

An IoT-Enabled Framework for Real-Time Air Quality Monitoring and AQI Prediction Using Machine Learning

Jaime Alfonso Flores Navas^{1,*}, Luigi Orlando Freire Martínez², Sara Toumi³, Moustafa Adel Darwish⁴, Ahmed Ali Qameeh⁵, Rajyalakshmi Pulicherla⁶

¹Department of Environmental Sciences, National Autonomous University of Mexico, Coyoacán, Mexico City, Mexico.

²School of Engineering and Applied Sciences, Technical University of Cotopaxi, Latacunga, Cotopaxi, Ecuador.

³Department of Hydrocarbon Production, Kasdi Merbah University, Ouargla, Ouargla Province, Algeria.

⁴Department of Physics, Faculty of Science, Tanta University, Tanta, Gharbia Governorate, Egypt.

⁵Department of Physics, University of Lahej, Sabr, Lahj Governorate, Yemen.

⁶Department of Electronics and Communication Engineering, Sri Venkatesa Perumal College of Engineering and Technology, Puttur, Andhra Pradesh, India.

greatwhiteshark@ciencias.unam.mx¹, luigi.freire@utc.edu.ec², sarahtoumi1992@gmail.com³, mostafa_ph@science.tanta.edu.eg⁴, ahmedqameeh@gmail.com⁵, rajyalakshmi.bose@gmail.com⁶

Abstract: Air pollution is a significant environmental issue that hurts the quality of life as well as the health of individuals. Air quality monitoring helps people understand pollution levels and take preventive measures. This paper develops an Internet of Things (IoT)-based air quality monitoring system that uses machine learning techniques to predict the Air Quality Index (AQI). As input data, the system considers environmental characteristics such as PM2.5, PM10, CO, NO2, and SO2 levels, as well as O2 levels, temperature, humidity, and wind speed. The system then processes these sensor values and feeds them into a machine learning model to predict the AQI. This paper uses the Random Forest regression algorithm to produce reliable forecasts. The predicted Air Quality Index (AQI) value is then categorized as Good, Moderate, Unhealthy, Very Unhealthy, or Hazardous. The Gradio programming language provides a user-friendly interface that makes it easy for users to enter sensor information. The system instantly displays the projected AQI and its category. This research shows how machine learning and Internet of Things concepts can be combined to build a smart air quality monitoring system.

Keywords: Air Quality Index (AQI); Machine Learning (ML); Random Forest; Internet of Things (IoT); Air Quality Monitoring; Wind Speed; Air Pollution; Sensor Information.

Received on: 15/06/2025, **Revised on:** 18/08/2025, **Accepted on:** 21/10/2025, **Published on:** 07/09/2026

Journal Homepage: <https://www.fmdbpublish.com/user/journals/details/FTSESS>

DOI: <https://doi.org/10.69888/FTSESS.2026.000786>

Cite as: J. A. F. Navas, L. O. F. Martínez, S. Toumi, M. A. Darwish, A. A. Qameeh, and R. Pulicherla, “An IoT-Enabled Framework for Real-Time Air Quality Monitoring and AQI Prediction Using Machine Learning,” *FMDB Transactions on Sustainable Environmental Sciences*, vol. 3, no. 3, pp. 143–152, 2026.

Copyright © 2026 J. A. F. Navas *et al.*, licensed to Fernando Martins De Bulhão (FMDB) Publishing Company. This is an open access article distributed under [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits copying, remixing, redistributing, transformation, and building upon the material in any medium so long as the original work is properly cited.

1. Introduction

*Corresponding author.

Air pollution is one of the most serious environmental issues for human health, ecosystems and overall quality of life in both urban and rural regions [1]. Rapid industrialization, increased vehicle numbers, population growth, construction activities, conventional farming methods, energy consumption and fossil fuel burning have contributed significantly to the degradation of atmospheric conditions. Monitoring harmful pollutant concentrations is increasingly important as cities expand and industrial activity grows [2]. Polluted air contains particulate matter and gaseous pollutants that adversely affect the respiratory and cardiovascular systems. Chronic exposure to poor air quality can cause breathing problems, asthma, respiratory infections, heart disease and other health problems. Therefore, timely monitoring and accurate air-quality prediction are essential to protect public health and enable effective environmental management [3]. The Air Quality Index (AQI) is an important indicator that communicates the overall air pollution condition in a simple, understandable form. Rather than requiring people to understand multiple pollutant concentrations at once, the AQI translates pollution data into a numerical value or category that indicates air quality in their area. Air quality is usually categorized as Good, Moderate, Unhealthy, Very Unhealthy and Hazardous, according to the relevant AQI standard [4]. These categories help people understand the health risks they may face and what to do. For example, when air quality is unhealthy, sensitive individuals may limit outdoor activities, and environmental authorities may investigate pollution sources and take appropriate control measures. Conventional air quality monitoring systems are usually based on fixed monitoring stations equipped with specialized instruments [5].

These stations can provide accurate pollutant measurements, but they can be expensive to install, operate and maintain. Infrastructure and operating costs may limit how many monitoring stations can be deployed. As a result, some sites may lack adequate spatial coverage to capture local variations in air pollution. But the emergence of the Internet of Things (IoT) offers a way around some of these constraints, allowing environmental sensors to gather and transmit data continuously. IoT-based monitoring systems can combine multiple sensors, communication technologies, data-processing platforms, and visualization interfaces to provide real-time or near-real-time environmental information [6]. An IoT-based air quality monitoring system consists of multiple connected components. Environmental sensors monitor pollutant concentrations and weather conditions. A processing unit collects the sensor measurements and formats the data for transmission or analysis. Communication can be via Wi-Fi, Bluetooth, cellular networks, LoRa or other wireless technologies, depending on the system design. The collected data can then be transmitted to a local computer, cloud platform or machine learning application for analysis [7]. The architecture supports continual monitoring and storage of historical observations that can later be used to develop predictive models. In this paper, researchers integrate IoT-based environmental monitoring and machine learning to predict the Air Quality Index (AQI). The system takes environmental parameters such as PM_{2.5}, PM₁₀, carbon monoxide (CO), nitrogen dioxide (NO₂), sulphur dioxide (SO₂), ozone (O₃), temperature, humidity, and wind speed as input features [8]. These variables provide information on both pollutant concentrations and meteorological conditions that can affect overall air quality. The collected data are processed and then fed into the machine learning model.

This preprocessing step is important because raw sensor measurements may have missing values, inconsistent measurements, noise, outliers or values that need transformation before they can be used effectively for prediction. Particulate matter is a key indicator of air pollution. PM_{2.5} refers to particles of aerodynamic diameter approximately 2.5 micrometres or less, while PM₁₀ refers to particles of aerodynamic diameter approximately 10 micrometres or less [9]. "Fine particles are sufficiently small to remain suspended in the atmosphere and to penetrate deep into the respiratory system. PM_{2.5} is especially important because long-term exposure to high levels can pose a serious health risk. PM₁₀ can also affect respiratory health and may come from sources such as road dust, construction activities, industrial processes and other combustion or mechanical activities [10]. Including both PM_{2.5} and PM₁₀ in the prediction system provides relevant information about particulate pollution. Another important pollutant the system looks at is carbon monoxide. CO is a colourless, odourless gas that is mainly produced by incomplete combustion. Increased CO concentrations can come from vehicle emissions, industrial combustion and fuel burning [11]. Nitrogen dioxide is linked to combustion processes and is often emitted from vehicles and industrial sources. High NO₂ levels can also form secondary pollutants and drive atmospheric reactions. SO₂ usually results from burning sulphur-containing fuels and some industrial processes [12]. Unlike many directly emitted pollutants, ground-level ozone is not emitted directly into the air. Instead, it forms in the atmosphere through chemical reactions among precursor pollutants under appropriate environmental conditions. Simultaneous monitoring of these pollutants allows for a more complete characterization of air quality [13].

Meteorological variables are also important because air pollution is not solely a function of emissions. Temperature, humidity and wind conditions can affect pollutant dispersion, chemical reactions and atmospheric stability. For example, wind speed can affect how quickly pollutants disperse from an emission source. When wind is low, pollutants tend to build up near their sources. Higher winds favour dispersion. Temperature can affect atmospheric chemical processes, and humidity can affect particulate behaviour and sensor measurements [14]. Thus, adding meteorological parameters to the machine learning model's features can improve performance by capturing complex correlations between environmental conditions and AQI. This approach helps identify relationships between input variables and AQI values. Traditional mathematical methods may require defining the relationships between variables beforehand. Machine learning algorithms can learn patterns from past data [15]. During training, the model receives environmental measurements and the corresponding AQI values. It learns the relationship between

the input features and the target variable. After training, the model can receive new sensor measurements and estimate the AQI. This prediction can then be translated into an understandable air quality category. For AQI prediction in this paper, researchers chose Random Forest regression. Random Forest is an ensemble machine learning method that builds multiple decision trees and combines their predictions to produce a result.

Instead of building a single decision tree, the algorithm builds a set of trees, each trained on a variation of the training data and different feature subsets [1]. Finally, it combines the individual predictions to produce the final regression output. This ensemble approach improves prediction stability and reduces overfitting compared with a single decision tree. Random Forest can also learn nonlinear relationships and interactions between environmental variables, which suits air quality prediction. One advantage of Random Forest is its ability to handle multiple input features without a simple linear relationship. Air quality results from complex interactions between pollutant concentrations and meteorological conditions [2]. For instance, the effect of one pollutant may depend on temperature, humidity or other pollutants. Random Forest can learn such nonlinear patterns from historical observations. The algorithm can also output feature importance, which identifies the most significant environmental parameters for AQI prediction [3]. This information may help explain the model's behaviour and improve future system development. The environmental dataset must be properly preprocessed before training the Random Forest model. First, identify missing values and invalid observations. Missing data can result from sensor failures, communication problems, maintenance activities or incomplete historical records. Depending on the type and amount of missing data, you can use appropriate imputation methods. Sometimes records with very few missing values may be discarded. Statistical or data-driven approaches may estimate numerical values [4]. The selected approach should minimize bias in the training dataset. After cleaning the data, perform exploratory analysis to understand the environmental variables' characteristics. You can calculate statistical measures such as min, max, mean, median and standard deviation for all parameters.

They can also use visualization techniques to examine distributions and identify unusual observations. Correlation analysis can identify relationships between pollutant concentrations and AQI. Such analysis can provide useful information before model training and could reveal potential data-quality issues [5]. It also helps researchers determine if some features contain redundant or highly related information. Researchers should examine outliers carefully during preprocessing. Environmental datasets may include very high pollutant concentrations during pollution episodes, and researchers should not automatically discard these as errors [6]. A real pollution event can be a good training example for the model to recognize dangerous conditions. Therefore, outlier detection should distinguish between measurement errors and valid extreme observations. An extreme value may be retained or rejected based upon sensor calibration, domain knowledge, statistical analysis, and comparison to neighbouring observations. The dataset is split into training and testing sets to evaluate the model's predictive capability [7]. The training dataset is used to learn the relationship between the environmental parameters and AQI; the testing dataset is reserved to evaluate the trained model's performance on unseen observations. Select a suitable train-test ratio based on the dataset size. Cross-validation can also improve model development by providing a more reliable performance estimate. Avoid data leakage, which occurs when information from the test set accidentally influences model training [8]. Various regression metrics can evaluate the Random Forest regression model's performance. The mean absolute error (MAE) is a measure of the absolute difference between the predicted AQI and the actual AQI. Mean Squared Error (MSE) is more sensitive to larger prediction errors because it squares the differences. To get an error measure in the same units as AQI, take the square root of MSE, which is called the root mean squared error (RMSE).

R-squared, or the coefficient of determination, measures how well the model explains variation around the outcome variable [9]. Using multiple metrics gives a more complete picture of model performance than a single metric. The system maps the predicted numerical AQI value to predefined air quality categories. The category thresholds must match the selected AQI standard and remain consistent across the system [10]. The numerical prediction provides a quantitative estimate, and the category offers a simpler interpretation for users. For example, a low AQI forecast may indicate Good air quality, while higher values may indicate Moderate, Unhealthy, Very Unhealthy, or Hazardous air quality. Define the exact category ranges based on the paper's adopted standard, rather than arbitrary thresholds. The user interface is a significant part of the proposed system because machine learning predictions are most useful when they are easily accessible. This paper uses Gradio to build a simple, intuitive user interface [11]. With Gradio, you can quickly build an interactive interface for your machine learning app. Users enter environmental sensor values in the respective input fields, and the system processes them to provide the predicted AQI. The interface displays the numerical AQI value and the corresponding air quality category [12]. This makes the system more user-friendly for users without technical knowledge of machine learning or environmental data analysis. The proposed interface can be extended to accept data from real IoT sensors, rather than requiring users to input measurements manually. In a fully integrated deployment, sensor nodes could continuously collect measurements of PM2.5, PM10, CO, NO2, SO2, O3, temperature, humidity, and wind speed. The data may be transferred to the machine-learning application over a suitable communication network [13]. The prediction system then automatically generates the corresponding updated AQI values at predefined intervals. This would transform the prototype into an operational, real-time or near-real-time air quality monitoring platform. Combining IoT and machine learning offers several potential benefits.

IoT technology enables distributed environmental data collection, and machine learning enables automated data analysis and prediction [14]. Together, they can reduce the need to interpret large datasets manually. A network of low-cost sensor nodes can be deployed at different locations to provide localized air-quality information. Machine learning algorithms could then use the incoming measurements to generate predictions for each monitoring location. Such systems could support smart-city initiatives, environmental monitoring, public-health applications, industrial pollution management and community-based environmental awareness. However, several challenges must be addressed when implementing an IoT-based air quality prediction system. Sensor accuracy and calibration are especially important, as inaccurate measurements can directly affect machine-learning predictions. Low-cost sensors can be sensitive to temperature, humidity, cross-sensitivity to other gases, and other environmental conditions. Regular calibration and validation against reference monitoring stations might enhance data reliability [15]. Communication failures may also cause missing observations. Limited computational resources may limit the ability to perform complex processing at the sensor level. The system design must account for these considerations. Another important consideration is model reliability. The quality and representativeness of the training data determine a machine learning model's performance. If the training dataset lacks enough examples of extreme pollution events, the model will not be able to predict hazardous conditions [1]. Similarly, a model trained on data from one geographic region may perform worse in another, as emission sources, climate, topography, traffic patterns, and seasonal conditions can differ. Models should therefore be periodically evaluated and retrained with updated local data as needed.

Meanwhile, IoT-based environmental monitoring must consider data security and privacy [3]. Air quality measurements are collected by IoT devices that are connected to networks and could pose potential cybersecurity risks, though generally not considered personal data. Unlawful access to sensor devices, measurement manipulation or communication disruption can compromise the monitoring system's reliability [5]. Therefore, the system architecture should include proper authentication, communication security, access control and device management mechanisms. The proposed paper demonstrates the power of combining environmental sensing, IoT concepts, machine learning and interactive visualization into a single application. Random Forest regression is a flexible approach to predicting AQI using multiple environmental variables [7]. The Gradio interface makes the prediction process easy to use. The system can also serve as a prototype for more advanced air quality monitoring platforms, where sensor data are acquired automatically, and predictions are made continuously. The paper also opens up further development possibilities. Other machine learning algorithms, such as linear regression, Decision Trees, Random Forests, Gradient Boosting, Support Vector Regression, XGBoost and artificial neural networks, can be tried and compared. Hyperparameter optimization can improve prediction accuracy [9]. Moreover, time-series data can be incorporated, since air pollution often shows temporal trends linked to traffic, industrial operations, weather conditions and seasons. Instead of estimating AQI measurements, this work could use recurrent neural networks and other temporal models to predict future AQI [2]. Another potential improvement is adding geographical information. Combining sensor measurements with location information could create air quality maps that show spatial variations in pollution.

These maps can help users find areas with more pollution and support environmental planning. Cloud infrastructure could store historical measurements, enable remote access and support large-scale sensor networks. Mobile applications can send personalized notifications when predicted AQI levels exceed certain thresholds [4]. To summarise, monitoring air pollution is essential to protecting human health and environmental quality. Traditional monitoring stations provide useful information, but integrating IoT technology can increase the flexibility and coverage of environmental data collection. Machine learning can further improve these systems by automatically identifying relationships among pollutant concentrations, meteorological conditions, and the AQI. The proposed system uses PM_{2.5}, PM₁₀, CO, NO₂, SO₂, O₃, temperature, humidity, and wind speed as input parameters and predicts the numerical AQI using Random Forest regression [6]. The system then converts the predicted value into a readable air quality category and displays it through a Gradio-based user interface. The proposed system shows that machine learning and IoT can be effectively combined to build an intelligent air quality monitoring platform. The system helps users understand current environmental conditions and take preventive measures by providing fast AQI predictions and easy, category-based information [8]. Key challenges remain sensor accuracy, data quality, model generalization, communication reliability and system security. However, these can be addressed through proper calibration, pre-processing, validation, safe communication, and continuous model improvement. Future integration with real-time IoT sensors, cloud platforms, advanced machine learning algorithms, geographical visualization and automated alerts can further enhance the system's usefulness [10]. Hence, the proposed approach provides a practical basis for developing smart, scalable and data-driven air quality monitoring systems that can support environmental management and public-health awareness.

2. Literature Survey

Air quality monitoring and prediction have become important research areas because air pollution increasingly affects human health and the environment. In many parts of the world, the atmosphere has worsened due to rapid urbanization, industrialization, increased vehicle use and energy consumption. Recent studies have therefore examined machine learning and Internet of Things (IoT) technologies for continuous environmental monitoring and AQI prediction. Kalia and Ansari [1] proposed an IoT-based air quality monitoring system in which sensors collect environmental parameters and send the

measurements to cloud platforms for real-time observation. Such systems improve the availability of pollution information and give users remote access to measurements. However, Mishra et al. [2] noted that traditional monitoring approaches primarily focus on collecting and displaying environmental measurements and may not offer adequate predictive capabilities. Thus, integrating predictive algorithms with IoT systems can enable more intelligent, proactive air quality assessment. Machine learning algorithms are widely used to predict air pollution levels because they can identify complex relationships among environmental variables. Unlike traditional statistical methods, machine learning methods are not limited to predefined input-output relationships and can learn patterns from historical observations.

Malleswari and Mohana [3] utilized Support Vector Machine (SVM) for AQI prediction and validated that the algorithm can efficiently analyze pollution patterns. SVM can handle non-linear relationships with suitable kernel functions and perform well when the dataset has clearly distinguishable patterns. However, its performance depends on proper parameter selection, kernel configuration and data preprocessing. Because the dataset is very large, computational requirements may also increase. Therefore, Decision Tree and Random Forest algorithms have been considered as alternative approaches. Mihirani et al. [4] showed that Random Forest can provide reliable prediction performance and effectively handle nonlinear relationships and multiple environmental features. Wireless sensor networks and IoT technologies have further extended environmental monitoring systems, enabling sensors to collect data continuously from various locations. Distributed sensor nodes can measure pollutant concentrations and transmit observations to a central processing platform via wireless networks. Saad et al. [5] proposed a smart air quality monitoring system with multiple sensors that measure pollutants such as PM_{2.5}, PM₁₀, and CO and provide real-time information. However, real-time measurements may not be enough to predict future pollution. Furthermore, massive sensor data can be hard to interpret manually, especially when observing multiple pollutants and meteorological parameters simultaneously. Hence, integrating IoT systems with machine learning algorithms can strengthen analytical capabilities by converting raw sensor observations into AQI predictions, pollution categories and possible early warnings.

Recent studies have also explored deep learning techniques for air quality forecasting because they can learn complex, non-linear relationships from large datasets. Neural networks can process numerous environmental variables and discover patterns that traditional machine learning techniques struggle to model. Ramadan et al. [6] used neural networks for pollution prediction and reported better performance than traditional approaches. Deep learning methods can be particularly helpful when large amounts of historical, time-dependent environmental data are available. However, deploying them typically requires substantial computational resources, careful parameter optimization and adequate training datasets. Muljana et al. [7] explored Random Forest and ensemble-based methods for AQI prediction using environmental parameters and obtained reliable performance with relatively lower computational requirements. This comparison indicates that choosing a prediction algorithm should consider factors such as accuracy, dataset size, complexity, interpretability, training time and available deployment hardware. Researchers have also proposed hybrid approaches that combine IoT-based environmental data collection with machine learning techniques to improve prediction performance. These frameworks can continuously collect sensor measurements and automatically feed them into predictive models for analysis. Ma et al. [8] proposed a hybrid framework that combines sensor measurements and machine learning algorithms to improve air pollution prediction. The real-time data acquisition combined with predictive analytics can provide a more responsive monitoring system, as the model can continuously process updated environmental conditions.

In addition to the traditional machine learning approaches, transformer-based architectures and advanced data analytics techniques have recently gained attention in environmental monitoring. Kaushik et al. [9] showed that more advanced machine learning approaches can improve prediction performance when large-scale datasets are available. But more advanced models may require more computational power, more training data and more complex implementation. Thus, practical systems face a trade-off among prediction accuracy, computational efficiency, data availability, model complexity and deployment requirements. Researchers have also addressed air quality classification issues with simple machine learning algorithms such as K-Nearest Neighbours (KNN). KNN classifies a new observation by examining its neighbours in the feature space. The algorithm is relatively simple and requires little model training. This makes it suitable for some small-scale classification applications. Mujtaba et al. [10] reported that KNN can perform well on small datasets when the feature characteristics are suitable. However, performance can decline as the number of environmental variables increases because distance calculations become more difficult in high-dimensional feature spaces. The choice of K and the distance measure can affect classification accuracy. Prediction can also be computationally costly when many training observations must be compared with each new sample. These limitations suggest that although KNN may be useful in some specific applications, it may not always be the most suitable choice for large-scale environmental datasets.

The selected articles collectively emphasize the importance of integrating IoT-based data collection with appropriate machine learning techniques for effective air quality monitoring and prediction. IoT provides the sensing and communication infrastructure to collect environmental measurements, and machine learning continuously provides the analytical power to detect trends and estimate current or future air quality conditions. Each algorithm has its own strengths and weaknesses. SVMs

can provide accurate predictions, but they require careful parameter tuning; decision tree models can provide interpretable classification rules, but they are prone to overfitting; and KNN is simple, but it can struggle with computational cost and dimensionality in large datasets. Deep learning methods can learn complex patterns but typically require massive amounts of data and computing power. Random Forest offers a useful compromise between prediction performance, nonlinear modelling ability, robustness, and computational efficiency. Therefore, a comparative study of different algorithms under the same experimental conditions is necessary to find the best model for a particular AQI dataset. This evaluation can help develop reliable, scalable and cost-effective intelligent air quality monitoring systems.

3. Methodology

The methodology predicts the Air Quality Index (AQI) using machine learning based on environmental sensor data. The system collects parameters such as PM2.5, PM10, CO, NO₂, SO₂, O₃, temperature, humidity, and wind speed. The system preprocesses these values and feeds them into a Random Forest regression model for training and prediction. The trained model predicts the AQI value and displays the pollution category through a user-friendly Gradio interface. Environmental parameters such as PM2.5, PM10, CO, NO₂, SO₂, O₃, temperature, humidity, and wind speed are collected as input data (Figure 1).

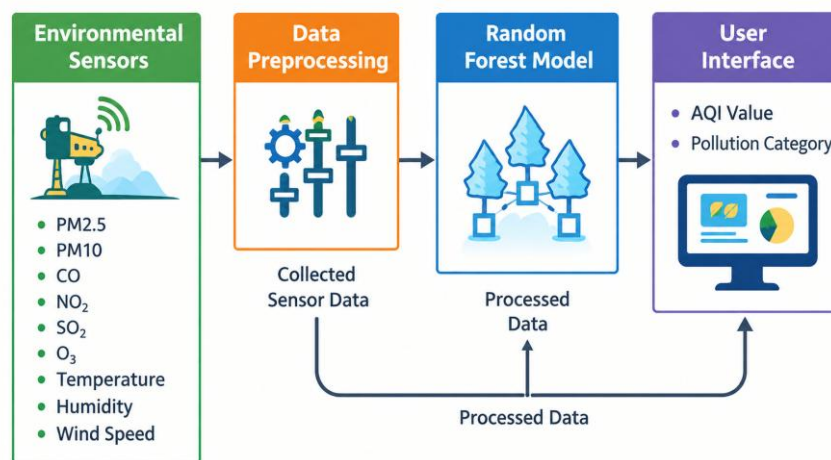


Figure 1: System architecture of the proposed air quality index (AQI) prediction system

The collected data is preprocessed and fed into a Random Forest regression model to analyze pollution patterns and predict AQI. The predicted AQI and pollution category are then displayed through a user-friendly interface for easy understanding of air quality conditions. This step is important in the air quality prediction system because it helps identify useful environmental parameters from the collected data. The system collects raw sensor data, including PM2.5, PM10, CO, NO₂, SO₂, O₃, temperature, humidity, and wind speed (Figure 2).

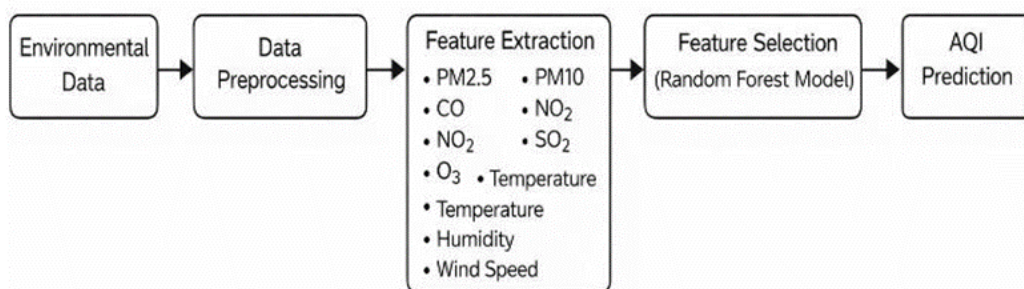


Figure 2: Feature extraction process

These parameters play a major role in determining environmental air pollution levels. Before feature extraction, the collected data is cleaned and normalized during the preprocessing stage to remove noise and inconsistencies. After preprocessing, the most relevant environmental parameters are selected as features. This process reduces unnecessary data and improves the machine learning model's efficiency and performance. The extracted features are then provided as input to the Random Forest

regression model. The model analyses the relationship between environmental parameters and pollution levels. Based on these features, the system predicts the Air Quality Index (AQI) and determines the corresponding pollution category (Figure 3).

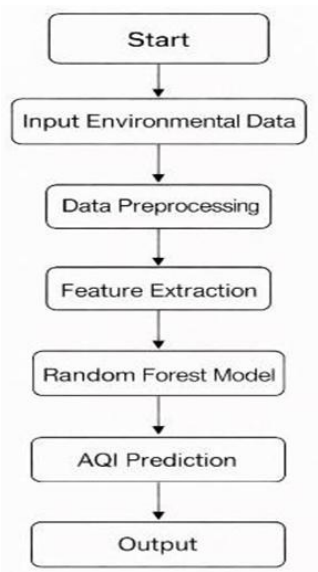


Figure 3: Workflow of the proposed air quality prediction system

The system receives environmental sensor data, including PM2.5, PM10, CO, NO₂, SO₂, O₃, temperature, humidity, and wind speed. The system first preprocesses the collected data by cleaning and normalizing the values. After preprocessing, the system extracts key environmental parameters and provides them to the Random Forest machine learning model. The trained model analyzes these features and predicts the Air Quality Index (AQI) value. Finally, the system displays the predicted AQI and the corresponding pollution category. This interface allows users to input environmental sensor values such as PM2.5, PM10, CO, NO₂, SO₂, O₃, temperature, humidity, and wind speed (Figure 4).

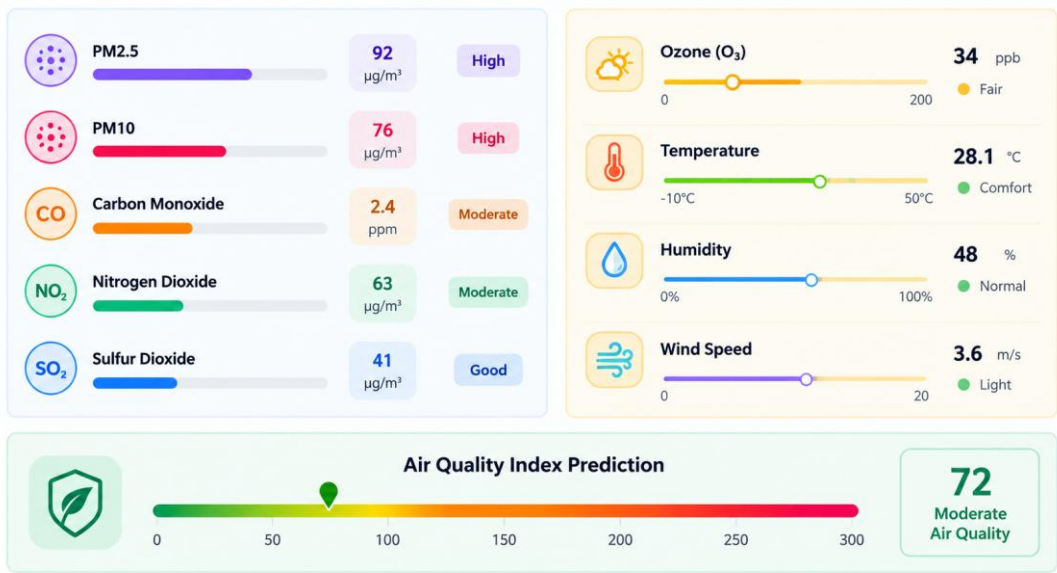


Figure 4: Gradio interface for AQI prediction

Users adjust the values using sliders to represent current environmental conditions. After entering the values, the user can click the “Predict AQI” button to generate the Air Quality Index prediction. The Air Quality Monitoring Dashboard is a modern, intuitive user interface that clearly and accurately presents essential environmental data. It shows key pollutants such as PM2.5, PM10, CO, NO₂, and SO₂ with measured values and status levels. Colourful progress bars and indicators let the user immediately understand whether pollution levels are high, moderate or good. The display also shows environmental parameters

such as ozone, temperature, humidity and wind speed. The bottom scale, called the Air Quality Index (AQI) prediction, provides a rough estimate of air quality. The soft colours, rounded cards, basic symbols and clear writing create a clean, visually pleasing experience.

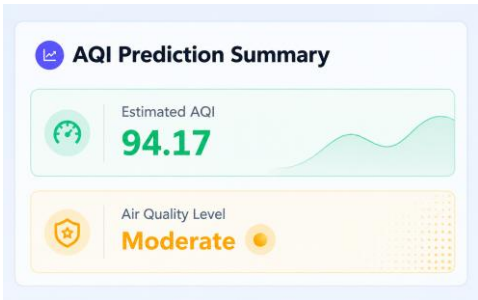


Figure 5: AQI prediction output interface

Finally, the system displays the predicted air quality index (AQI) value based on the input environmental parameters. It also shows the corresponding AQI category, helping users easily understand the current air quality level. Figure 5 shows the sleek, modern AQI Prediction Summary interface that clearly visualizes the air quality analysis findings. The main section displays the estimated AQI value of 94.17 in large, prominent typography. The expected outcome is shown with a visual gauge and subtle graphical elements. The second portion presents the Air Quality Level as Moderate, making the overall condition easy to understand. Soft green and yellow tones visually differentiate the information sections. A professional-looking and user-friendly dashboard with rounded cards, simple icons and minimal design features.

4. Results and Discussions

The Air Quality Prediction system used a Random Forest model and was evaluated using environmental data including PM2.5, PM10, CO, NO₂, SO₂, O₃, temperature, humidity, and wind speed.

Table 1: Comparison of environmental parameters and predicted AQI values

Input Environmental Parameters									AQI Prediction	
PM2.5	PM10	CO	NO ₂	SO ₂	O ₃	Temp	Humidity	Wind Speed	Predicted AQI	AQI Category
40	60	1.2	30	8	20	26	55	2.0	55.32	Moderate
93	79	2.6	75	46	35	27	60	2.4	94.17	Moderate
150	120	3.5	110	60	50	30	65	3.1	162.45	Unhealthy
250	200	5.2	180	90	80	32	70	4.0	275.63	Very Unhealthy

These parameters were used as features to analyze air pollution levels and predict the Air Quality Index (AQI). The collected data were preprocessed and used to train and test the machine learning model.

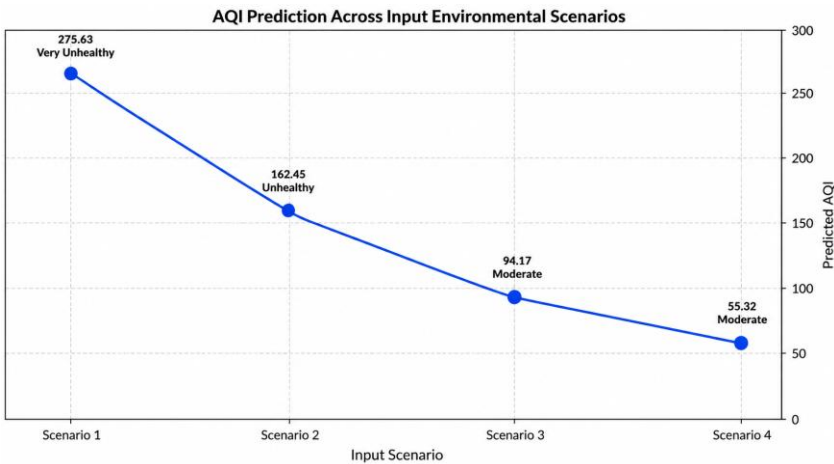


Figure 6: AQI prediction

The system successfully predicts the AQI value and classifies the pollution level, helping users easily understand the air quality condition. Table 1 illustrates how different environmental parameter values influence the predicted Air Quality Index (AQI). It clearly compares the input pollution levels with the corresponding AQI results generated by the model. This representation helps explain how variations in pollutant concentrations affect the overall air quality classification. Figure 6 shows changes in the projected Air Quality Index (AQI) values for four different environmental input scenarios. The AQI climbs from 55.32 in Scenario 1 to 94.17 in Scenario 2. Scenario 3 shows a considerable jump, with the anticipated AQI at 162.45, indicating unhealthy air quality (Figure 7).

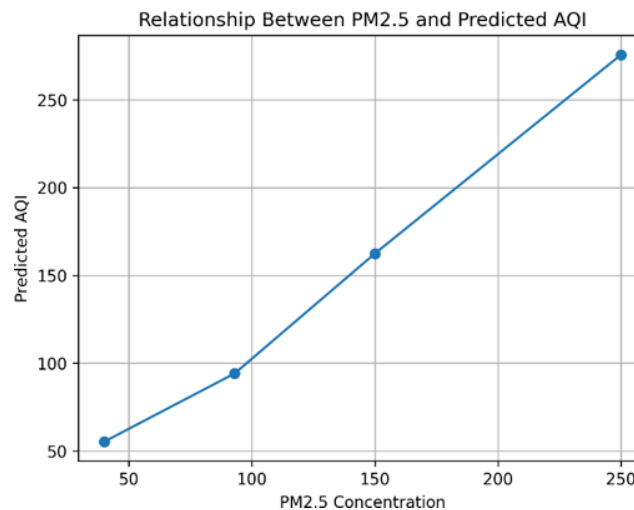


Figure 7: Relationship between PM2.5 concentration and predicted AQI

In Scenario 4, the AQI increases further to 275.63, indicating highly unhealthy air quality. The line graph clearly shows how changing environmental parameters affect the predicted AQI. The visualization shows how air quality deteriorates progressively across the scenarios. The graph illustrates the relationship between PM2.5 concentration and the predicted Air Quality Index (AQI). As PM2.5 levels increase, the AQI also rises significantly. Higher particulate matter concentrations result in poorer air quality. This analysis demonstrates PM2.5's impact on air pollution levels.

5. Conclusion

The proposed IoT-based air quality monitoring system provides a smart, efficient approach to monitoring and predicting atmospheric pollution. The system collects key environmental parameters such as PM2.5, PM10, carbon monoxide (CO), nitrogen dioxide (NO₂), sulphur dioxide (SO₂), ozone (O₃), temperature, humidity and wind speed. These parameters provide information on pollutant concentrations and prevailing environmental conditions. The collected data are preprocessed using appropriate data preparation techniques to remove inconsistencies and prepare the measurements for machine learning analysis. Then, a trained machine learning model is applied to investigate the connection between the environmental parameters and the Air Quality Index (AQI). The system generates an estimated AQI value based on the inputs, representing the overall level of air pollution. The system also provides an interactive, intuitive user interface that lets users enter environmental measurements and get the predicted AQI without advanced technical knowledge. The output provides the predicted AQI value and its corresponding air quality category, so the user can understand whether the surrounding air is good, moderate, unhealthy, very unhealthy or hazardous. The experimental results show that the proposed approach can efficiently estimate pollution levels and classify different air quality conditions. The system combines IoT-based sensing, machine learning and an interactive interface to enable continuous environmental monitoring and provide useful information for decision-making. It can also raise public awareness by helping people to understand changes in local air quality and to take appropriate preventive action. In conclusion, the presented system demonstrates the potential of intelligent IoT and ML technologies for developing practical, scalable, and accessible air quality monitoring solutions.

Acknowledgement: The authors sincerely acknowledge the support and contributions of the National Autonomous University of Mexico, Technical University of Cotopaxi, Kasdi Merbah University, Tanta University, University of Lahej and Sri Venkatesa Perumal College of Engineering and Technology.

Data Availability Statement: Data supporting this study are available from the corresponding author upon reasonable request.

Funding Statement: No external funding or financial support was received for this research.

Conflicts of Interest Statement: The authors declare no conflicts of interest, and all sources have been appropriately acknowledged.

Ethics and Consent Statement: The study complied with applicable ethical standards, with informed consent and confidentiality maintained throughout.

References

1. P. Kalia and M. A. Ansari, "IoT based air quality and particulate matter concentration monitoring system," *Materials Today: Proceedings*, vol. 32, no. 3, pp. 468–475, 2020.
2. R. K. Mishra, R. Rana, S. Tomar, S. Sidhant, and M. Sharma, "Air Quality Prediction Using Machine Learning Techniques," in *Blue Sky, Blue Water*, Springer, Cham, Switzerland, 2025.
3. S. M. S. D. Malleswari and T. K. Mohana, "Air pollution monitoring system using IoT devices: Review," *Materials Today: Proceedings*, vol. 51, no. 6, pp. 1147–1150, 2022.
4. M. Mihirani, L. Yasakethu, and S. Balasooriya, "Machine learning based air pollution prediction model," *IEEE IAS Global Conference on Emerging Technologies (GlobConET)*, London, United Kingdom, 2023.
5. S. M. Saad, A. R. M. Saad, A. M. Y. Kamarudin, A. Zakaria, and A. Y. M. Shakaff, "Indoor air quality monitoring system using wireless sensor network (WSN) with web interface," in *proc. International Conference on Electrical, Electronics and System Engineering (ICEESE)*, Kuala Lumpur, Malaysia, 2013.
6. M. N. A. Ramadan, M. A. H. Ali, S. Y. Khoo, M. Alkhedher, and M. Alherbawi, "Real-time IoT-powered AI system for monitoring and forecasting of air pollution in industrial environment," *Ecotoxicology and Environmental Safety*, vol. 283, no. 9, pp. 1–15, 2024.
7. R. Muljana, L. D. Ayuningtyas, R. P. Daksa, S. F. Djahhari, M. A. Fiezayyan, N. Noviyanti, and T. M. Sagala, "Air Pollution Prediction using Random Forest Classifier: A Case Study of DKI Jakarta," in *Proc. International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, Jakarta, Indonesia, 2023.
8. J. Ma, J. C. P. Cheng, C. Lin, Y. Tan, and J. Zhang, "Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques," *Atmospheric Environment*, vol. 214, no. 10, p. 116885, 2019.
9. R. Kaushik, E. Kelkar, A. Gupta, and A. Bhargava, "Development of an air pollution monitoring system," *Procedia Computer Science*, vol. 260, no. 5, pp. 645–654, 2025.
10. M. A. Mujtaba, M. A. Munir, S. Ali, J. Petrů, T. Ansar, W. Akhlaq, M. Ahmad, H. Iqbal, F. Ali, M. N. Bashir, and T. Alexander, "Using machine learning for air quality prediction and sustainable urban planning," *Sustainable Futures*, vol. 10, no. 12, pp. 1–13, 2025.
11. A. S. Moursi, N. El-Fishawy, S. Djahel, and M. A. Shouman, "An IoT enabled system for enhanced air quality monitoring and prediction on the edge," *Complex & Intelligent Systems*, vol. 7, no. 6, pp. 2923–2947, 2021.
12. A. Almalawi, F. Alsolami, A. I. Khan, A. Alkhatlan, A. Fahad, K. Irshad, S. Qaiyum, and A. S. Alfakeeh, "An IoT based system for magnify air pollution monitoring and prognosis using hybrid artificial intelligence technique," *Environmental Research*, vol. 206, no. 4, p. 112576, 2022.
13. D. Zhang and S. S. Woo, "Real time localized air quality monitoring and prediction through mobile and fixed IoT sensing network," *IEEE Access*, vol. 8, no. 5, pp. 89584–89594, 2020.
14. J. A. Martín-Baos, L. Rodríguez-Benítez, R. García-Ródenas, and J. Liu, "IoT based monitoring of air quality and traffic using regression analysis," *Applied Soft Computing*, vol. 115, no. 1, p. 108282, 2022.
15. V. Behal and R. Singh, "Personalised healthcare model for monitoring and prediction of air pollution: Machine learning approach," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 33, no. 3, pp. 425–449, 2021.

Publisher's Note: The publisher remains impartial concerning jurisdictional claims in published maps and institutional affiliations. Responsibility for the content rests entirely with the authors and does not necessarily reflect the publisher's perspectives.