

Edge-Driven CO₂ and Occupancy Prediction with Certificate Generation for Optimized Classroom Ventilation Systems

Chinthlagari Bhavitha
Chaitanya Bharathi Institute of Technology
Hyderabad, India

Sri Harsha Goruganti
Chaitanya Bharathi Institute of Technology
Hyderabad, India

Vangipuram Sree Raghu Vardhan
Chaitanya Bharathi Institute of Technology
Hyderabad, India

Eliganti Ramalakshmi
Chaitanya Bharathi Institute of Technology
Hyderabad, India

Abstract—Optimizing classroom ventilation is crucial for indoor air quality and occupant well-being. This paper presents an edge computing approach using Internet of Things (IoT) sensors and Long Short-Term Memory (LSTM) networks for intelligent ventilation management in educational settings. Low-cost IoT devices with environmental sensors collect and process data locally. A lightweight LSTM model on the edge predicts ventilation quality and occupancy patterns in real-time, enabling prompt adjustments. The system employs a hierarchical architecture with edge devices for local processing, a fog layer for aggregation, and cloud services for data storage and analysis. This edge-driven system demonstrates effective real-time monitoring and analysis of temperature, humidity, carbon dioxide levels, and occupancy, facilitating optimized classroom conditions and generating environmental quality certificates. The integration of edge computing, machine learning, and IoT offers precise environmental quality evaluation for educational institutions.

Index Terms—Edge Computing, Indoor Air Quality, CO₂ Prediction, LSTM, IoT, Occupancy Monitoring, Smart Ventilation, Certificate Generation

I. INTRODUCTION

Indoor air quality (IAQ) in educational settings significantly impacts student cognitive function, comfort, and overall health, particularly in densely occupied classrooms [1]. Traditional ventilation systems often rely on static thresholds or manual controls, proving inefficient and unresponsive to dynamic changes in occupancy and environmental conditions like Carbon Dioxide (CO₂) levels [2]. Such systems struggle to adapt to fluctuating occupancy, varying room sizes, and diverse usage patterns, leading to inconsistent air quality, potential health concerns, and suboptimal energy consumption.

The primary challenge lies in developing an intelligent, responsive ventilation system capable of real-time data acquisition, accurate prediction of environmental changes, and automated control actions. While cloud-based solutions exist, they often suffer from latency issues unsuitable for immediate environmental adjustments [3]. Edge computing, coupled with advancements in Internet of Things (IoT) sensors and deep learning models like Long Short-Term Memory (LSTM)

networks, offers a promising alternative by enabling localized data processing and predictive control with minimal delay.

This paper proposes an edge-driven system for real-time CO₂ and occupancy prediction to optimize classroom ventilation. Our approach integrates low-cost IoT sensors for environmental data collection (temperature, humidity, CO₂, occupancy), a lightweight LSTM model deployed on edge devices for predictive analysis, and certificate generation for environmental quality assurance. The system employs a hierarchical architecture, leveraging edge devices for immediate processing and a fog layer for aggregation, alongside cloud integration for long-term data storage and analysis.

The key contributions of this work are:

- Development and edge deployment of a lightweight LSTM model for real-time prediction of CO₂ levels and occupancy patterns in classrooms.
- Design and implementation of a hierarchical edge-fog-cloud architecture for efficient data processing, low-latency control, and scalability.
- Real-time generation of environmental quality certificates based on continuous monitoring and prediction.
- Demonstration of the system's effectiveness in maintaining optimal classroom conditions through adaptive ventilation adjustments.

The remainder of this paper is organized as follows: Section II reviews related work in smart ventilation and edge computing. Section III details the proposed system architecture and methodology. Section IV presents the experimental results and evaluation. Finally, Section V concludes the paper and discusses future work.

II. RELATED WORK

Research in smart ventilation and indoor air quality (IAQ) management has evolved significantly with the integration of Internet of Things (IoT) technologies and machine learning. Early systems often employed rule-based control triggered by simple sensor thresholds, which proved inadequate for

dynamic environments. Recent efforts have focused on leveraging IoT sensor networks for real-time data collection and more sophisticated control strategies.

Several studies have explored IoT-based monitoring of various IAQ parameters, including CO₂, temperature, humidity, and particulate matter (PM). Luo et al. [4] investigated natural ventilation potential using real-time sensor data, highlighting the gap between potential and actual usage due to occupant behavior. Rastogi et al. [5] developed a context-aware system using k-NN to detect poor ventilation based on multiple pollutants. Others, like Mahbub et al. [6], integrated IAQ monitoring with other building functions like lighting and fire safety within a single embedded system.

The application of machine learning, particularly deep learning models like Long Short-Term Memory (LSTM) networks, has shown promise for predicting IAQ trends. Tagliabue et al. [7] used Artificial Neural Networks (ANNs) and Markov models to forecast comfort conditions in educational settings based on CO₂, temperature, and humidity. Wang et al. [8] developed a Convolutional Transformer LSTM (CT-LSTM) model for predicting Air Quality Index (AQI), demonstrating superior performance over traditional models. Yang et al. [9] proposed a hybrid Bayesian Optimization-Empirical Mode Decomposition-LSTM (BO-EMD-LSTM) method for highly accurate long-term CO₂ prediction. These predictive capabilities enable proactive ventilation control rather than reactive adjustments.

The trend towards edge computing is prominent in recent IAQ research, aiming to reduce latency, enhance privacy, and improve responsiveness compared to purely cloud-based architectures. Idrees et al. [3] proposed an edge architecture for real-time monitoring using low-cost sensors and local processing. Taştan and Göközan [10] developed an IoT e-nose using edge computing for rapid anomaly detection. Zhang et al. [11] deployed an LSTM model directly on an ESP32 for CO₂ forecasting, enabling proactive control and energy savings. Similarly, Sun et al. [12] implemented model predictive control (MPC) on a Raspberry Pi for real-time optimization of CO₂ and energy use.

While existing research has explored IoT-based monitoring, predictive modeling, and edge deployment individually or in partial combinations, our work integrates these aspects into a cohesive system specifically targeting classroom environments. We combine real-time, multi-sensor data acquisition (including occupancy) with a lightweight LSTM model deployed at the edge for simultaneous CO₂ and occupancy prediction, coupled with automated environmental quality certificate generation. This holistic edge-driven approach aims to provide a responsive, scalable, and verifiable solution for optimizing ventilation in educational spaces, addressing the limitations of previous systems in terms of combined prediction, real-time edge control, and quality assurance.

III. SYSTEM DESIGN AND METHODOLOGY

Our proposed system employs a hierarchical Edge-Fog-Cloud architecture (illustrated in Figure 1) designed for real-

time monitoring, prediction, and control of classroom environmental conditions, primarily focusing on CO₂ levels and occupancy.

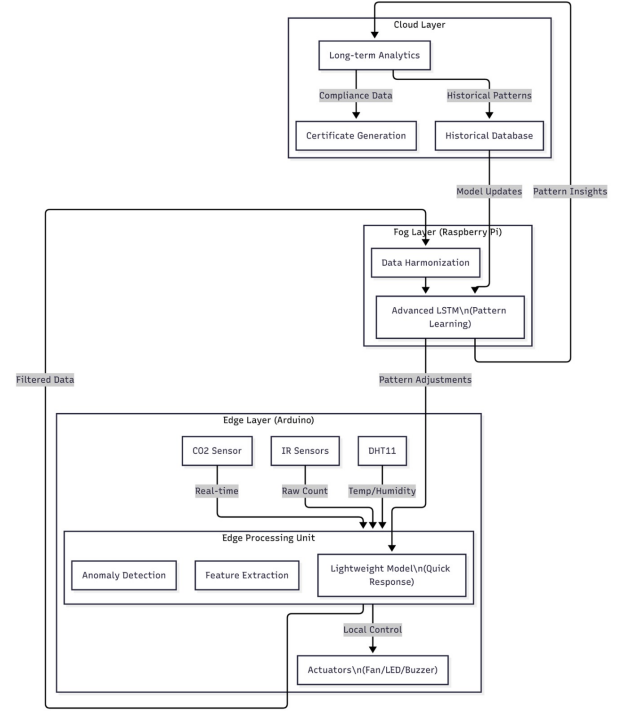


Fig. 1. Proposed System Flow Architecture (Edge-Fog-Cloud).

A. System Architecture

Edge Layer: Consists of Arduino Uno based nodes deployed in each monitored classroom. Each node is equipped with:

- An MQ135 sensor for CO₂ concentration measurement (400-2000 ppm range).
- A DHT11 sensor for temperature and humidity readings.
- Paired Infrared (IR) sensors at doorways for bidirectional occupancy counting.
- A 16x2 LCD for displaying real-time status (temperature, humidity, air quality).
- Actuators: 12V DC fans (controlled via relay modules), LEDs (status indicators), and buzzers (alerts).

The Arduino units perform initial data acquisition, basic filtering, local actuation (immediate response to critical thresholds), and transmit processed sensor readings to the Fog layer via serial communication (USB, 9600 baud).

Fog Layer: A Raspberry Pi 4 (4GB RAM) acts as the central fog controller. Its responsibilities include:

- Aggregating data streams from multiple Edge nodes.
- Executing the pre-trained LSTM model for CO₂ and occupancy prediction.
- Implementing decision logic based on predictions and custom metrics (KPIv, Room Quality Score).
- Sending control commands back to Edge nodes (e.g., activate fans).

- Communicating with the Cloud layer for data logging and synchronization.

Python is used for programming the Fog layer, utilizing libraries like TensorFlow/TensorFlow Lite, NumPy, Pandas, and Scikit-learn.

Cloud Layer: ThingSpeak is utilized for cloud functionalities:

- Long-term storage of historical sensor data and predictions.
- Data visualization through dashboards.
- Automated generation of environmental quality certificates based on aggregated performance data.
- Potential for remote monitoring and triggering model updates.

B. Data Collection and Preprocessing

Real-time data (CO_2 , temperature, humidity, occupancy count) is collected every few minutes (configurable, typically 3-5 mins) by the Edge nodes. Before model training (offline) and prediction (online on Fog), the data undergoes preprocessing:

- Handling Missing Values: Time-based interpolation is used for offline dataset preparation.
- Normalization: Min-Max scaling is applied to features to ensure values are within a consistent range (e.g., 0 to 1), preventing features with larger magnitudes from dominating the learning process.
- Sequencing: Data is structured into sequences using a sliding window approach (e.g., using the last 10-15 timesteps) to provide temporal context for the LSTM model.
- Train/Validation/Test Split: For offline training, the dataset is split chronologically (e.g., 70

C. LSTM Predictive Model

A Long Short-Term Memory (LSTM) network is employed for its ability to capture long-range temporal dependencies in time-series data.

Architecture: The model consists of:

- An input layer accepting sequences of features (CO_2 , temperature, humidity, occupancy).
- Two stacked LSTM layers (64 units followed by 32 units) to learn hierarchical temporal patterns.
- Dropout layers between LSTM layers to mitigate overfitting.
- A Dense output layer producing the predicted CO_2 concentration for the next timestep.

Training: The model was trained offline using the preprocessed dataset.

- Loss Function: Mean Squared Error (MSE) was used to minimize the difference between predicted and actual values.

$$\mathcal{L}_{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (1)$$

- Optimizer: Adam optimizer was used for efficient gradient descent.
- Epochs: Trained for up to 200 epochs with EarlyStopping based on validation loss to prevent overfitting.
- Learning Rate Scheduling: Adaptively adjusted the learning rate during training.

The trained model weights and scaler were saved (`pretrained_lstm.weights.h5`, `pretrained_scaler.joblib`) and converted to TensorFlow Lite format for efficient deployment on the Raspberry Pi (Fog layer).

D. Prediction Logic and Control

The Fog layer uses the deployed LSTM model to predict future CO_2 levels. Based on these predictions and current sensor readings, it calculates two key metrics:

KPIv (Key Performance Indicator for Ventilation): Quantifies ventilation quality based on CO_2 and occupancy.

$$KPIv = \begin{cases} 2.0 & \text{if } CO_2 > 1000 \text{ ppm} \\ \frac{CO_2 - 400}{20 \times \text{actual_people}} & \text{if } \text{actual_people} > 0 \\ 0.5 & \text{if } \text{est_people} > 0, \text{actual_people} = 0 \\ 0.0 & \text{otherwise} \end{cases} \quad (2)$$

where `actual_people` is the count from IR sensors, and `est_people` represents estimated occupancy.

Room Quality Score: Provides a holistic score incorporating multiple factors.

$$\text{Score} = 5 \times KPIv + 0.2 \times |\text{Temp} - 24| + 0.3 \times \frac{CO_2}{1000} + 0.1 \times \text{Occupancy} \quad (3)$$

This score helps compare classrooms and trigger broader actions.

Based on these metrics and predictions, the Fog layer sends commands to the Edge nodes to activate fans for ventilation, display status updates on LCDs, and trigger alerts (LEDs, buzzers) if thresholds are breached or predicted to be breached.

IV. RESULTS AND EVALUATION

This section presents the results obtained from the implementation and deployment of the proposed edge-driven classroom environmental monitoring and ventilation optimization system.

A. Experimental Setup

The system was deployed across two classrooms, utilizing the hardware architecture described in Section III (Arduino nodes, Raspberry Pi fog controller, sensors), shown in Figure 2. Environmental data (Temperature, Humidity, CO_2 , Occupancy) was collected continuously over a monitoring period (e.g., several days, as indicated by ThingSpeak data from April 13-19 in the thesis). Data was logged locally and synchronized with the ThingSpeak cloud platform at regular intervals (approx. 5-15 minutes, based on descriptions).

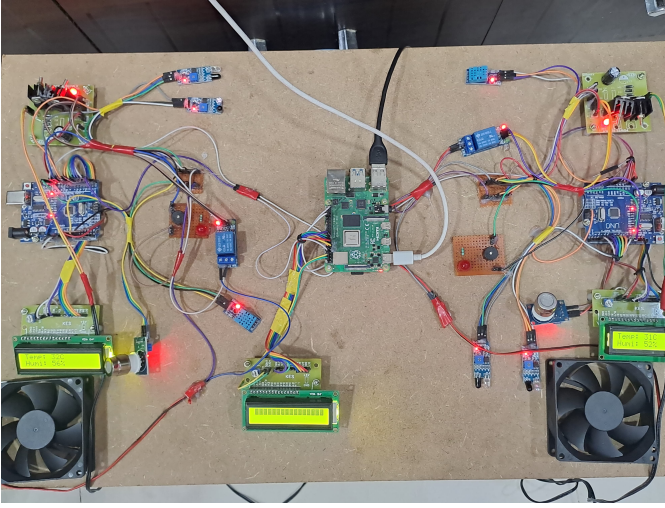


Fig. 2. Hardware Implementation of the Edge-Based IoT System.

B. Data Collection and System Operation

Real-time sensor data was successfully collected and processed by the Edge and Fog layers. ThingSpeak visualizations (Figures 3 and 4) confirmed consistent data logging. During the monitored period, observed environmental parameters included:

- Temperature: Fluctuating generally between 31.0°C and 32.0°C.
- Humidity: Varying between 36.0% and 40.0%.
- CO₂: Concentrations typically observed between 11-13 ppm in the provided sample, but showing potential spikes correlating with occupancy or ventilation changes (as suggested by the description of graphs in the thesis).
- Occupancy: Variable rates observed, reflecting classroom usage patterns.

The system demonstrated its ability to process this data in real-time, calculate the KPIv and Room Quality Score, and provide actionable recommendations. Log outputs and LCD displays (Figure 5) confirmed the system's capability to compare classroom conditions and suggest the optimal room based on calculated scores (e.g., recommending Classroom 1 with a score of 61.35 over Classroom 2 with 114.74 in one instance).

C. LSTM Model Performance

The LSTM model was trained offline on a harmonized dataset derived from multiple sources, preprocessed using forward-fill for missing values and Min-Max scaling, and structured into sequences of 10 timesteps. The training utilized the Adam optimizer and Mean Squared Error (MSE) loss, running for 5 epochs with early stopping and learning rate adjustment (as detailed in Section III). The pre-trained model, evaluated on the test set, achieved a final loss (MSE) of 0.0012 and a Mean Absolute Error (MAE) of 0.0143. This level of performance indicates the model's capability to predict environmental trends with reasonable accuracy. The trained model, deployed on the Fog layer (Raspberry Pi), successfully

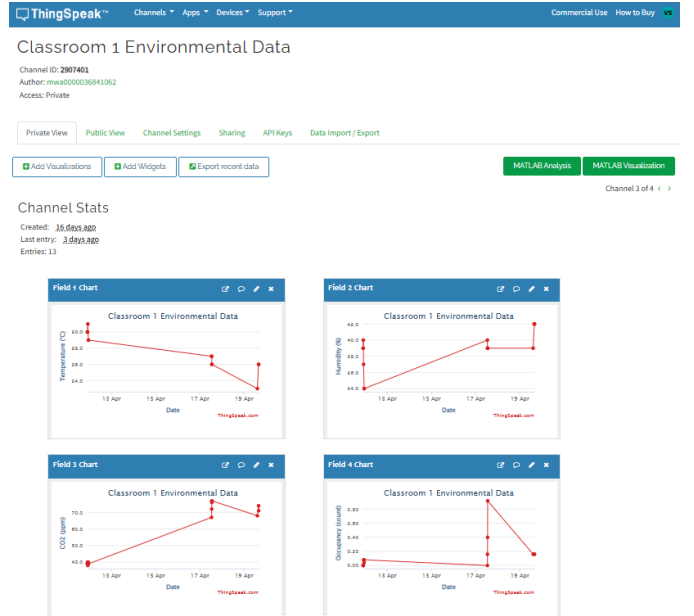


Fig. 3. ThingSpeak Data Visualization for Classroom 1.

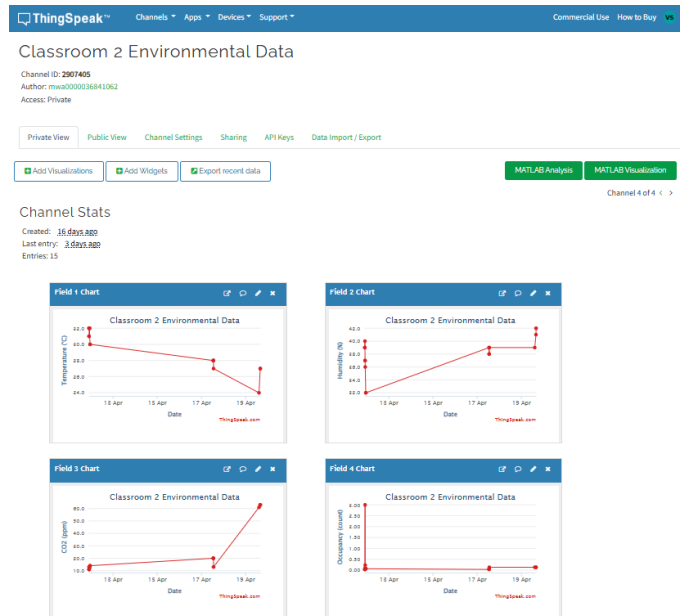


Fig. 4. ThingSpeak Data Visualization for Classroom 2.

processed real-time sensor sequences to generate predictions. The system's ability to utilize these predictions alongside current data for calculating KPIv and Room Quality Score confirms the practical integration of the ML model into the control loop.

D. Certificate Generation

The automated certificate generation component successfully processed the collected environmental data stored via ThingSpeak. A sample generated certificate (Figure 6) demonstrated the system's ability to:

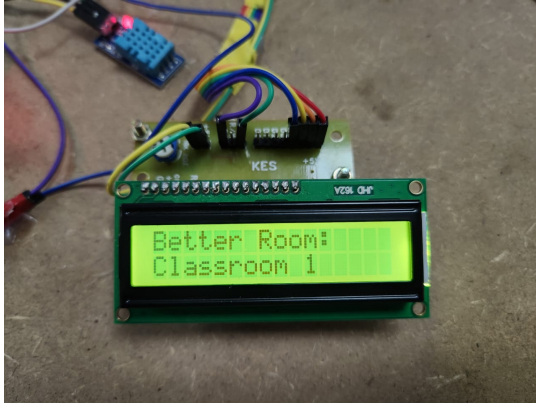


Fig. 5. LCD Output Display Indicating Recommendation of Optimal Classroom.

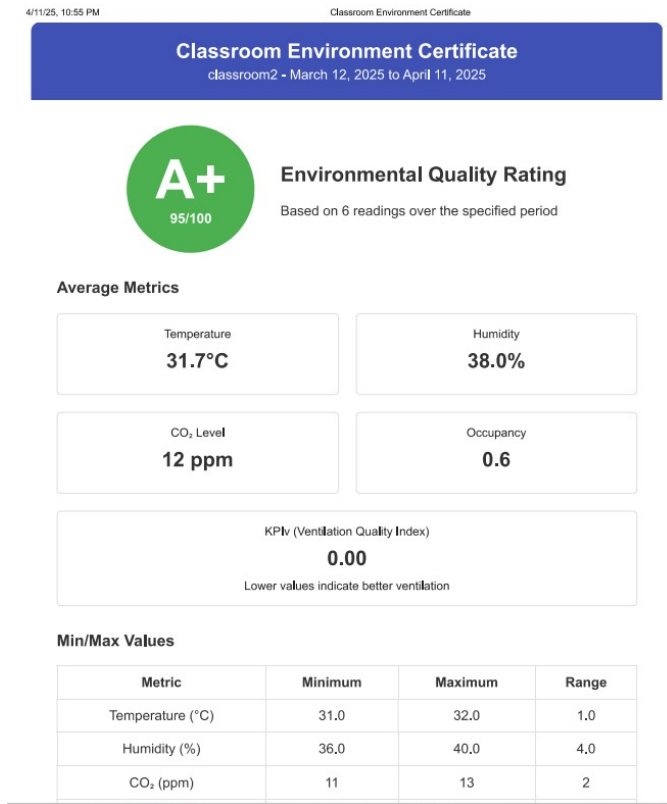


Fig. 6. Example Generated Classroom Environmental Quality Certificate.

- Aggregate data over a specified period (e.g., 6 readings shown).
- Calculate average, minimum, and maximum values for key parameters (Temp, Humidity, CO₂, Occupancy).
- Compute the KPIv based on the collected data.
- Assign an overall environmental quality grade (e.g., A+ with a score of 95/100 achieved in the sample).

This validates the end-to-end functionality of the system, from data collection and analysis to formal quality reporting.

Overall, the results demonstrate the feasibility and effec-

tiveness of the proposed edge-driven system in monitoring classroom environments, utilizing LSTM predictions for intelligent assessment, and providing automated control and quality certification.

V. CONCLUSION AND FUTURE WORK

This paper presented an edge-driven system for CO₂ and occupancy prediction aimed at optimizing classroom ventilation. By integrating real-time IoT sensor data (CO₂, temperature, humidity, occupancy) with an LSTM predictive model deployed on a Raspberry Pi (fog layer) controlling Arduino-based edge nodes, the system demonstrated effective environmental monitoring and proactive ventilation management. Key achievements include the successful implementation of the hierarchical edge-fog-cloud architecture, accurate prediction of environmental trends enabling calculations of KPIv and Room Quality Score, and the automated generation of environmental quality certificates via ThingSpeak integration.

The results validate the feasibility of using lightweight LSTM models on edge/fog devices for real-time analysis and control in resource-constrained environments like classrooms. The system effectively bridges low-cost IoT deployment with intelligent building management principles, offering a data-backed approach to ensuring healthier indoor air quality. The ability to provide verifiable environmental quality certificates adds a layer of accountability and transparency.

Future work will focus on enhancing the system's practicality and scalability. Key directions include:

- **Wireless Communication:** Replacing the current wired serial connection between edge and fog layers with wireless protocols like Zigbee or LoRaWAN to improve deployment flexibility.
- **Model Enhancement:** Training the LSTM model on larger, more diverse datasets incorporating data from various room types and conditions to improve generalizability and accuracy.
- **Expanded Sensing:** Incorporating additional sensors for parameters like Particulate Matter (PM2.5) and Volatile Organic Compounds (VOCs) to provide a more comprehensive IAQ assessment.
- **HVAC Integration:** Interfacing the system directly with building HVAC systems for fully automated and optimized ventilation control, moving beyond simple fan actuation.
- **Advanced ML Techniques:** Exploring techniques like federated learning for enhanced privacy and adaptive learning for real-time model fine-tuning directly on edge/fog devices.
- **Scalability Testing:** Deploying the system across a larger number of classrooms or zones to evaluate performance and management at scale.

By pursuing these enhancements, the system can evolve into a more robust, scalable, and intelligent solution, contributing significantly to the development of healthier, energy-efficient, and data-driven educational environments.

REFERENCES

- [1] K. Azuma, N. Kagi, U. Yanagi, and H. Osawa, "Effects of low-level inhalation exposure to carbon dioxide in humans: A systematic review," *Environment International*, vol. 121, pp. 1178–1190, 2018.
- [2] V. F. McNeill, G. Zivelonghi, and L. C. Tagliabue, "Indoor air quality optimization strategies in schools and universities," *Current Opinion in Environmental Science & Health*, vol. 25, p. 100313, 2022.
- [3] A. K. Idrees, A. Al-Maadid, A. K. Al-Ali, and M. S. Al-Muraikhi, "An edge computing based architecture for real-time air quality and ventilation monitoring in enclosed environments," *Sensors*, vol. 21, no. 18, p. 6086, 2021.
- [4] M. Luo, B. Cao, and Y. Zhu, "Assessing natural ventilation potential in buildings based on real-time sensor data," *Building and Environment*, vol. 200, p. 107959, 2021.
- [5] K. Rastogi, S. Srivastava, and V. Sharma, "An iot based context-aware system for indoor ventilation monitoring," *IEEE Internet of Things Journal*, 2021.
- [6] M. Mahbub, K. Biswas, and A. K. Das, "Design and implementation of a cloud-enabled multifunctional embedded iot system for indoor ventilation, lighting control and fire detection," *Internet of Things*, vol. 15, p. 100413, 2021.
- [7] L. C. Tagliabue, F. R. Cecconi, S. Rinaldi, and A. L. C. Ciribini, "A data-driven model based on sensor network data for indoor air quality prediction in educational facilities," *Building and Environment*, vol. 199, p. 107893, 2021.
- [8] J. Wang, S. Li, and C. Li, "Ct-lstm: A novel hybrid model for air quality index prediction," *IEEE Access*, vol. 9, pp. 113 438–113 447, 2021.
- [9] G. Yang, L. Zhang, and J. Zhang, "A hybrid bo-emd-lstm model for reliable long-term co2 concentration forecasting in hvac systems," *Sustainable Cities and Society*, vol. 72, p. 103018, 2021.
- [10] M. Taştan and H. Göközan, "An iot based electronic nose for real-time indoor air quality monitoring," *Air Quality, Atmosphere & Health*, vol. 12, no. 10, pp. 1177–1186, 2019.
- [11] Y. Zhang, B. Li, and W. Li, "Lstm-based co2 concentration forecasting model for edge computing devices in smart buildings," *IEEE Internet of Things Journal*, vol. 8, no. 16, pp. 13 229–13 238, 2021.
- [12] L. Sun, R. Jia, and H. Li, "Edge computing based model predictive control for optimal indoor co2 concentration and energy consumption in smart buildings," *Building and Environment*, vol. 217, p. 109063, 2022.