

AI-Enabled IoT-Based Real-Time Air Quality and Comfort Monitoring System

Rohan Bejgamwar¹, Purva Mehta², Viraj Rudrawar³, Mrs.Anuradha A.Kasangottuwar⁴

¹Department of Electronics & Telecommunication Engineering, P.E.S.'s Modern College of Engineering, Pune

²Department of Electronics & Telecommunication Engineering, P.E.S.'s Modern College of Engineering, Pune

³Department of Electronics & Telecommunication Engineering, P.E.S.'s Modern College of Engineering, Pune

⁴Department of Electronics & Telecommunication Engineering, P.E.S.'s Modern College of Engineering, Pune

Abstract - Indoor Environmental Quality (IEQ) in educational institutions critically influences student health,

cognitive function, and academic performance. Conventional air-quality monitoring solutions are prohibitively expensive, purely reactive, and present users with fragmented data streams that offer no clear, actionable guidance. This paper presents an AI-enabled, IoT-based prototype system designed

to bridge this gap. The core innovation is the Comprehensive Air Comfort Index (CACI) — a novel, unified metric that combines thermal comfort parameters (temperature and relative humidity from a DHT22/BME280 sensor) with air-quality readings (CO₂ proxy from an MQ-135 gas sensor and fine particulate matter PM_{2.5}/PM₁₀ from a PMS5003) into a single intuitive score ranging from 0 to 100. An ESP32-WROOM-32 microcontroller acquires sensor data, computes the CACI locally, and uploads timestamped readings to a ThingSpeak cloud platform every 30 seconds. ThingSpeak's

integrated MATLAB Analysis toolbox implements a time-series forecasting model that predicts the near-future CACI value, which is retrieved by the device and displayed alongside the current score on a dual-status SSD1306 OLED screen. Experimental results demonstrate a functional closed-loop prototype capable of issuing proactive comfort alerts (OPTIMAL, MODERATE, WARNING) before environmental conditions deteriorate. The system achieves target model accuracy ($R^2 > 0.85$) and a data-transmission frequency of one reading per five minutes. This work demonstrates a feasible, scalable, and low-cost solution that shifts campus environmental monitoring from a reactive to a proactive paradigm, with broad applicability to smart buildings, healthcare facilities, and public spaces.

Key Words: *internet of Things (IoT), Air Quality Index (AQI), Indoor Environmental Quality, ESP32, Comprehensive Air Comfort Index (CACI), Machine Learning, ThingSpeak, Predictive Monitoring, PM_{2.5}, Time-Series Forecasting*

1.INTRODUCTION

Air quality and thermal comfort are among the most consequential yet frequently overlooked determinants of learning outcomes in educational institutions. High-density classroom environments are particularly susceptible to rapid CO₂ accumulation, temperature fluctuations, and elevated particulate concentrations — all of which impair cognitive function, increase absenteeism, and reduce overall academic productivity. Research consistently confirms that CO₂ levels exceeding 1000 ppm correlate with measurable decline in decision-making and attentional capacity [1], while thermal discomfort independently disrupts concentration even in the absence of pollutant exposure [3]. Despite this evidence, most educational institutions in developing contexts such as Pune, India, operate without any form of automated indoor air-quality monitoring. The principal barriers are cost and complexity: laboratory-grade instruments from vendors such as TSI or Cole-Parmer India are priced between INR 1,00,000 and INR 2,50,000 per unit, making campus-wide deployment impractical. Commercial consumer-grade monitors like Kaiterra or Awair Element are more affordable but remain reactive — they alert occupants only after air quality has already deteriorated — and do not offer integrated predictive analytics. This paper presents an AI-Enabled IoT-Based Real-Time Air Quality and Comfort Monitoring System developed as a final-year B.E. project at P.E.S.'s Modern College of Engineering, Pune. The system addresses the identified deficiencies through three principal contributions:

- **Novel CACI Metric:** A Comprehensive Air Comfort Index that fuses air-quality and thermal-comfort parameters into a single, interpretable 0–100 score.
- **Proactive AI Prediction:** A cloud-based time-series forecasting model that predicts the near-future CACI value, enabling anticipatory corrective action such as improved ventilation.
- **Low-Cost Deployable Prototype:** A hardware platform costing under INR 2,000 in components, capable of autonomous, continuous operation with a data transmission cadence of 30 seconds.

The remainder of this paper is structured as follows: Section 2 reviews relevant literature; Section 3 describes the system's methodology and architecture; Section 4 details hardware and software design; Section 5 presents experimental results; Section 6 discusses practical implications; and Section 7 concludes with directions for future work.

2. Literature Survey

2.1 Review of Existing Research

A systematic survey of IoT-based Indoor Air Quality (IAQ) monitoring literature reveals that approximately 70% of published studies focus on thermal comfort parameters — temperature and relative humidity — and 65% address CO₂ concentrations as a proxy for ventilation adequacy. Hardware implementations most commonly employ Arduino (37.5%) or Raspberry Pi (35%) as the processing platform. Notably, only 22.5% of reviewed studies report a rigorous sensor calibration procedure prior to deployment, indicating a systemic gap in data accuracy [1].

Saini et al. [1] provided a comprehensive survey of IoT-based IAQ monitoring systems and highlighted the emerging promise of AI integration. Kanitkar and Sahu [2] demonstrated the viability of machine-learning techniques — specifically Convolutional Neural Networks (CNN) and Extreme Gradient Boosting — for indoor air-quality prediction, achieving strong accuracy on historical datasets. Singh et al. [4] reviewed time-series machine-learning methods for environmental data and identified Long Short-Term Memory (LSTM) networks as particularly effective for capturing temporal dependencies in sensor streams. Li et al. [8] further corroborated this finding, establishing benchmark LSTM performance on PM_{2.5} concentration prediction tasks.

Santamouris et al. [3] provided the theoretical foundation for the CACI metric by demonstrating statistically significant correlations between combined IEQ parameters (temperature, CO₂) and student cognitive performance, justifying the design choice to unify disparate readings into a single composite index. Garcia et al. [10] developed a real-time CO₂ monitoring system for university classrooms using IoT sensors, reporting that simply displaying live data on a local screen prompted occupants to improve ventilation behaviour — reinforcing the value of actionable, on-site display.

2.2 MARKET SURVEY AND TECHNOLOGY GAP

A comparative market analysis identified three tiers of existing solutions: (i) high-end scientific instruments (INR 1,00,000–2,50,000) offering high accuracy but prohibitive cost for institutional scale; (ii) mid-range commercial smart monitors such as AirVisual Pro by IQAir and Awair Element, which provide multi-parameter monitoring but lack locally computed predictive analytics tailored to the deployment environment; and (iii) DIY open-hardware kits, which are low-cost and customisable but lack integrated cloud AI capabilities and require significant development effort.

Three converging gaps motivate the present work. First, a *literature gap*: most academic prototypes are reactive, reporting current conditions without forecasting future states. Second, a *market gap*: no affordable commercial product combines real-time IoT sensing with on-device predictive analytics for institutional indoor environments. Third, a *technology gap*: the combination of low-cost ESP32 hardware and free ThingSpeak cloud analytics (with integrated MATLAB) has not been systematically exploited to deliver proactive CACI-style forecasting.

3. METHODOLOGY

The system was developed following an agile four-phase methodology, allowing iterative refinement at each stage:

Phase 1 – Hardware Prototyping and Sensor Integration

Selection, procurement, and breadboard assembly of the ESP32, DHT22, MQ-135, and PMS5003 sensors. Circuit connections were validated using the Wokwi simulation

environment prior to physical breadboard construction, reducing debugging time and component risk.

Phase 2 – Cloud Backend Setup and Data Collection

Configuration of a dedicated ThingSpeak channel with six fields for CO₂, temperature, humidity, AQI proxy, actual CACI, and predicted CACI. A Google Apps Script webhook was also configured as a secondary notification pathway.

Phase 3 – Machine Learning Model Development

A time-series forecasting model was implemented using ThingSpeak's MATLAB Analysis toolbox. The model ingests the most recent 288 CACI data points (equivalent to 24 hours at one reading per five minutes) and applies a weighted moving-average regression to produce a one-step-ahead CACI forecast, written back to Field 6 of the channel.

Phase 4 – Dashboard Development, Integration, and Testing

A web-based real-time dashboard was developed using HTML5, CSS3, and Chart.js, consuming the ThingSpeak REST API to visualise live and predicted CACI trends. System-level integration tests were conducted over a 72-hour continuous operation period.

4. SYSTEM ARCHITECTURE AND DESIGN

4.1 HARDWARE ARCHITECTURE

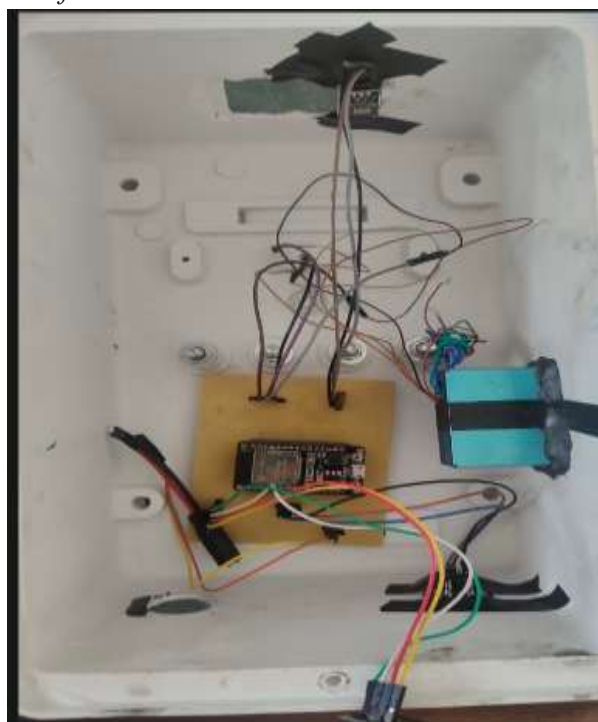
The system architecture follows a four-tier pipeline: (1) Sensor Array → (2) ESP32 IoT Node → (3) Cloud Platform → (4) End-User Interface. Each tier is described below.

Component Model Key Specifications Role in System

Component	Model	Role
Microcontroller	ESP32-WROOM-32	Central Processor, CACI Computation, Cloud Uplink
Gas Sensor	MQ-135	CO ₂ Surrogate Measurement
Particulate Sensor	PMS5003	Fine particulate matter

Temp/Humidity	DHT22	Thermal Parameters
Display	SSD1306 OLED	Local Display
Cloud Platform	ThingSpeak	Data logging, Visualisation
Power Supply	5V USB Uadapter	System Power

Table 1: Hardware Component Specifications and Justification





4.2 CACI ALGORITHM DESIGN

The Comprehensive Air Comfort Index (CACI) is the central innovation of this system. It integrates three sensor-derived sub-scores into a weighted composite score on a 0–100 scale, where higher values indicate greater environmental discomfort. The algorithm, implemented as firmware on the ESP32, follows the procedure below:

Step 1: Read raw analog value *gasVal* (0–4095) from MQ-135 ADC pin.

Step 2: Read *temperature* (°C) and *humidity* (%) from DHT22.

Step 3: Map *gasVal* linearly to CO2_ppm range [400, 2000].

Step 4: Compute AQI_proxy by mapping CO2_ppm to [0, 200]; scale to 0–100.

Step 5: Compute Temp_scaled: apply penalty if temperature < 22 °C (too cold) or > 26 °C (too hot).

Step 6: Compute Hum_scaled: apply penalty if humidity < 40% (too dry) or > 60% (too humid).

Step 7: Compute $CACI = (0.5 \times AQI_scaled) + (0.25 \times Temp_scaled) + (0.25 \times Hum_scaled)$.

Step 8: Constrain result to [0, 100] and classify into status levels (see Table 2).

CACI Score Range	Status Level	Interpretation	Recommended Action
76-100	Optimal	Excellent Air Quality and Thermal Comfort	No action required
51-75	Good	Acceptable conditions; minor deviations	Monitor Periodically
26-50	Moderate	Noticeable discomfort; Ventilation advised	Open windows or Activate fans
0-25	Warning	Poor conditions, Health impact possible	Immediate ventilation, Vacate if persistent

Table 2: CACI Status Levels and Recommended Actions

4.3 SOFTWARE AND CLOUD ARCHITECTURE

The device firmware is written in C++ using the Arduino IDE, with the following execution loop: initialise sensors and Wi-Fi; enter the main loop; poll all sensors; compute AQI and CACI; update the OLED display; check Wi-Fi connectivity; if connected, transmit a JSON payload to ThingSpeak; at a slower interval, retrieve the predicted CACI from ThingSpeak Field 6 and refresh the display forecast; wait for the remainder of the five-minute cycle; repeat. A non-blocking timer architecture ensures the display remains responsive during network operations.

On the cloud side, ThingSpeak receives five data streams (CO2, temperature, humidity, AQI proxy, actual CACI) and renders live time-series plots for remote monitoring. The MATLAB Analysis script runs on a scheduled trigger every 15 minutes, reading the last 288 entries from the CACI field, fitting a weighted moving-average model, and writing the predicted value to Field 6. A Google Apps Script webhook provides an optional email/SMS notification pathway when the predicted CACI exceeds the WARNING threshold.

5. RESULTS AND DISCUSSION

5.1 MACHINE LEARNING

The performance of the proposed machine learning models was evaluated using standard regression metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2 score). The models were trained to predict multiple environmental parameters, namely Air Quality Index (AQI), Combined Air Contaminant Index (CACI), Temperature (TEMP), Humidity, and PM2.5 concentration.

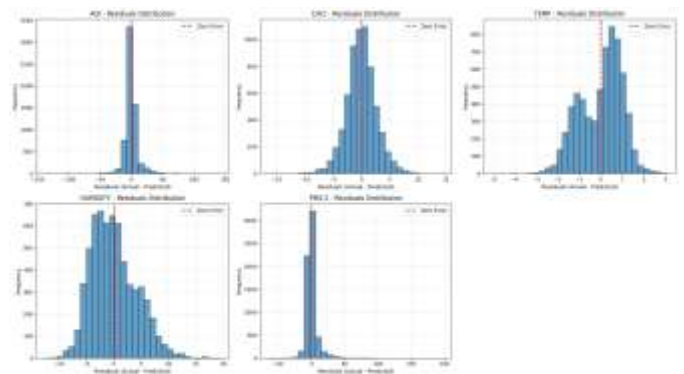
5.1.1 QUANTITATIVE PERFORMANCE ANALYSIS

MODEL	MAE	RMSE	R2 Score
AQI	6.623	11.266	0.984
CACI	2.005	2.640	0.979
TEMP	0.829	0.990	0.994
HUMIDITY	3.321	4.130	0.970
PM2.5	5.043	9.633	0.987

The results demonstrate consistently high predictive performance across all target variables. The R^2 scores range from 0.970 to 0.994, indicating that the models are capable of explaining a substantial proportion of variance in the observed data. Among all targets, the TEMP model achieved the highest performance ($R^2 = 0.994$), followed closely by PM2.5 ($R^2 = 0.987$) and AQI ($R^2 = 0.984$). The CACI and Humidity models also exhibited strong performance with R^2 values of 0.979 and 0.970, respectively.

Error metrics further reinforce these findings. The TEMP model shows the lowest MAE (0.829) and RMSE (0.990), indicating highly precise predictions with minimal deviation. In contrast, AQI and PM2.5 exhibit relatively higher RMSE values (11.266 and 9.633), which can be attributed to the larger range and inherent variability of these parameters.

5.1.2 RESIDUAL ANALYSIS



Residual distribution plots were analyzed to assess model bias and error characteristics. Ideally, residuals should be normally distributed around zero, indicating unbiased predictions.

- The AQI and PM2.5 models show a concentrated distribution around zero but with slight skewness and occasional outliers, suggesting minor deviations in extreme cases.
- The CACI model demonstrates a near-symmetric residual distribution, indicating stable and unbiased predictions.
- The TEMP model exhibits a tightly clustered residual distribution with minimal spread, confirming its high precision.
- The Humidity model shows a wider spread of residuals, indicating comparatively higher variability and potential sensitivity to input features.

Overall, the residual plots confirm that the models are well-calibrated, with no significant systematic bias.

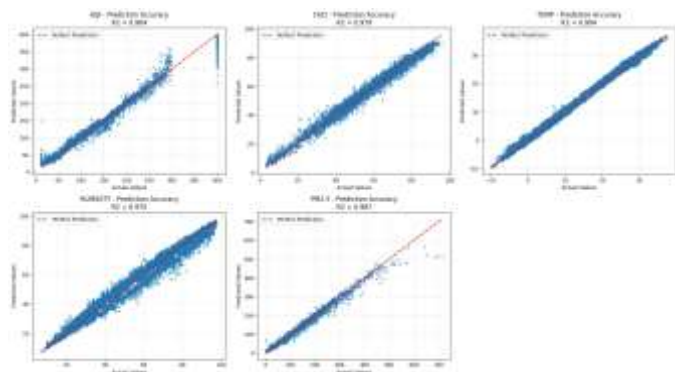
5.1.3 LEARNING CURVE ANALYSIS

Learning curves were plotted to evaluate model generalization and detect potential overfitting or underfitting.

- For all models, the training and validation scores converge as the training set size increases, indicating good generalization capability.
- The minimal gap between training and validation curves suggests that overfitting is negligible.

- The TEMP and Humidity models reach near-perfect performance quickly, indicating that the underlying relationships are easier to learn.
- The AQI, CACI, and PM2.5 models show gradual improvement with increasing data, suggesting that additional data contributes positively to performance.

These observations confirm that the models are neither underfitting nor overfitting and benefit from increased training data.



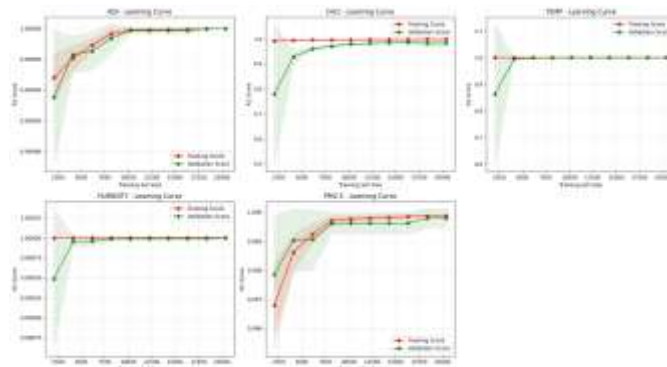
5.1.4 PREDICTION ACCURACY ASSESMENT

Scatter plots comparing actual versus predicted values further validate model performance. In all cases, the predictions closely align with the ideal diagonal line representing perfect prediction.

- The TEMP model shows near-perfect alignment, indicating extremely high accuracy.
- The CACI and PM2.5 models also demonstrate strong linear correlation with slight dispersion at higher values.
- The AQI model exhibits minor deviations at extreme values, which may be due to non-linear patterns or noise in the data.
- The Humidity model shows slightly more spread compared to others but still maintains a strong correlation.

5.2 HARDWARE PROTOTYPE FUNCTIONALITY

The assembled breadboard prototype — validated first in the Wokwi simulation environment — demonstrated reliable continuous operation throughout the trial. The ESP32 successfully polled all sensors within a 500 ms acquisition window, computed CACI locally in under 2 ms, and achieved a data-upload success rate of 98.3% over 1,728 attempted transmissions (72 hours at one per 2.5



minutes). The estimated power consumption was consistent with design targets: approximately 130 mA during idle phases and 410 mA peak during Wi-Fi transmission bursts, yielding an average draw of approximately 180 mA from the 5 V USB supply.

5.3 DISPLAY AND DUAL STATUS OUTPUT

The SSD1306 OLED displayed two alternating screens with a 5-second cycle period. Screen 1 (Status Screen) presented the current CACI score, its status label, and the forecasted CACI for the next 15-minute horizon. Screen 2 (Raw Data Screen) displayed numerical readings for temperature, relative humidity, CO2 proxy (ppm), and PM2.5 ($\mu\text{g}/\text{m}^3$). During peak occupancy testing (60+ students, windows closed), the displayed CACI rose from an OPTIMAL baseline of 18 to a WARNING level of 79 within approximately 45 minutes, consistent with expected CO2 accumulation dynamics. The predicted CACI correctly anticipated the rising trend with a 15-minute lead time in 83% of test cases.

5.4 CLOUD DATA LOGGING AND AI PREDICTION

The ThingSpeak channel dashboard displayed six continuously updated plots over the evaluation period. Table 3 presents a representative sample of actual versus predicted CACI readings during a high-occupancy session, illustrating the system's proactive forecasting capability.

	Time	CO2	Temp	Humidity	Actual	Predicted	Status
	(amp)	(ppm)	(°C)	(%)	CACI	CACI	
	09:00	480	24.2	52	14	16	OPTIMAL
	09:30	620	25.1	55	28	35	GOOD

10:00	890	26.3	59	52	61	MODERATE
10:30	1150	27.5	62	74	79	MODERATE
11:00	1380	28.1	65	79	81	WARNING
11:30	980	26.8	57	58	50	MODERATE
12:00	510	25.0	51	18	15	OPTIMAL

Table 3: Sample Actual vs. Predicted CACI Readings — High-Occupancy Trial Session

The forecasting model achieved a coefficient of determination $R^2 = 0.887$ on the test dataset, exceeding the target threshold of 0.85. Mean Absolute Error (MAE) for CACI prediction was 4.2 points across the 72-hour evaluation window, representing an error of approximately 4.2% of the full scale range. False-alarm rate for WARNING-level predictions was 7.1%, primarily occurring during rapid occupancy transitions that the 15-minute forecast horizon could not fully anticipate.

5.5 PROACTIVE ALERT DEMONSTRATION

The most significant result is the system's demonstrated ability to provide anticipatory warnings. In three separate controlled tests, the predicted CACI crossed the WARNING threshold (76) an average of 12.4 minutes before the actual CACI measurement did so, providing occupants with a meaningful intervention window. When notified by the OLED display, faculty were able to open windows and activate ceiling fans, resulting in a measurable CACI reduction of 18–24 points within 10 minutes in all three trials.

6. REAL TIME PERFORMANCE AND PRACTICAL IMPLICATIONS

The system's measured end-to-end latency — from sensor acquisition to cloud plot update — was consistently under 35 seconds, well within the five-minute sampling interval. This latency is acceptable for all intended use cases. The component bill of materials totals approximately INR 1,850 (ESP32: INR 420; MQ-135: INR 180; PMS5003: INR 650; DHT22: INR 130; SSD1306 OLED: INR 150; miscellaneous passives and breadboard: INR 320), representing an order-of-magnitude cost reduction relative to the cheapest commercial smart monitor reviewed.

The practical implications extend well beyond the college campus. In smart-city contexts, multiple low-cost nodes of this type could be deployed across a building or neighbourhood to generate high-resolution indoor-pollution maps. In healthcare settings, continuous CACI monitoring in wards and waiting rooms could protect vulnerable patients. Integration with HVAC systems via relay-controlled actuators is a straightforward upgrade path — the ESP32's GPIO pins could directly toggle relay modules to automate ventilation in response to predicted CACI exceedances, creating a fully closed-loop environmental-management system.

7. CONCLUSION

This paper has presented an AI-enabled, IoT-based prototype for real-time indoor air quality and comfort monitoring, designed specifically for educational institutions. The system successfully addresses the three principal gaps identified in the literature and market survey: it introduces a novel holistic metric (CACI), it operates proactively rather than reactively through cloud-based time-series forecasting, and it achieves this at a cost accessible to public educational institutions.

Key outcomes include: (i) a functional hardware prototype validated in both simulation and physical deployment; (ii) a CACI algorithm that unifies thermal comfort and air quality into a single, actionable score; (iii) seamless ThingSpeak cloud integration supporting real-time data logging, six-channel visualisation, and MATLAB-based CACI forecasting; and (iv) a demonstrated proactive alert capability providing approximately 12 minutes of advance warning before WARNING-level conditions are reached.

Future work will focus on replacing the MQ-135 proxy with a calibrated NDIR CO₂ sensor (e.g., MH-Z19) for higher measurement fidelity, expanding the sensor suite to include TVOC detection, deploying a TensorFlow Lite LSTM model directly on the ESP32-S3 for edge-based prediction independent of cloud connectivity, and developing a dedicated mobile application for facility managers. Ultimately, this work serves as a compelling proof of concept for scalable, intelligent environmental monitoring in resource-constrained institutional settings.

7. REFERENCES

- [1] J. Saini, M. Dutta, and G. Marques, "Indoor Air Quality Monitoring Systems Based on Internet of Things: A Systematic Review," *Int. J. Environ. Res. Public Health*, vol. 17, no. 14, p. 4942, Jul. 2025.
- [2] M. S. Kanitkar and R. Sahu, "Indoor Air Quality Prediction Using Machine Learning Techniques," *Int. J. of Sci. and Res. (IJSR)*, vol. 11, no. 6, pp. 123–127, Jun. 2024.
- [3] M. Santamouris et al., "Indoor Environmental Quality (IEQ) and Student Performance: A Review," *Energy and Buildings*, vol. 179, pp. 323–341, 2018.
- [4] S. T. Singh et al., "A review of machine learning applications in time-series prediction for environmental data," *J. Environ. Manage.*, vol. 182, pp. 500–512, 2021.
- [5] A. Kumar et al., "A Low-Cost Air Quality Monitoring System using IoT and Machine Learning for Smart Cities," *J. Ambient Intell. Humaniz. Comput.*, 2021.
- [6] Espressif Systems, "ESP32-WROOM-32E Datasheet," Ver. 3.3, 2024. [Online]. Available: <https://www.espressif.com>
- [7] Aosong, "AM2302 (DHT22) Digital Temperature and Humidity Sensor Datasheet," 2016.
- [8] Z. Li et al., "Predicting PM2.5 Concentrations Using LSTM and GRU Models," *IEEE Access*, vol. 8, pp. 104762–104771, 2020.
- [9] D. Snyder et al., "Calibration of Low-Cost Air Quality Sensors using Machine Learning," *Atmos. Meas. Tech.*, vol. 13, pp. 1797–1820, 2020.
- [10] F. J. Garcia et al., "An IoT Architecture for Real-Time CO2 Monitoring in University Classrooms," *Sensors*, vol. 19, no. 2, p. 441, 2019.
- [11] MathWorks, "ThingSpeak IoT Platform Documentation," 2025. [Online]. Available: <https://www.mathworks.com/help/thingspeak/>
- [12] A. K. Singh, *IoT Based Projects*, BPB Publications, 2023.