Edge AI-Based Multimodal Sensing for Real-Time Anomaly Detection in Intelligent Building Environments

Rochelle Hogg
Department of Computer Science, University of New Hampshire, Durham, NH, USA.
rochelle.hogg@unh.edu

Kiandra Yolanda
School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA.
yolanda1970@oregonstate.edu

Abstract

Intelligent buildings generate heterogeneous measurements from environmental sensors, power meters, occupancy detectors, cameras, acoustic devices, and building automation systems. The central problem is not simply detecting isolated outliers, but recognizing context-dependent and collective anomalies while respecting tight latency, energy, memory, privacy, and reliability constraints at the edge. This manuscript develops a methodological framework for edge AI-based multimodal sensing and real-time anomaly detection in intelligent building environments. The contribution is an integrated design-and-analysis model that connects sensing, synchronization, modality-specific feature extraction, uncertainty-aware fusion, anomaly scoring, risk-calibrated decision logic, and constrained deployment. The framework combines clock-aware preprocessing with temporal encoders, an adaptive gated fusion rule, a hybrid generative-predictive anomaly score, hysteresis-based alarm decisions, and a resource-constrained placement policy across sensors, gateways, and embedded accelerators. The analysis explains why static concatenation is fragile under missing modalities and sampling-rate mismatch, why asymmetric false-alarm and missed-fault costs require calibrated thresholds rather than fixed percentile cutoffs, and why occupancy and weather drift demand controlled adaptation. Boundary conditions include sensor drift, clock skew, label scarcity, model overconfidence, and privacy-sensitive visual data. The resulting framework provides a technically defensible basis for designing and evaluating edge-resident building anomaly detectors before field deployment.

Keywords

edge AI; multimodal fusion; anomaly detection; intelligent buildings; real-time sensing; uncertainty calibration

1. Introduction

Anomaly detection in intelligent buildings concerns the identification of deviations that matter for comfort, energy performance, equipment reliability, indoor air quality, and safety. The detection target may be a point anomaly such as a sudden power spike, a contextual anomaly such as overheating during unoccupied hours, or a collective anomaly such as a slowly developing chiller fault. Classical anomaly detection formulations emphasize statistical deviation from nominal behavior [1], while deep learning provides representation-

learning tools for high-dimensional and nonlinear sensor data [2]. Edge computing adds a deployment requirement: inference must occur close to sensors under bounded latency and limited energy [3], and the practical emergence of edge systems has made local computation a first-class architectural concern [4].

Multimodal sensing is essential because building states are rarely observable through one signal. Temperature, humidity, carbon dioxide, power, current, valve position, vibration, acoustic emissions, and occupancy provide complementary evidence. Multimodal machine learning studies how heterogeneous representations can be aligned, fused, and used jointly [5], and earlier fusion research distinguishes feature-level, decision-level, and intermediate strategies [6]. In buildings, however, fusion is complicated by asynchronous sampling, missing channels, modality-specific noise, and unequal diagnostic value across operating modes. A detector that performs well on a laboratory dataset with complete modalities may fail when a gateway loses a camera stream, a power meter reports at a different rate, or occupancy sensors become stale.

The central contribution of this manuscript is a connected methodological framework for edge AI-based multimodal anomaly detection in intelligent buildings. It is not an account of a completed deployment or a benchmark experiment. Rather, it formalizes the design problem from sensing to decision and derives the consequences of specific architectural choices. The framework treats synchronization, modality encoding, adaptive fusion, anomaly scoring, threshold calibration, and resource-aware placement as linked problems. It also specifies failure modes and evaluation logic that follow from the physical and computational setting. This approach is appropriate because the dominant risks in building anomaly detection are not merely model-capacity limitations; they are temporal misalignment, nonstationarity, asymmetric alarm costs, and deployment constraints that must be designed into the system.

2. Problem Formulation and Multimodal Sensing Model

An intelligent building can be represented by a physical state vector that includes zone temperature, humidity, airflow, occupancy, equipment status, and power demand. Sensors observe this state imperfectly. For a modality m at time t, the observation is modeled as:

$$y_m(t) = f_m\big(x(t), \eta_m(t)\big) + \varepsilon_m(t)$$

Here $y_m(t)$ is the measured value, $x(t)$ is the underlying physical state, $\eta_m(t)$ captures sensor response and environmental conditions, and $\varepsilon_m(t)$ is measurement noise. The function f_m is modality-specific and may be nonlinear, hysteretic, or saturating. This model matters because anomalies can arise from changes in $x(t)$, from sensor faults in η_m, or from noise bursts in ε_m. A detector that attributes every deviation to the building state will confuse sensor drift with equipment failure.

The available modalities include environmental channels, electrical measurements, occupancy indicators, acoustic or vibration signals, and visual or thermal imagery. They differ in sampling rate, latency, reliability, and privacy sensitivity. A power meter may report at one-second intervals, a temperature sensor at one-minute intervals, and a camera at several frames per second. The first preprocessing task is therefore temporal alignment. For each modality i, a local window is transformed into a feature representation:

$$z_i(t) = \varphi_i(y_i(t - \tau_i{:}\, t); \theta_i)$$

In this expression, τ_i is the lookback window, $y_i(t{-}\tau_i{:}t)$ is the retained sample history, φ_i is a modality-specific encoder, and θ_i are its parameters. The encoder may be a one-dimensional

convolutional network, a recurrent network, a small transformer, or a handcrafted statistical extractor. The choice should reflect sampling regularity and available computation. Recurrent structures are natural for irregular temporal dependencies [9], while attention mechanisms can weight informative time steps and cross-modal evidence [10].

The anomaly detection problem is not defined by a single score. Point anomalies are localized in time, contextual anomalies depend on operating conditions, and collective anomalies involve a pattern across a window. The detector must also handle nonstationarity. Weather, occupancy, set points, and equipment schedules change, so the data distribution shifts over time. Concept-drift research distinguishes gradual, abrupt, recurring, and incremental changes, each requiring different adaptation logic [11]. Edge intelligence provides an architectural route for local inference and selective collaboration [12], while building energy studies demonstrate that real-time anomaly detection is feasible when the detection target and operating context are defined explicitly [13]. Prior design work on multimodal perception systems for smart environments further supports treating sensor selection, placement, and fusion as a coordinated design problem rather than an isolated modeling step [14].

3. Edge AI Architecture and Fusion Mechanism

The proposed architecture distributes computation across three tiers: sensor nodes, local edge gateways, and optional building-level or cloud services. Sensor nodes perform sampling, basic validation, timestamping, and lightweight filtering. Edge gateways perform synchronization, feature extraction, fusion, anomaly scoring, and alarm generation. Building-level services support model updates, fleet monitoring, and long-horizon analytics. This hierarchy is consistent with fog and edge computing principles that place latency-sensitive processing near data sources while retaining centralized functions for coordination [15]. It also aligns with building-lifecycle evidence that machine learning value depends on data quality, commissioning, and maintenance workflows rather than on model choice alone [16].

Fusion is the pivotal design decision. Early concatenation of raw signals is simple but brittle when modalities have different noise levels, missingness, and dimensionalities. Late fusion is robust to missing streams but can discard cross-modal interactions that reveal faults. An intermediate adaptive fusion rule is preferable for building data:

$$z(t) = \sum_m \alpha_m(t) g_m\big(z_m(t)\big), \alpha_m(t) \geq 0, \sum_m \alpha_m(t) = 1$$

Here $z_m(t)$ is the encoded representation from modality m, g_m is a modality-specific projection, and $\alpha_m(t)$ is a normalized weight. The weight can depend on signal quality, freshness, predictive reliability, and cross-modal consistency. If a camera stream is occluded or privacy-disabled, its weight should fall rather than forcing the model to impute a confident but unsupported feature. If a power signal and an occupancy signal disagree during working hours, the fusion layer can preserve both as evidence rather than averaging them into a single ambiguous state.

The anomaly model can combine generative reconstruction with predictive residual analysis. A variational autoencoder provides a latent-variable representation and a reconstruction objective:

$$L = \mathbb{E}[\|y - D(E(y))\|^2] + \lambda \cdot KL(q(z|y)\|p(z))$$

In this expression, E is the encoder, D is the decoder, q(z|y) is the approximate posterior over latent variables, p(z) is the prior, λ controls the regularization strength, and KL denotes Kullback-Leibler divergence. This objective is established in variational inference [7]. Deep anomaly detection reviews emphasize that reconstruction error and density estimates have different inductive biases and failure modes [8]. A building detector should therefore not rely on one objective alone. The predictive component can capture temporal regularity, while the reconstruction component can capture unusual combinations of simultaneous sensor values.

A hybrid score can be written as:

$$s(t) = w_r \|r(t)\|^2 - w_p \log p(y(t)|y(t-H:t-1))$$

Here r(t) is the reconstruction residual, p(y(t)|y(t−H:t−1)) is a predictive density over the next observation given a context window H, and w_r and w_p are nonnegative weights. The first term responds to combinations that the generative model cannot reconstruct; the second responds to temporal patterns that the predictive model finds improbable. In practice, the weights should be selected by validation on normal operating periods and checked separately for occupied, unoccupied, heating, cooling, and economizer modes. Otherwise, a model may learn that all nighttime operation is anomalous simply because nighttime data were underrepresented.

4. Anomaly Scoring, Decision Logic, and Calibration

A continuous anomaly score is not yet an alarm. Real-time operation requires a decision rule that controls false alarms and missed detections. A basic rule with confidence gating is:

$$a(t) = 1 \; if \; s(t) > \kappa(t) \; and \; c(t) \geq c_{min}, \; else \; 0$$

Here a(t) is the alarm indicator, κ(t) is an adaptive threshold, c(t) is a calibrated confidence estimate, and c_{min} is the minimum confidence for actuation or operator notification. The confidence term prevents the system from raising alarms when input quality is poor, when a modality is stale, or when the model is operating outside its validated regime. Hysteresis should be applied so that an alarm does not chatter around the threshold. A common strategy is to require k consecutive exceedances for alarm onset and l consecutive non-exceedances for clearing, with k and l selected from the required response time and tolerated false-alarm rate.

Threshold selection should reflect asymmetric costs. Missing a chiller fault may be more costly than a brief false alarm, but excessive false alarms degrade operator trust and may cause alarm fatigue. If π is the prior probability of an anomaly, F_0 and F_1 are the score distributions under normal and anomalous conditions, and c_0 and c_1 are the costs of false positives and false negatives, the cost-minimizing threshold satisfies:

$$\kappa* = argmin_\kappa \left[c_0(1-\pi)F_0(\kappa) + c_1\pi(1-F_1(\kappa))\right]$$

This expression makes the calibration problem explicit. It also shows why a fixed percentile threshold is insufficient when the anomaly prior or the cost ratio changes across building zones. In practice, F_0 and F_1 are not known exactly, so thresholds must be estimated from labeled fault events, maintenance records, simulation studies, or conservative expert priors. Evaluation should report precision and recall at the selected operating threshold, the area under the precision-recall curve, detection delay, and false alarms per zone per week. Accuracy alone is inappropriate because anomalies are rare and because a detector that never alarms can appear numerically strong.

Calibration must also address uncertainty. A neural network may produce a score with low numerical variance but high epistemic uncertainty if the input modality combination was absent during training. Ensemble disagreement, temperature scaling, or distance to the training manifold can provide a confidence signal. The confidence gate should be validated separately from the anomaly score, because a detector can rank anomalies well while producing poorly calibrated probabilities. In buildings, the temporal dimension matters as well: an alarm raised after a delay may still be useful for maintenance but unacceptable for freeze protection or smoke-related safety functions. The decision rule must therefore be matched to the consequence of the event, not only to the statistical properties of the score.

5. Resource-Aware Deployment and Technical Trade-offs

Edge deployment introduces constraints that are absent from offline anomaly detection. A gateway has finite compute, memory, energy, and thermal headroom. It may serve multiple zones, run several models, and forward only selected events. The placement and scheduling problem can be expressed as a constrained policy choice:

$$min_\{\omega \in \Omega\}\, \mathbb{E}[L_\omega]\; s.t.\; \mathbb{E}[T_\omega] \le T_{max},\, \mathbb{E}[\varepsilon_\omega] \le \varepsilon_{max},\, \mathbb{E}[M_\omega] \le M_{max}$$

Here ω is a deployment policy from the feasible set Ω, L_ω is the detection loss, T_ω is end-to-end latency, ε_ω is energy consumption, and M_ω is memory use. The bounds T_{max}, ε_{max}, and M_{max} encode the operational envelope. This formulation clarifies why a high-accuracy model that exceeds the latency bound is not admissible for real-time control or fast maintenance response. Quantization, pruning, compact recurrent encoders, and event-triggered inference can reduce resource use, but each introduces a trade-off. Quantization may reduce numerical precision at decision boundaries; pruning may remove rare but diagnostically important features; event-triggered inference may miss short transients.

The architecture should also decide what data leave the edge. Raw video and audio are privacy-sensitive, and centralized transmission increases bandwidth and exposure. Local feature extraction can preserve diagnostic value while reducing raw data transfer. Wireless sensor networks make energy and communication constraints explicit [19], and occupancy inference from electricity data illustrates how a single signal can support multiple building functions without additional sensing infrastructure [17]. Supervisory and optimal control of HVAC systems further shows that building decisions are embedded in control loops with their own stability and comfort constraints [18]. An anomaly detector that ignores these loops may generate technically correct but operationally unusable alarms. The practical design rule is to place computation where it reduces latency or privacy risk, not merely where it is computationally convenient.

6. Reliability, Failure Modes, and Evaluation Logic

The framework's reliability depends on failure modes that are specific to building environments. Sensor drift changes $\eta_m(t)$ slowly and can be mistaken for a gradual building fault. Clock skew misaligns windows and creates spurious cross-modal correlations. Missing modalities can cause the fusion layer to over-weight a single noisy stream. Model overconfidence can produce high-confidence alarms on out-of-distribution inputs, especially after a schedule change or a sensor replacement. Concept drift can make a previously valid threshold obsolete. These are not secondary implementation details; they determine whether the detector remains useful after commissioning.

Evaluation should therefore be stratified by operating mode, season, occupancy state, and equipment type. It should include normal periods with known maintenance events, fault-injection tests where safe and permissible, and retrospective analysis of historical alarms. Temporal validation must avoid leakage between training and test windows. If labels are scarce, weakly supervised or unsupervised metrics can be used, but they should not be presented as substitutes for fault-level validation. The evaluation should also report the latency distribution, not only the mean, because building alarms often have deadline constraints. Finally, the detector should be tested under realistic missingness and noise, including dropped packets, stale values, and sensor saturation. A system that works only with complete and synchronized data is not an edge-ready system.

7. Conclusion

This manuscript has developed a methodological framework for edge AI-based multimodal sensing and real-time anomaly detection in intelligent buildings. The central argument is that detection quality emerges from the joint design of sensing, synchronization, adaptive fusion, hybrid scoring, risk-calibrated decision logic, and resource-aware deployment. The proposed formulation treats anomalies as context-dependent and multimodal, models heterogeneous sensor observations explicitly, and links anomaly scores to operational alarm costs. It also shows why static fusion and fixed thresholds are fragile under missing modalities, nonstationarity, and asymmetric error costs. The equations provide a basis for specifying a detector, selecting thresholds, and analyzing latency, energy, and memory constraints without relying on invented empirical results. The practical significance is a design discipline: building anomaly detectors should be evaluated against the operating conditions, failure modes, and decision consequences they will actually face. Future deployment work can instantiate the framework with specific sensors, validated fault records, and commissioning data, but the present contribution is the technical structure that such work requires.

References

1. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), Article 15. https://doi.org/10.1145/1541880.1541882

2. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

3. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. IEEE Internet of Things Journal, 3(5), 637–646. https://doi.org/10.1109/JIOT.2016.2579198

4. Satyanarayanan, M. (2017). The emergence of edge computing. Computer, 50(1), 30–39. https://doi.org/10.1109/MC.2017.9

5. Baltrusaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(2), 423–443. https://doi.org/10.1109/TPAMI.2018.2798607

6. Atrey, P. K., Hossain, M. A., El Saddik, A., & Kankanhalli, M. S. (2010). Multimodal fusion for multimedia analysis: A survey. Multimedia Systems, 16(6), 345–379. https://doi.org/10.1007/s00530-010-0182-0

7. Kingma, D. P., & Welling, M. (2014). Auto-encoding variational Bayes. International Conference on Learning Representations. https://arxiv.org/abs/1312.6114

8. Ruff, L., Kauffmann, J. R., Vandermeulen, R. A., Montavon, G., Samek, W., Kloft, M., Dietterich, T. G., & Müller, K.-R. (2021). A unifying review of deep and shallow anomaly detection. Proceedings of the IEEE, 109(5), 756–795. https://doi.org/10.1109/JPROC.2021.3052449

9. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

10. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (Vol. 30, pp. 5998–6008).

11. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. ACM Computing Surveys, 46(4), Article 44. https://doi.org/10.1145/2523813

12. Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., & Zhang, J. (2019). Edge intelligence: Paving the last mile of artificial intelligence with edge computing. Proceedings of the IEEE, 107(8), 1738–1762. https://doi.org/10.1109/JPROC.2019.2918951

13. Chou, J.-S., & Telaga, A. S. (2014). Real-time detection of anomalous power consumption. Renewable and Sustainable Energy Reviews, 33, 400–411. https://doi.org/10.1016/j.rser.2014.01.088

14. Liu, J., Luo, D., Fu, X., Lu, Q., & Kang, K. Y. (2022). Design strategy of multimodal perception system for smart environment. In Internet of Things for Smart Environments (pp. 93-115). Cham: Springer International Publishing.

15. Bonomi, F., Milito, R., Zhu, J., & Addepalli, S. (2012). Fog computing and its role in the Internet of Things. In Proceedings of the First MCC Workshop on Mobile Cloud Computing (pp. 13–16). ACM. https://doi.org/10.1145/2342509.2342513

16. Hong, T., Wang, Z., Luo, X., & Zhang, W. (2020). State-of-the-art on research and applications of machine learning in the building life cycle. Energy and Buildings, 212, 109831. https://doi.org/10.1016/j.enbuild.2020.109831

17. Kleiminger, W., Beckel, C., & Staake, T. (2014). Occupancy detection from electricity consumption data. In Proceedings of the 5th ACM Workshop on Embedded Systems for Energy-Efficient Buildings (pp. 1–8). ACM.

18. Wang, S., & Ma, Z. (2008). Supervisory and optimal control of building HVAC systems: A review. HVAC&R Research, 14(1), 3–32. https://doi.org/10.1080/10789669.2008.10390991

19. Alsheikh, M. A., Lin, S., Niyato, D., & Tan, H.-P. (2014). Machine learning in wireless sensor networks: Algorithms, strategies, and applications. IEEE Communications Surveys & Tutorials, 16(4), 1996–2018. https://doi.org/10.1109/COMST.2014.2320099