November 2026| Volume 05 | Issue 04 | Pages 63-73

Future Technology
Open Access Journal
ISSN 2832-0379

Journal homepage: https://fupubco.com/futech

https://doi.org/10.55670/fpll.futech.5.4.6

Future Publishing LLC

Article

Hybrid CNN-LSTM model for urban air pollution forecasting: towards early warning systems for public health: a case study of Baghdad and Basra

Israa Nadheer[1], Mohammed Imad Aal-Nouman[2]*

[1]Department of Automation and Artificial Intelligence Engineering, College of Information Engineering, AL-Nahrain University, Baghdad, Iraq
[2]Department of Information and Communications Engineering, College of Information Engineering, AL-Nahrain University, Baghdad, Iraq

ARTICLE INFO

Article history:
Received 10 February 2026
Received in revised form
19 June 2026
Accepted 21 July 2026

Keywords:
Air pollution forecasting, PM2.5, PM10, CNN-LSTM, Early warning system

*Corresponding author
Email address:
m.aalnouman@nahrainuniv.edu.iq

DOI: 10.55670/fpll.futech.5.4.6

ABSTRACT

Air pollution has become a critical issue in environmentally challenged regions due to rapid urbanization, especially in countries where forecasting tools remain underdeveloped. Despite the success of deep learning approaches in predicting air quality, there has been little research on multivariate forecasting of particulate matter in Iraqi cities using publicly accessible environmental information. This study proposes a CNN-LSTM architecture to predict the concentrations of $PM_{2.5}$ and PM_{10} the following day in Baghdad and Basra. Environmental data related to air quality, along with temperature, relative humidity, wind speed, and seasonal cyclical features, were used to generate multivariate time series with a 14-day lookback window. The proposed model was compared with an LSTM model and a conventional autoregressive integrated moving average model using 1,249 daily data samples collected from August 2022 to December 2025. The results reveal that the deep learning models outperformed the statistical baseline, and the LSTM approach exhibited the best generalization performance. A threshold-based pollution alert module was also implemented to demonstrate how continuous concentration forecasts can support operational air quality warning systems. The findings demonstrate the applicability of multivariate deep learning for short-term urban air pollution forecasting in data-constrained environments and provide a baseline framework for future environmental monitoring systems in Iraq.

1. Introduction

In recent years, air pollution has become one of the most critical environmental problems worldwide, posing severe threats to ecosystem sustainability and human health. Fast-paced urbanization, industrial growth, car emissions, and population growth have led to deteriorating air quality in many cities [1]. Various atmospheric pollutants have been identified as causing respiratory diseases, cardiovascular problems, reduced life expectancy, and mortality in people. Thus, predicting particulate matter levels, especially $PM_{2.5}$ and PM_{10}, has become part of environmental management, enabling authorities to take preventive measures to protect people from polluted air and issue advance warnings [2,3]. Because of artificial intelligence, the process of forecasting air quality has become more efficient. Traditional statistical methods were replaced by machine learning techniques and deep learning models, enabling the identification of nonlinear relationships between environmental parameters and pollutant levels [4-6]. Comparative analyses have revealed that Random Forest, Gradient Boosting, and CatBoost are effective at modeling interactions between meteorological variables and pollutants, yielding good results for air quality forecasting [7,8]. Recently, recurrent neural networks, including Long Short-Term Memory (LSTM) architectures and their hybrids, have become popular due to their ability to learn short- and long-term dependencies in multivariate environmental time series. These approaches have greatly enhanced air quality forecasting and enabled the use of forecasting analytics in practical environmental management [9,10]. However, despite numerous achievements, most related studies have been conducted in regions with dense monitoring networks and stable environmental conditions. In contrast, relatively few studies have focused on air pollution forecasting in Middle Eastern cities, where particulate matter levels depend on complex interactions among urban emissions, industrial activities, and frequent dust storms. Iraq is particularly difficult for research because rapid urbanization, increased transportation activity, industrial activity, and dust storms make the situation highly unpredictable. The two largest metropolises in the country, Baghdad and Basra, often experience elevated $PM_{2.5}$ and PM_{10} concentrations, which cause serious health problems for their inhabitants. However, a forecasting system capable of providing reliable short-term predictions for decision-making remains lacking [10,11].

Thus, this study aims to develop a multivariate deep learning framework for next-day prediction of particulate matter concentrations in Baghdad and Basra, based on publicly available air quality and meteorological data. A hybrid architecture combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory Networks (LSTMs) is used to capture local dependencies between variables via convolutional layers and long-term dependencies via LSTM units. The forecasting capabilities of the proposed model are compared to those of a standalone LSTM network and a classical ARIMA model, following the same chronological validation scheme. The framework uses the concentrations of pollutants together with meteorological variables, including temperature, humidity, wind speed, and cyclic seasonal descriptors based on the calendar information, as input. There are four main advantages of this study (Figure 1).

Firstly, a reproducible multivariate forecasting framework is proposed to simultaneously predict $PM_{2.5}$ and PM_{10} in Baghdad and Basra, based on publicly available environmental data. Secondly, the performance of a hybrid CNN-LSTM approach is evaluated against a standalone LSTM network and a classical ARIMA model, using the same chronological validation scheme. Thirdly, the effect of combining convolutional feature extraction with recurrent sequence learning on urban air pollution forecasting is explored. Finally, a threshold-based pollution alert system is designed in order to demonstrate the practical use of continuous concentration forecasts.

2. Related work

Environmental scientists use Recurrent Neural Networks (RNNs) and their advanced variants to create models that predict how environmental data will change over time. The Long Short-Term Memory (LSTM) network architecture enables accurate predictions of particulate matter concentrations by capturing long-term temporal patterns and mitigating the vanishing gradient problem [12,13]. Multiple studies from countries including China, India, and Europe found that LSTM-based systems achieved higher PM2.5 forecasting accuracy across various environmental settings [14,15]. The development of Convolutional Neural Networks (CNNs) for image processing led to their adaptation to time-series studies, enabling the extraction of local feature patterns from sequential environmental data.

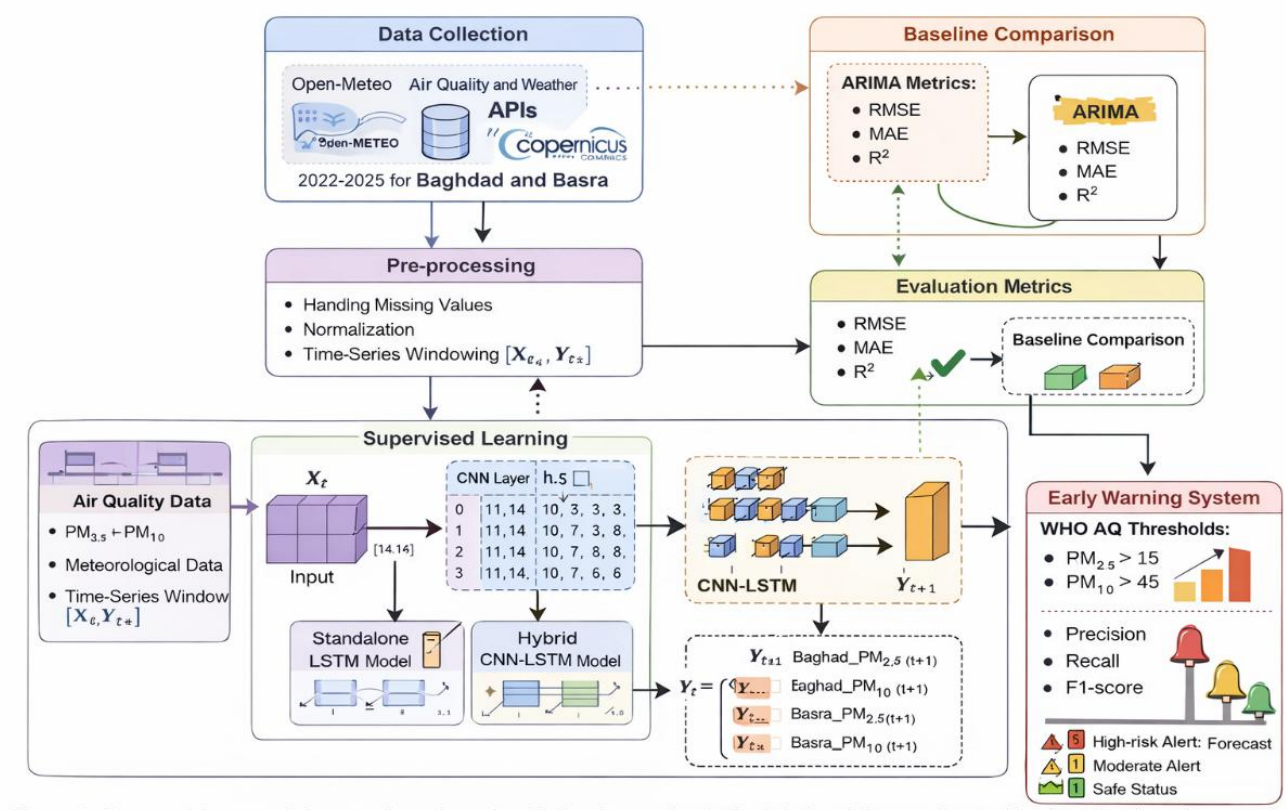

Figure 1. Overall framework of the proposed urban air pollution forecasting system for Baghdad and Basra

The hybrid CNN–LSTM framework demonstrates successful results across multiple environmental monitoring functions, which include particulate matter forecasting, atmospheric pollution modeling, and climate variability analysis [16,17]. Hybrid deep learning techniques have achieved better results in air pollution forecasting than previous methods used in different urban areas. The hybrid architecture, which combines convolutional layers with recurrent networks, has been used to forecast PM2.5 levels in multiple Asian metropolitan areas, achieving better results than its individual components [18,19]. The combination of CNN and LSTM with additional attention mechanisms has been implemented in frameworks to enhance air quality prediction systems, which monitor pollution levels across extensive urban areas [20,21]. New methods now use sophisticated neural network designs, including graph neural networks, attention-based systems, and multimodal architectures, to track how monitoring stations interact with atmospheric transport patterns [22,23].

The latest research shows that environmental management systems should use predictive models to achieve better results in forecasting future events. Machine learning-based systems have been designed to enable cities to monitor environmental conditions, assess health hazards, and develop pollution control measures [24,25]. The systems provide environmental agencies with tools to forecast pollution events and take preventive measures, which include controlling traffic, managing emissions, and issuing public health warnings. The development of dependable short-term prediction systems has become an essential element that supports sustainable environmental management in cities [26,27]. The use of deep learning methods for air quality forecasting yields inconsistent results across different geographic areas despite recent technological progress. The research has primarily focused on three metropolitan areas in East Asia, Europe, and North America, which have extensive monitoring systems that provide sufficient environmental data to train their models [28,29]. Research into air pollution forecasting for Middle Eastern cities has not yet developed sufficiently, as the region faces growing environmental problems from urban development, dust storms, and industrial pollution. The development of predictive systems that match these particular environments will support environmental management needs and public health emergency planning requirements [30,31]. Data-driven forecasting systems, which use current machine learning methods, will enable environmental monitoring and early warning systems in Iraqi urban areas to function effectively [32,33]. The final framework employs a threshold-based early-warning system to detect high-risk pollution events, helping organizations maintain their environmental management programs [34-40].

3. Materials and methods

This study developed a multivariate time-series forecasting framework for next-day prediction of particulate matter concentrations in Baghdad and Basra, Iraq. The methodological pipeline combined publicly available air quality and meteorological records, chronological preprocessing, supervised sequence generation, deep learning model development, statistical baseline comparison, and a threshold-based early warning module. The overall objective was to forecast four pollutant targets for day t+1 using the preceding 14 days of environmental observations.

3.1 Study area and data sources

The study examined two important Iraqi cities, Baghdad and Basra, because the two cities serve as urban centers that frequently experience particulate matter pollution. Baghdad is a densely populated city with three main sources of pollution: vehicle emissions, city activities, and seasonal weather patterns. Basra experiences pollution from three main sources: industrial operations, port activities, and dust storms originating from surrounding areas. The paper used online databases to gather daily air quality measurements and weather data that were publicly available [41-43]. The Open-Meteo air quality service provided researchers with PM2.5 and PM10 concentration data, while meteorological data was accessed through the Open-Meteo historical weather archive. The final predictor set included particulate matter concentrations, temperature at 2 m, relative humidity at 2 m, wind speed at 10 m, and seasonal encodings derived from calendar structure. The merged daily dataset from the observation period yielded 1,249 records between 1 August 2022 and 31 December 2025 after aggregation and preprocessing.

3.2 Variables and target definition

For each city, the retained variables were daily PM2.5, PM10, mean temperature, mean relative humidity, and mean wind speed. Cyclic temporal descriptors that maintained seasonal patterns throughout the study period were created. The system used sine and cosine transformations to represent both the day of the year and the month. The four prediction targets were defined as:

$$y(t+1) = \left[Bagd_{PM2.5(t+1)}, Bagd_{PM10(t+1)}, Bsa_{PM2.5(t+1)}, Bsa_{PM10(t+1)}\right] \tag{1}$$

Thus, the framework performed multivariate one-step-ahead forecasting for both cities simultaneously.

3.3 Data aggregation and preprocessing

The study used timestamp-based merging to achieve hourly-resolution alignment of air quality and meteorological data that they downloaded. The team used arithmetic averaging to convert hourly data into daily data after completing synchronization. The daily mean for a variable x, which is measured every hour on day d was calculated using:

$$x_{daily(d)} = \left(\frac{1}{H}\right) * \ sum(h = 1 \ to \ H) x(d,h) \tag{2}$$

where H denotes the number of hourly measurements available for that day.

Missing daily values were handled through linear interpolation to preserve continuity of the time series. For a missing observation located between two valid points x(t1) and x(t2), the interpolated value at time t was estimated as:

$$x(t) = x(t1) + \left(\frac{(t-t1)}{(t2-t1)}\right) * \left(x(t2) - x(t1)\right) \tag{3}$$

This step avoided sequence fragmentation and allowed subsequent window construction. Because seasonal patterns in air pollution are periodic rather than linear, temporal variables were encoded using trigonometric cyclic transformations. For day of year DOY and month M, the seasonal features were defined as:

$$day\ of\ year_{sin} = \sin\left(\frac{2*pi*DOY}{365.25}\right) \tag{4}$$

$$day\ of\ yearr_{cos} = \cos\left(\frac{2*pi*DOY}{365.25}\right) \tag{5}$$

$$month_{\sin} = \sin\left(\frac{2*pi*M}{12}\right) \quad (6)$$

$$month_{\cos} = \cos\left(\frac{2*pi*M}{12}\right) \quad (7)$$

These encodings preserve circular continuity, such that the transition from the end of one year or month to the beginning of the next is represented smoothly.

3.4 Feature matrix and scaling

After preprocessing, the input matrix X consisted of 14 predictor variables, while the output matrix Y consisted of four forecast targets. Standardization was applied separately to the input and target variables, using statistics computed only from the training subset. For any feature x, standardization was performed as:

$$x_{norm} = \frac{(x - mu_x)}{sigma_x} \quad (8)$$

where mu_x is the training-set mean and $sigma_x$ is the training-set standard deviation. This transformation ensured that predictors with different units and numerical ranges contributed comparably during optimization.

3.5 Supervised sequence construction

The forecasting problem was formulated as sliding-window supervised learning with a lookback horizon of 14 days. For each time step t, the model input consisted of the multivariate observations from days $t - 13$ to t, while the target corresponded to the pollutant concentrations on the day $t + 1$. Accordingly, each input sample was represented as:

$$X_t = [x(t-13), x(t-12), \ldots, x(t)] \quad (9)$$

and the corresponding target vector was:

$$Y_t = y(t+1) \quad (10)$$

where $x(t)$ is the 14-dimensional predictor vector at day t, and $y(t+1)$ is the four-dimensional output vector for the next day.

This transformation produced a three-dimensional tensor with the structure:

$$X_{seq} in\ R^{N\ x\ L\ x\ F} \quad (11)$$

where N is the number of sequences, $L = 14$ is the lookback window length, and $F = 14$ is the number of input features. The target matrix had the form:

$$Input(14,14) \rightarrow Conv1D\begin{pmatrix} 64, kernel_{size} = 3 \\ , padding = same, ReLU \end{pmatrix}$$
$$\rightarrow BatchNormalization$$
$$\rightarrow MaxPooling1D(pool_{size} = 2)$$
$$\rightarrow Conv1D(32, kernel_{size} = 3, padding = same, ReLU)$$
$$\rightarrow BatchNormalization$$
$$\rightarrow LSTM(64)$$
$$\rightarrow Dropout(0.25)$$
$$\rightarrow Dense(64, ReLU)$$
$$\rightarrow Dense(32, ReLU)$$
$$\rightarrow Dense(4)$$

$$Y_{seq} in\ R^{N\ x\ 4} \quad (12)$$

Following chronological splitting, the sequence counts were 859 for training, 173 for validation, and 174 for testing.

3.6 Chronological dataset partitioning

The dataset required a chronological division to maintain its time-series structure while safeguarding against information leakage. Let N_{total} denote the total number of usable daily samples after shifting the targets by one day. The first 70% of the samples were assigned to the training set, the next 15% to the validation set, and the final 15% to the test set. Formally,

$$N_{train} = floor(0.70 * N_{total}) \quad (13)$$

$$N_{val} = floor(0.15 * N_{total}) \quad (14)$$

$$N_{test} = N_{total} - N_{train} - N_{val} \quad (15)$$

This strategy ensured that future observations were never used to predict past observations and that the final evaluation mimicked real forecasting conditions.

3.7 Standalone LSTM model

The study developed the initial deep learning architecture through the creation of a self-contained Long Short-Term Memory network, which solved multivariate sequence problems by analyzing time-dependent relationships. The model received an input tensor of size 14 x 14 and processed it through two stacked LSTM layers followed by fully connected layers. The architectural design of the system has three main components, which can be described as follows:

$$Input(14,14) \rightarrow LSTM\left(64, return_{sequences} = true\right) \rightarrow Dropout(0.2) \rightarrow LSTM(32) \rightarrow Dropout(0.2) \rightarrow Dense(32, ReLU) \rightarrow Dense(4)$$

For an LSTM unit, the internal update equations at time step t were:

$$f_t = sigmoid\left(W_{fx_t} + U_{fh_{t-1}} + b_f\right) \quad (16)$$

$$i_t = sigmoid\left(W_{ix_t} + U_{ih_{t-1}} + b_i\right) \quad (17)$$

$$g_t = \tanh\left(W_{gx_t} + U_{gh_{t-1}} + b_g\right) \quad (18)$$

$$c_t = f_t.* c_{t-1} + i_t.* g_t \quad (19)$$

$$o_t = sigmoid\left(W_{ox_t} + U_{oh_{t-1}} + b_o\right) \quad (20)$$

$$h_t = o_t.* \tanh(c_t) \quad (21)$$

where x_t is the input vector at time t, h_t is the hidden state, c_t is the cell state, and W, U, and b are trainable parameters. The forget gate f_t, input gate i_t, and output gate o_t regulate information flow across time, enabling the network to preserve long-range dependencies.

3.8 Hybrid CNN-LSTM model

The second system used one-dimensional convolutional processing together with recurrent models, which handle time-based data. The convolutional layers extracted short-term local patterns that spread across the entire sequence before the system performed LSTM-based time-summary processing. The network structure was:

For a one-dimensional convolution, the k-th feature map at temporal location t can be written as:

$$z_{k(t)} = phi\left(sum(i = 1\ to\ C) sum(j = 0\ to\ K - 1) w_{k(i,j)} * x_{i(t-j)} + b_k\right) \quad (22)$$

where C is the number of input channels, K is the kernel length, w_k are convolution weights, b_k is the bias term, and $phi(.)$ is the ReLU activation function:

$$ReLU(a) = \max(0, a) \quad (23)$$

The max-pooling layer reduced temporal resolution while it enhanced the most important local pattern. After that, the LSTM layer extracted longer temporal dependencies from the reduced sequence representation.

3.9 ARIMA baseline model

To provide a classical statistical reference, autoregressive integrated moving average models were fitted independently for each target pollutant series. The $ARIMA(p, d, q)$ model can be defined through the following equation:

$$phi(B) * (1 - B)^d * y_t = theta(B) * epsilon_t \quad (24)$$

The equation includes B as the backshift operator, d as the differencing order, $phi(B)$ as the autoregressive polynomial, $theta(B)$ as the moving-average polynomial, and epsilon_t as the white-noise error term. The model can be expressed in its full form as:

$$y'_t = c + sum(i = 1 \; to \; p) phi_i * y'_{t-i} + epsilon_t + sum(j = 1 \; to \; q) theta_j * epsilon_{t-j} \quad (25)$$

where y'_t denotes the differenced series after applying order d. In implementation, ARIMA models were estimated separately for Baghdad PM2.5, Baghdad PM10, Basra PM2.5, and Basra PM10, and then used to generate forecasts over the test horizon.

3.10 Model training

The deep learning models were trained using the Adam optimizer with an initial learning rate of 0.001. The optimization objective was mean squared error. For n samples and m output targets, the loss function was:

$$MSE = \left(\frac{1}{(n*m)}\right) * sum(i = 1 \; to \; n) sum(j = 1 \; to \; m)\left(y_{ij} - yhat_{ij}\right)^2 \quad (26)$$

Mean absolute error was also monitored during training:

$$MAE = \left(\frac{1}{(n*m)}\right) * sum(i = 1 \; to \; n) sum(j = 1 \; to \; m) abs\left(y_{ij} - yhat_{ij}\right) \quad (27)$$

Training was performed with a batch size of 32 and a maximum of 40 epochs. Early stopping was used to terminate training when validation loss ceased improving, thereby limiting overfitting. In addition, adaptive learning-rate reduction was employed, halving the learning rate when the validation loss plateaued. The best model weights were restored according to the minimum validation loss.

3.11 Forecast inversion and post-processing

Because the output targets were standardized prior to training, the model predictions were transformed back to the original units after inference. For any predicted normalized value $yhat_{norm}$, the inverse transformation was:

$$yhat = yhat_{norm} * sigma_y + mu_y \quad (28)$$

where mu_y and $sigma_y$ are the mean and standard deviation of the corresponding target variable calculated from the training data.

3.12 Evaluation metrics

Forecasting performance was assessed using root mean squared error, mean absolute error, and coefficient of determination. For a set of n observations, these metrics were defined as:

$$RMSE = sqrt\left(\left(\frac{1}{n}\right) * sum(i = 1 \; to \; n)(y_i - yhat_i)^2\right) \quad (29)$$

$$MAE = \left(\frac{1}{n}\right) * sum(i = 1 \; to \; n) abs(y_i - yhat_i) \quad (30)$$

$$R2 = 1 - \left(\frac{sum(i=1 \; to \; n)(y_i - yhat_i)^2}{sum(i=1 \; to \; n)(y_i - ybar)^2}\right) \quad (31)$$

The equation includes the observed value y_i, and the predicted value $yhat_i$ and the mean value ybar of the observed values. Better forecasting accuracy is indicated by lower RMSE and MAE, while higher R2 values indicate better explanatory performance. The team calculated metrics for each individual pollutant target before combining results from all four targets to create an overall model comparison.

4. Early warning system

The system combined its ongoing forecasting work with a threshold-based early warning system, which detected days of increased particulate pollution. The warning rule used the WHO daily guideline thresholds to establish its criteria. A day was labeled high-risk if at least one of the following conditions was satisfied:

$$Baghdad_{PM2.5} > 15 \; or \; Basra_{PM2.5} > 15 \; or \; Baghdad_{PM10} > 45 \; or \; Basra_{PM10} > 45.$$

Thus, the binary warning label for day t+1 was defined as:

$$risk(t + 1) = 1, if \; any \; threshold \; exceedance$$

$$risk(t + 1) = 0, otherwise \quad (32)$$

Predicted pollutant concentrations from each model were converted into predicted warning labels using the same threshold logic. The resulting classification outputs were evaluated by precision, recall, and F1-score:

$$Precision = \frac{TP}{(TP + FP)} \quad (33)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (34)$$

$$F1 = \frac{2 * Precision * Recall}{(Precision + Recall)} \quad (35)$$

where TP, FP, and FN denote true positives, false positives, and false negatives, respectively.

The complete pipeline was created in Python, encompassing data acquisition, preprocessing, statistical analysis, and deep learning models/frameworks. The workflow included automated downloading of Baghdad and Basra data through daily data aggregation, data interpolation, data standardization, sequence creation, the training of LSTM and CNN-LSTM models, ARIMA forecasting, inverse scaling, and the computation of both regression and warning-based classification metrics. The final structure was designed to be reproducible and directly extensible to additional cities or pollutant variables.

5. Results and discussions

The final integrated dataset comprised 1,249 daily observations collected in Baghdad and Basra, merging air quality measurements with meteorological variables. The predictor matrix included $PM_{2.5}$, PM_{10}, temperature, relative humidity, wind speed, and cyclic temporal descriptors derived from the day of the year and month. Following chronological preprocessing and supervised sequence generation with a 14-day lookback window, the dataset was partitioned into 859 training, 173 validation, and 174 test sequences (Table 1). Each training sample consisted of an input tensor of dimensions 859 × 14 × 14, representing 14 consecutive days of observations with 14 predictor variables, and a corresponding output tensor of dimensions 859 × 4, representing next-day forecasts for Baghdad $PM_{2.5}$, Baghdad PM_{10}, Basra $PM_{2.5}$, and Basra PM_{10}.

Table 1. Dataset structure and chronological split used for next-day forecasting

Item	Value
Total daily records after preprocessing	1,249
Total variables in merged dataset	15
Forecast targets	4
Lookback window	14 days
Training sequences	859
Validation sequences	173
Test sequences	174
Input tensor shape (train)	859 × 14 × 14
Output tensor shape (train)	859 × 4

The learning behavior of both deep learning models is shown in Table 2 and illustrated in Figure 2. The independent LSTM model demonstrated consistent convergence throughout the learning process, with the training loss decreasing from 1.0104 to 0.6091, and a minimum validation loss of 1.0560 recorded at epoch 12 before early stopping was triggered due to overfitting. On the other hand, the CNN-LSTM model was faster at reducing training loss, with the loss falling to 0.3180 within 16 epochs. However, a minimum validation loss of 1.0753 was recorded at an earlier epoch (8), after which the validation performance degraded despite further decreases in training loss. This difference between training and validation loss indicates that the CNN-LSTM model performed better on the training data but showed poorer generalization compared to the LSTM. Therefore, although the convolutional layer made learning more efficient, it did not improve predictive performance.

Figure 2 shows the distinct convergence properties of the two architectures. Although the CNN-LSTM is characterized by fast training, the widening gap between training and validation losses starting around epoch 8 indicates increasing overfitting. Conversely, the standalone LSTM model converges much more slowly while maintaining stable performance on the validation set, suggesting better generalization given the dataset's size. Table 3 presents the model's forecasting performance for each pollutant. Given the different numerical ranges of the prediction targets, the model evaluation will be based on RMSE, MAE, and R^2 calculated separately for each target. No model can be singled out as performing better than others for each forecast. Generally, the standalone LSTM demonstrated the most consistent behavior on all four targets, while the CNN-LSTM performed better for Basra $PM_{2.5}$ and PM_{10} but worse for the Baghdad datasets. Although the average performance indicates overall model performance, the main point of interest lies in the target-specific results, as the four pollutants span significantly different numerical ranges. For Baghdad $PM_{2.5}$, the ARIMA showed the smallest MAE (7.13), followed by the LSTM (7.77) and CNN-LSTM (8.13). For Baghdad PM_{10}, the ARIMA model also had the smallest MAE, whereas the CNN-LSTM had the largest prediction error. On the other hand, for Basra $PM_{2.5}$ and Basra PM_{10}, the CNN-LSTM performed better than the standalone LSTM, although the difference was rather marginal.

Table 2. Training and validation loss progression for the two deep learning models

Model	First epoch training loss	Final training loss	Best validation loss	Epoch of best validation loss
LSTM	1.0104	0.6091	1.0560	12
CNN-LSTM	1.0178	0.3180	1.0753	8

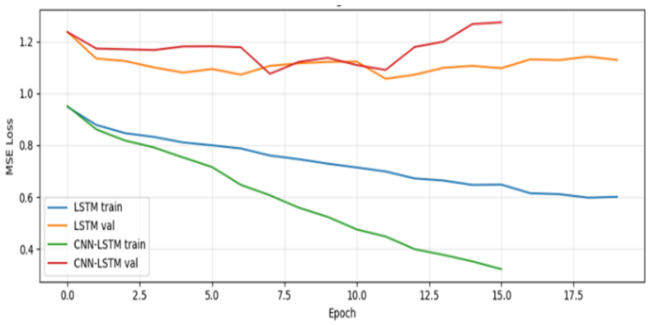

Figure 2. Training and validation loss curves for the standalone LSTM and the hybrid CNN-LSTM

The above shows that the hybrid architecture captures some specific aspects of Basra pollutant dynamics but cannot be regarded as consistently superior to the standalone LSTM for forecasting. The coefficient of determination (R^2) further highlights the limitations of the forecasting models. Negative R^2 values were obtained for Baghdad $PM_{2.5}$, Baghdad PM_{10}, and Basra $PM_{2.5}$ for all models, which means that the predictions were inferior to a simple mean value estimator for these variables. However, Basra PM_{10} was the only pollutant to yield positive R^2 values: 0.3356 for the standalone LSTM and 0.3295 for the CNN-LSTM. Thus, the temporal dynamics of this pollutant were comparatively easy to model. Low values of R^2 indicate that daily particulate matter concentrations vary widely and depend on many factors not included in the predictor set, such as dust storms, emissions, and weather conditions. The behavior of qualitative forecasts is shown in Figure 3 for the case of next-day prediction of Baghdad $PM_{2.5}$. LSTM and CNN-LSTM successfully reproduced the temporal trend but produced smoother forecasts than the actual values. Smoothing is inherent in models trained with mean squared error, as they place less weight on outliers during training. Hence, while both models could replicate the medium-term variability in pollution levels, they were unable to reproduce the largest peaks in pollution recorded in the test dataset.

Table 4 provides an overview of model performance using regression statistics, aggregated across all pollutants. The values given are for comparison purposes only, since aggregating regression errors across variables with different scales can obscure target-specific information. Thus, the results obtained in this study are discussed primarily using the data in Table 3. In this paper, the ability of deep learning models to predict the following day's particulate matter concentration in Baghdad and Basra, based on multivariate time series of environmental data, has been studied.

Table 3. Regression performance for pollutant forecasting

	Target	RMSE	MAE	R^2
ARIMA	Baghdad PM2.5	79.9689	7.1288	-0.0092
ARIMA	Baghdad PM10	360.6260	14.7804	-0.0129
ARIMA	Basra PM2.5	306.9074	13.4431	-0.3311
ARIMA	Basra PM10	22469.4783	110.3830	-0.1349
ARIMA	Average	5804.2451	36.4338	-0.1220
LSTM	Baghdad PM2.5	104.7549	7.7688	-0.3221
LSTM	Baghdad PM10	498.8855	17.6676	-0.4012
LSTM	Basra PM2.5	306.2708	13.5722	-0.3283
LSTM	Basra PM10	13154.3711	94.1899	0.3356
LSTM	Average	3516.0706	33.2996	-0.1790
CNN-LSTM	Baghdad PM2.5	107.0793	8.1316	-0.3514
CNN-LSTM	Baghdad PM10	628.6547	20.5754	-0.7657
CNN-LSTM	Basra PM2.5	282.6543	13.3766	-0.2259
CNN-LSTM	Basra PM10	13274.0576	93.2020	0.3295
CNN-LSTM	Average	3573.1115	33.8214	-0.2534

Table 4. Overall model ranking based on average regression performance

Rank	Model	Avg_RMSE	Avg_MAE	Avg_R^2	Precision	Recall	F1
1	LSTM	3516.0706	33.2996	-0.1790	1.0000	1.0000	1.0000
2	CNN-LSTM	3573.1115	33.8214	-0.2534	1.0000	1.0000	1.0000
3	ARIMA	5804.2451	36.4338	-0.1220	1.0000	1.0000	1.0000

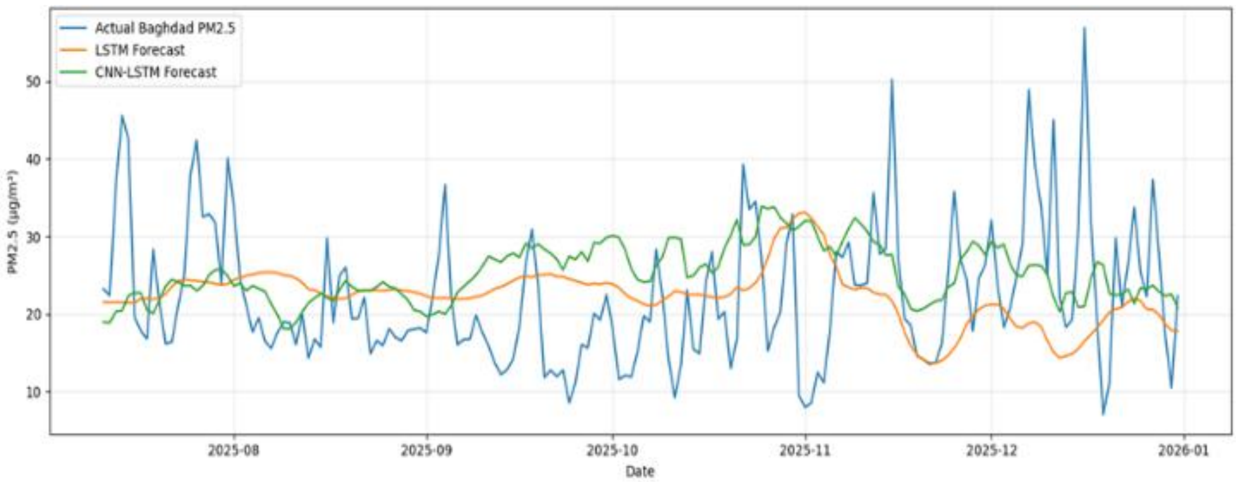

Figure 3. Next-day forecasting results for Baghdad PM2.5

The comparative analysis revealed several important conclusions regarding the applicability of recurrent and hybrid deep learning algorithms to air pollution forecasting when data are limited. Although the deep learning models had advantages over the classical statistical model on certain performance metrics, the expected advantages of the CNN-LSTM model were not observed. Based on the experiments, the LSTM model alone showed the best stability of forecasting performance among all three tested algorithms. In spite of the simpler structure of the network compared to CNN-LSTM, it always had a lower average prediction error than the latter and showed better generalization on the testing dataset. This conclusion is supported by existing literature describing the successful use of LSTM networks to model nonlinear environmental time series, owing to their long-term memory capabilities [5,32]. Air pollutant concentrations depend on the accumulation of meteorological factors and their delayed interactions over several days. Thus, recurrent memory seems to be more significant in this task than local feature extraction. On the other hand, the hybrid CNN-LSTM design achieved a much lower training loss but showed no improvement in validation or testing performance. Such an observation implies that the convolutional part made the model more complex without delivering any extra information. Previous studies have found better forecasting results when employing CNN-LSTM architectures, as convolutional layers enable efficient extraction of temporal and spatial features prior to recurrent learning [20,26]. However, such an advantage can be observed in datasets that contain rich spatial data, multiple monitoring stations, or larger training samples. In the current study, the predictor dataset included measurements from only two cities, along with meteorological data, indicating little spatial diversity in the data. Therefore, under such circumstances, a convolutional layer may extract redundant information and increase the number of trainable parameters, thereby encouraging overfitting. The regression analysis results presented illustrate the problem of forecasting particulate matter concentrations in urban areas of Iraq. Negative R^2 values indicate that daily pollution is very difficult to forecast using the predictor variables in the study. It is wrong to conclude that the neural networks did not work properly. The point is that particulate matter concentration exhibits high dynamics driven by dust storms, local emissions, traffic pollution, industrial activities, and rapid changes in meteorological conditions that cannot be accounted for by the predictor dataset used. Thus, the comparatively better results for Basra PM_{10} indicate that this particulate matter concentration exhibits more stable dynamics.

One further benefit of this research is the development of an early-warning module based on thresholds that converts pollutant concentration forecasts into binary pollution warnings. While all models achieved perfect precision, recall, and F1-score with the adopted thresholding technique, this success cannot be directly attributed to their performance, as it depends on the chosen threshold and the distribution of pollution events, rather than on the regression models' forecasting capabilities. Operational early warning systems usually employ probabilistic forecasts, uncertainty quantification, and event-based validation [13,37], and therefore, the proposed warning module should be viewed as an initial attempt to develop such a system.

In practical terms, the research results show that deep learning approaches are a viable method for short-term urban air quality forecasting in areas where traditional prediction systems are unavailable. Accurate forecasts would help environmental agencies issue warnings about polluted air and optimize both monitoring and preventive actions in the event of increased particulate matter concentrations. In addition, the developed forecasting systems could help protect public health by predicting pollution episodes in advance and issuing warnings to people with cardiovascular and respiratory problems [24,35].

A few more shortcomings are also worth mentioning. Firstly, the dataset was based on publicly accessible atmospheric products and included only two urban areas, resulting in limited spatial variability for model training. Secondly, several key features, such as traffic density, industrial emissions inventory, aerosol optical depth, land-use characteristics, and atmospheric transport indicators, were not included in the predictive algorithm. Thirdly, while a hybrid CNN-LSTM model was compared to both LSTM and ARIMA algorithms, other deep learning architectures, including transformers, graph neural networks, temporal convolutional networks, and attention-based recurrent models, were not examined within this project. The inclusion of the data sources and models mentioned above can significantly improve prediction accuracy, especially for extreme pollution events that remain difficult to predict with the currently considered features [33,40].

Some limitations should be pointed out. Firstly, this project is limited to two cities, and the analysis was conducted solely using publicly available atmospheric and meteorological datasets. Important predictor variables such as traffic density, industrial emissions, aerosol optical depth, and atmospheric transport indicators were excluded from the predictive algorithm. Besides, only ARIMA, LSTM, and CNN-LSTM models were considered, while more recent models, including attention-based models, temporal convolutional networks, graph neural networks, and transformer-based approaches, were beyond the scope of this study. Further research can focus on incorporating additional environmental predictor variables, high-resolution monitoring data, satellite data, and spatial information from various monitoring stations. It may result in improved prediction of extreme pollution events and allow for creating better operational air quality forecasting and warning systems in Iraq and regions with a similar environment.

6. Conclusion

This research designed and implemented a multi-variable deep learning framework for predicting the daily concentration levels of $PM_{2.5}$ and PM_{10} in Baghdad and Basra based on publicly accessible air quality and weather datasets. Two types of deep learning models – CNN-LSTM and LSTM were assessed together with an ARIMA baseline predictor using a 14-day historical window and a chronological test strategy. Both deep learning methods have proven to be better than ARIMA in the majority of forecasting tasks; however, the standalone LSTM model performed more consistently overall, whereas the CNN-LSTM was more prone to overfitting with a lower training loss value. The results show that more complex models do not necessarily yield higher forecasting quality when the environmental dataset is relatively small and unstable. The suggested warning system demonstrates how the pollutant forecasts can be applied to air quality monitoring. Nevertheless, further validation and uncertainty analysis are required for practical application and large-scale testing. Even though this study was restricted to only two cities in Iraq and a limited number of environmental variables, it offers a reproducible baseline for air pollution forecasting in Iraq. Future research should focus on using additional predictive variables, satellite imagery, and advanced deep learning models.

Ethical issue

The authors are aware of and comply with best practices in publication ethics, specifically with regard to authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with policies on research ethics. The authors adhere to publication requirements that the submitted work is original and has not been published elsewhere. All survey participants provided informed consent prior to participation, and the anonymity and confidentiality of all respondents were strictly maintained throughout the research process.

Data availability statement

The manuscript contains all the data. However, additional data will be provided by the corresponding author upon reasonable request.

Conflict of interest

The authors declare no potential conflict of interest.

References

[1] P. Jayaraman, G. N. Manikandan, and P. Partheeban, "Analysis of neural networks for air quality forecasting in populated areas," Environmental Engineering Research, vol. 31, no. 4, p. 250400, 2026. Doi:10.4491/eer.2025.400

[2] Y. He, Y. Jiang, Q. Cao, W. Li, and Z. Chen, "A study on the prediction and driving mechanisms of typical air pollutant concentrations in the Yunnan Plateau region using machine learning," Atmospheric Environment, vol. 372, p. 121880, 2026. Doi: 10.1016/j.atmosenv.2026.121880

[3] C. Erden, "Adaptive learning rate optimization in deep recurrent architectures for precision PM2.5 forecasting under climate variability," Chemosphere, vol. 397, p. 144875, 2026. Doi: 10.1016/j.chemosphere.2026.144875

[4] J. Bhatta, S. R. Acharya, and K. M. Yang, "Machine learning-enhanced air quality forecasting and trend analysis: A five-year comprehensive assessment of PM2.5 concentrations in Bangkok, Thailand," Environmental Challenges, vol. 22, p. 101442, 2026. Doi: 10.1016/j.envc.2026.101442

[5] L. Zeng, R. Dong, M. Yuan, L. Jing, and S. Jiao, "Evaluating deep learning time series models for PM2.5 forecasting across diverse horizons," iScience, vol. 29, no. 2, p. 114770, 2026. DOI: 10.1016/j.isci.2026.114770

[6] P. Karaiskos, Y. Munian, A. Martinez-Molina, and M. Alamaniotis, "Indoor air quality prediction modeling for a naturally ventilated fitness building using RNN-LSTM artificial neural networks," Smart and Sustainable Built Environment, vol. 15, no. 1, pp. 384–406, 2026. Doi: 10.1007/s11869-026-01901-0

[7] Q. Li, M. Li, C. Zhao, and X. Jiang, "Day ahead PM concentration forecast in Fenwei plain of China," Air Quality, Atmosphere and Health, vol. 19, no. 2, p. 37, 2026. Doi: 10.1108/SASBE-10-2023-0308

[8] S. A. Shahriar, Y. Choi, R. Islam, S. R. Shams, and A. K. Salman, "Forecasting PM2.5 and assessing health impacts in Texas using advanced deep learning models," Earth Systems and Environment, vol. 10, no. 1, pp. 819–840, 2026. Doi: 10.1007/s41748-025-00634-x.

[9] Y. Yin, W. Bao, H. Lu, Z. Huang, and X. Lin, "Mechanistic impacts of extra-long tunnel construction on eco-environmentally vulnerable zones: Coupled system dynamics and LSTM-Transformer modelling," Tunnelling and Underground Space Technology, vol. 167, p. 107084, 2026. DOI: 10.1016/j.tust.2025.107084.

[10] Y. Domalatov, K. Chezhimbayeva, B. Issakov, M. Sakitzhanov, and D. Tokseit, "Air pollution forecasting in Almaty based on meteorological data using machine learning for sustainable environmental management," International Journal of Sustainable Development and Planning, vol. 20, no. 12, pp. 5227–5246, 2025. DOI: 10.18280/ijsdp.201217.

[11] A. Adnane, A. Ajdour, and R. Leghrib, "Improving the forecasting of air pollutant gases applying hybrid machine learning methods: Application to extremes," Modeling Earth Systems and Environment, vol. 11, no. 6, p. 436, 2025. Doi: 10.1007/s40808-025-02597-4.

[12] T. Dangare, N. R. Matandirotya, G. Handizvihwe, E. Matandirotya, and E. Mashonjowa, "Leveraging low-cost sensors and machine learning for air quality insights in an urban location of Zimbabwe: A case study," Scientific African, vol. 30, p. e02992, 2025. Doi: 10.1016/j.sciaf.2025.e02992

[13] M. Rajesh, R. G. Babu, U. Moorthy, and S. V. Easwaramoorthy, "Machine learning-driven framework for real-time air quality assessment and predictive environmental health risk mapping," Scientific Reports, vol. 15, no. 1, p. 28801, 2025. Doi: 10.1038/s41598-025-14214-6.

[14] A. Iqbal and N. Mukherjee, "A systematic review and comparative study of machine learning techniques for air quality prediction," Water, Air, and Soil Pollution, vol. 236, no. 12, p. 789, 2025. Doi: 10.1007/s11270-025-08410-3.

[15] X. Gao, X. Wang, F. Li, A. Zhang, and L. Jiao, "PM2.5 concentration 7-day prediction in the Beijing–Tianjin–Hebei region using a novel stacking framework," Scientific Reports, vol. 15, no. 1, p. 20731, 2025. Doi: 10.1038/s41598-025-07719-7.

[16] W. Abuouelezz, N. Ali, Z. Aung, S. B. Shah, and D. Gliddon, "Exploring PM2.5 and PM10 ML forecasting models: A comparative study in the UAE," Scientific Reports, vol. 15, no. 1, p. 9797, 2025. Doi: 10.1038/s41598-025-94013-1.

[17] L. Sharma, H. Singh, and M. P. Choudhary, "Application of deep learning techniques for analysis and prediction of particulate matter at Kota city, India," EQA, vol. 66, pp. 107–115, 2025. Doi: 10.6092/issn.2281-4485/20687.

[18] M. Kaya and İ. Ö. Bucak, "Time series analysis and prediction of PM2.5 pollution concentration in Igdir province with deep learning models," Atmospheric Environment, vol. 361, p. 121491, 2025. Doi: 10.1016/j.atmosenv.2025.121491

[19] S. A. Asklany, D. Mohammed, I. K. Youssef, M. Nawaz, and W. A. Malwi, "Forecasting urban air quality in Paris using ensemble machine learning: A scalable framework for environmental management," PLOS ONE, vol. 20, no. 11, p. e0336897, 2025. Doi: 10.1371/journal.pone.0336897

[20] S. Raja, A. Damodaran, and G. Manohar, "Enhanced forecasting of air quality index through an integrated deep learning framework," Environmental Science and Pollution Research, vol. 32, no. 55, pp. 30503–30516, 2025. Doi: 10.1007/s11356-025-37269-4

[21] H. Bouzghiba, A. Ajdour, N. Omar, A. Mendyl, and G. Géczi, "A novel application of Choquet integral for multi-model fusion in urban PM10 forecasting," Atmosphere, vol. 16, no. 11, p. 1274, 2025. Doi: 10.3390/atmos16111274.

[22] M. E. Akiner, "A comprehensive modeling framework for air quality prediction in Istanbul and CatBoost-SHAP based explainability," Pure and Applied

Geophysics, vol. 182, no. 11, pp. 4771–4803, 2025. Doi: 10.1007/s00024-025-03840-w

[23] N. A. K. and A. Mathew, "Insights into airborne particulate matter: Artificial intelligence-driven PM2.5 modelling in Hyderabad district, India," Stochastic Environmental Research and Risk Assessment, vol. 39, no. 10, pp. 4575–4598, 2025. Doi: 10.1007/s00477-024-02728-w.

[24] C. Mountzouris, G. Protopsaltis, and J. Gialelis, "Short-term forecast of indoor CO2 using attention-based LSTM: A use case of a hospital in Greece," Sensors, vol. 25, no. 17, p. 5382, 2025. Doi: 10.3390/s25175382

[25] M. Su, J. Li, and W. Kim, "Port ship congestion and port-oriented cities air pollution: The role of machine learning models in transportation environmental governance," Transport Policy, vol. 171, pp. 896–915, 2025. Doi: 10.1016/j.tranpol.2025.07.023.

[26] N. Ahmad and V. Kumar, "Effective air pollution prediction using WaveNet deep learning with XGBoost (1DCNN-BiLSTM-XgRC) for urban US embassies," Theoretical and Applied Climatology, vol. 156, no. 9, p. 464, 2025. Doi: 10.1007/s00704-025-05715-5.

[27] Y. Liu and Y. Du, "A deep learning-random forest hybrid model for predicting historical temperature variations driven by air pollution: Methodological insights from Wuhan," Atmosphere, vol. 16, no. 9, p. 1056, 2025. Doi: 10.3390/atmos16091056

[28] R. S. Pathak, V. Pathak, and A. Rai, "A novel attention-based deep learning model for accurate PM2.5 concentration prediction and health impact assessment," Journal of Atmospheric and Solar-Terrestrial Physics, vol. 274, p. 106583, 2025. Doi: 10.1016/j.jastp.2025.106583

[29] A. Issakhov, N. Rysmambetov, and A. Abylkassymova, "Predicting particulate matter (PM2.5) air pollution levels in Almaty city using machine learning techniques," Modeling Earth Systems and Environment, vol. 11, no. 4, p. 236, 2025. Doi: 10.1007/s40808-025-02397-w

[30] S. Helyar and A. Alnaggar, "Air quality monitoring and mitigation through time series forecasting and stochastic optimization," Journal of Environmental Management, vol. 389, p. 125540, 2025. Doi: 10.1016/j.jenvman.2025.125540

[31] F. Zhang, J. Hu, and M. Zeng, "A spatiotemporal multimodal framework for air pollution prediction based on Bayesian optimization—Evidence from Sichuan, China," Atmosphere, vol. 16, no. 8, p. 958, 2025. Doi: 10.3390/atmos16080958

[32] A. Hasnain, A. Sohail, U. A. Bhatti, M. Asif, and M. A. Zia, "Prediction of particulate matter pollution using a long short-term memory model in Zhejiang Province, China," Environmental Earth Sciences, vol. 84, no. 16, p. 478, 2025. Doi: 10.1007/s12665-025-12463-2

[33] F. Illescas-Martinez, L. Garcia, A. J. Garcia-Sanchez, R. Asorey-Cacheda, and J. Garcia-Haro, "Air quality forecasting in non-monitored urban areas through machine and deep-learning model," Expert Systems with Applications, vol. 284, p. 127749, 2025. Doi: 10.1016/j.eswa.2025.127749

[34] S. R. Shams, Y. Choi, D. Singh, S. Kayastha, and J. Park, "Deep learning-based forecasting of daily maximum ozone levels and assessment of socioeconomic and health impacts in South Korea," Science of the Total Environment, vol. 983, p. 179684, 2025. Doi: 10.1016/j.scitotenv.2025.179684

[35] M. Komarudin, S. R. Sulistiyanti, S. Suharso, and A. Sophian, "Advancing precision in air quality forecasting through machine learning integration," International Journal of Artificial Intelligence, vol. 14, no. 3, pp. 2113–2122, 2025. Doi: 10.11591/ijai.v14.i3.pp2113-2122

[36] R. R. He, Y. Q. Chen, L. Tian, L. Shan, and X. S. Sang, "Constructing and evaluating predictors for data-driven PM2.5 forecasting models," International Journal of Environmental Research, vol. 19, no. 3, p. 99, 2025. Doi: 10.1007/s41742-025-00767-x

[37] I. Stergiou, N. Traka, D. Melas, E. Tagaris, and R. E. P. Sotiropoulou, "A deep learning method for improving community multiscale air quality forecast: Bias correction, event detection, and temporal pattern alignment," Atmosphere, vol. 16, no. 6, p. 739, 2025. Doi: 10.3390/atmos16060739

[38] P. Viswanathan and N. Subbu, "Revolutionizing air quality forecasts with optimal nonlinear regression self-organizing mapping recurrent network," Modeling Earth Systems and Environment, vol. 11, no. 3, p. 199, 2025. Doi: 10.1007/s40808-025-02381-4

[39] F. Hu, P. Xie, J. Xu, Y. Li, and W. Q. Liu, "Mapping regional meteorological processes to ozone variability in the North China Plain and the Yangtze River Delta, China," Environmental Science & Technology, vol. 59, no. 19, pp. 9632–9643, 2025. Doi: 10.1021/acs.est.4c11988

[40] S. Raj, J. Smith, and E. Hayes, "Hybrid graph convolutional LSTM model for spatio-temporal air quality transfer learning," Air Quality, Atmosphere and Health, vol. 18, no. 5, pp. 1425–1445, 2025. Doi: 10.1007/s11869-025-01713-8

[41] Open-Meteo, "Open-Meteo Air Quality API: Global Air Quality Data including PM2.5 and PM10," Open-Meteo Environmental Data Services, 2024. Available: https://open-meteo.com/en/docs/air-quality-api

[42] Open-Meteo, "Open-Meteo Historical Weather API: Global Weather Data including Temperature, Humidity, and Wind Speed," Open-Meteo Environmental Data Services, 2024. Available: https://open-meteo.com/en/docs/historical-weather-api

[43] European Centre for Medium-Range Weather Forecasts (ECMWF), "Copernicus Atmosphere Monitoring Service (CAMS): Global Atmospheric Composition Data for Air Quality Monitoring," Copernicus Programme, 2024. Available: https://atmosphere.copernicus.eu

Abbreviations

PM	Particulate Matter
PM2.5	Particulate Matter with diameter less than 2.5 µm
PM10	Particulate Matter with diameter less than 10 µm
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
ARIMA	Autoregressive Integrated Moving Average
RNN	Recurrent Neural Network
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
R^2	Coefficient of Determination
WHO	World Health Organization
API	Application Programming Interface
DOY	Day of Year
TP	True Positive
FP	False Positive
FN	False Negative
ReLU	Rectified Linear Unit