IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 23 September 2023, accepted 10 October 2023, date of publication 19 October 2023, date of current version 27 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3325896

TOPICAL REVIEW

Deep Learning Technologies for Time Series Anomaly Detection in Healthcare: A Review

XUE YANG[1,2], **XUEJUN QI**[3], **AND XIAOBO ZHOU**[4]
[1]Department of Pancreatic Surgery, Sichuan University, Chengdu 610041, China
[2]West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu 610041, China
[3]School of Architecture and Civil Engineering, Xihua University, Chengdu 610039, China
[4]Center for Computational Systems Medicine, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

Corresponding author: Xiaobo Zhou (Xiaobo.Zhou@uth.tmc.edu)

This work was supported in part by the Center of Excellence-International Collaboration Initiative Grant, West China Hospital, Sichuan University, under Grant 139170052; and in part by the 1·3·5 Project for Disciplines of Excellence–Clinical Research Incubation Project, West China Hospital, Sichuan University, under Grant 2019HXFH022.

ABSTRACT Medical time series data often exhibit intricate and dynamic patterns. With the rapid advancement of medical digitization, deep learning-based time series anomaly detection techniques have found extensive applications in the healthcare field, such as detecting irregular heart rhythms and monitoring patients' vital signs. To fully leverage digitized medical records to identify anomalies in healthcare and address key challenges in precise anomaly detection, this paper provides a comprehensive review of deep learning-based anomaly detection techniques applied to medical time series data. By reviewing and summarizing the relevant research, this paper explores the deep learning-based time series anomaly detection techniques within the medical and health domain, analyzing the strengths and limitations of different deep learning architectures and algorithms in tackling specific medical tasks. Lastly, we discuss the challenges faced by this field and outline future research directions. By reviewing and summarizing advanced deep learning methods for time series anomaly detection in medical applications in recent years, this study contributes to the advancement of healthcare analytics, aiming to enhance patient treatment outcomes.

INDEX TERMS Anomaly detection, artificial neural networks, deep learning, healthcare, time series.

I. INTRODUCTION

Anomaly detection [1] is the process of identifying observations or behaviors within a dataset that significantly deviate from most samples. Anomalies typically exhibit numerical deviations from the normal range or exhibit distinct differences in distribution or pattern compared to other samples. In anomaly detection, our objective is to recognize these outliers for further analysis, interpretation, or appropriate action.

The rapid digital transformation in the healthcare domain is reshaping medical practices, diagnostics, and treatments at an unprecedented pace. Time series data generated within medical practices has emerged as a pivotal resource for deepening our understanding of patient conditions, especially in the realm of anomaly detection. The significance of time series data in healthcare is undeniable. Medical data such as electrocardiograms, electroencephalograms, blood pressure, and body temperature not only offer comprehensive insights into patients' physiological states but also contain rich latent patterns and features. However, medical time series data often exhibit intricate structures, characterized by high individual variability and noise. This complexity poses limitations on traditional analytical methods when attempting to uncover concealed abnormal patterns. Deep learning models excel at automatically discovering complex relationships and patterns within time series data. In comparison to traditional machine learning techniques, deep learning can capture higher-level nonlinear feature representations, particularly when dealing with large-scale datasets. Additionally, deep learning exhibits

The associate editor coordinating the review of this manuscript and approving it for publication was Shuihua Wang.

superior robustness and generalization capabilities, adapting well to noise and unknown data, thus achieving more accurate and robust anomaly detection [2]. Consequently, deep learning technology presents new opportunities and prospects for time series anomaly detection in healthcare.

Despite significant advancements in deep learning techniques for time series anomaly detection, their application in healthcare still faces numerous technical and challenges. We need to consider many complex factors, such as the complexity of medical data, class imbalance, model interpretability, and so on. In the pursuit of high-quality medical anomaly detection, elevated demands are placed on algorithms. Precise anomaly detection necessitates highly sensitive models, efficient feature extraction, and accurate model training. As data volumes rapidly expand, anomaly detection algorithms must also possess the capability to handle high-dimensional data and conduct dynamic data analysis [2].

While recent review papers [3], [4], [5] have introduced the application of deep learning methods in time series anomaly detection, there is a lack of focus on deep time series anomaly detection specifically tailored for the healthcare domain. Medical time series data often comprises complex and dynamic patterns, presenting unique challenges when applying deep learning-based anomaly detection techniques to medical scenarios. Therefore, this paper aims to delve into the application, techniques, and challenges of deep learning-based time series anomaly detection in healthcare. We will comprehensively review existing research, analyze their strengths and limitations, address the challenges faced by current studies, and provide insights into future directions.

The main contributions of this paper are as follows:

1) We systematically summarize representative research on deep learning-based time series anomaly detection in healthcare.
2) We propose a taxonomy to categorize techniques and issues related to medical time series data anomaly detection, as illustrated in Fig. 1. Subsequently, we conduct a detailed analysis based on this taxonomy.
3) We discuss the challenges existing in this research and outline future research directions.

The remainder of this paper is organized as follows. In Sections II, the types of medical time series data are discussed respectively. In Section III, the types of anomalies in time series data are discussed. In Section IV, deep learning time series anomaly detection basic models are reviewed. In Section V, anomaly detection learning modes are listed. In Section VI, we summarize the medical datasets for time series anomaly detection. In Section VII, evaluation criteria are presented. In Section VIII, we discuss some special concerns for time series anomaly detection tasks in healthcare. In Section IX, research challenges and future directions in time series anomaly detection are presented. Finally, in Section X, conclusions are discussed.

Deep learning-based time series anomaly detection technologies in healthcare		
Types of Medical Data	**Types of Anomalies**	**Deep Learning Architecture**
• Physiological Signal • EMR • Medical Voice • Medical Video	• Point Anomaly • Contextual Anomaly • Collective Anomaly	• CNN/TCN • RNN/LSTM • TRANSFORMER • AE/VAE • GAN • GNN/GCN/GAT • HTM
Learning Modes	**Datasets**	**Evaluation Criteria**
• Supervised Learning • Unsupervised Learning Self-Supervised Learning Contrastive Learning • Semi-Supervised Learning • Weakly Supervised Learning	• MIMIC-III • PhysioNet • SEED • UCI MHEALTH • UCI ECG • UCI Apnea-ECG • PAMAP2 • PTB Diagnostic ECG Database	• Accuracy • Precision + Recall • F1 Score • AUC

FIGURE 1. Deep learning-based time series anomaly detection technologies in healthcare.

II. TYPES OF MEDICAL TIME SERIES DATA

The characteristics and types of medical data directly influence the methods, models, and strategies for anomaly detection. Therefore, understanding the types of medical data is crucial before conducting medical anomaly detection tasks.

A. MEDICAL PHYSIOLOGICAL SIGNAL

Medical physiological signals refer to the signals that record human physiological activities and are used to monitor and assess the health status and functions of the human body. Common physiological signals include Electrocardiogram (ECG), Electroencephalogram (EEG), Electromyogram (EMG), Galvanic Skin Response (GSR), and so on. By analyzing and processing physiological signal data, potential diseases can be identified and detected. Deep learning techniques have achieved significant advancements in fields such as abnormal ECG detection [6], [7], [8], [9], [10], [11], EEG anomaly detection [12], [13], [14], [15], [16], [17], [18], and diagnosis of sleep breathing disorders [19], [20], [21].

In time series anomaly detection tasks, medical physiological signals face challenges such as limited data volume, sample imbalance, and difficulties in labeling. Compared to general signal domains, anomaly detection in medical physiological signals is susceptible to noise and interference, such as motion artifacts and electromyographic interference. This complexity adds to the demand for models with robust denoising capabilities [22]. Furthermore, due to individual variations among patients, significant differences might exist in medical physiological signals across different populations and individuals. Building generalized anomaly detection models suitable for diverse populations can enhance model performance [23], [24].

B. ELECTRONIC MEDICAL RECORDS (EMRS)

EMRs are pivotal information resources in healthcare, encompassing a wealth of medical time series data, including vital signs, laboratory tests, medication records, and payment data [25]. These data reflect the evolution of patients' medical histories and health conditions. Leveraging deep learning models, EMRs can be harnessed to uncover latent anomalies, offer decision support and personalized healthcare services, facilitate data quality management, and enable medical fraud detection, among other applications [26].

Lack of precise anomaly labels often poses challenges in anomaly detection based on electronic medical records (EMRs). Since anomalies in EMRs are typically determined and defined by clinical experts, varying definitions and interpretations of anomalies might exist among different physicians and healthcare institutions, leading to subjectivity and discrepancies. Insufficient and inaccurate anomaly labels can influence the training and evaluation of anomaly detection models. Additionally, due to the uniqueness and sensitivity of EMR data involving patients' private and confidential information, access and utilization of such data are subject to strict legal, ethical, and privacy regulations, further complicating the acquisition of accurate anomaly labels.

To address these challenges, approaches such as semi-supervised learning [27], weakly supervised learning [28], and transfer learning [29] can be considered to leverage limited anomaly labeled data. Combining unsupervised learning and self-supervised learning techniques can be employed for enhanced anomaly detection in EMRs.

C. MEDICAL VOICE AND VIDEO

Medical voice data includes patients' voice recordings and doctors' voice notes, while medical video data includes records of surgical procedures and dynamic medical images such as Computed Tomography (CT) scans and Magnetic Resonance Imaging (MRI) images. By learning the features and patterns of normal voice data, abnormal vocal sounds like pronunciation irregularities and abnormal voice rates can be detected. This aids doctors in diagnosing and treating conditions like pathological voice disorders [30], [31], [32]. Similarly, by learning the temporal and spatial features of normal medical video data (e.g., surgical videos), abnormal surgical actions and unusual changes in lesions can be detected, enhancing the detection of abnormal operations and pathological changes [33], [34].

Medical voice and video data are often sourced from different devices, collection environments, and individuals, resulting in significant diversity within the data. This diversity can lead to changes and biases in data distribution, making it challenging for anomaly detection models to generalize to new datasets and scenarios.

Furthermore, the complexity of medical voice and video data is also a crucial factor impacting the effectiveness of deep learning models. Such data typically contains a multitude of variations and noise. Voice data may be influenced by background noise, variations in voice quality, and pronunciation differences. Video data may be affected by lighting conditions, camera angles, and motion blur. These factors contribute to the diversity and complexity of the data, which in turn increases the difficulty of anomaly detection.

III. TYPES OF ANOMALIES

Different types of anomalies may possess distinct temporal patterns and features, necessitating the targeted design and optimization of detection models. The data distribution and trend variations for different anomaly types might demand diverse deep learning architectures or feature extraction methods to achieve more accurate anomaly detection outcomes. The main categories of anomalies include point anomaly, contextual anomaly, and collective anomaly [1].

A. POINT ANOMALY

A point anomaly refers to an individual data point that stands out noticeably compared to other data points. It exists independently in the feature space from the rest of the data points and deviates significantly from normal data samples [1]. In the context of medical anomaly detection, point anomalies could signify individual patient data that differs from most patient data. For instance, if a patient's physiological parameter value (such as heart rate, blood pressure, or blood glucose level) is much higher or lower than the normal range, it might be considered a point anomaly. Fig. 2 illustrates an example of a point anomaly. In time series data, anomaly points may correspond to extreme values. In such cases, anomaly values closely resemble noise in their numerical values, leading to potential misclassification. Thus, distinguishing point anomalies from noise within time series is a critical challenge in time series anomaly detection.

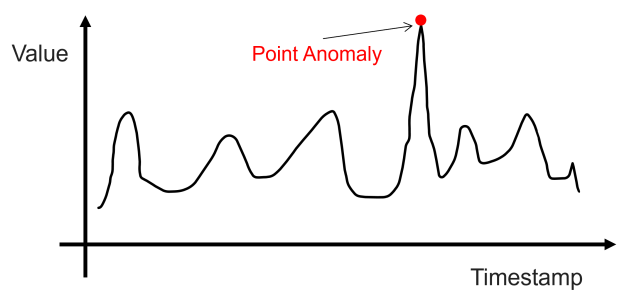

FIGURE 2. An example of point anomaly.

B. CONTEXTUAL ANOMALY

Contextual anomalies refer to data anomalies within specific contextual environments, where the value of a data point significantly deviates from the surrounding data under certain conditions [1]. These anomalies may be considered

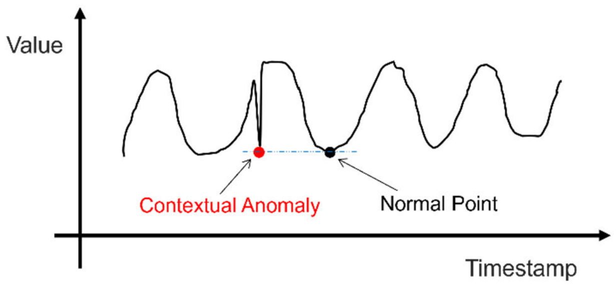

FIGURE 3. An example of contextual anomaly.

abnormal in certain specific contexts while being normal in other contexts. For instance, physiological data of a pregnant woman might undergo changes during pregnancy, such as elevated blood pressure or fluctuating blood glucose levels. These variations can be seen as contextual anomalies. Fig. 3 illustrates an example of a contextual anomaly. Medical time series data often involve intricate contextual environments. Factors like a patient's physiological state, disease condition, and treatment processes can all influence the definition and recognition of anomalies. This complexity presents more challenges and difficulties for detecting contextual anomalies in the medical and healthcare field.

C. COLLECTIVE ANOMALY

Collective anomalies refer to a subset or sequence within a whole dataset, data grouping, or time series that differs from other subsets or sequences [1]. This type of anomaly often indicates an abnormal pattern or event across the entirety of the data. A subset of data from patients with a specific disease that significantly differs from subsets of data from patients with other diseases can also be considered a collective anomaly. For instance, the distribution and trend of physiological parameters in a group of patients with a rare disease might noticeably differ from that of the general population, constituting a collective anomaly. Fig. 4 illustrates an example of a collective anomaly.

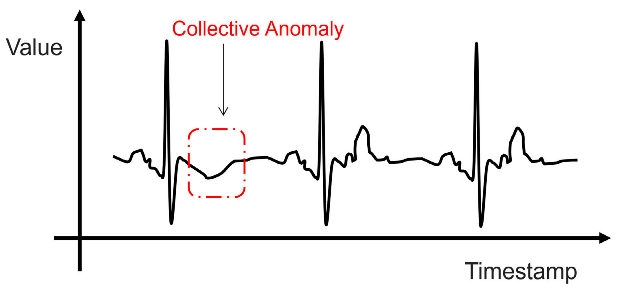

FIGURE 4. An example of ECG collective anomaly.

IV. DEEP LEARNING ARCHITECTURE USED IN TIME SERIES ANOMALY DETECTION

In medical anomaly detection, deep learning models are often built upon classical architectures and then adjusted and combined based on the task requirements and data characteristics. In this chapter, we will discuss some common foundational deep learning models for time series data anomaly detection. These foundational models can serve as backbones for medical anomaly detection tasks and can be further customized and extended according to specific application scenarios. Additionally, leveraging other techniques such as transfer learning and reinforcement learning can enhance the performance and robustness of anomaly detection.

A. CONVOLUTIONAL NEURAL NETWORKS (CNNS)

Although CNNs are primarily used in the field of image processing, in some cases, one-dimensional CNNs can be applied to time series data for anomaly detection. The roles of CNNs in time series anomaly detection include the following aspects:

1) Through convolutional operations, CNNs can be utilized to extract local features and patterns from time series data. The convolutional layers perform sliding window operations along the time dimension, which is like filtering in images, capturing local structural information in the data.
2) CNNs can gradually extract hierarchical features through multiple convolutional and pooling layers. For time series data, this can help capture patterns and variations at different time scales.
3) For certain long sequences of data, using traditional step-by-step sliding window methods might lead to extensive computations. CNNs can mitigate this by utilizing weight sharing and pooling operations, reducing data volume, and enhancing computational efficiency.
4) CNNs are particularly useful for dealing with complex relationships within data that possess both temporal and spatial features simultaneously.

In practice, CNN is often combined with other network structures such as Long Short-Term Memory (LSTM), GRU, etc., to better capture the complex relationships within time series data. For instance, BeatGAN [35] incorporates a one-dimensional CNN module into its adversarial learning reconstruction framework, enhancing the robustness of anomaly detection and achieving improved accuracy and rapid inference capability. The Temporal Convolutional Network (TCN) model [36] is a deep learning architecture designed for processing time series data. It utilizes convolutional layers to capture long-term dependencies and patterns within the data. Compared to traditional Recurrent Neural Networks (RNNs) and LSTMs, the TCN model efficiently captures features in time series data through parallelized convolution operations, without being limited by sequence length. This model has shown promising performance in time series data anomaly detection tasks. Kok et al. [37] applied TCN for automated sepsis detection, achieving impressive results.

B. RECURRENT NEURAL NETWORKS (RNNS)

RNN is a deep learning algorithm suitable for sequence data that possesses memory capabilities, enabling it to capture temporal dependencies. In medical anomaly detection, RNN can handle time series data such as ECG and EEG.

The LSTM model is a variant of the RNN structure. LSTM models have been extensively researched and have been found to excel in predicting time series data with long-term temporal dependencies [38], [39], [40], [41]. These studies suggest that LSTM offers better predictive accuracy compared to many other machine learning models and neural networks. However, the drawback of LSTM lies in its recurrent structure, making it challenging to effectively parallelize processing and resulting in lower learning efficiency. As a result, LSTM typically performs better when dealing with small-sample data.

C. TRANSFORMER

The Transformer was originally introduced in the field of natural language processing, but it has later found applications in the domain of time series data analysis [42], [43], [44]. The attention mechanism of the Transformer facilitates capturing the significance of various parts within a sequence, overcoming the limitations of recursive structures like LSTM. It is well-suited for modeling long-range dependencies in sequences. In terms of efficiency, the self-attention mechanism of the Transformer allows for efficient parallel computations, making it suitable for large-scale sequence data. Additionally, the Transformer exhibits strong scalability. When confronted with diverse tasks and data, its depth and number of attention heads can be adjusted to adapt to the specific requirements.

D. AUTOENCODER (AE)/ VARIATIONAL AUTOENCODER (VAE)

AE and VAE are neural network models used for unsupervised learning, and they are often combined with LSTM for time series data anomaly detection. AE employs encoding and decoding processes to learn a compact representation (encoding) of data and reconstruct the original data (decoding), aiding in extracting crucial features from time series data and capturing anomaly patterns. For instance, Liu et al. [45] devised an LSTM autoencoder model for arrhythmia detection, demonstrating improved performance.

Variational Autoencoder (VAE), a generative model, is not only used for encoding and decoding but also generates new data samples by sampling in the latent space. For anomaly detection, it helps identify samples that deviate from the normal data distribution. Chen et al. [46] proposed a semi-supervised VAE anomaly detection strategy based on LSTM. Experiments showed that their strategy outperformed existing baseline methods, detecting anomalies faster and more accurately while providing precise localization of the root causes of anomalies.

E. GENERATIVE ADVERSARIAL NETWORKS (GANS)

The GAN is an adversarial model composed of a generator and a discriminator. The generator aims to produce samples resembling normal data, while the discriminator endeavors to distinguish between the generated samples and real anomalous samples. In medical anomaly detection, GANs can be employed to generate synthetic data that closely resembles real patient data, thereby expanding the training dataset and enhancing the model's robustness and generalization capabilities. This approach also widens the training data to encompass various abnormal scenarios, aiding the model in capturing abnormal patterns more effectively. You et al. [47] utilized GANs to learn from behind-the-ear EEG of epilepsy patients, generating new samples for anomaly detection. The results demonstrated the effectiveness of this method for long-term epilepsy monitoring.

F. GRAPH NEURAL NETWORKS (GNNS)

Medical time series data often involve complex relationships among multiple variables, such as inter-patient interactions, medical history, clinical indicators, medications, and so on. GNN, or Graph Neural Network, is effective in capturing associations between multiple variables. It constructs a graph structure among patients and performs anomaly detection based on this graph. GNN incorporates temporal information into nodes by learning feature representations of nodes and edges. It leverages the connectivity within the graph for information propagation and aggregation, enhancing the discovery of abnormal patterns in medical data [48]. Graph Convolutional Network (GCN) [49] and Graph Attention Network (GAT) [50] are two common GNN models. GCN introduces convolution operations to facilitate information propagation and feature aggregation in graph data. GAT is a graph neural network model based on attention mechanisms, aiming to address the issue of varying importance among different nodes during information propagation.

Compared to other neural network models, GNN can establish connections between multiple data sources, aiding in comprehensive analysis across various domains. This capability makes GNN particularly effective in handling multi-dimensional time series data. In the medical field, multi-dimensional time series data can encompass various physiological indicators for patients, such as heart rate, body temperature, and blood pressure. GNN can help capture the intricate relationships among these indicators, leading to more precise anomaly detection, like disease outbreaks or changes in patient conditions. However, it's important to note that mapping time series data to a graph structure might introduce information loss or inaccuracies, potentially affecting the modeling of dynamic temporal characteristics. This is especially relevant for high-dimensional time series data where relationships between nodes could be more intricate. Adjusting more hyperparameters can also become complex, requiring expertise or extensive experimentation to determine the optimal configuration.

G. HIERARCHICAL TEMPORAL MEMORY (HTM)

The HTM is a computational model inspired by the neural circuitry of the human brain. This model aims to emulate the brain's processing of temporal data, enabling it to capture intricate temporal patterns and associations. The key to this model lies in its hierarchical structure, where each layer is capable of feature extraction and representation learning. The lower-level HTM layers can effectively capture temporal features of raw data, while deeper learning layers can abstract and combine these features to achieve higher-level pattern recognition and anomaly detection [51]. This architecture enhances the model's ability to identify rare anomalies in complex medical temporal data. For instance, Midani et al. [52] applied HTM to ECG signals, achieving exceptional anomaly detection in heart arrhythmias. Experimental results demonstrated that, compared to deep learning, HTM is more effective at detecting anomalies in ECG signals.

V. ANOMALY DETECTION LEARNING MODES

Based on the presence of label information in the training data and the availability of labels, we will introduce several learning modes for anomaly detection, including supervised learning, unsupervised learning, semi-supervised learning, and weakly supervised learning.

A. SUPERVISED LEARNING

For known anomaly types with labeled data, supervised learning trains samples from labeled training data to learn the differences between normal and abnormal patterns. In healthcare, supervised learning is extensively applied and finds utility in various time series anomaly detection tasks, such as detecting anomalies in electrocardiograms, electroencephalograms, medical images, diabetes management, medication dosage monitoring, and so on.

From an accuracy and reliability standpoint, supervised learning clearly outperforms unsupervised learning methods, thereby possessing higher accuracy and reliability in medical applications. For instance, in tasks involving the prediction of intensive care unit mortality using medical time series data, experiments have demonstrated highly robust and accurate outcomes [53]. However, in many instances, labeling data incurs substantial costs. Moreover, due to certain rare anomalous situations in healthcare, the number of normal samples often greatly exceeds that of abnormal samples, leading to class imbalance issues. Additionally, the dependency of supervised learning models on labeled data might impact their generalization ability, particularly when confronted with novel or rare anomaly patterns.

B. UNSUPERVISED LEARNING

In the field of anomaly detection, unsupervised algorithms typically judge the presence of anomalies by calculating the differences between reconstructed or generated samples and actual samples. Unsupervised learning does not require data labeling, but its drawback is a relatively weaker resistance to noise interference. VAEs and GANs are both typical examples of unsupervised learning.

Self-supervised learning is a specialized form of unsupervised learning where the model generates labels for itself to learn data feature representation. Algorithms use internally generated labels for training, eliminating the need for manual labeling and reducing labeling costs. Xu et al. [54] introduced a self-supervised learning approach and applied it to anomaly detection in brainwave signals. Experiments exhibited high robustness, addressing the challenge of difficult epilepsy data labeling.

Contrastive learning, also categorized as the unsupervised learning, typically employs pairs of samples divided into positive and negative samples [55]. The model learns to distinguish different categories or capture data features by comparing the differences between positive and negative samples. This learning process is conducted in an unsupervised context. Chen et al. [56] proposed a novel electrocardiogram signal contrastive learning scheme, CLECG, for mining effective information from unlabeled data. Experiments demonstrated superior performance compared to other self-supervised learning methods.

C. SEMI-SUPERVISED LEARNING

Semi-supervised learning involves training with a small amount of labeled data and a large amount of unlabeled data, utilizing the information from the unlabeled data to enhance the model's performance and generalization ability [57]. Thus, semi-supervised learning is well-suited for scenarios where only a small portion of the data carries anomaly labels. Label acquisition for medical data is often challenging, and medical anomalies tend to be diverse and complex. As a result, the application of semi-supervised learning in the medical domain is becoming increasingly common. For instance, Ying et al. [58] introduced a novel Federated Semi-Supervised Learning (FSSL) framework to predict abnormal signals in electrocardiogram records, achieving a 94.8% accuracy with only 50% labeled data. You et al. [59] designed a semi-supervised variational autoencoder network for epilepsy detection. Experiments demonstrated that the proposed algorithm improved seizure detection sensitivity to 90.4% and reduced the false alarm rate to 0.83 per hour.

As semi-supervised learning involves a substantial amount of unsupervised data during training, effectively managing the influence of noise becomes a significant challenge. Additionally, effectively integrating labeled and unlabeled data is a key issue. Designing appropriate model architectures and loss functions is essential to ensure that the model fully leverages the limited labeled information to enhance anomaly detection performance.

D. WEAKLY SUPERVISED LEARNING

Weakly supervised learning refers to training with data that has incomplete or inaccurate labels, utilizing partial label information for learning. It is suitable for datasets where only coarse anomaly labels are available, such as cases where the location or partial information about anomalies is known, but the specific anomaly labels are not entirely reliable [60]. Compared to supervised learning, weakly supervised learning can reduce the cost and complexity of the labeling process by utilizing incomplete anomaly labels for training. For instance, Liu et al. [61] proposed a weakly supervised model called FGSQA-Net for fine-grained cardiac monitoring. This model was applied to two real-world electrocardiogram databases and a synthetic dataset, effectively addressing the challenge of lacking fine-grained labels.

Relatively speaking, weakly supervised learning may not be as widely adopted in the medical domain as supervised or semi-supervised learning. However, it offers a flexible solution for anomaly detection in healthcare, particularly in situations where data annotation is difficult or expensive.

Since weakly supervised learning models need to handle both instance-level and sample-level label information, this complexity increases the model's intricacy. Hence, it's important to consider how to make the decision process of complex models more interpretable. Additionally, weakly supervised learning models might not perform well on new anomaly patterns, which places higher demands on the model's generalization capability.

VI. DATASETS

The application of time series anomaly detection techniques in healthcare relies heavily on high-quality datasets, and the selection of appropriate datasets is crucial for the performance and generalization ability of deep learning models. Many studies in the field of healthcare time series anomaly detection have utilized publicly available datasets.

Medical Information Mart for Intensive Care (MIMIC- III) [62] is a large-scale medical dataset widely used for clinical research. It includes clinical data, laboratory test data, medication records, and so on from ICU patients. PhysioNet [63] contains multiple medical time series datasets, covering electrocardiograms, respiratory signals, blood pressure data, and so on. Sleep-EDF Expanded Database (SEED) [64] is a dataset used for sleep monitoring research, containing EEG, EOG, and EMG signals from different sleep stages, making it valuable for sleep anomaly detection. Other datasets include UCI MHEALTH, UCI ECG, UCI Apnea-ECG, PAMAP2, and PTB Diagnostic ECG Database.

However, due to the challenges of annotating anomaly data, many researchers resort to simulating or synthesizing abnormal data to enhance model robustness and generalization. The synthesized or simulated anomaly data should ensure that the generated anomalies match the real data in statistical characteristics, temporal patterns, and distributions. Additionally, it's essential to evaluate the impact of the generated anomaly data on model performance, ensuring that the model performs well in detecting real anomalies. The use of Generative Adversarial Networks (GANs) to generate medical anomaly data is a typical approach [47]. During training, the generator attempts to generate anomaly data that is like the real data, while the discriminator distinguishes between the real and the generated data.

VII. EVALUATION CRITERIA

To accurately assess the performance of time series anomaly detection techniques, it's essential to establish appropriate evaluation criteria. Accuracy represents the proportion of correctly predicted samples to the total number of samples and is used as a measurement criterion in many studies of time series anomaly detection. While accuracy focuses on the overall predictive ability of the model, it can be misleading in cases of class imbalance. To measure model performance more effectively, some studies employ precision and recall as evaluation metrics, where their values determine the accuracy and comprehensiveness of anomaly detection. Precision indicates the proportion of true positives among the samples predicted as positives, evaluating the alignment between model predictions and actual anomaly cases. Recall represents the proportion of true positive predictions among all actual positive samples, assessing the model's ability to identify true anomalies. For medical data anomaly detection, high precision and recall are crucial to maximize the ability to detect potential health risks. In such cases, the F1 Score, which combines precision and recall, serves as a comprehensive evaluation metric.

Area Under the Curve (AUC) is another important criterion for measuring the effectiveness of time series anomaly detection techniques in healthcare. A higher AUC value, closer to 1, indicates better anomaly detection capability of the model. For imbalanced datasets and when measuring the performance of time series anomaly detection models at different thresholds, AUC provides a more robust assessment. It can be used in conjunction with precision, recall, F1 score, etc., to comprehensively evaluate the model's performance.

In the healthcare domain, the performance of anomaly detection models directly correlates with the health and life of patients. Therefore, high precision and high recall are particularly crucial in this field. Inaccurate anomaly predictions may lead to misdiagnosis, unnecessary treatments, or cause panic, emphasizing the criticality of high precision. Missing a true anomaly could have serious consequences as it might represent a potential health risk for the patient. Therefore, maintaining a high recall, ensuring the capture of as many anomalies as possible, is vital for enhancing patient safety. Overall, in the healthcare sector, a comprehensive consideration of high precision and high recall, especially utilizing composite evaluation metrics such as the F1 score, helps ensure that anomaly detection models can accurately and comprehensively identify potential health risks in real-world

TABLE 1. Summary of deep learning techniques for time series abnormality detection in healthcare.

Types of Data	Methods	Deep Learning Architecture	Learning Modes	Datasets	Evaluation Criteria
ECG	FedECG[58]	FSSL+ResNet-9	Semi-Supervised	PhysioNet(MIT-BIH)	Accuracy
	FGSQA-Net[61]	residual CNN+maxpooling+SE	Weakly supervised learning	PhysioNet/CinC Challenge databases, One synthetic dataset	AUC
	BeatGAN[35]	CNN/RNN+GAN	Unsupervised learning	PhysioNet(MIT-BIH)	Accuracy, Inference Speed
	LSTM Autoencoder[45]	LSTM+Autoencoder	Unsupervised learning	PhysioNet(MIT-BIH)	Accuracy, Sensitivity, Positive predictivity
	An automated detection system of MI[65]	CNN+LSTM+Ensemble technique(SMOTE + Tomek)	Supervised leamning	PTB, PhysioNet(MIT-BIH)	Accuracy, Confusion matrix, Loss history curve
	CLECG[56]	xresnet1d101	Contrastive Learning	ICBEB2018, PhysioNet 2017, PTB-XL	F1 Score, AUC
	HTM[52]	HTM	Supervised leamning	PhysioNet(MIT-BIH)	Precision, Recall, FPR, F1 Score
EEG	A novel semi-supervised network[59]	VAE	Semi-Supervised	Samsung Medical Center, Seoul, Korea	ROC, AUC
	GAN[47]	GAN	Unsupervised learning	Samsung Medical Center, Seoul, Korea	Accuracy, Sensitivity
EMR	TCN[37]	TCN	Supervised leamning	PhysioNet	Accuracy, AUC
	DETERRENT[50]	GNN+Attention	Supervised leamning	A health-related article dataset from 2014 to 2019	Accuracy, Precision, Recall, F1 Score
	RNN[66]	RNN	Supervised leamning(Transfer learning)	MIMIC-III	AUC
Endoscopic Videos	3D Sequential DenseConvLstm[33]	3DCNN+ConvLstm	Supervised leamning	Gastrointestinal Atlas	Recall, Precision, F-measure
Multimodal data	HAIM framework[67]	HAIM framework	Supervised leamning	PhysioNet(HAIM-MIMIC-MM)	AUC
Pathological Voice	SincNet[32]	Sinc Filters+CNN/DNN	Supervised leamning	Three different Far Eastern Memorial Hospital voice datasets	Accuracy, Sensitivity

applications. Table 1 categorically summarizes the reviewed deep learning techniques for time series anomaly detection applied in healthcare, as outlined in this paper.

VIII. SOME SPECIAL ISSUES IN HEALTHCARE

In the field of healthcare, the application of time series anomaly detection techniques comes with its own intricacies and challenges. Introducing deep learning into time series anomaly detection presents unique challenges and issues. When employing these techniques to enhance the accuracy and efficiency of clinical practices, it is essential to address specific challenges that arise when working with medical time series data.

The success of time series data anomaly detection relies significantly on high-quality datasets. However, the scarcity and imbalance of healthcare data often lead to a limited number of anomaly samples. To address this issue, synthetic techniques can be employed to generate additional anomaly samples, such as Synthetic Minority Over-sampling Technique (SMOTE) and GAN. For instance, Rai and Chatterjee [65] applied SMOTE to address dataset imbalance, resulting in improved accuracy for minority anomaly classes within ECG signals. Jakobsen et al. proposed a deep neural network that combines SMOTE with class balancing techniques to enhance accuracy in differentiating between depression patients and healthy controls using motion activity time series data. Methods like transfer learning [66] can also be utilized to enhance model performance, and improvements can be made by incorporating unsupervised learning, self-supervised learning, and weakly supervised learning methods.

The integration of multimodal data is also a significant focus of attention. Different modalities of information can complement each other, providing mutually reinforcing features. Fusion of these modalities can result in more representative feature representations. Moreover, multimodal data fusion can assist in identifying and filtering out false positives that might arise from a single data source, thereby reducing false alarm rates and enhancing the accuracy and generalizability of anomaly detection. For instance, Soenksen et al. [67] proposed and evaluated a Healthcare Artificial Intelligence Multimodal (HAIM) framework, demonstrating its consistency and robustness in generating models that outperform similar single-data-source methods across various healthcare demonstrations.

Many medical time series data encompass a multitude of features or dimensions, categorizing them as high-dimensional data. This includes physiological signals like electrocardiograms and electroencephalograms, medical images such as CT scans and MRI images, and diverse clinical indicators found in electronic medical records (EMRs). In high-dimensional spaces, the distances between data points become sparse, which might impact the performance of anomaly detection algorithms, making it more challenging to effectively identify anomalies. Additionally, high-dimensional data often contain substantial noise and redundant features, as anomalies may exhibit similarity to normal points in noisy and redundant features, placing higher demands on the model's denoising capabilities.

IX. RESEARCH CHALLENGES AND FUTURE DIRECTIONS

Significant progress has been made in the application of deep learning-based time series anomaly detection techniques in the field of healthcare. However, there are still a series of challenges that need to be addressed in future research endeavors.

A. HANDING REAL-TIME DATA

Medical time series data is often influenced by factors such as patients' physiological states and treatment processes, leading to dynamic changes and non-stationary patterns in the data. Moreover, the features observed in medical data can be affected by external factors such as seasons, geographical variations, and epidemics, causing the data characteristics to evolve. Consequently, anomaly detection models may struggle to capture the underlying patterns and trends in the data, necessitating the design of adaptable model architectures. Techniques like incremental learning and adaptive learning need to be introduced. Incremental learning [68], [69] can integrate new data into the model in real-time, enhancing adaptability while reducing computational and storage costs. On the other hand, adaptive learning [70] involves dynamically adjusting model parameters and features to accommodate the dynamic changes in the data.

B. PRIVACY AND SECURITY

Medical time series data contains sensitive personal health information, and using deep learning models for anomaly detection may risk exposing patient privacy, especially during processes like medical data sharing and centralized model training. Therefore, ensuring privacy protection and data security while performing joint model training is a significant challenge.

Federated learning presents a promising solution, allowing models to be trained on local devices and sharing only model parameters without revealing raw data. This approach helps safeguard privacy and reduces the risks associated with data transmission and centralized storage [71]. Additionally, privacy protection techniques like differential privacy and homomorphic encryption can be employed to encrypt data and further protect patient privacy.

C. INTERPRETABILITY

Understanding the rationale behind a model's decisions is crucial for healthcare professionals and decision-makers, making model interpretability particularly important in the medical field. However, the black-box nature of current deep learning models limits their practicality in clinical settings. Therefore, the pursuit of highly interpretable model designs is essential to meet clinical requirements.

Model interpretability techniques, such as Local Interpretable Model-Agnostic Explanations (LIME) [72] and SHapley Additive exPlanations (SHAP) [73], can be applied to explain the predictions of the model. These techniques generate interpretable explanations based on the relationship between model inputs and outputs. Combining deep learning models with traditional machine learning methods, such as decision trees and logistic regression, is also an effective strategy since these methods offer higher interpretability. In addition, introducing interpretable mechanisms like attention mechanisms and hierarchical structures should also be considered.

D. MULTIMODAL DATA

Effectively fusing diverse types of temporal data from different modalities presents a highly challenging task. Firstly, there may be data imbalance issues across different modalities, with some modalities having fewer samples. Handling data imbalance needs to be addressed before model training. Secondly, distinct data modalities could possess varying feature representations, making the alignment and extraction of useful information complex. Ensuring consistency in the fused data becomes a crucial issue. Furthermore, fused data from different modalities might encounter problems like missing data or amplified noise, which can impact the performance of anomaly detection. To address these challenges, potential solutions include adopting an adaptive fusion weight mechanism. This involves dynamically adjusting the fusion weights of different modalities based on data quality and the significance of feature information. Lastly, leveraging pre-trained multimodal models and employing transfer learning for anomaly detection tasks could also be considered.

E. HIGH DIMENSIONALITY

High-dimensional features may lead to the curse of dimensionality, making model training complex and time-consuming. This also complicates anomaly detection. Addressing the challenges posed by high-dimensional features remains a concern. To mitigate the impact of high-dimensional data, several dimensionality reduction techniques can be employed. These include selecting meaningful features, using methods like PCA and Autoencoders to map high-dimensional data to a lower-dimensional feature space, and incorporating attention mechanisms and adaptive learning to focus the model on key features and enhance its understanding of high-dimensional data.

F. DEFINITION OF ANOMALIES

The complexity of medical data makes it challenging to precisely define anomalies. Variability in different cases, pathologies, and patient needs can result in diverse definitions of anomalies, making it difficult for models to uniformly adapt. To address these situations, besides data augmentation and transfer learning, incorporating domain expertise is essential. By combining data-driven and knowledge-driven approaches, dynamically adjusting anomaly definitions can enhance model adaptability.

Techniques such as self-supervised learning and generative adversarial networks can be leveraged to achieve more flexible, dynamic, and adaptive anomaly definitions. This enhances the model's ability to recognize and generalize diverse anomalies.

Moreover, existing anomaly classification methods are based on supervised learning models. However, specific datasets often only label anomalies for certain diseases or symptoms, leaving out other disease-related anomalies. Therefore, a direction for future research and exploration is how to make anomaly detection tasks encompass all anomalies.

X. CONCLUSION

In this study, we review and summarize deep learning-based techniques for time series anomaly detection. We propose a classification framework to describe techniques related to anomaly detection in medical time series data. These techniques encompass the types of medical time series data, anomaly categories within time series data, fundamental deep learning models for time series anomaly detection, deep learning paradigms, medical time series datasets, and evaluation criteria. Through comprehensive analysis of relevant literature and research outcomes, this approach assists researchers in obtaining detailed insights into the latest advancements in medical time series anomaly detection technologies. Subsequently, the paper discusses specific concerns in the application of time series anomaly detection techniques within the medical and healthcare field. Finally, challenges and future directions are presented.

Despite the significant progress achieved through the application of deep learning-based time series anomaly detection techniques in the medical and healthcare sector, there remains ample room for development in this field. Continuous research and exploration hold the promise of achieving more precise, interpretable, powerful data mining-capable, real-time, and secure anomaly detection solutions.

REFERENCES

[1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–58, Jul. 2009.

[2] G. Pang, C. Shen, L. Cao, and A. Van Den Hengel, "Deep learning for anomaly detection: A review," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–38, 2021.

[3] K. Choi, J. Yi, C. Park, and S. Yoon, "Deep learning for anomaly detection in time-series data: Review, analysis, and guidelines," *IEEE Access*, vol. 9, pp. 120043–120065, 2021.

[4] Z. Z. Darban, G. I. Webb, S. Pan, C. C. Aggarwal, and M. Salehi, "Deep learning for time series anomaly detection: A survey," 2022, *arXiv:2211.05244*.

[5] G. Li and J. J. Jung, "Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges," *Inf. Fusion*, vol. 91, pp. 93–102, Mar. 2023.

[6] W. Qi and H. Su, "A cybertwin based multimodal network for ECG patterns monitoring using deep learning," *IEEE Trans. Ind. Informat.*, vol. 18, no. 10, pp. 6663–6670, Oct. 2022.

[7] G. Sivapalan, K. K. Nundy, S. Dev, B. Cardiff, and D. John, "ANNet: A lightweight neural network for ECG anomaly detection in IoT edge sensors," *IEEE Trans. Biomed. Circuits Syst.*, vol. 16, no. 1, pp. 24–35, Feb. 2022.

[8] E. B. Panganiban, A. C. Paglinawan, W. Y. Chung, and G. L. S. Paa, "ECG diagnostic support system (EDSS): A deep learning neural network based classification system for detecting ECG abnormal rhythms from a low-powered wearable biosensors," *Sens. Bio-Sens. Res.*, vol. 31, Feb. 2021, Art. no. 100398.

[9] K. K. Patro, A. J. Prakash, S. Samantray, J. Pławiak, R. Tadeusiewicz, and P. Pławiak, "A hybrid approach of a deep learning technique for real–time ECG beat detection," *Int. J. Appl. Math. Comput. Sci.*, vol. 32, no. 3, pp. 455–465, 2022.

[10] H. Zhu, C. Cheng, H. Yin, X. Li, P. Zuo, J. Ding, F. Lin, J. Wang, B. Zhou, Y. Li, S. Hu, Y. Xiong, B. Wang, G. Wan, X. Yang, and Y. Yuan, "Automatic multilabel electrocardiogram diagnosis of heart rhythm or conduction abnormalities with deep learning: A cohort study," *Lancet Digit. Health*, vol. 2, no. 7, pp. e348–e357, Jul. 2020.

[11] O. S. Lih, V. Jahmunah, T. R. San, E. J. Ciaccio, T. Yamakawa, M. Tanabe, M. Kobayashi, O. Faust, and U. R. Acharya, "Comprehensive electrocardiographic diagnosis based on deep learning," *Artif. Intell. Med.*, vol. 103, Mar. 2020, Art. no. 101789.

[12] S. Tahura, S. H. Samiul, M. S. Kaiser, and M. Mahmud, "Anomaly detection in electroencephalography signal using deep learning model," in *Proc. Int. Conf. Trends Comput. Cognit. Eng. (TCCE)*. Springer, 2021, pp. 205–217, doi: 10.1007/978-981-33-4673-4_18.

[13] A. M. Alvi, S. Siuly, and H. Wang, "Neurological abnormality detection from electroencephalography data: A review," *Artif. Intell. Rev.*, vol. 55, no. 3, pp. 2275–2312, Mar. 2022.

[14] D. Ahmedt-Aristizabal, T. Fernando, S. Denman, J. E. Robinson, S. Sridharan, P. J. Johnston, K. R. Laurens, and C. Fookes, "Identification of children at risk of schizophrenia via deep learning and EEG responses," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 1, pp. 69–76, Jan. 2021.

[15] M. Alhussein, G. Muhammad, and M. S. Hossain, "EEG pathology detection based on deep learning," *IEEE Access*, vol. 7, pp. 27781–27788, 2019.

[16] D. O. Nahmias and K. L. Kontson, "Easy perturbation EEG algorithm for spectral importance (easyPEASI): A simple method to identify important spectral features of EEG in deep learning models," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2020, pp. 2398–2406.

[17] S. L. Oh, Y. Hagiwara, U. Raghavendra, R. Yuvaraj, N. Arunkumar, M. Murugappan, and U. R. Acharya, "A deep learning approach for Parkinson's disease diagnosis from EEG signals," *Neural Comput. Appl.*, vol. 32, pp. 10927–10933, Aug. 2020.

[18] C. Chu, Z. Zhang, J. Wang, S. Liu, F. Wang, Y. Sun, X. Han, Z. Li, X. Zhu, and C. Liu, "Deep learning reveals personalized spatial spectral abnormalities of high delta and low alpha bands in EEG of patients with early Parkinson's disease," *J. Neural Eng.*, vol. 18, no. 6, Dec. 2021, Art. no. 066036.

[19] Y. Wang, M. Hu, Y. Zhou, Q. Li, N. Yao, G. Zhai, X.-P. Zhang, and X. Yang, "Unobtrusive and automatic classification of multiple people's abnormal respiratory patterns in real time using deep neural network and depth camera," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8559–8571, Sep. 2020.

[20] M. Murugesan, V. Abirami, M. Iswarya, and N. Kaviya, "Sleeping abnormalities detection using deep learning techniques," in *Proc. 2nd Int. Conf. Electron. Renew. Syst. (ICEARS)*, Mar. 2023, pp. 1283–1289.

[21] J. M. Perero-Codosero, F. Espinoza-Cuadros, J. Antón-Martín, M. A. Barbero-Álvarez, and L. A. Hernández-Gómez, "Modeling obstructive sleep apnea voices using deep neural network embeddings and domain-adversarial training," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 2, pp. 240–250, Feb. 2020.

[22] C. Yang, M. Hu, G. Zhai, and X.-P. Zhang, "Graph-based denoising for respiration and heart rate estimation during sleep in thermal video," *IEEE Internet Things J.*, vol. 9, no. 17, pp. 15697–15713, Sep. 2022.

[23] P. A. Noseworthy, Z. I. Attia, L. C. Brewer, S. N. Hayes, X. Yao, S. Kapa, P. A. Friedman, and F. Lopez-Jimenez, "Assessing and mitigating bias in medical artificial intelligence: The effects of race and ethnicity on a deep learning model for ECG analysis," *Circulat., Arrhythmia Electrophysiol.*, vol. 13, no. 3, Mar. 2020, Art. no. e007988.

[24] K. C. Siontis, P. A. Noseworthy, Z. I. Attia, and P. A. Friedman, "Artificial intelligence-enhanced electrocardiography in cardiovascular disease management," *Nature Rev. Cardiol.*, vol. 18, no. 7, pp. 465–478, Jul. 2021.

[25] J. Frankovich, C. A. Longhurst, and S. M. Sutherland, "Evidence-based medicine in the EMR era," *New England J. Med.*, vol. 365, no. 19, pp. 1758–1759, Nov. 2011.

[26] C. Xiao, E. Choi, and J. Sun, "Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review," *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 10, pp. 1419–1428, Oct. 2018.

[27] Y. Zhang et al., "High-throughput phenotyping with electronic medical record data using a common semi-supervised approach (PheCAP)," *Nature Protocols*, vol. 14, no. 12, pp. 3426–3444, Dec. 2019.

[28] M. Reuver, "Finding the smoke signal: Smoking status classification with a weakly supervised paradigm in sparsely labelled Dutch free text in electronic medical records," 2020. [Online]. Available: https://theses.ubn.ru.nl/handle/123456789/10278.

[29] A. Rios and R. Kavuluru, "Neural transfer learning for assigning diagnosis codes to EMRs," *Artif. Intell. Med.*, vol. 96, pp. 116–122, May 2019.

[30] S.-H. Fang, Y. Tsao, M.-J. Hsiao, J.-Y. Chen, Y.-H. Lai, F.-C. Lin, and C.-T. Wang, "Detection of pathological voice using cepstrum vectors: A deep learning approach," *J. Voice*, vol. 33, no. 5, pp. 634–641, Sep. 2019.

[31] J.-Y. Lee, "Experimental evaluation of deep learning methods for an intelligent pathological voice detection system using the Saarbruecken voice database," *Appl. Sci.*, vol. 11, no. 15, p. 7149, Aug. 2021.

[32] C.-H. Hung, S.-S. Wang, C.-T. Wang, and S.-H. Fang, "Using SincNet for learning pathological voice disorders," *Sensors*, vol. 22, no. 17, p. 6634, Sep. 2022.

[33] N. Ghatwary, M. Zolgharni, F. Janan, and X. Ye, "Learning spatiotemporal features for esophageal abnormality detection from endoscopic videos," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 1, pp. 131–142, Jan. 2021.

[34] H. Ali, M. Sharif, M. Yasmin, M. H. Rehmani, and F. Riaz, "A survey of feature extraction and fusion of deep learning for detection of abnormalities in video endoscopy of gastrointestinal-tract," *Artif. Intell. Rev.*, vol. 53, no. 4, pp. 2635–2707, Apr. 2020.

[35] S. Liu, B. Zhou, Q. Ding, B. Hooi, Z. Zhang, H. Shen, and X. Cheng, "Time series anomaly detection with adversarial reconstruction networks," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 4293–4306, Apr. 2023.

[36] L. Lin, B. Xu, W. Wu, T. W. Richardson, and E. A. Bernal, "Medical time series classification with hierarchical attention-based temporal convolutional networks: A case study of myotonic dystrophy diagnosis," in *Proc. CVPR Workshops*, 2019, pp. 83–86.

[37] C. Kok, V. Jahmunah, S. L. Oh, X. Zhou, R. Gururajan, X. Tao, K. H. Cheong, R. Gururajan, F. Molinari, and U. R. Acharya, "Automated prediction of sepsis using temporal convolutional network," *Comput. Biol. Med.*, vol. 127, Dec. 2020, Art. no. 103957.

[38] Y. Wang, X. Du, Z. Lu, Q. Duan, and J. Wu, "Improved LSTM-based time-series anomaly detection in rail transit operation environments," *IEEE Trans. Ind. Informat.*, vol. 18, no. 12, pp. 9027–9036, Dec. 2022.

[39] Y. Wei, J. Jang-Jaccard, W. Xu, F. Sabrina, S. Camtepe, and M. Boulic, "LSTM-autoencoder-based anomaly detection for indoor air quality time-series data," *IEEE Sensors J.*, vol. 23, no. 4, pp. 3787–3800, Feb. 2023.

[40] H. D. Nguyen, K. P. Tran, S. Thomassey, and M. Hamad, "Forecasting and anomaly detection approaches using LSTM and LSTM autoencoder techniques with the applications in supply chain management," *Int. J. Inf. Manage.*, vol. 57, Apr. 2021, Art. no. 102282.

[41] F. Kong, J. Li, B. Jiang, H. Wang, and H. Song, "Integrated generative model for industrial anomaly detection via bidirectional LSTM and attention mechanism," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 541–550, Jan. 2023.

[42] Z. Chen, D. Chen, X. Zhang, Z. Yuan, and X. Cheng, "Learning graph structures with transformer for multivariate time-series anomaly detection in IoT," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 9179–9189, Jun. 2022.

[43] J. Kim, H. Kang, and P. Kang, "Time-series anomaly detection with stacked transformer representations and 1D convolutional network," *Eng. Appl. Artif. Intell.*, vol. 120, Apr. 2023, Art. no. 105964.

[44] X. Wang, D. Pi, X. Zhang, H. Liu, and C. Guo, "Variational transformer-based anomaly detection approach for multivariate time series," *Measurement*, vol. 191, Mar. 2022, Art. no. 110791.

[45] P. Liu, X. Sun, Y. Han, Z. He, W. Zhang, and C. Wu, "Arrhythmia classification of LSTM autoencoder based on time series anomaly detection," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103228.

[46] N. Chen, H. Tu, X. Duan, L. Hu, and C. Guo, "Semisupervised anomaly detection of multivariate time series based on a variational autoencoder," *Appl. Intell.*, vol. 53, no. 5, pp. 6074–6098, 2023.

[47] S. You, B. H. Cho, S. Yook, J. Y. Kim, Y.-M. Shon, D.-W. Seo, and I. Y. Kim, "Unsupervised automatic seizure detection for focal-onset seizures recorded with behind-the-ear EEG using an anomaly-detecting generative adversarial network," *Comput. Methods Programs Biomed.*, vol. 193, Sep. 2020, Art. no. 105472.

[48] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Jan. 2009.

[49] L. Yao, C. Mao, and Y. Luo, "Graph convolutional networks for text classification," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 7370–7377.

[50] L. Cui, H. Seo, M. Tabar, F. Ma, S. Wang, and D. Lee, "DETERRENT: Knowledge guided graph attention network for detecting healthcare misinformation," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2020, pp. 492–502.

[51] J. Wu, W. Zeng, and F. Yan, "Hierarchical temporal memory method for time-series-based anomaly detection," *Neurocomputing*, vol. 273, pp. 535–546, Jan. 2018.

[52] W. Midani, Z. Fki, and M. BenAyed, "Online anomaly detection in ECG signal using hierarchical temporal memory," in *Proc. 5th Int. Conf. Adv. Biomed. Eng. (ICABME)*, Oct. 2019, pp. 1–4.

[53] V. Cerqueira, L. Torgo, and C. Soares, "Layered learning for early anomaly detection: Predicting critical health episodes," in *Discovery Science*. Split, Croatia: Springer, Oct. 2019, pp. 445–459.

[54] J. Xu, Y. Zheng, Y. Mao, R. Wang, and W.-S. Zheng, "Anomaly detection on electroencephalography with self-supervised learning," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2020, pp. 363–368.

[55] Y. Tian, C. Sun, B. Poole, D. Krishnan, C. Schmid, and P. Isola, "What makes for good views for contrastive learning?" in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 6827–6839.

[56] H. Chen, G. Wang, G. Zhang, P. Zhang, and H. Yang, "CLECG: A novel contrastive learning framework for electrocardiogram arrhythmia classification," *IEEE Signal Process. Lett.*, vol. 28, pp. 1993–1997, 2021.

[57] J. E. Van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Mach. Learn.*, vol. 109, no. 2, pp. 373–440, 2020.

[58] Z. Ying, G. Zhang, Z. Pan, C. Chu, and X. Liu, "FedECG: A federated semi-supervised learning framework for electrocardiogram abnormalities prediction," *J. King Saud Univ., Comput. Inf. Sci.*, vol. 35, no. 6, Jun. 2023, Art. no. 101568.

[59] S. You, B. Hwan Cho, Y.-M. Shon, D.-W. Seo, and I. Y. Kim, "Semi-supervised automatic seizure detection using personalized anomaly detecting variational autoencoder with behind-the-ear EEG," *Comput. Methods Programs Biomed.*, vol. 213, Jan. 2022, Art. no. 106542.

[60] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 44–53, Jan. 2018.

[61] H. Liu, T. Gao, Z. Liu, and M. Shu, "FGSQA-Net: A weakly supervised approach to fine-grained electrocardiogram signal quality assessment," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 8, pp. 3844–3855, Aug. 2023.

[62] A. E. W. Johnson, T. J. Pollard, L. Shen, L.-W.-H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, no. 1, pp. 1–9, May 2016.

[63] G. B. Moody, R. G. Mark, and A. L. Goldberger, "PhysioNet: A web-based resource for the study of physiologic signals," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 70–75, May/Jun. 2001.

[64] B. Kemp, A. Zwinderman, B. Tuk, H. Kamphuisen, and J. Oberyé, "Sleep-EDF database expanded," PhysioNet, 2018. [Online]. Available: https://doi.org/10.13026/C2X676

[65] H. M. Rai and K. Chatterjee, "Hybrid CNN-LSTM deep learning model and ensemble technique for automatic detection of myocardial infarction using big ECG data," *Appl. Intell.*, vol. 52, no. 5, pp. 5366–5384, Mar. 2022.

[66] P. Gupta, P. Malhotra, J. Narwariya, L. Vig, and G. Shroff, "Transfer learning for clinical time series analysis using deep neural networks," *J. Healthcare Informat. Res.*, vol. 4, no. 2, pp. 112–137, Jun. 2020.

[67] L. R. Soenksen, Y. Ma, C. Zeng, L. Boussioux, K. Villalobos Carballo, L. Na, H. M. Wiberg, M. L. Li, I. Fuentes, and D. Bertsimas, "Integrated multimodal artificial intelligence framework for healthcare applications," *NPJ Digit. Med.*, vol. 5, no. 1, p. 149, Sep. 2022.

[68] A. Gepperth and B. Hammer, "Incremental learning algorithms and applications," in *Proc. Eur. Symp. Artif. Neural Netw. (ESANN)*, 2016, pp. 1–13.

[69] L. Melgar-García, D. Gutiérrez-Avilés, C. Rubio-Escudero, and A. Troncoso, "A novel distributed forecasting method based on information fusion and incremental learning for streaming time series," *Inf. Fusion*, vol. 95, pp. 163–173, Jul. 2023.

[70] X. Yan, Y. Xu, X. Xing, B. Cui, Z. Guo, and T. Guo, "Trustworthy network anomaly detection based on an adaptive learning rate and momentum in IIoT," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 6182–6192, Sep. 2020.
[71] V. Mothukuri, P. Khare, R. M. Parizi, S. Pouriyeh, A. Dehghantanha, and G. Srivastava, "Federated-learning-based anomaly detection for IoT security attacks," *IEEE Internet Things J.*, vol. 9, no. 4, pp. 2545–2554, Feb. 2022.
[72] I. Neves, D. Folgado, S. Santos, M. Barandas, A. Campagner, L. Ronzio, F. Cabitza, and H. Gamboa, "Interpretable heartbeat classification using local model-agnostic explanations on ECGs," *Comput. Biol. Med.*, vol. 133, Jun. 2021, Art. no. 104393.
[73] L. Antwarg, R. M. Miller, B. Shapira, and L. Rokach, "Explaining anomalies detected by autoencoders using Shapley additive explanations," *Expert Syst. Appl.*, vol. 186, Dec. 2021, Art. no. 115736.

XUE YANG received the Ph.D. degree in management system engineering from Sichuan University, Chengdu, China, in 2020. She is currently a Postdoctoral Researcher with the West China Biomedical Big Data Center, West China Hospital, Sichuan University. Her research interests include data mining, knowledge management and discovery, and analysis in big data.

XUEJUN QI received the Ph.D. degree from the Huazhong University of Science and Technology, in 2014. He is currently an Associate Professor with Xihua University. His research interests include machine learning, thermal comfort, and built environment.

XIAOBO ZHOU received the Ph.D. degree in applied mathematics from Peking University, China.

He is currently a Professor and the Director of the Center for Systems Medicine, School of Biomedical Bioinformatics, University of Biomedical Bioinformatics, The University of Texas Health Science Center at Houston. His research interests include data sciences, imaging informatics, and clinical informatics.