



## PREDICTING AIR POLLUTION FROM MONITORING DATA USING ARTIFICIAL INTELLIGENCE MODELS

Minh Quy Cao<sup>1\*</sup>, Hong Hai Pham<sup>2</sup>

<sup>1</sup>University of Transport and Communications, No 3 Cau Giay Street, Hanoi, Vietnam

<sup>2</sup>Hanoi University of Natural Resources and Environment, 41A Phu Dien, Hanoi, Vietnam

### ARTICLE INFO

TYPE: Research Article

Received: 18/12/2023

Revised: 16/03/2024

Accepted: 31/03/2024

Published online: 15/04/2024

<https://doi.org/10.47869/tcsj.75.7>

\* *Corresponding author*

Email: [caominhquy@utc.edu.vn](mailto:caominhquy@utc.edu.vn);

**Abstract.** Air pollution is a challenging factor for the environment and human health. Pollutants such as microscopic dust particles and toxic gases arising from the operation of transport vehicles and industrial production processes are becoming a major threat to human health. Air pollution not only affects human health, but air pollution also contributes to global climate change, seriously affecting biodiversity on the earth. This paper presents a method for predicting air quality to provide early warning of pollution through the use of artificial intelligence models. The model receives pollutant data from monitoring stations as input and predicts the concentration of these pollutants in the near future. Three artificial intelligence models such as Long Short-Term Memory (LSTM), Artificial Neural Network (ANN), and Recurrent Neural Network (RNN), are used to predict pollution from some typical pollutants. The results from three evaluation methods including RMSE, R2 and MAE show that the LSTM model gives the best prediction results in predicting pollutants.

**Keywords:** air pollution, predict, artificial intelligence.



## DỰ BÁO Ô NHIỄM KHÔNG KHÍ TỪ DỮ LIỆU QUAN TRẮC SỬ DỤNG CÁC MÔ HÌNH TRÍ TUỆ NHÂN TẠO

Cao Minh Quý<sup>1\*</sup>, Phạm Hồng Hải<sup>2</sup>

<sup>1</sup>Trường Đại học Giao thông vận tải, Số 3, Cầu Giấy, Láng Thượng, Đống Đa, Hà Nội

<sup>2</sup> Trường Đại học Tài nguyên & Môi trường Hà nội, 41A Phú Diễn, Hà Nội, Việt Nam

### THÔNG TIN BÀI BÁO

CHUYÊN MỤC: Công trình khoa học

Ngày nhận bài: 18/12/2023

Ngày nhận bài sửa: 16/03/2024

Ngày chấp nhận đăng: 31/03/2024

Ngày xuất bản Online: 15/04/2024

<https://doi.org/10.47869/tcsj.75.7>

\* Tác giả liên hệ

Email: caominhquy@utc.edu.vn;

**Tóm tắt.** Ô nhiễm không khí đang tạo nên một thách thức đối với môi trường và sức khỏe của con người. Các chất gây ô nhiễm như các hạt bụi siêu nhỏ, khí độc hại phát sinh từ hoạt động của các phương tiện vận tải và quá trình sản xuất công nghiệp đang trở thành mối đe dọa lớn đối với sức khỏe con người. Không chỉ ảnh hưởng đến sức khỏe con người, ô nhiễm không khí còn góp phần vào tình trạng biến đổi khí hậu toàn cầu, gây ảnh hưởng nghiêm trọng đến sự đa dạng sinh học trên Trái Đất. Bài báo trình bày phương pháp dự báo chất lượng không khí nhằm cảnh báo sớm ô nhiễm thông qua sử dụng mô hình trí tuệ nhân tạo. Mô hình nhận dữ liệu các chất ô nhiễm từ trạm quan trắc và dự đoán nồng độ các chất ô nhiễm này trong thời gian kế tiếp.

**Từ khóa:** ô nhiễm không khí, dự báo, trí tuệ nhân tạo.

@ 2024 Trường Đại học Giao thông vận tải

### 1. ĐẶT VẤN ĐỀ

Sự bùng nổ về tăng trưởng kinh tế và công nghiệp trong thế kỷ 21 đã mang theo nhiều cơ hội phát triển đáng kể cho xã hội. Tuy nhiên, kèm theo đó cũng là sự gia tăng tình trạng ô nhiễm không khí, đang tạo nên một thách thức đối với môi trường và sức khỏe của chúng ta [1]. Ô nhiễm không khí không chỉ là vấn đề cục bộ mà còn là một thách thức toàn cầu, ảnh hưởng đến con người và các hệ thống sinh thái trên khắp thế giới. Các chất gây ô nhiễm như các hạt bụi mịn (PM2.5), khí độc hại phát sinh từ hoạt động của các phương tiện giao thông và quá trình sản xuất công nghiệp đang trở thành nguy cơ đe dọa đối với sức khỏe của cộng

đồng [2]. Hệ lụy của ô nhiễm không khí không chỉ giới hạn ở sức khỏe cá nhân mà còn lan rộng đến các lĩnh vực khác. Tác động của ô nhiễm không khí đã góp phần vào tình trạng biến đổi khí hậu toàn cầu, gây ảnh hưởng nghiêm trọng đến đa dạng sinh học, đất đai, nước và nguồn tài nguyên thiên nhiên. Vấn đề ô nhiễm không khí cũng đặt ra câu hỏi về bản chất của sự phát triển bền vững và trách nhiệm của con người đối với môi trường. Bên cạnh những biện pháp quyết liệt và cơ chế quản lý hiệu quả để giảm thiểu tác động của ô nhiễm không khí, bảo vệ cả sức khỏe con người và hệ thống sinh thái thì cần có những phương pháp kỹ thuật để dự báo ô nhiễm và cảnh báo sớm cho người dân [3, 4].

Hiện nay có rất nhiều phương pháp có thể được sử dụng để dự báo chất lượng của không khí, trong đó một số phương pháp nổi bật có thể kể đến như các mô hình thống kê như ARIMA (AutoRegressive Integrated Moving Average) hoặc SARIMA (Seasonal ARIMA) để dự đoán các chỉ số ô nhiễm dựa trên dữ liệu quá khứ [5, 6]. Ưu điểm của các phương pháp này là phù hợp với dữ liệu quan trắc theo thời gian, có thể xử lý dữ liệu thay đổi theo mùa và tích hợp tốt với phân tích thống kê. Tuy nhiên phương pháp này có một số hạn chế là cần phải đảm bảo dữ liệu thỏa mãn các giả định của mô hình và chưa thể dự báo hiệu quả các biến đa biến hoặc dữ liệu với cấu trúc phức tạp. Một phương pháp truyền thống phân tích dữ liệu quan trắc là sử dụng các phương pháp phân tích chuỗi thời gian để hiểu xu hướng và mô hình hoá sự biến đổi của các chỉ số ô nhiễm theo thời gian [7]. Tuy nhiên nhược điểm của phương pháp này là khó xử lý dữ liệu nhiễu và dự báo ít hiệu quả khi có sự biến đổi khó dự đoán được trong tương lai.

Với sự phát triển ngày càng nhanh của công nghệ phần cứng và dữ liệu thu thập ngày càng nhiều, các mô hình học sâu được phát triển và sử dụng trong dự báo ô nhiễm không khí. Bài báo đề xuất một phương pháp dự báo ô nhiễm bụi không khí từ dữ liệu quan trắc sử dụng các mô hình trí tuệ nhân tạo là bộ nhớ dài ngắn LSTM [8], mạng neural nhân tạo ANN [9] và mạng neural hồi quy RNN [10]. Bài báo sử dụng ba mô hình để dự đoán chỉ số của 7 chất ô nhiễm phát tán trong không khí. Kết quả dự đoán có thể được cập nhật hàng giờ dựa trên mức độ cập nhật đầu vào của mô hình. Độ chính xác dự đoán của mô hình cho thấy tính hiệu quả cao trong việc ứng dụng trí tuệ nhân tạo vào dự báo ô nhiễm không khí.

## 2. NỘI DUNG

### 2.1. Dự báo ô nhiễm không khí sử dụng trí tuệ nhân tạo

Bài toán dự báo chất lượng không khí là việc sử dụng các thông tin về dữ liệu thời tiết, dữ liệu ô nhiễm không khí và các yếu tố liên quan khác để dự đoán mức độ ô nhiễm không khí tại các vị trí và thời điểm trong tương lai. Mục tiêu của bài toán này là cung cấp thông tin dự đoán sớm về chất lượng không khí, qua đó giúp cảnh báo người dân và các cơ quan quản lý môi trường về tình trạng ô nhiễm không khí. Chỉ số chất lượng không khí được tính theo thang điểm (khoảng giá trị AQI) tương ứng với biểu tượng và các màu sắc để cảnh báo chất lượng không khí và mức độ ảnh hưởng tới sức khỏe con người. Các chỉ số ô nhiễm không khí thường được đưa vào dự báo bao gồm các chất như PM<sub>2.5</sub>, PM<sub>10</sub>, NO<sub>2</sub>, SO<sub>2</sub>, O<sub>3</sub> và CO. Để thực hiện dự báo, người ta thường sử dụng các phương pháp thống kê, học máy và trí tuệ nhân tạo như mạng nơ-ron học sâu để xây dựng mô hình dự đoán.

Dữ liệu quan trắc là các dữ liệu liên tục theo thời gian, là một dãy các giá trị quan sát  $X := \{x_1, x_2, \dots, x_n\}$ . Lựa chọn phương pháp dự báo chất lượng không khí phụ thuộc vào tính chất của dữ liệu, mục tiêu dự báo và tài nguyên có sẵn. Khi xây dựng mô hình, quan trọng là cân nhắc giữa độ phức tạp của phương pháp và hiệu suất dự báo để đảm bảo rằng mô hình đáp

ứng được yêu cầu của ứng dụng cụ thể. Bài báo sử dụng ba mô hình học sâu trong trí tuệ nhân tạo là ANN, RNN và LSTM để dự báo một số chất ô nhiễm không khí.

Mạng neural nhân tạo (ANN) là một mô hình toán học được xây dựng dựa trên các mạng neural sinh học [11]. Nó gồm có một nhóm các neural (nút) nối với nhau, và xử lý thông tin bằng cách truyền theo các kết nối và tính giá trị mới tại các nút. Mạng ANN là sự kết hợp của những tầng perceptron, thông thường gồm ba tầng chính là tầng đầu vào, các tầng ẩn và tầng đầu ra. Để có thể dự báo, mạng neural cần phải được huấn luyện (training) thông qua cơ chế truyền ngược. Sai số giữa tầng đầu ra và nhãn thực tế sẽ được truyền ngược trở lại nhằm tinh chỉnh các trọng số giữa các kết nối. Mạng neural hồi quy (RNN) là một loại mạng neural nhân tạo (ANN) có thể xử lý dữ liệu tuần tự, phù hợp với bài toán dự báo [12]. RNN hoạt động bằng cách lưu trữ thông tin từ các bước trước đó trong một trạng thái ẩn. Trạng thái ẩn này sau đó được sử dụng để tính toán đầu ra cho bước hiện tại. Điều này cho phép RNN học các mối quan hệ giữa các giá trị trong dữ liệu tuần tự. Mạng bộ nhớ dài ngắn LSTM (Long Short-Term Memory) là một loại mạng RNN có thể xử lý dữ liệu tuần tự dài. LSTM được thiết kế để khắc phục nhược điểm của RNN truyền thống, vốn có thể bị mất thông tin dài hạn trong quá trình xử lý [13]. LSTM hoạt động bằng cách sử dụng ba cổng forget, cổng input và cổng output để kiểm soát thông tin được lưu trữ trong trạng thái ẩn. Ba cổng này được sử dụng để duy trì thông tin dài hạn trong trạng thái ẩn, cho phép mạng học các mối quan hệ phức tạp giữa các giá trị trong dữ liệu tuần tự.

Dựa trên những đánh giá về bộ dữ liệu, tính chất của bài toán này, trong đề tài này ta sẽ sử dụng mô hình LSTM bởi vì LSTM có khả năng xử lý các chuỗi thời gian dài và phức tạp với sự biến đổi phi tuyến. Điều này rất hữu ích khi dự báo chất lượng không khí vì dữ liệu ô nhiễm không khí thường có tính chất thay đổi không đều và không tuyến tính. LSTM được thiết kế để duy trì thông tin từ quá khứ trong các "cổng" đặc biệt của nó, cho phép mô hình hiểu và mô hình hóa các mối quan hệ dài hạn trong dữ liệu chuỗi thời gian. LSTM có khả năng học các mối quan hệ và biểu diễn phức tạp trong dữ liệu. Điều này cho phép nó học các mẫu phức tạp của chất lượng không khí dựa trên các yếu tố thời tiết và môi trường. Để thực hiện tác vụ dự báo, trong nghiên cứu này ta sẽ xây dựng mạng với thành phần chính là 2 lớp LSTM, sau đó sử dụng lớp Dense được trình bày trong bảng 1.

Bảng 1. Cấu trúc mạng LSTM trong dự báo ô nhiễm không khí.

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 168, 128)	66560
lstm_1 (LSTM)	(None, 64)	49408
dense (Dense)	(None, 25)	1625
dense_1 (Dense)	(None, 1)	26

Trong mô hình này, lớp LSTM đầu tiên có 128 đơn vị LSTM và kích thước đầu vào là vector 2 chiều có kích thước 128x1 tương đương với mảng biểu diễn đặc trưng dữ liệu tại 128 giờ. Lớp LSTM thứ hai có 64 đơn vị LSTM, lớp sẽ chỉ trả về đầu ra ở thời điểm cuối cùng của chuỗi đầu vào. Sau hai lớp LSTM, là một lớp kết nối đầy đủ với 26 đơn vị. Lớp này giúp mô hình học các đặc trưng phức tạp từ đầu ra của các lớp LSTM trước đó. Cuối cùng, lớp kết nối đầy đủ với một đơn vị, được sử dụng để dự đoán giá trị kết quả.

## 2.2. Kết quả thực nghiệm và thảo luận

Dữ liệu về thông số không khí được quan trắc từ các trạm cố định mặt đất tại Trung tâm quan trắc Tài nguyên và Môi trường (Sở Tài Nguyên và Môi Trường Tỉnh Bắc Ninh). Thời gian thu thập được chia 2 giai đoạn: Giai đoạn 2000-2022 dùng để huấn luyện mô hình và giai đoạn bắt đầu từ năm 2022 được dùng làm dữ liệu kiểm thử độ chính xác của mô hình.

Dữ liệu được thu thập từ 18 trạm quan trắc ở tỉnh Bắc Ninh, dữ liệu được ghi nhận từ 26/02/2022 đến 01/11/2022 với tổng số lượng bản ghi của 18 trạm thu được là 35,012 bản ghi. Dữ liệu trong bảng bao gồm một số trường thông tin như: Nhiệt độ, NO<sub>x</sub>, SO<sub>2</sub>, O<sub>3</sub>, PM<sub>10</sub>, PM<sub>2.5</sub>, NO, NO<sub>2</sub>, CO,.... Bản đồ khu vực nghiên cứu được chỉ ra trong hình 1.



Hình 1. Bản đồ khu vực thu thập dữ liệu (<https://bacninh.gov.vn/ban-do>).

Trong bài báo này sẽ tiến hành dự báo 7 chỉ số là PM<sub>1</sub>, PM<sub>10</sub>, PM<sub>2.5</sub>, SO<sub>2</sub>, O<sub>3</sub>, CO và NO<sub>2</sub>. Một số mẫu dữ liệu sử dụng trong nghiên cứu được mô tả trong bảng 2.

Bảng 2: Dữ liệu chứa các thông số xác định chất lượng không khí (µg/m<sup>3</sup>).

Datetime	PM <sub>1</sub>	PM <sub>10</sub>	PM <sub>2.5</sub>	SO <sub>2</sub>	NO <sub>2</sub>	O <sub>3</sub>	CO
11/01/2022 19:00	60,91	93,34	71,43	47,81	1,36	6,15	2573,37
11/01/2022 20:00	53,73	88,39	64,05	51,21	1,26	3,61	2251,22
11/01/2022 21:00	45,85	75,17	54,36	55,06	1,41	2,18	2449,72
11/01/2022 22:00	70,88	107,41	83,83	45,38	1,30	4,11	2155,38
11/01/2022 23:00	88,24	124,91	101,22	35,65	1,49	4,18	1605,49
...	...	...	...	...	...	...	...

Sau khi thu thập dữ liệu, cần phải tiến hành phân tích và xử lý dữ liệu để hiểu thêm các đặc trưng của dữ liệu. Bảng thống kê dữ liệu thu thập được mô tả trong bảng 3. Theo kết quả phân tích, dữ liệu không bị khuyết thiếu. Tuy nhiên, dựa trên các số liệu thống kê trong bảng 3, có điểm ngoại lai tại một số các chỉ số, do đó cần phải xử lý các điểm ngoại lai trước khi đưa vào huấn luyện mô hình.

Bảng 3: Thống kê dữ liệu các các chỉ số ô nhiễm ( $\mu\text{g}/\text{m}^3$ ).

Thống kê	PM <sub>1</sub>	PM <sub>10</sub>	PM <sub>2.5</sub>	SO <sub>2</sub>	NO <sub>2</sub>	O <sub>3</sub>	CO
count	24192	24192	24192	24192	24192	24192	24192
mean	21,65	39,19	26,08	46,67	10,96	66,65	683,68
std	31,39	52,55	37,52	40,44	13,44	58,03	589,96
min	0,00	0,00	0,00	0,00	0,00	0,00	0,00
25%	4,91	9,33	6,03	16,26	2,61	27,97	349,57
50%	11,98	21,45	14,51	34,30	5,26	52,41	534,60
75%	28,06	52,97	33,98	72,74	13,81	89,80	815,52
max	738,71	1155,05	938,86	745,19	156,03	1045,92	18601,11

Về mục tiêu dự đoán, mô hình được kỳ vọng sẽ dự đoán được các mốc thời gian trong tương lai, ví như 12 giờ, 1 ngày, 7 ngày hay 1 tháng. Với tính chất dữ liệu chất lượng không khí đang ở dạng chuỗi thời gian với một số tính chất đặc trưng như tính mùa vụ (ví như chất lượng không khí thấp ở các thời điểm người dân đi làm). Do vậy, trong nghiên cứu này sử dụng dữ liệu của 168 giờ từ  $t - 169$  đến  $t - 1$  để dự báo chất lượng không khí cho ngày  $t$ . Chúng tôi cũng xây dựng 3 mô hình để dự báo giá trị tương lai cho 7 chỉ số PM<sub>1</sub>, PM<sub>10</sub>, PM<sub>2.5</sub>, SO<sub>2</sub>, NO<sub>2</sub>, O<sub>3</sub>, CO là LSTM, RNN và ANN để so sánh hiệu quả của từng mô hình. Kết quả mô tả trong bảng 4 là chỉ số của 7 chất ô nhiễm sau khi tiến hành huấn luyện trên tập dữ liệu về chất lượng không khí tại tỉnh Bắc Ninh.

Bảng 4. Dự báo chỉ số ô nhiễm ( $\mu\text{g}/\text{m}^3$ ).

DateTime	PM <sub>1</sub>	PM <sub>10</sub>	PM <sub>2.5</sub>	SO <sub>2</sub>	NO <sub>2</sub>	O <sub>3</sub>	CO	AQI	AQI	AQI	AQI	AQI	AQI	AQI	Summary
								SO <sub>2</sub>	CO	O <sub>3</sub>	NO <sub>2</sub>	PM <sub>10</sub>	PM <sub>2.5</sub>		
12/01/2022 11:00	44,2	39,8	119,5	54,8	6,4	8,1	348,4	21,9	1,7	2,5	3,2	76,0	171,5	171,5	Xấu
12/01/2022 12:00	44,3	31,3	121,3	62,6	6,4	7,3	284,9	25,0	1,4	2,3	3,2	66,1	171,9	171,9	Xấu
12/01/2022 13:00	44,5	26,9	123,2	58,7	6,4	6,3	242,4	23,5	1,2	2,0	3,2	57,2	172,2	172,2	Xấu
12/01/2022 14:00	44,8	27,1	125,4	40,4	6,4	6,8	233,8	16,2	1,2	2,1	3,2	51,6	172,6	172,6	Xấu
12/01/2022 15:00	45,2	29,2	127,9	29,8	6,4	8,8	245,5	11,9	1,2	2,7	3,2	52,0	173,3	173,3	Xấu

Để đánh giá kết quả dự báo, chúng tôi sử dụng ba phương pháp đánh giá là hệ số xác định ( $R^2$ ), sai số căn trung bình bình phương (RMSE) và sai số trung bình tuyệt đối (MAE) theo công thức (1), (2), (3).

$R^2$  được tính bằng tổng phương sai được giải thích bởi mô hình chia cho tổng phương sai của biến phụ thuộc.

$$R^2 = 1 - \frac{SSE}{SST} \quad (1)$$

Trong đó SSE (Sum of Squared Errors) là tổng bình phương sai số của mô hình, SST (Total Sum of Squares) là tổng bình phương sai số của trung bình mẫu

Sai số trung bình tuyệt đối là trung bình độ chênh lệch của giá trị thực và giá trị dự báo

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (2)$$

trong đó:  $y_i$  là giá trị thực,

$\hat{y}_i$  là giá trị dự báo,  $N$  là số lượng mẫu

Sai số căn trung bình bình phương tương ứng với căn bậc hai của trung bình cộng bình phương hiệu số giữa giá trị mục tiêu và giá trị được mô hình hồi quy dự đoán.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3)$$

Bảng 5 chỉ ra kết quả dự đoán các chỉ số ô nhiễm sử dụng ba mô hình ANN, LSTM và RNN được đánh giá qua ba phương pháp RMSE,  $R^2$  và MAE trên tập dữ liệu huấn luyện và tập dữ liệu kiểm thử.

Kết quả trong bảng 4 cho thấy độ chính xác phân loại đạt hơn 85% trong tập huấn luyện và hơn 90% đối với tập kiểm định. Với mỗi chỉ số được thực hiện huấn luyện thực nghiệm 3 mô hình là LSTM, RNN và ANN. Trong bài báo này,  $R^2$  được chọn ở tập huấn luyện là tiêu chí để lựa chọn một mô hình là phù hợp hay chưa. Theo kết quả ở bảng 4 cho thấy mô hình phù hợp với từng chỉ số là LSTM phù hợp cho dự đoán các chất  $PM_{10}$ ,  $PM_{2.5}$ ,  $SO_2$ ,  $NO_2$ ,  $O_3$ , mô hình ANN phù hợp với dự đoán các chất  $PM_{10}$  và CO.

Chỉ số chất lượng không khí được tính theo thang điểm (khoảng giá trị AQI) tương ứng với biểu tượng và các màu sắc để cảnh báo chất lượng không khí và mức độ ảnh hưởng tới sức khỏe con người, theo Quyết định số 1459/QĐ-TCMT ngày 12/11/2029 v/v ban hành hướng dẫn kỹ thuật tính toán và công bố chỉ số chất lượng không khí Việt nam (VN\_AQI) của Tổng cục Môi trường. VN\_AQI được tính toán bao gồm AQI giờ và AQI ngày. Số liệu sử dụng để tính toán VN\_AQI là giá trị quan trắc trung bình 1 giờ, trung bình 8 giờ và trung bình 24 giờ. VN\_AQI được tính toán cho dữ liệu của từng trạm quan trắc không khí tự động liên tục đối với môi trường không khí xung quanh. Đối với mỗi trạm quan trắc,  $AQI_x$  được tính toán cho từng thông số quan trắc, giá trị AQI cuối cùng là giá trị lớn nhất trong các giá trị  $AQI_x$  của mỗi thông số.

Bảng 5. So sánh kết quả dự đoán giữa 3 mô hình.

Chỉ số	Mô hình	Tập huấn luyện			Tập kiểm thử		
		RMSE	R2	MAE	RMSE	R2	MAE
PM-1	ANN	40,23	0,76	7,16	8,15	0,61	4,22
	LSTM	44,56	0,83	5,01	7,34	0,68	3,56
	RNN	43,79	0,82	5,11	8,09	0,62	3,76
PM-10	ANN	72,01	0,84	9,66	10,97	0,68	6,02
	LSTM	70,76	0,82	8,59	10,61	0,71	5,52
	RNN	79,50	0,78	9,75	11,66	0,64	6,00
PM-2.5	ANN	48,87	0,78	7,80	9,36	0,61	4,78
	LSTM	54,35	0,82	6,08	8,84	0,65	4,11
	RNN	51,26	0,80	6,81	9,75	0,58	5,26
SO2	ANN	54,92	0,71	16,02	10,76	0,19	9,21
	LSTM	58,02	0,87	5,46	5,48	0,79	3,65
	RNN	55,51	0,81	5,34	6,14	0,74	4,11
NO2	ANN	17,24	0,82	2,96	0,31	0,29	0,27
	LSTM	18,81	0,88	2,51	0,91	-5,34	0,90
	RNN	19,01	0,86	2,47	0,74	-3,21	0,69
O3	ANN	73,88	0,69	22,53	29,70	0,65	19,24
	LSTM	82,17	0,87	11,69	24,20	0,77	14,95
	RNN	79,12	0,85	12,66	23,64	0,78	14,82
CO	ANN	737,23	0,65	181,54	335,77	0,65	190,03
	LSTM	742,42	0,65	157,61	351,20	0,62	191,42
	RNN	759,06	0,60	191,39	368,25	0,58	220,59
AQI	Tổng hợp	18,52	0,88	8,94	13,94	0,77	6,98

Về kết quả của các mô hình huấn luyện con cho thấy các mô hình đều có chỉ số đạt ở ngưỡng chấp nhận được, cũng như có sai số không quá lớn so với giá trị trung bình của chỉ số đó. Về kết quả AQI, mặc dù AQI là kết quả phái sinh của việc dự báo các mô hình trên, tuy nhiên kết quả thu được là rất rất tốt. Cụ thể, R2 ở tập huấn luyện đạt hơn 88% với sai số trung bình chỉ khoảng 8. Đây là một kết quả rất tốt, đặc biệt nó còn thể hiện qua mức độ phân loại màu của dự báo đạt độ chính xác trên 90%.

#### 4. KẾT LUẬN

Ô nhiễm không khí đang là một thực trạng bức thiết mà toàn thể xã hội đang tìm giải pháp để khắc phục. Bài báo đưa ra được phương pháp dự báo các chỉ số ô nhiễm không khí dựa trên các mô hình trí tuệ nhân tạo. Kết quả dự báo được đo bằng ba phương pháp đánh giá là RMSE, R2 và MAE cho thấy mô hình mạng bộ nhớ dài ngắn (LSTM) có độ chính xác dự báo cao nhất với 5 chất ô nhiễm không khí. Phương pháp dự báo đề xuất thể hiện rõ tính ưu việt về chi phí, thời gian và công sức triển khai khi đã áp dụng thành công mạng LSTM vào bộ dữ liệu quan trắc đã được thu thập tại tỉnh Bắc Ninh (Việt Nam) trong nhiều năm. Kết quả



thu được sau khi huấn luyện mô hình được đánh giá tốt, cùng với đó là sự linh hoạt, đơn giản trong việc lựa chọn được mốc thời gian trong tương lai để dự báo.

## TÀI LIỆU THAM KHẢO

- [1]. WHO. World Health Statistics 2021: Monitoring Health for the SDGs, Sustainable Development Goals; WHO: Geneva, Switzerland, 2021.
- [2]. J. Zaheer, J. Jeon, S.B. LeeII, J.S. Kim, Effect of Particulate Matter on Human Health, Prevention, and Imaging Using PET or SPECT, *Prog. Med. Phys.*, 29 (2018) 81-91. <https://doi.org/10.14316/pmp.2018.29.3.81>
- [3]. Bộ TN&MT, Báo cáo Hiện trạng Môi trường Việt Nam giai đoạn 2016-2020, 2021.
- [4]. Phạm Ngọc Đăng, Thực trạng chất lượng môi trường không khí Việt Nam giai đoạn 2016 – 2020, xác định các thách thức và đề xuất các giải pháp cải thiện trong 5 năm tới, *Tạp chí Môi trường số tháng 1* (2022).
- [5]. V. I. Kontopoulou, A. D. Athanasios, I. Kakkos, G. K. Matsopoulos., A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks, *Future Internet*, 15 (2023) 255. <https://doi.org/10.3390/fi15080255>
- [6]. Peng Chen, Aichen Niu, Duanyang Liu, Wei Jiang, Bin Ma, Time Series Forecasting of Temperatures using SARIMA: An Example from Nanjing, *IOP Conference Series: Materials Science and Engineering*, 394 (2018) 052024. <https://doi.org/10.1088/1757-899X/394/5/052024>
- [7]. Q. Song, B. S. Chissom, Forecasting enrollments with fuzzy time series – Part I. *Fuzzy Sets and Systems*, 54 (1993) 1–9. [https://doi.org/10.1016/0165-0114\(93\)90355-L](https://doi.org/10.1016/0165-0114(93)90355-L)
- [8]. C. R. Staudemeyer, E. R. Morris, Understanding LSTM -- a tutorial into Long Short-Term Memory Recurrent Neural Networks, (2019) 1-42, <https://doi.org/10.48550/arXiv.1909.09586>
- [9]. Ahmed Tealab, Time series forecasting using artificial neural networks methodologies: A systematic review, *Future Computing and Informatics Journal*, 3 (2018) 334-340. <https://doi.org/10.1016/j.fcij.2018.10.003>
- [10]. Gjylapi, Dezdemonia & Proko, Eljona, *Recurrent Neural Networks in Time Series Prediction*, 2017.
- [11]. David Kriesel, *A Brief Introduction to Neural Networks*, NeuralNetworks, 2007.
- [12]. Hansika Hewamalage, Christoph Bergmeir, Kasun Bandara, Recurrent Neural Networks for Time Series Forecasting: Current status and future directions, *International Journal of Forecasting*, 37 (2021) 388-427. <https://doi.org/10.1016/j.ijforecast.2020.06.008>
- [13]. G. Van Houdt, C. Mosquera, G. Nápoles, A review on the long short-term memory model, *Artif Intell Rev*, 53 (2020) 5929–5955. <https://doi.org/10.1007/s10462-020-09838-1>