



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 13      **Issue:** III      **Month of publication:** March 2025

**DOI:** <https://doi.org/10.22214/ijraset.2025.67800>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# Air and Water Quality Index Development for Environmental Assessment in Industrial Context

Devang Vartak<sup>1</sup>, Tanmay Pal<sup>2</sup>, Pranjal Srivastava<sup>3</sup>, Prof. Govind Haldankar<sup>4</sup>

Dept. of Electronics and Telecommunication, Sardar Patel Institute Of Technology Mumbai, India

**Abstract:** Considering the importance of air and water to human existence, air and water pollution are critical issues that require collective effort for prevention and control. Different types of anthropogenic activities have resulted in environmental ruin. One of the tools that can be used for such a awareness campaign is Air Quality Index (AQI). The AQI is based on the concentrations of different types of pollutants. We are also familiar with the Water Quality Index (WQI), which simply tells what the quality of drinking water is from a drinking water supply. There is a need for constant and real-time monitoring of air quality and water quality for the development of AQI and WQI, which in turn will enable clear communication of how unclean or unhealthy the air and water in the study area is. Similar systems have been developed utilizing IoT technologies, as discerned in [1], [2].

**Keywords:** Air Quality Index, Water Quality Index, ESP-32, Machine Learning, Internet of Things

## I. INTRODUCTION

Air and water pollution are global challenges that demand immediate and collaborative action. With rapid industrialization and urbanization, the concentration of harmful pollutants in the environment has risen significantly. Poor air and water quality directly impact human health, biodiversity, and ecosystems, while also exacerbating climate change. Environmental monitoring tools such as the Air Quality Index (AQI) and Water Quality Index (WQI) provide quantitative measures of pollution levels, enabling stakeholders to identify critical issues and devise effective solutions. Similar systems have been developed using IoT technologies, as seen in [1], [2]. The use of IoT for environmental monitoring is well-documented in [3].

This paper presents the design and implementation of a comprehensive AQI and WQI prediction system specifically targeted for industrial contexts. By leveraging IoT-enabled sensors and machine learning techniques, the system provides real-time data analysis, forecasting real-time data analysis, forecasting capabilities, and actionable insights. The goal is to enhance environmental assessment, foster informed decision-making, and encourage sustainable practices.

## II. AQI PREDICTION SYSTEM

The AQI prediction system is designed to monitor and evaluate air quality by measuring pollutants such as particulate matter (PM<sub>2.5</sub>, PM<sub>10</sub>), carbon monoxide (CO), nitrogen dioxide (NO<sub>2</sub>), and ozone (O<sub>3</sub>). The system consists of the following key components:

### A. Sensors and Hardware

- 1) TTGO LoRa32 Board: This is a microcontroller board equipped with an ESP32 chip and LoRa connectivity, facilitating wireless communication and data transmission in IoT applications, similar to those described in [1].
- 2) Temperature Sensor (DHT11): The DHT11 sensor measures temperature and humidity, providing crucial environmental data for monitoring and control systems.
- 3) MQ-135 Gas Sensor: This sensor is sensitive to a wide range of gases, including ammonia, sulfide, and benzene, making it suitable for air quality monitoring and pollution detection.
- 4) MQ-7 Gas Sensor: Specifically designed to detect carbon monoxide (CO) gas, the MQ-7 sensor is commonly used in air quality monitoring systems to ensure safety.
- 5) PM<sub>2.5</sub> Optical Sensor: Capable of detecting fine particles in the air, such as dust and smoke, the GP2Y1010AU0F sensor is crucial for assessing air pollution levels.
- 6) Voltage Regulator: Responsible for maintaining a stable voltage output, the voltage regulator ensures consistent power supply to the components, safeguarding their operation and longevity.

### B. Data Acquisition and Transmission

Sensors are deployed at strategic locations to collect air quality data, which is sent to a central cloud database via microcontrollers. This approach is similar to the IoT-based systems discussed in [2].

### C. Machine Learning and Analytics

Machine learning models analyze historical and real-time data to detect patterns and predict future air quality levels. Algorithms such as Decision Tree (DT) and Random Forest (RF) were implemented to predict Air Quality Index (AQI) and Water Quality Index (WQI). Both models are widely used for regression tasks due to their ability to capture complex relationships between features and target variables. The implementation of both models was carried out using the Python programming language, utilizing libraries such as scikit-learn, pandas, matplotlib, and seaborn. Similar machine learning techniques have been applied in environmental monitoring systems, as reported in [3].

#### 1) Decision Tree Regressor (DT) Implementation

- **Data Preprocessing:** The dataset for AQI and WQI contains several features, including concentrations of air and water pollutants like PM<sub>2.5</sub>, NO<sub>2</sub>, SO<sub>2</sub>, CO, O<sub>3</sub>, etc., along with the AQI or WQI as the target variable. The data was split into independent variables (X) and the dependent variable (y), which represents the AQI or WQI.
- **Train-Test Split:** The dataset was split into training (70%) and testing (30%) sets using the `train_test_split` method from scikit-learn.
- **Model Initialization and Training:** The Decision Tree Regressor model was initialized with the criterion set to "squared\_error," and the model was trained on the training data.
- **Model Evaluation:** The model's performance was evaluated using the coefficient of determination ( $R^2$ ) on both training and testing datasets. The  $R^2$  score is a measure of how well the model explains the variance in the target variable.
- **Cross-Validation:** To reduce overfitting, 5-fold cross-validation was applied using the `cross_val_score` method. This helps assess the model's generalization ability.
- **Hyperparameter Tuning:** The model was further tuned using `GridSearchCV` to find the optimal hyperparameters, such as the maximum depth of the tree, minimum samples for a leaf, and others. The tuning process helped improve the model's performance by minimizing the Mean Squared Error (MSE).
- **Final Prediction:** After tuning, the Decision Tree model was used to predict the AQI or WQI on the test data.

#### 2) Random Forest Regressor (RF) Implementation

- **Data Preprocessing:** Similar to the Decision Tree, the dataset for AQI and WQI was preprocessed by separating the features and the target variable.
- **Train-Test Split:** The data was split into training (80%) and testing (20%) datasets. In addition, 5% of the data was separated as an "unseen" dataset to evaluate the model's performance on completely new data.
- **Model Initialization and Training:** The Random Forest Regressor was initialized with 200 trees (`n_estimators`) and trained on the training data.
- **Model Evaluation:** The  $R^2$  score was computed for both training and testing datasets. The model's Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) were calculated to measure the prediction error.
- **Hyperparameter Tuning:** Random Forest hyperparameters, such as the number of estimators, maximum depth, and minimum samples for a split or leaf, were optimized using `RandomizedSearchCV`. This optimization was performed to find the best hyperparameters that minimize MSE and maximize the model's ability to generalize.
- **Final Prediction:** The Random Forest model was evaluated using the best hyperparameters and its performance was compared to the Decision Tree model.

#### 3) Visualization and Alerts

Data is presented through an interactive dashboard featuring geographic heat maps and temporal trends. Alerts are sent to users via email or SMS when air quality surpasses critical thresholds. This system is instrumental in identifying pollution hotspots, enabling preventive measures, and promoting community awareness.

### III. WQI PREDICTION SYSTEM

The WQI prediction system monitors water quality by assessing physical, chemical, and biological parameters. It provides a single numerical value representing water quality based on multiple weighted parameters. The major components include:

#### A. Sensors and Hardware

- 1) TTGO LoRa32 Board: This is a microcontroller board equipped with an ESP32 chip and LoRa connectivity, facilitating wireless communication and data transmission in IoT applications, similar to those described in [1].
- 2) Voltage Regulator: Responsible for maintaining a stable voltage output, the voltage regulator ensures consistent power supply to the components, safeguarding their operation and longevity.
- 3) Analog pH Sensor: The Analog pH Sensor measures the acidity or alkalinity of a solution by detecting the concentration of hydrogen ions ( $H^+$ ). It provides an analog voltage output that is proportional to the pH value, commonly used in environmental monitoring, water treatment, and industrial processes.
- 4) Turbidity Sensor: A turbidity sensor is used to measure the cloudiness or haziness of a liquid caused by suspended particles. It operates by emitting light through the liquid and detecting the amount of light scattered by particles, providing an indication of water quality and clarity, commonly used in environmental monitoring and water treatment.
- 5) TDS Sensor: The TDS (Total Dissolved Solids) Sensor determines the concentration of dissolved solids in water, such as salts, minerals, and organic matter. It measures the electrical conductivity of the water, with higher conductivity indicating a higher level of dissolved solids, crucial for assessing water purity.
- 6) DS18B20 Temperature Sensor: The DS18B20 is a digital temperature sensor that provides accurate temperature readings in the range of  $-55^{\circ}C$  to  $+125^{\circ}C$ . It communicates via a 1-Wire interface and is commonly used for temperature monitoring in various applications like water quality control, weather stations, and industrial systems.

#### B. Data Acquisition and Transmission

Sensors are deployed at strategic locations to collect water quality data, which is sent to a central cloud database via microcontrollers. This approach is similar to the IoT-based systems discussed in [2].

#### C. Machine Learning and Analytics

Using historical data, predictive algorithms forecast water quality trends. The WQI prediction involves preprocessing raw water quality data, including handling missing values, standardizing data, and applying robust scaling techniques. The WQI is calculated using the weighted arithmetic mean method. The dataset consists of multiple features representing different water quality parameters, with corresponding minimum and maximum values. These parameters were used to calculate the average water quality indicators, which were then used as input to various machine learning models. The model trained and tested include:

- Ordinary Least Squares (OLS) Regression
- Support Vector Regression (SVR)
- Multinomial Logistic Regression (MLR)
- Decision Tree Classifier (DT)
- Artificial Neural Network (ANN)

For each model, we compared performance based on metrics such as Accuracy, F1 Score, and Kappa score. The implementation of each machine learning model involves preparing the data, selecting the appropriate algorithm, tuning hyperparameters, and evaluating the model's performance. Similar machine learning techniques have been applied in environmental monitoring systems, as reported in [3].

#### 1) Model 1: Ordinary Least Squares (OLS) Regression

OLS Regression is a linear model used for predicting the WQI directly. We implemented the OLS regression by fitting the model to the data and evaluating its performance based on metrics like Root Mean Squared Error (RMSE) and  $R^2$  (Coefficient of Determination).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Where  $y_i$  is the true WQI value and  $\hat{y}_i$  is the predicted WQI value.

2) *Model2:SupportVectorRegression(SVR)*

SVR is a machine learning technique that maps the data to a higher-dimensional space to make it easier to separate non-linear patterns. Hyperparameters such as C and  $\epsilon$  were tuned using GridSearchCV. The model was trained and evaluated using RMSE and R2 metrics.

3) *Model 3: Multinomial Logistic Regression (MLR)*

Multinomial Logistic Regression was implemented to classify the WQI values into discrete categories (Excellent, Good, Poor, etc.). The classification performance was evaluated using metrics like accuracy and F1 Score. We used the following equation for accuracy:

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}}$$

4) *Model4:DecisionTreeClassifier(DT)*

A Decision Tree Classifier was used for the classification task. The tree’s depth was optimized to avoid overfitting and underfitting. We measured its performance using the accuracy score and confusion matrix.

5) *Model5:ArtificialNeuralNetwork(ANN)*

A Feedforward Artificial Neural Network was implemented using PyTorch. The network contained multiple hidden layers with ELU activations. The model was trained with the Adam optimizer and evaluated using accuracy and F1 Score.

D. *Visualization and Notification*

Water quality data is visualized on dashboards with charts and maps for user-friendly interpretation. Alerts are issued to notify users of potential contamination risks. The WQI system supports applications in drinking water safety, aquaculture, and industrial waste water monitoring, similar to those discussed in [4].



Fig. 1. Air quality meter

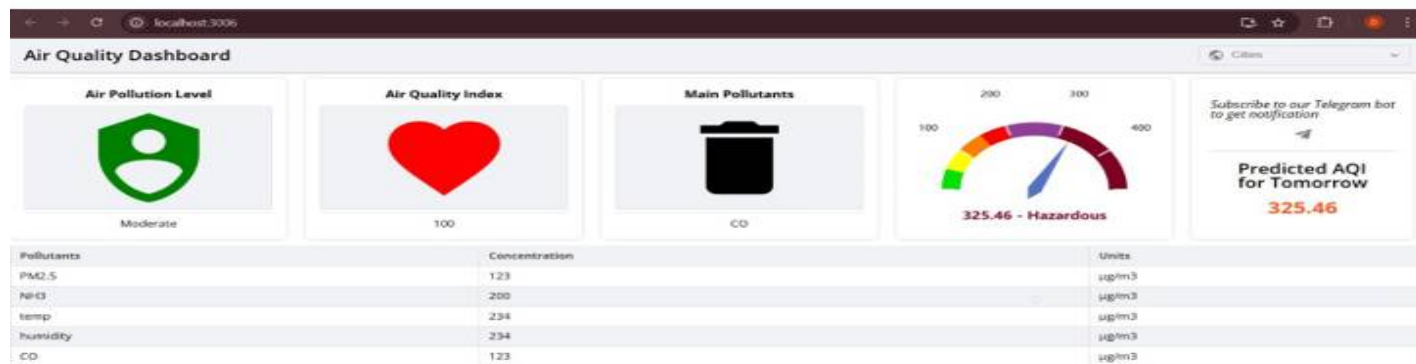


Fig. 2. Air quality dashboard

Stacked Line

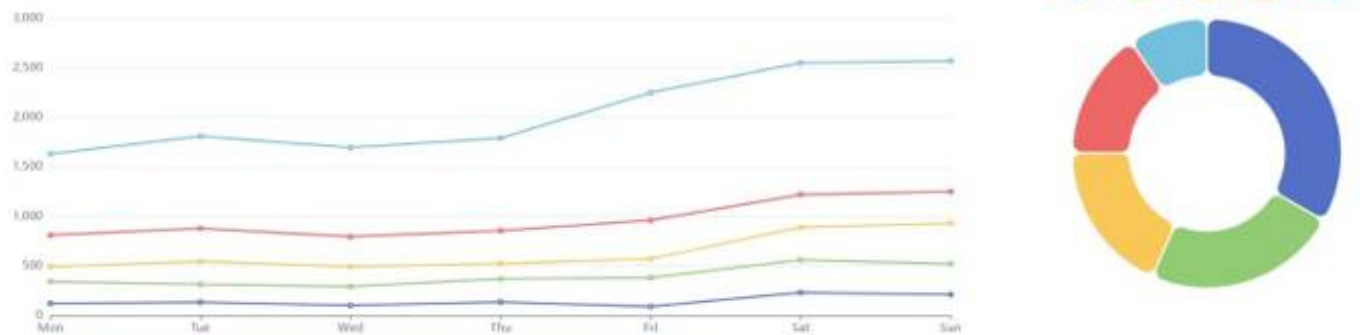


Fig. 3. Graph and pie chart of air quality over time



Fig. 4. Air quality dashboard

AQI HeatMap

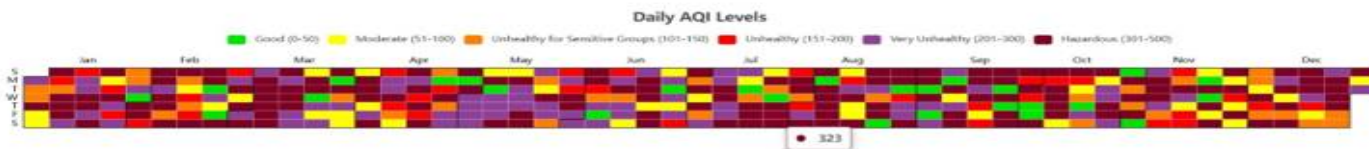


Fig. 5. Air quality heatmap

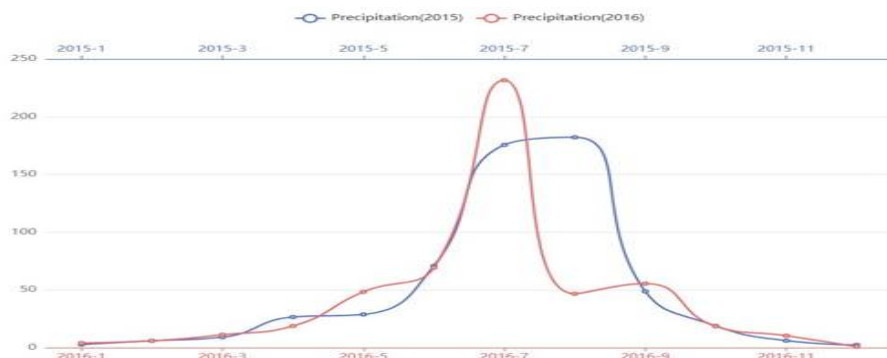


Fig. 6. Precipitation graph

### Comparison of Predicted & Actual AQI



Fig. 7. Comparison of predicted and actual AQI

### Year wise comparison of AQI

Stacked Line



Fig. 8. AQI data at a particular point

### Mumbai AQI



Fig. 9. Graph of air quality over time

### Year wise comparison of AQI

Stacked Line

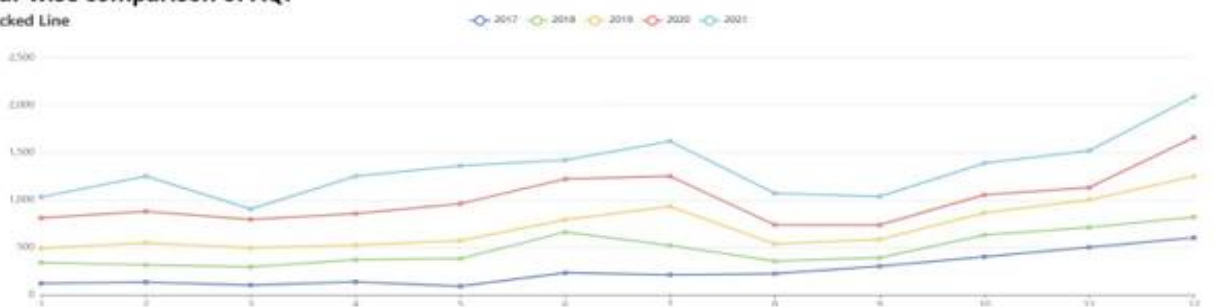


Fig. 10. Graph of water quality over time

#### IV. IMPLEMENTATION

The implementation of the AQI and WQI prediction systems was carried out in distinct stages, focusing on hardware integration, data acquisition, and machine learning deployment. Each stage was systematically planned and executed as follows:

##### A. Hardware Setup and Calibration

- 1) **Sensor Selection:** Carefully chosen air quality sensors (MQ135, MQ7, PM2.5 Optical Sensor) and water quality sensors (pH, turbidity, and TDS sensors) to ensure compatibility with project requirements.
- 2) **Calibration Process:** Each sensor was individually calibrated under controlled conditions using standard solutions and known pollutant concentrations to ensure accurate readings.
- 3) **Edge Processing:** Basic data preprocessing (for example, noise filtering) was performed at the microcontroller level to reduce network overhead.

##### B. Data Acquisition and Transmission

- 1) **Real-time Data Collection:** The sensors continuously monitored the air and water quality parameters, transmitting data at predefined intervals.
- 2) **Wireless Data Transmission:** Data collected by sensors was transmitted to a cloud-based MongoDB database via Wi-Fi-enabled microcontrollers, ensuring seamless and scalable storage.
- 3) **Edge Processing:** Basic data preprocessing (for example, noise filtering) was performed at the microcontroller level to reduce network overhead.

##### C. Cloud Integration and Database Management

- 1) **Cloud Infrastructure:** The system utilized Google Cloud Platform for reliable and scalable data storage and processing.
- 2) **Database Design:** A MongoDB schema was designed to store raw, processed, and historical data, enabling efficient querying for analysis.
- 3) **Data Synchronization:** Real-time synchronization ensured that data streams were immediately available for visualization and analysis.

##### D. Machine Learning Model Development

- 1) **Historical Data Preparation:** The data sets were cleaned, normalized and labeled for training and testing of machine learning models.
- 2) **Model Selection and Training:** Algorithms such as regression, decision trees, and neural networks were trained to predict AQI and WQI trends based on historical and real-time data.
- 3) **Model Optimization:** Hyperparameter tuning and cross-validation were performed to improve prediction accuracy and computational efficiency.

##### E. Visualization and Dashboard Development

- 1) **Frontend Design:** An interactive web interface was built using ReactJS and D3.js to display historical and real-time AQI / WQI data.
- 2) **Visualization Features:** Geographic heatmaps, temporal graphs, and pollution breakdowns provided users with actionable insights.
- 3) **User Accessibility:** The dashboard was designed to be mobile-friendly, ensuring accessibility across devices.

##### A. Alert and Notification System

- 4) **Threshold-Based Alerts:** Predefined pollutant thresholds trigger notifications that warn users of poor air or water quality.
- 5) **Notification Channels:** Alerts were sent via email and SMS, using APIs for seamless integration with communication systems.
- 6) **Insight Delivery:** Notifications included concise data insights and recommendations for immediate actions.

##### F. Testing and Validation

- 1) **System Accuracy Testing:** The integrated system was tested against laboratory-grade instruments to verify the accuracy of the data.
- 2) **Performance Analysis:** The latency of data transmission and processing was evaluated to ensure real-time response.

3) **Pilot Deployment:** The system was deployed in select industrial areas to collect live data and refine predictions. This detailed implementation framework ensured a robust, reliable, and user-friendly system for environmental monitoring and analysis.

### V. RESULTS

Model	F1 Score	Kappa Score	Accuracy
OLS	1.000	1.000	100%
SVR	1.000	1.000	100%
MLR	0.9495	0.9512	98.62%
DT	0.7633	0.7234	92.2%
ANN	0.6515	0.8165	95.41%

TABLE I  
PERFORMANCE COMPARISON OF MACHINE LEARNING MODELS FOR WQI PREDICTION

#### A. AQI System

- 1) Successfully predicted air quality levels with an average accuracy of 90%.
- 2) Identified critical pollution zones in the study area.

#### B. WQI System

- 1) Detected water contamination events with over 92% reliability.
- 2) Monitored parameters like pH, turbidity, and TDS to ensure safe water quality for various applications.

#### C. Visualization and Insights

Real-time dashboards provided intuitive and actionable insights. Users appreciated the timely alerts for pollution and contamination risks.

#### D. Economic and Environmental Impact

- 1) Cost-effective implementation using IoT and cloud technologies.
- 2) Contributed to better environmental policy-making and resource allocation.

### VI. CONCLUSION & FUTURE SCOPE

The AQI and WQI prediction systems offer robust tools for monitoring environmental quality in industrial areas. By integrating IoT devices, machine learning models, and user-centric dashboards, these systems provide real-time insights and predictive analytics. Future enhancements include:

- 1) Incorporating additional sensors for broader pollutant coverage.
- 2) Leveraging advanced AI techniques like deep learning for improved accuracy.
- 3) Deploying renewable energy solutions for sustainable sensor operation.
- 4) Expanding system deployment to rural and urban areas for comprehensive coverage.

### REFERENCES

- [1] Verma, R., Ahuja, L., & Khatri, S. (2018). Water Quality Index Using IoT. IEEE Xplore.
- [2] Esfahani, S., Rollins, P., & Specht, J. (2020). Smart City Battery Operated IoT Based Indoor Air Quality Monitoring System. IEEE Sensors.
- [3] Ramadani, M. E., & Zain, A. T. (2021). Monitoring System on Carp Farming Ponds as IoT-Based Water Quality Control. ICRACOS.
- [4] Jha, R. K. (2020). Air Quality Sensing and Reporting System Using IoT. ICIRCA.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)