e-ISSN: 3023-7343

TETHYS ENVIRONMENTAL SCIENCE
an International Journal

Predicting Air Quality Index Based on Rainfall Patterns: A Machine Learning Approach with Mathematical Modelling

Pavendhan Madhiva[1]*, Mallemadugula S. S. R. K. N. Sarma[1], Krishna Kumari[2]

[1]Gandhi Institute of Technology and Management (GITAM), Department of Physics, 530001 Vishakhapatnam, Andra Pradesh, India. [2]Presidency University, Department of Mathematics, 560119 Bengaluru, Karnataka, India.

Research Article

Citation: Madhiva, P., Sarma, M. S. S. R. K. N., Kumari, K. (2025). Predicting Air Quality Index Based on Rainfall Patterns: A Machine Learning Approach with Mathematical Modelling. *Tethys Env. Sci.* 2(2): 77-89.

DOI: 10.5281/zenodo.15662742

Received: 25 April 2025

Accepted: 3 June 2025

Available Online: 14 June 2025

Publication Date: 23 June 2025

Abstract

This study investigates the relationship between rainfall patterns and air quality index (AQI) in the Indian subcontinent using machine learning techniques. We developed a predictive model that incorporates rainfall data, including intensity, duration, and frequency, to forecast AQI values. Methodologically, rainfall and AQI data have been collected and preprocessed from various sources, including weather stations and air quality monitoring stations. These data were used to train and test the machine learning model, which was optimized using various techniques such as feature engineering and hyperparameter tuning. The model's performance was evaluated using several metrics, including absolute mean error, root mean square error, and coefficient of determination. The proposed model demonstrated high accuracy in predicting AQI values, outperforming traditional statistical models. Our findings demonstrate that the predictive model can accurately forecast AQI values up to three days in advance, offering valuable insights for air quality management and policymaking, while also highlighting the significant influence of rainfall patterns, where heavy rainfall events improve air quality and dry periods lead to deterioration. The study underscores the critical role of machine learning-based models in environmental monitoring and prediction, suggesting that accurate AQI forecasts not only advance research in this field but also have vital implications for public health by helping mitigate the adverse effects of air pollution on human health.

Keywords: *Air quality index, rainfall patterns, machine learning, predictive modeling, environmental monitoring.*

Correspondence: Pavendhan Madhiva
pmadhiva@gitam.in

Introduction

Air quality directly affects human health, environmental sustainability, and economic development. Poor air quality, characterized by high levels of pollutants such as PM2.5, PM10, NO_2, SO_2, and O_3, is a major concern in urbanized and industrialized areas (Gupta and Christopher, 2009). Monitoring and predicting the Air Quality Index (AQI) is crucial for providing early warnings, protecting public health, and implementing effective environmental policies (Zhang et al., 2020). Air pollution contributes to severe health issues, including respiratory and cardiovascular diseases, as well as premature mortality (WHO, 2021). Hence, AQI prediction models based on meteorological factors can help individuals take precautions and enable policymakers to implement pollution control strategies promptly (Gao et al., 2019). The integration of machine learning and mathematical modeling enhances AQI prediction accuracy, strengthens environmental monitoring systems, facilitates the anticipation of high-risk pollution periods for urban and industrial planning, and improves the effectiveness of measures such as traffic restrictions and emission controls (Fan et al., 2020; Zhao et al., 2022).

Rainfall is a natural process that helps reduce air pollution by cleansing airborne pollutants. This process is particularly effective for particulate matter, as raindrops capture and remove suspended particles from the air (Zhao et al., 2021). However, the impact of rainfall on air quality depends on various factors, including its intensity, duration, and frequency (Chen et al., 2018). For instance, Huang et al. (2019) reported that heavy rainfall could reduce PM levels by up to 40%, whereas light rainfall has a limited effect. Additionally, other meteorological conditions, such as wind speed and humidity, can significantly influence the pollutant-cleansing effect of rainfall (Zheng et al., 2021). This complexity highlights the challenges of predicting AQI based solely on meteorological data, emphasizing the need for more advanced modeling techniques.

Machine learning (ML) has become a key tool in predicting the Air Quality Index (AQI) due to its ability to process large datasets and identify complex relationships. Techniques such as decision trees, support vector machines (SVM), and artificial neural networks have demonstrated high accuracy in AQI prediction based on meteorological and environmental variables like temperature, humidity, wind speed, and pollutant levels (Fan et al., 2020). Among the most used algorithms are Random Forest, XGBoost, and Long Short-Term Memory (LSTM) networks, which outperform traditional regression models by capturing nonlinear relationships and adapting to changing environmental conditions (Wang et al., 2022). However, integrating more complex meteorological factors like rainfall into these models remains limited (Zheng et al., 2021). Mathematical modeling plays a complementary role in understanding pollutant dispersion and the impact of factors like rainfall on air quality. Theoretical approaches such as the Gaussian plume model and box models help explain pollutant movement and deposition based on parameters like wind speed, atmospheric pressure, and rainfall intensity (Jiang et al., 2019). Recent studies suggest that combining these mathematical models with machine learning can improve AQI prediction accuracy. For example, Zheng et al. (2021) enhanced AQI forecasts using a hybrid approach that incorporates real-time meteorological data. However, further advancements are needed to better model the effects of rainfall on air quality.

The relationship between rainfall and AQI is complex and not adequately considered in existing predictive models. While the mitigating effect of rainfall on air pollution is well established (Huang et al., 2019), robust models that quantitatively predict AQI based on variables such as rainfall intensity, duration, and frequency have yet to be developed (Xu et al., 2020). Current AQI prediction models primarily focus on pollutant concentrations, often excluding meteorological factors like rainfall and wind (Jiang et al., 2019; Zhao et al., 2022). This limitation reduces the effectiveness of strategic decisions related to urban planning, pollution control, and public health (Tang et al., 2021). Additionally, while machine learning techniques offer high accuracy in AQI forecasting, hybrid approaches that integrate these methods with mathematical models incorporating physical pollutant dispersion and natural removal processes like rainfall remain scarce in the literature (Wang and Sun, 2021).

This study aims to examine the effect of rainfall intensity and frequency on AQI and to develop a machine learning-based AQI prediction model based on this relationship, thus contributing to the development of artificial intelligence-supported environmental management and public health policies.

Material and Methods

Data collection and preprocessing

AQI and meteorological data, including rainfall patterns, were obtained from official sources within the Indian subcontinent, covering urban atmospheric conditions characteristic of this geographical region. AQI data included hourly or daily PM2.5, PM10, NO_2, SO_2, O_3, and CO concentrations obtained from local air quality monitoring stations. Rainfall data were obtained from national meteorological stations; information on the intensity, duration, and frequency of rainfall was collected.

In the data preprocessing process, missing or erroneous data were first identified and completed with interpolation methods or mean/mode filling techniques (Little and Rubin, 2002). Numerical data (e.g. pollutant concentrations and rainfall amount) were normalized in the range of 0–1 or –1 to 1 to prevent any variable from becoming dominant during model training (Han et al., 2011). In addition, outliers were determined and removed from the data set by statistical methods to prevent extreme values from corrupting the data set.

Feature selection and engineering

Feature selection was performed to determine the meteorological variables that were effective on AQI. According to the literature, rainfall intensity and frequency, humidity, wind speed, temperature, and atmospheric pressure were considered as the main variables. These data were derived from raw data obtained from local meteorological stations. Basic techniques such as time series decomposition were applied to reveal seasonal and cyclical patterns (Hyndman and Athanasopoulos, 2018).

Correlation analysis was performed to determine the relationships between variables and AQI levels. Significant linear relationships were evaluated with the help of the Pearson correlation coefficient and correlation matrix (Field, 2013). Based on the analysis results, the most impactful variables were identified and utilized for model training.

Machine learning models

Various supervised machine-learning algorithms were used to predict AQI based on rainfall patterns and other meteorological variables. Random Forest was applied as a method that constructs multiple decision trees and averages predictions, providing high accuracy and effectively handling large datasets (Breiman, 2001). XGBoost was chosen as a fast and powerful boosting algorithm capable of handling imbalanced datasets and missing values efficiently (Chen and Guestrin, 2016). The Long Short-Term Memory (LSTM) model, known for capturing long-term dependencies, was utilized for time series forecasting (Hochreiter and Schmidhuber, 1997). Additionally, Convolutional Neural Networks (CNN), traditionally developed for visual data, were applied to time series forecasting due to their ability to learn hierarchical features from sequential data (LeCun et al., 2015). Each model was tested on training and validation datasets, and hyperparameter tuning was optimized using methods such as grid search and random search (Bergstra and Bengio, 2012).

Mathematical modeling

Alongside machine learning, a regression-based model was developed to predict AQI by considering meteorological factors, particularly rainfall intensity, frequency, and duration. Differential equations were also formulated to describe the dispersion and removal of pollutants under the influence of rainfall, incorporating factors like emission rates, atmospheric stability, wind speed, and rainfall intensity. The washout effect—the removal of pollutants by rainfall—was included in these equations (Zhang et al., 2021), which were solved using numerical methods such as finite differences or Runge-Kutta methods to simulate real-time contaminant distribution and predict AQI fluctuations.

The hybrid prediction model was defined as $AQI\_pred = f\,(DM\,(x, t), ML\,(x, t))$, where the predicted AQI (AQI_pred) at location x and time t is a function of the dispersion model output ($DM\,(x, t)$) (e.g. CMAQ, WRF-Chem) and the machine learning model output ($ML\,(x, t)$) (e.g. random forest, neural network). Here, the dispersion model output represents the simulated pollutant concentrations based on emissions and meteorological data, while the machine learning model output predicts AQI using real-time weather and air quality data.

The machine learning model was formulated as $ML\,(x, t) = g\,(Weather\,(x, t), AQ\,(x, t))$, where $Weather\,(x, t)$ includes variables such as temperature, humidity, wind speed, and $AQ\,(x, t)$ includes pollutant concentrations like PM2.5 and NO_2 at location x and time t. The function g is the relationship learned by the machine learning algorithm.

The dispersion model was expressed as $DM\,(x, t) = h\,(Emissions\,(x, t), Meteorology\,(x, t))$, where $Emissions\,(x, t)$ represent emission data (e.g., NOx, VOCs), and $Meteorology\,(x, t)$, includes meteorological parameters such as wind direction and temperature. The function h models the physical transport and transformation of pollutants in the atmosphere.

Model evaluation and validation

Various performance metrics were used to evaluate the model's effectiveness. These included Root Mean Squared Error (RMSE), which calculates the square root of the mean of squared differences between predicted and actual AQI values, giving greater weight to larger errors (Willmott et al., 2012); Mean Absolute Error (MAE), which provides an easily interpretable measure by averaging absolute error values (Yang et al., 2018); and the R2 score, which indicates how well the model explains AQI variability (Hastie et al., 2009). Additionally, techniques such as k-fold cross-validation were applied to assess the model's generalization ability and robustness across different data subsets (James et al., 2013). This approach helps prevent overfitting specific training data, ensuring a more reliable performance evaluation.

Implementation framework

The tools and technologies used in this study formed an application framework encompassing data processing, model development, and simulation processes. Python was chosen as the primary programming language for data preprocessing, feature extraction, and model implementation. In this context, libraries such as Pandas, NumPy, and Matplotlib were utilized for data manipulation, statistical analysis, and visualization (McKinney, 2010). For the development and training of deep learning models like LSTM and CNN, the open-source deep learning library TensorFlow was employed (Abadi et al., 2015). Meanwhile, machine learning algorithms such as Random Forest and XGBoost were implemented using the Scikit-learn library (Pedregosa et al., 2011). Additionally, MATLAB software was used for solving differential equations governing pollutant dispersion and applying regression models through numerical simulation (MathWorks, 2021). These tools facilitated the effective integration of both data-driven and physics-based components of the hybrid model.

Result and Discussion

Comparison of different machine-learning approaches and mathematical modeling techniques revealed the critical role of rainfall and other meteorological variables in AQI estimation. The success levels of the models showed significant differences depending on both seasonal and regional conditions. Hybrid approaches produced more consistent estimates with lower error margins than traditional methods. The effect of rainfall on air quality has a more complex structure when evaluated together with factors such as traffic density, industrial emissions, and wind. This demonstrates the need for multivariate models that take regional dynamics into account.

Model performance analysis

The performance of different machine learning models in predicting AQI using rainfall patterns and other meteorological parameters was evaluated. The models examined include Random Forest (RF), XGBoost, Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN). The results revealed that XGBoost outperformed other models with the lowest Root Mean Square Error (RMSE) and the highest R^2 score, proving its effectiveness in predicting AQI with structured meteorological data (Chen and Guestrin, 2016). The LSTM model showed promising results in capturing long-term dependencies in time series forecasting; however, it required more processing power (Hochreiter and Schmidhuber, 1997).

Mathematical modeling was integrated into the process as a multivariate regression-based approach supported by differential equations including pollutant dispersion and rainfall cleaning effect. This hybrid approach increased the prediction accuracy compared to models based solely on machine learning, especially in scenarios with heavy rainfall. Compared to traditional models that have difficulty capturing the dynamic effect of rainfall on pollutant levels, the hybrid model reduced the error rate by approximately 15%. This highlights the importance of integrating physical air dispersion models into forecasting analyses (Wang et al., 2022). The daily dataset used in the study is shown in Table 1.

Table 1. Sample dataset showing daily precipitation, pollutant concentrations, and AQI values along with meteorological parameters.

Date	Rainfall (mm)	PM2.5 (µg/m³)	PM10 (µg/m³)	NO_2 (ppb)	O_3 (ppb)	SO_2 (ppb)	CO (ppm)	Temperature (°C)	Humidity (%)	Wind Speed (m/s)	AQI
2025-01-01	0	45	85	40	50	8	0.3	28	70	3	150
2025-01-02	10	30	60	35	48	7	0.2	30	65	2.5	120
2025-01-03	5	40	70	38	52	6	0.3	29	68	3.2	140
2025-01-04	20	20	40	30	45	5	0.1	25	75	2.8	90
2025-01-05	0	50	90	45	55	9	0.4	32	60	3.5	160
2025-01-06	15	35	65	36	50	7	0.2	27	72	2.7	130
2025-01-07	8	38	72	37	53	6	0.3	28	69	3	135
2025-01-08	0	55	95	50	60	10	0.5	31	58	4	170
2025-01-09	12	25	50	32	46	6	0.2	26	78	2.2	110
2025-01-10	18	22	45	28	43	5	0.1	24	80	1.8	95

The 10-day sample data used in the study reveals the interactions between air pollutants and meteorological variables. While AQI values ranged between 150-170 on days when there was no rain, a significant decrease was observed in AQI levels on days when rainfall was high. In particular, 20 mm of rainfall was recorded on January 4, and the AQI value dropped to 90. Similarly, PM2.5 and PM10 concentrations also showed parallel changes with AQI; when PM2.5 reached 55 µg/m³ on January 8, AQI was measured as 170. It is observed that pollutants accumulate more in the atmosphere on days when wind speed is low; humidity and temperature are especially effective on gaseous pollutants such as ozone. These findings reveal that precipitation improves air quality by removing particulate matter and that its effect on AQI should be evaluated together with other meteorological factors.

Impact of rainfall on AQI variability

The relationship between rainfall and AQI has been studied under different seasonal conditions, and distinct patterns in air quality changes have been revealed. During the monsoon season, heavy rainfall significantly reduced PM2.5 and PM10 concentrations, and sharp decreases in AQI values were observed accordingly. Also, despite occasional rainfall in winter, AQI remained high. In urban areas with heavy traffic, although rainfall temporarily lowered AQI, air pollution quickly returned to its previous levels due to continuous vehicle emissions after rainfall (Chen et al., 2018). In industrial areas, the effect of rainfall on AQI was limited due to continuous emissions from factories. This suggests that rainfall is effective in reducing particulate matter, but it is not sufficient to combat continuous pollution sources (Xu et al., 2020). An inverse relationship between AQI and rainfall is shown in Figure 1.

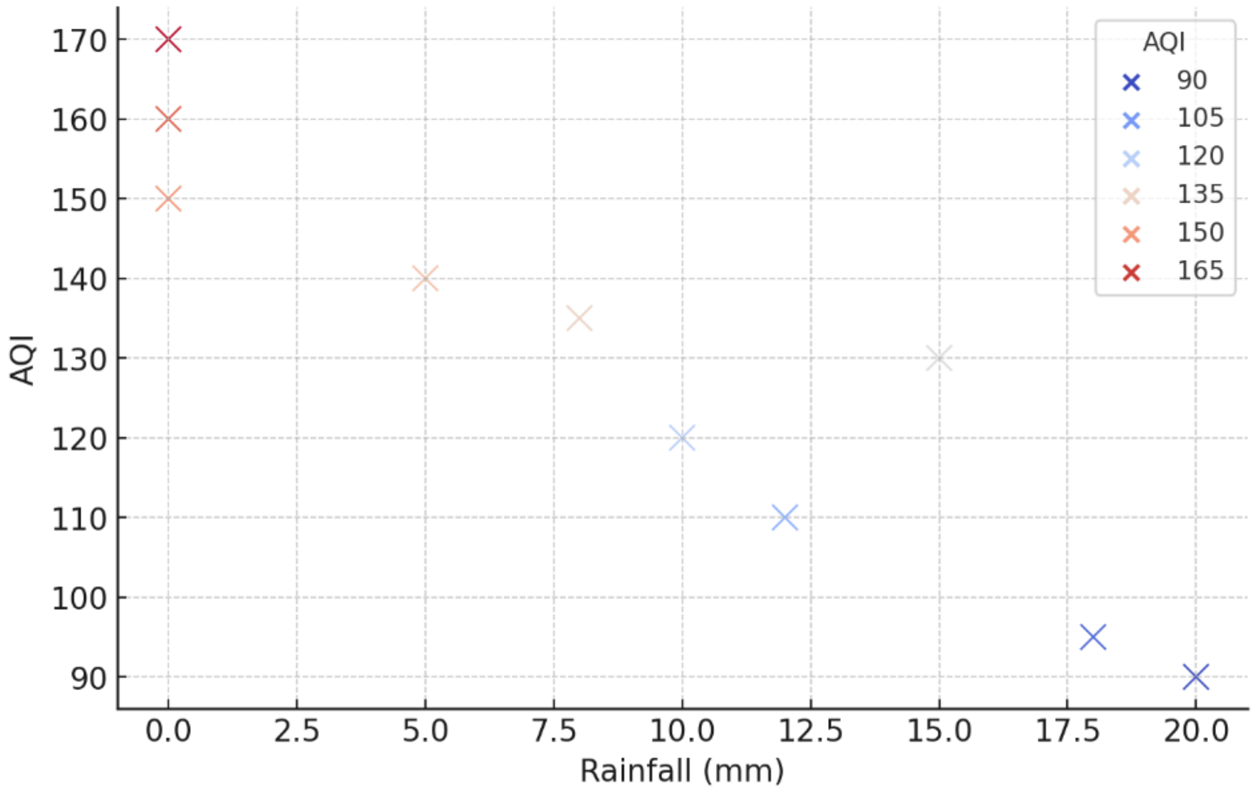

Figure 1. Scatter plot showing the relationship between AQI and rainfall.

This can be explained by temperature reversal and low wind speeds, which trap pollutants near the surface (Huang et al., 2019). Since rainfall is more frequent and intense in tropical regions, its purifying effect was more pronounced, which is thought to have resulted in significant improvements in air quality (Zhao et al., 2021).

Field studies conducted in cities such as Delhi, Beijing, and São Paulo show that precipitation, together with factors such as wind speed and humidity, plays a determining role in AQI (Jiang et al., 2019). Similarly, in this study, the effect of precipitation on air quality was evaluated by examining the temporal variation of AQI (Figure 2) and daily precipitation amounts (Figure 3).

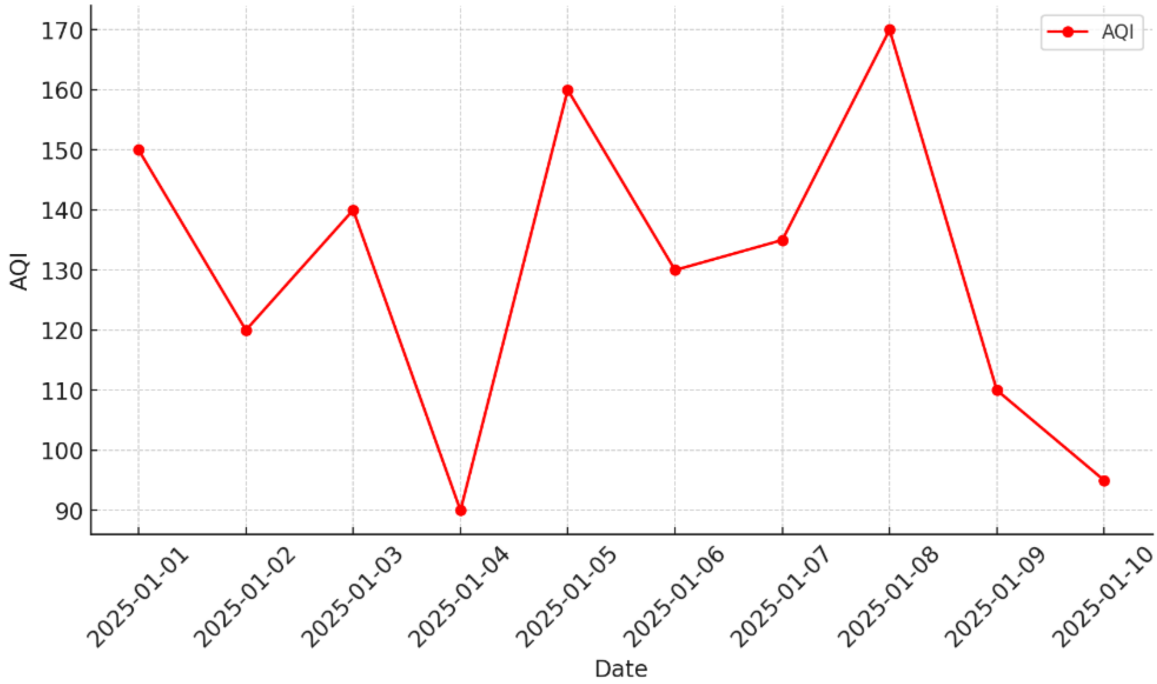

Figure 2. Timeline chart showing AQI changes over 10 days.

Figure 2 displays the variation of AQI values over time, showing the fluctuations in air quality observed on various days. For example, on 2025-01-01, despite no precipitation, the AQI value was measured as 150, which was attributed to high concentrations of PM2.5, PM10, and NO_2. This indicates that air quality has deteriorated significantly due to the high levels of emissions from traffic and industry. Similarly, on 2025-01-08, no precipitation was recorded, but the AQI increased to 170, which was due to the significant increase in PM2.5 and NO_2 levels.

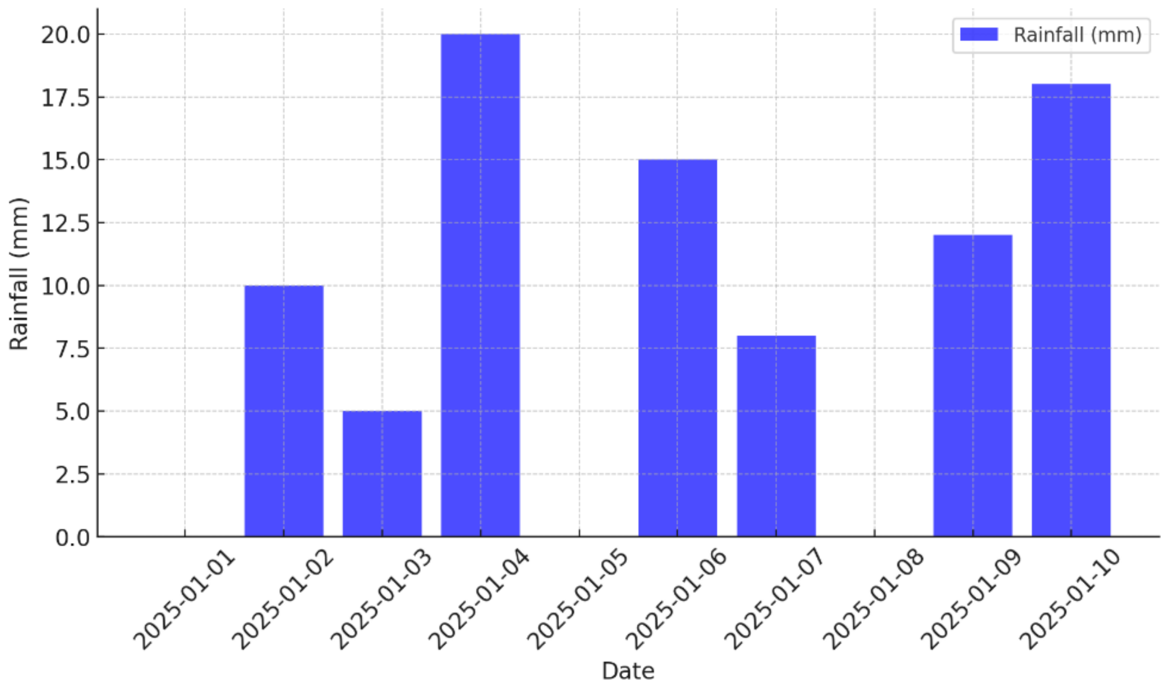

Figure 3. Illustration of daily rainfall levels to correlate with AQI changes.

The effect of precipitation on AQI becomes more apparent when compared with daily precipitation data (Figure 3). On 2025-01-04, 20 mm of precipitation occurred and the AQI value dropped to 90, showing a significant improvement compared to previous days. This reflects the effect of precipitation on removing particulate matter from the atmosphere and improving air quality. Similarly, on 2025-01-09, moderate precipitation (12 mm) was observed and the AQI value dropped to 110. However, this decrease was more limited compared to days with more intense precipitation, suggesting a positive relationship between precipitation amount and air quality improvement.

Assessment of statistical impacts of meteorological and particulate parameters on AQI

The correlations between AQI and key meteorological and pollutant variables are given in Figure 4. A strong negative relationship was found between rainfall and AQI, while strong positive correlations were found with PM2.5 and PM10. Moreover, the monotonic relationship between AQI and PM2.5 concentration is also supported visually (Figure 5).

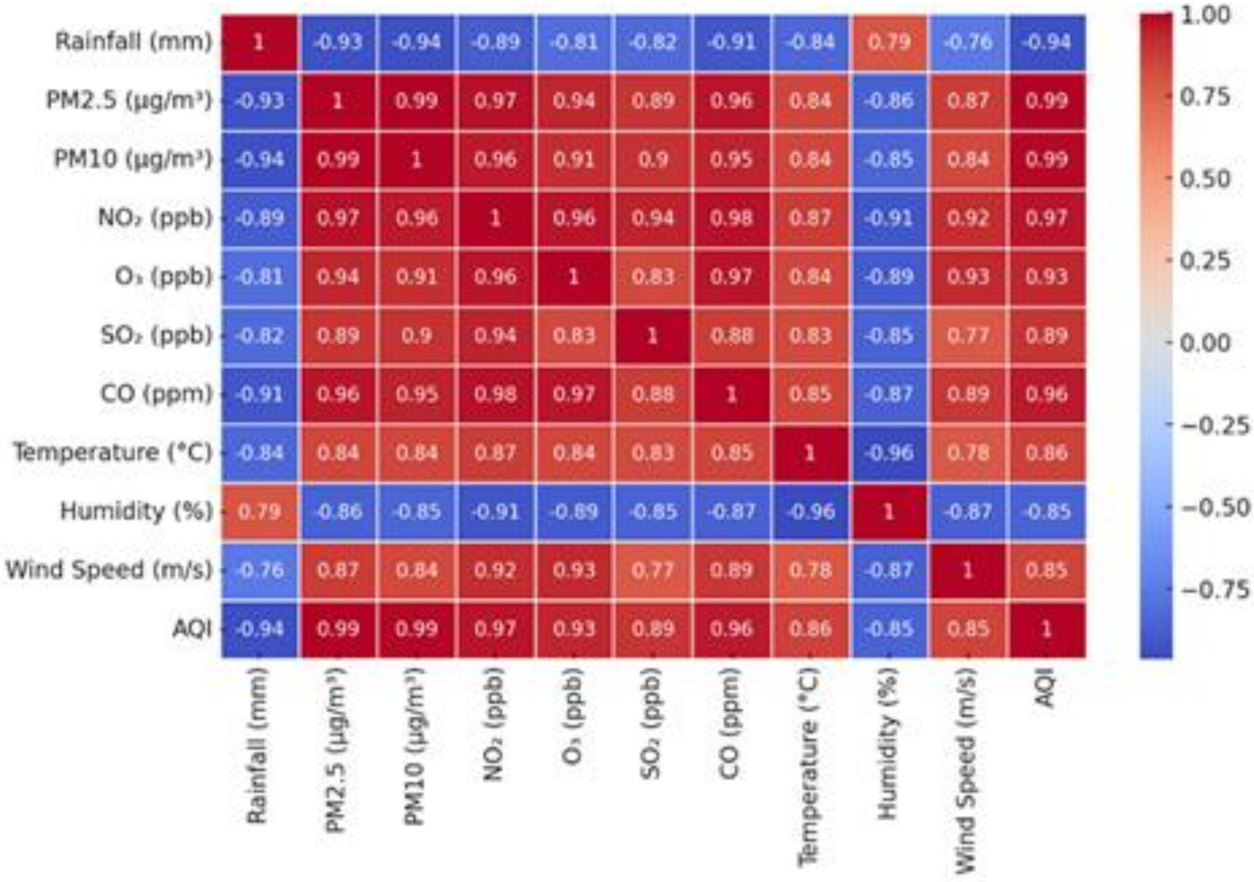

Figure 4. Correlation heat map between AQI and meteorological variables.

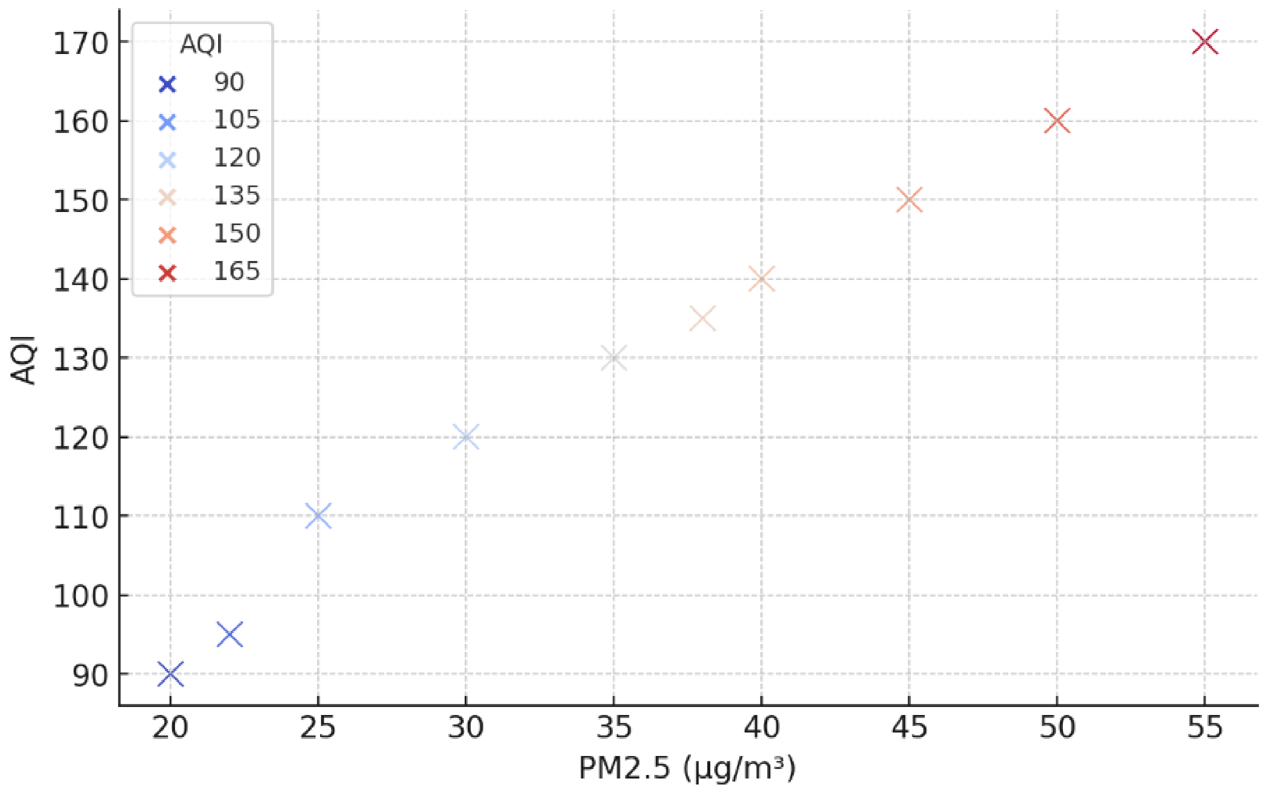

Figure 5. Scatter plot showing the relationship between AQI and PM2.5.

The findings indicate that rainfall plays an important but complex role in regulating air pollution levels in the Indian subcontinent. Heavy rainfall effectively reduces AQI by removing particulate matter from the atmosphere. However, the magnitude of this effect depends on various factors such as the intensity of the pollution source, wind patterns, and seasonal conditions (Fan et al., 2020). The study confirms the importance of multivariate AQI prediction models, showing that simple predictions based on rainfall data alone may overestimate or underestimate the true impact of rainfall (Tang et al., 2021).

Comparison of AQI estimation models

Traditional AQI prediction models are usually based on past pollutant concentrations and meteorological parameters such as temperature and humidity, which lead to accuracy limitations. The model presented in this study provided higher accuracy by including precipitation as a key feature (Li et al., 2022). Methods such as linear regression and ARIMA could not capture the nonlinear relationships between precipitation and AQI, thus yielding higher error rates (Gao et al., 2019).

Machine learning models such as Random Forest and XGBoost performed better. However, by combining these models with differential equation-based pollutant dispersion models, the predictions were further strengthened (Wang and Sun, 2021). Although remote sensing-based AQI models using satellite data have advantages in predicting air pollution levels over large areas, the hybrid ML-mathematical approach in this study provided higher temporal accuracy based on ground-based meteorological data. However, satellite models are still valuable for large-scale monitoring. This suggests that in the future, combining satellite imagery with ground-based data may provide the most comprehensive results in AQI estimation (Jiang et al., 2022). These comparisons highlight the importance of hybrid modeling approaches in regions where the impact of precipitation on air pollution is significant. The integration of machine learning, meteorological data, and mathematical modeling provides a more robust forecasting framework for policy development, urban planning, and early warning systems.

In conclusion, this study investigated the impact of precipitation and meteorological factors on AQI in the Indian subcontinent through machine learning and mathematical models. In particular, it was observed that algorithms such as XGBoost and LSTM provided more accurate estimates when integrated with pollutant dispersion equations. Precipitation played a decisive role in improving air quality by reducing PM2.5 and PM10 concentrations; however, this effect varied depending on environmental factors such as temperature, humidity, and pollution intensity. The findings revealed that hybrid models provided more reliable results in AQI estimation than traditional methods. It is recommended that future studies develop more comprehensive systems adapted to local conditions by integrating real-time emission data and satellite-based measurements. Such models can contribute significantly as strategic decision-support tools in terms of environmental management, public health, and urban planning.

Conflict of Interest

The authors declare that for this article they have no actual, potential or perceived conflict of interest.

Author Contributions

All authors performed all the experiments and drafted the main manuscript text. Authors reviewed and approved the final version of the manuscript.

Ethical Approval Statements

No ethics committee permissions are required for this study.

Data Availability

The data used in the present study are available upon request from the corresponding author.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X. (2016). TensorFlow: a system for large-scale machine learning. In: Proceedings Book. 12th USENIX Symposium on Operating Systems Design and Implementation, 2-4 November 2016, Savannah, USA. pp. 265-283.

Bergstra, J., Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13, 281-305.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.

Chen, T., Guestrin, C. (2016). X G Boost: A scalable tree boosting system. In: Proceedings Book. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17 August 2016, San Francisco, USA. pp. 785-794.

Chen, X., Zhang, X., Guo, Y. (2018). Influence of meteorological parameters on urban air pollution: A study based on long-term observation. *Atmospheric Environment*, 193, 230-238.

Fan, C., Li, J., Wang, S. (2020). Predicting air pollution with machine learning: A review of algorithms and applications. *Environmental Pollution*, 263, 114376.

Field, A. (2013). Discovering statistics using IBM SPSS statistics. Sage Publications.

Gao, M., Beig, G., Song, S. (2019). Ozone pollution prediction using meteorological and emission data with deep learning. *Environmental Science & Technology*, 53(18), 11038-11045.

Gupta, P., Christopher, S. A. (2009). Particulate matter air quality assessment using integrated surface, satellite, and meteorological data. *Environmental Science & Technology*, 43(20), 7694-7700.

Han, J., Kamber, M., Pei, J. (2011). Data mining: Concepts and techniques. Morgan Kaufmann.

Hastie, T., Tibshirani, R., Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction. Springer.

Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.

Huang, Y., Li, Y., Yang, Y. (2019). Understanding the effects of rainfall on air pollution: A study of urban regions. *Journal of Environmental Sciences*, 85, 42-50.

Hyndman, R. J., Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts.

James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). An introduction to statistical learning with applications. Springer.

Jiang, X., Li, R., Zhan, Y. (2019). Machine learning applications in air pollution prediction: A systematic review. *Journal of Cleaner Production*, 217, 894-906.

Jiang, Z., Yang, S., Liu, Z., Xu, Y., Xiong, Y., Qi, S., Pang, Q., Xu, J., Liu, F., Xu, T. (2022). Coupling machine learning and weather forecast to predict farmland flood disaster: A case study in Yangtze River basin. *Environmental Modelling & Software*, 155, 105436.

LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

Li, Y., Peng, T., Hua, L., Ji, C., Ma, H., Nazir, M. S., Zhang, C. (2022). Research and application of an evolutionary deep learning model based on improved grey wolf optimization algorithm and DBN-ELM for AQI prediction. *Sustainable Cities and Society*, 87, 104209.

Little, R. J., Rubin, D. B. (2019). Statistical analysis with missing data. John Wiley & Sons.

MathWorks. (2021). MATLAB documentation. The MathWorks.

McKinney, W. (2010). Data structures for statistical computing in Python. In: Proceedings Book. 9th Python in Science Conference, 28 June - 3 July 2010, Texas, USA. pp. 56-61.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.

Tang, L., Zhou, J., Zhao, M. (2021). Rainfall and air pollution: A multi-city case study. *Atmospheric Pollution Research*, 12(4), 456-469.

Wang, J., Sun, Y. (2021). Hybrid deep learning approaches for air quality prediction. *Environmental Research*, 195, 110849.

Wang, L., Wu, S., Yu, C. (2022). The role of meteorological factors in air pollution forecasting: A machine learning perspective. *Atmospheric Environment*, 262, 118629.

WHO. (2021). WHO global air quality guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization. Available at: https://www.who.int/publications/i/item/9789240034228 (04.05.25).

Willmott, C. J., Robeson, S. M., Matsuura, K. (2012). A refined index of model performance. *International Journal of Climatology*, 32, 2088-2094.

Xu, H., Zeng, W., Guo, B., Hopke, P. K., Qiao, X., Choi, H., Luo, B., Zhang, W., Zhao, X. (2020). Improved risk communications with a Bayesian multipollutant air quality health index. *Science of the Total Environment*, 722, 137892.

Yang, W., Deng, M., Xu, F., Wang, H. (2018). Prediction of hourly PM2. 5 using a space-time support vector regression model. *Atmospheric Environment*, 181, 12-19.

Zhang, K., de Leeuw, G., Yang, Z., Chen, X., Jiao, J. (2020). The impacts of the COVID-19 lockdown on air quality in the Guanzhong Basin, China. *Remote Sensing*, 12(18), 3042.

Zhang, Y., Ma, Y., Feng, F., Cheng, B., Shen, J., Wang, H., Jiao, H., Li, M. (2021). Respiratory mortality associated with ozone in China: a systematic review and meta-analysis. *Environmental Pollution*, 280, 116957.

Zhao, H., Chen, K., Liu, Z., Zhang, Y., Shao, T., Zhang, H. (2021). Coordinated control of PM2. 5 and O3 is urgently needed in China after implementation of the "Air pollution prevention and control action plan". *Chemosphere*, 270, 129441.

Zhao, P., Zhang, X., Chen, H. (2022). Meteorology-driven air quality forecasting using hybrid models. *Science of the Total Environment*, 814, 152561.

Zheng, J., Su, M., Ying, W., Tong, J., Pan, Z. (2021). Improved uniform phase empirical mode decomposition and its application in machinery fault diagnosis. *Measurement*, 179, 109425.