Artificial Intelligence and Computing Innovations

Intelligence Science Publishing

https://www.ispublishing.com

ISSN: 3070-4944

Vo.5, No.6, 2025

Adaptive Data Mining-Based Deep Detection for Streaming Time-Series Anomaly Detection

Shuai Li[1*]

[1]University of Sheffield, Sheffield, UK

**Corresponding author: lsqd93@hotmail.com*

Abstract: Addressing the challenges of anomaly sparsity, insufficient labeling, time-varying data distribution, and diverse anomaly patterns in streaming time-series scenarios, this paper proposes an adaptive data mining and energy model-based deep detection framework for anomaly detection. This method maps the input sequence to a latent representation space using a deep encoder. Self-supervised contrastive Learning enhances the robustness of the representation to noise and perturbations, and an online memory mechanism maintains a set of recent low-energy samples to characterize the current normal structure. Based on this, an energy scoring function based on normal prototypes is constructed, transforming the consistency between samples and dynamic normal centers into a continuously comparable anomaly metric. Simultaneously, a smoothly updated quantile threshold achieves stable online discrimination. This mechanism unifies structure maintenance, scoring scale, and alarm decision-making within a single process, enabling the model to maintain adaptability to changes in normal patterns and sensitivity to anomaly deviations in a continuously arriving data stream. Comparative experimental results demonstrate that the proposed method achieves superior overall performance across multiple evaluation metrics, validating the effectiveness and practical value of combining adaptive structure maintenance and energy measurement in time-series anomaly detection tasks.

Keywords: Energy scoring; online memory; self-supervised consistency; streaming alerts

1 Introduction

With the rapid development of cloud computing platforms, IoT systems, the Industrial Internet, and online transaction links, data generation has shifted from "static batch" to "continuous streaming," exhibiting typical characteristics of strong noise, strong heterogeneity, and strong coupling. Once a system malfunctions, it often triggers performance fluctuations, service degradation, and cascading failures, leading to business losses and security risks[1]. Therefore, building an algorithmic system for anomaly detection has become a crucial foundational capability for intelligent operation and maintenance, risk warning, and high-reliability system governance. Meanwhile, anomaly events are typically rare and irregular, and their patterns evolve continuously with changes in environment and load,

making traditional methods relying on fixed rules or static thresholds unsustainable in the long term. Against this backdrop, combining the representational capabilities of deep learning with the pattern discovery capabilities of "data mining" to form a sustainably adaptable anomaly detection framework has significant research value and application implications[2].

Existing anomaly detection research has generally evolved from statistical hypothesis testing and distance-density methods to deep representation learning. However, it still faces several common bottlenecks in real-world open environments: First, anomalies often lack sufficient labeling, or even lack usable labels altogether, limiting the supervision paradigm; second, data distribution drifts over time, making models prone to threshold failure and false alarm accumulation; third, anomaly categories are highly diverse, including point anomalies, contextual anomalies, and group anomalies, making it difficult to uniformly describe single discrimination boundaries; fourth, system noise and missing data collection amplify uncertainty and weaken detection stability[3]. To more clearly summarize these key challenges and required capabilities, this paper provides a summary in Table 1 and emphasizes the necessity of "adaptive data mining" and energy modeling in anomaly detection in the subsequent background discussion.

Table 1. Key Challenges and Capability Requirements for Anomaly Detection

Key Challenge	Capability Requirement
Limited or no labels	Self-supervised representations
Distribution drift	Online adaptation
Diverse anomaly types	Unified scoring mechanism
Noise and missing data	Robust modeling

At the methodological level, "adaptive data mining" emphasizes continuously discovering structural patterns and changes in data streams, transforming anomalies from "deviations under fixed rules" into "uninterpretability relative to the current data generation mechanism." This perspective naturally aligns with open environments: when system conditions, user behavior, or sensor states change, the model should have the ability to dynamically update representations and thresholds, thereby avoiding performance collapse caused by "concept drift." Simultaneously, anomaly detection requires not only a feature space capable of expressing complex patterns but also a stable, comparable, and calibrable anomaly scoring function to maintain a consistent discrimination scale across different time periods, data sources, and noise levels. Based on this, energy models, as a modeling approach that characterizes "sample fit" from an overall distribution perspective, provide a more unified form for anomaly scoring: by mapping normal patterns to low-energy regions and uninterpretable samples to high-energy regions, a more general anomaly measurement can be achieved without relying on precise category definitions, complementing adaptive mechanisms[4,5].

From a research perspective, adaptive data mining and energy model-based deep detectors for anomaly detection are expected to make fundamental contributions in three aspects: First, they will provide a more sustainable detection path for unlabeled or weakly labeled scenarios, reducing reliance on manual experience and rule maintenance; second, they will establish a unified scale for anomaly assessment across scenarios and time periods, alleviating difficulties in threshold migration and the propagation of false alarms; and third, they will provide interpretable structured clues for anomalies in

high-dimensional heterogeneous data, supporting subsequent localization, alarm classification, and risk management. More importantly, by incorporating pattern discovery from "data mining" and representation learning from "deep learning" into the same framework, anomaly detection can be advanced from a single-point algorithm problem to a long-term capability building approach for complex systems, providing a reusable algorithmic foundation for highly reliable intelligent systems, automated governance, and security protection[6].

2 Model Design

2.1 General Introduction

Let A be the continuously arriving data stream. The goal is to output anomaly scores in real time and maintain adaptability to distribution drift under weakly supervised or unlabeled conditions. The overall method consists of three parts: "adaptive data mining + deep representation learning + energy scorer". First, an encoder maps the original samples to a stable latent space. Then, through data mining-style online memorization and representative sample maintenance, the model continuously captures the structure of the "current normal pattern". Finally, an energy function characterizes the consistency between the sample and the current normal structure; the higher the energy, the greater the deviation from the normal. Let B be the encoder, and the latent representation be:

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

θ will be updated over time, but the update needs to be constrained by stability to avoid catastrophic drift under noise or short-term fluctuations. Therefore, a collaborative mechanism of "slowly updated reference representation" and "online threshold/prototype" will be introduced to make the scoring scale comparable over time and adjustable when the distribution changes. This paper also presents the overall architecture diagram of the model, as shown in Figure 1.

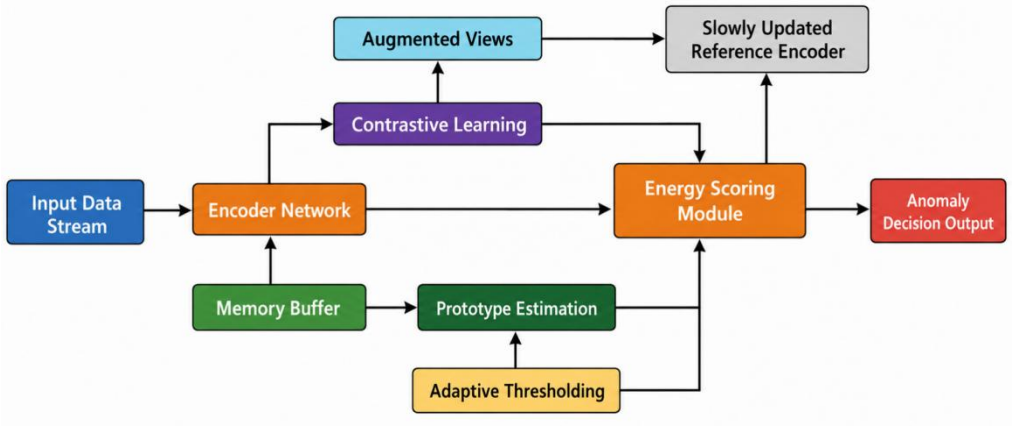


Figure 1. Overall model architecture diagram

2.2 Adaptive data mining and representation learning

To obtain transferable and robust representations without annotations, a lightweight self-supervised contrastive Learning approach is employed: two randomly augmented views $x_t^{(1)}, x_t^{(2)}$ are constructed for the same sample, and their latent representation differences are minimized, thereby depositing the invariant factors of the "normal structure" into the representation space.

$$L_{con} = \|f_\theta(x_t^1) - f_\theta^{(2)}\|_2^2 \quad (2)$$

Simultaneously, a slowly updated reference encoder parameter θ (which can be understood as an online stabilization anchor point) is introduced, and exponential moving averages are used to suppress drastic drift caused by short-term noise.

$$\theta_t = \mu\theta_{t-1} + (1-\mu)\theta_t \quad (3)$$

At the "data mining" level, a small memory set M_t is maintained to store representative representations of recent low-energy samples (more likely normal) to continuously characterize the backbone structure of the current distribution and provide a reference for the energy model to evolve.

The online representation mechanism further draws on selective memory retention and efficient adaptive updates. Key signal-aligned memory compression preserves information that remains discriminative over long contexts [7], while low-rank and soft-gated sparse updates offer a controlled route for adapting model parameters without destabilizing established representations [8]. Structural prompt injection informed by frequent subgraph mining [9] likewise motivates retaining recurring dependency motifs when the normal state evolves. Together, these principles support a compact memory buffer that prioritizes stable normal patterns rather than merely storing recent samples.

2.3 Energy Model Depth Detector

The energy scorer uses a "normal prototype" that is updated over time to define a consistency metric. The current normal center c_t is estimated using the representation in the memory set, and the distance from the sample to that center is used as the energy.

$$c_t = \frac{1}{|M_t|}\sum_{z\in M_t} z \quad (4)$$

$$E(x_t) = \|f_\theta(x_t) - c_t\|_2^2 \quad (5)$$

Here, reference encoder f_θ is used in the scoring to improve scoring stability, while c_t comes from online memory updates of "adaptive data mining," which can follow distribution drift, thus ensuring that the energy function is always sensitive to the "current normal pattern." The higher the energy, the more difficult it is for the sample to be explained by the current normal structure, and therefore the more likely it is to be an anomaly.

The representation space is also organized to capture dependencies that may not be apparent in individual measurements. Heterogeneous dynamic dependency graph learning [10] and self-supervised unified representation learning with structural dependency modeling [11] show how evolving relationships can reveal early risk signals. Collaborative discrimination between reconstruction residuals and latent distribution shifts [12] further supports the separation of persistent abnormal structure from transient noise. Accordingly, the slowly updated reference encoder and low-energy memory jointly act as temporal and structural stabilizers for the energy landscape.

This paper presents an example of the output of the Energy Model Depth Detector, and its experimental results are shown in Figure 2.

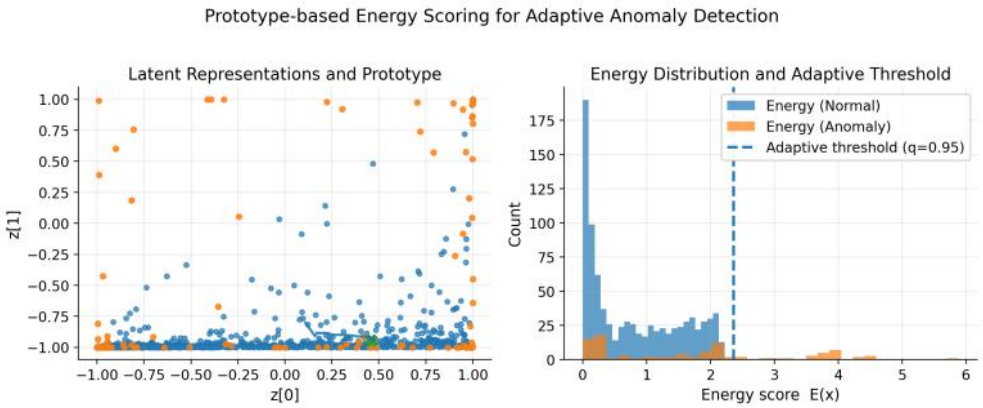


Figure 2 . Example of output from the Energy Model Depth Detector(The data comes from a random setting for demonstration purposes only, not actual experimental results.)

2.4 Online adaptive threshold and joint optimization objective

To prevent the fixed threshold from becoming invalid due to drift, the threshold τ_t is estimated online using the quantiles of the energy distribution and updated smoothly, making the alarm rate controllable and insensitive to short-term fluctuations.

$$\tau_t = (1-\alpha)\tau_{t-1} + \alpha Quantile_q(E(x) \mid x \in B_t) \tag{6}$$

Where A represents the current mini-batch window, and B represents the target quantile. During final training/update, self-supervised consistency, energy constraints, and stability regularization are combined into a single, simple online objective:

$$L = L_{con} + \lambda E_{x \in B_t}[E(x)] + \gamma \|\theta_t - \tilde{\theta}_t\|_2^2 \tag{7}$$

The second term encourages the model to maintain low-energy interpretability for samples "discovered as normal," while the third term restricts rapid parameter drift. Through "memory mining to update c_t and τ_t" and "slowly updated reference representations," the entire detector can achieve adaptive data mining-based anomaly detection under unlabeled, drift, and noise conditions.

For online alarm decisions, the detector also treats evidence accumulation and intervention risk as part of the same control loop. Temporal evidence aggregation with causal-contrastive Transformer learning [13] motivates combining observations across windows before assigning a persistent alarm, while time-varying operational graph reasoning with risk-constrained repair planning [14] supports conservative decisions when the available evidence is incomplete. This perspective complements the smoothed quantile threshold by coupling score calibration with the reliability of the current structural evidence.

3 Implementation Plan

3.1 Dataset

This paper uses the Numenta Anomaly Benchmark (NAB) as a defined open-source dataset to construct a research scenario for anomaly detection in streaming time series. NAB is designed for real-world time series anomaly detection tasks, covering univariate time series data from various sources and

patterns (such as service metrics, sensor readings, and business metrics), and provides anomaly interval labeling and a unified data organization standard, facilitating the reproduction and comparison of different anomaly detection algorithms under the same settings. The data uses timestamp sequences as the basic unit, with samples arranged chronologically, naturally conforming to the settings of online detection and distribution changes over time, making it suitable for studying the core issues of "normal patterns evolving with drift and diverse and sparse anomaly morphologies."

Within this framework, the streaming time series characteristics of NAB can be directly used to simulate continuously arriving input x_t: the model first windows or processes the sequence point by point to obtain a representation z_t, then maintains the "current normal prototype" through online memory, uses energy scores to characterize the consistency of samples with the current normal structure, and implements stable alarms in conjunction with an adaptive threshold mechanism. Because NAB contains multiple types of sequences and statistical property changes over different time periods, it supports the validation and analysis of adaptive data mining strategies: when the normal distribution drifts, the memory and prototype can be updated accordingly, thus maintaining the comparability of energy scales and sensitivity to outliers.

3.2 Experimental setup

This study was implemented and trained on a single machine with a single GPU. The operating system was Ubuntu 22.04 LTS, the development environment was Python 3.10, the deep learning framework was PyTorch 2.2, and the CUDA version was 12.1. The hardware configuration included an NVIDIA RTX 4090 24GB GPU, an Intel Core i9-13900K CPU, and 64GB of RAM. To ensure reproducibility, all randomization processes used a fixed random seed of 42, and the training and inference processes were repeatedly run under the same software version and hardware conditions to maintain consistency.

Regarding hyperparameters, a sliding window was used for the input timing, with a fixed window length of 64 and a stride of 1. Each feature was normalized using a z-score. The encoder employed a 2-layer MLP with tanh, a latent dimension of 128, and Dropout of 0.1. The energy model has a fixed online memory capacity of 1000, retains 0.60 samples from low-energy samples at each step, uses an energy quantile of 0.95 as the threshold, and is updated online with a smoothing coefficient of 0.20. The reference encoder uses EMA updates with a fixed coefficient of 0.95. The optimizer uses AdamW with a fixed learning rate of 0.001, weight decay of 0.0001, a batch size of 256, a fixed number of training epochs of 50, and a gradient clipping threshold of 1.0.

4 Experimental Results and Analysis

This paper first presents the experimental results compared with other models, as shown in Table 2.

Table 2. Comparative experimental results

Method	Precision	Recall	F1-Score	AUC	NAB Score
Anomaly Transformer (Xu et al.)[15]	0.842	0.811	0.826	0.945	71.2
TranAD (Tuli et al.)[16]	0.873	0.829	0.850	0.962	76.4
DCdetector (Yang et al.)[17]	0.861	0.803	0.831	0.953	73.9
USAD (Audibert et al.)[18]	0.792	0.755	0.773	0.921	65.1
OmniAnomaly (Su et al.)[19]	0.814	0.768	0.790	0.934	68.3

Donut (Xu et al.)[20]	0.781	0.742	0.761	0.912	63.7
DSEBM (Zhai et al.)[21]	0.764	0.701	0.731	0.898	59.4
Ours	0.901	0.873	0.887	0.975	82.6

The overall comparison reveals that the framework combining energy scoring and adaptive data mining exhibits greater stability and balance across various metrics. This advantage is evident not only in discrimination accuracy but also in its adaptability to changes in anomalous sparsity and normal patterns. Compared to approaches that rely solely on fixed discrimination boundaries or reconstruction errors, energy metrics based on normal prototypes directly map the consistency between samples and the current normal structure into continuously comparable anomaly scores, ensuring the model maintains a consistent discrimination scale across different time periods and states. Simultaneously, the slow update mechanism of online memory and the reference encoder help absorb the gradual drift of normal distributions into the prototype and representation space, thereby reducing the accumulation of false alarms caused by threshold failures or short-term noise fluctuations, aligning with the online adaptive anomaly detection objective emphasized in this paper.

Furthermore, the results in the table also demonstrate that this method achieves a better trade-off between accuracy and coverage, which is closely related to the unified scoring mechanism proposed in this paper. The energy model scores based on the current normal pattern, eliminating the need for explicit enumeration of anomaly types, thus exhibiting greater generalization ability when faced with diverse anomaly patterns. Meanwhile, self-supervised contrastive Learning enhances the robustness of the representation space to noise and missing data, making it easier to aggregate normal samples and separate anomalous samples, thereby improving overall detection reliability. In summary, these comparative results support the core argument of this paper: by maintaining a dynamic normal structure through adaptive data mining and using an energy function for calibrable anomaly measurement, both stronger detection performance and more robust cross-time generalization ability can be achieved in streaming scenarios.

The learning rate directly affects the update magnitude of the optimization process, thereby altering the model's stability in fitting normal patterns and its discrimination boundary for outliers. To test the robustness of our proposed method under different learning rate settings, we need to systematically perturb the learning rate and observe the trend of Precision. The experimental results are shown in Figure 3.

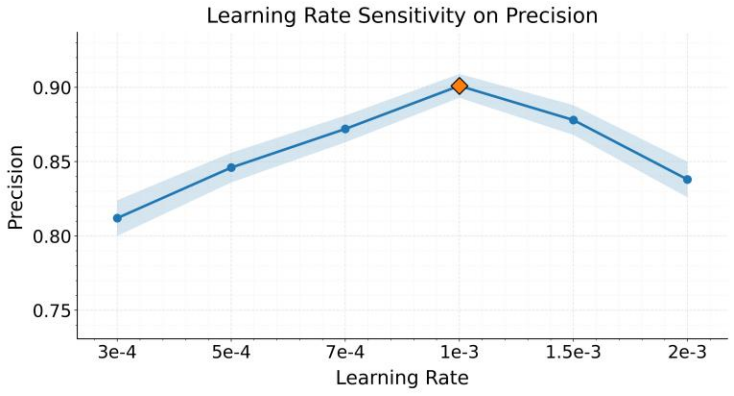


Figure 3. Sensitivity experiment of learning rate to precision

The trend in the figure shows that the proposed method has a relatively clear effective range for the learning rate. Within this range, precision gradually improves as the learning rate increases, but declines again when the learning rate continues to increase. This indicates that the learning rate not only affects the convergence speed of optimization but also directly impacts the stability of the representation space and the consistency of the energy scoring scale. A small learning rate makes the model updates too conservative, and the normal structures depicted by the encoder and online memory struggle to keep up with the slow drift of the data distribution. This results in insufficient adaptation of the energy score to newly emerging normal patterns, making it easier to misclassify boundary samples as anomalous or reduce discriminative purity.

On the other hand, the decrease in precision when the learning rate is large is more consistent with the stability constraint logic of the online adaptive mechanism in this framework. More aggressive updates amplify the impact of short-term noise and local perturbations, making the estimates of the reference encoder and normal prototypes more prone to jitter. This leads to insufficient smoothness in the scale of the energy distribution and adaptive threshold between adjacent windows, resulting in unnecessary fluctuations in anomalous scores. This sensitivity result indirectly supports the robust adaptation principle emphasized in this paper, which requires a balance between the response speed of following the drift and the stability of the representation in order to enable the energy model to maintain a high level of reliability in anomaly detection in streaming scenarios.

5 Conclusion

This paper addresses the core challenges of streaming time-series anomaly detection, such as insufficient labeling, distribution drift, and diverse anomaly morphologies, proposing an adaptive data mining and energy model-based deep detection framework. This framework utilizes self-supervised contrastive Learning to obtain robust representations and maintains the evolving normal structure through online memory. It then constructs a unified, continuous, and comparable anomaly metric using energy scores. Compared to traditional approaches relying on fixed thresholds or static discrimination boundaries, this method emphasizes continuously characterizing the current normal pattern in an open environment, shifting anomaly detection from one-time modeling to a sustainable structural update process. This provides a more stable risk perception capability and a more controllable alarm mechanism for long-term operating systems.

Comparative experimental results demonstrate that the proposed method exhibits stronger overall performance and consistency across key evaluation dimensions, further validating the effectiveness of combining data mining-based structural maintenance with energy modeling. The energy score, referencing a normal prototype, naturally adapts to the reality of unenumerable anomaly types. The online memory and slowly updated reference representation jointly ensure the stability of the scoring scale, enabling the model to maintain sensitivity to anomaly deviations and tolerance for normal changes even when data statistical characteristics change. Therefore, this paper not only presents an executable anomaly detection scheme but also provides a unified modeling perspective: centering on dynamic normal structures and achieving cross-time and cross-scenario anomaly identification through calibrable energy metrics.

At the application level, this work has direct significance for fields such as intelligent cloud service operation and maintenance, industrial process monitoring, IoT sensor networks, financial transaction risk control, and medical monitoring. These scenarios generally feature continuous data arrival, sparse and costly anomalies, and constantly changing system states. Traditional methods relying on manual rules and static thresholds are costly to maintain and prone to failure. The framework presented in this paper maintains a characterization of the latest normal patterns through adaptive data mining and supports alarm decisions with a unified energy score, helping to reduce operational noise and manpower burden caused by false alarms, and improving the ability to identify early weak anomalies and the level of business continuity assurance. More importantly, the framework has modular expansion space, which can be easily integrated with existing monitoring platforms, log analysis systems, and alarm orchestration processes, providing a general algorithmic foundation for building a sustainably evolving anomaly detection system.

Looking ahead, further research can focus on three directions: large-scale multi-source heterogeneous data fusion, more refined drift characterization and adaptive strategies, and enhanced interpretability and controllability. On one hand, energy modeling can be extended from a single prototype to multiple prototypes or hierarchical structures to better cover the normal operating modes of complex systems under various conditions. On the other hand, a more systematic drift detection and update triggering mechanism can be introduced to achieve better response and a stable balance between memory maintenance and threshold updates under different change rates. Furthermore, combining anomaly scoring with causal clues, event correlation, or interpretable attribution mechanisms holds promise for further advancing anomaly detection into a closed-loop governance system that is localizable, diagnosable, and interventionable, thereby generating a broader and longer-term practical impact on critical infrastructure and highly reliable business systems.

References

[1] Z. Z. Darban, et al., "CARLA: Self-supervised contrastive representation learning for time series anomaly detection", Pattern Recognition, vol. 157, Art. no. 110874, 2025.

[2] Y. Jeong, et al., "AnomalyBERT: Self-supervised transformer for time series anomaly detection using data degradation scheme", arXiv preprint arXiv:2305.04468, 2023.

[3] J. Zhu, et al., "METER: A dynamic concept adaptation framework for online anomaly detection", arXiv preprint arXiv:2312.16831, 2023.

[4] Q. Shentu, et al., "Towards a general time series anomaly detector with adaptive bottlenecks and dual adversarial decoders", arXiv preprint arXiv:2405.15273, 2024.

[5] W. Li, et al., "StackVAE-G: An efficient and interpretable model for time series anomaly detection", AI Open, vol. 3, pp. 101-110, 2022.

[6] Q. Liu and J. Paparrizos, "The elephant in the room: Towards a reliable time-series anomaly detection benchmark", Advances in Neural Information Processing Systems, vol. 37, pp. 108231-108261, 2024.

[7] J. Zhang, "Learning what to remember: Key signal-aligned memory compression for long-context LLMs", 2024.

[8] B. Chen, "Adaptive fusion of low-rank and soft-gated sparse updates for parameter-efficient instruction tuning", 2024.

[9] Z. Xiao, C. Moretti and Q. Li, "Frequent subgraph mining and structural prompt injection for large language model fine-tuning", 2025.

[10] F. Chen, "Heterogeneous dynamic dependency graph learning for anomaly detection in large-scale backend systems", 2024.

[11] T. Ying and C. Ding, "Self-supervised unified representation learning with structural dependency modeling for early risk detection in distributed systems", 2024.

[12] Z. Liu and D. Abdurehim, "Distributed system anomaly detection via collaborative discrimination of reconstruction residuals and latent distribution shifts", 2025.

[13] Z. Hu, "Temporal evidence aggregation with causal-contrastive transformer learning for intelligent cloud failure diagnosis", 2024.

[14] Y. Tang, "Time-varying operational graph reasoning and risk-constrained repair planning for tool-augmented large language model agents", 2024.

[15] J. Xu, et al., "Anomaly transformer: Time series anomaly detection with association discrepancy", arXiv preprint arXiv:2110.02642, 2021.

[16] S. Tuli, G. Casale and N. R. Jennings, "TranAD: Deep transformer networks for anomaly detection in multivariate time series data", arXiv preprint arXiv:2201.07284, 2022.

[17] Y. Yang, et al., "DCdetector: Dual attention contrastive representation learning for time series anomaly detection", Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023.

[18] J. Audibert, et al., "USAD: Unsupervised anomaly detection on multivariate time series", Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020.

[19] Y. Su, et al., "Robust anomaly detection for multivariate time series through stochastic recurrent neural network", Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019.

[20] H. Xu, et al., "Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications", Proceedings of the 2018 World Wide Web Conference, 2018.

[21] S. Zhai, et al., "Deep structured energy based models for anomaly detection", Proceedings of the International Conference on Machine Learning, PMLR, 2016.