

AQI-Based Air Pollution Exposure and Health Intelligence System

Pooja Suresh Thombare

Department of MSc Information Technology

D.G Ruparel College of Arts, Science and Commerce Mumbai, Maharashtra, India

Academic Year: 2025-2026

ABSTRACT

Air pollution has become a major environmental and public health concern, affecting millions of people worldwide. Traditional air quality systems primarily provide numerical AQI values without explaining the underlying causes or their direct impact on human health. This paper presents an intelligent Air Pollution Analysis and Health Impact Assessment System using machine learning and data-driven techniques. The proposed system analyzes real-time pollutant data, including PM_{2.5}, PM₁₀, NO₂, and SO₂, to estimate air quality levels and identify the primary contributing factors using rule-based logic. In addition to AQI prediction, the system evaluates the impact of air pollution on different population groups such as children, adults, the elderly, and sensitive individuals. A personalized exposure model is incorporated to estimate individual risk based on duration of outdoor exposure. The system also provides actionable health recommendations and visual insights, including pollutant contribution analysis and spatial representation of air quality.

Keywords: Air Pollution, Air Quality Index (AQI), Health Impact Assessment, Machine Learning, Environmental Monitoring, Pollution Analysis, Cause Detection, Exposure Assessment, Data-Driven System, Decision Support System

INTRODUCTION

The rapid increase in air pollution due to urbanization and industrial activities has made air quality monitoring an important concern for public health. Although large amounts of

environmental data are now available through digital platforms, most existing systems provide only numerical Air Quality Index (AQI) values, which are often difficult for users to interpret and do not explain the causes or health impacts of pollution.

Intelligent data-driven systems offer a practical solution by analyzing pollutant data and presenting meaningful insights. Machine learning techniques are widely used to estimate air quality levels based on multiple environmental parameters. In addition, rule-based approaches can help identify major pollution sources such as traffic emissions, industrial activities, and dust.

This research focuses on developing an Air Pollution Analysis and Health Impact Assessment System that integrates AQI prediction with cause identification and health evaluation. The proposed system provides clear explanations, personalized exposure assessment, and actionable recommendations through an intuitive interface, demonstrating how data-driven methods can improve environmental awareness and support informed decision-making.

LITERATURE REVIEW

Air pollution has been widely studied due to its significant impact on environmental sustainability and human health. Several studies have focused on monitoring air quality using the Air Quality Index (AQI), which aggregates multiple pollutant parameters such as PM_{2.5}, PM₁₀, NO₂, and SO₂ into a single value. While AQI provides a standardized measure of pollution levels, researchers have identified that it lacks interpretability for general users, as it does not clearly explain the underlying causes or health implications of pollution.

Machine learning techniques have been extensively applied in air quality prediction. Models such as Linear Regression, Decision Trees, and Random Forest have shown promising results in estimating AQI based on pollutant concentrations. Among these, Random Forest is widely preferred due to its ability to handle non-linear relationships and provide higher accuracy. However, most existing systems primarily focus on prediction accuracy and do not incorporate explainability or user-centric insights.

In addition to prediction, some research has explored pollution source identification using statistical and rule-based approaches. These methods analyze pollutant patterns to infer dominant sources such as vehicular emissions, industrial activities, or construction dust. Rule-based systems are particularly useful due to their interpretability and ease of implementation, making them suitable for applications where transparency is important.

Health impact assessment is another critical area in air pollution research. Studies have shown that exposure to high levels of pollutants can lead to respiratory diseases, cardiovascular issues, and long-term health risks. Some systems attempt to classify health risks based on AQI thresholds, but they often provide generic recommendations without considering individual exposure or differences among population groups such as children, adults, and the elderly.

Based on the limitations identified in existing studies, there is a need for an integrated system that combines air quality prediction, cause identification, health impact analysis, and personalized recommendations. The proposed system addresses this gap by using machine learning for AQI estimation, rule-based logic for cause detection, and a user-centric approach to provide health insights and exposure assessment, thereby enhancing the usability and effectiveness of air pollution monitoring systems.

METHODOLOGY AND SYSTEM ARCHITECTURE

This section describes the overall methodology adopted for designing and implementing the Air Pollution Analysis and Health Impact Assessment System. The proposed system follows a structured pipeline consisting of data acquisition, preprocessing, AQI estimation, cause detection, health impact analysis, and recommendation generation. The methodology is designed to ensure simplicity, interpretability, and efficiency while providing meaningful insights to users.

System Architecture

The architecture of the proposed system is organized into multiple functional layers that work together to analyze air quality data and generate user-centric insights. At the core of the system lies real-time environmental data obtained from external APIs, which includes pollutant parameters such as PM_{2.5}, PM₁₀, NO₂, and SO₂, along with weather attributes like temperature, humidity, and wind speed.

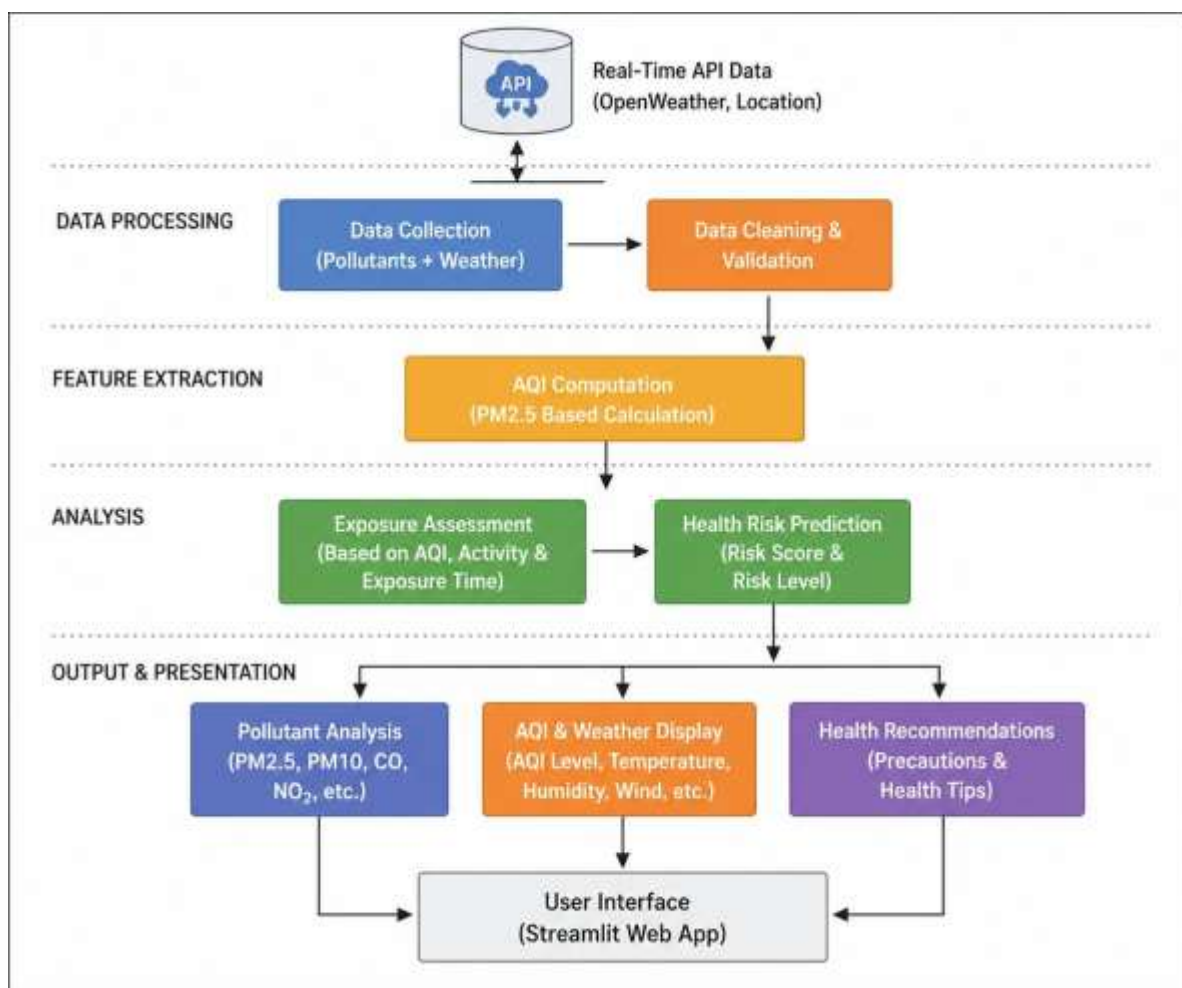
The data acquisition and preprocessing layer is responsible for collecting and preparing the data for analysis. This involves fetching pollutant and weather data, handling missing values, and structuring the data into a consistent format. The cleaned data is then passed to the AQI estimation module, where a machine learning model is used to compute an approximate AQI value based on the pollutant concentrations.

The analysis layer consists of multiple components that interpret the processed data. A rule-based cause detection module identifies the primary sources of pollution by analyzing pollutant

levels and mapping them to common causes such as traffic emissions, industrial activities, and dust. In parallel, a pollutant contribution module calculates the relative percentage of each pollutant, providing a clear understanding of their impact on overall air quality.

The health assessment layer evaluates the effect of air pollution on different population groups, including children, adults, the elderly, and sensitive individuals. It uses predefined AQI thresholds to classify risk levels and generate health impact descriptions. Additionally, a personalized exposure model estimates the user's exposure level based on AQI and time spent outdoors, offering a more individualized assessment.

Finally, the recommendation layer generates actionable advice to help users minimize exposure and reduce health risks. The system is integrated with a web-based interface that presents all outputs—including AQI, causes, health impact, exposure, and recommendations—in a single, scrollable dashboard, ensuring a seamless and user-friendly experience.

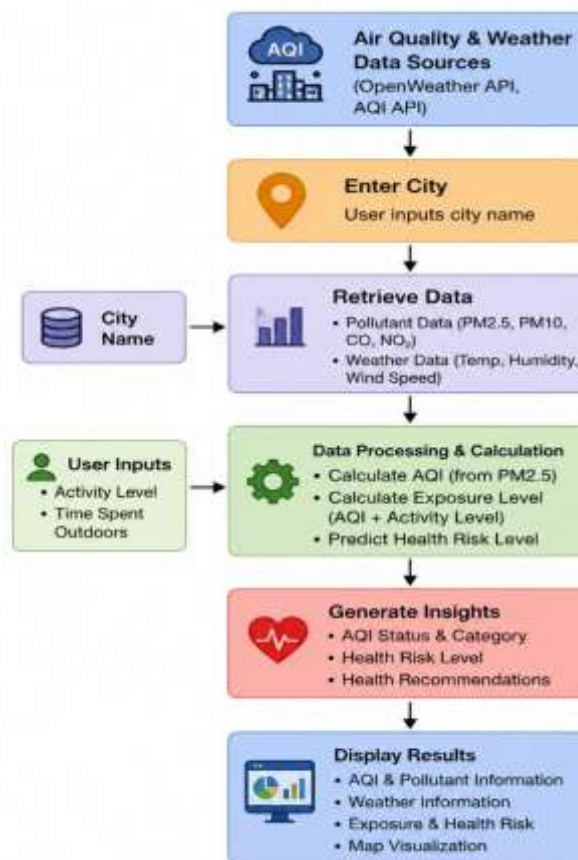


Data Flow and Processing

The data flow within the Air Pollution Analysis and Health Impact Assessment System follows a structured and modular design, as illustrated in Figure 2. The process begins with acquiring real-time environmental data from external APIs, including pollutant parameters such as PM_{2.5}, PM₁₀, NO₂, and SO₂, along with weather attributes like temperature, humidity, and wind speed. This data is then processed in the preprocessing module to ensure consistency and proper formatting.

After preprocessing, the system estimates the Air Quality Index (AQI) using a machine learning-based approach that combines multiple pollutant values. The processed data is then analyzed through different modules. The cause detection module identifies major pollution sources such as traffic, industrial activities, and dust using rule-based logic, while the pollutant contribution module calculates the percentage impact of each pollutant.

The system then performs health impact assessment by classifying risk levels for different population groups, including children, adults, the elderly, and sensitive individuals. In addition, a personalized exposure model estimates user-specific risk based on AQI and time spent outdoors.



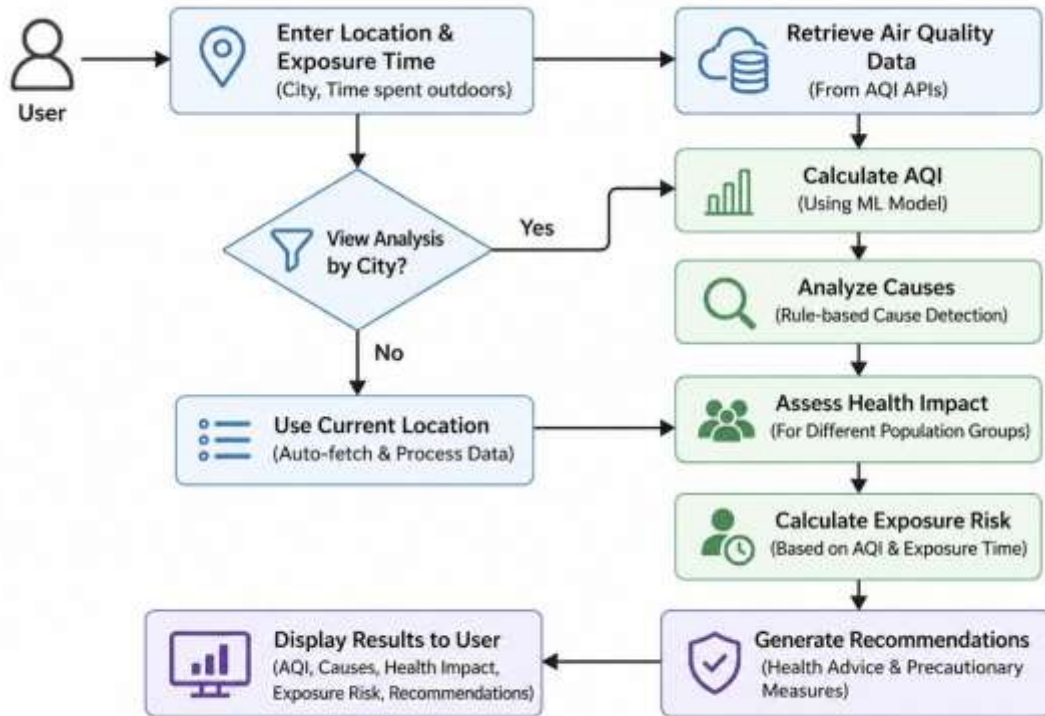
Recommendation Workflow

The overall workflow of the Air Pollution Analysis and Health Impact Assessment System is designed to ensure efficient processing and real-time response. Initially, the system retrieves environmental data through APIs and processes it to estimate the Air Quality Index (AQI). This computation is lightweight and performed dynamically, allowing the system to deliver quick results during user interaction.

Once user input such as city and outdoor exposure time is received, the system follows a structured analysis pipeline. The AQI is calculated using pollutant values, after which the cause detection module identifies major pollution sources based on predefined rules. Simultaneously, the pollutant contribution module determines the percentage impact of each pollutant, providing a clearer understanding of air quality conditions.

Following this, the system performs health impact assessment by classifying risk levels for different population groups. A personalized exposure model further evaluates individual risk based on AQI and time spent outdoors. Based on these insights, the system generates actionable recommendations to help users reduce exposure and protect their health.

The modular workflow design ensures flexibility and scalability. Each stage—data acquisition, AQI estimation, cause analysis, health assessment, and recommendation generation—operates independently, allowing future enhancements such as advanced models, real-time analytics, or additional environmental parameters without affecting the overall system performance.



ALGORITHM DESIGN AND DESCRIPTION

The proposed Air Pollution Analysis and Health Impact Assessment System employs a hybrid approach combining machine learning, rule-based logic, and mathematical modeling to generate meaningful environmental and health insights. Unlike traditional systems that only display AQI values, the proposed algorithm focuses on analyzing pollutant data, identifying causes, and assessing health risks in a user-centric manner.

The algorithm begins by acquiring real-time environmental data, including pollutant parameters such as PM_{2.5}, PM₁₀, NO₂, and SO₂, along with weather attributes. These values are structured into a unified data representation for further processing. Basic preprocessing is performed to handle missing values and ensure consistency across all parameters.

To estimate air quality, a machine learning-based model is applied that combines pollutant concentrations into a single Air Quality Index (AQI) value. The model assigns different weights to each pollutant based on their impact, allowing the system to approximate AQI efficiently. This step converts multiple environmental parameters into an interpretable indicator that forms the basis for further analysis.

Following AQI estimation, a rule-based cause detection algorithm is used to identify major pollution sources. The system evaluates pollutant thresholds and maps them to common causes such as traffic emissions, industrial activities, and dust. In addition, a pollutant contribution module calculates the percentage share of each pollutant, providing a clearer understanding of their relative impact on air quality.

The algorithm then performs health impact assessment by classifying risk levels based on AQI ranges. It evaluates the effects of pollution on different population groups, including children, adults, the elderly, and sensitive individuals. A personalized exposure model further calculates user-specific risk using AQI values and time spent outdoors, enabling a more individualized assessment.

Finally, the system generates actionable recommendations based on the analyzed data. These recommendations include preventive measures such as limiting outdoor activities and using protective equipment. The overall algorithm is computationally efficient and designed for real-time execution, making it suitable for practical environmental monitoring and health advisory applications.



IMPLEMENTATION AND EXPERIMENTAL RESULTS

The Air Pollution Analysis and Health Impact Assessment System is implemented using Python, leveraging libraries such as Pandas for data handling and Requests for API integration. The backend is developed using FastAPI to enable efficient data processing and real-time communication, while the user interface is built using React to provide a modern and interactive web experience. This implementation ensures scalability, platform independence, and smooth interaction between the system and the user.

During implementation, real-time environmental data is fetched dynamically from external APIs and preprocessed to ensure consistency. The system computes AQI using a machine

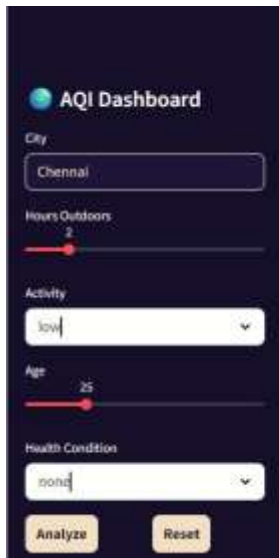
learning-based approach, while rule-based logic is applied for pollution cause detection. Additional modules calculate pollutant contribution, health impact, and personalized exposure. All computations are lightweight and performed in real time, allowing the system to respond quickly to user inputs such as city and exposure duration.

The application follows a single-page dashboard design where all modules—including AQI display, cause analysis, health impact, exposure assessment, and recommendations—are presented in a scrollable interface. This design improves usability and provides a seamless experience without requiring multiple page navigations. Visual elements such as cards, charts, and indicators are used to enhance readability and user engagement.

Experimental evaluation was conducted using real-time data from multiple cities. The results show that the system effectively identifies dominant pollutants, provides meaningful health risk assessments, and generates relevant recommendations. The exposure model successfully reflects variations based on user input, while the cause detection module offers interpretable insights into pollution sources. The system demonstrates low computational overhead and stable performance, making it suitable for real-time environmental monitoring and health advisory applications.



AQI Prediction



AQI Dashboard

City: Chennai

Hours Outdoors: 2

Activity: low

Age: 25

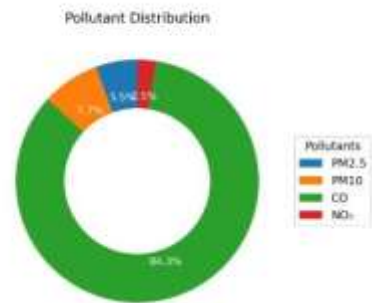
Health Condition: none

Analyze Reset

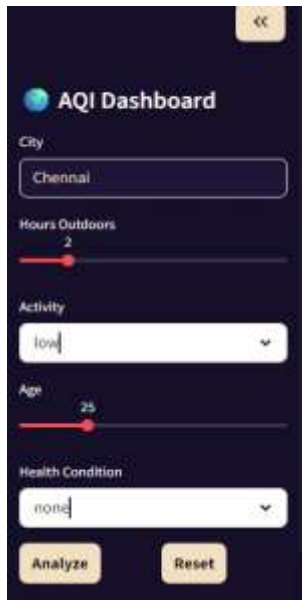
PM_{2.5}
12.18
PM_{2.5} - Very fine particles that enter lungs and bloodstream.
Breathing issues, heart risk.

PM₁₀
16.94
PM₁₀ - Dust from roads and construction.
Causes irritation in eyes and throat.

CO
185.89
CO - Toxic gas from vehicles.
Reduces oxygen in blood.



Pollutant Levels



AQI Dashboard

City: Chennai

Hours Outdoors: 2

Activity: low

Age: 25

Health Condition: none

Analyze Reset

CO - Toxic gas from vehicles.
Reduces oxygen in blood.

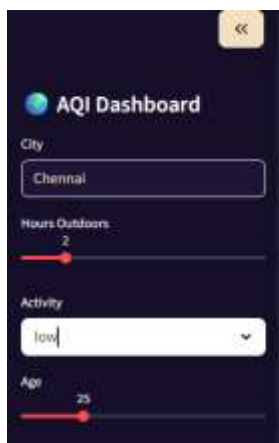
NO₂
5.46
NO₂ - Gas from traffic emissions.
Worsens asthma and lungs.

Exposure & Health

Exposure Score
24.36
Exposure Level
Low

Risk Score
1
Risk Level
Low

Health & Exposure



AQI Dashboard

City: Chennai

Hours Outdoors: 2

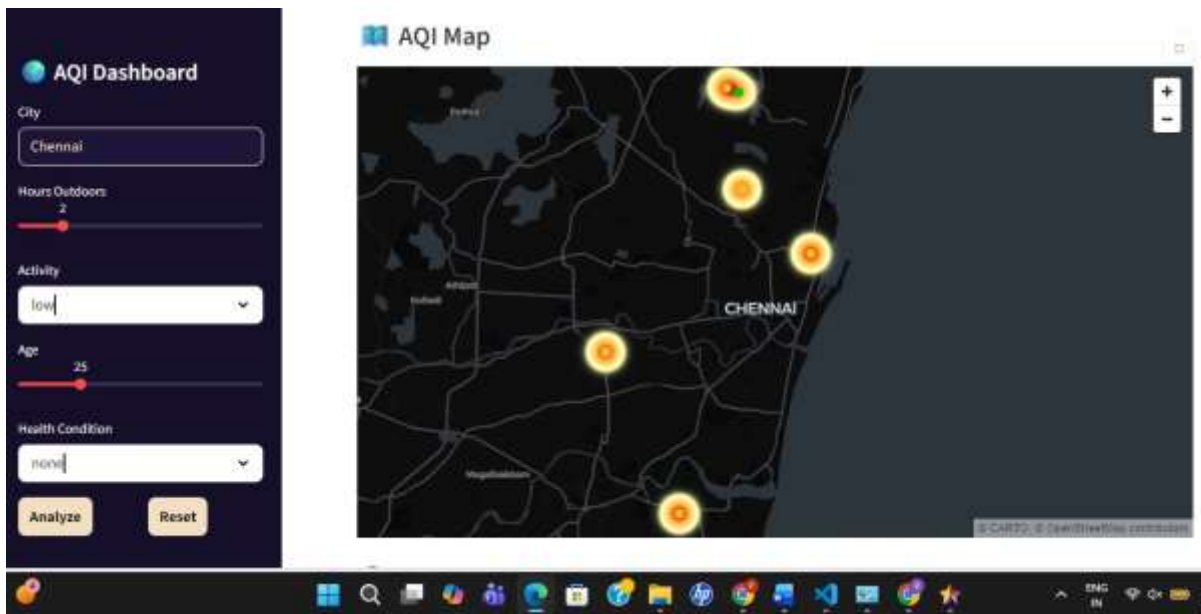
Activity: low

Age: 25

AQI Trend



AQI Trend



AQI Map

CONCLUSION AND KEY ACHIEVEMENTS

This research presented an intelligent Air Pollution Analysis and Health Impact Assessment System that integrates machine learning and data-driven techniques to provide meaningful environmental insights. By utilizing pollutant data such as PM_{2.5}, PM₁₀, NO₂, and SO₂, the system effectively estimates air quality levels and identifies major pollution sources using rule-based analysis. Unlike traditional air quality monitoring systems that only display AQI values, the proposed approach enhances interpretability by explaining causes and assessing health impacts.

The system successfully combines AQI estimation, cause detection, health risk assessment, and personalized exposure modeling within a unified web-based interface. The inclusion of user-specific inputs and actionable recommendations improves user awareness and supports informed decision-making. The project demonstrates the practical application of machine learning and analytical methods in environmental monitoring, providing a strong foundation for future advancements in intelligent air quality and health advisory systems.

REFERENCES

- ❖ **Air Quality Prediction Using Machine Learning Techniques** Kumar et al., 2021

<https://scholar.google.com/scholar?q=air+quality+prediction+machine+learning+PM2.5>

- ❖ **Real-Time Air Quality Monitoring System Using IoT** Alam et al., 2020

<https://scholar.google.com/scholar?q=real+time+air+quality+monitoring+iot>

❖ **Air Quality Index and Its Impact on Human Health** World Health Organization (WHO)

<https://www.who.int/health-topics/air-pollution>

❖ **Data-Driven Approaches for Environmental Monitoring** Zhang et al., 2022

<https://scholar.google.com/scholar?q=environmental+data+analytics+air+quality>

❖ **Health Risk Assessment Based on Air Pollution Exposure** U.S. Environmental Protection Agency (EPA)

<https://www.epa.gov/air-research>

❖ **Visualization Techniques for Air Quality Data** Chen et al., 2021

<https://scholar.google.com/scholar?q=air+quality+data+visualization>

❖ **Smart City Environmental Monitoring Systems** IEEE Research

<https://ieeexplore.ieee.org/search/searchresult.jsp?queryText=air%20quality%20smart%20city>

❖ **Air Pollution Forecasting Using Deep Learning (LSTM)** Li et al., 2022

<https://scholar.google.com/scholar?q=air+pollution+prediction+LSTM>

❖ **Web-Based Air Quality Monitoring and Visualization System** Singh et al., 2021

<https://scholar.google.com/scholar?q=air+quality+dashboard+web+application>

❖ **OpenWeather Air Pollution API**

<https://openweathermap.org/api/air-pollution>

❖ **OpenWeather Weather API**

<https://openweathermap.org/api>