IOT AND AI-BASED REAL-TIME AIR QUALITY MONITORING AND PREDICTION SYSTEM

Dr. K. Aravinda Shilpa Assistant Professor, Department of Electrical Engineering, Andhra University College of Engineering for Women, Visakhapatnam – 530017, Andhra Pradesh, India
Email: aravindashilpacf@andhrauniversity.edu.in
Dr. M. Nagavalli Assistant Professor, Department of Mathematics, Andhra University College of Engineering for Women, Visakhapatnam – 530017, Andhra Pradesh, India
Email: nagavallicf@andhrauniversity.edu.in
Mocherla Lakshmi Sai Satya Lavanya Student, Department of Electrical Engineering, Andhra University College of Engineering for Women, Visakhapatnam – 530017, Andhra Pradesh, India
Email: lavanyamocherla6@gmail.com
Sanapala Bhargavi Student, Department of Electrical Engineering, Andhra University College of Engineering for Women, Visakhapatnam – 530017, Andhra Pradesh, India
Email: bhargavisanapala5@gmail.com
Akiri Sravani Student, Department of Electrical Engineering, Andhra University College of Engineering for Women, Visakhapatnam – 530017, Andhra Pradesh, India
Email: sravani9277@gmail.com
Sariyam Kavya Student, Department of Electrical Engineering, Andhra University College of Engineering for Women, Visakhapatnam – 530017, Andhra Pradesh, India
Email: sariyamkavya@gmail.com

Abstract

Deteriorating urban air quality, driven by industrial emissions and rapid urbanization, poses serious risks to human health and ecological balance. Conventional air quality monitoring systems are constrained by high deployment costs and limited spatial coverage, rendering them inadequate for continuous, large-scale environmental surveillance. To address these limitations, this paper proposes an integrated Internet of Things (IoT) and Artificial Intelligence (AI)-based system for real-time air quality monitoring and prediction. The system employs MQ-series gas sensors along with temperature and humidity sensing modules, which transmit environmental data to a cloud platform via wireless communication.

Machine learning algorithms, including Random Forest, Gradient Boosting, and a Voting Classifier ensemble, are applied to classify air quality into distinct hazard categories. Additionally, an Artificial Neural Network (ANN) is employed to model temporal patterns in sensor data for predictive analysis of future pollution trends. Experimental results demonstrate that the proposed system achieves a maximum classification accuracy of 97.3% using the ensemble Voting Classifier, while offering a scalable, low-cost, and efficient solution for smart environmental monitoring in urban and industrial settings.

Keywords: Air Pollution Monitoring; Internet of Things; Air Quality Prediction; Machine Learning; Node MCU; Environmental Monitoring; Random Forest; Artificial Neural Networks; Smart Environmental System.

Introduction

Rapid urbanization, industrial expansion, and increasing vehicular emissions have contributed to a significant deterioration of air quality in cities worldwide. Exposure to airborne pollutants poses severe threats to human health, causing respiratory and cardiovascular diseases, while simultaneously disrupting ecological balance. Traditional air quality monitoring systems rely on precise but expensive analytical instruments that are typically deployed at fixed locations. This static infrastructure limits spatial coverage and restricts the availability of real-time, geographically distributed environmental data, thereby impeding effective pollution management.

The emergence of Internet of Things (IoT) technology has opened new avenues for the development of smart environmental monitoring systems. In parallel, the application of machine learning (ML) techniques to environmental data has demonstrated considerable promise in the accurate prediction of pollution levels. Ensemble methods such as Random Forest and Gradient Boosting have proven robust in multi-class classification tasks, while Artificial Neural Networks (ANNs) offer enhanced predictive accuracy by capturing complex non-linear relationships within environmental datasets.

Problem Statement

Air pollution has emerged as a serious environmental and health issue due to rapid urbanization, industrial emissions, and vehicular exhaust. Traditional air quality monitoring systems mainly focus on measuring individual gas concentrations such as CO, NO_2, and SO_2 using fixed and expensive monitoring stations. These systems are limited in number and cannot provide detailed, area-wise pollution analysis. Furthermore, they do not recognize complex odour patterns formed by mixed gases in real-world environments. As a result, they fail to capture the complete environmental impact of pollution, especially in densely populated urban areas.

Therefore, there is a need to develop a cost-effective, AI-based electronic nose system capable of recognizing complex smell patterns and providing real-time, area-wise pollution monitoring through IoT integration. Such a system should improve selectivity, adaptability, and reliability while being suitable for practical environmental deployment.

The integration of IoT technology with machine learning provides an effective approach to address this problem. IoT-based systems can collect real-time environmental data using sensors, while machine learning algorithms can analyse the collected data to identify patterns and classify pollution levels. By combining these technologies, it is possible to develop an efficient system capable of monitoring air quality, analysing pollution data, and predicting possible future trends.

Objectives of the Study

- To collect real-time sensor data and process it for identifying harmful gases and odour mixtures.
- To apply artificial intelligence/machine learning techniques for improving gas selectivity and reducing cross-sensitivity issues.
- To classify and estimate pollution levels based on combined sensor responses rather than individual gas readings.
- To integrate IoT technology for real-time monitoring and remote data visualization.
- To design a system capable of area-wise pollution analysis for smart city and environmental applications. To provide an intelligent, low cost, scalable, and adaptive solution for environmental based pollution monitoring.

Review of Literature

J. Burrell et al. (2004) proposed the use of wireless sensor networks for monitoring environmental conditions in agricultural fields. Their system collects real-time data from distributed sensor nodes. The collected data is transmitted to a central system for analysis and decision-making. It improves efficiency in data collection and decision-making. Kevin Ashton (2009) introduced the concept of the Internet of Things (IoT), where physical devices are connected to the internet. He explained how sensors and RFID can enable automatic data exchange. This reduces human intervention in monitoring systems. His work laid the foundation for modern smart monitoring applications.

Fanli Meng, et al. (2025) developed a semiconductor gas sensor under Dynamic Temperature Modulation (DTM) to extract unique "fingerprints" of industrial gases. **N. S.** Safiedze, et al. (2023) focuses on optimizing the raw data coming from gas sensors (MQ-2 and MQ-5) within a home environment. It uses the Kalman Filter algorithm to reduce "noise" in the sensor readings, ensuring that the identification of smoke or gas leaks is accurate and not triggered by random environmental fluctuations. Prof. Bal Gopal. et al.(2009) demonstrates how an E-nose mimics the human sense of smell using an array of chemical gas sensors. It focuses on generating a unique "smell Print" for various VOCs to classify substances and measure their concentrations in industrial and medical fields.

Jerome H. Friedman (2001) introduced Gradient Boosting for improving prediction accuracy. Models are built sequentially to correct errors. It enhances performance in complex datasets. Useful in classification problems. K. Mahalakshmi, et al. (2021) developed various ML algorithms to predict future air pollution levels based on historical sensor data. It focuses on the transition from simple monitoring to "intelligent forecasting" using regression-based AI models. Hafsa Binte Kibria, et al. (2025) focuses on creating a "lightweight" AI system. It uses machine learning to convert raw sensor data (PM {2.5}, CO, NOx, etc.) into a real-time Air Quality Index (AQI) value, specifically optimized to run on low-power microcontrollers.

Proposed System Architecture

The proposed system architecture integrates four functional layers - sensing, processing, communication, and data analytics to enable continuous real-time monitoring and prediction of ambient air quality.

A. Sensing Layer

The sensing layer comprises MQ-series gas sensors (MQ135, MQ7, and MQ2) for the detection of gaseous pollutants, including carbon monoxide, ammonia, benzene, and smoke. The DHT11 sensor module measures ambient temperature and relative humidity, both of which are known to influence pollutant concentration and dispersion dynamics. The overall system architecture is illustrated in Fig 1.

B. Processing Layer

A Node MCU microcontroller (ESP8266) serves as the central processing unit, responsible for acquiring Analog and digital sensor readings and conditioning the data for wireless transmission. The integrated Wi-Fi module of the Node MCU facilitates seamless connectivity to the cloud infrastructure.

C. Communication Layer

Sensor data is transmitted over Wi-Fi to the Thing Speak IoT cloud platform via the MQTT/HTTP protocol. This enables real-time remote monitoring through graphical dashboards accessible from any internet-connected device. The overall system architecture is illustrated in Fig 1.

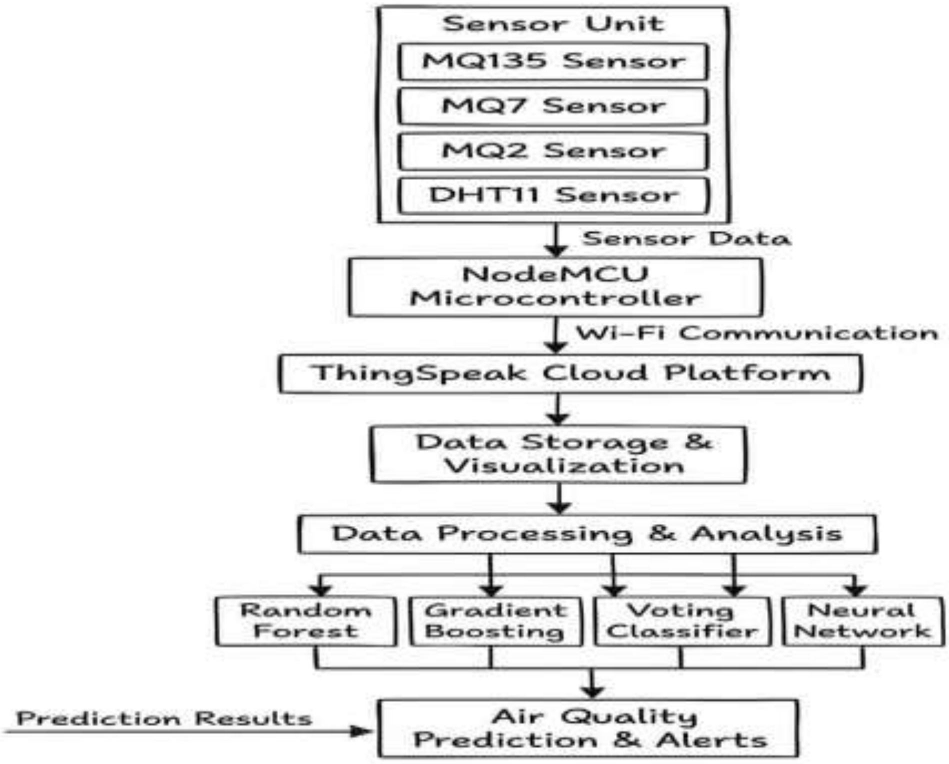

Fig 1. Proposed System Architecture

D. Data Analytics Layer

The analytics layer employs machine learning algorithms—Random Forest, Gradient Boosting, Voting Classifier, and ANN—to classify air quality into defined categories (safe, moderate, and hazardous) and to predict future pollution levels.

The proposed architecture provides an end-to-end, cost-effective solution for smart environmental monitoring, integrating distributed sensing, cloud communication, and intelligent analytics.

System Design

The system is designed to acquire, transmit, and analyse environmental data using an IoT framework augmented by machine learning models.

A. Hardware Design

The hardware configuration consists of MQ135, MQ7, and MQ2 sensors for gas detection and a DHT11 sensor for temperature and humidity measurement. All sensors interface with the Node MCU, which functions as the central data acquisition and communication hub.

B. Communication Design

The Node MCU [16] transmits aggregated sensor readings to the Thing Speak cloud platform over Wi-Fi, enabling real-time remote monitoring through graphical data visualizations.

C. Software Design

Firmware for the Node MCU is developed using the Arduino Integrated Development Environment (IDE). Machine learning models are implemented in Python using the Scikit-learn and TensorFlow libraries, which provide efficient tools for model training, validation, and deployment.

D. Data Processing

Data collected from IoT sensors is stored on the cloud platform and subsequently retrieved for analytical processing. The dataset is partitioned into training and testing subsets following standard protocols. These subsets are used to train and evaluate machine learning models for air quality classification and pollution trend forecasting.

Machine Learning Methodology

Environmental data acquired continuously from IoT sensors is stored in the cloud and retrieved for intelligent analysis. This facilitates real-time air quality assessment and data-driven decision-making.

A. Data Acquisition

Gas concentration, temperature, and humidity measurements are captured by dedicated sensors interfaced with the Node MCU. The microcontroller transmits the aggregated data wirelessly to the cloud server, where it is centrally stored and made available for subsequent analysis.

B. Data Preparation

Prior to model development, the acquired sensor data undergoes a comprehensive preprocessing pipeline to enhance data quality and consistency. Erroneous, duplicate, and missing values are identified and handled through appropriate imputation or removal strategies. Feature scaling techniques, such as min-max normalization or z-score standardization, are applied to ensure that all input variables contribute equitably to model training, thereby improving learning efficiency and predictive performance.

C. Machine Learning Model Implementation

Multiple machine learning models are employed to classify air quality levels with high accuracy. Random Forest and Gradient Boosting classifiers are selected for their robustness in managing high-dimensional, non-linearly separable data. An ensemble Voting Classifier is constructed by combining the predictions of multiple base learners, which enhances overall classification reliability through majority voting.

An Artificial Neural Network (ANN) employing a Multilayer Perceptron (MLP) architecture is additionally trained to capture complex non-linear patterns in the sensor data. This model is particularly effective in modelling temporal dependencies and supports accurate multi-step pollution forecasting under dynamically changing environmental conditions.

Results and Performance Analysis

A. Prototype Development

The hardware prototype integrates MQ-series gas sensors with temperature and humidity sensing modules, all connected to a Node MCU (ESP8266) microcontroller. The device continuously collects environmental data and transmits it to the Thing Speak cloud platform. An onboard LCD display provides immediate local feedback, enabling visual verification of sensor functionality during field operation. The Hardware Setup and MQ sensors outputs is shown in Fig 2.

Fig 2. Hardware setup and MQ sensors outputs

B. Environmental Data Monitoring

The sensing module continuously acquires real-time air quality measurements at defined sampling intervals. All readings are automatically transmitted to the cloud platform, enabling remote stakeholders to access live pollution data and track long-term environmental trends without interruption.

C. Performance Evaluation of Models

The classification performance of the implemented machine learning algorithms is evaluated using prediction accuracy as the primary metric. The comparative performance results are summarized in Table1.

Table 1: Comparative Classification Performance of Machine Learning Models

Model	Accuracy (%)
Random Forest Classifier	94.2
Gradient Boosting Classifier	95.6
Voting Classifier (Ensemble)	97.3
Artificial Neural Network (MLP)	96.1

D. Cloud-Based Visualization

The sensor readings are continuously plotted on the Thing Speak platform, which generates real-time graphical representations of environmental parameter variations. This functionality supports continuous remote monitoring and temporal trend analysis from geographically distributed locations. Cloud-based data visualizations are presented in Fig 3.

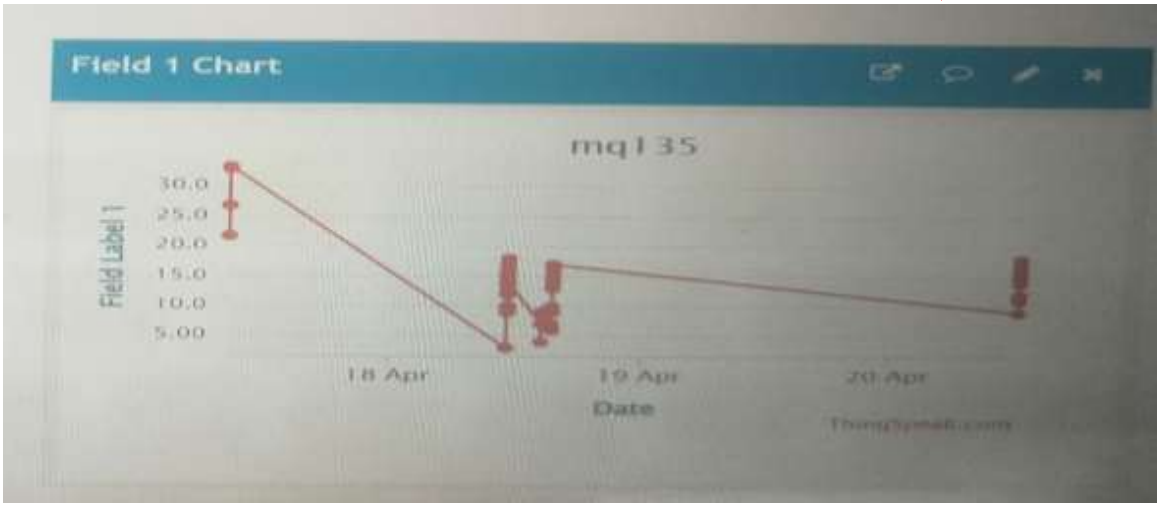

Fig 3. Real-Time Environmental Data on Thing Speak Platform

E. Output Classification

The trained classification models categorize ambient air quality into three discrete levels: safe, moderate, and hazardous. Among all evaluated approaches, the ensemble Voting Classifier consistently achieves the highest classification reliability. Detailed machine learning classification results are depicted in Fig 4.

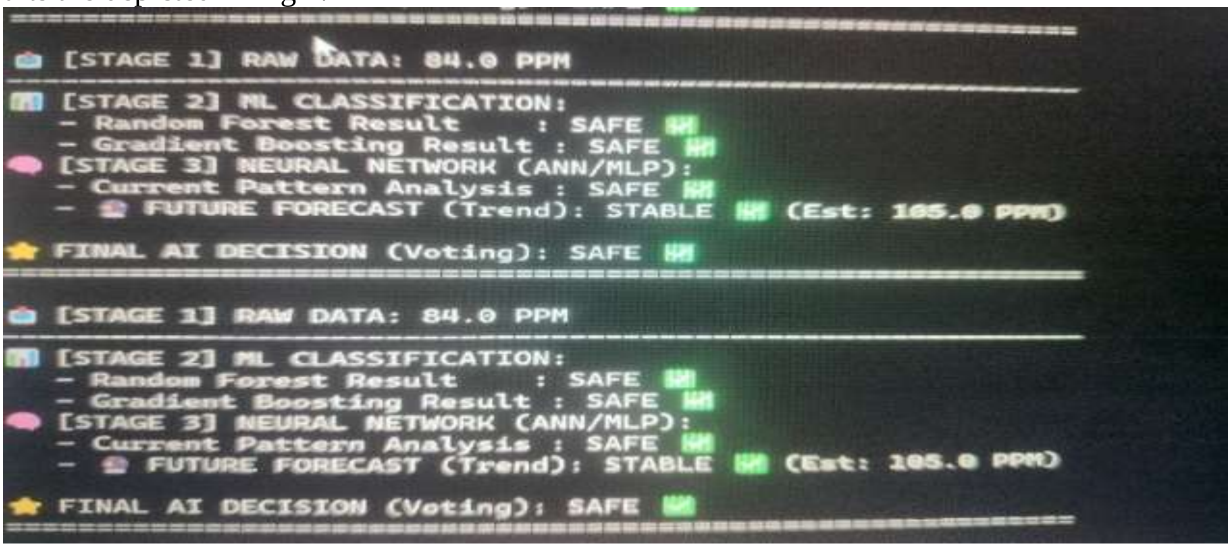

Fig 4. Machine Learning Classification Results

Applications and Future Scope

The proposed IoT and AI-based air quality monitoring system is well-suited for deployment in environments requiring continuous environmental surveillance. By integrating real-time sensing with intelligent predictive analytics, the system supports a broad range of practical applications.

In urban infrastructure management, the system enables the identification of pollution hotspots, supports evidence-based environmental planning, and assists regulatory authorities in formulating and enforcing air quality standards. In industrial environments, it facilitates continuous emissions monitoring to ensure regulatory compliance. The system also serves as a valuable tool for longitudinal pollution research, providing structured datasets that enable scientists to analyse pollution dynamics and forecast future trends with greater accuracy.

Several enhancements are envisioned to extend the capabilities of the current system. Integration of particulate matter sensors (PM2.5 and PM10) will broaden the pollutant detection spectrum and improve measurement comprehensiveness. Deploying a distributed network of sensor nodes will significantly enhance spatial coverage, enabling city-scale environmental monitoring.

Future work will incorporate advanced deep learning architectures, such as Long Short-Term Memory (LSTM) networks, to improve temporal prediction accuracy. Additionally, the development of mobile-based alert systems will enhance real-time user accessibility, enabling immediate notification of hazardous air quality conditions.

Conclusion

This paper presents a real-time air quality monitoring and prediction system that integrates Internet of Things sensing with machine learning-based analysis. The system employs gas, temperature, and humidity sensors interfaced with a Node MCU microcontroller, which transmits environmental data to the Thing Speak cloud platform for visualization, storage, and intelligent processing.

Four machine learning models - Random Forest, Gradient Boosting, Voting Classifier, and Artificial Neural Network were implemented and evaluated for air quality classification. The ensemble Voting Classifier achieved the highest classification accuracy of 97.3%, demonstrating the effectiveness of combining multiple learners for environmental decision-making. The integrated IoT-ML framework enables continuous real-time monitoring and accurate pollutant level forecasting.

The developed system offers a scalable, cost-effective, and reliable solution for smart environmental monitoring in urban and industrial contexts. Future enhancements, including the integration of deep learning models and expanded sensor arrays, are expected to further improve the system's predictive capability and deployment scope.

References

[1] J. Burrell, T. Brooke, and R. Beckwith, "Vineyard Computing: Sensor Networks in Agricultural Production," IEEE Pervasive Computing, vol. 3, no. 1, pp. 38–45, Jan.–Mar. 2004.
[2] K. Ashton, "That 'Internet of Things' Thing," RFID Journal, vol. 22, no. 7, pp. 97–114, 2009.
[3] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for Smart Cities," IEEE Internet of Things Journal, vol. 1, no. 1, pp. 22–32, Feb. 2014.
[4] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[5] J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.
[6] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.
[7] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
[8] T. Mitchell, Machine Learning. New York, NY, USA: McGraw-Hill Education, 1997.
[9] U. D. Singh and S. Kumar, "IoT Based Air Pollution Monitoring System Using Arduino," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 3, no. 1, pp. 123–129, 2018.
[10] M. F. Ramli, N. A. M. Zaini, and S. S. A. Zulkifli, "Air Quality Monitoring System Using IoT Based on Cloud Platform," International Journal of Engineering and Advanced Technology, vol. 8, no. 5, pp. 76–80, 2019.
[11] P. Rajalakshmi and S. D. Mahalakshmi, "IoT Based Air Pollution Monitoring System," International Journal of Advanced Research in Computer and Communication Engineering, vol. 5, no. 9, pp. 150–152, 2016.
[12] S. K. Gupta, P. Gupta, and S. Sharma, "Real-Time Air Quality Monitoring Using IoT," International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 6, pp. 235–240, 2019.
[13] D. Evans, "The Internet of Things: How the Next Evolution of the Internet is Changing Everything," Cisco White Paper, 2011.
[14] C. M. Bishop, Neural Networks for Pattern Recognition. Oxford, UK: Oxford University Press, 1995.
[15] ThingSpeak IoT Analytics Platform. [Online]. Available: https://thingspeak.com
[16] ESP8266 NodeMCU Technical Reference Manual, Espressif Systems, 2020.

