Data-lightweight particulate matter forecasting using temporal attention and Kolmogorov–Arnold Networks

Jarosław Bernacki
Department of Computer Science and Systems Engineering
Wrocław University of Science and Technology
Wrocław, Poland
jaroslaw.bernacki@pwr.edu.pl

Rafał Scherer
Faculty of Computer Science
AGH University of Kraków
Kraków, Poland
rscherer@agh.edu.pl
Department of Artificial Intelligence
Częstochowa University of Technology
Częstochowa, Poland

***Abstract*—Air pollution, particularly particulate matter (PM) concentration, poses a serious threat to public health and the environment. Accurate short-term forecasting of PM levels based on historical concentration data is essential for timely decision-making, public warning systems, and effective air quality management. In this study, we propose an interpretable and data-lightweight approach for predicting $PM_{2.5}$ and PM_{10} concentrations using a model that combines temporal attention with a Kolmogorov-Arnold Network. The proposed model is evaluated using real-world air quality data collected without using meteorological or other auxiliary data. Its performance is assessed using standard error metrics, including mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). The results demonstrate that the proposed model consistently outperforms state-of-the-art deep learning and statistical methods.**

***Index Terms*—forecasting, deep learning, neural networks, particulate matter**

I. Introduction

Air quality forecasting is crucial in protecting public health and ensuring environmental sustainability. Exposure to elevated levels of particulate matter (PM), especially $PM_{2.5}$ and PM_{10}, has been linked to a range of adverse health effects, including respiratory and cardiovascular diseases, as well as premature mortality. In urban areas, where pollutant concentrations often exceed safe limits, accurate and timely predictions of air pollution levels are essential for issuing health advisories, guiding individual behavior, and supporting governmental policy-making. As the impact of air pollution becomes an increasingly pressing global concern, the development of reliable forecasting models has become a vital area of research [1].

Accurate forecasting of particulate matter concentrations is essential in air quality management and public health protection. Elevated levels of $PM_{2.5}$ and PM_{10} are strongly associated with adverse health effects, including respiratory and cardiovascular diseases, particularly among vulnerable populations. Reliable short-term predictions enable early warning systems, support informed decision-making by environmental agencies, and facilitate the implementation of preventive measures aimed at reducing human exposure to harmful air pollutants [2]–[4].

In this study, we present an interpretable and data-lightweight deep learning TA-KAN model based on temporal attention (TA) and a Kolmogorov-Arnold Network (KAN) for $PM_{2.5}$ and PM_{10} concentrations forecasting. The proposed approach models temporal dependencies through a lag-aware attention mechanism, and the KAN component captures non-linear relationships in the PM time series without relying on recurrent or transformer-based architectures. The model is evaluated using real-world historical air quality data collected from multiple monitoring stations. The advantage of the method is that it requires only past PM measurements and does not use the meteorological data. This makes it simpler to implement in real-world scenarios. To assess prediction accuracy, the standard error metrics, including mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE), are used.

a) Contribution: The main contribution of this study is the development of a lightweight and interpretable forecasting framework for $PM_{2.5}$ and PM_{10} that combines temporal attention with a Kolmogorov-Arnold Network (TA-KAN). Unlike many existing approaches, the proposed method relies solely on historical PM measurements, eliminating the need for meteorological or other auxiliary data, while still achieving high predictive accuracy and capturing essential temporal dependencies.

b) Organization of the paper: The paper is organized as follows. Section II presents the proposed TA-KAN method for forecasting the concentration of the $PM_{2.5}$ and PM_{10}. In section III, the experimental evaluation of the proposed methods is performed. The section IV describes previous and related works. Finally, the last section concludes this work. Everywhere in the paper, bold font in formal descriptions denotes matrices or vectors.

II. PROPOSED METHOD

The proposed forecasting framework (TA-KAN) combines a temporal attention mechanism (TA) with a Kolmogorov-Arnold Network (KAN) to forecast PM concentration dynamics in a data-lightweight and interpretable manner. The model operates in a univariate setting and relies exclusively on historical PM observations, avoiding the use of exogenous meteorological variables. For each prediction step, a fixed-length vector of past PM measurements is constructed using a predefined number of temporal lags.

To capture the relative importance of individual temporal lags, a temporal attention mechanism is introduced prior to nonlinear regression. The attention module assigns learnable weights to past observations and produces a weighted aggregation of historical PM values. Unlike recurrent or convolutional architectures, this mechanism does not perform sequential processing but instead provides a direct and interpretable representation of temporal dependencies. As a result, dominant periodic patterns and memory effects in PM dynamics can be identified and analyzed through the learned attention weights.

The aggregated temporal representation is subsequently mapped to the forecasted PM concentration using a Kolmogorov-Arnold Network. KANs approximate nonlinear relationships through compositions of learnable univariate functions, enabling expressive modeling with a limited number of parameters. In the proposed framework, the KAN serves as a nonlinear regressor that captures complex dependencies between historical PM information and future concentrations. The clear separation between temporal dependency modeling and nonlinear function approximation yields a transparent architecture that balances predictive accuracy, interpretability, and computational efficiency, making it well-suited for air quality forecasting in data-constrained monitoring environments.

A. Problem formulation

Let $\{PM(t)\}_{t=1}^{T}$ denote a univariate time series representing particulate matter concentration measurements recorded at regular time intervals. The objective is to model the temporal dynamics of the series and predict future PM concentrations based solely on historical observations, without relying on exogenous variables such as meteorological data. This univariate and input-parsimonious formulation is motivated by the limited availability and reliability of auxiliary measurements in many air quality monitoring systems. Given a fixed maximum lag order K, the historical information available at time t is represented by a lag vector

$$\mathbf{x}_t = [PM(t-1), PM(t-2), \ldots, PM(t-K)]^\top, \qquad (1)$$

which serves as the input to the forecasting model. The value of K determines the temporal memory of the model and is selected to capture both short-term dependencies and dominant periodic patterns present in PM time series.

B. Forecasting task

The forecasting task is defined as a one-step-ahead prediction problem, where the goal is to estimate the future PM concentration $\widehat{PM}(t+1)$ based on the lag vector $\mathbf{x}_t$. Formally, the task consists of learning a nonlinear mapping

$$\widehat{PM}(t+1) = f(\mathbf{x}_t), \qquad (2)$$

where $f(\cdot)$ denotes the forecasting function parameterized by the proposed model. The parameters of $f(\cdot)$ are learned by minimizing a mean squared error loss function over the training dataset between predicted and observed PM concentrations.

C. Temporal attention and Kolmogorov-Arnold Network architecture

To model the relative importance of individual temporal lags, a temporal attention mechanism is introduced as the first stage of the forecasting function $f(\cdot)$. Given the lag vector $\mathbf{x}_t$, the attention module computes a vector of unnormalized relevance scores

$$\mathbf{e}_t = W\mathbf{x}_t + \mathbf{b}, \qquad (3)$$

where $W \in \mathbb{R}^{K\times K}$ and $\mathbf{b} \in \mathbb{R}^K$ are learnable parameters. The attention weights are obtained via a softmax normalization

$$\boldsymbol{\alpha}_t = \mathrm{softmax}(\mathbf{e}_t), \qquad (4)$$

ensuring that $\sum_{k=1}^{K} \alpha_{t,k} = 1$ and $\alpha_{t,k} \geq 0$. Using the attention weights, a weighted temporal representation of historical PM concentrations is computed as

$$\tilde{PM}_t = \sum_{k=1}^{K} \alpha_{t,k} \cdot PM(t-k). \qquad (5)$$

This scalar representation aggregates past observations while explicitly encoding their relative importance, providing an interpretable summary of temporal dependencies. The aggregated feature $\tilde{PM}_t$ is subsequently mapped to the forecasted PM concentration using a Kolmogorov-Arnold Network. Following the Kolmogorov-Arnold representation theorem, the KAN approximates the nonlinear mapping using compositions of learnable univariate functions. In the proposed architecture, the final prediction is given by

$$\widehat{PM}(t+1) = \mathrm{KAN}(\tilde{PM}_t), \qquad (6)$$

where $\mathrm{KAN}(\cdot)$ denotes the nonlinear regression function implemented by the KAN. The separation between temporal dependency modeling (attention) and nonlinear function approximation results in a transparent and computationally efficient forecasting architecture suitable for PM concentration prediction in data-constrained environments.

III. Experimental evaluation

In this section, we evaluate the efficacy of our TA-KAN forecasting model using real-world air quality data provided by the Chief Inspectorate of Environmental Protection in Poland (GIOS)[1]. The dataset contains historical measurements of different air pollutant concentrations, including particulate matter (expressed in $\mu\mathrm{g}/\mathrm{m}^3$) collected from multiple monitoring stations across various Polish cities. The goal is to predict future levels of $\mathrm{PM}_{2.5}$ and PM_{10} based on past concentration values.

A. Collected data and area description

The forecasting evaluation was conducted with the historical $\mathrm{PM}_{2.5}$ and PM_{10} concentrations. They were analyzed in the following air quality monitoring stations in Poland: KpZielBoryTu (Zielonka Bory Tucholskie), MpZakopaSien (Zakopane), SlCzestoZana ($\mathrm{PM}_{2.5}$) / SlCzestoBacz (PM_{10}) (Częstochowa), SlGliwicMewy ($\mathrm{PM}_{2.5}$) / SlKnurJedNar (PM_{10}) (Gliwice/Knurów), and ZpSzczec1Maj (Szczecin). These station codes, along with the city names in parentheses, are used throughout this paper. The data from these stations were chosen due to their complete accessibility, however in Częstochowa the monitoring of the $\mathrm{PM}_{2.5}$ is realized in SlCzestoZana air station; the PM_{10} in SlCzestoBacz station. Moreover, SlGliwicMewy monitors the concentrations of $\mathrm{PM}_{2.5}$; the PM_{10} is realized in SlKnurJedNar. However, this issue should not negatively affect the evaluation, because the stations SlCzestoZana and SlCzestoBacz are located physically very close to each other; a similar note is in the case of SlGliwicMewy and SlKnurJedNar. The locations of the air quality monitoring stations are shown in Fig. 1.

Fig. 1. The location of the air quality monitoring stations in Poland used in the experiment

[1]Data available at `https://www.gov.pl/web/gios-en`

We present the forecasting results based on daily data collected from January 1, 2022, to December 31, 2022. Consequently, the input data for each of the proposed methods consists of 365 daily observations.

B. Implementation

The meta-parameters were set as follows: the TA-KAN model was trained for up to 1000 epochs using the Adam optimizer with a learning rate of 0.001. The ReLU (Rectified Linear Unit) activation function, defined as $\mathrm{ReLU}(x) = \max(0, x)$, was applied in all layers to introduce nonlinearity while maintaining computational efficiency. A fixed forecasting window of the five most recent historical observations was used as input. To prevent overfitting and reduce unnecessary computation, an early stopping mechanism was employed: training was terminated if the validation loss did not decrease for at least 10 consecutive epochs. These parameters were determined experimentally and found to be optimal through multiple trials.

All experiments were conducted on a desktop personal computer equipped with an Intel Core i5-10400 CPU, 32 GB of RAM, and an Nvidia RTX 3060 GPU featuring 12 GB of dedicated video memory. The model was implemented under the PyTorch framework, with CUDA support. Scripts for statistical analysis were implemented in MATLAB.

C. State-of-the-art baseline models

To assess the effectiveness of the proposed hybrid approach, we compared its forecasting performance against both the original data from GIOS and state-of-the-art baseline methods, including a bidirectional LSTM (Bi-LSTM), bidirectional RNN (Bi-RNN), multilayer perceptron (MLP), and ARIMA algorithm.

a) Bidirectional LSTM: A bidirectional LSTM network is used as a baseline for forecasting time series of PM concentrations. It models both forward and backward temporal dependencies within an input window. The model accepts a sequence of historical measurements, which are processed by two Bi-LSTM layers, followed by a dense layer that generates a forecast for the next time step. This architecture allows for better capture of local trends and short-term fluctuations in PM concentrations, as each point in the sequence is encoded with the full temporal context within the window [5].

b) Bidirectional RNN: An alternative baseline is a bidirectional recurrent network (Bi-RNN) based on simple RNN cells that processes a sequence of historical PM concentrations in both temporal directions within a fixed input window. The Bi-RNN layer encodes each sample, taking into account both earlier and later observations in the sequence, and the resulting representation is then mapped by a dense layer to a predicted concentration value. Although classical RNNs are more susceptible to gradient decay and are less capable of modeling long-term dependencies than LSTMs, their low computational complexity and small number of parameters make them a useful benchmark for assessing the extent to which more advanced architectures improve the quality of PM concentration forecasting [6].

c) Multilayer perceptron: The multilayer perceptron (MLP) is a classical feedforward neural network composed of multiple fully connected layers, each followed by nonlinear activation functions. In our configuration, the MLP begins with a layer that expands the input to 64 neurons, followed by a deeper transformation to 128 neurons. Subsequently, the dimensionality is reduced back to 64 neurons, and the final layer performs a projection to the desired output length using a 1×1 convolution, enabling sequence-level predictions [7].

d) ARIMA: Autoregressive integrated moving average (ARIMA) is a widely used statistical model for analyzing and forecasting time series data. It combines three key components: an autoregressive part that models the relationship between current and past values of the series, an integrated part that involves differencing the data to achieve stationarity, and a moving average part that accounts for past forecast errors. The model is specified by three parameters: the order of autoregression p, the degree of differencing d, and the order of the moving average q [8]. In our experiments, we used $p = 5$, $d = 1$, and $q = 0$.

D. Evaluation results

To evaluate the performance of the forecasts generated by both the proposed and baseline methods, we calculate several key error metrics: mean absolute error (MAE, eq. 7), mean square error (MSE, eq. 8), and root mean square error (RMSE, eq. 9).

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - x_t}{n}\right| \tag{7}$$

where x_t is the actual value, y_t is a predicted value and n is the number of observations.

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(x_t - y_t)^2 \tag{8}$$

where x_t is the actual value, y_t is a predicted value and n is the number of observations.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_t - y_t)^2} \tag{9}$$

where x_t is the actual value, and y_t is a predicted value. In Tab. I and II, the error measure values for the $PM_{2.5}$ and PM_{10} forecasts for the proposed and state-of-the-art methods are listed.

The results show that the proposed TA-KAN model consistently outperforms Bi-LSTM, Bi-RNN, MLP, and ARIMA across all monitoring stations for $PM_{2.5}$ forecasting. The MAE of TA-KAN ranges from 0.05 to 0.09, with corresponding RMSE values also being the lowest among all methods, indicating both lower average errors and reduced large deviations. This demonstrates that the temporal attention effectively captures relevant lagged dependencies, while the KAN provides a flexible nonlinear mapping, resulting in robust and accurate predictions even using PM-only inputs.

The results indicate that the proposed method achieves the lowest errors across all monitoring stations for PM_{10} forecasting. The MAE values range from 0.02 to 0.05, with corresponding RMSE values also consistently below those of Bi-LSTM, Bi-RNN, MLP, and ARIMA, demonstrating both high accuracy and stability. These results confirm that the temporal attention effectively identifies relevant lagged dependencies, while the KAN captures nonlinear relationships, resulting in robust predictions using only historical PM measurements. In Fig. 2, the visual examination of the real data and forecasts obtained by the proposed and baseline models is presented. For clarity, we averaged the results achieved by baseline models and presented the chart for a short time step.

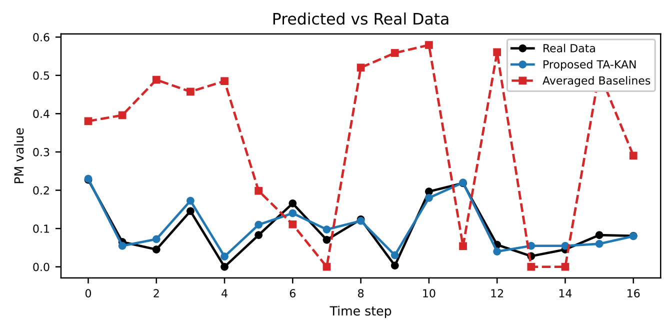

Fig. 2. Comparison of real PM measurements, predictions from the proposed model, and averaged baseline predictions for a partial time step

The proposed method's predictions closely follow the real measurements, whereas the baseline predictions deviate noticeably. This visual comparison highlights the superior accuracy of the proposed TA-KAN model.

IV. Previous and Related Work

The study [9] investigates long-term $PM_{2.5}$ forecasting using multiple machine learning and deep learning models, combined with explainable AI techniques to analyze influencing factors. The results show that ensemble methods outperform deep learning for shorter forecasting horizons, while recurrent neural networks achieve better performance for longer horizons. In [10], a hybrid deep learning model for hourly $PM_{2.5}$ forecasting using historical pollutant data, meteorological variables, and information from neighboring monitoring stations is proposed. Metaheuristic-optimized LSTM models combined with ensemble empirical mode decomposition demonstrate significant improvements in forecasting accuracy, achieving the best performance in terms of RMSE and MAE measures. The results confirm that utilizing neighboring $PM_{2.5}$ information substantially enhances prediction accuracy. The work [11] proposes a hybrid model for accurately forecasting organochlorine pesticide (OCP) concentrations in surface water. The two-stage decomposition combined with LSTM significantly outperforms benchmark and single-decomposition models in terms of SMAPE, demonstrating superior capability in capturing complex temporal dynamics. SHAP-based interpretability analysis further indicates that precipitation and

TABLE I
ERROR MEASURE VALUES FOR THE PROPOSED AND STATE-OF-THE-ART METHODS FROM ALL THE AIR QUALITY MONITORING STATIONS FOR $PM_{2.5}$

	Measure	Proposed	Bi-LSTM	Bi-RNN	MLP	ARIMA
KpZielBoryTu	MAE	0.07	0.15	0.19	0.23	0.31
	MSE	0.03	0.06	0.11	0.19	0.33
	RMSE	0.17	0.24	0.33	0.44	0.57
MpZakopaSien	MAE	0.05	0.15	0.21	0.26	0.28
	MSE	0.04	0.06	0.09	0.21	0.34
	RMSE	0.20	0.24	0.30	0.46	0.58
SlCzestoZana	MAE	0.08	0.16	0.20	0.24	0.34
	MSE	0.04	0.08	0.10	0.25	0.39
	RMSE	0.20	0.28	0.32	0.50	0.62
SlGliwicMewy	MAE	0.09	0.15	0.17	0.25	0.34
	MSE	0.05	0.07	0.11	0.24	0.35
	RMSE	0.22	0.26	0.33	0.49	0.59
ZpSzczec1Maj	MAE	0.08	0.17	0.19	0.22	0.31
	MSE	0.04	0.09	0.12	0.24	0.37
	RMSE	0.20	0.30	0.35	0.49	0.61

TABLE II
ERROR MEASURE VALUES FOR THE PROPOSED AND STATE-OF-THE-ART METHODS FROM ALL THE AIR QUALITY MONITORING STATIONS FOR PM_{10}

	Measure	Proposed	Bi-LSTM	Bi-RNN	MLP	ARIMA
KpZielBoryTu	MAE	0.05	0.22	0.21	0.26	0.31
	MSE	0.05	0.11	0.16	0.26	0.31
	RMSE	0.22	0.33	0.40	0.51	0.56
MpZakopaSien	MAE	0.03	0.17	0.19	0.25	0.33
	MSE	0.05	0.13	0.15	0.25	0.31
	RMSE	0.22	0.36	0.39	0.50	0.56
SlCzestoBacz	MAE	0.03	0.22	0.16	0.25	0.32
	MSE	0.06	0.09	0.16	0.29	0.32
	RMSE	0.24	0.30	0.40	0.54	0.57
SlKnurJedNar	MAE	0.02	0.22	0.20	0.28	0.32
	MSE	0.05	0.10	0.16	0.27	0.34
	RMSE	0.22	0.32	0.40	0.52	0.58
ZpSzczec1Maj	MAE	0.04	0.21	0.19	0.29	0.34
	MSE	0.05	0.11	0.18	0.28	0.33
	RMSE	0.22	0.33	0.42	0.53	0.57

total phosphorus are the most influential factors affecting OCP concentration variability.

In [12], the spatiotemporal variability of air quality index (AQI) using a seasonal ARIMA model is considered. Results show strong seasonal patterns in air pollution, with ARIMA achieving accurate AQI forecasts. The findings highlight both analytical and practical approaches to improving urban air quality in highly polluted developing regions. The study [13] analyzes five years of urban air pollution and meteorological data using multiple linear regression and principal component–based models (PCA) to predict PM concentrations. The model combines principal component analysis with regression, which improves prediction reliability and outperforms standard regression models through multiple error metrics. The results confirm that PCA-based hybrid models provide more accurate and robust air pollution forecasts than conventional linear regression alone. A hybrid CNN–LSTM model for $PM_{2.5}$ forecasting using spatiotemporal correlations between air quality and meteorological data is presented in [14]. CNN extracts latent spatial features while LSTM captures temporal dependencies, resulting in a PM predictor that outperforms standard MLP and LSTM models in overall forecasting. The results demonstrate that the proposed model provides improved accuracy, with only slight seasonal differences where LSTM performs comparably.

The study [15] proposes a stacked LSTM (SS-LSTM) architecture for accurate $PM_{2.5}$ and PM_{10} forecasting, utilizing historical pollutant and meteorological data. The hierarchical stacking of LSTM units enables the model to capture complex interdependencies among multiple pollutants and volatile PM patterns. Experimental results demonstrate that the SS-LSTM model outperforms traditional machine learning and deep learning approaches, providing more accurate and location-aware air quality predictions. In [16], multiple machine learning, statistical, and deep learning models for forecasting $PM_{2.5}$ concentrations are evaluated. An ensemble modeling approach is investigated, combining individual algorithms to improve predictive performance. Results show that hybrid models consistently outperform standalone methods, demonstrating their effectiveness for accurate air quality forecasting.

The research [17] proposes a hybrid AdaBoost-based model for real-time $PM_{2.5}$ prediction in subway environments, outperforming conventional methods such as AdaBoost, random forest, and neural networks. The study [18] presents a hybrid deep learning framework with spatial attention, graph convolution, and temporal convolution for short-term $PM_{2.5}$ and

PM_{10} forecasting, achieving superior accuracy, especially for peak concentrations. In [19], the Air-GAN, an autoencoder-based GAN with CNN-LSTM encoding, captures complex spatiotemporal patterns and reduces inference error by 31.7%, demonstrating robust $PM_{2.5}$ prediction.

In [20], a comprehensive software solution for predicting and forecasting urban air quality by integrating pollutant, meteorological, and traffic data is proposed. The solution was implemented in a web-based IoT platform with real-time alerts, demonstrating its practical applicability for urban air quality monitoring. The paper [21] introduces a dynamic synchronous graph transformer (DSGT) for urban air quality forecasting, capable of capturing multiscale spatiotemporal correlations from sensor networks. The model integrates synchronous graph attention and temporal attention within an encoder-decoder structure, and effectively learns both long-term auxiliary influences and short-term variations. Experiments on real-world datasets show that DSGT outperforms existing methods in both short- and long-term air quality prediction.

V. Conclusion

In this study, we proposed an interpretable and data-lightweight forecasting model based on temporal attention and a Kolmogorov-Arnold Network (TA-KAN) for predicting particulate matter concentrations $PM_{2.5}$ and PM_{10}. The proposed approach was comprehensively evaluated through multiple air quality monitoring stations using standard error metrics, including MAE, MSE, and RMSE. The experimental results demonstrate that the TA-KAN model consistently outperforms widely used baseline methods such as Bi-LSTM, Bi-RNN, MLP, and ARIMA, achieving lower prediction errors and improved stability in different locations. These findings highlight the effectiveness of explicitly modeling temporal dependencies through attention mechanisms combined with interpretable nonlinear regression, and indicate that the proposed approach is well-suited for practical air quality forecasting and environmental monitoring applications, particularly in data-constrained scenarios.

In future research, we plan to develop an alternative method for forecasting air pollutant concentrations, with a focus on long-term prediction. We also aim to further examine the PM concentrations and their components, which remain a major concern in Poland. An important direction will be the inclusion of meteorological data, as it significantly affects air quality.

Acknowledgements

This research received partial financing from the funds assigned to AGH University of Krakow by Polish Ministry of Science and Higher Education.

References

[1] S. S. Eddine, L. B. Drissi, N. Mejjad, J. Mabrouki, and A. A. Romanov, "Machine learning models application for spatiotemporal patterns of particulate matter prediction and forecasting over morocco in north of africa," *Atmospheric Pollution Research*, vol. 15, no. 9, p. 102239, 2024.

[2] A. Salcedo-Bosch, L. Zong, Y. Yang, J. B. Cohen, and S. Lolli, "Forecasting particulate matter concentration in shanghai using a small-scale long-term dataset," *Environmental Sciences Europe*, vol. 37, no. 1, p. 47, 2025.

[3] S. Tariq, J. Loy-Benitez, and C. Yoo, "Enhancing the sustainable management of fine particulate matter-related health risks at subway stations through sequential forecast and gated probabilistic transformer," *Building and Environment*, vol. 244, p. 110780, 2023.

[4] M. Qingsong, R. Xiao, W. Yang, X. Wang, and Y.-z. Kong, "Global burden of pneumoconiosis attributable to occupational particulate matter, gasses, and fumes from 1990˜ 2021 and forecasting the future trends: a population-based study," *Frontiers in Public Health*, vol. 12, p. 1494942, 2025.

[5] 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)*. IEEE, 2019, pp. 3285–3292.

[6] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Stacked bidirectional and unidirectional lstm recurrent neural network for forecasting network-wide traffic state with missing values," *Transportation Research Part C: Emerging Technologies*, vol. 118, p. 102674, 2020.

[7] P. Zhang, Y. Jia, J. Gao, W. Song, and H. Leung, "Short-term rainfall forecasting using multi-layer perceptron," *IEEE Transactions on Big Data*, vol. 6, no. 1, pp. 93–106, 2018.

[8] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "A comparison of arima and lstm in forecasting time series," in *2018 17th IEEE international conference on machine learning and applications (ICMLA)*. Ieee, 2018, pp. 1394–1401.

[9] Y. Zhang, Q. Sun, J. Liu, and O. Petrosian, "Long-term forecasting of air pollution particulate matter (pm2. 5) and analysis of influencing factors," *Sustainability*, vol. 16, no. 1, p. 19, 2023.

[10] N. Zaini, A. N. Ahmed, L. W. Ean, M. F. Chow, and M. A. Malek, "Forecasting of fine particulate matter based on lstm and optimization algorithm," *Journal of Cleaner Production*, vol. 427, p. 139233, 2023.

[11] J. Zhang, L. Dong, H. Huang, and P. Hua, "Elucidating and forecasting the organochlorine pesticides in suspended particulate matter by a two-stage decomposition based interpretable deep learning approach," *Water Research*, vol. 266, p. 122315, 2024.

[12] R.-R. Rahman and A. Kabir, "Spatiotemporal analysis and forecasting of air quality in the greater dhaka region and assessment of a novel particulate matter filtration unit," *Environmental Monitoring and Assessment*, vol. 195, no. 7, p. 824, 2023.

[13] M. T. Zateroglu, "Forecasting particulate matter concentrations by combining statistical models," *Journal of King Saud University-Science*, vol. 36, no. 3, p. 103090, 2024.

[14] U. Pak, Y. Son, K. Kim, J. Kim, M. Jang, K. Kim, and G. Pak, "Novel particulate matter (pm2. 5) forecasting method based on deep learning with suitable spatiotemporal correlation analysis," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 264, p. 106336, 2024.

[15] N. S. Muruganandam and U. Arumugam, "Seminal stacked long short-term memory (ss-lstm) model for forecasting particulate matter (pm2. 5 and pm10)," *Atmosphere*, vol. 13, no. 10, p. 1726, 2022.

[16] P. Srisuradetchai and W. Panichkitkosolkul, "Using ensemble machine learning methods to forecast particulate matter (pm2. 5) in bangkok, thailand," in *International Conference on Multi-disciplinary Trends in Artificial Intelligence*. Springer, 2022, pp. 204–215.

[17] J. Wang, Y. Lu, C. Xin, C. Yoo, and H. Liu, "Kernel pls with adaboost ensemble learning for particulate matters forecasting in subway environment," *Measurement*, vol. 204, p. 111974, 2022.

[18] A. Choudhury, A. I. Middya, and S. Roy, "Attention enhanced hybrid model for spatiotemporal short-term forecasting of particulate matter concentrations," *Sustainable Cities and Society*, vol. 86, p. 104112, 2022.

[19] S. Abirami and P. Chitra, "Regional spatio-temporal forecasting of particulate matter using autoencoder based generative adversarial network," *Stochastic Environmental Research and Risk Assessment*, vol. 36, no. 5, pp. 1255–1276, 2022.

[20] C.-M. Rosca, M. Carbureanu, and A. Stancu, "Data-driven approaches for predicting and forecasting air quality in urban areas," *Applied Sciences*, vol. 15, no. 8, p. 4390, 2025.

[21] H. Xia, X. Chen, B. Chen, and Y. Hu, "Dynamic synchronous graph transformer network for region-level air-quality forecasting," *Neurocomputing*, vol. 616, p. 128924, 2025.