \odot c_{t-1} + g_t \odot i_t \quad (d)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} W_o + \text{bias}_o) \quad (e)$$

$$h_t = \tanh(c_t) \odot o_t \quad (g)$$

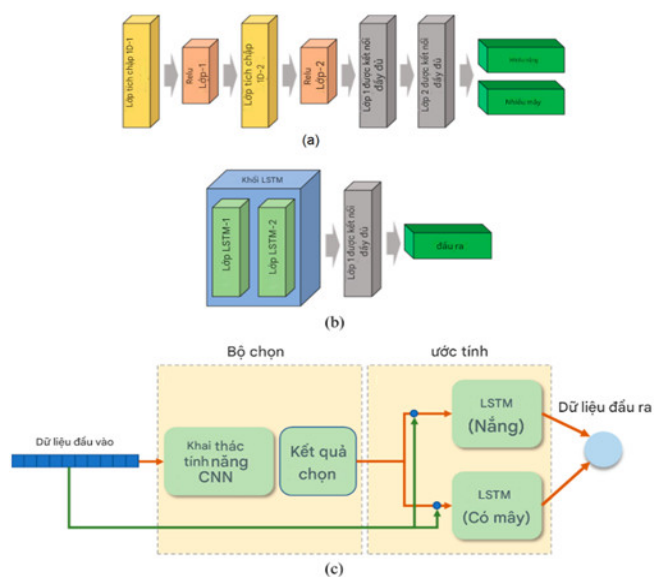
Trong đó: t là bước thời gian; W và U là ma trận tham số của mạng; x_t là vector đặc trưng được trích xuất tại bước thời gian t ; h_t là trạng thái ẩn tại bước thời gian t ; i_t là cổng đầu vào tại bước thời gian t ; f_t là cổng quên tại bước thời gian t ; o_t là cổng đầu ra tại bước thời gian t ; σ là hàm sigmoid; c_t tương ứng với ô nhớ tại bước thời gian t ; $\odot$ biểu thị phép nhân phần tử.

Các phương trình 1a, 1g được áp dụng cho mỗi cổng. Phương trình 1a cho thấy quá trình của cổng quên. Dữ liệu đầu vào được cập nhật thông qua các phương trình 1b-1d. Dữ liệu trải qua một hoạt động, như trong phương trình 1e, trong cổng đầu ra cuối cùng và được cập nhật theo phương trình cuối cùng 1g. Ở đây, σ là hàm sigmoid, w là ma trận trọng số, h và c_t lần lượt biểu diễn trạng thái ẩn và trạng thái ô.

Quy trình nghiên cứu này được chia thành hai giai đoạn: tiền xử lý dữ liệu CNN và dự báo LSTM

3.4. Mô hình CNN-LSTM lai

Sự ra đời của mô hình CNN-LSTM lai thể hiện một tiến bộ đáng kể trong lĩnh vực học sâu, kết hợp các thế mạnh của Mạng nơ-ron Convolutional (CNN) và mạng Bộ nhớ ngắn hạn dài (LSTM). Cách tiếp cận này tận dụng khả năng trích xuất đặc điểm không gian của CNN với khả năng học trình tự thời gian của LSTM, làm cho nó có hiệu quả cao cho nhiều ứng dụng khác nhau. Tính linh hoạt của mô hình được thể hiện trên nhiều lĩnh vực, bao gồm phát hiện hình ảnh, dự báo lưu lượng truy cập, tương tác giữa con người và máy tính,...



Hình 7. Mô hình CNN-LSTM lai

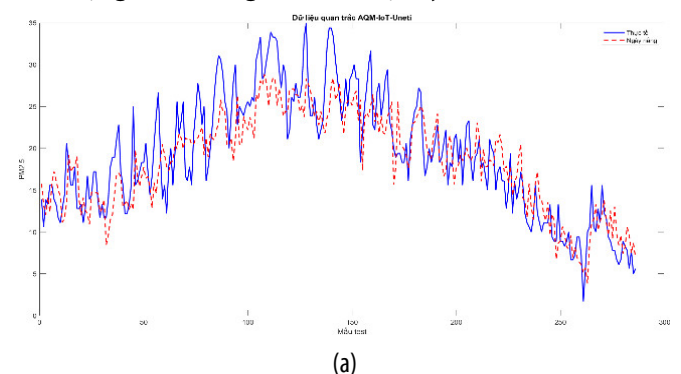
Hình 7a cho thấy kiến trúc CNN được sử dụng để phân loại các điều kiện thời tiết. Kiến trúc này bao gồm hai lớp tích chập với lớp relu và hai lớp được kết nối đầy đủ. Mô hình CNN được đào tạo để phân loại các ngày nắng và nhiều mây dựa trên mẫu dữ liệu về các điều kiện thời tiết. Lớp cuối cùng đưa ra kết quả của các điều kiện thời tiết. Trong quy trình bên trong của mạng nơ-ron, sự khác biệt giữa các giá trị dự đoán và nhãn được tính toán bằng cách sử dụng hàm mất mát. Sau đó, nó sử dụng thuật toán lan truyền ngược để giảm thiểu hàm mất mát sao cho các giá

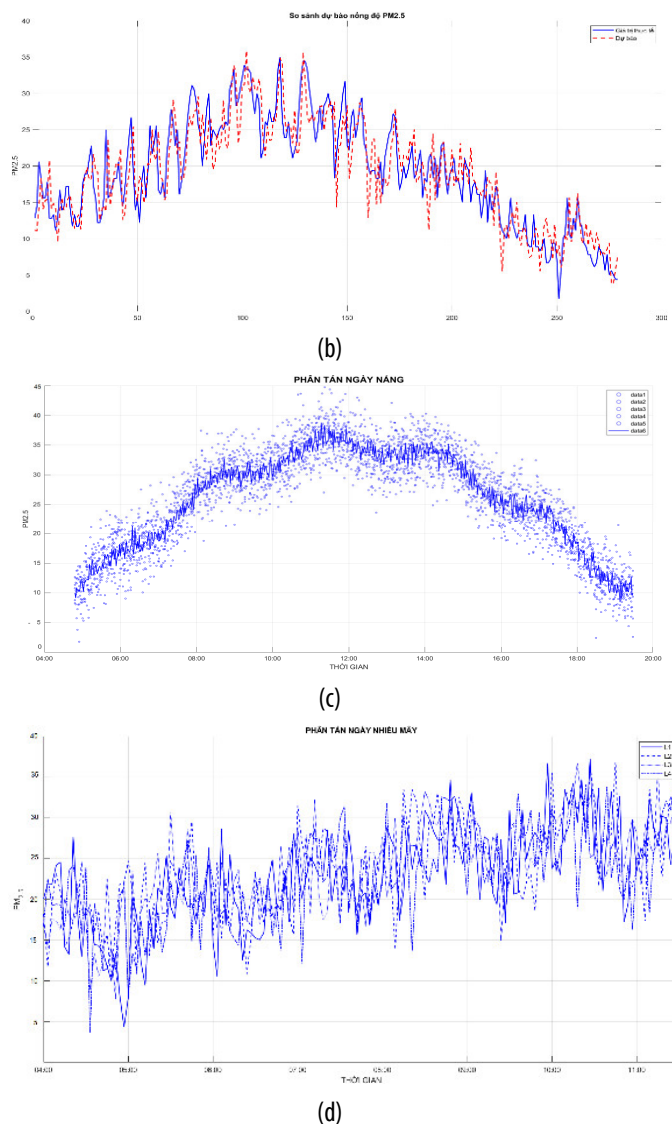
trị dự đoán càng gần với nhãn càng tốt. Trong quá trình đào tạo mạng nơ-ron, các tham số, bao gồm trọng số và độ lệch, được tinh chỉnh và cập nhật để so sánh các giá trị dự đoán với nhãn để tạo ra các dự đoán tốt hơn ở mọi kỳ nguyên. Bộ tối ưu hóa đã sử dụng Adam và đặt tốc độ học $= 0,01$, hệ số cho động lượng chính $\beta_1 = 0,9$, hệ số cho động lượng thứ cấp $\beta_2 = 0,999$ và epsilon $= 10^{-8}$. Khởi tạo trọng số sử dụng bộ khởi tạo thống nhất kaiming và áp dụng hàm entropy chéo làm hàm mất mát. Hình 7b cho thấy kiến trúc LSTM được sử dụng để dự báo nồng độ $PM_{2.5}$. Kiến trúc này bao gồm hai lớp LSTM và một lớp của lớp được kết nối đầy đủ. Lớp cuối cùng đưa ra nồng độ $PM_{2.5}$ dự đoán [19]. Hai mô hình LSTM có cùng cấu trúc được đào tạo trên dữ liệu ngày nắng và ngày nhiều mây. Trong trường hợp của mô hình LSTM, các cổng đầu vào, quên và đầu ra và trạng thái ô giúp mô hình xác định giá trị nào cần bảo toàn để duy trì bộ nhớ dài hạn trong quá trình tính toán. Các siêu tham số được sử dụng để đào tạo mạng như sau. Bộ tối ưu hóa sử dụng Adam và đặt tốc độ học $= 0,001$, hệ số cho động lượng chính $\beta_1 = 0,9$, hệ số cho động lượng phụ $\beta_2 = 0,999$ và epsilon $= 10^{-8}$. Khởi tạo trọng số sử dụng bộ khởi tạo thống nhất kaiming và áp dụng hàm MSE (Lỗi bình phương trung bình) làm hàm mất mát [19]. Hình 8c cho thấy cấu trúc tổng thể của mô hình lai CNN-LSTM. Chọn một mô hình LSTM theo đầu ra của CNN và dự đoán nồng độ $PM_{2.5}$ bằng cách sử dụng dữ liệu chuỗi thời gian.

4. THÍ NGHIỆM

4.1. Thu thập và tiền xử lý dữ liệu

Bài báo này đề xuất một mô hình CNN-LSTM lai để học các mẫu nồng độ $PM_{2.5}$ theo loại thời tiết, giúp giảm độ phức tạp và tính biến thiên của việc khớp dữ liệu để cải thiện độ chính xác của dự đoán. Chuỗi dữ liệu ban đầu cần được tiền xử lý bao gồm loại bỏ giá trị thiếu, chuẩn hóa về cùng thang đo và chuyển đổi thành chuỗi con (sliding window) với độ dài cố định. Các chuỗi này được định dạng thành tensor 4 chiều phù hợp với yêu cầu đầu vào của mạng CNN trong Matlab hoặc Python [20].

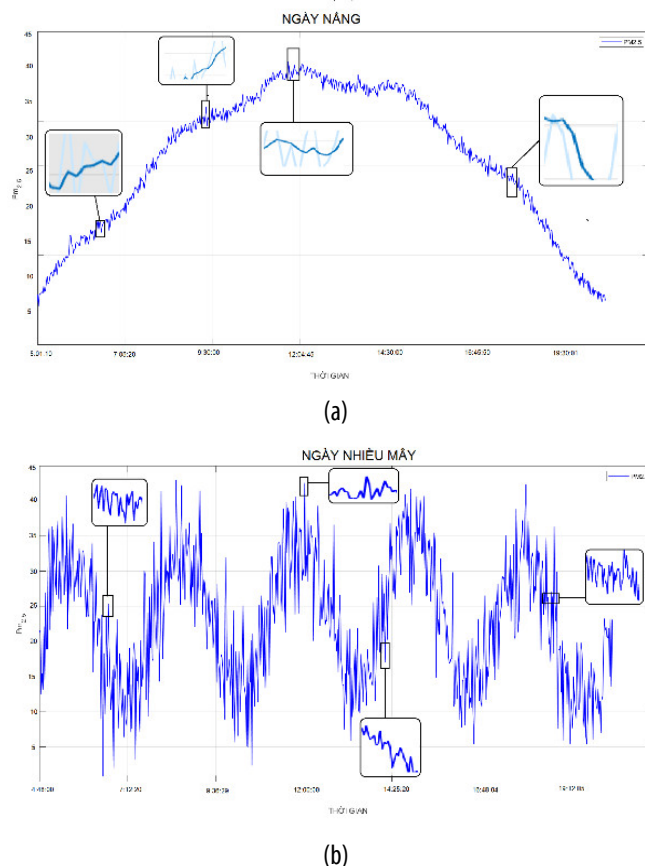




Hình 8. Biểu đồ nồng độ PM_{2.5}: (a) ngày nắng, (b) ngày nhiều mây, (c) biểu đồ phân tán ngày nắng, (d) biểu đồ phân tán ngày nhiều mây

Nồng độ PM_{2.5} khác nhau vào ngày nắng và ngày nhiều mây, như thể hiện trong hình 8a,b tương ứng. Nồng độ PM_{2.5} vào ngày nhiều mây thấp và không ổn định cao. Vào ngày nắng, biểu đồ có phân phối dữ liệu hình bán nguyệt bị ảnh hưởng bởi mặt trời mọc và lặn; nồng độ PM_{2.5} cao nhất vào khoảng giữa trưa, như được chỉ ra bởi độ cao lớn. Vào ngày nhiều mây, sự thay đổi của nồng độ PM_{2.5} rất nghiêm trọng do những thay đổi về bức xạ mặt trời. Hình 8c,d cho thấy dữ liệu nồng độ PM_{2.5} vào những ngày quang đãng và nhiều mây được chồng lên nhau và thể hiện dưới dạng biểu đồ phân tán. Trong hình 8c, có thể thấy rằng mô hình nồng độ bụi mịn PM_{2.5} vào ngày quang đãng được tạo ra đều đặn trong một khu vực nhất định. Trong hình này, có những trường hợp dữ liệu nồng độ PM_{2.5} nằm ngoài phạm vi một điểm. Đây là giá trị ngoại lệ do lỗi thiết bị hoặc nhiễu. Trong hình 8d, có thể

thấy rằng mô hình nồng độ PM_{2.5} vào thời điểm những ngày nhiều mây có sự thay đổi cao theo các đặc điểm phi tuyến tính.



Hình 9. Mô hình phân đoạn nồng độ bụi PM_{2.5}

Hình 9a,b cho thấy các mẫu PM_{2.5} theo từng phần trong những ngày nắng và nhiều mây. Vào ngày nắng, nồng độ PM_{2.5} tăng dần hoặc giảm dần, nhưng sự thay đổi đột ngột thay đổi vào ngày nhiều mây. Vì các mẫu dữ liệu như vậy cũng ảnh hưởng đến các mẫu chuỗi thời gian liên tục, nên việc sử dụng mô hình LSTM để đào tạo có thể làm phức tạp quá trình hội tụ dữ liệu. Do đó, các điều kiện thời tiết hiện tại trước tiên được phân loại theo các mẫu và nồng độ PM_{2.5} được dự báo bằng mô hình LSTM riêng lẻ theo kết quả phân loại. Các ô trong hình 9 cho thấy dạng mẫu chi tiết của dữ liệu PM_{2.5} đã thu thập được. Dữ liệu của ngày nắng được tạo ra với chu kỳ đều đặn, trong khi dữ liệu của ngày nhiều mây có phạm vi biến động lớn và không cho thấy tính chu kỳ.

Như đã trình bày ở phần 2, Bộ dữ liệu lịch sử về dữ liệu nồng độ bụi PM_{2.5}, nhiệt độ, độ ẩm... được thu thập trong 06 tháng (06/2022 - 11/2022 với 720 dữ liệu quan trắc hàng ngày) đã được sử dụng. Dữ liệu được xác định là tương ứng với điều kiện thời tiết nắng và nhiều mây của Tổng cục khí tượng thủy văn Việt Nam đã được sử dụng.

Tổng cục khí tượng thủy văn Việt Nam định nghĩa điều kiện thời tiết theo 11 bước. Trong nghiên cứu này, các bước 1 - 5 được phân loại là nắng, trong khi các bước 6 - 11 được phân loại là nhiều mây [15].

Bảng 2. Sơ đồ dữ liệu thu thập được ở KCN Hòa Xá - Nam Định

	Kiểu dữ liệu	Mặc định
Ngày	Ngày	Vô giá trị
Thời gian	Thời gian	Vô giá trị
Nồng độ PM _{2.5}	Double (22, 0)	"0"

Dữ liệu và thời gian trong bảng 2 biểu diễn thời điểm dữ liệu được thu thập; nồng độ PM_{2.5} biểu thị tổng nồng độ PM_{2.5}. Dữ liệu của 182 ngày đã được sử dụng cho thí nghiệm. Bộ dữ liệu bao gồm 146 và 36 điểm dữ liệu ngày nắng và ngày nhiều mây. Sự mất cân bằng dữ liệu có thể xảy ra vì số lượng dữ liệu ngày nắng lớn hơn 1,88 lần so với dữ liệu ngày nhiều mây. Do đó, 120 dữ liệu từ mỗi lớp đã được sử dụng để đào tạo mạng, trong khi dữ liệu còn lại được sử dụng để thử nghiệm.

Các yếu tố bất thường trong dữ liệu đầu vào của mô hình dự báo có thể gây ra lỗi dự báo cao. Việc xử lý trước dữ liệu đầu vào có thể cải thiện độ chính xác của mô hình bằng cách giảm chi phí tính toán và giải quyết vấn đề đào tạo không phù hợp. Các tập dữ liệu có thể chứa các giá trị ngoại lệ do sự cố thiết bị năng lượng mặt trời, lỗi hệ thống thu thập và sự cố hệ thống phần mềm. Do đó, để loại bỏ các giá trị ngoại lệ, các tập dữ liệu được làm sạch bằng cách tính toán phân vị 75%

$$Q_3 = t + \frac{\frac{3n}{4} - cf_{q-1}}{n_q} \quad (2)$$

trong đó:

+ t , l , n_q lần lượt là đầu mút trái, độ dài, tần số của nhóm q .

+ cf_{q-1} là tần số tích lũy của nhóm $q - 1$.

Dữ liệu nồng độ PM_{2.5} tích lũy trong bảng 1 được xử lý trước bằng cách sử dụng phạm vi liên tứ phân vị (IQR). Nếu giá trị dữ liệu cao hơn giá trị được cung cấp bởi phương trình (7), thì giá trị đó được xác định là giá trị ngoại lệ và bị loại bỏ.

Các giá trị bị thiếu trong dữ liệu thu thập được có thể xảy ra do sự cố ở bộ biến tần hoặc thiết bị thu thập dữ liệu. Vì mô hình đề xuất dự đoán dữ liệu của một thời điểm trong tương lai bằng cách sử dụng dữ liệu chuỗi thời gian liên tục, các giá trị bị thiếu tại x_t có thể ảnh hưởng đến hiệu suất của mô hình. Do đó, nội suy thời

gian sử dụng các giá trị tại thời điểm x_{t-1} và x_{t+1} đã được áp dụng.

Thang số của dữ liệu PM_{2.5} được sử dụng trong nghiên cứu này là đáng kể. Ví dụ, PM_{2.5} ít hơn 10 gần 06:00 sáng khi lượng bức xạ mặt trời thấp và nhiều hơn 45 giữa 12:00 - 14:00 khi lượng bức xạ mặt trời cao nhất. Sự khác biệt đáng kể giữa các giá trị tối thiểu và tối đa có thể ảnh hưởng đến tốc độ đào tạo và độ chính xác hội tụ của mạng. Do đó, sử dụng phương trình (3), chuẩn hóa đã được thực hiện để cải thiện tốc độ tính toán và tăng tốc độ hội tụ đào tạo mạng.

$$z = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (3)$$

Ở đây, x là dữ liệu thực tế và z biểu thị giá trị chuẩn hóa.

Dữ liệu được sử dụng trong nghiên cứu này là dữ liệu nồng độ bụi PM_{2.5} 15 phút/lần, thu được bằng cách chia dữ liệu nồng độ bụi PM_{2.5} tích lũy cho thời gian. Do đó, nghiên cứu này nhằm mục đích dự đoán nồng độ bụi PM_{2.5} chỉ dựa trên các mẫu thu thập từ trạm quan trắc. Mạng nhận được 20 dữ liệu chuỗi thời gian và xuất ra một dữ liệu. Các siêu tham số được sử dụng để đào tạo mạng như sau. Lỗi bình phương trung bình (MSE) được sử dụng làm hàm mất mát và trình tối ưu hóa Adam được sử dụng để tối ưu hóa. Tốc độ học là 0,01 được áp dụng để tối ưu hóa mạng.

4.2. Tham số đánh giá

Trong nghiên cứu này, một mô hình lai kết hợp CNN và LSTM được thiết lập. Các chỉ số chung được sử dụng để đánh giá hiệu suất của mô hình dự báo dữ liệu chuỗi thời gian bao gồm sai số tuyệt đối trung bình (MAE), sai số bình phương trung bình căn bậc hai (RMSE) và sai số phần trăm tuyệt đối trung bình (MAPE). Lấy N là số lượng dữ liệu thử nghiệm, x_{pred} là giá trị được dự đoán bởi thuật toán đề xuất và x_{act} là giá trị thực tế trong các chỉ số định lượng.

MAE đo lường lỗi giữa giá trị dự đoán và giá trị thực tế và có thể thay đổi tùy thuộc vào phép đo các biến liên tục, như thể hiện trong phương trình (4). Giá trị nhỏ hơn cho thấy độ chính xác cao hơn.

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_{\text{pred}} - x_{\text{act}}| \quad (4)$$

RMSE, được định nghĩa như trong phương trình (5), đo lường sự khác biệt giữa các giá trị dự đoán và giá trị thực tế; đây là thước đo độ lệch giữa các giá trị dự đoán và giá trị thực tế. RMSE gần 0 hơn cho thấy hiệu suất tốt hơn. Sự

khác biệt giữa RMSE và MAE là RMSE nhạy cảm với các giá trị ngoại lai, khiến nó dễ bị độ lệch đáng kể hơn. Vì lỗi được bình phương, lỗi lớn hơn dẫn đến trọng số được phản ánh ở mức độ lớn hơn.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{\text{predicted}} - x_{\text{real}})^2} \quad (5)$$

MAPE, được đưa ra bởi phương trình (6), chỉ ra mức độ lỗi trong giá trị dự đoán. Một giá trị gần 0 hơn có thể được hiểu là hiệu suất mô hình dự báo nổi bật hơn. MAPE mạnh mẽ với các giá trị ngoại lai, nhưng trong trường hợp này, việc kiểm tra lỗi trực quan khó khăn hơn so với trường hợp MAE.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_{\text{predicted}} - x_{\text{real}}}{x_{\text{real}}} \right| \times 100\% \quad (6)$$

Nồng độ PM_{2.5} dự báo sử dụng LSTM kéo theo lỗi, mức độ lỗi thay đổi tùy thuộc vào mật độ của dữ liệu nồng độ PM_{2.5} dự báo. Lỗi giữa giá trị dự báo và giá trị thực tế nhỏ hơn khi mật độ của giá trị dự báo dày đặc. Lỗi trở nên đáng kể hơn khi mật độ thưa thớt. Độ chính xác của mô hình dự báo cũng được xác định bởi mật độ. Như thể hiện trong phương trình (7), R² đo lường cường độ tương quan giữa giá trị dự đoán và giá trị thực tế. R² có phạm vi 0 ≤ R² ≤ 1, giá trị gần 0 biểu thị độ chính xác rất thấp và giá trị gần 1 biểu thị độ chính xác cao hơn của mô hình dự báo.

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_{\text{predicted}} - x_{\text{real}})^2}{\sum_{i=1}^n (x_{\text{real}} - \bar{x}_{\text{real}})^2} \quad (7)$$

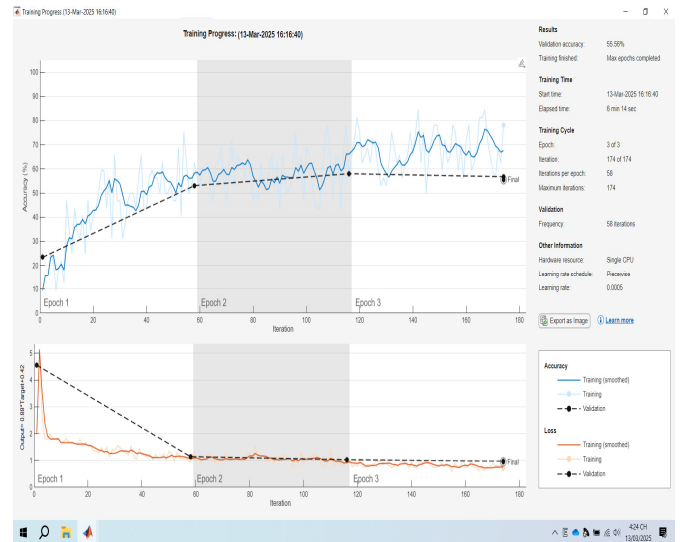
Ở đây $\bar{x}_{\text{real}}$ là giá trị trung bình của dữ liệu nồng độ bụi thực tế.

4.3. Kết quả thực nghiệm

Để đánh giá hiệu quả của mô hình sâu được đề xuất, chúng tôi đã thiết kế các thí nghiệm sau bằng cách sử dụng các tập dữ liệu nồng độ bụi PM_{2.5} đã thu thập. Mô hình lại được đề xuất đã được đào tạo và thử nghiệm bằng cách sử dụng khuôn khổ học sâu Matlab dựa trên Matlab R204a làm nền tảng phần mềm cùng với Intel(R) Core (TM) i7-6820HQ CPU @ 2.70GHz 2.70GHz [21, 22].

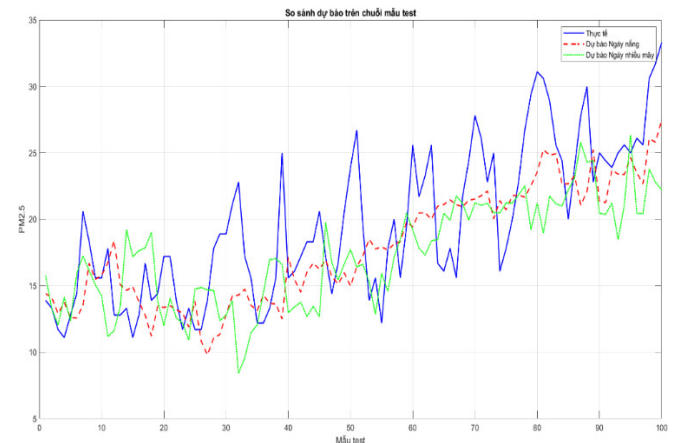
Mô hình CNN-LSTM lại được đề xuất thực hiện dự báo nồng độ PM_{2.5}, như thể hiện trong hình 10 ở bước 1, dữ liệu PM_{2.5} được thu thập từ trạm quan trắc AQM_IoT_Uneti được phân loại thành dữ liệu ngày nắng và ngày nhiều mây. Ở bước 2, dữ liệu thu thập được được xử lý trước để loại bỏ các yếu tố nhiễu, chẳng hạn như các

giá trị bị thiếu hoặc các giá trị ngoại lai ảnh hưởng đến thí nghiệm. Sau đó, chuẩn hóa được áp dụng để sử dụng dữ liệu làm đầu vào mạng. Ở bước 3, CNN và LSTM được đào tạo bằng cách sử dụng dữ liệu đào tạo. Mỗi mô hình chỉ sử dụng dữ liệu nồng độ PM_{2.5}. Ở bước 4, kết quả của mô hình dự báo được kiểm tra và xác minh bằng cách sử dụng nhiều ma trận khác nhau.



Hình 10. Mô hình CNN-LSTM chạy trên Matlab 2024a

Đánh giá định lượng và định tính được thực hiện bằng cách sử dụng dữ liệu ngày nắng và ngày nhiều mây để xác minh mô hình. Dữ liệu của cùng một tháng (ngày nắng: 28 tháng 06 năm 2022, ngày nhiều mây: 27 tháng 11 năm 2022) được sử dụng để thống nhất quá trình xác thực.



Hình 11. Kết quả dự báo nồng độ bụi PM_{2.5} ngày nắng và ngày nhiều mây

Đường màu xanh đậm hiển thị dữ liệu đo được từ thiết bị, trong khi đường xanh nhạt hiển thị dữ liệu dự đoán ngày nhiều mây và đường đỏ đứt quãng là dự báo cho ngày nắng. Đánh giá định lượng đã được thực hiện bằng cách sử dụng MAPE, RMSE và MSE để xác thực mô hình dự báo LSTM. Mô hình LSTM đã theo dõi đầy đủ xu hướng của dữ liệu quan sát để dự báo chính xác.

Bảng 3 trình bày kết quả đánh giá định lượng dữ liệu thử nghiệm cho những ngày nắng và nhiều mây. Dữ liệu ngày nắng cho thấy MAPE là 4,08, thấp hơn so với dữ liệu ngày nhiều mây là 7,12. Trong dữ liệu ngày nhiều mây, lỗi điểm đỉnh xảy ra trong biên độ dao động theo giai đoạn thu thập nồng độ bụi $PM_{2.5}$, làm tăng giá trị MAPE so với dữ liệu ngày nắng. RMSE và MAE thấp hơn trong dữ liệu ngày nhiều mây, do đó thể hiện kết quả nổi bật hơn. Giá trị R^2 đạt 0,9578 vào ngày nắng cho thấy dự đoán ngày nắng chính xác hơn ngày nhiều mây là 0,9421.

Bảng 3. Kết quả kiểm định định lượng của mô hình dự báo nồng độ $PM_{2.5}$

	MAPE	RMSE	MAE	R^2
Ngày nắng	4,08	41,87	31,00	0,9578
Ngày mây	7,12	9,09	6,87	0,9421

Ưu điểm của mô hình lai CNN- LSTM được nhấn mạnh thêm khi sử dụng các kiểu thời tiết bất lợi. Thuật toán để xuất toán thủ đẩy đủ các xu hướng trong nhiều kiểu thời tiết khác nhau để dự báo chính xác. Tuy nhiên, lỗi xảy ra tại các điểm đỉnh đối với dữ liệu nhiều mây với các kiểu không đều và biến động lớn. Tuy nhiên, thuật toán này mạnh mẽ trước những thay đổi đột ngột trong các kiểu và giai đoạn biến động, do đó tuân thủ hợp lý xu hướng để dự báo đáng tin cậy. Đánh giá định tính xác nhận hiệu quả cao và hiệu suất được cải thiện của mô hình để xuất trong việc giải quyết vấn đề dự báo nồng độ bụi $PM_{2.5}$.

Bài báo này đề xuất một mô hình lai CNN-LSTM để dự báo nồng độ bụi $PM_{2.5}$. Mô hình đề xuất khắc phục được những nhược điểm của từng mô hình riêng lẻ trong khi vẫn giữ được những ưu điểm của chúng. Vì việc đào tạo một mô hình LSTM sử dụng dữ liệu chuỗi thời gian khác nhau có thể ảnh hưởng đến sự hội tụ của mạng, nên các mô hình riêng biệt đã được xây dựng theo điều kiện thời tiết. Các kiểu thời tiết đã được phân loại bằng CNN và mô hình LSTM đã được áp dụng cho kết quả phân loại. Các kiểu nồng độ bụi $PM_{2.5}$ có thể phân biệt rõ ràng giữa những ngày nắng và những ngày nhiều mây. Mô hình dự báo đề xuất có thể phản ánh đầy đủ các biến động nồng độ bụi $PM_{2.5}$ để dự báo chính xác. Đánh giá định tính đã xác nhận rằng các tín hiệu đầu ra nồng độ bụi $PM_{2.5}$ dự báo phản ứng với các biến động và tuân theo đầy đủ xu hướng tín hiệu đầu ra nồng độ bụi $PM_{2.5}$ thực tế.

5. KẾT LUẬN

Nghiên cứu này đánh giá hiệu suất của một phương pháp dự báo và mô phỏng $PM_{2.5}$ tối đa trung bình ngày. Dữ liệu chuỗi thời gian về nồng độ $PM_{2.5}$ tối đa lịch sử của khu công nghiệp Hòa Xá - Nam Định trong sáu tháng (06/2022 - 12/2022) được sử dụng để xây dựng và đánh giá phương pháp đề xuất.

Phương pháp này bao gồm tiền xử lý dữ liệu và mô hình lai CNN-LSTM. Dữ liệu được chia thành hai tập: tập huấn luyện (70%) và tập kiểm tra (30%). Các kiểm định thống kê khác nhau đã được sử dụng để đánh giá phương pháp này. Kết quả cho thấy phương pháp có độ chính xác trong giai đoạn kiểm tra, với giá trị R^2 đạt 0,9578 ngày nắng và 0,9421 ngày mây mù (so với 1 là giá trị chuẩn). Nồng độ $PM_{2.5}$ cao vào ngày nhiều mây mù do giảm sự phân tán: Mây có thể làm giảm sự đối lưu, khiến các chất ô nhiễm không được phân tán hiệu quả. Các điều kiện khí tượng liên quan đến ngày nhiều mây (như độ ẩm cao, gió lặng) có thể tạo điều kiện cho sự tích tụ các chất ô nhiễm. Phương pháp CNN-LSTM để nội suy nồng độ $PM_{2.5}$ dựa trên dữ liệu AQM-Air _Uneti. Dự báo nồng độ $PM_{2.5}$ tại khu công nghiệp Hòa Xá - Nam Định chính xác trong khoảng 7 ngày.

Nghiên cứu này cung cấp thông tin hữu ích cho các cơ quan quản lý trong việc giám sát và kiểm soát ô nhiễm không khí tại khu công nghiệp Hòa Xá, góp phần bảo vệ sức khỏe cộng đồng. Trong các nghiên cứu tiếp theo cần thực hiện việc phân tích thành phần hóa học (bao gồm các nguyên tố và ion bám dính) của bụi $PM_{2.5}$ để xác định nguồn đóng góp và quỹ đạo di chuyển của bụi $PM_{2.5}$ bằng công cụ mô hình hóa như mô hình Hybrid Single-Particle Lagrangian Integrated Trajectory model (HYSPLIT).

TÀI LIỆU THAM KHẢO

- [1]. D. Heinrich, W. K. Stephan, K. Ulrich, "Epidemiological Analyses of the Relationship Between Environmental Pollution and Asthma," *Toxicology Letters*, 102, 307-316, 1998. [https://doi.org/10.1016/s0378-4274\(98\)00322-1](https://doi.org/10.1016/s0378-4274(98)00322-1).
- [2]. J. Nan, D. Zhe, X. Yuanqian, Y. Fei, Y. Shasha, Z. Ruiqin, T. Xiaoyan, "Characterization of PM10 and PM2.5 Source Profiles of Fugitive Dust in Zhengzhou, China," *Aerosol and Air Quality Research*, 18, 314-329, 2018. <https://doi.org/10.4209/aaqr.2017.04.0132>.
- [3]. Salman S A, Shahid S, Ismail T, Ahmed K, Wang X J, "Selection of climate models for projection of spatiotemporal changes in temperature of Iraq with uncertainties," *Atmospheric Research*, 213509-22, 2018.
- [4]. Zubaidi S L, Abdulkareem I H, Hashim K, Al-Bugharbee H, Ridha H M, Gharghan S K, AlQaim F F, Muradov M, Kot P, Al-Khaddar R, "Hybridised Artificial Neural Network Model with Slime Mould Algorithm: A Novel Methodology for Prediction of Urban Stochastic Water Demand," *Water*, 121-18, 2020.
- [5]. Jasim I A, Farhan S L, Al-Maliki L A, AL-Mamoori S K, "Climatic Treatments for Housing in the Traditional Holy Cities: A Comparison between

Najaf and Yazd Cities," in: *IOP Conference Series: Earth and Environmental Science*, IOP Publishing, 1-10, 2021.

[6]. Zubaidi S L, Kot P, Hashim K, Alkhaddar R, Abdellatif M, Muhsin Y R, "Using LARS –WG model for prediction of temperature in Columbia City, USA," in: *IOP Conference Series: Materials Science and Engineering*, 2019.

[7]. D. Heinrich, W. K. Stephan, K. Ulrich, "Epidemiological Analyses of the Relationship Between Environmental Pollution and Asthma," *Toxicology Letters*, 102, 307-316, 1998. [https://doi.org/10.1016/s0378-4274\(98\)00322-1](https://doi.org/10.1016/s0378-4274(98)00322-1).

[8]. H. A. Le, N. T. Huong, N. H. Phuc, "Influence of Traffic and Wind Speed on the PM₁₀, PM_{2.5} Levels at Truong Chinh Street, Hanoi," *VNU Journal of Science: Natural Sciences and Technology*, 31, 186-192, 2015.

[9]. A. Hajime, "Global Air Quality and Pollution," *Science*, 302, 1716-1719, 2003. <https://doi.org/10.1126/science.1092666>

[10]. Jasim I A, Farhan S L, Al-Maliki L A, AL-Mamoori S K, "Climatic Treatments for Housing in the Traditional Holy Cities: A Comparison between Najaf and Yazd Cities," in: *IOP Conference Series: Earth and Environmental Science*, IOP Publishing, 1-10, 2021.

[11]. WHO, *WHO Global Air Quality Guidelines: Particulate Matter (PM_{2.5} and PM₁₀), Ozone, Nitrogen, Sulfur Dioxide and Carbon Monoxide*. World Health Organization, 2021.

[12]. Nguyen Thi Hoa, "Concentration evaluation of fine particulate matter (PM_{2.5}) in Ho Chi Minh City in 2021," *Tạp chí Khí tượng Thủy văn*, 751, 68-77, 2023.

[13]. <https://iza.namdinh.gov.vn/kcn-hoa-xa>

[14]. Đỗ Quang Hiệp, *Nghiên cứu thiết kế và chế tạo thiết bị quan trắc tự động một số chỉ tiêu ô nhiễm không khí trên nền tảng IoT*. Báo cáo đề tài NCKH tỉnh Nam Định, 2023.

[15]. <https://www.nchmf.gov.vn/>

[16]. Song X., Wang J., Li R., Wang C., "A hybrid CNN-LSTM model for PM_{2.5} forecasting," *Atmospheric Environment*, 223, 117233, 2020. <https://doi.org/10.1016/j.atmosenv.2020.117233>

[17]. Li X., Peng L., Hu Y., Shao J., Chi T., Pan H., "A deep learning approach for air quality prediction based on convolutional neural network and long short-term memory," *Scientific Reports*, 7, 1-10, 2017. <https://www.nature.com/articles/s41598-017-16260-1>

[18]. Zhang H., Li Q., Wang X., Liu C., Huang H., "Deep Air Quality Forecasting Using Hybrid CNN-LSTM Based Deep Learning Model," *IEEE Access*, 9, 112330-112339, 2021. <https://ieeexplore.ieee.org/document/9377106>

[19]. Liang J., Zhang L., Huang K., Sun Y., "Deep learning models for air quality prediction: A review," *Journal of Cleaner Production*, 339, 130139, 2022. <https://doi.org/10.1016/j.jclepro.2022.130139>

[20]. Brownlee J., "How to Develop a CNN-LSTM Model for Multivariate Time Series Forecasting," *Machine Learning Mastery*, 2019. <https://machinelearningmastery.com/cnn-long-short-term-memory-networks/>

[21]. Olah C., "Understanding LSTM Networks," *Colah's blog*, 2015. <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

[22]. Y B Desire, "Air Pollution Forecasting using CNN-LSTM [Code Repository]," *GitHub*, 2021. <https://github.com/ybdesire/Air-Pollution-Forecasting-CNN-LSTM>

AUTHORS INFORMATION

Nguyen Thien Tan¹, Bui Van Hau¹, Do Quang Hiep²

¹Faculty of Electronics and Computer Engineering, University of Economics - Technology for Industries, Vietnam

²Faculty of Electrical-Automation Engineering, University of Economics - Technology for Industries, Vietnam