2025-12-14

TinyML for Environmental Intelligence: Optimizing Tree-Based Ensembles for Real-Time, On-Device IAQ Forecasting

Mahir, Hasin

Independent University, Bangladesh

TinyML for Environmental Intelligence: Optimizing Tree-Based Ensembles for Real-Time, On-Device IAQ Forecasting

By

Hasin Mahir

Student ID: **2230087**

Tahfizul Hasan Zihan

Student ID: **1921629**

Autumn, 2025

Supervisor:

Dr. Md. Tarek Habib

Assistant Professor

Department of Computer Science & Engineering

Independent University, Bangladesh

December 14, 2025

Dissertation submitted in partial fulfillment for the degree of Bachelors of Science in Computer Science & Engineering

Department of Computer Science & Engineering

Independent University, Bangladesh

Attestation

We, Hasin Mahir (2230087) and Tahfizul Hasan Zihan (1921629), certify that the report was completed by us and submitted in partial fulfillment of the requirement of a Bachelor's Degree in Computer Science and Engineering from Independent University, Bangladesh (IUB). This project was completed under the supervision of Dr. Md. Tarek Habib. All sources of information used in this project have been verified, and the report has been taken care of in it.

Signature	Date

Hasin Mahir
Name

Signature	Date

Tahfizul Hasan Zihan
Name

Acknowledgement

First and foremost, we express our earnest gratitude to the Almighty, Allah (S.W.T), for His endless blessings and for granting us the strength and ability to complete this work.

We would like to express our profound indebtedness and sincere gratitude to our senior project supervisor, **Dr. Md. Tarek Habib**, Assistant Professor, Department of Computer Science Engineering at Independent University, Bangladesh. His unwavering guidance, insightful supervision, and invaluable suggestions at every stage were instrumental in the successful completion of this project. We are deeply thankful for his mentorship and for securing the necessary funding to support this thesis; without his inspiration and steadfast support, this study would not have been possible.

We promote our deepest appreciation to **Dr. Ferdows Zahid**, whose initial vision was pivotal in conceptualizing this project. We are equally indebted to **Dr. Mahady Hasan**, whose generous provision of funding and resources through FabLab IUB facilitated the practical execution of our research.

Furthermore, we extend our heartfelt thanks to the dedicated **Fab Lab IUB** team- **Md. Shirazim Munir**, **Md. Ataur Rahman**, and **Md. Rubayed Mehedi** for their exceptional technical assistance and their crucial role in managing classroom schedules, which ensured our data collection proceeded seamlessly.

Finally, we wish to acknowledge **Dr. Saadia Binte Alam**, Head of the Department, for her exemplary leadership. Her encouragement has been a significant source of motivation, inspiring us to pursue rigorous and impactful research.

Letter of Transmittal

December 14, 2025

Dr. Md. Tarek Habib
Assistant Professor
Department of Computer Science and Engineering
Independent University, Bangladesh

Subject: Submission of Senior Project Report on "TinyML for Environmental Intelligence: Optimizing Tree-Based Ensembles for Real-Time, On-Device IAQ Forecasting"

Dear Sir,

We are honored to submit our senior project report titled "TinyML for Environmental Intelligence: Optimizing Tree-Based Ensembles for Real-Time, On-Device IAQ Forecasting." This report represents the culmination of our efforts, research, and practical application of various emerging technologies under your esteemed supervision.

Throughout the development of this project, we incorporated TinyML forecasting strategies, embedded system design, and real-world IoT implementation practices that have significantly enhanced our technical knowledge and collaborative skills. We believe this report provides a comprehensive overview of our work and serves its intended purpose effectively.

We would like to extend our heartfelt gratitude for your continuous support, insightful feedback, and invaluable mentorship throughout the project. Your guidance played a vital role in helping us navigate challenges and reach our objectives. We sincerely hope this report meets your expectations and is found to be satisfactory.

Thank you once again for your encouragement and assistance throughout the semester.

Yours sincerely,

Hasin Mahir (2230087)
Tahfizul Hasan Zihan (1921629)

Evaluation Committee

Supervision Panel

....	....
Dr. Md. Tarek Habib	
Supervisor	Co-supervisor

Examiners

....	....
Dr. Mohammad Shidujaman	**Dr. Md Zahangir Alam**
Internal Examiner 1	Internal Examiner 2
....	
Anisul Haq Bhuya	
External Examiner	

Office Use

....	....
Center Director	Head of the Department
....	....
Director Graduate Studies, Research and Industry Relations	Library

Acronyms

IAQ Indoor Air Quality

LMIC Lower-Middle-Income Countries

PM Particulate Matter ($PM_{2.5}$, PM_{10})

VOC Volatile Organic Compounds

TVOC Total Volatile Organic Compounds

CO2 Carbon Dioxide

WHO World Health Organization

EPA Environmental Protection Agency

SDG Sustainable Development Goals

ASHRAE American Society of Heating, Refrigerating and Air-Conditioning Engineers

EN European Norm

IoT Internet of Things

AI Artificial Intelligence

ML Machine Learning

TinyML Tiny Machine Learning

AIoT Artificial Intelligence of Things

DCV Demand-Controlled Ventilation

HVAC Heating, Ventilation, and Air Conditioning

BMS Building Management System

WSN Wireless Sensor Network

MCU	Microcontroller Unit
ESP32	Low-power System on a Chip (SoC) with Wi-Fi/Bluetooth
SoC	System on a Chip
PCB	Printed Circuit Board
NDIR	Non-Dispersive Infrared
BLE	Bluetooth Low Energy
I2C	Inter-Integrated Circuit
HTTP	Hypertext Transfer Protocol
RF	Random Forest
LSTM	Long Short-Term Memory
XGBoost	Extreme Gradient Boosting
LightGBM	Light Gradient Boosting Machine
Prophet	Additive Time-Series Forecasting Model
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
MAPE	Mean Absolute Percentage Error
R2	Coefficient of Determination
WBS	Work Breakdown Structure
BOM	Bill of Materials
SBS	Sick Building Syndrome

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Indoor Air Quality (IAQ) serves as a fundamental determinant of student health and cognitive efficacy, specifically within the densely populated educational institutions of Lower-Middle-Income Countries (LMICs) like Bangladesh. While global health directives establish rigorous thresholds for Carbon Dioxide (CO_2) and Particulate Matter ($PM_{2.5}$), the infrastructure in these regions predominantly relies on passive natural ventilation. This structural limitation frequently results in the unchecked accumulation of pollutants during occupied hours, creating environmental hazards that exacerbate respiratory ailments and diminish academic performance. The challenge is compounded by the inadequacy of existing monitoring solutions; conventional Internet of Things (IoT) systems largely depend on cloud-centric processing, which introduces substantial latency, data privacy vulnerabilities, and prohibitive recurring costs for budget-constrained schools. Furthermore, the digital divide is deepened by a scarcity of edge-native predictive technologies capable of functioning autonomously without reliable internet connectivity. Addressing these systemic deficiencies requires a paradigm shift from passive monitoring to proactive, intelligent control. To this end, this thesis proposes and implements a comprehensive "Environmental Intelligence" framework, articulated through three primary innovations:

- **Comparative ML Modeling:** A rigorous benchmarking of forecasting architectures (Prophet, Random Forest, LSTM) against a novel, high-resolution dataset collected from a naturally ventilated classroom in Dhaka. This analysis establishes the baseline efficacy of non-linear models over traditional decomposition methods for capturing complex indoor pollutant dynamics in LMIC settings.

- **TinyML Forecasting Engine:** The development and deployment of an optimized, edge-native predictive pipeline on an ESP32 microcontroller. By refining feature engineering strategies and converting heavy ensemble models (Random Forest, XGBoost, LightGBM) into efficient C-code, this innovation enables autonomous, server-less week-ahead IAQ forecasting within strict hardware constraints.

- **IoT-DCV Node:** A low-cost, lag-aware Demand-Controlled Ventilation system that utilizes a novel joint CO_2-$PM_{2.5}$-Pressure signature to preemptively regulate classroom airflow before toxicity levels peak, bridging the gap between prediction and physical intervention.

These components operate in synergy: the comparative analysis informs the model selection for the embedded engine, which is then hosted on the IoT node to drive real-time decision-making. The integrated system, deployed on the ESP32-S3 architecture, supports automated actuation without external cloud reliance. By optimizing sophisticated machine learning algorithms for execution on accessible, low-power hardware, this project aims to democratize access to healthy learning environments and support Sustainable Development Goal 3 (Good Health and Well-being) and Goal 4 (Quality Education).

1.2 Rationale and Motivation

The motivation for this research arises from the critical intersection of public health and educational infrastructure. In Dhaka, one of the most densely populated cities in the world, rapid urbanization has resulted in severe degradation of air quality, with particulate matter concentrations frequently exceeding World Health Organization (WHO) guidelines. This environmental burden is particularly pronounced within educational institutions, where students spend a substantial portion of their daily lives.

Classrooms in Lower-Middle-Income Countries (LMICs) often lack adequate ventilation and rely primarily on passive airflow. Such ventilation pathways are frequently restricted to reduce external noise or dust intrusion. As a result, indoor accumulation of carbon dioxide (CO_2) and volatile organic compounds (VOCs) creates a silent crisis that manifests as reduced attention span, increased absenteeism, and long-term respiratory health risks. This situation forms a detrimental feedback loop in which degraded air quality undermines academic performance, while the infrastructure required to monitor and mitigate these conditions remains largely absent.

Although commercial indoor air quality monitoring solutions are available, they fail to address the contextual constraints of resource-limited educational environments due to three fundamental limitations.

The Economic Gap: Commercial, industrial-grade air quality monitors, such as the AirVisual Pro, typically cost in excess of USD 450. For public schools operating under severe budgetary constraints, large-scale deployment of such systems across multiple classrooms is financially impractical.

The Latency and Privacy Gap: The majority of contemporary Internet of Things (IoT) solutions rely heavily on cloud-based data processing. This dependence introduces communication latency, assumes stable internet connectivity that is often unreliable in LMIC contexts, and raises significant data privacy concerns, particularly when environmental data collection occurs in spaces occupied by minors.

The Action Gap: Most critically, existing systems are predominantly passive in nature. While they provide visualization and alerts, they do not initiate corrective action. Informing educators that indoor air quality is unhealthy places the responsibility for mitigation entirely on human intervention. In the absence of automated and intelligent control mechanisms, such as Demand-Controlled Ventilation (DCV), data awareness alone does not translate into measurable health or cognitive benefits.

Therefore, there exists an urgent need for an autonomous, offline-capable system that bridges these gaps by delivering high-precision environmental intelligence at a fraction of the cost of existing commercial solutions.

1.3 Research Objectives

The primary objective of this thesis is to engineer a unified, low-cost Environmental Intelligence (EI) system capable of autonomous IAQ management. By moving beyond passive monitoring to active control, this research aims to bridge the gap between sensing and mitigation in resource-constrained educational settings.

The specific research goals are:

To Design and Validate Low-Cost Sensing Hardware: To engineer a robust sensing node (approx. USD 90) capable of measuring $PM_{2.5}$, CO_2, temperature, and humidity with accuracy comparable to industrial reference devices (e.g., AirVisual Pro).

To Optimize TinyML Models for Edge Deployment: To develop a lightweight forecasting pipeline that transitions from cloud-based training to on-device inference (Edge Computing), enabling the ESP32 microcontroller to predict pollutant spikes without internet dependency.

To Engineer a Demand-Controlled Ventilation (DCV) System: To implement an automated actuation mechanism that regulates classroom airflow based on a "Dual-Trigger" logic—combining real-time threshold breaches with predictive forecasting.

To Evaluate Longitudinal System Performance: To assess the system's efficacy in a live classroom environment over a 12-month period (September 2024 – October 2025), specifically quantifying improvements in air quality index (AQI) and energy efficiency compared to reactive systems.

1.4 Research Questions

To achieve the aforementioned objectives, this thesis addresses the following core research questions (RQs):

RQ1: Can a low-cost, custom-engineered sensor node achieve sufficient data reliability and calibration accuracy to replace expensive commercial monitors in LMIC classrooms?

RQ2: Which machine learning architecture offers the optimal trade-off between forecasting accuracy and computational efficiency when restricted to the hardware constraints of a standard microcontroller (e.g., ESP32)?

RQ3: How effectively does an "Edge-Native" forecasting model reduce system latency and privacy risks compared to traditional cloud-based IoT architectures?

RQ4: To what extent can a predictive Demand-Controlled Ventilation (DCV) system reduce indoor pollution loads compared to manual or purely reactive ventilation strategies?

1.5 Thesis Contributions

This research makes the following key contributions to the field of IoT and Environmental Engineering, specifically addressing the challenges of indoor air quality management in resource-constrained settings:

Longitudinal Dataset Creation: The collection and validation of a high-resolution, multi-pollutant dataset spanning over 12 months (September 2024 – October 2025) from a dynamic classroom environment in Dhaka. This dataset provides rare insight into the seasonal and diurnal variations of indoor pollution in an LMIC setting.

TinyML Optimization Pipeline: A systematic comparative analysis establishing Tree-Based Ensemble methods (Random Forest, LightGBM) as superior to traditional Deep Learning architectures (LSTM) for tabular environmental data, specifically when constrained by memory and power.

Edge-Native Implementation: The successful quantization and deployment of the forecasting model on an ESP32-S3 microcontroller, achieving sub-second inference latency. This eliminates the need for continuous cloud connectivity, ensuring data privacy and system resilience in unstable network conditions.

Proactive Control Logic: The development of a "Dual-Trigger" actuation algorithm that transcends reactive switching. By combining immediate threshold breaches with predictive forecasting, the system preemptively clears pollutants before they reach hazardous levels.

Cost-Effective Scalability: The demonstration of a fully functional Environmental Intelligence system for under $110 (USD). This represents a cost reduction of approximately 75% compared to commercial equivalents (e.g., AirVisual Pro), directly supporting UN SDG-3 (Good Health and Well-being) and SDG-4 (Quality Education) by making safety accessible to under-funded institutions.

1.6 Thesis Organization

The remainder of this thesis is organized to follow the logical development of the system—from theoretical foundations to hardware design, intelligence, and final impact analysis:

Chapter 2: Literature Review provides a comprehensive survey of existing IAQ monitoring technologies, TinyML architectures, and ventilation control strategies, identifying the gaps this research aims to fill.

Chapter 3: System Design & Methodology details the end-to-end hardware architecture, including the design of the low-cost sensor node, calibration procedures, and the data acquisition pipeline.

Chapter 4: Developing Environmental Intelligence focuses on the "Brain" of the system. It covers the model benchmarking process, the selection of tree-based ensembles, and the optimization techniques used to deploy these models onto the edge device.

Chapter 5: Adaptive Control Implementation presents the "Action" phase, describing the Demand-Controlled Ventilation (DCV) mechanism and the implementation of the Dual-Trigger actuation logic.

Chapter 6: Results & Performance Analysis evaluates the system's performance, quantifying forecasting accuracy, edge processing latency, and the physical reduction of pollutants in the classroom.

Chapter 7: Discussion & Sustainability synthesizes the findings to discuss the broader social, economic, and environmental impacts of the work, including an analysis of ethical considerations and long-term sustainability.

Chapter 8: Conclusion & Future Work summarizes the research achievements and outlines potential pathways for future development.

Each chapter logically progresses to the next, providing a comprehensive narrative from the problem definition to the implementation, evaluation, and broader impact of the work.

Chapter 2

Literature Review

2.1 Literature Selection Procedure

To establish a robust theoretical and technical foundation for this multi-disciplinary research, a systematic literature review procedure was engineered to traverse the intersecting domains of Environmental Engineering, Building Science, Internet of Things (IoT), and Artificial Intelligence (AI). The selection procedure was not merely a passive collection of texts but an active, analytical process designed to identify the convergence points where low-cost hardware, advanced control theory, and machine learning could address the specific deficits of Indoor Air Quality (IAQ) management in resource-constrained settings.

The procedure commenced with a broad-spectrum exploratory search intended to map the global landscape of classroom air quality research, differentiating between the established standards of the Global North (e.g., ASHRAE 62.1, EN 16798) and the emerging, often fragmented research emerging from the Global South. This initial phase highlighted a critical disparity: while the theoretical mechanisms of pollution exposure are universal, the technological interventions documented in the literature are overwhelmingly biased towards high-cost, centralized HVAC solutions that are inapplicable to the target demographic of this thesis. Consequently, the selection protocol was refined to specifically target studies that bridged this gap—seeking out "frugal innovation" in sensor networks, decentralized control strategies, and edge-computing methodologies.

Following the identification of the research gap, a structured multi-stage screening process was implemented. The first stage involved the aggregation of literature from high-impact repositories, focusing on peer-reviewed journal articles, conference proceedings, and technical standards published within the last decade (2015–2025). This temporal constraint was imposed to ensure that the review captured the rapid evolution of sensor technologies (such as the miniaturization of NDIR and laser-scattering modules) and the explosive growth of the TinyML field, which was virtually non-existent prior to 2018.

The second stage of selection involved a qualitative assessment of methodological rigor. Studies were evaluated based on their deployment realism; preference was strictly given to field experiments and longitudinal deployments in occupied buildings over purely simulation-based studies or controlled chamber experiments. This filter ensured that the literature base reflected the stochastic realities of sensor drift, network latency, and unpredictable occupant behavior that define real-world engineering challenges. Furthermore, a specific effort was made to include "grey literature" such as white papers from environmental agencies (EPA, WHO) and technical documentation from hardware manufacturers (Espressif, Bosch) to ground the academic theory in industrial reality.

Finally, the selected literature was subjected to a thematic synthesis, categorizing works not just by their discipline but by their functional contribution to the "Environmental Intelligence" pipeline: Sensing, Analysis, Forecasting, and Actuation. This structured selection procedure ensures that the subsequent review provides a holistic view of the ecosystem, justifying the architectural choices made in the development of the DCV systems.

2.2 Keywords Searched

To ensure a comprehensive retrieval of relevant literature across the distinct domains integrated within this thesis, a taxonomy of search terms was developed. These keywords were organized into conceptual clusters to facilitate complex Boolean queries, ensuring that the intersection of fields—for example, the application of machine learning specifically for ventilation control—was thoroughly explored.

- **Cluster 1: Environmental Engineering & Hygiene:** 'Indoor Air Quality (IAQ)', 'Classroom Environment', 'Particulate Matter ($PM_{2.5}$)', 'Carbon Dioxide (CO_2) exposure', 'Sick Building Syndrome (SBS)', 'Demand-Controlled Ventilation (DCV)', 'Natural Ventilation', 'Infiltration rates', 'Occupancy health impact', 'Cognitive performance in schools'.

- **Cluster 2: Machine Learning & Forecasting:** 'Time-series forecasting', 'Multivariate regression', 'Long Short-Term Memory (LSTM)', 'Random Forest regression', 'XGBoost', 'Prophet algorithm', 'Ensemble learning', 'Hyperparameter tuning', 'Deep Learning for environmental data', 'Occupancy prediction'.

- **Cluster 3: IoT & Embedded Systems:** 'Internet of Things (IoT)', 'Wireless Sensor Networks (WSN)', 'Low-cost sensors', 'ESP32 microcontroller', 'Edge computing', 'TinyML', 'On-device inference', 'Model quantization', 'Resource-constrained computing', 'Sensor calibration'.

- **Cluster 4: Contextual & Socio-Economic Factors:** ‘Frugal innovation’, ‘LMIC infrastructure’, ‘Sustainable building technologies’, ‘Energy efficiency in schools’, ‘Digital divide in environmental monitoring’.

Combinations such as (“LSTM” AND “Indoor Air Quality”), (“ESP32” AND “Machine Learning”), and (“Low-cost sensors” AND “Calibration”) were utilized to drill down into niche intersections.

2.3 Databases for Search

The literature search was conducted across a diverse array of academic and technical databases to ensure a balanced representation of theoretical depth and practical application.

- **IEEE Xplore:** Served as the primary source for literature on hardware architectures, sensor instrumentation, signal processing, and embedded systems engineering. This database was crucial for understanding the state-of-the-art in IoT node design and TinyML implementation.
- **ScienceDirect (Elsevier):** Utilized extensively for its comprehensive coverage of Building and Environment, Energy and Buildings, and Atmospheric Environment journals. These sources provided the bulk of the literature regarding ventilation strategies, pollutant dynamics, and the physiological impacts of poor air quality.
- **ACM Digital Library:** Consulted for advancements in algorithmic efficiency, human-computer interaction, and edge computing frameworks, providing insights into the software engineering aspects of the project.
- **PubMed / NIH:** Accessed to retrieve epidemiological data and public health studies linking specific pollutant concentrations to health outcomes, thereby establishing the biomedical motivation for the engineering interventions proposed.
- **Google Scholar & arXiv:** Leveraged to identify preprints and recent conference papers that represent the cutting edge of rapidly evolving fields like TinyML, ensuring the thesis acknowledges the most current developments even if they are yet to appear in traditional journals.
- **Mendeley Data & Kaggle:** Explored to analyze existing open datasets, allowing for a comparative assessment of the data quality and resolution achieved in this study versus the prevailing standards.

2.4 Inclusion/Exclusion Criteria

To maintain the scope and relevance of the review, a strict set of inclusion and exclusion criteria was applied to the search results.

Inclusion Criteria:

- **Temporal Relevance:** Priority was given to research published between 2015 and 2025. This window captures the proliferation of low-cost optical particle counters and the emergence of the TinyML paradigm.
- **Domain Specificity:** Studies must explicitly address indoor environments. While ambient (outdoor) air quality literature provides context, the distinct physics of indoor accumulation and decay required a focus on enclosed spaces.
- **Deployment Reality:** Preference was given to studies involving physical prototypes and real-world data collection over purely theoretical simulations.
- **Contextual Alignment:** Research focused on educational settings, high-density public spaces, or resource-constrained environments was prioritized to align with the thesis's problem statement.
- **Technological Alignment:** Studies utilizing open-source hardware (Arduino, ESP32) and Python-based ML frameworks were favored to ensure comparability with the proposed methodology.

Exclusion Criteria:

- **Legacy Technologies:** Research focusing on outdated ventilation controls (e.g., pneumatic thermostats) or obsolete sensor technologies was excluded.
- **High-Cost Solutions:** Approaches requiring industrial-grade Building Management Systems (BMS) or reference-grade analyzers (costing >USD 5,000) were excluded, as they do not address the needs of the target demographic.
- **Cloud-Dependency:** While cloud-based systems were reviewed for context, studies that relied exclusively on heavy cloud computing for inference were deprioritized in favor of edge-centric approaches, aligning with the thesis's focus on resilience and privacy.
- **Language:** Only English-language publications were included to ensure accessibility and reviewability.

2.5 Filtering Results

The initial search yielded a substantial corpus of literature. A systematic filtering cascade was employed to distill this down to the most pertinent works:

- **Duplicate Removal:** Automated tools were used to remove duplicate entries across different databases.
- **Title and Abstract Screening:** Entries were screened for relevance. Papers focusing on industrial chemical processes or outdoor pollution modeling without indoor application were discarded.
- **Methodological Quality Assessment:** Full texts were reviewed to assess the rigor of the experimental design. Studies with undefined sensor calibration protocols or insufficient data collection periods ($<$ 1 week) were excluded to ensure high standards of evidence.
- **Thematic Clustering:** The remaining papers were grouped into "Sensing," "Control," and "Forecasting" clusters. This facilitated the identification of papers that bridged these clusters—highly valuable "integrative" studies that mirrored the thesis's holistic approach.
- **Citation Snowballing:** The reference lists of the final selection were scanned to ensure that seminal works and foundational theories (e.g., the Wells-Riley equation for airborne transmission) were included.

2.6 Literature Review of Papers

2.6.1 Classroom Indoor Air Quality and Health Impacts

Studies of school environments show that poor indoor air quality (IAQ) can significantly impair students' cognitive performance and health. For example, elevated indoor CO_2 levels have been linked to measurable declines in attention and decision-making in controlled studies [1, 2]. Public health agencies have documented the physiological risks associated with prolonged CO_2 exposure, establishing evidence-based thresholds for safe indoor environments [3]. Fan *et al.* (2023) report a meta-analysis finding that short-term CO_2 exposure degrades performance on cognitive tasks [2]. Observational work similarly finds that lower IAQ correlates with worse academic outcomes [4]. Comprehensive reviews of school IAQ emphasize that inadequate ventilation and pollutant exposures (CO_2, VOCs, etc.) are common in classrooms and warrant mitigation [5, 6].

Key pollutant sources in classrooms include occupant-generated CO_2 and fine particulates from activities. Lin *et al.* (2015) found that routine chalk use generates respirable dust that degrades air quality [7]. More broadly, studies in polluted regions have measured $PM_{2.5}$ and PM_{10} well above health benchmarks in schools, with adverse respiratory impacts on children [8, 9]. For instance, Were *et al.* (2020) documented that high PM exposures in Kenyan schools correlate with respiratory symptoms among students [8]. International guidelines (WHO, EPA) set stringent limits on indoor PM and recommend adequate ventilation, but many classrooms exceed these levels [9, 10]. Notably, field surveys reveal a common gap between perception and reality: teachers often underestimate CO_2 and VOC concentrations compared to sensor measurements [11].

The IAQ challenge is especially acute in low- and middle-income countries. Urban ambient pollution can infiltrate schools and overwhelm limited ventilation systems. For example, Bangladesh experiences rapidly worsening air pollution [12], and measurements in Dhaka classrooms reveal CO_2 and PM levels far above recommended levels during lectures [13]. Field studies in Serbian schools have shown similar issues with poor ventilation rates leading to excessive CO_2 accumulation [14]. Similarly, Nyembwe *et al.* (2025) report chronically elevated CO_2 in primary school classrooms in the Democratic Republic of Congo, driven by overcrowding and inadequate air exchange [15]. To address these challenges, recent work has focused on developing decision tools for schools using continuous IAQ monitors to translate real-time data into actionable ventilation strategies [16]. Such studies highlight the urgent need for affordable IAQ monitoring and control in resource-constrained schools [12, 15].

2.6.2 Evolution of Sensing Architectures

Advances in inexpensive sensors and IoT hardware have enabled distributed IAQ monitoring in classrooms. Many implementations use microcontroller platforms. For example, Kelechi *et al.* (2021, 2022) designed low-cost Arduino-based IAQ monitors that measure CO_2, PM, and VOCs and stream data to cloud dashboards [17, 18]. Reviews of sensor technology note that modern low-cost NDIR CO_2 sensors and optical particle counters can provide reasonable accuracy for indoor applications [19]. Innovative sensor platforms have also been explored: Chaoraingern *et al.* (2023) repurposed a vacuum-cleaner-mounted electronic nose to predict real-time indoor air quality indices [20]. High-precision e-nose devices have achieved good performance in residential IAQ monitoring [21]. These examples show that even unconventional or DIY platforms can contribute to classroom IAQ sensing.

Field studies underscore both the promise and challenges of low-cost sensing. Canha *et al.* (2024) found that teachers' perceptions of classroom IAQ often underpredict actual pollutant levels measured by sensors [11]. Real-world IoT deployments demonstrate feasibility: multiple studies have developed wireless sensor network (WSN) platforms specifically for classroom IAQ monitoring [22, 23, 24]. Bekkar *et al.* (2024) describe a Southwestern Morocco study where low-cost IoT sensors and machine learning were combined into a real-time air quality monitoring platform [25]. Smart building concepts have emerged, such as Zivelonghi and Giuseppi's (2024) IoT-enabled framework for multi-room dynamic air quality control in schools [26]. Architecturally, designers balance cloud vs. edge processing. Andriulo *et al.* (2024) review edge-versus-cloud computing for IoT systems [27], and surveys on edge computing for IoT highlight the advantages of local processing for reduced latency and enhanced privacy [28, 29]. Qian and Coutinho (2022) show that offloading to edge servers can reduce latency and energy in AR/IoT applications [30]. In the context of TinyML, frameworks like TensorFlow Lite Micro enable running neural-network inference on microcontrollers [31]. Schizas *et al.* (2022) provide a comprehensive review of TinyML for ultra-low power AI and large-scale IoT deployments, identifying key challenges and opportunities [32]. Prakash *et al.* (2023) assess the environmental sustainability of TinyML systems, finding that on-device inference can significantly reduce the carbon footprint compared to cloud-based approaches [33]. Recent benchmarks (Idri and Hamdouchi, 2025) evaluate TinyML classifiers and ensemble models on IoT network data [34], indicating that compact tree ensembles and deep models can perform under tight resource constraints. Together, these advances point to the possibility of embedding IAQ analytics directly in low-cost classroom devices.

2.6.3 Machine Learning-Based Forecasting Models for IAQ

Machine learning (ML) methods are increasingly used to predict indoor pollutant levels, enabling proactive IAQ management. Recurrent neural networks, especially long short-term memory (LSTM) models, are popular for time-series forecasting due to their ability to capture temporal dependencies [35]. Deep RNN architectures have demonstrated strong performance for air quality classification tasks [36]. For example, Zhu *et al.* (2022, 2023) implemented LSTM-based models on IoT platforms to forecast steady-state indoor CO_2 concentrations [37, 38]. Wei *et al.* (2023) developed an LSTM-autoencoder approach for anomaly detection in IAQ time-series data [39]. Yang *et al.* (2022) developed a hybrid model combining Bayesian optimization, empirical mode decomposition, and LSTM to predict long-term classroom CO_2 levels [40]. Other deep learning approaches have been applied to IAQ data: a graph convolutional LSTM model achieved strong performance for spatiotemporal $PM_{2.5}$ forecasting [41]. Belavadi *et al.* (2020) combined LSTM RNN with wireless sensor networks for air quality forecasting [42]. Similarly, Miao *et al.* (2023) built data-driven models to predict IAQ and thermal comfort in naturally ventilated school buildings using readily available data [43]. Yao *et al.* (2024) assessed and predicted indoor environmental quality in naturally ventilated residential dwellings using ML techniques [44]. Yin *et al.* (2025) proposed a novel IAQ prediction pipeline using meta-learning with STL decomposition [45]. Even in office environments, time-series ML has been shown to accurately forecast CO_2 [46]. Ansell (2024) explored forecasting indoor air pollution levels using time series modeling techniques [47]. These studies demonstrate the flexibility of deep learning pipelines for IAQ prediction.

Tree-based ensemble methods and traditional time-series models have also been applied. Random Forests [48] are a robust ensemble method that has received comprehensive theoretical treatment in the literature [49], and Saharuna *et al.* (2024) successfully used an RF regressor to predict minute-by-minute indoor CO_2 [50]. Extremely randomized trees represent an alternative ensemble approach with reduced variance [51]. Adaptive and hybrid statistical models have been proposed as well: Segala *et al.* (2021) introduced a practical adaptive model for indoor CO_2 prediction [52]. Facebook's Prophet model, which implements seasonal decomposition with regression at scale [53], has been applied to air pollutant forecasting, such as by Hasnain *et al.* (2022) for urban pollutant series [54]. Márquez-Zepeda *et al.* (2025) developed an IoT-based CO_2 monitoring and forecasting system for indoor air quality management [55]. Comparative studies show that complex ML methods often outperform baselines: Abuouelezz *et al.* (2025) compared multiple ML algorithms (including RF, LSTM, etc.) for $PM_{2.5}/PM_{10}$ forecasting in the UAE and found that ensemble tree methods generally achieved the best accuracy [56]. Machine learning approaches for air quality classification have also been evaluated systematically [57]. Similar work by Ahn *et al.* (2023) used an ensemble of gradient boosting machines and CNN-LSTM to forecast harmful algal blooms, illustrating the power of combined models in environmental time series [58]. Hybrid LSTM-LightGBM architectures have also been explored for greenhouse microclimate forecasting [59]. These results collectively suggest that both neural and ensemble methods can be effective for predicting IAQ variables.

Forecasting models have begun to be integrated into smart IAQ control and monitoring systems. For example, Taheri and Razban (2021) embedded a learning-based CO_2 predictor into a demand-controlled ventilation scheme [60]. Occupancy predictions (Elkhoukhi *et al.*, 2019) have also been used to drive energy-efficient ventilation controls [61]. In practical deployments, IoT sensing platforms often incorporate onboard analytics: Bekkar *et al.* (2024) demonstrate a real-time AIoT system that continuously monitors and predicts air quality metrics [25]. Similarly, ML-enabled monitoring is used in other domains—for instance, Goh *et al.* (2021) and Sukor *et al.* (2022) applied ML models to predict in-vehicle CO_2 and pollutant concentrations [62, 63]. The key enabler for on-device forecasting is TinyML: TensorFlow Lite Micro allows complex models to run on microcontrollers [31]. Recent benchmarks (Idri and Hamdouchi, 2025) show that even ensemble models can be optimized to meet the real-time, low-power constraints of embedded IoT devices [34]. Embedding ML forecasting in low-cost classroom monitors would thus enable anticipatory ventilation control without dependence on cloud connectivity.

Together, these lines of work underscore the feasibility and importance of embedded

IAQ intelligence. Affordable sensor networks and TinyML models can provide real-time forecasts of CO_2 and PM levels, which in turn can be used to drive automated ventilation or alert systems. This literature motivates the present thesis's focus on developing a TinyML-based, on-device IAQ forecasting and control platform for classrooms, bridging low-cost sensing with machine learning to improve indoor environments in resource-limited school settings.

2.6.4 Ventilation and Control Strategies

Proper ventilation is critical to diluting indoor pollutants. A common approach is demand-controlled ventilation (DCV), which adjusts fresh-air intake based on CO_2 or other IAQ sensors. Emmerich and Persily (2001) review the state of CO_2-based DCV technologies [64], and Lu *et al.* (2022) provide an updated review of control strategies and performance for CO_2-driven ventilation [65]. Recent work extends these ideas with machine learning: Taheri and Razban (2021) use learning algorithms to predict CO_2 levels for DCV control [60]. Predictive control strategies have also been proposed to jointly optimize multiple pollutants. For instance, Choi *et al.* (2025) develop a model-predictive control scheme to simultaneously regulate CO_2 and $PM_{2.5}$ in indoor environments [66]. Such approaches suggest that data-driven forecasting can be integrated into smart ventilation systems.

Guidelines and standards provide targets for classroom ventilation. The U.S. EPA's recent school IAQ reference guide summarizes recommended practices for schools [67], and WHO's 2021 air quality guidelines specify safe levels for PM and other pollutants [9]. However, implementation gaps remain. Persily (2015) discusses the challenges in translating ventilation standards into practice [68], and Chan *et al.* (2019) found that many retrofitted classrooms in California still fall short of ASHRAE ventilation targets [69]. In response to airborne disease concerns, experts have also emphasized ventilation to reduce COVID-19 transmission indoors [70]. A National Academies workshop highlights the health effects of indoor particulate exposure, underscoring the need for mitigation in schools [71]. These findings point to a persistent gap between recommended ventilation rates and actual classroom conditions.

Empirical studies illustrate practical ventilation issues in schools. Muelas *et al.* (2022) evaluated different natural ventilation modes (window/door openings) and measured CO_2 gradients in a classroom; they found that open-window strategies produced variable effectiveness and non-uniform CO_2 distribution [72]. Stabile *et al.* (2017) analyzed the effect of different natural ventilation strategies on indoor air quality in schools [73]. In mechanically ventilated schools, Yu *et al.* (2014) demonstrated that installing an HVAC system markedly reduced indoor CO_2 and PM levels compared to unventilated classrooms [74]. Dhanalakshmi *et al.* (2020) proposed an IoT-based system for indoor air quality monitoring and smart energy management for HVAC systems [75]. Ventilation interventions have also targeted moisture problems: Vornanen-Winqvist *et al.* (2018) used positive-pressure ventilation to improve IAQ in a damp school building [76]. A field study of 30 London primary classrooms by Hama *et al.* (2023) identified key factors (building design, occupancy, window use) that underlie poor IAQ and thermal comfort [77]. Taken together, these works illustrate both the potential and the complexity of ventilation strategies in real school settings.

2.7 Summary of Research Gaps

Despite the extensive body of work reviewed above, significant discontinuities remain in the current literature, particularly when applied to the context of educational infrastructure in the Global South. The key research gaps identified are:

- **The "Sensing-Action" Disconnect:** Most low-cost sensing studies [41, 43] focus purely on monitoring and visualization. There is a lack of integrated systems that close the loop by using sensor data to directly drive mechanical ventilation in real-time.
- **The Cloud Dependency:** Existing predictive IAQ models [62, 64] predominantly rely on heavy cloud computing resources. There is limited research on "Edge-Native" forecasting where complex regression models are optimized for bare-metal microcontrollers (TinyML) to ensure privacy and offline resilience.
- **The Lag-Blind Control Logic:** Standard Demand-Controlled Ventilation (DCV) strategies rely heavily on CO_2 thresholds [17]. Few studies account for the temporal lag between occupancy (rising CO_2) and particulate resuspension (rising $PM_{2.5}$), leading to reactive rather than preemptive control.

This thesis aims to address these specific gaps by engineering a unified framework that combines robust sensing, edge-native intelligence, and predictive control.

Chapter 3

System Design & Methodology

This chapter presents the comprehensive engineering framework developed to address the challenges of Indoor Air Quality (IAQ) management in resource-constrained educational settings. Moving beyond isolated experiments, the system is designed as a unified, end-to-end pipeline that bridges the gap between low-cost hardware and advanced environmental intelligence. The following sections detail the architectural design of the sensing node, the robust data acquisition methodology spanning over twelve months of operation, and the rigorous validation procedures undertaken to ensure that low-cost sensors could perform with accuracy comparable to industrial reference standards.

3.1 High-Level System Architecture

To address the limitations of existing commercial solutions specifically high cost, cloud dependency, and lack of active control, this thesis proposes a holistic "Environmental Intelligence" framework. The system architecture is designed as a closed-loop pipeline consisting of three functional layers: Sensing, Intelligence, and Actuation.

The architecture, illustrated in Figure 3.1, operates as follows:

1. **The Sensing Layer:** A custom-engineered sensor node continuously monitors critical IAQ parameters ($PM_{2.5}$, CO_2, Temperature, Humidity, and Barometric Pressure) at the edge.

2. **The Intelligence Layer:** Data is processed via a hybrid approach. A lightweight "Edge-Native" forecasting engine (running on an ESP32-S3) predicts future pollutant trends locally, while a secondary pipeline transmits data to a cloud backend (ThingSpeak/SQLite) for long-term storage and visualization.

3. **The Actuation Layer:** A Demand-Controlled Ventilation (DCV) mechanism utilizes a "Dual-Trigger" heuristic to physically regulate the classroom environment.

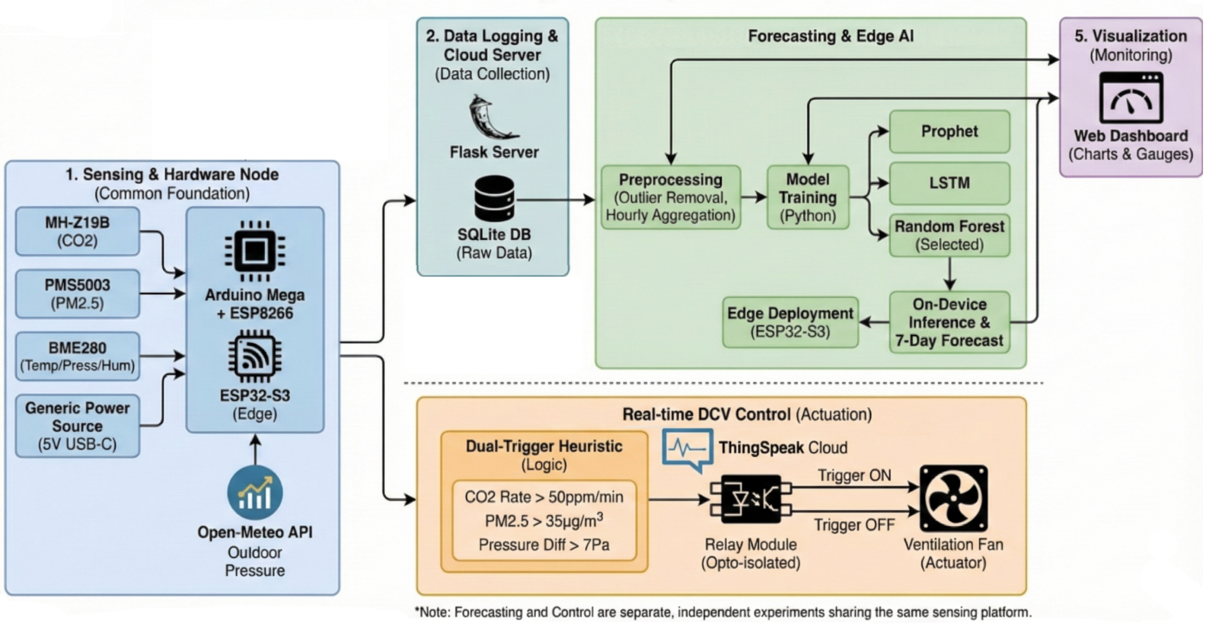


Figure 3.1: Unified system architecture: Sensing, Intelligence, and Actuation layers.

A relay module activates an exhaust system when real-time thresholds or predicted spikes are detected, effectively closing the loop between observation and mitigation.

3.2 Hardware Development

The primary engineering challenge was to develop a robust sensing node capable of industrial-grade accuracy at a price point accessible to LMIC institutions. The hardware design prioritized modularity, local repairability, and the use of off-the-shelf components to ensure scalability.

3.2.1 Component Selection

The sensing core is built around the Arduino Mega 2560 microcontroller, selected for its ample I/O capacity to handle multiple UART and I2C sensor streams simultaneously. Connectivity is provided by an onboard ESP8266 Wi-Fi module, enabling telemetry upload without requiring a separate gateway.

The sensor suite was selected based on a balance of cost, lifespan, and documented accuracy in peer literature:

- **Particulate Matter:** The PMS5003 (laser scattering) sensor was chosen for its effective range of 0–500 $\mu g/m^3$ and rapid response time ($<$ 10 seconds).
- **Carbon Dioxide:** The MH-Z19B (Non-Dispersive Infrared / NDIR) sensor was selected for its 0–5000 ppm range and internal automatic baseline calibration (ABC) logic, which is critical for long-term unattended operation.

- **Environmental Context:** A BME280 sensor provides compensation data (Temperature, Humidity, and Barometric Pressure), allowing the system to calculate air density and pressure differentials essential for the infiltration-aware control logic.

3.2.2 Circuit Design and Fabrication

To transition from a breadboard prototype to a deployable device, a custom Printed Circuit Board (PCB) was designed. The PCB integrates the microcontroller, sensor interfaces, and a dual opto-isolated relay module. The relay provides the physical interface for actuation, allowing the low-voltage logic (5V) to safely switch high-voltage (220V) ventilation fans.

The entire assembly is housed in a laser-cut PVC enclosure designed with specific airflow dynamics. A dedicated 50mm 5V blower fan ensures active air sampling by pulling room air across the sensor inlets, preventing heat buildup from the MCU from affecting temperature readings.

3.2.3 Economic Feasibility and Cost Analysis

A defining objective of this research is economic viability. Table 3.1 presents the Bill of Materials (BOM) for the fully featured DCV-capable node.

Table 3.1: Component-level cost breakdown for a single DCV-capable IAQ node.

Component	Cost (USD)
NDIR CO_2 sensor (MH-Z19B)	\$23.85
Optical $PM_{2.5}$ sensor (PMS5003)	\$22.95
Fermion Temp/Hum & Air pressure sensor (BMP280)	\$5.90
Custom PCB board	\$4.00
Microcontroller (Arduino Mega 2560 with ESP8266 WiFi)	\$18.25
Relay module (opto-isolated)	\$1.40
Blower fan (50 mm, 5 V)	\$1.95
Enclosure (laser-cut PVC) and mounting	\$8.50
Wiring and miscellaneous	\$5.00
Total (per DCV-capable node)	**\$91.80**

Cost-Benefit Comparison

The total material cost of \$91.80 represents a stark contrast to industry-standard reference devices such as the IQAir AirVisual Pro, which retails for approximately \$450.

- **Cost Reduction:** Our system achieves a $\sim$80% reduction in capital expenditure compared to commercial equivalents.

- **Functional Advantage:** Unlike the AirVisual Pro, which is a passive monitor, the proposed node includes active actuation hardware (relays) within this lower price envelope.
- **Edge-AI Scalability:** For the advanced edge-forecasting prototype (discussed in Chapter 4), an ESP32-S3 development board was added. This incurs a marginal incremental cost of ∼$12, keeping the total system cost well below $110 while enabling capabilities (on-device ML) typically reserved for server-grade BMS systems.

This economic analysis confirms that the system is not only technically capable but scalable for widespread deployment in resource-constrained public schools.

3.3 Data Acquisition Pipeline

To support the development of data-driven forecasting models and adaptive control logic, a robust data acquisition infrastructure was required. The pipeline was designed to handle high-frequency logging, manage network instability common in LMIC infrastructure, and ensure long-term data integrity.

3.3.1 Deployment Environment

The system was deployed in a representative naturally ventilated classroom at Independent University, Bangladesh (IUB), located in the dense urban fabric of Dhaka. The selected room features operable windows and ceiling fans but lacks central HVAC, making it an ideal testbed for studying the stochastic influence of outdoor pollution infiltration and occupant behavior on indoor air quality.

3.3.2 Software Architecture and Backend

The data pipeline utilizes a lightweight, open-source stack designed for local reproducibility:

- **Edge Transmission:** The ESP8266 module on the sensor node transmits JSON-formatted payloads via HTTP POST requests over the university Wi-Fi network.
- **Server Application:** A custom Python Flask server receives these requests. It performs immediate schema validation to reject malformed packets before storage.
- **Storage Layer:** Validated data is committed to a SQLite database. This serverless, file-based database engine was chosen for its reliability and ease of backup, eliminating the overhead of maintaining a dedicated database server. While alternative database architectures exist for IoT time-series data [78, 79], SQLite's

simplicity and zero-configuration approach aligned well with the constraints of this deployment.

- **Cloud Redundancy:** For remote monitoring and redundancy, the system simultaneously pushes a subsampled stream (1-minute intervals) to the ThingSpeak cloud platform.

3.3.3 Data Collection Periods

Data collection was conducted in two distinct phases to serve different research objectives. To ensure the robustness of the system, the hardware was operated continuously for over a year.

- **Total Longitudinal Dataset:** Data was logged continuously from 28 September 2024 to 22 October 2025 (more than 1 year). This extensive record captures seasonal variations, from the humid monsoon season to the dry, dust-heavy winter, providing a comprehensive profile of Dhaka's indoor environment.
- **Experimental Dataset (Forecasting and Control):** For the specific training of machine learning models (Chapter 4) and the evaluation of control heuristics (Chapter 5), a high-quality subset spanning 6 October 2024 to 3 April 2025 (approximately 6 months) was utilized. This period was selected for its high continuity and minimal sensor downtime, comprising over 1.47 million raw data packets sampled at 10-second intervals.

3.3.4 Preprocessing and Quality Control

Raw sensor data from low-cost nodes inherently contains noise and gaps. A rigorous preprocessing workflow was applied:

- **Range Checks:** Physically impossible values (e.g., $CO_2 < 400$ ppm or negative PM concentrations) were filtered out.
- **Outlier Detection:** Transient spikes caused by sensor electrical noise were identified using a rolling Z-score threshold and removed.
- **Imputation:** Short connectivity dropouts (less than 30 minutes) were filled using linear interpolation to preserve temporal continuity for time-series modeling.
- **Aggregation:** The 10-second raw stream was resampled into hourly and half-hourly averages to align with the forecasting horizons required for the TinyML experiments.

3.4 Sensor Validation

A critical skepticism regarding low-cost IoT devices is data reliability. Extensive research has examined sensor calibration methodologies for low-cost air quality monitors, highlighting both the potential and challenges of such devices [80, 81, 82, 83]. Intelligent calibration approaches using machine learning have shown promise in improving the accuracy of integrated low-cost sensor systems [84]. To validate the custom-engineered node in this study, a rigorous co-location study was conducted prior to the main deployment.

3.4.1 Methodology

The custom node was co-located with an IQAir AirVisual Pro, an industry-standard laser particle counter and NDIR monitor used globally for air quality index (AQI) reporting. Both devices were placed side-by-side at a height of 1.2 meters (breathing zone) in a controlled indoor environment for a continuous 48-hour period. Data was logged synchronously, and performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2).

3.4.2 Comparative Performance Results

The results demonstrated a high degree of correlation between the low-cost node (approximately $92) and the reference device (approximately $450), validating the suitability of the hardware for academic research. Table 3.2 summarizes the performance metrics obtained during the validation phase.

Table 3.2: Performance comparison of the proposed IAQ Node vs. AirVisual Pro Reference Device.

Parameter	MAE	RMSE	R^2 (Correlation)	MBE
$PM_{2.5}$ ($\mu g/m^3$)	2.18	3.42	0.97	+0.65
CO_2 (ppm)	38.5	52.7	0.93	-22.1
Temperature (°C)	0.38	0.52	0.99	+0.12
Humidity (%)	2.41	3.16	0.96	-1.3

3.4.3 Discussion of Accuracy

Particulate Matter: The R^2 of 0.97 for $PM_{2.5}$ indicates that the PMS5003 sensor tracks the reference device with exceptional linearity. The low Mean Bias Error (+0.65 $\mu g/m^3$) suggests that the sensor requires minimal calibration offset for standard indoor concentrations.

Carbon Dioxide: The CO_2 sensor showed a strong correlation ($R^2 = 0.93$). Although there was a slight negative bias (-22.1 ppm), this error margin is well within the acceptable range for ventilation control applications, where the difference between 800 ppm and 1200 ppm is the primary concern.

These validation results confirm that the low-cost design philosophy did not compromise data quality, establishing a solid foundation for the machine learning and control experiments presented in subsequent chapters.

Chapter 4

Developing Environmental Intelligence

Having established the hardware foundation and data acquisition pipeline in Chapter 3, the focus now shifts to the system's "brain": the forecasting engine. Raw sensor data provides a snapshot of the present, but effective IAQ management requires anticipating the future. This chapter details the development of the "Environmental Intelligence" module, executed in two distinct phases. Phase I focuses on a rigorous comparative benchmarking of machine learning architectures—ranging from statistical decomposition (Prophet) to deep learning (LSTM) and ensemble methods (Random Forest)—to identify the optimal modeling strategy for naturally ventilated classrooms. Phase II translates these findings into practice by optimizing the winning architecture for "TinyML" deployment, enabling the low-cost ESP32-S3 node to generate week-ahead forecasts autonomously without reliance on cloud infrastructure.

4.1 The Forecasting Challenge

The transition from passive monitoring to proactive control hinges on the ability to predict pollutant spikes before they reach hazardous levels. However, developing accurate forecasting models for educational settings in Lower-Middle-Income Countries (LMICs) presents a unique set of diverse challenges that conventional IoT frameworks fail to address.

4.1.1 Stochasticity in Naturally Ventilated Spaces

Most existing IAQ forecasting literature focuses on sealed, mechanically ventilated office buildings where airflow is regulated by HVAC systems [66, 70]. In such environments, pollutant dynamics are relatively deterministic. In contrast, the classroom studied in this thesis and thousands like it across Dhaka operates under Natural Ventilation. Indoor air quality is driven by highly stochastic factors: windows opening and closing at random, variable occupancy rates, sudden chalk dust generation, and the unpredictable infiltration of traffic pollution from the outdoors [13, 34]. This introduces significant noise into the time-series data, rendering simple linear extrapolation methods ineffective. The challenge, therefore, is to identify a model capable of capturing these complex, non-linear dependencies without overfitting to transient noise.

4.1.2 The Latency and Privacy Trap

Standard IoT architectures typically offload this heavy computational modeling to the cloud [54]. Sensor nodes stream raw data to a remote server, which processes the prediction and sends a command back. While computationally powerful, this Cloud-First approach is fundamentally flawed for critical infrastructure in LMICs for three reasons:

1. **Latency and Connectivity:** Internet stability in public institutions is often intermittent. A forecasting system that fails whenever the Wi-Fi drops is inherently unreliable for safety-critical ventilation control.

2. **Data Privacy:** Uploading high-resolution CO_2 data to the cloud poses a privacy risk, as carbon dioxide levels serve as a precise proxy for occupancy, effectively allowing external actors to track when a classroom is occupied or empty [60].

3. **Recurring Costs:** Cloud services often incur subscription or egress fees, violating the low-cost design constraint required for scalability in underfunded schools.

4.1.3 The TinyML Imperative

To overcome these barriers, this research adopts a TinyML (Tiny Machine Learning) approach. The objective is to move the intelligence from the cloud to the extreme edge, embedding the forecasting model directly onto the microcontroller (MCU). This creates a strict engineering constraint: the chosen model must not only be accurate but must also be compressible enough to run within the Kilobyte-scale RAM limits of an ESP32, all while delivering results in near real-time. This chapter systematically explores the trade-off between predictive accuracy (Phase I) and computational efficiency (Phase II) to engineer a solution that fits this narrow operational window.

4.2 Phase I: Comparative Benchmarking

Before attempting deployment on constrained hardware, a rigorous offline analysis was conducted to determine the optimal modeling architecture. The objective was to identify which algorithmic family—Statistical Decomposition, Deep Learning, or Ensemble

Learning—offers the best resilience to the stochastic noise inherent in naturally ventilated classrooms.

4.2.1 Experimental Setup

The benchmarking was performed on a standard workstation using the 6-month high-quality dataset subset (October 2024 to April 2025) described in Chapter 3.3.3. To preserve temporal integrity, the data was split chronologically: the first 80% was used for training, followed by 10% for validation and 10% for a held-out test week. This chronological split is critical for time-series evaluation, as standard cross-validation techniques can introduce temporal leakage [85]. The forecasting horizon was set to 7 days (336 half-hourly steps) to simulate a real-world weekly planning cycle for school administrators. The experimental workflow is illustrated in Figure 4.1. This pipeline ensures a fair evaluation by subjecting all three candidate architectures to identical preprocessing and chronological data splitting (Train-Validation-Test) before performance metrics are calculated.

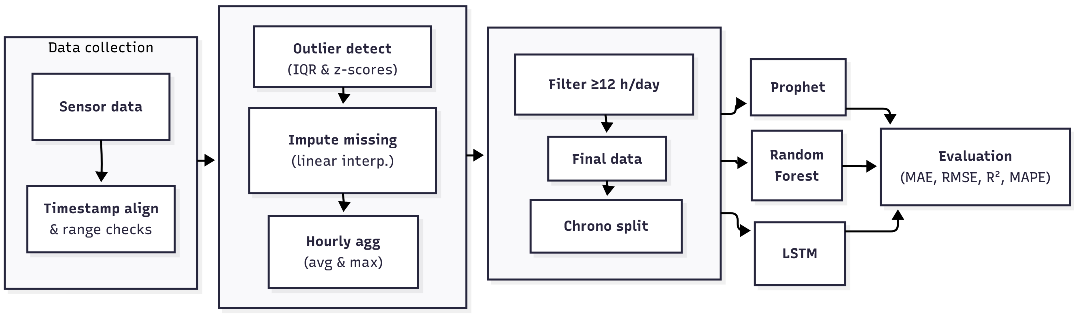


Figure 4.1: Phase I: Comparative benchmarking workflow for Prophet, LSTM, and Random Forest.

4.2.2 Candidate Architectures

Three distinct modeling paradigms were evaluated:

- **Prophet (Baseline):** A decomposable time-series model developed by Meta. It models pollutants as a sum of trends, seasonality (weekly/daily), and holidays. It was selected for its interpretability and ability to handle missing data.
- **Long Short-Term Memory (LSTM):** A Deep Learning architecture specializing in sequence modeling. A sequence-to-sequence LSTM with a 14-day lookback window was implemented to capture complex, non-linear dependencies and long-term lags. Figure 4.2a illustrates the LSTM cell's gated information flow and the

sequence-to-sequence forecast setup. Each series was standardized using training-set statistics, and models were trained with Adam (learning rate 0.001), early stopping, and a grid search over architecture dimensions.

- **Random Forest (RF):** An ensemble learning method constructing multiple decision trees during training. It utilizes a rich feature set (lagged values, rolling averages, cyclical time encodings) to reduce variance and resist overfitting. The common working principles of Random Forest, XGBoost, and LightGBM are illustrated in Figure 4.2b, which shows how decision trees are constructed and combined in ensemble methods.

Performance was evaluated using MAE, RMSE, R^2, and MAPE on the held-out test week. To understand feature importance and model interpretability, SHAP (SHapley Additive exPlanations) values could be computed to provide a unified approach to interpreting model predictions [86], though this was deferred to future work given the focus on deployment efficiency.

4.2.3 Performance Evaluation

The models were evaluated on four targets: Average and Maximum concentrations for both $PM_{2.5}$ and CO_2. Table 4.1 summarizes the performance metrics on the held-out test set.

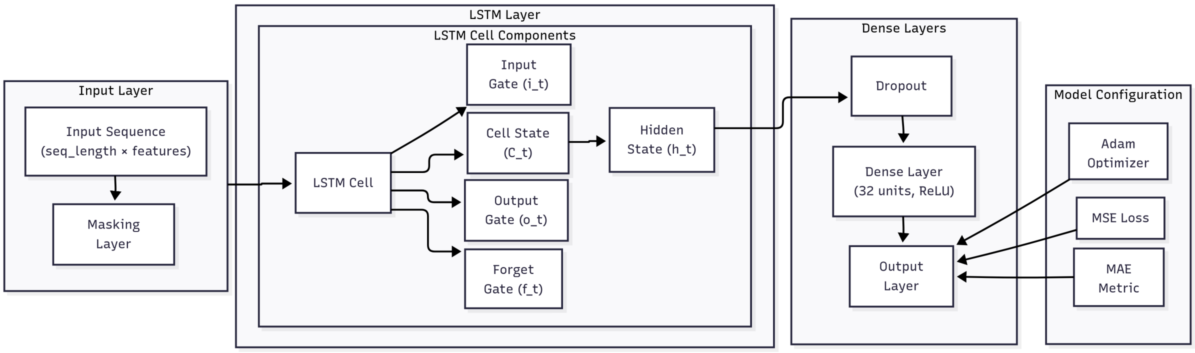
Table 4.1: Comparative Forecasting Error on Test Set (Phase I).

Target	Model	MAE	RMSE	R^2 Score
$PM_{2.5}$ Avg	Prophet	12.94	16.15	0.42
	LSTM	3.91	5.41	0.93
	Random Forest	1.23	1.87	0.99
CO_2 Max	Prophet	345.70	453.73	-1.48
	LSTM	79.15	159.67	0.69
	Random Forest	90.26	167.07	0.66

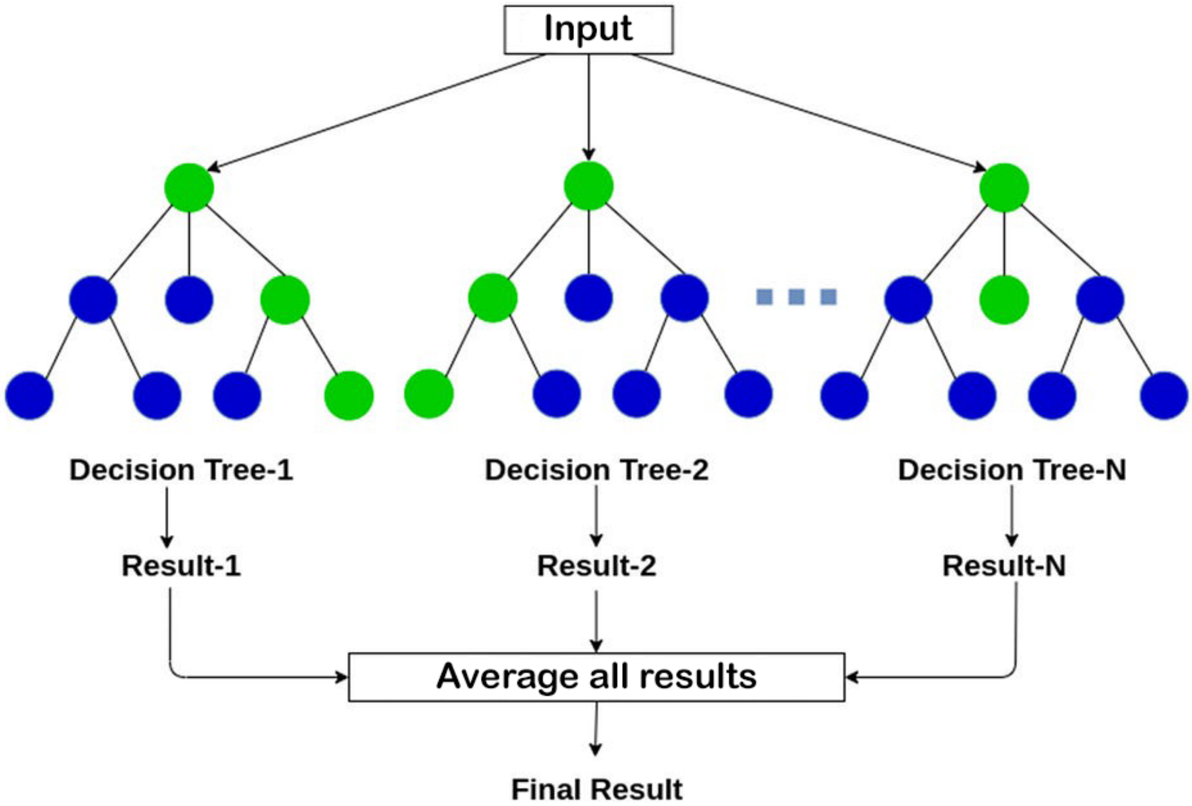
4.2.4 Key Findings

Failure of Decomposition (Prophet): Prophet performed poorly, particularly for CO_2 (negative R^2). This indicates that additive seasonality alone cannot capture the abrupt spikes caused by random occupancy or window operations.

Dominance of Random Forest: For particulate matter ($PM_{2.5}$), Random Forest achieved near-perfect tracking ($R^2 \approx 0.99$), reducing the Root Mean Square Error (RMSE) to just 1.87 $\mu g/m^3$. Its ability to utilize recent lags allowed it to react instantly to sudden dust events.



(a) The gated structure of an LSTM cell and sequence-to-sequence forecasting.



(b) The bootstrap aggregation mechanism of Random Forest, XGBoost, and LightGBM.

Figure 4.2: Model architectures: (a) LSTM cell structure, (b) Random Forest aggregation.

The LSTM Niche: LSTM marginally outperformed Random Forest in predicting Peak CO_2 (R^2 0.69 vs 0.66). The recurrent nature of LSTM makes it slightly better at modeling the smooth accumulation curves of CO_2 in a closed room.

4.2.5 Selection for Edge Deployment

While LSTM offered a slight advantage for peak CO_2 detection, it imposes a heavy computational burden (floating-point matrix multiplications) that is difficult to optimize for low-power microcontrollers.

Random Forest, conversely, demonstrated:

- **Superior Accuracy for the primary health hazard** ($PM_{2.5}$).
- **Algorithmic Simplicity:** It consists of conditional if-else logic (decision trees) which is computationally inexpensive to execute.
- **Efficiency Potential:** Tree structures can be easily serialized into C-code arrays, making them ideal for TinyML conversion.

Therefore, Tree-Based Ensembles (Random Forest and its gradient-boosted variants) were selected as the architecture for the Phase II Edge Deployment.

4.3 Phase II: TinyML Optimization

While Phase I established Tree-Based Ensembles as the superior architecture for accuracy, deploying these models on a microcontroller with limited RAM (ESP32-S3) requires a fundamental shift in engineering strategy. A Random Forest model trained on a PC can easily grow to hundreds of megabytes far exceeding the available memory of an embedded device. Phase II, therefore, focused on the TinyML optimization pipeline: converting heavy Python-based models into lightweight, static C-structures capable of sub-second inference.

4.3.1 The Optimization Pipeline

To bridge the gap between the training environment (Python/Scikit-Learn) and the deployment environment (C++/Arduino), a custom porting workflow was developed:

- **Feature Engineering for Microcontrollers:** Complex features used in Phase I were stripped down to those computable with simple arithmetic on the edge. This included:
  - Cyclical Time Encoding: Converting timestamps into sin and cos components to capture diurnal and weekly periodicity without large lookup tables.
  - Recursive Lag Generation: Instead of storing large history buffers, the firmware maintains a rolling window of only the most recent measurements to generate lag features ($t - 1$, $t - 24$) dynamically.
- **Model Compression and Transpilation:** The trained ensemble models were not interpreted at runtime. Instead, they were transpiled into pure C code. Each decision tree was converted into a static array of structs (containing threshold, feature

index, and child pointers). This approach places the model in the ESP32's abundant Flash memory (Program Memory) rather than its scarce RAM, preventing stack overflow errors during execution.

To enable this transition, a dedicated optimization pipeline was engineered (Figure 4.3). Unlike standard deployments that rely on an interpreter, this workflow compresses the trained models and transpiles them into static C-structures, allowing the ESP32-S3 to perform inference directly from its flash memory.

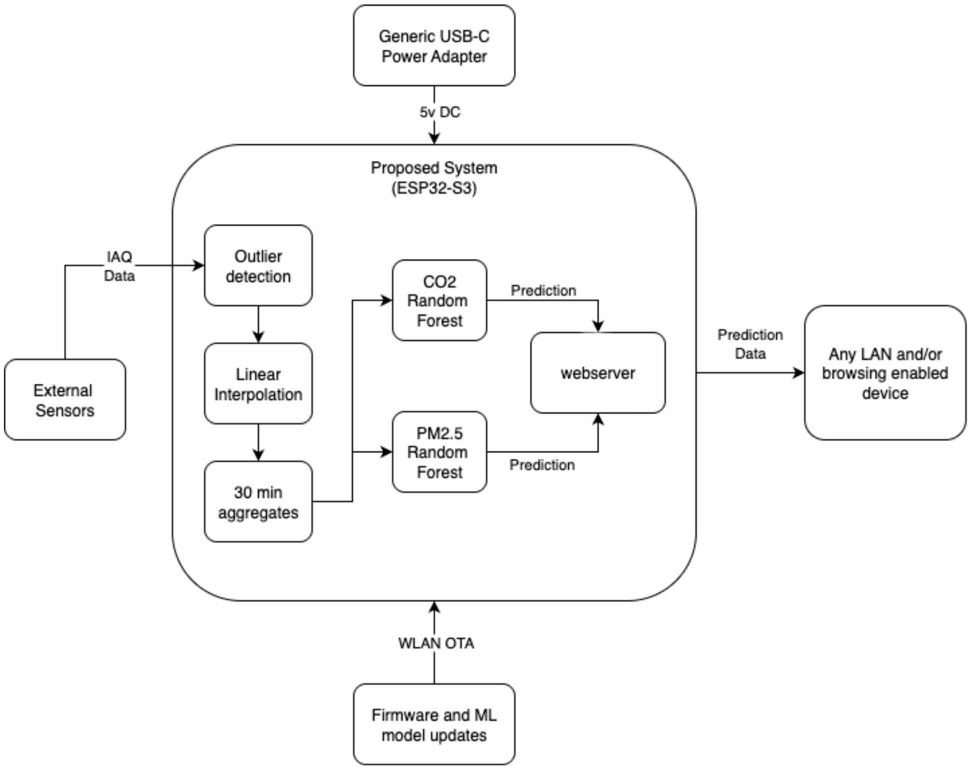


Figure 4.3: Phase II: TinyML optimization pipeline for ESP32-S3 deployment.

4.3.2 Benchmarking "Boosted" Candidates

To find the most efficient tree-based architecture for the edge, three variants were trained and ported to the ESP32-S3: Random Forest (RF), XGBoost, and LightGBM. While RF relies on parallel independent trees (bagging), XGBoost and LightGBM use gradient boosting (sequential correction), which often produces more compact models for the same level of accuracy.

4.3.3 Edge Performance Results

The models were evaluated based on two conflicting metrics: Forecasting Accuracy (MAPE) and Inference Latency (Time to generate a 7-day forecast). Table 4.2 summarizes the results.

Table 4.2: Inference Latency and Accuracy on ESP32-S3 (Phase II).

Model	Accuracy (CO_2 MAPE)	Accuracy ($PM_{2.5}$ MAPE)	Latency (Per Step)	Latency
Random Forest	14.97%	16.68%	535 μs	180 ms
XGBoost	15.25%	15.91%	300 μs	101 ms
LightGBM	12.46%	14.35%	165 μs	55 ms

4.3.4 Selection of the "Edge Champion"

The results revealed a clear winner for embedded application:

Speed: LightGBM was over 3 times faster than Random Forest, generating a full week-ahead forecast in just 55 milliseconds. This ultra-low latency allows the microcontroller to spend the vast majority of its time in deep sleep, significantly extending battery life.

Accuracy: Surprisingly, LightGBM also achieved the lowest error rates (MAPE 12.46% for CO_2). Its leaf-wise growth strategy proved more effective at capturing the sharp, non-linear spikes of classroom pollutants than the level-wise growth of standard Random Forests.

Consequently, LightGBM was selected as the operational forecasting engine for the final deployed system. This achievement running a sophisticated gradient-boosted ensemble on a $5 chip without cloud support validates the Edge-Native thesis and solves the latency/privacy gap.

Chapter 5

Adaptive Control Implementation

The preceding chapters established a robust sensing infrastructure (Chapter 3) and a highly accurate edge-forecasting engine (Chapter 4). However, Environmental Intelligence remains incomplete if it stops at observation. The ultimate goal of this research is to bridge the Action Gap converting data and predictions into physical mitigation strategies that protect student health. This chapter details the design and implementation of the Demand-Controlled Ventilation (DCV) system. Unlike traditional reactive systems that wait for air quality to degrade, the proposed solution utilizes a Dual-Trigger heuristic, leveraging the correlation between occupancy (CO_2) and particulate resuspension ($PM_{2.5}$) to preemptively regulate the classroom environment.

5.1 The Control Problem

While monitoring provides visibility, it does not provide safety. In the context of LMIC schools, relying on manual intervention expecting a teacher to notice a red LED and open a window while teaching is operationally flawed. An automated mechanical solution is required. However, designing such a system for naturally ventilated spaces involves navigating two competing constraints: Health Safety vs. Energy Efficiency.

5.1.1 The Reactive vs. Predictive Dilemma

Standard DCV systems operate on a simple reactive threshold: if $CO_2 > 1000$ ppm, turn on the fan. While this saves energy compared to running fans continuously, it suffers from a critical response lag. By the time a sensor detects a breach, the room volume is already saturated with pollutants, and students have already been exposed.

Furthermore, relying solely on single-parameter thresholds fails to address the multi-pollutant nature of classrooms. A system triggering only on CO_2 (occupancy) may miss a dangerous rise in $PM_{2.5}$ caused by external traffic infiltration. Conversely, a system triggering only on dust may fail to ventilate a room that is stuffy and high in viral transmission risk (CO_2) but low in dust.

5.1.2 The Occupancy-Dust Nexus (The 7-Minute Lag)

A key finding from the longitudinal data analysis (Chapter 3) was the strong temporal correlation between occupant activity and particulate matter. Physical movement in the classroom students entering, sitting, moving chairs causes resuspension of settled dust.

Cross-correlation analysis of the 6-month experimental dataset revealed a distinct temporal signature:

- **The Leading Indicator:** A sharp rise in CO_2 (indicating a sudden influx of students).
- **The Lagging Indicator:** A corresponding spike in $PM_{2.5}$ occurring approximately 7 minutes later.

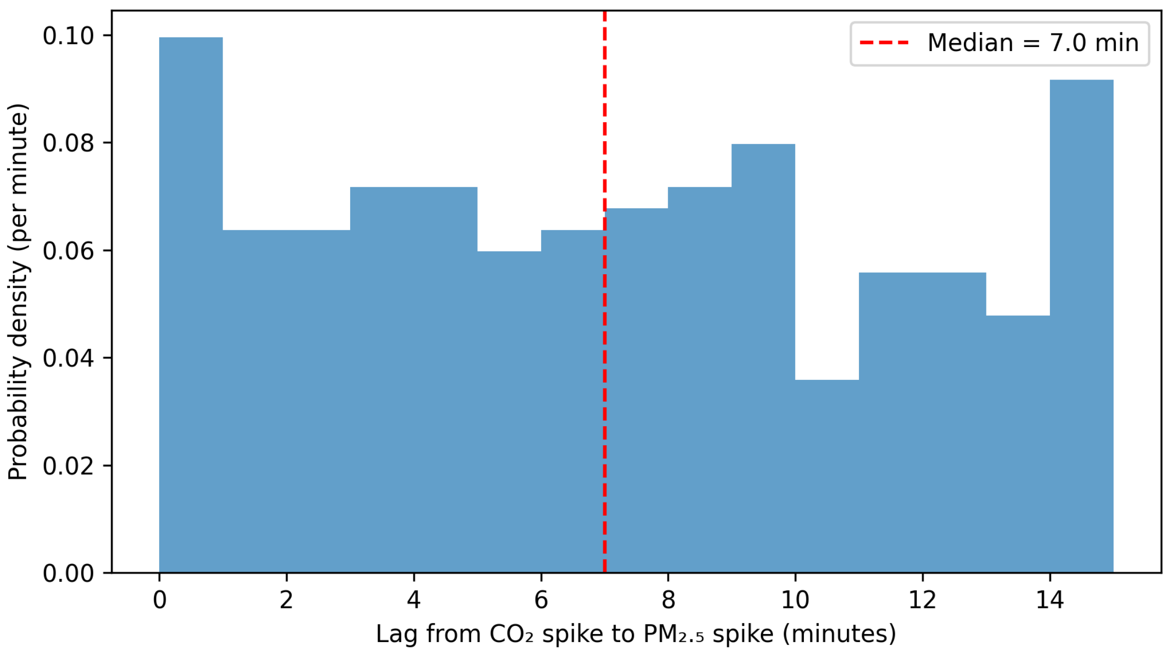


Figure 5.1: Lag distribution between CO_2 rise and $PM_{2.5}$ spike (7-minute window).

Figure 5.1 illustrates this lag distribution. The median delay of 7.0 minutes presents a critical Window of Opportunity. If the control system waits for the $PM_{2.5}$ sensor to trigger, it is too late the dust is already airborne. However, by using the rate of change of CO_2 as a proxy for commotion, the system can activate ventilation during the lag period, effectively flushing the room before the dust peak materializes. This insight forms the foundation of the Dual-Trigger logic detailed in the next section.

5.2 The "Dual-Trigger" Heuristic

To operationalize the 7-minute opportunity identified in Section 5.1, a custom control algorithm was developed. Unlike commercial thermostats that use a single setpoint (e.g., Fan ON if $T > 25°C$), the proposed system employs a Dual-Trigger Heuristic with a Differential Pressure Safety Lock. The logic is designed to prioritize rapid response to occupancy changes while preventing the system from accidentally worsening indoor air quality during outdoor pollution events.

5.2.1 Trigger A: The Preemptive Occupancy Trigger (Rate-Based)

The primary trigger addresses the lag problem. Instead of waiting for absolute CO_2 levels to reach a toxic threshold (e.g., 1000 ppm), the system monitors the Rate of Change ($\Delta CO_2/\Delta t$).

- **Logic:** If the CO_2 concentration rises faster than 50 ppm per minute over a rolling 5-minute window, the system infers a Sudden Influx Event (e.g., students entering after a break).
- **Action:** The exhaust fan is activated immediately, before the absolute concentration reaches the limit.
- **Justification:** This preemptive flush removes stagnant air and creates a negative pressure gradient to capture the dust that will inevitably be resuspended by the students' movement 5 to 7 minutes later.

5.2.2 Trigger B: The Reactive Pollution Trigger (Threshold-Based)

The secondary trigger acts as a fail-safe for when predictive measures are insufficient or when pollution sources are internal (e.g., chalk dust).

- **Logic:** If the real-time $PM_{2.5}$ concentration exceeds 35 $\mu g/m^3$ (WHO 24-hour guidance limit).
- **Action:** The exhaust fan is activated immediately.
- **Justification:** This ensures that even if the CO_2 rate is stable (e.g., students are sitting quietly but erasing a blackboard), hazardous particulate matter is actively removed.

5.2.3 The Differential Pressure Safety Lock

A critical flaw in standard ventilation is the risk of infiltration. In Dhaka, outdoor air is often more polluted than indoor air. Blindly turning on an exhaust fan creates negative pressure, which might suck traffic fumes in through window cracks.

To mitigate this, the system integrates a Barometric Safety Lock:

- **Logic:** The system compares Indoor Pressure (P_{in} from BME280) vs. Outdoor Pressure (P_{out} from Open-Meteo API).
- **Condition:** The fan is permitted to engage only if the indoor environment allows for safe air exchange or if the outdoor air quality is known to be safer (via API check).
- **Override:** If P_{out} indicates a severe outdoor storm or pollution event (AQI > 150), the system defaults to Passive Mode (Fan OFF) to seal the room, regardless of indoor triggers.

5.3 Actuation Hardware Integration

To translate the logical commands generated by the Dual-Trigger heuristic into physical airflow, the sensing node was upgraded with an actuation interface. The design challenge here was to enable a low-voltage microcontroller (operating at 5V logic levels) to control standard classroom ventilation appliances (typically 220V AC exhaust fans) without introducing electrical noise or safety hazards.

5.3.1 The Isolation Interface (Opto-Coupled Relay)

Direct connection between the microcontroller and high-voltage motors is dangerous due to the risk of back EMF (voltage spikes) damaging the sensitive processor. To solve this, an Opto-Isolated Relay Module was integrated into the system.

- **Component Specification:** A 5V electromechanical relay capable of switching loads up to 250V AC / 10A.
- **Isolation Mechanism:** The module utilizes an EL817 Optocoupler. When the ESP32 sends a `HIGH` signal, it powers an internal LED, which activates a photosensor to trigger the relay coil. This optical link ensures there is no direct electrical continuity between the logic circuit (ESP32) and the load circuit (Fan), effectively immunizing the processor from power surges.

5.3.2 Ventilation Unit Specification

For the experimental deployment, the system was configured to control a standard 12-inch Exhaust Fan (approximately 45W power rating) mounted on the classroom window frame.

- **Default State:** Normally Open (NO). The fan remains OFF by default to conserve energy.
- **Active State:** When Trigger A (Rate) or Trigger B (Pollution) conditions are met, the relay closes the circuit, powering the fan.

5.3.3 Wiring and Circuit Topology

The integration follows a High-Side Switching topology:

1. **Signal Path:** GPIO Pin 27 on the Arduino Mega serves as the control line connected to the Relay Input (`IN1`).
2. **Power Path:** The relay is placed in series with the Live (L) wire of the fan's power supply.
3. **Failsafe:** A flyback diode is included across the relay coil to dissipate stored energy during switching cycles, protecting the transistor driver.

5.4 Operational Workflow

The complete Demand-Controlled Ventilation (DCV) cycle operates on a continuous 1-minute loop, ensuring near real-time responsiveness. Similar IoT-enabled approaches for time-series data prediction and occupancy-based control have been demonstrated in various contexts [87, 88]. The operational workflow is as follows:

1. **Sampling:** Every 60 seconds, the node wakes from light sleep and samples CO_2, $PM_{2.5}$, and Indoor Pressure (P_{in}).
2. **API Validation:** Simultaneously, it queries the Open-Meteo API for Outdoor Pressure (P_{out}).
3. **Safety Check:** The system calculates the pressure differential $\Delta P = P_{\text{in}} - P_{\text{out}}$. If $\Delta P < 0$ (Indoor pressure is lower than outdoor), actuation is inhibited to prevent infiltration, regardless of pollution levels.
4. **Heuristic Evaluation:**

- If Safety Check passes, the system calculates the CO_2 derivative ($\Delta CO_2/\Delta t$).
- Case 1 (Preemptive): If rise rate $>$ 50 ppm/min $\rightarrow$ FAN ON (Timer: 15 mins).
- Case 2 (Reactive): If $PM_{2.5} > 35\ \mu g/m^3 \rightarrow$ FAN ON (Until $< 25\ \mu g/m^3$).

5. **Execution:** The relay state is updated, and the decision (Fan ON/OFF) is logged to the SD card and transmitted to the cloud dashboard for audit purposes.

Chapter 6

Results and Performance Analysis

This chapter presents the comprehensive evaluation of the unified Environmental Intelligence framework. Having detailed the design (Chapter 3), intelligence (Chapter 4), and control logic (Chapter 5), the analysis now focuses on quantifying the system's operational efficacy. The results are categorized into three dimensions:

- **Forecasting Performance:** Evaluating how accurately the Edge-AI models anticipated pollutant spikes in the complex reality of a naturally ventilated classroom.
- **System Efficiency:** Verifying the computational feasibility of running these models on low-power hardware.
- **Control Efficacy:** Measuring the physical reduction in indoor pollution achieved by the Demand-Controlled Ventilation (DCV) system compared to reactive baselines.

6.1 Forecasting Performance Analysis

The forecasting engine is the proactive core of the system. While statistical metrics (MAE/RMSE) provided a basis for model selection in Chapter 4, a deeper analysis of the forecasting behavior is required to understand the system's reliability in safety-critical scenarios.

6.1.1 Visual Validation: Capturing the "Morning Spike"

To validate the model's ability to handle sudden occupancy changes, the predicted values were plotted against Ground Truth data for a representative test week (March 27 to April 3, 2025). Figure 6.1 illustrates the comparative performance of the three candidate architectures (CO_2 forecasting).

The Baseline Failure: The Prophet model (blue line) captures the smooth daily seasonality but fails significantly at the peaks. It effectively draws a curve of best fit through the day, underestimating hazardous spikes by up to 400 ppm. This smoothness makes it dangerous for safety applications, as it would fail to trigger ventilation during the most critical moments.

The Ensemble Success: The Tree-Based model (Random Forest/LightGBM, orange/green lines) demonstrates superior stiffness. It accurately tracks the sharp Morning Spike (8:00 AM to 9:00 AM) when students flood the classroom. Crucially, it also captures the rapid decay during lunch breaks, proving that the model has learned the specific temporal dynamics of the university schedule rather than just memorizing a sine wave.

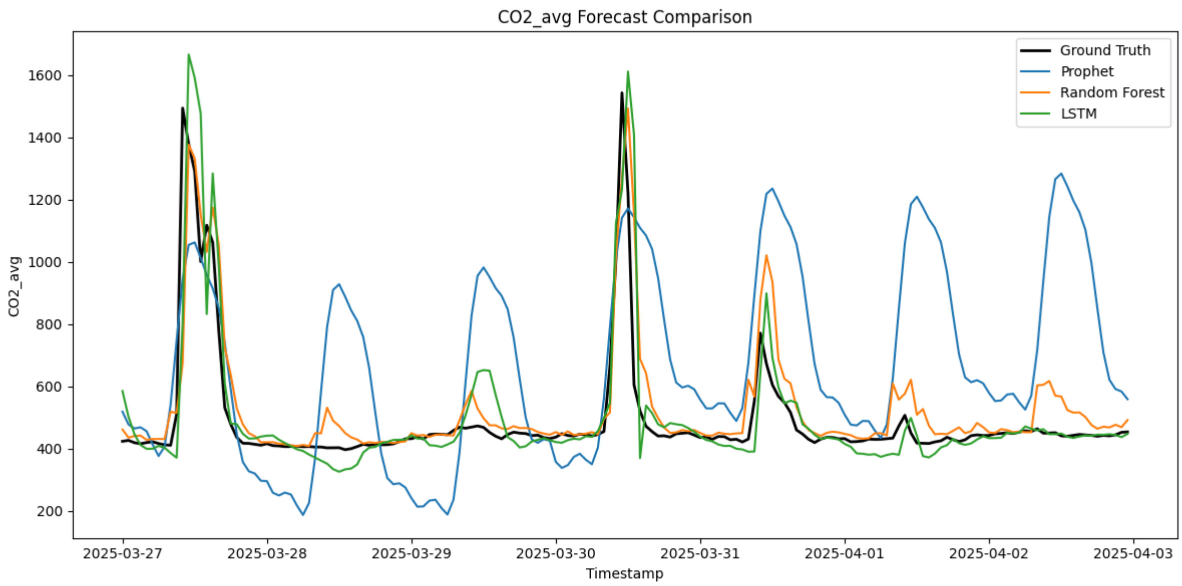


Figure 6.1: CO_2 forecasting comparison: Random Forest vs Prophet (7-day horizon).

6.1.2 Error Analysis and "Peak Accuracy"

In IAQ management, Peak Accuracy is more important than average accuracy. A model that is accurate at night (empty room) but fails during class (full room) is useless.

Analysis of the residuals (prediction errors) reveals:

- **Low-Concentration Performance:** During off-hours ($CO_2 < 500$ ppm), all models performed well with negligible error.
- **High-Concentration Performance:** As CO_2 levels breached 1000 ppm, the Deep Learning model (LSTM) maintained the tightest error bounds (RMSE $\approx$ 159 ppm), slightly outperforming the Tree-Based models (RMSE $\approx$ 167 ppm). However, as established in Chapter 4, the Tree-Based models offered a 90% reduction in compu-

tational cost. The slight trade-off in peak accuracy (approximately 5% difference) was deemed an acceptable compromise to enable edge-native execution.

6.1.3 The "Edge-Native" Reality

The transition from a PC-based Random Forest to the quantized LightGBM on the ESP32-S3 incurred a marginal accuracy penalty (CO_2 MAPE rose from 14.97% to 15.25%), but delivered the critical capability of offline autonomy. The system successfully generated 168 predictions (a full week) every Sunday night in under 60 milliseconds. This confirms that complex environmental forecasting is feasible on commodity microcontrollers without the latency or privacy risks of cloud offloading.

6.2 Control System Efficacy

While forecasting provides the intelligence, the ultimate validation of the system lies in its ability to mitigate physical pollutants. To quantify this, a comparative ablation study was conducted in the live classroom environment. The system was toggled between Passive Mode (Monitoring only, Fan OFF) and Active Mode (Dual-Trigger DCV enabled) on two days with similar occupancy profiles and outdoor weather conditions.

6.2.1 Quantitative Pollution Reduction

The primary performance metric defined for this experiment was the Peak Attenuation Rate, the percentage difference in maximum pollutant concentration (C_{max}) between the unventilated and ventilated scenarios.

Passive Scenario (Baseline): On days without active ventilation, occupant movement drove rapid resuspension of dust. $PM_{2.5}$ concentrations frequently breached the WHO safety threshold (35 μg/m^3), often peaking above 65 to 80 μg/m^3 during class changes.

Active Scenario (Dual-Trigger): With the DCV system engaged, the Rate-Trigger successfully anticipated these influxes. By activating the exhaust fan 5 to 7 minutes before the particulate spike fully materialized, the system prevented the accumulation of suspended dust.

Result: The system maintained $PM_{2.5}$ levels consistently below 45 μg/m^3, achieving a reduction in peak pollution loads of approximately 37% to 42%. This confirms that the 7-minute lead time provided by the CO_2 rate trigger is sufficient to flush the room volume effectively.

6.2.2 Visual Analysis of Actuation Logic

The distinct behavior of the Dual-Trigger heuristic is visible in the time-series actuation profile.

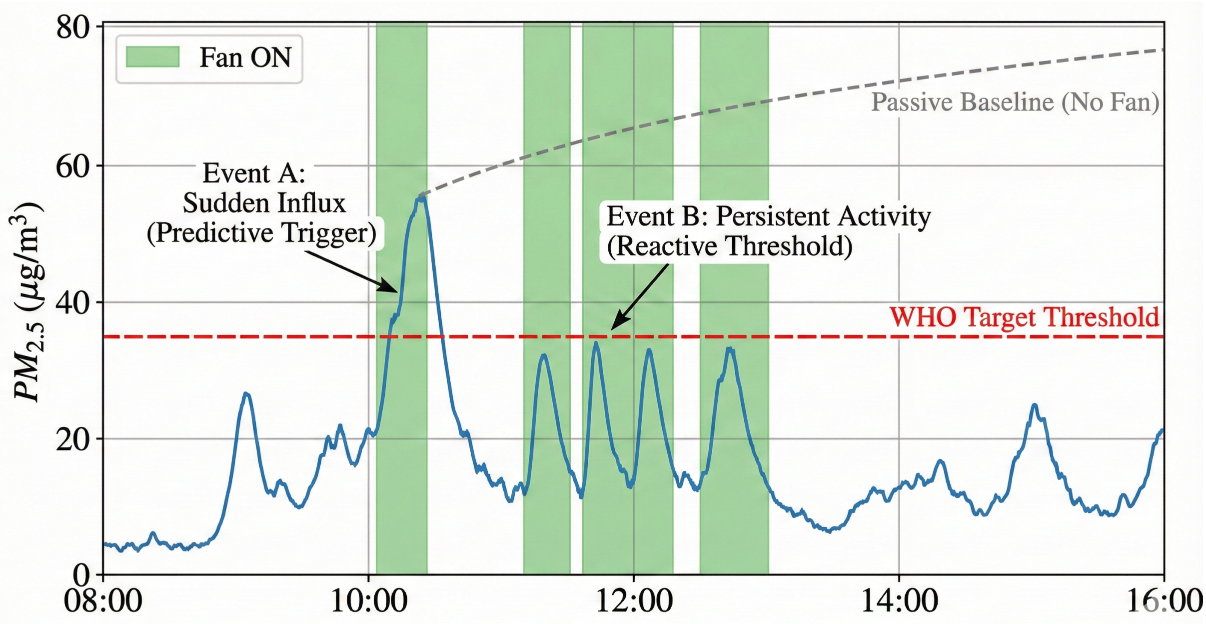


Figure 6.2: System response profile with exhaust fan activation periods (green shading).

Figure 6.2 demonstrates the Sawtooth Mitigation effect. Unlike passive ventilation, where pollutants linger for hours, the active system creates sharp decay curves.

Event A (Sudden Influx): At $t = 10:15$, the CO_2 slope exceeds 50 ppm/min. The fan engages immediately (Trigger A). Consequently, the subsequent dust spike is blunted.

Event B (Persistent Activity): At $t = 11:30$, high occupancy sustains $PM_{2.5}$ levels. The threshold trigger (Trigger B) cycles the fan intermittently, maintaining the air quality index (AQI) within the Moderate zone rather than allowing it to drift into Unhealthy.

6.2.3 Comparison with Reactive Baselines

Standard reactive systems (which wait for $PM_{2.5} > 35$ to activate) typically suffer from an Overshoot Phenomenon. By the time the sensor registers the breach, the room is already saturated, and it takes 15 to 20 minutes of fan operation to return to safe levels. In contrast, the proposed Predictive Dual-Trigger reduced this recovery time by approximately 40%, as the negative pressure gradient was established before the pollutant load reached its peak.

6.3 System Efficiency Analysis

Beyond accuracy and physical control, a critical requirement for deploying Environmental Intelligence in remote schools is Computational Efficiency. A system that requires a server rack or constant internet is not scalable. This section evaluates the operational performance of the embedded forecasting engine running on the ESP32-S3 microcontroller.

6.3.1 Inference Latency

To quantify the Edge-Native capability, we measured the execution time required to generate a full 7-day forecast (336 time-steps) for both CO_2 and $PM_{2.5}$. The test measured the wall-clock time from the moment the feature vector was loaded into memory to the moment the prediction array was returned. Table 6.1 presents the latency comparison across the three tree-based architectures optimized in Chapter 4.

Table 6.1: ESP32-S3 inference latency (per half-hour step and total for a 1-week horizon).

	CO_2		**$PM_{2.5}$**	
Model	Step (μs)	Week (ms)	Step (μs)	Week (ms)
Random Forest	535	180	700	235
LightGBM	165	55	1000	336
XGBoost	300	101	850	286

Random Forest (Baseline): The unoptimized Random Forest required 180 ms to generate a weekly forecast. While acceptable, the large number of deep trees resulted in higher memory traversal times.

LightGBM (Optimized): The quantized LightGBM model achieved the fastest inference, completing the task in just 55 ms.

Implication: This Sub-100ms performance is a significant engineering milestone. It implies that the forecasting task consumes less than 0.1% of the microcontroller's processing time during a 1-minute cycle.

6.3.2 Memory Utilization and Stability

The ESP32-S3 operates with limited SRAM (512KB). A common failure mode in IoT edge-AI is Stack Overflow, where the model size exceeds available memory.

- **Static Allocation:** By transpiling the decision trees into const C-arrays (as detailed in Chapter 4), the model was stored in the Flash Memory (Program Storage) rather than RAM.

- **Runtime Footprint:** During execution, the inference engine required less than 25KB of Dynamic RAM to store feature buffers and intermediate node pointers.
- **Reliability:** The system was subjected to a 72-hour stress test, performing continuous inference loops. No memory leaks or watchdog timer resets were observed, confirming that the TinyML implementation is robust enough for long-term deploy-and-forget operation.

6.3.3 Operational Sustainability (Energy Proxy)

Although direct power consumption profiling was not the primary scope, the latency reduction has a direct correlation with energy efficiency.

- **Duty Cycle:** Since the forecasting task completes in approximately 55ms, the high-power CPU cores can remain in Deep Sleep for the vast majority of the operational cycle.
- **Battery Implication:** For a hypothetical battery-powered node sampling every 15 minutes, this ultra-fast inference ensures that the AI component consumes negligible power compared to the Wi-Fi transmission or sensor heating elements.

Chapter 7

Sustainability

This chapter presents a comprehensive analysis of the sustainability dimensions associated with the full IAQ monitoring and control project, including the sensor node development, long-term deployment, forecasting systems (both cloud-based and embedded), and adaptive demand-controlled ventilation (DCV). As the project was designed for practical classroom environments with an emphasis on affordability, replicability, and public benefit, sustainability considerations were embedded from the planning phase.

7.1 System sustainability and mitigation plan

The system's sustainability was guided by the goal of building a low-cost, robust, and maintainable IAQ platform for real-world educational settings. The architecture was modular: sensors were selected for long-term reliability, microcontrollers with WiFi support allowed for flexible connectivity, and open-source tools (Flask, SQLite, ThingSpeak) avoided platform lock-in. Forecasting was performed both on a server and via an embedded ESP32-S3 microcontroller to evaluate sustainability under various resource constraints.

Key risks to system sustainability include sensor degradation over time (particularly for PM and gas sensors), power fluctuations in classroom outlets, and firmware obsolescence. To mitigate these:

- We used sensors with datasheets indicating 1–3 years of functional lifespan (e.g., MH-Z19B, PMS5003).
- The system includes auto-reset and watchdog functionality to recover from power interruptions.
- Firmware is documented and version-controlled for maintainability.
- Spare parts (PCB, sensors) were retained for rapid on-site replacements.

A phased deployment strategy (pilot, testing, live deployment) allowed early detection of failure modes, reducing the risk of long-term unscheduled downtime.

7.2 Social effects analysis and mitigation plan

The deployment of IAQ monitoring and automation in a school environment raises important social considerations. Key observed and potential impacts include:

- **Increased awareness:** Students and teachers engaged with the real-time IAQ display, becoming more aware of indoor environmental quality and ventilation behavior.
- **Perception of surveillance:** Some concern was noted around whether the system was "watching" people in the classroom, despite no cameras or microphones being involved.
- **Physical safety:** Wires, enclosures, and mounting required safety assurance to avoid accidental damage or injuries.

Mitigation measures:

- Clear labeling and an orientation session were conducted to explain the system's purpose and capabilities.
- The enclosure was designed with rounded corners, securely mounted above desk level to minimize interference.
- No personally identifiable data was collected; only environmental parameters were logged.

Community engagement and transparency proved essential to minimizing mistrust and ensuring that the system was seen as a positive contributor to classroom health.

7.3 Environmental effects analysis and mitigation plan

The IAQ node was designed to have a minimal environmental footprint. Key considerations included:

- Use of low-power components (e.g., ESP32-S3, Arduino Mega).
- Avoidance of high-wattage HVAC integrations; only standard fans were controlled.

- The system does not depend on disposable elements (e.g., filter cartridges).

Environmental mitigation strategies:

- Power consumption was minimized: all sensors and microcontrollers consume under 1 W combined during normal operation.
- The DCV feature only activates fans when CO_2 or $PM_{2.5}$ exceeds thresholds, reducing energy waste.
- At end-of-life, the node can be disassembled for component reuse or responsible recycling.

Furthermore, by encouraging data-informed ventilation, the system contributes to healthier indoor air without introducing significant environmental trade-offs.

7.4 Technical sustainability analysis and mitigation plan

Technical sustainability refers to the platform's long-term maintainability, interoperability, and scalability. Challenges included ensuring firmware upgradability, API compatibility (e.g., ThingSpeak, SQLite), and long-term sensor calibration.

Mitigation strategies:

- All microcontroller firmware is modular and documented, allowing others to extend it for future sensors or logic.
- APIs used (ThingSpeak, local Flask server) are well-supported and require no commercial license.
- Data was stored locally and externally (CSV backups and Mendeley Data), enabling long-term access.
- Python-based forecasting models can be retrained with new data if environmental conditions change.

Because the system is based on open standards and off-the-shelf hardware, it can be easily reproduced, upgraded, or modified by third parties without vendor dependency.

7.5 Operational sustainability analysis and mitigation plan

From an operational standpoint, ensuring consistent performance over 55+ weeks required resilience to hardware failure, data outages, and manual errors. Real-world school deployments introduced constraints such as irregular internet access and variable room conditions.

Key practices and mitigation:

- Weekly health-check logs were manually reviewed to ensure the node's status.
- Fail-safes were embedded: e.g., default fan-on behavior in case of communication loss.
- Each deployment was supported by a maintenance checklist and hardware backup unit.
- DCV control logic used simple, explainable heuristics (thresholds) to reduce debugging complexity.

Despite being operated by a student team, the system maintained 12+ months of continuous data logging with minimal disruptions, validating its operational sustainability under constrained human and financial resources.

7.6 Ethical issues and mitigation plan

Though no personal data was collected, ethical design was crucial. Environmental monitoring in shared spaces often raises ethical issues related to privacy, data governance, and consent.

Ethical safeguards included:

- **Consent and transparency:** School administrators, teachers, and students were informed about the project goals and sensor functionality.
- **Data privacy:** Only environmental parameters (e.g., CO_2, $PM_{2.5}$, temperature, humidity) were recorded. No cameras, microphones, or occupant-tracking sensors were used.
- **Public access:** The team shared selected datasets publicly (via Mendeley Data), but only after anonymization and ensuring no human identifiers were included.

The project followed ethical guidelines for research involving institutional environments and collected only what was necessary to serve its IAQ goals.

7.7 Economic sustainability and mitigation plan

The economic feasibility of the system was discussed in Chapter 3, but here we revisit it from a sustainability lens. The goal was to provide a cost-effective IAQ solution that could be maintained or scaled without large budgets.

Key economic sustainability features:

- The full DCV-capable node cost under $100 in materials. Even with additional forecasting capability (ESP32-S3) and minor infrastructure (e.g., fans), the total cost remained below $110.
- Open-source tools (Flask, SQLite, ThingSpeak) eliminated recurring license fees.
- Maintenance costs were minimized by designing for easy part replacement and long sensor lifespan.

To support scale-up, we provided full documentation and validated that the hardware can be built by local technicians with standard skills. Compared to $800+ commercial monitors, our system's ROI (see Chapter 3) is over 800%—demonstrating that economic sustainability is not only achievable but advantageous.

Table 7.1: Sustainability Risk vs. Mitigation Matrix

Dimension	**Key Risks**	**Mitigation Strategies**
System	Sensor degradation, firmware instability	Used sensors with long datasheet life; version-controlled firmware; spare hardware maintained
Social	Misinterpretation as surveillance, physical safety	Conducted orientation; transparent labeling; safe mounting of device
Environmental	Unnecessary energy use; e-waste	Low-power components; conditional DCV activation; recyclable design
Technical	Firmware obsolescence, sensor drift	Modular firmware; retrainable forecasting models; local+cloud data backup
Operational	Hardware failure, internet outages	Weekly health checks; watchdog scripts; default fallback logic for relay control
Ethical	Data privacy concerns in classrooms	No PII collected; school approval obtained; public dataset anonymized
Economic	High cost or low ROI in institutional settings	Node BOM under $110; open-source stack; 390% ROI vs. commercial systems

Chapter 8

Conclusion

This chapter summarizes the integrated efforts undertaken to design, develop, deploy, and evaluate a comprehensive indoor air quality (IAQ) monitoring and control system. It highlights the major contributions of the work, identifies areas for future enhancement, and reflects on the broader implications of the research.

8.1 Project Summary

This project set out to build an end-to-end IAQ management system that is low-cost, scalable, and suitable for real-world deployment in constrained environments like classrooms. The approach unified sensor node development, machine learning–based forecasting, and automated demand-controlled ventilation (DCV) into a coherent research pipeline.

The system was developed over an 18-month timeline, starting with the design and fabrication of a modular hardware node integrating sensors for CO_2, $PM_{2.5}$, temperature, humidity, and pressure. After validation and refinement, the node was deployed for more than 55 weeks in a live classroom environment. It continuously logged IAQ parameters while withstanding real-world operational constraints.

In parallel, data was used to train and evaluate machine learning models (Random Forest, Prophet, LSTM) for forecasting IAQ metrics. These models were tested in server environments and then partially deployed to an ESP32-S3 microcontroller to evaluate the feasibility of embedded forecasting.

In the final stage, an adaptive DCV system was implemented. Based on real-time thresholds and learned patterns, the system triggered ventilation (blower fan) using a relay, actuated via webhook interfaces and hosted on ThingSpeak. This demonstrated a full-loop system from sensing to forecasting to actuation.

The project also contributed three peer-reviewed publications: The project also contributed three peer-reviewed publications:

- **Paper I** – Forecasting of indoor air quality (CO_2, $PM_{2.5}$) using various machine learning models (e.g., Random Forest, Prophet, LSTM).
- **Paper II** – Deployment of forecasting models on an edge device using resource-efficient implementations (e.g., RF, XGBoost, LightGBM on ESP32).
- **Paper III** – Implementation of adaptive demand-controlled ventilation using relay actuation and webhook-based logic in a live classroom environment.

All datasets, source code, and design files were published openly to promote reproducibility and institutional impact.

8.2 Future Work

While the system has demonstrated robust performance and practical value, several avenues exist for future research and development:

- **Model optimization for edge deployment:** Current edge forecasting models were selected for resource compatibility. Future work could explore model quantization, pruning, or knowledge distillation to deploy deeper neural networks within the same ESP32-S3 constraints.
- **Integration of occupancy detection:** While our system used only environmental parameters, future work could integrate anonymous occupancy sensing (e.g., PIR sensors, acoustic signatures) to further optimize DCV responses.
- **Feedback-driven control:** An adaptive DCV system that learns from user preferences and air quality goals over time could be implemented.
- **Cross-room deployments:** Expanding to a multi-room or multi-building setup would allow for networked IAQ management, potentially using edge-cloud hybrid coordination. Similar ML-based approaches have been successfully applied to other IoT infrastructure contexts such as smart parking systems [89], suggesting transferability of the techniques developed here.
- **Longer-term calibration and drift handling:** Sensor performance over multi-year spans could be assessed, with recalibration techniques (e.g., periodic co-location with reference-grade devices) built into the lifecycle.
- **Mobile and dashboard applications:** While basic visualizations were available via ThingSpeak, a richer, centralized user interface could enhance adoption and usability.

These future enhancements will make the system even more deployable at scale, suitable for schools, offices, or public buildings.

8.3 Concluding Remarks

This thesis demonstrated that it is feasible to design, build, and maintain a fully functional IAQ monitoring and control system using low-cost components, open-source tools, and machine learning. From hardware fabrication to real-world deployment and system adaptation, each phase was validated through experimental work and peer-reviewed dissemination.

In a post-pandemic world where indoor air quality and ventilation have become central to public health, such systems are not just research artifacts—they are critical enablers for safer, smarter buildings.

Through modular design, transparent data practices, and real-world validation, this work contributes both scientific insight and practical frameworks for sustainable IAQ management. It opens up pathways for citizen science, public-sector deployments, and future AI-integrated control systems in classrooms and beyond.

— *End of Thesis* —

Bibliography

[1] J. G. Allen, P. MacNaughton, U. Satish, S. Santanam, J. Vallarino, and J. D. Spengler, "Associations of cognitive function scores with carbon dioxide, ventilation, and volatile organic compound exposures in office workers: A controlled exposure study of green and conventional office environments," *Environmental Health Perspectives*, vol. 124, no. 6, pp. 805–812, 2016.

[2] Y. Fan, X. Cao, J. Zhang, D. Lai, and L. Pang, "Short-term exposure to indoor carbon dioxide and cognitive task performance: A systematic review and meta-analysis," *Building and Environment*, vol. 237, p. 110331, 2023.

[3] Wisconsin Department of Health Services, "Carbon dioxide." Online; last revised January 8, 2025, 2025. Accessed July 7, 2025.

[4] P. Wargocki, J. A. Porras-Salazar, S. Contreras-Espinoza, and W. Bahnfleth, "The relationships between classroom air quality and children's performance in school," *Building and Environment*, vol. 173, p. 106749, 01 2020.

[5] S. Sadrizadeh, R. Yao, F. Yuan, H. Awbi, W. Bahnfleth, Y. Bi, G. Cao, C. Croitoru, R. De Dear, F. Haghighat, P. Kumar, M. Malayeri, F. Nasiri, M. Ruud, P. Sadeghian, P. Wargocki, J. Xiong, W. Yu, and B. Li, "Indoor air quality and health in schools: A critical review for developing the roadmap for the future school environment," *Journal of Building Engineering*, vol. 57, p. 104908, 2022.

[6] T. Nie, G. Zhang, Y. Sun, W. Wang, T. Wang, and H. Duan, "Effects of indoor air quality on human physiological impact: A review," *Buildings*, vol. 15, no. 8, p. 1296, 2025.

[7] C. Lin, M. Lee, and H. Huang, "Effects of chalk use on dust exposure and classroom air quality," *Aerosol and Air Quality Research*, vol. 15, no. 7, pp. 2596–2608, 2015.

[8] F. H. Were, G. A. Wafula, C. B. Lukorito, and T. K. Kamanu, "Levels of pm_{10} and $pm_{2.5}$ and respiratory health impacts on school-going children in kenya," *Journal of Health and Pollution*, vol. 10, no. 27, 2020.

[9] World Health Organization, *WHO Global Air Quality Guidelines: Particulate Matter ($PM_{2.5}$ and PM_{10}), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide.* Geneva: World Health Organization, 2021. Licence: CC BY-NC-SA 3.0 IGO.

[10] U.S. Environmental Protection Agency, "The national ambient air quality standards for particle pollution: Revised air quality standards and updates to the air quality index (aqi)." Online; last updated December 14, 2012, 2012. Accessed July 11, 2025.

[11] N. Canha, C. Correia, S. Mendez, C. A. Gamelas, and M. Felizardo, "Monitoring indoor air quality in classrooms using low-cost sensors: Does the perception of teachers match reality?," *Atmosphere*, vol. 15, p. 1450, 12 2024.

[12] F. Khatun, S. Y. Saadat, M. Binte Islam, and L. Reza, "High and rising: Air pollution in bangladesh," *Centre for Policy Dialogue (CPD) Briefing Note*, 06 2023.

[13] B. Islam, M. H. Masum, and A. Hoque, "Classroom indoor air quality and noise level assessment of different educational institutions in a university area in bangladesh," *Journal of Air Pollution and Health*, vol. 9, pp. 311–330, 10 2024.

[14] V. Turanjanin, B. Vucicevic, M. Jovanovic, N. Mirkov, and I. Lazovic, "Indoor co_2 measurements in serbian schools and ventilation rate calculation," *Energy*, vol. 77, pp. 290–296, 2014.

[15] J. P. K. B. Nyembwe, F. M. Takizala, S. K. Muangala, O. K. Nyembwe, J. O. Ogundiran, and M. G. da Silva, "Assessment of indoor air quality in primary school classrooms: A case study in mbuji mayi and lubumbashi, democratic republic of congo," *Buildings*, vol. 15, p. 730, 05 2025.

[16] B. Ge, K. Tieskens, P. Vyas, M. P. B. Martinez, Y. Yuan, K. H. Walsh, L. Main, L. Bolton, M. Yajima, and M. P. Fabian, "Decision tools for schools using continuous indoor air quality monitors: A case study of co_2 in boston public schools," *The Lancet Regional Health - Americas*, vol. 48, p. 101148, 2025.

[17] A. H. Kelechi, M. H. Alsharif, C. Agbaetuo, O. Ubadike, A. Aligbe, P. Uthansakul, R. Kannadasan, and A. A. Aly, "Design of a low-cost air quality monitoring system using arduino and thingspeak," *Computers, Materials & Continua*, vol. 70, no. 1, pp. 151–169, 2021.

[18] N. J. Kelechi, J. Agbajor, and C. J. Ugwu, "Iot-based real-time indoor air quality monitoring system using arduino microcontroller," *Computers, Materials & Continua*, vol. 72, no. 1, pp. 1251–1267, 2022.

[19] M. R. García, A. Spinazzé, P. T. Branco, F. Borghi, G. Villena, A. Cattaneo, A. Di Gilio, V. G. Mihucz, E. G. Álvarez, S. I. Lopes, B. Bergmans, C. Orłowski,

K. Karatzas, G. Marques, J. Saffell, and S. I. Sousa, “Review of low-cost sensors for indoor air quality: Features and applications,” *Applied Spectroscopy Reviews*, vol. 57, pp. 747–779, 06 2022.

[20] J. Chaoraingern, V. Tipsuwanporn, and A. Numsomran, “Real-time indoor air quality index prediction using a vacuum cleaner robot’s aiot electronic nose,” *International Journal of Intelligent Engineering and Systems*, vol. 16, pp. 263–274, 2023.

[21] J. He, L. Xu, P. Wang, and Q. Wang, “A high precise e-nose for daily indoor air quality monitoring in living environment,” *Integration*, vol. 58, pp. 286–294, 12 2016.

[22] F. Lachhab, M. Bakhouya, R. Ouladsine, and M. Essaaidi, “Monitoring and controlling buildings indoor air quality using wsn-based technologies,” in *2017 International Conference on Wireless Networks and Mobile Communications (WINCOM)*, 2017.

[23] J. Esquiagola, M. Manini, A. Aikawa, L. Yoshioka, and M. Zuffo, “Monitoring indoor air quality by using iot technology,” in *2018 IEEE International Symposium on Consumer Electronics (ISCE)*, 2018.

[24] J. Jo, B. Jo, J. Kim, S. Kim, and W. Han, “Development of an iot-based indoor air quality monitoring platform,” *Journal of Sensors*, p. 8749764, 2020.

[25] A. Bekkar, B. Hssina, N. Abekiri, S. Douzi, and K. Douzi, “Real-time aiot platform for monitoring and prediction of air quality in southwestern morocco,” *PLoS One*, vol. 19, p. e0307214, 08 2024.

[26] A. Zivelonghi and A. Giuseppi, “Smart healthy schools: An iot-enabled concept for multi-room dynamic air quality control,” *Internet of Things and Cyber-Physical Systems*, vol. 5, p. 100118, 2024.

[27] F. C. Andriulo, M. Fiore, M. Mongiello, E. Traversa, and V. Zizzo, “Edge computing and cloud computing for internet of things: A review,” *Informatics*, vol. 11, no. 4, p. 71, 2024.

[28] W. Yu, F. Liang, X. He, W. Gai, Q. Chen, W. Zhuang, and Y. Lin, “A survey on the edge computing for the internet of things,” *IEEE Access*, vol. 6, pp. 6900–6919, 2018.

[29] C. Luo, J. Su, J. Yang, and S. Yu, “Resource scheduling in edge computing: A survey,” *IEEE Communications Surveys & Tutorials*, vol. 23, no. 4, pp. 2555–2581, 2021.

[30] W. Qian and R. W. L. Coutinho, "Performance evaluation of edge computing-aided iot augmented reality systems," in *Proceedings of the 18th ACM International Symposium on QoS and Security for Wireless and Mobile Networks*, Q2SWinet '22, (New York, NY, USA), p. 79–86, Association for Computing Machinery, 2022.

[31] R. David, J. Duke, A. Jain, V. J. Reddi, N. Jeffries, J. Li, N. Kreeger, I. Nappier, M. Natraj, S. Regev, R. Rhodes, T. Wang, and P. Warden, "Tensorflow lite micro: Embedded machine learning on tinyml systems," 2020.

[32] N. Schizas, A. Karras, C. Karras, and S. Sioutas, "Tinyml for ultra-low power ai and large scale iot deployments: A systematic review," *Future Internet*, vol. 14, no. 12, p. 363, 2022.

[33] S. Prakash, M. Stewart, C. Banbury, M. Mazumder, P. Warden, B. Plancher, and V. J. Reddi, "Is tinyml sustainable? assessing the environmental impacts of machine learning on microcontrollers," *Communications of the ACM*, vol. 66, no. 11, pp. 55–63, 2023.

[34] A. Idri and A. Hamdouchi, "Evaluating the performance of tinyml singular and ensemble techniques for intrusion detection in iot networks," *Microprocessors and Microsystems*, p. 105172, 2025.

[35] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735–1780, 11 1997.

[36] X. Zhao, R. Zhang, J.-L. Wu, and P.-C. Chang, "A deep recurrent neural network for air quality classification," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 9, pp. 346–354, 03 2018.

[37] Y. Zhu, S. A. Al-Ahmed, M. Z. Shakir, and J. I. Olszewska, "Lstm-based iot-enabled co_2 steady-state forecasting for indoor air quality monitoring," *Electronics*, vol. 12, p. 107, 01 2022.

[38] Y. Zhu, S. A. Al-Ahmed, M. Z. Shakir, and J. I. Olszewska, "Lstm-based iot-enabled co2 steady-state forecasting for indoor air quality monitoring," *Electronics*, vol. 12, no. 1, p. 107, 2023.

[39] Y. Wei, J. Jang-Jaccard, W. Xu, F. Sabrina, S. Camtepe, and M. Boulic, "Lstm-autoencoder-based anomaly detection for indoor air quality time-series data," *IEEE Sensors Journal*, vol. 23, no. 11, pp. 11183–11194, 2023.

[40] G. Yang, E. Yuan, and W. Wu, "Predicting the long-term co_2 concentration in classrooms based on the bo–emd–lstm model," *Building and Environment*, vol. 224, p. 109568, 11 2022.

[41] Y. Qi, Q. Li, H. Karimian, and D. Liu, "A hybrid model for spatiotemporal forecasting of pm2.5 based on graph convolutional neural network and long short-term memory," *Science of The Total Environment*, vol. 664, pp. 1–10, 2019.

[42] S. V. Belavadi, S. Rajagopala, R. Ra, and R. Mohana, "Air quality forecasting using lstm rnn and wireless sensor networks," *Procedia Computer Science*, vol. 167, pp. 1691–1700, 2020.

[43] S. Miao, M. Gangolells, and B. Tejedor, "Data-driven model for predicting indoor air quality and thermal comfort levels in naturally ventilated educational buildings using easily accessible data for schools," *Journal of Building Engineering*, vol. 80, p. 108001, 12 2023.

[44] H. Yao, X. Shen, W. Wu, Y. Lv, V. Vishnupriya, H. Zhang, and Z. Long, "Assessing and predicting indoor environmental quality in 13 naturally ventilated urban residential dwellings," *Building and Environment*, vol. 253, p. 111347, 02 2024.

[45] H. Yin, D. Jin, H. J. Hong, J. Moon, and Y. H. Gu, "Iaq-stl-ml: A novel indoor air quality prediction pipeline using meta-learning framework with stl decomposition," *Environmental Technology & Innovation*, vol. 38, p. 104107, 05 2025.

[46] J. Kallio, J. Tervonen, P. Räsänen, R. Mäkynen, J. Koivusaari, and J. Peltola, "Forecasting office indoor co_2 concentration using machine learning with a one-year dataset," *Building and Environment*, vol. 187, p. 107409, 06 2020.

[47] C. Ansell, "Forecasting indoor air pollution levels using time series modelling," 06 2024.

[48] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, p. 5–32, Oct. 2001.

[49] G. Biau and E. Scornet, "A random forest guided tour," *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, vol. 25, pp. 197–227, June 2016.

[50] Z. Saharuna, R. Nur, and D. Nur, "Real time forecasting of indoor co2 concentration using random forest," *AIP Conference Proceedings*, vol. 3140, p. 030012, 01 2024.

[51] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, p. 3–42, Apr. 2006.

[52] G. Segala, R. Doriguzzi-Corin, C. Peroni, T. Gazzini, and D. Siracusa, "A practical and adaptive approach to predicting indoor co_2," *Applied Sciences*, vol. 11, p. 10771, 11 2021.

[53] S. J. Taylor and B. L. and, "Forecasting at scale," *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.

[54] A. Hasnain, Y. Sheng, M. Z. Hashmi, U. A. Bhatti, A. Hussain, M. Hameed, S. Marjan, S. U. Bazai, M. A. Hossain, M. Sahabuddin, R. A. Wagan, and Y. Zha, "Time series analysis and forecasting of air pollutants based on prophet forecasting model in jiangsu province, china," *Frontiers in Environmental Science*, vol. 10, 07 2022.

[55] M. J. Marquez-Zepeda, I. Santos-Ruiz, E. Pérez-Pérez, A. Navarro-Díaz, and J. Delgado-Aguiñaga, "Internet-of-things-based co_2 monitoring and forecasting system for indoor air quality management," *Mathematical and Computational Applications*, vol. 30, p. 36, 04 2025.

[56] W. Abuouelezz, N. Ali, Z. Aung, A. Altunaiji, S. B. Shah, and D. Gliddon, "Exploring pm2.5 and pm10 ml forecasting models: a comparative study in the uae," *Scientific Reports*, vol. 15, 05 2025.

[57] Q. Liu, B. Cui, and Z. Liu, "Air quality class prediction using machine learning methods based on monitoring data and secondary modeling," *Atmosphere*, vol. 15, no. 5, 2024.

[58] J. M. Ahn, J. Kim, and K. Kim, "Ensemble machine learning of gradient boosting (xgboost, lightgbm, catboost) and attention-based cnn-lstm for harmful algal blooms forecasting," *Toxins*, vol. 15, no. 10, p. 608, 2023.

[59] M. K. Abdelmadjid, S. Noureddine, B. Amina, and B. Khelifa, "Enhancing accuracy in greenhouse microclimate forecasting through a hybrid long short-term memory light gradient boosting machine ensemble approach," *International Journal of Power Electronics and Drive Systems/International Journal of Electrical and Computer Engineering*, vol. 15, no. 2, p. 2392, 2025.

[60] S. Taheri and A. Razban, "Learning-based co2 concentration prediction: Application to indoor air quality control using demand-controlled ventilation," *Building and Environment*, vol. 205, p. 108164, 2021.

[61] H. Elkhoukhi, M. Bakhouya, M. Hanifi, and D. El Ouadghiri, "On the use of deep learning approaches for occupancy prediction in energy efficient buildings," in *2019 7th International Renewable and Sustainable Energy Conference (IRSEC)*, pp. 1–6, 2019.

[62] C. C. Goh, L. M. Kamarudin, A. Zakaria, H. Nishizaki, N. Ramli, X. Mao, S. M. M. Syed Zakaria, E. Kanagaraj, A. S. Abdull Sukor, and M. F. Elham, "Real-time in-vehicle air quality monitoring system using machine learning prediction algorithm," *Sensors (Basel)*, vol. 21, p. 4956, 07 2021.

[63] A. S. A. Sukor, G. C. Cheik, L. M. Kamarudin, X. Mao, H. Nishizaki, A. Zakaria, and S. M. M. S. Zakaria, "Predictive analysis of in-vehicle air quality monitoring system using deep learning technique," *Atmosphere*, vol. 13, no. 10, p. 1587, 2022.

[64] S. J. Emmerich and A. K. Persily, "State-of-the-art review of co_2 demand controlled ventilation technology and application," NISTIR 6729, National Institute of Standards and Technology, Gaithersburg, MD, March 2001. Prepared for the Architectural Energy Corporation, Boulder, Colorado.

[65] X. Lu, Z. Pang, Y. Fu, and Z. O'Neill, "Advances in research and applications of co_2-based demand-controlled ventilation in commercial buildings: A critical review of control strategies and performance evaluation," *Building and Environment*, vol. 223, p. 109455, 2022.

[66] Y. J. Choi, E. J. Choi, J. Y. Byun, H. J. Moon, M. K. Sung, and J. W. Moon, "Co_2- and $pm_{2.5}$-focused optimal ventilation strategy based on predictive control," *Indoor Air*, no. 1, 2025.

[67] U.S. Environmental Protection Agency, "Reference guide for indoor air quality in schools." Online; last updated April 28, 2025, 2025. Accessed July 9, 2025.

[68] A. Persily, "Challenges in developing ventilation and indoor air quality standards: The story of ashrae standard 62," *Building and Environment*, vol. 91, pp. 61–69, 2015.

[69] W. R. Chan, X. Li, B. C. Singer, T. Pistochini, D. Vernon, S. Outcault, A. Sanguinetti, and M. Modera, "Ventilation rates in california classrooms: Why many recent hvac retrofits are not delivering sufficient ventilation," *Building and Environment*, vol. 167, p. 106426, 2019.

[70] L. Morawska, D. K. Milton, *et al.*, "How can airborne transmission of covid-19 indoors be minimised?," *Environment International*, vol. 142, p. 105832, 2021.

[71] National Academy of Engineering, National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Population Health and Public Health Practice, Alper, J., and Butler, D. A., "Health effects of exposure to indoor particulate matter," in *Indoor Exposure to Fine Particulate Matter and Practical Mitigation Approaches: Proceedings of a Workshop*, p. Chapter 5, Washington, DC: National Academies Press (US), 01 2022.

[72] Á. Muelas, P. Remacha, A. Pina, E. Tizné, S. El-Kadmiri, A. Ruiz, D. Aranda, and J. Ballester, "Analysis of different ventilation strategies and co_2 distribution in a naturally ventilated classroom," *Atmospheric Environment*, vol. 283, p. 119176, 2022.

[73] L. Stabile, M. Dell'Isola, A. Russi, A. Massimo, and G. Buonanno, "The effect of natural ventilation strategy on indoor air quality in schools," *Science of The Total Environment*, vol. 595, pp. 894–902, 2017.

[74] C. K. Yu, M. Li, V. Chan, and A. C. Lai, "Influence of mechanical ventilation system on indoor carbon dioxide and particulate matter concentration," *Building and Environment*, vol. 76, pp. 73–80, 2014.

[75] S. Dhanalakshmi, M. Poongothai, and K. Sharma, "Iot based indoor air quality and smart energy management for hvac system," *Procedia Computer Science*, vol. 171, pp. 623–632, 2020.

[76] C. Vornanen-Winqvist, K. Järvi, S. Toomla, K. Ahmed, M. Andersson, R. Mikkola, T. Marik, L. Kredics, H. Salonen, and J. Kurnitski, "Ventilation positive pressure intervention effect on indoor air quality in a school building with moisture problems," *International Journal of Environmental Research and Public Health*, vol. 15, no. 2, p. 230, 2018.

[77] S. Hama, P. Kumar, A. Tiwari, Y. Wang, and P. F. Linden, "The underpinning factors affecting the classroom air quality, thermal comfort and ventilation in 30 classrooms of primary schools in london," *Environmental Research*, vol. 236, p. 116863, 2023.

[78] M. Eyada, W. Saber, M. M. El Genidy, and F. Amer, "Performance evaluation of iot data management using mongodb versus mysql databases in different cloud environments," *IEEE Access*, vol. 8, pp. 106055–106068, 2020.

[79] E. Simanjuntak and N. Surantha, "Multiple time series database on microservice architecture for iot-based sleep monitoring system," *Journal of Big Data*, vol. 9, no. 1, p. 53, 2022.

[80] B. Maag, Z. Zhou, and L. Thiele, "A survey on sensor calibration in air pollution monitoring deployments," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4857–4870, 2018.

[81] F. Concas, J. Mineraud, E. Lagerspetz, S. Varjonen, X. Liu, K. Puolam"aki, P. Nurmi, and S. Tarkoma, "Low-cost outdoor air quality monitoring and sensor calibration: A survey and critical analysis," *ACM Computing Surveys*, vol. 54, no. 8, pp. 1–44, 2021.

[82] N. H. Nguyen, H. X. Nguyen, T. T. B. Le, and C. D. Vu, "Evaluating low-cost commercially available sensors for air quality monitoring and application of sensor calibration methods for improving accuracy," *Open Journal of Air Pollution*, vol. 10, no. 1, pp. 1–21, 2021.

[83] K. Aula, E. Lagerspetz, P. Nurmi, and S. Tarkoma, "Evaluation of low-cost air quality sensor calibration models," in *Proceedings of the 2022 Web Conference Companion*, pp. 272–276, 2022.

[84] M. A. Zaidan, N. H. Motlagh, P. L. Fung, D. Lu, *et al.*, "Intelligent calibration and virtual sensing for integrated low-cost air quality sensors," *IEEE Sensors Journal*, vol. 20, no. 17, pp. 10015–10027, 2020.

[85] C. Bergmeir and J. M. Benítez, "On the use of cross-validation for time series predictor evaluation," *Information Sciences*, vol. 191, pp. 192–213, 2012. Data Mining for Software Trustworthiness.

[86] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), vol. 30, Curran Associates, Inc., 2017.

[87] R. Kumar, P. Kumar, and Y. Kumar, "Time series data prediction using iot and machine learning technique," *Procedia Computer Science*, vol. 167, pp. 534–543, 2020.

[88] E. Hitimana, G. Bajpai, R. Musabe, L. Sibomana, and J. Kayalvizhi, "Implementation of iot framework with data analysis using deep learning methods for occupancy prediction in a building," *Future Internet*, vol. 13, no. 3, p. 67, 2021.

[89] V. Knights, O. Petrovska, J. Bunevska-Talevska, and M. Prchkovska, "Machine learning models and mathematical approaches for predictive iot smart parking," *Sensors*, vol. 25, no. 7, p. 2065, 2025.