A Review of Anomaly Detection Methods for Multivariate Time Series Based on Time-Domain and Frequency-Domain Perspectives

Jiaming Xu, Hanqi Liu, Jun Xu
College of Science, North China University of Science and Technology, Tangshan, China

Abstract: With the increasing proliferation of sensor fusion, industrial control systems (ICS), and internet services, multivariate time-series data has become ubiquitous in fields such as intelligent manufacturing, financial monitoring, network security, and transportation. Multivariate Time-Series Anomaly Detection (MTSAD) aims to identify patterns that deviate significantly from normal behavior within high-dimensional dynamic metrics, thereby providing critical support for system monitoring and early warning. Compared with static data and univariate sequences, multivariate time series are characterized by temporal dependencies, complex inter-variable interactions, and dynamic distribution shifts, posing strictly higher requirements for anomaly detection. This paper first introduces the definitions of time-series anomaly detection and categorizes common anomaly types. Secondly, it classifies existing methods from both time-domain and frequency-domain perspectives, providing a comprehensive analysis of their advantages, limitations, and application scenarios. Finally, the paper explores key research directions for the future design of anomaly detection methods, offering a reference for both theoretical and applied research in this domain.

Keywords: Industrial Internet of Things; Multivariate Time Series Data; Time-Domain Methods; Frequency-Domain Methods; Industrial Control Systems

1. Introduction

Multivariate time-series data is widely prevalent in domains such as industrial control, intelligent manufacturing, environmental monitoring, and financial analysis. In these applications, anomaly detection in time-series data is crucial for ensuring system stability and security. The primary objective of anomaly detection is to identify abnormal behaviors—distinguishable from regular patterns—within vast amounts of normal data. These anomalies may stem from equipment failures, cyberattacks, environmental changes, or human intervention. Unlike traditional univariate time-series anomaly detection, multivariate scenarios require the simultaneous modeling of temporal dependencies and cross-variable structural relationships, making algorithm design significantly more challenging[1].

Time-series anomaly detection refers to the identification of data segments or patterns that differ significantly from normal behavior within a time series. In multivariate contexts, the data not only reflects the laws of individual variables changing over time but also involves complex interactions and dependencies among variables. The goal is to provide early warning signals to the system by capturing these deviations from normal patterns. Since outliers often indicate potential system risks, faults, or abnormal events, their timely identification is vital for maintaining system stability and safety.

Traditional univariate anomaly detection methods [2] often fail to capture correlations between variables and system-level abnormal behaviors. Conversely, multivariate time-series anomaly detection must address dynamic changes in the temporal dimension while accounting for mutual influences in the variable dimension. In recent years, the rise of deep learning has provided new opportunities to address this challenge. The field is experiencing rapid development and evolution, transitioning from early statistical methods and traditional machine learning algorithms to modern deep learning models such as Recurrent Neural Networks (RNNs), Graph Neural Networks (GNNs), and attention mechanisms. Furthermore, frequency-domain analysis has demonstrated superior performance in handling time-series data with distinct periodicities and frequency components. By utilizing techniques such as the Fourier Transform and Wavelet Transform, frequency-domain methods convert time-series data into spectral information, thereby revealing periodic and frequency characteristics. This approach is particularly effective for detecting anomalies that are difficult to identify in the time domain, such as periodic faults and equipment vibration anomalies.

This paper surveys relevant literature in the field of time-series anomaly detection from recent years, summarizing the technical characteristics, applicable scenarios, and limitations of existing methods. Furthermore, it explores future development directions to provide a comprehensive reference for researchers and practitioners in related fields.

2. Definitions and Applications of Time-Series Anomaly Detection

2.1 Problem Definition

Given a multivariate time series X, as shown in Eq. (1),

$$X = \{x_1, x_2, ..., x_T\}, x_t \in \mathbb{R}^d \tag{1}$$

where the two parameters respectively denote the dimension of sensors or indicators and the sequence length. the main task is, in an unsupervised setting, to determine whether there exist potential anomalies within a sliding time window. During the inference stage, the weighted sum of the two absolute errors, prediction error and reconstruction error, is used. A threshold is set, and the observations are evaluated according to the anomaly score; that is, when the anomaly score obtained for a time window exceeds the threshold, the window is marked as anomalous. Here, a binary indicator is used to represent whether the data point at time stamp ttt is anomalous (1 for anomalous and 0 for normal).

2.2 Applications

Multivariate Time-Series Anomaly Detection (MTSAD) is extensively applied across various industries. The following are several typical application scenarios:

1)Intelligent Manufacturing: During the operation of industrial equipment, anomalous data may indicate equipment malfunctions or potential issues within the production process. Real-time monitoring enables the timely detection of anomalies, effectively reducing equipment downtime and production losses.

2)Cybersecurity: In network traffic monitoring, anomalous traffic patterns may reflect cyberattacks or malicious behaviors. The timely identification of traffic anomalies prevents security vulnerabilities from being exploited.

3)Financial Monitoring: In financial markets, abnormal transaction data can signal sudden market shifts, fraudulent activities, or irregular fluctuations. Effective anomaly detection methods assist financial institutions in robust risk management.

4)Transportation: Within transportation systems, anomalous data may reveal irregular changes in traffic flow. Timely early warnings facilitate the optimization of traffic management and the reduction of accidents.

2.3 Problems and Challenges in Time-Series Anomaly Detection

Constructing a robust framework for time-series anomaly detection remains difficult. It requires characterizing both long-term and short-term temporal dependencies within individual sequences while simultaneously modeling the coupling relationships across features. Furthermore, robustness against noise must be intrinsic to the model structure. Broadly speaking, time-series anomaly detection faces the following primary issues and challenges:

1)Temporal Dependencies and Inter-variable Correlations: Variables in multivariate time-series data typically exhibit temporal correlations, and complex dependencies often exist between variables. Traditional methods struggle to capture these high-dimensional, non-linear dependencies, resulting in insufficient model performance.

2)Noise and Perturbations: Real-world time-series data frequently contain noise, missing values, or other external perturbations. These factors can significantly compromise the stability and robustness of anomaly detection.

3)Multi-scale and Long-range Dependencies: Anomalies may manifest as short-term abrupt changes, long-term drifts, or periodic variations. Effectively capturing multi-scale features from data and modeling long-range dependencies remain significant challenges.

4)Real-time Performance and Computational Efficiency: With the increasing demand for real-time capabilities in industrial and internet applications, many existing methods suffer from high computational complexity, making them difficult to deploy for rapid detection in resource-constrained environments.

3. Classification of Anomalies and Datasets

3.1 Classification of Time-Series Anomalies

Time-series anomaly detection can be systematically categorized into three paradigms based on deviation morphology and discrimination criteria: Point Anomalies, Contextual Anomalies, and Collective Anomalies. This classification reveals a progressive complexity—ranging from isolated amplitude deviations to violations of conditional expectations, and finally to the disruption of pattern structures—directly corresponding to the need for models to possess modeling capabilities ranging from shallow statistics to deep semantics. Examples of anomaly types are shown in Fig. 1.

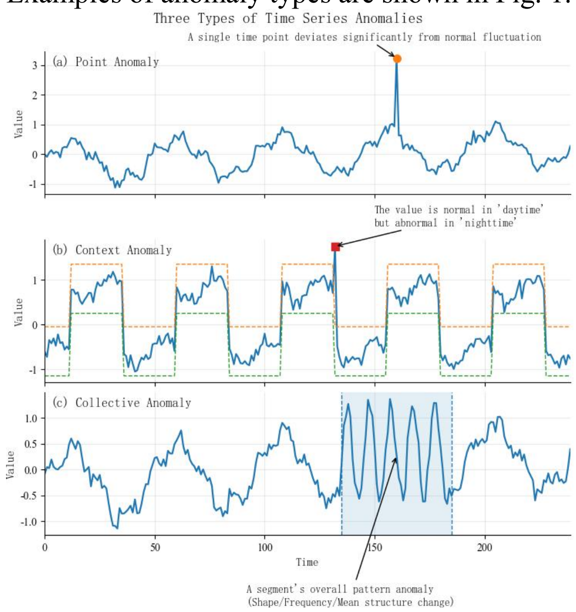

Figure 1. Examples of Time Series Anomaly Detection Types

Point Anomalies: These refer to observations at a specific time point that deviate significantly from the normal pattern. They typically occur as short-term anomalous events. These anomalies usually appear independently and can be determined without relying on preceding or succeeding temporal associations. The degree of deviation can be directly quantified via statistical thresholds, reconstruction errors, or prediction residuals. For example, instantaneous spikes in industrial sensors caused by electromagnetic interference, or a sudden surge in CPU usage to 100% on a server while adjacent timestamps remain around 20%.

Contextual Anomalies: These refer to anomalies that are conditional upon specific contexts; that is, the occurrence of the anomalous event depends on a specific time window or background. The determination of such anomalies must be combined with contextual information such as time, periodicity, and trends. Viewed in isolation (without context), the data point may fall entirely within a normal statistical range. For instance, a data point may be normal during certain time periods but considered anomalous during others. In practical industrial systems, these three types of anomalies often occur simultaneously. An ideal detection framework should possess multi-scale perception capabilities, enabling it to respond quickly to isolated point anomalies while capturing subtle deviations in context and patterns.

Collective Anomalies: These refer to a continuous subsequence that deviates from normal behavior in terms of overall dynamic patterns (e.g., shape, frequency, dependency structure, or variable covariance), even though the individual values within the subsequence may fall within the normal range. Such anomalies typically manifest as a collection of abnormalities across multiple time steps or variables. Single-point statistics are usually insignificant; therefore, judgment must be based on the sequence as a whole, often relating to the system's long-term trends or periodic fluctuations.

3.2 Datasets

Multiple public datasets are employed in time-series anomaly detection research. Common datasets include:

SWaT [3]: The Secure Water Treatment (SWaT) dataset, provided by the Singapore University of Technology and Design (SUTD), is designed to simulate cybersecurity issues in water treatment systems, with a particular focus on attacks and defenses in Industrial Control Systems (ICS). It contains data from multiple sensors in a water treatment facility and is suitable for researching anomaly detection in ICS.

WADI [4]: The Water Distribution (WADI) dataset is designed specifically for water distribution systems, aiming to provide a testbed for simulating cyber-physical attacks. It covers data from multiple sensors, assisting researchers in developing anomaly detection algorithms tailored for water distribution infrastructure.

SMD [5]: The Server Machine Dataset (SMD) comes from a large internet company (often used in smart manufacturing contexts in literature) and primarily contains data from various sensors (metrics). It is used to research fault diagnosis and anomaly detection in production equipment. This dataset is widely used in studies concerning industrial equipment fault detection and predictive maintenance.

MSL, SMAP [6]: These two datasets are provided by NASA. MSL (Mars Science Laboratory) and SMAP (Soil Moisture Active Passive) focus on telemetry data monitoring (such as soil humidity and freeze-thaw states in the context of SMAP). Acquired via aerospace sensors, these datasets are applicable to climate monitoring and anomaly detection in telemetry data.

SKAB [7]: The Skoltech Anomaly Benchmark (SKAB) is provided by Skoltech. It aims to provide benchmark data for time-series anomaly detection algorithms, specifically addressing multivariate point anomaly detection and changepoint detection problems. It is suitable for evaluating anomaly detection in industrial and environmental systems.

4. Classification of Time-Series Anomaly Detection Models

Research methods for Multivariate Time-Series Anomaly Detection (MTSAD) can be classified from various perspectives. This section surveys existing approaches across four dimensions: statistical methods, traditional machine learning methods, deep learning methods, and time-domain versus frequency-domain analysis. The classification framework is illustrated in Fig. 2.

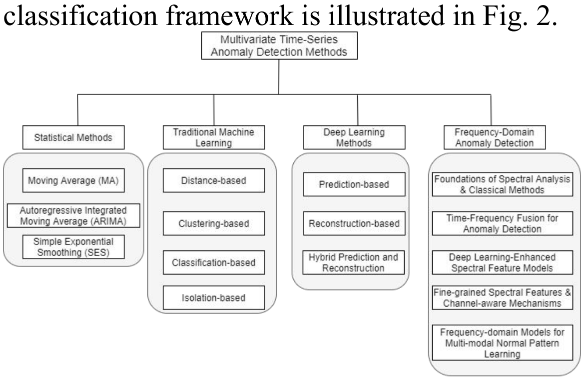

Figure 2. Classification Diagram of Temporal Anomaly Detection Model

4.1 Classification of Time-Series Anomaly Detection Based on Statistics

The fundamental premise of statistical-based time-series anomaly detection is that normal data follows a specific stochastic distribution (e.g., Gaussian distribution), whereas anomalies are low-probability events located at the tail of the distribution. In multivariate scenarios, the emphasis is placed on capturing the covariance relationships between variables, utilizing the statistical characteristics of time-series data to identify outliers. Typical methods include Moving Average (MA), smoothing filtering, Seasonal-Trend Decomposition, and Auto-Regressive Integrated Moving Average (ARIMA). The Moving Average method smooths random fluctuations by averaging data within a fixed window, thereby accentuating trends and seasonal patterns.

The ARIMA model integrates three components—Auto-Regressive (AR), Integrated (I), and Moving Average (MA)—to model time series. It predicts future values and compares them with actual observations; if the residual exceeds a specific confidence interval, the data point is flagged as an anomaly. Furthermore, Statistical Process Control (SPC) methods, such as Shewhart control charts, CUSUM, and EWMA, are widely used in industrial processes for online detection. These methods assume that monitoring metrics follow a statistical distribution and trigger alarms when real-time values deviate from the mean by several standard deviations or exhibit significant trend changes.

Reference [8] proposes the Seasonal Hybrid Extreme Studentized Deviate (S-H-ESD) algorithm. This method performs seasonal-trend decomposition on the time series and subsequently applies statistical tests to the residual component to detect anomalies. In industrial equipment monitoring, simply setting upper and lower control limits for measured variables can also detect out-of-limit anomalies in a timely manner. However, statistical methods often rely on stationarity and specific distribution assumptions, limiting their ability to detect anomalies in complex patterns and non-linear relationships. As the dimensionality of multivariate systems increases, univariate statistical monitoring cannot account for inter-variable correlations, often leading to missed detections of correlated anomalies. Additionally, statistical methods require the manual setting of thresholds or model orders. When data distributions or noise characteristics change, the model requires re-calibration, resulting in poor adaptability.

4.2 Classification of Time-Series Anomaly Detection Based on Machine Learning

Traditional machine learning-based methods typically distinguish between normal and anomalous points via clustering, distance metrics, or density estimation. These approaches utilize statistical learning and data mining algorithms to automatically learn normal behavioral patterns from historical data and identify anomalies that deviate from these patterns. Classical unsupervised algorithms include distance-based methods, density-based methods, and one-class classification methods.

(1) Distance-based Methods: These methods identify anomalies by measuring the distance between a new observation and representative samples (e.g., centroids or cluster centers). Points with excessive distances are considered anomalies. Reference [9] proposes a distance-based method using KNN (K-Nearest Neighbors) to analyze and detect anomalies in wireless sensor network data. However, such methods often suffer from high computational costs and low efficiency, making them infeasible in complex scenarios.

(2) Density-based Methods: Using local density as a criterion, these methods assume that normal data is densely distributed in the feature space, while anomalies are sparse. A representative algorithm is the Local Outlier Factor (LOF), which computes the local density deviation of each point relative to its neighbors. If a point's local density is significantly lower than that of its neighbors, it is judged as an anomaly.

(3) Clustering-based Methods: Algorithms such as K-Means or DBSCAN operate on the assumption that normal data belongs to large clusters, whereas anomalous data does not belong to any cluster or forms very small ones. Reference [10] applies K-Means clustering to network traffic anomaly detection. The study utilizes K-Means to analyze NetFlow records. The algorithm first clusters normal traffic patterns and detects anomalies (e.g., DoS attacks or port scanning) by calculating the distance of new traffic to these cluster centers. The paper demonstrates how clustering distance thresholds effectively distinguish normal traffic fluctuations from malicious attacks.

(4) Classification-based Methods (One-Class): Algorithms such as One-Class SVM (OCSVM) and Support Vector Data Description (SVDD) detect anomalies by fitting a hyperplane or hypersphere in the feature space that encompasses normal data. OCSVM and SVDD do not require anomalous samples for training; they build models based solely on normal data, making them highly suitable for anomaly detection tasks. Reference [11] proposes a DBSCAN ensemble method utilizing the internal structure of time series for adaptive parameter selection. Experiments on the Yahoo dataset indicate that this strategy effectively reduces detection variance but exhibits high sensitivity when handling imbalanced samples, limiting its robustness.

(5) Isolation-based Methods: The most typical algorithm is Isolation Forest (iForest). It partitions data by randomly constructing binary trees, based on the principle that anomalies are easier to "isolate" (i.e., they have shorter path lengths in the tree). Reference [12] applies Isolation Forest to streaming data anomaly detection. Addressing the characteristics of time-series data streams, the paper proposes a Sliding Window-based iForest scheme. By building iTrees within sliding windows, the method adapts to concept drift in data distribution and rapidly detects anomalies in the current window. The study proves that the isolation mechanism possesses high computational efficiency when processing high-speed, dynamic time-series data.

In supervised scenarios, if historically labeled anomaly samples exist, anomaly detection can be treated as a classification problem using traditional classifiers (e.g., SVM, Decision Trees, Neural Networks). However, in practice, anomaly samples are scarce and unevenly distributed, making fully supervised methods impractical. Consequently, semi-supervised or weakly supervised methods have emerged, such as synthesizing minority anomaly samples via data augmentation or employing active learning to focus on data near the decision boundaries.

Overall, traditional machine learning methods can capture more complex relationships compared to statistical methods. For instance, Clustering and Principal Component Analysis (PCA) can be used for dimensionality reduction and pattern extraction in multivariate data, detecting anomalies based on reconstruction error or principal component limits. Similarly, in network traffic monitoring, feature vectors can be extracted from time series and K-Means applied, where outlier cluster centers correspond to anomalous patterns. The disadvantage is that these methods often ignore the temporal dependency of time series, treating data as independent points; their effectiveness degrades in extremely high-dimensional feature spaces. Furthermore, in high-dimensional cases, they may face the "curse of dimensionality," causing distance metrics to fail and density estimation to become difficult. As system complexity increases, the effectiveness of traditional methods in multivariate scenarios drops significantly, prompting researchers to turn to more powerful models like deep learning.

Table 1. Machine Learning-Based Time-Series Anomaly Detection

Model Name	Advantageous Scenarios	Limitations
KNN	Simple logic;	High computational cost; Low efficiency;
LOF (Local Outlier Factor)	Identifying "local outliers" that are sparse relative to neighbors but not global extremes	Density estimation becomes difficult in high-dimensional spaces
K-Means / DBSCAN	Distinguishing normal fluctuations from attacks	Assumes normal data belongs to large, spherical clusters; Ignores temporal dependencies
OCSVM / SVDD	Novelty Detection; Training requires only normal samples	Ignores time dependencies;
Isolation Forest	High computational efficiency;	variable correlations compared to deep learning models

4.3 Classification of Time-Series Anomaly Detection Based on Deep Learning

Leveraging powerful feature extraction capabilities—particularly the ability to capture non-linearities and long-range dependencies—deep learning has made substantial strides in multivariate time-series anomaly detection. It has demonstrated exceptional performance in fields such as financial risk control, cybersecurity, and industrial equipment fault diagnosis. Based on the learning paradigm of the model, deep learning-based anomaly detection methods can be categorized into Reconstruction-based models, Prediction-based models, and Hybrid models.

4.3.1 Reconstruction-based Anomaly Detection Models

Reconstruction-based models operate by encoding input time-series data into a low-dimensional latent space and subsequently reconstructing the original data via decoding. The underlying assumption is that a model trained on normal data can effectively compress and restore normal samples, whereas it will fail to accurately reconstruct unseen anomalous samples.

These models utilize deep neural networks to learn low-dimensional representations of time-series data. If the input sequence conforms to the normal patterns learned during training, the model yields a low reconstruction error; conversely, anomalous sequences deviating from normal patterns result in significantly higher errors, which serve as the basis for anomaly detection. Typical models include Autoencoders (AE) and their variants. An AE employs an encoder network to map multivariate time series into a latent space and a decoder network to reconstruct them. Malhotra et al. (2016) [13] pioneered the use of LSTM-based autoencoders for multi-sensor anomaly detection, training the model to reconstruct normal states and using reconstruction error to detect various anomaly types. Numerous subsequent studies have adopted reconstruction error as a standard anomaly score.

To enhance the ability to model complex distributions, Variational Autoencoders (VAE) were introduced. In the encoding phase, VAEs generate random variables following a prior distribution, incorporating stochasticity and regularization to learn the probability distribution of the data. OmniAnomaly (Su et al., 2019) [5] is the first multivariate time-series anomaly detection model to utilize Stochastic Recurrent Neural Networks (SRNN)—introducing temporal dependencies into VAEs. It models temporal dependencies in the latent space via GRU networks and evaluates anomalies using the probability of the reconstructed sequence, effectively characterizing the temporal correlation of stochastic variables. This innovation improved performance on complex time series (e.g., periodic patterns superimposed with noise). Furthermore, some studies have integrated attention mechanisms to improve precision. For instance, MSCRED [14] utilizes multi-scale convolution and recurrent networks to extract feature correlations at different granularities, incorporating attention mechanisms to diagnose anomalies of varying severities effectively.

Another direction for reconstruction models involves integrating Generative Adversarial Networks (GANs) to enhance stability and reconstruction capability. A representative model, DAGMM [15], jointly trains a deep autoencoder and a Gaussian Mixture Model (GMM), achieving end-to-end optimization for dimensionality reduction and density estimation. While this ensures critical information is preserved, DAGMM relies on standard autoencoders and can suffer from local optima and high computational costs. Consequently, improved methods such as USAD [16] (Audibert et al., 2020) utilize adversarial training to boost the stability and speed of the autoencoder. Similarly, MAD-GAN [17] employs an adversarial framework where both the generator and discriminator are constructed using LSTMs to simultaneously capture temporal patterns and inter-variable interactions. While GAN-based models can learn complex distributions and discover hard-to-reconstruct anomalies via discriminator feedback, they often face instability issues (e.g., gradient vanishing or mode collapse). Subsequent research, such as TADGAN [18], has introduced techniques like Wasserstein loss and cycle consistency to mitigate these issues.

Overall, the advantage of reconstruction-based models lies in their ability to perform unsupervised learning using only normal data, automatically extracting key structural features with high sensitivity to out-of-distribution anomalies. Their disadvantage is the potential failure to reconstruct all complex patterns and the lack of interpretability regarding the reconstruction error itself. For highly non-linear multivariate sequences, simple autoencoders have limited capacity, whereas increasing capacity (via more layers or complex units) increases training difficulty and the risk of overfitting to anomalies. Thus, balancing generalization ability with reconstruction accuracy remains a key research direction.

4.3.2 Prediction-based Anomaly Detection Models

Prediction-based models utilize autoregressive principles to predict the next timestamp based on a historical window, using the prediction error as the anomaly score. Typical models include LSTM, GRU, and CNN. If the prediction error exceeds a certain threshold, the behavior at the current moment is deemed to deviate from historical laws and is flagged as anomalous. The core of these methods is to train a time-series forecasting model that achieves high accuracy on normal patterns; deviations result in significant residual spikes.

LSTM-based methods are among the most widely applied. Hundman et al. [6] proposed using multi-layer LSTMs for sequence prediction on NASA spacecraft telemetry data, designing a non-parametric dynamic thresholding (NDT) mechanism. This method (LSTM-NDT) adapts to residual distributions without manual threshold setting. However, it predicts only one dimension at a time, limiting its performance in high-dimensional, coupled scenarios. To address this, hybrid approaches like LGMAD [19] combine LSTM with GMM, using LSTM for temporal features and GMM for low-dimensional feature distribution modeling, thereby considering both temporal dynamics and variable correlations.

Convolution-based models, such as DeepAnt [20], use CNNs to automatically learn feature representations from time-series segments for prediction. Compared to RNNs, CNNs offer high parallel computing efficiency and excel at extracting local patterns. Architectures like Temporal Convolutional Networks (TCN) utilize causal and dilated convolutions to expand the receptive field, capturing long-range dependencies. TCNs often outperform traditional CNNs by learning multi-granular features. The limitation of CNNs lies in the finite kernel size, making global dependency modeling difficult, though this is partially mitigated by stacking layers.

Transformer-based models rely on self-attention mechanisms suitable for long-range dependencies. TimesNet [21] innovatively transforms time series into a 2D frequency domain, applying FFT to capture periodic components and 2D convolution for feature extraction. This integrates frequency-domain information into the prediction framework, enhancing generalization on complex periodic and trending sequences. Transformers and their hybrids represent the state-of-the-art, achieving record performance on datasets like SMD and SWaT.

Overall, prediction models directly exploit temporal dependencies and provide intuitive anomaly metrics via errors. They are particularly sensitive to point anomalies (sudden spikes/drops). However, they require comprehensive training on normal patterns; otherwise, unseen normal modes may generate false positives. Additionally, performance depends heavily on prediction accuracy, which is susceptible to noise. In multivariate scenarios, single-variable prediction ignores inter-variable linkages, while joint prediction increases complexity. Efficiently integrating multivariate information remains a challenge.

4.3.3 Hybrid Reconstruction and Prediction Models

Single prediction or reconstruction models often exhibit limitations. Hybrid models employ multi-task learning to enhance robustness: they learn normal patterns via reconstruction while capturing dynamic changes via prediction. This dual approach improves the capability to capture long-range dependencies and non-linear patterns. TranAD [22] is a representative hybrid model based on a Transformer encoder. It captures long-range dependencies via self-attention while introducing adaptive conditioning and adversarial training. TranAD computes a combined error from both reconstruction and masked prediction tasks. Experiments show it outperforms previous methods on multiple datasets (F1 score increased by up to 17%) with faster training. Its drawback lies in structural complexity, as the introduction of meta-learning and adversarial mechanisms may lead to over-generalization.

Anomaly Transformer [23] introduces an Anomaly-Attention mechanism. It hypothesizes that anomalies exhibit different association patterns with the global sequence compared to normal points. By calculating the association discrepancy between each time point and the overall sequence, and employing a Min-Max adversarial training strategy, it distinguishes anomalies from normal patterns. This method focuses on association differences rather than direct prediction values, making it highly effective for anomalies within a global context. However, it incurs high computational costs for long, high-dimensional sequences.

MTAD-GAT [24] applies Graph Attention Networks to multivariate time series, separately learning inter-variable and temporal dependencies, and combines prediction and reconstruction errors for detection. InterFusion [25] uses hierarchical VAEs to model these dependencies, also serving as a reconstruction-fusion model. In summary, hybrid models generally achieve higher detection accuracy through multi-task synergy but introduce increased model complexity and computational overhead. Designing efficient fusion mechanisms remains a critical research direction.

4.4 Classification Based on Time and Frequency Domains

From the perspective of data transformation, time-series anomaly detection can be divided into Time-domain methods and Frequency-domain methods. Most of the aforementioned approaches analyze data directly in the time domain. In contrast, frequency-domain methods first transform the time series into the frequency domain, utilizing spectral analysis to identify anomalous patterns. Recently, hybrid methods combining both domains have emerged to leverage complementary information for enhanced detection performance.

Table 2. Deep Learning-Based Time-Series Anomaly Detection

Model Name	Advantageous Scenarios	Limitations
LSTM-NDT / LSTM-AE	Automatically learns main feature structures;	Limited reconstruction capacity for highly complex patterns; Potential reconstruction bias
OmniAnomaly	Captures uncertainty;	Complex training
TranAD	Strong robustness;	Complex model structure
Anomaly Transformer	Detecting anomalies based on global associations rather than point values	High computational cost
TimesNet	Handles complex multi-periodicity effectively	Relies on the accuracy of period extraction via FFT
MTAD-GAT	Explicitly models inter-variable dependencies and temporal dependencies	Efficiency needs improvement on very large-scale datasets
USAD/MAD-GAN	Improving reconstruction stability (USAD);	GAN training is notoriously unstable (Mode Collapse, Gradient Vanishing)

4.4.1 Time-Domain Based Anomaly Detection Models

Time-domain methods detect anomalies by directly utilizing the variation of raw time-series values over time. This category encompasses nearly all the statistical, machine learning, and deep learning models discussed previously. For instance, using an ARIMA model to predict future values and detecting residual anomalies is a typical time-domain approach. Similarly, LSTM prediction models, Autoencoder (AE) reconstruction models, and Transformer attention models all perform modeling and analysis on the sequence itself within the temporal domain. In the time domain, the focus is placed on features such as trends, seasonality, and abrupt changes (mutations) of data points along the time axis. If a data point at a certain moment significantly deviates from the historical normal range, it is flagged as an anomaly. The advantage of these methods is that they are intuitive and minimize information loss, as all detections are performed directly on the raw time series. However, pure time-domain methods sometimes fail to detect anomalies that manifest primarily as changes in frequency components. For example, certain faults in machine vibration signals may result in increased energy in high-frequency components without significant changes in the amplitude of the time-domain waveform; in such cases, pure time-domain methods may miss the detection.

4.4.2 Frequency-Domain Based Anomaly Detection Models

Frequency-domain based anomaly detection methods map raw time series into the frequency domain to extract spectral features, thereby identifying anomalous changes in periodic components, shifts in frequency energy distribution, or significant alterations in frequency-domain structures. Borrowing from Fourier analysis in signal processing, these methods utilize tools such as the Discrete Fourier Transform (DFT) or Fast Fourier Transform (FFT) to decompose time-domain signals into a series of sinusoidal frequency components, enabling the observation of periodic patterns, noise characteristics, and frequency structures. The advantage of frequency-domain methods lies in their ability to reveal periodic anomalies and frequency shifts that are difficult to observe directly in the time domain, making them particularly suitable for detecting periodic faults, changes in vibration patterns, and periodic interference.

1) Fundamentals of Spectral Analysis and Classical Methods. The core of frequency-domain analysis lies in the extraction and comparison of frequency components. FFT allows for the rapid calculation of spectral information from time-series signals, discovering anomalies via changes in spectral amplitude or phase. Reference [26] proposed the Spectral Residual (SR) method, which first introduced the concept of visual saliency into time-series anomaly detection. By taking the logarithm of the amplitude spectrum of the Fourier-transformed sequence, calculating the residual spectrum, and then transforming it back to the time domain via inverse FFT, the method generates an anomaly score sequence. SR demonstrates robust response capabilities to periodic patterns and frequency jumps, showing better performance than traditional statistical algorithms in Microsoft's online monitoring services.

An important direction in frequency-domain detection is the significance analysis of frequency component changes. While traditional Fourier analysis reveals frequency distribution, it struggles to determine whether a change in a specific frequency is statistically significant. To address this, Reference [27] proposed applying a selective inference framework to detect frequency change points in time series, calculating p-values for frequency domain changes to measure statistical significance. This approach not only detects genuine structural changes in the frequency spectrum but also reduces false positives, providing a more reliable basis for anomaly judgment and aiding root cause analysis in complex systems.

2) Fusion of Time-Domain and Frequency-Domain. Since frequency-domain analysis alone does not preserve the temporal precision of anomaly occurrences (providing only statistical information on overall frequency components), it suffers from limitations in precisely locating anomalies. Recently, numerous studies have proposed frameworks that jointly model time and frequency domain information to improve detection accuracy and localization. Reference [28] proposed Dual-TF, a representative model that addresses the granularity difference between the two domains using a nested sliding window strategy: an outer window processes time-domain information, while an inner window constructs spectra for the corresponding period. By aligning anomaly scores from both domains, the frequency analysis results can be mapped finely to each time point, reducing localization errors caused by frequency granularity and achieving significantly improved performance on multiple benchmark datasets. Theoretically, this method solves the alignment problem between frequency-domain and time-domain scores.

Reference [29] proposed TFCLNet, which

adopts a dual-branch architecture to simultaneously extract time and frequency features, subsequently fusing this multi-domain information for anomaly discrimination. The frequency branch obtains spectral information via frequency transformation to capture periodic changes and frequency patterns, while the time branch extracts temporal dependency features. This design enables the model to recognize long-term periodic anomalies while accounting for short-term abrupt anomalies, enhancing overall accuracy and robustness.

Similarly emphasizing a time-frequency joint strategy is TFAD [30]. This architecture designs modules for time-series decomposition and time-frequency transformation, extracting complete features of normal sequences by utilizing both time-domain prediction and frequency-domain analysis to perform anomaly judgment under multi-task conditions. By combining features from both domains in its structural design, TFAD possesses stronger discriminative power for pattern anomalies, demonstrating distinct advantages when data contains complex external periodic variations.

In adversarial environments, anomalous data bias and distribution drift can affect model training. TFMAE [31] employs a dual-channel strategy with time-domain masking and frequency-domain masking within an autoencoder architecture to pre-suppress potential anomalies, thereby preventing anomalous samples from interfering with the learning of normal patterns. Furthermore, by replacing traditional reconstruction error with a contrastive learning objective, TFMAE achieves higher robustness regarding frequency features and has achieved state-of-the-art results on multiple real-world multivariate datasets (e.g., SWaT, SMD, SMAP, MSL).

3) Deep Learning Enhanced Frequency-Domain Models. To fully leverage frequency information and the expressive power of deep learning, recent research has incorporated frequency features into neural networks. The F-SE-LSTM [32] model first constructs a frequency-domain matrix representation using FFT, extracts weight relationships within and between frequency channels via a Squeeze-and-Excitation (SE) network, and then combines this with LSTM to learn temporal dependencies. Experimental results show that this structure better distinguishes between normal and anomalous patterns on benchmarks like Yahoo Webscope S5 and NAB, improving both accuracy and efficiency.

Models based on deep attention mechanisms also play a significant role. FDTAD [33] introduces frequency-domain augmentation and time-series decomposition mechanisms into a standard Transformer architecture, enabling the model to simultaneously focus on long-term temporal dependencies and frequency changes, thereby improving generalization on drifting data.

4) Fine-grained Frequency Features and Channel-Aware Mechanisms. In multivariate scenarios, frequency-domain correlations between different variable channels are crucial clues for identifying complex anomalies. CATCH [34] proposes decomposing the frequency domain into multiple "frequency patches" and introduces a Channel Fusion Module (CFM) to perceive spectral associations across different channels. Through patch-level mask generation and attention mechanisms, the model automatically learns the most relevant frequency features between channels, improving the identification of fine-grained frequency changes and local anomalies. This method has demonstrated superior performance on various real and synthetic datasets and represents a typical approach for comprehensively considering channel relationships in multivariate frequency-domain detection.

5) Frequency Models for Multi-Pattern Normality Learning. In complex systems like industrial and cloud services, normal patterns under different services or states may differ significantly, making it difficult for a single model to capture all of them. Addressing this, References [35] proposed MACE (Multi-Pattern Normalities in the Frequency Domain). This method establishes normal subspaces for different services or patterns in the frequency domain and detects anomalies by measuring the distance between observed data and these subspaces. This approach not only unifies the handling of multiple normal patterns but also significantly improves efficiency by exploiting the sparsity and parallelism of frequency features. To enhance the detection of short-term anomalies, Peak Convolution and Valley Convolution mechanisms were introduced, improving sensitivity to transient anomalies in industrial monitoring.

In summary, frequency-domain based MTSAD methods have developed rapidly—from traditional Fourier transforms and wavelet analysis to frequency-domain enhanced models incorporating deep learning, and further to time-frequency fusion and adversarial learning mechanisms. These diverse methods have merged to form a research direction that balances frequency features, temporal dependencies, and complex pattern modeling. They exhibit unique advantages in revealing periodic anomalies, detecting spectral change points, and modeling multi-pattern normal behaviors, providing solid technical support for improving the accuracy and robustness of anomaly detection in industrial monitoring, fault diagnosis, and complex systems.

Table 3: Frequency-Domain Based Time-Series Anomaly Detection Models (Note: Please insert your table here)

Model Name	Advantageous Scenarios	Limitations
TFCLNet	Environments with severe distribution shift and high noise levels	Computational complexity can be significantly high
TFMAE	Scenarios where training data contains a large number of unknown anomalies	Requires precise tuning of the masking ratio
FDTAD	IoT sensors and systems with distinct physical laws/principles	Decomposition effectiveness depends on the prominence of periodicity
CATCH	Complex CPS (Cyber-Physical Systems) with a massive number of sensors (>100)	Complex optimization process (e.g., bi-level optimization)
MACE	Cloud services with high real-time requirements and diverse patterns	May sacrifice subtle local features to gain processing speed
Dual-TF	Tasks requiring precise identification of anomaly onset and offset	Complex window alignment logic
Yamada SI	High-stakes domains (Healthcare or Finance) requiring P-value reporting	Calculating P-values may incur additional computational overhead

4.5 Comparative Summary of Different Methods

Synthesizing the classifications above, different methods have distinct applicable scenarios. Statistical methods are simple and fast but rely on strong assumptions; machine learning methods are flexible but require feature engineering; deep learning methods offer high precision but incur large training overheads; frequency-domain methods excel at periodic analysis, while time-domain methods are superior for trend detection. The aforementioned methods have been validated on public datasets such as SMAP, MSL, SMD, and SWaT. Among them, NASA's SMAP and MSL satellite telemetry data, the secure water treatment system SWaT data, and the server machine dataset SMD are widely used as benchmarks. It can be observed that early methods (e.g., LSTM-NAT, DAGMM) proposed solutions targeting either time-series prediction or deep compression individually, whereas, with research progression, models have gradually fused multiple mechanisms (e.g., OmniAnomaly combining VAE and RNN; TranAD and Anomaly Transformer combining reconstruction, prediction, and attention mechanisms). The overall trend indicates that while model complexity is increasing, detection performance and robustness are also improving significantly.

5. Conclusion and Future Prospects

Multivariate Time-Series Anomaly Detection (MTSAD) has made significant progress in industrial equipment monitoring applications, particularly on datasets such as SMAP, MSL, SMD, and SWaT. As the complexity of industrial equipment monitoring continues to increase, traditional statistical and machine learning methods have gradually exposed deficiencies in handling high-dimensional, non-linear, and multivariate data, whereas the emergence of deep learning methods has significantly enhanced detection accuracy and robustness. Statistical-based methods, such as ARIMA and control charts, while simple to implement and computationally efficient, rely on stationarity and linearity assumptions. They fail to effectively capture complex patterns, performing poorly especially in high-dimensional and non-linear data scenarios.

Machine Learning-based methods, such as distance-based, density-based, and one-class classification methods, can automatically learn data patterns without supervision. They are widely used in industrial applications—algorithms like LOF and Isolation Forest are particularly suitable for variable and complex data—but still require manual feature setting or distance metrics and face the "curse of dimensionality" in high-dimensional data.

Deep Learning-based methods, especially reconstruction-based, prediction-based, and reconstruction-prediction hybrid models, have significantly improved the capability to detect complex temporal patterns. Neural network models such as Autoencoders, LSTM, GRU, and Transformers effectively learn non-linear features and long-range dependencies, making them particularly suitable for detecting dynamic anomalies in industrial monitoring. In particular, hybrid models like TranAD and Anomaly Transformer, which combine reconstruction error and prediction residual, can simultaneously capture static patterns and dynamic changes in time-series data, further elevating anomaly detection precision.

Frequency-domain methods, as an emerging direction, effectively capture periodic or frequency component anomalies by analyzing data in the frequency domain. Hybrid methods combining time and frequency domains (e.g., Dual-TF and CATCH) demonstrate superior anomaly detection capabilities, proving more effective than pure time-domain methods, especially when dealing with periodic faults.

In conclusion, deep learning methods—particularly hybrid models based on reconstruction and prediction—have become the mainstream approach in the field of time-series anomaly detection, demonstrating greater flexibility and accuracy. However, in multivariate time-series data, as data dimensionality increases, model complexity and computational overhead also rise. Therefore, balancing model complexity with detection efficiency, especially in real-time monitoring scenarios, remains a hotspot for current research.

Future research could focus on: enhancing model interpretability, which is crucial for decision support in industrial scenarios; integrating more data sources (such as frequency domain, graph structures, etc.) to strengthen model robustness and generalization; and developing more efficient training methods to reduce dependency on massive labeled data and computational resources. Continuous innovation in MTSAD methods will contribute to improving the intelligence level of industrial equipment monitoring systems, driving developments in equipment fault diagnosis and production efficiency optimization.

References

[1] GARG A, ZHANG W Y, SAMARAN J, et al. An evaluation of anomaly detection and diagnosis in multivariate time series[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(6): 2508-2517.

[2] JIANG J R, KAO J B, LI Y L. Semi-supervised time series anomaly de-tection based on statistics and deep learning [J]. Applied Sciences, 2021, 11(15): 6698.

[3] Mathur A P, Tippenhauer N O. SWaT: A water treatment testbed for research and training on ICS security[C]//2016 international workshop on cyber-physical systems for smart water networks (CySWater). IEEE, 2016: 31-36.

[4] Ahmed C M, Palleti V R, Mathur A P. WADI: a water distribution testbed for research in the design of secure cyber physical systems[C]//Proceedings of the 3rd international workshop on cyber-physical systems for smart water networks. 2017: 25-28.

[5] Su Y, Zhao Y, Niu C, et al. Robust anomaly detection for multivariate time series through stochastic recurrent neural network[C]//Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019: 2828-2837.

[6] Hundman K, Constantinou V, Laporte C, et al. Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding[C]//Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. 2018: 387-395.

[7] Katser I D, Kozitsin V O. Skoltech anomaly benchmark (skab)[J]. Kaggle, 2020.

[8]. Vieira R G, Leone Filho M A, Semolini R. An Enhanced Seasonal-Hybrid ESD technique for robust anomaly detection on time series[C]//Simpósio Brasileiro De Redes De Computadores E Sistemas Distribuídos (SBRC). SBC, 2018: 281-294.

[9] ZHANG Y, MERATNIA N, HAVINGA P. Outlier detection techniques for wireless sensor networks: A survey [J]. IEEE Communications Surveys & Tutorials, 2010, 12(2): 159-170.

[10] Münz, G., Li, S., & Carle, G. (2007). Traffic anomaly detection using k-means

clustering. GI/ITG Workshop MMBnet, 13–14.

[11] CHESNOKOV M Y. Time series anomaly searching based on DB-SCAN ensembles[J]. Scientific and Technical Information Processing, 2019, 46(5): 299-305.

[12] Ding Z, Fei M. An anomaly detection approach based on isolation forest algorithm for streaming data using sliding window[J]. IFAC Proceedings Volumes, 2013, 46(20): 12-17.

[13] Malhotra P. LSTM-based Encoder-Decoder for Multi-sensor Anomaly Detection[J]. arXiv preprint arXiv:1607.00148, 2016.

[14] Zhang C, Song D, Chen Y, et al. A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data[C]//Proceedings of the AAAI conference on artificial intelligence. 2019, 33(01): 1409-1416.

[15] Zong B, Song Q, Min M R, et al. Deep autoencoding gaussian mixture model for unsupervised anomaly detection[C]//International conference on learning representations. 2018.

[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. 2020: 3395-3404.

[17] 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. Cham: Springer International Publishing, 2019: 703-716.

[18] Geiger A, Liu D, Alnegheimish S, et al. Tadgan: Time series anomaly detection using generative adversarial networks[C]//2020 ieee international conference on big data (big data). IEEE, 2020: 33-43.

[19] Ding N, Ma H X, Gao H, et al. Real-time anomaly detection based on long short-Term memory and Gaussian Mixture Model[J]. Computers & Electrical Engineering, 2019, 79: 106458.

[20] Munir M, Siddiqui S A, Dengel A, et al. DeepAnT: A deep learning approach for unsupervised anomaly detection in time series[J]. Ieee Access, 2018, 7: 1991-2005.

[21] Wu H, Hu T, Liu Y, et al. Timesnet: Temporal 2d-variation modeling for general time series analysis[J]. arXiv preprint arXiv:2210.02186, 2022.

[22] 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.

[23] Xu J, Wu H, Wang J, et al. Anomaly transformer: Time series anomaly detection with association discrepancy[J]. arXiv preprint arXiv:2110.02642, 2021.

[24] Zhao H, Wang Y, Duan J, et al. Multivariate time-series anomaly detection via graph attention network[C]//2020 IEEE international conference on data mining (ICDM). IEEE, 2020: 841-850.

[25] Hu J, Li T, Zhao M, et al. Hyperspectral image superresolution via deep structure and texture interfusion[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 8665-8678.

[26] REN H, XU B, WANG Y, et al. Time-series anomaly detection service at Microsoft[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, NY, USA: ACM, 2019: 3009-3017.

[27] AMADA A, SHIRAISHI T, NISHINO S, et al. Change point detection in the frequency domain with statistical reliability[J]. Transactions on Machine Learning Research, 2025.

[28] CHEN Y, WANG H, HONG M, et al. Dual-TF: A dual-domain time-frequency framework for time-series anomaly detection[J]. IEEE Transactions on Big Data, 2024, 10(2): 184-196.

[29] ZHANG Y, LI X, WANG J. TFCLNet: Time-frequency contrastive learning network for time series anomaly detection[J]. IEEE Transactions on Instrumentation and Measurement, 2025, 74: 1-12.

[30] Zhang C, Zhou T, Wen Q, et al. Tfad: A decomposition time series anomaly detection architecture with time-frequency analysis[C]//Proceedings of the 31st ACM international conference on information & knowledge management. 2022: 2497-2507.

[31] Fang Y, Xie J, Zhao Y, et al. Temporal-frequency masked autoencoders for time series anomaly detection[C]//2024 IEEE 40th International Conference on Data Engineering (ICDE). IEEE, 2024: 1228-1241.

[32] Lu Y X, Jin X B, Chen J, et al. F-SE-LSTM: A Time Series Anomaly Detection Method with Frequency Domain Information[J]. arXiv preprint arXiv:2412.02474, 2024.
[33] Li S, Wang Z, Wang X, et al. Frequency-enhanced and decomposed transformer for multivariate time series anomaly detection[J]. Applied Intelligence, 2025, 55(6): 1-18.
[34] Wu X, Qiu X, Li Z, et al. Catch: Channel-aware multivariate time series anomaly detection via frequency patching[J]. arXiv preprint arXiv:2410.12261, 2024.
[35] Chen F, Zhang Y, Qin Z, et al. Learning multi-pattern normalities in the frequency domain for efficient time series anomaly detection[C]//2024 IEEE 40th International Conference on Data Engineering (ICDE). IEEE, 2024: 747-760.

