Data-Driven Analysis and Predictive Modeling of Urban Air Quality for Environmental Management

Muhammad Raqib Hayat, Abu Bakar, Muhammad Bilal*, Muhammad Ramzan Shahid Khan
Department of Computer Science, Namal University, Pakistan
***Correspondence**: mbilal2292@gmail.com

Citation | Hayat. M. R, Bakar. A, Bilal. M, Khan. M. R. S, "Data-Driven Analysis and Predictive Modeling of Urban Air Quality for Environmental Management", IJIST, Special Issue pp 508-527, May 2026

Received | March 27, 2026 **Revised |** May 06, 2026 **Accepted |** May 11, 2025 **Published |** May 15, 2026.

This research study presents a detailed analysis of air quality data collected from an Italian city for one year. The study aimed to analyze air pollution trends, explore the relationships among various pollutants and environmental variables, and build predictive models for classifying air quality levels. The dataset contained measurements of various pollutants (CO, NO_x, NO_2 and C_6H_6), sensor readings, and environmental factors (temperature, humidity). Key findings include strong correlations between certain pollutants, clear seasonal and weekly patterns in pollution levels, and the successful development of classification models to predict high pollution events with up to 99.46% accuracy. The Support Vector Machine (SVM) model outperformed all others. Feature importance analysis consistently identified the CO sensor reading as the most significant predictor, along with seasonal factors. The study presents detailed visualizations that contribute to a better understanding to a better understanding of urban air pollution dynamics and provides a foundation for developing effective air quality management strategies and early warning systems.

Keywords: Air Quality; Data Analysis; Machine Learning; Time Series; Environmental Monitoring; Predictive Modeling

Introduction:

Air pollution is a significant environmental and public health concern in urban areas worldwide. The complex interactions between various pollutants, environmental factors, and human activities make air quality analysis a challenging but crucial field of study. Understanding the patterns, relationships, and dynamics of air pollution is essential for developing effective monitoring systems, implementing targeted interventions, and formulating evidence-based policies to improve air quality and protect public health.

This study focuses on the analysis of air quality data collected from an Italian city over one year (March 2004 to February 2005). The dataset contains hourly measurements of various pollutants, including Carbon Monoxide (CO), Nitrogen Oxides (NO_x), Nitrogen Dioxide (NO_2), and Benzene (C_6H_6), along with readings from corresponding chemical sensors. Additionally, the dataset includes environmental factors such as temperature and humidity, which can influence pollution levels and dispersion patterns.

Motivation:

The motivation for this research study stems from several key considerations:

Public health impact: Air pollution is associated with numerous adverse health effects, including respiratory and cardiovascular diseases.

Environmental monitoring: Effective air quality monitoring and prediction systems are essential for environmental management and protection.

Urban planning: Understanding air pollution patterns can inform traffic management, industrial zoning, and green space allocation decisions.

Policy development: Data-driven insights can support the formulation of evidence-based policies and regulations.

Technological advancement: Sensor networks and data analytics offer new opportunities for real-time air quality monitoring and prediction.

Research Questions:

The study aimed to address the following research questions:

What are the patterns and trends in air pollutant concentrations over different time scales (hourly, daily, weekly, monthly)?

How do different pollutants correlate with each other and with environmental factors such as temperature and humidity?

How reliable are the chemical sensor readings compared to the ground truth measurements of pollutants?

Can machine learning models effectively predict high pollution events based on sensor readings, environmental factors, and temporal features?

Which factors are most important in determining air pollution levels and predicting high pollution events?

Methodology Overview:

The research was structured into six distinct phases:

Scope and Objectives: Outlining research goals, defining project scope, and selecting a methodological approach.

Exploratory Data Analysis (EDA): Initial examination of the dataset to understand its structure, distribution, and quality.

Data Preprocessing: Cleaning, transformation, and feature engineering to prepare the data for analysis.

Correlation Analysis: Investigation of relationships between variables.

Time Series Analysis: Examination of temporal patterns and forecasting of pollution levels.

Modeling: Development and evaluation of classification models to predict high pollution events.

Related Work:

Air quality analysis has been a subject of extensive research across environmental science, public health, computer science, and data analytics.

Air Quality Monitoring and Analysis:

Traditional air quality monitoring relies on fixed monitoring stations that measure pollutant concentrations using reference methods [1]. While accurate, they are limited in spatial coverage due to high costs. Recent deployments of low-cost sensor networks offer greater spatial resolution, albeit with potential accuracy tradeoffs [2]. Principal component analysis and multiple linear regression have been applied to identify urban pollution sources [3]. City-based air quality has also been estimated using hybrid spatial–temporal modelling incorporating meteorological data and urban activity patterns [4].

Time Series Analysis of Air Pollution:

Seasonal decomposition is used to identify trends and irregular fluctuations in pollutant concentrations [5]. ARIMA models have been applied to forecast daily $PM_{2.5}$ with reasonable short-term accuracy [6]. Diurnal and weekly cycles in NO_x reveal clear traffic-related patterns [7], while decreasing long-term trends in UK pollutants have been attributed to policy interventions [8].

Machine Learning for Air Quality Prediction:

Ensemble methods like Random Forest often outperform single models in air quality prediction [9]. Long Short-Term Memory (LSTM)networks capture temporal dependencies in air quality data [10]. Classification models for ozone exceedance [11] and SVM-based air quality category classification [12] have also been explored.

Sensor Calibration and Validation:

Field calibrations show that appropriate calibration significantly improves low-cost sensor accuracy [13]. Electrochemical NO_2 sensors face challenges from cross-sensitivity and environmental factors [14].

Gaps in Existing Research:

Limited integration of sensor data with environmental and temporal features in predictive modeling.

Insufficient exploration of inter-pollutant relationships for monitoring strategies.

Need for a comprehensive evaluation of ML approaches for air quality classification.

Limited research on feature importance in air quality prediction models.

Methodology:

Dataset Description:

The dataset was collected from an Italian city air quality monitoring station with hourly measurements from March 2004 to February 2005. Variables include:

Date and Time: Timestamps for each measurement.

Ground truth: CO(GT), C_6H_6(GT), NO_x(GT), NO_2(GT).

Sensor readings: PT08.S1(CO), PT08.S2(NMHC), PT08.S3(NO_x), PT08.S4(NO_2), PT08.S5(O_3).

Environmental factors: Temperature (T), Relative Humidity (RH), Absolute Humidity (AH).

The dataset contains 9,357 observations with 15 variables. Missing values are encoded as −200; NMHC(GT) has over 90% missing values.

Data Preprocessing:

Missing value treatment: NMHC(GT) dropped. Regression-based imputation for CO(GT), NO_x(GT), NO_2(GT). Rolling mean with forward/backward fill for the rest.

Outlier treatment: IQR capping at Q_1−1.5× IQR and Q_3+1.5×IQR.

Standardization: Zero mean and unit variance via Standard Scaler.

Feature engineering: Hour, Day, Month, Year, Day of Week extracted; Weekend and Rush Hour binary flags; Season and pollution-level categories; sine/cosine cyclical encodings for Hour and Month; interaction features (e.g., pollutant × humidity).

Exploratory Data Analysis:

EDA covered: (i) univariate analysis (descriptive statistics, histograms, box plots); (ii) bivariate analysis (scatter plots, correlation matrices); (iii) temporal analysis (hourly, daily, monthly aggregations).

Correlation Analysis:

Pearson correlation coefficients were calculated for all numeric variable pairs. A heatmap identified patterns; scatter plots detailed key relationships among pollutants, sensor–pollutant pairs, and pollutant– environment pairs.

Time Series Analysis:

Phases: (i) pattern visualization at daily/weekly/monthly scales; (ii) additive decomposition into trend, seasonal, and residual components; (iii) ACF/PACF analysis; and (iv) ARIMA modeling for CO(GT) forecasting.

Classification Modeling:

Binary classification task: High Pollution ≡ CO(GT) > 75th percentile.

Features: Sensor readings, environmental variables, engineered temporal features.

Split: 80/20 train–test with StandardScaler.

Models: Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Recurrent Neural Network (RNN).

Evaluation: Accuracy, AUC, precision, recall, F1-score, confusion matrices, and ROC curves.

Feature importance: Scores from tree-based models analyzed.

Data Analysis and Results:

Exploratory Data Analysis Results Univariate Distributions:

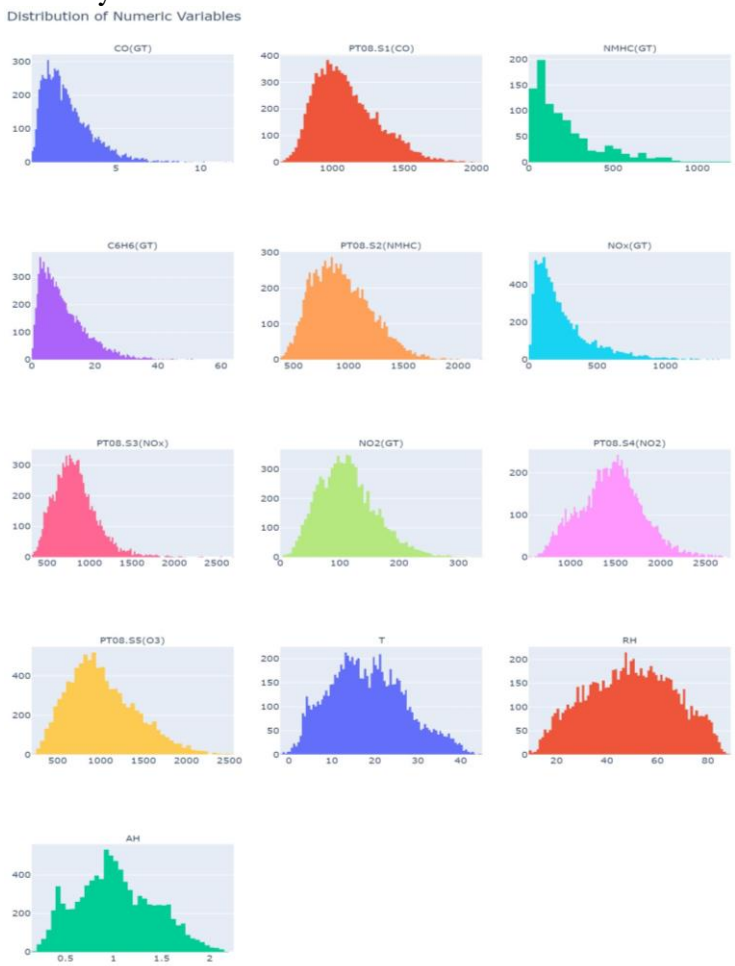

Figure 1. Histograms showing the distribution of each numerical variable.

Figure 1 shows histograms for all numerical variables. Most pollutant concentrations (CO(GT), C_6H_6(GT), NO_x(GT), NO_2(GT)) exhibit right-skewed distributions, indicating that lower values dominate, but high concentration events do occur. Temperature shows a bimodal distribution reflecting seasonal variation, while Relative Humidity is skewed towards higher values.

Figure 2 presents box plots confirming the skewness in pollutant variables, with medians closer to the lower quartile and numerous high-value outliers. These plots justify the need for outlier treatment before modeling.

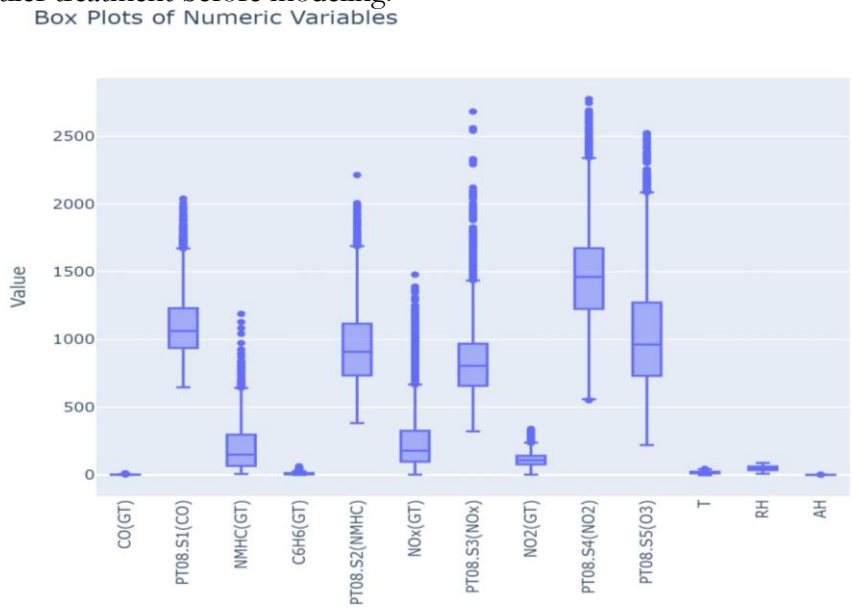

Figure 2. Box plots illustrating the distribution and potential outliers for each numerical variable.

Missing Value Analysis:

Figure 3 shows the percentage of missing values per variable. NMHC(GT) had over 90% missing data and was dropped. CO(GT), NO_x(GT), and NO_2(GT) had approximately 18% missing values; sensors and environmental variables had fewer than 4%.

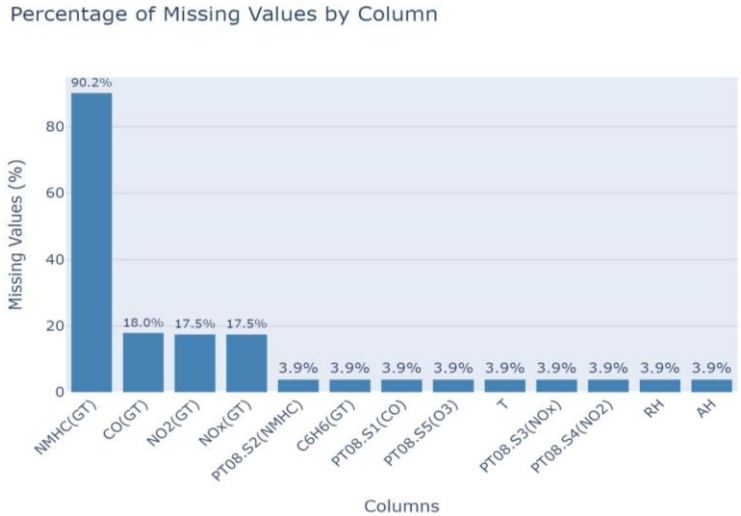

Figure 3. Bar chart showing the percentage of missing values for each variable.

Temporal Patterns Initial View:

Figure 4 shows the hourly variation of key variables across the full year. Seasonal trends are visible with denser clusters of high pollutant values during colder months, and rapid oscillations suggest strong diurnal cycles.

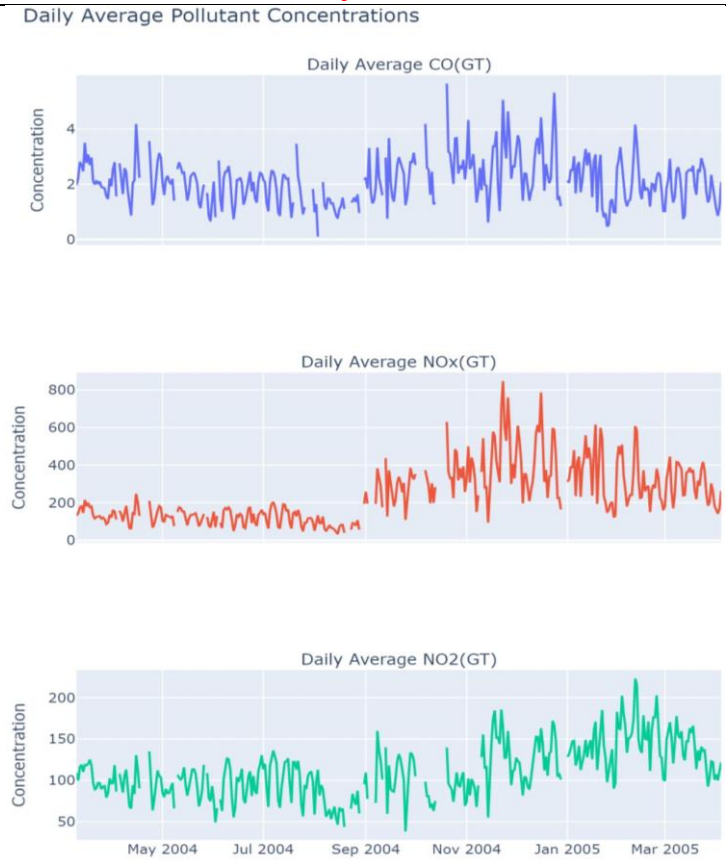

Figure 4. Time series plots showing the hourly variation of key variables over the year.

Data Preprocessing Results Outlier Treatment:

Figure 5 shows distributions *before* outlier treatment; Figure 6 shows distributions *after* IQR capping. Extreme points beyond the whiskers are replaced by Q_3+ 1.5×IQR or Q_1 −1.5×IQR, mitigating distortion while preserving the bulk distribution.

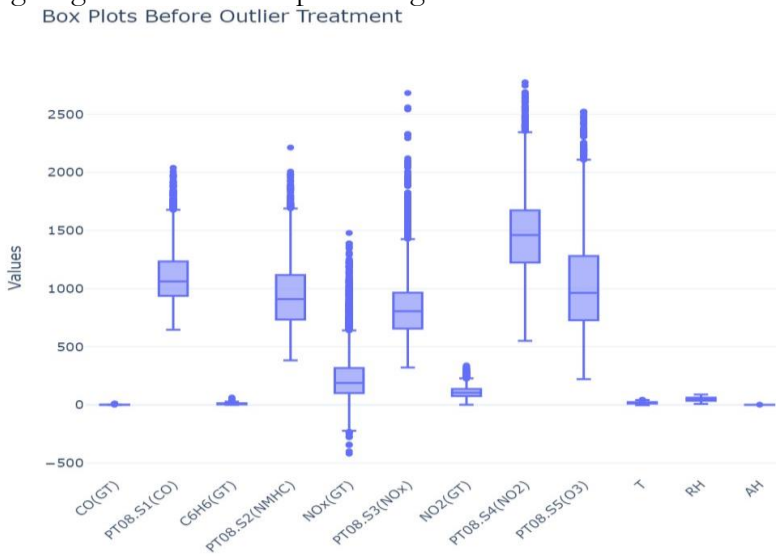

Figure 5. Box plots visualizing variable distributions and outliers **before** treatment.

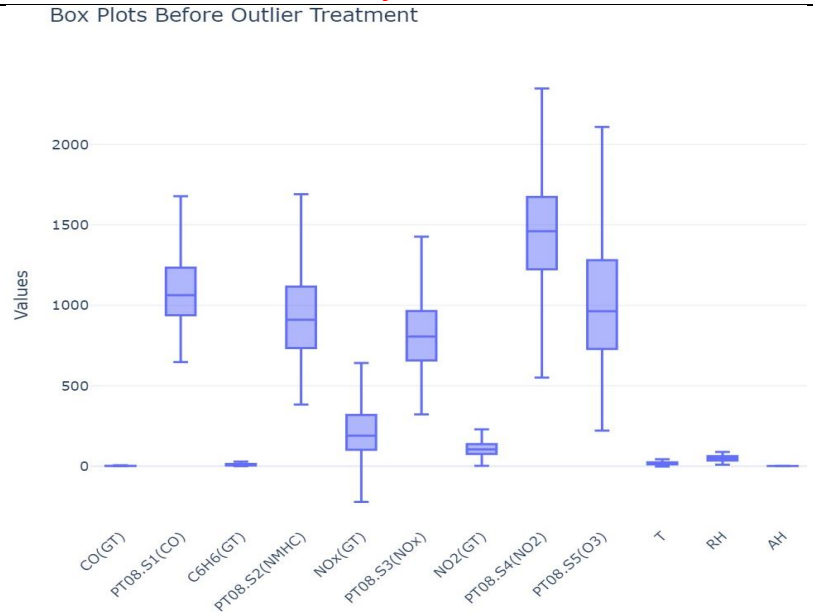

Figure 6. Box plots visualizing variable distributions **after** applying IQR capping.

Feature Engineering:

Figure 7 illustrates the sine/cosine cyclical encoding for Hour and Month, mapping them onto a unit circle to preserve their inherent periodicity and prevent artificial discontinuities at day/year boundaries.

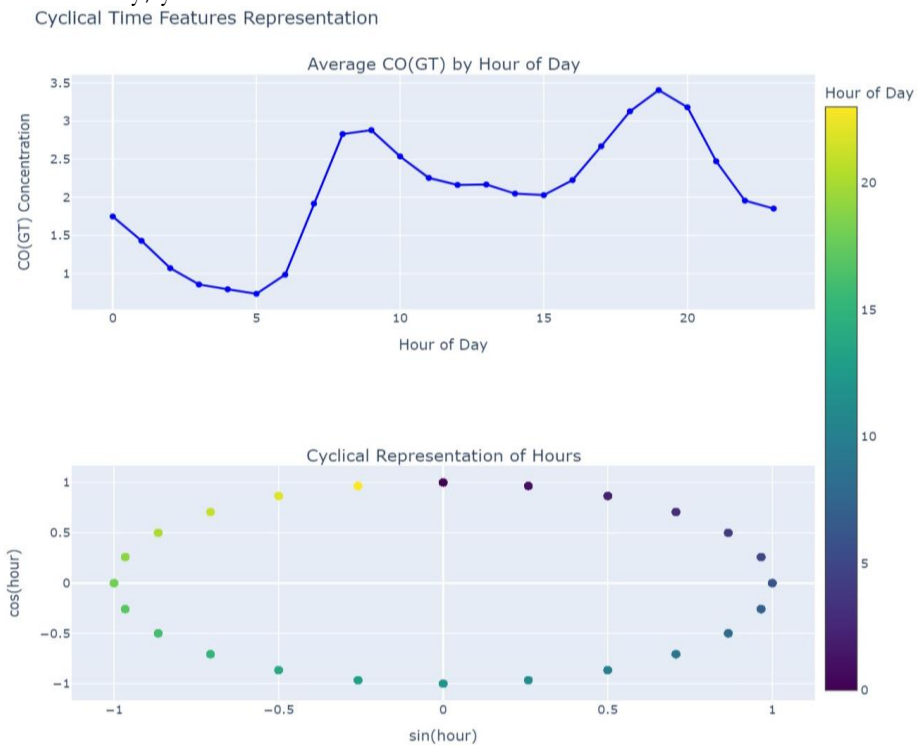

Figure 7. Visualization of cyclical encoding for Hour and Month features.

Figure 8 confirms lower median CO(GT) and reduced variability on weekends (Weekend =1) compared to weekdays, reflecting reduced traffic and commercial activity.

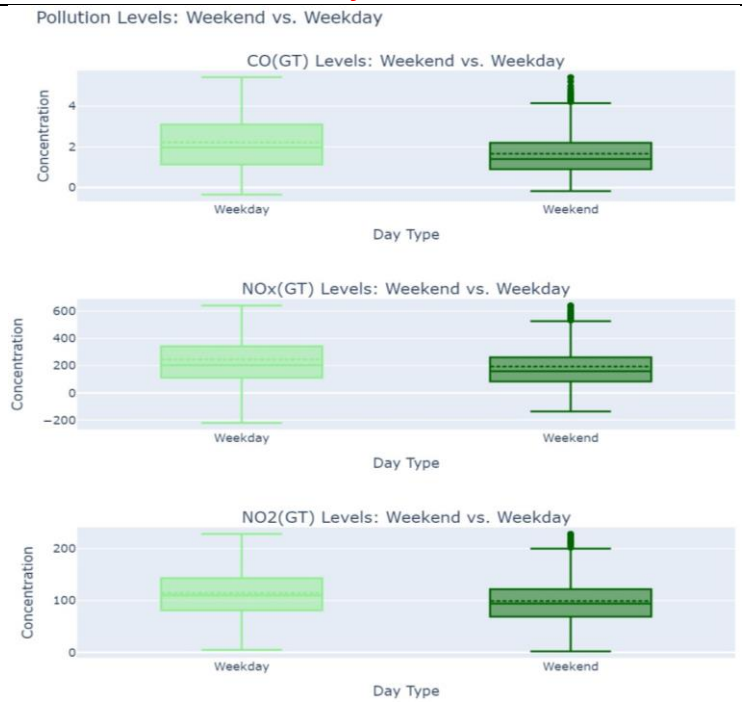

Figure 8. Box plot comparing CO(GT) distribution on weekends vs. weekdays.

Figure 9 shows significantly elevated median CO(GT) during rush hours (RushHour =1), isolating the contribution of peak commuting periods to CO pollution.

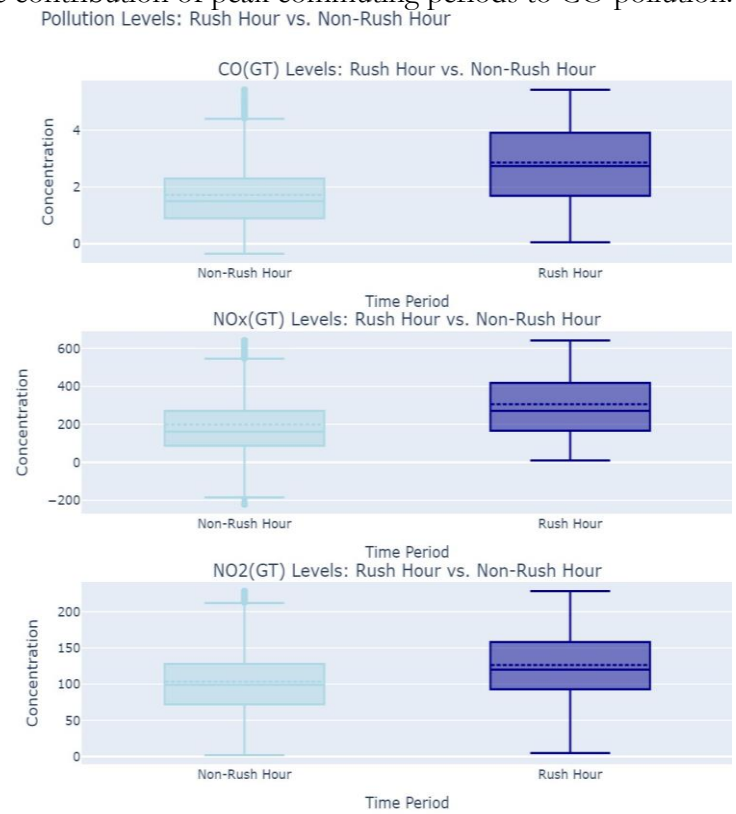

Figure 9. Box plot comparing CO(GT) distribution during rush hours vs. non-rush hours.

Figure 10 demonstrates the strong seasonal pattern:

Winter exhibits the highest CO(GT) concentrations, followed by Fall and Spring, with the lowest levels in Summer.

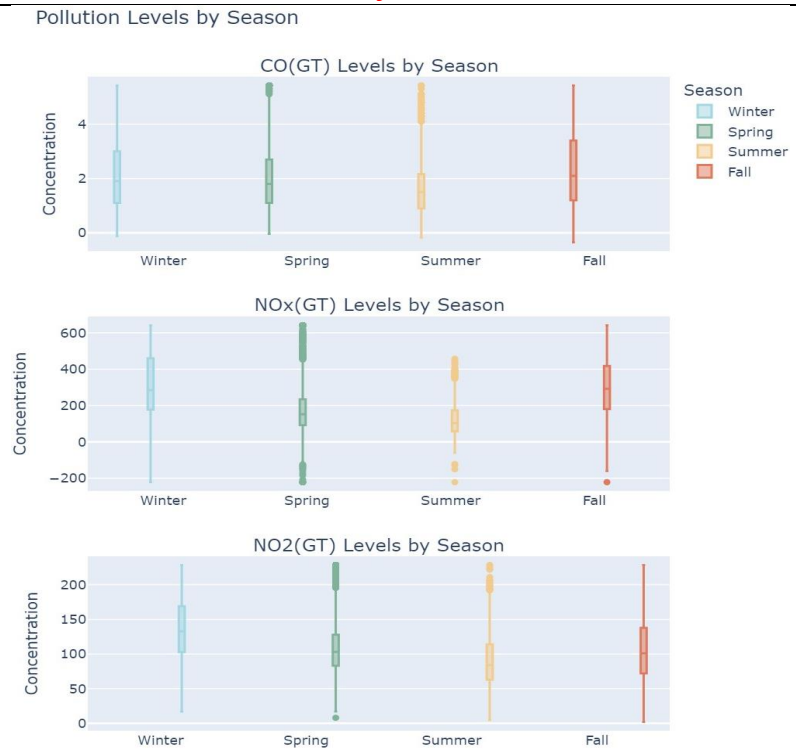

Figure 10. Box plot comparing CO(GT) distribution across seasons.

Figure 11 illustrates an example interaction feature, exploring how the CO(GT)–Temperature relationship varies with Relative Humidity. Such interaction terms allow models to capture non-additive effects.

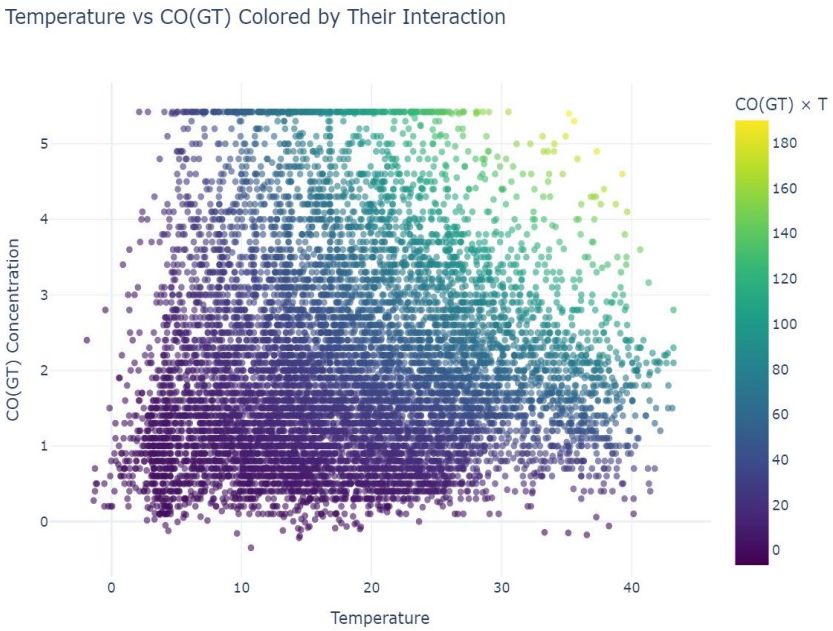

Figure 11. Scatter plot showing an example interaction feature (CO(GT) vs. T, colored by RH).

Figure 12 shows the correlation heatmap *after* preprocessing, including engineered features, to identify potential multicollinearity before modeling.

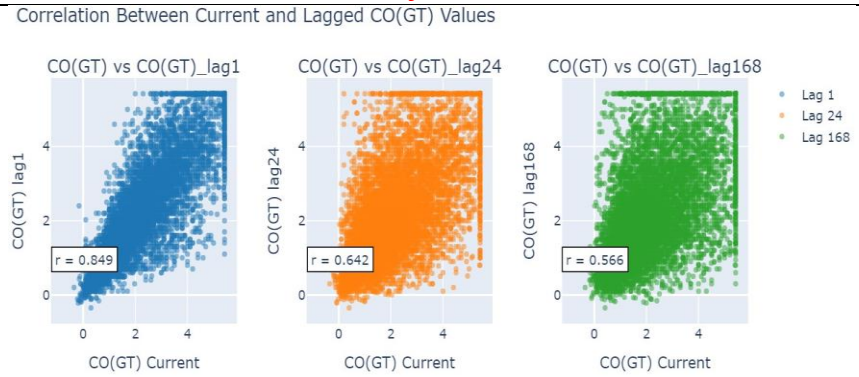

Figure 12. Correlation heatmap after preprocessing, including engineered features.

Correlation Analysis Results: Figure 13 shows the full Pearson correlation matrix. Strong positive correlations are evident among CO(GT), C_6H_6(GT), PT08.S1(CO), and PT08.S2(NMHC), suggesting common vehicular emission sources. NO_x(GT) and NO_2(GT) are also strongly correlated. PT08.S3(NO_x) shows strong *negative* correlations with several pollutants.

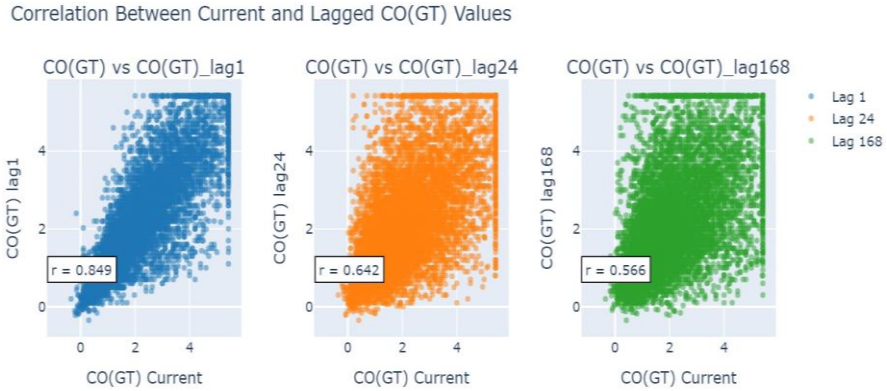

Figure 13. Heatmap of the Pearson correlation matrix for all numerical variables.

Figure 14 reveals a very strong positive linear relationship between CO and Benzene ($r \approx 0.90$), strongly implying co-emission from vehicular exhaust.

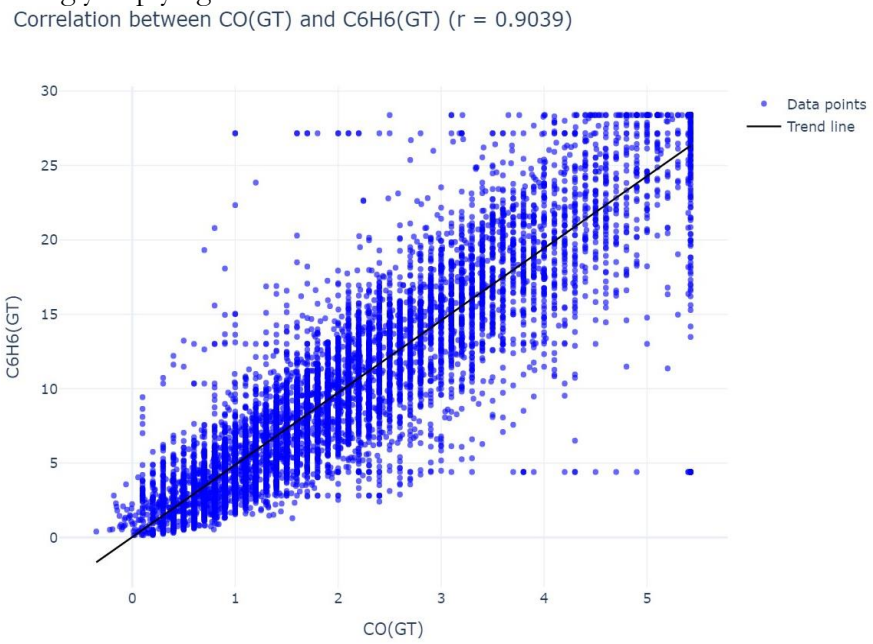

Figure 14. Scatter plot: CO(GT) vs. C_6H_6(GT) ($r \approx 0.90$).

Figure 15 illustrates the strong positive correlation between CO and NO_x ($r \approx 0.79$), reflecting their shared combustion origin.

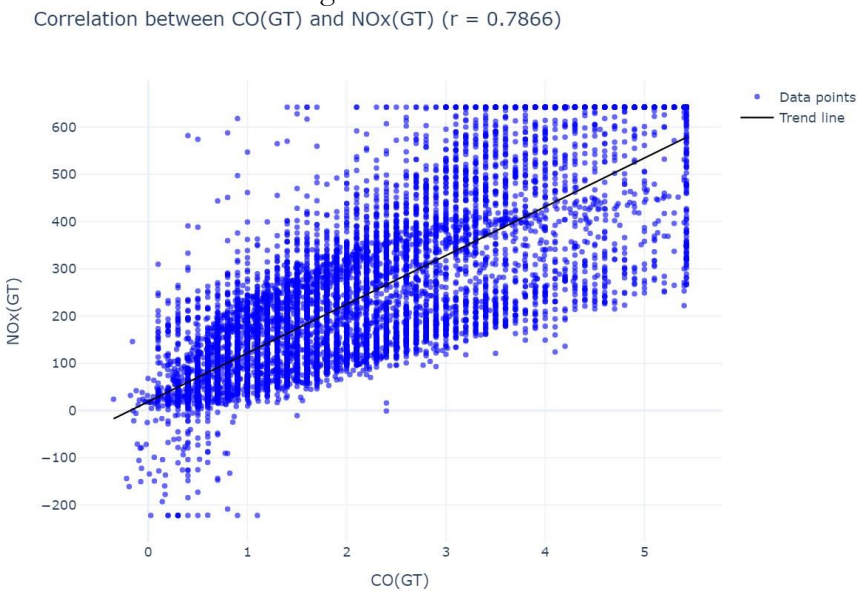

Figure 15. Scatter plot: CO(GT) vs. NO_x(GT) ($r \approx 0.79$).

Figure 16 shows the expected strong positive correlation between NO_x and NO_2 ($r \approx 0.72$); scatter reflects the variable NO_2/NO_x ratio due to photochemical conversion and atmospheric conditions.

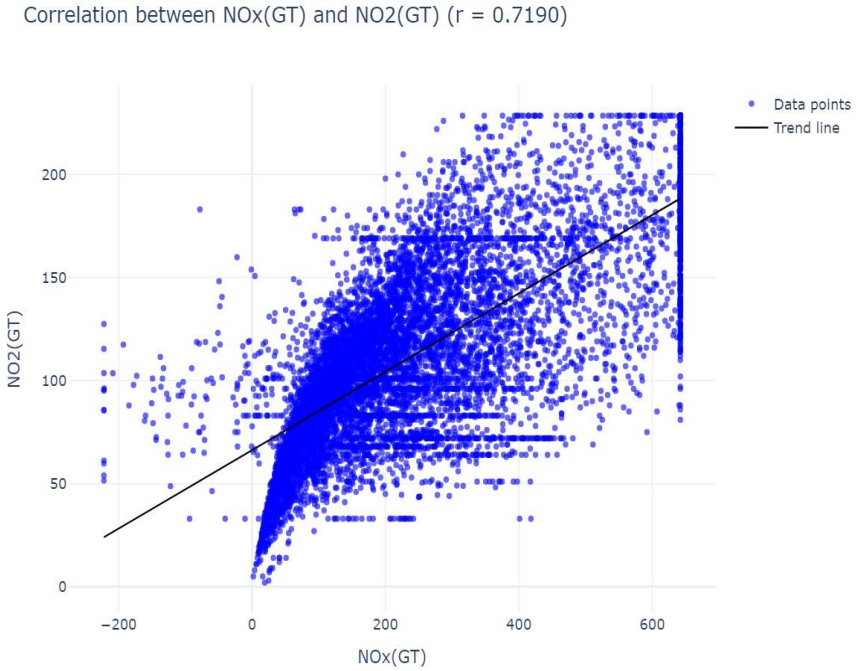

Figure 16. Scatter plot: NO (GT) vs. NO_2(GT)
$r \approx 0.72$

Figure 17 summarizes sensor–pollutant correlation performance. PT08.S1 ($r = 0.88$) and PT08.S2 ($r = 0.99$) are reliable proxies for CO and C_6H_6, respectively. PT08.S4 is unreliable for NO_2 ($r = 0.10$); PT08.S3 shows a strong *inverse* response to NO_x ($r = -0.75$).

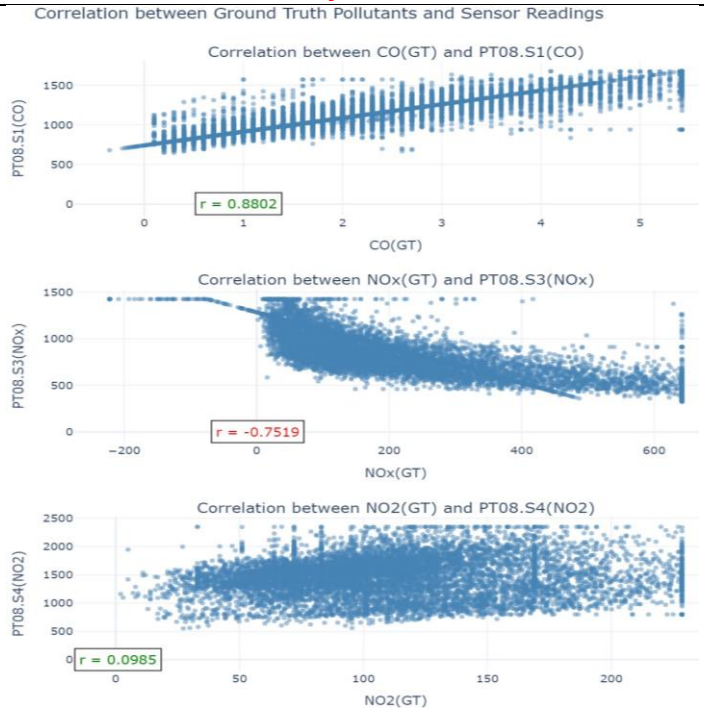

Figure 17. Correlations between ground truth pollutants and corresponding sensor readings.

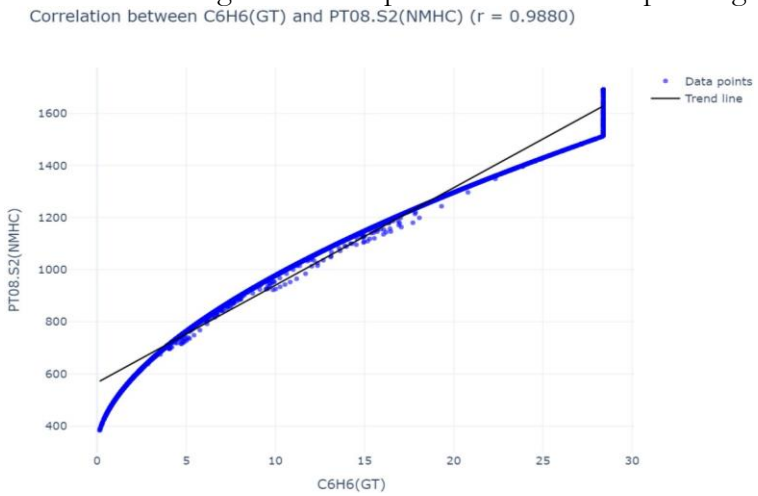

Figure 18. Scatter plot: C_6H_6(GT) vs. PT08.S2(NMHC) ($r \approx 0.99$).

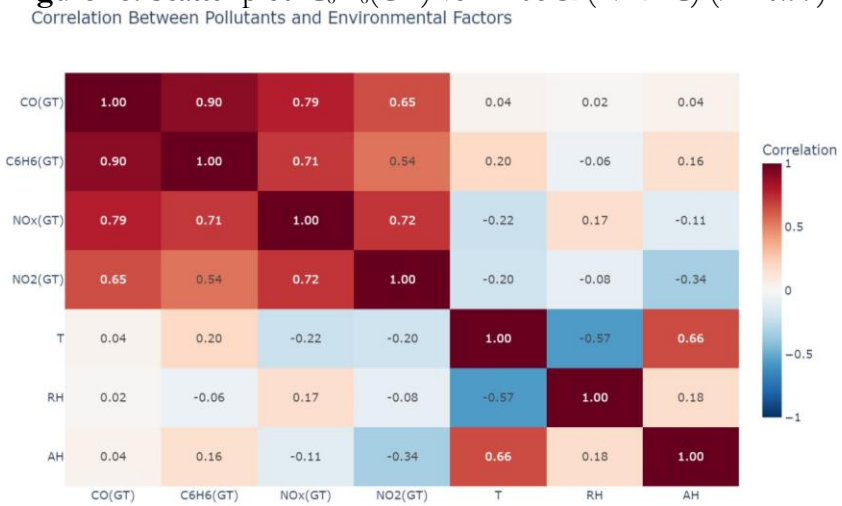

Figure 19. Correlations between pollutants and environmental factors (T, RH, AH).

Figure 18 provides a detailed view of the near-perfect relationship between C_6H_6(GT) and PT08.S2 ($r \approx 0.99$), making PT08.S2 an exceptionally strong Benzene proxy.

Figure 19 shows pollutant–environment correlations. Temperature has weak negative correlations with most pollutants; Absolute Humidity shows a weak negative correlation with NO_2 ($r = -0.34$). These weak linear relationships suggest non-linear or secondary environmental influences.

Figure 20 synthesizes the key correlation groups, highlighting traffic pollutant clusters, reliable and unreliable sensor pairs, and the weaker environmental factor influences.

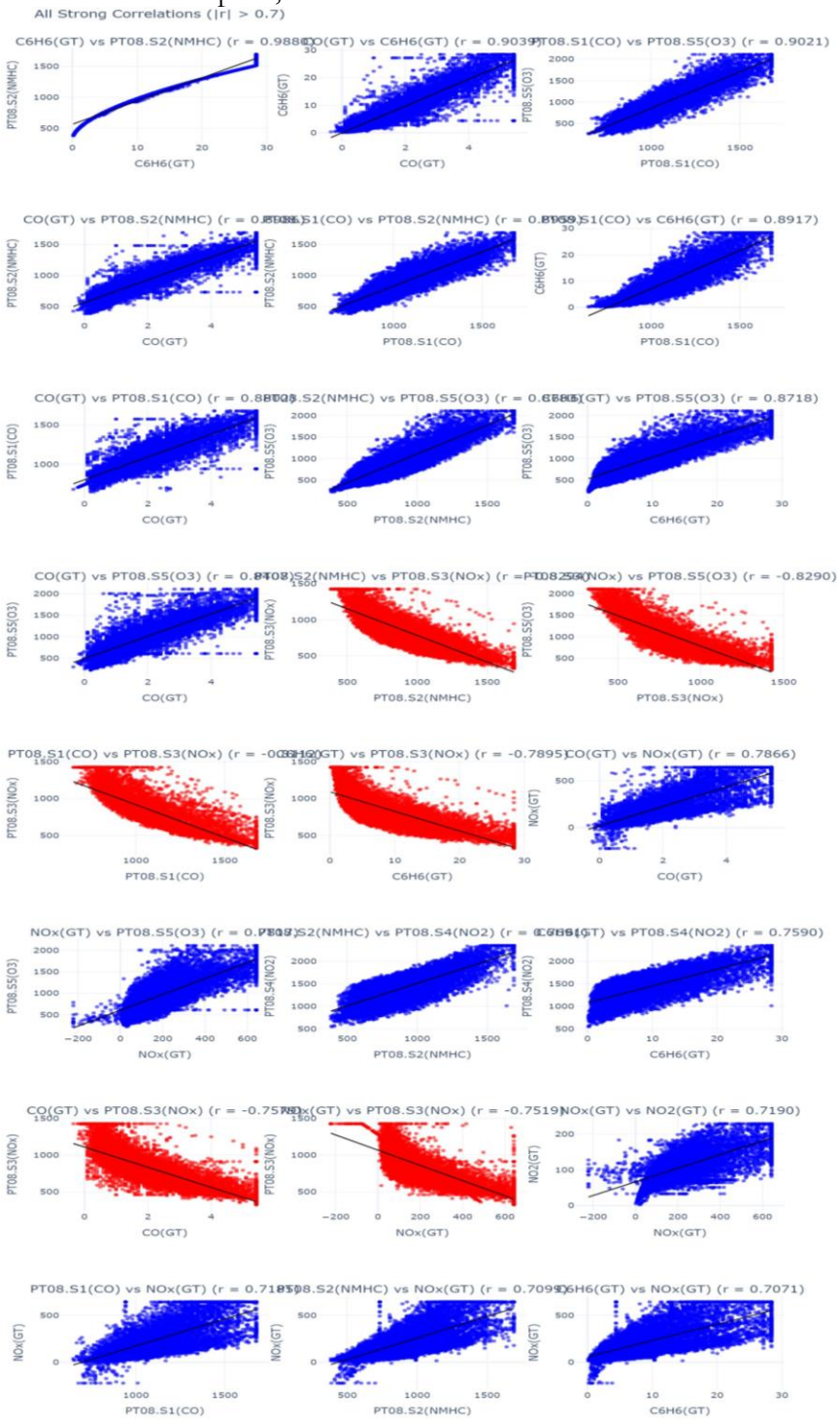

Figure 20. Combined visualization summarizing key correlation groups.

Time Series Analysis Results:

Figure 21 shows daily average concentrations over the year. CO, C_6H_6, and NO_x peak during winter months (Nov–Feb) and reach minima in summer (Jun–Aug), driven by meteorological conditions and seasonal emission sources such as residential heating.

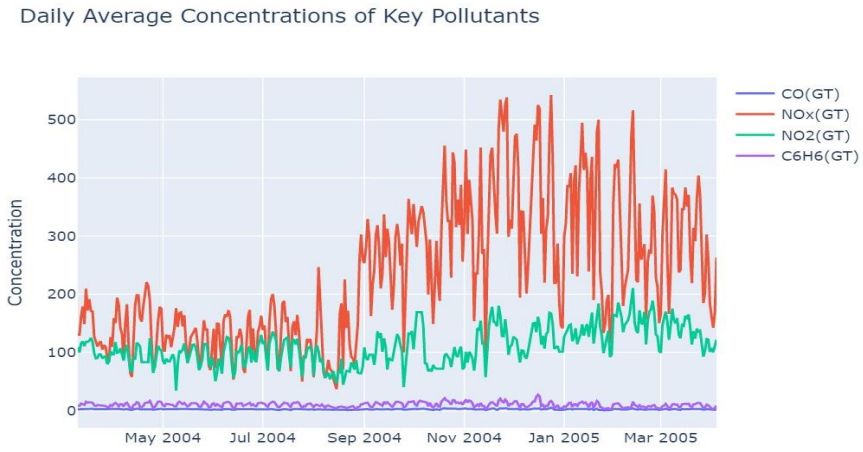

Figure 21. Time series plot of daily average pollutant concentrations.

Figure 22 further emphasizes the annual cycle through monthly averages: concentrations rise from fall to winter, peak in winter, and decline to summer lows.

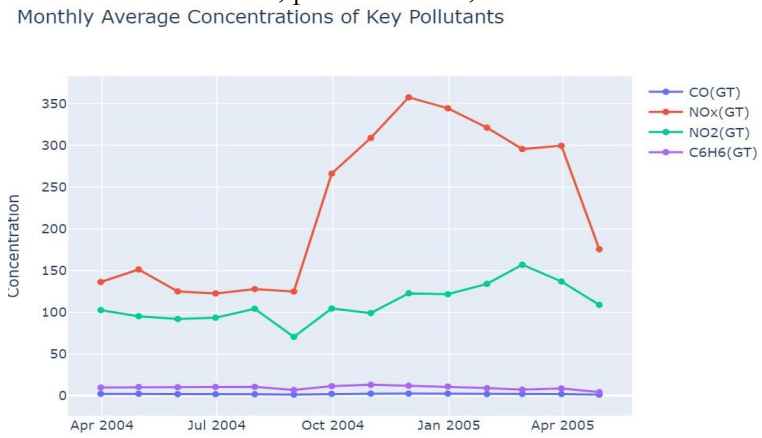

Figure 22. Time series plot of monthly average pollutant concentrations.

Figure 23 reveals the weekend effect: CO, C_6H_6, and NO_x average concentrations are significantly lower on Saturdays and Sundays, confirming the dominant contribution of weekday traffic and commerce to urban pollution.

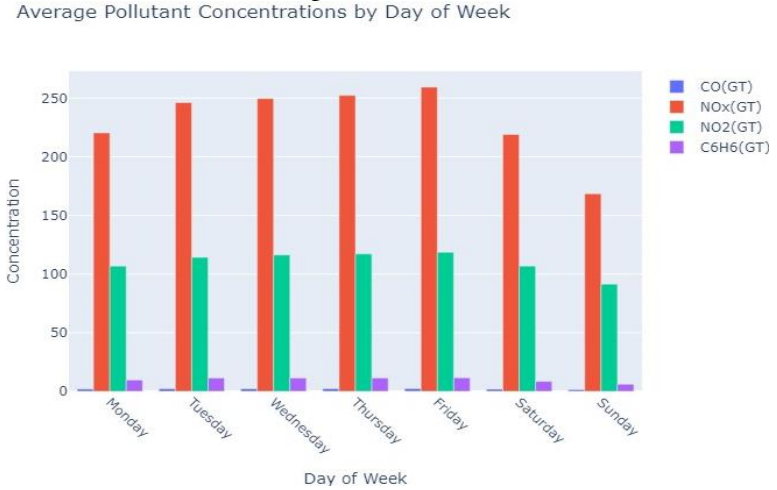

Figure 23. Average pollutant concentrations by day of the week.

Figure 24 decomposes daily CO(GT) into trend (slow seasonal variation), seasonal (repeating weekly cycle), and residual (irregular fluctuations) components, formally separating and quantifying the different time scales of variation.

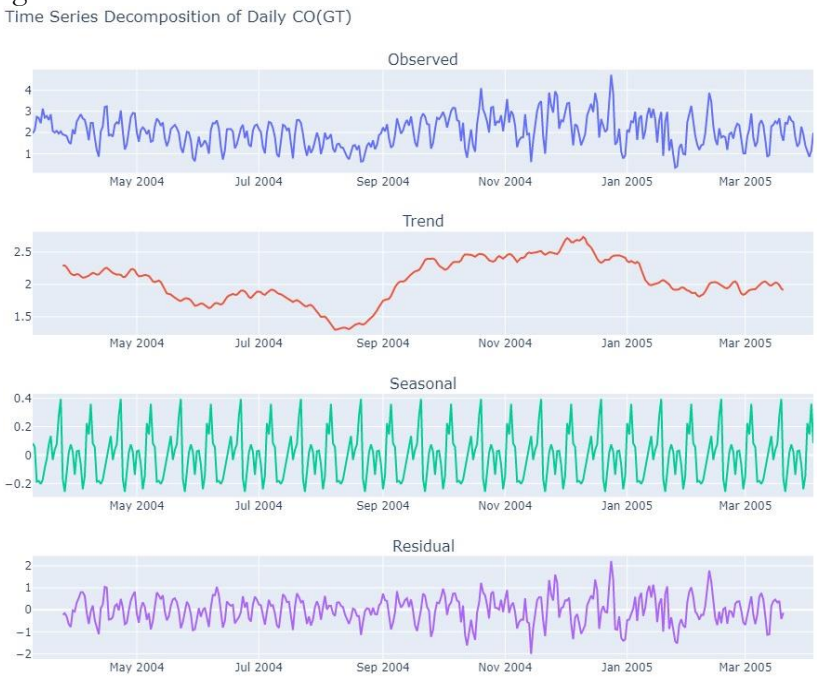

Figure 24. Additive decomposition of the daily average CO(GT) time series (weekly seasonality).

Figure 25 presents the ACF and PACF plots for daily CO(GT). Slow ACF decay confirms non-stationarity (d =1); spikes at lags 7, 14, 21 confirm weekly seasonality; PACF cut-off after lag 1 suggests AR (1). These guide ARIMA parameter selection toward ARIMA (1, 1, 1) or a seasonal variant.

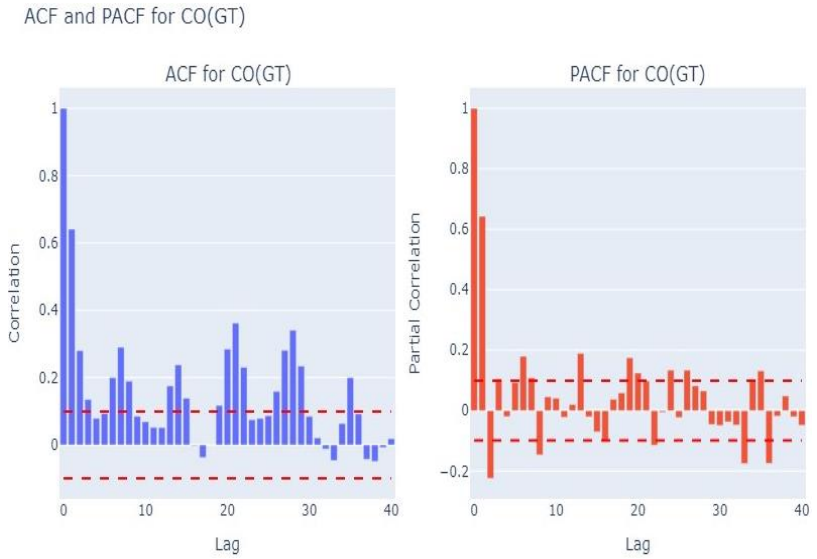

Figure 25. ACF and PACF plots for daily CO(GT).

Figure 26 shows the ARIMA(1,1,1) forecast against test data. While the model captures the general concentration level, it fails to replicate daily variability, yielding.
RMSE ≈ 0.7. A SARIMA or deep-learning model would better capture the complex dynamics.

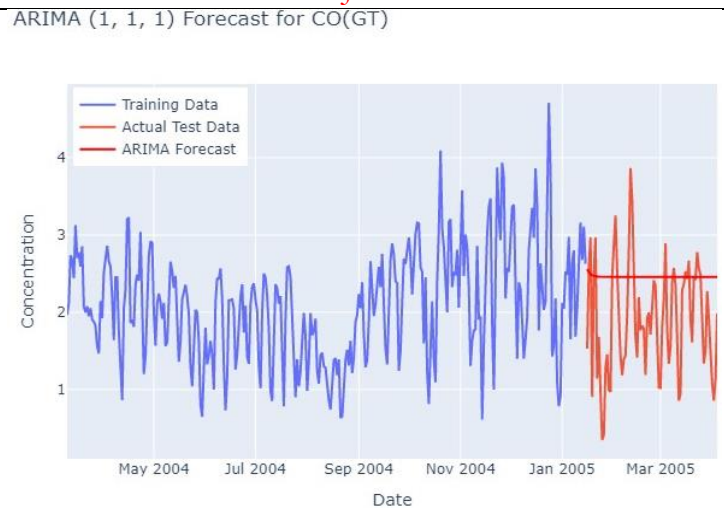

Figure 26. ARIMA (1,1,1) forecast for daily CO(GT) compared to actual test data.

Figure 27 shows an extended future forecast. As expected, the forecast quickly stabilizes and confidence intervals widen, highlighting the limitations of the simple ARIMA model for long-horizon forecasting.

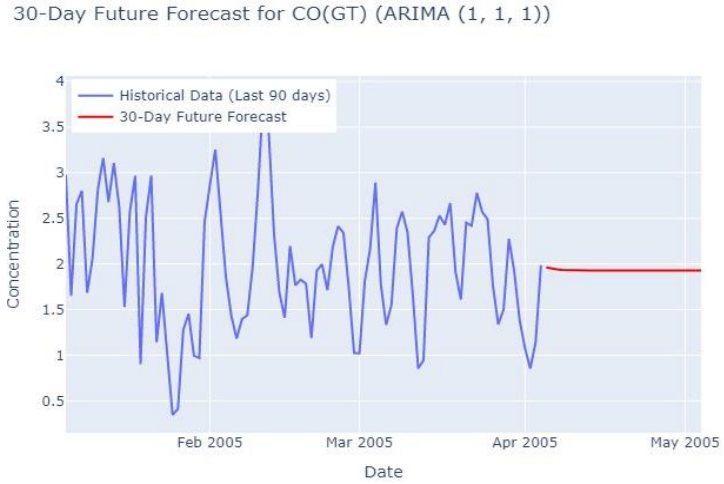

Figure 27. Extended future forecast using the fitted ARIMA (1,1,1) model.

Classification Modeling Results Target Variable Definition:

Figure 28 shows the CO(GT) distribution with the 75th-percentile threshold (≈2.63 mg/m^3) that defines the "High Pollution" binary class (25% positive, 75% negative).

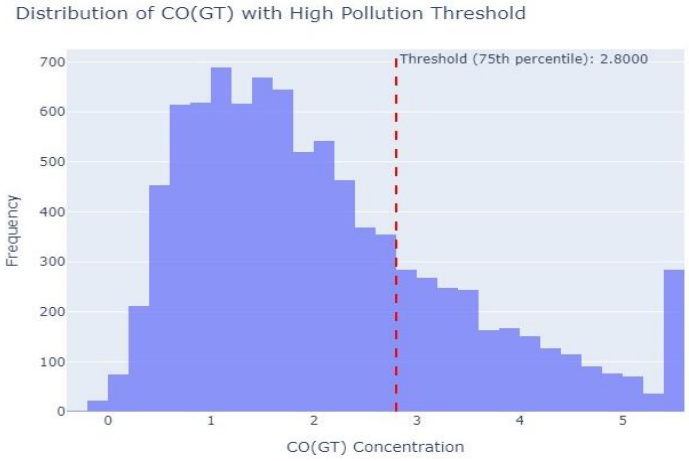

Figure 28. CO(GT) distribution with the 75th percentile threshold defining the "High Pollution" class.

Model Performance Comparison:

Figure 29 provides a comparative overview of all classifiers across key metrics. SVM and LR emerge as top performers; Table 1 gives the exact numbers.

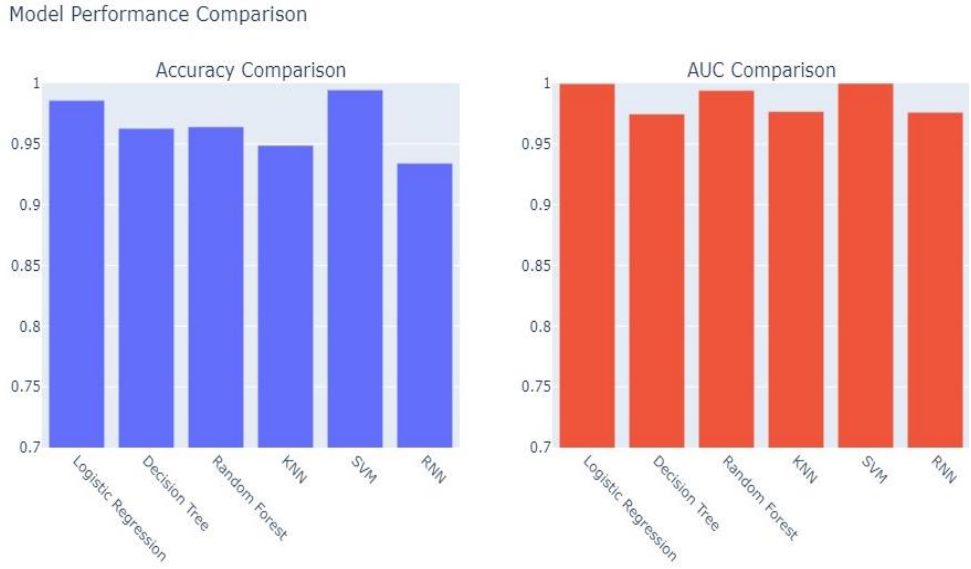

Figure 29. Comparison of Accuracy, Precision, Recall, F1-Score, and AUC across all classification models.

Table 1. Detailed performance metrics for all classification models.

Model	Acc.	Prec.	Rec.	F1	AUC
LR	0.9859	0.97	0.96	0.97	0.9998
DT	0.9630	0.90	0.91	0.91	0.9748
RF	0.9641	0.95	0.88	0.91	0.9941
KNN	0.9489	0.88	0.88	0.88	0.9767
SVM	**0.9946**	**0.99**	**0.99**	**0.99**	**0.9999**
RNN	0.9320	0.88	0.81	0.84	0.9772

Confusion Matrices and ROC Curves:

Figure 30 shows the Logistic Regression confusion matrix. High values on the main diagonal confirm strong performance on both Normal and High pollution classes.

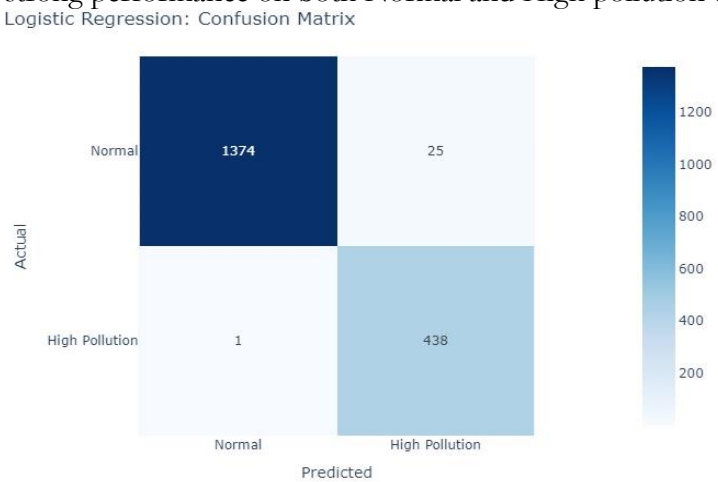

Figure 30. Confusion matrix for the Logistic Regression model.

Figure 31 displays the Logistic Regression ROC curve, which hugs the top-left corner (AUC ≈ 1.00), confirming excellent discrimination across all thresholds.

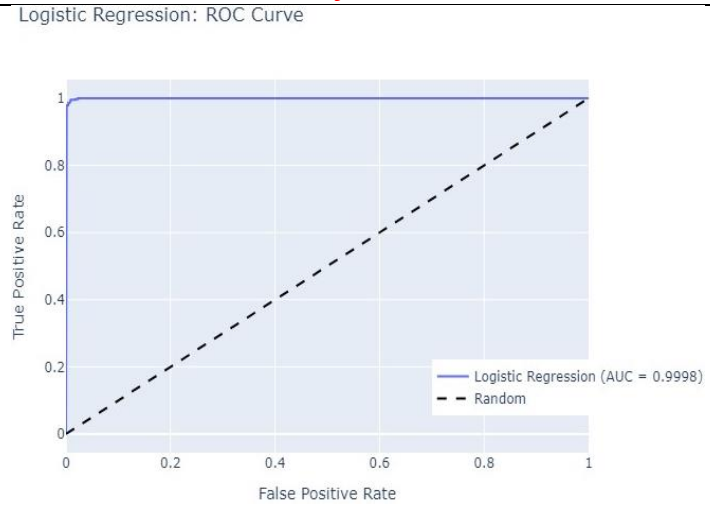

Figure 31. ROC curve and AUC for the Logistic Regression model.

Figure 32 overlays ROC curves for all six models. SVM and LR curves are virtually indistinguishable and closest to the ideal corner. RF is also strong, while DT, KNN, and RNN are slightly weaker.

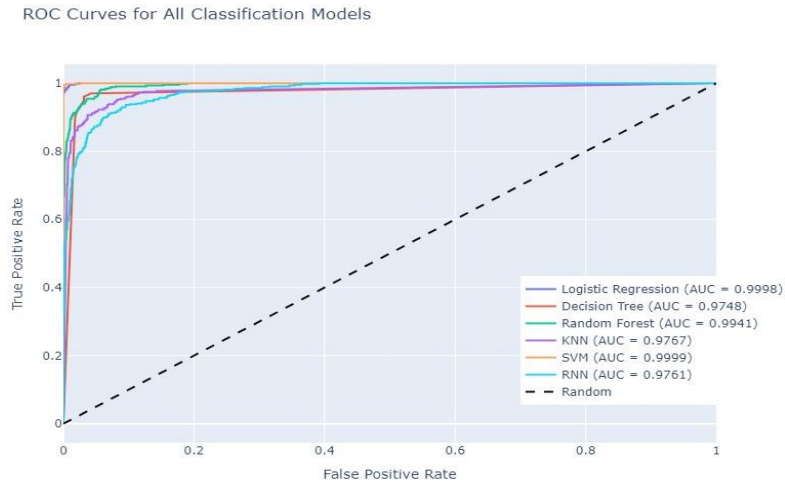

Figure 32. Combined ROC curves for all classification models.

Feature Importance:

Figure 33 shows the Decision Tree feature importance scores, providing an initial indication of influential variables.

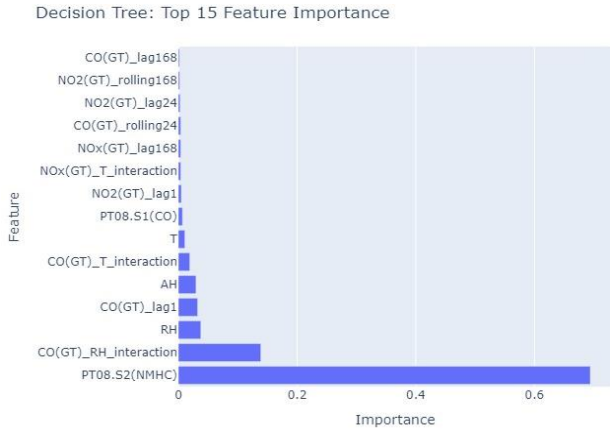

Figure 33. Feature importance scores from the Decision Tree model.

Figure 34 shows the more robust Random Forest feature importance scores. PT08.S1(CO) is consistently the most critical predictor. Temporal features Month and Hour

rank highly, followed by PT08.S5(O_3) and PT08.S2(NMHC). Environmental factors contribute but rank lower than sensor and temporal features.

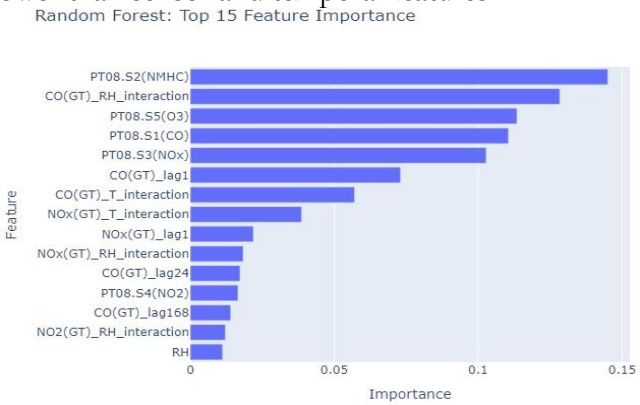

Figure 34. Feature importance scores from the Random Forest model.

Conclusion:

This comprehensive analysis has provided valuable insights into the patterns, relationships, and dynamics of urban air pollution. The study successfully addressed all five research questions and demonstrated the effectiveness of data-driven approaches for air quality understanding and prediction.

Key Findings:

Temporal patterns: Clear seasonal variations (winter peaks), distinct weekend effects in NO_x and CO, and pronounced diurnal rush-hour peaks.

Pollutant relationships: Strong positive correlations among CO, C_6H_6, and NO_x imply shared vehicular sources. PT08.S2 is a near-perfect C_6H_6 proxy; PT08.S4 is unreliable for NO_2.

Predictive modeling: SVM achieved 99.46% accuracy (AUC ≈1.00). PT08.S1(CO) is the dominant predictor; Month and Hour are key temporal features.

Implications:

Monitoring a subset of pollutants may suffice given strong inter-pollutant correlations. Clear temporal patterns support targeted seasonal and rush-hour interventions. High classification performance demonstrates the feasibility of near-real-time early-warning systems.

Limitations:

Single-location, single-year data limits generalizability. Relevant variables (wind speed, precipitation, traffic volume) are absent. The basic ARIMA model underperformed for operational forecasting, and the 75th percentile threshold is not aligned with regulatory standards.

Future Work:

Multi-site spatial analysis; incorporation of meteorological and traffic data; advanced forecasting (SARIMA, Prophet, LSTM); regulatory-threshold classification targets; formal source-apportionment studies.

Declaration:

Conflict of Interest: All authors declare that they have no conflict of interest.

Acknowledgment: The authors acknowledge the use of artificial intelligence-based tools for paraphrasing and improving the clarity and language of the manuscript.

Author Contribution:

Muhammad Raqib Hayat and Abu Bakar contributed to data collection, preprocessing, and analysis. Muhammad Bilal (corresponding author) led research design, modeling, and manuscript preparation. Muhammad Ramzan Shahid Khan contributed to the literature review and result validation. All authors reviewed and approved the final manuscript.

References:

[1] W. H. O. R. O. for Europe, "Air quality guidelines: global update 2005: particulate matter, ozone, nitrogen dioxide and sulfur dioxide," Dec. 2006, Accessed: Mar. 23, 2026. [Online]. Available: https://iris.who.int/handle/10665/107823

[2] "A Comparative Analysis of Monitored Ambient Hazardous Air Pollutant Levels with Modeled Estimates from the Assessment System for Population Exposure Nationwide | Request PDF." Accessed: Mar. 23, 2026. [Online]. Available: https://www.researchgate.net/publication/295184698_A_Comparative_Analysis_of_Monitored_Ambient_Hazardous_Air_Pollutant_Levels_with_Modeled_Estimates_from_the_Assessment_System_for_Population_Exposure_Nationwide

[3] Srishti Jain, S. K. Sharma, "Source apportionment of PM10 in Delhi, India using PCA/APCS, UNMIX and PMF," *Particuology*, vol. 37, pp. 107–118, 2018, doi: https://doi.org/10.1016/j.partic.2017.05.009.

[4] Yu Zheng, Furui Liu, "U-Air: when urban air quality inference meets big data," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2013, [Online]. Available: https://dl.acm.org/doi/10.1145/2487575.2488188

[5] R. Bhardwaj and D. Pruthi, "Time series and predictability analysis of air pollutants in Delhi," *Proc. 2016 2nd Int. Conf. Next Gener. Comput. Technol. NGCT 2016*, pp. 553–560, Mar. 2017, doi: 10.1109/NGCT.2016.7877476.

[6] Meng Du, Yixin Chen, "A Novel Hybrid Method to Predict PM2.5 Concentration Based on the SWT-QPSO-LSTM Hybrid Model," *Comput. Intell. Neurosci.*, 2022, doi: 10.1155/2022/7207477.

[7] Stuart K. Grange, David C. Carslaw, "Using meteorological normalisation to detect interventions in air quality time series," *Sci. Total Environ.*, vol. 653, pp. 578–588, 2019, doi: https://doi.org/10.1016/j.scitotenv.2018.10.344.

[8] "UK air quality showed clear improvement from 2015 to 2024 but breaching of targets remains very common - Environmental Science: Atmospheres (RSC Publishing) DOI:10.1039/D5EA00055F." Accessed: May 10, 2026. [Online]. Available: https://pubs.rsc.org/en/content/articlehtml/2025/ea/d5ea00055f

[9] Dr. Rais Abdul Hamid Khan, Mr. Kshirsagar Sopan Bapu, "A Review : Air Pollution Prediction using Machine Learning Techniques," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 10, no. 3, pp. 644–647, 2024.

[10] Xiang Li, Ling Peng, "Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation," *Environ. Pollut.*, vol. 231, pp. 997–1004, 2017, doi: https://doi.org/10.1016/j.envpol.2017.08.114.

[11] Héctor Jorquera, Ricardo Pérez, "Forecasting ozone daily maximum levels at Santiago, Chile," *Atmos. Environ.*, vol. 32, no. 20, pp. 3415–3424, 1998, doi: https://doi.org/10.1016/S1352-2310(98)00035-1.

[12] U. Mahalingam, K. Elangovan, H. Dobhal, C. Valliappa, S. Shrestha, and G. Kedam, "A machine learning model for air quality prediction for smart cities," *2019 Int. Conf. Wirel. Commun. Signal Process. Networking, WiSPNET 2019*, pp. 452–457, Mar. 2019, doi: 10.1109/WISPNET45539.2019.9032734.

[13] Md Masudur Rahman, "Recommendations on the measurement techniques of atmospheric pollutants from in situ and satellite observations: a review," *Arab. J. Geosci.*, vol. 16, no. 5, p. 326, 2023, [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10116117/

[14] Bas Mijling, Qijun Jiang, "Practical field calibration of electrochemical NO2 sensors for urban air quality applications," *Atmos. Meas. Tech. Discuss.*, 2017, doi: 10.5194/amt-2017-43.