



Artificial Intelligence and Machine Learning for Air Pollution Monitoring and Prediction

Kotipalli. Harshitha¹, Gudla. Pramodh², Lanka. Vasu³, Chukka. Mohan Krishna⁴

Email: harshithakotipalli007@gmail.com

1.ABSTRACT :

The consequences of air pollution caused by speedy industrialization and urbanization are indeed a serious issue for human health and environmental sustainability. Older monitoring methods are based on costly and widely-spaced physical monitoring stations with no predictive capabilities. Current approaches involve the use of Internet of Things (IoT) sensor networks and sophisticated machine learning (ML) and deep learning (DL) models to overcome these challenges. This review combines the results of several fundamental works related to environmental sensing through the Internet of Things, supervised learning methods (Random Forest, XGBoost), deep temporal learning methods (Long Short-Term Memory networks, CNN-LSTM networks), and specific industrial monitoring applications. The published materials demonstrate that AI-based architectures can significantly improve predictive accuracy, real-time data imputation, and proactively manage cases of pollution. However, more technological development is needed to address issues of sensor calibration drift, crossregional generalizability, and data class imbalances.

2.Keywords: *Air quality monitoring, Machine learning, Deep learning, Internet of Things (IoT), Spatio-temporal correlation, Sensor calibration*

Introduction

Pollution in the atmosphere, particularly particulate matter, carbon monoxide, nitrogen oxides and sulfur dioxide, has a negative health, ecosystem and climate effect on human health globally. High levels of toxic chemicals are found in urban areas and heavy industrial areas (chrome plating facilities) resulting in an increased risk of cardiovascular and respiratory damage to workers and surrounding populations.

Traditional static monitoring stations offer reference grade accuracy, but they have prohibitive maintenance costs, limited spatial coverage and are unable to give up-to-date forecasting data. Today, with the emergence of low cost IoT hardware, edge computing nodes and wireless cloud transmission, environmental tracking has been revolutionized. Raw data from IoT streams, however, can often be noisy, lack calibration data, and contain drift. As a result, AI plays a crucial role in environmental governance today, using data to help measure pollution data, cleanse it and predict its trends in real-time, changing the rules of the game from reacting to pollution after it happens to preventing it from happening in the first place.

3.Literature Review

The following is presented through a themed synthesis of the literature to show how data structures, modeling paradigms and applications in the real world interrelate through the literature.

The scientific aspects of the utilization of artificial intelligence for air quality assessment can be seen in a progressive line, spanning from the very basic collection of air quality data to the more complex forecasting of a spatiotemporal nature, and finally to the control of the production process in industry. At the bottom layer (ingestion), studies point to a significant paradigm shift in the traditional, large reference-station approach, towards the deployment of low-cost, decentralised IoT sensor networks. These nodes are created with microcontrollers like Arduino or ESP8266 and measure environmental parameters at high frequencies, but with limited functionality and at a lower price. The raw sensor output has been known to be susceptible to noise, missing data and calibration drift, however, making this reason why unsupervised ML algorithms, such as k-means clustering, DBSCAN and Autoencoders, are widely used in the preprocessing stage to clean up the data, remove outlier points, and apply anomaly correction. In addition, the architectures taking advantage of the edge computing phenomenon are more and more used to process these data streams locally and before sending them to the cloud, thus minimizing latency.

After removing any extraneous data, predictive modeling comes in two main philosophies: physical dispersion modeling and data-driven statistical and AI modeling. The physical models, such as AERMOD, CALPUFF, and WRF, were used to track horizontal and vertical transport of pollutants in the past, and used to be the gold standard. But their use of boundary inputs that were exhaustively heterogeneous, however, led researchers to turn to machine learning. In the supervised learning setting, tree-based ensemble methods, such as Random Forest (RF), XGBoost and CatBoost, excel in terms of performance and efficiency for structured, tabular environmental data signals.

Deep learning has taken over from traditional methods for capturing complex temporal sequences and non-linear interactions on a larger geographical scale. LSTM networks, CNN networks and hybrid networks, like CNN-LSTM, TCN-BiLSTM, have been shown to perform well when modelling the

analysis of spatial-temporal correlation. These models have been able to capture well the effects of variables such as meteorology (temperature and wind speed), traffic density and Land Use Land Cover (LULC) on pollutant dispersion.

This predictive ability has just extended beyond the outdoor urban air tracking to the specialized indoor and industrial environment. For example, the real-time IoT-AI deployments in high-risk manufacturing plants, such as chrome plating plants, show how continuous multi-parametric monitoring can be combined with the predictive algorithm (LSTM and Random Forest) to predict the occurrence of pollution spikes in smaller time frames. These systems can predict hazardous thresholds ahead of time and automatically activate factory ventilation systems before they become hazardous to protect worker health.

Although all these developments were already significant, the literature as a whole shows that there are some recurring research shortcomings:

Research production is highly skewed toward North America, Europe, and parts of Asia, and other parts of the world, such as Sub-Saharan Africa and the Middle East, are underrepresented, limiting the transferability of models. Furthermore, there is less attention to algorithmic fairness in a variety of socio-economic communities.

Sensor deployment frameworks and studies that introduce low-cost sensor implementations often fail to include important deployment metrics like inference latency, memory usage, or real system stability trade-offs.

Low cost sensing elements are vulnerable and prone to degradation due to moisture and have to be scaled up over large urban networks where data integration is a major challenge.

Spatial Resolution Limits: Satellite imagery used to model LULC is typically not fine enough to resolve micro-urban street canyons and industrial forecasting algorithms cannot ingest industrial factory batch capacities dynamically to account for sudden pollution anomalies.

4. Conclusion

The fusion of IOT sensor infrastructure with machine learning has completely revolutionized the use of air quality measurements from a passive reporting system to an active predictive system. The advantages of ensemble models and deep neural networks are that they can effectively model complex linear and non-linear interactions between the environment, but there are still some challenges to be overcome such as geographic bias, sensor drift, and real-time operational constraints. The importance of explainable AI (XAI) to promote policy transparency, light-weight edge-computing architectures and cross-regional validation frameworks should be explored in future research to guarantee that AI based environmental governance will safeguard public health on a global scale.

5. REFERENCES

1. Karmoude, M., et al. (2025). Machine learning for air quality prediction and data analysis: Review on recent advancements, challenges, and outlooks. *Science of the Total Environment*, 1002, 180593.
2. Garbagna, L., Saheer, L. B., & Dar Oghaz, M. M. (2025). AI-driven approaches for air pollution modelling: A comprehensive systematic review. *Environmental Pollution*, 373, 125937.
3. Ramadan, M. N. A., et al. (2024). Real-time IoT-powered AI system for monitoring and forecasting of air pollution in industrial environment. *Ecotoxicology and Environmental Safety*, 283, 116856.
4. Garcia, A., et al. (2025). Advancements in air quality monitoring: A systematic review of IoT-based air quality monitoring and AI technologies. *Artificial Intelligence Review*, 58, 275.
5. Ansari, M., & Alam, M. (2023). An intelligent IoT-cloud-based air pollution forecasting model using univariate time-series analysis. *Arabian Journal for Science and Engineering*, 49(3), 3135-3162.
6. Wang, X., et al. (2024). Air quality forecasting using a spatiotemporal hybrid deep learning model based on VMD-GAT-BiLSTM. *Scientific Reports*, 14(1), 18035.
7. Kalantari, E., et al. (2024). Machine learning for air quality index (AQI) forecasting: shallow learning or deep learning? *Environmental Science and Pollution Research*, 31(54), 62962-62982.
8. Vajs, I., Drajić, D., & Cica, Z. (2023). Data-driven machine learning calibration propagation in a hybrid sensor network for air quality monitoring. *Sensors*, 23(5), 2815.
9. Park, J., Seo, Y., & Cho, J. (2023). Unsupervised outlier detection for time-series data of indoor air quality using LSTM autoencoder with ensemble method. *Journal of Big Data*, 10(1), 74.