

Article

Not peer-reviewed version

Graph-Transformer Reconstruction Learning for Unsupervised Anomaly Detection in Dependency-Coupled Systems

Chong Zhang , [Chihui Shao](#) , [Junjie Jiang](#) , [Yinan Ni](#) , [Xiaoxuan Sun](#) *

Posted Date: 31 December 2025

doi: [10.20944/preprints202512.2788.v1](https://doi.org/10.20944/preprints202512.2788.v1)

Keywords: unsupervised anomaly detection; dependency graph modeling; temporal self-attention; reconstruction error scoring



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Graph-Transformer Reconstruction Learning for Unsupervised Anomaly Detection in Dependency-Coupled Systems

Chong Zhang ¹, Chihui Shao ², Junjie Jiang ³, Yinan Ni ⁴ and Xiaoxuan Sun ^{5,*}

¹ Carnegie Mellon University, Pittsburgh, USA

² Duke University, Durham, USA

³ Illinois Institute of Technology, Chicago, USA

⁴ University of Illinois at Urbana-Champaign, Urbana, USA

⁵ Independent Researcher, Mountain View, USA

* Correspondence: sunxiaox@usc.edu

Abstract

To address the practical challenges of diverse anomaly patterns, strongly coupled dependencies, and high labeling costs in large-scale complex infrastructures, this paper presents an unsupervised anomaly detection method that integrates graph neural networks with Transformer models. The approach learns normal system behavior and identifies deviations without relying on anomaly labels. Infrastructure components are abstracted as nodes in a dependency graph, where nodes are characterized by multiple source observability signals. A graph encoder aggregates neighborhood information to produce structure-enhanced node representations. Self-attention mechanisms are introduced along the temporal dimension to capture long-range dynamic dependencies. This design enables joint modeling of structural relations and temporal evolution. A reconstruction-based training strategy is adopted to constrain the learning of normal patterns. Reconstruction error is used to derive anomaly scores for detection. To ensure reproducibility and ease of deployment, complete specifications of data organization, training procedures, and key hyperparameter settings are provided. Comparative experiments on public benchmarks demonstrate overall advantages across multiple evaluation metrics and confirm the effectiveness of the proposed framework in representing anomaly propagation and temporal drift characteristics in complex systems.

Keywords: unsupervised anomaly detection; dependency graph modeling; temporal self-attention; reconstruction error scoring

I. Introduction

With the widespread adoption of cloud computing, microservice architectures, and distributed systems, modern information infrastructures continue to evolve toward large scale, heterogeneity, and high dynamism. Complex and tight dependencies have emerged among computation, storage, networking, and application services[1,2]. As a result, system states exhibit strong nonlinearity and coupling. Under such conditions, anomalies can rapidly propagate along dependency paths once they occur. This propagation may trigger cascading failures and lead to severe service disruptions and business losses. Therefore, timely and accurate identification of potential anomalies in complex large-scale infrastructures has become a critical issue for ensuring system stability and advancing intelligent operations and maintenance.

Most traditional anomaly detection approaches rely on manually defined rules or single metric thresholds to monitor system states through local and static judgments. These methods remain partially effective when the system scale is limited and operational patterns are relatively stable. However, they struggle in realistic environments characterized by high dynamics and heterogeneous

data sources[3]. On one hand, rule-based methods depend heavily on expert knowledge, incur high maintenance costs, and show poor transferability. On the other hand, methods based on independent time series often ignore structural relationships among components. They fail to capture how anomalies propagate within the system. This limitation constrains their ability to recognize complex failure patterns[4].

Advances in monitoring technologies enable modern infrastructures to continuously generate large volumes of observability signals. These include performance metrics, execution logs, and distributed traces. Such data reflect system states from multiple perspectives and provide a richer foundation for anomaly detection[5]. However, multi-source observability data differ substantially in temporal resolution, semantic granularity, and noise characteristics. Simple feature concatenation or isolated modeling is insufficient to exploit their latent correlations and dynamic patterns. In unsupervised settings, the lack of reliable anomaly labels further increases modeling difficulty. This situation calls for a unified modeling paradigm that can automatically learn intrinsic operational patterns while remaining sensitive to deviations.

From a system-level perspective, large-scale infrastructures can be abstracted as complex networks composed of functional components and their dependency relationships. Anomalies are rarely isolated events. They are often embedded as structural changes or temporal shifts within dependency graphs. Modeling individual components solely along the time dimension is inadequate for identifying anomalies caused by cross-component interactions[6]. Focusing only on structural relations while ignoring temporal evolution also fails to reflect how system states accumulate and change over time. Consequently, jointly modeling structural dependencies and temporal dynamics is essential for improving the accuracy and anticipatory capability of anomaly detection.

Against this backdrop, developing an unsupervised anomaly detection approach that simultaneously captures dependency graph structure and temporal dynamics is of significant theoretical and practical value. Such methods can overcome the limitations of threshold and rule-driven techniques by enabling adaptive modeling of complex system behaviors and more expressive risk representations. Moreover, early detection and precise localization of anomalous behaviors can effectively reduce the impact of fault propagation and improve infrastructure reliability and operational efficiency[7]. As system scale continues to expand and operational environments grow increasingly complex, exploring anomaly detection frameworks that integrate structure awareness with temporal modeling remains a crucial direction for intelligent operations and highly reliable system construction.

II. Background

Anomaly detection for cloud and microservice environments has been extensively studied, with surveys emphasizing the diversity of anomaly types, the heterogeneity of observability signals, and the difficulty of achieving reliable detection under evolving workloads. Early systematic reviews summarize common modeling paradigms and deployment constraints in cloud environments, highlighting that practical methods must balance detection accuracy with scalability and noise tolerance [8]. More recent surveys further stress that modern service-based systems require unified analysis across metrics, logs, and traces, and that dependency-aware modeling is crucial for identifying fault propagation and supporting root cause localization [9]. In addition to survey literature, recent applied work continues to reflect the breadth of anomaly detection designs in cloud settings, reinforcing the need for robust representation learning to handle distribution shift and complex operational patterns [10].

Methodologically, deep sequence modeling provides a strong basis for learning normal temporal behavior in an unsupervised or weakly supervised manner. Attention-augmented recurrent models demonstrate that selectively focusing on informative time segments can improve anomaly sensitivity while mitigating noise effects [11], and attention-based temporal modeling has also been used to capture long-range dependencies for risk-related prediction in complex time series [12]. Beyond purely temporal encoders, recent work introduces contrastive learning to strengthen

dependency representations by pulling consistent patterns together and separating abnormal behaviors in embedding space, which can improve robustness under changing system conditions [13]. These directions motivate combining temporal self-attention with reconstruction-style objectives to learn stable normality manifolds for anomaly scoring.

Another important trend is incorporating structure and explicit dependencies into the detection pipeline. Dynamic graph modeling and relational representation learning have been adopted to encode interactions among entities and to improve robustness under evolving dependencies, illustrating how graph-based inductive bias can complement temporal modeling when anomalies propagate across components [14]. Causal graph modeling and causally constrained representation learning further emphasize suppressing spurious correlations and improving explanation faithfulness through structured constraints, providing transferable methodology for making anomaly decisions more interpretable and traceable [15]. Complementarily, meta-learning has been explored to handle sample scarcity and evolving patterns, offering principles for rapid adaptation when anomaly characteristics shift over time [16]. Together, these works support the perspective that effective anomaly detection in complex infrastructures benefits from jointly modeling temporal evolution, dependency structure, and robustness-oriented representation learning.

Finally, recent advances in large language model ecosystems provide additional tools for reliability, interpretability, and privacy-aware deployment. Knowledge-augmented agent frameworks highlight mechanisms for integrating external structured knowledge into decision pipelines to improve explainability and controllability [17]. Risk-aware and uncertainty-aware summarization focuses on compressing evolving contexts while preserving critical signals, which can support efficient state summarization and alert reporting in monitoring systems [18]. Parameter-efficient fine-tuning methods, including privacy-preserving adaptation, demonstrate how to update large models with constrained parameter changes and privacy guarantees, offering a methodological blueprint for adapting components in production settings without full retraining or excessive data exposure [19]. Related studies on collaborative evolution in complex multi-component systems further emphasize coordinated adaptation and robustness as conditions evolve [20], while multi-scale LoRA fine-tuning illustrates lightweight adaptation across different granularities that can be useful when tuning auxiliary models for monitoring and diagnosis tasks [21].

III. Method

In complex infrastructure environments, a system can be abstracted as a dynamic graph structure composed of multiple components and their dependencies. Let the system's dependencies at time t be represented as a directed graph $G_t = (V, E_t)$, where V represents the set of nodes corresponding to service or resource components in the system, and E_t represents the set of dependency edges at time t . Each node $v_i \in V$ is associated with a set of multi-source observable signals, represented as a time series $x_i = \{x_i^1, x_i^2, \dots, x_i^T\}$. Through this graph-sequence joint representation, the system's structural dependencies and temporal evolution are modeled in a unified manner, providing a structured input foundation for subsequent anomaly modeling. This article presents the overall model architecture diagram, as shown in Figure 1.

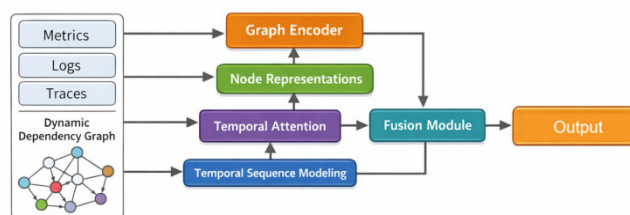


Figure 1. Overall model architecture diagram.

To effectively capture the dependencies and structural context among nodes, this study introduces and implements a graph structure modeling mechanism at each time step, which performs neighborhood aggregation on node features to encode local and global relationships. This aggregation process draws on the temporal-aware graph neural architecture proposed by Zhang et al. [22], which demonstrates the advantages of dynamically integrating structural context for comprehensive anomaly detection in microservices. Building on this, the model further leverages spatiotemporal aggregation strategies introduced by Qiu et al. [23] to jointly represent spatial dependencies and temporal correlations within distributed backend systems. To enhance adaptability in large-scale and dynamic environments, the neighborhood aggregation scheme incorporates collaborative feature propagation mechanisms inspired by Yao, Liu, and Dai [24], ensuring robust representation even under shifting topology and heterogeneous data sources. The formal aggregation operation is defined as follows. For node v_i , its structure-aware representation is obtained by weighted aggregation of neighboring node information, formally represented as:

$$h_i^t = \sigma(\sum_{j \in N(i)} \alpha_{ij} W_g x_j^t) \quad (1)$$

Where $N(i)$ represents the set of neighbors of node v_i , W_g is a learnable linear mapping matrix, and $\sigma(\cdot)$ represents a non-linear activation function. The weight coefficient α_{ij} is used to characterize the strength of the dependency between nodes, and it is defined as:

$$\alpha_{ij} = \frac{\exp(a^T [x_i^t || x_j^t])}{\sum_{k \in N(i)} \exp(a^T [x_i^t || x_k^t])} \quad (2)$$

Here, a represents a learnable parameter, and $||$ represents a vector concatenation operation. Through this structural modeling process, node representations can explicitly incorporate contextual information from the dependency graph.

After obtaining the structurally enhanced node representations, to further capture the evolution of the system's operating state over time, a time modeling module based on a self-attention mechanism is introduced to model the dynamic behavior of nodes in the time dimension. For node v_i , its time series representation is denoted as $H_i = [h_i^1, \dots, h_i^T]$. The time attention mechanism calculates the hidden state in the following form:

$$Z_i = \text{softmax}(\frac{Q_i K_i^T}{\sqrt{d}}) V_i \quad (3)$$

Where Q_i, V_i, K_i is the learnable parameter matrix, and d represents the feature dimension. This process can adaptively focus on the importance of different time segments to the current state, thereby modeling long-term time dependencies and potential anomalous evolution trends.

In unsupervised anomaly detection scenarios, the model identifies deviations by learning the normal operating patterns of the system. To this end, a reconstruction-based objective function is introduced to constrain the joint structural and temporal representations. The reconstruction error of the node v_i at time t is defined as:

$$e_i^t = ||x_i^t - \hat{x}_i^t||_2^2 \quad (4)$$

Where \hat{x}_i^t represents the model's reconstruction result of the original input. The overall optimization objective function is defined as:

$$L = \sum_{i \in V} \sum_{t=1} e_i^t \quad (5)$$

This objective prompts the model to fully characterize the system's normal structure and temporal patterns during training. When significant shifts occur in the system's operational state in terms of structural relationships or temporal behavior, the corresponding reconstruction error will change, thus providing a unified and interpretable metric basis for anomaly detection.

IV. Experimental Results

A. Dataset

This study employs the Server Machine Dataset, SMD, as the experimental dataset. SMD is an open-source multivariate time series dataset designed for anomaly detection in cloud server

monitoring scenarios. It captures state fluctuations and anomalous segments observed during long-term operation of real servers. The dataset effectively reflects metric-centered observability signals in large-scale infrastructures. Due to its public availability and strong reproducibility, SMD is well-suited as a unified evaluation benchmark for unsupervised anomaly detection. It is also highly aligned with the emphasis on temporal dynamic modeling in this work.

The dataset consists of monitoring metric sequences collected from multiple servers. Each server corresponds to a multivariate time series with dozens of dimensions. These dimensions typically include statistics related to CPU, memory, disk, and network usage. Point-wise anomaly annotations are provided for evaluation purposes. This design supports a consistent setting where training is unsupervised, and evaluation is supervised. The data are organized at the machine level and include separate training and testing splits. The training data are mainly used to learn normal operating patterns. The testing data contain more complex fluctuations and anomalous intervals, which allow assessment of model sensitivity and robustness to deviations.

To better support the joint modeling paradigm of dependency graphs and temporal dynamics, each server is treated as a node in a graph. The multivariate metric sequences within a temporal window are used as node features. Dependency edges are constructed based on statistical dependence or correlation across machines. Edge weights can be updated using sliding windows to form a dynamic dependency graph. In this way, the model can exploit graph structure to capture cross-node interaction patterns. It can also leverage temporal modeling to describe the evolution of both local and global system states. This design enables full support of the proposed methodology on a standardized and open dataset.

B. Experimental Setup

An unsupervised training paradigm is adopted for anomaly detection modeling in this study. During training, only normal sequences from the training set are used, and no anomaly labels are introduced. The model input consists of multivariate monitoring sequences segmented by sliding windows, together with a dependency graph constructed from statistical dependencies as a structural prior. At each time step, a graph encoder integrates neighborhood information to obtain structure-enhanced node representations. A temporal attention module is then applied to learn long-range temporal dependencies. The reconstruction network outputs reconstructed observations, and the reconstruction error is used as the training objective for end-to-end optimization. The AdamW optimizer is employed with a learning rate of 0.0001 and a batch size of 64. The window length is set to 100, the hidden dimension to 128, the weight decay to 0.01, and the dropout rate to 0.1. An early stopping strategy is applied when the validation reconstruction loss no longer decreases. A thresholding strategy is then used to map anomaly scores to anomaly decisions. This process enables the detection of deviations in operating states without relying on anomaly annotations.

C. Experimental Results

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

Table 1. Comparative experimental results.

Model	Acc(%)	Precision(%)	Recall(%)	F1-Score(%)
Glad[25]	47.3	44.8	42.1	43.4
Mambaad[26]	51.6	49.7	46.8	48.2
Dinomaly[27]	53.9	52.4	49.1	50.7
Geo-Hgan[28]	55.8	54.1	52.7	53.4
Autouad[29]	56.7	55.2	53.0	54.1
ADS-Bpois[30]	58.4	56.8	55.1	55.9

Ours	62.1	61.0	60.3	60.6
------	------	------	------	------

Baseline methods show incremental but uneven gains on complex infrastructure anomaly detection, reflecting the limits of single-perspective modeling for multifactor disturbances and cascading failures. In contrast, the proposed method leads across all metrics, achieving a more balanced precision–recall trade-off and stable discrimination by jointly modeling dependency graphs and temporal dynamics—aligning with how anomalies propagate and evolve in real systems. This yields more separable and threshold-robust anomaly scores, which is critical for unsupervised deployment and early warning. Finally, since representation capacity affects the ability to encode structure and long-range dynamics, we assess sensitivity to hidden dimension with other settings fixed; results are shown in Figure 2.

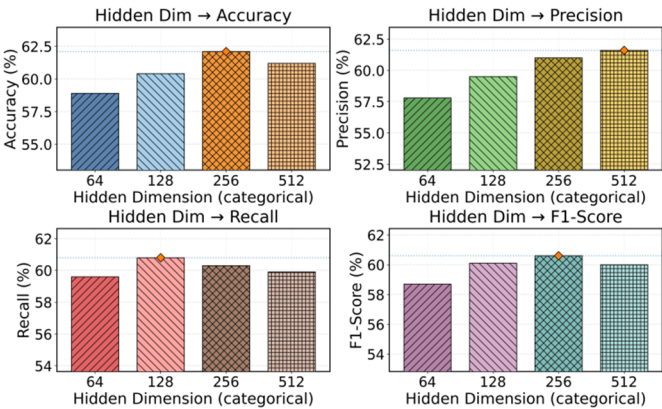


Figure 2. The impact of hidden dimensions on experimental results.

The subplots show that hidden dimension directly controls representational capacity and thus discrimination and class separation. Small dimensions over-compress structure and temporal cues, weakening cross-component interactions and blurring normal-behavior boundaries, which reduces sensitivity to subtle anomalies. A moderate dimension yields the best balance, allowing graph structure and temporal attention to jointly encode anomaly propagation and drift while suppressing noise, producing more stable unsupervised representations. Larger dimensions offer sharper discrimination for some patterns but increase flexibility and noise sensitivity, leading to overfitting and degraded coverage. Overall, hidden dimension is a critical hyperparameter requiring a trade-off between expressiveness and robustness; Figure 3 further examines the effect of time-window length.

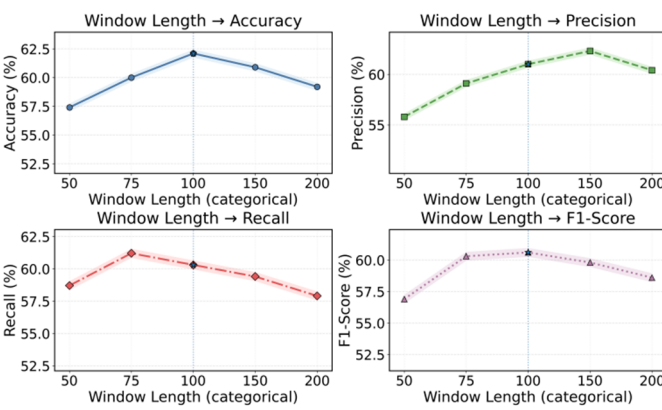


Figure 3. The effect of time window length on experimental results.

The subplots show that temporal window length strongly shapes anomaly detection: short windows focus on local fluctuations and miss delayed cross-component effects, making decisions

sensitive to noise, while moderate windows provide sufficient context for graph-based dependencies and temporal attention to capture anomaly propagation and cumulative drift. However, overly long windows introduce misaligned trends, workload cycles, and irrelevant fluctuations that dilute representations or cause over-smoothing under unsupervised reconstruction, creating a trade-off between coverage and precision. These results indicate that window length must be carefully aligned with dependency graph modeling—large enough to capture anomaly evolution paths, but not so large that noise dominates—to ensure robust early warning and reliability.

V. Conclusions

This work addresses the demand for unsupervised anomaly detection in large-scale and complex infrastructures. It proposes a unified modeling framework that integrates dependency graph structure with temporal dynamics to better reflect coupled propagation and evolution of anomalies in real systems. Structural context and temporal dependencies are jointly exploited at the representation level. This enables the model to learn clearer boundaries of normal behavior without anomaly labels and to map deviations into interpretable anomaly scores. Comparative results demonstrate overall advantages across multiple evaluation metrics. These findings indicate that coordinated design of graph structure awareness and long-range temporal modeling improves coverage of complex failure patterns and enhances decision stability. The framework provides a more transferable technical path for reliable detection in practical operations. From an application perspective, the proposed approach has direct relevance to cloud platforms, microservice systems, data center operations, and critical information infrastructures. Earlier and more robust anomaly identification can shorten fault discovery and response chains. It can reduce the cost of risk propagation caused by downtime and cascading failures. Service availability and resource utilization can be improved. Joint structural and temporal representations align naturally with system dependency relations as operational facts. They provide more reliable signals for alert aggregation, impact assessment, and troubleshooting decisions. In deployment, this unsupervised approach also offers low labeling cost and fast adaptation. It supports long-term usability in environments with multiple services, regions, and frequent version iterations. This contributes to the evolution of intelligent operations toward higher automation and reliability.

Looking ahead, several directions remain to further enhance generality and engineering readiness. First, unified representation learning can be extended to more complex observability modalities. Metrics, logs, and traces can be aligned more closely within a single framework to strengthen consistent modeling of multi-source anomaly cues. Second, stronger online updating and drift adaptation mechanisms can be incorporated. This helps maintain a stable normal pattern characterization under changing business policies, elastic scaling, and system upgrades. With these extensions, the proposed joint structural and temporal unsupervised detection paradigm can be applied to a broader range of critical infrastructures and continue to enhance system resilience and operational reliability.

References

1. J. Chen, F. Liu, J. Jiang *et al.*, "TraceGra: A Trace-Based Anomaly Detection for Microservice Using Graph Deep Learning," *Computer Communications*, vol. 204, pp. 109–117, 2023.
2. 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.
3. Z. Yu, C. Pei, X. Wang *et al.*, "Pre-Trained KPI Anomaly Detection Model through Disentangled Transformer," *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 6190–6201, 2024.
4. Z. Liu, X. Huang, J. Zhang *et al.*, "Multivariate Time-Series Anomaly Detection Based on Enhancing Graph Attention Networks with Topological Analysis," *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pp. 1555–1564, 2024.

5. H. Kang and P. Kang, "Transformer-Based Multivariate Time Series Anomaly Detection Using Inter-Variable Attention Mechanism," *Knowledge-Based Systems*, vol. 290, p. 111507, 2024.
6. L. Xu, K. Xu, Y. Qin *et al.*, "TGAN-AD: Transformer-Based GAN for Anomaly Detection of Time Series Data," *Applied Sciences*, vol. 12, no. 16, p. 8085, 2022.
7. J. Kim, H. Kang and P. Kang, "Time-Series Anomaly Detection with Stacked Transformer Representations and 1D Convolutional Network," *Engineering Applications of Artificial Intelligence*, vol. 120, p. 105964, 2023.
8. T. Hagemann and K. Katsarou, "A Systematic Review on Anomaly Detection for Cloud Computing Environments," *Proceedings of the 2020 3rd Artificial Intelligence and Cloud Computing Conference*, pp. 83–96, 2020.
9. J. Soldani and A. Brogi, "Anomaly Detection and Failure Root Cause Analysis in (Micro) Service-Based Cloud Applications: A Survey," *ACM Computing Surveys*, vol. 55, no. 3, pp. 1–39, 2022.
10. C. Nwachukwu, K. Durodola-Tunde and C. Akwiwu-Uzoma, "AI-Driven Anomaly Detection in Cloud Computing Environments," *International Journal of Science and Research Archive*, vol. 13, no. 2, pp. 692–710, 2024.
11. J. Li, Q. Gan, Z. Liu, C. Chiang, R. Ying and C. Chen, "An Improved Attention-Based LSTM Neural Network for Intelligent Anomaly Detection in Financial Statements," 2025.
12. Q. Xu, W. Xu, X. Su, K. Ma, W. Sun and Y. Qin, "Enhancing Systemic Risk Forecasting with Deep Attention Models in Financial Time Series," *Proceedings of the 2025 2nd International Conference on Digital Economy, Blockchain and Artificial Intelligence*, pp. 340–344, 2025.
13. Y. Xing, Y. Deng, H. Liu, M. Wang, Y. Zi and X. Sun, "Contrastive Learning-Based Dependency Modeling for Anomaly Detection in Cloud Services," *arXiv preprint, arXiv:2510.13368*, 2025.
14. C. F. Chiang, D. Li, R. Ying, Y. Wang, Q. Gan and J. Li, "Deep Learning-Based Dynamic Graph Framework for Robust Corporate Financial Health Risk Prediction," 2025.
15. J. Lai, C. Chen, J. Li and Q. Gan, "Explainable Intelligent Audit Risk Assessment with Causal Graph Modeling and Causally Constrained Representation Learning," 2025.
16. H. Fan, Y. Yi, W. Xu, Y. Wu, S. Long and Y. Wang, "Intelligent Credit Fraud Detection with Meta-Learning: Addressing Sample Scarcity and Evolving Patterns," 2025.
17. Q. Zhang, Y. Wang, C. Hua, Y. Huang and N. Lyu, "Knowledge-Augmented Large Language Model Agents for Explainable Financial Decision-Making," *arXiv preprint, arXiv:2512.09440*, 2025.
18. S. Pan and D. Wu, "Trustworthy Summarization via Uncertainty Quantification and Risk Awareness in Large Language Models," *arXiv preprint, arXiv:2510.01231*, 2025.
19. Y. Huang, Y. Luan, J. Guo, X. Song and Y. Liu, "Parameter-Efficient Fine-Tuning with Differential Privacy for Robust Instruction Adaptation in Large Language Models," *arXiv preprint, arXiv:2512.06711*, 2025.
20. Y. Li, S. Han, S. Wang, M. Wang and R. Meng, "Collaborative Evolution of Intelligent Agents in Large-Scale Microservice Systems," *arXiv preprint, arXiv:2508.20508*, 2025.
21. H. Zhang, L. Zhu, C. Peng, J. Zheng, J. Lin and R. Bao, "Intelligent Recommendation Systems Using Multi-Scale LoRA Fine-Tuning and Large Language Models," 2025.
22. Q. Zhang, N. Lyu, L. Liu, Y. Wang, Z. Cheng and C. Hua, "Graph Neural AI with Temporal Dynamics for Comprehensive Anomaly Detection in Microservices," *arXiv preprint, arXiv:2511.03285*, 2025.
23. Z. Qiu, F. Liu, Y. Wang, C. Hu, Z. Cheng and D. Wu, "Spatiotemporal Traffic Prediction in Distributed Backend Systems via Graph Neural Networks," *arXiv preprint, arXiv:2510.15215*, 2025.
24. G. Yao, H. Liu and L. Dai, "Multi-Agent Reinforcement Learning for Adaptive Resource Orchestration in Cloud-Native Clusters," *Proceedings of the 2nd International Conference on Intelligent Computing and Data Analysis*, pp. 680–687, 2025.
25. H. Yao, M. Liu, Z. Yin *et al.*, "GLAD: Towards Better Reconstruction with Global and Local Adaptive Diffusion Models for Unsupervised Anomaly Detection," *Proceedings of the European Conference on Computer Vision*, pp. 1–17, 2024.
26. H. He, Y. Bai, J. Zhang *et al.*, "MambaAD: Exploring State Space Models for Multi-Class Unsupervised Anomaly Detection," *Advances in Neural Information Processing Systems*, vol. 37, pp. 71162–71187, 2024.

27. J. Guo, S. Lu, W. Zhang *et al.*, "DiNomaly: The Less Is More Philosophy in Multi-Class Unsupervised Anomaly Detection," Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 20405–20415, 2025.
28. L. Ding, B. Chen, Y. Zhu *et al.*, "Geo-HGAN: Unsupervised Anomaly Detection in Geochemical Data via Latent Space Learning," *Computers & Geosciences*, vol. 192, p. 105703, 2024.
29. W. Dai and J. Fan, "AutoUAD: Hyper-Parameter Optimization for Unsupervised Anomaly Detection," Proceedings of the Thirteenth International Conference on Learning Representations, 2024.
30. P. Luo, B. Wang, J. Tian *et al.*, "ADS-BPoIs: Poisoning Attacks against Deep Learning-Based Air Traffic ADS-B Unsupervised Anomaly Detection Models," *IEEE Internet of Things Journal*, 2024.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.