International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Environmental Data Analysis for Air-Quality Monitoring and Control

Mr.R.Madhavan
Assistant Professor, Department of Information Technology
Panimalar Engineering College,
Chennai, Tamil Nadu, India

Ram Kumar M
B.Tech – Information Technology
Panimalar Engineering College
Chennai, Tamil Nadu, India
Ramadmire138@gmail.com

Mothish S
B.Tech – Information Technology
Panimalar Engineering College
Chennai, Tamil Nadu, India
Mothish302004@gmail.com

Mukesh M
B.Tech – Information Technology
Panimalar Engineering College
Chennai, Tamil Nadu, India
Mukeshmuki1075@gmail.com

Abstract: Air pollution has become one of the most serious environmental challenges affecting human health and ecological balance worldwide. Rapid industrialization, urban expansion, and increasing vehicular emissions have significantly contributed to the deterioration of air quality. Conventional air quality monitoring systems mainly focus on real-time measurement of pollutants and lack the capability to predict future conditions, which is essential for taking preventive actions.
In this study, a machine learning-based air quality prediction system is proposed to estimate the Air Quality Index (AQI) using historical environmental data. The system utilizes algorithms such as Random Forest to analyze parameters including benzene concentration (C6H6), nitrogen oxides (NOx), nitrogen dioxide (NO2), relative humidity (RH), and absolute humidity (AH). Furthermore, a web-based interface is developed to allow users to input data and visualize prediction results through graphical representations. The experimental results demonstrate that the proposed model achieves reliable accuracy and can be effectively used for environmental monitoring and decision-making.

I.INTRODUCTION

Air quality plays a crucial role in maintaining public health and environmental sustainability. In recent years, the rapid growth of industries, urbanization, and transportation systems has led to a significant increase in air pollution levels. Harmful pollutants such as particulate matter, carbon monoxide, nitrogen oxides, and volatile organic compounds contribute to various health issues, including respiratory diseases, cardiovascular disorders, and reduced life expectancy.

Traditional air quality monitoring systems rely on sensor-based measurements to provide real-time data. Although these systems are effective in detecting current pollution levels, they do not offer predictive insights that are essential for proactive decision-making. Predicting air quality in advance enables authorities to implement control measures and helps individuals take necessary precautions.

With the advancement of artificial intelligence, machine learning techniques have emerged as powerful tools for analyzing large datasets and identifying hidden patterns. These techniques can be used to develop predictive models that estimate future air quality levels with high accuracy.

The objective of this work is to develop a machine learning-based air quality prediction system that can analyze environmental parameters and provide accurate AQI predictions. The system also includes a user-friendly web interface for data input and visualization, making it accessible and practical for real-world applications.

II.LITERATURE SURVEY

Several research works have been carried out in the field of air quality monitoring and prediction using advanced computational techniques. Mujahid Hussain et al. (2020) proposed a low-cost and energy-efficient air quality monitoring system integrated with IoT technologies. Their approach utilized Support Vector Machines (SVM) for data analysis, which improved system efficiency but showed limitations in handling complex and large-scale datasets for accurate prediction.

Zhiwen Hu et al. (2019) developed a real-time fine-grained air quality sensing network designed for smart city applications. The system employed Convolutional Neural Networks (CNN) to analyze spatial and temporal variations in air quality data. Although the model achieved improved performance in capturing patterns, it required high computational resources and was not suitable for lightweight implementations.

In another study, Majid Mansouri et al. (2023) focused on enhancing fault detection in air quality monitoring networks using Deep Neural Networks (DNN). Their model improved detection accuracy and system reliability; however, the complexity of deep learning models increased training time and computational requirements.

Similarly, Hajer Lahdhiri et al. (2019) proposed a sensor fault detection approach using kernel-based techniques combined with machine learning methods. While the system was effective in identifying anomalies in sensor data, it faced challenges in processing large datasets efficiently.

Raoudha Baklouti et al. (2017) introduced a fault detection mechanism for air quality monitoring systems using the k-Nearest Neighbors (k-NN) algorithm. This approach was simple and easy to implement but lacked the capability to model complex nonlinear relationships in environmental data.

III.METHODOLOGY

The proposed system is designed to predict air quality using a structured approach that includes data collection, preprocessing, feature selection, model training, and prediction.

The dataset used in this study consists of environmental parameters such as benzene concentration (C6H6), nitrogen oxides (NOx), nitrogen dioxide (NO2), relative humidity (RH), and absolute humidity (AH). These parameters play a significant role in determining air quality levels.

Data preprocessing is performed to handle missing values, remove inconsistencies, and normalize the dataset. This step ensures that the data is clean and suitable for training the machine learning model. Feature selection techniques are then applied to identify the most relevant parameters, reducing complexity and improving model efficiency.

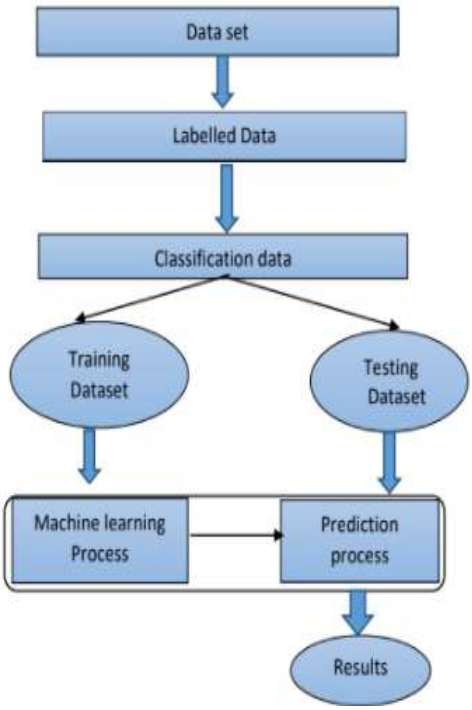
The Random Forest algorithm is used for model training due to its robustness and ability to handle complex data patterns. The dataset is divided into training and testing sets to evaluate the performance of the model.

Once the model is trained, it is used to predict AQI values based on new input data. The predicted results are displayed through a web-based interface, allowing users to easily interpret the outcomes.

A. System Architecture

The overall architecture of the proposed system is designed to illustrate the flow of data from user input to prediction output. The system consists of multiple components including data input, preprocessing, machine learning model, and result visualization.

The user provides environmental parameters through the interface, which are then processed and passed to the trained machine learning model. The model analyzes the input data and generates the predicted Air Quality Index (AQI). The output is displayed along with graphical representations for better understanding.



IV.MODEL FORMULATION

The model formulation stage plays a crucial role in developing an accurate and reliable air quality prediction system. In this work, a machine learning-based approach is adopted to model the relationship between environmental parameters and the Air Quality Index (AQI). The formulation of the model involves defining input variables, preprocessing the dataset, selecting an appropriate algorithm, and training the model for prediction.

A. Input Variable Selection

The selection of appropriate input variables is an essential step in developing an accurate prediction model. In this study, environmental parameters such as benzene concentration (C6H6), nitrogen oxides (NOx), nitrogen dioxide (NO2), relative humidity (RH), and absolute humidity (AH) are considered. These variables are chosen based on their significant influence on air quality and their availability in the dataset. The inclusion of relevant features ensures that the model captures the relationship between pollutant levels and the Air Quality Index (AQI) effectively.

B. Data Preprocessing

Data preprocessing is performed to improve the quality and consistency of the dataset before training the model. This process includes handling missing values, removing noise, and correcting inconsistencies in the data. Normalization techniques are applied to scale the features to a uniform range, which helps in improving the performance of the machine learning model. Proper preprocessing ensures that the dataset is suitable for accurate prediction.

C. Model Selection

The selection of an appropriate machine learning algorithm is crucial for achieving reliable prediction results. In this work, the Random Forest algorithm is used due to its ability to handle complex and nonlinear relationships in the data. It is an ensemble learning method that combines multiple decision trees to produce a more accurate and stable output. The algorithm also reduces the risk of overfitting, making it suitable for real-world datasets.

D. Training and Testing Process

The dataset is divided into two parts: training data and testing data. The training dataset is used to build the model by learning patterns from historical data, while the testing dataset is used to evaluate the model's performance. This approach ensures that the model can generalize well to new and unseen data. Performance metrics such as accuracy are used to assess the effectiveness of the model.

E. Prediction and Output Generation

After training, the model is used to predict the Air Quality Index based on new input data provided by the user. The predicted AQI value is then categorized into different pollution levels such as low, moderate, and high. The results are displayed through a user-friendly interface, along with graphical representations, making it easy for users to understand the air quality status.

V.EXPERIMENTAL ANALYSIS

A. Dataset Description

The experimental analysis is carried out using a dataset that contains various environmental parameters affecting air quality. The dataset includes features such as benzene concentration (C6H6), nitrogen oxides (NOx), nitrogen dioxide (NO2), relative humidity (RH), and absolute humidity (AH). These parameters are collected from publicly available environmental data sources and represent real-world pollution conditions. The dataset is used to train and evaluate the performance of the proposed machine learning model.

B. Data Splitting and Evaluation Setup

To evaluate the performance of the model, the dataset is divided into training and testing subsets. The training data is used to train the model, while the testing data is used to validate its performance on unseen data. This

approach ensures that the model is capable of generalizing well. A suitable split ratio is maintained to balance both training and evaluation processes.

C. Performance Metrics

The effectiveness of the model is evaluated using standard performance metrics such as accuracy and error rate. These metrics help in determining how closely the predicted values match the actual values. A lower error rate and higher accuracy indicate better model performance. The evaluation metrics provide a clear understanding of the model's reliability.

D. Results and Observations

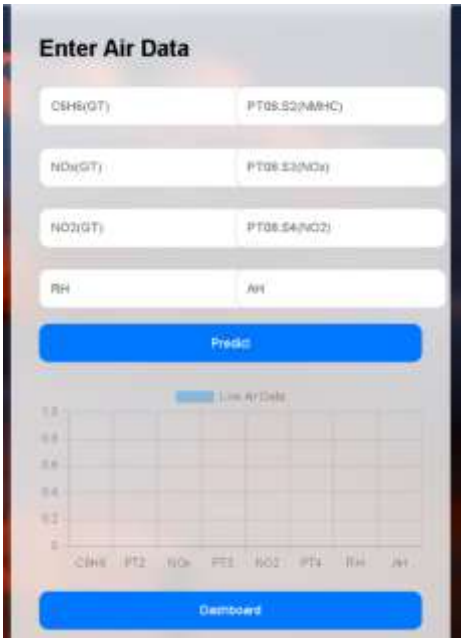
The experimental results show that the proposed system performs effectively in predicting air quality levels. The Random Forest model produces predictions that are closely aligned with actual AQI values. Minor deviations are observed due to variations in environmental data; however, the overall performance remains consistent.

It is observed that parameters such as NO2 and C6H6 have a significant impact on the predicted AQI. Higher values of these pollutants lead to increased air pollution levels. The model successfully captures the relationship between input features and output values.

VI.Results and Discussion

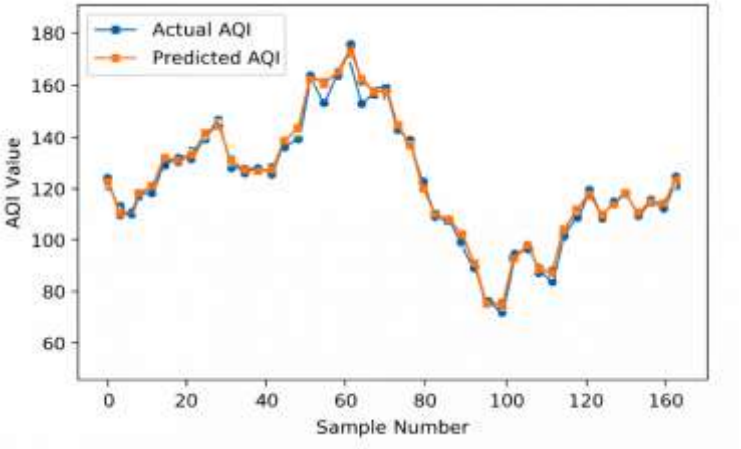
A. Prediction Performance

The proposed air quality prediction system was evaluated using a test dataset to measure its performance. The Random Forest model demonstrated strong predictive capability, producing AQI values that closely match the actual values. The model effectively captures the relationship between environmental parameters and air quality levels, resulting in reliable predictions. The overall performance indicates that the model is suitable for practical applications.



B. Accuracy Analysis

The accuracy of the model was assessed by comparing predicted AQI values with actual values from the dataset. The results show that the model achieves a high level of accuracy with minimal error. The use of ensemble learning in the Random Forest algorithm helps in reducing overfitting and improving generalization, which contributes to better prediction results.



C. Comparative Observation

A comparison between actual and predicted AQI values reveals that the model performs consistently across different data samples. Although minor deviations are observed in some cases due to fluctuations in environmental conditions, the predicted values remain close to the actual values. This demonstrates the robustness of the proposed system.

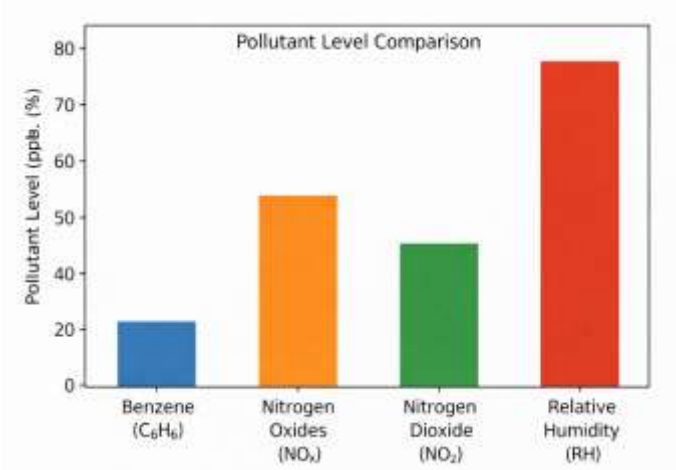
D. Impact of Environmental Parameters

The analysis shows that certain environmental parameters have a significant impact on air quality prediction. Pollutants such as nitrogen dioxide (NO2)

and benzene concentration (C6H6) contribute greatly to the increase in AQI values. Similarly, humidity levels also influence air quality. The model successfully identifies these relationships and incorporates them into the prediction process.

E. Graphical Representation

Graphical analysis plays an important role in understanding the results. The system generates visual representations such as bar charts and comparison graphs to display pollutant levels and predicted AQI values. These graphs help in identifying trends and patterns in the data.



VII. CONCLUSION AND FUTURE SCOPE

In this work, a machine learning-based air quality prediction system has been successfully developed to estimate the Air Quality Index (AQI) using environmental parameters. The system utilizes important features such as benzene concentration (C6H6), nitrogen oxides (NOx), nitrogen dioxide (NO2), relative humidity (RH), and absolute humidity (AH) to analyze air pollution levels.

The Random Forest algorithm is employed to build the predictive model due to its ability to handle complex data patterns and provide accurate results. The experimental analysis shows that the predicted AQI values are closely aligned with actual values, demonstrating the effectiveness and reliability of the proposed system.

Furthermore, the integration of a user-friendly web interface enhances the usability of the system by allowing users to input data and visualize results through graphical representations. Overall, the developed system provides an efficient and practical solution for air quality prediction and environmental monitoring.

Although the proposed system provides accurate predictions, there are several areas for further improvement and enhancement. Future work can focus on integrating real-time data using IoT sensors to improve the timeliness and accuracy of predictions.

Advanced machine learning and deep learning techniques such as neural networks can be implemented to further enhance prediction performance. The system can also be extended to include additional environmental parameters such as temperature, wind speed, and particulate matter (PM2.5 and PM10) for more comprehensive analysis.

In addition, the development of mobile and cloud-based applications can improve accessibility and allow users to monitor air quality from anywhere. The proposed system can also be integrated into smart city infrastructures to support large-scale environmental monitoring and decision-making.

REFERENCES

1.F. Chraim, Y. B. Erol and K. Pister, "Wireless gas leak detection and localization", IEEE Trans. Ind. Informat., vol. 12, no. 2, pp. 768-779, Apr. 2016.

2.C. Nandi, R. Debnath and P. Debroy, "Intelligent control systems for carbon monoxide detection in IoT environments" in Guide to Ambient Intelligence in the IoT Environment, Berlin, Germany:Springer, pp. 153-176, 2019.

3.N. H. Motlagh et al., "Low-cost air quality sensing process: Validation by indoor-outdoor measurements", Proc. 15th IEEE Conf. Ind. Electron. Appl., pp. 223-228, 2020.

4.N. H. Motlagh et al., "Toward massive scale air quality monitoring", IEEE Commun. Mag., vol. 58, no. 2, pp. 54-59, Feb. 2020.

5.F. Concas et al., "Low-cost outdoor air quality monitoring and sensor calibration: A survey and critical analysis", ACM Trans. Sensor Netw., vol. 17, no. 2, pp. 1-44, 2021.

6.B. Alfano et al., "A review of low-cost particulate matter sensors from the developers' perspectives", Sensors, vol. 20, no. 23, 2020.

7.S. Marco and A. Gutierrez-Galvez, "Signal and data processing for machine olfaction and chemical sensing: A review", IEEE Sensors J., vol. 12, no. 11, pp. 3189-3214, Nov. 2012.