

IoT Sensor-driven Autonomous Microclimate and Safety Monitoring System for Student Housing

Zhong-Jie Liu, Jia-Chen Cai, Cheng-Fei Xia, and Shih-Pang Tseng*

School of Software and Big Data, Changzhou University of Information Technology,
No. 22, Mingxin Middle Road, Changzhou, Jiangsu 213164, China

(Received February 10, 2026; accepted September 4, 2026)

Keywords: machine learning, anomaly detection, education data mining, comprehensive evaluation

Traditional housing assessment methods rely on manual, retrospective reporting, which is prone to subjectivity and logging errors. In addition, conventional remote sensing systems focus mainly on passive microclimate logging without administrative integration. To overcome these limitations, we developed an IoT sensor-driven autonomous monitoring and evaluation system that links continuous sensor data with institutional decision-making. Heterogeneous sensor arrays in the system are used to capture gas concentrations, ambient microclimates, and electrical current waveforms at edge sampling frequencies up to 10 Hz. Streamed data are processed through an asynchronous, multithreaded producer–consumer pipeline and evaluated using a hybrid unsupervised anomaly detection framework that fuses isolation forest and the local outlier factor using a conservative ANDlogic strategy. In contrast to previous systems that produce raw anomaly flags or subjective reports, the developed system converts multisource sensor streams into objective, rule-capped housing scores explicitly designed for governance integration. Validation across 185 housing units (53000 hourly records) showed that the hybrid detector achieved a 96.8% precision, a 1.2% false positive rate, and an inference latency of 0.85 ms/sample, outperforming single-model baselines. Data processing time was reduced from 180 to less than 2 min, with 90 times speedup with a 0% computational error. The results of this research show that by coupling multirate hardware acquisition, noise-resilient dual-voting analytics, and explainable behavioral mapping, a governance-ready framework can be constructed, differing from previous anomaly detection systems that are limited to technical validation or isolated sensor deployments.

1. Introduction

The student management system is essential for managing student development in universities. As educational digital transformation pursues intelligent empowerment,^(1,2) digital transformation in higher education governance requires automated solutions that integrate advanced computing technologies with domain-specific expertise.⁽³⁾ However, the complexity of heterogeneous multi-source data and manual workflows in digital transformation often leads to inefficiencies, human error, and delayed anomaly detection. Macro-concentration level system

*Corresponding author: e-mail: tsengshihpang@ccit.js.cn
<https://doi.org/10.18494/SAM6289>

architectures, theoretical evaluation systems,^(4–6) or conventional spreadsheet-based methods are widely used, but they cannot meet the growing demand for precision and intelligence in educational management. Even advanced analytics introduced to the campus management system mainly focus on online learning behavior analysis and evaluation. In general, the current systems' performance is limited by human-dependent data collection, which introduces delays and bias in data collection.

Remote sensing technology has long been used to monitor micro-environmental parameters such as temperature, humidity, and CO₂ (CO₂) concentration in remote locations.^(1–3) The results present the feasibility of distributed sensing but mainly focus on single parameters, static deployments, or retrospective data analysis. Most systems conduct environmental monitoring only, without integrating safety indicators, anomaly detection, or direct links to student housing governance.

To overcome the limitations of the previous systems, we integrate gas, smoke, humidity/temperature sensors, and non-invasive current transformers into a unified array that captures microclimate data and electrical safety conditions in real time in a micro-environment monitoring system. Second, instead of relying on retrospective reporting, we implement an asynchronous, multi-threaded method to continuously process collected data, ensuring immediate anomaly detection and administrative visibility. By integrating unsupervised machine learning methods such as isolation forest (IF) and local outlier factor (LOF) with a conservative AND-logic strategy voting mechanism, we developed a system in this study that delivers high-confidence alerts and objective housing scores.

The developed system transforms dormitory evaluation from manual, error-prone logging into a cyber-physical monitoring system that directly supports smart campus governance. Different from previous studies, the system developed in this study links environmental data with safety compliance and institutional evaluation, providing administrators with actionable, semester-based reports. By embedding anomaly detection and tiered scoring into everyday housing management, we realized this advanced system beyond prior micro-environment monitoring toward a scalable model for intelligent residential management in higher education.

2. System Architecture

2.1 System layers

The system is designed with a layered cyber-physical IoT architecture. Its data are collected and processed by the sensor layer (data acquisition), the IoT gateway and network layer (data transmission), and the cloud analytics layer (processing and persistence) (Fig. 1).

The system architecture ensures reliable sensor data collection and transmission, scalable cloud computation, and real-time administrative visibility. Table 1 shows the commercial components, manufacturers, parameters measured, operating ranges, and installation methods.

- Sensor layer (data acquisition): Each housing unit was equipped with non-visual, non-acoustic hardware modules to capture environmental and electrical signals without breaching occupant privacy.

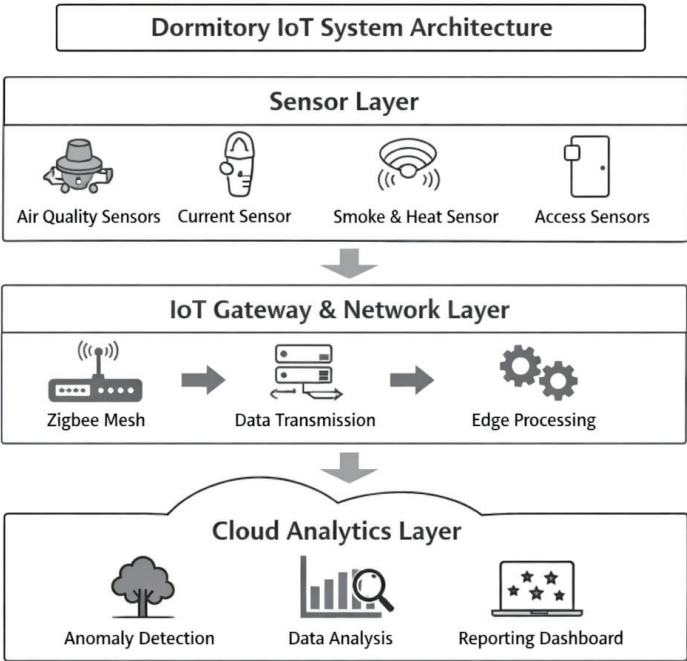


Fig. 1. Simplified system architecture with three layers responsible for data acquisition to cloud analytics (created in this study using AI).

Table 1
Sensors and their experimental configuration used in developed system.

Sensor (model)	Manufacturer	Measurement	Operating range (accuracy)	Installation and safeguard
MQ-135 (semiconductor gas sensor)	Zhengzhou Winsen Electronics, China	TVOCs, ammonia (NH ₃), (CO ₂), benzene	10–1000 ppm; high sensitivity to NH ₃ and organic gases	Mounted on central ceiling cluster; non-visual chemical sensing used to infer waste accumulation and ventilation status
DHT22/AM2302 (digital temperature and humidity sensor)	Guangzhou Aosong Electronics, China	Ambient temperature and relative humidity (RH)	–40 to +80 °C (±0.5 °C); 0–100% RH (±2%)	Mounted on central ceiling cluster; passive environmental profiling without visual monitoring
SCT-013-000 (non-invasive alternating current transformer)	Beijing Yhdc Technology, China	Alternating current load waveforms	0–30 A of AC; ±1% nonlinearity	Clipped onto main AC live line at breaker panel exterior; non-intrusive electrical metering outside living quarters
MQ-2 (combustible gas and smoke sensor)	Zhengzhou Winsen Electronics, China	Smoke, liquefied petroleum gas, propane, methane	200–10,000 ppm combustible gas concentration	Installed near ceiling fixtures in high-risk zones; detects fire hazards without visual surveillance
HC-SR501 [passive infrared (PIR) motion sensor]	Adafruit Industries, USA	Infrared thermal movement (occupancy detection)	<7 m range; 120° cone angle	Mounted at main door frame header; detects motion without video or audio recording
MC-38 (magnetic reed switch)	Shenzhen Sunfounder Intelligence Technology, China	Door's binary status (open/closed)	Operating gap 15–25 mm	Installed on door frame border; tracks entry counts without profiling indoor activity

- Air quality and microclimate array: MQ-135 gas sensors [Zhengzhou Winsen Electronics Technology Company Limited (Co., Ltd.), China] and DHT22 temperature/humidity modules (Guangzhou Aosong Electronics Co., Ltd., China) were mounted on central ceiling clusters. Waste accumulation was inferred indirectly from total volatile organic compounds (TVOCs) and NH_3 released by decaying organic matter, which increase MQ-135's resistance. Ventilation status was assessed by tracking CO_2 decay rates after occupancy peaks. No cameras or microphones were used, ensuring privacy.
- Electrical safety: SCT-013-000 non-invasive split-core current transformers (Beijing YHDC Technology Co., Ltd., China) were clipped onto the main AC live line inside breaker panels located outside student living quarters. These sensors measured current waveforms up to 30 A with a 0–1 V linear analog output, enabling the detection of prohibited high-wattage appliances.
- Fire-safety sensors: MQ-2 semiconductor sensors (Zhengzhou Winsen Electronics Technology Co., Ltd., China) were installed near ceiling fixtures to detect combustible gases and smoke particulates, providing early fire-risk alerts.
- Occupancy and entry access controls: HC-SR501 PIR motion sensors (Adafruit Industries, the United States of America) and MC-38 magnetic door reed switches (Shenzhen Sunfounder Intelligence Technology Co., Ltd., China) were mounted at entryway door frames. These devices logged entrance frequency and space utilization without monitoring indoor activity, preserving privacy.

Sensor signals were processed by Expressif system programmability 32-bit (ESP32) microcontrollers with 12-bit analog-to-digital converters, sampling at 10 Hz for current waveforms and 0.1 Hz for environmental gases. Sensor edge nodes relayed payload packets to a central campus IoT gateway via a secure Zigbee mesh network. The developed system employs an intuitive and user-friendly graphical user interface (GUI). This GUI is built with a modular design, using a tabbed main window to present score calculations, star-rating evaluations, data processing, and project documentation.

This system layout enables users to perform different tasks, making the system easy to learn and simple to operate. To ensure that the system presents the correct resolution and size, the elements on GUI are automatically adjusted using flexible vertical and horizontal layouts. The system also shows real-time progress bars and status updates, offering a smooth and responsive experience. To clarify how the system interprets its outputs, the primary environmental parameters monitored are defined as follows in this study.^(7,8)

- Ventilation status is defined as the real-time rate of indoor–outdoor air exchange. The system measures ventilation status by analyzing CO_2 concentration variations (in ppm), using the MQ-135 semiconductor sensor. Entry occupancy signals from passive infrared motion sensors provide contextual information.
- Adequate ventilation is defined as maintaining CO_2 concentrations below 1000 ppm during occupied periods. Poor ventilation is declared when continuous CO_2 accumulation exceeds 1500 ppm, indicating restricted airflow.
- Waste accumulation is measured to detect the presence of unemptied trash bins containing decaying food or organic materials. The system detects this condition through elevated levels

of TVOCs and NH_3 (in ppm). Decaying organic waste produces gaseous byproducts such as NH_3 and amine vapors, which cause sustained baseline drift in the MQ-135 sensor. These signals are particularly evident during unoccupied hours, identified by passive infrared and magnetic door reed sensors.

- Room cleanliness is quantified as a hygiene score, which is an integrated numerical index ranging from 0 to 100 points to evaluate indoor sanitation over a designated period. Instead of relying on manual visual checks, the score is calculated by integrating time-averaged concentrations of TVOCs and NH_3 , temperature, and RH collected by the DHT22 sensor. Baseline deviations in TVOC and NH_3 concentrations cause proportional deductions in the hygiene score, providing an objective measure of room cleanliness.

At the cloud analytics layer, high-volume data are decoupled from the user dashboard through an asynchronous thread-pool architecture built with PySide6 worker threads and Pandas multi-threading. PySide6 is the official Python binding for the Qt GUI. Qt applications run on a main thread (user interface thread), which handles window rendering, button clicks, and screen updates. Pandas is a data manipulation library designed for fast, vector-based data processing in Python. Incoming sensor payloads are processed using rolling-window time-series caches and stored in partitioned relational tables indexed by housing unit identity and timestamp. This design prevents thread contention, ensuring that database queries and machine learning inferences execute smoothly without freezing the administrative interface.

2.2 Algorithms

The telemetry data processing engine is constructed on the basis of Pandas, utilizing its high-performance multidimensional structures. It transforms raw, asynchronous sensor data into continuous, uniformly windowed time-series matrices. Rather than manipulating static text entries, the engine executes chronological linear interpolation to resolve missing packets, isolates and discards duplicate network telemetry waveforms, and performs scale standardization across heterogeneous sensory indicators. It mathematically aggregates continuous metrics, converting unstable raw data into structured mathematical data for machine learning algorithms.

The OpenPyXL report engine is used to abstract reporting parameters and compile summaries for decision-making support.⁽⁹⁾ Operating downstream from the machine learning layers, the engine processes sensory data matrices into standardized multi-sheet records, highlighting operational boundaries for the notifications of hazardous overcurrent or air quality degradation. To ensure the reliable monitoring of dormitory environments, the system processes high-frequency sensor data using a cloud interface that separates the visual dashboard from the backend computations. The system prevents the user interface from freezing when large datasets are analyzed. Administrative queries launch background worker threads that handle tasks such as rolling-window data cleaning, anomaly detection, and density estimation. Progress reports are continuously sent back to the dashboard, and upon completion, the system refreshes sensor heatmaps and issues verified alerts.

For model training and hyperparameter tuning, each unsupervised anomaly detection model was trained and evaluated on the same preprocessed data and feature matrix. IF was

implemented with 100 isolation trees and a contamination parameter of $\alpha = 0.05$.⁽¹⁰⁾ LOF was applied using $k = 20$ nearest neighbors with Euclidean distance.⁽¹¹⁾

2.2.1 IF

The IF method is used to identify rare, extreme sensor data by constructing random decision trees.^(12,13) Abnormal values $s(x, n)$ are isolated rapidly near the root of these trees.

$$s(x, n) = 2^{-E(h(x))/c(n)}, \quad (1)$$

where $E(h(x))$ is the expected path length of a reading in the tree and $c(n)$ is the normalization factor expressed as

$$c(n) = 2\ln(n-1) + 0.5772 - \frac{2(n-1)}{n}. \quad (2)$$

If $s(x, n) \rightarrow 1$, the reading is highly abnormal (e.g., sudden current spikes from restricted appliances).

2.2.2 LOF

LOF is used to compare each sensor's data with its neighbors to detect subtle local deviations.^(14,15)

$$LOF_k(p) = \frac{\sum_{o \in N_k(p)} \frac{lrd_k(o)}{lrd_k(p)}}{|N_k(p)|} \quad (3)$$

Here, p is the sensor reading we are checking, k is the number of nearest neighbors considered (five closest sensors), $N_k(p)$ is the set of those k nearest neighbors around p , and $d(p, o)$ is the distance between p and a neighbor o , measured as the Euclidean (straightline) distance. If $LOF_k(p) > 1$, the sensor is behaving differently from nearby sensors, for example, one room showing poor air quality, whereas others remain stable. A dormitory unit is considered abnormal only if sensor data exceed expected normal values. This dual check reduces false alarms caused by short-term fluctuations or sensor noise. Sensor data are converted into long-term performance scores as follows. Average sensor scores are calculated as

$$\bar{X}_{sensor} = \frac{1}{n} \sum_{i=1}^n X_{sensor,i}. \quad (4)$$

The safety violation frequency V_{safety} is calculated as

$$V_{safety} = \sum_{i=1}^n f(X_{sensor,i}), f(X) = \begin{cases} 1, & X \geq \text{Threshold Limit} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

Ratings follow the worst-case principle. The lowest performing score (environmental or safety) is used to determine the overall tier as follows.

$$Tier_{final} = G^{-1} \left\{ \max \left[G(L_{environment}), G(L_{safety}) \right] \right\}, \quad (6)$$

where G is used to convert descriptive levels, such as excellent, good, qualified, and unqualified, to numerical values.

Although IF and LOF are used to accurately identify multivariate anomalies, their mechanisms function as nonlinear black boxes. To understand the root cause of each flagged event, the system integrates a real-time explainability layer using Shapley additive explanations (SHAP) and treepath feature attribution.⁽¹⁶⁾ When an anomaly is confirmed by the conservative AND-logic strategy ($\hat{y}_{IF} \wedge \hat{y}_{LOF} = 1$), the system computes the local Shapley value ϕ_i for each input telemetry feature x_i .

$$\phi_i(x) = \sum_{S \subseteq F, \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f_x(S \cup \{i\}) - f_x(S)] \quad (7)$$

Here, F represents the full set of sensory features, S is a subset excluding feature i , and $f_x(S)$ is the anomaly score when only subset S is considered. The Shapley value $\phi_i(x)$ is used to quantify the contribution of feature i to the anomaly decision. By ranking these contributions, the system indicates which parameters, such as the concentrations of TVOCs, NH_3 , and CO_2 , temperature, humidity, or the root mean square (RMS) of electrical current, were responsible for pushing the anomaly score beyond the safety threshold. For example, a SHAP value of 70% means that 70% of the anomaly score was attributed to CO_2 accumulation and 20% came from the RMS of electrical current. This level of interpretability allows administrators to distinguish between ventilation problems and electrical safety violations, ensuring that corrective actions are targeted and effective.

2.3 Data security and privacy for IoT data

Considering the sensitive nature of continuous monitoring data in residential housing, we implement a multi-layered security framework spanning physical, transport, and storage tiers.⁽¹⁷⁾ We applied hardware-accelerated Advanced Encryption Standard (AES) 128-bit symmetric encryption to raw data payloads before transmission across the Zigbee mesh sub-network.

- Transport layer encryption (data in transit): Communication between local edge gateways and the central cloud analytics server is secured using Transport Layer Security version 1.3 protocols over Message Queuing Telemetry Transport Secure and Hypertext Transfer Protocol Secure. These measures prevent man-in-the-middle eavesdropping and packet injection attacks.⁽¹⁷⁾
- Storage encryption and data anonymization (data at rest): All time-series telemetry stored in the relational database is encrypted using AES 256-bit encryption. To safeguard student privacy, personally identifiable information is decoupled from sensory metrics. Physical room numbers are mapped to pseudonymous, cryptographically hashed identification tokens.
- Raw high-frequency current waveforms (10 Hz) are processed at the edge and discarded immediately; only compact statistical aggregates, such as the *RMS* of electrical current and peak power, are permanently logged.⁽¹⁷⁾
- Role-based access control (RBAC) and authentication: Access to the PySide6 administrative dashboard and analytics backend is governed by strict RBAC policies. All users must authenticate using multi-factor authentication and cryptographically signed JavaScript object notation web tokens with short-lived session lifetimes. Administrative permissions are segmented: residence hall advisors have read-only access to aggregated hygiene and safety scores for their assigned floors, whereas full data manipulation rights are restricted exclusively to system administrators.⁽¹⁷⁾

For instance, when a residence hall advisor logs into the dashboard, they can only see anonymized hygiene scores such as “Floor 3: Hygiene Score = 82/100” without access to raw telemetry or room identifiers. In contrast, system administrators monitor anomalies back to hashed identification tokens for diagnostic purposes, but never to personal student information. This layered approach ensures operational transparency, maintaining strict privacy protection.

3. Method

3.1 Anomaly detection

The system developed converts complex sensor data matrices into a standardized evaluation framework. This framework processes indicators into structured institutional metrics to assess long-term residential behavior through a multi-tier scoring model (Fig. 2).

- Sensor data monitoring and grade verification module: This module handles real-time monitoring, calibration checks, and baseline score calculations derived directly from incoming IoT telemetry pipelines.
- Longitudinal rating assessment module: This module manages multi-month evaluation cycles, processing historical sensor matrices across semesters.
- Data processing and analytics module: This module provides visualization interfaces for spatial heatmaps, sensor correlation plots, and campus-wide statistics, effectively mapping environmental profiles across the housing grid.

When processing large-scale time-series data (comprising thousands of hourly logs from hundreds of residential nodes), all machine learning models must be run asynchronously on

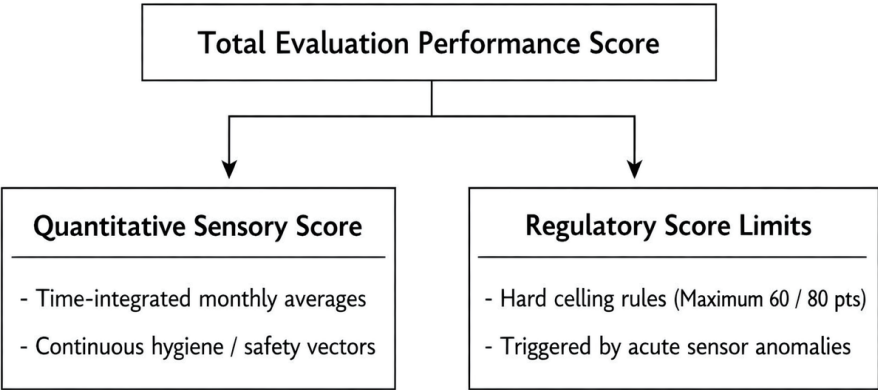


Fig. 2. Evaluation process of developed system.

background threads. This prevents the user interface from freezing, ensuring smooth data visualization and real-time continuity during peak periods of IoT data ingestion. The evaluation structure adopts a semester-based process with automated cloud computing. By continuously integrating the real-time sensor data, the system evaluates long-term performance, assigns ratings, and generates standardized reports without subjective bias. The evaluation structure is established to calculate an integrated score based on two metrics.

- Quantitative sensory score: This score represents the time-integrated monthly averages calculated from continuous sensor data streams. It is used to quantify long-term environmental and baseline power safety profiles, establishing a penalty system when metrics violate standard bounds, such as current waveforms exceeding safe power thresholds.
- Regulatory score limits: To notify the system of safety hazards, hard-coded compliance boundaries are enforced. If the sensors detect and confirm acute safety violations, such as sudden smoke spikes from the MQ-2 sensor or recurring electrical spikes from the SCT-013 transformer, the system automatically limits the total final score to a maximum of 60 or 80 points. This restriction overrides otherwise high quantitative metrics to reflect severe non-compliance.

By fusing real-time data, the system compiles an objective, quantitative performance report, providing an unalterable, data-validated foundation for semester-based student evaluations. Following the generation of baseline performance reports, a longitudinal, semester-based analysis is applied, using automated cloud computing. By continuously windowing and evaluating sensor data streams in a semester (for four months), the cloud analytics layer monitors long-term residential behavior. Housing units are assigned according to tiered ratings and standardized compliance reports, eliminating subjective human bias. To make the analytical processes accessible to administrators, the system provides a dashboard that shows sensor data and evaluation results, including scores, sensor calibration results, ratings of students, campus-wide statistics, and environmental parameters. To maintain system stability when handling large-scale time-series data comprising hourly collected sensor data from multiple dormitory units, algorithms are developed to manage efficient operations.

To complement the evaluation, the performance of different anomaly detection models was compared with that of the developed system using the 53000 h longitudinal telemetry dataset collected from 185 residential units. Raw sensor data, including concentrations of TVOCs and CO₂, RH, ambient temperature, and the *RMS* of electrical current, were aligned into hourly time-series windows using Pandas. To establish reliable ground-truth labels, domain experts and residential facility managers annotated 1200 hourly anomaly instances, such as prohibited high-wattage appliance usage, rapid gas accumulation from organic decomposition, and sensor drift. Each record was assigned a binary label (1 = anomaly, 0 = normal).

3.2 Performance comparison

One-class support vector machine (OC-SVM) was configured using a radial basis function kernel to map complex, nonlinear feature distributions into a higher-dimensional decision space. The kernel scale parameter was set to automatically compute the kernel coefficient on the basis of the variance of the input features to prevent overfitting. The marginal decision boundary parameter was set to 0.05, fixing the upper bound on the proportion of training errors (expected outliers) and the lower bound on the number of support vectors at 5%.⁽¹⁸⁾ The autoencoder was designed with a five-layer symmetric dense architecture (5 → 3 → 2 → 3 → 5 nodes), trained for 50 epochs using the Adam optimizer (learning rate = 0.001), and declared anomalies when reconstruction error exceeded the 95th percentile.⁽¹⁹⁾ The hybrid model (IF + LOF with AND-logic strategy) of the developed system computed ensemble predictions by evaluating IF and LOF independently and declaring anomalies only when both criteria were simultaneously violated.⁽²⁰⁾ Prediction accuracy was compared using ground-truth labels to calculate precision, recall, F1-score, and false positive rate using the following equations. Inference speed (ms/sample) was measured on a single Intel Core i7 processor by averaging 10000 consecutive single-sample prediction cycles.

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

Here, *TP* denotes the number of true positives, which are correctly identified sensor data anomalies, and *FP* denotes the number of false positives, which are normal instances incorrectly flagged as anomalies. Precision is used to measure the proportion of correctly detected sensor data out of all instances flagged as sensor data anomalies.

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

Here, *FN* is the number of false negatives, sensor data anomalies that were missed and classified as normal. Recall is used to measure the proportion of actual anomalies that were successfully detected.

$$F1\text{-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

F1-score is the harmonic mean of precision and recall, balancing the tradeoff between detecting anomalies and avoiding false alarms.

$$\text{False positive rate} = \frac{FP}{FP + TN} \quad (11)$$

Here, TN is the true negative, a correctly identified normal instance. The false positive rate indicates how often normal conditions are mistakenly classified as sensor data anomalies.

3.3 System application

To validate the stability and analytical accuracy, the developed dormitory evaluation system was deployed in the smart-campus infrastructure of the School of Software and Big Data, Changzhou University of Information Technology in China. The system validation was conducted in the second semester of the 2024–2025 academic year from February to July. Data were collected from 185 student housing units, accommodating more than 1400 students across 51 classes. In that semester, a longitudinal time-series dataset comprising more than 53000 hourly data was constructed. The dataset included continuous environmental and electrical data streams, including environmental indicators, safety parameters (smoke and temperature), current peaks, and air quality indicators (CO_2 and TVOC concentrations). The collected data were processed to identify anomalies. Table 2 shows the performance of the developed system during the validation period.

4. Results and Discussion

4.1 Dashboard of developed system

The developed system showed the high precision and correct identification of acute and chronic deviations. Sensor nodes stably monitored air quality and electrical safety parameters,

Table 2
Anomaly detection results of developed system on 19 February 2026.

Sensor node	Implementation location	Outcome of IF	Outcome of LOF	Status and diagnostic result
2025251101	Building A 101	Normal	Normal	Nominal–optimal air/power profiles
2025251102	Building A 102	Normal	Normal	Nominal–optimal air/power profiles
2025251103	Building A 103	Anomalous	Anomalous	Anomalous–severe indoor air degradation
2025251104	Building B 201	Normal	Normal	Nominal–optimal air/power profiles
2025251105	Building B 202	Anomalous	Anomalous	Anomalous–high wattage appliance violation
2025251106	Building C 301	Anomalous	Anomalous	Anomalous–hardware fault/edge disconnection
2025251107	Building C 302	Normal	Normal	Nominal–optimal air/power profiles
2025251108	Building D 104	Anomalous	Anomalous	Anomalous–transducer calibration error

accurately revealing excessive TVOC concentrations, illegal appliance usage, and sensor disconnections. The dual-warning framework (IF and LOF) ensured that anomalies were confirmed only when both algorithms agreed, minimizing false positives. By integrating multiple sensor data with unsupervised learning outputs, a real-time, non-invasive assessment of the safety and environmental conditions of the dormitory was conducted on the GUI of the system. The microclimate and safety analysis interface shows summaries with time-series graphs of environmental parameters. Real-time anomalies are presented in red letters and symbols for immediate intervention (Fig. 3). The sensor data dashboard provides engineering-level correlation analysis results. Line charts are presented to show environmental and electrical safety levels, and pie charts and bar graphs are displayed to show score distributions and weekly variations. To process real-time sensor streams without latency or packet loss, the data pipeline decouples the presentation layer from backend computations using an asynchronous, multi-threaded producer–consumer architecture. The system’s GUI presents administrative commands and updates live statuses, also conducting data cleaning, the linear interpolation of missing packets, path-length partitioning (IF), and data matrix updates (LOF). The dashboard metrics are continuously updated.

4.2 Performance evaluation

The performance of the developed system was compared with that of the traditional manual method under identical operational conditions using a longitudinal dataset of 53000 hourly data points collected across 185 residential units. The developed system demonstrated substantial improvements in administrative efficiency, data fidelity, and evaluation precision compared with manual workflows (Table 3).

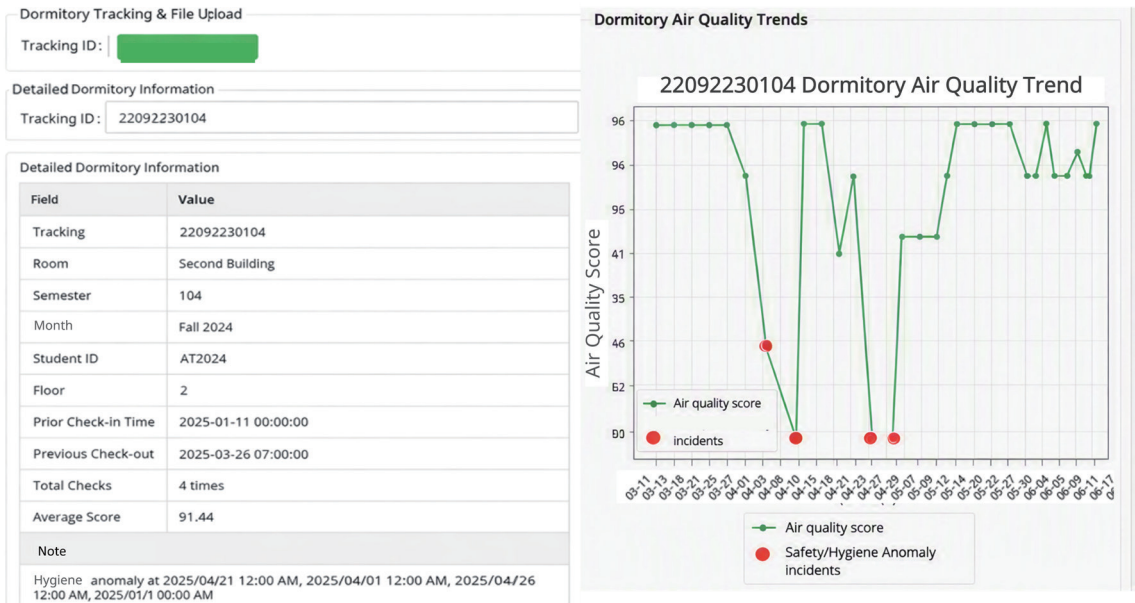


Fig. 3. (Color online) Sample GUI of developed system.

Table 3
Performance comparison of developed system and manual method.

Evaluation parameter	Manual method	Developed system	Advantage of developed system
Data processing time	>180 min	<2 min	Operational efficiency increased by more than 90 times
Error rate	Frequent human logging errors; repeated audits	0%	Complete mathematical precision (100% accuracy)
Reporting	Single static text-based sheets	Four automated, interactive dashboards	Rich, multidimensional analytics
Anomaly detection capability	Prone to human oversight; transient hazards missed	Real-time detection using IF + LOF	Continuous, fine grained tracking

Data processing time was reduced from more than 180 min in the manual method to less than 2 min in the automated system, representing an increase in operational efficiency by more than 90 times. This marked reduction highlights the system's ability to handle high-volume telemetry streams in near real time, a critical requirement for smart-campus deployments. Manual workflows were prone to frequent human logging errors and required repeated audits to maintain accuracy. In contrast, the developed system achieved complete mathematical precision, eliminating computational errors (0% error rate). This ensures that anomaly detection and hygiene scoring are based on reliable, reproducible data.

Whereas the manual method relied on static, text-based sheets, the automated system generated four interactive dashboards built on the PySide6 framework. These dashboards provided rich, multidimensional analytics, enabling administrators to visualize anomaly distributions, hygiene scores, and environmental trends in real time. In manual monitoring, transient hazards, such as short-lived electrical spikes or rapid gas accumulation, might not be detected. In contrast, the developed system employed a hybrid anomaly detection model that combines IF and LOF with conservative AND-logic. This approach ensured the continuous, fine-grained tracking of anomalies, reducing false positives while maintaining responsiveness.

The validation results showed the system's scalability, accuracy, and operational robustness. By transforming error-prone manual workflows into a real-time, data-driven monitoring ecosystem, the system provides faster data processing, zero errors, and continuous anomaly detection. A transient CO₂ accumulation event that might be overlooked in manual monitoring was detected in real time by the hybrid IF + LOF model. The system flagged the anomaly in milliseconds, attributed 70% of the anomaly score to CO₂ concentrations and 20% to the *RMS* of electrical current, and displayed the event on the dashboard with clear root-cause attribution. Such interpretability ensures that administrators can act immediately on the underlying cause rather than the symptom. These results show the system's readiness for large-scale smart-campus integration and its potential to serve as a benchmark model for intelligent residential management systems.

4.3 Comparison of anomaly detection ability

To evaluate anomaly detection accuracy, the performance of the hybrid IF and LOF model was compared with those of standalone unsupervised baselines using the 53000 h longitudinal dataset (Table 4). The conservative AND-logic strategy enabled the highest precision (96.8%)

Table 4
Performance comparison of anomaly detection frameworks on the validation dataset (53000 h records).

Anomaly detection method	Precision (%)	Recall (%)	F1-score (%)	False positive rate (%)	Inference speed (ms/sample)
IF ⁽¹⁰⁾	84.2	92.5	88.1	5.8	0.42
LOF ⁽¹¹⁾	81.5	90.1	85.6	6.9	0.68
OC-SVM ⁽¹⁸⁾	78.3	86.4	82.1	8.2	1.15
Autoencoder (deep learning) ⁽¹⁹⁾	88.6	89.2	88.9	3.9	2.8
IF + LOF with AND-logic strategy (this study)	96.8	88.7	92.6	1.2	0.85

and the lowest false positive rate (1.2%).⁽²⁰⁾ Although single models such as IF and LOF offered slightly higher recall, they generated frequent false alarms triggered by ambient noise and transient electrical spikes. Autoencoders delivered competitive F1-scores but required longer inference times, limiting their suitability for real-time dormitory monitoring. In contrast, the hybrid IF + LOF model of the developed system balanced accuracy and computational efficiency, ensuring that anomalies were flagged only when both detectors agreed. This minimized false positives while maintaining responsiveness, making the framework well-suited for high-density sensor deployments in smart campus environments.

4.4 Passive monitoring to active decision-making

Standard remote sensing systems can track microclimate parameters such as temperature, humidity, and gas concentrations in agricultural, structural, or outdoor environments.^(21,22) On a smart campus, IoT sensor nodes are deployed for facility maintenance,⁽²³⁾ but most sensors are only used to provide raw data or simple threshold alerts. In contrast, the developed system embeds sensor outputs into administrative decision-making. It couples high-frequency electrical load monitoring with gas and thermal sensing to capture both gradual air-quality decline and sudden electrical safety violations. A dual-voting anomaly detection pipeline (IF + LOF) transforms continuous time-series data into rule-capped institutional scores, ensuring that anomalies translate into actionable governance metrics. In this way, passive environmental monitoring becomes an active, behavior-aware management tool for student housing.

The autonomous sensor nodes used in the system significantly reduce reporting errors on air quality, humidity, temperature, space utilization, and CO₂ and TVOC concentrations. To process high-traffic IoT data streams without latency, the system implements a multi-threaded producer–consumer pipeline using Pandas,^(24,25) sampled at a rate of 0.1 Hz. Given the volatility of time-series microclimate logs in dense residential nodes,⁽²⁵⁾ the system employs an ensemble unsupervised pipeline to stabilize indicators before mapping them to institutional scores.^(12,13)

For the integrated AND-logic, IF and LOF are used for global isolation and local density analysis, respectively, counteracting individual limitations and suppressing false positives from hardware spikes or noise.⁽²⁶⁾ In addition, by combining environmental parameters with occupancy logs, the system ensures the accurate monitoring of anomalies in collected data. This is employed to avoid reliance on pre-labeled datasets, making it well-suited for

resource-constrained edge deployments.⁽²⁷⁾ The tiered grade-mapping function of the developed system is applied on the basis of the worst-case principle to ensure non-error notification for anomalies and potential hazards. The developed system shows how multi-sensor fusion, asynchronous computation, and unsupervised learning must be integrated into educational infrastructures. The system showed enhanced efficiency and data reliability, providing a reference for intelligent, behavior-aware management methods of smart campuses and other high-density living spaces.

The autonomous sensor nodes used in the system significantly reduce reporting errors on air quality, humidity, temperature, space utilization, and CO₂ and TVOC concentrations. To process high-traffic IoT data without latency, we implement a multi-threaded producer–consumer data stream in the developed system using Pandas that samples at a rate of 0.1 Hz. Considering the volatility of time-series microclimate data in dense residential nodes,⁽²⁵⁾ an ensemble unsupervised model is employed to stabilize data and convert them to institutional scores.^(12,13)

4.5 Scalability and network bandwidth optimization

Scalability and network bandwidth optimization are critical for the system when expanding from 185 nodes to campus-wide deployments involving thousands of active units. Large-scale deployments introduce risks of wireless channel congestion and cloud compute bottlenecks. To mitigate these bottlenecks, a dual-tier edge computing method is adopted. ESP32 microcontrollers are used to process raw current waveforms (10 Hz) using peak detection algorithms. Instead of transmitting continuous high-frequency arrays, sensor nodes send compact feature vectors, such as the *RMS* of electrical current and mean gas concentrations, to Zigbee mesh gateways, reducing network load.⁽²⁸⁾

Scaling limits must be considered when deploying smart campus infrastructure across thousands of distributed nodes. Local edge feature extraction reduces raw data transmission across Zigbee networks, preventing packet loss and channel saturation during peak usage.⁽²⁸⁾ At the cloud layer, multi-threaded worker pools and time-series indexing isolate database operations from graphical interface visualization, enabling the scalable processing of large telemetry streams. The comparison results show that the integration of global isolation with local density analysis enables a higher precision than single-model anomaly detection.⁽²⁰⁾

5. Conclusion

An autonomous IoT microclimate and safety monitoring system was developed and validated for high-density student housing, differing from previous systems that have emphasized single-parameter tracking, static deployments, or passive environmental logging. The developed system links heterogeneous sensor data, including gas concentrations, ambient microclimates, high-frequency electrical waveforms, and door-entry dynamics, with an automated institutional evaluation and scoring mechanism. Whereas previous systems yield raw anomaly flags or subjective reports, the developed system transforms continuous sensor streams into objective, rule-capped housing scores designed for governance integration. The hybrid anomaly detection

strategy of the developed system mitigates hardware drift, transient noise, and false positives by combining IF and LOF through a conservative AND-logic strategy. The system's explainable analytics enables evidence-based decision-making for acute hazards and gradual environmental degradation. Performance validation results across 185 residential units showed enhanced precision, reduced false positive rates, and a marked reduction in processing time from more than 180 to less than 2 min, eliminating human logging errors and enabling proactive, behavior-aware management.

By embedding deterministic rule-based score boundaries, we developed a system that advances technical validation to a governance-ready framework that supports institutional decision-making. Although the results of this study present significant operational gains, it is necessary to employ cloud-native architectures, integrate advanced sequence models for predictive forecasting, and standardize interfaces for seamless interoperability with campus systems. The developed system enables a scalable, precision-driven, and resilient paradigm by coupling multi-rate hardware acquisition with explainable, policy-enforceable analytics, thereby advancing the digital transformation of smart-campus and high-density urban residential management.

Acknowledgments

This research was supported by the Jiangsu Province Industry-University-Research Collaboration Project, “Tourist Situation Analysis and Intelligent Decision-Making System Development Based on Data Collection Bracelets (Project No. BY20240343)” and the “Education and Teaching Reform Project of Changzhou University of Information Technology, Research on the Ecosystem of AI Courses and Three-Dimensional Textbook Construction in Colleges Under the Background of Digital Intelligence Empowerment (Project No. 2026CXJG05).”

References

- 1 S. Yan and Y. Yang: ECNU Rev. Educ. **5** (2022) 1. <https://files.eric.ed.gov/fulltext/EJ1301603.pdf>
- 2 M. Bin: Int. J. Adv. Comput. Sci. Appl. **15** (2024) 289. https://thesai.org/Downloads/Volume15No5/Paper_30-Design_and_Implementation_of_an_Information_Management_System.pdf
- 3 Q. Wang, Z. Ma, and Q. Miao: Beijing Int. Rev. Educ. **7** (2025) 254. <https://doi.org/10.1177/25902547251395880>
- 4 L. Nong, G. Liu, C. Tang, and Y. Chen: Sustainability **15** (2023) 4732. <https://doi.org/10.3390/su15064732>
- 5 C. Han, B. Cheng, and Y. Jiang: Proc. 2025 2nd Int. Conf. Artificial Intelligence and Future Education (ACM, 2025) 514. <https://doi.org/10.1145/3785987.3786072>
- 6 Y. J. Wang, C. L. Gao, and X. D. Ye: Educ. Inf. Technol. **29** (2024) 11645. <https://doi.org/10.1007/s10639-023-12238-x>
- 7 K. W. Bae, E. J. Choi, Y. J. Choi, J. Y. Yun, G. Y. Yun, H. J. Moon, and J. W. Moon: Build. Environ. **285** (2025) 113568. <https://doi.org/10.1016/j.buildenv.2025.113568>
- 8 M. Szyłak-Szydłowski, W. Kos, R. Tarakowski, M. Tkaczyk, and P. Borowik: Sensors **25** (2025) 3153. <https://doi.org/10.3390/s25103153>
- 9 OpenPyXL: <https://openpyxl.readthedocs.io/> (accessed May 2026).
- 10 F. T. Liu, K. M. Ting, and Z.-H. Zhou: Proc. 2008 IEEE 8th Int. Conf. Data Mining (IEEE, 2008) 413.
- 11 A. Adesh, S. G. J. Shetty, and L. Xu: J. Parallel Distrib. Comput. **192** (2024) 104923. <https://doi.org/10.1016/j.jpdc.2024.104923>
- 12 G. Siemens and R. S. J. d. Baker: Proc. 2012 2nd Int. Conf. Learning Analytics and Knowledge (ACM, 2012) 252. <https://doi.org/10.1145/2330601.2330661>

- 13 Y. Bai, Z. Zhang, L. Tang, K. Yang, and R. Yang: Proc. 2023 3rd Int. Conf. Education, Information Management and Service Science (Atlantis, 2023) 74. https://doi.org/10.2991/978-94-6463-264-4_74
- 14 F. T. Liu, K. M. Ting, and Z.-H. Zhou: ACM Trans. Knowl. Discov. Data **6** (2012) 1. <https://doi.org/10.1145/2133360.2133363>
- 15 M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander: Proc. 2000 ACM SIGMOD Int. Conf. Management of data (ACM, 2000) 93. <https://doi.org/10.1145/342009.33538>
- 16 S. Park, Y. Chen, and K. Watanabe: Sens. Mater. **36** (2024) 1885. <https://doi.org/10.18494/SAM.2024.5310>
- 17 S. Thiruvengkatasamy, V. Boobalan, L. Gokul, V. Keerthi, and S. Logeshwaran: Proc. 2025 10th Int. Conf. Smart Structures and Systems (IEEE, 2025) 1. <https://doi.org/10.1109/ICSSS66939.2025.11346435>
- 18 H. J. Shin, D.-H. Eom, and S.-S. Kim: Comput. Ind. Eng. **48** (2005) 395. <https://doi.org/10.1016/j.cie.2005.01.009>
- 19 Z. P. Xu, Z. J. Cheng, and B. Guo: Proc. 2022 4th Int. Conf. System Reliability and Safety Engineering (IEEE, 2022) 231. <https://doi.org/10.1109/SRSE56746.2022.10067443>
- 20 E. Krzysztoń, I. Rojek, and D. Mikołajewski: Appl. Sci. **14** (2024) 11545. <https://doi.org/10.3390/app142411545>
- 21 N. Madrid, R. Boulton, and A. Knoesen: Build. Environ. **119** (2017) 128. <https://doi.org/10.1016/j.buildenv.2017.04.010>
- 22 M. Afrial, M. Rauf, M. Nouman, M. T. Khan, M. A. Rizwan, and N. Ahmad: Mater. Proc. **23** (2025) 12. <https://doi.org/10.3390/materproc2025023012>
- 23 S. Manokeaw, P. Khuwuthyakorn, Y.-C. Chan, N. Tengtrairat, M. Jintapitak, O. Thinnukool, C. Buachart, T. Sinthamrongruk, T. K. N. Ayutthaya, N. Suriyanon, S. Kanangkaew, and D. Rinchumphu: Technologies **13** (2025) 439. <https://doi.org/10.3390/technologies13100439>
- 24 QT: <https://www.qt.io/qt-for-python> (accessed May 2026).
- 25 X. Cao, J. Chen, Y. Zhang, and Y. Sun: ISA Trans. **47** (2008) 247.
- 26 A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab: IEEE Access **9** (2021) 78658. <https://doi.org/10.1109/ACCESS.2021.3083060>
- 27 C. Romero and S. Ventura: WIREs Data Min. Knowl. Discov. **3** (2012) 12. <https://doi.org/10.1002/widm.1075>
- 28 H. M. Saad, A. Shdefat, A. Nawaz, A. M. El-Sherbeeny, M. A. El-Meligy, and M. R. R. Rana: PeerJ Comput. Sci. **10** (2024) e2269. <https://doi.org/10.7717/peerj-cs.2269>

About the Authors



Zhong-Jie Liu received his B.S. degree from Xinyang Normal University, China, in 2006 and his M. E. degree from Changzhou University, China, in 2010. Since 2021, he has been an associate professor at Changzhou University of Information Technology. His research interests include AI agents, robot control systems, deep learning, and large model applications. (156247874@qq.com)



Jia-Chen Cai graduated from Changzhou University of Information Technology in 2025. He is currently employed as an engineer at an enterprise and concurrently serves as the corporate mentor for the AI Master Teacher Studio at Changzhou University of Information Technology. His research interests include AI, machine automation control, and IoT technology applications. (cai-jia-chen@foxmail.com)



Cheng-Fei Xia is an undergraduate student majoring in Software Technology at Changzhou University of Information Technology. He is a research assistant, contributing to projects on the application of large-scale model technologies. His research interests include data analysis, large model technology applications, as well as IoT application development and technical implementation. (1526196735@qq.com)



Shih-Pang Tseng received his B.S. and M.S. degrees in 1990 and 1992, respectively, from the Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan, and his Ph.D. degree in 2013. from the Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan. He is a professor at Changzhou University of Information Technology. His current research interests include deep learning, natural language processing, robotics, and learning technology. (tsengshihpang@ccit.js.cn)