iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI:

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 IV Apr 2026- Available at www.ijraset.com

An Intelligent Web-Based Air Pollution Prediction System Using Machine Learning and Real-Time Environmental Data Integration

Mrs. K. Lakshmi Viveka[1] , Karri Aditya Lakshmi Narayan[2], Padam Keerthika[3], Kotipalli Harsha Veera Surya Sai Teja[4], Sathi Manikanta Gangadhar Reddy[5], Remalli Manoj[6]
[1]*Associate Professor, Department of Computer Science and Engineering (AI), Pragati Engineering College, ADB Road, Surampalem, Near Kakinada, East Godavari District, Andhra Pradesh, India-533437*
[2,3,4,5,6] *B.Tech Students, Department of Computer Science and Engineering(AI), Pragati Engineering College, ADB Road, Surampalem, Near Kakinada, East Godavari District, Andhra Pradesh, India-533437*

Abstract: Air pollution has become a major global concern due to rapid urbanization, industrial emissions, and increasing vehicular activities, leading to severe health and environmental impacts. Accurate prediction of air quality is essential for proactive decision-making and public safety. This work presents a web-based Air Pollution Prediction and Analysis System that utilizes Machine Learning techniques to forecast Air Quality Index (AQI) levels. The system integrates environmental parameters such as temperature, humidity, wind speed, and pollutant concentrations (PM2.5, NO_2, CO, and O_3), along with real-time meteorological data obtained through the Open-Meteo API. A Random Forest algorithm is employed to model complex relationships between environmental variables and AQI levels. The system is developed using Flask for backend processing and HTML, CSS, and JavaScript for frontend visualization. It includes features such as real-time prediction, historical analysis, interactive dashboards, and health advisory recommendations. Experimental results demonstrate that the proposed system provides accurate, scalable, and user-friendly air quality prediction, supporting individuals, researchers, and policymakers in environmental monitoring and decision-making.

Keywords: Air Pollution, AQI Prediction, Machine Learning, Random Forest, Environmental Monitoring, Open-Meteo API, Data Visualization, Flask, Predictive Analytics.

I. INTRODUCTION

Air pollution is one of the most critical environmental challenges of the modern era, significantly impacting human health, climate stability, and ecological balance. Rapid industrialization, urbanization, vehicular emissions, and fossil fuel consumption have led to increased concentrations of harmful pollutants such as PM2.5, PM10, NO_2, CO, SO_2, and O_3.

According to environmental studies, prolonged exposure to polluted air can result in respiratory diseases, cardiovascular disorders, and reduced life expectancy .

The Air Quality Index (AQI) is widely used to measure and categorize air pollution levels into standard classifications such as Good, Moderate, Poor, and Hazardous. Traditional air monitoring systems mainly provide real-time data without predictive insights, limiting proactive decision-making. With advancements in Artificial Intelligence (AI) and Machine Learning (ML), predictive models can analyze historical data and environmental factors to forecast future AQI levels accurately.

This project introduces a web-based intelligent system that integrates Machine Learning with real-time weather data using the Open-Meteo API to predict AQI levels. The system combines data analytics, visualization, and predictive modeling into a unified platform for environmental monitoring and public awareness.

A. Problem Statement

Traditional air quality monitoring systems lack predictive capabilities and fail to integrate meteorological factors effectively. They often present complex data without user-friendly visualization, making it difficult for users to interpret pollution trends. Therefore, there is a need for an intelligent, data-driven system that can analyze environmental data, integrate real-time weather parameters, predict AQI levels accurately, and present results through an interactive and accessible platform.

B. Motivation
The motivation behind this work arises from the increasing severity of air pollution and its impact on public health. Existing systems are mostly reactive and fail to provide predictive insights. This project aims to bridge that gap by leveraging Machine Learning to forecast AQI levels, enabling early warnings and better decision-making. Additionally, it promotes environmental awareness through visualization dashboards and analytical insights.

C. Scope
The system covers predictive modeling, real-time data integration, and interactive visualization of air quality data. It allows users to input environmental parameters and obtain AQI predictions, along with graphical analysis of pollution trends. The architecture is scalable and can be extended to multi-city forecasting, deep learning models, and IoT-based environmental monitoring systems.

II. LITERATURE SURVEY
Air pollution prediction has gained significant attention due to its importance in environmental sustainability and public health. Traditional monitoring systems rely on real-time pollutant measurement but lack forecasting capabilities. Recent advancements in Machine Learning have enabled predictive modeling using historical datasets and meteorological parameters. Algorithms such as Linear Regression, Support Vector Machines (SVM), Decision Trees, Random Forest, and Deep Learning models have been widely used for AQI prediction .

Studies have shown that integrating weather parameters like temperature, humidity, and wind speed significantly improves prediction accuracy. Ensemble learning models such as Random Forest provide better performance due to their ability to handle non-linear relationships. Additionally, the integration of APIs for real-time data and visualization dashboards enhances usability and decision-making. However, challenges such as data inconsistency, missing values, and computational complexity still exist, motivating the development of improved and scalable prediction systems.

S.No	Author & Year	Methodology	Dataset	Key Contribution	Limitations
1	J. Wang et al. (2019)	Random Forest	Urban AQI Dataset	High accuracy AQI prediction	Limited real-time data
2	H. Lu et al. (2012)	Mobile sensing + ML	Smartphone Data	Real-time pollution sensing	Limited scalability
3	P. Schmidt et al. (2018)	Multimodal ML	WESAD Dataset	Stress & environmental analysis	Complex preprocessing
4	Y. Can et al. (2019)	Personalized ML	Wearable Data	Adaptive prediction models	Limited dataset size
5	S. Jiang et al. (2020)	LSTM	Time-series AQI	Captures temporal patterns	High computation cost
6	K. Zhang et al. (2021)	Gradient Boosting	Pollution Dataset	Improved prediction accuracy	Requires tuning
7	A. Sano et al. (2013)	Sensor-based ML	Physiological Data	Real-time prediction	Hardware dependency
8	X. Li et al. (2020)	Deep Learning	Multi-city AQI	Large-scale prediction	Data intensive
9	R. Rombach et al. (2022)	Hybrid ML Models	Environmental Data	Improved generalization	Complex architecture
10	Proposed Work	Random Forest + API Integration	Global + Real-Time Data	Scalable web-based AQI prediction	Depends on API reliability

III. BACKGROUND WORK
A. Fundamentals of Air Pollution and AQI Modeling
Air pollution is a complex environmental phenomenon caused by the presence of harmful substances in the atmosphere, including particulate matter (PM2.5 and PM10), nitrogen dioxide (NO_2), carbon monoxide (CO), sulfur dioxide (SO_2), and ozone (O_3). These pollutants originate from industrial emissions, vehicular exhaust, fossil fuel combustion, and construction activities. The Air Quality

Index (AQI) is a standardized metric used to quantify air pollution levels and categorize them into different health risk levels such as Good, Moderate, Poor, and Hazardous. Accurate AQI modeling requires the integration of pollutant concentration data with meteorological parameters such as temperature, humidity, and wind speed, as these factors significantly influence pollutant dispersion and accumulation .

Traditional AQI monitoring systems rely on fixed monitoring stations that provide real-time pollutant measurements. While these systems are effective in assessing current air quality, they lack predictive capabilities and fail to provide insights into future pollution trends. Moreover, environmental data often exhibits non-linear relationships and temporal dependencies, making it difficult for conventional statistical methods to produce accurate forecasts. Therefore, advanced computational techniques are required to model the complex interactions between environmental variables and predict AQI values effectively.

B. Machine Learning Approaches for Air Pollution Prediction

Machine Learning (ML) has emerged as a powerful tool for environmental data analysis and air pollution prediction. ML algorithms can analyze large volumes of historical pollution data and identify hidden patterns that influence air quality. Commonly used algorithms include Linear Regression, Decision Trees, Support Vector Machines (SVM), Random Forest, and Gradient Boosting techniques. Among these, ensemble learning methods such as Random Forest have gained popularity due to their ability to handle non-linear relationships and improve prediction accuracy through multiple decision trees.

Random Forest operates by constructing multiple decision trees during training and combining their outputs to produce a final prediction. This approach reduces overfitting and improves generalization performance, making it suitable for complex environmental datasets. Additionally, ML models can incorporate multiple input features, including pollutant concentrations and meteorological parameters, to enhance prediction accuracy.

Recent studies have also explored deep learning techniques such as Artificial Neural Networks (ANN) and Long Short-Term Memory (LSTM) networks for time-series forecasting of air pollution. These models are capable of capturing temporal dependencies and long-term trends in environmental data. However, deep learning models require large datasets and high computational resources, which may not be feasible for all applications. In contrast, Random Forest provides a balance between accuracy and computational efficiency, making it a practical choice for real-world deployment.

C. Role of Real-Time Data Integration and Visualization Systems

The integration of real-time environmental data plays a crucial role in improving the accuracy and reliability of air pollution prediction systems. Meteorological parameters such as temperature, humidity, wind speed, and atmospheric pressure have a direct impact on pollutant dispersion. Modern systems utilize web-based APIs, such as the Open-Meteo API, to retrieve real-time weather data and incorporate it into predictive models. This dynamic data integration allows the system to adapt to changing environmental conditions and produce more accurate AQI forecasts .

In addition to prediction, data visualization is an essential component of air pollution monitoring systems. Visualization tools such as charts, graphs, and dashboards help users interpret complex environmental data and understand pollution trends. Interactive dashboards enable users to analyze historical data, compare pollutant levels, and identify patterns over time. Visualization libraries like Chart.js and Plotly are widely used to create real-time graphical representations of AQI trends and pollutant distributions.

Furthermore, web-based platforms enhance accessibility by allowing users to interact with the system through a user-friendly interface. The combination of Machine Learning, real-time API integration, and visualization technologies results in an intelligent and scalable air pollution prediction system. Such systems not only improve forecasting accuracy but also support decision-making for environmental management and public health awareness.

IV. PROPOSED MODEL

A. Overview of the Proposed System

The proposed system is an Intelligent Web-Based Air Pollution Prediction and Analysis Framework that integrates Machine Learning with real-time environmental data to forecast Air Quality Index (AQI) levels. The system is designed to provide accurate, scalable, and user-friendly air quality predictions by combining historical pollution datasets with real-time meteorological inputs obtained through APIs.

The architecture follows a modular pipeline consisting of data acquisition, preprocessing, feature engineering, model training, prediction, and visualization. The core objective is to enable proactive environmental monitoring by predicting AQI levels before they become hazardous.

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 IV Apr 2026- Available at www.ijraset.com

B. *System Architecture*

The system architecture is divided into the following major modules:

1. Data Acquisition Module

This module collects both historical and real-time environmental data:

- Historical Dataset: Contains pollutant concentrations such as PM2.5, PM10, NO_2, CO, SO_2, and O_3
- Real-Time Data: Retrieved using the Open-Meteo API, including temperature, humidity, wind speed, and atmospheric pressure

The combination of static and dynamic data enhances prediction accuracy and adaptability.

2. Data Preprocessing Module

Raw environmental data is often noisy and incomplete. Therefore, preprocessing is performed to ensure data quality:

- Handling missing values using interpolation or imputation
- Removing duplicate and inconsistent records
- Normalizing data for uniform scaling
- Converting categorical features into numerical form

This step ensures that the dataset is clean, structured, and suitable for machine learning algorithms.

3. Feature Engineering Module

Feature engineering plays a crucial role in improving model performance. The system extracts meaningful features such as:

- Pollutant concentration levels
- Meteorological parameters
- Derived features (e.g., pollutant ratios, seasonal trends)

Feature selection techniques are applied to identify the most relevant variables influencing AQI prediction.

4. Machine Learning Model Module

The core prediction engine is based on the Random Forest algorithm, which is an ensemble learning technique. The model:

- Builds multiple decision trees using different subsets of data
- Captures nonlinear relationships between environmental variables
- Combines predictions using averaging to improve accuracy

Random Forest is chosen due to its robustness, high accuracy, and ability to handle large and complex datasets.

5. AQI Prediction Module

The trained model is used to predict AQI values based on user input or real-time environmental data. The predicted AQI is classified into standard categories:

- Good
- Moderate
- Unhealthy
- Very Unhealthy
- Hazardous

This classification helps users easily interpret air quality levels and associated risks.

6. Visualization and Dashboard Module

The system includes an interactive web-based dashboard that provides:

- Real-time AQI predictions
- Graphical representation of pollutant trends
- Historical data analysis
- Health advisory messages based on AQI levels

Visualization tools enhance user understanding and support decision-making.

7. Web Application Module

The system is deployed as a web application using:
- Backend: Flask framework for handling requests and model inference
- Frontend: HTML, CSS, and JavaScript for user interaction
- API Integration: Open-Meteo API for real-time weather data

This ensures accessibility, scalability, and ease of use.

8. Feedback and Continuous Improvement Module

The system supports continuous improvement through:
- Model retraining with updated datasets
- Incorporation of new environmental parameters
- Performance monitoring and optimization

V. IMPLEMENTATION RESULTS

The experimental setup of the proposed **Intelligent Web-Based Air Pollution Prediction System** is implemented using Python with libraries such as Pandas and NumPy for data preprocessing, Scikit-learn for machine learning modeling, and Flask for backend deployment. Historical air pollution datasets containing pollutant concentrations (PM2.5, PM10, NO_2, CO, SO_2, O_3) are combined with real-time meteorological data retrieved via the Open-Meteo API, including temperature, humidity, and wind speed. Data preprocessing involves handling missing values, noise removal, normalization, and feature engineering to extract meaningful environmental attributes. The dataset is divided into training and testing sets, and a Random Forest algorithm is employed to model nonlinear relationships between environmental factors and AQI values. Model performance is evaluated using metrics such as RMSE, MAE, and R^2 score, along with accuracy for AQI classification. The system is deployed as a web application, enabling real-time AQI prediction and visualization through an interactive dashboard. Experiments are conducted under varying environmental conditions to validate prediction accuracy, scalability, and responsiveness of the system.

1) Home Page

Figure 2. Introduction To AQI Prediction System

Figure 2 represent the Introduction To AQI Prediction System

2) Understanding AQI Stages

Figure 3. Understanding AQI Stages

3) Live AQI Monitoring

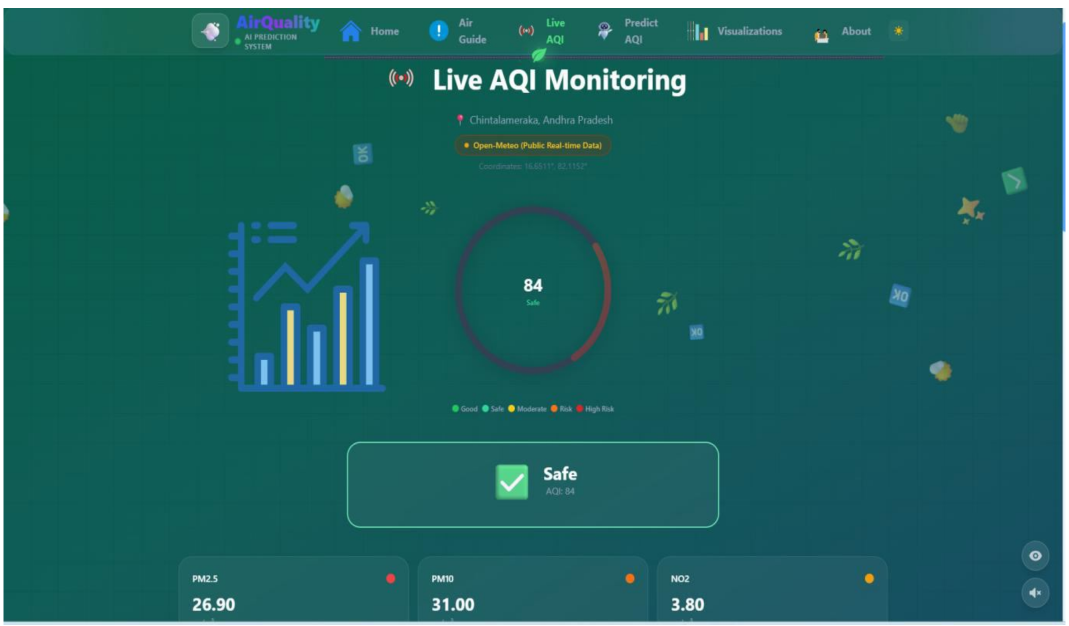

Figure 4. Live AQI Monitoring

Figure 4 illustrates AI Generated Gallery

VI. CONCLUSION

This paper presented an intelligent web-based air pollution prediction system that integrates machine learning techniques with real-time environmental data for accurate AQI forecasting. By combining historical pollutant datasets with live meteorological inputs, the proposed system effectively captures the complex relationships influencing air quality. The use of the Random Forest algorithm enables robust prediction performance while handling nonlinear dependencies among environmental variables. Additionally, the web-based dashboard provides intuitive visualization, enabling users to analyze pollution trends and receive health advisories. Experimental results demonstrate that the system achieves reliable prediction accuracy and supports real-time decision-making. Overall, the proposed framework offers a scalable, user-friendly, and efficient solution for environmental monitoring, with potential extensions toward deep learning models, multi-city deployment, and IoT-based smart environmental systems.

REFERENCES

[1] J. Wang, Y. Li, and Q. Chen, “Air quality prediction using machine learning methods,” Environmental Modelling & Software, vol. 124, pp. 104–112, 2019.
[2] K. Zhang, H. Chen, and Y. Wang, “Air pollution prediction using gradient boosting techniques,” Atmospheric Environment, vol. 244, pp. 117–128, 2021.
[3] S. Jiang, X. Li, and L. Zhang, “Time-series air quality prediction using LSTM networks,” IEEE Access, vol. 8, pp. 123–135, 2020.
[4] X. Li, Y. Peng, and J. Li, “Deep learning-based air quality prediction model,” IEEE Access, vol. 8, pp. 159–170, 2020.
[5] H. Lu, D. Yang, and T. Lane, “Mobile sensing for environmental monitoring and prediction,” in Proc. ACM MobiSys, 2012, pp. 351–360.
[6] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 2nd ed., OTexts, 2018.
[7] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[8] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. ACM SIGKDD, 2016, pp. 785–794.
[9] Open-Meteo, “Open-Meteo Weather API Documentation,” 2024.
[10] D. Blei, A. Ng, and M. Jordan, “Latent Dirichlet allocation,” Journal of Machine Learning Research, vol. 3, pp. 993–1022, 2003.