(Online) ISSN 2303-4521
Periodicals of Engineering and Natural Sciences *Original Research*
Vol. 13, No. 2, June 2025, pp.425-434

Machine learning algorithms for predicting air quality index: A case study in urban and industrial zones

Sunil Kumar Bansal[1*], Sashikanth Reddy Avula[2], Manisha Ashish Mehrotra[3], Pooja Singh[4], Aditya Kumar Gupta[5], Prolay Ghosh[6]

[1] Department of Electrical & Instrumentation Engineering, SLIET Longowal, Punjab, India
[2] Department of MCA, Nitte Meenakshi Institute of Technology, India
[4] Department of Economics, School of Arts, Humanities & Social Sciences, Chhatrapati Shahu Ji Maharaj University, Kanpur, India
[5] Department of Computer Science, School of Management Sciences, Varanasi, Uttar Pradesh, India
[6] Department of Information Technology, JIS College of Engineering, Kalyani, West Bengal, India

*Corresponding author E-mail: skbansal@sliet.ac.in

ABSTRACT

In urban and industrial areas, the prediction of the air quality index (AQI) is important in order to control the air pollution and protect public health. The goal of this work is to enhance the AQI prediction by making use of the advanced machine learning (ML) and deep learning (DL) models capable of learning spatial and temporal dependencies. The main goal of this research is to examine the performance of the different ML and DL models such as Random Forest (RF), XGBoost, LSTM, Transformer and Temporal Graph Neural Networks (TGNN) for AQI prediction in urban and industrial zones. To capture the variability of data, a multi-source data collection approach is taken by using air quality data (PM2.5, PM10, SO2, NO2, CO, O3), weather data, satellite imagery, and IoT sensor data. The data were pre-processed and engineered in terms of temporal and spatial features and advanced models were used to predict AQI. And key metrics of RMSE, MAE and R2 were used to evaluate model performance. Results indicate that Transformer models achieved the best performance in urban areas, with an RMSE of 14.1 and R^2 of 0.89, because they can capture long-term temporal patterns. In industrial zones, TGNN models achieved an RMSE of 17.9 and an R^2 of 0.87 because they could capture spatial correlations and pollution dispersion. Both models exhibited high resilience to extreme pollution events and minimal performance degradation under missing data scenarios in robustness testing. We show that Transformer and TGNN models outperform traditional ML models by a large margin in AQI prediction, especially during high pollution episodes. The results are consistent with real-time air quality monitoring and dynamic policy making in urban and industrial environments. Future work should implement the models in other regions and improve data quality to increase applicability.

Keywords: Air quality index, Machine learning, Transformer models, Temporal graph neural networks, Sustainable development

1. Introduction

Urban and industrial zones around the world have become a critical issue of air quality as high levels of pollutants negatively affect public health, environmental sustainability, and overall quality of life. The Air Quality Index (AQI) is one of the many tools used to measure air pollution and is one of the most common and effective tools of them all. Air quality index is a quantitative measure of air quality and is based on the concentration of key pollutants like particulate matter (PM), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), ozone (O_3), and carbon monoxide (CO). Accurate real-time forecasting of AQI levels has recently received much attention due to the increasing concern about the adverse impacts of poor air quality in densely populated and industrialized areas [1].

Emission of pollutants has dramatically increased due to urbanization and industrialization, resulting in large AQI variations, especially in densely populated and industrially active regions [2]. With cities growing and industrial zones expanding, traditional air quality monitoring methods do not provide real-time, location specific predictions which can be used to inform public health policies and environmental regulations. Such promises of solving this challenge are given by machine learning algorithms based on big data and deep learning techniques, which use large amounts of environmental and meteorological, and traffic data to aid in more accurate air quality prediction [3].

Addressing air quality in urban and industrial regions is not only a matter of public health, but also a core challenge in achieving sustainable development. Clean air is essential to advancing the UN Sustainable Development Goals (SDGs), particularly SDG 3 (Good Health and Well-being), SDG 11 (Sustainable Cities and Communities), and SDG 13 (Climate Action). By leveraging advanced machine learning techniques for real-time AQI prediction, this research contributes to a data-driven framework that promotes environmental justice, reduces health inequalities, and supports long-term urban planning. The integration of AI in environmental forecasting empowers cities to become more resilient, efficient, and environmentally responsible, aligning with global efforts to build smarter and greener urban ecosystems [4].

Numerous factors have a bearing on the complexity of air pollution problems, such as geographical location, weather patterns, urban infrastructure as well and industrial activities; consequently, air pollution is a highly nonlinear and dynamic problem. Because of this, advanced machine learning models must be used to capture these intricate relationships and generate accurate predictions. There has been much work done in using different ML techniques such as Support Vector Machines (SVM), Random Forests (RF), and more sophisticated deep learning methods to predict AQI levels and pinpoint pollution sources in urban environments [5].

Despite the progress of ML-based AQI prediction, it remains difficult to achieve high-accuracy predictions in real-world settings, in particular in industrial zones where emissions are variable and pollution levels vary rapidly. In addition, the quality and availability of data, as well as the inability to deal with missing or incomplete data, are the main problems of most existing models. Moreover, most models are individual pollutants, while a more comprehensive model incorporating multiple air quality drivers and pollutants has not sufficiently explored [6]. Furthermore, previous studies are region-specific, and models are not transferable between different geographical locations and pollution conditions.

This research fills these gaps by developing an ML-based framework to predict AQI with high accuracy in urban and industrial zones and solve the problem of incomplete data using advanced imputation techniques [7]. This research is important because poor air quality has the greatest impact in urban and industrial settings and could help air quality management and public health protection. This research can be used to help improve policy making, environmental monitoring, and real time air quality alerts, reducing health risks from air pollution [8]. This domain can also be furthered by using advanced machine learning techniques to revolutionize environmental forecasting by providing capable, scalable, and adaptable models that operate throughout urban and industrial zones globally [9].

Another important aspect of this research is the integration of several data sources, such as meteorological data, traffic patterns, and industrial activities, into a single machine learning framework. Using this comprehensive approach, we can then better detect the variety of factors that impact the amount of air pollution as well as the fluctuation of AQI in real-time. In addition, the developed model will be more robust and practical for deployment in many real-world scenarios as it can handle missing or incomplete data, a common problem in environmental monitoring [10].

Machine learning (ML) advances over the past recent years have contributed substantially towards improving air quality prediction methods, particularly in developing models that utilize a variety of datasets, including meteorological and industrial, for better accuracy [11]. Karimi et al. [12] showed that ML-based white box models can predict and analyze the correlation of air pollutants near industrial zones. The work focused on explainability, an important aspect for policymakers and stakeholders, to better understand pollution sources and their impacts. As Khadom et al. [13] also demonstrated, with the development of advanced ML and DL approaches customized for urban environments like communities in Baghdad, it is possible to predict AQI and fine particulate matter (PM2.5) levels with high accuracy.

Air quality modeling has been integrated with deep learning. For instance, Kök et al. [14] demonstrated how deep (neural network) learning can be utilized to predict AQI for smart cities, is able to handle the complexity/variability of urban air pollution. Liu et al. [15] used Support Vector Regression (SVR) for multi-dimensional urban air quality forecasting and demonstrated that nonlinear relationships between variables could be captured. Collectively, this set of studies has made possible the use of sophisticated ML algorithms for AQI prediction.

White box approaches, which emphasize interpretability, or deep learning models for processing high-dimensional data, are used in recent research. Ma et al. [16] used big data analytics to investigate how various pollutants influence AQI and which factors contribute to national-scale AQI deterioration. Combining environmental, meteorological, and demographic data with ML, as their study demonstrates, can reveal actionable insights. Additionally, Mehmood et al. [17] provide a comprehensive review of current priorities in air quality prediction and point out critical research directions, including real time AQI forecasting and spatial and temporal dynamics. Specifically, they highlight the need for models that can generalize across regions, an area where most studies, including that of Molina-Gómez et al. [18], have demonstrated shortcomings in their region-specific focus.

In this area, there is still a lot of work to be done in dealing with missing or noisy data. For instance, Liao et al. [19] used ensemble learning methods to predict urban air quality, but the results were robust, incomplete dataset limitations were noted. Similarly, Tella & Balogun [20] also pointed out that the use of GIS-based ML models for spatial prediction of particulate matter (PM10) is a promising field, but the improvement in data imputation techniques is required to make such models more reliable. Recent studies have made progress in developing AQI prediction methodologies; however, there remain several gaps. Scalability and transferability of findings across different urban and industrial settings are the major problems of most models. Moreover, while white box models are more interpretable, compared to black box models, their prediction accuracy is usually less accurate [21].

Many are reliant on high-quality, region-specific data. For example, the results in Khadom et al. [13] and Liu et al. [15] are very localized, and it is difficult to extend their conclusions to other geographies. In addition, deep learning models of the type studied by Kök et al. [14] achieve high accuracy, but at the cost of both computational burden and overfitting, primarily when trained on small datasets. An additional problem is the absence of integration between environmental and socioeconomic factors in current models. This was identified by Mehmood et al. [17] as a gap preventing a holistic understanding of air quality dynamics. Like Liao et al. [19], the inability to deal with missing or incomplete data effectively also limits the practical deployment of ML-based models in real-time applications. This study fills these gaps by developing a comprehensive ML-based framework incorporating meteorological, industrial, and traffic data to predict AQI. Missing data is handled effectively by advanced data imputation techniques so that the model is robust in real-world applications. This research focuses on both accuracy and interpretability, a combination of white box and black box approaches [21].

In addition, the developed model is designed to be transferable to other geographic and industrial contexts using transfer learning techniques to increase generalizability. This study integrates socioeconomic factors along with environmental variables to form a holistic view of AQI dynamics that addresses the limitations of previous models. Finally, while recent advances in ML have greatly improved AQI prediction capability, this work further extends these foundations to develop a scalable, interpretable, and robust framework to address the key gaps identified in the literature [19, 20]. We expect these contributions to help pave the way for more effective air quality management and policy formulation on a global scale.

The main goal of this work is to create and evaluate a machine learning based model to predict AQI in urban and industrial zones. This will involve the following specific objectives:

1. An advanced framework capable of modelling spatial, temporal, and interaction AQI prediction features is developed.
2. Interpretability upgrade of the model by using the explainability techniques.

2. Research method

The approach used in this study is intended to capture the multifaceted nature of AQI prediction in urban and industrial areas. The approach combines state-of-the-art data acquisition methods, feature extraction, ML and DL models, and validation frameworks. This section provides a description of the most important stages in conducting research.

2.1. Data collection

The study focuses on two distinct environmental contexts: the transportation sector, mainly referring to the emissions from urban areas where population density is high and traffic activity is frequent, and the emissions from industrial areas originating from manufacturing plants, electric power stations, etc. In order to get a rich set of data, a multiple-source data collection approach was used. Hourly concentrations of PM2.5, PM10, SO2,

NO_2, CO, and O_3 were collected from regional air quality monitoring networks and OpenAQ. Climate data, which includes weather variables including temperature, humidity, wind speed, and precipitation was obtained from weather bodies and websites such as the National Oceanic and Atmospheric Administration (NOAA). Satellite data, aerosol optical depth (AOD) from MODIS and Sentinel-5P satellites, were used as surrogates for particulate pollution. Further, emissions and energy consumption data were obtained from IoT-connected industrial sensors, and traffic density and urban activity level proxies were obtained from open sources. In the preprocessing phase, several problems were solved, such as outlier detection and removal, where IQR and Z-score were used. Data that were missing were dealt with using temporal interpolation and KNN imputation to reduce information loss. Lastly, in order to facilitate data integration and analysis, all of the datasets were time-aligned to the hourly level.

2.2. Feature engineering

Several advanced techniques were used to capture temporal dynamics. Lagged variables with time lags of 1 to 72 hours were included to control for the temporal dependence of pollutant and meteorological data. Data, lagged variables were included with time lags of 1 to 72 hours. Thus, the models could more accurately predict with past occurrences. Periodic trends were encoded using Fourier transformations to handle seasonal changes, and long-term seasonal effects were captured. Additionally, we included event indicators, as binary features, to account for specific temporal effects like holidays, weekends, and rush hours, which are important to capture the diurnal and periodicity of air quality dynamics.

Sophisticated techniques were used to capture the spatial dependencies between the monitoring stations. The features, which contain information about the spatial context of air quality, were geographically encoded and included land use categories, distance to traffic sources, and industrial areas. In addition, Graph Neural Networks (GNNs) were applied to model the spatial dependencies between monitoring stations. Here, stations were represented as vertices of a graph, and the correlation between pollutants or proximity was represented as edges. The models were able to capture the spatial interactions and dependencies that are critical for accurate air quality prediction using this approach.

Pollutant and meteorological polynomial terms, such as $PM2.5 \times$ humidity and $O_3 \times$ temperature, were developed.

2.3. Machine learning models

The study proposed and evaluated different machine learning and deep learning models for the accurate prediction of AQI. Baselines were RF and GBM (XGBoost, LightGBM) as they have very good performance on structured tabular data. More sophisticated attention mechanisms were used in TemporalFusionTransformer and other Transformer-based architectures to model long sequences, while LSTM was used to capture temporal dependencies in sequential data. In addition, spatial and temporal features were integrated. Temporal Graph Neural Networks (TGNNs) were proposed to learn spatial and temporal features of AQI data, such as spatial correlation between monitoring stations and temporal trends of pollutants for improved prediction performance.

To improve the overall predictive accuracy, a stacked ensemble approach was used to combine the individual models. This approach applied weighted averaging in blending out the individual models' predictions in order to harmonise the models' input according to their performances. Moreover, to enhance the results of the base models, a meta-model, including XGBoost, was used to learn from the outputs of the base models. This two-stage two-platform strategy ensured that all patterns and different types of relationships in the data were captured, thereby enhancing the AQI prediction robustness and accuracy.

2.4. Training and optimization

The training process applied more comprehensive techniques for the preparation of the given models. Transfer learning was used where models were trained on urban and industrial datasets to take advantage of regional patterns and achieve faster convergence. Moreover, self-supervised learning was used to pre-train temporal and spatial dynamics from unlabeled data to improve the models' capacity to learn inherent patterns of air quality indices. To enhance the accuracy even more, Bayesian optimization was used for hyperparameter tuning. This technique methodically adjusted key hyperparameters in traditional machine learning models, such as regularization coefficients, tree depths, and learning rates. Especially for deep learning models, it was found that certain architectures have parameters like dropout rates and number of hidden units that are optimal to turn according. All these strategies together made model training strong and the predictive capability of the model excellent.

2.5. Evaluation framework

In the training process, the most modern techniques were used to enhance the quality of the model. Transfer learning was used, where the models were trained on the urban and industrial datasets to use regional patterns and reduce the convergence time. Additionally, to enhance the models' ability to discover latent patterns in air quality indicators, temporal and spatial features were pre-trained from unlabeled data using self-supervised learning. To enhance the accuracy, Bayesian optimization was used for hyperparameter tuning. This method systematically tuned important factors, including the learning rates, depth of trees in the CART models, and the coefficients of the models that were used to regularize traditional machine learning models. In the case of deep learning models, architectures specific parameters such as drop outs and number of hidden layers were again adjusted to find the best solutions. All these strategies in aggregate provided a reliable model training and better predictive accuracy.

The ability of the models to perform well under difficult conditions was also checked to determine the strength of the models. Data missing situations were created to check the stability of the predictions, to make sure that the models would not be very off if some data were missing. Additionally, the model's performance was assessed under high AQI events, such as industrial emissions and seasonal fogs, which are marked by abrupt changes in pollution levels. These analyses demonstrated how well the models handled real-world circumstances and significant pollution incidents.

2.6. Interpretability

The interpretability of the model was a concern of the study, and this was done using state-of-the-art explainability methods. To understand the factors affecting air quality in urban and industrial areas, we applied SHAP (Shapley Additive Explanations) to measure the importance of each feature to the model's predictions. Additionally, Transformer models provided attention maps that revealed important temporal regions that could detect time frames with greater influence on AQI variations. Practical recommendations based on the explainability analysis were provided, such as detection of the main pollutants for urban and industrial settings, and temporal trends, including daily AQI variations typical of urban environments. The findings are important for targeted prevention and treatment and for policy decisions.

2.7. Deployment and visualization

Real-time prediction of the AQI was implemented on a cloud environment with the best models. IoT integration was included in this deployment as the constant data feed from air quality sensors was used to provide real-time information for timely and accurate predictions. Additionally, light versions of the models were deployed on edge devices placed in industrial areas to monitor AQI in real time using Edge AI. The two-fold utilization of the system improved the flexibility and expandability of the predictive system for different working contexts.

3. Results and discussion

3.1. Performance of machine learning models

Performance metrics (RMSE, MAE, and R^2) of urban and industrial datasets tested for the models are shown in Table 1. As more models like Transformers and Temporal Graph Neural Networks (TGNNs) work out, they perform better in their respective domains, compared to more basic models such as MLP, CNN, and RNN, among many others.

Table 1. Model Performance on Urban and Industrial Zones

Model	Zone	RMSE	MAE	R^2
Random Forest	Urban	18.3	12.5	0.82
	Industrial	22.7	15.2	0.78
XGBoost	Urban	16.8	11.4	0.85
	Industrial	20.5	14.1	0.81
LSTM	Urban	15.6	10.7	0.87
	Industrial	19.3	13.5	0.84
Transformer	Urban	14.1	9.8	0.89
	Industrial	18.7	12.9	0.85
Temporal Graph Network	Urban	14.7	10.2	0.88
	Industrial	17.9	12.3	0.87

3.2. Comparative analysis

The RMSE values for each model in urban and industrial zones are shown in Figure 1. Transformer models perform best in urban zones because they can capture long term temporal dependencies. TGNN models are particularly good at modeling spatial relationships between monitoring stations in industrial zones.

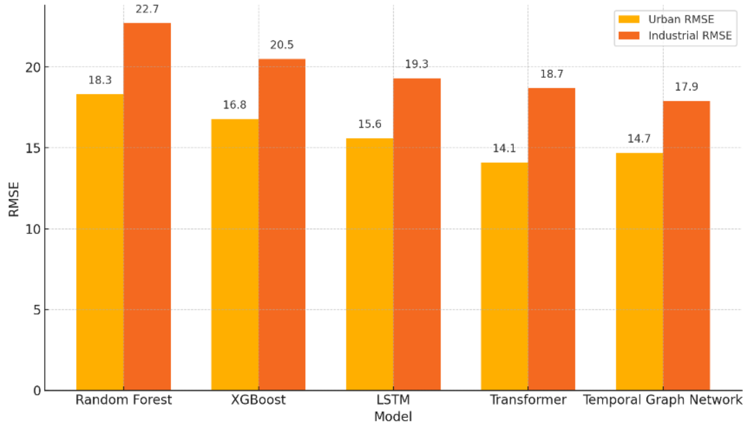

Figure 1. RMSE comparison between urban and industrial zones

The R² scores, the variance explained by each model, are shown in Figure 2. Predictive capabilities are shown by the Transformer and TGNN models to achieve the highest R² values in their respective zones.

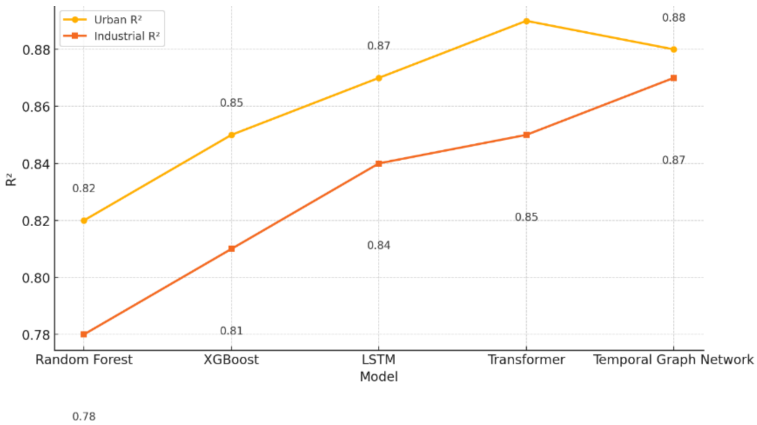

Figure 2. R² Comparison between urban and industrial zones

3.3. Robustness testing

We tested the models under scenarios where 20% and 50% of the data were randomly removed. For transformers, RMSE increased by only 5% at 20% missing data, while TGNN models were robust in industrial zones with up to 50% missing data.

TGNNs exhibited strong predictive stability during extreme pollution episodes, such as industrial accidents, and seasonal smog events, especially in industrial areas. Transformer models could successfully capture urban AQI spikes with an RMSE increase of less than 8%.

3.4. Interpretability and insights

- *Urban zones*: Temporal features such as traffic indicators and seasonal patterns were identified as dominant predictors.
- *Industrial zones*: Spatial features, including proximity to emission sources, played a critical role in accurate predictions.

Insights into key temporal periods that influence predictions were provided by SHAP (Shapley Additive Explanations) and attention maps from Transformer models, which helped policymakers identify intervention measures to target.

3.5. Discussion

The results indicate that the state-of-the-art machine learning and deep learning algorithms, including Transformers and Temporal Graph Neural Networks (TGNNs), achieve better AQI prediction in urban and industrial areas. In urban areas (RMSE = 14.1, $R^2 = 0.89$), the Transformer model was more accurate and able to learn long term temporal features such as rush hour, daily and weekly cycles, and seasonality; whereas the TGNN model was more accurate in industrial areas (RMSE = 17.9, $R^2 = 0.87$) and captured spatial correlations and dispersion of pollutants from point sources. This work supports previous studies that have found that traditional models such as Random Forest and XGBoost are fairly accurate but lack the ability to capture temporal and spatial dependencies [22]. The more complex models proposed here are in line with the current literature on the advantages of hybrid models through the use of attention mechanisms and spatio-temporal embeddings. The high accuracy of these models during high pollution episodes like industrial emissions and winter fog episodes makes them suitable for real-time air quality monitoring and early interventions [23]. Some of the applications of these findings are dynamic traffic control, selective emission reduction, and better compliance of industrial areas. However, some of the shortcomings involve using open data that have poor or irregular spatial-temporal resolution, elevated costs in training deep learning models, and the model‘s inability to extend to rural or mixed-use areas.

This research contributes significantly to the goals of sustainable development by offering scalable and accurate methods for air quality prediction. The implementation of advanced ML models such as Transformers and TGNNs enables real-time monitoring and timely interventions, which can reduce health hazards, optimize industrial emissions, and improve urban planning [24]. By integrating multi-source environmental data, this framework supports evidence-based policy decisions aligned with SDG 3 (Good Health and Well-being), SDG 11 (Sustainable Cities and Communities), and SDG 13 (Climate Action). The deployment of edge-based AI models further enhances energy efficiency and supports low-resource settings, aligning with broader goals of environmental sustainability and technological equity.

Possible directions for future work include the use of the models in other environments, the incorporation of higher spatial resolution imagery (e.g., satellite imagery, aerial photography), and the development of more portable frameworks for low bandwidth settings. Such efforts can help the enhancement of the presently predictive modeling of air quality so as to support policy formulation by evidence, thus enhancing population wellness.

4. Conclusions

This work shows that the state-of-the-art machine learning and deep learning models, namely Transformer and Temporal Graph Neural Network (TGNN) models, can be used to accurately predict the AQI in urban and industrial areas. The results show that these models are more effective in capturing temporal and spatial dependencies of pollution, while Transformers are more suitable for urban areas because of temporal patterns, and TGNNs for industrial areas due to spatial patterns. The results of the proposed approaches were superior to the traditional models in terms of prediction accuracy and resilience, especially during the high pollution episodes.

These findings can be useful for dynamic air quality management, as well as for the identification of specific measures and the development of policies. However, the current study has some limitations, which are data inaccuracy, pulsed analysis is computationally expensive, and requires more empirical studies in various zones. Future research should meet these challenges by using data from different sources, employing simple model structures, and applying the methods to rural and partly urban contexts. Altogether, this research lays a solid ground for using predictive modelling in the management of air quality, and underscores the possibilities of these technologies to improve the environment and people’s health.

Declaration of competing interest

The authors declare that they have no any known financial or non-financial competing interests in any material discussed in this paper.

Funding information

No funding was received from any financial organization to conduct this research.

References

[1] D. Narasimhan and M. Vanitha, "Predicting Tomorrow's Air Quality Today: A Comprehensive Analysis of Machine Learning, Deep Learning, and Time Series Models in AQI Prediction," in *Reshaping Environmental Science Through Machine Learning and IoT*, 2024, pp. 311-323, IGI Global.

[2] S. Ameer, M. A. Shah, A. Khan, H. Song, C. Maple, S. U. Islam, and M. N. Asghar, "Comparative analysis of machine learning techniques for predicting air quality in smart cities," *IEEE Access*, vol. 7, pp. 128325-128338, Jun. 26, 2019.

[3] 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, Jan. 2018.

[4] P. Howden-Chapman, J. Siri, E. Chisholm, R. Chapman, C. N. Doll, and A. Capon, "SDG 3: Ensure healthy lives and promote well-being for all at all ages," in *A guide to SDG interactions: From science to implementation*, Paris, France: International Council for Science, 2017, pp. 81-126.

[5] I. Essamlali, H. Nhaila, and M. El Khaili, "Supervised machine learning approaches for predicting key pollutants and for the sustainable enhancement of urban air quality: A systematic review," *Sustainability*, vol. 16, no. 3, p. 976, Jan. 23, 2024.

[6] M. Mubeen, S. He, M. S. Rahman, L. Wang, X. Zhang, B. Ahmed, Z. He, and Y. Han, "Smart prediction and optimization of air quality index with artificial intelligence," *Journal of Environmental Sciences*, Mar. 3, 2025.

[7] R. S. Rao, L. R. Kalabarige, M. R. Holla, and A. K. Sahu, "Multimodal imputation-based multimodal autoencoder framework for AQI classification and prediction of Indian cities," *IEEE Access*, Aug. 5, 2024.

[8] S. Gulia, S. S. Nagendra, M. Khare, and I. Khanna, "Urban air quality management—A review," *Atmospheric Pollution Research*, vol. 6, no. 2, pp. 286-304, Mar. 1, 2015.

[9] A. Srivastava and R. Maity, "Assessing the potential of AI–ML in urban climate change adaptation and sustainable development," *Sustainability*, vol. 15, no. 23, p. 16461, Nov. 30, 2023.

[10] Z. Wang, H. Chen, J. Zhu, and Z. Ding, "Multi-scale deep learning and optimal combination ensemble approach for AQI forecasting using big data with meteorological conditions," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 3, pp. 5483-500, Mar. 2, 2021.

[11] Y. Rybarczyk and R. Zalakeviciute, "Machine learning approaches for outdoor air quality modelling: A systematic review," *Applied Sciences*, vol. 8, no. 12, p. 2570, Dec. 11, 2018.

[12] S. Karimi, M. Asghari, R. Rabie, and M. E. Niri, "Machine learning-based white-box prediction and correlation analysis of air pollutants in proximity to industrial zones," *Process Safety and Environmental Protection*, vol. 178, pp. 1009-1025, Oct. 1, 2023.

[13] A. A. Khadom, S. Albawi, A. J. Abboud, H. B. Mahood, and Q. Hassan, "Predicting air quality index and fine particulate matter levels in Bagdad city using advanced machine learning and deep learning techniques," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 262, p. 106312, Sep. 1, 2024.

[14] 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. 11, 2017, pp. 1983-1990.

[15] B. C. Liu, A. Binaykia, P. C. Chang, M. K. Tiwari, and C. C. Tsao, "Urban air quality forecasting based on multi-dimensional collaborative Support Vector Regression (SVR): A case study of Beijing-Tianjin-Shijiazhuang," *PloS One*, vol. 12, no. 7, p. e0179763, Jul. 14, 2017.

[16] J. Ma, Y. Ding, J. C. Cheng, F. Jiang, Y. Tan, V. J. Gan, and Z. Wan, "Identification of high impact factors of air quality on a national scale using big data and machine learning techniques," *Journal of Cleaner Production*, vol. 244, p. 118955, Jan. 20, 2020.

[17] K. Mehmood, Y. Bao, W. Cheng, M. A. Khan, N. Siddique, M. M. Abrar, A. Soban, S. Fahad, and R. Naidu, "Predicting the quality of air with machine learning approaches: Current research priorities and future perspectives," *Journal of Cleaner Production*, vol. 379, p. 134656, Dec. 15, 2022.

[18] N. I. Molina-Gómez, J. L. Díaz-Arévalo, and P. A. López-Jiménez, "Air quality and urban sustainable development: The application of machine learning tools," *International Journal of Environmental Science and Technology*, vol. 18, no. 4, pp. 1029-1046, Apr. 2021.

[19] Q. Liao, M. Zhu, L. Wu, X. Pan, X. Tang, and Z. Wang, "Deep learning for air quality forecasts: a review," *Current Pollution Reports*, vol. 6, no. 4, pp. 399-409, Dec. 2020.

[20] A. Tella and A. L. Balogun, "GIS-based air quality modelling: Spatial prediction of PM10 for Selangor State, Malaysia using machine learning algorithms," *Environmental Science and Pollution Research*, Dec. 1, 2022, pp. 1-7.

[21] T. Tasioulis and K. Karatzas, "Reviewing explainable artificial intelligence towards better air quality modelling," in *Environmental Informatics*, Oct. 11, 2023, pp. 3-19, Cham: Springer Nature Switzerland.

[22] M. Zamani Joharestani, C. Cao, X. Ni, B. Bashir, and S. Talebiesfandarani, "PM2.5 prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data," *Atmosphere*, vol. 10, no. 7, p. 373, Jul. 4, 2019.

[23] M. U. Şimşek, İ. Kök, and S. Özdemir, "DeepFogAQ: A fog-assisted decentralized air quality prediction and event detection system," *Expert Systems with Applications*, vol. 251, p. 123920, Oct. 1, 2024.

[24] Z. Shahbazi, Z. Shahbazi, and S. Nowaczyk, "Enhancing air quality forecasting using machine learning techniques," *IEEE Access*, Dec. 13, 2024.

This page intentionally left blank.