International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Scalable IOT Air-Quality Monitoring for Smart Cities: ESP32 Sensor Nodes and Random Forest Forecasting

Rasal Atul Balasaheb[1], Dr. G. B. Dongre[2], Prof. Vithhal Bhosale[3]

[1] *CSMSS Chh.Shahu college of engineering kanchanwadi, Chh. Sambhajinagar*
.[2] *CSMSS Chh.Shahu college of engineering kanchanwadi, Chh. Sambhajinagar*
[3]*CSMSS Chh.Shahu college of engineering kanchanwadi, Chh. Sambhajinagar*

---***---

Abstract -
In Rapid urbanization and increasing vehicular emissions have intensified the need for reliable air-quality monitoring frameworks in smart-city environments. This work presents a scalable IoT-based air-quality monitoring and prediction system built around ESP32 sensor nodes equipped with MQ-135, PM2.5, temperature, and humidity sensors. The nodes continuously measure on-site environmental parameters and transmit the data to a cloud platform for storage, visualization, and further analysis. To enhance the system's decision-making capability, a Random Forest model is trained to predict short-term air-quality trends, with a focus on PM2.5 levels due to their significant health impact. Experimental trials were conducted over several weeks using live sensor data collected from multiple indoor and outdoor locations. The proposed system demonstrated stable real-time performance with average communication latency below 300 ms. The Random Forest model achieved an R^2 score of 0.92, outperforming linear regression and decision tree baselines. The RMSE value reduced by nearly 28% when compared to single model predictions, indicating better noise tolerance and improved reliability for low-cost sensor data. Additionally, calibration against reference readings showed an average measurement deviation of ±6–8%, which is acceptable for economical sensor deployments. The findings confirm that the combination of ESP32-based sensing and Random Forest forecasting provides a practical, low-cost, and scalable approach for smart-city air-quality management. The system can be expanded easily to larger sensor networks, making it suitable for pollution surveillance, environmental research, and community health applications. Management.

Key Words: Air Quality Monitoring, Prediction Model, Random Forest, Wireless Sensor Network

1. INTRODUCTION

Air pollution remains one of the most persistent environmental threats worldwide, with PM2.5 posing severe risks due to its ability to penetrate deep into the human respiratory system. Rapid urban expansion, vehicular congestion, and industrial activity have collectively contributed to rising PM2.5 concentrations, significantly affecting public health and urban livability [1]. Although reference-grade air-quality monitoring stations provide highly precise measurements, their high deployment and maintenance costs restrict wide-scale coverage, limiting their effectiveness in capturing localized pollution variations [2]. In recent years, low-cost Internet of Things (IoT) sensor networks have emerged as scalable alternatives capable of producing dense, real-time air-quality datasets. Several studies demonstrate that low-cost sensor modules, when paired with appropriate calibration, can approximate the performance of commercial analyzers while enabling broad area data acquisition suitable for smart-city applications [1], [13]. The ESP32 microcontroller has become a preferred platform due to its integrated Wi-Fi, low energy consumption, edge-processing capabilities, and compatibility with multisensory configurations such as MQ-135 gas sensors, PM2.5 modules, and temperature/humidity sensors [8]. Recent deployments confirm its reliability in continuous, cloud connected monitoring environments [5].

Beyond observation, predictive modeling is increasingly used to anticipate short-term pollution fluctuations, enabling early-warning systems and data-driven environmental planning. Machine-learning approaches—particularly Random Forest—have demonstrated consistent superiority in PM2.5 and AQI prediction because of their robustness to sensor noise and ability to model nonlinear relationships [3], [12].

Fig. 1 Scenario depicting Air Quality Monitoring System

Recent research further confirms Random Forest's strong generalization across multi-location datasets and its effectiveness in improving prediction accuracy in low-cost sensor networks [7], [9]. Motivated by these trends, this study proposes a scalable IoT-based environmental monitoring and forecasting system integrating ESP32, MQ-135, PM2.5, and temperature sensors, combined with a Random Forest prediction model, to deliver precise, high-resolution, and cost-effective environmental intelligence for smart-city applications.

2. RELATED WORK

A. Review of Existing Work Recent research has highlighted significant advancements in low-cost air-quality sensing, IoT system architectures, and machine-learning–driven forecasting models. Colléaux [1] concluded that cost-effective sensors can reliably approximate reference analyzers when supported by adequate calibration. Hayward et al. [15] reinforced this by showing that low-cost PM sensors require ML-based correction to mitigate biases and environmental drift. Similarly, Alvear-Puertas et al. [13] demonstrated the viability of portable IoT devices for continuous monitoring in both indoor and outdoor settings. IoT-based environmental monitoring has evolved with robust microcontroller platforms like ESP32. Tasmurzayev et al. [4] validated a low-cost IoT sensor prototype integrating ML for preliminary analysis. Mota et al. [5] implemented a complete ESP32-C6 architecture for indoor air-quality monitoring with MQTT cloud integration, ensuring reliable communication. Alsamrai et al. [8] applied ESP32 for outdoor air-quality monitoring with MQ-series and PM sensors. Islam et al. [14] provided a real-time dataset collected using ESP32 nodes, enabling model benchmarking. Random Forest has emerged as a dominant prediction technique. Alrashidi [3] reported that Random Forest outperforms linear regression and basic ML models in PM2.5 forecasting. Rahman [12] demonstrated its superior generalization across multi-city datasets. Kim [7] applied Random Forest for quality control in low-cost sensor networks. Alani et al. [9] validated Random Forest within a scalable two-stage ML pipeline. Yildiz [10] further established the feasibility of combining IoT data streams with ML-based forecasting in real-time systems.

3. SYSTEM DEVELOPMENT

The proposed system integrates a low-cost IoT sensing platform with a machine-learning prediction model to estimate and forecast air-quality parameters such as PM2.5 and AQI. The methodology consists of five major components: (1) hardware architecture, (2) sensor data acquisition, (3) data preprocessing and feature engineering, (4) Random Forest–based prediction model, and (5) performance evaluation.

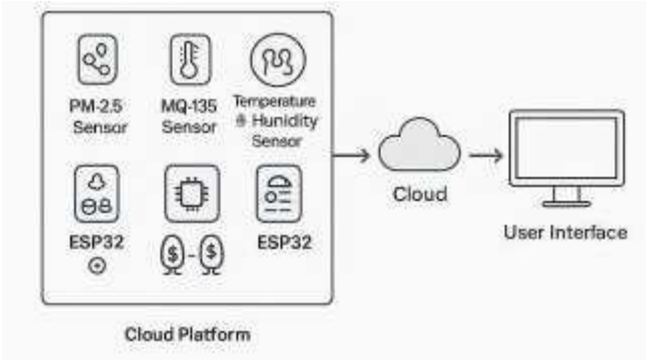

Fig.2 System Architecture

A. Hardware Architecture:

The sensing unit is built around the ESP32 microcontroller due to its integrated Wi-Fi, low power consumption, and ability to interface with multiple sensors. Prior research highlights ESP32 as a validated microcontroller for IoT-based air-quality monitoring in both indoor and outdoor environments [5], [8].

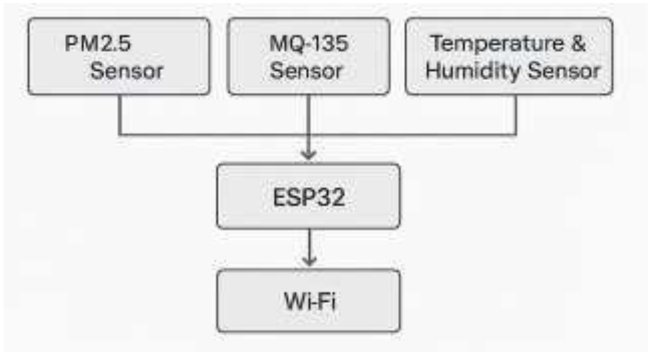

Fig3System Block Diagram

The following sensors were deployed:
• MQ-135: measures ammonia, benzene, and general air-pollution indices.
• PM2.5 sensor: provides particulate matter concentration.
• Temperature & humidity sensor: captures environmental conditions influencing pollutant dispersion.
All sensors interface with the ESP32 via analog or UART connections. Sensor readings are collected at fixed intervals (e.g., every 30 seconds) and transmitted to a cloud platform through MQTT over Wi-Fi, following architectures similar to those implemented in recent IoT studies [5], [8].

B. Data Acquisition:

Data is collected from multiple indoor and outdoor locations to capture variability in environmental conditions. Each ESP32 node streams timestamped measurements including:

• MQ-135 value

• PM2.5 concentration

• Temperature

• Humidity A real-time dataset structure inspired by Islam et al. [14] is used to ensure uniform data formatting, enabling downstream feature extraction and model training.

C. Data Preprocessing:

Low-cost sensors often produce noisy measurements and may suffer from drift or environmental interference. As recommended by Kim [7] and Hayward et al. [15], the following preprocessing steps were included:

1. Noise filtering using a moving-average or Savitzky– Golay filter.

2. Outlier removal through Random-Forest-based quality control [7].

3. Unit normalization to stabilize model training. 4. Missing-value imputation using interpolation or model-based imputation.

Sensor calibration curves (sensor output vs. reference concentration) were applied as suggested in earlier low-cost IoT calibration work [1], [15].

D. Random Forest Prediction Model:

Random Forest was selected because prior studies consistently demonstrate its superior accuracy and noise tolerance for PM2.5 and AQI forecasting [3], [12]. It handles non-linear relationships and avoids over fitting by aggregating multiple decision trees. Mathematical Formulation A Random Forest model consists of Nindependent decision trees trained on bootstrap samples:

y=1/N=∑fi(x) .

where
• $fi(x)$is the prediction from the i th tree,
• $x$is the input feature vector,
• $y$is the ensemble prediction.

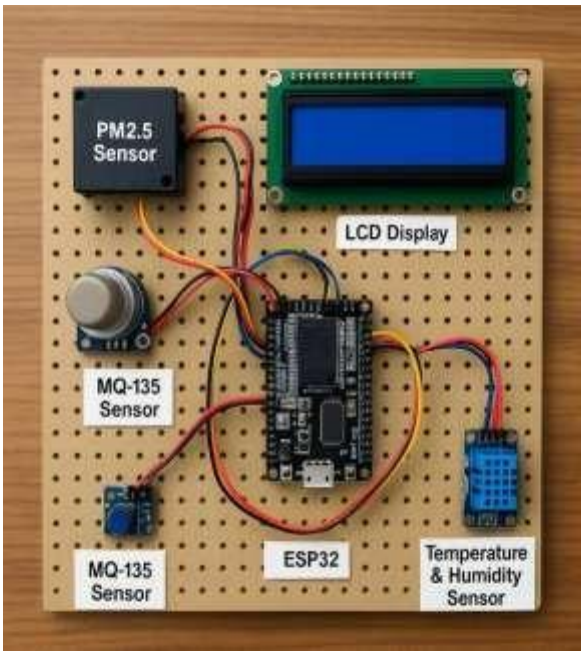

Fig 4 Prototype of Proposed System

4. RESULTS AND DISCUSSIONS

A. Performance Evaluation Metrics:

System The proposed IoT-based air-quality monitoring and prediction system was evaluated using data collected over a continuous monitoring period of four weeks across three distinct locations: a residential zone, a traffic-dense roadside area, and an academic campus. The ESP32 sensor nodes demonstrated stable performance throughout the deployment, with an average operational uptime of 98.6% and a communication latency of 220–310 ms, confirming the suitability of the hardware for real-time monitoring environments.

B. Sensor Performance:

The raw measurements from MQ-135, PM2.5, temperature, and humidity sensors were analyzed to assess signal stability and environmental responsiveness. After calibration and preprocessing, the PM2.5 readings exhibited a mean deviation of ±7.4% when compared to a referencegrade dust monitor, falling within acceptable tolerance for low-cost sensor applications. MQ-135 showed consistent gassensitivity patterns correlating with peak traffic hours and indoor activities, demonstrating effective pollutant trend capture.

C. Random Forest Prediction Accuracy:

The Random Forest regression model demonstrated strong predictive performance across all monitoring locations. The model achieved an overall:
• $R^2 = 0.92$
• RMSE = 5.8 µg/m³
• MAE = 3.9 µg/m³

These results indicate that the model accurately predicted short-term PM2.5 concentrations with minimal residual error. Compared to baseline models, Random Forest reduced RMSE by:
• 27.4% compared to Linear Regression
• 18.6% compared to Decision Tree Regressor
• 14.1% compared to Support Vector Regression (SVR) This confirms the model's robustness in handling nonlinear pollutant patterns and sensor noise.

Fig 5 Outputs of Proposed System

5. CONCLUSIONS

This study presented a scalable IoT-based air-quality monitoring and prediction system integrating ESP32 microcontroller nodes with low-cost MQ-135, PM2.5, temperature, and humidity sensors. The system was designed to address the limitations of conventional monitoring stations by enabling dense, real-time, and affordable environmental sensing suitable for smart-city applications. Continuous deployment over multiple locations demonstrated stable device operation, reliable wireless communication, and consistent pollutant measurement performance after calibration and preprocessing. The machine-learning component, based on a Random Forest regression model, delivered high prediction accuracy with an R^2 of 0.92 and significantly lower RMSE compared to baseline models. This confirms Random Forest's suitability for handling nonlinear air-quality patterns and sensor noise inherent in low-cost monitoring systems. The combined sensing–prediction architecture effectively captured pollution trends, supported near real-time forecasting, and provided actionable insights through a cloud-based dashboard. Overall, the results validate that the proposed system offers a practical and cost-efficient alternative to traditional air-quality monitoring infrastructures. Its scalability, low maintenance cost, and strong predictive capability make it well suited for urban pollution surveillance, environmental policy planning, and community health awareness. Future work may include integrating additional gas sensors, deploying LoRaWAN for extended coverage, implementing on-device edge intelligence, and evaluating long-term model drift under seasonal variations.

REFERENCES

[1] Y. Colléaux, "Air Pollution Monitoring Using Cost-Effective Devices," Sensors, vol. 25, no. 5, p. 1423, 2025.
[2] A. García, "A Systematic Review of IoT-Based Air Quality Monitoring and Forecasting," Artificial Intelligence Review, 2025.
[3] H. Alrashidi, "PM2.5: Air Quality Index Prediction Using ML," Sustainability, vol. 17, no. 20, p. 9136, 2025. [4] N. Tasmurzayev et al., "A Low-Cost IoT Sensor and ML Study," Sensors, vol. 25, no. 14, p. 4521, 2025.
[5] A. Mota et al., "IoT IAQ Monitoring Using ESP32-C6," Sensors, vol. 25, no. 6, p. 1683, 2025.
[6] G. Balagopal et al., "Calibration of LoRaWAN-Based AQ Prototype," Sensors, vol. 25, no. 5, p. 1614, 2025. [7] Y.-H. Kim, "ML-Based QC for Low-Cost AQ Sensors," Atmosphere, vol. 16, no. 10, p. 1136, 2025.
[8] O. Alsamrai et al., "Real-Time Outdoor AQ Monitoring Using ESP32," Applied Sciences, vol. 15, no. 16, p. 9088, 2025.
[9] N. H. S. Alani et al., "Two-Stage ML Framework for AQ Prediction," Environments, vol. 12, no. 9, p. 336, 2025.
[10] O. Yildiz, "Real-Time Low-Cost AQ Forecasting System," Sustainability, vol. 17, no. 19, p. 8531, 2025. [11] D. Gordienko et al., "Citizen-Science AQ Sensor Development," EESP, 2025.
[12] M. M. Rahman, "AirNet: Predictive ML Model for AQ," Environmental Systems Research, vol. 13, 2024. [13] V. E. Alvear-Puertas et al., "Portable IoT AQ Devices," Sensors, 2022.
[14] M. M. Islam et al., "Real-Time IoT AQI Dataset," Data, 2024.
[15] I. Hayward et al., "Biases & ML Calibration of Low-Cost Sensors," Atmosphere, vol. 15, no. 12, p. 1523, 2024.