


Imbalance-Sensitive Broad Learning System for Industrial Multivariate Time Series Anomaly Detection



Shenglong Tan 

Department of Engineering, International College, Krirk University, Bangkok 10220, Thailand

Corresponding Author Email: 67130156@krirk.ac.th

Copyright: ©2026 The author. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ijss.160308>

ABSTRACT

Received: 7 February 2026

Revised: 19 March 2026

Accepted: 24 March 2026

Available online: 31 March 2026

Keywords:

anomaly detection, broad learning system, class imbalance, industrial control system, multivariate time series

The operational reliability of modern industrial control systems (ICS) is vital to critical infrastructure security. However, in complex industrial systems deeply integrating the Industrial Internet of Things (IIoT), multivariate time-series data exhibits extreme complexity. Physical equipment sampling inevitably contains dynamic environmental noise. Moreover, industrial systems operate normally most of the time, making physical faults or cyberattacks extremely rare, which leads to extreme class imbalance. This combination of intense noise and extreme imbalance often traps traditional deep learning anomaly detection methods in a "majority class trap", causing severe false negatives and frequent false alarms. To address these dual challenges, we propose a robust anomaly detection method combining task-oriented feature reconstruction and an imbalance-sensitive broad learning system (IS-BLS). First, a sliding window mechanism extracts statistical, kinetic, and local trend features, reducing noisy high-frequency dynamic series into robust static feature vectors. Second, a large margin slack factor and a position hyperparameter are introduced to construct an adaptive asymmetric fuzzy membership allocation mechanism, seamlessly embedded into the weighted closed-form solution of the BLS. Extensive evaluations on the real-world secure water treatment (SWaT) dataset demonstrate the method's superiority. Compared to conventional approaches, IS-BLS significantly boosts the anomaly Recall to 0.8443 while maintaining an exceptionally high Precision of 0.9857, achieving an outstanding comprehensive F1-score of 0.9087. Furthermore, the optimal network configuration requires only 6.79 seconds per iteration. These quantitative results confirm that IS-BLS successfully achieves a Pareto optimal balance between high detection accuracy and real-time industrial computational efficiency.

1. INTRODUCTION

With the rapid development of Industry 4.0 and the Industrial Internet of Things (IIoT), the complexity of modern industrial systems spanning critical sectors, such as aerospace, energy networks, and smart manufacturing has become increasingly prominent [1-3]. To ensure the security and stability of these core infrastructures, anomaly detection technology driven by multivariate time series data from sensors has emerged as the primary method for equipment health management and intrusion detection. Researchers have established that data-driven anomaly detection based on multivariate time-series data is an effective strategy for equipment health prognostics and intrusion identification [4, 5]. Artificial intelligence algorithms can continuously analyze large-scale sensor data, such as pressure and temperature readings, together with actuator information, to provide early warnings before severe physical or cyber events occur [6]. The utility of machine learning in security contexts has also been supported by recent empirical studies on public video surveillance, advanced network defense mechanisms, and spatiotemporal tracking systems [7-9].

However, deploying data-driven anomaly detection in real-world industrial environments faces two major practical bottlenecks. The first is extreme class imbalance. Because industrial systems usually operate in safe and stable states under strict protocols, genuine physical faults or cyberattacks are rare. This highly skewed data distribution often causes traditional global-error-minimizing classifiers, such as SVMs and MLPs, to become biased toward the majority normal class [10]. As shown in evaluations of representative anomaly detection frameworks, models that pursue high overall accuracy may sacrifice sensitivity to minority anomaly classes, leading to missed detections, namely false negatives, for highly destructive events [11, 12].

The second challenge is high-frequency dynamic noise interference. High-frequency sampling of physical equipment often captures environmental noise, which may reduce the stability of anomaly detection models [13, 14]. If a model is overly sensitive to such interference, it may generate frequent false alarms and cause alarm fatigue among field operators [15]. Existing deep neural networks, such as long short term memory (LSTM) and convolutional neural network (CNN), also struggle with these coupled challenges. Algorithms

relying on repeated backpropagation iterations are prone to overfitting environmental noise [16, 17]. Moreover, traditional data-level oversampling methods interpolate directly on temporal waveforms [18, 19]. Because industrial time-series data are constrained by physical dependencies, such direct interpolation may violate inertia-related characteristics and generate unrealistic synthetic samples.

Recently, broad learning systems (BLS), including weighted BLS, fuzzy BLS, and other imbalance-oriented variants, have demonstrated notable effectiveness in addressing class imbalance in generic static classification tasks. However, their investigation and practical deployment in industrial multivariate time-series anomaly detection remain limited. A key challenge lies in the intrinsic characteristics of industrial time-series data, which are typically affected by high-frequency dynamic environmental noise generated by continuously operating physical equipment, as well as severe data imbalance due to the rarity of real physical faults or cyberattacks. Although existing imbalance-oriented BLS variants perform well in static settings, their direct application to such dynamic industrial environments often leads to biased decision boundaries and elevated false alarm rates.

To address these challenges, this paper proposes an application-oriented imbalance-sensitive broad learning system (IS-BLS) tailored for industrial control systems (ICS). The proposed method is a unified, data-driven anomaly detection framework that integrates data reconstruction with imbalance-sensitive discrimination. The main contributions are summarized as follows:

(1) A task-oriented sliding window feature reconstruction mechanism specifically designed for industrial multivariate time series. To address the dimensionality curse and noise amplification caused by directly inputting high-frequency raw waveforms, a 10-dimensional structured feature extraction flow comprising statistical distribution, first-order kinetics, and local trends is designed. This mechanism successfully flattens dynamic temporal fluctuations into a solid, noise-resistant feature representation, laying a crucial data foundation for subsequent anomaly discrimination.

(2) An IS-BLS tailored for industrial anomaly detection. While existing imbalance-oriented BLS variants often rely on static weighting for generic imbalanced data, the core application novelty of our method lies in designing an asymmetric fuzzy membership function specifically to address the complex coupling of intense noise and extreme imbalance in ICS. Inheriting the backpropagation-free ultra-fast analytical advantage of BLS, combined with large margin theory, a slack factor and a position tolerance hyperparameter are integrated. This specific construction assigns exponentially amplified penalty weights to extremely scarce subtle cyber-physical anomalies while applying a tolerant decay to majority-class baseline noise, effectively breaking the "majority class trap".

(3) An integrated closed-loop iterative analytical solution with comprehensive industrial scenario validation. By embedding the designed asymmetric fuzzy membership directly into the objective function, a strict matrix analytical solution based on weighted least squares is derived, and an alternating iterative inverse closed-loop is constructed to ameliorate the tendency of traditional algorithms to fall into local optima. Extensive experiments on a real-world large-scale industrial control security dataset (secure water treatment, SWaT) confirm that this combined architecture significantly outperforms existing mainstream algorithms in core metrics

(Recall, F1-score, and AUC-PR). It effectively mitigates the contradiction between high missed detections and high false alarms.

2. RELATED WORK

2.1 Industrial time series anomaly detection

In industrial time-series anomaly detection, early studies mainly relied on physics-based models [20]. However, with the growth of sensor networks and high-dimensional multivariate data, traditional methods have become increasingly insufficient [21]. Recent studies show that data-driven approaches have gradually become mainstream [22, 23]. These methods usually construct a baseline from historical normal data and identify anomalies when real-time monitoring data deviates from this baseline [24, 25].

However, in practical applications, most collected data represent normal operating conditions, while data related to faults or cyberattacks are relatively scarce. This leads to a serious class imbalance problem [26, 27], causing models to focus excessively on normal patterns during training and reducing their sensitivity to critical minority faults [28]. In addition, dynamic noise caused by electromagnetic interference, sensor aging, and other factors further complicates anomaly detection [29, 30]. Such noise may mask weak anomalous signals and cause missed detections, or be incorrectly identified as anomalies and trigger false alarms [31]. Therefore, developing a robust anomaly detection model with high detection accuracy and a low false alarm rate remains a major challenge.

2.2 Broad learning system

To address the computational bottlenecks of traditional DNN when processing massive industrial time series, the BLS [32], is a significant expansion of the traditional Random Vector Functional-Link (RVFL) neural network in terms of topological structure and solution paradigm. Its core idea is to abandon the deeply stacked hierarchical structure in DNN and instead adopt a "horizontally expanded" flat topology. Specifically, BLS first maps the original input data into feature nodes through randomly initialized weights, and subsequently performs nonlinear transformations on the feature nodes to generate enhancement nodes. These two groups of nodes are concatenated to form a high-dimensional generalized state matrix. During the parameter solution phase, BLS employs a pseudo-inverse algorithm based on Ridge Regression to directly analytically determine the optimal output layer weights. Compared to deep learning methods relying on backpropagation (BP), the BLS exhibits significant theoretical advantages: first, due to its single-layer structure and closed-form algebraic solution, it completely avoids gradient vanishing and local optima problems in multi-layer networks; second, the matrix inversion process demonstrates extremely high computational efficiency when handling large-scale industrial data, and naturally supports rapid online incremental learning of network structures (new nodes) or data samples [33, 34].

To intuitively illustrate the structural and computational differences, Figure 1 presents a comparison of the topology and solution mechanism between traditional DNN and the proposed BLS framework. As depicted in Figure 1(a), the

traditional DNN heavily relies on a deeply stacked hierarchical architecture, where parameter updating necessitates time-consuming error gradient backpropagation, making it highly susceptible to vanishing gradients. Conversely, as shown in Figure 1(b), the BLS adopts a horizontally expanded flat architecture. By concatenating the feature nodes group and the enhancement nodes group, the model calculates the output weights directly through a one-step analytical pseudo-inverse solution. This fundamental structural innovation effectively bypasses the gradient descent dilemmas and significantly accelerates the training process.

However, the standard broad learning model still exhibits obvious limitations in real industrial scenarios. Its objective function defaults to unweighted Ordinary Least Squares (OLS) for uniform optimization. When confronted with severely imbalanced industrial data, the global mean square error is easily dominated or "hijacked" by the massive majority class (e.g., normal operating state). To minimize the overall error, the model tends to sacrifice the fitting accuracy of the minority class (e.g., weak anomalies), causing the classification hyperplane to shift and severely compress towards the minority class space, leading to anomalous samples being incorrectly enclosed within the majority class boundary [35, 36].

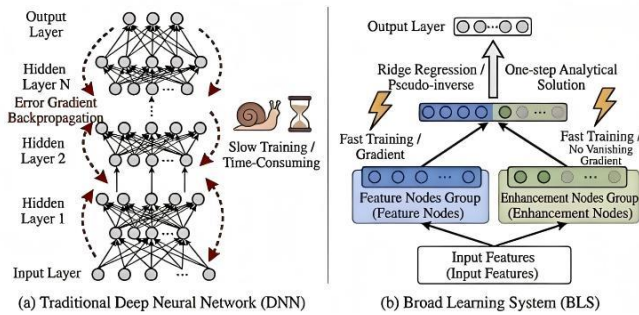


Figure 1. Comparison of topology and solution mechanism between traditional deep neural network (DNN) and broad learning system (BLS)

2.3 Class imbalance learning strategies

To resolve the aforementioned "majority class hijacking" vulnerability inherent in standard BLS and other global classifiers, current processing strategies for class imbalance are mainly divided into data-level and algorithm-level systems. Data-level methods, represented by the Synthetic Minority Over-sampling Technique (SMOTE) and its evolutionary variants (e.g., ADASYN), generate synthetic samples by linearly interpolating neighboring minority samples in the feature space to forcefully balance the data categories [37]. However, for multivariate time series generated by industrial systems, such data—which possesses strong physical constraints and spatiotemporal coupling characteristics—is particularly sensitive to interpolation. Simple linear interpolation easily destroys original temporal dependencies and physical causality logic, and is highly prone to generating numerous "fake samples" and extra noise that deviate from physical reality in the sparse regions of minority decision boundaries, subsequently leading to a significant increase in false alarm rates during industrial deployment.

Therefore, algorithm-level strategies have proven to be a more feasible path when handling rigorous industrial time-series data [38, 39]. Algorithm-level methods do not alter the

physical topology of the original dataset but rather implement interventions by reconstructing the optimization objective function. Among them, cost-sensitive learning balances the loss by assigning a higher preset misclassification cost to the minority class. Fuzzy weighting learning goes a step further by dynamically assigning continuous membership weights to each sample based on its distribution density or classification difficulty (e.g., margin from the hyperplane). This soft adjustment mechanism based on fuzzy membership can guide the model to autonomously strip away baseline noise and strengthen focus on difficult samples and minority anomaly boundaries, thereby significantly enhancing discriminative sensitivity to anomalous states while preserving the genuine temporal coupling of the data [40].

Figure 2 provides a conceptual diagram illustrating the decision boundary shifting and reshaping process under extreme class imbalance. As shown, the standard OLS boundary (dashed line) is severely compressed towards the minority class (anomaly) area due to the overwhelming numerical superiority of the majority class (normal) samples, inherently leading to a high probability of missed detections. In contrast, the proposed IS-BLS constructs a weighted boundary (solid line) fortified by a fuzzy membership-based "resistance ring," which assigns exponentially amplified penalty weights to misjudged minority anomalies. This targeted mechanism successfully corrects the boundary offset, ensuring a robust and clear separation between the rare anomalies and the massive normal baseline.

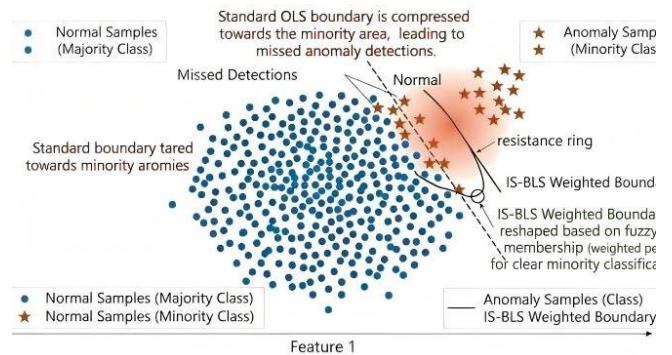


Figure 2. Class imbalance decision boundary shifting and reshaping conceptual diagram

2.4 Section summary

In summary, while existing studies have independently explored time-series feature extraction and imbalanced learning, directly applying existing BLS-based variants to industrial anomaly detection is sub-optimal. Conventional imbalance-oriented BLS models are designed for static data distributions; when exposed to raw, high-frequency industrial time series, dynamic environmental noise can easily disrupt their fuzzy memberships or distance-based weights, leading to severe performance degradation. Therefore, this study adopts a specific architectural combination: sliding-window feature reconstruction coupled with an IS-BLS. The sliding-window mechanism serves as a necessary physical shield, flattening dynamic waveforms into a stable, noise-resistant macroscopic feature space. Operating exclusively on this stabilized space, the IS-BLS can then fully leverage its large-margin asymmetric membership to severely penalize extreme minority anomalies without the computational burden of backpropagation. This combination explicitly decouples the

dual challenges of intense noise and extreme class imbalance in real-world ICS environments.

3. METHODOLOGY

This section details the core mathematical principles of the proposed anomaly detection method. The overall logic consists of a sliding window-based preprocessing procedure and an imbalance-aware efficient discriminative mechanism (IS-BLS).

To intuitively illustrate the holistic pipeline of our approach, Figure 3 presents the overall architecture of the proposed IS-BLS framework. As depicted in the figure, the raw high-

frequency multivariate time-series data initially undergoes a task-oriented sliding window feature reconstruction module. This module systematically extracts statistical, kinetic, and local trend representations to form a noise-resistant static feature vector. Subsequently, these reconstructed features are fed into the IS-BLS discriminative module. Within this mechanism, the inputs are mapped into broad feature and enhancement nodes, while an asymmetric fuzzy membership function dynamically assigns penalty weights based on the large margin slack factor. Finally, the system outputs robust anomaly detection labels via a weighted closed-loop analytical solution, which effectively filters out baseline noise and reshapes the decision boundary for the minority class anomalies.

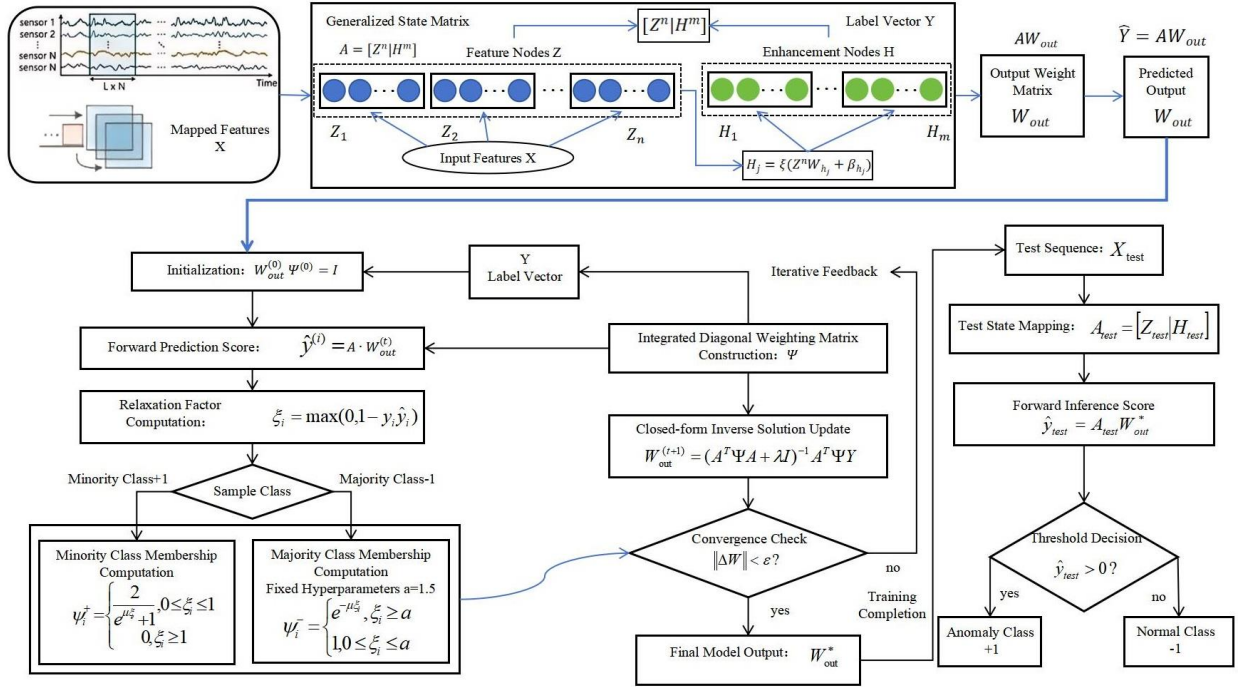


Figure 3. Overall architecture of the proposed imbalance-sensitive broad learning system (IS-BLS) framework for industrial anomaly detection

3.1 Task-oriented sliding window feature reconstruction

In ICS, multivariate time-series data collected by sensors is not only high-dimensional but also typically accompanied by high-frequency dynamic environmental noise. If the raw high-dimensional time-series tensor is directly input into a classifier for analysis, it not only substantially increases computational complexity but also allows high-frequency noise fluctuations to severely interfere with the solution of the classification hyperplane. Therefore, this paper proposes a task-oriented sliding window feature reconstruction mechanism, aiming to extract a noise-resistant static feature representation through dimensionality reduction strategies.

Assume the system contains N sensor channels, and its collected multivariate time-series data can be represented by tensor $X_{raw} \in \mathbb{R}^{T \times N}$, where T is the total time steps. This paper sets the fixed length of the sliding window as L and the sliding step as S . For any given observation time t , the system intercepts a local multivariate time window data subset along the time axis, denoted as $W_t \in \mathbb{R}^{L \times N}$. To extract stable physical state information from dynamic time-series fluctuations, for the local time-series fragment $w_n = \{x_1, x_2, \dots, x_L\}$ of the n -th

physical channel in W_t , this paper systematically extracts 10 key structured features from three dimensions: statistical distribution, first-order kinetics, and local trends.

First, statistical distribution features are primarily used to evaluate the physical baseline state and its dispersion during that period; such features naturally smooth transient impulse noise. Six statistics are specifically extracted: the mean (μ) reflects the DC component and stable operating point of the signal, defined as shown in Eq. (1). The standard deviation (σ) measures the severity of signal fluctuations, calculated in Eq. (2).

$$\mu = \frac{1}{L} \sum_{i=1}^L x_i \quad (1)$$

$$\sigma = \sqrt{\frac{1}{L} \sum_{i=1}^L (x_i - \mu)^2} \quad (2)$$

Additionally, the maximum (x_{\max}) and minimum (x_{\min}) values of the sequence are extracted to capture local physical extreme anomalies, and the range (R) is used to reflect the amplitude span of fluctuations, defined as $R = x_{\max} - x_{\min}$. Meanwhile, to obtain a statistical center more robust than the

mean, the median of the sequence is extracted.

Second, to capture transients in equipment states, first-order kinetic features are introduced. By performing a first-order forward difference operation $\Delta x_i = x_{i+1} - x_i$ on the original sequence, the mean (μ_Δ) and standard deviation (σ_Δ) of the difference sequence are extracted. These two features can acutely capture abnormal high-frequency jitter states caused by cyberattacks involving sensor numerical step tampering or mechanical faults.

Finally, considering the continuity of industrial system state evolution, local trend and dependency features are extracted. To quantify the overall evolution direction of data within the window, the linear slope (k) of the sequence is fitted based on the least squares method. Setting the mean of the time index as $\bar{i} = (L+1)/2$, the calculation of the linear slope is shown in Eq. (3).

$$k = \frac{\sum_{i=1}^L (i-\bar{i})(x_i-\mu)}{\sum_{i=1}^L (i-\bar{i})^2} \quad (3)$$

Simultaneously, to measure the internal inertial dependency degree of the sequence's current state on its previous state, the first-order autocorrelation coefficient (ACF) is extracted, defined in Eq. (4).

$$ACF = \frac{\sum_{i=1}^{L-1} (x_i-\mu)(x_{i+1}-\mu)}{\sum_{i=1}^L (x_i-\mu)^2} \quad (4)$$

Through the reconstruction across these three dimensions, the redundant and noisy high-frequency time-series fluctuations in a single sensor channel are effectively transformed into 10 compact static feature indicators. Ultimately, the entire multivariate time window W_t is reconstructed into a feature matrix of dimension $N \times 10$, which, after flattening, yields the feature vector $X_{feat} \in \mathbb{R}^{1 \times D}$ (where $D=10N$), providing a solid and noise-resistant input foundation for the subsequent anomaly discrimination model.

From a physical monitoring perspective, statistical (μ, σ) and trend (k, ACF) indicators are particularly effective in capturing sustained state deviations—such as continuous pressure drops caused by a stuck physical valve—or slow-rate stealthy degradation, including gradual sensor drift. Concurrently, kinetic difference features $\mu_\Delta, \sigma_\Delta$ are highly responsive to sudden step-change spoofing, rapidly isolating the anomalous high-frequency jitter induced by discrete cyber interventions.

However, the inherent temporal aggregation characteristics within a specific time window may also introduce detection blind spots. For instance, ultra-short transient pulse attacks propagate rapidly without altering the macroscopic physical properties of the system, and are therefore prone to being inadvertently smoothed out during statistical averaging processes. Furthermore, these explicitly defined statistical indicators and first-order dynamic parameters may be insufficient to fully capture the highly complex and nonlinear coupling correlations among multiple sensors.

3.2 Robust imbalance-sensitive discriminative mechanism (IS-BLS)

3.2.1 Generalized state matrix construction and objective function definition

The IS-BLS model constructed in this paper adopts a flat

topology of "feature mapping \rightarrow enhancement nodes \rightarrow generalized output". Let the reconstructed training set feature matrix be $X \in \mathbb{R}^{M \times D}$ (where M is the number of samples), and the label be $Y \in \mathbb{R}^{M \times C}$.

First, the input features X pass through n groups of random weights W_{e_i} and biases β_{e_i} , and are mapped into feature nodes via activation function $\phi(\cdot)$:

$$Z_i = \phi(XW_{e_i} + \beta_{e_i}), i=1,2,\dots,n \quad (5)$$

Concatenating them yields the global feature mapping set $Z^n = [Z_1, Z_2, \dots, Z_n]$.

Next, taking Z^n as input, m groups of enhancement nodes are mapped through random weights W_{h_j} and activation function $\zeta(\cdot)$:

$$H_j = \zeta(Z^n W_{h_j} + \beta_{h_j}), j=1,2,\dots,m \quad (6)$$

Concatenating them yields the enhancement node matrix $H^m = [H_1, H_2, \dots, H_m]$.

Combining both constitutes the system's generalized state matrix $A = [Z^n | H^m]$.

To implement targeted imbalance interventions, we introduce an adaptive fuzzy error weighted diagonal matrix $\Psi = \text{diag}(\psi_1, \psi_2, \dots, \psi_M)$ to reconstruct the objective function into a weighted least squares form:

$$J(W_{out}) = \|\Psi^{1/2}(AW_{out} - Y)\|_2^2 + \lambda \|W_{out}\|_2^2 \quad (7)$$

3.2.2 Slack factor and fuzzy membership based on large margin theory

To quantify the "classification difficulty" of each sample, a large margin slack factor ξ_i is introduced:

$$\xi_i = \max(0, 1 - y_i \cdot \hat{y}_i) \quad (8)$$

where, $\hat{y}_i = A_i W_{out}$ is the continuous predicted value of the model for the i -th sample. Based on this, an asymmetric fuzzy membership function ψ_i is designed:

Fuzzy membership for the minority class (anomaly):

For scarce minority class samples, if a misjudgment tendency appears, an exponentially amplified penalty is applied. When $\xi_i \geq 1$ (indicating a significant prediction deviation occurs), the model treats it as anomalous noise caused by severely damaged sensors and applies truncation set to 0, to reduce its excessive interference with global hyperplane construction.

$$\psi_i^+ = \begin{cases} \frac{2}{e^{\beta \xi_i + 1}}, & 0 \leq \xi_i < 1 \\ 0, & \xi_i \geq 1 \end{cases} \quad (9)$$

Position-aware fuzzy membership for the majority class (normal):

Normal samples are massive in number and often accompanied by equipment baseline white noise. A fixed position tolerance hyperparameter a (empirically set to 1.5 in this paper) is introduced as a delay threshold for decay. When $\xi_i < a$, the weight remains 1; once the threshold is breached, it is deemed as noise interfering with the decision boundary, and the weight decays exponentially. This is equivalent to establishing a smooth decay barrier outside the majority class

boundary.

$$\psi_i = \begin{cases} e^{-\beta \xi_i}, & \xi_i \geq a \\ 1, & 0 \leq \xi_i < a \end{cases} \quad (10)$$

3.2.3 Partial derivatives of the weighted objective function and closed-loop optimization

Taking the partial derivative of the scalar $J(W_{out})$ strictly with respect to the matrix W_{out} and setting the derivative to a zero matrix, the global analytical solution equation can be derived:

$$W_{out} = (A^T \Psi A + \lambda D)^{-1} A^T \Psi Y \quad (11)$$

Since a circular dependency exists in the analytical solution—the weighted matrix Ψ depends on ξ_i , ξ_i depends on the predicted output, and the predicted output is determined by W_{out} —an alternating iterative closed-loop algorithm is constructed to solve this problem effectively:

- Step 1: Construct generalized matrix A , initialize weight matrix $\Psi^{(0)} = I$, and calculate the first-round baseline hyperplane $W_{out}^{(0)} = (A^T A + \lambda D)^{-1} A^T Y$.
- Step 2: Set the maximum number of iterations MaxIter , and start the iteration loop (current iteration count is denoted as $t = 0$).
- Step 3: Forward infer the predicted values of the current model, and calculate the slack factor $\xi_i^{(t)}$ for each sample.
- Step 4: Based on the class to which the sample belongs, utilize the corresponding membership formula to update the diagonal weighted matrix $\Psi^{(t+1)}$.
- Step 5: Utilize the weighted pseudo-inverse method to update the output layer weight matrix $W_{out}^{(t+1)} = (A^T \Psi^{(t+1)} A + \lambda D)^{-1} A^T \Psi^{(t+1)} Y$.
- Step 6: Calculate the weight residual $\|\Delta W\| < \epsilon$ between two adjacent iterations; if the convergence condition is met, end the training, otherwise set $t = t + 1$ and return to Step 3 to execute the loop.

This alternating iterative mechanism effectively mitigates the sensitivity of traditional models to the initial weight distribution when processing imbalanced data.

3.3 Efficient inference and dynamic decision mechanism for anomalous states

After completing the closed-loop training of IS-BLS and obtaining the optimal output weight matrix W_{out}^* , the system can enter the inference and anomaly determination stage for real-time industrial data streams. For any newly collected test feature vector X_{test} , same-dimensional mapping is performed through frozen hidden layer basis mapping parameters (W_e, β_e) and (W_h, β_h) to construct the test set generalized state matrix $A_{test} = [Z_{test}^n | H_{test}^m]$. Subsequently, forward inference is executed to obtain the continuous model prediction score vector \hat{Y}_{test} :

$$\hat{Y}_{test} = A_{test} W_{out}^* \quad (12)$$

After obtaining the continuous prediction score, it needs to be mapped to a discrete physical equipment state label. Because this paper has deeply penalized and dynamically compensated for class imbalance through slack factors and asymmetric fuzzy memberships during the training phase, the classification hyperplane has been effectively reshaped and

pushed back to a reasonable decision space. Therefore, in the inference phase, there is no need to rely on complex dynamic threshold searching or post-processing methods; a standard separator can be directly adopted to determine the final decision boundary.

Specifically, for a binary classification industrial scenario using $\{-1, +1\}$ encoding (representing normal and anomalous states respectively), the standard sign function $\text{sgn}(\cdot)$ is directly utilized for discrete mapping:

$$\text{Label}_{test} = \begin{cases} +1 & (\text{Anomaly}), \hat{Y}_{test} \geq 0 \\ -1 & (\text{Normal}), \hat{Y}_{test} < 0 \end{cases} \quad (13)$$

Directly adopting this standard separator not only simplifies the inference process but further highlights the inherent effectiveness of the IS-BLS algorithm in reshaping the hyperplane in the feature space. Due to the flat structure of the BLS, the entire real-time inference process involves only one forward feature mapping, one pure linear matrix multiplication, and one standard sign determination, yielding overall low computational complexity. This characteristic effectively fulfills the stringent timeliness requirements for real-time early warnings in ICS.

4. EXPERIMENTS AND ANALYSIS

4.1 Dataset description and model initialization parameter settings

This study adopts the SWaT dataset, a recognized benchmark dataset in the field of industrial control security. This dataset is generated by a modern water treatment testbed encompassing 51 multi-dimensional sensors and actuators. The data records 7 days of normal continuous operating states and 4 days of anomalous states subjected to specific cyberattacks (36 attack scenarios). The anomalous samples account for a very small proportion (less than 12%), presenting a significant class imbalance characteristic, making it highly suitable as the industrial validation object for this study.

Strictly Isolated Validation Scheme (Preventing Data Leakage): To rigorously evaluate the proposed IS-BLS model and strictly prevent the temporal data leakage commonly associated with overlapping sliding windows, we implemented a robust validation scheme defined as Chronological Hold-out Testing with Internal Cross-Validation. Specifically, the data partitioning was executed strictly at the raw temporal sequence level rather than at the randomized window level. The continuous industrial multivariate time series was split based on absolute chronological order: the initial 80% of the temporal sequence was designated as the global training and validation pool, while the subsequent 20% of the sequence was strictly reserved as the unseen hold-out test set. Because the split point is applied to the raw timeline prior to window generation, there is absolutely zero physical overlap or information leakage between the training windows and the testing windows.

Model initial parameter suggestions and settings: To ensure model convergence efficiency and prevent overfitting, combining validation set grid search and the asymmetric topology principles of the BLS, this paper sets the following recommended initial parameters for the feature dimensions based on the SWaT dataset ($D = 51 \times 10 = 510$ dimensions):

Network node scale: In practical applications, to fully extract diversified representations of high-dimensional input data while controlling the redundancy of the enhancement space, the number of feature mapping nodes should generally be set larger than the number of enhancement nodes. Therefore, the number of feature mapping node groups is set to $m_{\text{ap}} = 40$, the enhancement node groups to $e_{\text{henc}} = 20$, and the batch size per group is 20. Under this configuration, the feature nodes and enhancement nodes are 800 and 400 respectively (forming a generalized state matrix of 1200 dimensions), achieving a good balance between model capacity and computational overhead.

Imbalance weighting parameters: The majority class position tolerance hyperparameter α is set to 1.5, and the fuzzy membership decay coefficient β is set to 0.5.

Training optimization parameters: The regularization coefficient λ is set to 0.001, the maximum number of closed-loop iterations MaxIter is set to 20, and the weight convergence residual threshold ϵ is set to 1×10^{-3} .

To eliminate data dimension impacts, input data is first standardized using Z-score. As emphasized in our strictly isolated validation scheme, the Stratified 5-Fold Cross Validation (CV) was adopted exclusively within the boundaries of the 80% training pool for robust internal hyperparameter optimization. The final comprehensive evaluation metrics reported in this study were derived entirely from the isolated 20% future hold-out test set.

4.2 Comparative experiments and result analysis

To verify the comprehensive discriminative capability of the proposed IS-BLS algorithm, representative advanced baseline algorithms were selected for comparison, including deep autoencoder models (LSTM-AE [41], DAGMM, USAD) and graph neural network-based models (GCN, GAT [42], VGAE [43], GAE-AD [44], GDN).

To ensure a strictly transparent and fair comparison, all baseline models were evaluated under equivalent experimental conditions. First, regarding input processing, all models ingested data from the exact same chronological split (80% training, 20% unseen testing) and utilized the identical temporal context (window length "L=60") after Z-score standardization. Second, regarding the supervision level and

thresholding, since the deep autoencoder and graph-based baselines are primarily unsupervised reconstruction or forecasting models, they were trained exclusively on the normal samples within the shared training pool.

Hardware and software environment: To ensure computational efficiency and a fair comparison, all experiments were conducted on a 64-bit Windows platform equipped with an Intel Core i7-14650HX CPU (2.20 GHz), 32 GB RAM, and an NVIDIA GeForce RTX 4070 GPU (8 GB VRAM). The algorithms were implemented in Python 3.9, utilizing PyTorch for deep learning baselines and NumPy/SciPy for the matrix analytical derivations of IS-BLS.

This study utilizes four core indicators to comprehensively evaluate the model's anomaly detection performance: Precision, Recall, AUC-PR, and F1-score. The experimental result comparison on the SWaT dataset is summarized in Table 1.

An in-depth analysis of the data in Table 1 reveals that because the SWaT dataset has severe class imbalance, traditional deep reconstruction models and graph learning models perform excellently in Precision by fitting the majority class, but generally perform poorly in the Recall metric (mainly hovering between 0.60 and 0.68). This indicates that the aforementioned comparison models are severely constrained by the "majority class trap", leading to over 30% of actual industrial anomalies being treated as normal data and missed, which is unacceptable in actual industrial security scenarios.

Table 1. Performance comparison of different methods on secure water treatment (SWaT) dataset

Algorithm Model	Precision	Recall	AUC-PR	F1-Score
LSTM-AE [41]	0.9624	0.5991	0.5546	0.7405
GCN [6]	0.9603	0.6775	0.8364	0.7941
GAE-AD [44]	0.9582	0.6768	0.8363	0.7934
GDN [6]	0.9561	0.6645	0.8362	0.7927
USAD [12]	0.9851	0.6771	0.8385	0.8074
DAGMM [14]	0.9866	0.6879	0.8433	0.8106
VGAE [43]	0.9955	0.6878	0.8437	0.8135
GAT [42]	0.9977	0.6878	0.8438	0.8143
IS-BLS (Ours)	0.9857	0.8443	0.9207	0.9087

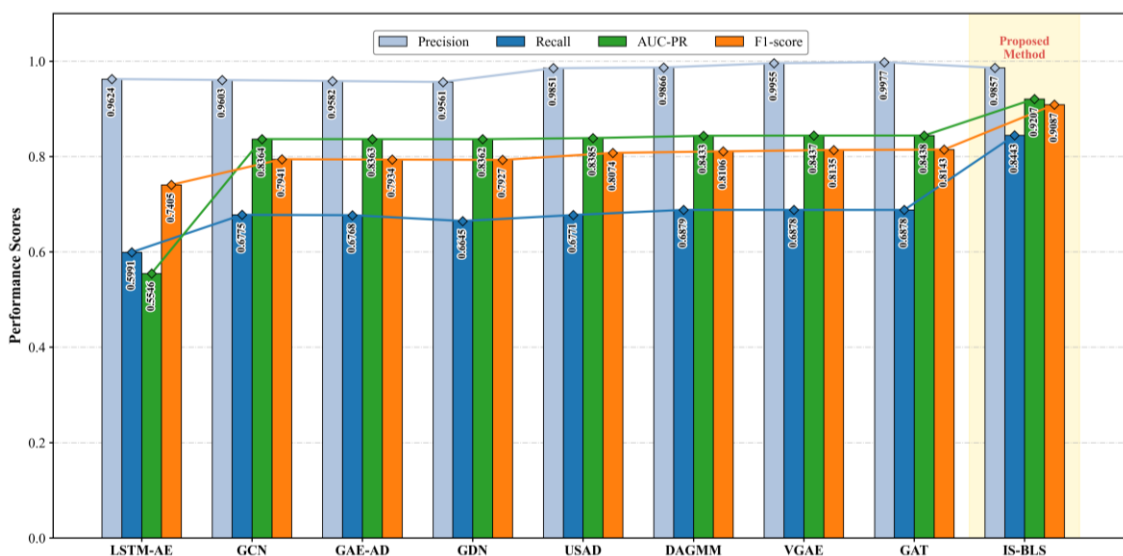


Figure 4. Comprehensive performance comparison of different methods on secure water treatment (SWaT)

In contrast, the IS-BLS detection algorithm proposed in this paper implements targeted compensation for the minority class by introducing a slack factor and fuzzy membership weighting mechanism. Under the premise of maintaining high precision (0.9857), the anomaly recall is boosted to 0.8443, and the AUC-PR and F1-score metrics reach 0.9207 and 0.9087 respectively, outperforming all comparative baseline models. This demonstrates that the method effectively reshapes the decision boundary, well balancing the core industrial demands of "low missed detection" and "low false alarms".

To visually demonstrate this performance gap, Figure 4 displays the comprehensive performance comparison of different methods on the SWaT dataset. The grouped bar chart intuitively reveals that while most deep reconstruction and graph learning models maintain high precision, their recall bars drop significantly, reflecting their vulnerability to the majority class trap. Conversely, the IS-BLS method achieves a highly balanced and superior performance across all metrics, particularly highlighting its distinct advantage in recall and overall F1-score.

The significant enhancement in Recall is fundamentally driven by the synergy of three coupled mechanisms within the IS-BLS architecture. Specifically, the window-based feature reconstruction smooths out high-frequency environmental noise. By flattening transient signal fluctuations, it secures a highly stable normal baseline, which is crucial for maintaining high Precision. Concurrently, the minority-class penalization restructures the optimization objective. Rather than being dominated by the overwhelming volume of normal data, the network assigns exponentially higher costs to undetected anomalies. Most critically, the asymmetric fuzzy membership drives a direct adjustment of the decision boundary. By deliberately relaxing the margin around the majority class, the classification hyperplane is pushed closer to the normal data space. This targeted boundary reshaping ensures that marginal cyber-physical attacks, which typically disguise themselves near the normal baseline, are effectively captured. This direct boundary adjustment serves as the primary catalyst for the substantial surge in Recall, effectively mitigating the missed-detection bottleneck in industrial anomaly detection.

4.3 Analysis of feature reconstruction effectiveness and label mechanisms

4.3.1 Validation of time window feature mapping effectiveness

To verify the necessity of the "task-oriented sliding window feature reconstruction" foundation, an ablation experiment was designed. We compared the performance difference between directly inputting high-frequency raw time-series waveforms into the model (No Windowing) and adopting the feature reconstruction mechanism (With Windowing, $L = 60$, step $S = 10$). The results are shown in Figure 5.

Figure 5 directly illustrates the magnitude of performance improvement achieved by the sliding window feature reconstruction in the ablation study. The data indicates that directly inputting raw high-frequency dynamic time series severely degrades the model's discriminative capability, performing particularly poorly in Recall (only 0.6533) with an overall F1-score of merely 0.7278. This deficiency is primarily attributed to the pervasive environmental white noise and sensor jitter in the raw data, which severely disrupt the analytical solving process. Conversely, after introducing the feature reconstruction mechanism, the model

systematically extracts 10-dimensional static statistical and kinetic features, effectively filtering out transient glitches and capturing macroscopic evolutionary patterns. This mechanism results in substantial improvements across all evaluative metrics: Precision increases from 0.8214 to 0.9857, and AUC-PR increases from 0.7650 to 0.9207. Most crucially, Recall experiences a remarkable surge of +29.2% (reaching 0.8443), which ultimately drives the overall F1-score up by +24.9% (reaching 0.9087). This significant data span compellingly verifies that the feature reconstruction step is an indispensable prerequisite for constructing a highly robust industrial anomaly detection baseline.

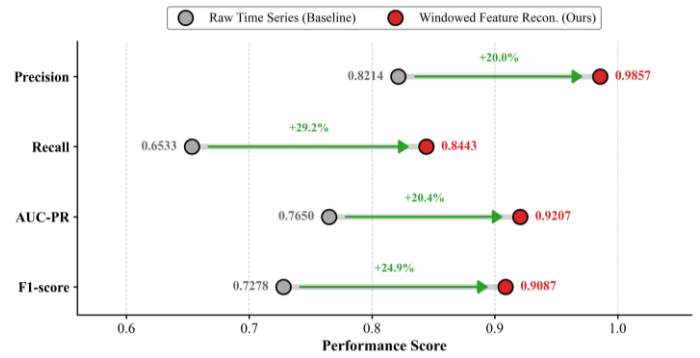


Figure 5. Performance gain via windowed feature reconstruction in the ablation study

4.3.2 Comparison of window anomaly label allocation mechanisms

When slicing continuous time series into windows, how to define the binary classification label (normal or anomalous) for the entire sliding window is a critical step. It is crucial to clarify that selecting a label allocation mechanism is not merely an algorithmic hyperparameter choice, but a fundamental definition of the anomaly detection task itself. This paper designed and compared two mainstream allocation mechanisms:

- Any Mechanism: As long as the window contains at least 1 anomalous sample point, the window is marked as an anomaly (minority class).
- Majority Mechanism: Only when over 50% (more than half) of the data points within the window are confirmed as anomalous is the entire window marked as an anomaly.

Figure 6 quantitatively demonstrate the multi-dimensional performance trade-offs when transitioning from the lenient 'Any' mechanism to the strict 'Majority' mechanism. The data reveals a distinct "give-and-take" characteristic. Specifically, adopting the 'Majority' mechanism incurs a marginal decline in theoretical Recall (decreasing from "0.8950" to "0.8443"). However, this is accompanied by a massive leap in Precision by "+15.8%" (surging from "0.8512" to "0.9857").

To cleanly interpret these metric changes, we must explicitly link them to the physical realities of ICS. In an actual ICS environment, a single or highly transient anomalous data point (e.g., lasting merely milliseconds) is highly likely to be environmental electromagnetic noise or a momentary sensor glitch rather than a destructive attack. By transitioning to the 'Majority' mechanism, we are effectively redefining the mathematical task to align with the true industrial definition of a sustained "System-level Event" or a continuous cyber-physical attack.

Therefore, the substantial improvement in Precision is not an arbitrary metric shift, but a direct result of this task redefinition. The 'Majority' mechanism successfully eliminates the contamination of massive normal baseline signals (and transient glitches) within the anomaly feature representations. By forcing the model to cease chasing transient noise and correctly focus on purifying genuine, sustained anomaly patterns, this mechanism drives a net positive gain of "+4.1%" in the comprehensive F1-score (rising from "0.8725" to "0.9087"). This firmly validates that redefining the task via strict anomaly purification is decisive for suppressing system false alarms and enhancing practical detection efficacy in complex industrial settings.

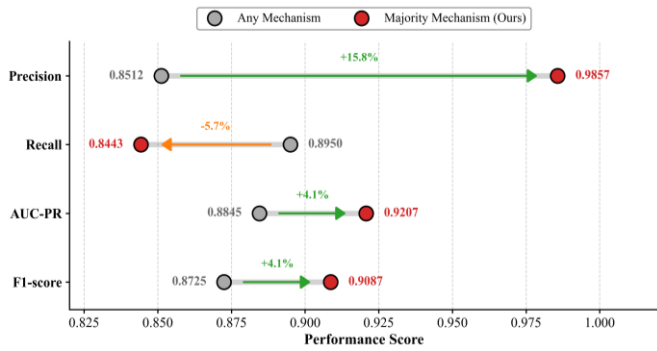


Figure 6. Performance trade-offs between "Any" and "Majority" window label allocation mechanisms

4.4 Core parameter sensitivity and computational efficiency analysis

4.4.1 Sliding window length (L) experimental analysis

The sliding window length L determines the temporal receptive field for feature extraction. To analyze its impact on performance, this experiment set the window length L to 10, 30, 60, 90, 120, 150 and 180 time steps respectively.

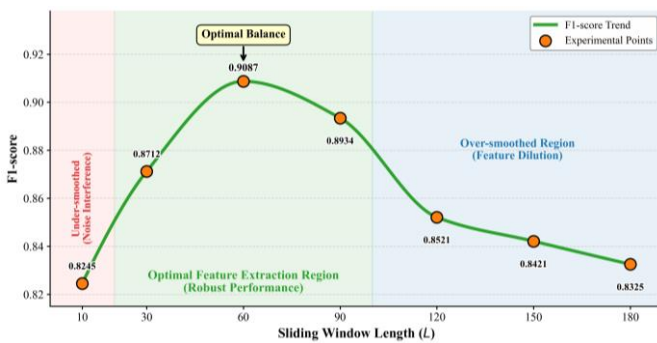


Figure 7. Sensitivity analysis of sliding window length with physical smoothing regimes

Figure 7 illustrates the sensitivity of the model's F1-score to the sliding window length L , systematically categorized into distinct physical smoothing regimes. The experimental trajectory manifests a prominent inverted-U non-linear curve. In the "Under-smoothed" region (e.g., $L = 10$), the insufficient sample capacity within a brief window fails to generate robust statistical and kinetic features. This renders the model highly susceptible to transient environmental noise interference, yielding a suboptimal F1-score of only 0.8245. As the temporal window expands, the model transitions into the "Optimal Feature Extraction Region," reaching its

performance zenith at $L = 60$ (F1-score = 0.9087). At this exact scale, a perfect equilibrium is struck between microscopic local information representation and macroscopic noise reduction. However, when the window is blindly extended deeper into the "Over-smoothed" region (e.g., as L scales from 120 all the way to 180), the excessively long temporal span forcibly incorporates massive volumes of normal baseline data. Consequently, short and abrupt anomaly features are severely diluted by this "over-averaging" effect, leading to a continuous, long-tail degradation in detection sensitivity (with the F1-score progressively dropping to 0.8325). This comprehensive data span firmly corroborates that configuring a moderate sliding window ($L = 60$) is the optimal strategic choice for balancing noise robustness and anomaly sensitivity.

4.4.2 Trade-off between network node scale and computational efficiency

The discriminative performance of the BLS is closely related to the network width. To find the optimal scale that guarantees recognition accuracy while meeting industrial computational timeliness requirements, this paper kept the batch size = 20 constant and gradually expanded the mapping groups and enhancement groups roughly at a 2:1 ratio. Experimental results are shown in Table 2.

As observed in Table 2, maintaining an asymmetric topological structure helps the model strike a reasonable balance. When the mapping groups reach 40 and enhancement groups reach 20, the F1-score hits 0.9087, and its actual average single iteration running time stabilizes at around 6.79 seconds. After this, if the node scale continues to increase, the accuracy gain is very limited (only an increase of 0.0025), but the single iteration time escalates to over 36 seconds. Therefore, adopting a configuration of $maptimes = 40$, $enhencetimes = 20$ is an ideal configuration point balancing recognition accuracy and computational efficiency.

Table 2. Impact analysis of asymmetric network node scale on computational efficiency

Node Groups	Total Matrix Dimension	F1-Score	Avg Iteration Time (s)
10 groups: 5 groups	300	0.8521	0.45 s
20 groups: 10 groups	600	0.8873	1.82 s
40 groups: 20 groups	1200	0.9087	6.79 s
60 groups: 30 groups	1800	0.9104	18.53 s
80 groups: 40 groups	2400	0.9112	36.21 s

To intuitively illustrate the data trends presented in Table 2, Figure 8 employs a Pareto frontier bubble chart to visually unpack the multi-dimensional trade-off among node scale (represented by bubble size), detection performance, and time cost. As depicted, during the initial network expansion, the model's performance improves significantly with marginal time overhead. However, once surpassing the optimal configuration of 40:20 (highlighted by the red bubble), the model abruptly enters the "Diminishing Returns Region" (shaded area). Within this regime, the computational time escalates exponentially, whereas the F1-score curve virtually flatlines. This striking visual trajectory perfectly echoes the

quantitative findings in Table 2, firmly validating that the 40:20 configuration achieves the optimal equilibrium between detection precision and industrial real-time computational feasibility.

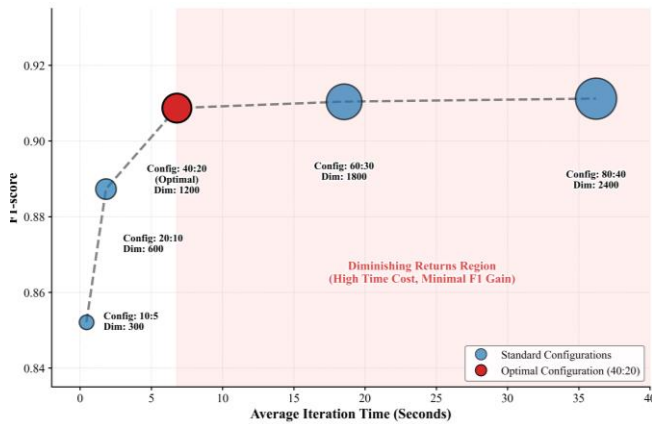


Figure 8. Pareto frontier analysis of network node scale versus computational efficiency

5. CONCLUSIONS

To address the challenges of environmental noise interference and class imbalance in multivariate time-series anomaly detection for complex industrial systems, this paper proposes an IS-BLS. The proposed approach integrates a sliding window feature reconstruction technique with a broad learning model that incorporates a large margin slack factor and an asymmetric fuzzy membership mechanism. During the closed-loop iterative solving process, this mechanism aims to reduce the weight of majority-class baseline noise and directly adjust the decision boundary for the minority class.

Evaluations on the real-world SWaT dataset validate the effectiveness of the proposed method. Comparative experiments indicate that, compared to traditional deep reconstruction and graph learning models, IS-BLS improves the Recall to 0.8443 while maintaining a Precision of 0.9857, yielding an overall F1-score of 0.9087. Furthermore, ablation and sensitivity analyses show that feature reconstruction with a window length of $L = 60$ increases the F1-score by approximately 24.9%; the 'Majority' label allocation mechanism trades a 5.7% decrease in Recall for a 15.8% increase in Precision; and the 40:20 asymmetric network node configuration achieves a reasonable Pareto balance between detection accuracy and computational time (6.79 seconds per iteration).

In conclusion, the IS-BLS method alleviates the trade-off between missed detections and false alarms in industrial anomaly detection to a certain extent. Despite these promising results, this study has certain limitations that warrant future investigation. First, regarding dataset scope and scenario transferability, the current validation is primarily based on the SWaT dataset, which inherently represents a continuous fluid process control system. The direct transferability and effectiveness of the proposed IS-BLS on fundamentally different physical systems—such as high-speed rotating mechanical equipment or discrete manufacturing lines—remain to be fully explored. Second, regarding parameter dependence, the macroscopic feature reconstruction currently relies on a fixed sliding window length. This static setting may

struggle to simultaneously capture ultra-short transient cyberattacks and extremely slow, long-term mechanical degradation anomalies. Future work will focus on integrating adaptive or multi-scale window mechanisms and extending the empirical validation to a broader, more diverse range of industrial infrastructures.

STATEMENT ON THE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE

During the preparation of this manuscript, the author utilized generative artificial intelligence technology (Large Language Models) strictly for language and presentation support, specifically to improve grammatical correctness, stylistic clarity, and sentence structure. The AI tools were not used to generate substantive academic content, data, or experimental results. After using these tools, the author meticulously reviewed and edited the content, taking full responsibility for the final publication.

REFERENCES

- [1] Li, S., Sun, W., Liu, J., Zhang, H., Wang, Y., Zhang, P. (2026). Progress in aero-engine fault signal recognition and intelligent diagnosis. *Machines*, 14(1): 118. <https://doi.org/10.3390/machines14010118>
- [2] Xu, C., Gui, X., Zhao, Y. (2024). Digital twin-assisted multiview reconstruction enhanced domain adaptation graph networks for aero-engine gas path fault diagnosis. *IEEE Sensors Journal*, 24(13): 21694-21705. <https://doi.org/10.1109/JSEN.2024.3400249>
- [3] Chalapathy, R., Chawla, S. (2019). Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*. <https://doi.org/10.48550/arXiv.1901.03407>
- [4] Ruff, L., Kauffmann, J.R., Vandermeulen, R.A., Montavon, G., Samek, W., Kloft, M., Dietterich, T.G., Müller, K.R. (2021). A unifying review of deep and shallow anomaly detection. *Proceedings of the IEEE*, 109(5): 756-795. <https://doi.org/10.1109/JPROC.2021.3052449>
- [5] Zhao, Y., Nasrullah, Z., Li, Z. (2019). PyOD: A python toolbox for scalable outlier detection. *Journal of Machine Learning Research*, 20(96): 1-7. <https://doi.org/10.48550/arXiv.1901.01588>
- [6] Deng, A., Hooi, B. (2021). Graph neural network-based anomaly detection in multivariate time series. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(5): 4027-4035. <https://doi.org/10.1609/aaai.v35i5.16523>
- [7] Chidananda, K., Kumar, A.P.S. (2024). VidAnomalyNet: An efficient anomaly detection in public surveillance videos through deep learning architectures. *International Journal of Safety and Security Engineering*, 14(3): 953-966. <https://doi.org/10.18280/ijssse.140326>
- [8] Abdallah, A.A., El Sayed Abdallah, M.S., Aslan, H., Azer, M.A., Cho, Y.I., Abdallah, M.S. (2024). Enhancing mobile ad hoc network security: An anomaly detection approach using support vector machine for black-hole attack detection. *International Journal of Safety and Security Engineering*, 14(4): 1015-1028. <https://doi.org/10.18280/ijssse.140401>

- [9] Pangavhane, M., Patil, R., Bharati, R., Gupta, D., Ahire, P., Patil, P., Rahane, W., Dharrao, D. (2025). Real-time deep learning-driven surveillance with spatiotemporal feature extraction for detection of anomalous human behavior across dynamic environments. *International Journal of Safety and Security Engineering*, 15(1): 105-111. <https://doi.org/10.18280/ijssse.150112>
- [10] Johnson, J.M., Khoshgoftaar, T.M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1): 1-54. <https://doi.org/10.1186/s40537-019-0192-5>
- [11] Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., Pei, D. (2019). Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage AK, USA, pp. 2828-2837. <https://doi.org/10.1145/3292500.3330672>
- [12] Audibert, J., Michiardi, P., Guyard, F., Marti, S., Zuluaga, M.A. (2020). USAD: UnSupervised anomaly detection on multivariate time series. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Virtual Event CA, USA, pp. 3395-3404. <https://doi.org/10.1145/3394486.3403392>
- [13] Wu, X., Keogh, E. (2021). Current time series anomaly detection benchmarks are flawed and are creating the illusion of progress. *IEEE Transactions on Knowledge and Data Engineering*, 35(3): 2421-2429. <https://doi.org/10.1109/TKDE.2021.3112126>
- [14] Zong, B., Song, Q., Min, M.R., Cheng, W., Lumezanu, C., Cho, D., Chen, H. (2018). Deep autoencoding Gaussian mixture model for unsupervised anomaly detection. *International Conference on Learning Representations (ICLR)*, pp. 1-19.
- [15] Hundman, K., Constantinou, V., Laporte, C., Colwell, I., Soderstrom, T. (2018). Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 387-395. <https://doi.org/10.1145/3219819.3219845>
- [16] Lyu, D., Li, C., Han, Z., Song, Z., Yang, Z., Ma, Y., Hong, J. (2025). Position-agnostic aeroengine intershaft bearing fault diagnosis via condition-guided multitask learning. *IEEE Transactions on Instrumentation and Measurement*, 74: 1-17. <https://doi.org/10.1109/TIM.2025.3575181>
- [17] Wang, M., Ge, Q., Jiang, H., Yao, G. (2019). Wear fault diagnosis of aeroengines based on broad learning system and ensemble learning. *Energies*, 12(24): 4750. <https://doi.org/10.3390/en12244750>
- [18] Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16: 321-357. <https://doi.org/10.1613/jair.953>
- [19] Fernández, A., Garcia, S., Herrera, F., Chawla, N.V. (2018). SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research*, 61: 863-905. <https://doi.org/10.1613/jair.1.11192>
- [20] Venkatasubramanian, V., Rengaswamy, R., Yin, K., Kavuri, S.N. (2003). A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *Computers & Chemical Engineering*, 27(3): 293-311. [https://doi.org/10.1016/S0098-1354\(02\)00160-6](https://doi.org/10.1016/S0098-1354(02)00160-6)
- [21] Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., Nandi, A.K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing*, 138: 106587. <https://doi.org/10.1016/j.ymssp.2019.106587>
- [22] Ren, P., Chen, J., Hu, Y., Yuan, H. (2016). Research on typical wear fault diagnosis of electro-hydraulic servo valve element. In *2016 Prognostics and System Health Management Conference (PHM-Chengdu)*, Chengdu, China, pp. 1-6. <https://doi.org/10.1109/PHM.2016.7819853>
- [23] Liao, Z., Zhan, K., Zhao, H., Deng, Y., Geng, J., Chen, X., Song, Z. (2024). Addressing class-imbalanced learning in real-time aero-engine gas-path fault diagnosis via feature filtering and mapping. *Reliability Engineering & System Safety*, 249: 110189. <https://doi.org/10.1016/j.ress.2024.110189>
- [24] Vairetti, C., Assadi, J.L., Maldonado, S. (2024). Efficient hybrid oversampling and intelligent undersampling for imbalanced big data classification. *Expert Systems with Applications*, 246: 123149. <https://doi.org/10.1016/j.eswa.2024.123149>
- [25] Yin, X., He, W., He, Q., Chen, D., Zhang, B., Zhao, H. (2026). Data imbalanced fault diagnosis of aviation fuel pumps based on adaptive weighting using an enhanced broad learning system. *Measurement*, 257: 118612. <https://doi.org/10.1016/j.measurement.2025.118612>
- [26] Liao, W., Zhu, R., Ge, L., Cao, D., Yang, Z. (2025). Mitigating class imbalance issues in electricity theft detection via a sample-weighted loss. *IEEE Transactions on Industrial Informatics*, 21(2): 1754-1763. <https://doi.org/10.1109/TII.2024.3485813>
- [27] Yao, S., Sun, S., Zhang, Y., Xu, C., Chen, W. (2025). Handling class imbalance in SAGIN heterogeneous devices via location-slack-fuzzy broad learning system. *Transactions on Emerging Telecommunications Technologies*, 36(1): e70242. <https://doi.org/10.1002/ett.70242>
- [28] Chen, W., Yang, K., Yu, Z., Nie, F., Chen, C.L.P. (2025). Adaptive broad network with graph-fuzzy embedding for imbalanced noise data. *IEEE Transactions on Fuzzy Systems*, 33(6): 1949-1962. <https://doi.org/10.1109/TFUZZ.2025.3543369>
- [29] Li, Y., Gao, Y., Jin, J., Nan, J., Meng, Y., Wang, M., Chen, C.L.P. (2025). Adaptive weights-based relaxed broad learning system for imbalanced classification. *Digital Signal Processing*, 156: 104869. <https://doi.org/10.1016/j.dsp.2024.104869>
- [30] Li, Y., Wang, Y., Jin, J., Zhang, W., Tao, H., Wu, H., Chen, C.L.P. (2025). Imbalanced broad learning system with label relaxation and sample weight adaptation. *Applied Soft Computing*, 182: 113543. <https://doi.org/10.1016/j.asoc.2025.113543>
- [31] Chen, W., Yu, Z., Yang, K., Jiang, J., Zhang, F., Chen, C.P. (2025). Minimum variance weighted broad cascade network structure for imbalanced classification. *Knowledge-Based Systems*, 324: 113803. <https://doi.org/10.1016/j.knosys.2025.113803>
- [32] Chen, C.P., Liu, Z. (2017). Broad learning system: An effective and efficient incremental learning system without the need for deep architecture. *IEEE Transactions on Neural Networks and Learning Systems*, 29(1): 10-24. <https://doi.org/10.1109/TNNLS.2017.2716952>
- [33] Gong, X., Zhang, T., Chen, C.P., Liu, Z. (2022).

Research review for broad learning system: Algorithms, theory, and applications. IEEE Transactions on Cybernetics, 52(9): 8922-8950. <https://doi.org/10.1109/TCYB.2021.3061094>

[34] Zhao, H., Zheng, J., Deng, W., Song, Y. (2020). Semi-supervised broad learning system based on manifold regularization and broad network. IEEE Transactions on Circuits and Systems I: Regular Papers, 67(3): 983-994. <https://doi.org/10.1109/TCSI.2019.2959886>

[35] Yang, K., Chen, W., Shi, Y., Yu, Z., Chen, C.L.P. (2024). Simplified kernel-based cost-sensitive broad learning system for imbalanced fault diagnosis. IEEE Transactions on Artificial Intelligence, 5(12): 6629-6644. <https://doi.org/10.1109/TAI.2024.3478191>

[36] Chen, W., Yang, K., Yu, Z., Zhang, W. (2022). Double-kernel based class-specific broad learning system for multiclass imbalance learning. Knowledge-Based Systems, 253: 109535. <https://doi.org/10.1016/j.knosys.2022.109535>

[37] Gao, Y., Dong, J. (2026). Robust imbalanced learning for aero-engine bearing anomaly detection via a hybrid SMOTE-BLS framework. Aerospace Engineering Communications, 1(1): 47-56. <https://doi.org/10.62762/AEC.2026.599020>

[38] Liu, L., Guo, J., Yin, Z., Chen, R., Huang, G. (2025). A novel three-way distance-based fuzzy large margin distribution machine for imbalance classification. Complex & Intelligent Systems, 11(3): 176. <https://doi.org/10.1007/s40747-025-01797-w>

[39] Wang, T., Qiu, Y., Hua, J. (2020). Centered kernel alignment inspired fuzzy support vector machine. Fuzzy Sets and Systems, 394: 110-123. <https://doi.org/10.1016/j.fss.2019.10.005>

[40] Goh, J., Adepu, S., Tan, M., Lee, Z.S. (2016). A dataset to support research in the design of secure water treatment systems. In Critical Information Infrastructures Security. CRITIS 2016, Lecture Notes in Computer Science, Springer, Cham, pp. 88-99. https://doi.org/10.1007/978-3-319-71368-7_8

[41] Malhotra, P., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., Shroff, G. (2016). LSTM-based encoder-decoder for multi-sensor anomaly detection. arXiv:1607.00148. <https://doi.org/10.48550/arXiv.1607.00148>

[42] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y. (2018). Graph attention networks. arXiv:1710.10903. <https://doi.org/10.48550/arXiv.1710.10903>

[43] Kipf, T.N., Welling, M. (2016). Variational graph auto-encoders. arXiv:1611.07308. <https://doi.org/10.48550/arXiv.1611.07308>

[44] Ding, K., Li, J., Bhanushali, R., Liu, H. (2019). Deep anomaly detection on attributed networks. In Proceedings of the 2019 SIAM International Conference on Data Mining (SDM), pp. 594-602. <https://doi.org/10.1137/1.9781611975673.67>

NOMENCLATURE

a	Majority class position tolerance hyperparameter
A	High-dimensional generalized state matrix
A_{test}	Test set generalized state matrix

ACF	First-order autocorrelation coefficient
C	Total number of classes
D	Reconstructed static feature dimension
H_j	Enhancement nodes of the j -th group
H^m	Global enhancement node matrix
I	Identity matrix
$J(\cdot)$	Weighted least squares objective function
k	Linear slope of local trend
L	Sliding window length
$Label_{test}$	Discrete physical equipment state label
m	Number of enhancement node groups
M	Total number of generated window samples
n	Number of feature mapping node groups
N	Number of sensor channels
N_e	Total number of enhancement nodes
N_f	Total number of feature nodes
R	Range of signal fluctuations
S	Sliding step size
T	Total time steps of raw data
W_{e_i}	Randomly initialized weights for feature nodes
W_{h_j}	Randomly initialized weights for enhancement nodes
W_{out}	Output layer weight matrix
W_t	Local multivariate time window data subset
X_{raw}	Raw multivariate time-series data tensor
X_{feat}	Reconstructed static feature vector
Y	True label matrix
\hat{Y}_{test}	Continuous model prediction score vector
Z_i	Feature mapping nodes of the i -th group
Z^n	Global feature mapping matrix

Greek symbols

β	Fuzzy membership exponential decay coefficient
β_{e_i}	Random biases for feature nodes
β_{h_j}	Random biases for enhancement nodes
Δ	First-order forward difference operator
ϵ	Weight convergence residual threshold
λ	L_2 Regularization coefficient
μ	Mean of statistical distribution
ξ_i	Large margin slack factor
$\xi(\cdot)$	Nonlinear activation function for enhancement nodes
σ	Standard deviation of signal fluctuations
Ψ	Adaptive fuzzy error weighted diagonal matrix
ψ_i	Asymmetric fuzzy membership weight for the i -th sample
$\phi(\cdot)$	Nonlinear activation function for feature nodes

Subscripts and superscripts

$+$	Minority class (anomalous state)
$-$	Majority class (normal state)
Δ	Difference sequence
i	Sample index or feature node group index
j	Enhancement node group index
\max	Maximum value
\min	Minimum value
(t)	Current iteration step