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

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121232

Vento Aureo: IoT-Based Pollution Detection with ML Insights

Syed Muteeb Bakshi[1], Saakshi S Urs[2], Nishmitha Shetty B.S[3], Poornima H N[4]

Student, IoT Cybersecurity and Blockchain, K S Institute of Technology, Bengaluru, India[1-3]

Professor, IoT Cybersecurity and Blockchain, K S Institute of Technology, Bengaluru, India[4]

Abstract: Air pollution has emerged as a major environmental and public health concern, particularly in urban areas where pollutant concentrations vary significantly across locations and time. Traditional air quality monitoring systems rely on centralized monitoring stations and lack real-time localized sensing and predictive capabilities. To address these limitations, this paper presents the design and implementation of *Vento Aureo*, an IoT-based air quality monitoring system integrated with machine learning for Air Quality Index (AQI) forecasting. The proposed system utilizes multiple environmental sensors interfaced with an ESP32 microcontroller to measure key air quality parameters such as particulate matter, gaseous pollutants, temperature, and humidity. Sensor data is transmitted to a cloud-hosted backend through REST APIs, where AQI values are computed using standardized pollutant-weighted formulas. Historical data is stored and processed to enable forecasting using time-series machine learning models. A Prophet-based model is employed for short-term AQI prediction, while a hybrid Prophet–LSTM model is implemented to support long-term forecasting. A web-based dashboard is developed to visualize real-time air quality data, historical trends, and predicted AQI levels in an intuitive manner. Experimental evaluation demonstrates that the system effectively provides localized air quality monitoring along with meaningful AQI forecasts. The proposed solution offers a scalable and practical approach for environmental monitoring, public awareness, and smart city applications.

Keywords: Internet of Things (IoT), Air Quality Index (AQI), Air Pollution Monitoring, Machine Learning, Time-Series Forecasting, Prophet Model, LSTM, Cloud Computing, Real-Time Monitoring, Smart Cities.

I. INTRODUCTION

Air pollution has become one of the most critical environmental challenges affecting public health, climate conditions, and overall quality of life, especially in urban regions. Pollutants such as particulate matter (PM2.5 and PM10), carbon dioxide, and volatile organic compounds have been directly linked to respiratory diseases, cardiovascular disorders, and reduced life expectancy.[10] Continuous monitoring of air quality is therefore essential for timely awareness and preventive action.

Conventional air quality monitoring systems rely heavily on centralized monitoring stations, which are costly to deploy and provide limited spatial resolution. These systems often fail to capture localized variations in pollution levels and typically lack predictive capabilities that could help users anticipate hazardous conditions. As a result, there is a growing need for decentralized, low-cost, and intelligent air quality monitoring solutions.[2],[6]

Recent advancements in the Internet of Things (IoT) have enabled the development of compact sensor-based systems capable of collecting real-time environmental data from multiple locations.[2],[4] When combined with cloud computing, such systems can efficiently process, store, and analyze large volumes of sensor data. Furthermore, machine learning techniques, particularly time-series forecasting models, offer the potential to predict future air quality trends based on historical data patterns.[1],[3]

This paper presents the design and implementation of *Vento Aureo*, an IoT-based air quality monitoring system integrated with machine learning models for AQI forecasting. The system employs environmental sensors interfaced with an ESP32 microcontroller to collect real-time pollution data, which is transmitted to a cloud-hosted backend for AQI computation and storage. Short-term AQI prediction is performed using a Prophet model, while a hybrid Prophet–LSTM approach is used for long-term forecasting. A web-based dashboard provides real-time visualization, historical analysis, and forecasted AQI trends. The proposed system aims to offer a scalable and practical solution for localized air quality monitoring and predictive environmental analysis.

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

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121232

II. LITERATURE REVIEW

Several studies have explored the use of IoT technologies for air quality monitoring. Many researchers have proposed sensor-based systems to measure pollutants such as PM2.5, PM10, and gaseous emissions using low-cost sensors and microcontrollers. These systems demonstrate the feasibility of real-time environmental data collection but often lack predictive analysis and long-term data utilization.[2],[4],[6]

Machine learning techniques have been increasingly applied to air quality prediction problems. Time-series forecasting models such as autoregressive methods, neural networks, and deep learning approaches have shown promising results in predicting AQI trends. Long Short-Term Memory (LSTM) networks, in particular, have been widely used due to their ability to capture temporal dependencies in environmental data.[1],[3],[7] However, such models often require large datasets and careful tuning to maintain stability.

Some studies have employed hybrid approaches that combine statistical models with deep learning techniques to improve forecasting accuracy. Prophet-based models have been used effectively for short-term forecasting due to their robustness in handling seasonality and missing data.[8] Hybrid models integrating Prophet with LSTM have been shown to enhance long-term prediction performance by leveraging both trend modeling and deep temporal learning.[3],[5]

Despite these advancements, many existing systems focus either on monitoring or prediction, but not both in an integrated manner. Additionally, limited emphasis is placed on end-to-end implementation involving real-time sensing, cloud-based processing, machine learning forecasting, and user-friendly visualization. This work addresses these gaps by implementing a complete IoT–cloud–ML pipeline that provides real-time air quality monitoring along with short-term and long-term AQI forecasting.[4],[6]

III. SYSTEM ARCHITECTURE AND OVERALL WORKFLOW

The system is designed as a layered architecture integrating IoT-based environmental sensing, cloud-hosted backend processing, machine learning–based forecasting, and a web-based visualization dashboard.[4],[6]

Fig. 1 presents the block diagram of the complete system, illustrating the interaction between hardware components, cloud services, machine learning models, and the frontend interface.

A. IoT Sensing Layer

The IoT sensing layer is responsible for real-time acquisition of environmental parameters related to air quality. Multiple low-cost sensors are interfaced with an ESP32 microcontroller to capture pollutant concentrations and ambient conditions.

The sensing unit consists of the following components:

- PMS7003 sensor for measuring particulate matter concentrations, specifically PM2.5 and PM10.
- MQ135 gas sensor for detecting volatile organic compounds (VOCs) and gaseous pollutants.
- BME280 sensor for measuring temperature and humidity, which influence air quality behavior and sensor accuracy.

All sensors are connected to the ESP32 microcontroller, which performs periodic data acquisition. The ESP32 aggregates raw sensor readings, performs basic formatting, and transmits the collected data to the cloud backend over Wi-Fi using HTTP-based REST API calls.

This layer enables continuous and localized air quality data collection, forming the foundation for subsequent processing and analysis.

B. Cloud Backend Layer (Google Cloud)

The cloud backend layer serves as the central processing and data management component of the system. It is deployed on Google Cloud to ensure scalability, reliability, and remote accessibility. The backend consists of the following modules:

1) REST API Module

A RESTful API is implemented using the Flask framework to receive sensor data transmitted from the ESP32. The API validates incoming requests, parses sensor payloads, and forwards the data for further processing.

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

International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025
DOI: 10.17148/IARJSET.2025.121232

2) AQI Computation Module
The AQI computation module converts raw pollutant concentrations into standardized Air Quality Index (AQI) values based on official AQI guidelines.[9] Individual pollutant sub-indices are calculated and combined to generate an overall AQI score. This ensures that sensor data is transformed into meaningful and interpretable air quality metrics.

3) Cloud Database
A cloud-hosted database is used to store:
- Raw sensor readings
- Computed AQI values
- Timestamped historical records
- Forecasted AQI outputs generated by machine learning models

Persistent storage enables long-term analysis, model training, and retrieval of historical and forecast data.
The cloud backend acts as a bridge between the IoT devices, machine learning layer, and frontend dashboard.

C. Machine Learning Layer
The machine learning layer is responsible for analyzing historical AQI data and generating air quality forecasts. This layer enhances the system by providing predictive insights in addition to real-time monitoring. The following models are implemented:

1) Prophet Model (Short-Term Forecasting)
The Prophet model is used for short-term AQI forecasting due to its ability to model trends, seasonality, and missing data effectively. It provides stable short-range predictions suitable for near-future air quality estimation.[8]

2) LSTM Model (Long-Term Forecasting)
A Long Short-Term Memory (LSTM) neural network is employed to capture long-term temporal dependencies in AQI data. The LSTM model is trained on historical AQI time-series data to predict extended future trends.[7]

3) Hybrid Forecasting Model
A hybrid forecasting approach is implemented by combining outputs from the Prophet and LSTM models. This hybrid model leverages the strengths of both statistical and deep learning methods, improving forecasting robustness and accuracy across different time horizons.[3],[5]

Forecasted AQI values generated by the machine learning layer are stored back into the cloud database for visualization and analysis.

D. Frontend Dashboard Layer
The frontend dashboard provides a user-friendly interface for visualizing air quality data. It retrieves data from the cloud backend and presents it in an intuitive manner.
The dashboard supports:
- Real-time AQI visualization
- Historical AQI trends
- Short-term and long-term AQI forecasts

This layer enables users to easily interpret air quality conditions and predictive insights without requiring technical expertise.

E. End-to-End Workflow
The overall workflow of the system operates as follows:
1. Sensors collect environmental data at regular intervals.
2. The ESP32 aggregates and transmits sensor data to the cloud backend.
3. The backend computes AQI values and stores data in the cloud database.
4. Historical AQI data is used by machine learning models to generate forecasts.
5. Real-time, historical, and forecasted AQI values are displayed on the frontend dashboard.

This integrated architecture ensures seamless data flow from physical sensing to predictive analytics and visualization.

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

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121232

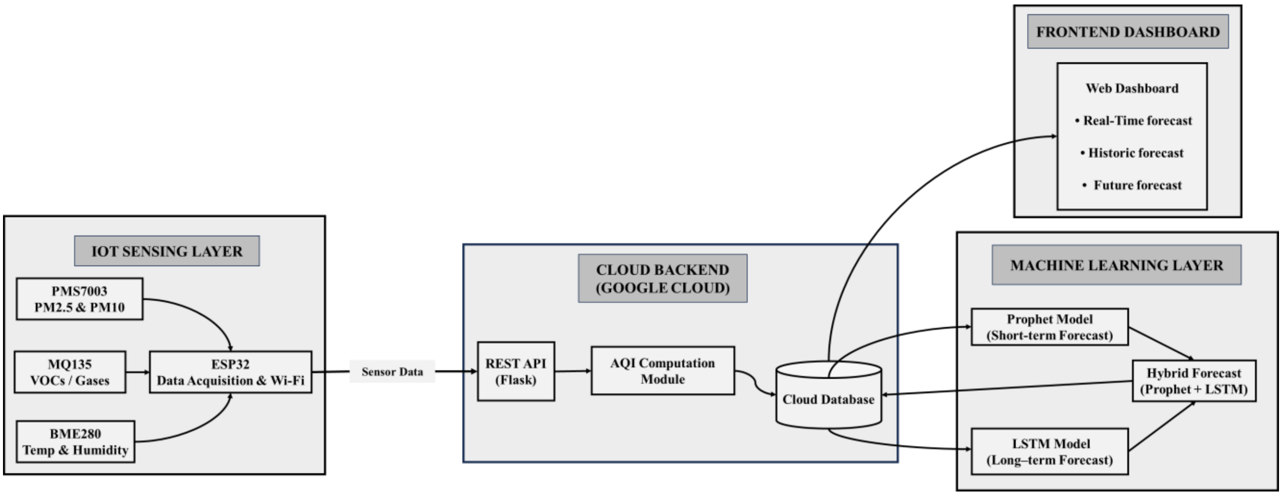

Fig. 1: Block Diagram of the proposed System for Vento Aureo

IV. AQI COMPUTATION AND MACHINE LEARNING IMPLEMENTATION

This section presents the complete implementation methodology adopted for converting raw IoT sensor data into standardized Air Quality Index (AQI) values and generating reliable short-term and long-term air quality forecasts using machine learning models. The methodology integrates environmental sensing, cloud-based processing, statistical modelling, and deep learning to form an end-to-end intelligent air quality monitoring system.

A. AQI Computation Framework

The IoT sensing layer continuously measures pollutant concentrations and environmental parameters. However, these raw values are not directly interpretable for air quality assessment. Hence, a standardized AQI computation framework is implemented at the cloud backend.

The system computes AQI values based on pollutant concentration breakpoints defined by internationally accepted AQI standards. Pollutants considered include PM2.5, PM10, and gas concentration values obtained from MQ-series sensors. For each pollutant, a sub-index is computed using linear interpolation between predefined concentration ranges:

$$\boldsymbol{AQI_p} = \frac{(\boldsymbol{I_{high}} - \boldsymbol{I_{low}})}{(\boldsymbol{C_{high}} - \boldsymbol{C_{low}})} \times (\boldsymbol{C_p} - \boldsymbol{C_{low}}) + \boldsymbol{I_{low}}$$

where:

- C_p represents the measured pollutant concentration
- C_{low} and C_{high} denote pollutant breakpoint limits
- I_{low} and I_{high} represent AQI index limits

The final AQI value is selected as the maximum among all computed sub-indices:

$$AQI = \max\,(AQI_{PM2.5}, AQI_{PM10}, AQI_{Gas})$$

This approach ensures that the pollutant posing the highest health risk determines the overall air quality classification. Computed AQI values are categorized into standard AQI levels such as *Good*, *Moderate*, *Poor*, and *Severe* before being stored in the cloud database.

B. Cloud-Based Data Processing Pipeline

The backend is implemented as a cloud-hosted RESTful service, responsible for processing sensor data received from ESP32 devices. Each incoming data packet undergoes the following steps:

1. Payload validation and timestamp verification
2. Noise filtering and missing value handling
3. AQI computation and classification

DOI: 10.17148/IARJSET.2025.121232

4. Persistent storage in the cloud database

The backend architecture is designed to support continuous data ingestion from multiple IoT devices, ensuring scalability and fault tolerance.

C. Data Preparation for Machine Learning

Historical AQI data accumulated in the cloud database is periodically extracted for machine learning training. Prior to model training, the dataset undergoes extensive preprocessing to ensure stability and accuracy.

Preprocessing steps include:

- Removal of invalid or incomplete records
- Temporal resampling and alignment
- Normalization of AQI values
- Sliding window sequence generation for time-series modeling

This processed dataset serves as the input for both statistical and deep learning models.

D. Short-Term AQI Forecasting Using Prophet

For short-term AQI prediction, the Prophet time-series forecasting model is implemented due to its robustness and interpretability. Prophet models AQI trends using additive components:

$$\boldsymbol{y(t) = g(t) + s(t) + h(t) + \epsilon_t}$$

where:

- $g(t)$captures long-term trends
- $s(t)$models seasonal patterns
- $h(t)$accounts for irregular variations
- $\epsilon_t$represents residual noise

Prophet is particularly effective for short-term forecasting as it handles missing data gracefully and adapts well to sudden fluctuations in environmental conditions. The model is trained using historical AQI values and produces forecasts for near-future intervals.

E. Long-Term AQI Forecasting Using LSTM

To capture complex temporal dependencies and long-term pollution patterns, a Long Short-Term Memory (LSTM) neural network is employed. LSTM networks are well-suited for sequential data due to their gated memory architecture.

The LSTM model learns temporal relationships using forget, input, and output gates, enabling it to retain long-term contextual information. This allows the system to forecast AQI trends over extended durations, reflecting seasonal and gradual environmental changes.

The LSTM model is trained using normalized AQI sequences generated from historical data stored in the cloud backend.

F. Hybrid Forecasting Model

While Prophet excels in short-term forecasting and LSTM performs well for long-term predictions, each model has limitations when used independently. To overcome this, a hybrid forecasting approach is implemented.

The hybrid model combines outputs from both Prophet and LSTM:

$$\boldsymbol{AQI_{hybrid} = \alpha \cdot AQI_{Prophet} + (1 - \alpha) \cdot AQI_{LSTM}}$$

where $\alpha$is an empirically determined weighting factor.

This hybrid approach improves forecast stability, reduces prediction error, and ensures consistency across multiple forecasting horizons.

G. Forecast Integration and Visualization

All computed AQI values and forecasted results are stored in the cloud database. The frontend dashboard retrieves this data through API endpoints and visualizes:

- Real-time AQI values
- Historical AQI trends
- Short-term and long-term AQI forecasts

DOI: 10.17148/IARJSET.2025.121232

This enables users to gain actionable insights into current and future air quality conditions.

V. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

This section discusses the experimental evaluation of the proposed **Vento Aureo** system, focusing on the performance of AQI computation and forecasting models under real-time and historical data scenarios. The evaluation aims to validate the effectiveness of integrating IoT sensing, cloud-based processing, and machine learning forecasting into a unified system.

A. Experimental Setup

The experimental setup consists of an IoT sensing unit equipped with particulate matter and gas sensors connected to an ESP32 microcontroller. Sensor data was periodically transmitted to the cloud backend over WiFi using REST APIs. The backend processed incoming data to compute AQI values and store them in a cloud database.

Historical AQI datasets were generated using a combination of collected sensor data and publicly available air quality records. These datasets were used to train and evaluate the forecasting models.
All experiments were conducted in a cloud-hosted environment to simulate real-world deployment conditions, including continuous data ingestion and remote access.

B. AQI Computation Validation

The AQI computation module was validated by testing it with multiple pollutant concentration scenarios, including normal, elevated, and extreme pollution levels. The computed AQI values were verified against standard AQI category ranges.
The system consistently classified air quality into appropriate categories such as *Good*, *Moderate*, and *Poor* based on dominant pollutant concentrations. This confirmed the correctness of the AQI computation logic and its suitability for real-time monitoring.

C. Forecasting Model Evaluation

Three forecasting approaches were evaluated:

- Prophet-based short-term forecasting
- LSTM-based long-term forecasting
- Hybrid Prophet–LSTM forecasting

The Prophet model demonstrated stable short-term predictions and effectively captured daily and weekly seasonal patterns in AQI data. It performed well in handling missing values and irregular sampling intervals.
The LSTM model showed improved performance for long-term trend prediction by learning temporal dependencies across extended time windows. However, it required careful preprocessing and tuning to maintain prediction stability.
The hybrid forecasting model combined the strengths of both approaches. By integrating Prophet's trend modeling with LSTM's temporal learning, the hybrid model produced smoother and more consistent predictions across different forecasting horizons.

D. Comparative Analysis

A qualitative comparison of the three forecasting approaches revealed that:

- Prophet performs best for short-term AQI forecasting
- LSTM captures long-term pollution trends more effectively
- The hybrid model reduces prediction fluctuations and improves overall robustness

The hybrid approach demonstrated improved consistency when compared to standalone models, especially under varying pollution conditions and seasonal transitions.

E. Visualization and System Output

The frontend dashboard successfully displayed real-time AQI values, historical trends, and forecasted AQI results. Visual representations enabled intuitive interpretation of air quality patterns and prediction outputs.The integration of real-time monitoring with predictive analytics enhanced situational awareness and demonstrated the practical applicability of the proposed system in environmental monitoring scenarios.

F. Discussion

The experimental results confirm that the proposed system effectively integrates IoT sensing, cloud processing, and machine learning forecasting. While standalone models provide useful insights, the hybrid forecasting approach offers improved reliability and adaptability.

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

International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025
DOI: 10.17148/IARJSET.2025.121232

The system demonstrates scalability and flexibility, making it suitable for smart city deployments, environmental monitoring applications, and public health awareness systems.

VI. CONCLUSION AND FUTURE SCOPE

A. Conclusion

This paper presented the design and implementation of Vento Aureo, an IoT-based air quality monitoring and forecasting system integrated with cloud computing and machine learning techniques. The system combines real-time environmental sensing with advanced data processing to provide meaningful air quality insights.

Low-cost IoT sensors were used to collect pollutant concentration and environmental data, which were transmitted to a cloud-hosted backend for AQI computation and storage. Machine learning models were implemented to perform both short-term and long-term AQI forecasting. The use of a hybrid forecasting approach enabled improved stability and consistency in prediction results.

Experimental results demonstrated that the proposed system can effectively monitor air quality, analyze historical trends, and generate reliable AQI forecasts. The integration of IoT sensing, cloud-based processing, and machine learning makes the system suitable for real-world environmental monitoring applications and smart city initiatives.[6],[10]

B. Future Scope

The proposed system can be further enhanced in several ways to improve accuracy, scalability, and usability:

1. Additional air quality sensors can be integrated to monitor a wider range of pollutants and environmental parameters.
2. Edge computing techniques can be incorporated to perform preliminary data processing directly on IoT devices, reducing cloud dependency and latency.
3. Mobile and web applications can be developed to provide real-time alerts and personalized air quality insights to users.
4. Advanced deep learning models and adaptive learning techniques can be explored to further improve forecasting accuracy.
5. The system can be deployed across multiple geographic locations to support large-scale smart city and public health monitoring initiatives.

REFERENCES

[1]. Agbehadji, I. E., & Obagbuwa, I. C. (2024). Systematic Review of Machine Learning and Deep Learning Techniques for Spatiotemporal Air Quality Prediction. Atmosphere, 15(11), 1352. MDPI.
[2]. Canales-Calero, A. J., Solis-Rosales, S. S., & Carrera-Salas, E. A. (2025). IoT-based Mobile App for Real-Time Monitoring and Reporting of Air Pollution. Engineering Research Express. IOP Publishing.
[3]. Rahman, M. M., Nayeem, M. E. H., Ahmed, M. S., Tanha, K. A., Sakib, M. S. A., Uddin, K. M. M., & Babu, H. M. H. (2024). AirNet: Predictive Machine Learning Model for Air Quality Forecasting Using Web Interface. Environmental Systems Research, 13(1), 44. Springer.
[4]. Moursi, A. S., El-Fishawy, N., Djahel, S., & Shouman, M. A. (2021). An IoT-Enabled System for Enhanced Air Quality Monitoring and Prediction on the Edge. Complex & Intelligent Systems, 7(6), 2923–2947. Springer.
[5]. Ramadan, M. N. A., Ali, M. A. H., Khoo, S. Y., Alkhedher, M., & Alherbawi, M. (2024). Real-time IoT-Powered AI System for Monitoring and Forecasting of Air Pollution in Industrial Environment. Ecotoxicology and Environmental Safety, 283, 116856. Elsevier.
[6]. Toma, C., Alexandru, A., Popa, M., & Zamfiroiu, A. (2019). IoT Solution for Smart Cities' Pollution Monitoring and the Security Challenges. Sensors, 19(15), 3401. MDPI.
[7]. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780.
[8]. Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. The American Statistician, 72(1), 37–45.
[9]. U.S. Environmental Protection Agency. (2018). Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI).
[10]. World Health Organization. (2023). Ambient (Outdoor) Air Pollution.