Forecasting $PM_{2.5}$ and NO_2 concentrations in Patras using low-cost sensors and machine learning

Ioannis D. Apostolopoulos
Institute of Chemical Engineering Sciences
Foundation for Research and Technology Hellas
Patras, Greece
japostol@iceht.forth.gr

George Fouskas
Institute of Chemical Engineering Sciences
Foundation for Research and Technology Hellas
Patras, Greece
fouskas@iceht.forth.gr

Spyros N. Pandis
Institute of Chemical Engineering Sciences
Foundation for Research and Technology Hellas, Department of Chemical Engineering, University of Patras
Patras, Greece
spyros@chemeng.upatras.gr

***Abstract*—We present a machine learning methodology for forecasting next day's $PM_{2.5}$ and NO_2 concentrations in Patras, Greece, using a sequence-to-sequence LSTM neural network architecture trained on data from a low-cost ENSENSIA sensor system and a meteorological station in Patras. The model integrates recent pollutant trends, temporal variables, and forecasted meteorology through an exogenous-aware encoder–decoder design with an attention mechanism. We used historical $PM_{2.5}$ and NO_2 concentration measurements from ENSENSIA and meteorological variables from a local weather station to build and train the model. Validation over December 2024 showed promising results, achieving a fractional error of 0.45 and bias of 0.01 for $PM_{2.5}$, and 0.17 and 0.02, respectively, for NO_2. The model underestimated $PM_{2.5}$ peaks, highlighting limitations in modelling pollution events driven by atmospheric chemistry. Our results underscore the potential of combining low-cost sensing and machine learning for urban air quality monitoring.**

Keywords—Air Quality; Machine Learning; Forecast Models

I. INTRODUCTION

The World Health Organization (WHO) reports that 92% of the global population is exposed to high air pollution [1]. $PM_{2.5}$ particles (smaller than 2.5 μm) pose a significant health and environmental risk [2]. Prolonged exposure to $PM_{2.5}$ is linked to heart disease, lung development issues, and increased chances of strokes and respiratory illnesses. NO_2, primarily emitted from vehicle exhaust and industrial activities, has been linked to increased airway inflammation, reduced lung function, and heightened asthma symptoms [3]. Long-term exposure to NO_2 may also contribute to the development of asthma and an elevated risk of respiratory infections [4]. Monitoring these pollutants is challenging due to the limited number of ground stations and the complexity of accurately modelling air quality across diverse regions and timeframes. In Greece, urban centers like Patras regularly experience episodes of elevated particulate and gaseous pollutant concentrations due to vehicular emissions, domestic heating, cooking and other sources. [5].

This gap highlights the need for scalable and affordable air quality monitoring and forecasting solutions in urban environments, where residents are disproportionately exposed to pollution spikes.

Air quality forecasting has relied on statistical models and Chemical Transport Models (CTMs) [6]. Statistical approaches use historical air quality and weather data and are computationally efficient, but cannot capture real-time changes or detailed atmospheric processes. CTMs simulate physical and chemical processes in the atmosphere and track pollutant sources and secondary pollutant formation [5]. However, their accuracy depends on up-to-date emission inventories and precise weather forecasting, both of which introduce uncertainties and biases [7].

Low-cost sensor (LCS) technologies have advanced air quality monitoring by enabling high-resolution, real-time data collection at a significantly lower fraction of the cost of regulatory instrumentation [8]. However, the raw data from LCS often requires calibration and correction due to inherent measurement uncertainties and bias errors [9].

Machine learning (ML) is a promising solution for air quality forecasting, offering powerful data-driven methods to capture the complex temporal dynamics of pollutants [10], [11]. ML algorithms learn directly from historical data to provide predictions of pollutant concentrations. Time-series forecasting models (e.g. recurrent neural networks and gradient-boosted ensembles) have demonstrated the ability to anticipate short-term fluctuations in particulate matter (PM) and gaseous pollutants by recognizing patterns in past measurements [12].

In Thessaloniki, Greece, researchers deployed a network of twelve PurpleAir $PM_{2.5}$ LCSs and trained a long short-term memory (LSTM) neural network [12]. The LSTM, supplemented with meteorological predictors, learned the diurnal and weekly pollution patterns unique to each neighborhood. It achieved good agreement with held-out sensor measurements (correlation 0.67–0.94) and was able to predict over 70% of local $PM_{2.5}$ pollution exceedance events before they occurred.

Another study in Lima, Peru, a city with very limited official monitoring, developed an air quality forecasting system using a network of LCSs [13]. The artificial intelligence system produced real-time pollution maps and short-term forecasts of PM levels across the 3000 km² metropolitan area. Evaluation against reference stations in Lima showed satisfactory accuracy of the ML forecasts.

A recent study in the Puget Sound region (Washington, USA) developed a spatiotemporal random forest model to predict NO_2 concentrations by integrating 117 low-cost electrochemical NO_2 sensors with regulatory monitor data [14]. Including the crowdsourced sensor data boosted the model's spatial coverage by 24%, and while overall cross-validated accuracy remained high ($R^2 \approx 0.85$), predictions in residential zones improved dramatically. The model with LCS inputs captured persistently higher NO_2 levels at downwind suburban sites (owing to local traffic and heating emissions)

2025 IEEE International Smart Cities Conference (ISC2) | 979-8-3315-5773-7/25/$31.00 ©2025 IEEE | DOI: 10.1109/ISC266238.2025.11293331

that a model trained only on sparse official stations would underestimate.

We propose a novel exogenous-aware sequence-to-sequence LSTM framework for next-day $PM_{2.5}$ and NO_2 forecasting that uses low-cost sensor measurements from the ENSENSIA device, developed by FORTH/ICEHT [15], meteorological and temporal variables. Our method uses and inputs the previous 48 hours of pollutant concentrations and the other variables via a bidirectional LSTM (Figure 1). It predicts concentrations of pollutants for the next 24 hours using the forecasted meteorology (temperature, humidity, wind speed/direction). This is the first time ENSENSIA systems are used for air quality forecasting. Moreover, the proposed LSTM framework employs a true sequence-to-sequence encoder-decoder architecture with an additive attention mechanism.

II. Methods

A. Environment Sensing Appliance (ENSENSIA)

The Environment Sensing Appliance (ENSENSIA) is a compact, low-cost air quality monitoring system produced by FORTH/ICEHT for both indoor and ambient environments [8], [9], [15]. ENSENSIA integrates multiple low-cost sensors to measure atmospheric aerosol and gaseous pollutants, including particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2), ozone (O_3), carbon monoxide (CO), and volatile organic compounds (VOCs). The device employs electrochemical and optical sensing technologies to provide real-time data on air quality parameters. ENSENSIA has been utilized in studies focusing on sensor calibration and inter-unit consistency assessment [8], [9], demonstrating its potential for widespread application in environmental monitoring networks.

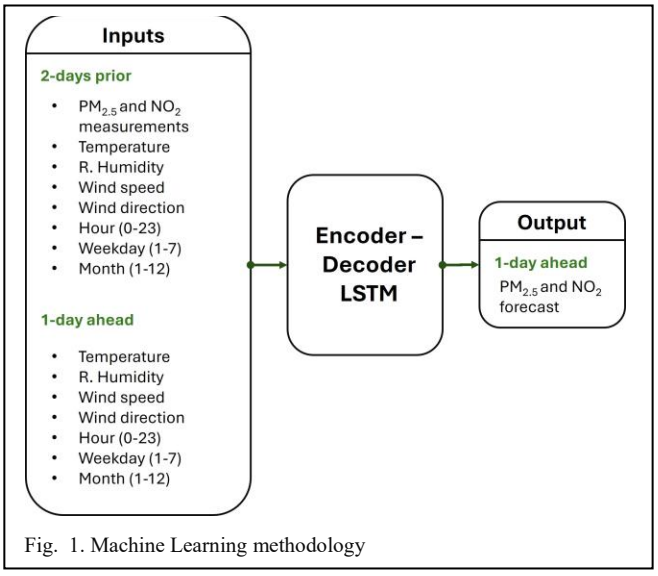

Fig. 1. Machine Learning methodology

B. Machine Learning

The core of our forecasting system is a sequence-to-sequence recurrent neural network that integrates both historical pollutant and meteorological measurements, along with meteorological forecasts via an exogenous-aware encoder–decoder structure (Fig.1.). The model is trained separately for $PM_{2.5}$ and for NO_2, yielding two pollutant-specific instances of the same architecture.

Each training instance consists of two components. First, the encoder receives the previous 48 hourly observations (i.e., a two-day window) of the target pollutant, historical meteorological variable observations (temperature, relative humidity, wind speed, wind direction, and precipitation), and contemporaneous calendar encodings (hour of day, weekday, and month). Second, the decoder is conditioned on the next 24 hours of meteorological forecasts (temperature, relative humidity, wind speed, wind direction, precipitation) plus the corresponding calendar variables. The historical window (48 h) was chosen based on a tradeoff between model's accuracy, complexity, and the use of resources.

The encoder is implemented as a bidirectional long short-term memory (Bi-LSTM) network [16] with 128 memory units in each direction. Bidirectionality enables the encoder to form representations that account for recent temporal trends (e.g., the build-up of a morning peak) and longer-term context (e.g., multi-day shifts due to synoptic changes). The forward and backward hidden and cell states are concatenated, producing a 256-dimensional latent state vector that succinctly summarizes the preceding two-day pollutant trajectory.

The decoder is a unidirectional LSTM with 256 units. It is initialized by the encoder's concatenated states, thereby inheriting all information about past pollutant dynamics. At each of the 24 forecast time steps, the decoder processes the exogenous meteorology and calendar signals, generating its own hidden output. An additive attention mechanism [17] The network then computes a context vector by attending to all 48 encoder time-step outputs, allowing the decoder to dynamically weight which past hours are most informative for each future hour. For instance, if a cold front approaches hour 11 of the forecast, the network can learn to attend most heavily to historical hours that shared similar meteorological precursors.

For each forecast hour, the context vector and the decoder's hidden state are concatenated and passed through a small time-distributed multilayer perceptron: first a 128-unit dense layer with ReLU activation, then a dropout layer (dropout rate 0.3) to combat overfitting, and finally a linear unit to produce the pollutant concentration estimate. This design ensures that each hour's prediction benefits from contextual memory and the decoder's dynamic computation, while the dropout regularizes against memorizing autocorrelated noise.

The model is trained to minimize mean absolute error (MAE), which is robust to occasional high-magnitude pollution spikes compared to mean squared error. We use the Adam optimizer with an initial learning rate of 1×10^{-3}. Early stopping (patience of five epochs) and learning-rate reduction on plateau (factor of 0.5, patience of three) are employed to prevent overfitting and to extract maximum performance from the data. Model checkpoints ensure that the best validation MAE weights are retained.

C. Measurement site

The urban measurement site is in the Patras center, Greece, at latitude 38° 14' 45.976" and longitude 21° 44' 8.036". The ENSENSIA device was placed approximately 6 m above ground on the roof of a small structure used as a regulatory air quality monitoring station by the Region of Western Greece.

D. Validation

For model training, we used historical ENSENSIA measurements and meteorological data in hourly resolution over the period of January 2023 to November 2024. Training features included the measured $PM_{2.5}$ and NO2 from the ENSENSIA device, measured wind speed, direction, and rainfall from the meteorological station, and calendar variables (month, weekday, hour of the day).

For validation, we carried out a pseudo-operational simulation for December 10–19, 2024. At 00:01 EET each day, the trained model issued a 24-hour ahead forecast of hourly $PM_{2.5}$ and NO_2 concentrations, conditioned on the preceding 48 h of observed pollutants, forecasted meteorological inputs, and calendar features. By retaining only the forecast generated at 00:01 for each valid hour, we constructed a continuous time series of 576 hourly predictions (24 days × 24 h), mirroring a real-world daily forecasting routine. Ground truth was provided by the ENSENSIA system at the selected site. The fractional bias (FBIAS) and fractional error (FERROR) were used to evaluate the model performance:

$$\text{FBIAS} = \frac{2}{N}\sum_{i=1}^{N}\frac{(P_i - O_i)}{(P_i + O_i)}$$

$$\text{FERROR} = \frac{2}{N}\sum_{i=1}^{N}\frac{|P_i - O_i|}{(P_i + O_i)}$$

where N is the total number of measurements, P_i is the predicted concentration, and O_i is the corresponding reference concentration.

All simulations and metric computations were performed in Python, using TensorFlow 2 for model inference and standard scientific libraries for data processing. The December 2024 forecasts used model weights corresponding to the lowest validation FERROR during training (on data from 2021–November 2024).

III. Results

A. $PM_{2.5}$

The LSTM model had a fractional error (FERROR) of 0.45 and an almost zero fractional bias (FBIAS) of 0.01 (Table I). The above errors were computed for December 2024 using the forecast model's hourly predictions.

TABLE I. FERROR AND FBIAS OF THE PREDICTION FOR DECEMBER 2024

Pollutant	December 2024	
	FERROR	*FBIAS*
$PM_{2.5}$	0.45	0.01
NO_2	0.17	0.02

As shown in Figure 2a, the model captured the broad temporal structure of $PM_{2.5}$ evolution over the 10-day window, including diurnal variation and multi-day trends. For instance, it identified rising pollution on December 11–12 and a relative dip on December 13 correctly based on the ENSENSIA measurements. However, it underestimated peak values on elevated-pollution days such as December 15, 18, and 19, when observed values exceeded 30 μg/m³ while predictions remained below 27 μg/m³. This is possibly due to the limited representation of extreme events in the training data or sudden local emissions that could not be forecasted.

Table II presents the average daily observed and predicted $PM_{2.5}$. The model showed daily mean $PM_{2.5}$ errors of up to 11 μg/m³ (e.g., December 15 and 18). It successfully preserved rank-order information across days: higher pollution days (e.g., December 14–15 and 19) were still predicted as relatively high, even if the absolute values were underestimated.

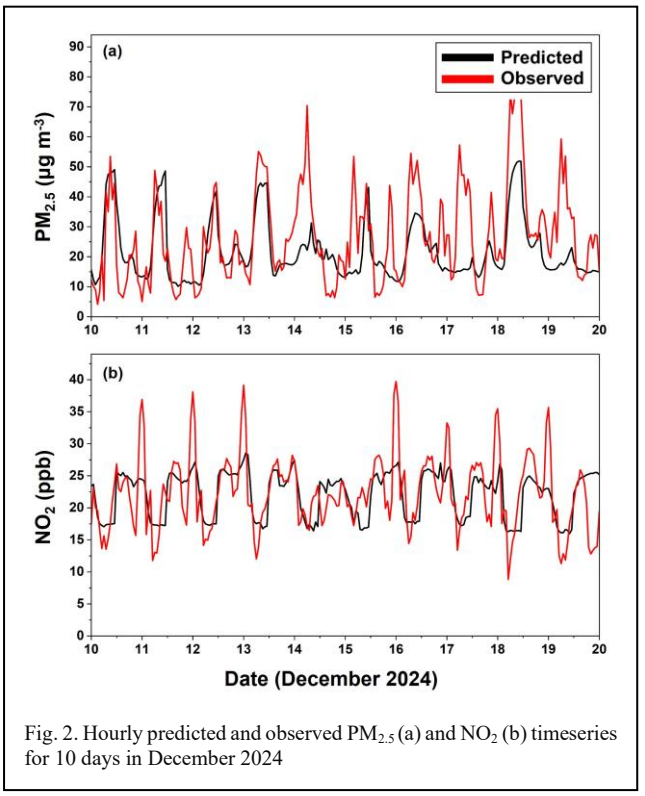

Fig. 2. Hourly predicted and observed $PM_{2.5}$ (a) and NO_2 (b) timeseries for 10 days in December 2024

TABLE II. PREDICTED AND OBSERVED HOURLY-AVERAGE $PM_{2.5}$ AND NO_2 FOR TEN DAYS OF DECEMBER 2024

Day (December 2024)	$PM_{2.5}$ (μg m^{-3})		NO_2 (ppb)	
	Predicted	*Observed*	*Predicted*	*Observed*
10	18	12	22	19
11	22	21	21	19
12	25	19	22	22
13	17	19	23	22
14	27	28	23	23
15	21	32	22	23
16	20	22	21	21
17	20	25	23	24
18	19	31	23	24
19	27	35	21	22

The scatter plot in Figure 3a reflects this underestimation of high values, with predicted points deviating below the identity line at concentrations >30 μg/m³.

Predictions for low-pollution days (e.g., December 10, 13, and 16) tended to be closer to the observed ones.

B. NO_2

Forecasting of NO_2 exhibited a FERROR of 0.17 and a low FBIAS of 0.02 (Table I).

The temporal agreement between predicted and observed NO_2 concentrations was strong ($R^2 = 0.73$). Unlike $PM_{2.5}$, the model showed degradation on high or low concentration days. Daily means in Table II show errors within ±2 ppb, with near-perfect matches on multiple days (e.g., December 14 and 16–18). The model was also robust to minor fluctuations and captured subtle variations in traffic-related peaks, likely due to NO_2's strong temporal structure and direct association with daily human activity cycles.

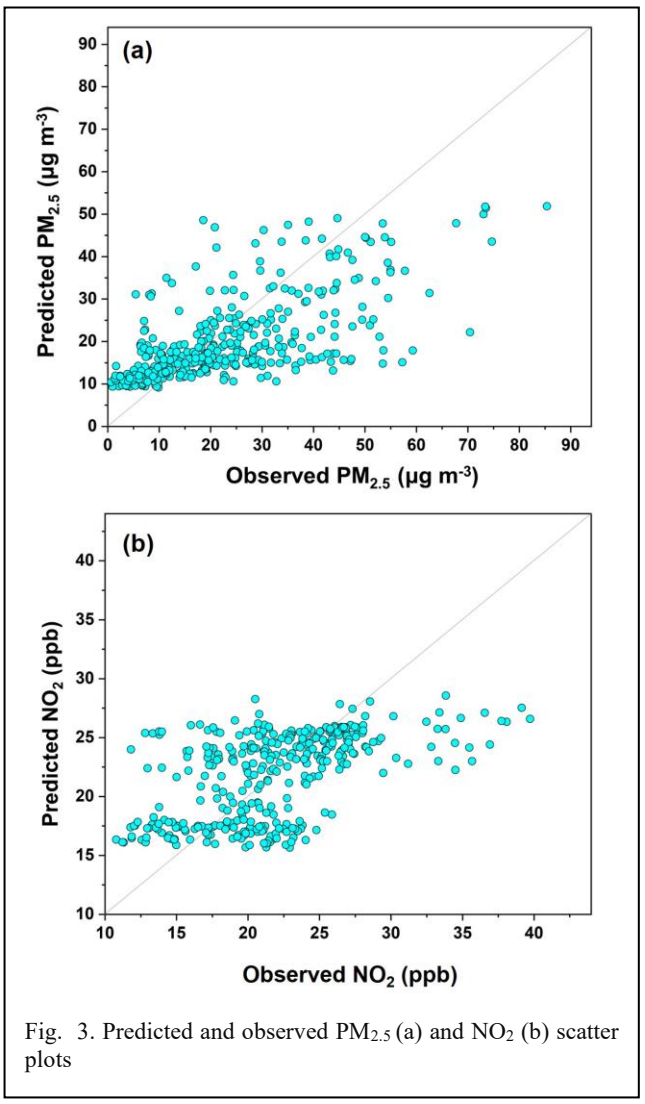

Fig. 3. Predicted and observed $PM_{2.5}$ (a) and NO_2 (b) scatter plots

NO_2 is less sensitive to long-range transport than $PM_{2.5}$, and more influenced by predictable short-term factors like traffic intensity and temperature inversions. The model's reliance on temporal and recent pollutant features aligns well with the pollutant's causal structure.

C. Feature importance

Feature attribution analysis was conducted using both permutation importance and drop-column importance approaches, evaluating the contribution of input features to forecast accuracy.

For $PM_{2.5}$, past pollutant measurements were by far the most critical, contributing 58% of the total relative importance. Omitting this feature led to a ~250% increase in FERROR. Temporal features (e.g., hour of day, weekday) and meteorological inputs also contributed significantly, with future meteorology responsible for a 43% increase in FERROR when removed.

TABLE III. RELATIVE IMPORTANCE OF EACH FORECAST MODEL INPUT AND FERROR INCREASE IN THE ABSENCE OF IT

Feature type	$PM_{2.5}$		NO_2	
	Relative Importance (%)	FERROR increase (%)	Relative Importance (%)	FERROR increase (%)
Temporal	18	23	36	39
Past meteorology	11	25	14	28
Future meteorology	13	43	11	20
Past measurements	58	~250	39	65

NO_2 forecasting displayed a balanced dependency across features. Past concentrations were dominant at 39% relative importance, but both temporal variables and past meteorology also played a strong role, with 36% and 14% relative importance, respectively. Dropping temporal features increased the NO_2 FERROR by 39%.

While pollutant history drives performance, exogenous features such as meteorology and calendar context for the next day significantly affect the prediction, especially for NO_2, which is more directly tied to traffic and time-of-day effects.

D. Comparison against other methods

Table IV presents the performance of the proposed LSTM model alongside a hybrid CNN-LSTM and a multi-layer perceptron (MLP) for December 2024 forecasts. For $PM_{2.5}$, the LSTM achieved a FERROR of 0.45 and a negligible FBIAS of 0.01, slightly outperforming both the CNN-LSTM (FERROR = 0.49, FBIAS = 0.02) and the MLP (FERROR = 0.46, FBIAS = 0.20). In NO_2 forecasting, the LSTM's FERROR of 0.17 and FBIAS of 0.02 closely matched the CNN-LSTM's lower FERROR of 0.15 (but with higher bias) and significantly outperformed the MLP's FERROR of 0.25 and FBIAS of 0.10

TABLE IV. FERROR AND FBIAS OF THE PREDICTION FOR DECEMBER 2024

Method	Pollutant	December 2024	
		FERROR	*FBIAS*
LSTM	$PM_{2.5}$	0.45	0.01
CNN-LSTM		0.49	0.02
MLP		0.46	0.2
LSTM	NO_2	0.17	0.02
CNN-LSTM		0.15	0.1
MLP		0.25	0.1

IV. DISCUSSION AND CONCLUSIONS

This study developed and evaluated a deep learning-based air quality forecasting system for $PM_{2.5}$ and NO_2 concentrations in the city of Patras. A sequence-to-sequence LSTM model was built, trained on ENSENSIA sensor data, and conditioned on forecasted meteorological variables and calendar information.

The results confirm that short-term forecasting of urban air pollutants is feasible using only local low-cost sensor measurements and data-driven models. However, the performance difference between NO_2 and $PM_{2.5}$ points to fundamental limitations in the current approach. NO_2 concentrations are heavily influenced by local and short-term factors like traffic. The model learned traffic patterns in the historical data and projected them for the next day each time. In contrast, $PM_{2.5}$ concentrations are dictated by more complex processes including secondary aerosol formation, long-range transport, and chemical transformations, processes that the LSTM can't model.

Feature importance analysis supports this interpretation. While past concentrations were dominant drivers for both pollutants, $PM_{2.5}$ predictions were significantly more sensitive to future meteorology inputs (43% FERROR increase when excluded). Forecast errors in meteorological inputs will directly propagate into $PM_{2.5}$ prediction errors. This is especially concerning under volatile or extreme conditions, where the model systematically failed to reproduce observed concentration spikes. This points to the potential benefit of hybrid systems that integrate outputs from Chemical Transport Models (CTMs), which simulate the physics and chemistry of the atmosphere, into the deep learning pipeline. Incorporating CTM-based features (e.g., predicted $PM_{2.5}$) may help the LSTM learn and represent chemical pathways contributing to $PM_{2.5}$ variability.

The model currently generates forecasts for a single location, determined by the placement of the ENSENSIA sensor. Urban air quality, however, is spatially heterogeneous; thus, generalizing this approach city-wide would require dense sensor deployments and spatial modeling capabilities (e.g., graph neural networks or geospatial embeddings). Moreover, because the model depends on historical pollutant data for conditioning, deploying it to new locations would necessitate a cold-start period to collect sufficient training data, followed by local retraining or transfer learning strategies.

In conclusion, our study demonstrates that next-day, hourly $PM_{2.5}$ and NO_2 forecasts can be achieved with encouraging accuracy using low-cost sensors. The framework is computationally lightweight, adaptable, and suitable for integrating urban environmental alert systems. Nonetheless, future research should address the incorporation of atmospheric chemical dynamics, expand spatial applicability, and explore the integration of CTM predictions to enhance robustness and scalability.

FUNDING

This work was carried out within the framework of the Action 'Flagship Research Projects in challenging interdisciplinary sectors with practical applications in Greek industry', implemented through the National Recovery and Resilience Plan Greece 2.0 and funded by the European Union – NextGenerationEU (project code: TAEDR-0536642)

References

[1] World Health Organization, "Ambient (outdoor) air pollution," World Health Organization. Accessed: Sep. 09, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

[2] EPA, "Health and Environmental Effects of Particulate Matter (PM)." Accessed: Dec. 07, 2024. [Online]. Available: https://www.epa.gov/pm-pollution/health-and-environmental-effects-particulate-matter-pm

[3] B. Ritz, B. Hoffmann, and A. Peters, "The Effects of Fine Dust, Ozone, and Nitrogen Dioxide on Health," *Deutsches Ärzteblatt international*, Dec. 2019, doi: 10.3238/arztebl.2019.0881.

[4] S. Huang *et al.*, "Long-term exposure to nitrogen dioxide and mortality: A systematic review and meta-analysis," *Science of The Total Environment*, vol. 776, p. 145968, Jul. 2021, doi: 10.1016/j.scitotenv.2021.145968.

[5] E. Siouti, K. Skyllakou, I. Kioutsioukis, D. Patoulias, G. Fouskas, and S. N. Pandis, "Development and Application of the SmartAQ High-Resolution Air Quality and Source Apportionment Forecasting System for European Urban Areas," *Atmosphere*, vol. 13, no. 10, p. 1693, Oct. 2022, doi: 10.3390/atmos13101693.

[6] B. T. Dinkelacker, P. Garcia Rivera, J. D. Marshall, P. J. Adams, and S. N. Pandis, "High-resolution downscaling of source resolved PM2.5 predictions using machine learning models," *Atmospheric Environment*, vol. 310, p. 119967, Oct. 2023, doi: 10.1016/j.atmosenv.2023.119967.

[7] A. Pappa, E. Siouti, S. N. Pandis, and I. Kioutsioukis, "High-resolution WRF forecasts in the SmartAQ system: Evaluation of the meteorological forcing used for PMCAMx predictions in an urban area," *Atmospheric Research*, vol. 296, p. 107041, Dec. 2023, doi: 10.1016/j.atmosres.2023.107041.

[8] I. D. Apostolopoulos, G. Fouskas, and S. N. Pandis, "Field Calibration of a Low-Cost Air Quality Monitoring Device in an Urban Background Site Using Machine Learning Models," *Atmosphere*, vol. 14, no. 2, p. 368, Feb. 2023, doi: 10.3390/atmos14020368.

[9] I. D. Apostolopoulos *et al.*, "Monitoring of Indoor Air Quality in a Classroom Combining a Low-Cost Sensor System and Machine Learning," *Chemosensors*, vol. 13, no. 4, p. 148, Apr. 2025, doi: 10.3390/chemosensors13040148.

[10] C. Bellinger, M. S. Mohomed Jabbar, O. Zaïane, and A. Osornio-Vargas, "A systematic review of data mining and machine learning for air pollution epidemiology," *BMC Public Health*, vol. 17, no. 1, p. 907, Dec. 2017, doi: 10.1186/s12889-017-4914-3.

[11] M. Delavar *et al.*, "A Novel Method for Improving Air Pollution Prediction Based on Machine Learning Approaches: A Case Study Applied to the Capital City of Tehran," *IJGI*, vol. 8, no. 2, p. 99, Feb. 2019, doi: 10.3390/ijgi8020099.

[12] S.-A. Logothetis, G. Kosmopoulos, O. Panagopoulos, V. Salamalikis, and A. Kazantzidis, "Forecasting the Exceedances of PM2.5 in an Urban Area," *Atmosphere*, vol. 15, no. 5, p. 594, May 2024, doi: 10.3390/atmos15050594.

[13] L. Montalvo, D. Fosca, D. Paredes, M. Abarca, C. Saito, and E. Villanueva, "An Air Quality Monitoring and Forecasting System for Lima City With Low-Cost Sensors and Artificial Intelligence Models," *Front. Sustain. Cities*, vol. 4, p. 849762, Jul. 2022, doi: 10.3389/frsc.2022.849762.

[14] C. Zuidema *et al.*, "Leveraging low-cost sensors to predict nitrogen dioxide for epidemiologic exposure assessment," *J Expo Sci Environ Epidemiol*, vol. 35, no. 2, pp. 169–179, Apr. 2025, doi: 10.1038/s41370-024-00667-w.

[15] I. D. Apostolopoulos, G. Fouskas, and S. N. Pandis, "An IoT Integrated Air Quality Monitoring Device Based on Microcomputer Technology and Leading Industry Low-Cost Sensor Solutions," in *Future Access Enablers for Ubiquitous and Intelligent Infrastructures*, vol. 445, D. Perakovic and L. Knapcikova, Eds., in Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications

Engineering, vol. 445. , Cham: Springer International Publishing, 2022, pp. 122–140. doi: 10.1007/978-3-031-15101-9_9.

[16] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "The Performance of LSTM and BiLSTM in Forecasting Time Series," in *2019 IEEE International Conference on Big Data (Big Data)*, Los Angeles, CA, USA: IEEE, Dec. 2019, pp. 3285–3292. doi: 10.1109/BigData47090.2019.9005997.

[17] D. Soydaner, "Attention mechanism in neural networks: where it comes and where it goes," *Neural Comput & Applic*, vol. 34, no. 16, pp. 13371–13385, Aug. 2022, doi: 10.1007/s00521-022-07366-3.