



OPEN Temporal trends and predictive modeling of air pollutants in Delhi: a comparative study of artificial intelligence models

Omer A. Alawi^{1,2}, Haslinda Mohamed Kamar¹, Ali Alsuwaiyan^{3,4} & Zaher Mundher Yaseen⁵✉

Air pollution monitoring and modeling are the most important focus of climate and environment decision-making organizations. The development of new methods for air quality prediction is one of the best strategies for understanding weather contamination. In this research, different air quality parameters were forecasted, including Carbon Monoxide (CO), Nitrogen Monoxide (NO), Nitrogen Dioxide (NO₂), Ozone (O₃), Sulphur Dioxide (SO₂), Fine Particles Matter (PM_{2.5}), Coarse Particles Matter (PM₁₀), and Ammonia (NH₃). Hourly datasets were collected for air quality monitoring stations near Delhi, India, from November 25, 2020 to January 24, 2023. In this context, five intelligent models were developed, including Long Short-Term Memory (LSTM), Bidirectional Long-Short Term Memory (Bi-LSTM), Gated Recurrent Unit (GRU), Multilayer Perceptron (MLP), and Extreme Gradient Boosting (XGBoost). The modelling results revealed that Bi-LSTM model had the best predictability performance for forecasting CO with ($R^2 = 0.979$), NO with ($R^2 = 0.961$), NO₂ with ($R^2 = 0.956$), SO₂ with ($R^2 = 0.955$), PM₁₀ with ($R^2 = 0.9751$) and NH₃ with ($R^2 = 0.971$). Meanwhile, GRU and LSTM models performed better in forecasting O₃ and PM_{2.5} with ($R^2 = 0.9624$) and ($R^2 = 0.973$), respectively. The current research provides illuminating visuals highlighting the potential of deep learning to comprehend air quality modeling, enabling improved environmental decisions.

Keywords Air quality forecasting, Air pollution monitoring, Deep learning, Particulate matter, Environmental assessment

General background of study

Climate change has been a significant issue recently due to its impact on weather conditions and land temperatures, which individuals or natural phenomena can cause. The main causes of air pollution include coal and gas combustion for power generation, transportation, industrial and residential developments¹. Global warming is caused by greenhouse gases (GHG), while climate change is caused by global warming. Emerging nations such as India are facing numerous issues related to air pollution and its detrimental environmental and public health consequences². India is experiencing substantial concerns regarding air quality degradation, such as the massive population growth, industrial companies, the use of fossil fuels for power generation, poor agricultural methods, and motor vehicle emissions^{3,4}. Some particulate matter and gaseous pollutants are produced directly from the source and cause air pollution such as in the size of 2.5 and 10 microns (PM_{2.5} and PM₁₀), carbon dioxide (CO₂), Sulphur dioxide (SO₂), Nitrogen oxide (NO₂), Ammonia (NH₃), benzene, volatile organic compounds (VOCs), carbon monoxide (CO), and ozone (O₃)⁵. The initial air pollutants are used in the general equation to calculate the air quality index (AQI), depending on the geographical region. Besides, these pollutants are the cause of the following issues such as air pollution, depletion of ozone layer, global warming, increased average land temperature, climate change, and acid rain⁶. The air pollutants concentration changes based on the main atmospheric factors such as precipitation (snowfall, rain, sleet or ice pellets, drizzle, hail, freezing rain, frost, and rime), wind speed (WS), wind direction (WD), relative humidity (RH), solar radiation

¹Department of Thermofluids, Department of Mechanical Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor Bahru, Malaysia. ²Department of Power Mechanics Engineering Techniques, Technical Engineering College, Al-Bayan University, Baghdad 10011, Iraq. ³Department of Computer Engineering, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia. ⁴Interdisciplinary Research Center for Intelligent Secure Systems, KFUPM, Dhahran, Saudi Arabia. ⁵Civil and Environmental Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia. ✉email: z.yaseen@kfupm.edu.sa; zaheryaseen88@gmail.com

(SR), and air temperature (T)⁷. In this context, AI approaches such as machine learning and deep learning can be developed using various datasets collected from different monitoring sites and stations to create strong connections between inputs and outputs.

Literature review

As per the open literature^{8,9}, complex AI models were developed by various researchers to predict the air pollutant levels^{10,11}. For instance, the air quality index of four stations such as New Delhi, Bangalore, Kolkata, and Hyderabad, was predicted by using three different regression algorithms such as support vector regression (SVR), random forest (RF), and CatBoost (CR)¹². The RF model showed the lowest root mean square error (RMSE) in Bangalore, Kolkata, and Hyderabad, with a value of 0.5674, 0.1403, and 0.3826, respectively. Meanwhile, the CR model obtained the lowest RMSE value (0.2792) in New Delhi. Additionally, the CR was the superior model in terms of accuracy in New Delhi ($R^2 = 79.8622\%$) and Bangalore ($R^2 = 68.6860\%$), respectively. Two different approaches such as deep learning and Holt-Winters statistical model were compared to predict the $PM_{2.5}$ and PM_{10} concentrations¹³. The Holt-Winters statistical model exhibited lower RMSE and MSE values than the deep learning model. Also, the $PM_{2.5}$ concentration was forecasted using LSTM, which was integrated to the particle swarm optimization (PSO) method¹⁴. They collected their datasets from 15 monitoring stations and achieved a strong accuracy ranging from $R^2 = 0.86$ to $R^2 = 0.99$. Also, the training and testing sets showed an average percentage error over 15 locations of about 6.6% and 6.9%, respectively. In Chennai, the AQI values were classified using a novel expert system integrated from support vector regression (SVR) and LSTM algorithms¹⁵. Their novel expert system showed a superior performance with $R^2 = 0.97$ and RMSE = 10.9.

A case study in Delhi, a dataset was collected from January 2018 to October 2021 to predict the $PM_{2.5}$ concentration using deep learning model that combined neural networks, fuzzy inference systems (ANFIS), and wavelet transforms¹⁶. Their novel model showed an outstanding accuracy, such as: $0.95 < R^2 < 0.99$ for 1-day short term prediction, $0.85 < R^2 < 0.94$ for 2-day short term prediction, and $0.81 < R^2 < 0.93$ for 3-day short term prediction. A novel hybrid model from GRU and LSTM was developed to predict $PM_{2.5}$ concentration in Delhi¹⁷. Also, the dataset was validated using five different standalone models, such as LSTM, linear regression (LR), GRU, K-Nearest Neighbour (KNN), and support vector machine (SVM). The LSTM-GRU was superior to the individual models with an MAE-value of 36.11 and R^2 -value of 0.84. Moreover, the AQI values were predicted in Chennai using historical data from meteorological locations from 2017 to 2022⁶. The authors developed four different tree-based models such as XGBoost, RF, Bagging Regressor, and LGBM. XGBoost model showed the following prediction metrics: $R^2 = 0.9935$, MAE = 0.02, MSE = 0.001, and RMSE = 0.04. Twelve pollutants and ten meteorological parameters from July 2017 to September 2022 were collected over Visakhapatnam, Andhra Pradesh, India¹⁸. This dataset was used to estimate the AQI value using five models such as LightGBM, RF, CatBoost, Adaboost, and XGBoost. The CatBoost model showed the following metrics: $R^2 = 0.9998$, MAE = 0.60, MSE = 0.58, and RMSE = 0.76. Meanwhile, the Adaboost model presented the following metrics with an $R^2 = 0.9753$. Encoder-Decoder (ED) layers were connected to GRU deep learning model for predicting 1-hour, 8-hour, and 24-hour of $PM_{2.5}$ concentrations in New Delhi, India, and the dataset was collected from 2008 to 2010¹⁹. The hybrid method showed superior performance over the standalone models such as (RF, XGBoost, ANNs, and LSTM). Standalone versus stacking models were used to estimate 1-hr and 24-hr $PM_{2.5}$ ²⁰. XGBoost showed higher accuracy than RF and LightGBM with $R^2 = 0.73$. Meanwhile, stacked model (XGBoost as a meta-regressor) improved the accuracy of standalone XGBoost with $R^2 = 0.77$. The eastern region exhibited the best 1-hr prediction with $R^2 = 0.80$ and substantial reduction in Mean Bias ($MB = -0.03 \mu g m^{-3}$), followed by the northern region with $R^2 = 0.63$ and $MB = -0.10 \mu g m^{-3}$. In Chandigarh, eight AI models such as RF, KNN, LR, LASSO regression, Decision Tree (DT), SVR, XGBoost, and Deep Neural Network (DNN) with 5-layers were used to predict 24-hr air pollution and outpatient visits for Acute Respiratory Infections (ARI)²¹. On ARI patients, the RF model performed best, with $R^2 = 0.606$, 0.608 without lag, and 24-hr lag, respectively. Also, on total patients, $R^2 = 0.872$, 0.871 without lag, and 24-hr lag, respectively.

China and India, covering 35% of the global population, face widespread urban air pollution, and real-time and remote sensing assessments remain insufficient. Air quality issues in these rapidly growing economies are increasingly being addressed by the application of machine and deep learning in recent studies. Six models i.e., MLR, SVR, RF, ANN, XGBoost, and LSTM, were used to predict LST for Hyderabad city, India using five-year (2018–2022) data on air pollution and meteorological parameters (from ambient air quality monitoring stations) and MODIS LST data²². Considerable influence of $PM_{2.5}$ and CO (during summer) and SO_2 (during winter) on LST was observed which demonstrated high sensitivity of these parameters on LST. ANN method demonstrated better accuracy with lower error metrics, comprising of RMSE, MAPE, and MSE, compared to the other approaches with ranking in the order ANN > RF > SVR > XGBoost > LSTM > MLR. Based on hourly observations from 2018 in India, integrating temporal and regional features into the LightGBM model led to a notable enhancement in its performance, achieving a 21% reduction in RMSE for $PM_{2.5}$ estimation and a 19% reduction for PM_{10} ²³. The ML model predicted an annual nationwide concentration of $68.3 \mu g/m^3$ for $PM_{2.5}$, which was consistent with high satellite aerosol optical depth (AOD) values. A real-time assessment of hazardous atmospheric pollutants across cities in China (Shanghai, Nanjing, Jinan, Zhengzhou and Beijing) and India (Kolkata, Asansol, Patna, Kanpur and Delhi) was conducted using ground observations, Sentinel-5P and NASA satellite data from 2012 to 2023²⁴. GMAO's SO_2 , NO_2 and CO predictions showed high accuracy with near-perfect PC values and low NRMSE proving model reliability. A Unified Spectro-Spatial Graph Neural Network (USS-GNN) designed for forecasting O_3 - NO_2 concentrations for New Delhi, utilized hourly observations for the years 2021 and 2022²⁵. The proposed model achieved R^2 values of 0.650 and 0.618, RMSE of 13.950 and 16.120 $\mu g/m^3$, MAE of 10.730 and 12.930 $\mu g/m^3$ for O_3 and NO_2 , respectively. Different artificial intelligence models were proposed to simulate climate parameters (1 January 1951–31 December 2022) of Jinan city in China, include ANN, RNN, LSTM, CNN, and CNN-LSTM²⁶. The hybrid CNN-LSTM model significantly

reduced the forecasting error compared to the models for the one-month time step ahead. The RMSE values of the ANN, RNN, LSTM, CNN, and CNN-LSTM models for monthly average atmospheric temperature in the forecasting stage were 2.0669, 1.4416, 1.3482, 0.8015 and 0.6292 °C, respectively.

Research objectives and novelty

This study addresses the critical need for an innovative approach to air quality forecasting, focusing on predicting concentrations of eight key air pollutants that significantly impact human health, environmental integrity, and everyday life. The primary objectives are twofold: (1) to develop accurate forecasting models for Carbon Monoxide (CO), Nitrogen Monoxide (NO), Nitrogen Dioxide (NO₂), Ozone (O₃), Sulphur Dioxide (SO₂), Fine Particulate Matter (PM_{2.5}), Coarse Particulate Matter (PM₁₀), and Ammonia (NH₃); and (2) to rigorously assess and compare the predictive performance of five advanced standalone artificial intelligence (AI) models tailored for univariate pollutant forecasting.

The novelty of this research lies in its comparative analysis across five AI models—Long Short-Term Memory (LSTM), Bidirectional Long-Short Term Memory (Bi-LSTM), Gated Recurrent Unit (GRU), Multilayer Perceptron (MLP), and Extreme Gradient Boosting (XGBoost)—in forecasting hourly pollutant concentrations in a high-pollution urban setting. Using extensive data from monitoring stations in Delhi, collected between 25/11/2020 and 24/01/2023, the study performs in-depth statistical analysis to explore patterns and dynamics of each pollutant. Through the use of scatter plots, Taylor diagrams, and forecasting performance matrices, this study provides comprehensive insights into the distinct strengths and limitations of each model.

In highlighting model efficiency and comparing forecast accuracy metrics, this study provides valuable guidance for advancing air quality management strategies. The findings provide practical implications for urban planning and environmental policy, demonstrating that accurate AI-driven forecasting models can be instrumental in environmental mitigation efforts and health risk assessments in rapidly urban areas.

Data collection and description

As outlined above, the datasets of eight air pollutants were collected from the capital city of Delhi, India, spanning from 25/11/2020 to 24/01/2023 including the following elements: Carbon Monoxide (CO), Nitrogen Monoxide (NO), Nitrogen Dioxide (NO₂), Ozone (O₃), Sulphur Dioxide (SO₂), Fine Particles Matter (PM_{2.5}), Coarse Particles Matter (PM₁₀), and Ammonia (NH₃). The dataset contains features recorded at an hourly interval, with 18,776 samples. It appears that the datasets are complete, and no missing values were observed. Table 1 shows summary statistics for the eight air pollutants, using various measurements such as Mean, Standard Deviation, Variance, Skewness, Kurtosis, Coefficient of Variation, Median, Interquartile Range (Q₃ - Q₁), Range (Maximum - Minimum), Median Absolute Deviation, and Robust Coefficient of Variation. Besides, Fig. 1 shows the numerical distributions of eight air pollutants through boxplot formats. Boxplots highlighted extreme outliers that significantly changed the common relationship of the dataset. In this regard, the Winsorizing technique was implemented to manage outliers by replacing extreme values with values closer to the upper or lower limits, resulting in a more robust and reliable analysis of the air quality data.

Applied artificial intelligence models

This section outlines the five models employed for predicting air pollutant concentrations. The development and configuration processes of the models are then reviewed. Finally, the section concludes with a summary of the forecasting metrics used to evaluate the accuracy of the model predictions.

Multilayer perceptron (MLP)

An MLP is a fundamental component of ANNs²⁷. It can model complex systems in engineering and data sciences. An MLP is characterized by the number of layers of interconnected neurons and the number of neurons per layer. Every layer transforms its input data and feeds it to the next layer in a purely feedforward fashion, contributing to the network's decision-making process²⁸. Figure 2a illustrates an MLP with an input layer, a single hidden layer, and an output layer. Every neuron implements a weighted sum with a bias term fed to an activation function, e.g., tanh. Figure 2b shows the structure of a neuron²⁹.

The architecture of an MLP, including the number of layers and neurons in every layer, significantly affects its performance. Additionally, the use of hyperparameters, such as the learning algorithm, is crucial in training MLPs to minimize their prediction errors effectively. The objective of the learning algorithm is to solve for the weights and biases that minimize an objective function³⁰. MLPs are capable of capturing non-linear data, making them suitable for various applications, including air pollutant prediction. One crucial property of MLPs is that the signals propagate from the inputs to the outputs in one direction, i.e., MLPs do not contain feedback loops. This property of MLPs limits them from capturing long-lasting relationships of time series. This is precisely why the prediction error rates are higher than the other models.

Long short-term memory (LSTM)

To understand the improvements offered by Long Short-Term Memory (LSTM) networks, the limitations of recurrent neural networks (RNNs) are first examined, as these are addressed by the LSTM architecture. RNNs contain both feedforward and feedback loops that facilitate capturing complicated temporal dynamics of sequential time series data²⁷. The main limitation of RNNs is that they fail to capture long-term dependency in the time series, and this is due to the vanishing and exploding gradients²⁷.

Schmidhuber and Hochreiter³¹ proposed the LSTM model in 1997 to solve the issue of vanishing and exploding gradients in RNNs. One of the proposed ideas is the cell state, which is intended to enhance the hidden state and to propagate the effect of some data through longer sequences. Also, LSTM introduces three

	<i>N</i> total	<i>N</i> missing	Mean	Standard deviation	Variance	Skewness	Kurtosis	Coefficient of variation	Median	Interquartile range (<i>Q</i> ₃ – <i>Q</i> ₁)	Range (maximum–minimum)	Median absolute deviation	Robust coefficient of variation
CO	18,776	0	2929.229	2854.524	8148304.444	2.005	4.340	0.975	1842.500	2616.880	20888.330	974.660	0.784
NO	18,776	0	33.661	62.127	3859.779	2.808	9.087	1.846	5.250	35.080	500.680	5.250	1.483
NO ₂	18,776	0	66.221	48.527	2354.917	2.034	6.487	0.733	54.150	49.700	456.340	23.300	0.638
O ₃	18,776	0	60.346	80.465	6474.605	1.997	5.744	1.333	27.180	92.640	801.090	27.180	1.483
SO ₂	18,776	0	66.694	49.439	2444.234	2.670	12.074	0.741	52.930	47.210	574.580	21.460	0.601
PM _{2.5}	18,776	0	238.130	226.534	51317.483	1.949	4.456	0.951	157.445	228.560	1696.260	92.630	0.872
PM ₁₀	18,776	0	300.093	267.166	71377.579	1.868	4.023	0.890	209.705	269.205	1954.860	114.565	0.810
NH ₃	18,776	0	25.110	26.402	697.071	3.568	18.752	1.051	17.480	20.770	287.770	9.250	0.785

Table 1. The statistical properties of air pollutants dataset.

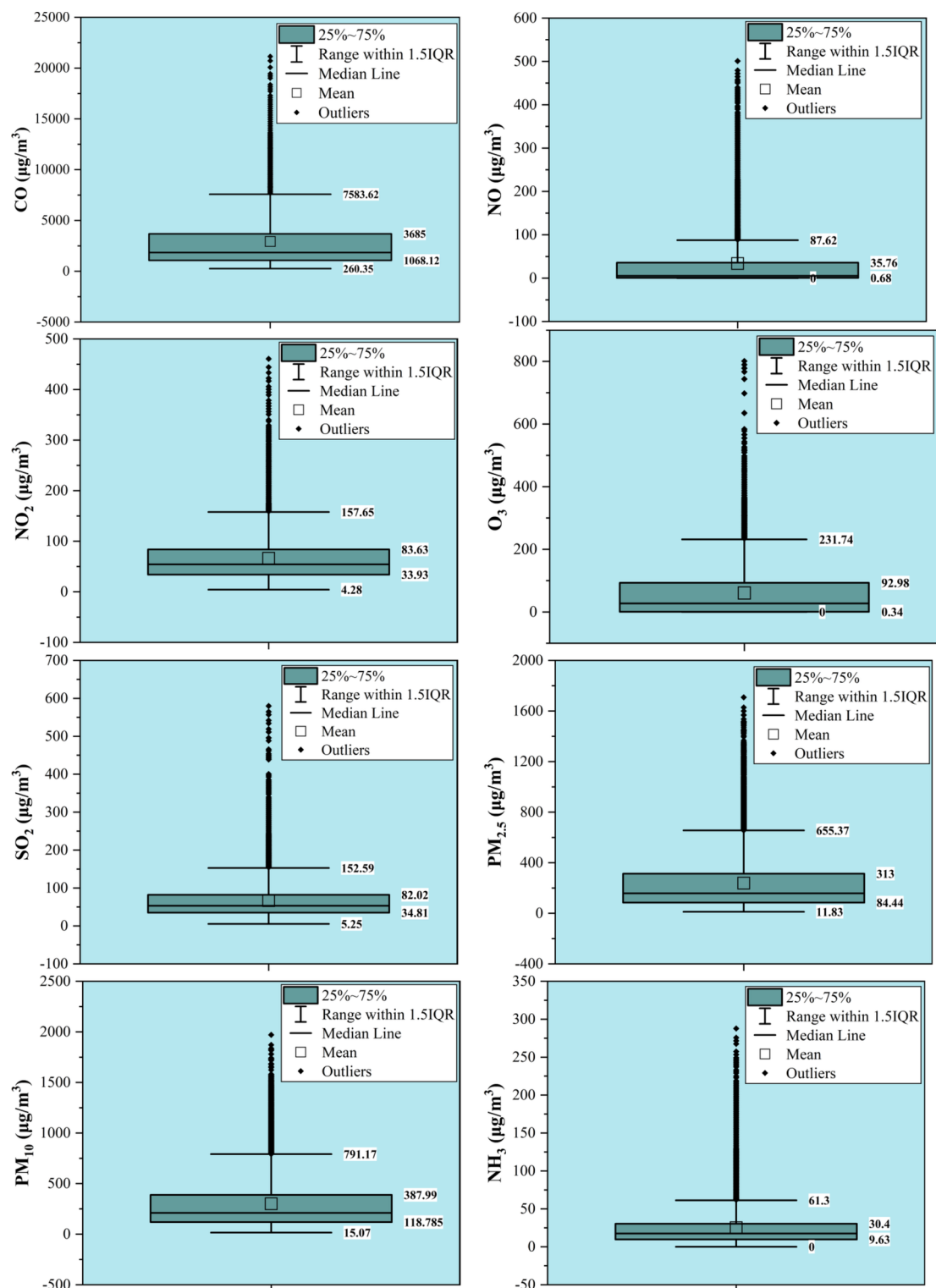


Fig. 1. Boxplots distributions of eight air pollutants.

elements that are intended to filter the data flowing through the network in such a way to prevent the vanishing and exploding gradients, and thus maintain that series of long-term dependence. These elements are known as input gate, forget gate, and output gate, as shown in Fig. 3. In particular, the forget gate selects what information to remember or forget. The value of the forget gate is a value between 0 and 1, where 0 means to forget the information and 1 means to remember, and a value in between means partially forget/remember.

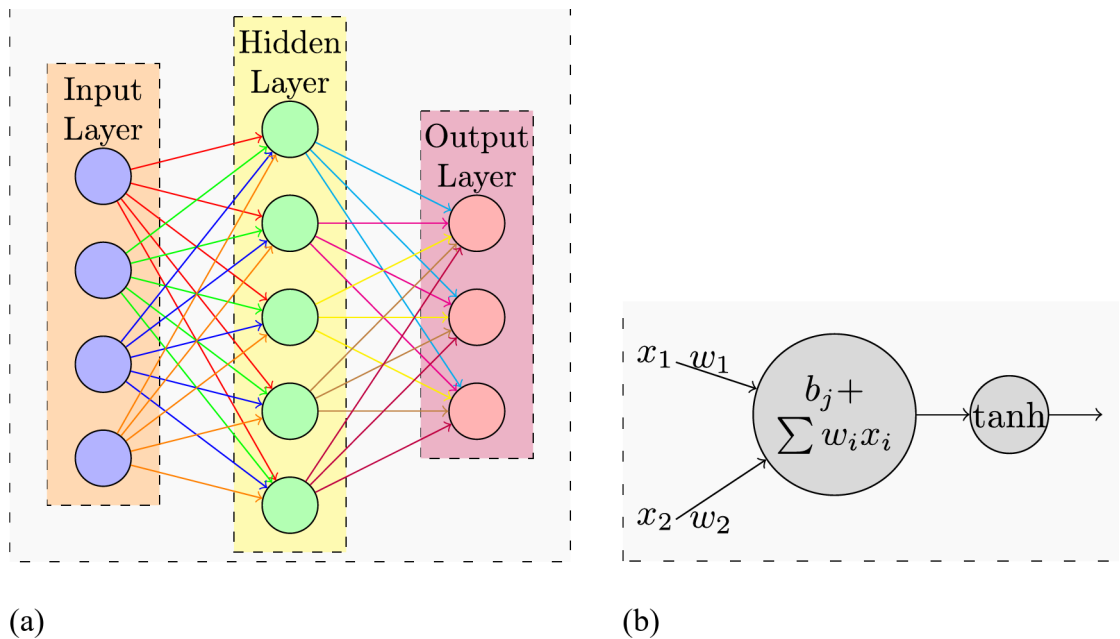


Fig. 2. (a) An example architecture of an MLP consisting of an input layer, a single hidden layer, and an output layer of neurons (b) An example architecture of a neuron with two inputs and tanh activation function.

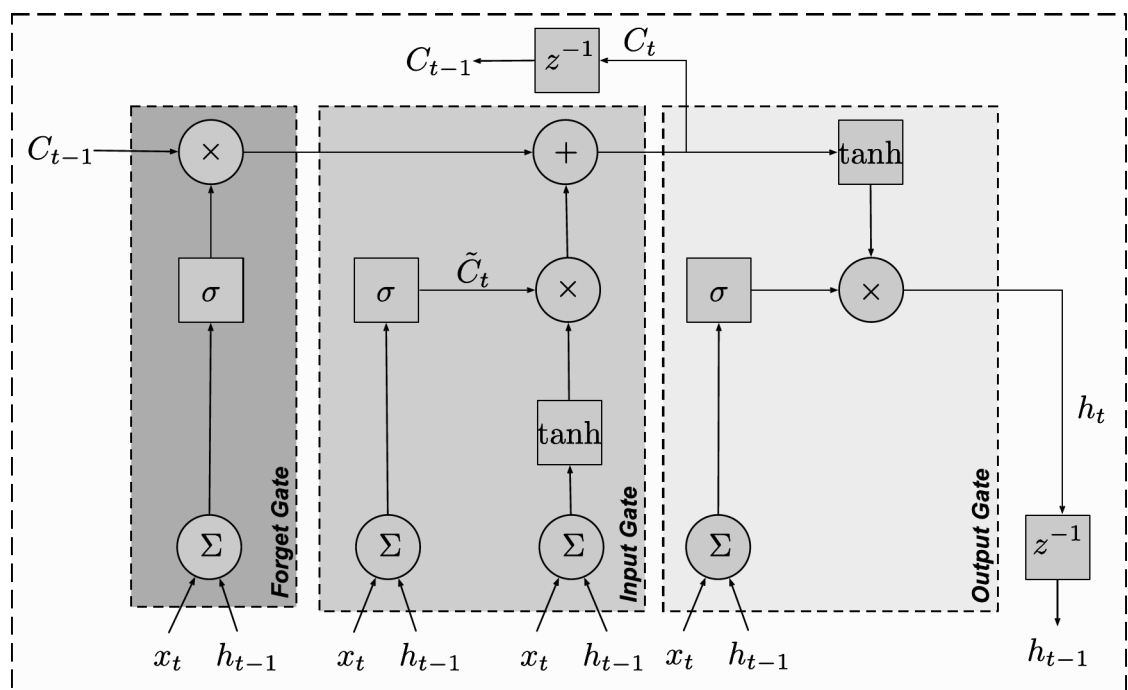


Fig. 3. Detailed LSTM Cell Architecture. Arrows indicate the flow of information, including the feedback loops for the cell state (C_t) and hidden state (h_t). Note that the summation nodes include bias terms.

Bidirectional long-short term memory (Bi-LSTM)

As proposed by Schuster and Paliwal in³², Bi-LSTM operates two instances of LSTM, one for processing the sequence in the forward direction and the other for processing the sequence in the reverse direction, enhancing the model to learn from the sequence in both directions. Bi-LSTM then combines the responses of both directions to generate a unified output sequence that can be used to predict data more accurately³³.

Gated recurrent unit (GRU)

GRU improves LSTM by reducing the number of gates from three to two³⁴. One gate is the update gate, and the other is the reset gate. The primary objective of these gates is to capture the long-term dependency between the data elements of a time series. The main objective of the update gate is to determine how many dependencies to remember, whereas for the reset gate, the main objective is to determine how many dependencies to forget. GRU utilizes a sequence of computations to generate the new hidden state, given the input and the current hidden state. This sequence of computations, i.e., Eqs. (1)–(4), shown in the following table, is then repeated in the form of a loop, to improve the accuracy of GRU to predict new data and to effectively capture the long-lasting relationship between the elements of the time series.

The update gate decides how much to remember

$$u^{(t)} = \sigma \left(W_{uh} h^{(t-1)} + W_{ux} x^{(t)} + b_u \right) \quad (1)$$

The reset gate decides how much to forget

$$r^{(t)} = \sigma \left(W_{rh} h^{(t-1)} + W_{rx} x^{(t)} + b_r \right) \quad (2)$$

The candidate hidden state computation according to reset gate feedback

$$\tilde{h}^{(t)} = \tanh \left(W_{hc} \left(r^{(t)} \odot h^{(t-1)} \right) + W_{hx} x^{(t)} + b_h \right) \quad (3)$$

The new hidden state computation, according to feedback from the update gate and candidate hidden state

$$h^{(t)} = (1 - u^{(t)}) \odot h^{(t-1)} + u^{(t)} \odot \tilde{h}^{(t)} \quad (4)$$

Extreme gradient boosting (XGBoost)

All the previous models are proposed to optimize the performance of the gradient boosting machine (GBM)³⁵, XGBoost integrates a number of weak learners to form an efficient predictive model as illustrated in Fig. 4³⁶. XGBoost generates k decision trees and validates them using a subset of the training data such that the model generalizes well to new data. Each decision tree creates a residual, i.e., the error in prediction of the current

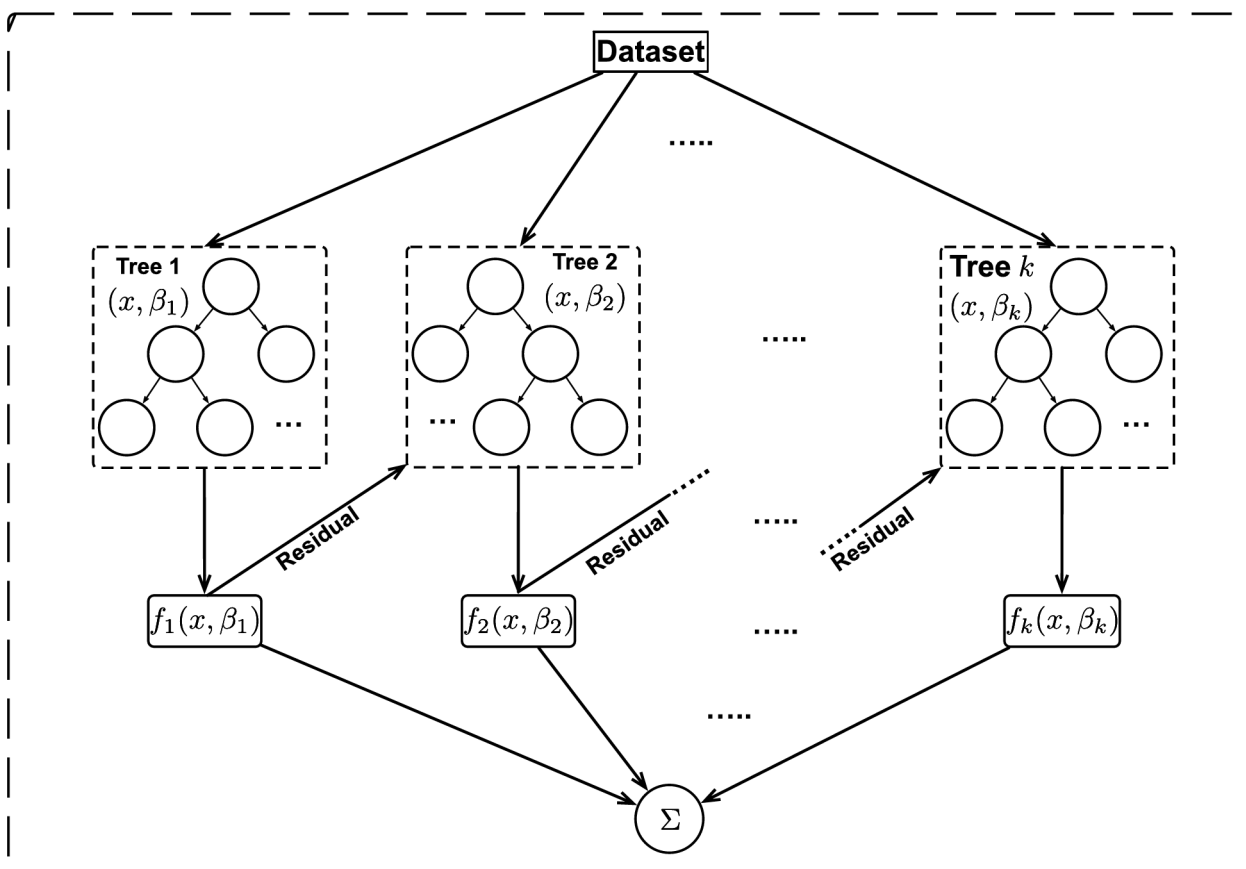


Fig. 4. XGBoost architecture consists of k decision trees that iteratively improves the prediction accuracy.

decision tree, and this residual becomes the target variable for prediction in the next decision tree. XGBoost has been proven to be effective in predictions requiring high levels of accuracy, e.g., see^{28,37–43}.

Models development and configuration

Figure 5 shows the development and configuration process. The implementation platform was Google Collaboratory, utilizing Python and powered by GCC 9.4.0. Many libraries were utilized to complete this project, including NumPy, Pandas, Scikit-learn, and the APIs for TensorFlow Keras. The input dataset contains the hours of concentrations of eight air pollutants in Delhi between November 25th, 2020, and January 24th, 2023. Prior to any actions being executed on the input dataset, it was standardized using scikit-learn and then split into 80–20% training and testing sets.

The training set uses a filter to substitute outliers by values that are closer to the upper or lower bound of the dataset when these outliers are excluded. After that, the filtered data is processed to each of the five models to complete the learning process. The test data is utilized to evaluate the accuracy of the trained models by generating a set of six performance metrics, such as RMSE, R^2 , MAE, MSE, standard deviation and WI.

Prior to the training, hyperparameter optimization and tuning were conducted to maximize the performance of each model. This tuning process was conducted using scikit-learn algorithms specifically developed for such purposes. The GridSearchCV algorithm was employed for neural network-based models, while for the XGBoost model, the RandomizedSearchCV algorithm was employed. The optimized hyperparameters for all models are presented in Table 2.

Forecasting metrics

There are several statistical performance metrics were adopted for the forecasting accuracy evaluation for each developed predictive model. The mathematical expression of each metric are as follows:

Coefficient of Determination (R^2)

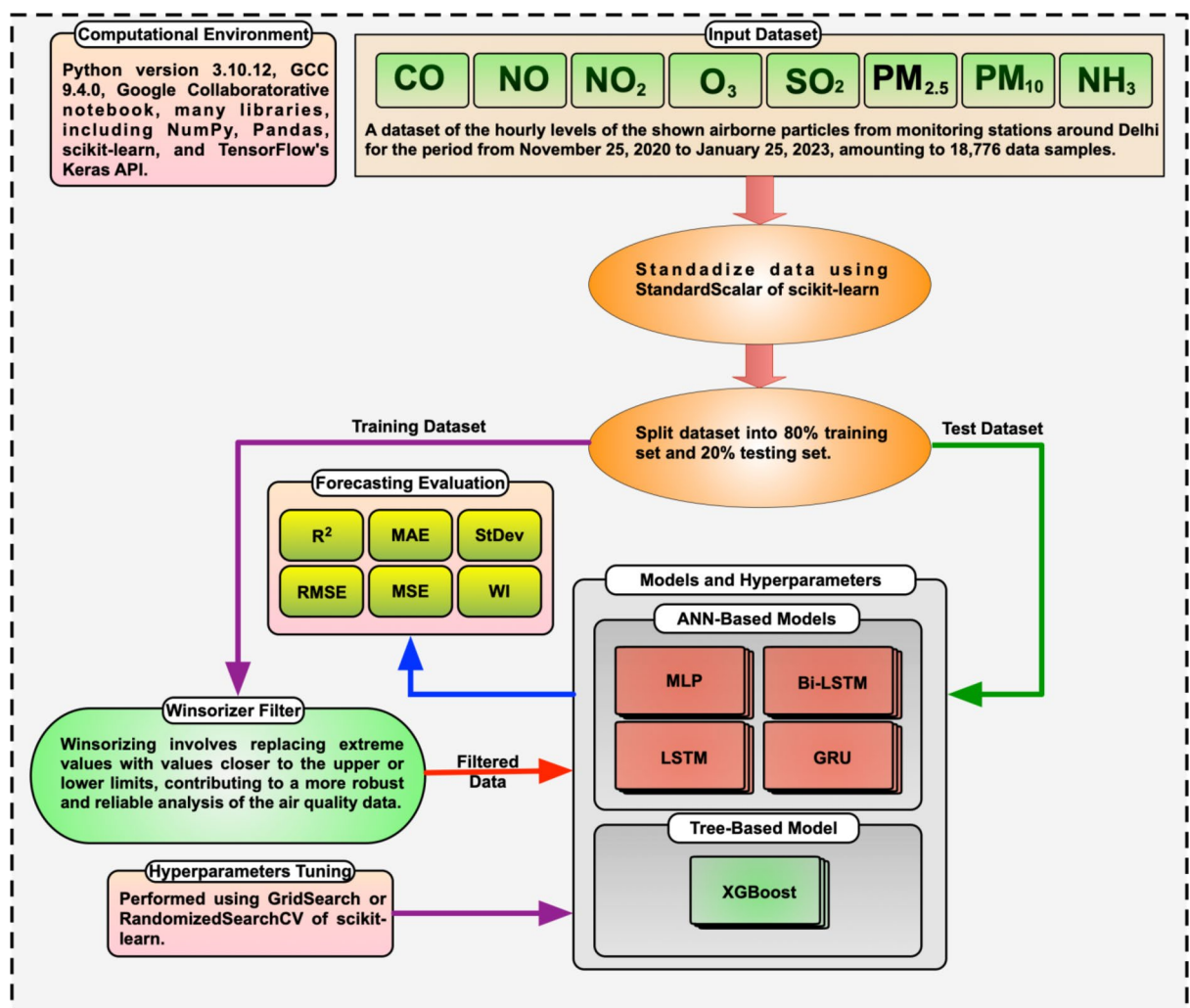


Fig. 5. Models development and configuration.

Model	Hyperparameter	values tested	Best value
XGBoost	n_estimators	[50, 100, 150]	Best: 100
	learning_rate	[0.01, 0.1, 0.2]	Best: 0.1
	max_depth	[3, 5, 7]	Best: 5
	min_child_weight	[1, 3, 5]	Best: 1
LSTM	neurons	[50, 100]	Best: 50
	epochs	[50, 100]	Best: 50
	batch_size	[32, 64]	Best: 32
Bi-LSTM	neurons	[50, 100]	Best: 50
	epochs	[50, 100]	Best: 50
	batch_size	[32, 64]	Best: 32
GRU	neurons	[50, 100]	Best: 50
	epochs	[50, 100]	Best: 50
	batch_size	[32, 64]	Best: 32
MLP	neurons	[50, 100]	Best: 50
	epochs	[50, 100]	Best: 50
	batch_size	[32, 64]	Best: 32

Table 2. Hyperparameters, along with the best values, were tested for LSTM, Bi-LSTM, GRU, ANN-MLP, and XGBoost models.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \tag{5}$$

Root Mean Square Error (RMSE)

$$RMSE(y, \hat{y}) = \sqrt{\frac{\sum_{i=0}^{N-1} (y_i - \hat{y})^2}{N}} \tag{6}$$

Mean Absolute Error (MAE)

$$MAE(y, \hat{y}) = \frac{\sum_{i=0}^{N-1} |y_i - \hat{y}_i|}{N} \tag{7}$$

Mean Squared Error (MSE)

$$MSE(y, \hat{y}) = \frac{\sum_{i=0}^{N-1} (y_i - \hat{y}_i)^2}{N} \tag{8}$$

Standard Deviation (σ)

$$\sigma = \frac{\sqrt{\sum_{i=1}^N (y_i - \mu)^2}}{N} \tag{9}$$

Willmott Index (WI)

$$WI(y, \hat{y}) = 1 - \frac{\sum_{i=0}^{N-1} (\hat{y}_i - y_i)^2}{\sum_{i=0}^{N-1} (|\hat{y}_i - mean(y)| + |y_i - mean(y)|)^2} \tag{10}$$

where \hat{y} is the predicted air pollutant, y is the actual air pollutant, $mean(y)$ is the mean of actual target output, σ is the Standard Deviation, μ is the air pollutant mean, N is the number of observations.

Results and discussion

This section presents the results and analysis of various ML models (LSTM, Bi-LSTM, GRU, MLP, and XGBoost) applied to forecast concentrations of eight airborne particles: Carbon Monoxide (CO), Nitrogen Monoxide (NO), Nitrogen Dioxide (NO₂), Ozone (O₃), Sulphur Dioxide (SO₂), Fine Particulate Matter (PM_{2.5}), Coarse Particulate Matter (PM₁₀), and Ammonia (NH₃). Each model was evaluated both graphically and statistically. To assess the predictive performance, a Taylor diagram was used to compare the models based on three key statistical metrics—correlation, RMSE, and standard deviation—in relation to the benchmarked observational

dataset. Additionally, scatter plots were generated to visualize the deviations between actual and forecasted values along the 1:1 line.

Carbon monoxide (CO)

Carbon monoxide was forecasted due to its highly poisonous concern for human health⁴⁴. Figure 6 presents a scatter plot and Taylor diagram along with the forecasting metrics. According to the scatter plot, the CO concentration was varying between (0–23000). In general, all models showed a good variation around the 45° line. However, MLP and XGBoost revealed some observations from the identical line. According to Fig. 6a, the obtained determination coefficient based on the regression scatter formula was (LSTM=0.97, Bi-LSTM=0.97, GRU=0.97, MLP=0.94 and XGBoost=0.96). Moreover, the Taylor diagram revealed that the observational dataset of the CO was located at the standard deviation of 1 as per Fig. 6b. Overall, Bi-LSTM surpassed others with the highest R^2 (0.979), lowest RMSE (438.898), and lowest MAE (244.543), indicating precise and reliable predictions (Fig. 6c). GRU also performed well, with an R^2 of 0.977 and RMSE of 456.729. LSTM followed closely with an R^2 of 0.971. XGBoost and MLP had a relatively lower accuracy, with MLP having the lowest R^2 (0.941) and the highest RMSE (740.927). Therefore, Bi-LSTM demonstrated the best performance, making it the most effective model for accurate air quality forecasting in this research.

Nitrogen monoxide (NO)

NO is an extremely reactive gas that is initiated during high-temperature fuel burning. It is emitted by automobiles and non-road vehicles (e.g., boats and construction equipment). Breathing NO with a high concentration can cause respiratory diseases such as asthma, which could lead to respiratory infections⁴⁴. According to the scatter plots visualization (Fig. 7a), the proposed models differed from the identical best-fit line. However, LSTM, GRU, and MLP demonstrated some observation scatters from a similar line. Numerically, the attained determination coefficient based on the regression scatter formula was (LSTM=0.94, Bi-LSTM=0.96, GRU=0.95, MLP=0.91 and XGBoost=0.94). In addition, based on the Taylor diagram presentation (Fig. 7b), the observational dataset of the NO was located at the standard deviation of 1. In this comparison of air quality forecasting models (Fig. 7c), Bi-LSTM emerged as the top performer with the highest R^2 (0.961), lowest RMSE (13.922), and MAE (7.854), indicating accurate and consistent predictions. GRU and LSTM followed, with R^2 values of 0.956 and

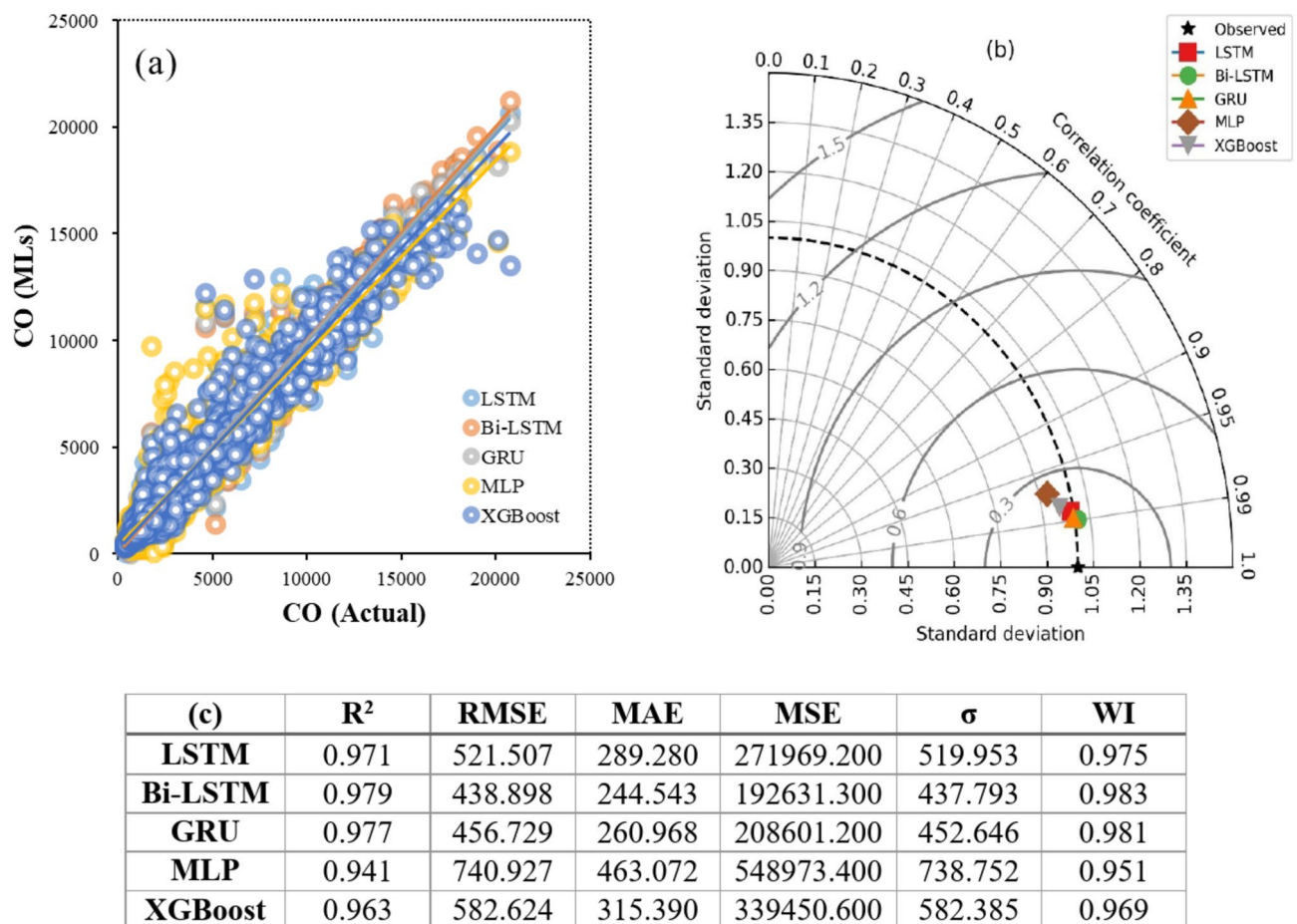


Fig. 6. Actual and forecasted values of CO using different AI-models; (a) Scatter plots, (b) Taylor diagram, (c) Forecasting metrics.

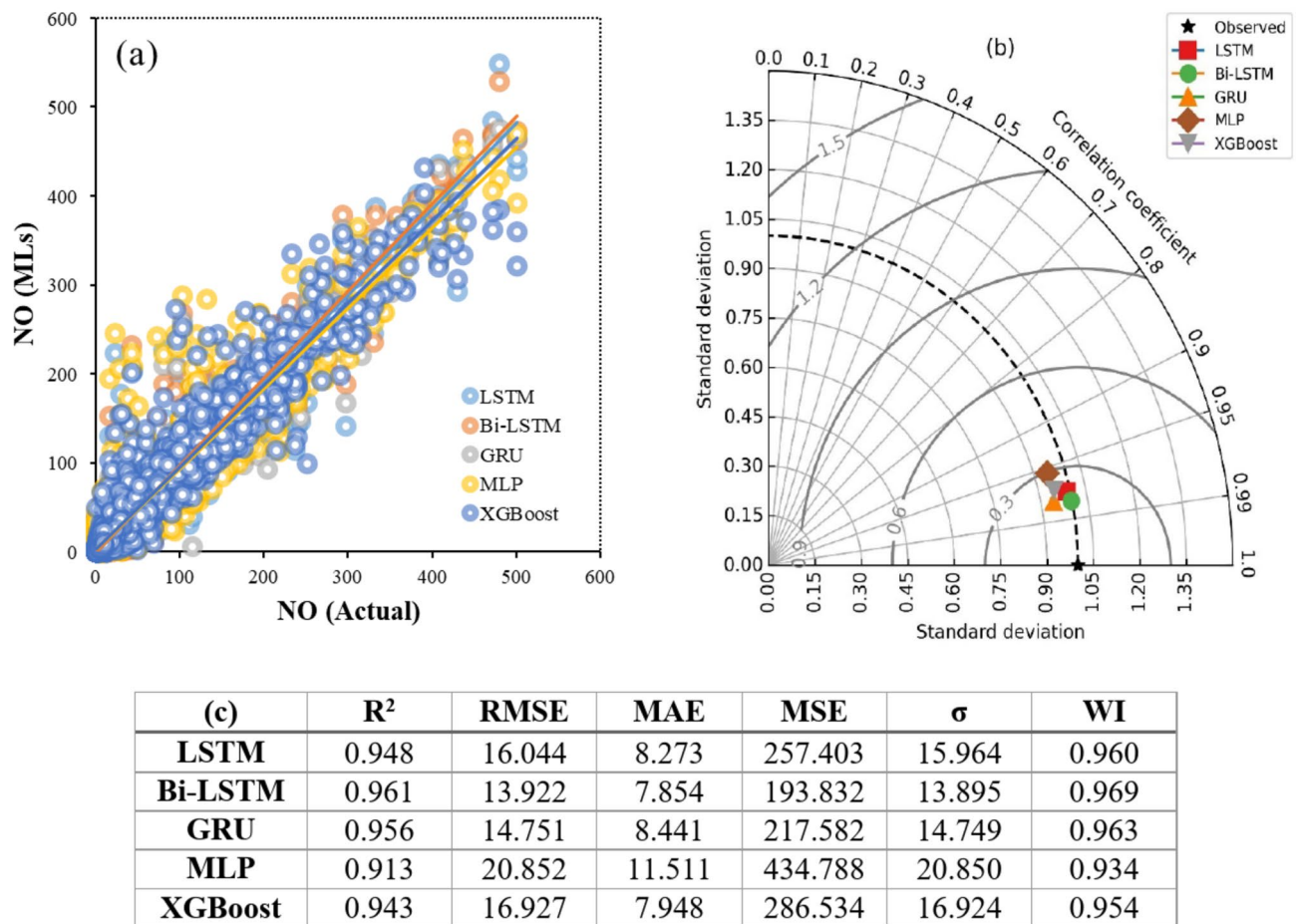


Fig. 7. Actual and forecasted values of NO using different AI-models; (a) Scatter plots, (b) Taylor diagram, (c) Forecasting metrics.

0.948, respectively, and slightly higher RMSEs and MAEs. XGBoost performed well, but had a slightly lower accuracy ($R^2 = 0.943$). The lowest performance was achieved by MLP, with an R^2 of 0.913 and the highest RMSE (20.852). Overall, Bi-LSTM proved the most reliable model for precise air quality forecasting in this study.

Nitrogen dioxide (NO_2)

Increases in mortality and hospital admissions for respiratory diseases are also associated with air nitrogen dioxide concentrations⁴⁴. The lungs' ability to combat microorganisms can be compromised by nitrogen dioxide, leaving the air more vulnerable to illnesses. Additionally, it may cause asthma to be worse. However, LSTM, GRU, and MLP revealed some observation scatters from the identical line. According to (Fig. 8a), the attained determination coefficient based on the regression scatter formula was (LSTM = 0.95, Bi-LSTM = 0.95, GRU = 0.94, MLP = 0.94, and XGBoost = 0.94). The Taylor diagram in Fig. 8b indicated that the observational dataset of the NO_2 was located at the standard deviation of 0.90 to 1. In this analysis of air quality forecasting models (Fig. 8c), Bi-LSTM achieved the highest R^2 (0.956) and the lowest RMSE (11.055) and MSE (122.215), suggesting its superior accuracy. Also, LSTM performed well, with an R^2 of 0.953 and a similar RMSE (11.452), making it a strong contender. GRU followed closely, though with slightly lower accuracy ($R^2 = 0.946$) and higher error metrics. XGBoost and MLP had a lower accuracy, with MLP having the lowest R^2 (0.937) and the highest RMSE (13.275). Overall, Bi-LSTM provided the most precise and consistent air quality predictions in this comparison.

Ozone (O_3)

Many health problems, such as congestion, coughing, throat irritation, and chest pain, can be caused by breathing in ground-level ozone⁴⁴. Bronchitis, asthma, and emphysema can all become worse due to it. Additionally, ozone can irritate the lining of the lungs and impair lung function. Long-term lung tissue scarring could result from repeated exposure. Increased vulnerability to diseases, pests, and other stressors such as severe weather results from elevated ozone levels, which also reduce the yields of commercial forests and crops and the growth and survival of tree seedlings. As per Fig. 9a, XGBoost and MLP revealed some observation scatters from the identical line. Numerically, the attained determination coefficient based on the regression scatter formula was (LSTM = 0.95, Bi-LSTM = 0.96, GRU = 0.96, MLP = 0.95, and XGBoost = 0.96). The observational dataset of the

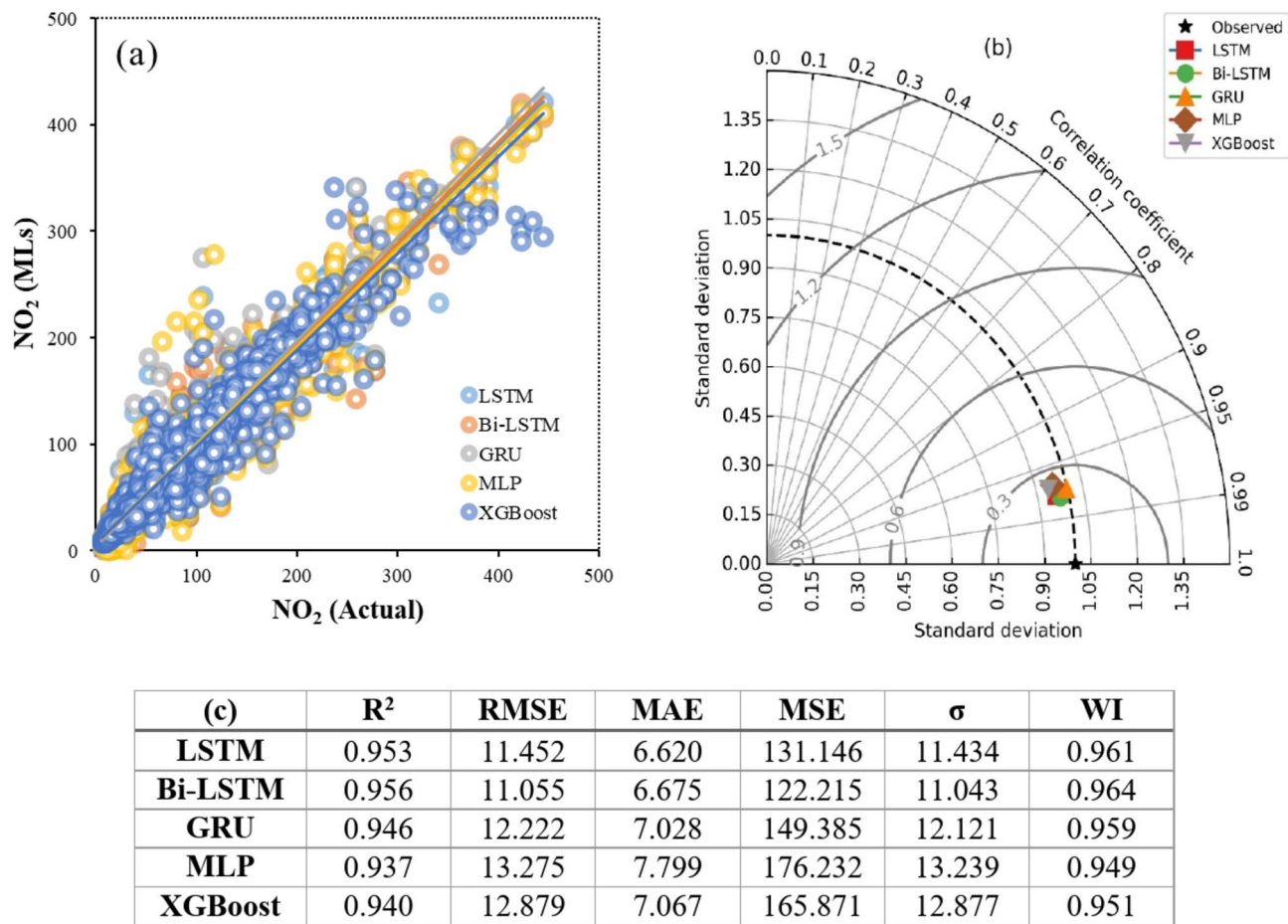


Fig. 8. Actual and forecasted values of NO₂ using different AI-models; (a) Scatter plots, (b) Taylor diagram, (c) Forecasting metrics.

O₃ was located at the standard deviation of 1 according to Taylor diagram in Fig. 9b. In this air quality forecasting comparison (Fig. 9c), both GRU and XGBoost excelled with the highest R² values (0.962). GRU achieved a lower RMSE (12.050) and MSE (145.191), while XGBoost had the lowest MAE (6.409), indicating excellent accuracy and minimal error. Bi-LSTM and LSTM also performed well with R² values of 0.959 and 0.958, respectively, but with slightly higher RMSE and MAE values. MLP, while accurate (R² = 0.957), showed the highest RMSE (12.869). In this part, GRU and XGBoost proved the most effective models for accurate and consistent air quality predictions.

Sulphur dioxide (SO₂)

The combustion of sulphur-containing fuels releases sulphur dioxide (SO₂), impacting ecosystems and human health. SO₂ harms plants, streams, and forests, and in humans, it worsens respiratory conditions like asthma and bronchitis, especially during exercise⁴⁴. Studies also link SO₂ exposure to higher cardiovascular disease risks. However, XGBoost and MLP revealed some observation scatters from the identical line. As per Fig. 10a, the attained determination coefficient based on the regression scatter formula was (LSTM=0.95, Bi-LSTM=0.95, GRU=0.95, MLP=0.92, and XGBoost=0.91). Based on the Taylor diagram in Fig. 10b, the observational dataset of the SO₂ was located at the standard deviation of 1 except MLP (SD=0.86). In this comparison of air quality forecasting models (Fig. 10c), Bi-LSTM achieved the highest performance with the highest R² (0.955), the lowest RMSE (11.110), and MAE (6.076), indicating superior accuracy and low error. LSTM also performed well with an R² of 0.950 and slightly higher RMSE (11.648). GRU followed closely, with similar metrics (R² = 0.949). The accuracy of MLP and XGBoost was lower, with MLP achieving an R² of 0.928 and XGBoost the lowest R² (0.916) and highest RMSE (15.214). In conclusion, Bi-LSTM was the most effective model for precise air quality predictions in this study.

Fine particles matter (PM_{2.5})

Many scientific investigations have demonstrated that particulate matter decreases visibility and negatively impacts materials, ecosystems, and climate. PM, particularly PM_{2.5}, alters how light is absorbed and scattered in the atmosphere, which can affect visibility^{45,46}. Long-term exposure to fine particles may also increase the risk of heart disease and be linked to a higher incidence of chronic bronchitis, deteriorated lung function, and

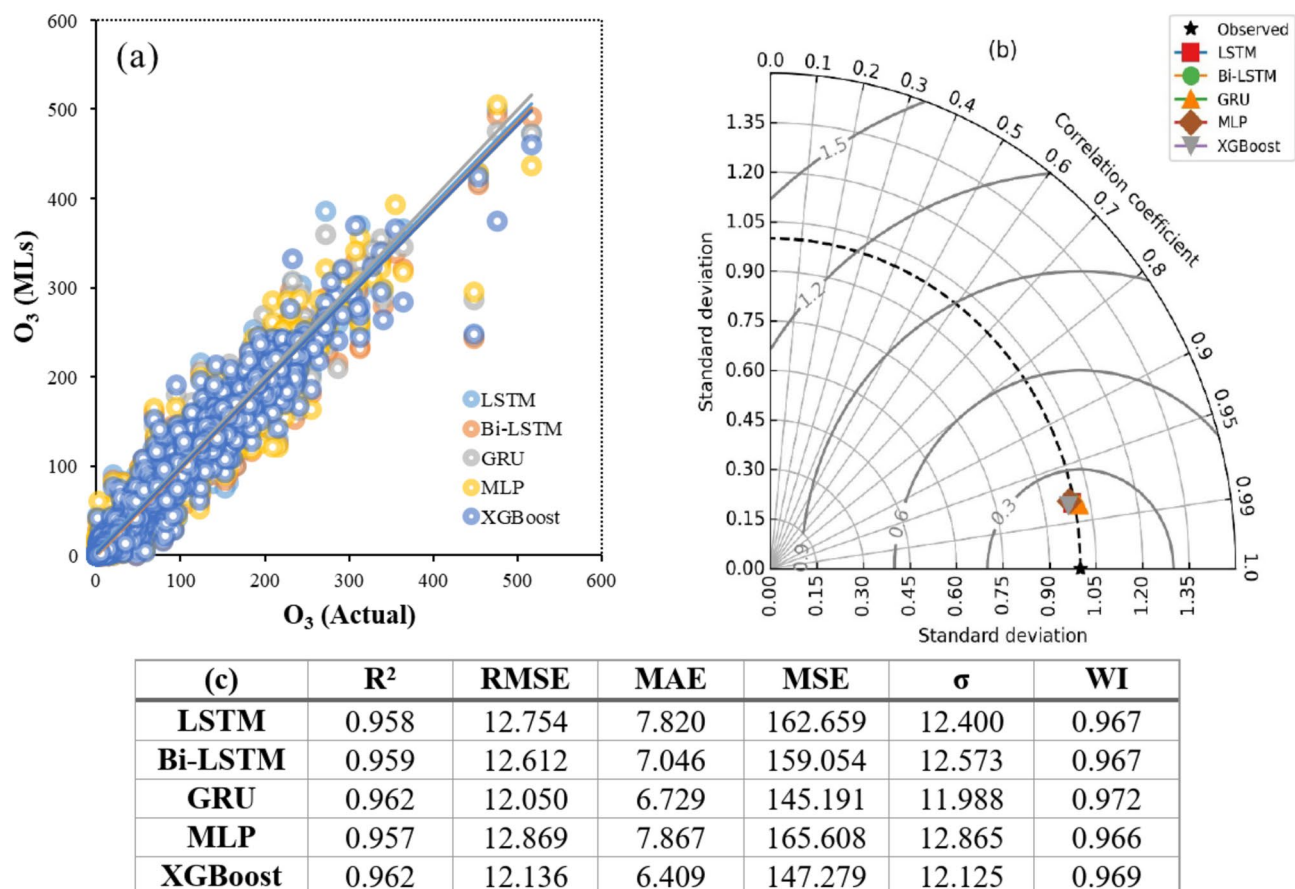


Fig. 9. Actual and forecasted values of O_3 using different AI-models; (a) Scatter plots, (b) Taylor diagram, (c) Forecasting metrics.

lung cancer, according to studies. However, XGBoost and Bi-LSTM revealed some observation scatters from the identical line. According to Fig. 11a, the attained determination coefficient based on the regression scatter formula was (LSTM = 0.97, Bi-LSTM = 0.97, GRU = 0.96, MLP = 0.93, and XGBoost = 0.96). The Taylor diagram illustration (Fig. 11b) demonstrated that, the observational dataset of $PM_{2.5}$ was located at the standard deviation of 1.05. In this analysis of air quality forecasting models (Fig. 11c), LSTM demonstrated the best performance with the highest R^2 (0.973), the lowest RMSE (37.549), and MAE (22.183), indicating strong accuracy and precision. Bi-LSTM closely followed with an R^2 of 0.971 but had a slightly higher RMSE (38.519). GRU also performed well, with an R^2 of 0.969 and comparable metrics. XGBoost achieved a satisfactory R^2 of 0.961 but with higher RMSE (45.189) and MAE (25.134). MLP had the lowest overall performance, with an R^2 of 0.936 and significantly higher RMSE (57.665) and MAE (37.627). In this study, LSTM emerged as the most effective model for accurate air quality predictions.

Coarse particles matter (PM_{10})

Particle size contributes to the health and ecological impacts of particulate matter (PM). PM_{10} particles (10 micrometer or smaller) can reach the lungs upon inhalation, resulting in serious health risks to heart and lung health. Additionally, PM deposition affects ecosystems by impairing water quality and altering plant growth, especially due to the metal and organic compounds within PM_{10} . However, XGBoost revealed some observation scatters from the identical line. As per Fig. 12a, the attained determination coefficient based on the regression scatter formula was (LSTM = 0.97, Bi-LSTM = 0.97, GRU = 0.97, MLP = 0.94, and XGBoost = 0.96). Also, Taylor diagram showed in Fig. 12b that, the observational dataset of PM_{10} was located at the standard deviation of 1.05. In this comparison of air quality forecasting models (Fig. 12c), Bi-LSTM achieved the highest performance with the highest R^2 (0.975), lowest RMSE (43.126), and MAE (24.414), indicating excellent accuracy. LSTM and GRU were closely followed, both with R^2 values of 0.974, but slightly higher RMSE and MAE values. XGBoost had a moderate performance, with an R^2 of 0.962 but a higher RMSE (53.085). MLP showed the lowest accuracy with an R^2 of 0.949 and the highest RMSE (61.782). Overall, Bi-LSTM proved to be the most accurate model for air quality prediction.

Ammonia (NH_3)

Ammonia is a major contributor to nitrogen pollution. The effects of nitrogen buildup on plant species diversity and composition within impacted environments are an essential component of ammonia pollution's impacts

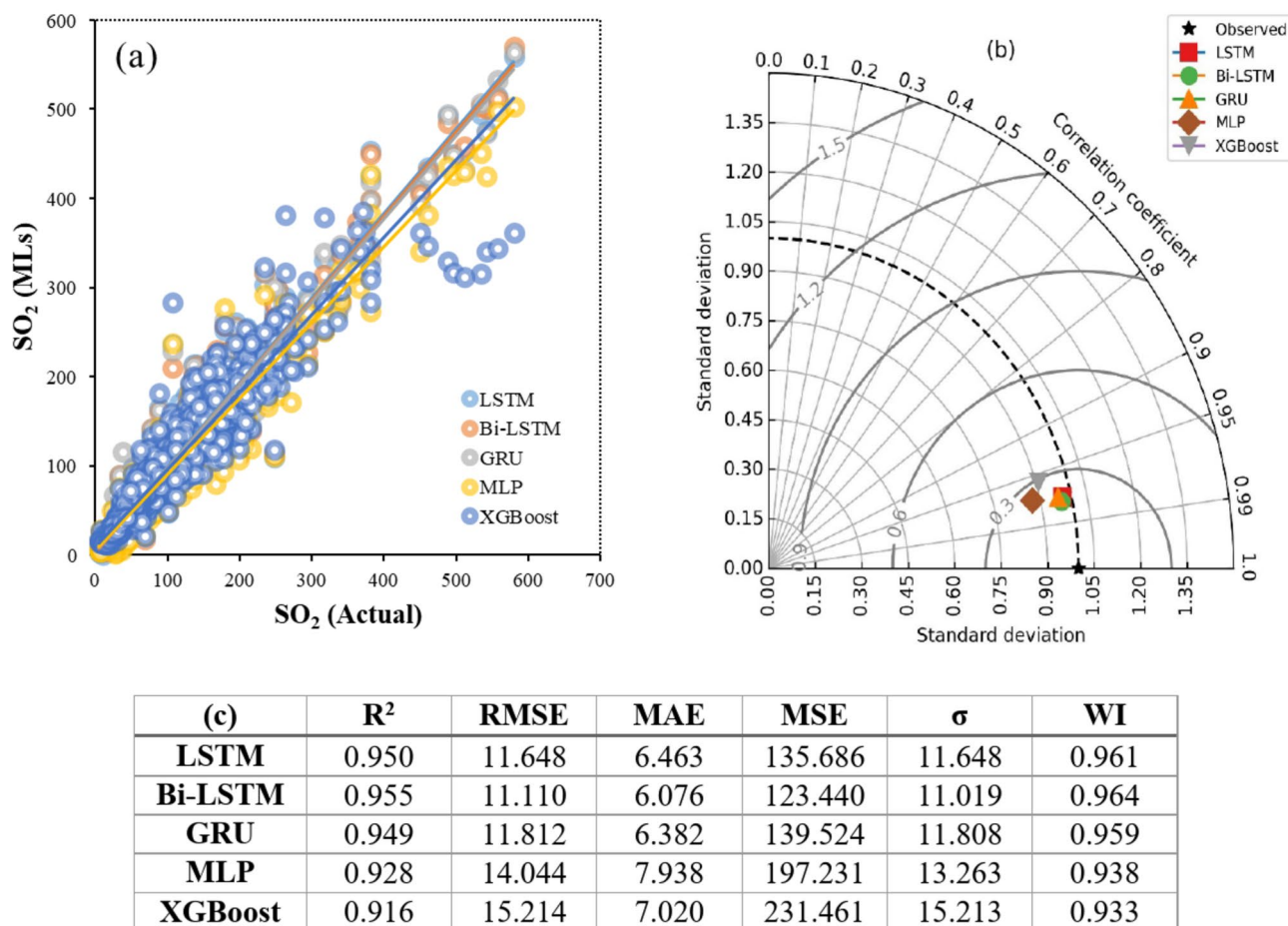


Fig. 10. Actual and forecasted values of SO_2 using different AI-models; (a) Scatter plots, (b) Taylor diagram, (c) Forecasting metrics.

on biodiversity. Ammonia can cause irritation to the skin or eyes or irritation to them when it comes into contact with it. If ammonia is inhaled, it can cause coughing, wheezing, and shortness of breath, as it can cause irritation to the respiratory system. Furthermore, ammonia inhalation may cause irritation to the throat and nose. However, XGBoost revealed some observation scatters from the identical line. According to Fig. 13a, the attained determination coefficient based on the regression scatter formula was (LSTM=0.96, Bi-LSTM=0.97, GRU=0.96, MLP=0.93, and XGBoost=0.93). As per Fig. 13b, Taylor diagram showed that the observational dataset of NH_3 was located at the standard deviation of 1.05. In this assessment of air quality forecasting models (Fig. 13c), Bi-LSTM demonstrated the highest performance with an R^2 of 0.971, the lowest RMSE (5.283), and the lowest MAE (2.743), indicating superior accuracy and minimal error. LSTM and GRU were followed closely, with R^2 values of 0.963 and 0.965, respectively, but slightly higher RMSE and MAE scores. The accuracy of MLP and XGBoost was lower, with MLP having an R^2 of 0.939 and the highest RMSE (7.645), while XGBoost had the lowest R^2 (0.933) and a high RMSE (8.036). In conclusion, Bi-LSTM was the most effective model for precise air quality predictions in this study.

Exposure to air pollution is a global public health hazard, with a considerable body of evidence linking short-term and long-term exposures to a range of health outcomes, including all-cause and cause-specific mortality, respiratory and cardiovascular conditions, neurodevelopmental deficiencies, and adverse pregnancy and birth outcomes⁴⁴. The current research was fueled by the robustness of deep learning models that predictability performance of recent research development on machine learning establishment, deep learning revealed superior performance to the other ML models. The eight air quality parameters were chosen due to their seriousness as environmental and atmospheric indices. In this regard, the study's unique methods for predicting air pollution provide valuable results in various circumstances. Models that are considered to be suitable and beneficial for daily prediction include LSTM, Bi-LSTM, GRU, MALP and XGBoost. In order to discuss the reliability and validity of the current intelligence models, Table 3 summarizes the latest air quality studies using multiple machine and deep learning models under different meteorological conditions.

This study's significance lies in its model variety, particularly Bi-LSTM, which achieved higher R^2 values than other methods. Comparatively, previous studies achieved comparable R^2 values but were limited to fewer pollutants or simpler seasonal analyses. In addition, studies in Hyderabad (2022) and Kolkata (2019–2022) showed strong results with models such as ANN and XGBoost, but focused on fewer pollutants and seasonal

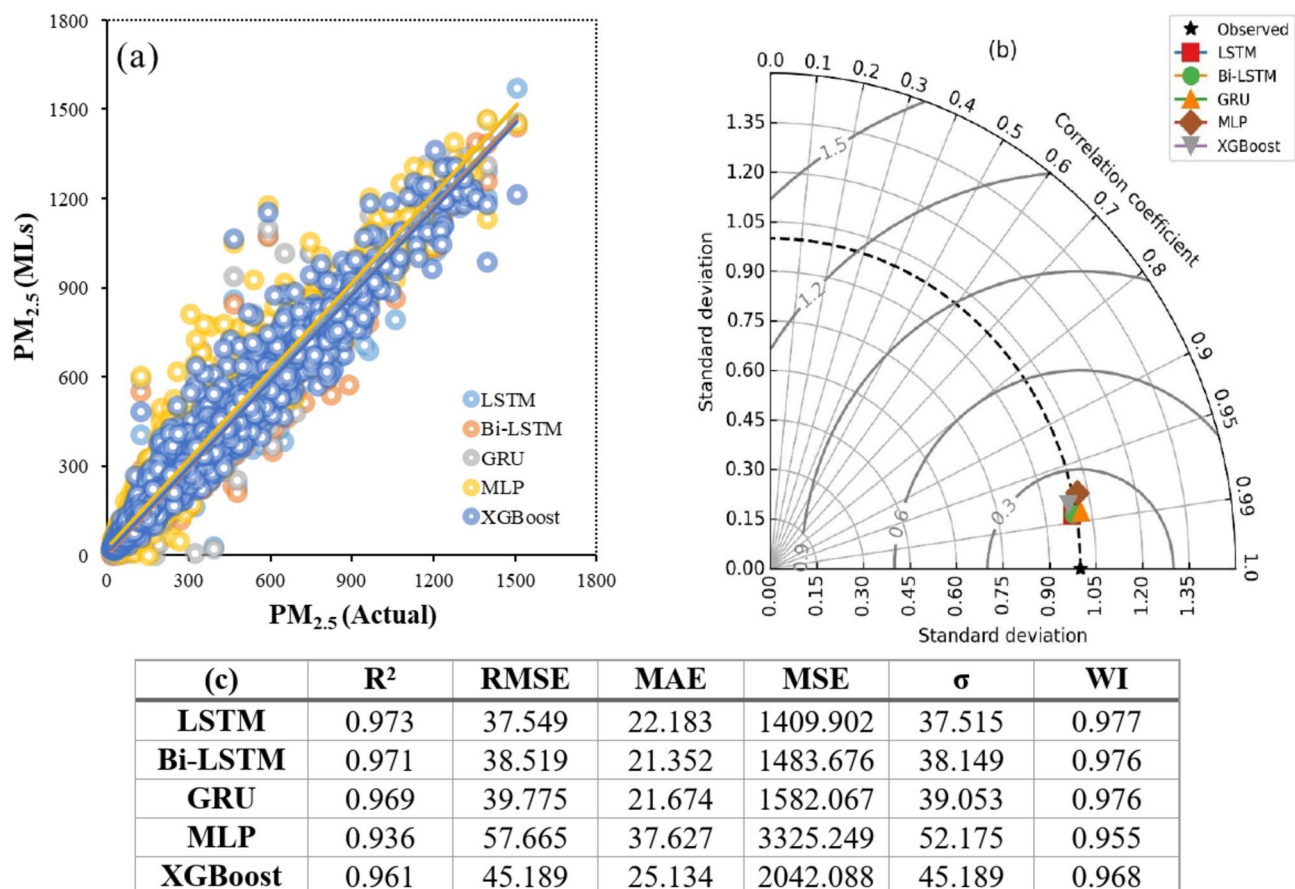


Fig. 11. Actual and forecasted values of $PM_{2.5}$ using different AI-models; (a) Scatter plots, (b) Taylor diagram, (c) Forecasting metrics.

variations. In contrast, the Delhi study utilized Bi-LSTM, LSTM, and GRU extensively across pollutants and extended the model's applicability across multiple seasons, enhancing predictive power for dynamic air quality conditions in a complex urban area. The high performance and adaptability of the Bi-LSTM model across pollutants highlight its potential as a robust choice for multi-pollutant air quality forecasting, enabling policymakers and environmental agencies to make data-driven decisions and mitigate pollution's impact on public health.

Conclusion, limitations and future directions

In this study, different air quality parameters were proposed, including CO, NO, NO_2 , O_3 , SO_2 , $PM_{2.5}$, PM_{10} , and NH_3 . Datasets were collected for monitoring stations located near Delhi, India, for the duration of (25-11-2020/24-01-2023) with an hourly rate. For this purpose, various AI models were introduced, including Long Short-Term Memory (LSTM), Bidirectional Long-Short Term Memory (Bi-LSTM), Gated Recurrent Unit (GRU), Multilayer Perceptron (MLP), and Extreme Gradient Boosting (XGBoost). The following findings can be drawn from the current study:

- In CO forecasting, the AI models showed the following accuracy: Bi-LSTM=0.979, GRU=0.977, LSTM=0.971, XGBoost=0.963, and MLP=0.941, respectively.
- In NO forecasting, the AI models showed the following accuracy: Bi-LSTM=0.961, GRU=0.956, LSTM=0.948, XGBoost=0.943, and MLP=0.913, respectively.
- In NO_2 forecasting, the AI models showed the following accuracy: Bi-LSTM=0.956, GRU=0.953, LSTM=0.946, XGBoost=0.940 and MLP=0.937, respectively.
- In O_3 forecasting, the AI models showed the following accuracy: GRU=0.9624, XGBoost=0.9619, Bi-LSTM=0.9588, LSTM=0.9579 and MLP=0.9571, respectively.
- In SO_2 forecasting, the AI models showed the following accuracy: Bi-LSTM=0.955, LSTM=0.950, GRU=0.949, MLP=0.928 and XGBoost=0.916, respectively.
- In $PM_{2.5}$ forecasting, the AI models showed the following accuracy: LSTM=0.973, Bi-LSTM=0.971, GRU=0.969, XGBoost=0.961 and MLP=0.936, respectively.
- In PM_{10} forecasting, the AI models showed the following accuracy: Bi-LSTM=0.9751, LSTM=0.9744, GRU=0.9737, XGBoost=0.9622 and MLP=0.9488, respectively.

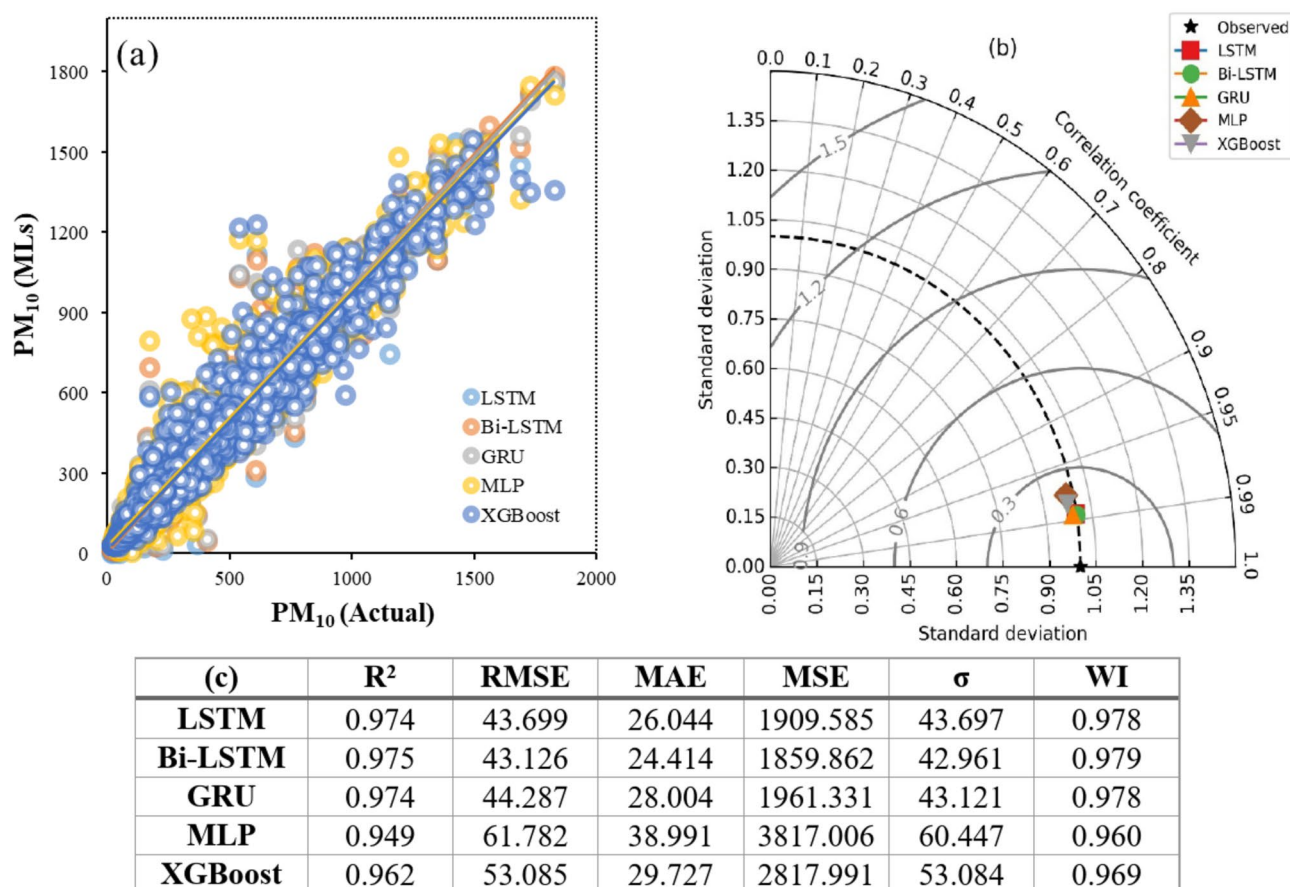


Fig. 12. Actual and forecasted values of PM_{10} using different AI-models; (a) Scatter plots, (b) Taylor diagram, (c) Forecasting metrics.

viii. In NH_3 forecasting, the AI models showed the following accuracy: Bi-LSTM=0.971, GRU=0.965, LSTM=0.963, MLP=0.939 and XGBoost=0.933, respectively.

This study highlights the potential of AI-based models in air quality forecasting, essential for proactive urban pollution management. Accurate predictions support real-time alert systems, preventing public health as cities expand. While high model accuracy was achieved, limitations remain, such as reliance on pollutant data alone, lacking real-time adaptability to sudden changes in pollution. Future research should incorporate meteorological and socioeconomic data, enabling more robust, responsive models. Real-time monitoring and long-term trend analyses could further enhance air quality management, enabling sustainable policy decisions. Interdisciplinary collaboration will be crucial in advancing these models for efficient, data-driven air quality solutions.

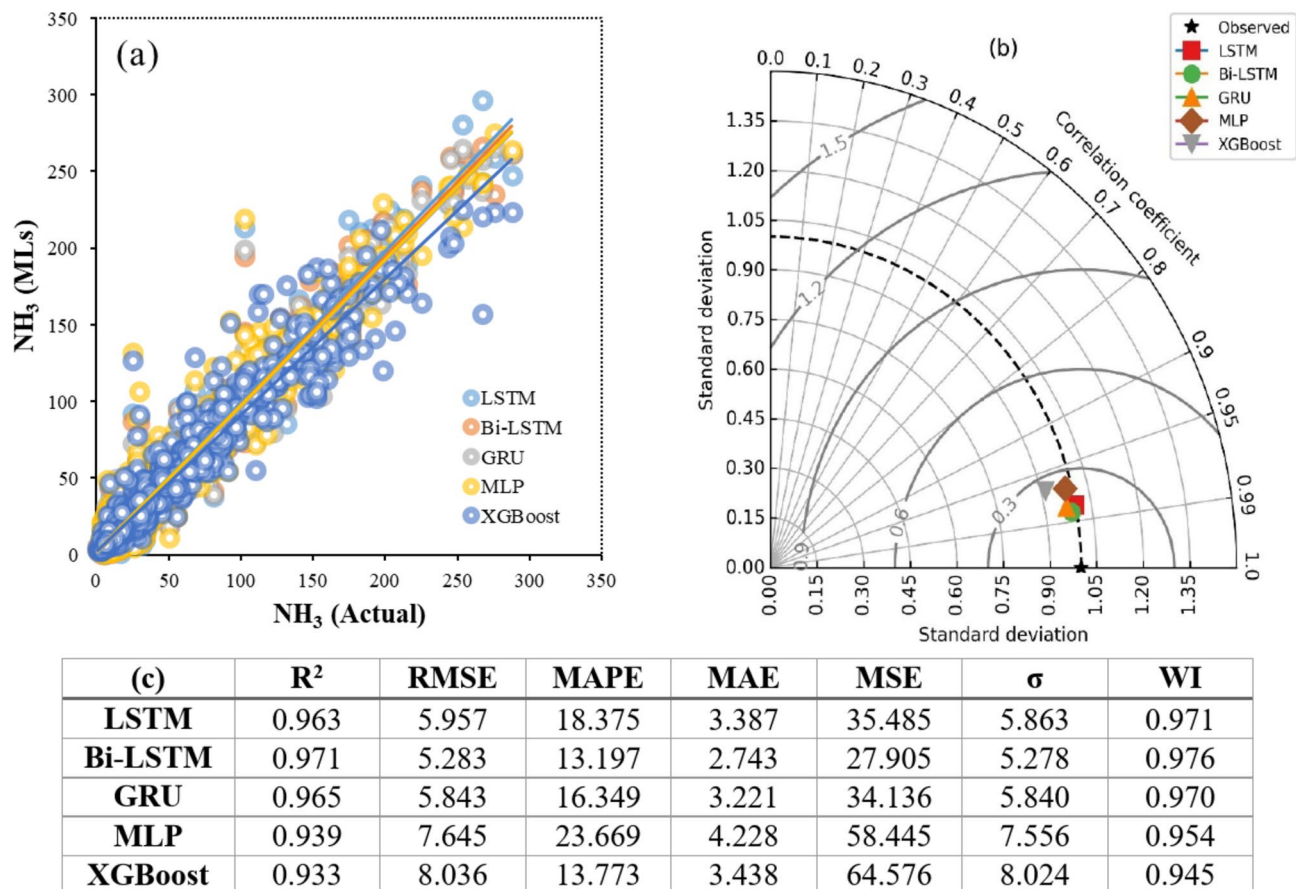


Fig. 13. Actual and forecasted values of NH_3 using different AI-models; (a) Scatter plots, (b) Taylor diagram, (c) Forecasting metrics.

Reference	Study area	Air pollutant	ML/DL model	Highlights
6	Chennai, using the historical data available from 2017 to 2022	AQI	XGBoost, RF, Bagging Regressor, and XGBRegressor	XGBoost achieved $R^2 = 0.9935$, $\text{MAE} = 0.02$, $\text{MSE} = 0.001$, and $\text{RMSE} = 0.04$
8	Ahmedabad, January 2015, to January 2021	AQI	SARIMA, SVM and LSTM	LSTM showed $R^2 = 0.951$, $\text{MSE} = 656.623$ and $\text{RMSE} = 25.625$.
18	Visakhapatnam, Andhra Pradesh, from July 2017 to September 2022	AQI	LightGBM, RF, Catboost, Adaboost, and XGBoost	Catboost model yielded ($R^2 = 0.9998$) and ($\text{RMSE} = 0.76$).
19	New Delhi, 2008 to 2010	$\text{PM}_{2.5}$	GRU-ED, RF, XGBoost, ANNs, and LSTM	GRU-ED showed ($R^2 = 0.959$, $\text{NSE} = 0.953$, $\text{MAE} = 1.770$, $\text{RRMSE} = 0.002$, and $\text{MAPE} = 0.190$).
22	Hyderabad using five-year (2018–2022) data	$\text{PM}_{2.5}$, NH_3 , CO , O_3 , NO_2 , and SO_2	MLR, SVR, RF, ANN, XGBoost, and LSTM	$\text{ANN} > \text{RF} > \text{SVR} > \text{XG Boost} > \text{LSTM} > \text{MLR}$. ANN outperformed other model ($R^2 > 0.90$) in both summer and winter seasons.
23	India using regional and temporal data in 2018	$\text{PM}_{2.5}$ and PM_{10}	LightGBM	$\text{PM}_{2.5}$ showed $R^2 = 0.79, 0.80, 0.86$ and 0.87 , for original, spatial, temporal and spatial-temporal, respectively. PM_{10} showed $R^2 = 0.81, 0.82, 0.87$ and 0.88 , for original, spatial, temporal and spatial-temporal, respectively.
25	New Delhi, hourly observations for 2021 and 2022	O_3 and NO_2	USS-GNN	$R^2 = 0.650$ and 0.618 , $\text{RMSE} = 13.950$ and 16.120 , $\text{MAE} = 10.730$ and 12.930 for O_3 and NO_2 , respectively.
48	Kolkata from 2019 to 2022 (4 years)	$\text{PM}_{2.5}$, PM_{10} , NO_2 , NH_3 , SO_2 and CO	RF, DT, KNN, SVR, Ridge, Lasso, and XGBoost	XGBoost showed $R^2 = 0.858$ (Winter), $R^2 = 0.919$ (Pre- Monsoon), $R^2 = 0.844$ (Monsoon) and $R^2 = 0.862$ (Post Monsoon).
49	Hyderabad, from January 2018 to December 2019	$\text{PM}_{2.5}$	MLR, DT, KNN, RF, XGBoost and LSTM	XGBoost: $R^2 = 0.82$ and $\text{MAE} = 7.01 \mu\text{g}/\text{m}^3$. LSTM: $R^2 = 0.89$ and $\text{MAE} = 5.78 \mu\text{g}/\text{m}^3$.
Current	Delhi, from 25/11/2020 to 24/01/2023	CO , NO , NO_2 , O_3 , SO_2 , $\text{PM}_{2.5}$, PM_{10} , and NH_3	LSTM, Bi-LSTM, GRU, MLP, and XGBoost	Bi-LSTM model was the best mode CO with ($R^2 = 0.979$), NO with ($R^2 = 0.961$), NO_2 with ($R^2 = 0.956$), SO_2 with ($R^2 = 0.955$), PM_{10} with ($R^2 = 0.9751$) and NH_3 with ($R^2 = 0.971$). Meanwhile, GRU and LSTM models performed better in forecasting O_3 and $\text{PM}_{2.5}$ with ($R^2 = 0.9624$) and ($R^2 = 0.973$), respectively.

Table 3. Summary of previous studies on air pollutant prediction using intelligent learning approaches across India.

Data availability

The data can be shared upon request from the corresponding author.

Received: 26 August 2024; Accepted: 2 December 2024

Published online: 28 December 2024

References

- Perera, F. Pollution from fossil-fuel combustion is the leading environmental threat to global pediatric health and equity: solutions exist. *Int. J. Environ. Res. Public Health*. **15**, 16 (2018).
- Manisalidis, I., Stavropoulou, E., Stavropoulos, A. & Bezirtzoglou, E. Environmental and health impacts of air pollution: a review. *Front. Public Health*. **8**, 14 (2020).
- Ravindra, K. Emission of black carbon from rural households kitchens and assessment of lifetime excess cancer risk in villages of North India. *Environ. Int.* **122**, 201–212 (2019).
- Ravindra, K., Singh, T., Pandey, V. & Mor, S. Air pollution trend in Chandigarh city situated in Indo-Gangetic Plains: understanding seasonality and impact of mitigation strategies. *Sci. Total Environ.* **729**, 138717 (2020).
- Villanueva, F. et al. Ambient levels of volatile organic compounds and criteria pollutants in the most industrialized area of central Iberian Peninsula: intercomparison with an urban site. *Environ. Technol.* **37**, 983–996 (2016).
- Ravindiran, G. et al. Impact of air pollutants on climate change and prediction of air quality index using machine learning models. *Environ. Res.* **239**, 117354 (2023).
- Bodor, Z., Bodor, K., Keresztesi, Á. & Szép, R. Major air pollutants seasonal variation analysis and long-range transport of PM 10 in an urban environment with specific climate condition in Transylvania (Romania). *Environ. Sci. Pollut. Res.* **27**, 38181–38199 (2020).
- Maltare, N. N. & Vahora, S. Air Quality Index prediction using machine learning for Ahmedabad city. *Digit. Chem. Eng.* **7**, 100093 (2023).
- Yadav, V., Yadav, A. K., Singh, V. & Singh, T. Artificial neural network an innovative approach in air pollutant prediction for environmental applications: a review. *Results Eng.*, **102305** (2024).
- Kumar, K. & Pande, B. Air pollution prediction with machine learning: a case study of Indian cities. *Int. J. Environ. Sci. Technol.* **20**, 5333–5348 (2023).
- Guo, Q., He, Z. & Wang, Z. The characteristics of Air Quality changes in Hohhot City in China and their relationship with Meteorological and Socio-economic factors. *Aerosol Air Qual. Res.* **24**, 230274 (2024).
- Gupta, N. S. et al. Prediction of air quality index using machine learning techniques: a comparative analysis. *J. Environ. Public Health* **2023** (2023).
- Nath, P., Saha, P., Middya, A. I. & Roy, S. Long-term time-series pollution forecast using statistical and deep learning methods. *Neural Comput. Appl.*, 1–20 (2021).
- Aggarwal, A. & Toshniwal, D. A hybrid deep learning framework for urban air quality forecasting. *J. Clean. Prod.* **329**, 129660 (2021).
- Janarthanan, R., Partheeban, P., Somasundaram, K. & Elamparithi, P. N. A deep learning approach for prediction of air quality index in a metropolitan city. *Sustainable Cities Soc.* **67**, 102720 (2021).
- Pruthi, D. & Liu, Y. Low-cost nature-inspired deep learning system for PM_{2.5} forecast over Delhi, India. *Environ. Int.* **166**, 107373 (2022).
- Sarkar, N., Gupta, R., Keserwani, P. K. & Govil, M. C. Air Quality Index prediction using an effective hybrid deep learning model. *Environ. Pollut.* **315**, 120404 (2022).
- Ravindiran, G., Hayder, G., Kanagarathinam, K., Alagumalai, A. & Sonne, C. Air quality prediction by machine learning models: a predictive study on the Indian coastal city of Visakhapatnam. *Chemosphere* **338**, 139518 (2023).
- Shakya, D., Deshpande, V., Goyal, M. K. & Agarwal, M. PM_{2.5} air pollution prediction through deep learning using meteorological, vehicular, and emission data: a case study of New Delhi, India. *J. Clean. Prod.* **427**, 139278 (2023).
- Dhandapani, A., Iqbal, J. & Kumar, R. N. Application of machine learning (individual vs stacking) models on MERRA-2 data to predict surface PM_{2.5} concentrations over India. *Chemosphere* **340**, 139966 (2023).
- Ravindra, K. et al. Application of machine learning approaches to predict the impact of ambient air pollution on outpatient visits for acute respiratory infections. *Sci. Total Environ.* **858**, 159509 (2023).
- Suthar, G., Singh, S., Kaul, N. & Khandelwal, S. Prediction of land surface temperature using spectral indices, air pollutants, and urbanization parameters for Hyderabad City of India using six machine learning approaches. *Remote Sens. Appl. Soc. Environ.*, **101265** (2024).
- Wang, S. et al. Extracting regional and temporal features to improve machine learning for hourly air pollutants in urban India. *Atmos. Environ.* **338**, 120834 (2024).
- Rahaman, S., Tu, X., Ahmad, K. & Qadeer, A. A real-time assessment of hazardous atmospheric pollutants across cities in China and India. *J. Hazard. Mater.* **479**, 135711 (2024).
- Mandal, S., Boppani, S., Dasari, V. & Thakur, M. A bivariate simultaneous pollutant forecasting approach by Unified Spectro-Spatial Graph Neural Network (USSGNN) and its application in prediction of O₃ and NO₂ for New Delhi, India. *Sustainable Cities Soc.* **114**, 105741 (2024).
- Guo, Q., He, Z. & Wang, Z. Monthly climate prediction using deep convolutional neural network and long short-term memory. *Sci. Rep.* **14**, 17748 (2024).
- Haykin, S. *Neural Networks: A Comprehensive Foundation* (Prentice Hall PTR, 1998).
- Jamei, M. et al. Air quality monitoring based on chemical and meteorological drivers: application of a novel data filtering-based hybridized deep learning model. *J. Clean. Prod.* **374**, 134011 (2022).
- Ehteram, M., Salih, S. Q. & Yaseen, Z. M. Efficiency evaluation of reverse osmosis desalination plant using hybridized multilayer perceptron with particle swarm optimization. *Environ. Sci. Pollut. Res.* **27**, 15278–15291 (2020).
- Tur, R. & Yontem, S. A comparison of soft computing methods for the prediction of wave height parameters. *Knowl.-Based Eng. Sci.* **2**, 31–46 (2021).
- Schmidhuber, J. & Hochreiter, S. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
- Schuster, M. & Paliwal, K. K. bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **45**, 2673–2681 (1997).
- Yaseen, Z. M. et al. Development of advanced data-intelligence models for radial gate discharge coefficient prediction: modeling different flow scenarios. *Water Resour. Manage.* **37**, 5677–5705 (2023).
- Wang, S. et al. Air pollution prediction via graph attention network and gated recurrent unit. *Computers Mater. Continua.* **73**, 673–687 (2022).
- Chen, T. & Guestrin, C. Xgboost: a scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* Vol. 2016, 785–794 (ACM, 2016).
- Fan, J. et al. Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agric. For. Meteorol.* **263**, 225–241 (2018).

37. Zhang, W., Wu, C., Zhong, H., Li, Y. & Wang, L. Prediction of undrained shear strength using extreme gradient boosting and random forest based on bayesian optimization. *Geosci. Front.* **12**, 469–477 (2021).
38. Ferreira, L. B. & da Cunha, F. F. New approach to estimate daily reference evapotranspiration based on hourly temperature and relative humidity using machine learning and deep learning. *Agric. Water Manage.* **234**, 106113 (2020).
39. Ni, L. et al. Streamflow forecasting using extreme gradient boosting model coupled with gaussian mixture model. *J. Hydrol.* **586**, 124901 (2020).
40. Xu, C. et al. A study of predicting irradiation-induced transition temperature shift for RPV steels with XGBoost modeling. *Nuclear Eng. Technol.* **53**, 2610–2615 (2021).
41. Feigl, M., Lebedzinski, K., Herrnegger, M. & Schulz, K. Machine learning methods for stream water temperature prediction. *Hydrol. Earth Syst. Sci. Dis.* **2021**, 1–35 (2021).
42. Sikorska-Senoner, A. E. & Quilty, J. M. A novel ensemble-based conceptual-data-driven approach for improved streamflow simulations. *Environ. Model. Softw.* **143**, 105094 (2021).
43. Tao, H. et al. Development of new computational machine learning models for longitudinal dispersion coefficient determination: case study of natural streams, United States. *Environ. Sci. Pollut. Res.* **29**, 35841–35861 (2022).
44. de Bont, J. et al. Ambient air pollution and daily mortality in ten cities of India: a causal modelling study. *Lancet Planet. Health.* **8**, e433–e440 (2024).
45. Guo, Q., He, Z. & Wang, Z. Predicting of daily $PM_{2.5}$ concentration employing wavelet artificial neural networks based on meteorological elements in Shanghai, China. *Toxics* **11**, 51 (2023).
46. Guo, Q., He, Z. & Wang, Z. Simulating daily $PM_{2.5}$ concentrations using wavelet analysis and artificial neural network with remote sensing and surface observation data. *Chemosphere* **340**, 139886 (2023).
47. Guo, Q., He, Z. & Wang, Z. Prediction of hourly $PM_{2.5}$ and PM_{10} concentrations in Chongqing City in China based on artificial neural network. *Aerosol Air Qual. Res.* **23**, 220448 (2023).
48. Mondal, S., Adhikary, A. S., Dutta, A., Bhardwaj, R. & Dey, S. Utilizing machine learning for air pollution prediction, comprehensive impact assessment, and effective solutions in Kolkata, India. *Results Earth Sci.* **2**, 100030 (2024).
49. Gokul, P., Mathew, A., Bhosale, A. & Nair, A. T. Spatio-temporal air quality analysis and $PM_{2.5}$ prediction over Hyderabad City, India using artificial intelligence techniques. *Ecol. Inf.* **76**, 102067 (2023).

Acknowledgements

Ali Alsuwaiyan and Zaher Mundher Yaseen would like to thank King Fahd University of Petroleum & Minerals, Saudi Arabia for its support.

Author contributions

Omer A. Alawi: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing. Haslinda Mohamed Kamar: Formal analysis, Resources, Supervision, Project administration, Funding acquisition. Ali Saleh Mohammed AlSuwaiyan: Methodology, Software, Formal analysis, Writing - Original Draft. Zaher Mundher Yaseen: Methodology, Validation, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Z.M.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024