Metaheuristic Optimization Review

♦ ISSN: 2692-4048 Vol: 03, No. 02, 2024 ♦ PP. 16–20

Article DOI: https://doi.org/10.54216/MOR.030204



REVIEW ARTICLE

A Review of Machine Learning Models for Predicting Air Quality in Urban Areas

El-Sayed M. El-kenawy[1,*]

[1] School of ICT, Faculty of Engineering, Design and Information & Communications Technology (EDICT), Bahrain Polytechnic, PO Box 33349, Isa Town, Bahrain

Emails: sayed.elkenawy@polytechnic.bh

Received: November 14, 2024 Revised: December 22, 2024 Accepted: January 14, 2025 ⋆ Corresponding author

ABSTRACT

Air pollution is a critical environmental issue that threatens much of the world, and public health, ecosystems, and the sustainability of cities are affected by its severe impacts. Urbanization and industrialization have been accelerating, with escalating pollution levels. Hence, air monitoring and air quality prediction are necessary for addressing such challenges. This review discusses advanced machine learning (ML), deep learning (DL) techniques, and IoT-based hybrid frameworks for air-quality prediction in urban settings. By integrating different datasets such as meteorological parameters, pollutant concentrations, and satellite imagery, these technologies provide strong and scalable solutions for real-time monitoring and forecasting. Key advancements include IoT-enabled sensors, convolutional and recurrent neural networks, and location-specific predictive models. Despite significant evolution, challenges involving data sparsity, computational requirements, and model adaptability remain. This paper frames these technologies as tools for transforming cities into smart and green cities and calls for continuous innovation and interdisciplinary collaboration to strengthen their effectiveness. These findings add to the advancement of knowledge on air-quality prediction methodologies and their crucial role in sustainable urban development.

Keywords: Air pollution ▪ Machine learning ▪ Deep learning ▪ IoT ▪ Air quality prediction ▪ Urban sustainability

1. INTRODUCTION

Pollution of the environment through emissions of greenhouse gases is one of the biggest concerns in the world today. It has far-reaching implications for people's health, the environment, and the future sustainability of urban centers. As urbanization and industrialization intensify, cities experience significant difficulties in regulating air quality, which is crucial for supporting public health. Consequently, ML and DL applications have been proposed as practical approaches for solving the problem accurately in various urban settings. Dominant studies concerning innovations in ML and DL models combined with IoT data collection, hybrid approaches, and location-based prediction algorithms are reviewed in this paper. These methodologies enhance the accuracy of air-quality forecasts and offer useful information for the design of more intelligent and sustainable cities. Therefore, by using these sophisticated technologies, researchers and policymakers can address air-quality issues sustainably and support the SDG goals in resilient urban environments.

1.1 The Growing Burden of Air Pollution in Urban Centers

Significant attention should be paid to air pollution as one of the key environmental problems of the present day, affecting large cities disproportionately. Increased emissions from vehicles, industries, and other human activities have made urban air unsafe and associated with serious health hazards. Densely populated urban regions with significant

© 2024 The Author(s). Published by ASPG. This article is licensed under CC BY 4.0.

industrial activity and vehicle traffic are currently among the highest emitters of dangerous pollutants, including PM2.5, PM10, NO_2, and SO_2. This has drastic impacts on public health and can lead to respiratory diseases, cardiovascular diseases, and premature deaths. Secondary effects such as acid rain and climate change exacerbate the problem through environmental cascading effects that influence ecosystems and biodiversity. Economic implications are also significant, including health-related costs and pollution-control expenditures. Impairment of air quality in urban areas also affects visibility, property value, and quality of life. Meeting this growing burden calls for sound, effective, and efficient solutions to monitor, forecast, and guide policies that help cities grow sustainably while protecting population health [1].

1.2 Machine Learning Innovations in Air Quality Prediction

Machine learning advancements have significantly facilitated modern air-quality prediction, giving the field tools to address complex and constantly changing pollution problems in cities. Conventional statistical paradigms often struggle with large and heterogeneous datasets, while ML algorithms can capture complex relationships among environmental variables. These models can consume data from satellite images, IoT sensors, and historical air-quality records to produce accurate and timely forecasts. Models such as random forests, support vector machines, gradient boosting, deep neural networks, and hybrid architectures have demonstrated strong potential for forecasting air-quality indices and pollutant concentrations [2, 3].

ML systems can also provide location-aware forecasts that help policymakers and citizens act before pollution episodes become severe. Predictive models can support traffic control, industrial-emission regulation, and public-health alerts. By linking forecasts to mobile applications and dashboards, citizens can decide whether to change commuting routes, postpone outdoor activities, or use cleaner transport modes. From a planning perspective, predictive modeling helps identify high-impact intervention points and supports long-term policy simulation. In summary, predictive modeling is a breakthrough direction for improving urban sustainability. Such models encourage evidence-based decisions and help governments, businesses, and citizens work toward smart, green, and resilient cities [1].

2. LITERATURE REVIEW

Continued advancements in industrialization and urbanization have placed environmental health under pressure, and air pollution has become one of the major problems of modern cities. Air quality in cities is therefore a significant challenge to human, ecological, and social well-being, creating a need for better methods for estimating and forecasting air quality. Advanced tools such as machine learning and deep learning can study massive environmental datasets and predict air quality with higher accuracy. The reviewed works cover deep learning structures used independently or in combination, IoT devices for data gathering and assessment, and learning models tailored for particular cities. These improvements not only enhance prediction accuracy but also offer meaningful data for policymakers and urban developers seeking sustainable responses to air pollution.

Ambient air pollution has been ranked among the leading environmental risk factors for health because it claims millions of lives annually, mostly in urban settings. Khan et al. [4] proposed an attention-enhanced deep multitask spatiotemporal learning model based on LSTM networks for megacities, where populations exceed 10 million and hazardous AQI levels can persist for weeks. The approach predicts major pollutants such as sulfur dioxide and carbon monoxide while capturing complicated patterns and variations that can guide policymakers.

Air pollution supervision is essential because it directly affects quality of life. Nandini and Fathima [5] presented an approach for predicting air-pollutant levels using K-means clustering, Multinomial Logistic Regression, and Decision Tree classifiers in R. Their comparison indicated that Multinomial Logistic Regression outperformed the Decision Tree model for air-pollution analysis. Blanco et al. [3] built machine and deep learning models to predict future concentrations of five key pollutants using Sentinel-5P satellite observations, climate, and topography. Experiments in Milan showed that predicted values achieved approximately 30% error for next-day pollutant levels and could be applied where monitoring stations are unavailable.

IoT-based pollution data collected in real time can support air-quality prediction and improve traffic-control systems. Ameer et al. [2] used measures such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to assess regression methods for pollution forecasting. Experiments on Apache Spark datasets showed the effectiveness of hybrid modeling and hyperparameter tuning in reducing prediction errors. Han et al. [6] proposed Deep-AIR, a hybrid CNN-LSTM framework for modeling complicated relationships between air pollution and urban dynamics. Using 1×1 convolution layers, the framework learned spatial characteristics of pollutants along with road density, building density, and street-canyon effects, outperforming baseline models in Hong Kong and Beijing.

Smart cities increasingly utilize IoT-based services to support sustainable urban lifestyles and improve awareness, interactivity, and efficiency. Kok et al. [7] developed an LSTM-based deep learning model for forecasting air-quality values in smart cities, demonstrating utility in several prediction scenarios. Naz et al. [8] compared LSTM, GRU, and statistical models for predicting NO_2, O_3, SO_2, PM2.5, and PM10 using data from Belfast. Their evaluation used RMSE, MAE, and R^2, strengthening evidence for deep learning as a solution for urban-pollution forecasting.

Bhushan et al. [9] provided an exploratory study of ML and DL methodologies for air-pollution and air-quality prediction, highlighting architectural ideas and emerging trends. Gururaj et al. [10] focused on mathematical models for air-quality prediction and analysis using real-time data including CO_2 concentration, temperature, and humidity. Mahalingam et al. [11] predicted AQI for smart cities using neural networks and SVM, while Barthwal and Goel [12] integrated DCNN and LSTM architectures for AQI time-series classification in Delhi, reporting high classification performance.

Recent work has also addressed city-specific and hybrid prediction. Shahbazi et al. [13] studied enhanced air-quality forecasting using ML techniques in mobility-related settings.

Chaturvedi [14] developed ML models for Varanasi AQI prediction using Random Forest, Decision Tree, and SVM. Sharma et al. [15] compared Random Forest, XGBoost, and hybrid models in Indian cities, showing that hybrid models can improve prediction accuracy and identify major pollutants. Dawar et al. [16] applied ML to predict AQI in Dehradun using historical pollutant and meteorological data; the Lasso Regressor achieved strong performance, including an R^2 value of 0.9999.

Table 1. Summary of Literature Review

Study	Key Focus	Methodology	Data Source	Findings
[4]	Predicting AQI in megacities using deep learning	Attention-enhanced deep multitask spatiotemporal model based on LSTM networks	Emissions data from urban megacities	Achieved high accuracy in predicting pollutants such as SO_2 and CO; guided policymaking for urban air-quality improvement.
[5]	Urban air-quality analysis and prediction	K-means clustering, Multinomial Logistic Regression, and Decision Tree classifiers	R programming and air-pollution data	Multinomial Logistic Regression outperformed Decision Tree in accuracy.
[3]	Urban air-pollution forecasting	ML models leveraging satellite data, meteorology, and topography	Sentinel-5P satellite data	Predicted pollutant levels with about 30% error; applicable in areas lacking monitoring stations.
[2]	IoT-based air-quality prediction in smart cities	Real-time IoT data and regression techniques assessed by MAE and RMSE	IoT devices and Apache Spark datasets	Hybrid models showed superior error minimization and prediction accuracy.
[6]	Deep learning framework for urban air-quality modeling	Deep-AIR hybrid CNN-LSTM framework	Urban road density, building density, and meteorological factors	Achieved better performance than baseline models in Hong Kong and Beijing.
[7]	Deep learning for air-quality forecasting in smart cities	LSTM-based forecasting model	IoT-enabled sensors in urban environments	Effective in various smart-city prediction scenarios.
[8]	Comparative analysis of air-pollutant prediction models	LSTM, GRU, and statistical methods	Air-quality monitoring station in Belfast	LSTM models demonstrated strong predictive performance.
[9]	Machine learning in air-pollution forecasting	ML/DL approaches for pollution-data analysis	Pollutant concentration and meteorological data	Highlighted effectiveness and future improvements in forecasting methodologies.
[10]	Air-quality analysis using mathematical models	Regression models and real-time data integration	CO_2 concentration, temperature, and humidity	Identified the need to balance heat transfer and ventilation for indoor air quality.
[11]	Smart-city AQI prediction	Neural networks and SVM	Central Pollution Control Board data in India	Random Forest and Decision Tree achieved very high AQI-prediction accuracy.
[12]	Deep learning for Delhi air conditions	DCNN-LSTM model	Meteorological and AQI data from 14 stations	DCNN-LSTM achieved 97.48% classification accuracy.
[13]	Smart mobility and air-quality alternatives	Real-time data analytics for air quality	Urban commuter and pollutant data	Encouraged eco-friendly commuting through predictive tools and incentives.
[14]	ML in Varanasi AQI prediction	Random Forest, Decision Tree, and SVM	Local pollutant datasets from 2019–2023	Random Forest and Decision Tree outperformed other models in accuracy.
[15]	Hybrid ML models in Indian cities	Random Forest, XGBoost, and hybrid models	Data from Ahmedabad, Delhi, Mumbai, and other cities	Hybrid models improved prediction accuracy and identified major pollutants.
[16]	AQI forecasting in Dehradun city	Lasso Regressor model	Historical AQI and meteorological parameters	Lasso Regressor achieved a 0.9999 R^2 score, demonstrating superior predictive capability.

The reviewed literature emphasizes that future air-pollution prediction and management will likely benefit from more complex machine learning and deep learning procedures. By coupling various environmental parameters and using data analysis, these models show the potential to improve monitoring and prediction efficiency in urban environments. Innovations such as hybrid models and IoT integration provide solutions that can be applied to various geographical and urban conditions. The literature also reveals the need to continually improve models as environmental conditions evolve. Altogether, this research direction supports wiser and more environmentally friendly municipalities that contribute to local, national, and global attempts to provide sustainable urban life and enhance population well-being.

3. DISCUSSION

Air pollution has been a challenge everywhere in recent years, but it has become more pronounced in cities because of population density, industrial activities, and vehicular emissions. If pollution sources are known, high-precision advanced tools can be used for prediction and mitigation. ML and DL have changed air-quality forecasting by drawing upon complex datasets and producing actionable insights. These applications integrate various datasets, including IoT-enabled sensors, satellite imagery, and meteorological parameters, to define urban air quality comprehensively.

The problem persists especially in metropolitan cities, where population density, industrial processes, and vehicle emissions significantly lower air quality. Addressing this problem involves combining knowledge of pollution sources with advanced tools for predicting and managing impacts. New ML and DL technologies promise a revolution in air-quality forecasting by analyzing and interpreting complex datasets into highly accurate actionable insights. Developed models integrate IoT sensors, satellite images, and meteorological parameters to assess urban air quality robustly [17].

Incorporating IoT technology has further advanced air-quality monitoring through real-time data capture and dynamic analysis. Hybrid models based on ML and IoT improve prediction accuracy and scalability for the complex diversity of cities. They also facilitate localized solutions, ranging from traffic optimization to industrial-emission control, thereby informing SDGs and building healthier, more sustainable cities [18].

Community engagement plays a vital role in addressing air-quality issues. ML-powered interfaces can provide real-time air-quality updates and personalized recommendations to citizens on reducing pollutant exposure. Making sustainable behavior, such as eco-friendly transport, more accessible promotes collective responsibility and active community participation [19].

Much progress has been made, but challenges remain, including data sparsity, limited model adaptability, and inadequate resources in developing countries. Stakeholders must collaborate to address these gaps. Such collaboration can bring air-quality management systems to a new level of efficiency and can help cities become sustainable and healthy places to live [20, 21].

4. CONCLUSION

Air pollution is a global threat, particularly in cities shaped by industrialization and vehicular emissions. Its effects on health, ecosystems, and sustainable development require reliable tools for monitoring and forecasting. This review examined machine learning, deep learning, and IoT-based models for predicting air quality in urban areas. The reviewed studies show that ML and DL techniques can integrate pollutant concentrations, meteorological information, satellite observations, and IoT sensor data to generate accurate and scalable forecasts.

The literature confirms the value of models such as LSTM, GRU, CNN-LSTM, Random Forest, XGBoost, SVM, and Lasso regression for AQI prediction and pollutant forecasting. Hybrid models and IoT-enabled systems are especially promising because they support real-time monitoring, localized interventions, and smart-city decision-making. These technologies can guide traffic planning, industrial-emission control, public-health warnings, and citizen engagement.

Despite these advances, several issues remain unresolved. Data sparsity, heterogeneous monitoring infrastructure, computational cost, and limited model adaptability continue to restrict practical deployment. Future research should emphasize robust transfer learning, explainable AI, low-cost sensor networks, and interdisciplinary collaboration among environmental scientists, engineers, urban planners, and policymakers. Continued innovation in this area will help cities become smarter, greener, more resilient, and better prepared to protect public health.

REFERENCES

[1] Y. Zheng, X. Yi, M. Li, R. Li, Z. Shan, E. Chang, T. Li, T.-Y. Liu, and H. Xiong, "Machine learning for urban air quality analytics: A survey," *Journal of the ACM*, vol. 37, no. 111, p. 35, 2023.

[2] S. Ameer *et al.*, "Comparative analysis of machine learning techniques for predicting air quality in smart cities," *IEEE Access*, vol. 7, pp. 128 325–128 338, 2019.

[3] G. Blanco, L. Barco, L. Innocenti, and C. Rossi, "Urban air pollution forecasting: A machine learning approach leveraging satellite observations and meteorological forecasts," in *2024 IEEE International Workshop on Metrology for Living Environment*, 2024, pp. 421–426.

[4] H. Khan *et al.*, "Novel approach for predicting the air quality index of megacities through attention-enhanced deep multitask spatiotemporal learning," 2024.

[5] K. Nandini and G. Fathima, "Urban air quality analysis and prediction using machine learning," in *1st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing and Communication Engineering*, 2019, pp. 98–102.

[6] Y. Han, Q. Zhang, V. O. K. Li, and J. C. K. Lam, "Deepair: A hybrid cnn-lstm framework for air quality modeling in metropolitan cities," 2021.

[7] I. Kok, M. U. Simsek, and S. Ozdemir, "A deep learning model for air quality prediction in smart cities," in *2017 IEEE International Conference on Big Data*, 2017, pp. 1983–1990.

[8] F. Naz *et al.*, "Comparative analysis of deep learning and statistical models for air pollutants prediction in urban areas," *IEEE Access*, vol. 11, pp. 64 016–64 025, 2023.

[9] M. Bhushan, I. Dawar, S. Sharma, T. K. Bawaniya, U. Anand, and A. Negi, "Air quality prediction using machine learning and deep learning: An exploratory study," in *2023 7th International Conference on Computing, Communication, Control and Automation*, 2023.

[10] A. Gururaj, V. Nagaraj, A. S. Rajesh, K. N. Achyuth, M. U. Somesh, and B. A. Dubay, "Air quality prediction and analysis using machine learning," in *2024 5th International Conference on Mobile Computing and Sustainable Informatics*, 2024, pp. 543–550.

[11] 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*, 2019, pp. 452–457.

[12] A. Barthwal and A. K. Goel, "Advancing air quality prediction models in urban india: A deep learning approach integrating dcnn and lstm architectures for aqi time-series classification," *Modeling Earth Systems and Environment*, vol. 10, no. 2, pp. 2935–2955, 2024.

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

[14] P. Chaturvedi, "Air quality prediction system using machine learning models," *Water, Air, and Soil Pollution*, vol. 235, no. 9, pp. 1–16, 2024.

[15] G. Sharma, S. Khurana, N. Saina, Shivansh, and G. Gupta, "Comparative analysis of machine learning techniques in air quality index (aqi) prediction in smart cities," *International Journal of System Assurance Engineering and Management*, vol. 15, no. 7, pp. 3060–3075, 2024.

[16] I. Dawar, M. Singal, V. Singh, S. Lamba, and S. Jain, "Predicting air quality index using machine learning: A case study of the himalayan city of dehradun," *Natural Hazards*, pp. 1–27, 2024.

[17] A. Gangwar, S. Singh, R. Mishra, and S. Prakash, "The state-of-the-art in air pollution monitoring and forecasting systems using iot, big data, and machine learning," *Wireless Personal Communications*, vol. 130, no. 3, pp. 1699–1729, 2023.

[18] I. Gryech, C. Assad, M. Ghogho, and A. Kobbane, "Applications of machine learning and iot for outdoor air pollution monitoring and prediction: A systematic literature review," 2024.

[19] N. Bandara, S. Hettiarachchi, and P. Athukorala, "Airspec: An iot-empowered air quality monitoring system integrated with a machine learning framework to detect and predict defined air quality parameters," 2021.

[20] M. D. and R. V., "Discretized linear regression and multiclass support vector based air pollution forecasting technique," *International Journal of Engineering Trends and Technology*, vol. 70, no. 11, pp. 315–323, 2022.

[21] I. Gryech, C. Assad, M. Ghogho, and A. Kobbane, "Applications of machine learning and iot for outdoor air pollution monitoring and prediction: A systematic literature review," 2024.