iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VI **Month of publication:** June 2025

DOI: https://doi.org/10.22214/ijraset.2025.72606

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VI June 2025- Available at www.ijraset.com

Review on Air Quality Index Prediction and Management Using Machine Learning Techniques

Ankur Goswami[1], Sanjay Kumar Sharma[2]
[1]*PG Student,* [2]*Professor, Department of Civil Engineering, NITTTR, Chandigarh*

Abstract: ***The Air Quality Index (AQI) is a standardized tool that simplifies complex air pollutant data into an accessible format, aiding public awareness and policy-making. It incorporates key pollutants like PM2.5, PM10, NO2, SO2, CO, and Ozone, with values ranging from 0 (Good) to 500+ (Hazardous). This paper reviews the role of machine learning (ML) in enhancing AQI prediction, particularly in urban areas like Jaipur, which face unique pollution challenges due to rapid urbanization, vehicular emissions, industrial activities, and natural phenomena like dust storms. The study synthesizes findings from recent research on ML-based AQI forecasting, focusing on predictive modeling and real-time alert systems. It also identifies gaps in region-specific modeling and handling extreme pollution events, emphasizing the need for tailored solutions for Jaipur's semi-arid climate. The review highlights the potential of advanced ML techniques to improve air quality management and proposes directions for future research.***

Keywords: Air Quality Index, machine learning, air pollution, predictive modeling, Jaipur

I. INTRODUCTION

Air pollution is a critical global issue, particularly in urban areas where dense populations, transportation, and industrial activities exacerbate pollutant levels. The Air Quality Index (AQI) serves as a vital tool to communicate air pollution severity to the public, translating complex pollutant data into a simplified scale. AQI considers major pollutants—PM2.5, PM10, NO2, SO2, CO, and Ozone—and categorizes air quality into six levels, from Good (0–50) to Hazardous (301+), each with specific health implications. Effective AQI prediction is essential for proactive air quality management, enabling timely health advisories and mitigation strategies. Machine learning (ML) has emerged as a transformative approach, offering superior predictive capabilities over traditional methods by analyzing multidimensional datasets. This paper reviews the application of ML in AQI prediction, with a focus on Jaipur, a rapidly urbanizing city with unique pollution challenges due to its semi-arid climate and diverse pollution sources.

A. Air Quality Index(AQI)

The AQI is a composite measure that quantifies air pollution levels by aggregating sub-indices of key pollutants: PM2.5, PM10, NO2, SO2, CO, and Ozone. The index ranges from 0 to 500+, with higher values indicating worse air quality and greater health risks. The six AQI categories and their health implications are shown in Table I

TABLE I
AQI Categories and Health Implications

Range	Category	Health Implications
0-50	Good	Minimal health impact.
51-100	Moderate	Acceptable; slight impact on sensitive groups.
101-150	Unhealthy for Sensitive Groups	May affect children, elderly, and individuals with respiratory conditions.
151-200	Unhealthy	General population may experience adverse effects.
201-300	Very Unhealthy	Increased risk of respiratory and cardiovascular conditions.
301+	Hazardous	Severe health impacts across all population groups.

©IJRASET: All Rights are Reserved

B. Air Pollution in Urban Areas

.Urban centers, like Jaipur, are pollution hotspots due to high population density, vehicular traffic, industrial emissions, and construction activities. Jaipur's semi-arid climate amplifies challenges, with dust storms contributing to elevated PM10 levels. Key pollution sources include:

- *Transportation*: Diesel and petrol vehicles emit NOx, CO, and particulates, worsened by high traffic density and outdated vehicles.
- *Industrial and Commercial Activities*: Factories and small-scale industries release SO2, VOCs, and NOx, often with inadequate emission controls.
- *Construction and Dust*: Urban expansion and seasonal dust storms increase PM10 concentrations.
- *Residential Sources*: Biomass burning in households contributes CO and PM2.5. Health impacts range from short-term respiratory irritation to long-term risks like COPD, cardiovascular diseases, and cancer. Environmental consequences include biodiversity loss and water contamination.

C. Role of Machine Learning in Air Pollution Management

Machine learning enhances air quality management by leveraging advanced algorithms to process complex datasets, including pollutant concentrations and meteorological variables. Key applications include:

- Predictive Modeling: ML models forecast AQI using historical and real-time data.
- Real-Time Alerts: ML systems provide timely warnings when AQI reaches critical thresholds.

D. Jaipur City: Pollution and Challenges

Jaipur's air quality is influenced by its rapid urbanization, vehicular growth, industrial activities, and natural dust storms. The city's three monitoring stations provide real-time data but lack predictive capabilities. Key contributors to pollution include vehicular emissions, unregulated industrial activities, and seasonal PM10 spikes.

E. Need for Predictive Modeling in AQI Management

Current monitoring systems in Jaipur offer real-time data but cannot forecast future AQI trends. Predictive models are crucial for:

- Forecasting AQI to enable early warnings.
- Analyzing long-term trends for urban planning.
- Identifying pollution hotspots for targeted mitigation.

II. LITERATURE SURVEY

A comprehensive review of recent studies on AQI prediction using ML techniques highlights advancements and identifies research gaps. Below is a summary of key studies:

1) Soni et al. (2018) used satellite AOD and meteorological parameters to estimate PM concentrations in Jaipur, achieving a correlation of R = 0.8 with linear regression models [1].
2) Singh et al. (2022) developed an LSTM model for PM2.5 prediction in Jaipur, reporting RMSE below 10 μg/m³, showcasing LSTM's strength in handling temporal data [2].
3) Suri et al. (2023) employed ANN for AQI prediction across Indian cities, achieving 91% accuracy, outperforming linear regression [3].
4) Bhati et al. (2024) analyzed air quality during the COVID-19 lockdown using a hybrid AI model, noting a 30% improvement in air quality [4].
5) Gupta et al. (2023) compared ML models for AQI prediction, with Gradient Boosting achieving $R^2 = 0.95$ [5].

III. RESULTS AND DISCUSSION

The literature reveals significant advancements in ML-based AQI prediction. Key findings include:

1) Model Performance: Models like Gradient Boosting, LSTM, and ANN consistently achieve high accuracy ($R^2 > 0.9$) for AQI and pollutant concentration predictions.
2) Regional Specificity: Most models are generalized and do not account for Jaipur's unique semi-arid climate and pollution sources, reducing their local applicability.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

3) Extreme Scenarios: Models often fail to maintain accuracy during abrupt pollution changes, such as haze or lockdown-induced reductions, necessitating adaptable algorithms.

Table II summarizes the performance of various ML models for AQI prediction.

TABLE II
Performance of ML Models for AQI Prediction

Author	Model Type	Metric	Value
Soni et al (2018)	Linear Regression	Correlation(R)	0.8
Singh et al. (2022)	LSTM	RMSE	<10 µg/m³
Suri et al. (2023)	ANN	Accuracy	91%
Gupta et al. (2023)	Gradient Boosting	R^2	0.95

IV. CONCLUSION

This review underscores the transformative potential of ML in AQI prediction and air quality management. Techniques like LSTM, ANN, and Gradient Boosting have demonstrated high accuracy in forecasting pollutant concentrations. However, the lack of region-specific models for cities like Jaipur and challenges in handling extreme pollution events highlight areas for future research. Developing tailored ML models that incorporate Jaipur's environmental and pollution dynamics will enhance predictive accuracy and support proactive air quality management.

V. ACKNOWLEDGEMENT

The authors wish to acknowledge the support of NITTTR, Chandigarh for providing access to air quality data and research facilities.

REFERENCES

[1] Soni, Manish, Swagata Payra, and Sunita Verma (2018) [1] conducted a study titled "Particulate Matter Estimation over a Semi-Arid Region Jaipur, India Using Satellite AOD and Meteorological Parameters". This study uses satellite Aerosol Optical Depth (AOD) and meteorological parameters to estimate PM concentrations in Jaipur. The authors employed a linear regression model to establish relationships between AOD and PM, supported by meteorological factors such as temperature, humidity, and wind speed. Results indicated a strong correlation (R = 0.8), demonstrating that satellite data can effectively complement ground-based PM measurement.

[2] Singh, Uday Pratap et al. (2022) [2] conducted a study titled "Unraveling the Prediction of Fine Particulate Matter over Jaipur, India Using Long Short-Term Memory Neural Network". This research develops an LSTM model for predicting PM2.5 concentrations in Jaipur. The model considers meteorological parameters and historical PM2.5 data as inputs. The study reports high predictive accuracy with RMSE values below 10 µg/m³ and highlights the robustness of LSTM in handling temporal dependencies in air quality data.

[3] Suri, Raunaq Singh et al. (2023) [3] conducted a study titled "Air Quality Prediction—A Study Using Neural Network-Based Approach". This research employs an artificial neural network (ANN) to predict AQI based on meteorological and pollution data from multiple Indian cities. The study highlights the capability of ANN to outperform linear regression models with an accuracy of 91%, emphasizing its suitability for air quality forecasting.

[4] Bhati, Vikram Singh et al. (2024) [4] conducted a study titled "Exploring Air Quality Dynamics and Predictive Modeling Using Artificial Intelligence During COVID-19 Lockdown Over the Western Part of India". This work analyzes air quality dynamics during the COVID-19 lockdown period. Using a hybrid AI model combining regression and clustering techniques, the study provides a comparative analysis of air quality before and during the lockdown. Results show a 30% improvement in air quality, validating the effectiveness of reduced anthropogenic activities.

[5] Gupta, N. Srinivasa et al. (2023) [5] conducted a study titled "Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis". This research compares various ML models, including Random Forest, SVM, and Gradient Boosting, for AQI prediction. The Gradient Boosting model achieved the highest accuracy (R^2 = 0.95), demonstrating its effectiveness in handling complex nonlinear relationships in air quality data.

[6] Sethi, Jasleen Kaur and Mamta Mittal (2021) [6] conducted a study titled "Prediction of Air Quality Index Using Hybrid Machine Learning Algorithm". This study integrates neural networks with statistical techniques to predict AQI. The hybrid model outperformed standalone approaches, achieving RMSE values as low as 5.2 µg/m³ for PM2.5 prediction. The research underscores the potential of hybrid methods for enhanced prediction accuracy.

[7] Gokul, P. R. et al. (2023) [7] conducted a study titled "Spatio-Temporal Air Quality Analysis and PM2.5 Prediction Over Hyderabad City, India Using Artificial Intelligence Techniques". The study employs AI techniques, including Decision Trees and Neural Networks, to analyze spatio-temporal variations of PM2.5 concentrations. Results demonstrate that Neural Networks perform best, achieving an R^2 value of 0.92 for PM2.5 predictions.

[8] Suthar, Gourav et al. (2024) [8] conducted a study titled "Predicting Land Surface Temperature and Examining Its Relationship with Air Pollution and Urban Parameters in Bengaluru: A Machine Learning Approach". The study integrates machine learning algorithms to investigate the relationship between land surface temperature (LST), air pollution, and urbanization. The Gradient Boosting model achieved a prediction accuracy of 89%, revealing a significant correlation between LST and PM2.5 concentrations.

[9] Suthar, Gourav et al. (2024) [9] conducted a study titled "Prediction of Land Surface Temperature Using Spectral Indices, Air Pollutants, and Urbanization Parameters for Hyderabad City of India Using Six Machine Learning Approaches". This research compares six ML models, including Random Forest and XGBoost, to predict LST based on spectral indices and pollution parameters. XGBoost achieved the best performance, with an RMSE of 1.2°C.

[10] Natarajan, Suresh Kumar et al. (2024) [10] conducted a study titled "Optimized Machine Learning Model for Air Quality Index Prediction in Major Cities in India". This research optimizes ML models using hyperparameter tuning for AQI prediction. Results indicate that the Random Forest algorithm outperformed other models, achieving an accuracy of 96%.

[11] Goyal, S. and R. Sharma (2023) [11] conducted a study titled "Prediction of the Concentrations of PM2.5 and NOx Using Machine Learning-Based Models". This study utilizes regression and neural network models to predict PM2.5 and NOx concentrations. Neural networks performed better than regression models, achieving a correlation coefficient of 0.89 for PM2.5 predictions.

[12] Dey, Sweta et al. (2024) [12] conducted a study titled "APICT: Air Pollution Epidemiology Using Green AQI Prediction During Winter Seasons in India". This work develops a novel Green AQI prediction model incorporating meteorological and pollution data. The model demonstrated a significant reduction in false alarms compared to traditional AQI models, achieving a sensitivity of 94%.

[13] Choudhary, Arti et al. (2023) [13] conducted a study titled "Evaluating Air Quality and Criteria Pollutants Prediction Disparities by Data Mining Along a Stretch of Urban-Rural Agglomeration". This research highlights the disparities in pollution levels across urban and rural regions using data mining techniques. The study reveals significant variations in NO2 and PM2.5 concentrations, emphasizing the need for location-specific models.

[14] Bajpai, Mann et al. (2023) [14] conducted a study titled "Air Quality Index Prediction Using Various Machine Learning Algorithms". This work evaluates multiple ML algorithms, including SVM, ANN, and k-NN, for AQI prediction. The SVM model exhibited the highest accuracy, achieving a mean squared error of 18.2.

[15] Suthar, Gourav et al. (2023) [15] conducted a study titled "Spatiotemporal Variation of Air Pollutants and Their Relationship with Land Surface Temperature in Bengaluru, India". This study examines the spatiotemporal dynamics of air pollutants and their correlation with LST. Results show that urban heat islands significantly contribute to increased pollutant concentrations, with PM10 showing the highest correlation.

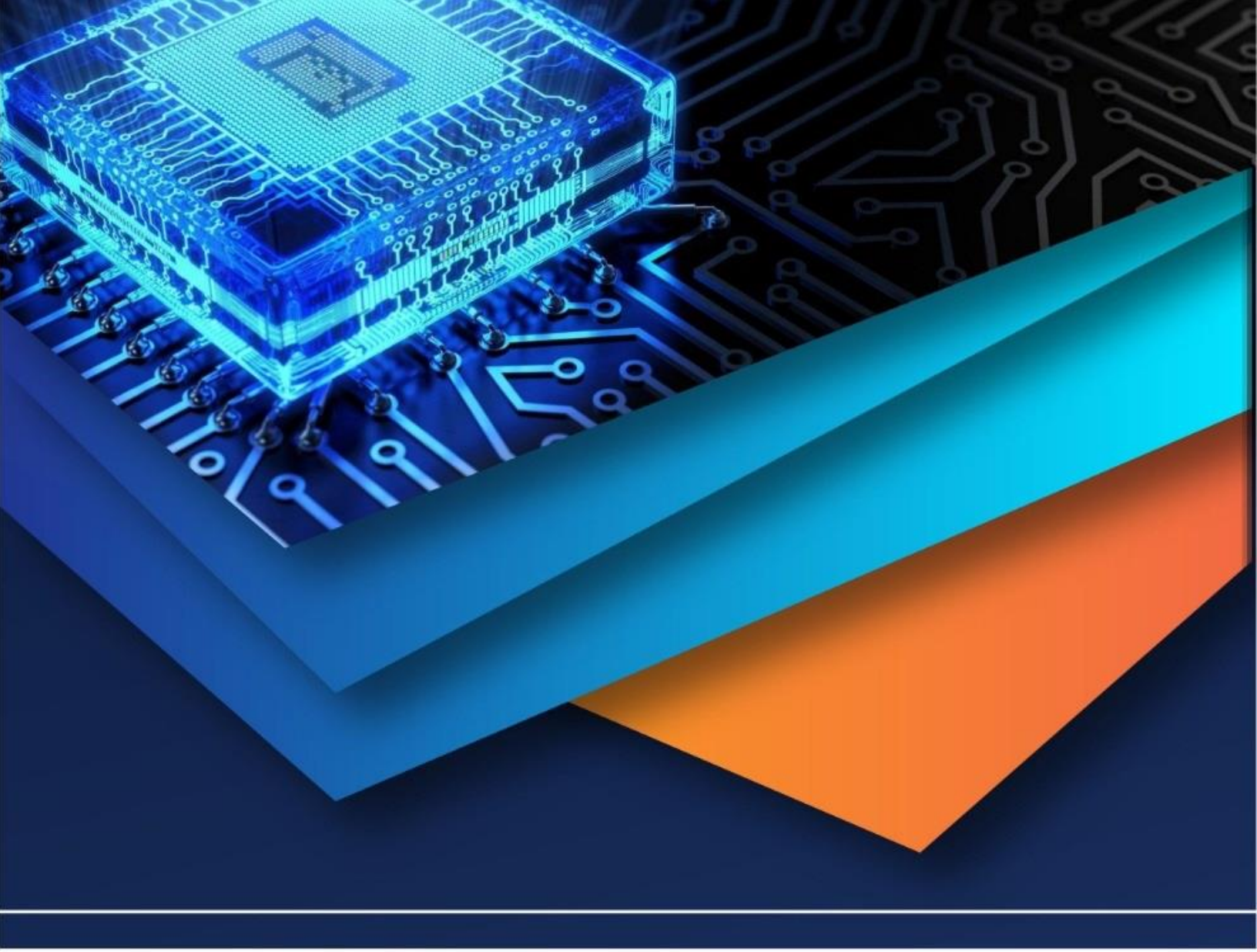
[16] Venkateswaran, R. et al. (2024) [16] conducted a study titled "Optimized Air Quality Index and Meteorological Predictions with Machine Learning and IoT". This study integrates IoT-based data collection with ML models for AQI prediction. Results demonstrate a 20% improvement in prediction accuracy using IoT-enhanced datasets.

[17] Halder, S. and S. Bose (2024) [17] conducted a study titled "Ecological Quality Assessment of Five Smart Cities in India: A Remote Sensing Index-Based Analysis". This research uses remote sensing indices to assess ecological quality in five smart cities. Results indicate a strong inverse correlation between urbanization and air quality, with PM2.5 concentrations being the primary determinant of ecological quality.

[18] Al-Hamdan, Mohammad Z. et al. (2009) [18] conducted a study titled "Methods for Characterizing Fine Particulate Matter Using Ground Observations and Remotely Sensed Data". This foundational study combines ground-based measurements with satellite data to monitor PM2.5. The methodology, leveraging hybrid statistical and machine learning techniques, set a precedent for air quality surveillance.

[19] Blanco, Giacomo et al. (2024) [19] conducted a study titled "Urban Air Pollution Forecasting: A Machine Learning Approach Leveraging Satellite Observations and Meteorological Forecasts". This research integrates satellite data and ML algorithms for urban air pollution forecasting, achieving RMSE values below 5 $\mu g/m^3$ for PM10 predictions.

[20] Goyal, P. and Sidhartha (2015) [20] conducted a study titled "Modeling and Prediction of Hourly Ambient Ozone (O_3) and Oxides of Nitrogen (NO_x) Concentrations Using Artificial Neural Network and Decision Tree Algorithms". This study evaluates the efficacy of ANN and Decision Trees for O_3 and NO_x prediction. ANN outperformed Decision Trees, achieving an R^2 value of 0.88 for hourly concentration predictions.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)