Deep Learning for Anomaly Detection in IoT Time Series

Jiange Jiang[a,f,*], Chen Chen[a], Yanwei Xu[c], Pu Li[c], Fan Jin[d], Dingye Ning[e], Ilir Murturi[f], and Schahram Dustdar[f]

[a]School of Telecommunications Engineering, Xidian University, Xi'an, China;
[b]College of Intelligence and Computing, Tianjin University, Tianjin, China;
[c]College of Information Science & Electronic Engineering, Zhejiang University, Hangzhou, China;
[d]National Key Laboratory of Science and Technology on Space Microwave, CAST, Xi'an, China;
[e]State Grid Jingdezhen Power Supply Company, Jingdezhen Jiangxi, China;
[f]Distributed Systems Group, TU Wien, Vienna, Austria;
[*]Corresponding author.

ARTICLE HISTORY
Compiled October 17, 2024

ABSTRACT
With the widespread adoption of the Internet of Things (IoT), time series data is being generated in massive quantities across various industries. In IoT systems, detecting anomalies in time series data is crucial for ensuring reliability, security, and efficiency. This review provides a comprehensive overview of anomaly detection techniques in IoT time series data, categorizing anomalies into three main types: point anomalies, collective anomalies, and contextual anomalies. Firstly, it discusses the significance of anomaly detection in IoT, highlighting its importance in diverse applications such as smart grid, network, financial, etc. Subsequently, it surveys commonly used anomaly detection methods including statistical approaches, machine learning algorithms, and deep learning models, outlining their principles, advantages, and limitations. Furthermore, challenges and future directions in IoT anomaly detection are discussed, addressing issues such as data heterogeneity, scalability, interpretability, and real-time processing. This review serves as a valuable resource for researchers, practitioners, and stakeholders interested in understanding and deploying anomaly detection techniques for IoT time series data.

KEYWORDS
Anomaly detection; Time series; Deep learning; Internet of Things

1. Introduction

In the Internet of Things (IoT), interconnected devices generate vast amounts of time series data. This data provides insights into operational performance, environmental conditions, and human behavior. Effective time series anomaly detection is crucial for identifying deviations from expected patterns within these sequences of data points. Across domains such as finance [45, 27], industrial manufacturing [41, 34], healthcare

Email: JiangeJiang@stu.xidian.edu.cn; cc2000@mail.xidian.edu.cn; xuyanwei@tju.edu.cn; 11931084@zju.edu.cn; jinf@cast504.com; 295264174@qq.com; imurturi@dsg.tuwien.ac.at; dustdar@dsg.tuwien.ac.at.

[67], and cybersecurity [52, 19], where time series data pervades every facet of operations, understanding and detecting anomalies within these data streams are crucial for ensuring the integrity, reliability, and security of systems and processes. For instance, in sectors like manufacturing and healthcare, where system reliability is paramount, real-time anomaly detection ensures the smooth operation of processes [26], reduces downtime, and enhances overall productivity. Furthermore, by promptly detecting anomalies, businesses can mitigate potential risks such as fraud, equipment failures, network intrusions [46], or health complications, thereby minimizing financial losses and operational disruptions.

However, the inherent temporal correlation in time series render anomaly detection more challenging compared to image anomaly detection. Firstly, time series data often contains high levels of noise and volatility, including various forms of random disturbances or periodic fluctuations, making it more difficult to identify anomalies accurately within the data. Furthermore, anomalies in time series data are typically rare events relative to normal conditions, leading to data imbalance, which makes models more susceptible to the influence of normal data, thereby reducing the accuracy and reliability of anomaly detection. Lastly, Some anomalies may arise from long-term dependencies, necessitating models that can effectively capture and understand long-term sequence patterns. However, time series data often suffers from missing or incomplete data, which may result from sensor failures, communication interruptions, or data recording errors. Addressing this data incompleteness requires appropriate techniques for data imputation or interpolation to ensure the accuracy and reliability of anomaly detection.

The traditional methods for anomaly detection are usually based on statistical principles and machine learning algorithm, which learn patterns and rules from time series data [37]. However, there are several limitations when applied to anomaly detection in IoT time series. Firstly, these methods often require manual feature engineering [80], where domain expertise is necessary to select and engineer relevant features from the raw data. This process can be time-consuming, labor-intensive, and may not fully capture the complex temporal patterns present in IoT time series data. Secondly, traditional machine learning algorithms typically assume that data instances are independent and identically distributed (i.i.d.), which may not hold true for time series data where sequential dependencies exist [44, 49]. This can lead to suboptimal performance and difficulty in capturing the temporal dynamics inherent in IoT environments. Additionally, traditional machine learning methods may struggle with handling high-dimensional and heterogeneous time series data, such as those generated by diverse sensors in IoT systems [3]. These methods may not effectively capture the diverse characteristics and correlations present in such data, leading to reduced accuracy in anomaly detection.

The emergence of deep learning techniques has revolutionized time series anomaly detection, addressing many of these challenges. Deep learning models, such as Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), and Convolutional Neural Networks (CNNs), can automatically learn intricate patterns and dependencies from raw time series data without the need for manual feature engineering. In [81], Zhang et al. proposed a novel deep learning-based anomaly detection algorithm, the Deep Convolutional Autoencoding Memory network (CAE-M), which combines a Deep Convolutional Autoencoder with a Memory Network to capture spatial and temporal dependencies in multi-sensor time-series data. It can achieve better performance when intra-class data exhibit similar distributions or regular patterns. Similarly, authors introduced an architecture that integrates a CNN with a LSTM-based autoencoder to enhance anomaly detection in time series data [76]. These models excel at capturing complex temporal dynamics and dependencies, thus overcoming the limitations of traditional machine learning methods in IoT time series anomaly detection. Additionally, deep learning models can handle high-dimensional and heterogeneous time series data more effectively, leading to improved accuracy in anomaly detection tasks. As a result, the application of deep learning in time series anomaly detection represents a promising direction for enhancing anomaly detection capabilities in IoT environments.

Several survey papers have delved into the realm of anomaly detection in time series using Artificial Intelligence (AI) methodologies, including notable works such as the study by Cook et al.[11] and Li et al.[36]. However, existing literature often falls short in providing a holistic exploration of the nuanced challenges and advancements specific to this domain. This paper endeavors to address this gap by conducting a meticulous review and synthesis of deep learning techniques tailored for anomaly detection in IoT time series data. Through this endeavor, it aims to elucidate the unique intricacies and emerging trends in this interdisciplinary field, offering valuable insights to researchers, practitioners, and stakeholders alike.

The reminder of this paper is organized as follows. Section 2 explore the categories of time series anomalies and their definitions, providing a comprehensive understanding of the different types of anomalies encountered in time series data. In Section 3, we summarize the methods for time series anomaly detection, encompassing both traditional statistical approaches and modern machine learning techniques. Section 4 focus on the application of time series anomaly detection in IoT. Section 5 examine the challenges inherent in time series anomaly detection and propose future research directions. Finally, Section 6 concludes this paper.

2. Taxonomy of Time Series Anomalies

Time series anomalies refer to events or data points within time series data that deviate significantly from the expected patterns. These anomalies may manifest as sudden spikes, persistent irregular patterns, or significant deviations from the data distribution. We categorize time series anomalies into three distinct types based on their characteristics and occurrence patterns: point anomalies, collective anomalies, and contextual anomalies.

2.1. Point Anomalies

Within the domain of time series anomaly detection, point anomalies serve as pivotal indicators, spotlighting individual data points that veer markedly from the anticipated trajectory of the time series. These anomalies stand out for their solitary occurrences, where a sole data point demonstrates an aberration from the customary behavior observed in the time series data. This inherent characteristic of singularity underscores the distinctive nature of point anomalies, wherein a solitary data point exhibits a deviation from the typical pattern observed in the time series data. Figure 2.1 shows an example of point anomalies.

From a mathematical standpoint, the identification of point anomalies entails a rigorous examination of each data point x_i within the time series $X = \{x_1, x_2, ..., x_n\}$ against a predefined threshold ϵ, against which its deviation from the expected pattern is measured. Specifically, if the absolute difference between x_i and the mean or median

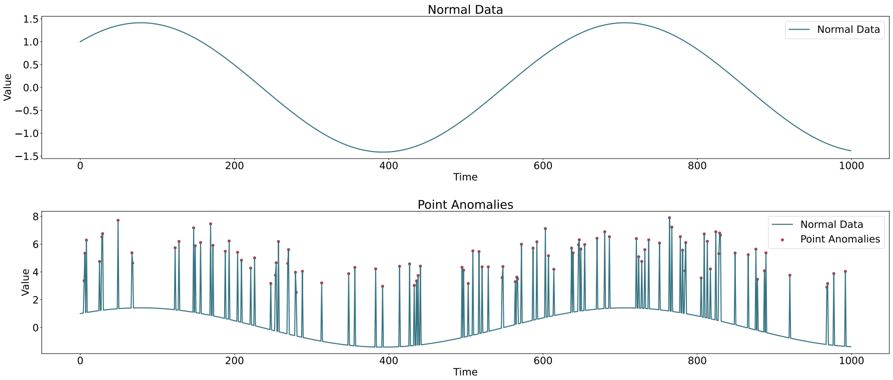

Figure 1. Example of point anomalies.

μ of the time series surpasses ϵ, as expressed by $|x_i - \mu| > \epsilon$, x_i earns the classification of a point anomaly.

The significance of point anomaly detection transcends disciplinary boundaries, finding applications across diverse domains. For example, in cybersecurity, they might signify unauthorized access attempts or suspicious network activity [29]. In financial markets, sudden fluctuations in stock prices or trading volumes could indicate anomalies warranting further investigation [72].

2.2. Collective Anomalies

Within the realm of time series anomaly detection, collective anomalies serve as a critical focal point, shedding light on deviations from anticipated patterns within specific temporal windows or segments of the data. Unlike point anomalies, which pinpoint individual aberrant data points, collective anomalies encapsulate anomalies that manifest within defined time intervals, offering insights into broader trends or patterns of irregularity within the time series. Figure 2.2 shows an example of collective anomalies.

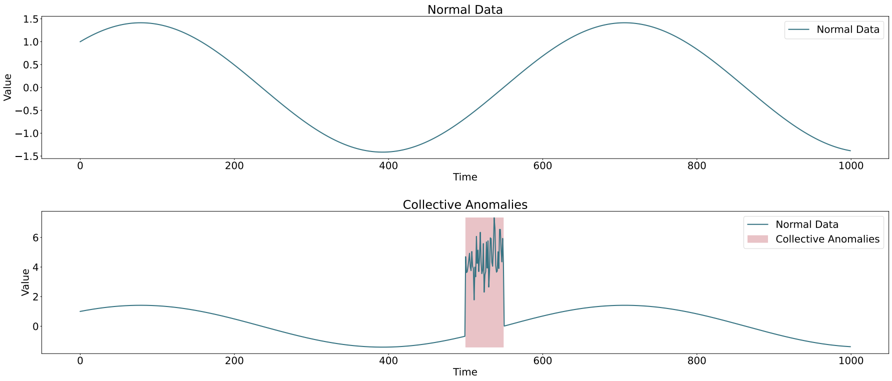

Figure 2. Example of collective anomalies.

From a mathematical perspective, the identification of collective anomalies hinges on discerning deviations in the aggregate behavior or statistical properties of the time series data within a specified time interval. This involves computing the aggregated value, such as the average or sum, of data points within the interval and comparing it to the expected value. If the disparity between the aggregated value A and the expected value μ exceeds a predefined threshold ϵ, as expressed by $|A - \mu| > \epsilon$, the interval is classified as anomalous.

Applications of collective anomaly detection span diverse domains, each presenting unique challenges and opportunities for analysis. For instance, in healthcare monitoring systems, anomalies in patient vital signs may alert healthcare providers to potential medical emergencies or adverse reactions to treatment [62]. Varone et al. investigated the power spectrum density (PSD), in resting-state electroencephalography (EEG), to evaluate the abnormalities in Psychogenic Non-Epileptic Seizures (PNES) affected brains [61]. The detection and analysis of collective anomalies offer valuable insights into temporal patterns of deviation within time series data, facilitating proactive decision-making, and risk management across various applications and industries.

2.3. Contextual Anomalies

Additionally, contextual anomalies emerge as a nuanced category, delineating instances where a data point deviates from the anticipated pattern within a specific context or subset of the time series data. Unlike point and collective anomalies, which focus on individual data points or predefined temporal windows, contextual anomalies delve deeper into the underlying context of the data, discerning deviations within specific subsets based on contextual information. Figure 2.3 shows an example of contextual anomalies.

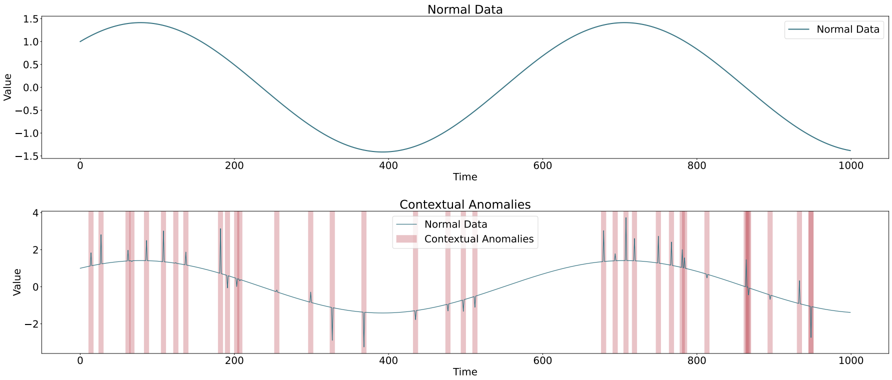

Figure 3. Example of contextual anomalies.

From a mathematical perspective, the identification of contextual anomalies hinges on scrutinizing each data point x_i within the time series X against a predefined threshold ϵ, against which its deviation from the expected behavior within a specific context or subset C is measured. Specifically, if the absolute difference between x_i and the mean or median μ_C of the subset C surpasses ϵ, as expressed by $|x_i - \mu_C| > \epsilon$, x_i garners the classification of a contextual anomaly.

The application landscape of contextual anomaly detection spans diverse domains, where contextual information plays a pivotal role in anomaly identification and interpretation. For instance, in network security, contextual anomalies may manifest as unusual patterns of network activity within specific user groups, departments, or geographical locations, indicating potential security threats or policy violations [51].

Similarly, in supply chain management, contextual anomalies could signify irregularities in product demand or inventory levels within specific regions or market segments, necessitating adjustments in production and distribution strategies [6].

In summary, different time series anomaly scenarios can exhibit diverse characteristics and nuances. For instance, in cybersecurity, point anomalies might indicate isolated instances of malicious access attempts or suspicious network activity, whereas collective anomalies could signify abnormal patterns of data transfer rates within specific time periods, potentially signaling network congestion or security breaches. Similarly, in financial markets, point anomalies may represent anomalous trading behaviors or market irregularities, while contextual anomalies might involve deviations in trading patterns across different geographical regions or user groups. In manufacturing, point anomalies could denote sudden deviations in machine sensor readings, indicating equipment malfunctions or production line interruptions. Meanwhile, collective anomalies might reveal prolonged periods of decreased productivity within specific shifts or production cycles, hinting at underlying process inefficiencies or resource constraints. In contrast, contextual anomalies may manifest as irregularities in product quality across different manufacturing plants or geographical locations, highlighting variations in production processes or environmental factors. Therefore, when describing the application scenarios of different anomaly types, it is essential to consider the specific requirements and nuances of each domain, ensuring that each anomaly type's cases adequately showcase their uniqueness and significance.

3. Methods of anomaly detection

3.1. Statistical

Statistical-based time series anomaly detection methods typically involve establishing a baseline model to describe the normal patterns of time series data and then identifying anomalies based on the deviation or residuals between the data and the model.

3.1.1. Methods Based on Threshold

The threshold-based method relies on setting a predefined threshold, above or below which data points are classified as anomalous. Anomalies are identified based on whether they exceed or fall below this threshold. This approach assumes that anomalies exhibit significant deviations from normal data and can be captured using a straightforward threshold. Let TPR represent the True Positive Rate and FPR represent the False Positive Rate. The threshold is set to T, then:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{1}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}, \tag{2}$$

where TP represents True Positives (correctly detected anomalies), FP represents False Positives (false alarms, normal data incorrectly classified as anomalies), TN represents

True Negatives (correctly detected normal data), and FN represents False Negatives (missed anomalies).

This method is commonly applied in scenarios where anomalies are expected to display extreme values compared to normal data. In [14], Akande et al. propose an anomaly detection approach using a threshold to discriminate between regular and aberrant log files. It finds utility in domains where anomalies are distinct and easily distinguishable from normal behavior. Furthermore, the threshold-based method offers simplicity in implementation and understanding. It provides an intuitive approach by offering a clear distinction between normal and anomalous data points. Additionally, it demands minimal computational resources, enhancing computational efficiency. However, the method is sensitive to threshold selection, with its performance highly reliant on the chosen threshold, which may vary across different datasets. Moreover, it lacks flexibility, making it unsuitable for detecting anomalies with complex patterns or subtle deviations from normal behavior. Additionally, it overlooks contextual information, potentially leading to false alarms or missed anomalies.

3.1.2. Methods Based on Time Series Decomposition

Time series decomposition involves breaking down a time series into its constituent components, such as trend, seasonality, and residual. Anomalies are typically identified in the residual component, as it represent the part of the data that cannot be explained by trend and seasonality. This method compares the residual with normal residual patterns to detect anomalies. Let $y(t)$ represent the original time series, $T(t)$ represent the trend component, $S(t)$ represent the seasonality component, and $R(t)$ represent the residual component, then:

$$y(t) = T(t) + S(t) + R(t). \tag{3}$$

In time series decomposition, a commonly used method is STL (Seasonal-Trend Decomposition using Loess), where the residual is obtained through Local Weighted Regression (Loess). The formula for Loess is:

$$\hat{R}_t = y(t) - \hat{T}(t) - \hat{S}(t), \tag{4}$$

where $\hat{R}_t$ is the residual, $\hat{T}(t)$ is the estimated trend, and $\hat{S}(t)$ is the estimated seasonality. It finds utility in applications where anomalies are expected to manifest as deviations from expected patterns after accounting for trend and seasonality. Wu et al. propose a comprehensive analysis framework for anomaly detection in water supply systems, integrating time series decomposition, outlier detection, and quantitative evaluation [69]. Applied to a monitored water supply zone with over 8,000 pipes, the solution achieved a 75% true positive rate and successfully detected 90% of field events with a lead time of more than 24 hours, demonstrating its effectiveness in improving operational management and maximizing the return of investment in smart water grids. It is commonly applied in domains such as finance, environmental monitoring, and energy consumption analysis. This approach allows for the separation of anomalous patterns from underlying trends and seasonal variations, enhancing interpretability. In [82], authors utilize Seasonal-Trend Decomposition to simplify analysis and improve detection performance, demonstrating state-of-the-art results across diverse anomalies. Besides, authors develop a decomposition approach to separate a time series into its latent component in time series anomalies detection analysis, which offers control over type-I errors and robustness against anomalies [9]. It offers robustness and can handle time series data with complex temporal structures and multiple seasonalities. Additionally, it provides insights into the nature of anomalies by analyzing residual patterns. However, time series decomposition relies on the assumption that anomalies are captured in the residual component, which may not always hold true. It is sensitive to decomposition techniques, and performance may vary depending on the choice of method and parameters. Moreover, it requires additional computational resources compared to threshold-based methods due to decomposition processes.

3.1.3. Methods Based on Statistical Modeling

Statistical modeling assumes that time series data follows a specific probability distribution, such as normal or exponential distribution. Anomalies are identified as data points that deviate significantly from the expected distribution according to the model. This method involves fitting a statistical model to the data and assessing the likelihood of observed data under the model. This approach is suitable for applications where anomalies are characterized by deviations from expected statistical properties, such as mean, variance, or distribution shape. In [79], zeng et al. propose a method for anomaly detection in wind turbine gearbox oil temperature, which utilizes Sparse Bayesian Learning (SBL) and hypothesis testing (HT) to estimate variation ranges for different operating conditions and detect anomalies with high reliability. It also finds applications in quality control, network monitoring, and healthcare. Authors utilize denoising diffusion probabilistic models (DDPMs) for anomaly detection, leveraging a novel partial diffusion strategy named AnoDDPM, which outperforms existing methods by incorporating multi-scale simplex noise diffusion for improved anomaly detection [70]. Statistical modeling provides a formal framework for quantifying the likelihood of anomalies based on probability theory. It offers adaptability, as it can accommodate various types of anomalies by selecting appropriate probability distributions or model parameters. Additionally, it is capable of detecting anomalies with diverse patterns and characteristics. However, the performance of statistical modeling heavily depends on the accuracy of the assumed probability distribution, which may not always reflect the true data-generating process. It entails complexity and requires expertise in statistical modeling and model selection, particularly for non-standard distributions or complex data patterns. Moreover, its performance may degrade in the presence of data quality issues, such as outliers or missing values in the dataset.

3.2. Traditional machine leaning

Traditional machine learning methods for time series anomaly detection typically involve extracting relevant features from the time series data and training a model, such as a decision tree (DT), support vector machine (SVM), or neural network, to differentiate between normal and anomalous patterns based on these features. The underlying principle is to capture the distinctive characteristics of anomalies through feature-based representation and learn to distinguish them from normal patterns. These methods rely on the assumption that anomalies exhibit unique feature patterns that can be effectively learned by machine learning models.

3.2.1. Methods Based on Feature

The feature-based method involves extracting various features from time series data and training traditional machine learning models using these features. These models differentiate between normal and abnormal patterns based on the extracted features. This approach is particularly useful in scenarios where anomalies exhibit distinct feature patterns that can be effectively captured and distinguished by machine learning models. Feature-based methods find wide applications across various domains, such as intrusion detection in cybersecurity [12, 54, 75], and fraud detection in financial transactions [53]. Besides, in industrial systems, Dhiman et al. introduce an anomaly detection model for wind turbine gearboxes based on a twin support vector machine (TWSVM), using SCADA data from wind farms in the U.K [16]. They are employed in situations where specific characteristics or patterns serve as indicators of anomalies. These methods offer flexibility in accommodating a wide range of feature types and extraction techniques, allowing for the interpretation of detected anomalies based on the importance of individual features. They can also generalize well to unseen data if the selected features adequately represent the underlying patterns. However, feature-based methods have their drawbacks. They heavily rely on domain knowledge and feature engineering expertise, making them sensitive to the quality and relevance of the selected features. Additionally, they may suffer from the curse of dimensionality when dealing with high-dimensional feature spaces, which can impact their performance.

3.2.2. Methods Based on Clustering

Clustering methods aim to group time series data into different clusters and identify clusters that differ from others as anomalies. The principle behind this approach is that anomalies typically manifest as clusters that are distinct from the majority of the data. These methods find applications in scenarios where anomalies are characterized by unique cluster patterns, such as network traffic analysis, sensor data monitoring, and healthcare monitoring systems. Clustering methods are suitable for scenarios where anomalies can be identified based on their deviation from the majority of data points and are expected to form distinct clusters. In [56], Subudhi et al. introduces a novel approach for detecting fraud in automobile insurance claims using Genetic Algorithm-based Fuzzy C-Means clustering, combined with various supervised classifier models. The method effectively identifies genuine, malicious, and suspicious claims, leading to improved fraud detection in the automobile insurance domain. One advantage of clustering methods is their unsupervised nature, as they do not require labeled data for training, making them suitable for unsupervised anomaly detection scenarios. Ghezelbash et al. propose a genetic algorithm-based clustering method, known as genetic K-means clustering (GKMC) [22], to delineate multi-elemental patterns in stream sediment geochemical data for mineral exploration, demonstrating its effectiveness and reliability in geochemical anomaly recognition. They also offer flexibility in detecting anomalies with diverse patterns and characteristics without relying on predefined anomaly labels. Additionally, clustering methods can handle large volumes of time series data efficiently. However, these methods have their limitations. They require careful selection of the number of clusters, which can be challenging and subjective. Performance may also vary based on the initialization of cluster centroids and the choice of clustering algorithm. Furthermore, the interpretation of detected anomalies may be challenging due to the lack of explicit labels and the inherent complexity of cluster structures.

3.2.3. Methods Based on Distance

Distance-based methods identify anomalies by computing the distance or similarity measures between time series data points, where anomalies are typically identified as data points that are significantly distant from others. The principle behind this approach is that anomalies often exhibit dissimilarities or distant relationships compared to normal data points. Let x_i and x_j represent two data points in the time series data, and $\text{dist}(x_i, x_j)$ represent the distance between them. A common distance measure is the Euclidean distance:

$$\text{dist}(x_i, x_j) = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2}, \tag{5}$$

where n is the number of dimensions/features. These methods find applications in scenarios where anomalies can be detected based on their dissimilarity or distance from normal patterns, such as anomaly detection in sensor networks, environmental monitoring, and outlier detection in financial data. Laskar et al. propose an intrusion detection system for industrial computer networks that combines the K-Means algorithm with the Isolation Forest [35]. Distance-based methods are suitable for scenarios where anomalies can be identified based on their deviation from normal data points and are expected to exhibit distinct distance relationships. Henriques et al. introduce a distance-based anomaly detection model applied in the field of computing and networking systems, effectively processing large log file datasets to identify anomalous events, combining K-Means and XGBoost Models [28]. One advantage of distance-based methods is their intuitiveness, as anomalies are identified based on their distance from normal data points. They also offer flexibility, as they can be applied to various types of time series data without restrictions on specific data distributions. Additionally, distance-based methods are robust and can handle noisy data and outliers effectively. However, these methods have their limitations. Performance heavily depends on the choice of distance metrics, which may vary based on the characteristics of the data. They may also suffer from the curse of dimensionality in high-dimensional feature spaces, leading to degraded performance. Furthermore, computing distances between data points can be computationally intensive, especially for large datasets and high-dimensional feature spaces.

3.3. Deep learning

In the field of deep learning, there are various methods for time series anomaly detection [41]. These methods can be categorized into six main types based on their core concepts and technical characteristics: RNNs-based, CNN-based, Autoencoders-based, GANs-based, Attention-based, and GNNs-based methods.

3.3.1. Methods Based on RNNs

RNNs are a class of neural network models capable of processing sequential data. Their structure includes recurrent connections, allowing the network to remember past information and apply it to current predictions. In time series anomaly detection, RNNs can incrementally process each time step of sequential data and identify abnormal patterns by learning the temporal dependencies between data points [1, 47]. Let x_t represent the t-th data point of the time series and h_t represent the hidden state of the RNN, then the propagation formula for RNN is:

$$h_t = f(x_t, h_{t-1}), \tag{6}$$

where f is the non-linear activation function of the RNN. RNNs are widely applied in scenarios requiring consideration of sequential data, such as natural language processing, speech recognition, and time series prediction [65]. In anomaly detection, RNNs can be applied to various scenarios requiring capturing temporal dependencies, such as power system monitoring [85], network traffic analysis [63], and intelligent health monitoring [17]. In power system fault detection, RNNs can analyze power sensor data to identify abnormal power waveform patterns, such as voltage spikes or frequency anomalies [43]. Additionally, in intelligent health monitoring, utilizing RNNs can analyze patients' physiological signal data, such as heart rate, respiration rate, and body temperature, to detect abnormal physiological patterns like arrhythmias or sudden temperature rises. RNNs are suitable for processing sequential data with long-term dependencies and are sensitive to the temporal order of data. Due to their recurrent structure, RNNs can handle variable-length sequences and demonstrate excellent performance in modeling and predicting sequential data. A common variant of RNNs is LSTM, which captures long-term dependencies by introducing gating units. Another variant is Gated Recurrent Units (GRU), which incorporates gating mechanisms but with a simpler structure and fewer parameters. In time series anomaly detection, these variant models often outperform traditional RNN models as they better capture long-term dependencies in sequential data [40]. Research indicates that using variant RNN models like LSTM or GRU for time series anomaly detection can effectively identify various types of abnormal patterns and generally outperform traditional RNN models in terms of performance and generalization ability [66]. These methods can capture long-term dependencies in time series data, handle sequences of different lengths, and process variable-length sequences. Moreover, RNN-based methods exhibit flexibility and generality in handling various types of time series data. However, these methods also have some drawbacks, including difficulties in training with longer sequences, the problem of vanishing or exploding gradients, and the requirement for large amounts of data and computational resources.

3.3.2. Methods Based on CNNs

CNNs are deep learning models specifically designed for processing image data, but their application to time series data involves capturing local features of sequential data. Through a combination of convolutional and pooling layers, CNNs can effectively extract local patterns and periodic variations in sequential data. CNNs are suitable for time series anomaly detection scenarios requiring capturing local patterns or periodic variations, such as power system fault detection, sensor data anomaly detection, and speech recognition [60]. In sensor networks, CNNs can analyze sensor data streams to detect abnormal sensor measurement patterns, such as abnormal vibration patterns or temperature changes. Additionally, in network traffic analysis, CNNs can analyze network traffic data packets to detect abnormal network communication patterns, such as DDoS attacks or anomalous data transmission behavior [21]. CNNs are suitable for cases where time series data exhibit distinct local features or periodic patterns, and they are insensitive to the overall sequence order. A common variant of CNNs is one-dimensional Convolutional Neural Networks (1D-CNN), specifically designed for processing sequential data. Another variant is CNN models with residual connections, such as ResNet, which can learn deeper features from data and often exhibit better performance and stability. Research suggests that using variant CNN models like 1D-CNN or ResNet for time series anomaly detection can effectively capture local patterns and periodic variations in sequential data, thereby improving the accuracy and robustness of anomaly detection.

3.3.3. Methods Based on Autoencoders

Autoencoders are unsupervised learning models consisting of an encoder and a decoder, capable of encoding input data into a low-dimensional representation and attempting to decode it back to the original data. In time series anomaly detection, the reconstruction error of autoencoders is used to identify abnormal patterns. Autoencoders are widely applied in scenarios requiring dimensionality reduction and reconstruction of data, such as image anomaly detection, financial fraud detection, and abnormal behavior detection [76]. In financial fraud detection, autoencoders can model customer transaction data to identify abnormal transaction patterns, such as abnormal transaction amounts or frequencies. In manufacturing, autoencoders can analyze sensor data from production equipment to detect abnormal production process patterns, such as abnormal temperature changes or pressure fluctuations [38]. Autoencoders are suitable for cases where time series data have a potential low-dimensional representation and abnormal patterns differ significantly from normal patterns. Due to their unsupervised learning nature, autoencoders do not require labeled anomaly data for training, thus exhibiting a degree of flexibility and generality in anomaly detection. A common variant of autoencoders is Variational Autoencoders (VAE), which introduce latent variables to learn the underlying distribution of data and often have better generation and sampling capabilities [42]. Another variant is sparse autoencoders, which learn more compact data representations by introducing sparsity constraints. Research indicates that using variant autoencoder models like VAE or sparse autoencoders for time series anomaly detection can effectively capture the latent features of data and generally have better anomaly detection performance and generalization ability.

3.3.4. Methods Based on GANs

Generative Adversarial Networks (GANs) consist of a generator and a discriminator. The generator attempts to generate data samples similar to the normal data distribution, while the discriminator tries to distinguish between generated fake samples and real samples. Through adversarial training, the generator gradually learns to generate samples closer to real data, while the discriminator becomes better at distinguishing between real and fake samples. Anomalies are considered data that deviates from the generated data. GANs are widely applied in fields such as image generation, image restoration, and style transfer. In time series anomaly detection, they can be used to generate data samples similar to normal data, thereby detecting anomalous data [20]. A typical application is Anomaly Detection with Generative Adversarial Networks (AnoGAN) in time series anomaly detection [5]. AnoGAN learns the distribution of normal data by training the generator and discriminator to generate samples similar to normal data. During the generation process, anomalous data typically cannot generate samples similar to normal data, thereby enabling the identification of anomalies based on the output of the discriminator. These methods are suitable for cases where time series data have complex data distributions and significant differences between abnormal and normal patterns. GAN-based models improve the performance and stability of GANs by introducing conditional information, improving loss functions, or implementing cross-domain data transformation. GANs-based models have achieved good results in some time series anomaly detection tasks. They can effectively generate data samples similar to the normal data distribution and distinguish anomalous data. However, the training process of GANs is typically complex, requiring fine-tuning of hyperparameters and training techniques, and also demanding high-quality and feature-rich data.

3.3.5. Methods Based on Attention

Attention mechanisms allow models to dynamically focus on the most important parts when processing sequential data, thereby enhancing the model's ability to detect anomalies. Anomalies typically manifest as abnormal patterns that attract attention. Attention mechanisms are widely applied in natural language processing, speech recognition, bioinformatics, etc. In time series anomaly detection, they can be used to capture key features in sequential data and improve anomaly detection accuracy [30]. A common application is attention-based RNN models like the Transformer model. The Transformer model dynamically captures dependencies between different positions in sequence data through self-attention mechanisms, thereby enhancing the perception of anomalies [59]. These methods are suitable for cases where time series data contain key features or patterns, and there are significant differences between abnormal and normal patterns. Variant attention mechanism models include multi-head attention mechanisms, variants of self-attention mechanisms, etc. These variant models further enhance the performance and generalization ability of models by introducing different attention mechanism designs. Models based on attention mechanisms have achieved significant results in tasks such as language modeling, sequence generation, text classification, etc. These can dynamically focus on important parts of sequence data, improving the accuracy of anomaly detection. However, these models typically require substantial computational resources and large-scale datasets for training and are sensitive to the selection of model structures and hyperparameters. These methods can dynamically focus on important parts of sequence data, improving the accuracy of anomaly detection. However, they require a large amount of training data and complex model structures and may suffer from overfitting, requiring a high level of understanding and modeling capability for sequence data.

3.3.6. Methods Based on GNNs

Graph neural networks can capture relationships and features between nodes in sequential data, identifying anomalous patterns by learning information from the graph structure. In time series data, sequential data can be represented as a graph structure, where nodes represent data points and edges represent relationships or similarities between data points. Graph neural networks are widely applied in social network analysis, traffic flow analysis, bioinformatics, etc [15, 68]. In time series anomaly detection, they can be used to capture node relationships and features in sequential data, thereby identifying anomalous patterns. A common application is the Graph Attention Network (GAT) based on graph neural networks. The GAT model effectively captures relationships and features between nodes in the graph structure through self-attention mechanisms, thus capturing abnormal patterns in sequential data [84]. These methods are suitable for cases where sequential data can be represented as a graph structure and abnormal patterns typically manifest as anomalous relationships between nodes or edges. Variant graph neural network models include Graph Convolutional Networks (GCNs) [4], GraphSAGE [7], Deep Graph Infomax [18], etc. These models improve the performance and robustness of the model by introducing different graph convolution operations, attention mechanism designs, or graph structure modeling methods. Models based on graph neural networks have achieved significant results in fields such as social network analysis, image segmentation, bioinformatics, etc [8]. These methods can effectively capture node relationships and features in sequential data, suitable for handling complex sequential data structures. However, they may require significant training and inference times for large graph structures and complex relationships, requiring a high level of understanding and processing capability for graph structures.

Overall, these six categories of methods share the common goal of utilizing deep learning techniques to achieve time series anomaly detection. They employ different approaches and technical means to address various types of anomaly detection problems. Depending on factors such as data characteristics, task requirements, and model performance, one can choose a suitable method for conducting research and application in time series anomaly detection.

4. Application

4.1. Smart Grid

The power grid is one of the most crucial infrastructures in modern society, and its stable operation is vital for ensuring social livelihoods and economic development. With the rapid development of smart grids and renewable energy sources, the scale and complexity of power grid data are continuously increasing. Power grid data exhibits high complexity and heterogeneity, including real-time monitoring data from sensors, measuring devices, and surveillance systems, as well as multidimensional time series covering aspects such as power grid status, load conditions, energy production, and consumption. These data are often influenced by various factors such as natural phenomena, human interference, and technical faults. Therefore, time series anomaly detection has widespread applications in the field of power grids, including but not limited to abnormal load detection, voltage anomaly detection, fault diagnosis, and prediction. By promptly identifying and addressing abnormal situations, the reliability, safety, and economic efficiency of the power grid can be improved, ensuring stable and sustainable power supply. For example, Wei et al. introduce Grid Load Anomaly Detection (GLAD), a method for detecting load anomalies in microgrids using the enhanced ESD test [64]. GLAD shows promising results in simulating anomaly detection through statistical analyses. Expanding upon anomaly detection in power grids, Liang et al. present a fusion algorithm combining Isolation Forest and K-means to accurately detect abnormal Access IP in power grid dispatching platforms [55]. By addressing limitations of Isolation Forest, such as binary classification and threshold setting, the proposed approach demonstrates effectiveness and advancement in anomaly detection, as validated through evaluation metrics using data from the southern power grid dispatching platform. DYN-WATCH, a domain knowledge-based and topology-aware algorithm, effectively detects anomalies in power grid sensor data [39]. It accommodates regular topology adjustments, offering both accuracy and speed, with an average processing time of less than 1.7 ms per time tick per sensor on a laptop computer, scaling linearly with the size of the graph. In addition, leveraging a CNN within the Digital Twin framework, William et al. propose a method for detecting physical faults in power systems [13]. By analyzing high-fidelity data from benchmark power systems, the approach not only detects faults but also classifies the affected bus. In [48], Oprea et al. explore the detection of errors and fraud in smart metering data using machine learning algorithms, aiming to alert utility companies to suspicious consumption behavior. A Spectral Residual-Convolutional Neural Network (SR-CNN) approach is proposed, which achieves a 90% accuracy, 0.875 precision score, and 0.894 F1 score in identifying suspicious consumers. Zhang et al. propose a self-supervised framework for anomaly detection in smart grid time series data [83]. The framework captures dynamic correlations between different sensors, effectively detecting anomalies in smart grid data. In [24], authors present a deep generative model for anomaly detection in power grid data and applies it to transient stability assessment in the IEEE NewEngland-39 bus. Xiao et al. propose the Swin Transformer model for detecting abnormal power time-series data, enhancing prediction accuracy in power grid security [71]. Park et al. introduce GridCAL, a context-aware anomaly detection algorithm for real-time power system SCADA data [50]. By considering the impact of network topology and load/generation changes, GridCAL converts power flow measurements to context-agnostic values, enabling accurate anomaly detection across different grid contexts.

4.2. Network

In modern society, cybersecurity is paramount to safeguarding sensitive information and ensuring the uninterrupted operation of various digital systems. With the proliferation of networked devices and the increasing sophistication of cyber threats, the volume and complexity of cybersecurity data continue to grow exponentially. Cybersecurity data encompasses a wide range of heterogeneous sources, including network traffic logs, system event logs, user activity logs, and application logs, each capturing different aspects of system behavior and potential security incidents. These data are susceptible to diverse threats such as malware infections, unauthorized access attempts, distributed denial-of-service (DDoS) attacks, and data breaches, among others. As a result, time series anomaly detection plays a critical role in cybersecurity, offering a proactive approach to identify and mitigate emerging threats before they escalate into full-fledged cyberattacks. By analyzing patterns and deviations in time series data, anomaly detection algorithms can detect unusual behavior indicative of potential security breaches or malicious activities. Common applications of time series anomaly detection in cybersecurity include but are not limited to intrusion detection, network anomaly detection, malware detection, and fraud detection. By promptly identifying and responding to anomalies, organizations can enhance their cyber resilience, protect sensitive data, and maintain the integrity and availability of their digital assets. For instance, Goetz et al. propose an unsupervised anomaly detection approach for cyber-physical production systems using graph neural networks to model the system as a graph, incorporating correlations and interactions [23]. The introduced reconstruction-based graph neural network effectively detects anomalies in real industrial cyber-physical production systems. In [74], Yang et al. introduce ADT, an agent-based dynamic thresholding method using deep reinforcement learning (DQN), for time series anomaly detection in cyber-physical systems (CPS). Besides, Sun et al. propose MTS-DVGAN, an unsupervised dual variational generative adversarial model, for anomaly detection in multivariate time series data in CPS [57]. Yin et al. introduces MSCBL-ADN, a novel model combining multi-scale CNN and bidirectional Long-short Term Memory (bi-LSTM) arbitration dense network, for detecting Low-rate Distributed Denial of Service (LDDoS) attacks in cloud computing networks [77]. In [78], the SAnDet (SDN anomaly detector) architecture is proposed as an anomaly-based intrusion detection system, leveraging software-defined networking (SDN) capabilities. The system utilizes RNN and LSTM-based encoder-decoder (EncDecAD) methods to detect unknown attacks using flow features from OpenFlow switches. Reference [32] introduces a deep learning-based approach utilizing a GAN for cybersecurity threat detection in IoT-driven cyber-physical systems, with superior performance in detecting various types of attacks. Thiruloga et al. introduce the TENET framework, which employs temporal convolutional neural networks with an integrated attention mechanism for anomaly detection in vehicles, addressing cybersecurity threats in modern vehicle systems [58].

4.3. Financial

In the realm of finance, time series anomaly detection holds immense significance for ensuring the integrity and stability of financial markets and institutions. With the rapid evolution of financial technologies and the increasing complexity of trading systems, financial data has become voluminous and heterogeneous, comprising market prices, trading volumes, transaction records, and economic indicators, among others. These data streams are susceptible to various anomalies, including market manipulation, insider trading, fraudulent activities, and systemic risks, which can have far-reaching consequences on market stability and investor confidence. Time series anomaly detection techniques play a pivotal role in identifying and mitigating such anomalies by analyzing patterns and deviations in financial data. By leveraging statistical models, machine learning algorithms, and advanced data analytics techniques, financial institutions can detect irregularities in trading behavior, detect fraudulent transactions, and identify emerging market trends or anomalies that may pose risks to financial stability. Common applications of time series anomaly detection in finance include fraud detection in payment transactions, market surveillance for detecting abnormal trading activities, credit risk assessment, and algorithmic trading monitoring. By promptly detecting and responding to anomalies in financial time series data, organizations can enhance market transparency, investor protection, and financial resilience, thereby fostering trust and confidence in financial markets and institutions. For examole, Aftabi et al. introduce a novel approach for fraud detection in financial statements using GAN and ensemble models [2]. It achieves competitive performance with supervised models and outperforming unsupervised methods. In [33], Khodabandehlou et al. propose FiFrauD, an unsupervised and scalable approach for detecting fraudulent traders and suspicious behaviors in real-time financial transactions in e-markets. Besides, Chullamonthon et al. propose an ensemble model combining supervised LSTM network and unsupervised LSTM-based autoencoder deep learning techniques for detecting stock price manipulations [10]. In [31], Jiang et al. propose the UAAD-FDNet framework for credit card fraud detection, applying unsupervised attentional anomaly detection networks. It is designed to detect fraudulent transactions from large-scale transaction datasets and shows good performance. Habibpour et al. propose three uncertainty quantification (UQ) techniques—Monte Carlo dropout, ensemble, and ensemble Monte Carlo dropout—for credit card fraud detection using transaction data [25]. In [73], Yan et al. introduce the CEEMD-PCA-LSTM model for stock market anomalies detection, combining complementary ensemble empirical mode decomposition (CEEMD), principal component analysis (PCA), and LSTM networks.

5. Challenges and Future Works

Despite the wide-ranging applications of time series anomaly detection, significant challenges persist in its implementation and deployment.

5.1. Model Interpretability

Many time series anomaly detection models adopt complex deep learning structures such as RNNs, LSTMs, and CNNs. These models possess a large number of parameters and layers, making it difficult to intuitively understand the inner workings and decision-making processes of the models. Additionally, deep learning-based time series anomaly detection models are often perceived as black-box models, where their internal logic and decision-making processes are invisible or opaque to the user. This lack of transparency prevents users from understanding how the model processes data and detects anomalies, leading to a lack of trust in practical applications and making the model results difficult to interpret and validate. Future research can focus on the following methods to improve model interpretability:

- **Model Simplification:** Enhance model interpretability by designing simplified model structures or adding specialized interpretability components. Attention mechanisms allow models to focus on the most relevant parts of the data, making the decision-making process more understandable. For example, attention mechanisms can be used in time series anomaly detection to highlight anomalous data points or patterns. Interpretability layers are special layers used to transform the model's output into a more understandable form, such as probability distributions or interpretable rules. Rule engines allow domain expert knowledge to be directly integrated into the model to form rule-based interpretable models.
- **Construction of Interpretable Models:** Developing specialized interpretable time series anomaly detection models is an effective method to improve model interpretability. Interpretable models are typically based on simple algorithms or rules, such as rule-based methods, decision trees, or linear models. Compared to complex deep learning models, these models are easier to understand and interpret because their decision-making processes are more transparent. By constructing specialized interpretable models, users can more intuitively understand anomaly detection results and comprehend the workings and decision logic of the model.

5.2. Large-Scale and High-Dimensional Data

With the proliferation of the internet and the development of IoT, an increasing amount of data is being generated and collected, leading to exponential growth in data volume. Additionally, data in many fields not only have a large number of samples but also may have a large number of features, resulting in increased data dimensionality. For example, data collected in sensor networks may have thousands or even millions of features, making traditional anomaly detection algorithms ineffective. Future research can focus on utilizing the statistical properties of data and redundancy information to address the challenges of large-scale and high-dimensional data processing:

- **Distributed Computing:** Distributed computing processes data distributed across multiple computing nodes, significantly reducing computational complexity and processing time. By employing parallel computing and distributed algorithms, large-scale datasets can be divided into multiple small batches, each processed on different computing nodes, and then the results are merged. This allows for effective utilization of computational resources and handling datasets larger than the memory capacity of individual computing nodes.
- **Compression and Dimensionality Reduction:** Compression and dimensionality reduction techniques reduce data redundancy and feature dimensionality, greatly reducing storage requirements and computational overhead. For example, compression methods based on sampling can reduce the number of data samples through random sampling or importance sampling, thereby reducing storage requirements. Dimensionality reduction techniques such as Principal Component Analysis (PCA) or Singular Value Decomposition (SVD) project high-dimensional data into low-dimensional subspaces, reducing the number of features and computational complexity. Therefore, future developments will focus on developing efficient data processing techniques, including algorithms based on distributed and parallel computing, data compression and dimensionality reduction techniques, and GPU acceleration. Additionally, for high-dimensional data, attention should be focused on feature selection and feature extraction techniques to reduce data dimensionality and improve algorithm efficiency.

5.3. Model Lifelong Learning

Traditional anomaly detection models are often trained on static or representative training datasets, which may not fully cover various situations and changes in the real world. When the model encounters data distributions different from the training data, its performance may significantly deteriorate because the model cannot learn sufficiently generalized representations or patterns from the training data. Additionally, model parameters are often fixed and cannot be dynamically adjusted based on data changes. This means that once the model is trained, its behavior remains unchanged in subsequent applications, and it cannot adapt flexibly to changes in data distribution or concept drift. Future research can focus on developing anomaly detection models with continual learning capabilities to address changes in data distribution and concept drift:

- **Online Learning:** Utilizing online learning techniques, models can continuously learn from new data and adjust model parameters in a timely manner based on data changes. By updating the model in real-time, it can adapt to changes in data distribution, thereby maintaining model performance and accuracy.
- **Transfer Learning:** Leveraging transfer learning principles, previously learned knowledge and models can be transferred to new data domains, accelerating the learning process for new data. Through transfer learning, models can leverage existing knowledge and experience to quickly adapt to the characteristics and distribution of new data, improving model generalization and adaptability.
- **Incremental Learning:** Adopting incremental learning techniques allows models to continuously learn from new data and gradually accumulate new knowledge and experience. Through incremental learning, models can continuously update

their parameters or structures to adapt to changes in data distribution and concept drift, thereby maintaining model performance and robustness.

5.4. Data Imbalance and Class Skewness:

In many practical scenarios, the normal time series data often exhibits a higher frequency than anomalous data. For instance, in monitoring systems, the normal operational state typically dominates the majority of time, with anomalous events being sporadic occurrences. This inherent data distribution imbalance results in issues of data imbalance and class skewness. In real-world applications, normal data instances are usually far more abundant than anomalous ones, leading to problems of data imbalance and class skewness. Future research can concentrate on the following methodologies to develop anomaly detection algorithms tailored to imbalanced data:

- **Cost-sensitive Learning:** This methodology involves assigning varying weights or costs to samples from different classes during model training to mitigate the data imbalance problem. By adjusting the loss function, the model imposes greater penalties for misclassifying anomalous samples. The weights assigned to normal and anomalous samples in the loss function can be set differently, allowing the model to focus more on the accurate classification of anomalous samples during prediction. Additionally, distinct cost weights can be assigned to samples of each class during training, enabling the model to prioritize classes that are challenging to classify, thereby enhancing performance in imbalanced environments.
- **Meta-learning:** Meta-learning is an approach that enhances model performance and generalization by learning the process of learning itself. In imbalanced environments, meta-learning can improve performance by acquiring the ability to adapt to imbalanced data distributions. By incorporating meta-learning algorithms during model training, the model can dynamically adjust its learning strategy to accommodate the characteristics of imbalanced data. By learning how to design loss functions suitable for imbalanced data, the model can effectively handle anomalous samples and enhance performance in imbalanced environments.

5.5. Multi-modal Data Fusion

Time series data may originate from various data sources and sensors, such as text descriptions, images, sensor measurements, etc. Each data source possesses its unique data modality and feature representation, resulting in the generation of multi-modal data. With the extensive adoption of multi-modal data across various domains, future research can explore novel methods of multi-modal data fusion to fully exploit the correlation and complementarity among different data sources.

- **Large Language Models (LLMs):** LLMs, such as GPT, BERT, etc., acquire language representations by pre-training on extensive text corpora, which contain abundant semantic and syntactic information. In multi-modal fusion, the textual representation capability of LLMs can be harnessed to fuse text data with other modalities. LLMs can transform textual data into high-dimensional semantic vector representations, offering a unified semantic space that facilitates semantic correlation and fusion across different data modalities.

- **Knowledge Graphs:** Knowledge graphs are graphical structures employed to represent and organize knowledge, containing rich entity and relationship information. In multi-modal fusion, the structured representation capability of knowledge graphs can be leveraged to map data from different modalities onto the knowledge graph and exploit the entities and relationships therein for data correlation and fusion. Knowledge graphs provide semantically rich relationship information, aiding the model in better understanding and analyzing multi-modal data.

6. Conclusion

In conclusion, the detection of anomalies in IoT time series data is of paramount importance for ensuring the reliability, security, and efficiency of IoT systems across various domains. This review has provided a comprehensive overview of anomaly detection techniques in IoT, encompassing statistical methods, machine learning algorithms, and deep learning models for point anomalies, collective anomalies, and contextual anomalies. Despite the progress made in this field, several challenges such as data heterogeneity, scalability, interpretability, and real-time processing persist. Future research efforts should focus on addressing these challenges and exploring innovative approaches to enhance the effectiveness and efficiency of anomaly detection in IoT time series data. By overcoming these challenges and advancing the state-of-the-art techniques, we can unlock the full potential of IoT applications and facilitate the realization of a smarter and more connected world.

Funding

This work was supported by the National Natural Science Foundation of China (62072360, 62172438), the key research and development plan of Shaanxi province(2021ZDLGY02-09, 2023-GHZD-44, 2023-ZDLGY-54), the Natural Science Foundation of Guangdong Province of China(2022A1515010988), Key Project on Artificial Intelligence of Xi'an Science and Technology Plan(23ZDCYJSGG0021-2022, 23ZDCYYYCJ0008, 23ZDCYJSGG0002-2023), and the Proof-of-concept fund from Hangzhou Research Institute of Xidian University (GNYZ2023QC0201).

References

[1] Joseph M Ackerson, Rushit Dave, and Naeem Seliya. Applications of recurrent neural network for biometric authentication & anomaly detection. *Information*, 12(7):272, 2021.

[2] Seyyede Zahra Aftabi, Ali Ahmadi, and Saeed Farzi. Fraud detection in financial statements using data mining and gan models. *Expert Systems with Applications*, 227:120144, 2023.

[3] Elena-Simona Apostol, Ciprian-Octavian Truică, Florin Pop, and Christian Esposito. Change point enhanced anomaly detection for iot time series data. *Water*, 13(12):1633, 2021.

[4] Oliver Atkinson, Akanksha Bhardwaj, Christoph Englert, Vishal S Ngairangbam, and Michael Spannowsky. Anomaly detection with convolutional graph neural networks. *Journal of High Energy Physics*, 2021(8):1–19, 2021.

[5] Md Abul Bashar and Richi Nayak. Tanogan: Time series anomaly detection with generative adversarial networks. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1778–1785. IEEE, 2020.

[6] Alinne Beteto, Vidal Melo, Jessica Lin, Marwan Alsultan, Eduardo Mario Dias, Elizabeth Korte, DeAndre A Johnson, Negin Moghadasi, Thomas L Polmateer, and James H Lambert. Anomaly and cyber fraud detection in pipelines and supply chains for liquid fuels. *Environment Systems and Decisions*, 42(2):306–324, 2022.

[7] Chuchu Chen, Qiang Li, Li Chen, Yule Liang, and Hui Huang. An improved graphsage to detect power system anomaly based on time-neighbor feature. *Energy Reports*, 9:930–937, 2023.

[8] Zekai Chen, Dingshuo Chen, Xiao Zhang, Zixuan Yuan, and Xiuzhen Cheng. Learning graph structures with transformer for multivariate time-series anomaly detection in iot. *IEEE Internet of Things Journal*, 9(12):9179–9189, 2021.

[9] Sunav Choudhary, Gaurush Hiranandani, and Shiv Kumar Saini. Sparse decomposition for time series forecasting and anomaly detection. In *Proceedings of the 2018 SIAM International Conference on Data Mining*, pages 522–530. SIAM, 2018.

[10] Phakhawat Chullamonthon and Poj Tangamchit. Ensemble of supervised and unsupervised deep neural networks for stock price manipulation detection. *Expert Systems with Applications*, 220:119698, 2023.

[11] Andrew A Cook, Göksel Mısırlı, and Zhong Fan. Anomaly detection for iot time-series data: A survey. *IEEE Internet of Things Journal*, 7(7):6481–6494, 2019.

[12] Mingjian Cui, Jianhui Wang, and Meng Yue. Machine learning-based anomaly detection for load forecasting under cyberattacks. *IEEE Transactions on Smart Grid*, 10(5):5724–5734, 2019.

[13] William Danilczyk, Yan Lindsay Sun, and Haibo He. Smart grid anomaly detection using a deep learning digital twin. In *2020 52nd North American Power Symposium (NAPS)*, pages 1–6. IEEE, 2021.

[14] Toluwalope David Akande, Barjinder Kaur, Sajjad Dadkhah, and Ali A Ghorbani. Threshold based technique to detect anomalies using log files. In *Proceedings of the 2022 7th International Conference on Machine Learning Technologies*, pages 191–198, 2022.

[15] Ailin Deng and Bryan Hooi. Graph neural network-based anomaly detection in multivariate time series. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 4027–4035, 2021.

[16] Harsh S Dhiman, Dipankar Deb, SM Muyeen, and Innocent Kamwa. Wind turbine gearbox anomaly detection based on adaptive threshold and twin support vector machines. *IEEE Transactions on Energy Conversion*, 36(4):3462–3469, 2021.

[17] Koustav Dutta, Rasmita Lenka, Soumya Ranjan Nayak, Asimananda Khandual, and Akash Kumar Bhoi. Med-net: a novel approach to ecg anomaly detection using lstm auto-encoders. *International Journal of Computer Applications in Technology*, 65(4):343–357, 2021.

[18] Peng Gao, Haotian Zhang, Ming Wang, Weiyong Yang, Xinshen Wei, Zhuo Lv, and Zengzhou Ma. Deep temporal graph infomax for imbalanced insider threat detection. *Journal of Computer Information Systems*, pages 1–11, 2023.

[19] Astha Garg, Wenyu Zhang, Jules Samaran, Ramasamy Savitha, and Chuan-Sheng Foo. An evaluation of anomaly detection and diagnosis in multivariate time series.

IEEE Transactions on Neural Networks and Learning Systems, 33(6):2508–2517, 2021.

[20] Alexander Geiger, Dongyu Liu, Sarah Alnegheimish, Alfredo Cuesta-Infante, and Kalyan Veeramachaneni. Tadgan: Time series anomaly detection using generative adversarial networks. In *2020 ieee international conference on big data (big data)*, pages 33–43. IEEE, 2020.

[21] Maryam Ghanbari and Witold Kinsner. Detecting ddos attacks using an adaptive-wavelet convolutional neural network. In *2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 1–7. IEEE, 2021.

[22] Reza Ghezelbash, Abbas Maghsoudi, and Emmanuel John M Carranza. Optimization of geochemical anomaly detection using a novel genetic k-means clustering (gkmc) algorithm. *Computers & Geosciences*, 134:104335, 2020.

[23] Christian Goetz and Bernhard G Humm. Unsupervised correlation-and interaction-aware anomaly detection for cyber-physical production systems based on graph neural networks. *Procedia Computer Science*, 232:2057–2071, 2024.

[24] Dibyajyoti Guha, Rajdeep Chatterjee, and Biplab Sikdar. Anomaly detection using lstm-based variational autoencoder in unsupervised data in power grid. *IEEE Systems Journal*, 2023.

[25] Maryam Habibpour, Hassan Gharoun, Mohammadreza Mehdipour, AmirReza Tajally, Hamzeh Asgharnezhad, Afshar Shamsi, Abbas Khosravi, and Saeid Nahavandi. Uncertainty-aware credit card fraud detection using deep learning. *Engineering Applications of Artificial Intelligence*, 123:106248, 2023.

[26] Weijie Hao, Tao Yang, and Qiang Yang. Hybrid statistical-machine learning for real-time anomaly detection in industrial cyber–physical systems. *IEEE Transactions on Automation Science and Engineering*, 20(1):32–46, 2021.

[27] Ezz El-Din Hemdan and DH Manjaiah. Anomaly credit card fraud detection using deep learning. *Deep Learning in Data Analytics: Recent Techniques, Practices and Applications*, pages 207–217, 2022.

[28] João Henriques, Filipe Caldeira, Tiago Cruz, and Paulo Simões. Combining k-means and xgboost models for anomaly detection using log datasets. *Electronics*, 9(7):1164, 2020.

[29] Truong Thu Huong, Ta Phuong Bac, Dao Minh Long, Tran Duc Luong, Nguyen Minh Dan, Bui Doan Thang, Kim Phuc Tran, et al. Detecting cyberattacks using anomaly detection in industrial control systems: A federated learning approach. *Computers in Industry*, 132:103509, 2021.

[30] Abdul Rehman Javed, Muhammad Usman, Saif Ur Rehman, Mohib Ullah Khan, and Mohammad Sayad Haghighi. Anomaly detection in automated vehicles using multistage attention-based convolutional neural network. *IEEE Transactions on Intelligent Transportation Systems*, 22(7):4291–4300, 2020.

[31] Shanshan Jiang, Ruiting Dong, Jie Wang, and Min Xia. Credit card fraud detection based on unsupervised attentional anomaly detection network. *Systems*, 11(6):305, 2023.

[32] Irfan Ali Kandhro, Sultan M Alanazi, Fayyaz Ali, Asadullah Kehar, Kanwal Fatima, Mueen Uddin, and Shankar Karuppayah. Detection of real-time malicious intrusions and attacks in iot empowered cybersecurity infrastructures. *IEEE Access*, 11:9136–9148, 2023.

[33] Samira Khodabandehlou and Alireza Hashemi Golpayegani. Fifraud: Unsupervised financial fraud detection in dynamic graph streams. *ACM Transactions on Knowledge Discovery from Data*, 18(5):1–29, 2024.

[34] Fanhui Kong, Jianqiang Li, Bin Jiang, Huihui Wang, and Houbing Song. In-

tegrated generative model for industrial anomaly detection via bidirectional lstm and attention mechanism. *IEEE Transactions on Industrial Informatics*, 19(1):541–550, 2021.

[35] Md Tahmid Rahman Laskar, Jimmy Xiangji Huang, Vladan Smetana, Chris Stewart, Kees Pouw, Aijun An, Stephen Chan, and Lei Liu. Extending isolation forest for anomaly detection in big data via k-means. *ACM Transactions on Cyber-Physical Systems (TCPS)*, 5(4):1–26, 2021.

[36] Gen Li and Jason J Jung. Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges. *Information Fusion*, 91:93–102, 2023.

[37] Jinbo Li, Hesam Izakian, Witold Pedrycz, and Iqbal Jamal. Clustering-based anomaly detection in multivariate time series data. *Applied Soft Computing*, 100:106919, 2021.

[38] Longyuan Li, Junchi Yan, Haiyang Wang, and Yaohui Jin. Anomaly detection of time series with smoothness-inducing sequential variational auto-encoder. *IEEE transactions on neural networks and learning systems*, 32(3):1177–1191, 2020.

[39] Shimiao Li, Amritanshu Pandey, Bryan Hooi, Christos Faloutsos, and Larry Pileggi. Dynamic graph-based anomaly detection in the electrical grid. *IEEE Transactions on Power Systems*, 37(5):3408–3422, 2021.

[40] Haoqiang Liu, Hongbo Zhao, Jiayue Wang, Shuai Yuan, and Wenquan Feng. Lstm-gan-ae: A promising approach for fault diagnosis in machine health monitoring. *IEEE Transactions on Instrumentation and Measurement*, 71:1–13, 2021.

[41] Yi Liu, Sahil Garg, Jiangtian Nie, Yang Zhang, Zehui Xiong, Jiawen Kang, and M Shamim Hossain. Deep anomaly detection for time-series data in industrial iot: A communication-efficient on-device federated learning approach. *IEEE Internet of Things Journal*, 8(8):6348–6358, 2020.

[42] Li Longyuan, Yan Junchi, Wang Haiyang, and Jin Yaohui. Anomaly detection of time series with smoothness-inducing sequential variational auto-encoder. *IEEE transactions on neural networks and learning systems*, 32(3):1177–1191, 2020.

[43] Srinidhi Madabhushi and Rinku Dewri. A survey of anomaly detection methods for power grids. *International Journal of Information Security*, 22(6):1799–1832, 2023.

[44] Kursat Rasim Mestav, Xinyi Wang, and Lang Tong. A deep learning approach to anomaly sequence detection for high-resolution monitoring of power systems. *IEEE Transactions on Power Systems*, 38(1):4–13, 2022.

[45] Giulia Moschini, Régis Houssou, Jérôme Bovay, and Stephan Robert-Nicoud. Anomaly and fraud detection in credit card transactions using the arima model. *Engineering Proceedings*, 5(1):56, 2021.

[46] Viraaji Mothukuri, Prachi Khare, Reza M Parizi, Seyedamin Pouriyeh, Ali Dehghantanha, and Gautam Srivastava. Federated-learning-based anomaly detection for iot security attacks. *IEEE Internet of Things Journal*, 9(4):2545–2554, 2021.

[47] M Murugesan and S Thilagamani. Efficient anomaly detection in surveillance videos based on multi layer perception recurrent neural network. *Microprocessors and Microsystems*, 79:103303, 2020.

[48] Simona-Vasilica Oprea, Adela Bâra, Florina Camelia Puican, and Ioan Cosmin Radu. Anomaly detection with machine learning algorithms and big data in electricity consumption. *Sustainability*, 13(19):10963, 2021.

[49] Guansong Pang, Longbing Cao, and Charu Aggarwal. Deep learning for anomaly detection: Challenges, methods, and opportunities. In *Proceedings of the 14th*

ACM international conference on web search and data mining, pages 1127–1130, 2021.

[50] SangWoo Park and Amritanshu Pandey. Anomaly detection in power grids via context-agnostic learning. *arXiv preprint arXiv:2404.07898*, 2024.

[51] Sergio Iglesias Pérez, Santiago Moral-Rubio, and Regino Criado. A new approach to combine multiplex networks and time series attributes: Building intrusion detection systems (ids) in cybersecurity. *Chaos, Solitons & Fractals*, 150:111143, 2021.

[52] ANM Bazlur Rashid, Mohiuddin Ahmed, Leslie F Sikos, and Paul Haskell-Dowland. Anomaly detection in cybersecurity datasets via cooperative co-evolution-based feature selection. *ACM Transactions on Management Information Systems (TMIS)*, 13(3):1–39, 2022.

[53] Naoufal Rtayli and Nourddine Enneya. Selection features and support vector machine for credit card risk identification. *Procedia Manufacturing*, 46:941–948, 2020.

[54] Iqbal H Sarker. Cyberlearning: Effectiveness analysis of machine learning security modeling to detect cyber-anomalies and multi-attacks. *Internet of Things*, 14:100393, 2021.

[55] Liang Shouyu, Kun Zhang, Wenchong Fang, Zhifeng Zhou, Rong Hu, Wen Zhu, Yingchen Li, Yichang Wang, and Jian Hou. Anomaly detection of power grid dispatching platform based on isolation forest and k-means fusion algorithm. In *Journal of Physics: Conference Series*, volume 1601, page 022010. IOP Publishing, 2020.

[56] Sharmila Subudhi and Suvasini Panigrahi. Use of optimized fuzzy c-means clustering and supervised classifiers for automobile insurance fraud detection. *Journal of King Saud University-Computer and Information Sciences*, 32(5):568–575, 2020.

[57] Haili Sun, Yan Huang, Lansheng Han, Cai Fu, Hongle Liu, and Xiang Long. Mts-dvgan: Anomaly detection in cyber-physical systems using a dual variational generative adversarial network. *Computers & Security*, 139:103570, 2024.

[58] Sooryaa Vignesh Thiruloga, Vipin Kumar Kukkala, and Sudeep Pasricha. Tenet: Temporal cnn with attention for anomaly detection in automotive cyber-physical systems. In *2022 27th Asia and South Pacific design automation conference (ASP-DAC)*, pages 326–331. IEEE, 2022.

[59] Shreshth Tuli, Giuliano Casale, and Nicholas R Jennings. Tranad: Deep transformer networks for anomaly detection in multivariate time series data. *arXiv preprint arXiv:2201.07284*, 2022.

[60] Waseem Ullah, Amin Ullah, Ijaz Ul Haq, Khan Muhammad, Muhammad Sajjad, and Sung Wook Baik. Cnn features with bi-directional lstm for real-time anomaly detection in surveillance networks. *Multimedia tools and applications*, 80:16979–16995, 2021.

[61] Giuseppe Varone, Wadii Boulila, Michele Lo Giudice, Bilel Benjdira, Nadia Mammone, Cosimo Ieracitano, Kia Dashtipour, Sabrina Neri, Sara Gasparini, Francesco Carlo Morabito, et al. A machine learning approach involving functional connectivity features to classify rest-eeg psychogenic non-epileptic seizures from healthy controls. *Sensors*, 22(1):129, 2021.

[62] Kilian Vos, Zhongxiao Peng, Christopher Jenkins, Md Rifat Shahriar, Pietro Borghesani, and Wenyi Wang. Vibration-based anomaly detection using lstm/svm approaches. *Mechanical Systems and Signal Processing*, 169:108752, 2022.

[63] Guanglu Wei and Zhonghua Wang. Adoption and realization of deep learning in network traffic anomaly detection device design. *Soft Computing*, 25(2):1147–

1158, 2021.

[64] Qiyu Wei, Rui Ma, Yiqiu Wang, Mingyu Chen, Yanru Sun, Mingjie Liu, and Xiaoyong Lin. Glad: A method of microgrid anomaly detection based on esd in smart power grid. In *2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS)*, pages 103–107. IEEE, 2020.

[65] Xin Wei, Lulu Zhang, Hao-Qing Yang, Limin Zhang, and Yang-Ping Yao. Machine learning for pore-water pressure time-series prediction: Application of recurrent neural networks. *Geoscience Frontiers*, 12(1):453–467, 2021.

[66] Yuanyuan Wei, Julian Jang-Jaccard, Wen Xu, Fariza Sabrina, Seyit Camtepe, and Mikael Boulic. Lstm-autoencoder-based anomaly detection for indoor air quality time-series data. *IEEE Sensors Journal*, 23(4):3787–3800, 2023.

[67] Julia Wolleb, Florentin Bieder, Robin Sandkühler, and Philippe C Cattin. Diffusion models for medical anomaly detection. In *International Conference on Medical image computing and computer-assisted intervention*, pages 35–45. Springer, 2022.

[68] Yulei Wu, Hong-Ning Dai, and Haina Tang. Graph neural networks for anomaly detection in industrial internet of things. *IEEE Internet of Things Journal*, 9(12):9214–9231, 2021.

[69] Zheng Yi Wu and Yekun He. Time series data decomposition-based anomaly detection and evaluation framework for operational management of smart water grid. *Journal of Water Resources Planning and Management*, 147(9):04021059, 2021.

[70] Julian Wyatt, Adam Leach, Sebastian M Schmon, and Chris G Willcocks. Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 650–656, 2022.

[71] Xiongbo Xiao, Zhonglin Yang, and Xueping Gao. Anomaly detection of power time-series data based on multi-dimensional transformer network. *Comput.-Aided Des. Applic*, 15, 2023.

[72] Zhiwen Xiao and Jianbin Jiao. Explainable fraud detection for few labeled time series data. *Security and Communication Networks*, 2021:1–9, 2021.

[73] Binbin Yan, Memon Aasma, et al. A novel deep learning framework: Prediction and analysis of financial time series using ceemd and lstm. *Expert systems with applications*, 159:113609, 2020.

[74] Xue Yang, Enda Howley, and Michael Schukat. Adt: Time series anomaly detection for cyber-physical systems via deep reinforcement learning. *Computers & Security*, page 103825, 2024.

[75] Asad Yaseen. The role of machine learning in network anomaly detection for cybersecurity. *Sage Science Review of Applied Machine Learning*, 6(8):16–34, 2023.

[76] Chunyong Yin, Sun Zhang, Jin Wang, and Neal N Xiong. Anomaly detection based on convolutional recurrent autoencoder for iot time series. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52(1):112–122, 2020.

[77] Xiaochun Yin, Wei Fang, Zengguang Liu, and Deyong Liu. A novel multi-scale cnn and bi-lstm arbitration dense network model for low-rate ddos attack detection. *Scientific Reports*, 14(1):5111, 2024.

[78] Sultan Zavrak and Murat Iskefiyeli. Flow-based intrusion detection on software-defined networks: a multivariate time series anomaly detection approach. *Neural Computing and Applications*, 35(16):12175–12193, 2023.

[79] XJ Zeng, M Yang, and YF Bo. Gearbox oil temperature anomaly detection

for wind turbine based on sparse bayesian probability estimation. *International Journal of Electrical Power & Energy Systems*, 123:106233, 2020.

[80] Jiuqi Elise Zhang, Di Wu, and Benoit Boulet. Time series anomaly detection for smart grids: A survey. In *2021 IEEE Electrical Power and Energy Conference (EPEC)*, pages 125–130. IEEE, 2021.

[81] Yuxin Zhang, Yiqiang Chen, Jindong Wang, and Zhiwen Pan. Unsupervised deep anomaly detection for multi-sensor time-series signals. *IEEE Transactions on Knowledge and Data Engineering*, 35(2):2118–2132, 2021.

[82] Zhenwei Zhang, Ruiqi Wang, Ran Ding, and Yuantao Gu. Unravel anomalies: an end-to-end seasonal-trend decomposition approach for time series anomaly detection. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5415–5419. IEEE, 2024.

[83] Zhenyu Zhang, Lin Zhao, Dongyang Cai, Shuming Feng, Jiawei Miao, Yirun Guan, Haicheng Tao, and Jie Cao. Time series anomaly detection for smart grids via multiple self-supervised tasks learning. In *2022 IEEE International Conference on Knowledge Graph (ICKG)*, pages 392–397. IEEE, 2022.

[84] Hang Zhao, Yujing Wang, Juanyong Duan, Congrui Huang, Defu Cao, Yunhai Tong, Bixiong Xu, Jing Bai, Jie Tong, and Qi Zhang. Multivariate time-series anomaly detection via graph attention network. In *2020 IEEE International Conference on Data Mining (ICDM)*, pages 841–850. IEEE, 2020.

[85] Ming Zhou and Petr Musilek. Real-time anomaly detection in distribution grids using long short term memory network. In *2021 IEEE Electrical Power and Energy Conference (EPEC)*, pages 208–213. IEEE, 2021.