

NGHIÊN CỨU XÂY DỰNG HỆ THỐNG NHÚNG IoT TÍCH HỢP AI GIÁM SÁT CHẤT LƯỢNG KHÔNG KHÍ MÔI TRƯỜNG THEO THỜI GIAN THỰC - ỨNG DỤNG THỬ NGHIỆM TẠI HÀ NỘI

Nguyễn Văn Hách*, Nguyễn Văn Tuyên, Trương Mạnh Đạt, Phạm Hồng Hải
Trường Đại học Tài nguyên và Môi trường Hà Nội

Tóm tắt

Ô nhiễm không khí, đặc biệt là bụi mịn PM2.5 là một vấn đề nghiêm trọng tại các thành phố lớn như Hà Nội. Nghiên cứu này phát triển một hệ thống giám sát không khí sử dụng công nghệ IoT, giúp thu thập và phân tích dữ liệu thời gian thực về chất lượng không khí, bao gồm các chỉ số như PM2.5, CO₂, nhiệt độ và độ ẩm. Hệ thống được thử nghiệm tại Hà Nội và được tích hợp các mô hình dự báo chuỗi thời gian ARIMA và Long Short Term Memory (LSTM), cung cấp các dự báo về xu hướng ô nhiễm trong tương lai. Kết quả thử nghiệm chỉ ra khả năng giám sát và dự báo chính xác, hỗ trợ các cơ quan chức năng trong việc đưa ra các biện pháp phòng ngừa kịp thời.

Từ khóa: Internet vạn vật (IoT); Giám sát chất lượng không khí; PM2.5; Dự báo chuỗi thời gian; ARIMA; Bộ nhớ ngắn - dài hạn (LSTM); Phân tích dữ liệu thời gian thực; Giám sát môi trường.

Abstract

Research on developing an embedded IoT system integrated with AI for real-time environmental air quality monitoring - a pilot application in Hanoi

Air pollution, particularly fine particulate matter (PM2.5), poses a serious challenge in major cities such as Hanoi. This study develops an air quality monitoring system based on Internet of Things (IoT) technology, enabling real-time collection and analysis of air quality data, including parameters such as PM2.5, CO₂, temperature, and humidity. The system was tested in Hanoi and integrated with time series forecasting models AutoRegressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM) to provide predictions of future pollution trends. Experimental results demonstrate the system's capability for accurate monitoring and forecasting, thereby supporting governmental agencies in implementing timely preventive measures.

Keywords: Internet of Things (IoT); Air quality monitoring; PM2.5; Time series forecasting; ARIMA; Long Short-Term Memory (LSTM); Real-time data analysis; Environmental monitoring.

BBT nhận bài: 19/8/2025; Phản biện xong: 17/9/2025 Chấp nhận đăng: 30/10/2025

*Tác giả liên hệ, Email: nvhach@hunre.edu.vn

DOI: <https://doi.org/10.63064/khtnmt.2025.726>

1. Giới thiệu

Ô nhiễm không khí là một trong những vấn đề môi trường nghiêm trọng nhất hiện nay, đặc biệt tại các đô thị lớn. Theo Tổ chức Y tế Thế giới [9] ô nhiễm không khí gây ra khoảng 7 triệu ca tử vong sớm mỗi năm, trong đó bụi mịn PM2.5 được coi là tác nhân nguy hiểm hàng đầu do khả năng xâm nhập sâu vào phổi và hệ tuần hoàn, gây ra các bệnh tim mạch, hô hấp và ung thư.

Tại Việt Nam, Hà Nội thường xuyên ghi nhận nồng độ PM2.5 vượt xa mức khuyến nghị 15 $\mu\text{g}/\text{m}^3$ trung bình năm của WHO. Số liệu từ AirVisual [10] cho thấy, vào mùa Đông, chỉ số PM2.5 tại nhiều khu vực của thành phố đạt 100 - 150 $\mu\text{g}/\text{m}^3$, ảnh hưởng nghiêm trọng đến sức khỏe cộng đồng. Các trạm quan trắc truyền thống tuy có độ chính xác cao nhưng chi phí lắp đặt và vận hành lớn, phạm vi bao phủ hạn chế [1]. Trong bối cảnh đó, công nghệ Internet of Things (IoT) kết hợp trí tuệ nhân tạo (AI) mở ra giải pháp khả thi, cho phép triển khai mạng lưới cảm biến phân tán, chi phí hợp lý, thu thập dữ liệu thời gian thực và dự báo xu hướng ô nhiễm [5, 6].

Trong nghiên cứu này xây dựng hệ thống nhúng IoT tích hợp mô hình ARIMA và LSTM nhằm giám sát và dự báo chất lượng không khí tại một số điểm ở Hà Nội, với tập trung giám sát bốn chỉ số môi trường chính gồm nồng độ bụi mịn PM2.5 ($\mu\text{g}/\text{m}^3$), nồng độ CO₂ (ppm), nhiệt độ (°C) và độ ẩm (%), góp phần hỗ trợ quản lý môi trường và nâng cao khả năng cảnh báo sớm và hỗ trợ ra quyết định ngăn chặn.

2. Mục tiêu và phạm vi nghiên cứu

Mục tiêu chính của nghiên cứu là xây dựng một hệ thống giám sát không khí theo

thời gian thực ứng dụng IoT, có khả năng đo các chỉ số môi trường như PM2.5, CO₂, nhiệt độ và độ ẩm. Dữ liệu thu thập được xử lý, lưu trữ và trực quan hóa trên nền tảng IoT, đồng thời được phân tích bằng các mô hình dự báo chuỗi thời gian ARIMA và mạng nơ-ron LSTM nhằm dự đoán xu hướng ô nhiễm trong tương lai gần.

Phạm vi triển khai thử nghiệm giám sát không khí trong 24 giờ từ 09:00 ngày 07/8/2025 đến 09:48 ngày 08/8/2025 (giờ địa phương) tại khu vực xung quanh Trường Đại học Tài nguyên và Môi trường Hà Nội. Dữ liệu được ghi nhận bởi cảm biến nhiệt độ môi trường, truyền qua ESP32 và lưu trữ trên Firebase Realtime Database.

3. Phương pháp nghiên cứu và xây dựng hệ thống

Hệ thống giám sát không khí bao gồm các cảm biến PM2.5, CO₂, nhiệt độ và độ ẩm, kết nối với nền tảng IoT để thu thập và truyền tải dữ liệu. Dữ liệu thu thập được từ các cảm biến này được xử lý và phân tích trên nền tảng đám mây để hiển thị trực quan và đưa ra các cảnh báo khi mức độ ô nhiễm vượt ngưỡng an toàn. Thiết bị phần cứng để xây dựng giám sát bằng IoT:

- *Vi điều khiển ESP32 & KIT WIFI ESP-32 ESP-WROOM-32S*: Bộ vi điều khiển 32-bit tích hợp Wi-Fi và Bluetooth, xử lý dữ liệu và điều khiển toàn bộ hệ thống IoT.

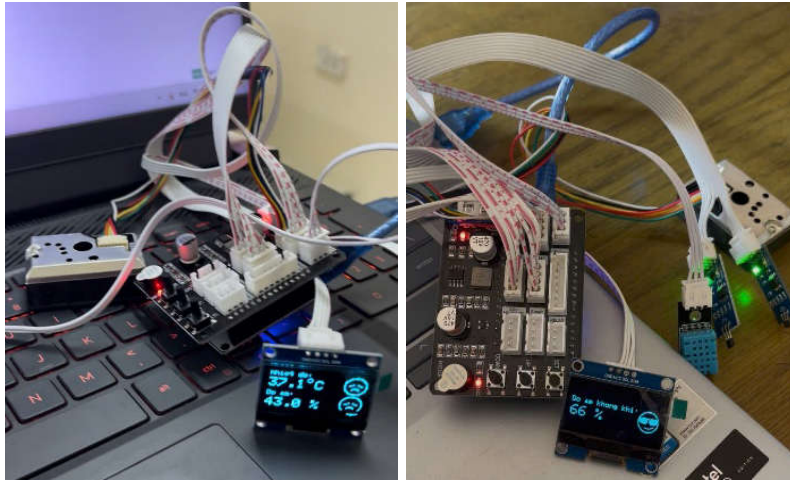
- *Cảm biến nhiệt độ và độ ẩm (DHT11/DHT22)*: Đo nhiệt độ (°C) và độ ẩm (%) của môi trường xung quanh.

- *Cảm biến bụi mịn PM2.5 (GP2Y1010AU0F)*: Đo nồng độ bụi mịn trong không khí, đánh giá chất lượng không khí.

- Màn hình OLED 1.3" 128 × 64 I²C (OLED-13-I2C-B): Hiển thị dữ liệu đo theo thời gian thực với độ tương phản cao.

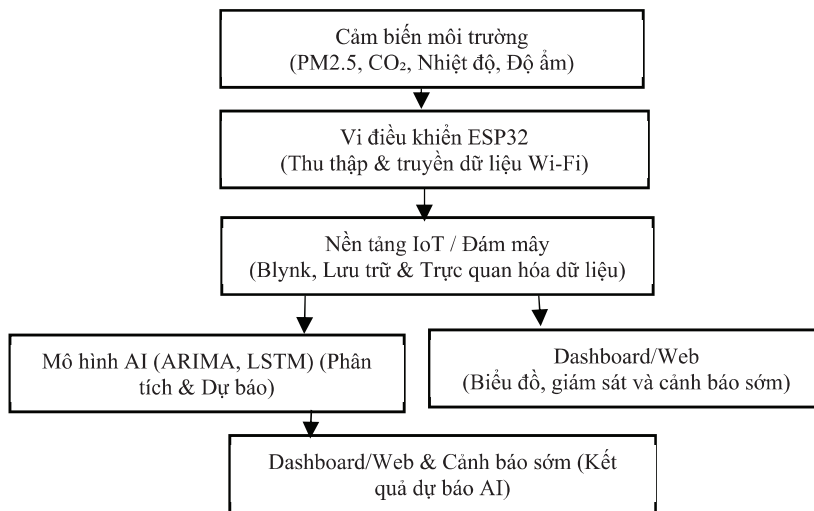
- Dây nối & cáp Micro - USB: Cung cấp nguồn điện và kết nối dữ liệu giữa các mô-đun trong hệ thống.

Sau khi gắn kết nối và xây dựng hệ thống nhúng kết hợp với nền tảng giám sát IoT (<https://blynk.io/>) và viết code chương trình trên Arduino để nhúng vào phần cứng, chúng ta thu được kết quả trên phần cứng như sau:



Hình 1: Thiết bị phần cứng giám sát chất lượng không khí

Tiếp theo xây dựng kiến trúc tổng thể của hệ thống, thể hiện mối liên kết giữa cảm biến, vi điều khiển, nền tảng IoT và mô hình AI trong quá trình giám sát và dự báo chất lượng không khí.



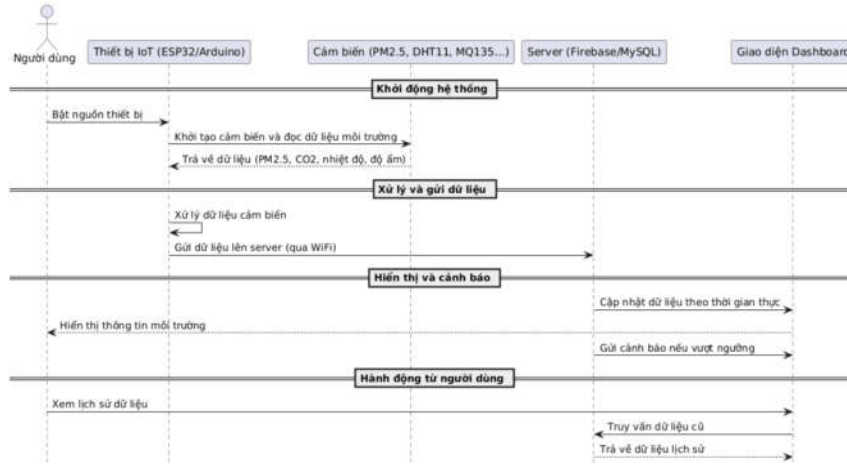
Hình 2: Sơ đồ kiến trúc hệ thống IoT tích hợp AI giám sát chất lượng không khí

Ngoài ra, mô tả sơ đồ luồng dữ liệu tổng thể của hệ thống giám sát chất lượng không khí, trong đó các thành phần phần cứng, nền tảng IoT và mô hình AI phối hợp với nhau để tạo thành một quy trình khép kín. Dữ liệu môi trường được thu

thập liên tục từ các cảm biến (PM2.5, CO₂, nhiệt độ, độ ẩm), sau đó được vi điều khiển ESP32 xử lý sơ bộ và truyền không dây lên nền tảng IoT/đám mây. Tại đây, dữ liệu vừa được lưu trữ và trực quan hóa theo thời gian thực, vừa

Nghiên cứu

được chuyển tiếp tới các mô hình dự báo ARIMA và LSTM để phân tích xu hướng trong ngắn hạn. Kết quả phân tích và dự báo được hiển thị trên dashboard/web, đồng thời cung cấp cảnh báo sớm cho người dùng.



Hình 3: Sơ đồ tổng thể luồng dữ liệu kết nối giữa người dùng, phân cứng và giao diện chương trình

Trong bài báo này, nhóm tác giả còn tích hợp và phát triển hai mô hình dự báo để phân tích dữ liệu và dự đoán xu hướng ô nhiễm:

(1). Mô hình ARIMA (AutoRegressive Integrated Moving Average): Mục tiêu của mô hình này được sử dụng để dự báo các chuỗi thời gian, giúp phân tích các xu hướng biến động

của các chỉ số môi trường gồm nồng độ PM2.5, CO₂, nhiệt độ và độ ẩm, thu thập từ các cảm biến IoT trong dữ liệu đã ghi nhận và xu hướng biến động ngắn hạn.

Mô hình bao gồm ba thành phần chính:

AR (AutoRegressive): Thành phần tự hồi quy mô hình hóa giá trị hiện tại dựa trên p giá trị trước đó:

$$AR(p) : z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + \epsilon_t$$

MA (Moving Average): Thành phần trung bình trượt mô tả giá trị hiện tại như hàm tuyến tính của sai số ngẫu nhiên từ các bước trước

$$MA(q) : z_t = \mu + \sum_{i=1}^q \theta_i \epsilon_{t-i}$$

I (Integrated): Thành phần sai phân giúp biến chuỗi không dừng thành chuỗi dừng:

- Sai phân bậc 1: $I(1) = z_t - z_{t-1}$
- Sai phân bậc d: $I(d) = \Delta^d(z_t)$

(2). LSTM (Long Short Term Memory): Mô hình học sâu này có khả năng xử lý các chuỗi thời gian dài, thích hợp cho việc dự báo ô nhiễm không khí dựa trên các đặc trưng phi tuyến của dữ liệu. LSTM là một biến thể của mạng nơ-ron hồi tiếp (RNN) được thiết kế để ghi nhớ/loại bỏ thông tin theo thời gian thông qua các cổng (gates), khắc phục hiện tượng biến mất/ bùng nổ gradient khi xử lý chuỗi dài [11]. Tại mỗi thời điểm t , với đầu vào x_t , trạng thái ẩn h_{t-1} và trạng thái ô nhớ c_{t-1} , LSTM cập nhật như sau:

$$\begin{aligned} \mathbf{i}_t &= \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i) && \text{(cổng vào)} \\ \mathbf{f}_t &= \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f) && \text{(cổng quên)} \\ \mathbf{o}_t &= \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o) && \text{(cổng ra)} \\ \tilde{\mathbf{c}}_t &= \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c) && \text{(ứng viên bộ nhớ)} \\ \mathbf{c}_t &= \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t \\ \mathbf{h}_t &= \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \end{aligned}$$

$$\underbrace{[\mathbf{x}_{t-L+1}, \dots, \mathbf{x}_t]}_{\text{input } L \times d} \longrightarrow \underbrace{[y_{t+1}, \dots, y_{t+H}]}_{\text{target } H \times 1}$$

trong đó y là biến đích (ví dụ nhiệt độ). Với dự báo nhiều biến đích, mở rộng vector mục tiêu tương ứng.

(iii) chia train/validation/test theo thứ tự thời gian với tỷ lệ khuyến nghị 70/15/15 % trên 298 mốc:

train (208 mốc): 09:00 07/8/2025 → 02:15 08/8/2025

validation (45 mốc): 02:20 08/8/2025 → 06:00 08/8/2025

test (45 mốc): 06:05 08/8/2025 → 09:45 08/8/2025

(iv) xây dựng mạng 1 - 2 lớp LSTM (32 - 64 nút/lớp) + Dropout (0,2 - 0,5) + tầng Dense đầu ra.

Huấn luyện bằng Adam, hàm mất mát MSE, Early Stopping theo lỗi validation. Hai cách dự báo đa bước: *direct multi-output* (Dense trả về vector 48 bước) hoặc *recursive/rolling* (dự báo từng bước và hồi tiếp đầu vào).

LSTM học các mẫu phi tuyến giữa PM2.5, CO₂, nhiệt độ, độ ẩm theo thời gian, từ đó dự báo 48 bước tiếp theo cho từng điểm đo. Hiệu năng được đánh giá trên tập kiểm thử sau khi khôi phục (inverse-transform) về thang đo gốc, theo hai chỉ số MAE và RMSE:

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

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

4. Kết quả nghiên cứu và thực nghiệm

Kết quả thực nghiệm, hệ thống giám sát môi trường được triển khai kết nối với các cảm biến PM2.5, CO₂, nhiệt độ và độ ẩm, đồng thời truyền dữ liệu thời gian

trong đó $\sigma(\cdot)$ là sigmoid, \odot là nhân Hadamard. Nhờ cơ chế cổng, LSTM học được quan hệ phi tuyến và phụ thuộc dài hạn, phù hợp dự báo chuỗi môi trường biến động phức tạp như PM2.5 [8, 11].

Dữ liệu đầu vào gồm các thuộc tính theo thời gian: PM2.5, CO₂, nhiệt độ, độ ẩm (có thể bổ sung giờ trong ngày, ngày trong tuần).

Quy trình thực hiện:

(i) tiền xử lý và chuẩn hóa (Min-Max hoặc Z-score): Dữ liệu mẫu như chuỗi đo 1 phút từ 09:00 07/8/2025 đến 09:48 08/8/2025 có tổng $n_1 = 1.489$ mẫu. Để giảm nhiễu đo, dữ liệu được tái lấy mẫu 5 phút (lấy trung bình), thu được $n_5 = 298$

(ii) tạo cửa sổ trượt *look-back* L (ví dụ 24 - 72 bước) và nhãn *horizon* H (48 bước dự báo):

Với chuỗi đa biến $\mathbf{x}_t \in \mathbb{R}^d$ (d là số đặc trưng), tạo cặp (*đầu vào, nhãn*) theo cửa sổ trượt:

Look-back L : Số bước quá khứ, ví dụ $L \in [24, 72]$

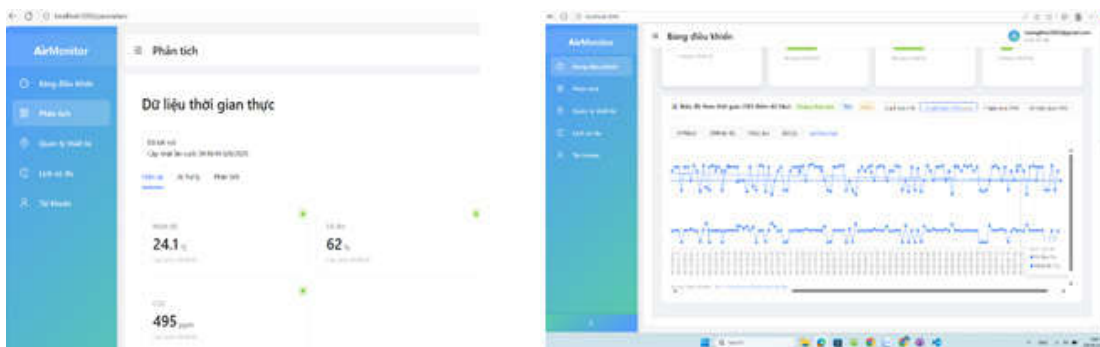
Horizon H : Số bước dự báo phía trước, mặc định $H = 48$

trong đó y_i là giá trị đo thực và \hat{y}_i là dự báo tại thời điểm i (đơn vị °C cho nhiệt độ, ppm cho CO₂, µg/m³ cho PM2.5); kết quả được so sánh với ARIMA để chứng minh lợi ích khi chuỗi có quan hệ phi tuyến/chu kỳ bất quy tắc [8].

Nghiên cứu

thực lên nền tảng web. Kết quả thu được cho thấy hệ thống hoạt động ổn định, dữ liệu được cập nhật liên tục và trực quan

hóa dưới dạng biểu đồ, giúp người dùng dễ dàng theo dõi các thông số môi trường theo thời gian.

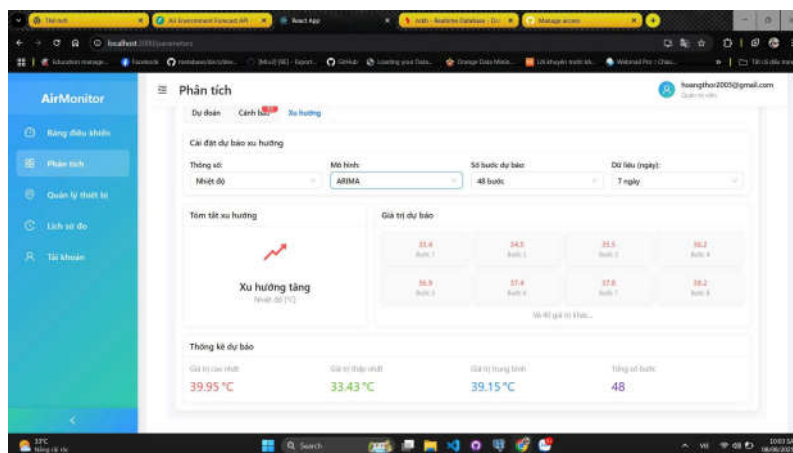


a. Giám sát môi trường theo thời gian thực b. Biểu đồ hiển thị dữ liệu theo thời gian thực

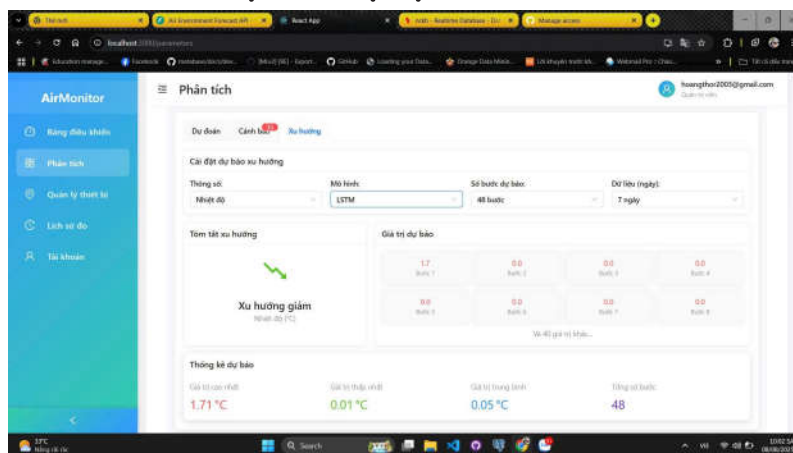
Hình 4: Kết quả giám sát môi trường trên nền Web

Mô hình ARIMA: Kết quả từ mô hình ARIMA cho thấy xu hướng tăng dần của nhiệt độ trong vòng 48 bước dự báo, với giá trị cao nhất đạt 39,95 °C. Cảnh

báo này giúp các cơ quan quản lý sớm phát hiện và có kế hoạch điều chỉnh các biện pháp giảm thiểu tác động của nhiệt độ cao và ô nhiễm.



Hình 5: Dự báo nhiệt độ với mô hình ARIMA



Hình 6: Dự báo nhiệt độ với mô hình LSTM

Mô hình LSTM: Kết quả thử nghiệm từ mô hình LSTM cho thấy xu hướng giảm nhiệt độ trong các bước dự báo, với giá trị thấp nhất là 0,01 °C trong Hình 6 dưới được coi là nhiều hiển thị chứ không phải biến thiên có ý nghĩa vật lý, điều này nhất quán với xu hướng ngắn hạn ổn định; Sai số MAE/RMSE được trình bày cho H = 48. Mô hình này có khả năng xử lý và dự đoán các xu hướng thay đổi nhỏ của dữ liệu, giúp phát hiện các tình huống ô nhiễm tiềm ẩn từ những biến động nhỏ của dữ liệu.

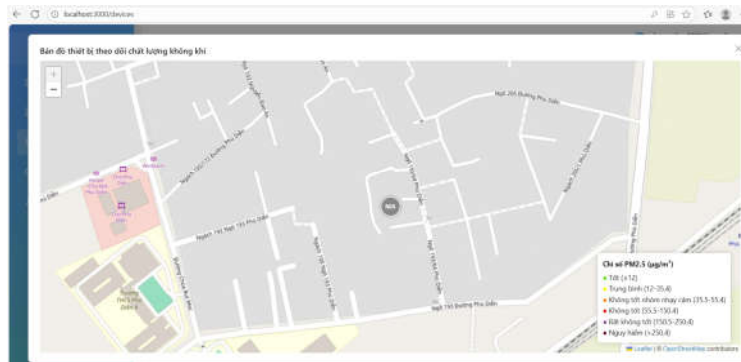
5. Đóng góp của nghiên cứu

Hệ thống giám sát không khí sử dụng IoT giúp theo dõi chất lượng không khí trong thời gian thực và cung cấp các cảnh báo kịp thời khi nồng độ ô nhiễm vượt

ngưỡng an toàn. Hệ thống còn tích hợp các mô hình ARIMA và LSTM để dự đoán xu hướng ô nhiễm, giúp đưa ra các biện pháp phòng ngừa hiệu quả. Các kết quả này góp phần quan trọng trong việc quản lý ô nhiễm không khí và bảo vệ sức khỏe cộng đồng.

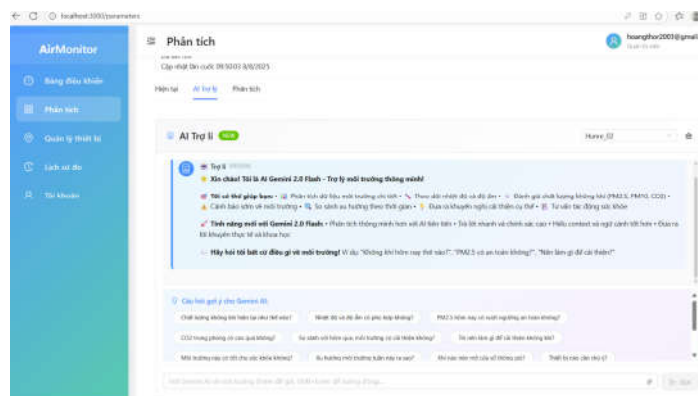
- **Ứng dụng công nghệ IoT và AI:** Hệ thống không chỉ sử dụng công nghệ IoT để thu thập dữ liệu mà còn ứng dụng các mô hình AI để dự báo xu hướng ô nhiễm, hỗ trợ các cơ quan quản lý đưa ra quyết định kịp thời.

- **Phát triển hệ thống giám sát môi trường:** Hệ thống cung cấp nền tảng để nghiên cứu và phát triển các giải pháp giám sát môi trường thông minh, có thể mở rộng ứng dụng ở các khu vực khác.



Hình 7: Tích hợp bản đồ và giám sát theo vị trí

Ngoài ra nhóm tác giả còn phát triển tích hợp AI trong trợ lý ảo hỗ trợ người dùng khi muốn đặt các câu hỏi cho hệ thống



Hình 8: Phát triển trợ lý môi trường thông minh

6. Kết luận và kiến nghị

Bài báo đã trình bày quy trình thiết kế và triển khai hệ thống IoT giám sát chất lượng không khí, tích hợp cảm biến chi phí thấp với nền tảng lưu trữ và hiển thị dữ liệu thời gian thực. Hệ thống không chỉ thu thập thông tin về PM2.5, CO₂, nhiệt độ và độ ẩm mà còn tích hợp mô hình dự báo chuỗi thời gian ARIMA và LSTM, cho phép dự báo xu hướng biến động của các thông số môi trường trong ngắn hạn. Kết quả thử nghiệm cho thấy phương pháp học sâu LSTM có tiềm năng cải thiện độ chính xác dự báo so với ARIMA trong nhiều trường hợp, đặc biệt đối với dữ liệu biến động phức tạp theo thời gian. Ngoài ra, trong nghiên cứu của bài báo này cũng hạn chế như: Phần cứng xây dựng hệ thống IoT mới chỉ mang tính ứng dụng mô phỏng nhỏ, các cảm biến (sensor) đo chưa độ nhạy cao, chưa nhiều thiết bị kết nối, các mẫu huấn luyện còn ít và so sánh giữa các mô hình còn hạn chế.

Tiến hành hiệu chuẩn cảm biến PM2.5 và CO₂ với thiết bị chuẩn hoặc trạm quan trắc để nâng cao độ tin cậy dữ liệu. Mở rộng bộ dữ liệu huấn luyện và thử nghiệm nhằm cải thiện khả năng khái quát của mô hình, đồng thời thử nghiệm các mô hình lai (hybrid) kết hợp ARIMA và LSTM.

TÀI LIỆU THAM KHẢO

[1]. Nguyen, T. M., & Nguyen, H. T. (2020). *Phát triển hệ thống giám sát chất lượng không khí sử dụng cảm biến và IoT tại Việt Nam*. Tạp chí Khoa học và Công nghệ Việt Nam.

[2]. Nguyễn Văn Quân (2018). *Ứng dụng công nghệ IoT trong giám sát và cảnh báo môi trường không khí*. Hội thảo quốc gia về công nghệ thông tin và truyền thông.

[3]. Bộ Tài nguyên và Môi trường (2013). *QCVN 05:2013/BTNMT - Quy chuẩn kỹ thuật quốc gia về chất lượng không khí xung quanh*. Nxb. Tài nguyên - Môi trường và Bản đồ Việt Nam, Hà Nội.

[4]. Ashton, K., (2009). *That 'Internet of Things' Thing*. *RFID Journal*.

[5]. Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M., (2013). *Internet of Things (IoT): A vision, architectural elements, and future directions*. *Future Generation Computer Systems*, 29(7), 1645 - 1660.

[6]. Bandyopadhyay, D., & Sen, J., (2011). *Internet of Things: Applications and Challenges in Technology and Standardization*. *Wireless Personal Communications*, 58(1), 49 - 69.

[7]. Zhou, G., He, T., Krishnamurthy, S., & Stankovic, J., (2006). *Models and solutions for radio irregularity in wireless sensor networks*. *ACM Transactions on Sensor Networks*, 2(2), 221 - 262.

[8]. Ong, B. T., Sugiura, K., & Zettsu, K., (2016). *Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting PM2.5*. *Neural Computing and Applications*, 27(6), 1553 - 1566.

[9]. World Health Organization (WHO) (2021). *Air pollution*. [https://www.who.int/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health).

[10]. AirVisual (2023). *Global air quality index data platform*. <https://www.iqair.com/air-quality-map>.

[11]. Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J., (2021). *Dive into Deep Learning*. Cambridge University Press. Available: <https://d2l.ai/>.