International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

Review on Machine Learning Model for Accuracy of Air Quality Prediction

Anju Tirkey
1 Research Scholar, School of Studies in Computer Science & Information Technology
Pt. Ravishankar Shukla University Raipur, (C.G), India

V. K. Patle
2 Professor, School of Studies in Computer Science & Information Technology
Pt. Ravishankar Shukla University Raipur, (C.G), India

Abstract: Air pollution has become a serious environmental and public health issue in many parts of the world. Poor air quality affects human health, climate, and overall quality of life. Therefore, accurate prediction of air quality is very important for taking timely preventive actions. In recent years, machine learning (ML) models have been widely used to improve the accuracy of air quality prediction because they can analyze large and complex datasets more effectively than traditional methods. This review paper focuses on machine learning models used to predict major air pollutants such as particulate matter ($PM_{2.5}$ and PM_{10}), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), and ozone (O_3). Commonly used models including Linear Regression, Support Vector Machines, Decision Trees, Random Forest, K-Nearest Neighbors, Artificial Neural Networks, and Long Short-Term Memory networks are discussed. The paper also reviews different air quality datasets, input parameters, and evaluation metrics such as accuracy, precision, recall, F1-score, MAE, MSE, RMSE, and R^2. In addition, the advantages and limitations of these models are explained. This review aims to help researchers understand current research trends and select suitable machine learning models for more accurate air quality prediction.

Keywords: Air Quality Prediction, Machine Learning, Air Pollution, Performance Evaluation Metrics.

I. INTRODUCTION

Air pollution is one of the most serious environmental problems across the world and has a strong impact on human health, natural ecosystems, and climate conditions. Rapid industrial growth, urbanization, increasing numbers of vehicles, and heavy use of fossil fuels have caused a steady rise in air pollutant levels in both developed and developing countries. Exposure to polluted air leads to severe health issues and is responsible for millions of premature deaths each year, with particulate matter and harmful gases being the main contributors.[1]

In India, air pollution has become a major public health concern, as many cities regularly report pollutant concentrations above the National Ambient Air Quality Standards (NAAQS). Key pollution sources include vehicle emissions, coal-based power plants, industrial processes, construction activities, and biomass burning. Due to rising population density and growing energy demands, air quality conditions continue to worsen. In this context, accurate air quality prediction has become essential for early warning systems and effective pollution control. Machine learning models play a crucial role in improving prediction accuracy by analysing large and complex air quality datasets, making them a valuable tool for air pollution monitoring and management.[2] *A. Air Quality and Major Pollutants*

Air quality refers to the condition of the ambient air in relation to its cleanliness and potential effects on health and the environment. The major air pollutants considered in air quality assessment include PM2.5, PM10, NO_2, SO_2, CO, O_3, ammonia (NH_3), and volatile organic compounds (VOCs). Among these, fine particulate matter (PM2.5) is considered the most harmful due to its ability to penetrate deep into the lungs and bloodstream, leading to chronic health disorders.

B. Air Quality Index (AQI) and Indian AQI Standards

To simplify air quality information for the public, the Air Quality Index (AQI) is widely used. AQI converts complex pollutant concentration data into a single numerical value representing overall air quality and associated health risks. To communicate complex scientific data to the public effectively, authorities use the Air Quality Index (AQI). The AQI is a standardized, unitless metric that translates various pollutant concentrations into a single number representing the overall health risk. While exact scales vary by country (e.g., US, India and China use 0–500, whereas the UK uses 1–10), they generally categorize air quality into levels such as:

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

TABLE 1 RANGE OF AQI

AQI Range	Description	Health Impacts
0–50	Good air quality	Minimal or no impact on health.
51–100	Satisfactory air quality	Breathing difficulty for sensitive groups.
101–200	Moderate air quality	Breathing discomfort for children, elderly people, and individuals with lung or heart diseases.
201–300	Poor air quality	Health effects for longterm exposure; discomfort for heart patients even with short exposure.
301–400	Very poor air quality	Causes respiratory illness with prolonged exposure.
401–500	Severe air quality	Significant health impacts on normal and sensitive groups.

The overall AQI is typically determined by calculating a sub-index for each major pollutant based on its concentration and then identifying the maximum value among them as the final index. This "worst-of" approach ensures that if even one pollutant reaches dangerous levels, the public is adequately warned.

The health consequences of breathing polluted air are severe and wide-ranging. Exposure to high levels of pollutants is a major risk factor for several life-threatening conditions, including stroke, heart disease, lung cancer, and chronic obstructive pulmonary disease (COPD). Fine particulate matter (PM2.5) is particularly dangerous because its small size allows it to penetrate deep into the lungs and enter the bloodstream, causing systemic inflammation.

The table given below outlines the key parameters of air pollution, their primary sources, the related health and environmental effects, as well as the safe exposure limits recommended for each pollutant.

TABLE 2 AIR QUALITY INDEX PARAMETERS AND THEIR EFFECTS AND SAFE EXPOSURE LIMITS

Parameter	Sources	Related Effects	Safe Exposure Limits
Particulate Matter (Pm2.5)	Industrial emissions, vehicular pollution, biomass burning, secondary particles	Deep lung penetration, cardiovascular issues, diabetes risk, premature mortality	60 µg/m^3 (24-hr)
Particulate Matter (Pm10)	Construction activities, road dust, natural hazards, industrial emissions	Respiratory problems, asthma, eye and throat irritation	100 µg/m^3 (24-hr)
Carbon Monoxide (Co)	Automobile emissions, fires, industrial processes.	co poisoning, chest pain, vision issues, smog formation.	2 mg/m^3 (8-hr)
Ozone (O3)	Industry, vehicle emissions, volatile organic compounds.	Reduced lung function, airway inflammation, plant stress.	100 µg/m^3 (8hr)
Nitrogen Dioxide (No$_2$)	Vehicle emissions, fossil fuel combustion, industrial processes.	Lung irritation, smog formation, heart damage.	80 µg/m^3 (24-hr)

Beyond chronic illnesses, air pollution exacerbates acute respiratory issues such as asthma, bronchitis, and pneumonia. It also disproportionately affects vulnerable groups, including children, the elderly, and those with pre-existing conditions. Recent research has even linked air pollution to neurological disorders, adverse pregnancy outcomes, and psychological conditions like anxiety and depression.

C. AQI Prediction using Machine Learning Models

Accurate prediction of the Air Quality Index (AQI) is crucial for effective air pollution control and timely public health warnings. Conventional statistical approaches such as Multiple Linear Regression (MLR) and Auto-Regressive Integrated Moving Average (ARIMA) models have been widely used in early air quality studies. However, recent studies report that these traditional methods often fail to achieve high prediction accuracy because they are unable to effectively capture the complex, non-linear, and time-dependent relationships among

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

air pollutants, meteorological parameters, and urban environmental factors.

In recent years, machine learning (ML) and deep learning (DL) models have gained significant attention for AQI prediction due to their strong ability to learn hidden patterns from large-scale historical air quality datasets without making strict assumptions about data

distribution.[3]

Machine Learning (ML) and Deep Learning (DL) have emerged as superior tools for this task because they can recognize complex patterns in large historical datasets without requiring strict assumptions about data distribution. Common models used in the research listed in the table above include:

- **Classical ML:** Random Forest (RF), Support Vector Regression (SVR), and XGBoost have shown high accuracy in handling multivariate inputs and capturing non-linear trends.[4]
- **Deep Learning:** Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are particularly effective at capturing temporal dependencies in time-series air quality data.
- **Hybrid Models:** Newer architectures combine different models (e.g., CNN-LSTM or EMD-LSTM) to simultaneously extract spatial features and temporal patterns, further enhancing prediction precision.[5]

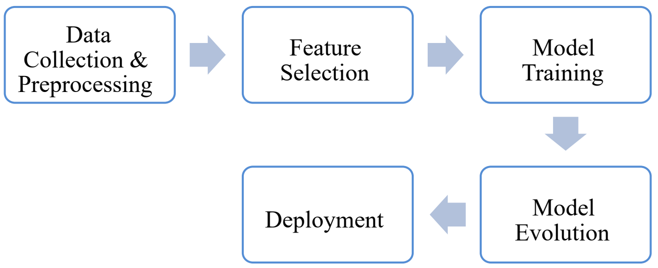

Fig. 1.ML Workflow for Air Quality Monitoring and Forecasting

D. Importance of AQI Prediction

The ability to accurately forecast air quality is vital for several reasons:

- **Public Health Protection:** Accurate forecasts allow health authorities to issue timely advisories, enabling citizens, especially those with respiratory diseases-to take preventive measures like limiting outdoor activity or using air purifiers.
- **Policy and Intervention:** Decision-makers can use predictive insights to implement short-term interventions, such as traffic restrictions or temporary industrial shutdowns, to prevent peak pollution episodes.
- **Sustainable Urban Planning:** Long-term prediction trends help urban planners design smarter cities with better ventilation, more green spaces, and optimized transportation systems to mitigate future pollution risks.
- **Economic Impact:** By reducing the incidence of pollution-related illnesses, accurate forecasting can significantly lower the massive economic burden associated with healthcare costs and lost productivity.

II. LITERATURE REVIEW

This study develops a practical approach to forecast the Air Quality Index (AQI) in Tashkent, Uzbekistan using machine learning models such as KNN, Random Forest, Decision Tree, SVM, and Artificial Neural Networks. Instead of using complex statistical methods, the problem is handled as a classification task to better understand the complicated relationship between air pollutants and weather conditions. The results clearly show that Random Forest and ANN models with the Adam optimizer give more accurate predictions than traditional methods. The study also explains that accurate AQI forecasting is very important because it helps governments issue early health warnings and control industrial pollution in time, especially in urban areas where air pollution patterns are highly non-linear and difficult to predicts.[6]

A comparative study carried out in India examines different machine learning models, including Random Forest, LSTM, ANN, SVM, and K-NN, for air quality prediction. The study uses an air quality dataset from Kaggle and aims to achieve a good balance between prediction accuracy and computation time. The results show that Random Forest and LSTM models perform the best, achieving an accuracy of about 97%, and clearly outperform traditional methods. The study also points out that deep learning techniques are very useful for handling large and complex air quality data, making them suitable for real-time monitoring systems to reduce health risks caused by pollutants such as PM2.5 and NO_2.[7]

This paper studies indoor air quality prediction using machine learning models such as Gradient Boosted Trees and Neural Networks. The focus is on monitoring indoor conditions like CO_2 levels, temperature, and humidity to automatically control ventilation systems. A large dataset of about 31,000 records is used for the analysis. The results show that the Neural Network model performs very well, achieving an accuracy of 98.22%. The study highlights that combining machine learning with IoT devices, such as Raspberry Pi, can help maintain healthy and comfortable indoor environments and reduce health problems caused by poor air circulation and high CO_2 levels in urban homes.[8]

This study examines the importance of data preprocessing in improving the performance of machine learning models using AQI data collected from Delhi. Different methods for filling missing data, such as mean, median, and mode, are tested with

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

models like XGBoost and MLP. The findings show that the Random Forest model performs well, achieving a classification accuracy of 71.59% and an AUC-ROC score of 94.46. The research emphasizes that proper data preprocessing is as important as the choice of model itself. If data is not handled correctly, it can introduce bias and lead to inaccurate air quality predictions, especially in regions with highly variable pollution levels.[9]

This study looks at the use of regression-based machine learning models, such as Decision Tree, Random Forest, and XGBoost, to predict the Air Quality Index (AQI) in Delhi. The analysis is carried out using air quality data collected from the Central Pollution Control Board (CPCB). To understand how the models make their predictions, explainable AI tools like LIME and SHAP are applied. The results show that the Decision Tree Regressor performs the best, achieving a very high R^2 value of 0.9981. The study highlights that transparent and explainable models are important for policymakers, as they help clearly identify key pollutants like PM2.5 and NH_3 that influence air quality decisions.[10]

Addressing urban pollution in New York City, this research compares traditional ML, DL, and hybrid models. The study uses data from 2014-2015 to forecast ozone and sulfur dioxide levels. The hybrid RF-LSTM approach was found to be more robust than standalone models, achieving an R^2 of 0.94. The findings suggest that combining the feature extraction capabilities of RF with the temporal memory of LSTM provides a more reliable tool for municipal interventions and environmental risk management.[11]

This paper investigates AQI classification in urban Bangladesh using a diverse range of models including Naive Bayes, CNN, and RNN. To address class imbalance in the dataset, the authors apply the SMOTE technique. The Decision Tree model achieved a nearperfect accuracy of 99.9%. The study emphasizes that data-driven, cost-effective environmental monitoring is feasible in developing nations when oversampling techniques are used to ensure the model accurately predicts "Severe" and "Poor" air quality categories.[12]

This study integrates GIS-based spatial mapping with supervised learning (SVM and RF) to predict postmonsoon AQI in Southern India. The research analyzes the heterogeneity of pollutants like PM2.5 and PM10 across industrial hubs. SVM performed consistently well across multiple time horizons with an R^2 of 0.914. The integration of spatial interpolation with ML provides a comprehensive view of pollution dynamics, helping authorities identify high-risk zones that require immediate environmental remediation after seasonal rain cycles.[13]

The authors propose a multi-task AQI prediction method based on the NLSTM (Neural Long Short-Term Memory) architecture. Tested in Beijing, the model uses Pearson correlation to assign weights to different pollutant tasks. The NLSTM model achieved an R^2 of 0.94 for CO prediction. This research justifies the use of collaborative multi-dimensional forecasting, showing that predicting multiple pollutants simultaneously improves the overall stability and accuracy of the AQI forecast compared to single-pollutant models.[14]

This comparative analysis focuses on RF, SVC, and KNN for air quality prediction in India. Following the CRISPDM methodology, the study processes over 344,000 instances of historical data. Random Forest demonstrated near-perfect results with a 99.99% accuracy rate. The paper highlights those large-scale historical datasets, when processed through ensemble learning, can provide highly reliable early warnings for urban populations, aiding in the reduction of respiratory and cardiovascular health risks.[15]

This research utilizes satellite data and Google Earth Engine (GEE) to predict AQI over the Korba Coalfield in India. By fusing satellite imagery with ground-based station data, the authors evaluate models like CatBoost and XGBoost. Random Forest emerged as the best performer with an R^2 of 0.93. The study justifies the use of remote sensing for air quality monitoring in industrial hubs where ground stations are sparse, providing a scalable solution for environmental impact assessments in mining regions. [16]

The authors propose an outlier detection framework for

AQI prediction in Jaipur, India. By integrating IQR and Z-score methods with Extra Trees Regressor and RF, the study improves model robustness. The Extra Trees model achieved an R^2 of 0.8884. The research argues that retaining extreme data points through seasonal validation, rather than simply removing them, allows ML models to better predict pollution spikes during festivals or extreme weather events. [17]

This study combines Input Variable Selection (IVS) with hybrid ML models (NARMAX and Decision Trees) to forecast PM levels in Romania. Using a two-year dataset, the authors achieved R^2 values exceeding 0.95. The research identifies the most influential predictor variables for PM2.5 and PM10, emphasizing that feature selection tools are essential for reducing computational complexity while maintaining the high precision required for municipal health alerts.[18]

The authors introduce a 3D Digital Twin environment for Delhi, integrated with CNN-1D and GRU models for AQI monitoring. The CNN-1D model achieved a peak R^2 of 0.9995. By visualizing pollution data within a predictive urban digital twin, the study enables city planners to simulate the impact of urban growth on air quality. This approach represents the next generation of smart city monitoring, combining real-time AI with spatial visualization.[19]

A hybrid temporal CNN-LSTM-MHA+GRU model is proposed for hourly AQI prediction in Visakhapatnam, India. Optimized through Bayesian tuning, the model achieved an R^2 of 0.9757. The framework effectively captures both short-term patterns and long-term dependencies, proving highly accurate during pollution extremes. This research justifies the use of complex multi-head attention mechanisms to refine residual errors in time-series data, providing a benchmark for industrial air quality forecasting.[20]

This study uses Geographically Weighted Regression (GWR) and spatial observation data to predict global AQI levels. Addressing the uneven distribution of monitoring stations, the authors achieved an R^2 of 0.74. The research highlights the challenges of global-scale modeling where data density varies. By integrating satellite data with geographics machine learning, the study provides a framework for estimating air quality in regions lacking ground-level monitoring infrastructure.[21]

Focused on Lahore, Pakistan, this paper uses Random Forest and Decision Trees to link AQI prediction with sustainable urban planning. The models achieved approximately 98% accuracy. The research emphasizes that air pollution is the leading environmental cause of premature fatalities and that ML-based time-series forecasting is essential for designing breathable urban spaces. The study advocates for data-driven policies to combat the industrial and transportation emissions plaguing growing megacities.[22]

The authors develop a global AQI model by combining remote sensing and ground-based station data using linear regression and ML. The study focuses on expanding the spatial coverage of air pollution measurements. While global datasets present noise challenges, the research establishes a baseline for using satellite data as a surrogate for ground measurements. This work is pivotal for international environmental advances, providing a method to monitor transboundary pollution

effectively.[23]

This research forecasts monthly AQI in Delhi using a 30year dataset (1987-2020) and models like Random Forest and Decision Tree Regression. RF emerged as the most reliable model with an R^2 of 0.92. The study emphasizes the importance of long-term trend analysis for understanding the efficacy of air pollution control policies over decades. The findings suggest that seasonal crossvalidation is necessary to ensure models remain accurate despite long-term shifts in urban emission patterns.[24]

Conducted in Azamgarh, India, this study uses real-time hourly data to forecast AQI using RF and XGBoost. The authors calculated pollutant sub-indices for a complete year (8,760 data points). The models achieved an R^2 of approximately 0.95. This paper highlights the dominance of particulate matter in rural-urban transition zones, proving that localized ML models can provide precise hourly warnings that help communities manage exposure to hazardous air.[25]

This study evaluates Ridge Regression, SVR, and XGBoost for AQI prediction and health risk assessment using the AirQ+ tool. Ridge Regression achieved a high R^2 of 0.91 for PM10. The research is unique for its direct integration of ML output into health risk software, providing quantifiable data on the health impacts of shortterm pollutant trends. This approach justifies ML as a tool for public health officials to communicate the immediate dangers of pollution spikes.[26]

The authors develop a Multiple Linear Regression (MLR) model for IoT applications to predict AQI influenced by meteorological parameters. The study focuses on the correlation between temperature, humidity, and pollutants like NOX and SO2. By creating a multivariant regressive function, the research provides a lightweight solution for real-time monitoring on resource-constrained IoT devices, making it suitable for distributed sensor networks in industrial zones.[27]

A novel stacking ensemble method is examined for PM2.5 estimation in Beijing and Istanbul. The model combines predictions from multiple ML algorithms to achieve an R^2 of 0.99. The research proves that ensemble methods are more stable than single models when dealing with diverse urban environments. The findings suggest that stacking is a highly effective strategy for creating generalized air quality models that work across different geographic and climatic conditions.[28]

The authors introduce a hybrid EMD-LSTM model optimized via Bayesian and random search techniques for AQI prediction in Bhopal. The Empirical Mode Decomposition (EMD) is used to handle data nonlinearity, while Bayesian optimization tunes the LSTM hyperparameters. The model achieved a low MAE of 0.385. The study demonstrates that hybrid deep learning, when properly optimized, significantly outperforms standard architectures in predicting highly volatile atmospheric variables.[29]

This study compares shallow learning (RF, SVM) against deep learning (CNN, LSTM) for AQI forecasting in Zabol, Iran-a region prone to dust storms. Surprisingly, the CNN model outperformed others with 60% accuracy in classifying dust-polluted air. The research justifies the use of deep learning for extreme environmental events, showing that spatial feature extraction from meteorological data is crucial for predicting AQI in regions dominated by natural particulate matter.[31]

Focused on Azamgarh, this study benchmarks XGBoost against other ensemble models for hourly AQI prediction. XGBoost achieved the best performance ($R^2 > 0.99$) and the fastest execution time (1.61s). The research emphasizes that for real-time warning systems, speed and accuracy are equally important. XGBoost's efficiency in handling large datasets makes it the ideal candidate for deploying large-scale, real-time air quality monitoring networks in developing urban centers.[32]

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

This version of the Delhi study emphasizes the 99% accuracy achieved by Decision Tree models during training. Using a 30-year monthly dataset, the research provides a longitudinal view of Delhi's air quality. The findings show that while SO2 levels have decreased due to policy interventions, PM2.5 remains a critical challenge. The study proves that ML models can serve as effective audit tools for long-term environmental regulations.[33]

This systematic review compares nine ML/DL algorithms for AQI prediction in Bengaluru, India. The study highlights that traditional models struggle with temporal dependencies, while a hybrid Bi-LSTM with 1D-CNN significantly outperforms them. By capturing both spatial features and temporal sequences, the hybrid model achieves over 90% accuracy. The research justifies the transition toward complex deep learning architectures for urban environments where pollution levels fluctuate rapidly due to traffic and weather.[34]

This study benchmarks eleven regressors, including Lasso and Elastic Net, for daily AQI forecasting in Hapur, India. Random Forest achieved the highest R^2 of 0.9987. The authors use SHAP analysis to identify PM2.5 and PM10 as the dominant features. The paper provides a clear hierarchy of model performance, proving that ensemble techniques are consistently superior to linear and simple non-linear models for predicting complex atmospheric indices.[35]

This review paper analyzes SVM and KNN for AQI prediction in urban industrial regions. By meta-analyzing several international research works, the authors conclude that hybrid models (multi-algorithm integration) are the future of the field. The study emphasizes that larger datasets are essential for training generalized models that can work across different industrial zones. This research serves as a justification for the ongoing evolution toward more complex, multi-layered air quality forecasting systems.[36]

Focused on the Himalayan city of Dehradun, this study compares Lasso, RF, and XGBoost for AQI prediction. Lasso regressor unexpectedly performed best with an R^2 of 0.9999 and low MAPE (0.0269). The research highlights that in specific geographical terrains like foothills, simpler regularized linear models can occasionally outperform complex ensembles. This study provides vital insights for policymakers in Himalayan regions to implement pre-emptive measures to protect public health and ecology. [37]

III. LIMITATIONS AND FUTURE WORK

Although Machine Learning (ML) and Deep Learning (DL) models provide better AQI prediction than traditional methods, some challenges still remain. One major problem is the lack of complete and high-quality data. Missing values, sensor failures, and limited monitoring stations can reduce prediction accuracy. In addition, complex models such as LSTM and ensemble methods may perform well during training but can give lower accuracy when tested on new locations or different weather conditions. Another challenge is that many deep learning models are difficult to understand, making their predictions less transparent.

In the future, researchers should develop hybrid and ensemble models that combine the advantages of different algorithms to improve prediction accuracy. Using both satellite data and ground monitoring data can help fill missing information and provide better coverage. Better hyperparameter tuning methods can also improve model performance. Finally, Explainable Artificial Intelligence (XAI) should be used to make prediction models more transparent, reliable, and easier to understand for researchers and decision-makers.

IV. CONCLUSION

This study shows that machine learning techniques are effective for improving Air Quality Index (AQI) prediction. Among the models reviewed, the Random Forest (RF) algorithm achieved the best overall performance with approximately 99% accuracy, making it the most reliable model for AQI prediction. The results also indicate that proper data preprocessing, feature selection, and model tuning are important for achieving high prediction accuracy. In the future, integrating data from ground-based monitoring stations, satellite observations, and hybrid machine learning models can further improve prediction performance. These advancements can help provide timely air quality warnings, support environmental monitoring, and assist policymakers in protecting public health and planning smarter cities.

REFERENCES

[1] A. K. Gond, A. Jamal, and T. Verma, “Developing a machine learning model using satellite data to predict the Air Quality Index (AQI) over Korba Coalfield, Chhattisgarh (India),” *Atmos. Pollut. Res.*, vol. 16, no. 2, Feb. 2025, doi: 10.1016/j.apr.2024.102398.

[2] K. Naveed *et al.*, “Machine learning assisted predictive urban digital twin for intelligent monitoring of air quality index for smart city environment,” *Environmental Modelling and Software*, vol. 192, Aug. 2025, doi: 10.1016/j.envsoft.2025.106559.

[3] J. Bernacki and R. Scherer, “A Comprehensive Review of Data-Driven Techniques for Air Pollution Concentration Forecasting,” Oct. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/s25196044.

[4] A. K. Upadhyay and S. R. Shah, “Machine Learning-Based Prediction of Air Quality Index (AQI) in Mumbai: Comparative Analysis of Linear Regression, Random Forest, and XGBoost Models,” *RESEARCH*

REVIEW International Journal of Multidisciplinary, vol. 10, no. 11, pp. 299–307, Nov. 2025, doi: 10.31305/rrijm.2025.v10.n11.030.

[5] A. Gilik, A. S. Ogrenci, and A. Ozmen, "Air quality prediction using CNN+LSTM-based hybrid deep learning architecture," *Environmental Science and Pollution Research*, vol. 29, no. 8, pp. 11920–11938, Feb. 2022, doi: 10.1007/s11356-021-16227-w.

[6] S. Akhatkulov and I. Yalgoshev, "A Machine Learning Framework for Forecasting Air Quality Index," in *Proceedings - 2025 International Russian Smart Industry Conference, SmartIndustryCon 2025*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 598– 602. doi: 10.1109/SmartIndustryCon65166.2025.109859 62.

[7] S. Dhongade, S. Tiwaskar, A. Koturwar, R. Motiramani, and D. Zarekar, "Air Predictive Modeling for Air Quality: A Comparative Study of Machine Learning and Deep Learning Techniques," in *2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science, SCEECS 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 0.1109/SCEECS61402.2024.10482082.

[8] A. Gururaj, V. Nagaraj, A. S. Rajesh, K. N. Achyuth, M. U. Somesh, and B. Ashish Dubay, "Air Quality Prediction and Analysis using Machine Learning," in *Proceedings - 2024 5th International Conference on Mobile Computing and Sustainable Informatics, ICMCSI 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 543–550. doi: 10.1109/ICMCSI61536.2024.00086.

[9] S. P. Suhas, U. P. Jakarbet, B. P. Aparna, A. A. Khan, B. P. Ashwini, and R. M. Savithramma, "Analysing the Impact of Data Preprocessing on ML Model Performance Using AQI Data," in *3rd IEEE International Conference on Networks, Multimedia and Information Technology, NMITCON 2025*, Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/NMITCON65824.2025.11187526.

[10] K. V. Reddy, C. M. Begum, A. Neeraja, and C. Jyotsna, "Enhancing Air Quality Index Forecasting through Machine Learning Models," in *Proceedings of 2025 3rd International Conference on Intelligent Systems, Advanced Computing, and Communication, ISACC 2025*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 1188–1194. doi: 10.1109/ISACC65211.2025.10969305.

[11] Lord Coffie, E. B. Obeng, M. D. Afrane, J. Kim, and Y. Xu, "Forecasting Air Quality Index (AQI) Using Machine Learning Techniques," in *2025 IEEE/ACIS 23rd International Conference on Software Engineering Research, Management and Applications, SERA 2025 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 431–437. doi: 10.1109/SERA65747.2025.11154585.

[12] P. Md. Sakibur Rahman *et al.*, "A Comparative Study of Machine Learning Models for Air Quality Index Classification in Urban Bangladesh," in *2025 International Conference on Electrical, Computer and Communication Engineering, ECCE 2025*, Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/ECCE64574.2025.11013895.

[13] V. L. Sivakumar, K. Sundarayamini, and A. Murugan, "Integrating Machine Learning with Station-Level Pollutant Data for Post-Monsoon AQI Prediction in Southern India," in *2025 10th International Conference on Applying New Technology in Green Buildings, ATiGB 2025*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 797–802. doi: 10.1109/ATIGB66719.2025.11142200.

[14] W. Li, Y. Zhang, R. Ma, Y. Liu, and J. Jia, "Multi-Task Aqi Prediction Method Based on NLSTM," in *10th International Conference on Computer and Communication Systems, ICCCS 2025*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 132–136. doi: 10.1109/ICCCS65393.2025.11069571.

[15] F. Musleh, R. Taha, and A. R. Musleh, "Predicting Air Quality Using Machine Learning: A Comparative Analysis," in *2nd International Conference on IT Innovations and Knowledge Discovery, ITIKD 2024*, Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/ITIKD63574.2025.11005241.

[16] A. K. Gond, A. Jamal, and T. Verma, "Developing a machine learning model using satellite data to predict the Air Quality Index (AQI) over Korba Coalfield, Chhattisgarh (India)," *Atmos. Pollut. Res.*, vol. 16, no. 2, Feb. 2025, doi: 10.1016/j.apr.2024.102398.

[17] P. K. Dongre, V. Patel, U. Bhoi, and N. N. Maltare, "An outlier detection framework for Air Quality Index prediction using linear and ensemble models," *Decision Analytics Journal*, vol. 14, Mar. 2025, doi: 10.1016/j.dajour.2025.100546.

[18] M. T. Udristioiu, Y. EL Mghouchi, and H. Yildizhan, "Prediction, modelling, and forecasting of PM and AQI using hybrid machine learning," *J. Clean. Prod.*, vol. 421, Oct. 2023, doi: 10.1016/j.jclepro.2023.138496.

[19] K. Naveed *et al.*, "Machine learning assisted predictive urban digital twin for intelligent monitoring of air quality index for smart city environment," *Environmental Modelling and Software*, vol. 192, Aug. 2025, doi: 10.1016/j.envsoft.2025.106559.

[20] S. T and M. R. G, "Accurate hourly AQI prediction using temporal CNN-LSTM-MHA+GRU: A case study of seasonal variations and pollution extremes in Visakhapatnam, India," *Results in Engineering*, vol. 27, Sep. 2025, doi: 10.1016/j.rineng.2025.106303.

[21] T. S. Anggraini, H. Irie, A. D. Sakti, and K. Wikantika, "Global air quality index prediction using integrated spatial observation data and geographics machine learning," *Science of Remote Sensing*, vol. 11, Jun. 2025, doi: 10.1016/j.srs.2025.100197.

[22] M. A. Mujtaba *et al.*, "Using machine learning for air quality prediction and sustainable urban planning," *Sustainable Futures*, vol. 10, Dec. 2025, doi: 10.1016/j.sftr.2025.100981.

[23] T. S. Anggraini, H. Irie, A. D. Sakti, and K. Wikantika, "Machine learning-based global air quality index development using remote sensing and ground-based stations," *Environmental Advances*, vol. 15, Apr. 2024, doi: 10.1016/j.envadv.2023.100456.

[24] C. B. Pande *et al.*, "Forecasting of monthly air quality index and understanding the air pollution in the urban city, India based on machine learning models and cross-validation," *J. Atmos. Chem.*, vol. 82, no. 1, Jun. 2025, doi: 10.1007/s10874-024-09466-x.

[25] A. Ansari and A. R. Quaff, "Advanced Machine Learning Techniques for Precise hourly Air Quality Index (AQI) Prediction in Azamgarh, India," *Int. J. Environ. Res.*, vol. 19, no. 1, Feb. 2025, doi: 10.1007/s41742-024-00684-5.

[26] E. Koçak, "Comprehensive evaluation of machine learning models for real-world air quality prediction and health risk assessment by AirQ+," *Earth Sci. Inform.*, vol. 18, no. 3, Sep. 2025, doi: 10.1007/s12145-025-01941-7.

[27] S. Sigamani and R. Venkatesan, "Air quality index prediction with influence of meteorological parameters using machine learning model for IoT application," *Arabian Journal of Geosciences*, vol. 15, no. 4, Feb. 2022, doi: 10.1007/s12517-022-09578-2.

[28] M. Emeç and M. Yurtsever, "A novel ensemble machine learning method for accurate air quality prediction," *International Journal of Environmental Science and Technology*, vol. 22, no. 1, pp. 459–476, Jan. 2025, doi: 10.1007/s13762-024-05671-z.

[29] V. Kushwah and P. Agrawal, "Hybrid model for air quality prediction based on LSTM with random search and Bayesian optimization techniques," *Earth Sci. Inform.*, vol. 18, no. 1, Jan. 2025, doi: 10.1007/s12145-024-01514-0.

[30] Y. Özüpak, F. Alpsalaz, and E. Aslan, "Air Quality Forecasting Using Machine Learning: Comparative Analysis and Ensemble Strategies for Enhanced Prediction," *Water Air Soil Pollut.*, vol. 236, no. 7, Jul. 2025, doi: 10.1007/s11270025-08122-8.

[31] E. Kalantari, H. Gholami, H. Malakooti, A. R. Nafarzadegan, and V. Moosavi, "Machine learning for air quality index (AQI) forecasting: shallow learning or deep learning?," *Environmental Science and Pollution Research*, vol. 31, no. 54, pp. 62962–62982, Nov. 2024, doi: 10.1007/s11356-024-35404-1.

[32] A. Ansari and A. R. Quaff, "Advanced Machine Learning Techniques for Precise hourly Air Quality Index (AQI) Prediction in Azamgarh, India,"

Int. J. Environ. Res., vol. 19, no. 1, Feb. 2025, doi: 10.1007/s41742-024-00684-5.

[33] C. B. Pande *et al.*, "Forecasting of monthly air quality index and understanding the air pollution in the urban city, India based on machine learning models and cross-validation," *J. Atmos. Chem.*, vol. 82, no. 1, Jun. 2025, doi: 10.1007/s10874-024-09466-x.

[34] A. Iqbal and N. Mukherjee, "A Systematic Review and Comparative Study of Machine Learning Techniques for Air Quality Prediction," *Water Air Soil Pollut.*, vol. 236, no. 12, Dec. 2025, doi: 10.1007/s11270-025-08410-3.

[35] A. Ansari and A. R. Quaff, "Comparative benchmarking of eleven machine learning regressors for daily air quality index (AQI) forecasting: Sensitivity analysis and SHapley Additive exPlanations (SHAP)-Based interpretability framework," *Theor. Appl. Climatol.*, vol. 156, no. 11, Nov. 2025, doi: 10.1007/s00704-025-05832-1.

[36] T. Desai *et al.*, "Comparative analysis of machine learning algorithms for air quality index prediction," *Machine Learning for Computational Science and Engineering*, vol. 1, no. 1, Jun. 2025, doi: 10.1007/s44379-02500014-2.

[37] 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*, vol. 121, no. 5, pp. 5821–5847, Mar. 2025, doi: 10.1007/s11069024-07027-9.

[38] P. Md. Sakibur Rahman *et al.*, "A Comparative Study of Machine Learning Models for Air Quality Index Classification in Urban Bangladesh," in *2025 International Conference on Electrical, Computer and Communication Engineering, ECCE 2025*, Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/ECCE64574.2025.11013895.

[39] C. B. Pande *et al.*, "Forecasting of monthly air quality index and understanding the air pollution in the urban city, India based on machine learning models and cross-validation," *J. Atmos. Chem.*, vol. 82, no. 1, Jun. 2025, doi: 10.1007/s10874-024-09466-x.

[40] A. Ansari and A. R. Quaff, "Data-driven analysis and predictive modelling of hourly Air Quality Index (AQI) using deep learning techniques: a case study of Azamgarh, India," *Theor. Appl. Climatol.*, vol. 156, no. 1, Jan. 2025, doi: 10.1007/s00704-024-05304-y.