SenseLess: Minimal Vision, Maximum Insight for Smart Homes

***Abstract*—We present SenseLess, a hybrid anomaly detection framework for smart homes that, during the training phase, automatically labels images without manual annotation by combining sensor-guided detection, self-supervised visual clustering, and unsupervised multi-sensor delay estimation for precise alignment. During operation, the system relies primarily on non-vision sensors and activates a confidence-aware vision model only under low-confidence, thereby preserving privacy while maintaining adaptability. Evaluated in real home monitoring, SenseLess achieved an average label coverage of 97.65% with 94.9% accuracy and reduced vision usage to less than 4% of wall-clock operating time. Calibration mechanisms and minimal configuration requirements support scalability and deployment across diverse residential environments.**

***Index Terms*—smart homes, anomaly detection, self-supervised learning, sensor fusion, image labeling, privacy preservation**

I. INTRODUCTION

Anomaly detection in home care is critical for supporting older adults who live alone, where unnoticed events such as forgotten appliances or unexpected exits pose risks. Monitoring systems can enable early detection and intervention, but they must operate continuously while preserving privacy and minimizing caregiver workload. Effective systems must adapt to behavioral changes, reduce false positives, and provide actionable insights without intruding on daily life. Achieving this balance is essential for safe, respectful, and scalable in-home care.

Non-vision sensor approaches are widely adopted in home care for their low cost, unobtrusiveness, and ability to run continuously. Motion and door sensors detect abnormal activity and nonresponse [1], while environmental sensors monitor air quality, including elevated CO_2 levels [2]. Smart plugs, tags, and smoke or gas detectors support safety by identifying open doors, unattended stoves, or unauthorized access [3]. However, these systems often suffer from high false positives, limited context, and data drift [4]–[6].

Vision-based systems, by contrast, offer high-resolution contextual information and have been used to monitor illness progression [7], detect falls using RGB and 3D video [8], and recognize behavioral anomalies such as agitation in residents with cognitive impairments [9]. However, these models require extensive annotated datasets for training and are sensitive to privacy constraints, lighting variability, and limited camera coverage. Relying solely on visual input typically requires continuous monitoring, which incurs high computational costs and further complicates deployment on resource-constrained devices. As a result, the use of vision systems in personal living spaces remains limited, particularly in scenarios where trust and privacy are essential.

In response, hybrid approaches have emerged, combining data from multiple sensing modalities to improve robustness and coverage [10]. Yet, these systems often inherit the drawbacks of their constituent sensors, and can introduce computational overhead and new challenges in data fusion and synchronization. Addressing these trade-offs requires a design that not only integrates complementary sensing streams, but also adapts to uncertainty, limits privacy exposure, and operates effectively with minimal supervision.

To address these limitations, we present a hybrid anomaly detection framework that fuses non-vision sensor data with self-supervised learning to build an adaptive, privacy-conscious vision model. During training, the system uses an unsupervised sensor-based model to identify anomalous patterns in environmental data, which are then aligned with corresponding camera frames to produce initial image labels. In parallel, a contrastive self-supervised algorithm learns visual representations from the same unlabelled images and generates pseudo-labels via clustering. These two streams are merged through a confidence-weighted refinement process, yielding a high-quality, self-labeled dataset for training a vision-based anomaly detector. At deployment, the non-vision model operates continuously as the primary detector, while the vision model is selectively activated when sensor predictions are uncertain or potentially drifting. This design enables accurate and scalable anomaly detection with minimal visual exposure, no manual annotations, and continuous adaptation to evolving home environments, forming the basis of SenseLess, our proposed hybrid framework.

While designing SenseLess, we addressed four key challenges that arise in deploying anomaly detection systems in real-world home environments:

- **Scarcity of annotated image data.** Vision-based anomaly detection typically requires large volumes of labeled images, which are difficult to obtain in home settings. Manual labeling is time-consuming, costly, and often impractical for rare or privacy-sensitive events, as it typically requires human supervision [11]. To overcome this, we use a dual-labeling strategy:
  - *Sensor-guided image labeling:* Non-vision sensors detect anomalies, which are aligned with camera frames to generate initial labels automatically.
  - *Self-supervised visual labeling:* A contrastive learning model learns representations from unlabelled

images and clusters them to generate pseudo-labels based on visual similarity.

- **Temporal misalignment between sensor and image data.** Sensors such as temperature or CO_2 respond with delay to environmental changes, hindering alignment of sensor events with visual observations [12]. Our analysis confirmed these delays (see Section II-B). We address them with a backward synchronization mechanism that estimates and compensates for response lag to improve labeling accuracy.
- **Balancing privacy with contextual awareness.** Always-on cameras are intrusive. Our system activates vision only when sensor predictions are uncertain, limiting exposure while retaining interpretability.
- **Drift in sensor performance.** Sensor readings can shift over time with environmental or behavioral changes, reducing accuracy [6]. We address this with a self-healing mechanism where the vision model retrains the sensor model when drift is detected.

This work is guided by three research questions:

- **RQ1:** Can non-vision sensors be used to automatically generate accurate labels for training vision-based anomaly detection models?
- **RQ2:** How can self-supervised learning be combined with non-vision data to refine these labels and eliminate the need for manual annotation?
- **RQ3:** How can multi-modal sensor streams with variable response times be calibrated and aligned for accurate, synchronized anomaly detection?

To address these questions, we contribute the following:

- We propose *SenseLess*, a hybrid anomaly detection system that uses sensor-based and self-supervised pseudo-labels to train vision models without manual annotation.
- We introduce the Hierarchical Event-Driven Synchronization (HEDS) algorithm, which performs unsupervised multi-sensor delay estimation and compensates for response lags during sensor-to-image alignment. This improves label precision across heterogeneous modalities, ensuring reliable training data without manual calibration.
- We present a self-healing feedback mechanism in which the vision model retrains and recalibrates the non-vision model when drift or low-confidence conditions are detected, enabling long-term robustness.

II. SYSTEM DESIGN

A. Problem Definition

We consider the problem of automatically labeling images captured in indoor environments to train a vision-based anomaly detection model, without relying on manual annotation. Let $I = \{i_1, i_2, \ldots, i_n\}$ be a set of unlabelled images, and $S = \{s_1, s_2, \ldots, s_m\}$ be the corresponding aligned non-vision sensor readings associated with each image. The goal is to generate a refined label set $L^* = \{\ell_1, \ell_2, \ldots, \ell_n\}$, where each label $\ell_k \in \{\texttt{Normal}, \texttt{Anomaly}, \texttt{Unknown}\}$, such that the resulting dataset $\{(i_k, \ell_k)\}$ can be used to train a robust, privacy-aware vision-based anomaly detector.

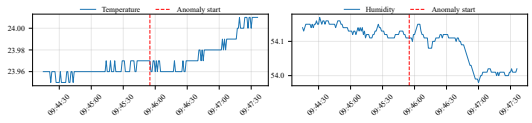

Fig. 1. Delayed temperature and humidity response after door opens. The red dashed line marks the event onset.

This task involves several key challenges:

- **Sensor-to-image alignment:** Sensor and image data must be temporally aligned despite variable response delays across sensor types.
- **Sensor label uncertainty:** Labels generated from unsupervised sensor models may be noisy or uncertain due to environmental drift or hardware variability.
- **Unsupervised vision labeling:** Visual features extracted via self-supervised learning must be clustered and interpreted without access to ground truth.
- **Deployment constraints:** The final system must operate under privacy constraints, manage data drift, and provide confidence-aware decisions with minimal supervision.

B. Preliminary Investigation

Before system design, we ran experiments with real-world home care sensor data. They revealed two key challenges that shaped *SenseLess*: the limited accuracy of non-vision anomaly detection without labels, and sensor response delays that complicate alignment with visual input.

a) Limitations of non-vision anomaly detection.: We first trained an unsupervised anomaly detection model using only non-vision environmental sensor data. While this approach worked in controlled settings, it performed poorly in realistic home environments. The absence of ground truth made it difficult to verify the timing or cause of anomalies, and many sensor readings appeared ambiguous without contextual information. These findings motivated the use of an additional image stream and a self-supervised learning model to provide complementary cues for labeling.

b) Sensor delay and misalignment.: We examined the temporal patterns of sensor responses around events and observed, through visual inspection, that some environmental changes occurred gradually rather than immediately. For example, temperature and humidity values often shifted several seconds after a door was opened, as shown in Figure 1. This misalignment increases the risk of associating anomalies with incorrect visual frames. Based on these observations, we designed the HEDS algorithm to systematically detect and compensate for sensor delays during alignment with image timestamps.

C. System Overview

SenseLess, illustrated in Figure 2, operates in two sequential phases: a training phase that generates a self-labeled image dataset, and a deployment phase that performs anomaly detection with minimal visual exposure.

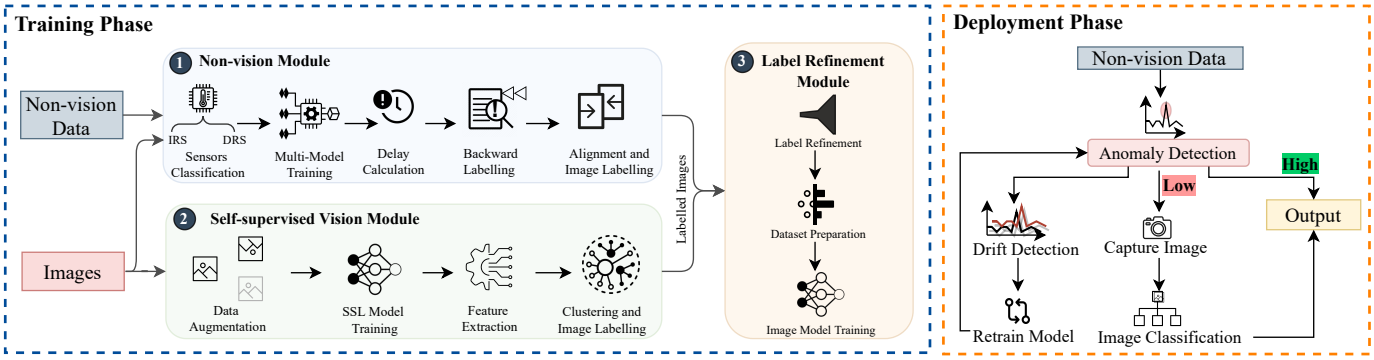

Fig. 2. System architecture of *SenseLess*. The training phase integrates non-vision anomaly detection, self-supervised vision, and label refinement to generate labeled data. The deployment phase relies primarily on non-vision sensing and selectively activates the vision module when confidence is low.

During training, the system integrates three modules. First, the non-vision module is trained on normal environmental data to detect anomalies from sensor readings. Detected events are then temporally aligned with image frames using the Hierarchical Event-Driven Synchronization (HEDS) algorithm, which performs unsupervised multi-sensor delay estimation and compensates for response lags. This ensures that anomalies are matched to their true visual context, improving label accuracy across heterogeneous modalities. Second, the self-supervised vision module applies contrastive learning to extract visual features from unlabelled images and clusters them into pseudo-labels. Third, the label refinement module fuses the sensor- and vision-derived labels using a confidence-weighted strategy, producing a high-quality dataset for training a vision-based anomaly classifier.

In deployment, the non-vision module runs continuously as the primary anomaly detector. When predictions are uncertain, the vision module is selectively activated to validate the decision. This selective activation preserves privacy by minimizing visual exposure and supports adaptation through feedback-driven retraining. Together, these modules enable *SenseLess* to deliver accurate, scalable, and privacy-aware anomaly detection in dynamic home environments.

D. Training Phase

1) Non-Vision Module: The non-vision module generates initial labels for image data by detecting anomalies in environmental sensor streams. This process is implemented using the HEDS algorithm, which synchronizes sensor data with image timestamps while accounting for response delays and alignment uncertainty. The algorithm is outlined in Algorithm 1 and described in detail below.

a) Sensor Classification: Sensors are grouped into instant-response (IRS) and delayed-response (DRS) categories based on their characteristics and observed response behavior. IRS sensors (e.g., motion detectors) react immediately and are assigned zero delay ($\Delta t_s = 0$). DRS sensors (e.g., temperature or humidity) exhibit measurable latency and are assigned a calibrated delay $\Delta t_s > 0$ estimated during delay calculation.

b) Model Training: We train a multi-model ensemble of an autoencoder, Isolation Forest, and Elliptic Envelope, each capturing different aspects of anomaly behavior. Sensor data is cleaned in two stages: values outside physical ranges are replaced by interpolation, and statistical outliers are removed using a configurable method, with median absolute deviation as the default and Z-score filtering as an alternative. The autoencoder is trained on clean, normal-only data using mean squared error loss, with architecture scaled to sensor dimensionality. Features are normalized with either RobustScaler or StandardScaler, and thresholds are derived from the training error distribution and calibrated on validation data. During evaluation, anomaly scores from the three detectors are combined, and sensitivity is tuned by requiring agreement from one, two, or all detectors. This mechanism is optimized to maximize recall while limiting false positives. Confidence scores $c_k^{(s)}$ are computed from reconstruction errors and sensor-specific error distributions, aggregated across sensors, and calibrated with isotonic regression to produce reliable probabilities. When one detector achieves higher validation performance than the others, the system can use it alone. If no method is clearly superior, the full ensemble is retained.

c) Delay Estimation and Calibration: For DRS, the system estimates the delay Δt_s between the anomaly onset and the point when reconstruction error first exceeds the threshold. Anomalies are detected with the trained model, which produces per-sensor error traces and thresholds. For each event, the anomaly start time t^{start} is identified by tracing backward from the error rise to where values stabilize near baseline, and the threshold crossing time t^{cross} is the first index where the error exceeds the threshold. The onsite delay is then computed as $\Delta t_s = t^{\text{cross}} - t^{\text{start}}$. The final Δt_s for each sensor is the mean of the filtered values, while IRS sensors are assigned $\Delta t_s = 0$. If ground truth event times are available, an optional lab-based calibration can be performed. In this case, delays are measured as the average difference between known event times and the first detected anomalies. Lab and onsite values may then be blended using configurable weights or adjusted relative to a reference sensor. Human-in-the-loop review is also supported, allowing experts to inspect

Algorithm 1 Hierarchical Event-Driven Synchronization (HEDS)

Input: Sensor dataset S, image dataset I
Output: Initial labeled image dataset L_s
Step 1: Sensor classification
for all sensor s in sensor modalities **do**
  if s is instant-response (IRS) **then**
    Assign $\Delta t_s = 0$
  else
    Assign s to delayed-response sensor (DRS) group
  end if
end for
Step 2: Model training
Train anomaly detection model f on a normal-only subset of S
Use f to compute sensor-based confidence scores $c_k^{(s)}$ for detected anomalies
Step 3: Anomaly detection and delay estimation
for all sensor s in available sensors **do**
  Detect anomaly events with start time t^{start}
  if s is DRS **then**
    Find threshold crossing t^{cross} (first error > threshold)
    Compute delay $\Delta t_s = t^{\text{cross}} - t^{\text{start}}$
    Set Δt_s to mean of valid delays
  end if
end for
Step 4: Backward labeling and multi-sensor aggregation
for all sensor s in sensor modalities **do**
  for all anomaly periods P_s **do**
    Shift start time backward by Δt_s to obtain P'_s
    for all $t_k \in P'_s$ **do**
      Assign sensor-level label $\ell_k^{(s)} = \texttt{Anomaly}$
    end for
  end for
end for
for all timestamps t_k **do**
  if $\exists s : \ell_k^{(s)} = \texttt{Anomaly}$ **then**
    Set global label $\ell_k = \texttt{Anomaly}$
  else
    Set $\ell_k = \texttt{Normal}$
  end if
end for
Step 5: Sensor-to-image alignment
for all image timestamps $t_i \in I$ **do**
  Define primary window: $[t_i - 2,\ t_i + 2]$ seconds
  Search for anomaly event within primary window
  if match found **then**
    Assign $\ell_k^{(s)}$ and $c_k^{(s)}$; mark as `Primary`
  else
    Define fallback window: $[t_i - d,\ t_i + d]$, where $d = \max(\Delta t_s, 5\,\text{s})$
    Search for anomaly event within fallback window
    if match found **then**
      Assign $\ell_k^{(s)}$ and $c_k^{(s)}$; mark as `Fallback`
    else
      Assign $\ell_k^{(s)} = \texttt{Unknown}$ with $c_k^{(s)} = 0$
    end if
  end if
end for
return L_s

and correct selected cases. This multi-stage calibration process ensures accurate temporal alignment between sensor events and images, improving label precision across all sensor types.

d) Backward Labeling: For each sensor s, the calibrated delay Δt_s is applied by shifting the start time of detected anomaly periods P_s backward to obtain P'_s. All records $t_k \in P'_s$ are then labeled anomalous ($\ell_k^{(s)} = \texttt{Anomaly}$). This process is repeated for all sensors, and the overall label ℓ_k is set to `Anomaly` if any sensor produces an anomalous record ($\ell_k^{(s)} = \texttt{Anomaly}$); otherwise, ℓ_k is set to `Normal`.

e) Image Alignment: To align images with sensor data, the system applies a dual-window strategy. For each image at timestamp t_i, it first searches within a primary window $[t_i - 2,\ t_i + 2]$ seconds, a tolerance chosen to accommodate jitter and timestamp offsets and to provide a margin of roughly two sensor sampling intervals at 1 Hz. If no match is found, a fallback window is applied with size $d = \max(\Delta t_s, 5\,\text{s})$ to capture delayed sensor responses, where 5 s was chosen as the maximum delay observed across all tested sensors. The closest sensor record within the selected window is used, and the image is marked as `Primary` or `Fallback`. If no record is found, the image is labeled `Unknown` with zero confidence.

Matched images inherit the sensor-derived label $\ell_k^{(s)}$ and its confidence $c_k^{(s)}$. Unknown images are resolved during label refinement using SSL pseudo-labels or optionally escalated for human annotation. This process yields the labeled dataset $L_s = \{(i_k, \ell_k^{(s)}, c_k^{(s)})\}$, enabling supervised training without manual annotation.

2) Self-Supervised Vision Module: The self-supervised vision module learns visual representations from unlabelled images and produces cluster-based pseudo-labels independent of the sensor subsystem.

We adopt a SimCLR framework with a ResNet-18 encoder and projection head. Each image is stochastically augmented twice (random crops, color jitter, grayscale, flipping) and the paired views are trained with the normalized temperature-scaled cross-entropy (NT-Xent) loss. This pulls representations of the same image together while separating others.

After training, the SimCLR encoder generates feature embeddings for the dataset. These embeddings are clustered using a Bayesian Gaussian Mixture Model (GMM), which employs a Dirichlet prior to adapt the number of active components while aligning with the binary objective of distinguishing Normal from Anomaly. For each image, we compute confidence using a hybrid score that combines posterior probability with normalized distance to the assigned cluster, providing well-calibrated values in [0,1].

The result is a pseudo-labeled dataset (L_{ssl}) where each image has a cluster index and confidence score. These labels are later fused with sensor-derived labels during refinement.

3) Label Refinement Module: Each image i_k receives two candidate labels: a sensor label $\ell_k^{(s)}$ with confidence $c_k^{(s)}$ and an SSL label $\ell_k^{(\text{ssl})}$ with confidence $c_k^{(\text{ssl})}$. Confidence values are calibrated using isotonic regression models trained for each source and reused in later runs. Labels are then refined by the following rules:

1) If both confidences fall below thresholds (τ_s, τ_{ssl}), assign `Unknown`.
2) If both labels agree and at least one confidence exceeds its threshold, accept the shared label.
3) If only one source exceeds its threshold and is more confident, select its label.
4) Otherwise, assign `Unknown`.

Thresholds τ_s and τ_{ssl} are obtained through isotonic calibration against ground truth correctness flags, with the final cutoff selected by grid search to balance accuracy and coverage. Optional human review is supported for `Unknown` cases. Reviewed labels are stored to avoid duplication in future runs. The result is the final dataset $L^* = \{(i_k, \ell_k)\}$ with $\ell_k \in \{\texttt{Normal}, \texttt{Anomaly}, \texttt{Unknown}\}$.

4) Vision Model Training: We use the final labeled dataset L^* to train a supervised convolutional neural network for the

detection of visual anomalies, excluding `Unknown` samples. We adopt MobileNetV2 for its favorable balance between classification performance and computational efficiency. The model is initialized with ImageNet weights, with the final layer adapted for binary classification. Training is performed using cross-entropy loss and the Adam optimizer, with early stopping based on validation loss. To improve generalization, we apply standard data augmentations such as random cropping, horizontal flipping, and color jitter.

E. Deployment Phase

During deployment, *SenseLess* runs a two-stage decision pipeline. The non-vision module operates continuously as the primary anomaly detector, and the vision module is selectively activated when additional validation is required. This design minimizes visual exposure while maintaining robustness under uncertainty and sensor drift.

1) Primary Non-Vision Module: The non-vision module operates continuously, analyzing sensor streams and identifying outliers using an ensemble model trained on normal data. Sensor values are cleaned using a Z-score filter, and anomaly predictions are produced only for non-corrupted segments. Each prediction is assigned a label (`Normal`, `Anomaly`, or `Sensor_Error`) and accompanied by a confidence score estimated from reconstruction error. When the confidence is high, the prediction is accepted without further processing. Sensor errors and the names of affected sensors are logged per instance to support inspection.

2) Fallback Vision Module: When the non-vision module produces low confidence predictions or sensor errors, the system activates the vision module to perform secondary validation. In this case, a new image is captured and processed by a pre-trained classification model. The image is preprocessed and evaluated by the model, which returns a predicted class label (`Normal` or `Anomaly`) along with a confidence score derived from softmax outputs or object count estimation. Vision-based predictions that exceed a configured confidence threshold are logged and can override the original decision. This selective activation strategy improves robustness in ambiguous cases while minimizing unnecessary visual processing. The captured image is processed solely for fallback inference and is not retained, ensuring privacy is preserved.

3) Data Drift Handling: To support long-term reliability, the system maintains a reference baseline that characterizes normal sensor behavior under stable conditions. This baseline is computed during an initial data collection period or periodically updated using recent deployment logs filtered to exclude sensor errors. For each sensor, the system calculates mean, standard deviation, and percentiles, storing these as reference statistics. Once deployed, incoming sensor readings are compared against this baseline to detect data drift. Drift is flagged when recent samples exhibit significant deviation in statistical properties, such as shifts in mean or variance, or an increased outlier rate relative to the baseline. In parallel, the system monitors reconstruction error from the autoencoder to detect structural changes in sensor patterns that may not affect first-order statistics. When drift is detected consistently across multiple runs, the system activates a self-healing feedback loop. If the vision module is triggered due to low confidence and returns a high-confidence prediction that contradicts the non-vision decision, this prediction is logged and used as a trusted label. These vision-verified events are then used to retrain the non-vision model, ensuring that the system remains adaptive to changing environments without relying on manual annotations.

4) Model Retraining: When persistent data drift is detected, the system triggers a retraining workflow to update the non-vision anomaly detector. A data collector module first gathers a retraining set composed of recent sensor data, fallback vision-labeled samples, replay buffer entries, and a small portion of the original training data. These sources are combined and filtered to ensure statistical quality and coverage diversity. A new non-vision model is then trained using the same ensemble configuration as in the original deployment, and its performance is evaluated against a validation set. If the new model outperforms the current one by a predefined margin (e.g., ≥ 0.02 macro-F1 improvement), it is deployed to replace the old model after creating a backup. To prevent overfitting to recent events, the replay buffer maintains representative samples across time and label types. This modular pipeline ensures that retraining occurs only when sufficient, high-quality data is available and performance gains are measurable.

III. IMPLEMENTATION AND EVALUATION

A. Evaluation Setup and Dataset Overview

We evaluated *SenseLess* in a real home with sensors and cameras installed. The system ran continuously for four weeks, capturing non-vision readings and image data. Deployment targeted key safety cases: prolonged door openings, unattended kitchen appliances, abnormal occupancy, and obstructed pathways. The setup included temperature, humidity, pressure, CO_2, and ultrasonic distance sensors, and both RGB and thermal cameras. Non-vision sensor data were collected every second during 12.5 hours of daily operation. To broaden applicability, we repurposed common household sensors for secondary detection of these cases, even though each case already has dedicated primary sensors. To support model training and evaluation, the collected dataset was split into 24 days for training and 6 days for testing. The image dataset comprised 2,516 door images, 2,575 appliance images, 1,511 occupancy images, and 1,960 abnormal object images, covering diverse conditions across the four scenarios. In real deployments, the system should begin with a 1–2 week period of collecting normal data, which can then be used to train the initial sensor model and establish baseline patterns.

B. Non-vision Sensor Model Performance

Table I reports the results on temporally held-out test sets. Non-vision models achieved high accuracy and macro-F1 on appliance and object use cases. Occupancy and door detection were less reliable. Across door, appliance, and occupancy

TABLE I
NON-VISION MODEL PERFORMANCE BY USE CASE

Use Case	Method	Accuracy	Macro-F1
Door	Single Autoencoder	86.0%	0.73
Appliance	Single Autoencoder	100%	0.98
Occupancy (Incremental)	Single Autoencoder	98.0%	0.95
Abnormal Object	Rule-based	100%	1.00

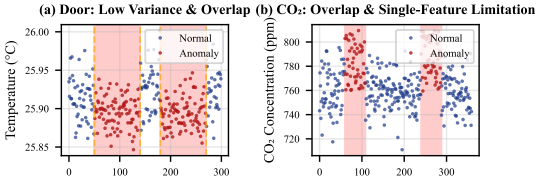

Fig. 3. Signal factors affecting detection. (a) Door anomalies show subtle, overlapping changes (b) CO_2 anomalies overlap strongly with normal values, highlighting the limitation of using a single feature. Red: anomaly periods.

cases, a single autoencoder consistently outperformed the ensemble variant, indicating that additional methods introduced more noise than benefit in these settings.

In the door case (Figure 3a), temperature, humidity, and pressure exhibited low variance and strong distributional overlap between normal and anomalous states. For example, anomaly means were close to normal averages. Correlations across features were weak, and temporal patterns followed smooth trends rather than abrupt shifts. These subtle changes reduced the contrast available to the autoencoder, making boundary precision and anomaly detection more difficult.

Occupancy detection using only CO_2 (Figure 3b) was hindered by strong overlap between normal and anomalous ranges. Anomalies averaged 791 ppm (SD = 67.7) and normal periods 725 ppm (SD = 42.4), making elevated levels common in both states. Additonally, this single-feature setup limited the autoencoder's ability to distinguish anomalies. Temporal patterns showed smooth drifts rather than sharp jumps, further reducing discriminability. Conventional full-dataset training produced unstable boundaries, but incremental training in 5-day cycles with replay memory improved robustness, raising macro-F1 from 0.72 to 0.95.

C. Delay Calculation

Delay baselines were established using labeled datasets and a supervised Random Forest classifier (RF), which served as a proxy ground truth for anomaly onset. For each use case, the classifier was trained to detect the first anomaly, and the average gap between event onset and first detection defined the lab delay. These values provide a reference baseline but are not directly usable in deployment since true physical onset cannot be measured at scale. When evaluated against the RF baseline, our delay estimation method achieved RMSE = 1.83s and MAE = 1.60s (Figure 4). Door events showed strong agreement, with deviations below 1.2s for temperature and humidity, and 2.6s for pressure. Appliance sensors tracked short delays with errors between 1.7s and 2.5s, while the occupancy case differed by only 0.62s. These results confirm that onsite estimation provides physically plausible and deployment-ready delays, even in scenarios where environmental responses are gradual.

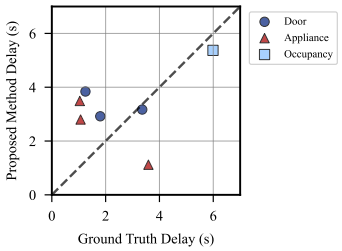

Fig. 4. Comparison between proxy ground truth (RF) delays and the proposed delay estimation method. Each point represents a sensor measurement: door sensors (circles), appliance sensors (triangles), and occupancy sensor (square). The dashed line indicates close agreement ($y = x$). Our method achieved RMSE = 1.83s and MAE = 1.60s.

Door sensors showed close agreement between RF baselines (3.36s, 1.80s, 1.25s for temperature, humidity, and pressure) and onsite estimates (3.17s, 2.92s, 3.84s). For appliances, temperature and humidity delays were near-instantaneous (RF: 1.07s, 1.04s; onsite: 2.80s, 3.49s), while CO_2 exhibited slightly longer but consistent lags (RF: 3.59s, onsite: 1.12s). The occupancy case showed the slowest dynamics, with CO_2 delays of 5.99s (RF) and 5.37s (onsite), reflecting gradual diffusion compared to the sharper transitions of appliance events. If available, final delays can be blended with RF baselines, reference-sensor adjustments, or human-in-the-loop review to ensure robust alignment across deployment settings.

Across all use cases, onsite estimates stayed within seconds of RF-derived ground truth, while alternative unsupervised baselines such as CUSUM or cross-correlation often misestimated delays by minutes or produced negative lags. The stability of our method ensures physically plausible estimates, which is essential for sensor–image alignment in multimodal anomaly detection systems.

D. Sensor-Image Alignment

TABLE II
SENSOR–IMAGE ALIGNMENT RESULTS ACROSS USE CASES

Use Case	Match Rate (%)	Accuracy (%)
Door	100	67.2
Appliance	100	97.1
Occupancy	100	78.6
Abnormal Object	100	100

Alignment accuracy was evaluated by comparing the image labels transferred from non-vision records with the ground truth image labels (Table II). Since the alignment procedure relies on previously labeled sensor data, its performance reflects the effectiveness of the non-vision anomaly detector. Appliances achieved 97.1% accuracy, consistent with strong upstream sensor performance. Occupancy reached 78.6%, where the autoencoder struggled to capture CO_2 deviations, leading to reduced image-level agreement. Door alignment was lowest at 67.2%, mirroring the difficulty of detecting gradual environmental transitions in sensor data. Abnormal object events were straightforward to align, achieving 100% accuracy due to the clear signal from instant-response distance sensors. Overall, the observed performance variations reflect the inherent challenges of the upstream anomaly detection task,

and therefore the alignment procedure performs best for events producing immediate and unambiguous sensor changes.

E. Self-Supervised Vision Module Performance

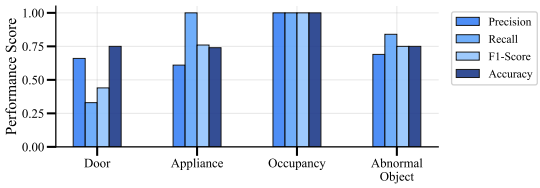

Fig. 5. Performance of the SimCLR and GMM pipeline across the four evaluation use cases. Metrics include precision, recall, and F_1-score for anomaly and normal classes, and overall accuracy.

We evaluated multiple SSL models, including Barlow Twins, DINO, FastSiam, MoCo, and SMoG. SimCLR achieved the best balance of accuracy, training time, and model size, and was selected for our framework. For clustering, we compared K-Means, GMM, DBSCAN, and Agglomerative methods. The GMM was adopted as the default, as it adapts the number of active components through a Dirichlet prior, provides well-calibrated confidence scores, and offered the most consistent trade-off between accuracy and efficiency across use cases.

Figure 5 shows the performance of the selected SimCLR + GMM pipeline. Occupancy detection achieved perfect accuracy (1.00), with both classes classified correctly. Appliance detection reached 0.74 accuracy; anomaly recall was perfect (1.00) but precision was lower (0.61), indicating that some normal frames were misclassified as anomalies, likely due to thermal interference from nearby appliances. Abnormal object detection reached 0.75 accuracy with balanced F_1-scores for both classes (anomaly 0.75, normal 0.71). Door detection was the most challenging, reaching 0.75 accuracy. Anomaly precision was high (0.66) but recall was lower (0.33).

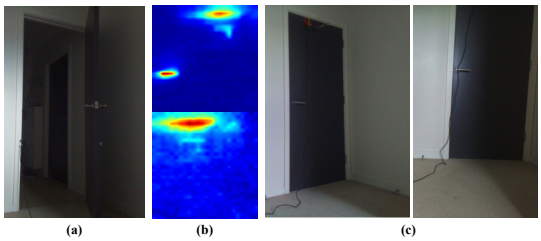

Fig. 6. Failure cases in SSL clustering: (a) overlapping door states, (b) thermal overlap from nearby appliances, (c) shadows misclassified as objects.

The observed errors were primarily due to visual ambiguity. In the door case, shadows and overlapping door states reduced class separability (Figure 6a). Appliance detection was affected by thermal interference from nearby devices (Figure 6b), while abnormal object detection was hindered by shadows misinterpreted as obstructions (Figure 6c). These failures illustrate how limited visual separability and intra-class variability constrain SSL clustering quality.

F. Label Refinement

Table III compares the non-vision, SSL, and hybrid methods. The non-vision approach achieved high coverage but variable accuracy, ranging from 67.2% in the door case to 100% for abnormal objects. SSL clustering offered more balanced performance, with stronger results for occupancy (100%) but lower accuracy in appliances and abnormal objects (74.0% and 75.2% respectively).

The hybrid method improved reliability by combining both sources. Overall, in automated mode, it achieved 97.65% coverage with 94.94% accuracy, yielding an effective accuracy of 92.50%. When low-confidence cases (`Unknown`) were reviewed, coverage reached 100% and accuracy rose to 95.84%, with only 9.4% of door images requiring manual annotation. The appliance, occupancy, and abnormal object cases required no human input, as automated refinement already provided near-perfect performance.

Overall, the hybrid strategy reduced dependence on a single modality. Door events benefited from fusion, where SSL compensated for weak sensor signals, while appliance and abnormal object cases leveraged sensor reliability to correct SSL misclassifications. Occupancy detection benefited strongly from visual clustering, with SSL and the hybrid approach achieving perfect accuracy. These results demonstrate that multimodal refinement not only recovers accuracy lost in individual methods but also achieves robustness across diverse event types.

G. Image Classification Model Performance

We evaluated the MobilenetV2 classifier on the door, appliance, and abnormal object use cases, while the occupancy case employed a pre-trained EfficientNet-based crowd counting model without further training. Since CO_2 levels indicate occupancy, people counting validates the sensor-based occupancy anomaly detection. In each case, only the classification head was trained, with the backbone frozen, using refined image labels containing a small proportion of mislabelled samples from the automated sensor–SSL labeling process. As shown in Fig. 7, appliance achieved the highest accuracy (95.35%) and balanced macro-recall (0.95) across both classes. Door and abnormal object reached accuracies of 88.10% and 89.80%, respectively, with anomaly recall exceeding normal recall, suggesting a bias toward detecting anomalous cases even under imperfect label conditions. This bias may be advantageous in safety-critical scenarios, where missing an anomaly carries greater cost than a false positive.

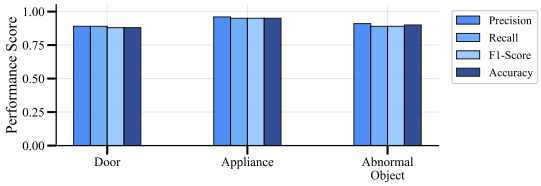

Fig. 7. Performance of the MobilenetV2 classifier across use cases.

H. End-to-End Labeling Duration

Table IV reports the execution time of each stage in the hybrid labeling pipeline. Across the four use cases, anomaly detection was the most time-consuming step, ranging from

TABLE III
COMPREHENSIVE LABELING PERFORMANCE COMPARISON

Use Case	Non-Vision			SSL			Hybrid (Ours)					
							Automated			+ Human		
	Coverage (%)	Accuracy (%)	Effective Acc. (%)	Coverage (%)	Accuracy (%)	Effective Acc. (%)	Coverage (%)	Accuracy (%)	Effective Acc. (%)	Coverage (%)	Accuracy (%)	Effective Acc. (%)
Door	100	67.2	67.2	100	74.6	74.6	90.6	82.81	75.02	100	84.4	84.4
Appliance	100	97.1	97.1	100	74.0	74.0	100	96.97	96.97	100	96.97	96.97
Occupancy	100	78.6	78.6	100	100	**100**	100	**100**	**100**	100	**100**	**100**
Abnormal Object	100	**100**	**100**	100	75.2	75.2	100	**100**	**100**	100	**100**	**100**
Average	100.00	85.73	85.73	100.00	80.95	80.95	97.65	94.94	92.50	100.00	95.84	95.84

Notes: Coverage = Percentage of images labeled. Accuracy = Percentage correct among labeled images. Effective Acc. = Coverage × Accuracy. Automated = Fusion without human input. + Human = Including optional human labeling for ambiguous cases. **Bold** = Best performing method.

TABLE IV
LABELING PIPELINE TIMING PERFORMANCE (SECONDS)

Use Case	Non-Vision	Alignment	SSL	Refinement
Door	142.95	35.09	7.36	73.99*
Appliance	136.82	54.90	7.39	0.26
Occupancy	65.09	17.51	10.06	0.24
Abnormal Object	27.72	35.54	11.02	0.31
Average	**93.15**	**35.76**	**8.96**	**18.70**

*Includes human annotation for 9.4% (Door). Total pipeline time: 71.38–259.39 seconds.

27.72 seconds (abnormal object) to 142.95 seconds (door). Image-sensor alignment required between 17.51 and 54.90 seconds, with larger datasets incurring longer alignment times. SSL labeling was comparatively fast, completing in 7.36–11.02 seconds across all scenarios. Label refinement varied depending on dataset complexity: while door required 73.99 seconds (including human labeling), appliance, occupancy, and abnormal object cases required less than one second each. Overall, full dataset annotation was achieved in 71.38–259.39 seconds per use case, representing a significant efficiency improvement over manual labeling, which would require hours to days. These times exclude model training (non-vision anomaly detector and SimCLR).

I. Ablation on Delay Compensation

To quantify the contribution of the HEDS algorithm, we conducted an ablation study comparing alignment performance with and without delay compensation. Table V reports the accuracy of sensor-to-image label alignment across the four evaluation use cases.

HEDS improved label alignment in cases where sensor delays were present. The door scenario showed the largest gain, with accuracy increasing by 4.7%. Appliances benefited only marginally (+0.1%), reflecting their strong and immediate thermal signatures. Occupancy showed a small improvement (+0.4%). No improvement was observed for abnormal object detection, since instant-response distance sensors incurred negligible delays. These results demonstrate that backward delay compensation is most effective in scenarios affected by gradual environmental responses.

TABLE V
ALIGNMENT ACCURACY WITH AND WITHOUT HEDS DELAY COMPENSATION.

Use Case	HEDS Alignment (%)	No-Delay Alignment (%)
Door	67.2	62.5
Appliance	97.1	97.0
Occupancy	78.6	78.2
Abnormal Object	100.0	100.0

TABLE VI
VISION USAGE DURING DEPLOYMENT (6 DAYS, 12.5 H/DAY).

Use Case	Fallback Rate (%)	Inf. (ms)	Sec/day	Usage (%)
Door	27.2	39.5	483.7	1.08
Appliance	29.9	35.4	476.7	1.06
Occupancy	29.2	35.4	458.2	1.03
Abnormal Object	8.0	31.4	112.8	0.25
Overall	94.3	141.7	1531.4	3.42

J. Deployment Evaluation

We evaluated *SenseLess* over the final six days of deployment, covering 270k records per use case. Figure 8 and Table VI summarise accuracy, fallback behaviour, and vision usage. Appliance monitoring maintained strong performance (88.5%) with modest fallback use, while occupancy detection started at 21.1% accuracy but improved to 76.7% after retraining, though still requiring frequent fallback (29.2%). Door detection initially achieved 55.2% accuracy but recovered to 69.3% following automatic retraining, and abnormal object detection achieved perfect accuracy with minimal fallback.

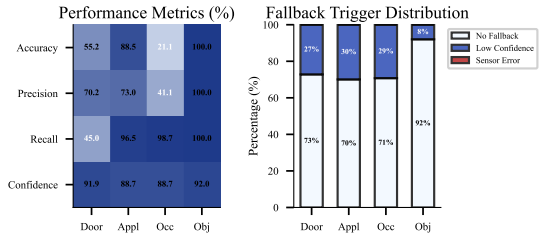

Fig. 8. Deployment performance across four use cases: (left) detection metrics and confidence scores; (right) vision fallback triggers showing the share of decisions needing secondary validation due to low sensor confidence or errors.

Across all scenarios, fallback activation scaled with sensor uncertainty and drift, providing secondary validation only

when required. Door and occupancy triggered the highest fallback rates, but each inference added just tens of milliseconds. Normalized over 270,000s of deployment (6 days, 12.5 h/day), the cumulative vision overhead was 9188s, amounting to 3.42% of operating time.

K. Cross-Home Validation

We further evaluated the framework in a new home using the door and occupancy use cases. For the door case, training on two days of data and testing on a third yielded 96.0% accuracy with an F1-score of 0.84. During deployment, the initial model achieved 24.4% accuracy, but automatic retraining adapted within the same environment and improved performance to 80.8% without manual intervention. For the occupancy case, training on 2.5 days and testing on 1.5 days achieved 93.0% accuracy with an F1-score of 0.91, and deployment maintained 89.6% accuracy with anomaly recall of 99.2%.

IV. DISCUSSION

SenseLess demonstrates that a hybrid sensor–vision pipeline can support scalable, privacy-preserving, and adaptive anomaly detection in home environments without manual image annotation. The HEDS algorithm compensates for sensor delays and improves label accuracy in scenarios with gradual responses. The hybrid labeling strategy achieves high coverage (97.65%) and accuracy (94.94%) while reducing annotation effort. Selective fallback limits visual processing to less than 4% of operating time, preserving privacy.

The self-healing loop enables adaptation under data drift by retraining the non-vision model using vision-derived corrections. Door and occupancy performance improves after retraining, with accuracy rising from 55.2% to 69.3% and from 21.1% to 76.7%, respectively. Cross-home evaluation highlights both generalizability and its limits. Physics-driven signals such as CO_2 transfer well across environments, while layout-dependent signals require adaptation. These results show that generalization is scenario-dependent but achievable through adaptation.

Performance variations reflect differences in non-vision sensor signal characteristics and availability. Strong and immediate signatures from appliances and distance sensors support accurate detection, whereas gradual transitions in door and occupancy cases limit model sensitivity. In practice, sensor types, placement, sampling rates, and reliability vary across homes, and while the framework can operate with different sensor subsets, detection accuracy depends on the quality and stability of the available signals.

SenseLess focuses on anomalies manifested through environmental changes. Human-centered abnormalities, such as falls or behavioral shifts, rely on sensing dynamics that differ from environment- and object-centric events and often do not produce clear environmental signatures. Detecting such events would therefore require different sensing modalities or explicit behavioral modeling beyond the design assumptions of this framework.

Environmental variability also impacts performance. Differences in layout, ventilation, and occupant routines influence baseline sensor behavior. Abrupt changes, such as renovations or sensor relocation, may temporarily degrade accuracy until adaptation occurs.

The delay estimation algorithm assumes a clear anomaly onset followed by a delayed sensor response at one-second sampling resolution. Weak, intermittent, or overlapping anomalies may violate this assumption. Higher sampling intervals, such as minute-level data, reduce temporal resolution and would require architectural changes to the algorithm.

The vision component of SenseLess is designed to operate with selective activation and multimodal validation limiting exposure to visual uncertainty. Image augmentations applied during training, including zooming through random resized cropping, improve robustness to lighting variation and partial visibility. Despite these design choices, inherent limitations remain. Camera coverage is spatially constrained, and anomalies may still be fully occluded by furniture or occupants. In addition, the system infers anomalies from single images, which limits its ability to distinguish events whose interpretation depends on temporal evolution rather than a single visual snapshot. Addressing these limitations fully would require multi-camera configurations.

Privacy perception is a major concern in camera-based systems. Prior studies show low acceptance of always-on cameras, with higher acceptance when video use is restricted to specific situations or time windows and combined with privacy-preserving measures [13]. SenseLess follows this model by activating vision only when non-vision predictions are uncertain. While user preferences are not directly evaluated, this design aligns with reported privacy expectations and balances contextual awareness with privacy preservation.

V. RELATED WORK

A. Sensor-Based Anomaly Detection

Unsupervised and semi-supervised methods are widely used to detect anomalies in smart homes with ambient sensors. Common approaches include autoencoders [14], Isolation Forest [15], and One-Class SVM [16], with ensembles improving robustness and reducing false positives [17], [18]. Federated learning has also been explored to support distributed training without raw data sharing [19], though most work is limited to homogeneous deployments and scalar signals, without camera fusion or delay handling [18].

Vision-based systems provide detailed spatial context and are used for fall detection [4], [20] and other indoor anomalies [21], but require large annotated datasets and face privacy and computational challenges [22].

A recent survey highlights persistent gaps in multimodal fusion and adaptability [11]. Existing systems often treat sensor and vision data independently, lack continuous adaptation to drift, and require manual tuning for seasonal or layout changes. To our knowledge, no prior work integrates ambient sensing and static vision in a temporally aligned, privacy-preserving framework with adaptive, bidirectional training.

B. Image Labeling in Vision-Based Systems

Labeling large-scale visual data remains a major bottleneck, particularly for anomaly detection. Manual annotation yields high accuracy but is slow and costly, with datasets such as ImageNet requiring years of effort [23]. Semi-automated methods combine model suggestions with human input [24], [25], while automated approaches use clustering [26], pre-trained models [27], or domain-specific pipelines [28], often integrated into tools like V7 and CVAT.

Most approaches assume abundant, well-structured categories, whereas anomalies are rare, context-dependent, and hard to define. Cross-modal labeling with wearable sensors has been explored in activity recognition [29]–[31], but remains limited for anomaly detection in images [32], [33]. We extend this line by leveraging ambient non-vision signals (e.g., motion, temperature, CO_2) to label indoor images, enabling scalable, privacy-aware training without human supervision.

VI. CONCLUSIONS

We presented SenseLess, a hybrid anomaly detection framework that generates image labels without manual annotation by combining sensor-guided and self-supervised learning with delay-aware alignment. The system integrates confidence-aware vision fallback, drift detection, and selective vision activation to preserve privacy while maintaining adaptability across scenarios. Evaluation across four home monitoring tasks demonstrated high labeling coverage and competitive accuracy, with scalability to new environments supported by modular design and minimal configuration requirements.

REFERENCES

[1] J. H. Shin, B. Lee, and K. S. Park, "Detection of abnormal living patterns for elderly living alone using support vector data description," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, pp. 438–448, 5 2011.

[2] B. Maag, Z. Zhou, and L. Thiele, "W-air enabling personal air pollution monitoring onwearables," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, pp. 1–25, 3 2018.

[3] S. Ramapatruni, S. N. Narayanan, S. Mittal, A. Joshi, and K. Joshi, "Anomaly detection models for smart home security," in *2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*. IEEE, 5 2019, pp. 19–24.

[4] T. Han, W. Kang, and G. Choi, "Ir-uwb sensor based fall detection method using cnn algorithm," *Sensors (Switzerland)*, vol. 20, pp. 1–23, 10 2020.

[5] Z. A. Almusaylim and N. Zaman, "A review on smart home present state and challenges: linked to context-awareness internet of things (iot)," *Wireless Networks*, vol. 25, pp. 3193–3204, 8 2019.

[6] M. Aly and M. H. Behiry, "Enhancing anomaly detection in iot-driven factories using logistic boosting, random forest, and svm: A comparative machine learning approach," *Scientific Reports*, vol. 15, pp. 1–17, 7 2025.

[7] E. Vildjiounaite, S.-M. Mäkelä, T. Keränen, V. Kyllönen, V. Huotari, S. Järvinen, and G. Gimel'farb, "Unsupervised illness recognition via in-home monitoring by depth cameras," *Pervasive and Mobile Computing*, vol. 38, pp. 166–187, 7 2017.

[8] G. Leite, G. Silva, and H. Pedrini, "Fall detection in video sequences based on a three-stream convolutional neural network," in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*. IEEE, 12 2019, pp. 191–195.

[9] S. S. Khan, P. K. Mishra, N. Javed, B. Ye, K. Newman, A. Mihailidis, and A. Iaboni, "Unsupervised deep learning to detect agitation from videos in people with dementia," *IEEE Access*, vol. 10, pp. 10 349–10 358, 2022.

[10] F. M. Noori, M. Riegler, M. Z. Uddin, and J. Torresen, "Human activity recognition from multiple sensors data using multi-fusion representations and cnns," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 16, pp. 1–19, 5 2020.

[11] D. Fährmann, L. Martín, L. Sánchez, and N. Damer, "Anomaly detection in smart environments: A comprehensive survey," *IEEE Access*, vol. 12, pp. 64 006–64 049, 2024.

[12] S. Birnbach, S. Eberz, and I. Martinovic, "Haunted house: Physical smart home event verification in the presence of compromised sensors," *ACM Transactions on Internet of Things*, vol. 3, pp. 1–28, 8 2022.

[13] T. Mujirishvili, C. Maidhof, F. Florez-Revuelta, M. Ziefle, M. Richart-Martinez, and J. Cabrero-García, "Acceptance and privacy perceptions toward video-based active and assisted living technologies: Scoping review," *Journal of Medical Internet Research*, vol. 25, p. e45297, 5 2023. [Online]. Available: https://www.jmir.org/2023/1/e45297

[14] D. Gonzalez, M. A. Patricio, A. Berlanga, and J. M. Molina, "Variational autoencoders for anomaly detection in the behaviour of the elderly using electricity consumption data," *Expert Systems*, vol. 39, p. e12744, 5 2022.

[15] M. E. Bilgin, H. Kilinc, and A. H. Zaim, "An anomaly detection study for the smart home environment," in *2022 7th International Conference on Computer Science and Engineering (UBMK)*. IEEE, 9 2022, pp. 31–36.

[16] S. W. Yahaya, C. Langensiepen, and A. Lotfi, "Anomaly detection in activities of daily living using one-class support vector machine," *Advances in Intelligent Systems and Computing*, vol. 840, pp. 362–371, 2019.

[17] N. I. Haque, M. A. Rahman, and H. Shahriar, "Ensemble-based efficient anomaly detection for smart building control systems," pp. 504–513, 7 2021.

[18] M. J. Reis and C. Serôdio, "Edge ai for real-time anomaly detection in smart homes," *Future Internet 2025, Vol. 17, Page 179*, vol. 17, p. 179, 4 2025.

[19] R. A. Sater and A. B. Hamza, "A federated learning approach to anomaly detection in smart buildings," *ACM Transactions on Internet of Things*, vol. 2, pp. 1–23, 8 2021.

[20] I. Ahmed, G. Jeon, and F. Piccialli, "A deep-learning-based smart healthcare system for patient's discomfort detection at the edge of internet of things," *IEEE Internet of Things Journal*, vol. 8, pp. 10 318–10 326, 7 2021.

[21] J.-W. Hsieh, C.-H. Chuang, S. Alghyaline, H.-F. Chiang, and C.-H. Chiang, "Abnormal scene change detection from a moving camera using bags of patches and spider-web map," *IEEE Sensors Journal*, vol. 15, pp. 1–1, 2014.

[22] C. Zhu, W. Sheng, and M. Liu, "Wearable sensor-based behavioral anomaly detection in smart assisted living systems," *IEEE Transactions on Automation Science and Engineering*, vol. 12, pp. 1225–1234, 10 2015.

[23] C. Sager, C. Janiesch, and P. Zschech, "A survey of image labelling for computer vision applications," pp. 91–110, 7 2021. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/2573234X.2021.1944691

[24] S. Bianco, G. Ciocca, P. Napoletano, and R. Schettini, "An interactive tool for manual, semi-automatic and automatic video annotation," *Computer Vision and Image Understanding*, vol. 131, pp. 88–99, 2 2015.

[25] P. Denzler, M. Ziegler, A. Jacobs, V. Eiselein, P. Neumaier, and M. Koppel, "Multi-sensor data annotation using sequence-based active learning," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 10 2022, pp. 258–263.

[26] B. J. Boom, P. X. Huang, J. He, and R. B. Fisher, "Supporting ground-truth annotation of image datasets using clustering," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, 2012, pp. 1542–1545.

[27] R. Aljundi, P. Chakravarty, and T. Tuytelaars, "Who's that actor? automatic labelling of actors in tv series starting from imdb images," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10113 LNCS, pp. 467–483, 2017.

[28] X. Zhuo, F. Fraundorfer, F. Kurz, and P. Reinartz, "Automatic annotation of airborne images by label propagation based on a bayesian-crf model," *Remote Sensing 2019, Vol. 11, Page 145*, vol. 11, p. 145, 1 2019.

[29] M. Barz, M. M. Moniri, M. Weber, and D. Sonntag, "Multimodal multisensor activity annotation tool," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM, 9 2016, pp. 17–20.

[30] F. Cruciani, I. Cleland, C. Nugent, P. McCullagh, K. Synnes, and J. Hallberg, "Automatic annotation for human activity recognition in free living using a smartphone," *Sensors (Switzerland)*, vol. 18, 7 2018.

[31] A. Diete, T. Sztyler, and H. Stuckenschmidt, "Exploring semi-supervised methods for labeling support in multimodal datasets," *Sensors 2018, Vol. 18, Page 2639*, vol. 18, p. 2639, 8 2018.

[32] D. Cook, K. D. Feuz, and N. C. Krishnan, "Transfer learning for activity recognition: a survey," *Knowl. Inf. Syst.*, vol. 36, no. 3, p. 537–556, Sep. 2013.

[33] N. P. Owoh, M. M. Singh, and Z. F. Zaaba, "Automatic annotation of unlabeled data from smartphone-based motion and location sensors," *Sensors (Switzerland)*, vol. 18, 7 2018.