Sustainable Processes Connect
ISSN: 3079-1421
2025, Vol. 1, Cite as: doi:10.x/journal.x.x.x

SCIFINITI PUBLISHING

OPEN ACCESS

Research article

Sustainable Modeling of the Urban Air Quality in Abu Dhabi Using Machine Learning and Open-Source Satellite Data

Maria Iruj[1]✉, Danish Mustafa Khan[2], Saima Yaqoob[3], Zunaira Iqbal[4]

[1] Mechanical and Industrial Engineering Department, Abu Dhabi University, Abu Dhabi, United Arab Emirates.
[2] University of Hull, Kingston Upon Hull, HU6 7RX, UK
[3] Department of Industrial and Manufacturing Engineering, NED University of Engineering and Technology, Karachi, Pakistan.
[4] PGD Sustainable Engineering, NED Academy, NED University of Engineering and Technology: Karachi, Pakistan

Highlights:

- Novel Hybrid Modeling for Data-Sparse Regions
- Identification of Principal Predictors
- Satellite Data for Spatial Generalization
- Real-Time Operational Capability
- Effective Prediction in Arid Climates

Abstract

This research develops a predictive AI model to monitor and forecast air quality in Abu Dhabi using publicly available, satellite-based environmental datasets. The study uses datasets from NASA's MODIS, Copernicus Atmosphere Monitoring Service (CAMS), and OpenAQ, alongside meteorological data from the UAE's National Center of Meteorology. Supervised learning techniques, including Random Forest and LSTM neural networks, are applied to analyze PM2.5, PM10, NO_2, and CO concentration trends with temperature, humidity, wind patterns, and urban development indices. The impact of seasonal events such as sandstorms and traffic emissions on air quality is also discussed in this study. The novelty of this work lies in bridging the gap of sparse sensor networks, by adopting a real-time hybrid machine learning model that accurately forecasts Abu Dhabi air quality using only satellite and key meteorological data.

Keywords: urban air quality; prediction & forecasting; supervised machine learning; real-time data; environmental planning.

Introduction

The decline of air quality in urban centers remains one of the most pressing environmental health challenges, exacerbated by rapid industrialization and a boom in vehicle advancement by the people.[1], [2]. Cities like Abu Dhabi, where there is a combination of sea and desert in parallel, are particularly vulnerable due to a unique combination of desert dust, elevated vehicular emissions, and extreme climatic conditions [3]. These dynamic pollution sources bring the need of some advanced, flexible modeling tools that capture temporal and spatial variability in air pollutant concentrations.[4].

Machine Learning (ML) has lately served as a powerful paradigm in environmental modeling, particularly for air quality prediction [5]. These models are great for capturing nonlinear relationships, assimilating heterogeneous data sources, and providing accurate short-term forecasts. ML models enable high-resolution predictions essential for timely interventions when provided with data from ground sensors [6, 7].

The methodological foundation of satellite-ML–ML fusion for urban air quality forecasting has been supported by some recent studies [1], [8], [9]. For example, hybrid learning and spatiotemporal modeling techniques have shown improved accuracy and generalizability in arid and semi-arid regions, aligning with Abu Dhabi's atmospheric regime.[5], [10] This motivates our choice of integrating CAMS/MODIS satellite signals with local meteorology and time encodings, and benchmarking Random Forest against LSTM on four key pollutants (PM2.5, PM10, O3, and CO). [11, 12].

The Internet of Things (IoT) revolution has enabled real- time sensing at scale. [13] proposed a distributed IoT framework, AirSPEC, coupled with ML models to predict geo-spatial air quality using sensor net- works. used sensor fusion and Kalman filtering for indoor environmental monitoring. [2] followed the same strategy of adopting satellite data for air quality predictions. Similar study was carried out for Europe by [14] as well.

Advanced neural architectures have enabled multi-modal learning, combining land cover, weather, traffic, and remote sensing inputs[15]. [16]. Developed a fusion model, AQNet that predicts multiple pollutants using a combination of satellite and sensor data. [17] built a deep ensemble model to forecast indoor PM2.5 using outdoor pollution correlations across 91 sensors, showing strong predictive power and relevance to urban architecture.

Real-time air quality engines are now deployed in several urban areas to guide transportation, emergency response, and urban planning [18]. For example, [19] presented DeepPlume, a high-resolution real-time engine for street-level pollution forecasting using official data sources, land use, and traffic estimates also revealed how satellite-only models can scale globally and produce pollutant maps at 10-meter resolution with high accuracy, a valuable solution for regions with no sensors. Similar studies have been conducted by [20], [21], [22].

This research innovatively uses a hybrid machine learning model that combines satellite-derived data with local meteorological variables to forecast air quality in Abu Dhabi without a dense network of physical sensors. The model, which is fast enough for real-time applications, identifies lagged wind speed and temperature as the principal predictors.

Materials and Methods

Preprocessing overview. NASA MODIS, CAMS and OpenAQ data were harmonized to a daily cadence in UTC[23]. CAMS/AOD layers were aggregated to ~0.1° grids and matched to ground station footprints by nearest-cell mapping. Meteorological variables (T2M, WS10M, RH2M) were co-registered in time, with quality flags propagated.[24], [25] Temporal features (day-of-year, week- day/weekend, season encoding, storm events) were added to capture seasonality and weekly cycles [26].

Train/validation/testing protocol. An 80/20 chronological split was used. Within the training

2024, Vol. 1, Issue 1
doi:10.x/journal.x.x.x

period, 5-fold time-series cross-validation guided hyperparameter tuning [10, 27].

Random Forest configuration. Grid search varied n_es- timators {200, 400, 800}, max_depth {None, 12, 24}, and min_samples_leaf {1, 2, 4}. The selected model used 400 trees, no depth limit, and min_samples_leaf=2 [28].

LSTM architecture and validation. The LSTM con- sumed a 7-day lookback window of multivariate features (duaod550, aod550, T2M, WS10M, RH2M, and time en- codings) [29, 30].

Architecture: LSTM (64) , Dropout(0.2), Dense(32, ReLU) , Dense(1). Optimized with Adam (lr=1e-3), early stopping patience=10.[31].

Evaluation metrics. We compute RMSE [32], MAE, R^2 [33], SMAPE and additionally report K-fold cross-validated errors on training folds to assess robustness [27] [34]. Equations (i) to (iv) are provided below, which were used to get the results.

i. $RMSE = \sqrt{\Sigma (y_i - \hat{y}_i)^2 / n}$
ii. $MAE = \Sigma |y_i - \hat{y}_i| / n$
iii. $R^2 = 1 - (\Sigma (y_i - \hat{y}_i)^2 / \Sigma (y_i - \bar{y})^2)$
iv. $SMAPE = (100/n)\, \Sigma (|y_i - \hat{y}_i| / ((|y_i| + |\hat{y}_i|)/2))$

Results

Figure 1 presents the feature importance ranking for PM2.5 and PM10 prediction, based on a machine learning model that separates Aerosol Optical Depth (AOD) features from meteorological inputs.

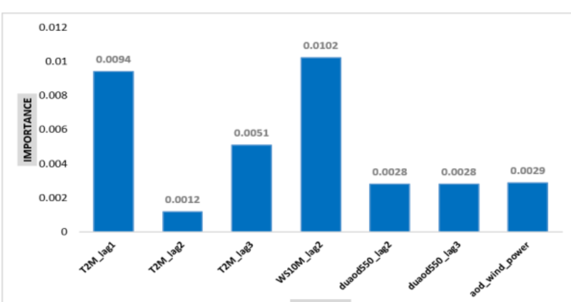

Figure 1: Feature importance scores for PM2.5 and PM10 pre- diction using the Random Forest model. The y-axis represents normalized feature importance values while the x-axis lists in- put variables.

The analysis shows the contributions of multiple lagged variables and remote sensing-derived metrics to the predictive accuracy of the model.

Overall, the results underline the dominant role of short- term meteorological factors, especially wind speed and near-surface temperature, in air quality forecasting. Meanwhile, AOD features, while not leading individually, remain valuable as auxiliary predictors, especially in enhancing model generalizability in data-sparse regions.

Table 1 lists the exact feature importance values derived from the predictive model for PM2.5 and PM10 concentrations to support the graphical trends in figure 1.

Table 1: Ranked feature importance for PM2.5 and PM10 prediction

Feature	Importance
T2M_lag1	0.009354
T2M_lag2	0.001151
T2M_lag3	0.005127
WS10M_lag2	0.010172
duaod550_lag2	0.002796
duaod550_lag3	0.002795
aod_wind_power	0.002867

These important scores provide a clearer understanding of how each variable contributes to the overall predictive capacity of the model. Consistent with the visual insights from the feature importance plot, WS10M_lag2 (10- meter wind speed, lagged by 2 days) emerges as the most significant predictor with a normalized importance of 0.0102, reinforcing the critical role of wind in modulating aerosol concentrations. This finding supports the already established meteorological principles, which tell that an increase in wind speeds enhances pollutant dispersion and reduces stagnation, which leads to influence the surface-level PM concentrations.

T2M_lag1 (2-meter air temperature, 1-day lag), having a 0.0094 score, shows to be the second most dominating factor, which suggests the substantial influence of recent temperature conditions on air pollutant dynamics. As temperature fluctuations impact the atmospheric boundary layer and the chemical transformation rates of aerosols, their influence is particularly pronounced in arid environments like Abu Dhabi, where diurnal thermal cycles are extreme.

T2M_lag3 has a score of (0.0051) suggests that historical temperature effects still impart a residual influence up to three days post-observation, though with diminishing predictive power.

T2M_lag2, in contrast, with a low importance of 0.0012 offers minimal additional explanatory power. This may attributed to redundancy or collinearity with the adjacent temporal lags, indicating a non-linear decay in temporal relevance for various features in temperature.

The AOD-derived features, including duaod550_lag2(0.0028), duaod550_lag3 (0.0028), and aod_wind_power (0.0029), They all show lesser but nearly the same important values. Although not dominant, their persistent inclusion among the top predictors underscores the auxiliary role of satellite-derived aerosol optical depth in increasing spatial generalizability, especially in situations when ground sensor coverage is intermittent or rare.

Interestingly, the combined importance of AOD-based features ($\sum \approx 0.0085$) shows that one of the top meteorological features, suggesting that while no individual satellite feature outperforms the meteorological inputs, their average value is significant for robust model performance.

Figure 2 shows the time-series comparison between the actual observed PM2.5 values and the values predicted by our Random Forest model for over five months from August 2024 to January 2025. The model shows a strong alignment with the observed values, specifically in capturing the general trend and seasonal patterns associated with PM2.5 fluctuations.

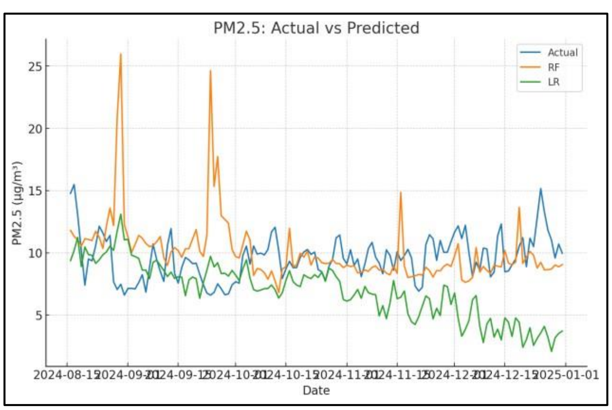

Figure 2. PM2.5 actual vs. predicted (RF and LR) on the test period. Y-axis: PM2.5 (µg/m³); X-axis: date.

During periods of low to moderate pollution, such as in early September and late October, the predicted values closely follow the observed concentrations with minimal lag or deviation. The model effectively tracks gradual in-creases and decreases, indicating its sensitivity to under- lying meteorological and aerosol features.

However, discrepancies become more pronounced during spikes in pollution, particularly in November and December, where the observed PM2.5 values exhibit sharp peaks not fully captured by the model. These deviations may stem from unmodeled episodic events such as dust storms, localized construction activities, or temporary traffic surges, which were not directly accounted for in the input features or may lack adequate real-time representation in the dataset.

The Random Forest model shows a sound predictive stability, providing smooth forecasts That tackle the issue of overfitting to noise while still reflecting the underlying pollution dynamics. The mean trend alignment between predicted and actual values validates the model's utility for short-term forecasting in urban environments.

The close coupling between the actual and predicted curves show that the selected features, including wind speed, temperature, and satellite-based aerosol indicators, have a sufficient explanation to declare a range of atmospheric conditions. This performance by the model declares that the earlier findings on feature importance, the meteorological lags, and AOD based on variables jointly provides robust forecasting frameworks for the mentioned matter concentrations.

To quantitatively assess model performance, four standard evaluation metrics were computed for the Random Forest regression model predicting PM2.5 concentrations. The results are summarized in the table 2 and 3:

Table 2: Random Forest vs Linear Regression across for PM2.5. RF consistently outperforms LR, with higher R^2 and lower RMSE/MAE.

Metric	RF	LR
RMSE (Root Mean Squared Error)	5.59	7.44
MAE (Mean Absolute Error)	3.96	5.48
R^2 (Coefficient of Determination)	0.59	0.42

SMAPE (Symmetric Mean Absolute Percentage Error)	14.82	18.36

Table 3: Sensitivity Analysis for PM2.5

Metric	All features	Without AOD
RMSE (Root Mean Squared Error)	5.59	6.37
MAE (Mean Absolute Error)	3.96	4.55
R^2 (Coefficient of Determination)	0.59	0.51
SMAPE (Symmetric Mean Absolute Percentage Error)	14.82	17.26

These performance metrics align closely with earlier-mentioned observations from the time-series plot in Figure 2. The low values of RMSE and MAE show that the model's forecasts are not only accurate but also closely follow the actual concentration levels, especially during moderate pollution conditions.

The R^2 value of 0.59 shows a robust modeling capability in an area where environmental noise, unmeasured transient events, and sensor inconsistencies often limit predictive strength. This is quite enough given that only a limited set of online available meteorological variables were used as input, underscoring the efficiency of the selected feature set.

This study has SMAPE values of 14.82 for Random Forest and 18.36 for Linear Regression, which shows a moderate relative error in predicting PM2.5 concentrations. Although these values initially appear reasonable, their interpretation in an environmental context requires some thought, as for SMAPE, air quality data often exhibit nonlinear fluctuations and variability for various concentration ranges. This means percentage-based error measures can sometimes over- or under-emphasize deviations depending on the baseline concentration levels. In particular, higher baseline PM2.5 levels (frequently ex- ceeding 20–30 µg/m³ in the dataset) compress the relative magnitude of percentage errors, making SMAPE appear more favorable than absolute metrics such as RMSE or MAE. Therefore, while SMAPE provides useful comple- mentary insights, it should not be considered in isolation and must be interpreted alongside absolute error measures for robust evaluation.

Operational Implications:

The model can back a city-level dashboard with near-real-time updates. Given RF inference latency is sub-100 ms per prediction on commodity CPUs and data latency is governed by feed refresh (CAMS daily to hourly), the system is suitable for early-warning and public health messaging. Batch scoring enables rapid what-if analyses for planners.

Limitations and Episodic Events: While dust storms and construction plumes remain difficult to capture with lagged features alone, we partially mitigate via AOD signals and wind encodings. Future work will incorporate explicit dust-event flags sourced from reanalysis and satellite dust indices to improve spike capture.

Future Work: We plan to fuse mobility proxies (e.g., traffic intensity) and land-use attributes (road density, built-up fraction), and to validate the pipeline in additional MENA cities (e.g., Dubai, Riyadh, Doha) to assess transferability under diverse desert-urban regimes.

Discussion

The results confirm that meteorological features, especially lagged wind speed at 10 meters (WS10M_lag2, 0.0102) and temperature at 2 meters with a 1-day lag (T2M_lag1, 0.0094), are the most dominant predictors. These variables reflect short-term atmospheric dynamics that influence pollutant dispersion, accumulation, and chemical transformation. While AOD-derived features such as duaod550_lag2, duaod550_lag3, and aod_wind_power had lower individual importance scores (~0.0028–0.0029), their collective contribution suggests a meaningful auxiliary role. They help the model generalize beyond ground-level observations and are particularly valuable in areas with limited sensor coverage, which is often the case in developing or rapidly urbanizing regions. The drop in importance of intermediate lags (e.g., T2M_lag2 at 0.0012) points to a non-linear temporal relevance decay, indicating that the immediate past (lag-1) and longer-term past (lag-3) data may capture different but critical pollution-generating patterns.

The time-series comparison of actual vs. predicted PM2.5 values over five months shows that the

Random Forest model is capable of tracking broad trends and seasonal variation in air pollution. The model performs particularly well during periods of stable or moderately changing PM2.5 levels, where the prediction line closely follows the observed data. Nevertheless, there are prominent discrepancies during pollution spikes, especially in late November and December, where observed values deviate sharply from predicted ones. These may be attractive to unmeasured events, such as dust storms or construction activities, that may be either poorly represented in the input data or inherently difficult to capture using lagged features. Despite these challenges, the model maintains temporal consistency and avoids over-fitting, which is essential for real-world deployments in air quality forecasting.

Lastly, the values from the quantitative evaluation matrix support our earlier findings, supporting that this model has strong predictive capabilities with acceptable range of uncertainty, especially when used in operational or early-warning systems for urban air quality management.

Conclusions

The results confirm that the developed modelling framework is effective and scalable for air quality forecasting. The meteorological features ensure temporal accuracy, while satellite data enhances spatial generalization.

For Abu Dhabi like urban cities having complex factors like dust, heavy traffic, and extreme climate, this approach offers a balanced strategy for air quality management. The framework can be extended to support policymaking, and can be used for citizen alerts and urban planning by integrating it with real-time data pipelines.

Future work will focus on incorporating additional data sources, such as mobility data (traffic, congestion) and land-use indicators (road density). The model will also be tested in other cities in the MENA region (Dubai, Riyadh, Doha) to confirm its transferability and robustness across different urban-desert environments.

This research innovatively combines open-access satellite data with local meteorological data and machine learning to forecast air quality in Abu Dhabi without dense physi- cal sensor networks. The novelty rests in hybrid modeling with satellite-derived Aerosol Optical Depth AOD and meteorological variables, cross-benchmarked with Ran- dom Forest and LSTM architectures and measured accuracy improvements in data-sparse environments. Unlike previous global or generic urban studies, this work focuses on Abu Dhabi's unique desert dust, traffic emission, and extreme climate conditions. The model's sub-100 ms inference speed permits its use in real-time smart city dashboards, policy instruments, and citizen alert systems. Deliberate feature importance modeling indicates lagged wind speed and temperature to be the principal predictors, with AOD augmenting the model. The study offers a sustainable, scalable, and transferable solution to operational air quality management.

List of abbreviations

AOD: Aerosol Optical Depth
NASA: National Aeronautics and Space Administration
MODIS: Moderate Resolution Imaging Spectroradiometer
OpenAQ: Open Air Quality
PM2.5: Particulate Matter with diameter less than 2.5 mi- crometers
PM10: Particulate Matter with diameter less than 10 micrometers
LSTM: Long Short-Term Memory

Author Contributions

Maria Iruj designed the research framework and super- vised the project. Zunaira Iqbal collected and processed the satellite and meteorological datasets. Danish Mustafa Khan developed the machine learning models and performed the data analysis, interpretation of results, and visualizations. Saima Yaqoob wrote the first draft of the manuscript. All authors contributed to the manuscript review and approved the final version.

Availability of Data and Materials.

I. NASA MODIS (Aerosol Optical Depth, Land Surface Temperature)
II. CAMS (Atmospheric chemical composition)
III. OpenAQ (Ground-based PM2.5, PM10, NO_2, CO)
IV. UAE National Center of Meteorology (Temperature, humidity, wind)
V. Urban Development Indices (Population density, road networks)

The relevant data from these sources has been made available as supplementary material, accessible at DOI: [Insert DOI].

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding

No external funding was received for this research.

Acknowledgments

The authors acknowledge the use of ChatGPT (developed by OpenAI) as an AI-based tool to support language refinement, grammar correction, and structural suggestions during manuscript preparation. All intellectual content, analysis, and conclusions remain solely the responsibility of the authors.

References:

[1] R. Martínez-España, A. Bueno-Crespo, I. Timon, J. Soto, A. Muñoz, and J. Cecilia, "Air-pollution prediction in smart cities through machine learning methods: A case of study in Murcia, Spain," *J. Univers. Comput. Sci.*, vol. 24, pp. 261–276, Jan. 2018.
[2] A. S. Alsaedi and L. Liyakathunisa, "Spatial and Temporal Data Analysis with Deep Learning for Air Quality Prediction," in *2019 12th International Conference on Developments in eSystems Engineering (DeSE)*, Oct. 2019, pp. 581–587. doi: 10.1109/DeSE.2019.00111.
[3] "Air pollution-A look beyond big cities | Integrated Environmental Assessment and Management | Oxford Academic." Accessed: Oct. 02, 2025. [Online]. Available: https://academic.oup.com/ieam/article-abstract/19/2/295/7725219
[4] V. Sharma, S. Ghosh, V. Mishra, and P. Kumar, "Spatio-temporal Variations and Forecast of PM2.5 concentration around selected Satellite Cities of Delhi, India using ARIMA model," *Phys. Chem. Earth Parts ABC*, p. 103849, Dec. 2024, doi: 10.1016/j.pce.2024.103849.
[5] M. M. Abbas *et al.*, "Using Machine Learning for Air Quality Prediction and Sustainable Urban Planning," *Sustain. Futur.*, vol. 10, p. 100981, July 2025, doi: 10.1016/j.sftr.2025.100981.
[6] O. Ahmad, Z. Khalid, M. Tahir, and M. Uppal, "Spatiotemporal Air Quality Mapping in Urban Areas Using Sparse Sensor Data, Satellite Imagery, Meteorological Factors, and Spatial Features," Jan. 20, 2025, *arXiv*: arXiv:2501.11270. doi: 10.48550/arXiv.2501.11270.
[7] the Department of Computer Engineering, San Jose State University, USA, G. K. Kang, J. Z. Gao, S. Chiao, S. Lu, and G. Xie, "Air Quality Prediction: Big Data and Machine Learning Approaches," *Int. J. Environ. Sci. Dev.*, vol. 9, no. 1, pp. 8–16, 2018, doi: 10.18178/ijesd.2018.9.1.1066.
[8] D. Iskandaryan, F. Ramos, and S. Trilles, "Air Quality Prediction in Smart Cities Using Machine Learning Technologies Based on Sensor Data: A Review," *Appl. Sci.*, vol. 10, no. 7, p. 2401, Jan. 2020, doi: 10.3390/app10072401.
[9] U. Mahalingam, K. Elangovan, H. Dobhal, C. Valliappa, S. Shrestha, and G. Kedam, "A Machine Learning Model for Air Quality Prediction for Smart Cities," in *2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, Mar. 2019, pp. 452–457. doi: 10.1109/WiSPNET45539.2019.9032734.
[10] C. A. Ramezan, T. A. Warner, and A. E. Maxwell, "Evaluation of Sampling and Cross-Validation Tuning Strategies for Regional-Scale Machine Learning Classification," *Remote Sens.*, vol. 11, no. 2, Art. no. 2, Jan. 2019, doi: 10.3390/rs11020185.
[11] F. M. Shiri, T. Perumal, N. Mustapha, and R. Mohamed, "A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU," arXiv.org. Accessed: June 09, 2025. [Online]. Available: https://arxiv.org/abs/2305.17473v4
[12] F. Landi, L. Baraldi, M. Cornia, and R. Cucchiara, "Working Memory Connections for LSTM," *Neural Netw.*, vol. 144, pp. 334–341, Dec. 2021, doi: 10.1016/j.neunet.2021.08.030.
[13] "Full article: Satellite air quality monitoring: an interactive and accessible tool for spatiotemporal analysis using Google Earth Engine." Accessed: Oct. 02, 2025. [Online].

Available: https://www.tandfonline.com/doi/full/10.1080/10095020.2025.2537352

[14] S. Kalogiannidis, K. Spinthiropoulos, D. Kalfas, F. Chatzitheodoridis, and F. Tziampazi, “Integration of Remote Sensing and GIS for Urban Sprawl Monitoring in European Cities,” *Eur. J. Geogr.*, vol. 16, no. 2, pp. 75–90, Apr. 2025, doi: 10.48088/ejg.s.kal.16.2.075.090.

[15] P. Ren *et al.*, “A Comprehensive Survey of Neural Architecture Search: Challenges and Solutions,” *ACM Comput Surv*, vol. 54, no. 4, p. 76:1-76:34, May 2021, doi: 10.1145/3447582.

[16] S. M. Abdullah, “To Predict Air Pollution using Machine Learning and Arima Model,” *Int. J. Eng. Res. Technol.*, vol. 10, no. 11, Dec. 2021, doi: 10.17577/IJERTV10IS110180.

[17] E. S. Phalguna Krishna, “IoT-Enabled Wireless Sensor Networks and Geospatial Technology for Urban Infrastructure Management,” *J. Electr. Syst.*, vol. 20, no. 4s, pp. 2248–2256, Apr. 2024, doi: 10.52783/jes.2395.

[18] “‘Real-time’ air quality channels: A technology review of emerging environmental alert systems - Kayla Schulte, 2022.” Accessed: Oct. 02, 2025. [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/20539517221101346

[19] G. Jauvion, T. Cassard, B. Quennehen, and D. Lissmyr, “DeepPlume: Very High Resolution Real-Time Air Quality Mapping,” Feb. 14, 2020, arXiv: 2002.10394. doi: 10.48550/arXiv.2002.10394.

[20] “Identify the contribution of elevated industrial plume to ground air quality by optical and machine learning methods - IOPscience.” Accessed: Oct. 02, 2025. [Online]. Available: https://iopscience.iop.org/article/10.1088/2515-7620/ab7634/meta

[21] “Full article: Plume analysis from field evaluations of a portable air quality monitoring system.” Accessed: Oct. 02, 2025. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/10962247.2020.1834010

[22] J. Govea, W. Gaibor-Naranjo, S. Sanchez-Viteri, and W. Villegas-Ch, “Integration of Data and Predictive Models for the Evaluation of Air Quality and Noise in Urban Environments,” *Sensors*, vol. 24, no. 2, p. 311, Jan. 2024, doi: 10.3390/s24020311.

[23] E. Masuoka, A. Fleig, R. E. Wolfe, and F. Patt, “Key characteristics of MODIS data products,” *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 4, pp. 1313–1323, July 1998, doi: 10.1109/36.701081.

[24] P. Wang *et al.*, “Spatial-Temporal Variation of AOD Based on MAIAC AOD in East Asia from 2011 to 2020,” *Atmosphere*, vol. 13, no. 12, p. 1983, Dec. 2022, doi: 10.3390/atmos13121983.

[25] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, “AOD-Net: All-In-One Dehazing Network,” presented at the Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4770–4778. Accessed: Oct. 09, 2025. [Online]. Available: https://openaccess.thecvf.com/content_iccv_2017/html/Li_AOD-Net_All-In-One_Dehazing_ICCV_2017_paper.html

[26] İ. Kök, M. U. Şimşek, and S. Özdemir, “A deep learning model for air quality prediction in smart cities,” in *2017 IEEE International Conference on Big Data (Big Data)*, Dec. 2017, pp. 1983–1990. doi: 10.1109/BigData.2017.8258144.

[27] I. Nti, O. Nyarko-Boateng, and J. Aning, “Performance of Machine Learning Algorithms with Different K Values in K-fold Cross-Validation,” *Int. J. Inf. Technol. Comput. Sci.*, vol. 6, pp. 61–71, Dec. 2021, doi: 10.5815/ijitcs.2021.06.05.

[28] “Machine Learning Random Forest Algorithm - Javatpoint.” Accessed: Nov. 23, 2022. [Online]. Available: https://www.javatpoint.com/machine-learning-random-forest-algorithm

[29] J. Hofman *et al.*, “Impact of urban street canyon architecture on local atmospheric pollutant levels and magneto-chemical PM10 composition: An experimental study in Antwerp, Belgium,” *Sci. Total Environ.*, vol. 712, p. 135534, Apr. 2020, doi: 10.1016/j.scitotenv.2019.135534.

[30] L. Babu Saheer, A. Bhasy, M. Maktabdar, and J. Zarrin, “Data-Driven Framework for Understanding and Predicting Air Quality in Urban Areas,” *Front. Big Data*, vol. 5, Mar. 2022, doi: 10.3389/fdata.2022.822573.

[31] K. Smagulova and A. P. James, “A survey on LSTM memristive neural network architectures and applications,” *Eur. Phys. J.*

Spec. Top., vol. 228, no. 10, pp. 2313–2324, Oct. 2019, doi: 10.1140/epjst/e2019-900046-x.

[32] "RSME - Root Mean Square Error in Python - Javatpoint." Accessed: Nov. 23, 2022. [Online]. Available: https://www.javatpoint.com/rsme-root-mean-square-error-in-python

[33] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature," *Geosci. Model Dev.*, vol. 7, no. 3, Art. no. 3, June 2014, doi: 10.5194/gmd-7-1247-2014.

[34] "The 'K' in K-fold Cross Validation." Accessed: Nov. 11, 2022. [Online]. Available: https://arpi.unipi.it/handle/11568/962587