International Journal of Engineering Research and Science & Technology

IJERST

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2(1) (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com

Peer Reviewed,Referred & Indexed Journal

Research Paper

URBAN AMBIENT AIR QUALITY INDEX CLASSIFICATION USING MULTI-POLLUTANT SENSOR READINGS

[1]K Srikanth, [2] D Akhila,[3] A Ashritha , [4] J Gunguli
[1]AssistantProfessor, [234]Students
Department of Computer Science and Technology
Siddhartha Institute of Technology & Sciences, Narapally
srikanthk@siddhartha.org.in, 23TQ1A05C8@siddhartha.co.in, 23TQ1A05G6@siddhartha.co.in, 23TQ1A05F4@siddhartha.co.in,

Abstract

Urban air pollution has emerged as a critical environmental and public health challenge, especially in rapidly growing cities. Continuous monitoring and accurate assessment of air quality are essential for mitigating its adverse effects. This project presents a machine learning-based approach for the classification of the Urban Ambient Air Quality Index (AQI) using data obtained from multi-pollutant environmental sensors.

The proposed system analyzes key atmospheric pollutants, including particulate matter (PM2.5 and PM10), carbon monoxide (CO), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and ozone (O_3). These pollutant concentrations are collected through sensor networks and used as input features for model development. Prior to model training, the dataset undergoes preprocessing steps such as missing value treatment, normalization, and feature optimization to enhance accuracy and reliability.

Various classification algorithms, such as Decision Trees, Random Forest, Support Vector Machines, and Neural Networks, can be employed to categorize air quality into standard AQI levels (e.g., Good, Moderate, Poor, etc.). The system aims to provide real-time and accurate predictions of air quality, enabling authorities and individuals to take informed decisions for health protection and environmental management.

Overall, this work contributes to the development of intelligent air quality monitoring systems by leveraging machine learning techniques, thereby supporting sustainable urban living and pollution control strategies.

I. Introduction

Air pollution has become one of the most critical environmental challenges in urban areas, mainly due to rapid industrialization, population growth, and the increasing number of vehicles. The release of harmful pollutants into the atmosphere significantly affects air quality and poses serious health risks to humans. Exposure to polluted air can lead to respiratory illnesses, cardiovascular diseases, and other long-term health complications. Therefore, monitoring and analyzing air quality is essential to ensure public safety and promote a sustainable environment.

The Air Quality Index (AQI) is a widely used indicator that represents the level of air pollution in a specific region. It is calculated based on the concentration of major pollutants such as particulate matter (PM2.5 and PM10), carbon monoxide (CO), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and ozone (O_3). These pollutants originate from various sources, including vehicle emissions, industrial processes, construction activities, and the burning of fossil fuels. Higher AQI values indicate deteriorating air quality and increased health hazards.

Traditional air quality monitoring systems depend on large and expensive monitoring stations, which often provide limited coverage and are not suitable for real-time, large-scale monitoring. With recent advancements in sensor technologies and data analytics, it has become possible to collect pollution data from multiple low-cost sensors and process it efficiently. This has opened new opportunities for applying machine learning techniques to analyze air quality data and improve prediction accuracy.

This project, titled *Urban Ambient Air Quality Index Classification Using Multi-Pollutant Sensor Readings*, aims to develop a system that utilizes sensor data to classify air quality into different AQI categories such as Good, Satisfactory, Moderate, Poor, Very Poor, and Severe. Machine learning algorithms are employed to analyze pollutant concentrations and predict the corresponding AQI category with improved efficiency and accuracy.

The proposed system provides a scalable and cost-effective solution for air quality monitoring. It can assist environmental agencies, urban planners, and the general public by delivering timely insights into pollution levels. Additionally, it helps raise awareness about environmental conditions and supports informed decision-making to reduce pollution and improve urban living conditions. Future enhancements may include real-time data integration, advanced deep learning models, and deployment through web or mobile platforms for continuous monitoring.

II. Literature Survey

Air pollution has become a major environmental concern affecting urban populations across the globe, driven by factors such as industrialization, vehicular emissions, and rapid urban expansion. As a result, significant research efforts have been directed toward monitoring, analyzing, and predicting air quality to minimize its impact on human health and the environment. Earlier studies primarily relied on traditional air quality monitoring systems, which used large-scale monitoring stations and statistical methods to assess pollution levels. Although these systems provided reliable data, they were often expensive, had limited spatial coverage, and were not suitable for real-time analysis.

With the advancement of sensor technologies and data analytics, recent research has shifted toward the use of multi-pollutant sensor data combined with machine learning techniques for air quality analysis. Researchers have widely used pollutant parameters such as PM2.5, PM10, CO, NO_2, SO_2, and O_3 to estimate and classify the Air Quality Index (AQI). Various machine learning algorithms, including Decision Trees, Random Forest, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN), have been applied to model the relationship between pollutant concentrations and AQI levels. Among these, ensemble methods like Random Forest and Gradient Boosting have demonstrated higher accuracy and robustness in classification tasks.

In addition to traditional machine learning approaches, recent studies have explored deep learning techniques such as Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks for air quality forecasting. These models are particularly effective in analyzing time-series data and predicting future pollution trends based on historical records. Furthermore, hybrid models combining multiple algorithms and preprocessing techniques have been proposed to enhance prediction performance and improve classification accuracy.

Several research works have also conducted comparative analyses of different algorithms, concluding that advanced models such as XGBoost, LightGBM, and CatBoost provide improved accuracy and efficiency for AQI prediction using real-time environmental data. These studies emphasize the importance of integrating machine learning with sensor-based monitoring systems to develop intelligent and scalable air quality assessment solutions. Overall, the existing literature highlights the growing role of data-driven approaches in environmental monitoring and provides a strong foundation for developing efficient systems for Urban Ambient Air Quality Index Classification using multi-pollutant sensor readings.

III. System Analysis

System analysis is an essential phase in software development that involves understanding the existing system, identifying its limitations, and defining the requirements for the proposed system. In the context of air quality monitoring, traditional systems rely on large-scale monitoring stations that collect pollution data from limited locations. Although these systems provide accurate measurements, they suffer from several drawbacks such as high installation and maintenance costs, lack of real-time analysis, and limited geographical coverage. Additionally, conventional methods mainly use statistical approaches, which may not efficiently handle large and complex environmental datasets.

To overcome these limitations, the proposed system introduces a machine learning-based approach for classifying the Air Quality Index (AQI) using multi-pollutant sensor readings. The system is designed to collect and process data from multiple environmental sensors measuring pollutants such as PM2.5, PM10, CO, NO_2, SO_2, and O_3. These readings are used as input for analysis and prediction.

The functional requirements of the system include data collection, data preprocessing, feature selection, model training, AQI classification, and result visualization. Data preprocessing involves handling missing values, removing noise, and normalizing the data to improve model accuracy. The system then applies machine learning algorithms to classify air quality into predefined categories such as Good, Satisfactory, Moderate, Poor, Very Poor, and Severe.

Existing system

The existing air quality monitoring system mainly depends on large-scale monitoring stations installed at specific locations in urban areas. These stations collect data related to various pollutants such as PM2.5, PM10, CO, NO_2, SO_2, and O_3, and the collected data is analyzed using traditional statistical methods to determine the Air Quality Index (AQI). Government agencies and environmental organizations use these systems to monitor pollution levels and publish reports.

However, these systems are limited in number due to their high cost and infrastructure requirements. They provide data only for specific locations and often lack real-time analysis capabilities. As a result, they may not accurately represent pollution levels across different areas of a city.

Disadvantages of Existing System

- High installation and maintenance cost
- Limited geographical coverage
- Lack of real-time monitoring and prediction
- Dependency on centralized monitoring stations
- Inefficient handling of large and complex datasets
- Delayed data processing and reporting
- Inability to provide localized air quality information

Proposed System

The proposed system aims to develop a machine learning-based model for classifying the Urban Ambient Air Quality Index (AQI) using multi-pollutant sensor readings. Instead of relying only on large monitoring stations, the system utilizes data collected from multiple low-cost sensors that measure pollutants such as PM2.5, PM10, CO, NO_2, SO_2, and O_3.

The system processes the collected data using preprocessing techniques like data cleaning, normalization, and feature selection. Machine learning algorithms such as Decision Trees, Random Forest, Support Vector Machines (SVM), and others are used to classify air quality into categories like Good, Satisfactory, Moderate, Poor, Very Poor, and Severe.

The system can be implemented using tools such as Python, Pandas, and Scikit-learn, and it provides faster and more accurate classification of air quality levels.

Advantages of Proposed System

- Cost-effective compared to traditional systems
- Provides real-time or near real-time air quality classification
- Covers a wider geographical area using multiple sensors
- Improved accuracy using machine learning algorithms
- Efficient handling of large datasets
- Scalable and adaptable for future enhancements
- Helps in better decision-making for environmental management
- Increases public awareness about air pollution

IV. Methodology

The methodology of the proposed system consists of the following steps:

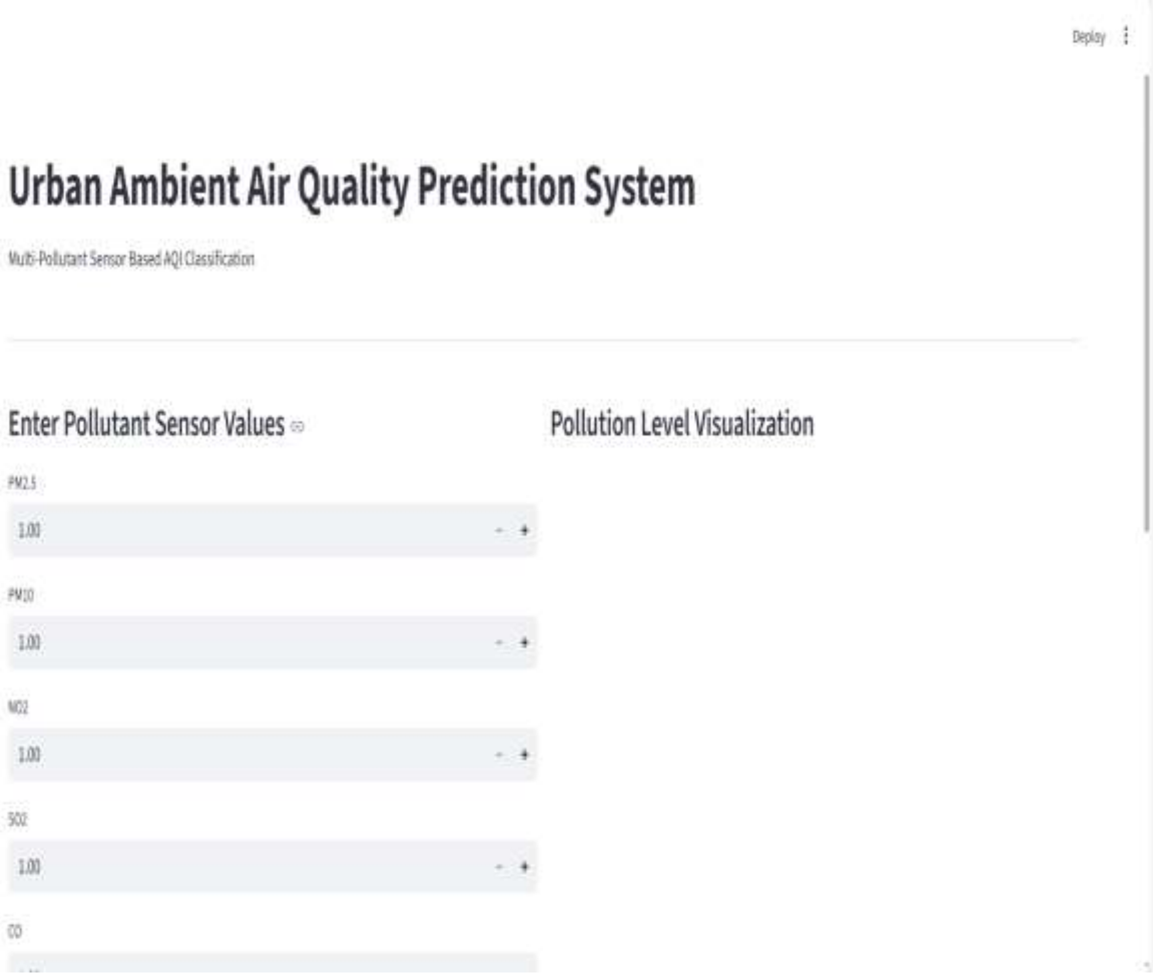
1. **Data Collection:**
   Air quality data is collected from multiple sensors measuring pollutants such as PM2.5, PM10, CO, NO_2, SO_2, and O_3.
2. **Data Preprocessing:**
   The collected data is cleaned by handling missing values, removing noise, and normalizing the dataset to improve accuracy.
3. **Feature Selection:**
   Important features (pollutants) are selected to enhance model performance and reduce complexity.
4. **Model Selection and Training:**
   Machine learning algorithms such as Decision Trees, Random Forest, and Support Vector Machines are applied to train the model using the prepared dataset.
5. **AQI Classification:**
   The trained model classifies air quality into predefined categories such as Good, Satisfactory, Moderate, Poor, Very Poor, and Severe.
6. **Evaluation:**
   The model's performance is evaluated using metrics like accuracy, precision, recall, and confusion matrix.
7. **Visualization:**
   Graphs and charts are generated using visualization tools to analyze pollution trends and model performance.

System Architecture

The system architecture for the Urban Ambient Air Quality Index (AQI) Classification Using Multi-Pollutant Sensor Readings project is designed as a multi-layered structure that enables efficient data collection, processing, analysis, and visualization. The process begins with the data collection layer, where air quality data is gathered from multiple environmental sensors measuring pollutants such as PM2.5, PM10, CO, NO_2, SO_2, and O_3. This raw data is then passed to the data preprocessing layer, where it undergoes cleaning, handling of missing values, noise removal, normalization, and feature selection to ensure data quality and improve model performance.

After preprocessing, the refined data is fed into the machine learning layer, which serves as the core component of the system. In this layer, various algorithms such as Decision Trees, Random Forest, and Support Vector Machines are used to train models and analyze the relationships between pollutant concentrations and air quality levels. The trained model then classifies the input data into predefined AQI categories such as Good, Satisfactory, Moderate, Poor, Very Poor, and Severe in the classification layer.

V. Result and Output



VI. Conclusion

Air pollution remains a major challenge in urban environments, posing serious risks to human health and ecological balance. Effective monitoring and accurate assessment of air quality are essential to address this issue and support sustainable development. The *Urban Ambient Air Quality Index (AQI) Classification Using Multi-Pollutant Sensor Readings* project presents a modern and efficient approach to air quality analysis by integrating sensor technology with machine learning techniques.

The proposed system utilizes multiple pollutant parameters such as PM2.5, PM10, CO, NO_2, SO_2, and O_3 to classify air quality into standard AQI categories. By applying data preprocessing methods and machine learning algorithms, the system improves the accuracy and efficiency of air quality classification compared to traditional monitoring methods. It overcomes limitations such as high cost, limited coverage, and delayed analysis associated with conventional systems.

Furthermore, the system provides a scalable and cost-effective solution that can support real-time monitoring and decision-making. It can assist environmental agencies, urban planners, and the general public in understanding pollution levels and taking necessary preventive measures. The use of data visualization also enhances user awareness and interpretation of air quality conditions.

References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time Object Detection in Drone Surveillance Using YOLOv5,” in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, “Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks,” in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, “Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology,” in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, “An advanced movie recommender using collaborative filtering and sentiment analysis,” International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, “Automatic crop recommendation system using LightGBM and decision tree machine learning models,” Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, “Smart agriculture through IoT and machine learning for analyzing carbon footprints,” in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, “Soil image classification using transfer learning approach: MobileNetV2 with CNN,” SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.