Journal of Scientific Insight
E-ISSN: 3062-8571
DOI: https://doi.org/10.69930/jsi.v1i3.183
Review Article
Vol. 1 (3), 2024
Page: 119-130

The Application of Machine Learning and Intelligent Sensors for Real-Time Air Quality Monitoring: A Literature Review

Syaiful Bachri Mustamin [1,*], Muhammad Atnang [1], Sahriani [1], Nurhikmah Fajar [2], Sri Kurnian Sari [3], Muammar Reza Pahlawan [1], Mujahidin Amrullah [1]

[1] Department of Information Technology, Faculty of ScienceTechnology and Health, Institute of Science Technology and Health 'Aisyiyah Kendari, Indonesia
[2] Department of Computer Engineering, Faculty of Engineering, Muhammadiyah Institute of Technology and Science North Kolata, Indonesia
[3] Departement of Informatics Engineering, Faculty of Computer Science, Eastern Indonesia University, Indonesia

*Email (corresponding author): sbm@istekaisyiyah.ac.id

Abstract
Air pollution is a global issue that has major consequences for human health and the environment. Accurate air quality prediction plays an important role in mitigating and preventing the negative impacts of air pollution. The thirteen sources analyzed in this literature study show a growing trend in the use of machine learning for air quality prediction, driven by the limitations of traditional methods and machine learning capabilities in efficiently processing complex data. This literature study examines a variety of commonly used machine learning models, such as Support Vector Regression (SVR), Random Forest, Gradient Boosting, and Long Short-Term Memory (LSTM), and evaluates their performance based on metrics such as RMSE, MAE, and R^2. The sources also highlight the importance of understanding the factors that affect air quality, including concentrations of various pollutants (PM2.5, PM10, NO2, CO, SO2, and ozone), meteorological data (temperature, humidity, wind speed, air pressure, precipitation, and temperature inversion), traffic data, and spatial-temporal variations. The integration of the Internet of Things (IoT) and machine learning is the main focus in the development of real-time air quality monitoring systems. IoT sensors enable the collection of real-time air quality and meteorological data, which are then processed using machine learning models to generate predictions. This literature study identifies several challenges in air quality prediction, such as data limitations, the complexity of air pollution dynamics, and ethical & privacy considerations. However, machine learning offers great potential to improve the accuracy of air quality predictions and monitoring, thus contributing to a healthier and more sustainable environment.

Keywords: *Air pollution, Air quality, machine learning, IoT, sensor*

1. Introduction

Air pollution has become a global public health crisis, contributing to millions of premature deaths each year (Khomenko et al., 2021). Increasing urbanization, industrialization, and vehicle emissions have led to an increase in The concentrations of harmful air pollutants such as PM2.5, PM10, NO2, and ozone. To overcome these challenges, accurate and timely Monitoring and forecast of air quality. is essential (Chae et al., 2021; Ramli et al., 2023).

Traditional air quality monitoring systems, which rely on fixed monitoring stations, are frequently expensive, have limited coverage, and do not provide real-time predictions. (Samadi & Kumawat, 2023). However, the rise of *machine learning* has revolutionized the field of air quality prediction, offering a cost-effective, scalable, and predictive alternative (Pandithurai et al., 2023; Saikiran et al., 2021).

Machine Learning Algorithms have shown tremendous potential in predicting air pollutant concentrations by utilizing complex patterns in historical data and predictive variables (Anggriani et al., 2024; Oğuz & Pekin, 2024). A number of studies have explored the efficiency of various machine learning algorithms, such as Support Vector Regression. (SVR), *Random Forest*, *Artificial Neural Network* (ANN), and *Extreme Gradient Boosting* (XGBoost), in the context of air quality prediction (Ashraf et al., 2022; Lei et al., 2023; Li et al., 2020).

The study consistently highlights the ability of *machine learning* models to achieve high prediction accuracy, outperforming traditional statistical methods. In addition, advances in *deep learning*, particularly models such as *Long Short-Term Memory* (LSTM), have further improved predictive capabilities by capturing complex temporal dependencies in air quality data (Mao, Jiao, et al., 2021; Mao, Wang, et al., 2021; Zhou et al., 2019).

One of the growing research focuses is the ability of *machine learning* models to predict pollutant concentrations in locations different from the training data locations. This is important given the need for wide monitoring coverage and the limited number of physical monitoring stations (Balamurugan et al., 2023; Ren et al., 2022).

The research of Samad et al., for example, shows that *machine learning* models can be successfully applied to predict pollutant concentrations in different locations, albeit with slightly lower accuracy compared to predictions at training data locations (Samad et al., 2023). The study emphasizes the importance of incorporating spatial data, such as pollutant concentrations from nearby monitoring stations, to improve the model's accuracy on different location predictions (Miasayedava et al., 2023; Wardana et al., 2021).

In addition to *machine learning* algorithms, the right selection of features and data processing are essential for the performance of prediction models (Bhimavarapu & Sreedevi, 2022; Fang et al., 2022; Neagu et al., 2002; Ul-Saufie et al., 2022). Meteorological parameters, such as temperature, humidity, wind speed, and precipitation, have been shown to affect pollutant concentrations, and are often included as predictive variables(Dai et al., 2021). Traffic data, such as the number of vehicles, is also an important factor in predicting pollutant concentrations in urban areas (Ul-Saufie et al., 2022).

Autocorrelation analysis and *principal component analysis* (PCA) are commonly used techniques to identify the most relevant features and reduce the dimensions of the data, thereby improving the efficiency and accuracy of the model (Liu et al., 2023; Lu et al., 2022).

Despite the great potential *of machine learning* in air quality prediction, some challenges and opportunities remain (Aggrawal & Bhushan, 2023). Data limitations, spatial and temporal complexity of air pollution, and the need for generalizable models are some of the challenges that need to be addressed (Sharma et al., 2022).

Future research should focus on developing more sophisticated models that can address these challenges, combining data from various sources, including IoT sensors, satellite imagery, and social media platforms, and leveraging cutting-edge *deep learning* techniques (Mengara Mengara et al., 2022). Additionally, emphasis should be placed on developing real-time air quality prediction systems that can provide actionable information for policymakers, public health authorities, and the general public (Parkavi & Rathi, 2021a).

2. Methods

Research on air quality prediction using *machine learning* has grown rapidly, with various approaches and models being implemented to improve prediction accuracy and aid in environmental decision-making. In India, (Ravindiran et al., 2023) developed an air quality prediction model for the city of Visakhapatnam using a dataset consisting of 12 air pollutants and 10 meteorological factors. Although the model used is not described in detail, this study shows the importance of long-term monitoring. In contrast, (Ramadan et al., 2024) emphasize monitoring the intelligent environment using IoT and *deep learning*, although the specific *deep learning* model used is not explained . The study shows that sensors can provide valuable *real-time* data in air quality management, especially with parameters such as PM2.5 and other harmful gases.

A systematic review by (Z. Zhang et al., 2024) revealed that *deep learning* models such as CNNs, RNNs, and LSTMs have shown superiority in air quality prediction. They emphasized the importance of selecting the right model to handle complex pollution data. (Samad et al., 2023) developed an air pollution prediction model using LSTM combined with a genetic algorithm for hyperparameter optimization, showing that this optimization technique can improve prediction performance.

In Stuttgart, the study using several (Parakawi el al., 2021)*machine learning models* such as *ridge regressors*, *random forests*, and *support vector regressors*, managed to estimate pollutant concentrations with high accuracy. Another study by (Ravindiran et al., 2023) in Bangladesh introduced the use of IoT and *machine learning for* real-time *air quality monitoring*, with regression and classification models used to predict AQI based on pollutants such as SO2, CO, and NO2.

The LSTM approach was also used in studies that tested the accuracy of air pollution predictions with a larger dataset. On the other hand, (Parakawi el al., 2021) (Castelli et al., 2020) used *Support Vector Regression* (SVR) to predict pollutants in California, with data pre-processing that includes imputation of lost data to improve prediction accuracy. This study shows that SVR with RBF kernel is an effective model in predicting air pollution in urban areas.

A combined technique that combines statistical and *machine learning* methods for air quality prediction is proposed by Prediction of Air Quality and Pollution using *Linear Regression* and *Adaboosting*. Meanwhile, Real Time Localized Air Quality Monitoring uses fixed and mobile IoT sensors to provide local predictions in *real-time*, with models such as SVR and *Random Forest* being used to improve prediction accuracy.

In Taiwan, the study used models such as (Liang et al., 2020)*Random Forest* and *Adaboost* predict AQI, suggesting that a combination of multiple models can provide more accurate results. (Hardini et al., 2023) in their study on AQI prediction using *ensemble machine learning* also emphasized the importance of integrating sensors with data processing platforms to generate better predictions. Finally, research by combining CNN and LSTM to predict PM2.5 concentrations in Beijing. This hybrid model is optimized for (Wardana et al., 2021)*edge computing devices*, demonstrating the potential of the technology for more efficient applications in the field.

Overall, these studies show that the use of *machine learning models* such as LSTM, CNN, SVR, and *ensemble learning* has made a significant contribution to improving the accuracy of air quality predictions, especially when combined with *real-time* data from IoT sensors. These approaches provide deeper insights into how air pollution can be managed more efficiently with advanced technology.

Table 1. Summary of research on air quality prediction using machine learning and deep learning methods

Researchers	Heading	Method	Information
(Ravindiran et al., 2023)	Air quality prediction using machine learning models: A predictive research on the Indian coastal city of Visakhapatnam.	*Gradient Boosting Machine* (GBM)	This article focuses on AQI prediction and lists GBM as one of the methods used, but does not detail other models.
(Ramadan et al., 2024)	Smart environment monitoring for chrome plating plant using internet of things and deep learning	LSTM, *Random Forest*, Regresi Linear	LSTM is used to predict temperature and humidity, *Random Forest* for PM2.5, and linear regression for other parameters.
(Z. Zhang et al., 2024)	A thorough survey of air quality prediction using deep learning.	TDGTN, ATGCN, *Hierarchical table bidirectional encoder representations from transformers* (BERT), *Hybrid sequence-to-sequence* (seq2seq), *Multi-Task Learning* (MTL)	This article is a systematic *review* of air quality prediction using *deep learning* and discusses the various models used in other studies.
(Samad et al., 2023)	Machine learning-based estimation of PM2.5, PM10, and NO2 concentrations at two Stuttgart sites	*Ridge regressor*, *Support Vector Regressor*, *Random Forest*, *Extra Trees Regressor*, dan *Extreme Gradient Boosting*	This article evaluates the performance of these five models in predicting pollutant concentrations.
(Islam et al., 2024)	A real-time dataset of air quality index monitoring using IoT and machine learning in the perspective of Bangladesh	Regression, Classification	The article mentions the use of *machine learning algorithms* for regression and classification.
(Parkavi & Rathi, 2021a)	A real-time dataset of air quality index monitoring utilising IoT and machine learning from the perspective of Bangladesh.	*Long Short-Term Memory* (LSTM)	The article discusses the use of LSTM for air pollution prediction, but does not mention the title or researcher specifically.

(Castelli et al., 2020)	A Machine Learning Approach for Predicting Air Quality in California	*Support Vector Regression* (SVR)	SVR is used to predict various pollutants in California.
(Liang et al., 2020)	Prediction of Air Quality and Pollution using Statistical Methods	Regresi Linear, *Adaboosting*	This article combines statistical and *machine learning methods.*
(D. Zhang & Woo, 2020)	Real-time localised air quality monitoring and prediction using mobile and fixed IoT sensing networks.	*Support Vector Regression* (SVR), *Random Forest Regression* (RFR), *Gradient Boosting* (GB)	This article compares the performance of the three models in predicting PM2.5 and PM10.
(Liang et al., 2020)	Air Quality Prediction Using Machine Learning	*Random Forest, Adaboost, Support Vector Machine* (SVM), *Stacking Ensemble Learning*	This article discusses AQI predictions in Taiwan, but does not mention specific titles or researchers.
(Hardini et al., 2023; Liang et al., 2020)	predicting-air-quality-index-using-ensemble-machine-learning	*Ensemble Machine Learning*	This article discusses AQI prediction using *ensemble machine learning* but does not mention specific models.
(Wardana et al., 2021)	Optimising deep learning at the edge for accurate hourly air quality prediction	CNN 1D and LSTM	The article discusses CNN-LSTM for PM2.5 prediction in Beijing, but does not mention the specific title.

3. Results and Discussion

Monitoring and predicting air quality is an important challenge amid increasing air pollution that endangers health and the environment. As technology advances, various machine learning-based approaches have been used to predict air quality more accurately. This article combines findings from 13 recent studies that address *machine learning* models for air quality prediction. Discussions will include model comparisons, multi-source data integration, real-time system implementation, impact on decision-making, ethical challenges, and future research directions.

3.1. Comparison of Machine Learning Models for Air Quality Prediction

Numerous machine learning models have been utilized to forecast air quality, such as Support Vector Regression (SVR), Random Forest, Gradient Boosting Machine (GBM), and Long Short-Term Memory (LSTM). Each model has its advantages and disadvantages.

i. SVRs, as used by Castelli et al. (2020), excel at handling non-linear data and are highly effective on small datasets. The main advantage of SVR is the ability to handle non-linear data with customizable kernels, such as RBF kernels that improve pollutant prediction performance in urban areas.
ii. Random Forest offers the advantage of handling high-dimensional data and providing results that are easy to interpret. The Estimation of PM2.5, PM10, and

NO2 shows that Random Forest can be used effectively in short-term predictions for harmful pollutants such as PM2.5 and NO2.

iii. LSTM is a popular model for processing time serial data due to its ability to handle long-term dependencies on data. Drewil & Al-Bahadili (2022) successfully showed that the combination of LSTM with genetic algorithms for hyperparameter optimization resulted in an increase in air quality prediction accuracy. However, LSTMs require more complex parameter optimization than other models.

iv. The Gradient Boosting Machine (GBM) and other ensemble learning techniques, used by Hardini et al. (2023), offer greater flexibility and predictive capabilities, but require greater computational time and resources for model tuning.

The effectiveness of these models is evaluated using metrics like Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Each model shows different performance based on the data and environmental characteristics of each study site. Liang et al. (2020) used RMSE and MAE to evaluate the accuracy of AQI predictions in Taiwan, finding that *stacking ensembles* and *random forests* provided the most accurate results.

3.2. Multi-Source Data Integration for More Accurate Air Quality Prediction

The accuracy of air quality predictions can be enhanced by incorporating diverse data sources, including meteorological information, traffic data, pollutant data from IoT sensors, satellite imagery, and industrial emission data. Zhang et al. (2024) highlight the importance of combining multi-source data to improve the accuracy of *deep learning* models in air quality prediction. Data from various sources, such as pollutant historical data, meteorological data, and topographic data, provide a more complete picture of the factors that affect air quality.

Techniques such as *data fusion* and *multi-modal learning* can be used to integrate data from multiple sources, but multi-source data integration faces challenges, including data heterogeneity, inappropriate scale, and missing *value issues*. For example, (Pasupuleti et al., 2020) showed that the integration of real-time data with historical data improves the accuracy of AQI predictions in Bangladesh.

3.3. Application of Machine Learning Models in Real-time Air Quality Monitoring Systems

The integration of IoT sensors with cloud computing platforms has facilitated the creation of precise, real-time air quality monitoring systems. These systems gather environmental data, such as PM2.5 levels and other pollutants, through sensors and apply machine learning models to deliver real-time air quality forecasts.

(Pasupuleti et al., 2020) developed an IoT-based real-time system in Bangladesh to monitor air quality, where sensors collect data on key pollutants such as SO2, CO, and NO2. Challenges in real-time data processing, such as latency and throughput, need to be considered in order for the system to provide timely information. Data visualization is also an important part of this system. Techniques such as heatmaps and *visual dashboards, such as those used in* Liang et al.'s (2020) study**,** help present air quality data in an informative manner to users.

3.4. Impact of Air Quality Prediction on Decision-Making and Public Health

Air quality prediction systems have a significant impact on decision-making, especially in protecting public health. Exposure to pollutants such as PM2.5, NO2, and O3 has harmful effects on health, including increasing the risk of respiratory and cardiovascular diseases.

By using the AQI prediction model, as conducted by (Hardini et al., 2023)**,** governments and communities can receive early warning when air pollution reaches dangerous levels. This allows for preventive measures, such as the use of masks or restrictions on outdoor activities, which can reduce the negative impact of air pollution exposure.

In addition, accurate air quality information can help governments formulate more effective public policies to reduce air pollution and protect public health, such as limiting vehicle emissions or stricter industry regulations.

3.5. Ethical and Privacy Challenges in Air Quality Monitoring and Prediction

Air quality monitoring systems that use IoT sensors also present ethical challenges, especially related to privacy and data protection. The collection of location data and other personal data through IoT sensors can raise concerns regarding user privacy. It is important for system developers to ensure that the data collected is protected and not misused.

In addition, *machine learning* models can inherit biases from the data used to train them. Therefore, it is necessary to evaluate the prediction model to ensure that it is not discriminatory against certain groups and provides fair predictions for all populations.

3.6. Future Research Directions

Although *ML* models have yielded promising results in air quality prediction, there are still many challenges that need to be addressed. Further research is required to create models that are more accurate and dependable. One promising area of research is the development of *Explainable AI (XAI)* models, which allow users to understand and interpret predictions generated by *machine learning models*.

The integration of air quality monitoring systems with other *smart city* systems is also an important step forward. By connecting air quality prediction systems with smart transportation systems and energy management systems, cities can optimize their infrastructure to create healthier and more sustainable urban environments.

Through a combination of various *machine learning* models, multi-source data integration, and the deployment of real-time systems, air quality monitoring and prediction has made significant progress. Each model has its advantages and disadvantages, and is suitable for different applications based on data characteristics. Obstacles like data privacy, algorithmic bias, and security still need to be addressed, but with more research and integration with *smart city technologies*, more effective solutions to reduce the impact of air pollution can be achieved.

Conclusions

Based on the reviewed studies, it is evident that air quality prediction through machine learning has become a swiftly advancing trend. This growth is fueled by the limitations of traditional methods and the strengths of machine learning in processing complex data effectively. Air pollution remains a worldwide issue with substantial effects on human health and the environment, making precise air quality predictions essential for reducing and preventing the harmful consequences of pollution..

A variety of machine learning models have been applied, with widely used examples including Support Vector Regression (SVR), Random Forest, Gradient Boosting, and Long Short-Term Memory (LSTM). Choosing an appropriate model depends on data characteristics, the complexity of relationships between variables, and the study's objectives. Model performance is often assessed through metrics like Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R^2. Some studies also emphasize the value of statistical models, such as Autoregressive Integrated Moving Average (ARIMA), which can be combined with machine learning for enhanced predictions.

Several critical factors influence air quality, including pollutants like PM2.5, PM10, NO_2, CO, SO_2, and ozone, along with meteorological variables such as temperature, humidity, wind speed, air pressure, and precipitation. Temperature inversion's effect on PM concentration is notable, especially in winter. In urban areas, high traffic density significantly increases pollutant levels. Due to spatial and temporal variations in air quality, predictive models must consider both factors to improve accuracy.

The integration of the Internet of Things (IoT) and machine learning has made real-time air quality monitoring more achievable. IoT sensors facilitate real-time collection of air quality and meteorological data, which is processed by machine learning models and displayed on online platforms or mobile apps, enabling people to access timely information and respond to air pollution more effectively.

However, the development of air quality prediction models faces challenges, including the limitations of accurate data, the complexity of air pollution dynamics, and the challenge of building models that can capture these complexities. Continuous research aims to improve the accuracy of *machine learning* models and integrate various data sources. In addition, ethical and privacy aspects need to be considered in the collection and processing of data used in air quality monitoring systems. Overall, *machine learning* has great potential in improving air quality prediction and monitoring, especially with IoT integration enabling *real-time* system development. The system provides useful information that can support communities and policymakers in reducing the impact of air pollution.

Funding

This research is financed independently and does not receive support from any external funding sources.

Acknowledgments

We would like to express our deepest appreciation and sincere thanks to the Institute of Science Technology and Health 'Aisyiyah Kendari, the Institute of Technology and Sciences Muhammadiyah North Kolaka, and the Department of Informatics Engineering, Faculty of Computer Science, Eastern Indonesia University for their invaluable support. Their contributions have been incredibly meaningful and have significantly aided in reaching our collective goals. The collaborations established have not only enhanced institutional partnerships but also fostered shared progress in the fields of science and technology. We are truly grateful for every effort and contribution made.

Conflicts of Interest

The authors declare no conflict of interest.

References

Aggrawal, S., & Bhushan, B. (2023). Machine Learning for Air Pollution. *2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN)*, 1–6. https://doi.org/10.1109/VITECON58111.2023.10157028

Anggriani, D., Bachri Mustamin, S., Atnang, M., Fatmah, S., Azaliah Mar, N., Fajar, N., dan Kesehatan, T., & Kendari, A. (2024). Transforming the Diabetes Mellitus Diagnosis and Treatment Using Data Technology: Comprehensive Analysis of Deep Learning and Machine Learning Methodologies. *Journal of Scientific Insights*, *1*(1), 26–32. https://doi.org/10.69930/JSI.V1I1.71

Ashraf, M. U., Akram, F., & Usman, S. (2022). Comparative Analysis of Machine Learning Techniques for Predicting Air Pollution. *Lahore Garrison University Research Journal of Computer Science and Information Technology*, *6*(2), 40–54. https://doi.org/10.54692/LGURJCSIT.2022.0602270

Balamurugan, V., Chen, J., Wenzel, A., & Keutsch, F. N. (2023). Spatiotemporal modeling of air pollutant concentrations in Germany using machine learning. *Atmospheric Chemistry and Physics*, *23*(17), 10267–10285. https://doi.org/10.5194/ACP-23-10267-2023

Bhimavarapu, U., & Sreedevi, M. (2022). Kurtosis-Based Feature Selection Method using Symmetric Uncertainty to Predict the Air Quality Index. *Comput. Sci. J. Moldova*, *30*(3), 360–375. https://doi.org/10.56415/CSJM.V30.19

Castelli, M., Clemente, F. M., Popovič, A., Silva, S., & Vanneschi, L. (2020). A Machine Learning Approach to Predict Air Quality in California. *Complexity*, *2020*. https://doi.org/10.1155/2020/8049504

Chae, S., Shin, J., Kwon, S., Lee, S., Kang, S., & Lee, D. (2021). PM10 and PM2.5 real-time prediction models using an interpolated convolutional neural network. *Scientific Reports*, *11*(1). https://doi.org/10.1038/S41598-021-91253-9

Dai, H., Huang, G., Zeng, H., & Yang, F. (2021). PM2.5 Concentration Prediction Based on Spatiotemporal Feature Selection Using XGBoost-MSCNN-GA-LSTM. *Sustainability*, *13*(21). https://doi.org/10.3390/SU132112071

Fang, L., Jin, J., Segers, A., Lin, H. X., Pang, M., Xiao, C., Deng, T., & Liao, H. (2022). Development of a regional feature selection-based machine learning system (RFSML v1.0) for air pollution forecasting over China. *Geoscientific Model Development*, *15*(20), 7791–7807. https://doi.org/10.5194/GMD-15-7791-2022

Hardini, M., Sunarjo, R. A., Asfi, M., Riza Chakim, M. H., & Ayu Sanjaya, Y. P. (2023). Predicting Air Quality Index using Ensemble Machine Learning. *ADI Journal on Recent Innovation (AJRI)*, *5*(1Sp), 78–86. https://doi.org/10.34306/ajri.v5i1sp.981

Islam, M. M., Jibon, F. A., Tarek, M. M., Kanchan, M. H., & Perbhez Shakil, S. U. (2024). A real-time dataset of air quality index monitoring using IoT and machine learning in the perspective of Bangladesh. *Data in Brief*, *55*. https://doi.org/10.1016/j.dib.2024.110578

Khomenko, S., Cirach, M., Pereira-Barboza, E., Mueller, N., Barrera-Gómez, J., Rojas-Rueda, D., de Hoogh, K., Hoek, G., & Nieuwenhuijsen, M. (2021). Premature mortality due to air pollution in European cities: a health impact assessment. *The Lancet. Planetary Health*, *5*(3), e121–e134. https://doi.org/10.1016/S2542-5196(20)30272-2

Lei, T. M. T., Ng, S. C. W., & Siu, S. W. I. (2023). Application of ANN, XGBoost, and Other ML Methods to Forecast Air Quality in Macau. *Sustainability*, *15*(6). https://doi.org/10.3390/SU15065341

Li, Z., Yim, S. H. L., & Ho, K. F. (2020). High temporal resolution prediction of street-level PM2.5 and NOx concentrations using machine learning approach. *Journal of Cleaner Production*, *268*. https://doi.org/10.1016/J.JCLEPRO.2020.121975

Liang, Y. C., Maimury, Y., Chen, A. H. L., & Juarez, J. R. C. (2020). Machine learning-based prediction of air quality. *Applied Sciences (Switzerland)*, *10*(24), 1–17. https://doi.org/10.3390/app10249151

Liu, M., Hu, H., Zhang, L., Zhang, Y., & Li, J. (2023). Construction of Air Quality Level Prediction Model Based on STEPDISC-PCA-BP. *Applied Sciences*, *13*(14). https://doi.org/10.3390/APP13148506

Lu, T., Cen, M., Huo, S., & Luo, L. (2022). Principal Component Analysis-Improved Fuzzy Genetic Algorithm. *Proceedings of the 2022 2nd International Conference on Control and Intelligent Robotics*, 257–264. https://doi.org/10.1145/3548608.3559203

Mao, W., Jiao, L., Wang, W., Wang, J., Tong, X., & Zhao, S. (2021). A hybrid integrated deep learning model for predicting various air pollutants. *GIScience & Remote Sensing*, *58*(8), 1395–1412. https://doi.org/10.1080/15481603.2021.1988429

Mao, W., Wang, W., Jiao, L., Zhao, S., & Liu, A. (2021). Modeling air quality prediction using a deep learning approach: Method optimization and evaluation. *Sustainable Cities and Society*, *65*. https://doi.org/10.1016/J.SCS.2020.102567

Mengara Mengara, A. G., Park, E., Jang, J., & Yoo, Y. (2022). Attention-Based Distributed Deep Learning Model for Air Quality Forecasting. *Sustainability*, *14*(6). https://doi.org/10.3390/SU14063269

Miasayedava, L., Kaugerand, J., & Tuhtan, J. A. (2023). Lightweight Open Data Assimilation of Pan-European Urban Air Quality. *IEEE Access*, *11*, 84670–84688. https://doi.org/10.1109/ACCESS.2023.3302348

Neagu, C. D., Avouris, N., Kalapanidas, E., & Palade, V. (2002). Neural and neuro-fuzzy integration in a knowledge-based system for air quality prediction. *Applied Intelligence*, *17*(2), 141–169. https://doi.org/10.1023/A:1016108730534

Oğuz, K., & Pekin, M. A. (2024). Prediction of Air Pollution with Machine Learning Algorithms. *Turkish Journal of Science and Technology*, *19*(1), 1–12. https://doi.org/10.55525/TJST.1224661

Pandithurai, O., Bharathiraja, N., Pradeepa, K., Meenakshi, D., Kathiravan, M., & Vinoth Kumar, M. (2023). Air Pollution Prediction using Supervised Machine Learning Technique. *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, 542–546. https://doi.org/10.1109/ICAIS56108.2023.10073821

Parkavi, P., & Rathi, S. (2021a). Deep Learning Model for Air Quality Prediction Based on Big Data. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 170–175. https://doi.org/10.32628/CSEIT217332

Parkavi, P., & Rathi, S. (2021b). Deep Learning Model for Air Quality Prediction Based on Big Data. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 170–175. https://doi.org/10.32628/cseit217332

Pasupuleti, V. R., Uhasri, Kalyan, P., Srikanth, & Reddy, H. K. (2020). Air Quality Prediction Of Data Log By Machine Learning. *2020 6th International Conference on Advanced*

Computing and Communication Systems (ICACCS), 1395–1399. https://doi.org/10.1109/ICACCS48705.2020.9074431

Ramadan, M. N. A., Ali, M. A. H., Khoo, S. Y., Alkhedher, M., & Alherbawi, M. (2024). Real-time IoT-powered AI system for monitoring and forecasting of air pollution in industrial environment. *Ecotoxicology and Environmental Safety*, *283*. https://doi.org/10.1016/j.ecoenv.2024.116856

Ramli, N., Abdul Hamid, H., Yahaya, A. S., Noor, N. M., & Elena, H. (2023). Short-term Predictions of PM10 Using Bayesian Regression Models. *E3S Web of Conferences*, *437*. https://doi.org/10.1051/E3SCONF/202343701006

Ravindiran, G., Hayder, G., Kanagarathinam, K., Alagumalai, A., & Sonne, C. (2023). Air quality prediction by machine learning models: A predictive study on the indian coastal city of Visakhapatnam. *Chemosphere*, *338*. https://doi.org/10.1016/j.chemosphere.2023.139518

Ren, X., Mi, Z., Cai, T., Nolte, C. G., & Georgopoulos, P. G. (2022). Flexible Bayesian Ensemble Machine Learning Framework for Predicting Local Ozone Concentrations. *Environmental Science & Technology*, *56*(7), 3871–3883. https://doi.org/10.1021/ACS.EST.1C04076

Saikiran, K., Lithesh, G., Srinivas, B., & Ashok, S. (2021). Prediction of Air Quality Index Using Supervised Machine Learning Algorithms. *2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)*, 1–4. https://doi.org/10.1109/ACCESS51619.2021.9563323

Samad, A., Garuda, S., Vogt, U., & Yang, B. (2023). Air pollution prediction using machine learning techniques – An approach to replace existing monitoring stations with virtual monitoring stations. *Atmospheric Environment*, *310*. https://doi.org/10.1016/j.atmosenv.2023.119987

Samadi, S., & Kumawat, A. K. (2023). IoT Enabled Low-Cost Real-Time Remote Air Quality Monitoring and Forecasting System. *2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT)*, 253–258. https://doi.org/10.1109/DICCT56244.2023.10110229

Sharma, S., Tanwar, P., Yadav, A., Kesava Sairam, B., & Jaswal, S. (2022). Critical Review of Air Quality Prediction using Machine Learning Techniques. *2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 1–7. https://doi.org/10.1109/I-SMAC55078.2022.9987252

Ul-Saufie, A. Z., Hamzan, N. H., Zahari, Z., Shaziayani, W. N., Noor, N. M., Zainol, M. R. R. M. A., Sandu, A. V., Deak, G., & Vizureanu, P. (2022). Improving Air Pollution Prediction Modelling Using Wrapper Feature Selection. *Sustainability*, *14*(18). https://doi.org/10.3390/SU141811403

Wardana, I. N. K., Gardner, J. W., & Fahmy, S. A. (2021). Optimising deep learning at the edge for accurate hourly air quality prediction. *Sensors (Switzerland)*, *21*(4), 1–28. https://doi.org/10.3390/s21041064

Zhang, D., & Woo, S. S. (2020). Real Time Localized Air Quality Monitoring and Prediction through Mobile and Fixed IoT Sensing Network. *IEEE Access*, *8*, 89584–89594. https://doi.org/10.1109/ACCESS.2020.2993547

Zhang, Z., Zhang, S., Chen, C., & Yuan, J. (2024). A systematic survey of air quality prediction based on deep learning. In *Alexandria Engineering Journal* (Vol. 93, pp. 128–141). Elsevier B.V. https://doi.org/10.1016/j.aej.2024.03.031

Zhou, Y., Chang, F. J., Chang, L. C., Kao, I. F., & Wang, Y. S. (2019). Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. *Journal of Cleaner Production*, *209*, 134–145. https://doi.org/10.1016/J.JCLEPRO.2018.10.243