



## OPEN ACCESS

## EDITED BY

Lisheng Xu,  
Northeastern University, China

## REVIEWED BY

Chunsheng Li,  
Shenyang University of Technology,  
China  
Yang Yao,  
Northeastern University, China  
Cai Chen,  
Shandong Institute of Advanced  
Technology (CAS), China

## \*CORRESPONDENCE

Shijie Chang,  
sjchang@cmu.edu.cn

<sup>†</sup>These authors have contributed equally  
to this work

## SPECIALTY SECTION

This article was submitted to  
Computational Physiology and  
Medicine,  
a section of the journal  
Frontiers in Physiology

RECEIVED 05 June 2022

ACCEPTED 02 August 2022

PUBLISHED 02 September 2022

## CITATION

Shan L, Li Y, Jiang H, Zhou P, Niu J, Liu R,  
Wei Y, Peng J, Yu H, Sha X and Chang S  
(2022), Abnormal ECG detection based  
on an adversarial autoencoder.  
*Front. Physiol.* 13:961724.  
doi: 10.3389/fphys.2022.961724

## COPYRIGHT

© 2022 Shan, Li, Jiang, Zhou, Niu, Liu,  
Wei, Peng, Yu, Sha and Chang. This is an  
open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which does  
not comply with these terms.

# Abnormal ECG detection based on an adversarial autoencoder

Lianfeng Shan<sup>1†</sup>, Yu Li<sup>2†</sup>, Hua Jiang<sup>3</sup>, Peng Zhou<sup>2</sup>, Jing Niu<sup>2</sup>,  
Ran Liu<sup>2</sup>, Yuanyuan Wei<sup>2</sup>, Jiao Peng<sup>2</sup>, Huizhen Yu<sup>2</sup>,  
Xianzheng Sha<sup>2</sup> and Shijie Chang<sup>2\*</sup>

<sup>1</sup>Department of Intelligent Computation, School of Intelligent Medicine, China Medical University, Shenyang, China, <sup>2</sup>Division of Biomedical Engineering, School of Intelligent Medicine, China Medical University, Shenyang, China, <sup>3</sup>Department of Cardiovascular Medicine, The First Affiliated Hospital of China Medical University, Shenyang, China

Automatic detection and alarm of abnormal electrocardiogram (ECG) events play an important role in an ECG monitor system; however, popular classification models based on supervised learning fail to detect abnormal ECG effectively. Thus, we propose an ECG anomaly detection framework (ECG-AAE) based on an adversarial autoencoder and temporal convolutional network (TCN) which consists of three modules (autoencoder, discriminator, and outlier detector). The ECG-AAE framework is trained only with normal ECG data. Normal ECG signals could be mapped into latent feature space and then reconstructed as the original ECG signal back in our model, while abnormal ECG signals could not. Here, the TCN is employed to extract features of normal ECG data. Then, our model is evaluated on an MIT-BIH arrhythmia dataset and CMUH dataset, with an accuracy, precision, recall, F1-score, and AUC of 0.9673, 0.9854, 0.9486, 0.9666, and 0.9672 and of 0.9358, 0.9816, 0.8882, 0.9325, and 0.9358, respectively. The result indicates that the ECG-AAE can detect abnormal ECG efficiently, with its performance better than other popular outlier detection methods.

## KEYWORDS

outlier detection (OD), autoencoder (AE), generative adversarial network (GANs), ECG, temporal convolutional network (TCN)

## 1 Introduction

Cardiovascular diseases (CVDs) are leading causes of human death (R.L. Sacco et al., 2016), and ECG is an important method of diagnosing CVDs. Earlier detection of abnormal ECG is the key step in prevention, identification, and diagnosis of CVDs. Portable ECG could detect sudden abnormal ECG events in the early stage (Dong and Zhu, 2004) and activate warning; it is expected to reduce the mortality rate. Therefore, automatic identification of abnormal ECG events is the first important part of an ECG monitoring system.

Currently, popular artificial intelligence (AI) ECG diagnosis methods, including machine learning (feature extraction and classifiers) and deep networks, always detect abnormal ECG events using classification models. In machine learning, self-organizing map (SOM) (M.R. Risk et al., 1997), C-means clustering (Özbay et al.,

2011), etc. are some of the successful machine learning methods for ECG classification. They extract features, such as wavelet coefficients (P. De Chazal et al., 2000) and autoregressive coefficients (N. Srinivasan et al., 2002), as ECG presentation. Other research studies focus on deep learning for ECG analysis, including convolutional neural networks (CNNs) (U.R. Acharya et al., 2017) and recurrent neural networks (RNNs) (H.M. Lynn et al., 2019). Xia used a deep convolutional neural network (DCNN) (Xia et al., 2018) for atrial fibrillation detection from short ECG signals (<5s) without any designed feature extraction procedure. Martin used long a short-term memory network (LSTM) (H. Martin et al., 2021) to detect myocardial infarction from a single lead ECG signal. Onan, 2020 proposed a CNN-LSTM framework for sentiment analysis of product review on Twitter. Onan and Tocoglu (2021) proposed a three-layer stacked bidirectional LSTM architecture to identify sarcastic text documents. Deep ECG (C. Li et al., 2021) takes ECG images as inputs and performs arrhythmia classification using the DCNN and transfer learning. Furthermore, a new method combining a recurrence plot (RP) and deep learning in two stages (B.M. Mathunjwa et al., 2021) is proposed to detect arrhythmias.

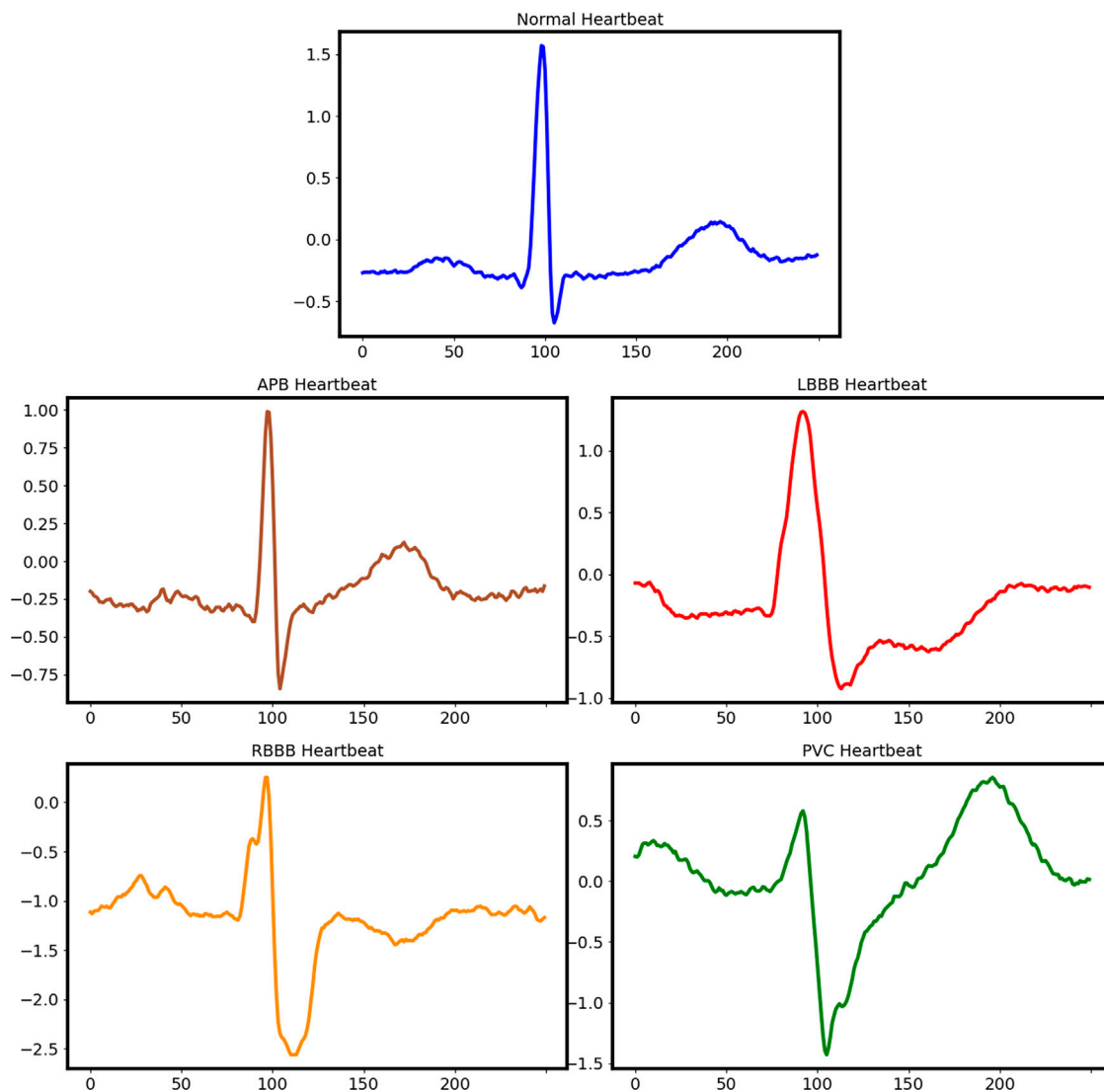
These aforementioned supervised learning ECG interpreting methods have achieved sound performance in previous studies. But these classification frameworks require the dataset to include all types of heart disease data with accurate manual annotation by professional doctors. The clinical ECG data are always imbalanced with fewer abnormal ECG samples, which makes it difficult to establish an effective classification model. Moreover, it is difficult to establish a large dataset including all types of abnormal ECG for clinical purposes in practice. Therefore, the sensitivity and specificity of abnormal ECG detection cannot meet clinical requirements (O. Faust et al., 2018). An outlier detection method (G. Pang et al., 2021) is more suitable for abnormal ECG in an early warning system, only based on normal data in clinical applications.

The outlier detection methods are unsupervised machine learning methods including clustering and semi-supervision including deep learning. In unsupervised methods, statistical methods usually focus on modeling the distribution of normal categories by learning the parameters of the probability model, to identify abnormal categories as outliers with low probability. The distance-based outlier detection methods assume that the normal categories are close to each other, while the abnormal samples are far away from the normal ones. Thus, outliers could be identified by calculating the distance between the abnormal and normal samples. Bin Yao and Hutchison (2014) proposed a density-based local outlier detection method (LOF) for uncertain data. H. Shibuya and Maeda (2016) developed an anomaly detection method based on multidimensional time-series sensor data and using normal state models. Principal

component analysis (Li and Wen, 2014) could be used for linear models; and the Gaussian mixture model (GMM) (Dai and Gao, 2013), isolation forest (F.T. Liu et al., 2008), and one-class support vector machine (OC-SVM) (B. Schölkopf et al., 2000) are used in actual outlier detection applications. But these machine learning algorithms often require the manual design of effective features.

Performance of an outlier detection method based on deep learning has been proved well, including Auto-Encoder (Zhou and Paffenroth, 2017), LSTM (P. Malhotra et al., 2015), and VAE (Wang et al., 2020), and widely used in AI-aided diagnosis (T. Fernando et al., 2021) such as X-ray film, MRI, CT, and other medical images, and in the detection of EEG, ECG, and other timing signals as well. Y. Xia et al. (2015) eliminated abnormal data from noisy data by reducing reconstruction errors of the autoencoder, and applying gradients of the autoencoder to make reconstruction errors discriminatory to positive samples. By using deep neural networks (autoencoders) as feature extractors, a deep hybrid model (DHM) has been applied for outlier detection to input extracted features into traditional outlier detection algorithms, such as OC-SVM (Mo, 2016). L. Ruff et al. (2018) used deep one-class classification for end-to-end outlier detection, effectively customizing trainable targets for outlier detection to extract features. K. Li et al. (2012) proposed a transfer learning framework for detecting abnormal ECG; however, this method requires manual coding of features and relies on labeled data for all different types of abnormalities. Due to diversity of diseases and different waveforms collected from different abnormal diseases, such data are not easy to obtain. Time series outlier detection technology is also used in ECG signal processing; Lemos and Tierra-Criollo, 2008; Chauhan and Vig, 2015 proposed an outlier detection method based on LSTM. An abnormal condition is considered when the difference between the predicted value of LSTM and normal value exceeds a given threshold. Latif et al. (2018) used a recurrent neural network (RNN) to detect abnormal heartbeats in the PCG signal detection of the heart sound, which needs a large amount of calculation. K. Wang et al. (2016) used an autoencoder to reconstruct normal ECG data, determine the threshold according to the reconstruction error, and finally, to detect the test set.

Recently, a GAN-based framework has been applied to outlier detection (T. Schlegl et al., 2017). The model generates new data according to the input; if the input was similar to the training data (as normal data), the output would be similar to the input, otherwise, the input would be an outlier. T. Schlegl et al. (2017) used a GAN-based model (AnoGAN) to identify anomalies in medical images. However, the aforementioned methods have the problems of overfitting (C. Esteban et al., 2017) or instability (D. Li et al., 2019) when they deal with abnormal ECG detection problems.



**FIGURE 1**  
Typical sketch of normal and abnormal heartbeats.

An autoencoder is another method of simply “memorizing” the training data and reproducing them. The parameters of the intermediate hidden layer would completely fit the training set, and the content of its memory will be completely output at the time of the output, resulting in identity mapping of the neural network and data overfitting. Problems such as instability and poor controllability occur with the latent model based on the GAN method.

In this study, we proposed a novel method named ECG-AAE for detecting abnormal ECG events, based on an adversarial autoencoder and TCN (L. Sun et al., 2015). It consists of three parts: 1) an autoencoder, 2) a discriminator, and 3) an outlier

detector. Our method was evaluated on the MIT-BIH and our CMUH datasets and compared with several other popular outlier detection methods.

## 2 Materials

### 2.1 Electrocardiogram datasets

1) Massachusetts Institute of Technology Arrhythmia Dataset (MIT-BIH). The dataset consists of 48 double-lead ECG recordings from 47 subjects; each set lasts 30 min at a

TABLE 1 Number of heartbeats involved in each dataset and the division of datasets.

Dataset	Type	Type of heartbeats	Number of heartbeats	Number of cases	Sample size	Number of training set	Number of test set
MIT-BIH	Normal	N	74,962	40	15,000	10,000	5,000
		Abnormal	A	2,545	—	5,000	0
		L	8,068	—			
		R	7,254	—			
		V	7,034	—			
		Total	99,863	47	20,000	10,000	10,000
CMUH	Normal	N	20,000	20,000	15,000	10,000	5,000
		Abnormal	A	6,811	6,811	5,000	0
		L	1,247	1,247			
		R	8,268	8,268			
		V	7,847	7,847			
		Total	44,173	44,173	20,000	10,000	10,000

sample rate of 360 Hz, with approximately 110,000 beats. A set of beat labels is equipped at the peak of R.

2) A CMUH dataset supported by the First Affiliated Hospital of China Medical University. The dataset contains 12-lead ECG records of inpatients in the First Affiliated Hospital of China Medical University from January 2013 to December 2017, with a sampling rate of 560 Hz.

Two or three cardiologists annotated all heartbeats for both datasets independently. Only lead II ECG signals are used in this study.

## 2.2 Data preprocessing

A total of four types of arrhythmia and normal beats are selected from datasets: right bundle branch block (R), left bundle branch block (L), atrial premature beat (A), ventricular premature beat (V), and normal sinus rhythm (N). ECG signals are split into single heartbeats which are normalized to a range of  $[-1, 1]$  for network training. Five typical heartbeats are shown in [Figure 1](#).

## 2.3 MIT-BIH dataset

In this study, 45 lead II signal records are selected from the MIT-BIH dataset (records 102, 104, and 114 were excluded, as they do not include the lead II data or the type of heart disease in our experiments). Wavelet transform is used to reduce noise and baseline drift ([Alfaouri and Daqrouq, 2008](#)). Then, the ECG data are split into single heartbeats using the marked R peak location. A total of 250 points (100 points before the R

peak and 150 points after the R peak) are included in a heartbeat.

## 2.4 CMUH dataset

ECG data of 44,173 people from the CMUH dataset have been selected for this study. Data are resampled at 360 Hz to maintain consistency with MIT-BIH data. The beat segmentation method is the same as the one mentioned previously.

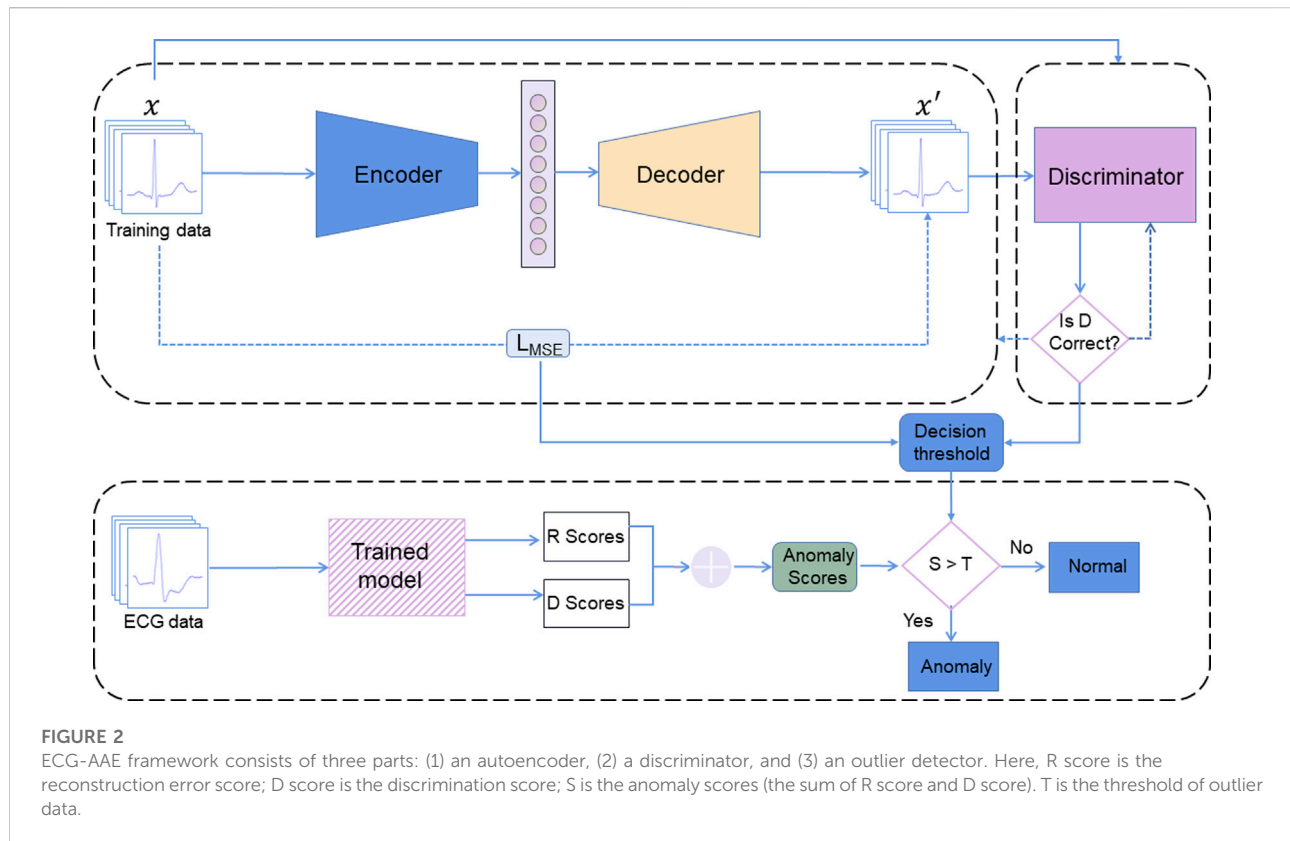
For each dataset, 10,000 normal ECG data are randomly selected as the training set, and 5,000 normal ECG data and 5,000 abnormal ECG data are randomly selected as the test set, as shown in [Table 1](#).

## 3 Methods

### 3.1 ECG-AAE framework

The ECG-AAE framework consists of three parts: 1) an autoencoder, 2) a discriminator, and 3) an outlier detector, as shown in [Figure 2](#). The autoencoder tries to minimize reconstruction errors to generate ECG signals similar to input signals. The discriminator uses reconstructed and original data as the input, and is trained to distinguish normal data from reconstructed data. Both the autoencoder and discriminator update simultaneously to improve the reconstruction performance of the autoencoder.

Finally, the combination of reconstruction errors and discriminant scores (probability output of discriminator) is used to evaluate normal ECG. Test data are mapped back to



potential space, and loss between reconstructed test samples and actual test samples has been applied to calculate the corresponding reconstruction loss.

A detailed network of the ECG-AAE is shown in Table 2. The encoder is composed of three TCN blocks, three MaxPooling1D layers, a flatten layer, and a dense layer. The decoder is composed of a dense layer, three TCN blocks, three UpSampling1D layers, and a Conv1D layer. The discriminator is composed of three TCN blocks, three MaxPooling1D layers, a flatten layer, and two dense layers. The activation function for the last dense layer is sigmoid. A large discriminator can make the data overflow easily, while a shallow autoencoder cannot generate enough real data to defeat the discriminator. A small number of hidden units is chosen as the starting point, and the number of hidden units has been gradually increased in each successive layer, which is effective for the training of the model in this study. Also, three TCN blocks are used in the encoder, decoder, and discriminator.

In this study, stochastic gradient descent (Adam) (Kingma and Ba, 2015) is adopted to conduct alternating update training for each lost component, and parameters of the network model are obtained through training and learning.

### 3.2 Temporal convolutional network

Atemporal convolutional network (TCN) (L. Sun et al., 2015) could capture long-term dependence in an ECG sequence more effectively. A TCN block is superimposed by two causal convolution layers with the same expansion factor, followed by normalization, ReLU, and dropout layers, as shown in Figure 3.

The TCN is used to extract features of ECG time series data. The TCN module has shown competitiveness in many sequence-related modeling tasks (W. Zhao et al., 2019). It can capture dependencies in sequences more effectively than recurrent neural networks (Graves et al., 2013; Z. Huang et al., 2015; J. Chung et al., 2014). The TCN convolution kernel is shared in the same layer, with lower requirement memory.

The TCN is mainly composed of dilated causal convolution. Figure 4 shows a simple structure of TCNs, where  $x_i$  represents the characteristics of the  $i$ th moment. Expanded convolution enables input interval sampling during convolution, and the sampling rate is controlled by  $d$ . The parameter  $d = 1$  in the bottom layer means that every point is sampled as input, and  $d = 2$  in the middle layer means that every two points are sampled as input. Generally, the higher the level, the larger will be the value

TABLE 2 Detailed overview of the proposed ECG-AAE model.

Modules	Layers	Types	Activation function	Output shapes	Kernel size	No. of filters
Encoder	0	Input	—	$250 \times 1$	—	—
	1	TCN block	ReLU	$250 \times 32$	9	32
	2	MaxPooling1D	—	$50 \times 32$	—	—
	3	TCN block	ReLU	$50 \times 16$	9	16
	4	MaxPooling1D	—	$10 \times 16$	—	—
	5	TCN block	ReLU	$10 \times 8$	9	8
	6	MaxPooling1D	—	$2 \times 8$	—	—
	7	Flatten	—	16	—	—
	8	Dense	ReLU -	8	—	—
Decoder	0	Input	—	8	—	—
	1	Dense	ReLU	16	—	—
	2	Reshape	—	$2 \times 8$	—	—
	3	UpSampling1D	—	$10 \times 8$	—	—
	4	TCN block	ReLU	$10 \times 8$	9	8
	5	UpSampling1D	—	$50 \times 16$	—	—
	6	TCN block	ReLU	$50 \times 16$	9	16
	7	UpSampling1D	—	$250 \times 16$	—	—
	8	TCN block	ReLU	$250 \times 32$	9	32
	9	Conv1D	ReLU	$250 \times 1$	9	1
Discriminator	0	Input	—	$250 \times 1$	—	—
	1	TCN block	ReLU	$250 \times 32$	9	32
	2	MaxPooling1D	—	$50 \times 32$	—	—
	3	TCN block	ReLU	$50 \times 16$	9	16
	4	MaxPooling1D	—	$10 \times 16$	—	—
	5	TCN block	ReLU	$10 \times 8$	9	8
	6	MaxPooling1D	—	$2 \times 8$	—	—
	7	Flatten	—	16	—	—
	8	Dense	ReLU	8	—	—
	9	Dense	sigmoid	1	—	—

of  $d$  used, with the size of the effective window of dilated convolution increasing exponentially with the number of levels. Convolution networks can obtain a larger receptive field with fewer layers.

The TCN uses a residual block structure which is similar to that in ResNet to solve problems such as a deeper network structure causing gradient disappearance, to make the model more generic. A residual block superimposes multiple causal convolutional layers with the same expansion factor, followed by normalization, ReLU, and dropout. In this study, a residual block containing two layers of convolution and nonlinear mapping is constructed, and normalization and dropout to each layer are added to regularize the network, as shown in Figure 3.

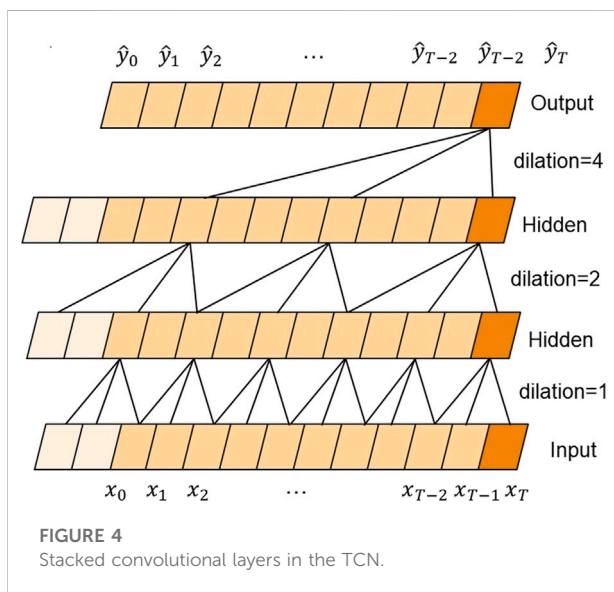
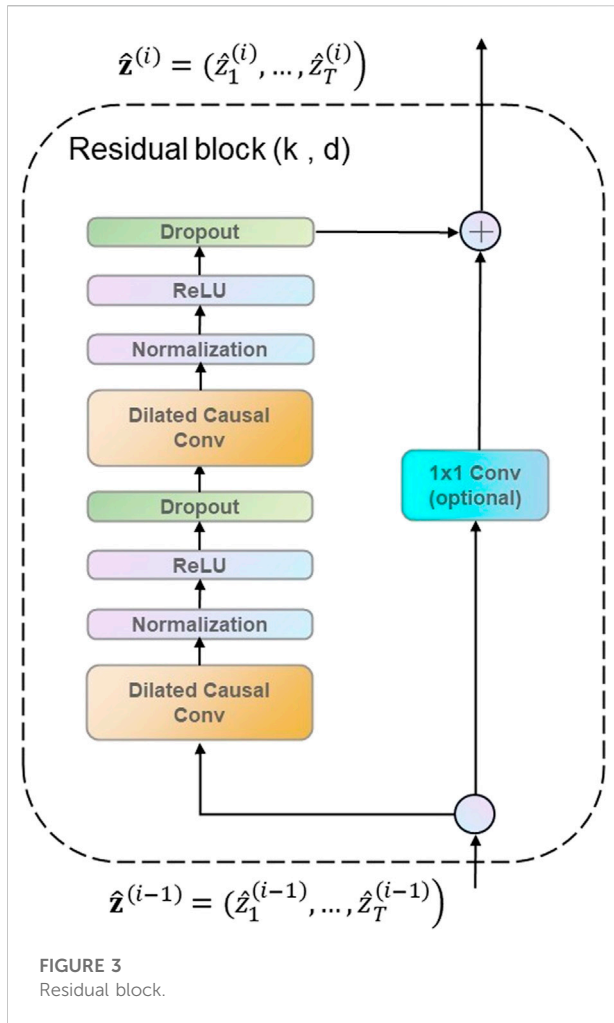
### 3.3 Autoencoder module

An autoencoder module consists of three parts: an encoder, a hidden layer, and a decoder. Only normal ECG data are used for training. First, input data  $x$  are compressed and encoded into the hidden layer data, and then hidden layer data are decoded to obtain reconstructed ECG data  $X'$ . The loss function during training is the reconstruction error between input data  $x$  and output data  $X'$ :

$$\text{Loss}(X, X') = \|X - X'\|^2 \quad (1)$$

The encoder and decoder are optimized to minimize reconstruction errors of normal ECG using training data  $X$ .





The activation functions of the encoder and decoded neural networks are shown as follows:

$$Z = \delta(WX + b) \tag{2}$$

$$X' = \delta'(W'Z + b') \tag{3}$$

where,  $\delta$  and  $\delta'$  are non-linear exciting functions, and  $W, b, W',$  and  $b'$  are weights and offsets of linear transformations.

Minimizing the loss function to optimize the parameters in the encoder and decoder is equivalent to a nonlinear optimization problem:

$$\min_{\delta, w, b} \text{Loss}(X, X') = \|X - \delta'(\delta(WX + b)) + b'\|^2 \tag{4}$$

### 3.4 Discriminator module

The discriminator (D) is to distinguish reconstructed ECG data  $X'$  generated by the autoencoder (AE) from real data  $X$  during the training process, and to make reconstructed data similar to the input data. Thus, the autoencoder tries to minimize the reconstruction error, while the discriminator tries to maximize it. During training, the two modules optimize themselves and improve refactoring and discrimination. The autoencoder is trained to minimize the difference between reconstructed and input samples, and the discriminator is trained to maximize confidence in discriminating the difference between reconstructed and real samples. After training, the discriminator assigns correct labels to real and fake ECG data as sensitively as possible, while the autoencoder generates real ECG data as much as possible to deceive the discriminator, and the two reach a balance (D. Li et al., 2019). The conditional autoencoder and discriminator are trained following a two-player minimax game:

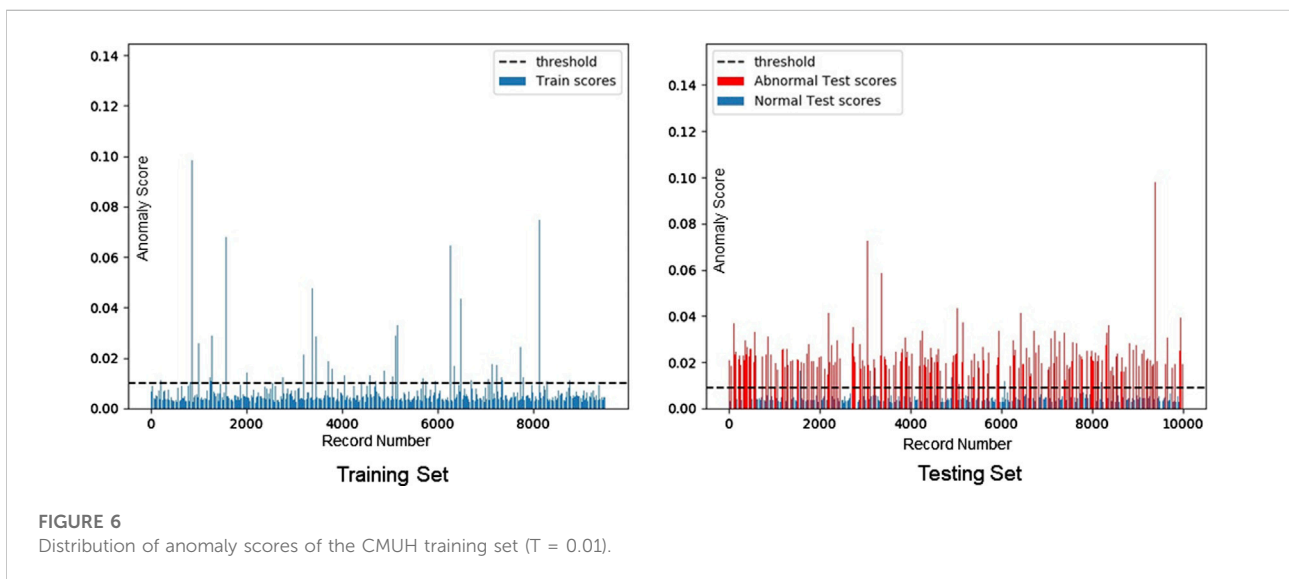
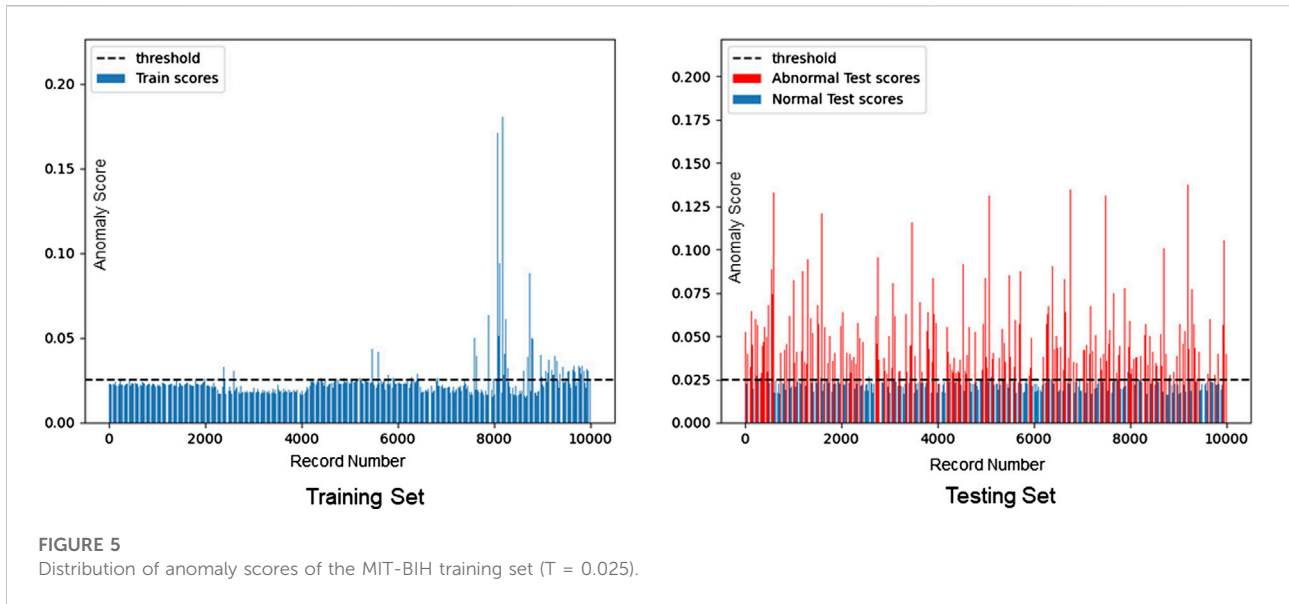
$$\min_{\text{AE}} \max_{\text{D}} V(D, \text{AE}) = \epsilon_{x \sim \text{pdata}(X)} [\log D(x)] + \epsilon_{x \sim \text{pz}(Z)} [\log(1 - D(\text{AE}))] \tag{5}$$

### 3.5 Outlier detection module

The combination of reconstruction errors and discriminant scores is used to define the abnormal score. Reconstruction loss  $R(x)$  makes a higher score on abnormal ECG data and a lower score on normal ECG data. The discrimination score  $D(x)$  produces lower scores on abnormal ECG data and higher scores on normal ECG data.

Therefore, the anomaly score  $a(x)$  formula is expressed as

$$a(x) = (1 - \lambda)R(x) + \lambda \frac{1}{D(x)} \tag{6}$$



$\lambda = 0$ , according to our experience. The threshold is decided following one standard deviation above the mean. ECG data with  $a(x)$  greater than the threshold are abnormal.

## 4 Results

### 4.1 Evaluation indexes

Accuracy (ACC), precision (Pre), recall (Rec), F1-score ( $F_1$ ), and AUC value (area under the ROC curve) are used to evaluate the performance of our ECG-AAE and compare it with other

methods. In the confusion matrix, abnormal ECG is defined as positive, normal ECG is defined as negative, and true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are calculated.

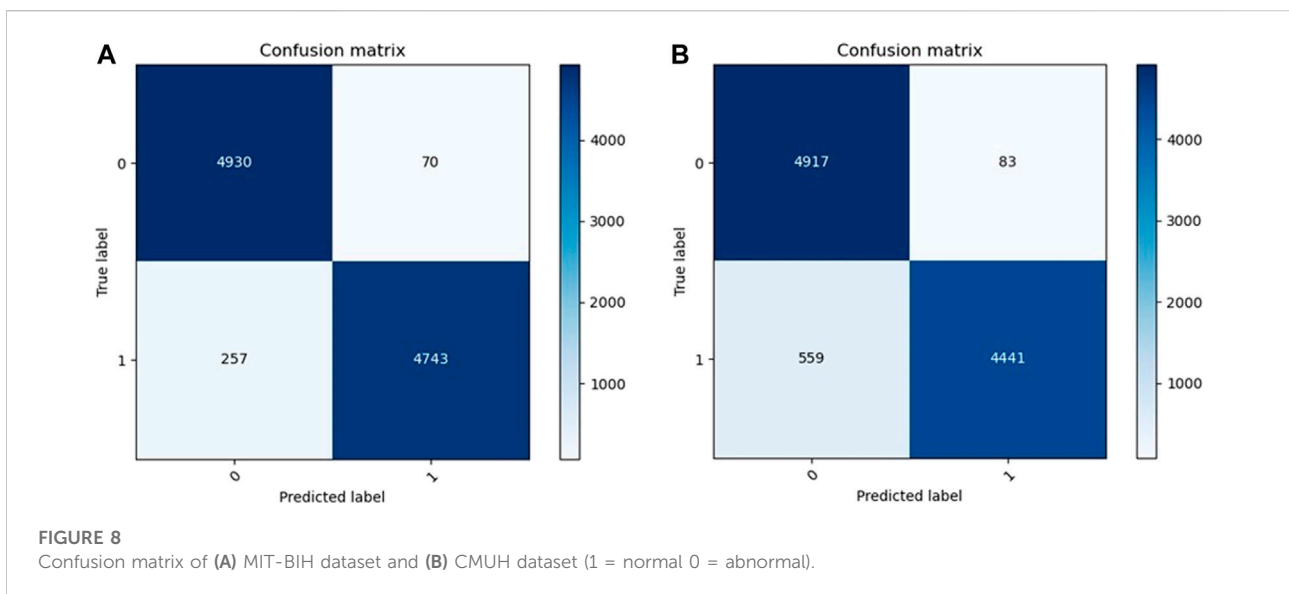
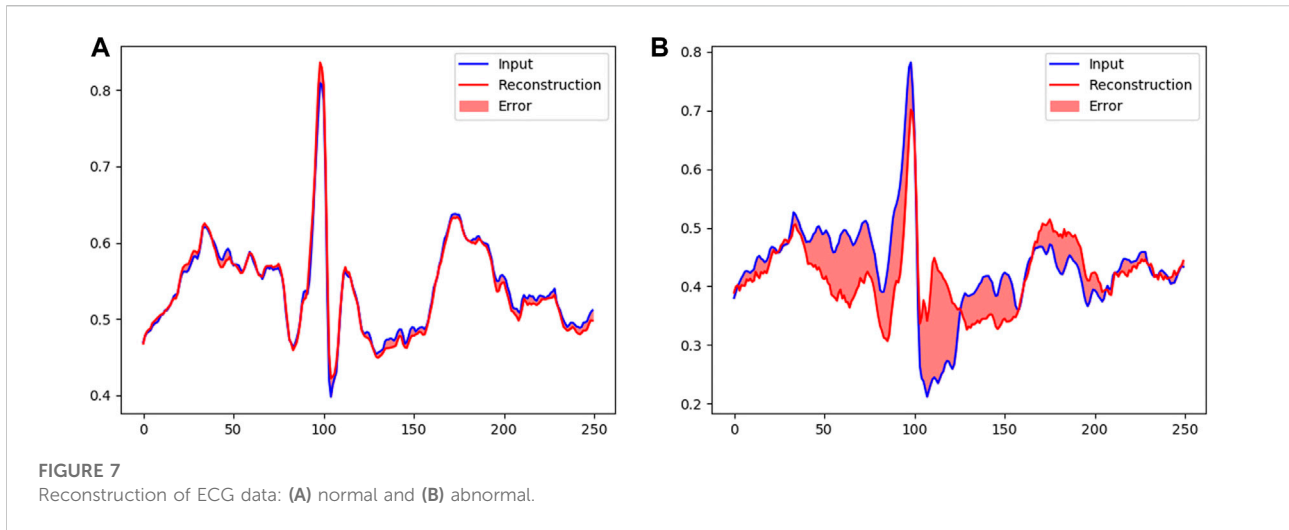
$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Pre = \frac{TP}{TP + FP}$$

$$Rec = \frac{TP}{TP + FN}$$

$$F_1 = 2 \times \frac{Pre \times Rec}{Pre + Rec}$$





In clinical practice, the precision rate represents the proportion of patients with true ECG abnormalities, while recall rate represents the proportion of patients with true ECG abnormalities. The high-precision detection model could prevent misdiagnosis, while the detection model with a high recall rate could avoid missed diagnosis. The F1-score is a weighted harmonic average of the recall rate and accuracy rate; the F1-score and AUC value are used as the main indicators to measure the performance of outlier detection in this study.

The experiment was implemented on a workstation (Dell T7600, Xeron 2,650 × 2, 256 GB RAM, 1080Ti×2), with Linux 18.04, Python 3.6, Keras 2.3.1, and TensorFlow 2.0.

### 4.1.1 Experiment 1

Both MIT-BIH and CMUH datasets have been used to verify the performance of our framework. The threshold value is selected as one standard deviation above the mean according to the abnormal score in the training set. T values of MIT-BIH and CMUH datasets can be obtained as 0.025 and 0.01, respectively. When the training set includes normal data only, its abnormal scores are within the range of the threshold T (Figures 5A, 6A), while, in the test dataset including both normal and abnormal ECG data, the abnormal scores are less than the threshold T for normal ECG data, but are greater than the threshold T (Figures 5B, 6B) for abnormal ECG data.

TABLE 3 Average classification performance for different methods on the MIT-BIH dataset.

Methods	Acc $\pm$ SD	Pre $\pm$ SD	Rec $\pm$ SD	F1-score $\pm$ SD	AUC $\pm$ SD
OURS	<b>0.9673 <math>\pm</math> 0.0005</b>	0.9854 $\pm$ 0.0003	0.9486 $\pm$ 0.0001	<b>0.9666 <math>\pm</math> 0.0014</b>	<b>0.9672 <math>\pm</math> 0.0015</b>
AnoGAN (Schlegl et al.)	0.9257 $\pm$ 0.0101	0.8829 $\pm$ 0.0167	0.9876 $\pm$ 0.0027	0.9323 $\pm$ 0.0085	0.9283 $\pm$ 0.0101
AE (K.Wang et al.)	0.9282 $\pm$ 0.0180	0.8733 $\pm$ 0.2042	<b>0.9902 <math>\pm</math> 0.0233</b>	0.9281 $\pm$ 0.1490	0.9233 $\pm$ 0.0049
VAE (X.Wang et al.)	0.8048 $\pm$ 0.0028	0.7196 $\pm$ 0.0029	0.9874 $\pm$ 0.0002	0.8325 $\pm$ 0.0157	0.8013 $\pm$ 0.0028
Stack LSTM (Chauhan et al.)	0.8875 $\pm$ 0.0017	0.8313 $\pm$ 0.0021	0.9740 $\pm$ 0.0007	0.8970 $\pm$ 0.0019	0.8882 $\pm$ 0.0052
GRU (Cowton et al.)	0.8764 $\pm$ 0.0040	0.8128 $\pm$ 0.0064	0.9746 $\pm$ 0.0017	0.8864 $\pm$ 0.0031	0.8751 $\pm$ 0.0040
RNN (Latif et al.)	0.8568 $\pm$ 0.0031	0.7826 $\pm$ 0.0040	0.9798 $\pm$ 0.0003	0.8702 $\pm$ 0.0024	0.8538 $\pm$ 0.0031
DEEP-SVDD (Ruff et al.)	0.8039 $\pm$ 0.0035	0.7221 $\pm$ 0.0037	0.8342 $\pm$ 0.0002	0.8342 $\pm$ 0.0025	0.8037 $\pm$ 0.0033
AE + OCSVM (Mo et al.)	0.8624 $\pm$ 0.0036	0.7965 $\pm$ 0.0046	0.9788 $\pm$ 0.0003	0.8783 $\pm$ 0.0029	0.8644 $\pm$ 0.0050
DAGMM (Song et al.)	0.7646 $\pm$ 0.0007	<b>0.9992 <math>\pm</math> 0.0008</b>	0.5304 $\pm$ 0.0004	0.6930 $\pm$ 0.0019	0.7650 $\pm$ 0.0019
GMM (Dai et al.)	0.6462 $\pm$ 0.0463	0.9986 $\pm$ 0.1603	0.2924 $\pm$ 0.0068	0.4524 $\pm$ 0.0274	0.6460 $\pm$ 0.0042
OCSVM (Schölkopf et al.)	0.8376 $\pm$ 0.0009	0.9982 $\pm$ 0.0006	0.6760 $\pm$ 0.0005	0.8061 $\pm$ 0.0018	0.8374 $\pm$ 0.0019
iForest (Liu et al.)	0.6521 $\pm$ 0.0106	0.9987 $\pm$ 0.2119	0.3046 $\pm$ 0.3468	0.4668 $\pm$ 0.1334	0.6521 $\pm$ 0.0106
LOF (Bin Yao et al.)	0.5050 $\pm$ 0.0006	0.5027 $\pm$ 0.0007	0.9170 $\pm$ 0.0025	0.6494 $\pm$ 0.0018	0.5050 $\pm$ 0.0020

SD, standard deviation.  
The bold values mean maximum.

TABLE 4 Average classification performance for different methods on the CMUH dataset.

Methods	Acc $\pm$ SD	Pre $\pm$ SD	Rec $\pm$ SD	F1-score $\pm$ SD	AUC $\pm$ SD
OURS	<b>0.9358 <math>\pm</math> 0.0004</b>	0.9816 $\pm$ 0.0002	0.8882 $\pm$ 0.0010	<b>0.9325 <math>\pm</math> 0.0008</b>	<b>0.9358 <math>\pm</math> 0.00010</b>
AnoGAN (Schlegl et al.)	0.8985 $\pm$ 0.0092	0.8396 $\pm$ 0.0128	0.9852 $\pm$ 0.0018	0.9066 $\pm$ 0.0078	0.8985 $\pm$ 0.0092
AE (K.Wang et al.)	0.9103 $\pm$ 0.0181	0.8504 $\pm$ 0.0253	<b>0.9946 <math>\pm</math> 0.0012</b>	0.9169 $\pm$ 0.0148	0.9098 $\pm$ 0.0181
VAE (X.Wang et al.)	0.7744 $\pm$ 0.0040	0.6885 $\pm$ 0.0039	0.9910 $\pm$ 0.0015	0.8125 $\pm$ 0.0027	0.7713 $\pm$ 0.0041
Stack LSTM (Chauhan et al.)	0.8754 $\pm$ 0.0033	0.8097 $\pm$ 0.0051	0.9772 $\pm$ 0.0019	0.8856 $\pm$ 0.0025	0.8738 $\pm$ 0.0033
GRU (Cowton et al.)	0.8779 $\pm$ 0.0038	0.8156 $\pm$ 0.0052	0.9748 $\pm$ 0.0019	0.8881 $\pm$ 0.0030	0.8772 $\pm$ 0.0038
RNN (Latif et al.)	0.8221 $\pm$ 0.0037	0.7414 $\pm$ 0.0041	0.9860 $\pm$ 0.0021	0.8464 $\pm$ 0.0026	0.8210 $\pm$ 0.0037
DEEP-SVDD (Ruff et al.)	0.7649 $\pm$ 0.0050	0.6794 $\pm$ 0.0047	0.9908 $\pm$ 0.0017	0.8061 $\pm$ 0.0032	0.7616 $\pm$ 0.0050
AE + OCSVM (Mo et al.)	0.8245 $\pm$ 0.0030	0.7436 $\pm$ 0.0034	0.9864 $\pm$ 0.0022	0.8479 $\pm$ 0.0021	0.8231 $\pm$ 0.0030
DAGMM (Song et al.)	0.7260 $\pm$ 0.0012	0.9991 $\pm$ 0.0004	0.4520 $\pm$ 0.0024	0.6224 $\pm$ 0.0023	0.7258 $\pm$ 0.0012
GMM (Dai et al.)	0.6057 $\pm$ 0.0037	<b>1.0000 <math>\pm</math> 0.0005</b>	0.2148 $\pm$ 0.0074	0.3536 $\pm$ 0.0101	0.6074 $\pm$ 0.0037
OCSVM (Schölkopf et al.)	0.7600 $\pm$ 0.0024	0.9985 $\pm$ 0.0010	0.5208 $\pm$ 0.0048	0.6845 $\pm$ 0.0041	0.7600 $\pm$ 0.0024
iForest (Liu et al.)	0.6303 $\pm$ 0.0043	<b>1.0000 <math>\pm</math> 0.0009</b>	0.2606 $\pm$ 0.0086	0.4135 $\pm$ 0.0107	0.6303 $\pm$ 0.0043
LOF (Bin Yao et al.)	0.5767 $\pm$ 0.0059	<b>1.0000 <math>\pm</math> 0.0010</b>	0.1700 $\pm$ 0.0117	0.2906 $\pm$ 0.0174	0.5850 $\pm$ 0.0059

SD, standard deviation.  
The bold values mean maximum.

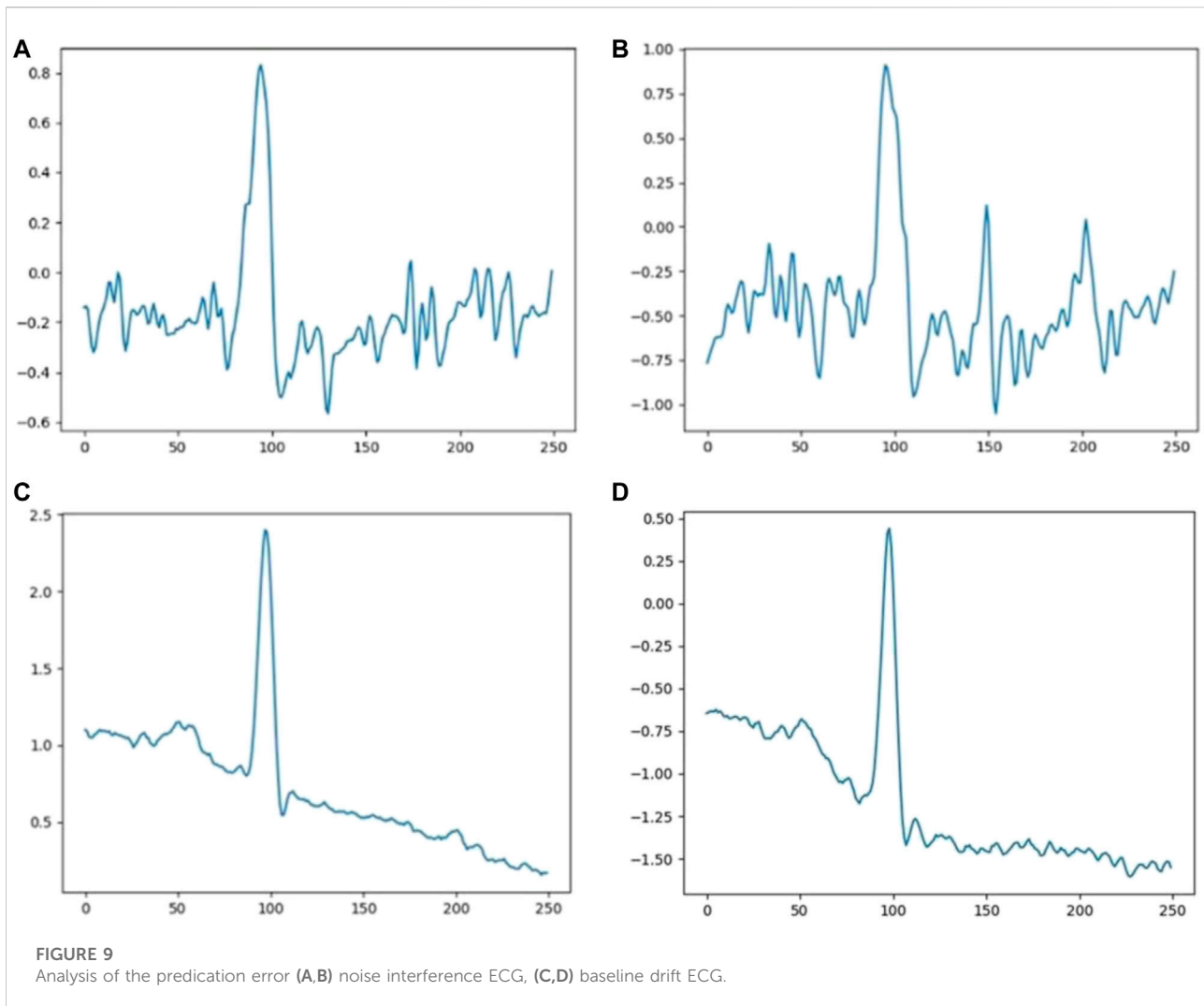
An example of normal and abnormal ECG data reconstructed by our model is shown in Figure 7. For normal ECG data, reconstructed data are continuous, and the shape of the reconstructed waveform is basically the same as the input one, with an error range of  $0.0063 \pm 0.0098$ . For abnormal ECG data, the shape of the reconstructed waveform differs greatly from that of the input waveform. Although the reconstructed data are continuous, the error range reaches  $0.0289 \pm 0.0264$ .

The confusion matrixes of detection results are shown in Figure 8. In the MIT-BIH dataset, 4,930 abnormal ECGs were

detected, and 257 normal ECGs were predicted as abnormal. In our CMUH dataset, 4,917 abnormal ECG data were detected, and 559 normal ECGs were predicted as abnormal. The accuracy, recall, F1 score, and AUC of our model are 0.9673, 0.9854, 0.9486, 0.9666, and 0.9672, and 0.9358, 0.9816, 0.8882, 0.9325, and 0.9358, respectively.

#### 4.1.2 Experiment 2

Our method was compared with 13 popular outlier detection methods using MIT datasets, as shown in Table 3. Among the five



evaluation indicators, our model achieves the highest score of 0.9673 in accuracy. DAGMM achieves the highest score of 0.9992 in precision, but its recall is 0.5304. This shows that DAGMM tries to predict the sample as a positive sample when it is “more certain,” but misses many unsure positive samples due to its excessive conservativeness. The AE achieves the highest recall score of 0.9902, but the precision is 0.8829, indicating that the AE produces more false positives. The ECG-AAE model achieves the highest scores of 0.9673, 0.9666, and 0.9672 in accuracy, F1-score, and AUC value, respectively, better than other models.

#### 4.1.3 Experiment 3

We further verify the robustness and generalization of the model with our CMUH dataset, as shown in Table 4.

Our model achieves the highest scores of 0.9358, 0.9325, and 0.9358 in accuracy, F1-score, and AUC, respectively. GMM, iForest, and LOF models achieve the highest score of 1.000 in

precision, but the recall was lower. The AE achieves the highest recall of 0.9946, but the F1-score and AUC value are lower.

## 5 Discussion

To solve problems that the classification model cannot effectively detect in abnormal ECGs, we propose the ECG-AAE, a framework for detecting abnormal ECG signals. Its performance is verified and compared with the AE, AnoGAN, and other 11 popular outlier detection methods on the MIT-BIH arrhythmia dataset and our CMUH dataset.

The four kinds of machine learning outlier detection algorithms with low performance scores were GMM (Dai and Gao, 2013), OCSVM (B. Schölkopf et al., 2000), iForest (F.T. Liu et al., 2008), and LOF (S.H. Bin Yao and Hutchison, 2014). Among them, GMM enjoys the best performance, whose AUC values are 0.6460 and 0.6074 on the MIT-BIH and CMUH

datasets respectively; LOF, the worst model, shows AUC values are 0.5050 and 0.5850, respectively. It suggests that the machine learning methods might not be the best choice for abnormal ECG detection; they may not extract abnormal ECG effectively. Moreover, the subsequent classifiers could not fit the boundary functions in high-dimension feature space, while, the deep learning models could make ECG feature extraction more elastic to fit the nonlinear feature distribution, and finally improve the detection rate of abnormal ECG while ensuring accuracy.

Among deep learning models, generative models based on AE or GAN are better than hybrid models of machine learning and deep learning (e.g., AE + OCSVM (Mo, 2016) deep-SVDD (L. Ruff et al., 2018), DAGMM (Q. Song et al., 2018), RNN and its variants, LSTM, GRU, and other recurrent neural network models). The autoencoder Cowton et al., (K. Wang et al., 2016) encodes one-dimensional signal data into a lower dimension to learn the general distribution of data and then decodes to a higher dimension to reconstruct data. In this experiment, the AE performs well on both the MIT-BIH and CMUH datasets.

The ECG-AAE combines the autoencoder and discriminator, and it uses the autoencoder to realize reconstruction of the ECG and the discriminator to improve the generation ability of the autoencoder. The TCN could obtain ECG features at different scales with different receptive fields, which helps accurately reconstruct the normal ECG. In addition, the TCN avoids problems of gradient disappearance or gradient explosion. We use the combination of reconstruction errors and discriminant scores as the anomaly score, which effectively reduces the impact of the AE overfitting and instability of the GAN model. Compared with methods dealing with two leads or more, Liu F (2020) provided an accuracy of 97.3% in ECG anomaly detection; Thill et al. (2021) designed a temporal convolutional network autoencoder (TCN-AE) based on dilated convolutions for time series data.

Experiments 2 and 3 suggest that the CMUH dataset is about 0.3% lower than the MIT-BIH dataset on each model. The reason is that all the heartbeats in the MIT-BIH dataset are only from 48 people. These independent heartbeats are obtained through heartbeat segmentation, with very similar characteristics which are not enough for generalization, while each heartbeat in our CMUH dataset comes from a signal person, which is more in line with reality.

False positive data are largely affected by noise interference, as shown in Figures 9A,B. At the same time, false negative data in the experiment have also been analyzed with the finding that a baseline exists in most cases, as shown in Figures 9C,D. The ECG-AAE model can tolerate noise and baseline drift of conventional static ECG, but the form of input data in these error cases is quite different from that of normal ECG data. This situation might occur when patients move in a large range. Although noise filtering and baseline drift are carried out in the data preprocessing stage, an ideal effect is not achieved on the ECG data with large variation, which leads to a false positive and

negative output of the model. In clinical practice, false positives and negatives can be avoided by analyzing several continuous heartbeats, and when the several continuous heartbeats are judged to be abnormal ECGs, abnormal ECGs can be diagnosed.

## 6 Conclusion

Detection and early warning of sudden abnormal ECG is an important procedure in an ECG monitoring and alarm system. The ECG-AAE framework proposed in this study could efficiently detect abnormal ECG signals, and provide better performance on several indicators in our tests. It also suggests that outlier detection performs better than the classical classification framework in clinical practices. As far as we know, this is the first study to combine the adjournment network of abnormal ECG detection, which solves all types of abnormal ECG data and data balance problems and effectively improves the detection rate of abnormal ECG in the open set condition while ensuring accuracy.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

## Author contributions

LS and YL contributed equally to design of the study, performed the statistical analysis, and wrote the first draft of the manuscript. HJ organized the dataset and interpreted our ECG data. PZ contributed to debug the program. JN, RL, YW, JP, and HY organized the dataset. SC contributed to the concept and revised the manuscript. All authors contributed to approve the submitted version.

## Funding

This work was supported by Liaoning Natural Science Funds for Medicine and Engineering Interdisciplines 2021 (1600779161987) and Big Data Research for Health Science of China Medical University (Key Project No. 6).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Acharya, U. R., Fujita, H., Oh, S. L., Hagiwara, Y., Tan, J. H., and Adam, M. (2017). Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Inf. Sci. (N. Y.)* 415–416, 190–198. doi:10.1016/j.ins.2017.06.027
- Alfaouri, M., and Daqrouq, K. (2008). ECG signal denoising by wavelet transform thresholding. *Am. J. Appl. Sci.* 5 (3), 276–281. doi:10.3844/ajassp.2008.276.281
- Bin Yao, S. H., and Hutchison, D. (2014). *Density-based local outlier detection on uncertain data*. doi:10.11896/j.issn.1002-137X.2015.5.046
- Chauhan, S., and Vig, L. (2015). Anomaly detection in ECG time signals via deep long short-term memory networks. *Proc. 2015 IEEE Int. Conf. Data Sci. Adv. Anal. DSAA 2016*. doi:10.1109/DSAA.2015.7344872
- Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). *Empirical evaluation of gated recurrent neural networks on sequence modeling*. Available at: <http://arxiv.org/abs/1412.3555>.
- Cowton, J., Kyriazakis, I., Plötz, T., and Bacardit, J. (2018). A combined deep learning GRU-autoencoder for the early detection of respiratory disease in pigs using multiple environmental sensors. *Sensors Switz.*, E2521. doi:10.3390/s18082521
- Dai, X., and Gao, Z. (2013). From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis. *IEEE Trans. Ind. Inf.* 9, 2226–2238. doi:10.1109/TII.2013.2243743
- De Chazal, P., Celler, B. G., and Reilly, R. B. (2000). Using wavelet coefficients for the classification of the electrocardiogram. *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.* 1, 64–67. doi:10.1109/IEMBS.2000.900669
- Dong, J., and Zhu, H. H. (2004). Mobile ECG detector through GPRS/Internet. *Proc. IEEE Symp. Comput. Med. Syst.* 17, 485–489. doi:10.1109/cbms.2004.1311761
- Esteban, C., Hyland, S. L., and Rättsch, G. (2017). *Real-valued (medical) time series generation with recurrent conditional GANs*. Available at: <http://arxiv.org/abs/1706.02633>.
- Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S., and Acharya, U. R. (2018). Deep learning for healthcare applications based on physiological signals: A review. *Comput. Methods Programs Biomed.* 161, 1–13. doi:10.1016/j.cmpb.2018.04.005
- Fernando, T., Gammulle, H., Denman, S., Sridharan, S., and Fookes, C. (2021). Deep learning for medical anomaly detection – a survey. *ACM Comput. Surv.* 54, 1–37. doi:10.1145/3464423
- Graves, A., Mohamed, A., and Hinton, G. (2013). *Speech recognition with deep recurrent neural networks*, 3. Department of Computer Science, University of Toronto, 45–49. Dep. Comput. Sci. Univ. Toronto. Available at: <https://ieeexplore.ieee.org/document/6638947>.
- Huang, Z., Xu, W., and Yu, K. (2015). *Bidirectional LSTM-CRF models for sequence tagging*. Available at: <http://arxiv.org/abs/1508.01991>.
- Kingma, D. P., and Ba, J. L. (2015). Adam: A method for stochastic optimization, 3rd. *Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, 1–15.
- Latif, S., Usman, M., Rana, R., and Qadir, J. (2018). Phonocardiographic sensing using deep learning for abnormal heartbeat detection. *IEEE Sens. J.* 18, 9393–9400. doi:10.1109/JSEN.2018.2870759
- Lemos, C. W. M., and Tierra-Criollo, C. J. (2008). ECG anomalies identification using a time series novelty. *Detect. Tech.* 18, 766–769. doi:10.1007/978-3-540-74471-9
- Li, C., Zhao, H., Lu, W., Leng, X., and Xiang, J. (2021). DeepECG: Image-based electrocardiogram interpretation with deep convolutional neural networks. *Biomed. Signal Process. Control* 69, 102824. doi:10.1016/j.bspc.2021.102824
- Li, D., Chen, D., Jin, B., Shi, L., Goh, J., and Ng, S. K. (2019). MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks. *Lect. Notes Comput. Sci. Incl. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma.* 11730 LNCS, 703–716. doi:10.1007/978-3-030-34090-4\_56
- Li, K., Du, N., and Zhang, A. (2012). Detecting ECG abnormalities via transductive transfer learning. *ACM Conf. Bioinforma. Comput. Biol. Biomed. BCB 2012*, 210–217. doi:10.1145/2382936.2382963
- Li, S., and Wen, J. (2014). A model-based fault detection and diagnostic methodology based on PCA method and wavelet transform. *Energy Build.* 68, 63–71. doi:10.1016/j.enbuild.2013.08.044
- Liu, F. T., Ting, K. M., and Zhou, Z. H. (2008). Isolation forest. *Proc. - IEEE Int. Conf. Data Min. ICDM.*, 413–422. doi:10.1109/ICDM.2008.17
- Lynn, H. M., Pan, S. B., and Kim, P. (2019). A deep bidirectional GRU network model for biometric electrocardiogram classification based on recurrent neural networks. *IEEE Access* 7, 145395–145405. doi:10.1109/ACCESS.2019.2939947
- Malhotra, P., Vig, L., Shroff, G., and Agarwal, P. (2015). Long short term memory networks for anomaly detection in time series, 23rd Eur. Symp. Artif. Neural Networks. *Comput. Intell. Mach. Learn. ESANN 2015 - Proc.*, 89–94.
- Martin, H., Izquierdo, W., Cabrerizo, M., Cabrera, A., and Adjouadi, M. (2021). Near real-time single-beat myocardial infarction detection from single-lead electrocardiogram using Long Short-Term Memory Neural Network. *Biomed. Signal Process. Control* 68, 102683. doi:10.1016/j.bspc.2021.102683
- Mathunjwa, B. M., Lin, Y. T., Lin, C. H., Abbod, M. F., and Shieh, J. S. (2021). ECG arrhythmia classification by using a recurrence plot and convolutional neural network. *Biomed. Signal Process. Control* 64, 102262. doi:10.1016/j.bspc.2020.102262
- Mo, J. (2016). *Detecting anomalous data using auto-encoders*. doi:10.18178/ijmlc.2016.6.1.565
- Mogren, O. (2016). *C-RNN-GAN: Continuous recurrent neural networks with adversarial training*. Available at: <http://arxiv.org/abs/1611.09904>.
- Onan, A. (2020). Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks. *Concurr. Comput. Pract. Exper.* 33, 1–12. doi:10.1002/cpe.5909
- Onan, A., and Tocoglu, M. A. (2021). A term weighted neural language model and stacked bidirectional LSTM based framework for sarcasm identification. *IEEE Access* 9, 7701–7722. doi:10.1109/ACCESS.2021.3049734
- Özbay, Y., Ceylan, R., and Karlik, B. (2011). Integration of type-2 fuzzy clustering and wavelet transform in a neural network based ECG classifier. *Expert Syst. Appl.* 38, 1004–1010. doi:10.1016/j.eswa.2010.07.118
- Pang, G., Shen, C., Cao, L., and Van Den Hengel, A. (2021). Deep learning for anomaly detection: A review. *ACM Comput. Surv.* 54, 1–38. doi:10.1145/3439950
- Risk, M. R., Sobh, J. F., and Saul, J. P. (1997). Beat detection and classification of ECG using self organizing maps. *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.* 1, 89–91. doi:10.1109/iembs.1997.754471
- Ruff, L., Vandermeulen, R. A., Binder, A., Emmanuel, M., and Kloft, M. (2018). *Deep one-class classification deep one-class classification*.
- Sacco, R. L., Roth, G. A., Reddy, K. S., Arnett, D. K., Bonita, R., Gaziano, T. A., et al. (2016). The heart of 25 by 25: Achieving the goal of reducing global and regional premature deaths from cardiovascular diseases and stroke: A modeling study from the American heart association and world heart federation. *Glob. Heart* 11, 251–264. doi:10.1016/j.gheart.2016.04.002
- Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., and Langs, G. (2017). Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. *Lect. Notes Comput. Sci. Incl. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma.* 10265 LNCS, 146–157. doi:10.1007/978-3-319-59050-9\_12
- Schölkopf, B., Williamson, R., Smola, A., Shawe-Taylor, J., and Platt, J. (2000). Support vector method for novelty detection. *Adv. Neural Inf. Process. Syst.*, 582–588.
- Shibuya, H., and Maeda, S. (2016). Anomaly detection method based on fast local subspace classifier. *Electron. Comm. Jpn.* 99, 32–41. doi:10.1002/ecj.11770
- Song, Q., Zong, B., Wu, Y., Tang, L. A., Zhang, H., Jiang, G., et al. (2018). TGNet: Learning to rank nodes in temporal graphs. *Int. Conf. Inf. Knowl. Manag. Proc.*, 97–106. doi:10.1145/3269206.3271698
- Srinivasan, N., Ge, D. F., and Krishnan, S. M. (2002). Autoregressive modeling and classification of cardiac arrhythmias. *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.* 2, 1405–1406. doi:10.1109/IEMBS.2002.1106452

Sun, L., Jia, K., Yeung, D. Y., and Shi, B. E. (2015). "Human action recognition using factorized spatio-temporal convolutional networks," in Proc. IEEE Int. Conf. Comput. Vis. 2015 Inter, 4597–4605. doi:10.1109/ICCV.2015.522

Thill, M., Konen, W., Wang, H., and Back, T. (2021). Temporal convolutional autoencoder for unsupervised anomaly detection in time series. *Appl. Soft Comput.* 2021 (3), 107751. doi:10.1016/j.asoc.2021.107751

Wang, K., Zhao, Y., Xiong, Q., Fan, M., Sun, G., Ma, L., et al. (2016). Research on healthy anomaly detection model based on deep learning from multiple time-series physiological signals. *Sci. Program.* 2016, 1–9. doi:10.1155/2016/5642856

Wang, X., Du, Y., Lin, S., Cui, P., Shen, Y., and Yang, Y. (2020). adVAE: A self-adversarial variational autoencoder with Gaussian anomaly prior knowledge for anomaly detection. *Knowl. Based. Syst.* 190, 105187. doi:10.1016/j.knsys.2019.105187

Xia, Y., Cao, X., Wen, F., Hua, G., and Sun, J. (2015). "Learning discriminative reconstructions for unsupervised outlier removal," in Proc. IEEE Int. Conf. Comput. Vis. 2015 Inter, 1511–1519. doi:10.1109/ICCV.2015.177

Xia, Y., Wulan, N., Wang, K., and Zhang, H. (2018). Detecting atrial fibrillation by deep convolutional neural networks. *Comput. Biol. Med.* 93, 84–92. doi:10.1016/j.compbiomed.2017.12.007

Zhao, W., Gao, Y., Ji, T., Wan, X., Ye, F., and Bai, G. (2019). Deep temporal convolutional networks for short-term traffic flow forecasting. *IEEE Access* 7, 114496–114507. doi:10.1109/ACCESS.2019.2935504

Zhou, C., and Paffenroth, R. C. (2017). Anomaly detection with robust deep autoencoders. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. Part F1296*, 665–674. doi:10.1145/3097983.3098052