Simen Owen Academic Proceedings Series

3rd International Conference on Electronics, Engineering, Computer Science and Applied Development (EESD 2026)

Article

Semi-Supervised Anomaly Detection in Industrial IoT Sensors via Contrastive Predictive Coding

Litong Xu[1,*]

[1] New Channel Jinqiu International, Qingdao, China
* Correspondence: Litong Xu, New Channel Jinqiu International, Qingdao, China

Abstract: Industrial Internet of Things sensor streams support continuous monitoring of equipment conditions and process states, but anomaly detection remains difficult because abnormal events are rare, sensor variables are strongly coupled, and reliable labels are limited. Existing reconstruction- and one-class methods may fail to distinguish abnormal patterns from valid operating transitions, while fully supervised models depend on costly fault annotations. This study proposes SS-CPC-AD, a semi-supervised anomaly detection framework based on contrastive predictive coding. The method combines a multiscale temporal encoder, autoregressive future-state prediction, a label-aware contrastive objective, confidence-weighted pseudo-labeling, and an adaptive anomaly score integrating predictive inconsistency, deviation from the normal representation center, and classifier probability. Experiments on the SWaT and WADI datasets used 5% labeled training windows and five random seeds. SS-CPC-AD achieved F1-scores of 0.892±0.009 and 0.731±0.014, exceeding the strongest baseline by 3.1 and 4.0 percentage points, respectively. Ablation results showed that removing label-aware contrast reduced F1 by 2.8 points on SWaT and 4.0 points on WADI. Under 20% Gaussian sensor noise, the method retained 94.2% and 91.6% of its original F1-score. These results indicate that limited supervision can improve predictive representation learning while supporting more stable anomaly detection and model-level sensor attribution.

Keywords: industrial internet of things; anomaly detection; contrastive predictive coding; semi-supervised learning; multivariate sensor data

Received: 31 May 2026
Revised: 17 July 2026
Accepted: 28 July 2026
Published: 01 August 2026

1. Introduction

Industrial Internet of Things systems continuously collect multivariate measurements from sensors, actuators, and control devices for process monitoring and equipment management [1]. These data streams support the detection of sensor faults, equipment degradation, control deviations, and security-related incidents in cyber-physical systems. However, anomaly detection remains challenging because abnormal events are rare, process variables exhibit strong temporal dependencies, and sensors are physically coupled through operational constraints. Reliable anomaly labels typically require expert inspection or controlled experimental interventions, making fully supervised training impractical in many industrial settings.

Existing methods mainly rely on reconstruction, one-class representation learning, or supervised classification. Reconstruction-based models learn normal patterns from unlabeled observations and identify deviations through reconstruction errors. However, high-capacity models may reconstruct abnormal inputs with low fidelity loss, reducing the discriminative margin between normal and anomalous windows. One-class methods construct compact representations of normal behavior but may generate false alarms when operating conditions shift. Supervised classifiers directly learn decision boundaries, yet their performance often degrades significantly when only a limited number of labeled anomalies are available [2]. In addition, many approaches focus exclusively on reconstructing current observations rather than evaluating whether historical states can reliably predict future sensor behavior. Fixed anomaly scoring schemes and uniformly weighted pseudo-labels may further constrain model adaptability and propagate erroneous predictions during semi-supervised training.

To address these limitations, this study proposes a semi-supervised anomaly detection framework based on contrastive predictive coding, termed SS-CPC-AD. Its main objectives and contributions are as follows: (1) A label-aware contrastive predictive objective jointly learns from unlabeled sensor windows and a small labeled subset. Unlabeled data capture temporal predictability, while labeled normal and anomalous windows constrain the latent boundary. This component is evaluated through ablation experiments and comparison with an unsupervised CPC baseline [2]. (2) A confidence-weighted pseudo-label mechanism reduces the influence of uncertain predictions. Training weights are assigned according to prediction confidence, and low-confidence samples are excluded. Its effectiveness is examined through ablation experiments and tests under different labeled-data ratios. (3) An adaptive anomaly score combines predictive inconsistency, deviation from the normal representation center, and supervised anomaly probability. The score is calibrated on validation data, while gradient-based attribution identifies sensors associated with each alarm. This mechanism is evaluated through scoring ablations, noise-robustness tests, and attribution cases.

The technical route includes sensor synchronization, missing-value processing, sliding-window construction, multiscale temporal encoding, autoregressive context modeling, contrastive future-state prediction, confidence-controlled optimization, and validation-based anomaly decisions. The temporal encoder captures changes at different durations, while the autoregressive network summarizes historical states [3]. Contrastive predictive coding distinguishes the observed future representation from alternative candidates, and limited labels provide class information without replacing the predictive objective.

Academically, this study examines how predictive representation learning and limited supervision can be integrated for industrial anomaly detection. Practically, the framework aims to reduce dependence on extensive fault annotation while maintaining stability under limited labels and sensor noise. Retained data partitions, random seeds, threshold-selection procedures, score components, and sensor attribution records also support audit-oriented analysis [1]. However, attribution reflects model-level input contributions rather than physical causality, and compliance with specific industrial standards requires separate assessment.

2. Related Works

2.1. Reconstruction- and Boundary-Based Anomaly Detection

Reconstruction-based methods are widely used for industrial sensor anomaly detection because they can learn normal operating patterns without requiring extensive fault annotations. Autoencoders, recurrent reconstruction networks, and related variants encode multivariate sensor windows and reconstruct the original observations. Anomalies are identified when reconstruction errors exceed a predefined threshold [2]. These methods can model nonlinear sensor relationships and are suitable for systems with abundant normal data.

However, reconstruction quality does not always ensure reliable anomaly discrimination [4]. High-capacity models may also reconstruct abnormal inputs, especially when anomalies resemble common operating states or persist over extended durations. Their scores are sensitive to window length, feature scaling, and threshold selection. Boundary-based methods instead learn a compact normal region, but a single boundary may not adequately represent multiple operating modes, causing normal transitions to be misclassified.

Predictive representation learning differs by examining whether future sensor behavior is consistent with historical states rather than whether the current input can be reconstructed. This is relevant because some industrial faults initially affect temporal dependencies before producing large point-wise deviations. Nevertheless, many reconstruction- and boundary-based methods either ignore limited anomaly annotations or use them only in a separate classifier [3]. The proposed framework integrates labeled guidance directly into predictive representation learning, and its impact is evaluated through ablation experiments.

2.2. Contrastive Learning for Multivariate Sensor Data

Contrastive learning enables representation learning from unlabeled time-series data without requiring reconstruction of individual sensor values [5]. It promotes similarity between related samples while enhancing dissimilarity between unrelated ones. Predictive contrastive approaches further leverage historical context to distinguish the correct future latent representation from a set of alternatives. This capability supports modeling of delayed dependencies, periodic patterns, and cross-sensor correlations.

Performance critically depends on how positive and negative sample pairs are constructed. Random signal augmentations risk distorting the physical interpretation of sensor measurements, and temporally distant windows may not constitute valid negatives, as industrial processes can cyclically return to similar operational conditions. Moreover, purely self-supervised objectives do not inherently enforce separation between verified normal and anomalous instances [4]. Current predictive methods typically apply statistical distance metrics post-hoc, resulting in weak alignment between the representation learning objective and the final anomaly decision.

This study integrates contrastive future-state prediction with label-informed boundary shaping. A small labeled subset guides the formation of discriminative latent structure, while unlabeled temporal windows sustain robust sequence modeling. The approach bridges unsupervised predictive learning and low-label classification tasks. Evaluation includes comparison against an unsupervised CPC baseline and ablation of the label-informed component.

2.3. Semi-Supervised and Explainable Industrial Detection

Semi-supervised anomaly detection provides a practical balance between unsupervised learning and fully supervised classification. It leverages abundant unlabeled observations alongside a limited set of confirmed fault instances. Common methodological approaches include pseudo-labeling, consistency regularization, and joint reconstruction-classification learning [6].

A key limitation lies in error accumulation: incorrect pseudo-labels can progressively distort the decision boundary, especially under class imbalance or shifting operating conditions. Many existing methods treat all pseudo-labeled samples uniformly, disregarding substantial variations in prediction confidence. Moreover, anomaly decisions often depend on a single scalar output—such as reconstruction error or class probability—which may exhibit inconsistent sensitivity across different anomaly types [4]. Explainability is frequently restricted to attention weights or feature rankings that lack direct interpretability with respect to the final anomaly score.

The proposed method mitigates these issues by weighting pseudo-labels according to prediction confidence and excluding low-confidence samples from the semi-supervised loss. It integrates predictive inconsistency, deviation from the normal representation center, and calibrated anomaly probability into a validation-calibrated scoring function. Gradient-based sensor attribution explicitly identifies influential input variables contributing to the final decision [2]. These components are systematically evaluated through pseudo-labeling and scoring ablation studies, experiments under label scarcity, robustness tests against noise, and illustrative attribution cases. The study thus assesses whether predictive learning, judicious label utilization, and score-level interpretation collectively enhance performance on the selected industrial datasets.

3. Methodology

3.1. Overall Architecture

This study proposes SS-CPC-AD, a semi-supervised contrastive predictive coding framework for anomaly detection in multivariate industrial sensor streams [7]. The method jointly learns temporal process representations, class-aware boundaries, and confidence-controlled pseudo-labels, integrating representation learning and classification into a unified optimization objective.

As shown in Figure 1, the framework comprises five interconnected components: sensor synchronization and window construction, multiscale temporal encoding, autoregressive context modeling, label-aware contrastive predictive learning, and adaptive anomaly scoring. Raw sensor and actuator records are aligned to a common timeline, preprocessed to handle missing values, and segmented into overlapping windows. Each window is encoded via parallel causal convolution branches. A gated recurrent unit then aggregates sequential latent representations to predict future latent states. Predictive learning leverages unlabeled windows, labeled windows enforce class separation, and high-confidence pseudo-labels from unlabeled data are incorporated with adaptive weighting [4]. During inference, anomaly scores are generated by fusing predictive inconsistency, representation deviation, and classifier probability, enabling both instance-level detection and sensor-level attribution. Solid paths in Figure 1 denote operations active in both training and inference; dashed paths indicate label-dependent training-only operations.

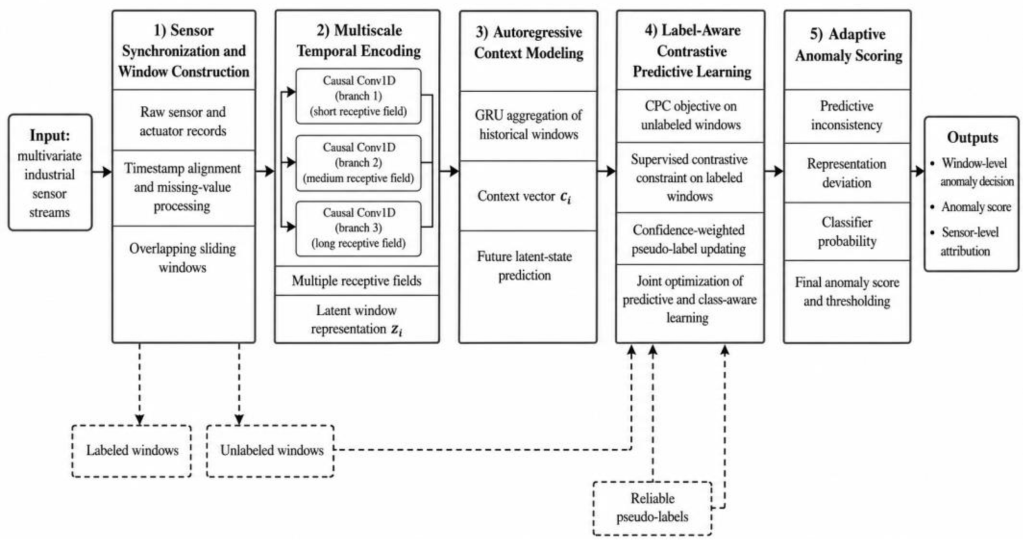

Figure 1. Overall Architecture of the Proposed Ss-cpc-ad Framework.

3.2. Sensor Synchronization and Window Labeling

Industrial sensor records often exhibit asynchronous timestamps, missing observations, and variables spanning disparate scales. All sensor channels are resampled to a uniform sampling interval [8]. Missing segments shorter than five consecutive time steps are imputed via linear interpolation; windows containing longer gaps are excluded from further processing. Continuous variables are standardized using statistics computed from the training set, whereas binary actuator states remain unchanged.

Let $x_t = [x_t^{(1)}, \dots, x_t^{(D)}]^\top \in \mathbb{R}^D$ denote the synchronized observation at time t, where D is the number of variables [9]. The i-th window and its corresponding label are defined as

$$X_i = [\tilde{x}_i, \dots, \tilde{x}_{i+L-1}] \in \mathbb{R}^{L\times D}, y_i = \begin{cases} 1, & X_i \text{ overlaps an anomaly,} \\ 0, & X_i \text{ is normal,} \\ -1, & X_i \text{ is unlabeled,} \end{cases} \quad (1)$$

where L is the window length and $\tilde{x}_t$ is the processed observation [10]. Unlabeled windows contribute to predictive modeling and pseudo-label generation but are excluded from supervised loss computation.

3.3. Multiscale Temporal Encoder

Industrial anomalies may manifest as short impulses, gradual drifts, or sustained deviations [11]. To capture these diverse temporal patterns, the encoder employs three parallel one-dimensional causal convolution branches with kernel sizes $k_r \in \{3,5,9\}$.

$$h_i^{(r)} = \phi\left(\mathrm{Conv1D}_{k_r}(X_i)\right), z_i = W_z[h_i^{(1)} \parallel h_i^{(2)} \parallel h_i^{(3)}] + b_z, \quad (2)$$

Here, $\phi(\cdot)$ denotes the GELU activation function, $\parallel$ represents concatenation, and $z_i \in \mathbb{R}^{d_z}$ is the resulting latent representation. Layer normalization and dropout are applied after linear projection, and causal convolution ensures that only past and current observations influence the representation [12].

3.4. Autoregressive Context Modeling

A single window may fail to distinguish an anomaly from a normal process transition; thus, a gated recurrent unit aggregates H historical representations.

$$c_i = \mathrm{GRU}(z_{i-H+1}, \dots, z_i), \quad (3)$$

Here, $c_i \in \mathbb{R}^{d_c}$ denotes the context vector. For each future step k, a linear predictor estimates the corresponding latent state.

$$\hat{z}_{i+k} = W_k c_i + b_k, k = 1, \dots, K, \quad (4)$$

In this formulation, K represents the prediction horizon, while W_k and b_k are trainable parameters.

3.5. Contrastive Predictive Learning

The context vector is trained to distinguish the actual future representation from a set of candidate alternatives. The true pair $(\hat{z}_{i+k}, z_{i+k})$ constitutes the positive sample, while representations from other samples in the batch serve as negative candidates:

$$\mathcal{L}_{\mathrm{CPC}} = -\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=1}^{K}\log\frac{\exp[\mathrm{sim}(\hat{z}_{i+k}, z_{i+k})/\tau]}{\sum_{j=1}^{N}\exp[\mathrm{sim}(\hat{z}_{i+k}, z_{j+k})/\tau]}, \quad (5)$$

where N denotes the batch size, $\mathrm{sim}(\cdot,\cdot)$ represents cosine similarity, and τ is the temperature parameter. To avoid false negatives, temporally adjacent windows drawn from the same operating sequence are excluded from the negative set. Minimizing this loss enhances the model's ability to capture normal temporal dependencies; conversely, anomalies that deviate from expected process dynamics yield larger prediction errors.

3.6. Label-Aware Contrastive Constraint

Pure CPC learning captures temporal structure but does not explicitly distinguish between confirmed normal and anomalous samples. A supervised contrastive objective is therefore introduced:

$$\mathcal{L}_{\mathrm{sup}} = -\frac{1}{|\mathcal{I}_L|}\sum_{i\in\mathcal{I}_L}\frac{1}{|\mathcal{P}(i)|}\sum_{p\in\mathcal{P}(i)}\log\frac{\exp[\mathrm{sim}(z_i, z_p)/\tau_s]}{\sum_{a\in\mathcal{I}_L\setminus\{i\}}\exp[\mathrm{sim}(z_i, z_a)/\tau_s]}, \quad (6)$$

where $\mathcal{I}_L$ is the labeled index set, $\mathcal{P}(i)$ contains labeled samples sharing the same class as i, and τ_s is the supervised contrastive temperature parameter. This objective encourages intra-class cohesion and inter-class separation between normal and anomalous samples in the predictive latent space.

3.7. Confidence-Weighted Pseudo-Label Updating

The classifier outputs an anomaly probability $p_i = \sigma(w_c^{\top} z_i + b_c)$. For each unlabeled window, the pseudo-label, confidence, and training weight are computed as follows.

$$\hat{y}_i = \mathbb{I}(p_i \geq 0.5), q_i = 2\,|\,p_i - 0.5\,|, \omega_i = \mathbb{I}(q_i \geq \delta)q_i, \quad (7)$$

Here, $\mathbb{I}(\cdot)$ denotes the indicator function, $q_i \in [0,1]$ represents the confidence score, and δ is the acceptance threshold. Predictions with confidence below this threshold are excluded from the pseudo-label loss [1]. Pseudo-label learning begins after a 10-epoch warm-up phase. The confidence threshold starts at 0.90 and decreases by 0.02—down to a minimum of 0.80—whenever validation F1 remains unchanged for three consecutive epochs. This adaptive strategy mitigates early error propagation while gradually incorporating more unlabeled samples as model representations mature.

3.8. Joint Optimization and Adaptive Anomaly Scoring

The complete training objective is

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\mathrm{CPC}} + \lambda_2 \mathcal{L}_{\mathrm{sup}} + \lambda_3 \frac{1}{|\mathcal{I}_L|} \sum_{i \in \mathcal{I}_L} \mathrm{BCE}(p_i, y_i) + \lambda_4 \frac{1}{|\mathcal{I}_U|} \sum_{i \in \mathcal{I}_U} \omega_i \, \mathrm{BCE}(p_i, \hat{y}_i), \quad (8)$$

where $\mathcal{I}_U$ is the unlabeled index set, BCE denotes binary cross-entropy loss, and λ_1–λ_4 are weighting coefficients for the respective loss components. Model parameters are updated using the AdamW optimizer as $\Theta \leftarrow \Theta - \eta \nabla_{\Theta} \mathcal{L}$, with η representing the learning rate.

During inference, the final anomaly score is

$$s_i = \alpha e_i^{\mathrm{pred}} + \beta e_i^{\mathrm{dev}} + \gamma p_i, \alpha + \beta + \gamma = 1, \quad (9)$$

where e_i^{pred} is the future-state prediction error, e_i^{dev} is the Mahalanobis distance from the normal representation center, and p_i is the classifier output probability. The weighting coefficients and decision threshold are determined exclusively on the validation set [13]. A time window is classified as anomalous when $s_i > \theta$. Sensor attribution is derived from the absolute value of the input-gradient product and reflects model-level feature sensitivity rather than physical causality.

Algorithm 1. Training and Inference of SS-CPC-AD

Input: Sensor sequence X, labeled set DL, unlabeled set DU,
window length L, horizon K, confidence threshold δ
Output: Parameters Θ, anomaly scores S, decisions Y, attributions A
1: Synchronize timestamps and process missing observations
2: Normalize variables and construct sliding windows
3: Initialize encoder, GRU, predictors, and classifier
4: for epoch = 1 to E do
5: Sample labeled batch BL and unlabeled batch BU
6: Encode windows into latent representations z
7: Aggregate historical representations into contexts c
8: Predict the next K latent representations
9: Compute CPC loss using valid contrastive pairs
10: Compute supervised contrastive and classification losses
11: Generate pseudo-labels and confidence weights for BU
12: Update Θ by minimizing Equation (8)
13: Adjust δ using validation performance
14: end for
15: Calibrate score weights and threshold; output S, Y, and A

4. Results and Analysis

4.1. Experimental Setup

Experiments were conducted on the SWaT and WADI industrial control datasets. SWaT contains seven days of normal operation and four days of anomalous event data, with 51 sensor and actuator variables and 41 recorded anomalous events. WADI contains 14 normal days and two anomalous event days, with 123 variables and 15 anomalous events. SWaT represents a staged water-treatment process, whereas WADI models a larger water-distribution system with more complex variable interactions. Because both datasets originate from the same research centre, the evaluation examines performance consistency across two industrial water-system datasets rather than cross-institutional generalization. The datasets were obtained through the official research-access procedure, and their raw files cannot be redistributed. Reproducibility materials include preprocessing scripts, configuration files, dataset hashes, partition indices, and preprocessing logs.

All variables were aligned to one-second intervals. Missing sequences of no more than five samples were filled by linear interpolation, while windows containing longer gaps were removed. Continuous variables were standardized using training-set statistics, and binary actuator states retained their original values. The normal sequence and the first 20% of the anomalous event sequence were used to construct the training and validation sets, while the remaining 80% of the anomalous event sequence was reserved for testing. The combined training data were divided chronologically into 80% for training and 20% for validation. Only 5% of the training windows retained their normal or anomalous labels, while the remainder were treated as unlabeled [10]. A one-window gap was inserted between partitions to prevent overlap leakage.

Four baselines were included: LSTM-AE, Deep SVDD, USAD, and TRL-CPC, representing reconstruction, one-class learning, adversarial reconstruction, and unsupervised predictive contrastive detection. All methods used identical data partitions, windows, and validation-based threshold selection [14]. SS-CPC-AD was trained with AdamW, a learning rate of 1×10^{-3}, batch size 128, latent dimension 128, a three-step prediction horizon, and early stopping with a patience of 10 epochs. Experiments were performed on an Intel Core i7-12700K CPU, 32 GB RAM, and an NVIDIA GeForce RTX 3060 GPU. Five runs used seeds 13, 21, 42, 87, and 102. Precision, Recall, and F1-score were reported as mean ± standard deviation, with F1 as the primary metric. Statistical significance was assessed using two-sided paired t-tests on seed-matched F1 values, with $p<0.05$.

4.2. Performance Comparison

Table 1 compares SS-CPC-AD with four anomaly detection baselines on SWaT and WADI. All results are reported as mean ± standard deviation over five independent runs (n=5). SS-CPC-AD achieved the highest F1-score on both datasets, though baseline methods exhibited varying balances between Precision and Recall [15].

Table 1. Performance Comparison on SWaT and WADI, Mean ± SD over Five Runs.

Method	SWaT Precision	SWaT Recall	SWaT F1	WADI Precision	WADI Recall	WADI F1
LSTM-AE	0.834 ± 0.018	0.801 ± 0.022	0.817 ± 0.015	0.654 ± 0.026	0.601 ± 0.031	0.626 ± 0.023
Deep SVDD	0.812 ± 0.021	0.836 ± 0.019	0.824 ± 0.014	0.628 ± 0.030	0.668 ± 0.027	0.647 ± 0.021
USAD	0.858 ± 0.015	0.841 ± 0.017	0.849 ± 0.011	0.693 ± 0.021	0.661 ± 0.024	0.676 ± 0.018
TRL-CPC	0.873 ± 0.012	0.850 ± 0.014	0.861 ± 0.010	0.718 ± 0.019	0.666 ± 0.022	0.691 ± 0.016
SS-CPC-AD	0.901 ± 0.010	0.884 ± 0.012	0.892 ± 0.009	0.748 ± 0.016	0.715 ± 0.019	0.731 ± 0.014

SS-CPC-AD achieved the highest F1-score on both datasets. Compared with TRL-CPC, it improved F1 by 3.1 percentage points on SWaT and 4.0 points on WADI [4]. Two-sided paired t-tests confirmed statistically significant differences on SWaT (p=0.008) and WADI (p=0.011).

The baselines exhibited differing Precision–Recall trade-offs. Deep SVDD yielded higher Recall than LSTM-AE but lower Precision, indicating more detected anomalies accompanied by increased false positives [4]. USAD produced relatively balanced outcomes but remained below TRL-CPC, suggesting that future-state prediction captures temporal deviations more effectively than reconstruction alone.

The larger improvement on WADI is attributed to the label-aware and confidence-weighted mechanisms. Its higher dimensionality increases overlap between normal operational transitions and unsupervised anomaly scores. Limited labeled data help refine the latent decision boundary, while confidence weighting suppresses unreliable pseudo-labels. However, the lower absolute performance on WADI indicates that complex sensor interdependencies remain challenging to model.

4.3. Ablation Study and Mechanism Verification

Table 2 quantifies the contribution of each component under identical data partitions, model parameters, and random seeds.

Table 2. Ablation Results on SWaT and WADI, Mean ± SD over Five Runs.

Variant	SWaT F1	WADI F1
Full SS-CPC-AD	0.892 ± 0.009	0.731 ± 0.014
w/o Multiscale Encoder	0.873 ± 0.012	0.704 ± 0.018
w/o Label-Aware Contrast	0.864 ± 0.013	0.691 ± 0.019
w/o Confidence Weighting	0.879 ± 0.011	0.708 ± 0.017
Predictive Score Only	0.858 ± 0.015	0.682 ± 0.021
w/o CPC Learning	0.846 ± 0.017	0.671 ± 0.023

Excluding the future-state prediction objective led to the largest performance drop, reducing F1 scores by 4.6 points on SWaT and 6.0 points on WADI. This confirms that modeling temporal dynamics through future-state prediction serves as the primary mechanism for representation learning. In its absence, the model relies more heavily on scarce labeled examples and captures underlying process dynamics less effectively [5, 8].

Omitting the class-aware contrastive objective reduced F1 scores by 2.8 and 4.0 points on SWaT and WADI, respectively. The larger decline on WADI suggests that discriminative guidance between classes is especially beneficial when normal instances occupy a broader region in latent space. Removing confidence-based weighting yielded smaller but consistent performance losses, with a more pronounced effect on WADI due to higher frequency of uncertain pseudo-labels [9].

The multiscale encoder delivered moderate improvements—1.9 points on SWaT and 2.7 points on WADI—indicating that capturing patterns across multiple temporal scales enhances detection of anomalies with varying durations. The variant using predictive scores alone also underperformed, as gradual operational shifts may retain local predictability while drifting away from the central region of normal representations. Thus, predictive inconsistency, latent-space deviation, and classifier probability jointly provide complementary anomaly signals.

Paired statistical comparisons between the full model and each ablated variant yielded statistically significant differences ($p < 0.05$) on both datasets. The smallest significance margin occurred between the full model and the SWaT variant without confidence weighting ($p = 0.041$).

4.4. Convergence and Statistical Analysis

Figure 2 shows the validation-loss curves of SS-CPC-AD, TRL-CPC, the model without confidence weighting, and the model without CPC learning. Curves represent the mean over five independent training runs, with shaded regions indicating ±1 standard deviation.

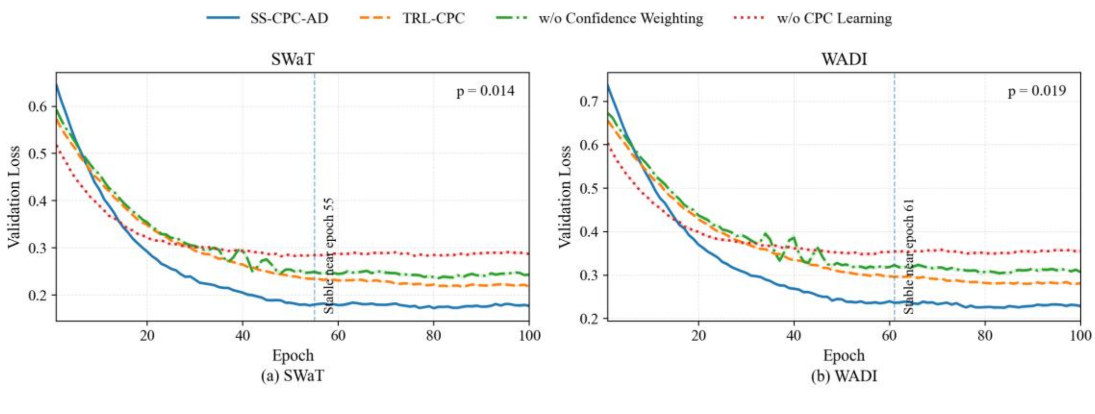

Figure 2. Validation-loss Convergence on SWaT and WADI, Mean ± SD over Five Runs: (A) SWaT and (B) WADI.

SS-CPC-AD did not achieve the lowest loss in every early epoch due to joint optimization of predictive, supervised contrastive, and classification objectives. After approximately 15 epochs, its loss decreased more consistently and stabilized near epoch 55 on SWaT and epoch 61 on WADI.

TRL-CPC exhibited smoother early optimization but converged to a higher final loss [8]. The model without confidence weighting displayed greater epoch-to-epoch fluctuations between epochs 30 and 50, especially on WADI, as uncertain pseudo-labels repeatedly perturbed the decision boundary. The model without CPC learning converged more rapidly at first but attained the highest stable loss, demonstrating that faster convergence does not necessarily reflect superior representation quality.

The mean loss over the final ten epochs was significantly lower for SS-CPC-AD than for TRL-CPC on both SWaT ($p=0.014$) and WADI ($p=0.019$). Its late-stage standard deviation was also smaller, indicating that confidence-based filtering enhanced optimization stability [2, 12].

4.5. Stability, Robustness, and Interpretability

Figure 3 evaluates SS-CPC-AD under varying labeled-data ratios and Gaussian sensor-noise levels. All results are reported as mean F1 score ± standard deviation over five independent runs.

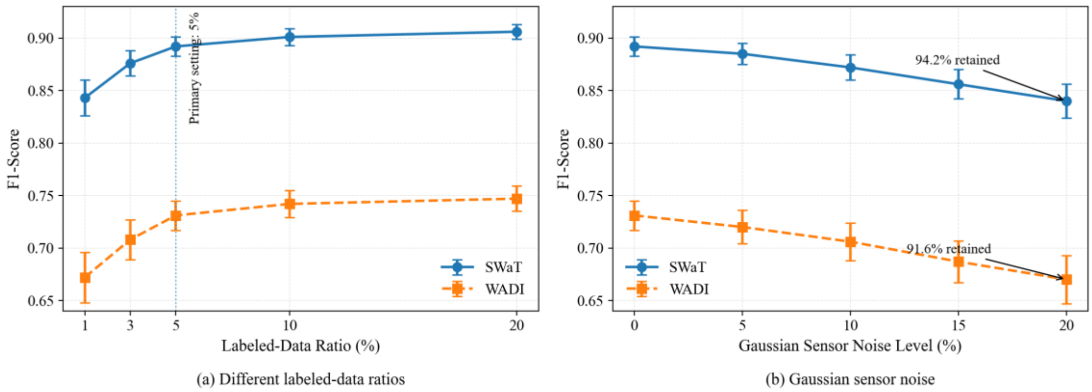

Figure 3. Stability and Robustness of Ss-cpc-ad: (A) Different Labeled-Data Ratios and (B) Gaussian Sensor Noise, Mean ± SD over Five Runs.

In Figure 3(a), reducing the labeled ratio from 5% to 1% decreased the F1 score from 0.892±0.009 to 0.843±0.017 on SWaT and from 0.731±0.014 to 0.672±0.024 on WADI. Increasing the ratio to 10% improved the F1 score to 0.901±0.008 and 0.742±0.013, respectively, while further increasing to 20% yielded only marginal gains of 0.906±0.007 and 0.747±0.012, indicating diminishing returns. Figure 3(b) introduces Gaussian noise at levels ranging from 5% to 20% of each sensor's training-set standard deviation. At 20% noise, the F1 score remained at 0.840±0.016 on SWaT and 0.670±0.023 on WADI, retaining 94.2% and 91.6% of the original performance. The larger performance decline observed on WADI reflects its more complex sensor interdependencies.

Sensor contributions were computed using the input-gradient product with respect to the final anomaly score. Highly ranked variables correlated with shifts in predictive inconsistency and classifier probability, whereas gradual anomalies exhibited stronger deviations from the normal-center distribution. These attributions reflect model sensitivity rather than physical causality.

Trustworthiness was enhanced through chronological data partitioning, seed-level result reporting, threshold selection restricted to validation data, confidence-controlled pseudo-labeling, and explicit decomposition of the anomaly score. These methodological safeguards improve reproducibility but do not constitute formal certification against any industrial standard.

Evaluation on SWaT and WADI demonstrates consistent performance across two water treatment process testbeds [14]. As both datasets originate from the same research institution, generalization across institutions or industrial sectors remains unverified.

5. Conclusion

This study evaluated whether semi-supervised contrastive predictive coding could improve anomaly detection in multivariate industrial sensor streams under limited labeling. On SWaT and WADI, SS-CPC-AD achieved F1-scores of 0.892±0.009 and 0.731±0.014, exceeding the strongest baseline by 3.1 and 4.0 percentage points, respectively. These gains were statistically significant across five seed-matched runs.

The results correspond directly to the three contributions introduced in Chapter 1. First, the label-aware contrastive objective improved class separation within the predictive latent space; removing this component reduced F1 by 2.8 points on SWaT and 4.0 points on WADI. Second, confidence-weighted pseudo-labeling limited the influence of uncertain unlabeled samples; its removal caused consistent performance degradation and larger fluctuations during training, particularly on WADI. Third, the combined anomaly score outperformed predictive inconsistency alone by integrating temporal prediction error, deviation from the normal representation center, and classifier probability. Sensor-level attribution further identified variables that contributed to each decision, although these contributions should not be interpreted as physical causality.

The method also retained 94.2% of its original F1-score on SWaT and 91.6% on WADI under 20% Gaussian sensor noise. Performance gains became limited when the labeled ratio exceeded 10%, suggesting that the framework can use small labeled subsets effectively without assuming diminishing returns from additional labels.

Several limitations remain. The evaluation used two water-system testbeds from the same research centre, so cross-sector and cross-institutional generalization has not been established. Fixed-length windows may not represent anomalies with highly variable durations, and contrastive batch construction increases training cost. Future work should examine adaptive windowing, process-aware structural considerations, online threshold updating under operational drift, and validation on manufacturing and energy-system datasets.

References

1. Á. L. P. Gómez, L. F. Maimó, A. H. Celdrán, and F. J. G. Clemente, "Madics: A methodology for anomaly detection in industrial control systems," *Symmetry*, vol. 12, no. 10, p. 1583, 2020.
2. V. Shukla, A. Shukla, S. P. SK, and S. Shukla, "A systematic survey: role of deep learning-based image anomaly detection in industrial inspection contexts," *Frontiers in Robotics and AI*, vol. 12, p. 1554196, 2025.
3. W. Song, X. Li, P. Chen, J. Chen, J. Ren, and Y. Xia, "A novel graph structure learning based semi-supervised framework for anomaly identification in fluctuating IoT environment," *CMES-Computer Modeling in Engineering and Sciences*, vol. 140, no. 3, pp. 3001–3016, 2024.
4. R. Xiao, W. Li, J. Lu, and S. Jin, "Contexlog: non-parsing log anomaly detection with all information preservation and enhanced contextual representation," *IEEE Transactions on Network and Service Management*, vol. 21, no. 4, pp. 4750–4762, 2024.
5. M. Raeiszadeh, A. Ebrahimzadeh, R. H. Glitho, J. Eker, and R. A. Mini, "Real-time adaptive anomaly detection in industrial IoT environments," *IEEE Transactions on Network and Service Management*, vol. 21, no. 6, pp. 6839–6856, 2024.
6. F. Arellano-Espitia, M. Delgado-Prieto, A. D. Gonzalez-Abreu, J. J. Saucedo-Dorantes, and R. A. Osornio-Rios, "Deep-compact-clustering based anomaly detection applied to electromechanical industrial systems," *Sensors*, vol. 21, no. 17, p. 5830, 2021.
7. M. E. Mumcuoglu, S. M. Farea, M. Unel, S. Mise, S. Unsal, E. Cevik, et al., "Detecting APS failures using LSTM-AE and anomaly transformer enhanced with human expert analysis," *Engineering Failure Analysis*, vol. 165, p. 108811, 2024.
8. N. Liao and Q. Li, "Anomaly Detection Method Based on Semi-Supervised Deep Learning for Internet of Things Traffic," in *2024 2nd International Conference on Machine Vision, Image Processing & Imaging Technology (MVIPIT)*, Sep. 2024, pp. 236–242.
9. R. Riaz, G. Han, K. Shaukat, N. U. Khan, and H. Zhu, "A robust anomaly detector for imbalanced industrial internet of things data," *Journal of Computational Design and Engineering*, vol. 12, no. 9, pp. 46–60, 2025.
10. A. Singh, S. Singh, M. N. Alam, and G. Singh, "Deep Learning for Anomaly Detection in IoT Systems: Techniques, Applications, and Future Directions."

11. W. Hu and Y. Cai, "A semi-supervised method for digital twin-enabled predictive maintenance in the building industry," *Neural Computing and Applications*, vol. 36, no. 25, pp. 15759–15775, 2024.
12. X. Wang, S. Garg, H. Lin, J. Hu, G. Kaddoum, M. J. Piran, and M. S. Hossain, "Toward accurate anomaly detection in industrial internet of things using hierarchical federated learning," *IEEE Internet of Things Journal*, vol. 9, no. 10, pp. 7110–7119, 2021.
13. Y. Wu, H. N. Dai, and H. Tang, "Graph neural networks for anomaly detection in industrial Internet of Things," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9214–9231, 2021.
14. A. De Benedictis, F. Flammini, N. Mazzocca, A. Somma, and F. Vitale, "Digital twins for anomaly detection in the industrial internet of things: Conceptual architecture and proof-of-concept," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 12, pp. 11553–11563, 2023.
15. G. Q. Zeng, Y. W. Yang, K. D. Lu, G. G. Geng, and J. Weng, "Evolutionary adversarial autoencoder for unsupervised anomaly detection of industrial internet of things," *IEEE Transactions on Reliability*, vol. 74, no. 3, pp. 3454–3468, 2025.