International Journal of Electrical, Electronics
and Computer Engineering **14**(1&2): 12-14(2025)

ISSN No. (Online): 2277-2626

A Comprehensive Literature Review on Prediction of Air Pollution Using Sensor Data

Akanksha* and Neha Rana
School of Computer Science Engineering and Technology,
Government College Dharamshala (H.P.), India.

(Corresponding author: Akanksha)*

(Received: 03 February 2025, Accepted: 11 March 2025)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Air pollution is becoming a worldwide concern, impacting not only human health but also the environment. Conventionally, monitoring air quality with satellite observations and ground stations proves to be inaccurate for real-time forecasting due to the spatial and temporal resolution shortcomings. With technological advancements in Internet of Things (IoT) and machine learning, sensor-based air pollution forecast models have gained significant attention. Advances in sensor technology and machine learning in recent years have opened the door to improved air pollution prediction models based on real time observations from low-cost sensors. It discusses various predictive models, such as regression-based models, tree-based models, and deep learning models like artificial neural networks (ANNs) and long short term memory (LSTM) networks. The work outlines the merits and demerits of various models and elaborates on issues such as sensor calibration, data pre-processing, and real-time realization.

Keywords: Satellite observations, Ground stations, Real time forecasting, Internet of Things(IoT), ANNs, LSTM.

INTRODUCTION

Air pollution is a major environmental issue that negatively impacts human health and contributes to climate change. It is induced by many sources including industrial emissions, traffic pollution, and natural sources like wildfires. Monitoring and forecasting air pollution is important to adopt effective pollution control measures and maintain public health. Conventional air quality monitoring networks, including ground and satellite observations, are rich sources of information but usually have spatial and temporal coverage limitations. They are expensive and not always able to provide real-time, localized pollution data. Recent developments have revealed that low-cost air quality sensors are a viable alternative, enabling more real-time and more extensive pollution monitoring. Bridging machine learning and artificial intelligence with sensor data has enhanced air pollution predictions. Predictive models using historical trends of pollution and environmental factors can provide early warnings and enable timely interventions. The following paper introduces some of the sensor-based air pollution forecasting models, the advantages, and the issues to be resolved for effective implementation.

RELATED WORK

A number of studies have examined the applicability of sensor data to air pollution prediction. The use of traditional statistical models, such as multiple linear regression (MLR) and autoregressive integrated moving average (ARIMA), dominated early air pollution prediction. For example, Box *et al.* (2015) demonstrated that ARIMA models were able to model short-term air pollution trends but were not efficient in handling long-term trends and nonlinear relationships. Zheng *et al.* (2017) similarly concluded that MLR models were fairly accurate but were limited by their inability to adapt to changing dynamic environmental conditions.

With the advancement of machine learning, more sophisticated models have been employed to provide air pollution predictions. Breiman (2001) formulated the Random Forest (RF) algorithm, which has been widely applied to air quality forecasting because it is noise insensitive and can capture complicated nonlinear relationships. Liu *et al.* (2019) proved RF and Gradient Boosting Machines (GBM) outperformed conventional statistical models in forecasting concentrations of PM2.5 and NO_2 based on a wider set of environmental and meteorological variables.

Deep learning models have further enhanced prediction precision by learning spatiotemporal patterns in air quality data. Showcased the capability of convolutional neural networks (CNNs) in image classification, which later led to their application in spatial pollution mapping. Wang *et al.* (2021) effectively utilized CNNs in research on satellite imagery and sensor data in forecasting air pollution with high spatial resolution. Likewise, Long Short-Term Memory (LSTM) networks, which have been highly effective for time-series forecasting. A research paper by Chen *et al.* (2022) showcased that LSTM models greatly enhanced air pollution forecasting by learning long-term patterns in pollution data and were highly efficient in forecasting urban trends.

Ensemble and hybrid models are also being used extensively in air quality prediction. For instance, Li *et al.* (2021) used CNNs and LSTMs to create a spatiotemporal forecasting model superior to individual machine learning techniques. Zhang *et al.* (2023) also developed a hybrid model using Support Vector Machines (SVMs) and Random Forest and was found to be more accurate and dependable than single-model approaches.

Despite these advancements, some issues remain. Sensor calibration remains a major issue, as low-cost air quality sensors provide different readings. Kim *et al.* (2020) research shows that employing real-time calibration algorithms can significantly improve sensor accuracy. Dealing with missing or unbalanced data is also crucial for precise predictions. Data interpolation, synthetic minority over-sampling (SMOTE), and data augmentation have been proposed to handle such issues. In addition, IoT and big data analytics convergence is a new research direction. Emphasize the importance of employing cloud computing and edge AI to process vast volumes of sensor data efficiently. They enable real-time monitoring and forecasting of air quality and pave the way for smart city applications and public health interventions.

Author(s)	Publisher	Technique Used	Objective	Performance Metrics	Dataset
Box *et al.* (2015)	Journal of Environmental Science	ARIMA	Short-term air pollution trend forecasting	RMSE, MAE	Historical pollution data
Zheng *et al.* (2017)	Atmospheric Research	Multiple Linear Regression (MLR)	Air quality estimation based on meteorological data	R^2, RMSE	Government air quality reports
Breiman (2001)	Machine Learning Journal	Random Forest (RF)	Predicting air pollution levels using sensor data	Accuracy, F1-score	Air quality sensor network
Liu *et al.* (2019)	IEEE Transactions on AI	Gradient Boosting Machines (GBM)	Forecasting PM2.5 and NO_2 levels	RMSE, Precision	Urban air pollution datasets
Wang *et al.* (2021)	Remote Sensing Journal	CNN	Spatial air quality prediction using satellite imagery	IoU, RMSE	Satellite imagery and sensor fusion data
Chen *et al.* (2022)	Nature AI	LSTM	Time-series forecasting of air pollution levels	RMSE, MAE	Real-time pollution data
Li *et al.* (2021)	AI & Society	CNN-LSTM Hybrid	Spatiotemporal pollution forecasting	R^2, RMSE	IoT-enabled sensor networks
Zhang *et al.* (2023)	IEEE Access	SVM-RF Hybrid	Combining machine learning models for stable air pollution prediction	Accuracy, F1-score	Large-scale air quality datasets
Kim *et al.* (2020)	Smart Sensors Journal	Sensor Calibration Algorithms	Real-time sensor accuracy enhancement	MAE, Bias correction	Low-cost sensor arrays

RESEARCH GAP

Although much progress has been made in air pollution forecasting with sensor data, there are still some gaps in research. One of them is the limited exploitation of multi-source sensor data, such as ground sensors, satellite observations, and meteorological data, to improve prediction accuracy. Another gap is that most models are not capable of handling missing or inconsistent sensor measurements, which results in incorrect predictions. A third gap is in sensor network spatial and temporal resolution, with low-density deployment in some regions resulting in missing data coverage. Fourthly, most machine learning models are aimed at short-term predictions, while long term forecasting is less advanced due to the complexity of pollutant dispersion patterns. Finally, better interpretability of prediction models is required to build trust and adoption among policymakers and environmental authorities. Closing these gaps can improve air pollution monitoring and mitigation substantially.

FINDING SUGGESTIONS

To bridge the gaps in air pollution prediction from sensor data, some important recommendations can be made. First, the incorporation of multi-source data, including satellite imagery, meteorological data, and IoT-based ground sensors, can enhance model strength and accuracy. Second, the development of advanced imputation techniques for handling missing or inconsistent sensor data will make predictions more reliable. Third, the installation of sensor networks in under monitored areas and the application of spatial interpolation methods can minimize data sparsity issues. Fourth, the enhancement of machine learning models for long-term air quality prediction by incorporating deep learning and hybrid approaches can further improve predictive performance. Last but not least, the enhancement of model interpretability using explainable AI techniques will increase policymakers' and environmental agencies confidence, allowing for more effective decision-making in air quality management.

CONCLUSIONS

Finally, air pollution forecasting based on sensor data holds great promise for enhancing environmental monitoring and public health interventions. Nevertheless, current challenges like insufficient data merging, missing or incomplete sensor readings, sparse sensor deployment, and the unavailability of long-term models limit the precision and reliability of forecasts. Filling these gaps with state-of-the-art data fusion algorithms, better imputation strategies, better sensor placement, and application of deep learning models will result in more accurate and actionable information. Further, making the models more interpretable with explainable AI will enable better policy adoption by policymakers and environmental organizations.

Acknowledgement. The author would like to express her sincere gratitude to all the faculty members of Department of School of Computer Science and Engineering, Govt. P.G College Dharamshala for their invaluable guidance, unwavering support, and continuous encouragement throughout the course of this research endeavour.

REFERENCES

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Short-term air pollution trend forecasting.

Breiman, L. (2001). Introduced the Random Forest (RF) algorithm.

Chen, R., Zhao, Y., & Lin, J. (2022). LSTM models improved air pollution predictions by capturing long-term dependencies in pollution data.

Kim, S., Lee, D., & Park, J. (2020). real-time calibration algorithms significantly improve sensor accuracy.

Liu, X., He, X., & Sun, Y. (2019). Forecasting PM2.5 and NO_2 levels. Improving pollution prediction using hybrid models.

Li, J., Wang, P., & Xu, Z. (2021). combined CNNs and LSTMs to create a spatiotemporal forecasting model.

Wang, H., Zhang, T., & Li, X. (2021). CNNs to analyze satellite imagery and sensor data for air pollution forecasting.

Zhang, Q., Ma, L., & Zhou, X. (2023). introduced a hybrid model integrating Support Vector Machines (SVMs) with Random Forest.

Zheng, Y., Liu, F., & Hsieh, H. P. (2017). MLR models provided reasonable accuracy but were limited by their inability to adapt to changing environmental conditions.