From Reconstruction to Generation: A Survey of Multivariate Time Series Anomaly Detection

Danyang Cao*, Hanyu Cui, Jiayi Fu

School of Artificial Intelligence and Computer Science, North China University of Technology, Beijing 100043, China

*Corresponding author's e-mail: ufocdy@163.com

Abstract. Multivariate time series (MTS) anomaly detection is a core task for ensuring system reliability in industrial AIOps. Traditional methods typically rely on reconstruction or prediction errors, which often fail to capture complex data distributions. This paper reviews the transition of detection paradigms from discriminative to generative approaches. We categorize existing techniques into four groups: prediction, reconstruction, explicit relationship modeling, and generative frameworks. Specifically, we analyze the capability of Graph Neural Networks (GNNs) in capturing variable dependencies and discuss the trade-offs of diffusion models in distribution fitting. Additionally, we evaluate modern backbones, such as ModernTCN, regarding their feature extraction efficiency, and benchmark these paradigms across standard datasets. Recent research has increasingly focused on generative models, particularly diffusion models, due to their superior performance in modeling complex data distributions.

Keywords: Multivariate Time Series; Anomaly Detection; Diffusion Models; Graph Neural Networks; Large Language Models; Contrastive Learning.

1 Introduction

The growth of the Industrial Internet of Things (IIoT) has produced extensive Multivariate Time Series (MTS) data, creating a demand for robust anomaly detection. Methodologies in this domain have evolved through three main phases, as illustrated in Figure 1. First, reconstruction and prediction-based methods, such as RNNs and autoencoders, were used to learn normal patterns. However, these models often generalize too well, reconstructing anomalies as accurately as normal data, which obscures detection.

To address this, the second phase focused on explicit relationship modeling. Techniques like GNNs were introduced to capture spatial dependencies and variable interactions that earlier models missed. Recently, generative models, including diffusion models, have gained significant attention for their ability to improve anomaly detection in multivariate time series. Distribution-based models, including DAGMM[1] and OmniAnomaly[2], and more recently diffusion models, have become prominent. Diffusion models leverage the manifold hypothesis to improve distribution fitting, quantifying anomalies via denoising or likelihood estimation. In parallel, efficient backbones such

N. Malik et al. (eds.), *Proceedings of the 6th International Conference on Public Management and Intelligent Society (PMIS 2026)*, Advances in Engineering Research 301,
https://doi.org/10.2991/978-94-6239-721-7_8

as ModernTCN[3] and TimesNet[4], along with Large Language Models (LLMs)[5], are redefining feature extraction. This paper surveys this trajectory, analyzing the shift from discriminative reconstruction to generative understanding.

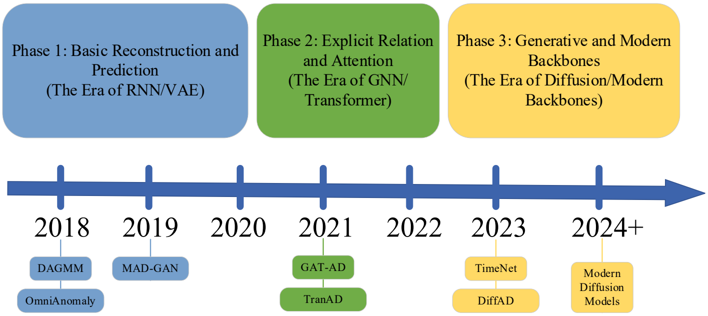

Fig. 1. Three Phases of Technological Evolution in Multivariate Time Series Anomaly Detection (2018-2024+).

2 Problem Definition

The task of MTS anomaly detection is formally defined as follows. Let $\mathcal{X} = \{x_1, x_2, \dots, x_N\}$ be a multivariate time series sequence of length N, where each $x_t \in \mathbb{R}^D$ represents a D-dimensional vector of observations at time step t. The goal is to learn a function $f: \mathbb{R}^{W \times D} \rightarrow \mathbb{R}$ that maps a sliding window of historical data W to an anomaly score s_t.

For a given timestamp t, the anomaly score is typically computed as:

$$s_t = f(x_{t-W+1:t}) \tag{1}$$

Where $x_{t-W+1:t}$ denotes the input sequence window. The system then determines the binary label $y_t \in \{0,1\}$ (0 for normal, 1 for anomaly) by comparing s_t against a threshold τ:

$$y_t = \begin{cases} 1, & if\ s_t \geq \tau \\ 0, & otherwise \end{cases} \tag{2}$$

Depending on the paradigm, s_t is derived differently. For reconstruction-based methods, it is the reconstruction error:

$$s_t = ||x_t - \hat{x}_t||^2 \tag{3}$$

For generative frameworks, s_t is often derived from the negative log-likelihood (NLL) of the observation under the learned distribution $p_\theta(x)$:

$$s_t = -\log p_\theta\,(x_t | x_{t-1}, \dots) \tag{4}$$

This probabilistic formulation allows the model to handle stochasticity more effectively than deterministic errors.

3 Methodology Survey

To systematically organize existing research, this paper establishes a classification framework as illustrated in Figure 2, and categorizes existing methods into the following four categories.

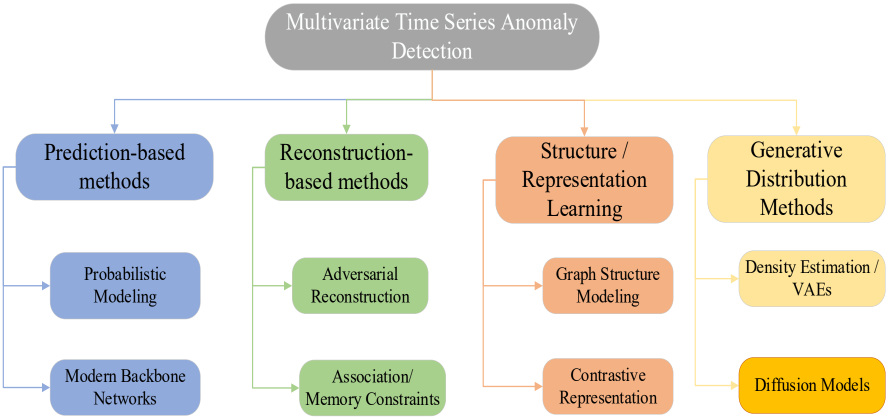

Fig. 2. Classification Framework of Multivariate Time Series Anomaly Detection Methods Proposed in This Paper

3.1 Prediction-based Methods

Prediction-based methods detect anomalies by comparing actual values with forecasted trends. Moving beyond simple MSE optimization used in early RNNs, DeepAR[6] introduces probabilistic forecasting to better handle data randomness. To capture complex temporal variations, TimesNet transforms 1D time series into 2D tensors via FFT to model multi-periodicity. Furthermore, ModernTCN utilizes large-kernel convolutions and parameter decoupling to efficiently model long-sequences, offering a pure CNN alternative that rivals Transformers in feature extraction efficiency. Prediction-based scoring is sensitive to sudden changes but may underperform under strong seasonality shifts or concept drift. It is often preferred when forecasting is already available in the monitoring stack and low-latency inference is required.

3.2 Reconstruction-based Methods

These methods assume that models trained on normal data cannot reconstruct anomalies effectively. To prevent the "generalization pitfall" (where anomalies are also well-reconstructed), USAD[7] and TranAD[8] employ adversarial training strategies to amplify reconstruction errors of outliers. Anomaly Transformer[9] utilizes a correlation discrepancy mechanism to highlight anomalies based on association differences, while TFAD[10] and MEMTO[11] impose constraints via time-frequency decomposition and memory networks, respectively, to strictly limit the model's capacity to represent abnormal patterns. Reconstruction models are effective for pattern deviations but may suffer from over-generalization, motivating constraints (memory, frequency priors) or adversarial objectives. They are practical when training data is mostly clean and the system allows offline model updates.

3.3 Explicit Relationship Modeling: Graph Neural Networks

GNNs address the spatial blind spots of traditional models by explicitly modeling topological dependencies. GDN[12] learns a sensor embedding graph to capture causal relationships and deviations. To further break spatiotemporal boundaries, MTAD-GAT[13] employs a dual-stream graph attention architecture that simultaneously models inter-variable spatial correlations and temporal dependencies, significantly enhancing the perception of fine-grained anomaly signals in Cyber-Physical Systems. Graph modeling improves detection when sensor dependencies are stable and physically meaningful, but performance can degrade if the learned graph is noisy or changes over time. It is especially suitable for cyber-physical systems where topology priors or causal relations are available.

3.4 Probability Distribution-based and Generative Models

Generative models focus on characterizing the inherent probability distribution of data. Early works like DAGMM and OmniAnomaly use joint optimization and VAEs to estimate density in low-dimensional spaces, while InterFusion[14] introduces latent MCMC imputation to simultaneously model inter-metric dependencies and data distributions. Recently, diffusion-based methods have reported strong performance on several public benchmarks. Methods like DiffAD[15], Imputed Diffusion[16], and DTE[17] leverage the reverse diffusion process to repair masked sequences or calculate diffusion steps. Distribution-based models better capture complex manifolds but often incur higher computational cost, especially for iterative diffusion sampling. They are attractive when accuracy is prioritized over latency or when acceleration techniques are available.

4 Indicative Performance Overview Across Paradigms

4.1 Motivation and Scope

To provide a high-level overview of representative approaches spanning from reconstruction-based to generative paradigms, this section summarizes reported results on widely used public benchmarks in multivariate time series anomaly detection.

It is important to clarify that the performance values presented in this section are collected directly from the corresponding original publications. Since preprocessing pipelines, sliding window configurations, threshold selection strategies, and point adjustment (PA) mechanisms differ across studies, the reported F1-scores are not strictly comparable under a unified experimental protocol. Therefore, Table 1 should be interpreted as an indicative overview of paradigm-level performance trends, rather than a rigorous head-to-head benchmark comparison.

4.2 Experimental Datasets and Evaluation Metric

Four commonly used datasets—SMAP, MSL, SWaT, and PSM—are considered, as they are widely adopted in the literature for evaluating multivariate time series anomaly detection methods.

The F1-score is reported as the primary metric in most referenced studies. Many works apply the Point Adjustment (PA) strategy to account for continuous anomaly segments, which aligns better with industrial monitoring requirements. However, it should be noted that the implementation details of PA and threshold calibration vary across papers, further limiting strict comparability.

Table 1 summarizes representative models from different methodological categories.

Table 1. Reported F1-Scores of Representative Methods on Public Benchmarks

Type	Model	SMAP	MSL	SWaT	PSM
Recon/Prob	OmniAnomaly	86.92	87.67	82.83	80.83
Adversarial	USAD	86.29	88.54	80.44	82.65
GNN	GDN	-	-	81.00	92.90
Transformer	Anomaly Trans	96.69	93.59	94.07	97.89
CNN	ModernTCN	71.26	84.92	93.86	97.23
Diffusion	DiffAD	96.95	94.19	97.66	97.95

Note: Results are compiled from the original papers and may not be strictly comparable due to differences in preprocessing, point adjustment strategies, and thresholding choices. "-" indicates not reported.

4.3 Discussion and Observations

Generative and diffusion-based models demonstrate strong performance across multiple datasets, suggesting that probabilistic distribution modeling is advantageous for capturing complex data structures, albeit at the cost of increased computational overhead due to iterative sampling. In contrast, discriminative architectures such as ModernTCN remain competitive in scenarios where inference efficiency is critical. Graph-based approaches like GDN perform well in systems with stable physical dependencies; however, they face challenges including topology drift, graph learning instability, and substantial computational and memory overhead in high-dimensional settings. Overall, existing methods reflect a methodological transition from reconstruction-based paradigms to probabilistic generative formulations, while also highlighting practical trade-offs among accuracy, efficiency, and scalability.

5 Cutting-edge Trends

Based on a comprehensive review of existing literature, multivariate time series anomaly detection (MTS-AD) is exhibiting four key trends, shifting from single-model paradigms to cross-modal fusion and evolving from discriminative computation to generative understanding.

5.1 Resurgence of General-purpose High-performance Backbone Networks

Recent studies have demonstrated that well-designed pure convolutional networks possess the potential to outperform Transformers in feature extraction efficiency. TimesNet captures multi-periodic variations via time-series folding, while ModernTCN enables low-cost long-sequence modeling using large convolution kernels. Future research will increasingly revert to these general-purpose backbone networks, rather than relying on elaborate module designs tailored for specific tasks.

5.2 Efficiency Optimization of Generative Diffusion Models

Diffusion models have become a prominent alternative to GAN/VAE-style generative models, showing improved flexibility in fitting complex distributions; however, their iterative sampling introduces latency. The key direction for future research lies in breaking through the real-time performance bottleneck while maintaining high accuracy, through techniques such as model distillation or fast sampling algorithms.

5.3 Cross-domain Empowerment of Large Language Models (LLMs)

Leveraging the general reasoning capabilities and pre-trained knowledge bases of LLMs to address semantic understanding challenges in time-series analysis has emerged as a research hotspot. Future trends will focus on utilizing LLMs to enable few-shot or zero-shot anomaly detection, and providing automated anomaly explanations through natural language interaction, thereby lowering the threshold for operation and maintenance.

5.4 Integration of Interpretability and Root Cause Analysis

To address the limitation of current deep generative models characterized by "high accuracy yet black-box mechanisms", future research is committed to developing unified frameworks that achieve both detection accuracy and interpretability. The goal is to bridge the gap between graph models and generative models, enabling the model to output attribution paths of anomalous variables alongside anomaly alerts, thus supporting rapid decision-making.

6 Conclusion

This paper presents a comprehensive review of the research progress in multivariate time series anomaly detection, sorting out the technological evolution trajectory from traditional prediction-based and reconstruction-based methods, to graph neural network methods for explicit modeling of variable relationships, and further to the state-of-the-art probabilistic generative and diffusion models.

Early reconstruction models and prediction models laid the foundation for unsupervised detection, yet they faced bottlenecks in handling complex dependencies and preventing overfitting. Subsequently, graph neural networks and contrastive learning frameworks significantly improved the model's ability to perceive multivariate interactions by introducing structural priors and discriminative constraints.

Currently, the field is transitioning toward probabilistic and generative modeling frameworks. While diffusion-based approaches have demonstrated competitive results on multiple benchmarks, their computational overhead and deployment complexity remain open challenges. Future research should aim to balance accuracy, efficiency, scalability, and interpretability, rather than focusing solely on performance gains. Developing unified frameworks that integrate efficient backbone networks, topology reasoning, and probabilistic modeling remains a promising yet non-trivial direction.

References

1. Zong B, et al. (2018) Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection. ICLR. https://openreview.net/forum?id=BJJLHbb0-
2. Su Y, et al. (2019) Robust Anomaly Detection for Multivariate Time Series through Stochastic Recurrent Neural Network. KDD. https://doi.org/10.1145/3292500.3330672
3. Luo D, et al. (2024) ModernTCN: A Modern Pure Convolution Structure for General Time Series Analysis. ICLR. https://openreview.net/forum?id=vpJMJerXHU
4. Wu H, et al. (2023) TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. ICLR. https://arxiv.org/abs/2210.02186
5. Zhang X, et al. (2024) Large Language Models for Time Series: A Survey. arXiv preprint. https://arxiv.org/abs/2402.01801
6. Salinas D, et al. (2020) DeepAR: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting. https://arxiv.org/abs/1704.04110
7. Audibert J, et al. (2020) USAD: UnSupervised Anomaly Detection on Multivariate Time Series. KDD. https://doi.org/10.1145/3394486.3403392
8. Tuli S, et al. (2022) TranAD: Deep Transformer Networks for Anomaly Detection in Multivariate Time Series. VLDB. https://doi.org/10.14778/3514061.3514067
9. Xu J, et al. (2022) Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy. ICLR. https://openreview.net/forum?id=LzQQ89U1qm_
10. Zhang C, et al. (2022) TFAD: A Time-Frequency Decomposition Architecture for Time Series Anomaly Detection. NeurIPS. https://arxiv.org/abs/2210.09693
11. Song J, et al. (2024) MEMTO: Memory-Guided Transformer for Multivariate Time Series Anomaly Detection. IEEE TNNLS. https://arxiv.org/abs/2312.02530
12. Deng A, et al. (2021) Graph Neural Network-Based Anomaly Detection in Multivariate Time Series. AAAI. https://arxiv.org/abs/2106.06947

13. Zhao H, et al. (2020) Multivariate Time-series Anomaly Detection via Graph Attention Network. ICDM. https://doi.org/10.1109/ICDM50108.2020.00093
14. Li Z, et al. (2021) Multivariate Time Series Anomaly Detection and Interpretation using Hierarchical Inter-Metric and Temporal Embedding. KDD. https://doi.org/10.1145/3447548.3467075
15. Xiao C, et al. (2023) Imputation-based Time-Series Anomaly Detection with Conditional Weight-Incremental Diffusion Models. KDD. https://doi.org/10.1145/3580305.3599391
16. Chen Y, et al. (2023) Imputed Diffusion Models for Multivariate Time Series Anomaly Detection. arXiv. https://arxiv.org/abs/2306.02062
17. Livernoche V, et al. (2024) On Diffusion Modeling for Anomaly Detection. ICLR. https://arxiv.org/abs/2305.18593