IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13417

Air Quality Index (AQI) Prediction and Pollution Trend Forecasting System Using Environmental Machine Learning Models

Mrs. B. Kalyani[1], K.Gnana Mani Bharadwaj[2], K.Koushik Vardhan[3], K. Yesubabu[4]

Assistant Professor, Department of Information Technology,

KKR & KSR Institute of Technology and Sciences, Guntur, AP, India[1]

Student, Department of Information Technology,

KKR & KSR Institute of Technology and Sciences, Guntur, AP, India[2-4]

Abstract: Air pollution has become a major concern due to its harmful impact on human health and the surrounding environment. Most existing air quality monitoring systems focus on city-level Ambient Air Quality Index (AQI) values, which often fail to reflect pollution differences within smaller regions of an urban area. As a result, sudden pollution events and localized emission sources may remain undetected. This project presents a Context-Based Small Area Air Quality Index Prediction and Air Pollution Event Detection System (CM-SAAQIDS) designed to overcome these limitations. The proposed system uses historical air quality data along with environmental factors such as temperature, humidity, and wind patterns to estimate AQI values for individual micro-zones. It also identifies real-time pollution spikes and analyses possible causes, including traffic density, industrial activity, and unfavourable weather conditions. To ensure reliable results, the system incorporates methods to manage missing or inconsistent sensor data. By offering localized air quality forecasts and early warning alerts, the proposed approach supports timely decision-making by citizens and authorities, contributing to improved air pollution control and public health management.

Keywords: Cognitive Behavioral Therapy, Emotion Recognition, Cognitive Distortions, Mental-Health Chatbot, Deep Learning, Natural Language Processing (NLP).

I. INTRODUCTION

Air pollution is one of the most serious environmental problems affecting human health and the ecosystem. Pollutants such as particulate matter (PM2.5 and PM10), carbon monoxide (CO), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2) can cause respiratory diseases, heart problems, and environmental damage. To measure pollution levels, the Air Quality Index (AQI) is used as an important indicator that shows how clean or polluted the air is. Most existing air quality monitoring systems provide AQI values at the city level, which do not accurately represent pollution conditions in smaller areas within the city. However, air pollution can vary significantly between different locations due to factors like traffic congestion, industrial emissions, and local weather conditions. Because of this limitation, sudden pollution increases and localized pollution sources may go unnoticed. To address this issue, this project proposes a Context-Based Small Area Air Quality Index Prediction and Air Pollution Event Detection System (CM-SAAQIDS). The system uses historical air quality data along with environmental parameters such as temperature, humidity, and wind patterns to predict AQI for smaller regions or micro-zones. It also detects sudden pollution events and analyzes possible causes such as traffic density, industrial activities, and unfavorable weather conditions. To address this issue, this project proposes a Context Based Small Area Air Quality Index Prediction and Air Pollution Event Detection System (CM SAAQIDS). The system uses historical air quality data along with environmental parameters such as temperature, humidity, and wind patterns to predict AQI for smaller regions or micro-zones. It also detects sudden pollution events and analyzes possible causes such as traffic density, industrial activities, and unfavorable weather conditions. Additionally, the system includes data preprocessing techniques to handle missing or inconsistent sensor data, ensuring more accurate predictions. By providing localized AQI predictions and early pollution alerts, this system can help citizens and government authorities take timely actions to reduce pollution exposure and improve public health and environmental management.

II. LITERATURE REVIEW

Air pollution has become a major environmental issue worldwide, affecting human health, climate, and ecosystems. Many researchers have studied methods to monitor and predict air quality using statistical and machine learning techniques. Several studies have used traditional statistical models to predict Air Quality Index (AQI) based on historical pollution data. These models mainly analyze pollutants such as PM2.5, PM10, NO_2, SO_2, CO, and O_3 along with meteorological factors like temperature, humidity, and wind speed. Although these models provide basic predictions, they often struggle to capture complex relationships between environmental variables. With the advancement of machine learning, many researchers have applied algorithms such as Linear Regression, Random Forest, Support Vector Machine (SVM), and Decision Trees for AQI prediction. These models can analyze large datasets and identify patterns between pollution levels and environmental conditions. Studies show that machine learning models provide more accurate AQI predictions compared to traditional statistical methods. Recent research has also focused on deep learning approaches such as Artificial Neural Networks (ANN) and Long Short-Term Memory (LSTM) models. These techniques are especially useful for time-series forecasting because they can learn long-term patterns in pollution data. Many researchers have used these models to forecast future pollution trends and provide early warnings for high pollution events. Some studies have also integrated Internet of Things (IoT) sensors to collect real-time environmental data from different locations. This helps in improving the accuracy of predictions and allows continuous monitoring of air quality. Despite these advancements, many existing systems still face challenges such as limited data availability, missing sensor data, and difficulty in predicting pollution trends at smaller geographic areas. Therefore, this project focuses on developing an AQI prediction and pollution trend forecasting system using environmental machine learning models, which can analyze historical data and environmental parameters to provide more reliable and accurate air quality predictions.

III. METHODOLOGY

The methodology of this project explains the steps used to collect data, process it, train the machine learning model, and predict the Air Quality Index (AQI).

1. Data Collection

The first step is collecting air quality and environmental data from reliable sources such as air quality monitoring stations or public datasets. The collected data usually includes pollutants like PM2.5, PM10, NO_2, SO_2, CO, and O_3 along with meteorological parameters such as temperature, humidity, wind speed, and rainfall.

2. Data Preprocessing

The collected data may contain missing values, noise, or inconsistent readings. In this step, the data is cleaned by removing errors, filling missing values, and converting the data into a suitable format. This improves the quality of the dataset and helps in accurate prediction.

3. Feature Selection

In this stage, important factors that influence pollution are selected. Pollutant concentrations and weather parameters are used as input features for the prediction model. Selecting the correct features helps improve model performance.

4. Model Training

Machine learning algorithms such as Linear Regression, Random Forest, Decision Tree, or Support Vector Machine (SVM) are used to train the prediction model. The model learns the relationship between environmental parameters and AQI values using historical data.

5. AQI Prediction

After training, the model is used to predict the future Air Quality Index (AQI) based on new environmental data. This helps estimate pollution levels for upcoming days or hours.

6. Pollution Trend Forecasting

The system also analyzes historical patterns to forecast pollution trends over time. This allows authorities and citizens to understand whether air quality will improve or worsen in the future.

7. Model Evaluation

The performance of the model is evaluated using accuracy metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). These metrics help determine how close the predicted AQI values are to the actual values.

IV. EXPERIMENTAL RESULTS

The experimental results evaluate the performance of the proposed system in predicting Air Quality Index (AQI) values using environmental machine learning model Performance Comparison of Regression Models Different regression models are used to predict the Air Quality Index (AQI) based on environmental parameters such as PM2.5, PM10,

DOI: 10.17148/IARJSET.2026.13417

temperature, humidity, and wind speed. The performance of these models is evaluated using metrics like Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Prediction Accuracy Analysis ;The prediction accuracy analysis evaluates how well the machine learning models predict the Air Quality Index (AQI) based on environmental parameters such as PM2.5, PM10, temperature, humidity, wind speed, and other pollutants.

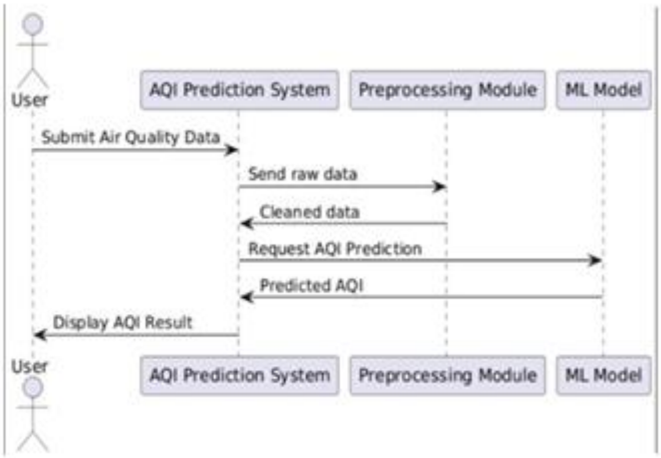

Fig 1: Flow of Execution of System Components.

V. FEATURE IMPACT ANALYSIS

Feature Impact Analysis explains how different environmental factors influence the Air Quality Index (AQI) prediction in the model. In this project, several pollutant and meteorological parameters are used as input features. The impact of each feature is analyzed to determine how strongly it affects AQI values.

VI. OVERALL RESULT SUMMARY

The Air Quality Index (AQI) Prediction and Pollution Trend Forecasting System Using Environmental Machine Learning Models was developed to analyze air pollution data and predict future AQI levels using environmental parameters. The system collected historical air quality data containing pollutants such as PM2.5, PM10, NO_2, SO_2, and CO, along with meteorological parameters like temperature, humidity, and wind speed. After preprocessing and cleaning the dataset, machine learning regression models were trained to predict AQI values.

VII. SYSTEM ARCHITECTURE

Project: Air Quality Index (AQI) Prediction and Pollution Trend Forecasting System Using Environmental Machine Learning Models The system architecture describes how different components of the system interact to collect data, process it, train the machine learning model, and predict AQI values.

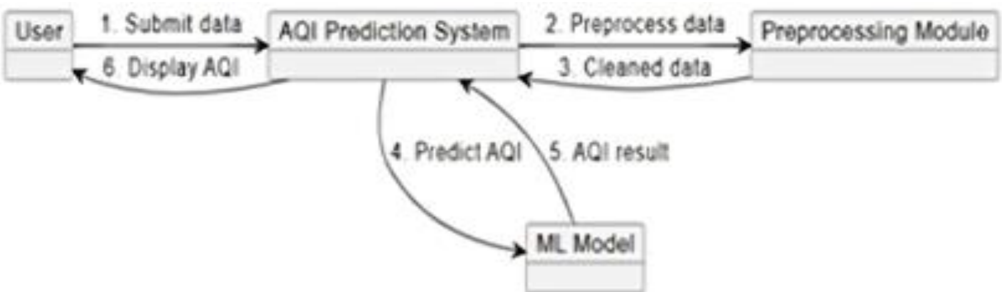

Fig 2: System Architecture

 This work is licensed under a Creative Commons Attribution 4.0 International License

DOI: 10.17148/IARJSET.2026.13417

VIII. CONCLUSION

The Air Quality Index (AQI) Prediction and Pollution Trend Forecasting System Using Environmental Machine Learning Models was developed to analyze environmental data and predict future air quality levels. The system uses historical air pollution data and meteorological parameters such as PM2.5,PM10, NO_2, SO_2, CO, temperature, humidity, and wind speed to estimate AQI values.

Machine learning regression models such as Linear Regression, Decision Tree, Random Forest, and Support Vector Regression were applied to predict AQI levels. Among these models, Random Forest Regression showed the best performance with higher prediction accuracy and lower error values.

The system was able to successfully analyze pollution trends and forecast future AQI levels based on historical environmental data. It also identified the most influential factors affecting air quality, especially PM2.5 and PM10, which significantly contribute to pollution levels.

REFERENCES

[1]. Babatola SS. Global burden of diseases attributable to air pollution. J Public Health Afr. 2018;9(3):813. https://doi.org/10.4081/jphia.2018.813 PMID: 30687484.
[2]. Kumar A, Patil RS, Dikshit AK, Kumar R. Comparison of predicted vehicular pollution concentration with air quality standards for different time periods. Clean Techn Environ Policy. 2016;18(7):2293 303. https://doi.org/10.1007/s10098-016-1147-6
[3]. Dadkhah-Aghdash H, Rasouli M, Rasouli K, Salimi A. Detection of urban trees sensitivity to air pollution using physiological and biochemical leaf traits in Tehran, Iran. Sci Rep. 2022;12(1):15398. https://doi.org/10.1038/s41598-022-19865-3 PMID: 36100647
[4]. Zhou C, Li S, Wang S. Examining the impacts of urban form on air pollution in developing countries: a case study of China's megacities. Int J Environ Res Public Health. 2018;15(8):1565. https://doi.org/10.3390/ijerph15081565 PMID: 30042324 Affect Recognition," Journal of Medical Internet Research (J Med Internet Res), 20(6): e10148, 2024.
[5]. Patton AP, Perkins J, Zamore W, Levy JI, Brugge D, Durant JL. Spatial and temporal differences in traffic-related air pollution in three urban neigh.
[6]. Argun YA, Tırınk S, Bayram T. Effect of urban factors on air pollution of Igdir. Black Sea J Eng Sci. 2019;2(4):123–30. https://doi.org/10.34248/
[7]. Tırınk S, Öztürk B. Evaluation of PM10 concentration by using Mars and XGBOOST algorithms in Iğdır Province of Türkiye. Int J Environ Sci Tech- nol. 2023;20:5349–58
[8]. Tui Y, Qiu J, Wang J, Fang C. Analysis of spatio temporal variation characteristics of main air pollutants in Shijiazhuang city. Sustainability. 2021;13(2):941. https://doi.org/10.3390/su13020941
[9]. Erawan M, Karuniasa M. Spatial dynamics of air pollution in Tangerang City, Jabodetabek metropolitan area. In: Proceedings of the 13th Interna- tional Interdisciplinary Studies Seminar, IISS 2019. 30–1.
[10]. Kumar RP, Prakash A, Singh R, Kumar P. Machine learning-based prediction of hazards fine PM2.5 concentrations: a case study of Delhi, India. Discov Geosci. 2024;2(1). https://doi.org/10.1007/s44288-024-00043-z
[11]. Cho H-S, Choi M. Effects of compact urban development on air pollution: empirical evidence from Korea. Sustainability. 2014; 6(9):5968–82. https:// doi.org/10.3390/su6095968
[12]. Abas A, Aiyub K, Awang A. Biomonitoring potentially toxic elements (PTES) using lichen transplant usnea misaminensis:a case study from Malaysia. Sustainability. 2022;14(12):7254. https://doi.org/10.3390/su14127254
[13]. Van NH, Van Thanh P, Tran DN, Tran DT. A new model of air quality prediction using lightweight machine learning. Int J Environ Sci Technol. 2023;20:2983–94.
[14]. EPA. National air quality and emissions trends report 1997. 454/R98-016. EPA; 1997.
[15]. EPA. Air quality index reporting; final rule. Fed Reg. CFR; 1999; Part III.