Artificial Intelligence and Computing Innovations

Intelligence Science Publishing

https://www.ispublishing.com

ISSN: 3070-4944

Vo.4, No.2, 2024

Self-Supervised Unified Representation Learning with Structural Dependency Modeling for Early Risk Detection in Distributed Systems

Ting Ying[1], Chris Ding[2]*

[1]*Northwestern University, Evanston, USA*

[2]*Boston University, Boston, USA*

**Corresponding author: Chrisddding@gmail.com*

Abstract: This study addresses the challenge of accurately modeling system states in distributed systems under highly dynamic workloads, multi-indicator coupling, and continuously evolving structures. A self-supervised unified representation framework is proposed for capturing multi-indicator association features. The method first constructs a high-dimensional temporal input matrix from multi-source monitoring data and uses a nonlinear encoder to extract multi-scale features, including local trends, short-term disturbances, and long-term dependencies. A dependency graph based on structural consistency is then used to describe functional relations and potential propagation paths between nodes. A structural propagation mechanism integrates multi-indicator coupling relationships into a unified latent representation space. Predictive, contrastive, and reconstruction-based self-supervised tasks are further designed to enable the model to learn stable structural patterns and statistical regularities of normal operation without labels. Early risk signals are identified by detecting shifts in the latent space, providing efficient and reliable support for system safety. Experimental results show that the framework exhibits strong robustness and adaptability under conditions such as structural disturbances, node fault injection, and data noise. It can capture hidden early anomaly features and significantly enhance early risk detection. The method integrates temporal features, structural dependencies, and self-supervised signals in a unified modeling manner and offers a new technical pathway for understanding complex association behaviors in large-scale distributed systems.

Keywords: Self-supervised representation learning; distributed systems; multi-metric coupling; early risk detection

1. Introduction

In highly digitalized and service-oriented infrastructure environments, distributed systems have become the core computational foundation for cloud computing, online services, intelligent manufacturing, financial technology, and complex business platforms. As service scale continues to grow, these systems exhibit highly dynamic workloads, parallel scheduling of multiple resource types, frequent cross-node interactions, and continuously evolving topologies[1]. Their operating states show strong nonlinearity, multiple disturbances, and high-dimensional coupling. In this context, even minor performance drift, resource bottlenecks, or local anomalies can propagate along node chains and develop into broader system risks. This process may lead to rising latency, service degradation, or even global failures. Accurately understanding overall system behavior and identifying potential risks at an early stage has therefore become a key scientific and engineering problem for maintaining the stability of modern distributed systems[2].

At the same time, with the continuous enhancement of monitoring infrastructures, the volume of operational data generated by distributed systems has grown exponentially. CPU, memory, disk, IO, bandwidth, container load, tracing data, resource contention metrics, and application-level behavior jointly form a multi-dimensional, multi-granular, and high-frequency monitoring ecosystem. These indicators carry not only temporal dependencies but also complex structural coupling and implicit causal relations[3]. As a result, the evolution of anomalies often spans multiple nodes and resource types. Traditional single-indicator or weak-structure modeling methods cannot capture such cross-dimensional and multi-path risk propagation patterns. They cannot meet the needs of anomaly and risk detection in complex systems. In addition, many real-world anomalies lack standardized labels, and the cost of manual labeling is very high. Supervised methods cannot cover the long-term and multi-scenario distributional changes present in real environments. It is therefore necessary to develop models that can autonomously understand system states and learn indicator coupling patterns under unlabeled conditions[4].

In recent years, self-supervised learning has shown strong generalization ability and robustness in modeling high-dimensional time series. Its core advantage is that it does not rely on manual labels. Instead, it extracts intrinsic structural patterns from raw monitoring data through prediction, reconstruction, or contrastive tasks. This paradigm brings new opportunities to distributed systems[5,6]. During normal states, systems tend to maintain stable coupling patterns. The emergence of potential risks disrupts these patterns. Self-supervised tasks can therefore capture internal structural consistency, dependency relations, and coordinated changes across indicators. This provides a modeling basis that is more generalizable and more aligned with real operational behaviors for early risk detection. In addition, self-supervised methods can continuously use newly generated unlabeled data. This allows them to adapt to long-term trends, version updates, and environmental shifts. Such adaptability is difficult for traditional approaches to achieve.

In this context, risks in distributed systems often do not appear as isolated anomalies. More commonly, they propagate across links in a structured manner. Resource contention may affect multiple interdependent nodes. Microservice topology changes may trigger dynamic shifts in request paths. Cluster load balancing strategies may spread local performance fluctuations across the entire system[7]. These cross-node and cross-indicator interactions create complex spatiotemporal propagation behaviors. Early risk detection, therefore, cannot rely on local thresholds or static models. Effective methods must understand coupling patterns across indicators, interactions between nodes, and dynamic evolution caused by structural changes. Building a unified state representation model that captures multi-indicator coupling and uses self-supervised mechanisms to analyze internal associations has become a key direction for improving risk detection.

Based on the above background, developing a method that can automatically learn operational patterns from large-scale multi-source monitoring data, detect disruptions to structural consistency, and

provide early warnings of potential risks has important theoretical and practical value. Such methods can move beyond the label-dependent paradigm of traditional supervised learning. They can better match real operational scenarios where indicator distributions and system structures are complex and constantly changing. By capturing deep coupling features across indicators and cross-node propagation patterns, this line of research can advance predictive maintenance, intelligent scheduling, capacity planning, and adaptive control in distributed systems. It can help maintain stability under highly dynamic workloads, resource fluctuations, and evolving topologies. Furthermore, it provides a theoretical foundation for next-generation intelligent operations and offers essential support for the stable operation of high-concurrency services and large-scale cloud platforms.

Recent studies further clarify the modeling requirements for early risk detection in distributed systems. Generalizable latent representation learning for multi-source cloud time series supports the construction of transferable state embeddings from heterogeneous monitoring streams[8]. Structure-aware root cause localization based on joint graph representation learning and causal inference strengthens the interpretation of dependency relations and propagation paths among microservice backends[9]. Embedding and sequence modeling for microstructure-aware risk detection shows how weak risk signals can be captured from high-frequency dynamic observations through compact sequential representations[10]. Semi-supervised robust optimization for low-resource noisy data mining provides a methodological basis for learning stable decision boundaries when labels are scarce and operational signals contain disturbance[11]. Backend latency forecasting with multi-scale convolution and adaptive attention demonstrates the value of combining local fluctuation extraction with long-range dependency modeling[12]. Secure parameter aggregation and distributed anomaly detection in cross-cloud data center networks extend this direction toward privacy-preserving and distributed risk analysis[13]. Multi-scale Transformer forecasting for CPU utilization further supports hierarchical temporal modeling in cloud resource scenarios[14], while differentiable mixed-integer optimization for graph-based node decision-making offers a structured optimization perspective for risk-aware system control[15]. These studies collectively support the proposed framework by linking multi-source representation learning, structural dependency modeling, robust semi-supervised optimization, temporal feature extraction, and graph-based decision logic.

2. Proposed Framework

To achieve a unified representation of the multi-indicator coupled structure in distributed systems, the method first constructs a multi-dimensional time-series input matrix based on monitoring data, integrating system indicators from different sources into a unified sequential representation according to time order. Let the observation vector of the system at time t be $x_t \in R^d$, and the entire monitoring window can be represented as a matrix $X = [x_1, \ldots, x_T]$. To ensure that the model can capture local trend changes and high-dimensional dependencies, a non-linear transformation is used in the encoding stage to map the original indicators to a latent representation space, forming a basic state representation:

$$h_t = f_\theta(x_t) \tag{1}$$

Here, f_θ represents a learnable multi-layer nonlinear structure. This representation not only provides a compressed expression of the instantaneous system state but also offers a unified feature basis for subsequent structure-dependent modeling. The overall model architecture is shown in Figure 1.

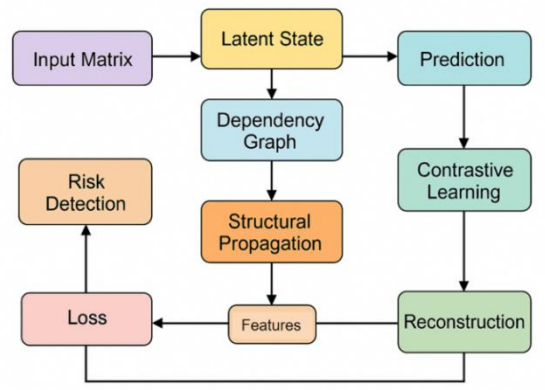


Figure 1. Overall model architecture

To characterize the dynamic coupling relationships between different indicators, the method constructs a dependency graph based on correlation and structural consistency. By calculating the degree of association between any two time points or any two dimensions of indicators, a dynamic adjacency matrix can be obtained:

$$A_{ij} = g(h_i, h_j) \tag{2}$$

Matrix $g(\bullet)$ is used to measure the similarity or functional dependence between features. This matrix changes dynamically over time, reflecting structural changes within the system due to resource competition, link interactions, or topological drift. Based on this, the method further constructs a structural propagation operator:

$$z_t = \sum_{j=1}^{T} A_{ij} h_j \tag{3}$$

This is used to integrate local representations with structure-dependent information into a unified system state representation, enabling the model to perceive potential risk factors propagating across nodes in the representation space.

In self-supervised learning mechanisms, methods enable models to continuously learn the stable internal structure of the system even without labeled data by constructing predictive, contrastive, and consistency-based objectives. For sequence-level future state prediction, the self-supervised objective can be expressed as:

$$L_{pred} = \|\hat{x}_{t+k} - x_{t+k}\|_2^2 \tag{4}$$

Here, $\widehat{x}_{t+k}$ represents the future vector inferred by the model based on the current state. On the other hand, to strengthen the system's robustness and consistency during structural changes, the method designs a feature comparison objective to distinguish between similar states under normal structures and structural deviations caused by potential risks. Its form is as follows:

$$L_{cont} = -\log \frac{\exp(sim(z_i, z_i^+)/\tau)}{\sum_j \exp(sim(z_i, z_j)/\tau)} \tag{5}$$

Here, z_i^+ represents the positive samples constructed under structural consistency conditions, and τ represents the temperature coefficient. This mechanism prompts the model to automatically learn the statistical regularities of normal operating conditions and capture structural disturbances that deviate from these regularities.

Finally, to construct a stable latent space suitable for early risk detection, the method obtains a joint encoder by minimizing a multi-objective self-supervised loss. The overall optimization objective is defined as:

$$L = \lambda_1 L_{pred} + \lambda_2 L_{cont} + \lambda_3 L_{recon} \tag{6}$$

Here, L_{recon} represents an auxiliary consistency loss based on state reconstruction, used to improve the completeness and structural stability of the representation. Through this joint optimization strategy, the model obtains a state representation capable of uniformly characterizing the coupling relationships of multiple indicators, capturing structural propagation patterns, and being sensitive to abnormal structural deviations, enabling it to form an early identification capability for potential risks even under unsupervised conditions.

3. Experimental Analysis

3.1 Dataset

This study uses the public dataset named Multi-Source Distributed System Data for AI-Powered Analytics, or MSDS. The dataset is generated by a complex distributed system during its operation. It covers three typical monitoring modalities used in system operations. These include system metrics, application logs, and distributed call traces. The metrics contain numerical data sampled over time across multiple monitoring dimensions. The log files record application-level events and anomaly messages. The trace data describes the call chain of requests across system service components. This design matches the multi-indicator and multi-modality characteristics of modern distributed system monitoring and provides strong practical relevance and research value.

From a structural perspective, the MSDS dataset supports in-depth analysis of coupling relations and propagation paths during system operation. It provides time series data for each metric. It also reveals node interactions, service dependencies, and resource sharing patterns through logs and call chains. This cross-dimensional, multi-node, and multi-modality observation setting aligns well with the focus of this study on multi-indicator coupling relations and structural propagation mechanisms. The dataset enables the construction of latent system state representations, the understanding of coupling patterns both across indicators and across nodes, and the tracing of how micro-level anomalies evolve into system-level risks.

In terms of significance, using the MSDS dataset supports the proposed framework for self-supervised, structured, and coupling-aware modeling. The dataset contains rich, unlabeled operational data, including long-term logs and metric records, as well as multi-path propagation patterns observed in real systems. This makes it suitable for self-supervised learning. The model first learns the coupling patterns and propagation structures that appear during normal operation and then detects deviations from these patterns for early risk warning. Moreover, as a public dataset, MSDS ensures reproducibility and community-level comparability and enhances the generalizability of the research contributions. Overall, MSDS provides a solid data foundation and validation basis for this study.

3.2 Experimental Results

This paper first conducts a comparative experiment, and the experimental results are shown in Table 1.

Table 1. Comparative experimental results

Method	Acc	Precision	F1-Score	AUC
MLP[16]	0.873	0.861	0.854	0.892
XGBoost[17]	0.891	0.878	0.872	0.903
Transformer[18]	0.914	0.907	0.902	0.931
1DCNN[19]	0.902	0.895	0.888	0.918
Ours	0.948	0.942	0.939	0.967

From the overall performance, traditional methods such as MLP and XGBoost achieve a certain level of recognition in distributed systems with multi-indicator coupling. However, these methods mainly rely on independent features or shallow structures. Their ability to model cross-indicator relations and temporal dynamics is limited. As a result, both F1 Score and AUC are clearly lower than those of deep structural models. For example, MLP reaches an F1 Score of 0.854, and XGBoost shows only a slight improvement. This indicates that static features or weak dependency modeling cannot adequately capture the coupling patterns or potential structural disturbances in complex systems.

In contrast, deep representation models such as Transformer and 1DCNN have stronger capability in capturing high-dimensional temporal features and local dynamic behaviors. Their accuracy, F1 Score, and AUC therefore show significant improvement. Transformer achieves an AUC of 0.931, which indicates that the multi-head attention mechanism can effectively model long-range dependencies across indicators. This ability helps the model identify hidden cross-node fluctuations and weak early signs of anomalies. The 1DCNN uses local convolutions to capture short-term changes and also surpasses traditional baseline models in accuracy.

The method proposed in this study achieves the best performance across all metrics. The F1 Score increases from 0.902 in Transformer to 0.939, and the AUC improves to 0.967. This demonstrates that the method captures the coupling structure of multi-indicator distributed systems more effectively. Through self-supervised objectives, the model learns structural consistency and propagation patterns under normal system operation. Potential risks can therefore be detected through feature shifts before they become explicit. The performance advantage shows that the proposed method not only provides higher quality representations but also maintains higher sensitivity to early anomaly signals in complex systems. This enables more reliable risk detection under highly dynamic and high-dimensional operating environments.

A hyperparameter study on how the hidden layer dimensionality affects the AUC is reported in Figure 2.

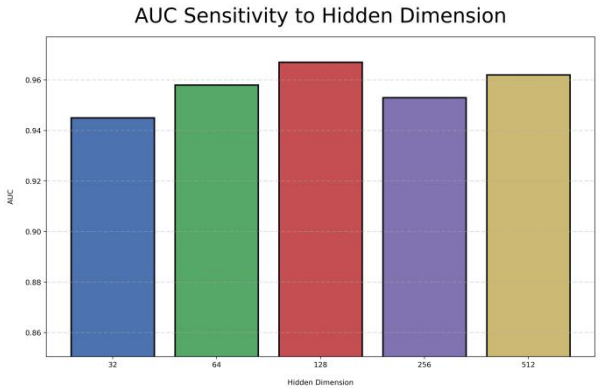


Figure 2. Experiment on the sensitivity of the AUC metric to hyperparameters related to hidden layer dimensions.

As the hidden layer dimension is adjusted, the AUC changes in a non-monotonic manner. This indicates that the representation capacity required for multi-indicator coupling is not simply improved by increasing dimensionality. When the dimension is low, such as 32 and 64, the model can only capture local patterns in the operating states of distributed systems. It cannot fully represent cross-indicator or cross-node structural dependencies. As a result, the AUC remains in a relatively low range. When the dimension increases to 128, the latent space gains stronger expressive power. It can model coupling structures and propagation patterns more accurately. This leads to the peak AUC value.

When the hidden dimension continues to grow to 256, the AUC decreases significantly. This shows that excessive dimensionality introduces additional noise. The model becomes more likely to capture unstable, irrelevant, or short-term random fluctuations in system states. This weakens overall discriminative ability. This phenomenon suggests that, in multi-indicator coupling scenarios, an overly expanded latent space may break structural consistency. It makes self-supervised learning less stable in capturing essential system patterns. In addition, high-dimensional features may increase the burden on the structural propagation module. This further reduces its ability to model true coupling relationships.

Overall, a medium-sized dimension such as 128 achieves a good balance between expressive power and structural stability. The model can capture temporal dependencies while preserving cross-indicator coupling patterns. This leads to the best AUC performance. The result shows that enlarging network size alone is not sufficient to improve performance in self-supervised modeling of distributed system states. The key is to select a dimensionality that maintains structural consistency and compact representation. This ensures that the model can reliably identify potential risks under highly dynamic and disturbance-rich environments.

Figure 3 illustrates how varying the node failure injection rate influences the model's performance.

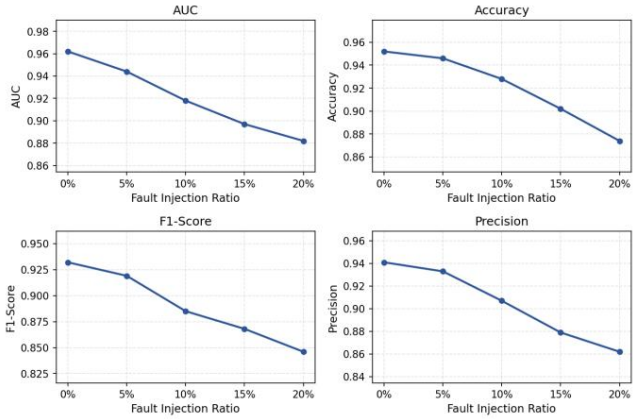


Figure 3. The impact of node failure injection rate on experimental results

When the node fault injection ratio increases from 0% to 20%, all four metrics consistently deteriorate. This indicates that when structural disturbances intensify, the predictability and stability of the distributed system weaken. The decline in AUC and Accuracy is more pronounced. This reflects that higher node failure frequency disrupts structural consistency and shifts cross-node coupling relationships. The model faces greater interference when distinguishing normal states from risk states. This pattern is consistent with the characteristics of multi-indicator coupling systems. A single node fault can propagate along call chains or resource dependency paths and influence the decision boundaries of the entire state space.

A closer look at the declining trends of F1 Score and Precision shows that the performance drop is not only caused by reduced overall prediction capability. It also results from the blurring of early anomaly-related features within the system. When the fault injection ratio increases, the noise level and fluctuation range of some indicators expand. This makes the latent representation space constructed through self-supervised learning unstable. Clear structural patterns become weaker, and the model becomes less capable of detecting potential risk signals at an early stage. This reflects the vulnerability of distributed systems in high-disturbance environments. In addition, the decrease in Precision implies a higher false alarm rate, which aligns with real systems where fault-related noise can confuse judgment.

Overall, the experiment reveals the sensitivity of the model when facing structural disturbances. Under low fault injection ratios, the model maintains high robustness by relying on learned multi-indicator coupling structures. As the disturbance range expands, structural propagation paths are damaged. The self-supervised representations gradually deviate from the statistical patterns of normal states. This

leads to continuous performance degradation. The results indicate that in real deployment scenarios, improving the model's adaptability to structural fluctuations is essential for effective early risk detection. Approaches such as strengthening structural consistency constraints, enhancing robust representation learning, or using dynamic structure modeling can help maintain reliable risk identification under highly dynamic and disturbance-rich system environments.

4. Conclusion

This study addresses the understanding of system states and early risk detection in distributed systems under highly dynamic workloads, multi-indicator coupling, and continuously evolving structures. A unified self-supervised modeling framework is proposed. The framework integrates multi-dimensional temporal representation, structural dependency learning, and cross-node propagation mechanisms. This design enables deep characterization of internal behaviors in complex systems. The latent representation space constructed in this study captures long-term trends and short-term disturbances in system states. It also uncovers coupling structures hidden behind multiple indicators and provides a more interpretable modeling perspective for understanding system behavior. By combining predictive, contrastive, and structural consistency self-supervised tasks, the framework can continuously learn stable system patterns in an unlabeled environment and build a strong foundation for risk identification.

The results show that the framework demonstrates strong robustness and adaptability when facing structural variations, node fault injection, and multi-source noise disturbances. Risks in distributed systems often originate from subtle initial deviations. These weak signals are easily overwhelmed in traditional representation methods. The structured latent space constructed in this study can capture such deviations with high sensitivity. It enables the model to identify potential risks before they become explicit. This capability is valuable for core tasks such as resource scheduling, service self-healing, and intelligent operations. It provides a new technical pathway to improve system availability, performance stability, and service continuity.

At the application level, the proposed modeling paradigm can serve a wide range of critical scenarios. These include cloud computing platforms, microservice architectures, large-scale distributed storage, financial-grade backend systems, and industrial internet environments. As system scale expands and operational structures become more complex, traditional risk detection approaches that rely on rules or supervised data can no longer meet long-term operational requirements. The self-supervised mechanism in this study removes the dependence on manual labels and provides a scalable and reproducible foundation for future large-scale intelligent operation systems. Its adaptability to distribution shifts and environmental disturbances also allows the method to maintain high reliability in high-concurrency and high uncertainty systems. This offers potential support for advancing automation and autonomy in intelligent system operations.

Future work can focus on dynamic structure modeling, multimodal data fusion, and cross-system knowledge transfer to further improve generalization and real-time performance in complex scenarios. One direction is to explore dynamic structure models built from finer-grained call chains, event streams, and resource contention graphs. This can make system state modeling more consistent with real operational mechanisms. Another direction is to introduce cross-domain self-supervised transfer strategies. This would allow the model to transfer its learned structural knowledge across heterogeneous systems and support risk prediction and system understanding in larger and more complex environments. As the methodology continues to develop, the proposed framework has the potential to become an essential foundational module for next-generation intelligent operations in distributed systems and to provide sustained support for future autonomous system management.

References

[1] Zhang Y, Wang J, Chen Y, et al. Adaptive memory networks with self-supervised learning for unsupervised anomaly detection[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(12): 12068-12080.

[2] Fu Y, Xue F. Mad: Self-supervised masked anomaly detection task for multivariate time series[C]//2022 International Joint Conference on Neural Networks (IJCNN). IEEE, 2022: 1-8.

[3] Mozaffari M, Doshi K, Yilmaz Y. Self-supervised learning for online anomaly detection in high-dimensional data streams[J]. Electronics, 2023, 12(9): 1971.

[4] Pazho A D, Noghre G A, Purkayastha A A, et al. A survey of graph-based deep learning for anomaly detection in distributed systems[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 36(1): 1-20.

[5] Li D, Chen D, Jin B, et al. MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks[C]//International Conference on Artificial Neural Networks. Springer, 2019: 703-716.

[6] Yang W, Zhang K, Hoi S C H. A causal approach to detecting multivariate time-series anomalies and root causes[J]. arXiv preprint arXiv:2206.15033, 2022.

[7] Carvalho J, Zhang M, Geyer R, et al. Invariant anomaly detection under distribution shifts: a causal perspective[J]. Advances in Neural Information Processing Systems, 2023, 36: 56310-56337.

[8] Huang W. Generalizable Latent Representation Learning for Multi-Source Time Series Forecasting in Cloud Systems[J]. 2024, 3(1).

[9] Wang L, Chang C Y. Structure-Aware Root Cause Localization for Microservice Backends via Joint Graph Representation Learning and Causal Inference[J]. 2024, 3(1).

[10] Su Z. Microstructure-Aware Risk Detection in High-Frequency Trading Using Embedding and Sequence Modeling[J]. 2024, 4(3).

[11] Mei Y. Unified Semi-Supervised and Robust Optimization Framework for Low-Resource Noisy Data Mining[J]. 2024, 3(7).

[12] Pi S. Intelligent Backend Latency Forecasting with Multi-Scale Convolution and Adaptive Attention Mechanisms[J]. 2024, 3(5).

[13] Zhang S. Secure Parameter Aggregation and Distributed Anomaly Detection in Cross-Cloud Data Center Networks[J]. 2024, 3(4).

[14] Lim B, Arık S Ö, Loeff N, et al. Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting[J]. International Journal of Forecasting, 2021, 37(4): 1748-1764.

[15] Ren X. A Unified Framework for Social Graph Modeling and Node Decision-Making via Differentiable Mixed-Integer Optimization[J]. 2024, 3(5).

[16] Audibert J, Michiardi P, Guyard F, et al. USAD: UnSupervised Anomaly Detection on Multivariate Time Series[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2020: 3395-3404.

[17] Chen Z, Li Z W, Huang J, et al. An effective method for anomaly detection in industrial Internet of Things using XGBoost and LSTM[J]. Scientific Reports, 2024, 14(1): 23969.

[18] Tuli S, Casale G, Jennings N R. Tranad: Deep transformer networks for anomaly detection in multivariate time series data[J]. arXiv preprint arXiv:2201.07284, 2022.

[19] Athar A, Mozumder M A I, Ali S, et al. Deep learning-based anomaly detection using one-dimensional convolutional neural networks (1D CNN) in machine centers (MCT) and computer numerical control (CNC) machines[J]. PeerJ Computer Science, 2024, 10: e2389.