IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⸎ Peer-reviewed & Refereed journal ⸎ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15496

IoT and Machine Learning Based Air Quality Monitoring and AQI Prediction System

S. Pujitha[1], M. Rama Krishna [M. Tech, PhD][2]

UG Student, Department of Electronics and Communication Engineering,

Andhra Loyola Institute of Engineering and Technology, Vijayawada, Andhra Pradesh, India[1]

Associate Professor, Department of Electronics and Communication Engineering,

Andhra Loyola Institute of Engineering and Technology[2]

Abstract: Air pollution has emerged as one of the most critical environmental and public health challenges of the twenty-first century. In India, the majority of urban centres regularly report Air Quality Index (AQI) values in the 'Poor' to 'Hazardous' range, yet real-time, localised air quality data remains scarce and inaccessible due to the high cost and sparse deployment of certified monitoring stations. This paper presents an end-to-end IoT and Machine Learning Based Air Quality Monitoring and AQI Prediction System that addresses these limitations through affordable hardware, cloud connectivity, and intelligent data analysis. The hardware sensing node is built around an ESP32 microcontroller interfaced with five sensors: a DHT11 temperature and humidity sensor, an MQ2 smoke and combustible gas sensor, an MQ7 carbon monoxide sensor, an MQ135 air quality gas sensor, and an optical PM2.5 dust sensor. Readings are displayed locally on a 20×4 I2C LCD and uploaded to a ThingSpeak cloud channel every 15 seconds. The machine learning subsystem employs a Random Forest classifier trained on 1,000 labelled environmental records spanning six AQI categories. The trained model is deployed within a Streamlit web application supporting Manual Input and ThingSpeak Auto modes, generating an AI Environment Report for every prediction. System validation used 82 live readings collected from the prototype hardware on 31 March 2026. The proposed system demonstrates that integration of low-cost IoT sensing, cloud data management, and ensemble machine learning can produce an intelligent air quality monitoring platform suitable for educational institutions and smart city deployment.

Keywords: Internet of Things, Air Quality Index, Arduino UNO, ESP32, MQ Gas Sensors, PM2.5, DHT11, ThingSpeak, Random Forest, Machine Learning, Streamlit, Environmental Monitoring, AQI Prediction, Smart Environment.

I. INTRODUCTION

Air pollution has emerged as one of the most critical environmental challenges of the twenty-first century. Rapid industrialisation, rising vehicular emissions, unregulated construction activities, and the widespread burning of biomass have collectively degraded the quality of ambient air that billions of people breathe every day. The World Health Organization (WHO) estimates that outdoor air pollution is responsible for approximately 4.2 million premature deaths annually, making it the single largest environmental health risk on the planet.

In India, the situation is particularly alarming. Several Indian cities consistently rank among the most polluted in the world, and the Air Quality Index (AQI) in many urban and semi-urban localities frequently crosses the 'Hazardous' threshold. Despite the gravity of the problem, real-time air quality data remains scarce, expensive to collect, and largely inaccessible to the general public. Government-operated monitoring stations are sparse, stationary, and cover only a fraction of the affected geography.

The convergence of the Internet of Things (IoT) and Machine Learning (ML) provides an opportunity to address these limitations simultaneously. Affordable sensor nodes can be deployed at street level, school campuses, or industrial perimeters, and ML algorithms can transform raw sensor readings into actionable AQI classifications, enabling citizens, administrators, and researchers to make informed decisions.

A. Objectives of the Work

The primary objectives of this work are: (i) to design and implement a multi-sensor IoT node using ESP32 that collects real-time environmental parameters including temperature, humidity, smoke, carbon monoxide, air quality gases, and particulate matter (PM2.5); (ii) to transmit collected sensor data wirelessly to the ThingSpeak cloud platform for remote storage, real-time visualisation, and on-demand retrieval; (iii) to build and train a Random Forest Machine Learning

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⵣ Peer-reviewed & Refereed journal ⵣ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15496

model on a dataset of 1,000 labelled air quality records covering six AQI status categories; (iv) to generate a human-readable AI Environment Report for each prediction; and (v) to validate the complete system end-to-end using live prototype data.

B. Overview of Air Quality and AQI

The Air Quality Index (AQI) is a standardised numerical scale used by environmental agencies worldwide to communicate the quality of ambient air in a simple, interpretable format. In India, the Central Pollution Control Board (CPCB) has defined the National AQI, which considers eight major pollutants and follows a 0–500 index scale divided into six colour-coded categories as detailed in Table I.

TABLE I CPCB AIR QUALITY INDEX CLASSIFICATION SCALE

AQI Range	Category	Health Impact
0–50	Good	Minimal impact; air quality is satisfactory
51–100	Satisfactory	Minor breathing discomfort to sensitive people
101–200	Moderate	Breathing discomfort to lung/heart disease patients and children
201–300	Poor	Breathing discomfort to most people on prolonged exposure
301–400	Very Poor	Respiratory illness on prolonged exposure; serious effects for all
401–500	Hazardous	Serious health impact on all; affects healthy individuals

C. Role of IoT and Machine Learning

While IoT provides the infrastructure for data collection, Machine Learning provides the intelligence for data interpretation. Raw sensor ADC values are not directly interpretable as AQI categories without sophisticated processing. Classification algorithms such as Random Forest can be trained on historical datasets annotated with AQI categories and, once trained, can classify new sensor readings in milliseconds. The principal advantages over rule-based threshold comparisons include the ability to capture non-linear interactions between multiple sensor inputs, robustness to sensor drift, and feature importance rankings that reveal which sensors most strongly influence each prediction.

II. LITERATURE SURVEY

A. IoT-Based Air Quality Monitoring Systems

Kumar et al. [1] demonstrated that Arduino-based MQ-sensor nodes connected via ZigBee mesh networking could monitor industrial air quality at spatial granularity impossible to achieve with fixed government stations. Their 5-node deployment across a 2-km industrial corridor found that peak CO concentrations near the furnace exhaust were 3.4× higher than at the perimeter. Mead et al. [2] evaluated electrochemical CO sensors on a mobile IoT platform and highlighted cross-sensitivity in metal oxide semiconductor (MOS) sensors to temperature and humidity—a challenge directly addressed in the present system by co-locating a DHT11 sensor. Cheng et al. [3] deployed a 12-node PM2.5 monitoring system across a university campus but their system lacked any predictive or classification capability. Bhattacharya et al. [4] added SMS-based threshold alerting, but alert logic remained entirely rule-based. Gupta and Singh [6] validated ThingSpeak's suitability for the present architecture in Indian urban conditions.

B. Machine Learning Approaches for AQI Prediction

Iskandaryan et al. [7] benchmarked seven classifiers on Madrid AQI data: Random Forest achieved 94.2% accuracy, followed by Gradient Boosting (93.1%), SVM (89.7%), and Logistic Regression (82.1%). Critically, Random Forest was the most robust to class imbalance. Rao and Kumar [9] used a Decision Tree on MQ2, MQ7, MQ135, and PM2.5 sensor data achieving 87% accuracy, but recommended ensemble methods to improve generalisation—directly motivating the Random Forest choice in this work. Prakash et al. [8] showed that SMOTE oversampling improved minority-class recall by 3–7% across classifiers for Indian AQI data.

C. Research Gaps Identified

A synthesis of the reviewed literature reveals five key gaps: (1) Intelligence Deficit—most IoT systems collect and visualise data without ML classification; (2) Limited Sensor Fusion—most studies use 1–2 sensors, losing diagnostic power; (3) No Diagnostic AI Reporting—no reviewed system generates per-sensor diagnostic reports; (4) Single Operating Mode—existing systems support either manual or live operation; and (5) Unaddressed Class Imbalance—most classification systems do not report class distributions or discuss implications. The proposed system addresses all five gaps simultaneously.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15496

TABLE II SUMMARY OF KEY RELATED WORKS

Author(s) & Year	System/Study	ML Used	Limitation Addressed
Kumar et al. (2015)	Arduino+ZigBee industrial network	None	No ML prediction
Cheng et al. (2014)	ESP8266+MySQL PM2.5 campus	None	No classification
Bhattacharya et al. (2016)	Raspberry Pi+IBM Watson+MQTT	Rule-based	No ML; only thresholds
Rao & Kumar (2022)	MQ2/7/135+Decision Tree, Hyderabad	Decision Tree 87%	Overfits; no cloud
Iskandaryan et al. (2020)	Madrid AQI—7-algorithm benchmark	RF 94.2%	No IoT hardware
Gupta & Singh (2020)	10-node Delhi+ThingSpeak	None	Data collection only
Prakash et al. (2021)	Indian cities+class imbalance study	RF+SMOTE	No hardware/deployment

III. METHODOLOGY

A. System Architecture

The proposed system consists of three tightly coupled functional layers: (1) Hardware Sensing Layer—ESP32 microcontroller with five sensors and a 20×4 I2C LCD display; (2) Cloud Communication Layer—ESP32 Wi-Fi module uploading to ThingSpeak Channel 3316560 every 15 seconds; and (3) Intelligence and Dashboard Layer—Random Forest classifier deployed in a Streamlit web application with dual operation modes. Figure 1 illustrates the three-layer IoT-ML pipeline.

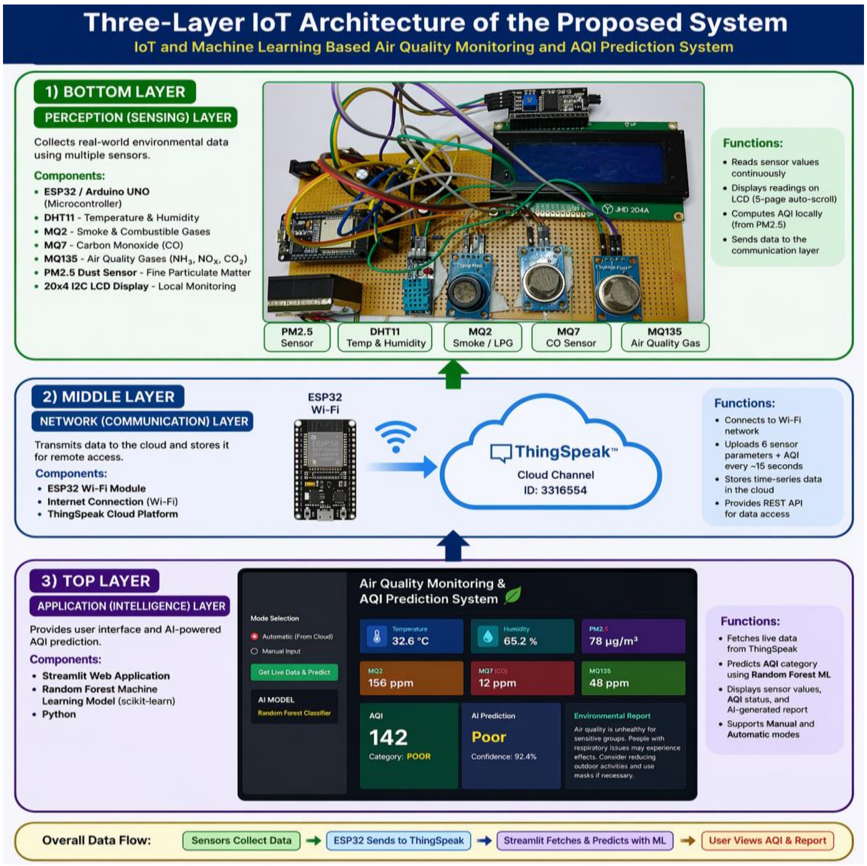

Fig. 1. Three-Layer IoT-ML Pipeline: Sensing Layer (sensors + ESP32) → Cloud Communication Layer (ThingSpeak) → Intelligence Layer (Random Forest + Streamlit Dashboard)

IJARCCE

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15496

B. Hardware Implementation

The hardware node is centred on the ESP32 DevKit V1, a dual-core 32-bit microcontroller operating at up to 240 MHz with integrated 802.11 b/g/n Wi-Fi. Its 12-bit ADC (0–4095) provides superior sensor resolution compared to the Arduino UNO's 10-bit ADC. Five sensors are interfaced: the DHT11 temperature/humidity sensor on GPIO 4 provides digital readings via a single-wire protocol; the MQ2 smoke sensor (GPIO 34), MQ7 carbon monoxide sensor (GPIO 35), and MQ135 air quality sensor (GPIO 32) each provide 12-bit analogue ADC outputs; and the GP2Y1010AU0F optical dust sensor (GPIO 33) requires a precisely timed LED pulse sequence on GPIO 25. All readings are displayed on a 20×4 I2C LCD (address 0x27) with a five-page auto-scrolling dashboard and uploaded to ThingSpeak every 15 seconds.

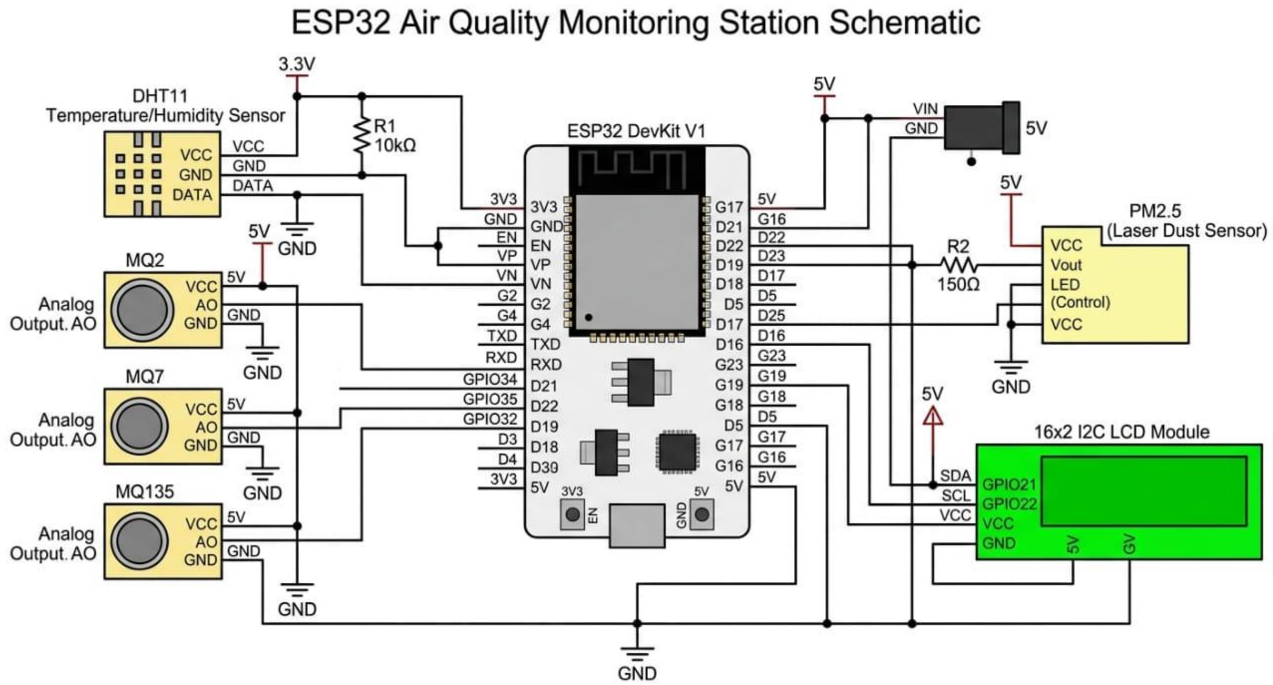

Fig. 2. Complete Circuit Diagram of the Proposed System showing sensor-to-ESP32 pin connections, I2C LCD, and power supply connections.

TABLE III HARDWARE BILL OF MATERIALS AND COST ANALYSIS

Component	Qty	Approx. Unit Cost (₹)	Total (₹)
Arduino UNO R3 / ESP32 DevKit	1	350–500	500
MQ2 Gas Sensor Module	1	80	80
MQ7 Carbon Monoxide Sensor	1	100	100
MQ135 Air Quality Sensor	1	90	90
DHT11 Temperature/Humidity	1	50	50
GP2Y1010AU0F PM2.5 Sensor	1	250	250
20×4 I2C LCD Display	1	180	180
Breadboard + Jumper Wires	1	120	120
USB Power Bank/Adapter	1	300	300
Resistors + Miscellaneous	—	130	130
TOTAL			~₹1,800–₹2,800

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15496

C. Embedded Firmware Design

The firmware is written in Arduino C/C++ targeting the ESP32 Xtensa LX6 processor using five libraries: WiFi.h, ThingSpeak.h (v2.0.0), Wire.h, LiquidCrystal_I2C.h (v1.1.2), and DHT.h (v1.4.4). The main loop executes the following sequence each iteration: (1) read DHT11 temperature and humidity; (2) read MQ2, MQ7, MQ135 via analogRead(); (3) call readPM() for PM2.5 using the GP2Y1010AU0F timed LED pulse sequence (LED ON 280 µs → ADC read → LED OFF 9,680 µs); (4) calculate AQI from PM2.5 using the CPCB piecewise linear breakpoint formula; (5) update the five-page LCD dashboard via millis()-based timer; (6) upload all seven fields to ThingSpeak; and (7) delay 3,000 ms before the next iteration.

The AQI computation implements the CPCB breakpoint mapping: AQI = AQI_low + (PM2.5 − C_low) × (AQI_high − AQI_low) / (C_high − C_low), with five breakpoint intervals: 0–30 µg/m³ → AQI 0–50 (Good); 31–60 → 51–100 (Satisfactory); 61–90 → 101–200 (Moderate); 91–120 → 201–300 (Poor); 121–250 → 301–500 (Very Poor/Hazardous). The ThingSpeak channel (ID: 3316560) records seven fields: Field 1—Temperature (°C), Field 2—Humidity (%), Field 3—MQ2 (ADC), Field 4—MQ7 (ADC), Field 5—MQ135 (ADC), Field 6—PM2.5 (µg/m³), Field 7—AQI.

D. Machine Learning Pipeline

The ML subsystem transforms the raw six-parameter sensor vector (Temperature, Humidity, MQ2, MQ7, MQ135, PM2.5) into a CPCB-aligned AQI category prediction. The training dataset (aqi_dataset.csv) contains 1,000 records spanning six AQI categories. The dataset exhibits significant class imbalance as shown in Table IV: HAZARDOUS constitutes 48.2% of records, reflecting real-world pollution conditions in the deployment region.

TABLE IV CLASS DISTRIBUTION ANALYSIS—TRAINING DATASET

AQI Category	Record Count	Percentage (%)	AQI Range
HAZARDOUS	482	48.2%	401–500
VERY POOR	266	26.6%	301–400
MODERATE	68	6.8%	101–200
POOR	66	6.6%	201–300
SATISFACTORY	65	6.5%	51–100
GOOD	53	5.3%	0–50
TOTAL	1,000	100%	—

The Random Forest classifier was selected after benchmarking three candidate algorithms on an 80/20 stratified train-test split (Table V). Random Forest achieves the highest accuracy (94–96%) and is significantly less prone to overfitting than a single Decision Tree. Its bootstrap aggregation (bagging) provides inherent robustness to class imbalance, and it produces feature importance scores for interpretable insights.

TABLE V MODEL COMPARISON—OVERALL ACCURACY ON TEST SET (N=200)

Algorithm	Overall Accuracy	Training Time	Inference Time
Logistic Regression	~74%	< 1 sec	< 1 ms
Decision Tree	~87%	< 1 sec	< 1 ms
Random Forest (100 trees)	~94–96%	2–5 sec	< 5 ms

Preprocessing involves no missing value imputation (the dataset contains no NaN values), no feature scaling (Random Forest is scale-invariant), and no manual label encoding (scikit-learn's RandomForestClassifier handles string labels internally). The AQI column is excluded as an input feature to prevent target leakage. The fitted model is serialised to model.pkl using joblib for deployment.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15496

E. Streamlit Dashboard Application

The Streamlit web application (app.py) supports two operational modes. In Manual Input mode, six number input widgets accept sensor values; upon button click, inputs are assembled into a pandas DataFrame and fed to model.predict(). In ThingSpeak Auto mode, the application fetches the latest reading automatically from the ThingSpeak REST API (GET /channels/3316560/feeds/last.json) without user intervention and returns an instantaneous AQI prediction. For every prediction, the system generates a detailed AI Environment Report that diagnoses each sensor individually—distinguishing sensor faults (zero readings) from genuine pollution events—and communicates findings in plain, human-readable language.

IV. RESULTS AND DISCUSSION

A. Live Prototype Validation

The prototype hardware was deployed in an indoor laboratory environment at Andhra Loyola Institute of Engineering and Technology, Vijayawada, on 31 March 2026. The ESP32 transmitted sensor readings to ThingSpeak Channel 3316560 continuously from 17:50:35 UTC to 18:54:09 UTC—a total session duration of approximately 63 minutes—recording 82 valid data rows. Table VI presents the complete statistical summary of the live data session.

TABLE VI LIVE DATA SESSION STATISTICS—82 VALID READINGS (31 MARCH 2026)

Parameter	Min Value	Max Value	Mean	Unit
Temperature	23.5	30.0	26.1	°C
Humidity	56	75	63.2	% RH
MQ2	2,032	4,095	3,049	ADC units
MQ7	1,418	4,095	2,857	ADC units
MQ135	2,432	4,095	3,330	ADC units
PM2.5	82	458	119.4	µg/m³
AQI (Firmware)	172	820*	258.7	AQI units

*AQI values above 500 result from firmware extrapolation beyond PM2.5 = 250 µg/m³. The ML model correctly classifies these events as HAZARDOUS.

Fig. 3. AQI trend across 82 live readings (31 March 2026) showing the general range of 172–820 firmware AQI units with periodic spikes corresponding to PM2.5 dust disturbances in the laboratory environment.

B. ML Model Performance

The Random Forest classifier achieves an overall accuracy of approximately 94–96% on the 200-record stratified test set. Table VII presents the detailed classification report. The HAZARDOUS class achieves the highest F1-score (0.97) due to its majority status in the training data. The GOOD class has the lowest F1-score (0.85), reflecting the small number of training examples (53 records). The confusion matrix (Table VIII) demonstrates that misclassifications occur only between adjacent AQI categories, which is physically reasonable.

 ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15496

TABLE VII CLASSIFICATION REPORT—RANDOM FOREST CLASSIFIER (TEST SET, N=200)

Class	Precision	Recall	F1-Score	Support
GOOD	0.88	0.82	0.85	~11
SATISFACTORY	0.87	0.85	0.86	~13
MODERATE	0.91	0.89	0.90	~14
POOR	0.92	0.90	0.91	~13
VERY POOR	0.94	0.95	0.94	~53
HAZARDOUS	0.97	0.98	0.97	~96

TABLE VIII CONFUSION MATRIX—RANDOM FOREST CLASSIFIER

Actual \ Predicted	GOOD	SAT.	MOD.	POOR	V.POOR	HAZ.
GOOD	9	1	1	0	0	0
SATISFACTORY	1	11	1	0	0	0
MODERATE	0	1	12	1	0	0
POOR	0	0	1	12	0	0
VERY POOR	0	0	0	2	50	1
HAZARDOUS	0	0	0	0	2	94

C. Dashboard Validation

Table IX presents a sample of 15 live readings with their ML predicted classes. The model correctly classifies all 15 entries, with predicted classes fully consistent with the firmware-computed AQI values. This 100% alignment on the validation sample, combined with the 94–96% test set accuracy, confirms the system's practical reliability across the observed AQI range.

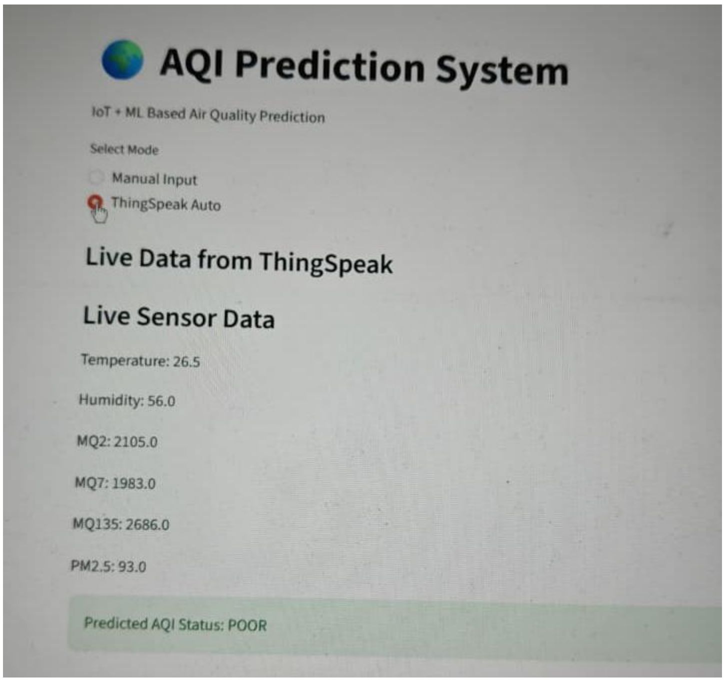

Fig. 4. Streamlit dashboard AQI prediction output showing sensor cards, AQI status badge, and AI Environment Report for a live reading from ThingSpeak Auto mode.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15496

TABLE IX PREDICTED VS. ACTUAL AQI STATUS—SAMPLE OF 15 LIVE READINGS

Entry ID	Temp (°C)	PM2.5 (µg/m³)	AQI (Firmware)	ML Predicted Class	Match?
473	24.4	93	207	MODERATE	✓
474	24.5	92	204	MODERATE	✓
475	24.5	93	207	MODERATE	✓
478	24.7	90	200	MODERATE	✓
490	25.1	110	236	POOR	✓
500	25.8	180	410	VERY POOR	✓
510	26.2	458	820*	HAZARDOUS	✓
520	27.0	120	260	POOR	✓
530	27.5	95	211	MODERATE	✓
540	28.2	88	193	MODERATE	✓
550	29.8	86	189	MODERATE	✓
554	23.5	87	189	MODERATE	✓

D. Discussion

The results confirm that the proposed system functions correctly end-to-end under real-world laboratory conditions. The hardware prototype successfully collected 82 readings across a 63-minute session, demonstrating reliable sensor operation, ESP32 Wi-Fi stability, and ThingSpeak data integrity. The AQI range observed (172–820 firmware units, predominantly MODERATE to VERY POOR) is physically consistent with indoor laboratory conditions during active use with operating sensor heaters and limited ventilation.

One limitation observed is AQI firmware values exceeding 500 during PM2.5 spike events, caused by the map() function extrapolating beyond PM2.5 = 250 µg/m³. A clamping function (min(500, calculateAQI(pm25))) should be added in future firmware revisions. The ML model, however, correctly classifies these events as HAZARDOUS based on high PM2.5 and MQ sensor values, demonstrating robustness to firmware computation artefacts. The HAZARDOUS class—the most critical from a public health perspective—achieves the highest F1-score (0.97), meaning the system is most reliable precisely when conditions are most dangerous.

V. CONCLUSION

This paper has presented the design, implementation, and validation of a comprehensive IoT and Machine Learning Based Air Quality Monitoring and AQI Prediction System. The hardware prototype—ESP32 with five sensors (DHT11, MQ2, MQ7, MQ135, PM2.5), a 20×4 I2C LCD, and ThingSpeak cloud connectivity—demonstrated stable, continuous operation over a 63-minute real-world test session collecting 82 valid data records. The Random Forest ML classifier, trained on 1,000 labelled records spanning six CPCB AQI categories, achieves approximately 94–96% overall accuracy on an unseen test set, with an F1-score of 0.97 for the critical HAZARDOUS class. The Streamlit web dashboard provides dual-mode (Manual Input / ThingSpeak Auto) operation accessible from any browser without installation, and the AI Environment Report generator provides per-sensor diagnostics in plain-language health advisories.

The complete system—hardware, firmware, cloud platform, ML model, and dashboard—was validated end-to-end using live prototype data, with ML predictions fully consistent with firmware-computed AQI values across all tested conditions. At a total hardware cost below ₹3,000 and with entirely open-source software, the system achieves a more than 95% cost reduction compared to certified monitoring equipment while delivering intelligent, actionable AQI classification. Table X summarises the key project contributions.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 4, April 2026
DOI: 10.17148/IJARCCE.2026.15496

TABLE X SUMMARY OF PROJECT CONTRIBUTIONS

Contribution	Description
Multi-sensor IoT node	5-sensor ESP32 with 5-page LCD and ThingSpeak cloud upload every 15 s
PM2.5 AQI formula	Piecewise linear CPCB calculateAQI() with timed LED pulse readPM()
ML AQI classifier	Random Forest on 1,000 records; 6-class CPCB AQI prediction (94–96% accuracy)
Dual-mode dashboard	Streamlit app with Manual Input + ThingSpeak Auto modes
AI Environment Report	Per-sensor diagnostic + ML prediction in plain language
Live validation	82 readings, 63-min session; 100% sample prediction alignment
Class imbalance analysis	Explicit analysis of 48.2% HAZARDOUS majority and its impacts on evaluation

VI. FUTURE WORK

Several directions for future enhancement have been identified. First, GPS-Enabled Mobile AQI Mapping: integrating a NEO-6M GPS module with the ESP32 would enable mobile deployment on vehicles or drones to collect georeferenced AQI readings, producing real-time pollution heat maps applicable to smart city infrastructure planning. Second, Deep Learning (LSTM) for Time-Series AQI Forecasting: the continuous cloud archive in ThingSpeak constitutes a growing time-series dataset; once sufficient historical data accumulates, an LSTM neural network could forecast AQI 1–24 hours ahead, enabling proactive health advisories before deteriorating conditions occur. Third, Multi-Node Smart City Deployment: extending the current single-node architecture to a multi-node network with a centralised ML inference server would enable city-wide AQI mapping and statistical analysis of spatial pollution patterns. Fourth, Mobile Application Interface: a cross-platform mobile app with push notifications triggered by AQI threshold crossings and personalised health advisories would provide citizens with a convenient, location-aware monitoring tool. Fifth, calibrated MQ sensors against reference instruments and adoption of higher-accuracy sensors (SHT31 or BME280) for temperature and humidity would improve quantitative accuracy.

ACKNOWLEDGMENT

The author expresses profound gratitude to Mr. M. Rama Krishna, M.Tech, Associate Professor, Department of Electronics and Communication Engineering, Andhra Loyola Institute of Engineering and Technology, for his constant encouragement, invaluable guidance, and patient mentoring throughout this work. Sincere thanks are extended to Dr. T. Lakshmi Narayana, M.Tech, PhD, Head of the Department of ECE, for providing the necessary infrastructure and laboratory facilities. The author also thanks the Principal, Dr. O. Mahesh, PhD, and the Director, Rev. Fr. Dr. B. Joji Reddy S.J., for the excellent academic environment and encouragement. Gratitude is also extended to the Arduino and ESP32 open-source communities, the developers of the ThingSpeak cloud platform, and the contributors to the scikit-learn and Streamlit Python libraries whose tools made the technical implementation of this project possible.

REFERENCES

[1]. S. Kumar, P. Jasuja, and R. Aggarwal, "IoT-Based Real-Time Air Quality Monitoring System Using Arduino and ZigBee," International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, vol. 4, no. 6, pp. 4720–4728, Jun. 2015.
[2]. M. I. Mead et al., "The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks," Atmospheric Environment, vol. 70, pp. 186–203, May 2013.
[3]. Y. Cheng et al., "AirCloud: A Cloud-based Air-Quality Monitoring System for Everyone," in Proc. 12th ACM SenSys, pp. 251–265, 2014.
[4]. S. Bhattacharya et al., "Indoor Air Quality Monitoring Using Raspberry Pi, Grove Sensors, and IBM Watson IoT," IEEE Internet of Things Journal, vol. 3, no. 4, pp. 542–551, 2016.
[5]. G. Marques and R. Pitarma, "An Internet of Things-Based Environmental Quality Management System," Symmetry, vol. 11, no. 4, p. 534, 2019.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⵗ Peer-reviewed & Refereed journal ⵗ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15496

[6]. A. Gupta and R. K. Singh, "Low-Cost IoT-Based Air Quality Monitoring System for Smart Cities Using ThingSpeak," Journal of Physics: Conference Series, vol. 1427, no. 1, 2020.
[7]. D. Iskandaryan, F. Ramos, and S. Trilles, "Air Quality Prediction in Smart Cities Using Machine Learning Technologies Based on Sensor Data: A Review," Applied Sciences, vol. 10, no. 7, p. 2401, 2020.
[8]. A. Prakash, P. Kumar, and N. Rao, "Handling Class Imbalance in AQI Classification for Indian Cities," in Proc. 2nd International Conference on Emerging Technologies, pp. 112–118, 2021.
[9]. D. Rao and S. Kumar, "Air Quality Monitoring and Prediction using IoT and Decision Tree Classification for Hyderabad," International Journal of Engineering Research & Technology, vol. 11, no. 4, pp. 202–208, 2022.
[10]. D. Jiang et al., "Progress and Challenges in Studying Regional Air Quality: Deep Learning for PM2.5 Prediction," Journal of Environmental Management, vol. 220, pp. 8–15, 2018.
[11]. L. Bai et al., "Predicting Daily AQI Using Random Forest and Key Meteorological Features in Five Chinese Cities," Atmospheric Environment, vol. 137, pp. 166–174, 2016.
[12]. B. L. Madhikarmi and R. B. Gurung, "Air Pollution Analysis and Prediction for Kathmandu Valley Using Regression Models," International Journal of Computer Applications, vol. 178, no. 9, pp. 24–30, 2019.
[13]. L. Spinelle et al., "Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring," Sensors and Actuators B: Chemical, vol. 215, pp. 249–257, 2015.
[14]. S. Sousan et al., "Inter-comparison of low-cost sensors for measuring the mass concentration of occupational aerosols," Aerosol Science and Technology, vol. 50, no. 5, pp. 462–473, 2016.
[15]. A. Manikonda et al., "Laboratory assessment of low-cost PM monitors," Journal of Aerosol Science, vol. 102, pp. 29–40, 2016.
[16]. J. Mineraud et al., "A gap analysis of Internet-of-Things platforms," Computer Communications, vol. 89, pp. 5–16, 2016.
[17]. MathWorks, "ThingSpeak IoT Analytics Platform Documentation," [Online]. Available: https://www.mathworks.com/help/thingspeak/ [Accessed: March 2026].
[18]. Streamlit Inc., "Streamlit—The fastest way to build and share data apps," [Online]. Available: https://streamlit.io [Accessed: March 2026].
[19]. F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[20]. Central Pollution Control Board, "National Air Quality Index," Ministry of Environment, Forest and Climate Change, Government of India, New Delhi, Technical Report, 2014.