International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: I Month of publication: January 2026

DOI: https://doi.org/10.22214/ijraset.2026.77048

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue I Jan 2026- Available at www.ijraset.com

A Survey on AQI Prediction and Analysis of Traffic Impact on Air Quality

Ashwini Bhavsar[1], Damayanti Kale[2], Rajvi Damrekar[3], Mrunali Bawane[4], Arya Madhe[5]
[1, 2, 3, 4, 5]*Department of Computer Engineering, Pimpri Chinchwad College of Engineering and Research, Pune, Maharashtra, India*

Abstract: Air pollution, a matter of grave concern, has caught the cities by storm, and transportation is taking a major role in that process. Therefore, accurate Air Quality Index prediction is necessary for successful management of air pollution, urban planning, and public awareness. A number of approaches, mostly using machine learning algorithms, have been proposed recently for precise prediction of the Air Quality Index by describing complex correlations between air pollutant concentration and atmospheric variables. However, the role of traffic-related variables in Air Quality Index prediction is often ignored.

This survey provides a comprehensive overview of the relevant methods of machine learning and deep learning applied to air quality index (AQI) forecasting, primarily focusing on the works that involve or address parameters of vehicle traffic, like density and speed. The relevant literature is provided with a systematic arrangement on the basis of methods of models, sources of data, and methods of combination of traffic parameters. A comparative study of the research conducted earlier shows the key results and current deficiencies. In addition to that, the survey points out the key research deficiencies and research avenues that can be addressed to develop effective traffic-aware AQI forecasting systems.

Keywords: Air Quality Index (AQI), Machine Learning, Traffic Analysis, Air Pollution, Predictive Modeling.

I. INTRODUCTION

Air Pollution has undoubtedly turned out to become an, in fact, indisputable issue within the category of enviromnental as well as public health concerns within the present century. The terrible results of industrialization, population growth, as well as urbanization, have caused a substantial increase in pollutant emission levels, ultimately affecting air quality in metropolises negatively. Among several air pollutants, motor vehicles, as well as traffic, are one of the key sources, releasing considerable amounts of toxic pollutants such as particulate matter (PM2.5, PM10), nitrogen oxides (NOx), carbon monoxide (CO), as well as volatile organic compounds (VOCs). Prolonged exposure to these air pollutants has already been recognized to cause respiratory problems, cardiovascular disease, decreased life expectancy, as well as increased death rates. Hence, it has become a pressing concern for air quality index prediction to become highly significant for, as well as among, authorities, as well as citizens, within the awareness as well as prevention process.

The AQI models employed so far have been seriously dependent on data related to meteorology and the environment, including the levels of temperatures, humidness, air speed, and concentration of pollutants. Even if these models are quite proficient in recognizing the overall variation of pollutants, they overlook the fluctuating effect of traffic congestion caused by rush hours. The effect of rush hour traffic congestion causes sudden hikes in the levels of pollutants, which are not recognizable by models related to the environment. The deficiency in accuracy and timeliness of predictions has restricted the use of AQI models in the context of decision-making.

ML has gone through a lot of novel ways such that the researchers can now combine various sources of data with an aim of predicting Air Quality Index better. Different researches have given performance of models like Random Forest, Support Vector Regression, Artificial Neural Networks, and Long Short-Term Memory networks in discovering nonlinear relationships between concentrations of the pollutants and environmental parameters. However, proper and adequate integration of factors related to traffic like the density of the vehicles, congestion rate, and average speed of traffic in AQI prediction models still forms a major lacuna. It is emphasized that this lacuna needs to be filled to satisfy one of the principal requirements for smart monitoring of an urban area.

II. RELATED WORK

Air Quality Management related research has evolved in two distinct but complementary directions: (i) Air Quality Index Modeling by Machine Learning Techniques, and (ii) analyzing air pollution levels in terms of certain external factors such as traffic flow.

A. Machine Learning–Based AQI Prediction Models

Recent research has shown that ML techniques far exceed traditional statistical methods, identifying complex nonlinear relationships and temporal dependencies of air pollutant data. In a review of more than seventy recent studies, ML approaches were categorized into non-neural models-such as Random Forest and XGBoost-deep learning models including Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN)-and hybrid models. The review concluded that the ensemble methods work exceptionally well on structured datasets, while deep learning models are better suited to spatiotemporal patterns. However, challenges related to generalizability and interpretability remain. Various hybrid modeling methods have been proposed to enhance AQI prediction accuracy. A recent study has proposed ANN-LSTM and XGBoost, which could reach a value of up to 92.77% R^2 using deep learning for feature extraction and ensemble techniques for error correction from pollutant PM2.5 and meteorological data from urban monitoring stations. Considering the enhanced accuracy of the technique, its quality was affected by three major negative factors: critical issues with quality; high computational complexity; low potential for cross-regional implementation due to dependence on data from IoT sensors in specific locations. To make model design simpler, some authors have tried to concentrate on fewer input features. For instance, a regression model including PM2.5, PM10, and CO as features employed Gaussian Process Regression and Ensemble Regression methods, resulting in an R^2 value of 0.9937 on a dataset from 2022 in Gazipur, Bangladesh. But lack of meteorological features can limit the robustness of the model in different environments. Models designed with interpretability in mind have also been researched. In AQI prediction using air quality data from India, a hybrid model using both Random Forest and ARIMA was considered that resulted in an MSE value of 508.46 with a corresponding R-squared value of 0.94. SHAP values indicated PM2.5 to have the maximum impact of all factors. Models like these would have difficulty in modeling the non-linear effects associated with lockdowns due to external events or massive gatherings. Some comparative studies were conducted based on shallow as well as deep learning models to identify trade-offs between their performances. In Zabol, Iran, CNN-based models showed greater accuracy (0.60) for PM10-based AQI classification using daily data from 2013 to 2022, but there were certain limitations due to imbalances in classes related to unhealthy AQI levels. Similarly, deep learning models like ANN and LSTM were applied for air quality data from Macau from 2013 to 2021, during which LSTM surpassed all models by producing a Pearson Correlation Coefficient of 0.87 for air pollutants like PM10, CO, etc., but these models can cause certain region-related dependencies due to inadequate interpretability. A comparative analysis was done on air quality data from 2020-2023 for Bhubaneswar, India, which revealed that models developed on CNN performed better ($R^2 = 0.97$) compared to regression models because they addressed spatial features. In another study, the model developed on LSTM-Bootstrap for data from 2022-2023 for Lanzhou, China, obtained R^2 of 0.978. The model determined PM10 and O3 to be significant pollutants by applying a cumulative logit model. The model's applicability was constrained by its restricted timeframe.

B. Traffic Impact–Aware AQI Prediction

Integration of real-time traffic information with AQI prediction has been selected as one of the latest key research areas that are significant in reducing air pollution within cities. A research conducted by Setiawan et al. (2023) used IoT sensing techniques and machine learning algorithms such as Linear Regression and Decision Trees that analyze CO_2 and CH_2O emissions measured from 2021-2022 in Jakarta. These researches gave a negative correlation between traffic flow, air quality, and pollution. Pollution sources from industries were excluded in the study. Similarly, a traffic-influenced AQI model was proposed for air quality index prediction by Praveena Kumari et al. using regression algorithms (Linear Regression, Ridge Regression, and Lasso Regression) and ensemble learning algorithms (Bagging and Stacking algorithms). The results gave evidence of a model at R^2 0.89 for PM2.5 prediction influenced by traffic. However, this approach was again limited to region-specific data and failed to address important dynamic factors in the environment, such as road work and diversions of traffic.

TABLE 1: SUMMARY OF LITERATURE SURVEY

Sr. No.	Author(s) & Year	Title	Methodology	Limitations
1	Yıldırım Özüpak, Feyyaz Alpsalaz, Emrah Aslan (2025)	Air Quality Forecasting Using Machine Learning: Comparative Analysis and Ensemble Strategies for Enhanced Prediction	Developed AQI prediction models using 10 ML regression algorithms. Bayesian hyperparameter optimization and stacking ensemble techniques were applied to improve pollutant prediction accuracy.	Study limited to a single urban area; lack of multi-region validation reduces generalizability.

Sr. No.	Author(s) & Year	Title	Methodology	Limitations
2	Praveena Kumari, M. K. Manjaiha, D. H. Ashwini, K. M. (2024)	Traffic Prediction Based on Air Quality in IoT-Based Smart City Using Regression and Ensemble Techniques	Used air quality parameters (PM2.5, NO_2, CO, etc.) with regression models (Linear, Ridge, Lasso, Polynomial) combined with Bagging and Stacking ensemble techniques to predict traffic patterns.	Limited dataset and insufficient generalization techniques reduce robustness of results.
3	Setiawan et al. (IEEE, 2023)	The Impact of Traffic Conditions on Air Quality: A Case Study of IoT Monitoring and Machine Learning Analysis	Implemented IoT sensors with ML models (Linear Regression, Decision Tree) to analyze traffic congestion impact on pollutants such as CO_2 and CH_2O.	Simple models may miss complex patterns; external pollution sources were not considered.
4	Nilesh Maltare, Safvan Vahora (2023)	Air Quality Index Prediction Using Machine Learning for Ahmedabad City	Compared SVM (various kernels) and LSTM models for AQI prediction using city-level air quality data.	Model trained on a limited geographic region; requires retraining for other cities.
5	Fareena Naz et al. (2023)	Comparative Analysis of Deep Learning and Statistical Models for Air Pollutants Prediction in Urban Areas	Compared deep learning models (LSTM, GRU) with statistical models for single-step prediction of NO_2, O_3, SO_2, PM2.5, and PM10 using real-world data.	Restricted to single-step forecasting; no advanced feature engineering or hyperparameter tuning.
6	Dobrea et al. (2020)	Air Quality Prediction Using Machine Learning Algorithms: A Review	Reviewed ML techniques such as SVR and LSTM applied to time-series pollutant data (PM10, PM2.5).	LSTM showed lower performance compared to SVR; limited evaluation scope.
7	Tanisha Madan, Shraddha Sagar, Deepali Virmani (2020)	Air Quality Forecasting and Classification Using Machine Learning Based on Indian Urban Pollutants	Evaluated multiple ML algorithms for AQI prediction and classification across pollutants and regions, emphasizing health impact mitigation.	Challenges remain in sensor data quality, real-time analysis, and efficient multi-pollutant prediction.

III.METHODOLOGY

The survey is performed via a systematic literature review methodology to identify, analyze, and synthesize existing research regarding the prediction of AQI, with a particular emphasis on the impact of vehicular traffic. The methodology was designed in such a way that it ensures comprehensive coverage of relevant studies, minimizing selection bias.

A. Literature Search Strategy

A literature search has been done in several academic databases, including IEEE Xplore, Scopus, Springer Link, ScienceDirect, along with Google Scholar, with the aim of acquiring relevant information concerning the prediction of AQI and traffic impact analysis. The search covers conference papers, articles, as well as review articles.

The following keywords and their combinations were used during the search process:

“Air Quality Index prediction”, “air quality forecasting”,“traffic impact on air pollution”, “machine learning”,“deep learning”, “smart city air quality”, “traffic congestion and pollution”

The use of Boolean operators like AND and OR was undertaken to refine the searches and ensure that the results include literature that addressed both AQI prediction and traffic factors.

B. Study Selection Process

The process of study selection was done in a series of steps. Firstly, the titles and abstracts of the retrieved articles were screened to exclude irrelevant study articles. In the second stage, full-text analysis was done to determine the relevance of remaining articles for inclusion according to the predefined criteria of inclusion.

However, after the filtering process, more than fifty high-quality research articles were selected for the analysis. These articles comprise the main body of the survey since the research covers different geographical areas, sources, and models.

C. Data Extraction and Analysis

Relevant information was systematically extracted for each selected study, including:

1) Machine learning or deep learning models used
2) Type of input data: pollutant, meteorological, traffic, or hybrid
3) Traffic parameters considered, whether vehicle density, congestion index, speed, flow rate etc.
4) Geographical region and source of the dataset
5) Reported evaluation metrics such as RMSE, MAE, and coefficient of determination (R^2)
6) Key findings and reported limitations

The information extracted was tabulated and analyzed for recurring patterns across the studies, both in terms of strengths and challenges.

D. Classification and Synthesis Approach

The studies were grouped according to:

1) Machine learning method: This consists of traditional machine learning, ensemble learning methods,
2) Use of data sources (pollutant-only models).
3) Traffic parameter integration strategy

A comparative synthesis was then conducted to assess the impact of varying approaches used in modeling and traffic variables on the accuracy of predictions of the air quality indicator. There was focus on the trends and limitations associated with the models developed and used in varying studies.

E. Identification of Research Gaps

Based on the results of the comparative analysis and synthesis, the gaps in the research were determined based on the unresolved questions related to the inconsistencies in the studies. These questions include:

1) Lack of real-time data on traffic incidents
2) Region-specific model dependency
3) Interpretability of model result
4) Heterogeneous data sources integration issues.

IV. DISCUSSION

Analysis of the existing literature highlights that the application of machine learning algorithms has been successful in enhancing the accuracy level for predicting the values of Air Quality Index (AQI) rather than using statistical algorithms. It is noted in the studies that the application of Random Forest, XGBoost, or any other ensemble model is more efficient in comparison to linear models due to their ability to efficiently model the nonlinear associations existing in the environment for air quality. Models based on LSTMs are efficient in modeling the temporal associations for air quality.

One of the most important things that have been concluded from the literature review of the above-mentioned studies is the fact that the AQI forecasting models primarily concentrate on the pollutant concentration values and the meteorological factors. The traffic factors were given minimal consideration. The studies where the traffic factors like vehicle density, congestion level, and average speed of the vehicles were considered demonstrated a marked enhancement in the values of the model results measured in terms of an increase of 5 to 10% in the R^2 values.

Despite these innovations, there are a few areas that are universal in the existing literature. Most of these models, particularly the complex ones using deep learning, are non-interpretable, thereby hindering their applicability in practical policy-making contexts. Secondly, most models are region-specific, given that their training data is driven by local conditions, thereby reducing their applicability elsewhere in cities. Availability of realistic traffic data also remains a concern in practical, massive implementations.

In conclusion, the reviewed studies identify that to provide an integrated air quality forecast, the strategy should involve the combination of traffic, meteorological, and pollutant variables. This will aid the creation of a precise and local air quality assessment to help create decisions on traffic regulation and overall pollution.

V. CONCLUSION

This review paper comprehensively discussed machine learning-based approaches for the prediction of Air Quality Index, with the primary concentration on the impact that vehicle traffic patterns have on air quality. By integrating these research results from over fifty peer-reviewed publications, a conclusion was drawn that vehicle congestion level, vehicle density, and speed-related factors improve the accuracy of AQI prediction when combined with prevailing AQI data.

From the literature review, it has been found that ensemble methods and deep learning techniques have greater accuracy in nonlinear and dynamic variations of air quality than compared with traditional statistical methods. Some issues still exist, such as a lack of interpretability for proposed methods, regional adaptability of methods, or real-time lack of standardized traffic data. Further studies are required in improving regional model interpretability, traffic data merging methods, or discovering relationships between traffic patterns and concentration levels of certain air pollutants.

VI. ACKNOWLEDGMENT

The authors of this survey paper would like to express their heartfelt thanks to those who have helped in completing this paper named “*AQI prediction and analysis of traffic impact on air quality*.”

The authors would like to extend sincere thankfulness to their revered guide, Prof. Ashwini Bhavsar, who provided them with invaluable advice, motivation, and support in the preparation of this survey. The knowledge and intellectual inputs provided by Prof. Ashwini Bhavsar added considerable quality to the work.

Secondly, the authors are also obliged to express their thanks to the Head of the Department, Department of Computer Engineering, for giving an academic ambience in order to carry out this work.

Lastly, they would like to thank all faculty members and staff of the Department of Computer engineering, Pimpri Chinchwad College of Engineering & Research, for all support and cooperation in completion of this survey paper.

REFERENCES

[1] Yıldırım Özüpak, F., Alpsalaz, F., & Aslan, E. (2025). “Air quality forecasting using machine learning: Comparative analysis and ensemble strategies for enhanced prediction”. Volume 236, Article 464.

[2] Kumari, P. M. K., Manjaiha, D. H., & Ashwini, K. M. (2024). “Traffic prediction based on air quality in IoT-based smart city using regression and ensemble techniques”. In IIP Series, Volume 3, Book 2, Part 1, Chapter 7 (e-ISBN: 978-93-6252-830-8).

[3] Naz, F., Yousuf, M., Zia, I., & Tasadduq, I. A. (2023). “Comparative analysis of deep learning and statistical models for air pollutants prediction in urban areas”. IEEE Access. https://doi.org/10.1109/ACCESS.2023.3289153

[4] Maltare, N., & Vahora, S. (2023). “Air quality index prediction using machine learning for Ahmedabad city”. Data in Chemical Engineering, 100093. https://doi.org/10.1016/j.dche.2023.100093

[5] Setiawan, A., et al. (2023). “The impact of traffic conditions on air quality: A case study of IoT monitoring and machine learning analysis”. https://doi.org/10.1109/ICISS59129.2023.10291616

[6] A. Bose, I. Roy Chowdhury “Investigating the association between air pollutants' concentration and meteorological parameters in a rapidly growing urban center of West Bengal, India: a statistical modeling-based approach” Model. Earth Syst. Environ., 1 (2023), pp. 1-16, 10.1007/S40808-022-01670-6/FIGURES/9

[7] K. Chaudhari, D. Joshi, P. Harugade, K. Jambusariya and V. Tiwari, “Predicting Mumbai's Air Quality Index by Machine Learning,” INCET, Belgaum, India, 2023, pp. 1–5.

[8] M. S. Ram, C. Reshmasri, S. Shahila and J. V. P. Saketh, “Air Quality Prediction using Machine Learning Algorithm,” ICSCDS, Erode, India, 2023, pp. 316–321.

[9] Alkabbani, H., Ramadan, A., Zhu, Q. and Elkamel, A., 2022. An improved air quality index machine learning-based forecasting with multivariate data imputation approach. Atmosphere, 13 (7), p. 1144.

[10] Balogun A.-L., Tella A. “Modelling and investigating the impacts of climatic variables on ozone concentration in Malaysia using correlation analysis with random forest, decision tree regression, linear regression, and support vector regression” Chemosphere, 299 (2022), Article 134250, 10.1016/j.chemosphere.2022.134250

[11] A. Bekkar, B. Hssina, S. Douzi, K. Douzi “Air-pollution prediction in smart city, deep learning approach” J. Big Data, 8 (2021), pp. 1-21, 10.1186/S40537-021-00548-1/FIGURES/17

[12] Saikiran, K., Lithesh, G., Srinivas, B. and Ashok, S., 2021, September. Prediction of Air Quality Index Using Supervised Machine Learning Algorithms. In 2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS) (pp. 1 - 4). IEEE.

[13] Kekulanadara, K. M. O. V. K., Kumara, B. T. G. S., & Kuhaneswaran, B. (2021). “Air quality forecasting and classification using machine learning based on Indian urban pollutant”.

[14] Gopu P., Panda R.R., Nagwani N.K.Time series analysis using ARIMA model for air pollution prediction in hyderabad city of India Advances in Intelligent Systems and Computing, Springer, Singapore (2021), pp. 47-56, 10.1007/978-981-33- 6912-2_5

[15] Murugan, R., & Palanichamy, N. (2021). “Smart city air quality prediction using machine learning”. In 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). https://doi.org/10.1109/ICICCS51141.2021.9432074

[16] Gowri, D., & Anandhasilambarasan, D. (2021). “Prediction of air pollution in smart cities using machine learning techniques. International Journal for Research in Applied Science and Engineering Technology”, 9(12). https://doi.org/10.22214/ijraset.2021.39241

[17] Madan, T., Sagar, S., & Virmani, D. (2020). “Air quality forecasting and classification using machine learning based on Indian urban pollutant”. In 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN).

https://doi.org/10.1109/ICACCCN51052.2020.9362912

[18] Dobrea, D., et al. (2020). "Air quality prediction using machine learning algorithms – A review".

[19] Halsana, S. (2020). "Air quality prediction model using supervised machine learning algorithms. International Journal of Scientific Research in Computer Science, Engineering and Information Technology", 8, 190–201. https://doi.org/10.32628/CSEIT206435

[20] Chang Y.-S., Chiao H.-T., Abimannan S., Huang Y.-P., Tsai Y.-T., Lin K.-M. "An LSTM-based aggregated model for air pollution forecasting Atmos". Pollut. Res., 11 (8) (2020), pp. 1451-1463, 10.1016/j.apr.2020.05.015

[21] World Health Organization, "Ambient (outdoor) air pollution," WHO, 2020. [Online]. Available: https://www.who.int/airpollution/ambient/about/en/. [Accessed: 06-Jan-2026].

[22] Microsoft, "A fresher breeze: How AI can help improve air quality," Microsoft News, 20-Aug-2019. [Online]. Available: https://news.microsoft.com/europe/2019/08/20/a-fresherbreeze-how-ai-can-help-improve-air-quality/. [Accessed: 06-Jan-2026].

[23] G. Kaur, J. Gao, S. Chiao, S. Lu, and G. Xie, "Air Quality Prediction: Big Data and Machine Learning Approaches," International Journal of Environmental Science and Development, vol. 9, no. 1, pp. 8-16, 2018.

[24] "Artificial Intelligence for Cleaner Air in Smart Cities," PDX Engineering, 28-Mar-2019. [Online]. Available: https://www.pdxeng.ch/2019/03/28/artificial-intelligence-for-cleanerair-in-smart-cities/.[Accessed: 06-Jan-2026].

[25] S. Garzon, S. Walther, B. Deva, and A. Küpper, "Urban Air Pollution Alert Service for Smart Cities," 2018.

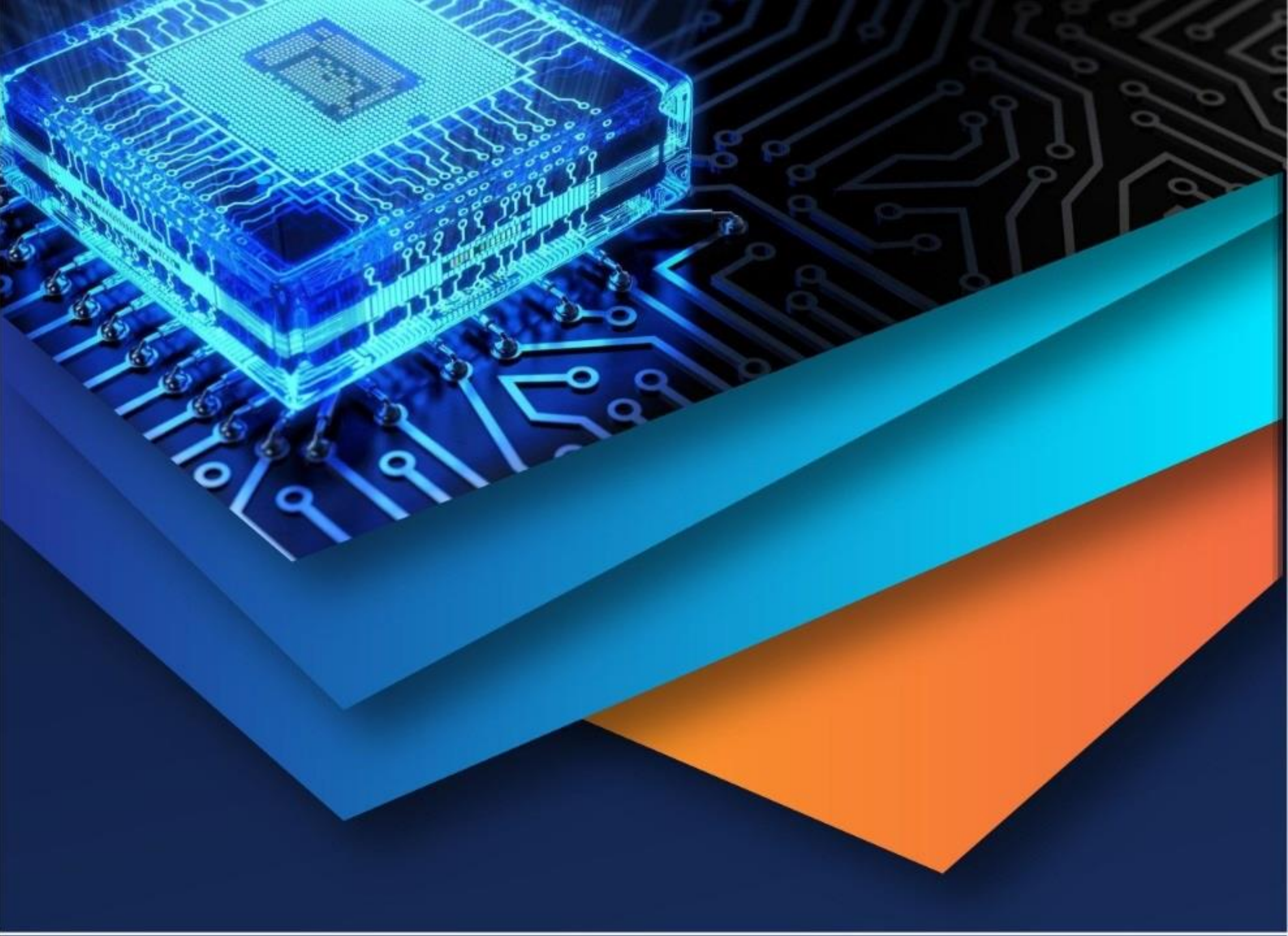
[26] A. Dorogush, V. Ershov, and A. Gulin, "CatBoost: Gradient Boosting with Categorical Features Support," arXiv preprint, vol. arXiv:1810.11363, 2018.

[27] N. Alharbi and B. Soth, "Roles and Challenges of Network Sensors in Smart Cities," IOP Conference Series: Earth and Environmental Science, vol. 322, no. 1, 2019.

[28] R. Sánchez-Corcuera, A. Nuñez-Marcos, J. Sesma-Solance, A. Bilbao Jayo, R. Mulero, U. Zulaika, and A. Almeida, "Smart Cities Survey: Technologies, Application Domains and Challenges for the Cities of the Future," International Journal of Distributed Sensor Networks, vol. 15, no. 6, 2019.

[29] "Smart Cities Air Quality," Aeroqual. [Online]. Available: https://www.aeroqual.com/smart-cities-air-quality. [Accessed: 06-Jan-2026].

[30] "Beijing Multi-Site Air Quality Data," UCI Machine Learning Repository. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Beijing+Multi-Site+Air-Quality+Data. [Accessed: 06-Jan-2026].



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY