Sustainable Processes Connect
ISSN: 3079-1421
2025, Article ID. 2025.0020, Cite as: https://doi.org/10.69709/SusProc.2025.120032

SCIFINITI PUBLISHING

Research Article

OPEN ACCESS

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

Article History
Received: May 22, 2025 Accepted: December 25, 2025 Published: December 29, 2025

Abstract
This study 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 trends in PM2.5, PM10, NO2, and CO concentrations in relation to temperature, humidity, wind patterns, and urban development indices. This study also discusses the impact of seasonal events, such as sandstorms and traffic emissions, on air quality. The novelty of this work lies in addressing the limitations of sparse sensor networks by employing a real-time hybrid machine learning model that accurately forecasts air quality in Abu Dhabi using only satellite and key meteorological data.

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

1. Introduction

The decline in air quality in urban centers remains one of the most pressing environmental health challenges, exacerbated by rapid industrialization and growth in vehicle ownership and traffic [1,2]. Cities such as Abu Dhabi, characterized by the coexistence of sea and desert environments, are particularly vulnerable due to the unique combination of desert dust, elevated vehicular emissions, and extreme climatic conditions [3]. These dynamic pollution sources create a need for 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 well suited for capturing nonlinear relationships, integrating 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 data fusion for urban air quality forecasting has been supported by 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 the decision to integrate CAMS/MODIS satellite signals with local meteorology and time encod-

* **Corresponding Author:**
Maria Iruj, Mechanical and Industrial Engineering Department, Abu Dhabi University, Abu Dhabi, United Arab Emirates, maria.iruj@gmail.com

ings, and to benchmark 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. A distributed IoT framework (AirSPEC) was proposed that integrates machine learning models to predict geospatial air quality using sensor networks. The framework employed sensor fusion and Kalman filtering for indoor environmental monitoring [13]. Alsaedi et al. [2] followed the same strategy of adopting satellite data for air quality predictions. A similar study was conducted for Europe [14].

Advanced neural architectures have enabled multimodal learning, combining land cover, weather, traffic, and remote sensing inputs [15,16]. A fusion model, AQNet, was developed to predict multiple pollutants using a combination of satellite and sensor data. A deep ensemble model was developed to forecast indoor PM2.5 using outdoor pollution correlations across 91 sensors, demonstrating strong predictive power and relevance to urban architecture [17].

Real-time air quality engines are now deployed in several urban areas to guide transportation, emergency response, and urban planning [18]. For example, DeepPlume is a high-resolution real-time engine for street-level pollution forecasting that uses official data sources, land use, and traffic estimates, and demonstrates how satellite-only models can be scaled globally to generate pollutant maps at 10-m resolution with high accuracy, providing a valuable solution for regions without ground sensors [19]. Some of the similar studies have been conducted by [20–22].

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 sufficiently fast for real-time applications, identifies lagged wind speed and temperature as the primary predictors.

2. 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, weekday/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 for training and testing. During training, hyperparameters were tuned using 5-fold time-series cross-validation [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 consumed a 7-day lookback window of multivariate features (duaod550, aod550, T2M, WS10M, RH2M, and time encodings) [29,30].

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

Evaluation metrics. RMSE [32], MAE, R^2 [33], SMAPE were computed, and K-fold cross-validated errors on the training folds were also reported to assess robustness [27,34]. Equations (i)–(iv) below were used to obtain 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. $SMAPE = (100/n)\ \Sigma (|y_i - \hat{y}_i| / ((|y_i| + |\hat{y}_i|)/2))$

3. 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.

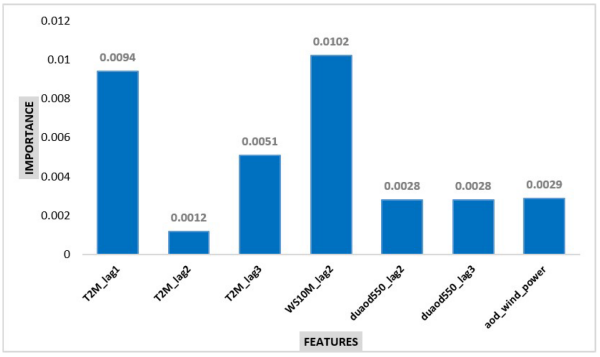

Figure 1: Feature importance scores for PM2.5 and PM10 prediction using the random forest model. The y-axis represents normalized feature importance values, while the x-axis lists input variables.

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

2025, Vol. 1, Article ID. 2025.0020
https://doi.org/10.69709/SusProc.2025.120032

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 individually leading, remain valuable as auxiliary predictors, especially for 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 model's overall predictive capacity. Consistent with the visual insights from the feature importance plot, WS10M_lag2 (10-m 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 established meteorological principles, which indicate that increased wind speeds enhance pollutant dispersion and reduce stagnation, thereby influencing surface-level PM concentrations.

T2M_lag1 (2-m air temperature, 1-day lag), with a 0.0094 score, is the second most dominant factor, suggesting a substantial influence of recent temperature conditions on air pollutant dynamics. Temperature fluctuations affect the atmospheric boundary layer and the chemical transformation rates of aerosols, with their influence being particularly pronounced in arid environments such as 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 be attributed to redundancy or collinearity with the adjacent temporal lags, indicating a nonlinear decay in temporal relevance across various temperature features.

The AOD-derived features, including duaod550_lag2 (0.0028), duaod550_lag3 (0.0028), and aod_wind_power (0.0029), all exhibit slightly lower but nearly comparable importance values. Although not dominant, their consistent presence among the top predictors underscores the auxiliary role of satellite-derived aerosol optical depth in enhancing spatial generalizability, particularly in scenarios with intermittent or sparse ground sensor coverage.

Interestingly, the combined importance of the AOD-based features ($\Sigma \approx 0.0085$) is comparable to that of a top meteorological feature. This suggests that while no individual satellite feature outperforms meteorological inputs, their combined contribution supports robust model performance.

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

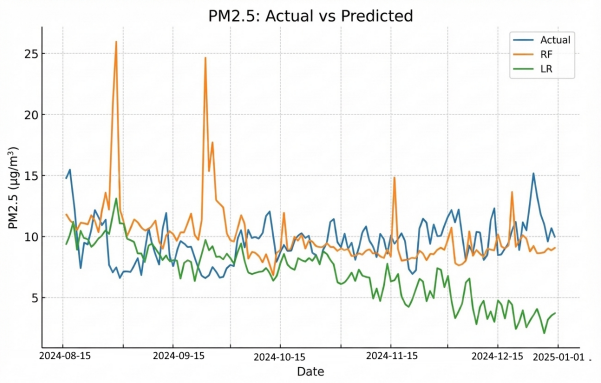

Figure 2: PM2.5 actual vs. predicted (RF and LR) on the test period. Y-axis: PM2.5 ($\mu g/m^3$); X-axis: date.

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

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

The Random Forest model demonstrates robust predictive stability, providing smooth forecasts that mitigate overfitting to noise while still capturing 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 alignment between the observed and predicted curves indicates that the selected features, including wind speed, temperature, and satellite-based aerosol indicators, provide sufficient explanatory power to represent a range of atmospheric conditions. This performance indicates that the identified feature-importance patterns, including meteorological lags and AOD-based variables, jointly support a robust forecasting framework for particulate-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 Tables 2 and 3:

Table 2: Random forest vs linear regression 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
$\mathbf{R^2}$ (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 fea- Tures	Without AOD
RMSE (Root Mean Squared Error)	5.59	6.37
MAE (Mean Absolute Error)	3.96	4.55
$\mathbf{R^2}$ (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 indicate that the model's forecasts are not only accurate but also closely track actual concentration levels, especially during moderate pollution conditions.

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

This study reports SMAPE values of 14.82 for Random Forest and 18.36 for Linear Regression, indicating moderate relative error in predicting PM2.5 concentrations. Although these values initially appear reasonable, their interpretation in an environmental context requires some thought, as with SMAPE, air quality data often exhibit nonlinear fluctuations and variability across 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 exceeding 20–30 $\mu g/m^3$ 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 complementary insights, it should not be considered in isolation and must be interpreted alongside absolute error measures for robust evaluation.

2025, Vol. 1, Article ID. 2025.0020
https://doi.org/10.69709/SusProc.2025.120032

4. Operational Implications

The model can support 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 are difficult to capture using lagged features alone, their effects are partially mitigated through 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: The pipeline will be extended to incorporate mobility proxies (e.g., traffic intensity) and land-use attributes (road density, built-up fraction), and validated in additional MENA cities (e.g., Dubai, Riyadh, Doha) to assess its transferability across diverse desert-urban regimes.

5. Discussion

The results confirm that meteorological features, particularly lagged wind speed at 10 m (WS10M_lag2, 0.0102) and temperature at 2 m with a 1-day lag (T2M_lag1, 0.0094), are the most influential 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 the importance of intermediate lags (e.g., T2M_lag2 at 0.0012) points to a non-linear temporal relevance decay, suggesting that the immediate past (lag-1) and the longer-term past (lag-3) may capture different but critical pollution-generating patterns.

A time-series comparison of actual vs. predicted PM2.5 values over 5 months shows that the Random Forest model can track 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 significant discrepancies during pollution spikes, especially in late November and December, when observed values deviate sharply from predicted values. These may be attributable to unmeasured events, such as dust storms or construction activities, that are either poorly represented in the input data or inherently difficult to capture with 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 quantitative evaluation metrics support the earlier findings, indicating that the model has strong predictive capability with an acceptable level of uncertainty, particularly for operational or early-warning systems in urban air quality management.

6. Conclusions

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

For Abu Dhabi, as in other urban areas with complex factors such as dust, heavy traffic, and extreme climate, this approach provides 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 including mobility data (traffic and 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 physical sensor networks. The novelty lies in the hybrid modeling approach that integrates satellite-derived Aerosol Optical Depth (AOD) and meteorological variables, cross-benchmarked against Random Forest and LSTM architectures, and demonstrates improved accuracy in data-sparse environments. Unlike previous global or generic urban studies, this work focuses on Abu Dhabi's unique desert dust, traffic emissions, 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. Feature-importance analysis indicates that lagged wind speed and temperature are the principal predictors, with AOD features augmenting the model. The study offers a sustainable, scalable, and transferable solution to operational air quality management.

List of Abbreviations

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

Author Contributions

Conceptualization, supervision: M.I.; Data curation, resources: Z.I.; Methodology, software, formal analysis, visualization: D.M.K.; Writing—original draft preparation: S.Y. All authors have read and agreed to the published version of the manuscript.

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)

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

The study did not receive any external funding and was conducted using only institutional resources.

Acknowledgments

All intellectual content, analysis, and conclusions remain solely the responsibility of the authors.

AI Declaration

The authors acknowledge using ChatGPT (developed by OpenAI) as an AI tool to support language refinement, grammar correction, and structural suggestions during manuscript preparation.

References

[1] Martínez-España, R.; Bueno-Crespo, A.; Timon, I.; Soto, J.; Muñoz, A.; Cecilia, J. Air-Pollution Prediction in Smart Cities through Machine Learning Methods: A Case of Study in Murcia, Spain. *J. Univers. Comput. Sci.* **2018**, *24*, 261–276. [CrossRef]

[2] Alsaedi, A.S.; Liyakathunisa, L. Spatial and Temporal Data Analysis with Deep Learning for Air Quality Prediction. In *Proceedings of the 2019 12th International Conference on Developments in eSystems Engineering (DeSE)*; IEEE: New York, NY, USA, 2019; pp. 581–587. [CrossRef]

[3] da Silva Júnior, F.M.R.; de Moura, F.R.; de Lima Brum, R.; Tavella, R.A. Air Pollution—A Look beyond Big Cities. *Integr. Environ. Assess. Manag.* **2023**, *19*, 295–306. [CrossRef] [PubMed]

[4] Sharma, V.; Ghosh, S.; Mishra, V.; Kumar, P. Spatio-Temporal Variations and Forecast of PM2.5 Concentration around Selected Satellite Cities of Delhi, India Using ARIMA Model. *Phys. Chem. Earth Parts ABC* **2024**, *138*, 103849. [CrossRef]

[5] Mujtaba, M.; Munir, M.A.; Ali, S.; Petrů, J.; Ansar, T.; Akhlaq, W.; Ahmad, M.; Iqbal, H.; Ali, F.; Bashir, M.N.; et al. Using Machine Learning for Air Quality Prediction and Sustainable Urban Planning. *Sustain. Futur.* **2025**, *10*, 100981. [CrossRef]

[6] Ahmad, O.; Khalid, Z.; Tahir, M.; Uppal, M. Spatiotemporal Air Quality Mapping in Urban Areas Using Sparse Sensor Data, Satellite Imagery, Meteorological Factors, and Spatial Features. *arXiv* **2025**, arXiv:2501.11270. [CrossRef]

[7] Kang, G.K.; Gao, J.Z.; Chiao, S.; Lu, S.; Xie, G. Air Quality Prediction: Big Data and Machine Learning Approaches. *Int. J. Environ. Sci. Dev.* **2018**, *9*, 8–16. [CrossRef]

[8] Iskandaryan, D.; Ramos, F.; Trilles, S. Air Quality Prediction in Smart Cities Using Machine Learning Technologies Based on Sensor Data: A Review. *Appl. Sci.* **2020**, *10*, 2401. [CrossRef]

[9] Mahalingam, U.; Elangovan, K.; Dobhal, H.; Valliappa, C.; Shrestha, S.; Kedam, G. A Machine Learning Model for Air Quality Prediction for Smart Cities. In *Proceedings of the 2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*; IEEE: New York, NY, USA, 2019; pp. 452–457. [CrossRef]

[10] Ramezan, C.A.; Warner, T.A.; Maxwell, A.E. Evaluation of Sampling and Cross-Validation Tuning Strategies for Regional-Scale Machine Learning Classification. *Remote Sens.* **2019**, *11*, 185. [CrossRef]

[11] Shiri, F.M.; Perumal, T.; Mustapha, N.; Mohamed, R. A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU. *arXiv* **2023**, arXiv:2305.17473v4. [CrossRef]

[12] Landi, F.; Baraldi, L.; Cornia, M.; Cucchiara, R. Working Memory Connections for LSTM. *Neural Netw.* **2021**, *144*, 334–341. [CrossRef] [PubMed]

[13] Ayek, A.A.E.; Loho, M.A.; Karmoka, S.F. Satellite Air Quality Monitoring: An Interactive and Accessible Tool for Spatiotemporal Analysis Using Google Earth Engine. *Geo-Spat. Inf. Sci.* **2025**, *28*, 1–15. [CrossRef]

[14] Kalogiannidis, S.; Spinthiropoulos, K.; Kalfas, D.; Chatzitheodoridis, F.; Tziampazi, F. Integration of Remote Sensing and GIS for Urban Sprawl Monitoring in European Cities. *Eur. J. Geogr.* **2025**, *16*, 75–90. [View Online]

[15] Ren, P.; Xiao, Y.; Chang, X.; Huang, P.-Y.; Li, Z.; Chen, X.; Wang, X. A Comprehensive Survey of Neural Architecture Search: Challenges and Solutions. *ACM Comput. Surv.* **2021**, *54*, 76:1–76:34. [CrossRef]

[16] Abdullah, S.M. To Predict Air Pollution Using Machine Learning and ARIMA Model. *Int. J. Eng. Res. Technol.* **2021**, *10*, 574–579. [View Online]

[17] Phalguna Krishna, E.S. IoT-Enabled Wireless Sensor Networks and Geospatial Technology for Urban Infrastructure Management. *J. Electr. Syst.* **2024**, *20*, 2248–2256. [CrossRef]

[18] Schulte, K. 'Real-Time' Air Quality Channels: A Technology Review of Emerging Environmental Alert Systems. *Big Data Soc.* **2022**, *9*, 1–14. [CrossRef]

[19] Jauvion, G.; Cassard, T.; Quennehen, B.; Lissmyr, D. DeepPlume: Very High Resolution Real-Time Air Quality Mapping. *arXiv* **2020**, arXiv:2002.10394. [CrossRef]

[20] Feng, L.; Yang, T.; Wang, D.; Wang, Z.; Pan, Y.; Matsui, I.; Chen, Y.; Xin, J.; Huang, H. Identify the Contribution of Elevated Industrial Plume to Ground Air Quality by Optical and Machine Learning Methods. *Environ. Res. Commun.* **2020**, *2*, 055001. [CrossRef]

[21] Marto, J.P.; Zhang, J.; Schwab, J.J. Plume Analysis from Field Evaluations of a Portable Air Quality Monitoring System. *J. Air Waste Manag. Assoc.* **2021**, *71*, 431–445. [CrossRef]

[22] Govea, J.; Gaibor-Naranjo, W.; Sanchez-Viteri, S.; Villegas-Ch, W. Integration of Data and Predictive Models for the Evaluation of Air Quality and Noise in Urban Environments. *Sensors* **2024**, *24*, 311. [CrossRef]

[23] Masuoka, E.; Fleig, A.; Wolfe, R.E.; Patt, F. Key Characteristics of MODIS Data Products. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1313–1323. [CrossRef]

[24] Wang, P.; Tang, Q.; Zhu, Y.; He, Y.; Yu, Q.; Liang, T.; Zheng, K. Spatial-Temporal Variation of AOD Based on MAIAC AOD in East Asia from 2011 to 2020. *Atmosphere* **2022**, *13*, 1983. [CrossRef]

[25] Li, B.; Peng, X.; Wang, Z.; Xu, J.; Feng, D. AOD-Net: All-In-One Dehazing Network. In *Proceedings of the IEEE International Conference on Computer Vision*; 2017; pp. 4770–4778. Available online: https://openaccess.thecvf.com/content_iccv_2017/html/Li_AOD-Net_All-In-One_Dehazing_ICCV_2017_paper.html (accessed on 9 October 2025).

[26] Kök, İ.; Şimşek, M.U.; Özdemir, S. A Deep Learning Model for Air Quality Prediction in Smart Cities. In *Proceedings of the 2017 IEEE International Conference on Big Data (Big Data)*; IEEE: New York, NY, USA, 2017; pp. 1983–1990. [CrossRef]

[27] Nti, I.; Nyarko-Boateng, O.; Aning, J. Performance of Machine Learning Algorithms with Different K Values in K-Fold Cross-Validation. *Int. J. Inf. Technol. Comput. Sci.* **2021**, *13*, 61–71. [CrossRef]

[28] Ismail, A.A.M.; Ali, N.; Amirul, M.S.; Endut, R.; Aljunid, S.A. Prediction Model for Spectroscopy Using Python Programming. *Int. J. Nanoelectron. Mater.* **2021**, *14*, 353–363. [View Online]

[29] Hofman, J.; Castanheiro, A.; Nuyts, G.; Joosen, S.; Spassov, S.; Blust, R.; De Wael, K.; Lenaerts, S.; Samson, R. 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.* **2020**, *712*, 135534. [CrossRef]

[30] Babu Saheer, L.; Bhasy, A.; Maktabdar, M.; Zarrin, J. Data-Driven Framework for Understanding and Predicting Air Quality in Urban Areas. *Front. Big Data* **2022**, *5*, 822573. [CrossRef] [PubMed]

[31] Smagulova, K.; James, A.P. A Survey on LSTM Memristive Neural Network Architectures and Applications. *Eur. Phys. J. Spec. Top.* **2019**, *228*, 2313–2324. [CrossRef]

[32] Sun, K.; Zhou, R.; Kim, J.; Hu, Y. PyGRF: An Improved Python Geographical Random Forest Model and Case Studies in Public Health and Natural Disasters. *Trans. GIS* **2024**, *28*, 2476–2491. [CrossRef]

[33] Chai, T.; Draxler, R.R. Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)?—Arguments against Avoiding RMSE in the Literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [CrossRef]

[34] The 'K' in K-Fold Cross Validation. CINECA. Available online: https://arpi.unipi.it/handle/11568/962587 (accessed on 11 November 2022).