Research article

Convolutional transformer-driven robust electrocardiogram signal denoising framework with adaptive parametric ReLU

Jing Wang[1,†], Shicheng Pei[2,†], Yihang Yang[1] and Huan Wang[2,*]

[1] School of Computer Science, Xi'an Polytechnic University, Xi'an 710021, China
[2] Glasgow College, University of Electronic Science and Technology of China, Chengdu 611731, China

* Correspondence: Email: wh.huanwang@gmail.com.

† The authors contributed equally to this work.

Abstract: The electrocardiogram (ECG) is a widely used diagnostic tool for cardiovascular diseases. However, ECG recording is often subject to various noises, which can limit its clinical evaluation. To address this issue, we propose a novel Transformer-based convolutional neural network framework with adaptively parametric ReLU (APtrans-CNN) for ECG signal denoising. The proposed APtrans-CNN architecture combines the strengths of transformers in global feature learning and CNNs in local feature learning to address the inadequacy of learning with long sequence time-series features. By fully exploiting the global features of ECG signals, our framework can effectively extract critical information that is necessary for signal denoising. We also introduce an adaptively parametric ReLU that can assign a value to the negative information contained in the ECG signal, thereby overcoming the limitation of ReLU to retain negative information. Additionally, we introduce a dynamic feature aggregation module that enables automatic learning and retention of valuable features while discarding useless noise information. Results obtained from two datasets demonstrate that our proposed APtrans-CNN can accurately extract pure ECG signals from noisy datasets and is adaptable to various applications. Specifically, when the input consists of ECG signals with a signal-to-noise ratio (SNR) of -4 dB, APtrans-CNN successfully increases the SNR to more than 6 dB, resulting in the diagnostic model's accuracy exceeding 96%.

Keywords: ECG signal; signal denoising; convolutional neural network; transformer

1. Introduction

The electrocardiogram (ECG) signal is a crucial health indicator that enables physicians to diagnose cardiovascular diseases and perform biometric identification [1]. However, the accuracy of ECG recordings can be severely compromised by baseline wander, power line interference, and physiological artifact during the recording process [2]. Furthermore, in telemedicine applications involving the transmission and storage of ECGs, poor channel conditions can introduce additional noise [3]. Noise can obscure critical clinical features in ECG signals, rendering their identification and diagnosis challenging [4]. As a result, obtaining pure ECG signals has become an important task in ECG signal processing.

In ECG signal processing, traditional denoising methods mainly adopt filtering or decomposition techniques to separate the signal and noise. Among them, empirical mode decomposition (EMD) [5,6], adaptive filters [7], and the wavelet transform [8,9] have been widely studied and applied for a long time. For instance, Weng et al. [10] proposed a noise reduction method based on EMD and validated its effectiveness on the MIT-BIH Arrhythmia Database. Chandrakar et al. [11] proposed an adaptive filter using a recursive least squares algorithm, which outperformed traditional LMS-based adaptive filters. Reddy et al. [12] proposed an improved threshold denoising method, which combined hard and soft thresholding to achieve superior denoising performance. However, these traditional methods have some limitations in ECG signal denoising, such as being effective only under specific noise conditions [13]. For instance, EMD could suffer from modal aliasing when the sampling frequency of ECG signals is too low. Moreover, these methods require frequent parameter adjustment to achieve satisfactory denoising performance across different ECG signals due to the inconsistent attenuation features of the high-frequency and low-frequency components in the signal [14].

According to the above analysis, the existing ECG noise reduction methods still need improvement. Deep learning possesses a vast array of applications and showcases superior robustness [15,16]. Hong et al. [17] reviewed existing deep learning methods and proved that deep learning methods generally perform better than traditional methods by testing them on multiple datasets. Deep learning is highly dependent on data, meaning that larger amounts of data make the network perform better. The widespread use of telemedicine and wearable ECG devices provides a large database for ECG studies. This provides a good precondition for applying deep learning to the field of ECG. Antczak [18] presented a novel deep recurrent neural network for ECG noise reduction. The architecture is tested on a real dataset. Experimental results show that the proposed method is better than traditional methods, such as the band pass filter and wavelet filter in signal-to-noise ratio (SNR) from 0 to -10 dB. In [13], Arsene et al. proposed two deep learning models and a traditional wavelet-based noise reduction method. The results on different datasets prove the superiority of the CNN. In addition, autoencoders are a classic and effective architecture for signal denoising [19]. For example, Chiang et al. [20] proposed a denoising auto-encoder (DAE) based on full convolution, which can effectively reduce the noise of ECG signals. Singh et al. [21] proposed a novel attention-based convolutional denoising autoencoder (ACDAE) model that utilizes skip links and attention modules to reliably reconstruct ECG signals from extreme noise conditions. Samann et al. [22] proposed a running DAE (RunDAE) for denoising short ECG segments without relying on the R-peak detection algorithm for alignment. Xiong et al. [23] designed a DAE architecture that combines the wavelet transform (WT), which improved the performance of DAE on removing baseline drift. Fotiadou et al. [24] established a convolutional encoder-decoder network structure with skip connections, which achieved remarkable results in fetal cardiac noise reduction.

Although the CNN has strong feature extraction capability, there are still some problems when it is applied to ECG signal noise reduction.

1) As the CNN is developed based on image recognition, it rarely cares about negative value information in data. The ECG signal is different from the image data. The ECG data fluctuates based on zero values. Therefore, it contains much valuable negative information. The commonly used ReLU function is unsuitable for ECG signal processing.

2) The CNN developed based on image recognition, focuses on learning local neighborhood information. The ECG signal is a long-term sequence signal, and learning global features is essential for ECG analysis. Limited by the structure of the CNN itself, it is difficult for CNNs to perform well in such tasks.

3) The primary objective of ECG denoising is to restore the original ECG signal while preserving its valuable feature information. However, in the process of optimizing the denoising network, it is possible to inadvertently minimize the distance between the denoised signal and the original signal. This may cause the CNN-based region network to overlook some weak but important features that could reflect disease information. For example, irregular R-R intervals, missing or abnormal P-Waves, variation in QRS complex width, ST segment and T-Wave abnormalities, and atrioventricular block all reflect specific related disease information. Consequently, valuable information may be lost, even if the denoising indicators appear to be excellent.

For encoding long-term global information, Transformer has demonstrated excellent performance and has been successfully applied in ECG analysis. For example, Hu et al. [25] proposed a novel transformer-based deep learning neural network that detects arrhythmia on single-lead ECG segments. Meng et al. [26] proposed a novel lightweight fussing transformer, which can achieve dynamic ECG heartbeat classification using fewer parameters. Xia et al. [27] proposed a generative adversarial network based on Transformer and convolution to solve the data imbalance problem. Xia et al. [28] proposed a novel framework based on a lightweight Transformer and combining CNN and DAE, which uses a DAE to extract local features from a single heartbeat and adopts a lightweight Transformer to focus on global features. In addition, the Transformer also plays an important role in ECG signal-denoising applications. For example, Yin et al. [29] proposed a GAN-guided parallel CNN and Transformer network, significantly outperforming existing networks in various artifact removal tasks. Pu et al. [30] proposed a new transformer-based EEG signal denoising network, which can effectively remove eye artifacts and muscle artifacts.

Based on the problems CNN faces in ECG noise reduction, this paper aims to propose a deep learning-based ECG signal noise reduction algorithm to solve them. Therefore, this paper constructs a novel Transformer-based convolutional neural network framework with adaptively parametric ReLU (APtrans-CNN). APtrans-CNN can accurately learn local detail features and important global context features from the noise signal while retaining the disease information contained in the ECG signal while reducing noise. To be specific, the Transformer is combined with CNN, which preserves the CNN's excellent local detail perception and greatly improves the ability to capture global features of ECG signals. This improves the inadequacy of CNN long-sequence time-series feature learning. In addition, we introduced the activation function AP-ReLU, especially for one-dimensional ECG signal processing. It solves the problem that the traditional activation function cannot preserve the negative features of ECG signals. Inspired by [31,32], a dynamic feature aggregation module (DAM) was proposed in this paper. The module enhances the network's ability to capture valuable features, which allows the network to complete the noise reduction task while retaining valuable disease information.

The contributions of this paper are summarized as follows:

1) The proposed method introduces AP-ReLU to replace the general ReLU and was embedded into the network. AP-ReLU realizes the full utilization of ECG information by assigning weight values to negative information.

2) The proposed method inherits the advantages of CNN and Transformer through the organic combination of the CNN and Transformer, thus having efficient local and global feature perception ability.

3) This paper proposes a novel feature extraction architecture APtrans-CNN. By combining the CNN and Transformer, the local-global feature fusion extraction is realized. The framework can be applied to ECG signal noise reduction for people with different health conditions due to its excellent disease information retention ability.

4) The proposed method is verified on two real datasets, and the results show that it has excellent performance under different noises and has good application potential.

The paper is organized as follows: In Section 2, the proposed ECG denoising method is introduced in detail. In Section 3, the effectiveness and superiority of APtrans-CNN are verified. Section 4 summarizes this paper.

2. Methodology

2.1. Problem definition

The general form of the ECG signal is defined as follows:

$$s = p + n$$

<div align="right">(1)</div>

where s represents the observed ECG signal, and $p = \{p_z\}_{z=1}^{\mathcal{Z}}$ represents the original ECG signal without noise. $n = \{n_z\}_{z=1}^{\mathcal{Z}}$ represents the noise generated during signal acquisition and storage. $\mathcal{Z}$ represents the length of each ECG signal. The main task of denoising is to extract pure ECG signals from noise-polluted signals. For the training of the proposed method, we set $M = \{(s_i, p_i)\}_{i=1}^{N}$ as the training set, where N is the number of training samples. $S = \{s_i\}_{i=1}^{N} \in \mathbb{R}^{N \times \mathcal{Z}}$ represents the ECG signal matrix used for denoising and $P = \{p_i\}_{i=1}^{N} \in \mathbb{R}^{N \times \mathcal{Z}}$ represents the pure ECG signal matrix we want to get. We take the noise-doped ECG signal as the input signal, and the original ECG signal as the target for training. The main training method of the network is to optimize the parameters of the network by minimizing the noise reduction error, which is defined as

$$\underset{\theta}{arg\min} \sum_{i=1}^{N} L(p_i, \hat{p}_i)$$

<div align="right">(2)</div>

where L represents the loss function, and $\hat{p}_i$ represents the ECG signal predicted by the network. We selected the mean square error (MSE) as the loss function. This formula can be expressed as:

$$L(p_i, \hat{p}_i) = \|p_i - \hat{p}_i\|_2^2$$

<div align="right">(3)</div>

2.2. Transformer-based ECG signal denoising framework

Figure 1 illustrates the overall architecture of the proposed model, which comprises two stages: signal compression and signal reconstruction. In the compression stage, convolutional layers are

employed to capture the essential details of ECG signals. Additionally, a Transformer module is incorporated at the bottom of the encoding layer to extract valuable global feature information. In the reconstruction stage, deconvolution and feature aggregation are performed on the feature image to recover the information of the ECG signal. Moreover, the skip connection mechanism is integrated into the network to facilitate the restoration of missing details and obtain a pure ECG signal.
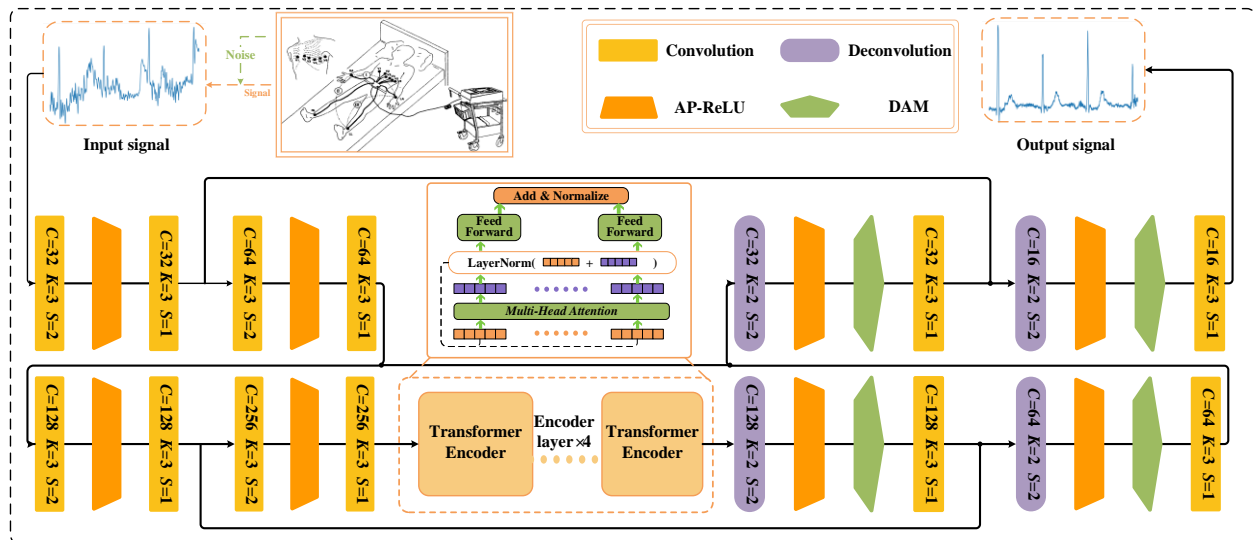

Figure 1. The network architecture of APtrans-CNN. *C* represents that the number of channels, *K* represents the convolution kernel, and *S* represents the step size.

In the signal compression stage, a typical convolutional network architecture is employed along with the AP-ReLU [33] to replace the modified linear unit, thereby endowing the model with more flexible linear transformation capabilities. The batch normalization layer is also included to expedite the training process of the network. Furthermore, the Transformer [34] module is added at the end of the encoding layer to compensate for the convolutional network's limitations in global signal feature extraction and extract the essential features of ECG signals more effectively. In the signal reconstruction stage, we chose the transpose convolution over upsampling since it can carry more information. Additionally, a DAM module is added after each transpose convolution layer to enable the network to learn crucial features of ECG signals and mitigate the noise's impact on feature extraction. Furthermore, because some edge features of ECG signals may be lost during signal compression due to sampling, we utilize skip connections to extract the features of each convolution layer and connect them with the corresponding features of the deconvolution layers. This enables the retrieval of some edge features of ECG signals.

2.3. Adaptively parametric ReLU

In conventional CNNs, ReLU is often used as the classical activation function. However, due to the definition of the function, when the input information is negative, the function will be in an inactive state. As a result, only half of the information contained in the ECG signal can be utilized by the network. Although some improved activation functions, such as PReLU and Leaky ReLU, have been

proposed in subsequent studies to give a small output to the negative part, they do not consider the essential property that both the negative and positive parts of the ECG signal are equally important. Furthermore, since each individual has different physical conditions and external interference during signal acquisition, their ECG signals may vary. The ReLU function is the same for all input transformations, which is not conducive to improving the universality of the ECG noise reduction model.
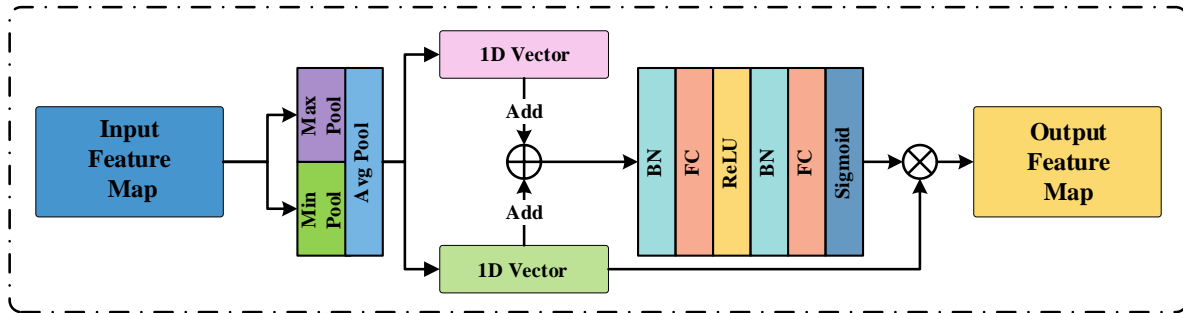

Figure 2. The network architecture of adaptively parametric ReLU (AP-ReLU).

To solve this problem, inspired by [33], we introduced AP-ReLU as the activation function of this model. As shown in Figure 2, the input feature $F_{in} \in \mathbb{R}^{C \times L}$ is first divided into positive and negative parts $F^{+} \in \mathbb{R}^{C \times L}$ and $F^{-} \in \mathbb{R}^{C \times L}$. To improve the stability of the method, the positive and negative feature maps go through a global average pooling layer respectively, and the dimension of feature maps is changed into $C \times 1$.

$$F_v^{+} = Avg\left(F_i^{+}\right) \in \mathbb{R}^{C \times 1} \tag{4}$$

$$F_v^{-} = Avg\left(F_i^{-}\right) \in \mathbb{R}^{C \times 1} \tag{5}$$

After that, the two new feature maps are combined according to the time dimension to get a new feature $F_0 \in \mathbb{R}^{C \times 2}$, which is then input into a nonlinear mapping module. The feature map $F_0 \in \mathbb{R}^{C \times 2}$ first passes through a BN layer and then maps F_0 to $F_1 \in \mathbb{R}^{C \times 1}$ through a fully connected and ReLU layer. Then, the new features pass through a BN layer, a fully connected layer, and sigmoid layer again to get $F_2 \in \mathbb{R}^{C \times 1}$. Finally, the dimension of F_2 is extended to obtain the feature weights $\alpha \in \mathbb{R}^{C \times L}$ of negative signals.

Finally, the feature weight α is assigned to the negative feature F^{-}, which is then added to the positive feature F^{+} to obtain the final output feature *y*. The parameter transfer of the whole AP-ReLU module is shown as follows, where *x* and *y* represent input signals and output signals respectively. α represents the feature weight obtained by the FCN module.

$$F_{out} = max\left(x, 0\right) + \alpha \times min\left(x, 0\right) \tag{6}$$

The AP-ReLU module facilitates the involvement of negative features in ECG signals in network learning. The module's global feature weight varies for each input, producing distinct output sets, which enhances the network's nonlinear transformation capability. This feature also results in significant improvements in the network's denoising ability.

2.4.   Dynamic feature aggregation module (DAM)

Inspired by the SENet [31], this paper designed the dynamic feature aggregation module (DAM) to overcome the limitations of the CNN. The architecture of the DAM is shown in Figure 3. It consists of two continuously connected sub-modules: the channel feature aggregation module (CFAM) and the time feature aggregation module (TFAM). Since convolution operations extract information features by both channel and time information, DAM guides the network on what and where to pay attention by processing channel and time dimension information.
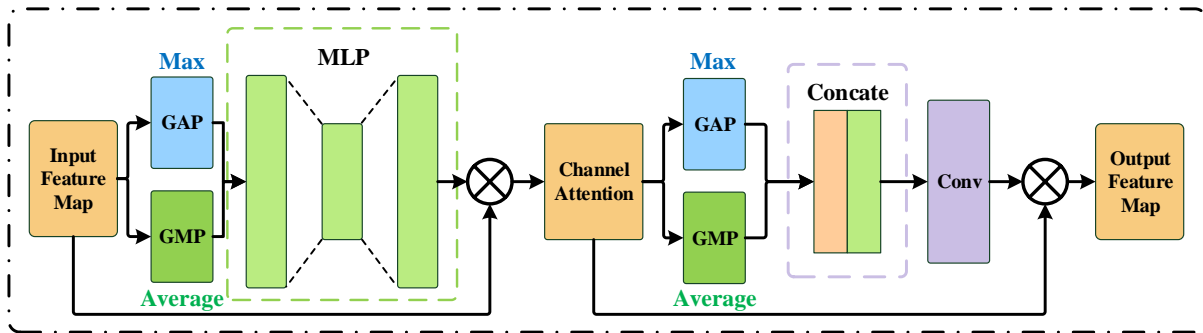

Figure 3. The architecture of dynamic feature aggregation module (DAM).

In CFAM, it compresses the time dimension of the signal to make the network pay more attention to the channel information. The input features $F_{in} \in \mathbb{R}^{C \times L}$ pass through the global average pooling (GAP) and global max pooling (GMP) to get two size of $C \times 1$ feature vectors. GAP generates feature vectors by calculating the average value of each channel feature, which is spatially invariant. GMP generates feature vectors by calculating the maximum value of each channel feature, which can retain the most significant features in the feature map. A weighted feature map of channel attention is obtained by adding the two feature maps through a linear network and applying a set of nonlinear transformations based on the sigmoid activation function. Then, the output $F_1 \in \mathbb{R}^{C \times L}$ of CFAM is obtained by the following calculation:

$$F_1 = \boldsymbol{\delta}\left(W_0\left(W_1\left(F_{in}^{avg}\right)\right) + W_0\left(W_1\left(F_{in}^{\max}\right)\right)\right) \otimes F_{in} \tag{7}$$

$\boldsymbol{\delta}$ represents the sigmoid activation function, and $\otimes$ means the element-wise multiplication is used to calculate the number between two matrices. F_{in}^{avg} and $F_{in}^{\max}$ represent the features after average pooling and max-pooling in the time dimension. W_0 and W_1 represent the weight dimensions in the MLP, which are shared by the two inputs.

To make the network better aggregate the important features in the time dimension, the F_1 is average pooled and maximum pooled based on the channel dimension, which generates two feature vectors with a size of $1 \times L$. Then, they are connected according to the channel dimension to obtain $F_2 \in \mathbb{R}^{2 \times L}$, which is propagated through a convolution layer with a kernel size of 7. Finally, the weight coefficient is generated through a sigmoid function. Then, the output $F_{out} \in \mathbb{R}^{C \times L}$ of TFAM is obtained by the following calculation:

$$F_{out} = \boldsymbol{\delta}\left(k^7\left[F_1^{avg}; F_1^{\max}\right]\right) \otimes F_1 \tag{8}$$

k^7 means that the convolutional kernel size is 7.

2.5. Transformer for global feature encoding

CNNs have achieved remarkable success in various fields. However, their convolution operations can only capture local information near the convolution kernel, whereas global features often contain critical information for ECG signals. On the other hand, Transformers, a model originally designed for natural language processing tasks, have been extensively employed in computer vision tasks and have shown significant performance improvements. Thanks to its self-attention mechanism, Transformers can model long-range correlated features effectively. Thus, we propose to integrate CNNs and Transformers in our network to leverage the complementary strengths of these models.

As shown in Figure 4, in this module, features extracted by the CNN are used as inputs, and more abundant output features are obtained through a multi-head attentional mechanism module. Skip connection operation is performed on the input and output signals to recover the details of lost ECG signals, and normalization is carried out to speed up network operation. The multi-attentional mechanism comprises multiple parallel scaled dot product attentional mechanisms.
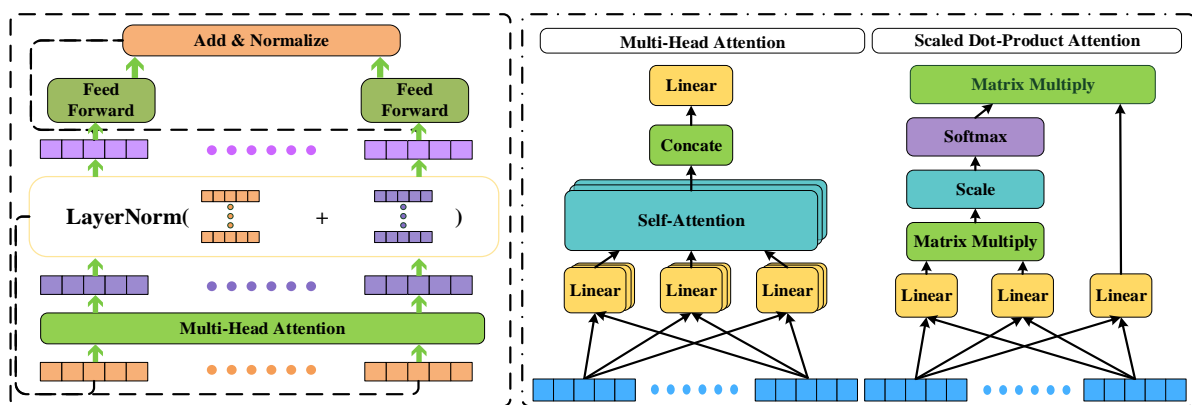

Figure 4. The network architecture of a single transformer encoder layer.

2.5.1. Scaled dot product attention (SDPA)

In the SDPA module, the calculation of the attention can be divided into three steps. First, three learnable weight matrices are defined by three different linear transformation layers. Then, the input features S_{in}^{n} are mapped to Query (Q), Key (K), and Value (V), respectively, using the three weight matrices, which are defined as follows:

$$Q = \text{Linear}_q(F_I^n) = F_I^n W^Q \tag{9}$$

$$K = \text{Linear}_k(F_I^n) = F_I^n W^K \tag{10}$$

$$V = \text{Linear}_v(F_I^n) = F_I^n W^V \tag{11}$$

After that, the similarity between Query and Key is calculated by the dot product of Q and K^T. To improve numerical stability, the similarity first needs to undergo a scale transformation, that is, divide

by the square root of the scale factor d_k. Then, the similarity is input into the softmax function ϕ for normalization. Finally, these values are multiplied by the output weight to get the final Attention.

$$Attention\,(\mathrm{Q,K,V}) = \phi\,(\frac{\mathrm{QK^T}}{\sqrt{\mathrm{d_k}}})V \tag{12}$$

2.5.2. Multi-head attention (MHA)

In MHA, it is proposed to define multiple groups of Q, K, and V, and put them into the self-attention mechanism module respectively, hoping that they can obtain ECG information from the subspace of different aspects, thus extracting richer and more comprehensive ECG features than a single SDPA module. As shown in Figure 4, in this part of the module, we generated different output matrices $head_i$ by using eight parallel SDPA, and these SDPA modules are independent of each other and do not share their parameters. The process can be expressed as follows:

$$\begin{aligned}\mathrm{Multihead(Q,K,V)} = \\ \mathrm{Concat}\left[\phi\,(\frac{\mathrm{Q_1 K_1^T}}{\sqrt{\mathrm{d_{k_1}}}})V_1;\phi\,(\frac{\mathrm{Q_2 K_2^T}}{\sqrt{\mathrm{d_{k_2}}}})V_2;...\phi\,(\frac{\mathrm{Q_i K_i^T}}{\sqrt{\mathrm{d_{k_i}}}})V_i\right]\mathrm{W^O}\end{aligned} \tag{13}$$

where i represents the number of heads in MHA, and W^O are learnable parameter mapping matrices used for the aggregation of ECG signal features.

2.5.3. Feedforward network (FFN)

After the MHA module, the skip connections are added to reduce the degradation of the network. Then the obtained features are normalized to accelerate the training speed of the network.

$$F_A{}^n = LayerNorm\,(F_I{}^n + F_O{}^n) \tag{14}$$

$$LayerNorm\,(F) = \frac{F_{ij} - \mu(F_j)}{\sigma^2(F_j) + \varepsilon} \tag{15}$$

where μ represents mean calculation, and σ^2 represents variance calculation. And ε is added to prevent the denominator from being zero. Then, the FFN is used to enhance the feature extraction capability of the Transformer. The FFN consists of two linear connection layers with weights W_0, W_1 and a ReLU activation function. After skip connection and layer normalization, the encoder will enter the next sublayer. This can be expressed as follows:

$$F_W{}^n = LayerNorm\,(F_A{}^n + W_0(\max(0, (W_1(F_A{}^n))))) \tag{16}$$

3. Experimental verification

3.1. Experimental setup

To improve the reliability of the experiment, real ECG signals from the MIT-BIH arrhythmia database [35] and MIT-BIH atrial fibrillation database [36] were used in the experiment [37]. The

arrhythmia dataset included 48 dual-channel electrocardiogram recordings from 47 subjects at a sampling frequency of 360 Hz, each recording lasting 30 minutes. The atrial fibrillation dataset consisted of 23 dual-channel ambulatory electrocardiogram recordings from 25 subjects sampled at 250 Hz, each recording lasting 10 hours.

For these two datasets, we add noise manually to simulate signal doping in the process of signal acquisition. To make the results more convincing, we selected three kinds of noise that are most likely to occur in ECG signal extraction, including baseline drift, muscle artifacts , and motion artifact noise. All the source of the noise is randomly selected from the MIT-BIH Noise Stress Test Database (NSTDB) [38]. Based on the length of the ECG cycle, we segmented every 250 data points, and the duration of each ECG signal was approximately one ECG cycle. To simulate the randomness of noise generation in the real world, we randomly select three kinds of noise from the NSTDB dataset. The signal intensity of the noise was changed to meet the certain SNR_{dB}, and was added to the ECG signal fragments randomly selected from the MIT-BIH arrhythmia database and atrial fibrillation database. Then, the mixed noisy ECG signal is used as the input signal. The original pure ECG signal as the target signal. Then, all the data are normalized so that the amplitude of the data can be kept in the range of (0, 1). In order to have a quantitative standard for the added noise signal intensity, we define SNR_{dB} via the following formula:

$$SNR_{dB} = 10 \log_{10} \left(P_{signal} / P_{noise} \right) \tag{17}$$

P_{signal} represents the power of the signal after adding noise, and P_{noise} represents the power of the noise signal.

The Transformer model requires a large amount of data for training, so we use the sliding segmentation method to obtain more training samples. In addition, we use 80% of the samples in the two data sets as the training set, 10% of the samples as the validation set, and 10% of the samples as the test set. In the arrhythmia data set, 81,882 ECG segments and 9094 ECG segments were randomly selected for training and validation, and 9602 ECG segments were selected for the test. For the atrial fibrillation data set, 224,730 ECG segments and 24,971 ECG segments were randomly selected for training and validation, and 26,299 ECG segments were selected for the test. The training set, validation set, and test set shared the same input of -4, -2, 0 and 4 dB, respectively. The proposed model is implemented by Pytorch and Python. During the model training process, the Adam optimization algorithm was adopted, the learning rate was set to 0.0001, and the batch size was set to 64. Each experiment was repeated three times, and the average value was used as the final result to avoid accidental experimental results.

3.2. Evaluation metrics

In this study, the signal-to-noise ratio (SNR) and the MSE were used to evaluate the network denoising performance. SNR_{out} represents the SNR of the output predicted signal. When the input SNR_{dB} is the same, the larger the SNR_{out} is, the better the denoising performance is. SNR_{out} can be expressed as:

$$SNR_{out} = 10 \times \log_{10} \frac{\sum_{i=1}^{N} p_i^2}{\sum_{i=1}^{N} (\hat{s}_i - p_i)^2} \tag{18}$$

MSE is used to determine the variance between the predicted ECG signal and the pure ECG signal. The smaller the MSE value is, the difference between the reconstructed signal and the original pure signal is smaller. MSE can be expressed as:

$$MSE = \frac{\sum_{i=1}^{N}(\hat{s}_i - p_i)^2}{N} \tag{19}$$

where p_i represents the original pure ECG signal, $\hat{s}_i$ represents the reconstructed and enhanced ECG signal, and N represents the sample number of input ECG signals.

3.3. Denoising performance for arrhythmia database

3.3.1. Effectiveness of proposed method

To explore the effect of the three main modules that we proposed on the performance improvement of the traditional autoencoder noise reduction network, we select the autoencoder network as the basic network and add three modules, AP-ReLU, DAM, and Transformer, respectively, on the network and compare them with the original network and the proposed method APtrans-CNN. These comparison methods are CNN, AP-CNN, DAM-CNN, and Trans-CNN. CNN: The convolutional autoencoder architecture with skip connections consists of four encoding and decoder layers. AP-CNN: Replacing ReLU in the CNN encoding and decoding layers with the AP-ReLU. DAM-CNN: Adding DAMs in each decoding layer of CNN. Trans-CNN: The encoder part of the Transformer network is added at the bottom of the encoder module of CNN.

To prove the versatility of the network model for ECG signals under different noise conditions, four SNR_{dB} were selected for input ECG signals, which were -4, -2, 0, and 4 dB, respectively. The SNR and mean square error of the enhanced ECG signal were compared. The experimental results are shown in Table 1 and Figure 5.
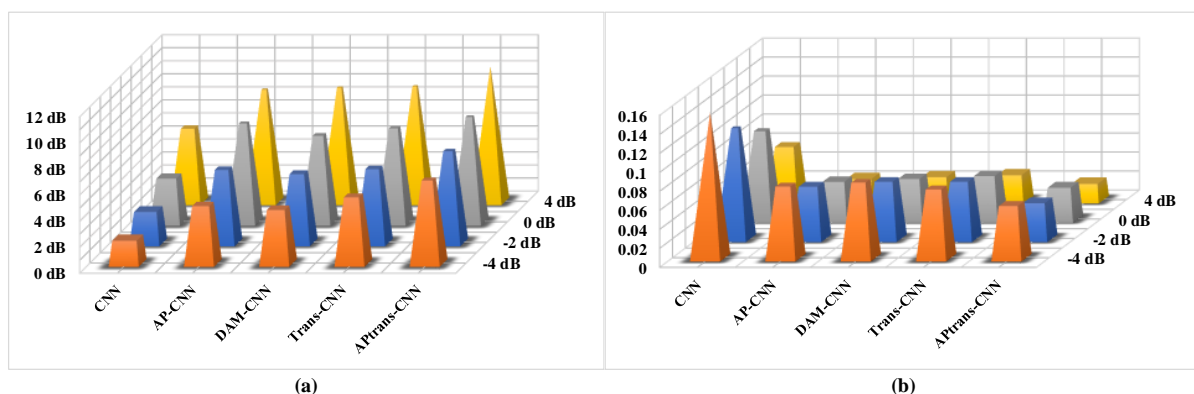

Figure 5. The comparison among denoising performance. (a) SNR, (b) MSE of different architectures under different noise environments.

Obviously, the proposed method, particularly the Trans-CNN architecture, has demonstrated remarkable performance in all four tested SNR$_{dB}$ conditions. Incorporating the AP-ReLU and DAM techniques into the CNN network resulted in significant improvements. For SNR$_{dB}$ = -4 dB, the SNR$_{Out}$

values of AP-CNN and DAM-CNN were 4.53 and 4.25 dB, respectively. The use of AP-ReLU facilitated the inclusion of valuable negative information in the ECG signal. The DAM-CNN network bolstered the feature extraction capability of the network, allowing for the automatic extraction of valuable feature information and the elimination of useless noise. Notably, Trans-CNN exhibited the best performance in high-noise environments, with a SNR_{Out} of 5.19 dB at SNR_{dB} = -4 dB, which can be attributed to the Transformer's self-attention mechanism. Networks that integrate this module can learn more valuable global features that are typically ignored during convolution operations. From another evaluation index, MSE, the performance of the three proposed architectures are very similar, and each has its advantages under different noise conditions. Through tests in different noise environments, we also obtained that the proposed APtrans-CNN can combine the advantages of each module. It can maintain an SNR_{Out} of more than 6 dB in a noise environment from -4 to 4 dB, and can adapt to different noise environments.

Table 1. Experiment results under four noise environments (MIT-BIH arrhythmia database).

Method	SNR_{Out}				MSE			
Noise	-4 dB	-2 dB	0 dB	4 dB	-4 dB	-2 dB	0 dB	4 dB
CNN	1.95 dB	2.56 dB	3.52 dB	5.70 dB	0.153	0.117	0.094	0.058
AP-CNN	4.53 dB	5.70 dB	7.63 dB	8.62 dB	0.077	0.057	0.042	0.025
DAM-CNN	4.25 dB	5.39 dB	6.71 dB	8.78 dB	0.081	0.062	0.045	0.027
Trans-CNN	5.19 dB	5.76 dB	7.26 dB	8.87 dB	0.074	0.062	0.048	0.029
APtrans-	**6.45 dB**	**7.09 dB**	**8.12 dB**	**10.38 dB**	**0.057**	**0.040**	**0.036**	**0.020**

3.3.2. Visualization of denoising results

To further explore the noise reduction effect of each module, we visualized the predicted signal under different noise conditions. The results in Figure 6 showed that by adding AP-ReLU, DAM, and combining the CNN with Transformer, the data quality after noise reduction was significantly improved.

The data processed by the CNN network has restored some data from the original ECG signal, but it can be seen that the denoised signal still has many high-frequency noise components. In addition, the signal's peak does not reach the peak of the original signal wave. Although AP-CNN has good performance in terms of SNR and MSE, the visualized signal contains a lot of useless signals. However, the ECG data after the AP-CNN processing still preserved most of the information in the original ECG signal. The R-wave peaks also reached the original expectations. This is because AP-ReLU enables the network to learn more valuable feature information. With the addition of the DAM module, the network is more likely to eliminate those irrelevant high-frequency noise components and separate the original pure signal from the noise signal. After adding the DAM module, the network can automatically aggregate ECG features, eliminating a large part of the high-frequency noise. The combination of CNN and Transformer greatly improves the ECG signal's smoothness and restores each part's waveform very well. The combination of CNN and Transformer enables the network to improve the ability to extract global features and can restore the original waveform signal well. We also found that the proposed three methods have advantages, but also have some shortcomings. Our proposed method successfully combines the advantages of the above methods. It has excellent performance during the whole SNR_{dB} throughout the test, and the denoised signals successfully preserve most of the features of disease information.

In our study, we considered very severe and extreme experimental circumstances. For example, when the signal-to-noise ratio is 0 dB, the energy of the noise signal is the same as the energy of the original signal. This means that the ECG signal is seriously interfered by noise. Generally speaking, under good experimental conditions, the signal-to-noise ratio of ECG signals can reach more than 20 to 30 dB. Therefore, in actual scenarios, we basically will not encounter noise interference around 0 dB. In addition, it can be seen from Figure 6 that, when the signal-to-noise ratio is 4 dB, the denoised signal obtained by APtrans-CNN is consistent with the original signal. This shows that in actual scenarios, the proposed method can have excellent performance and will not cause medical personnel to ignore important information.
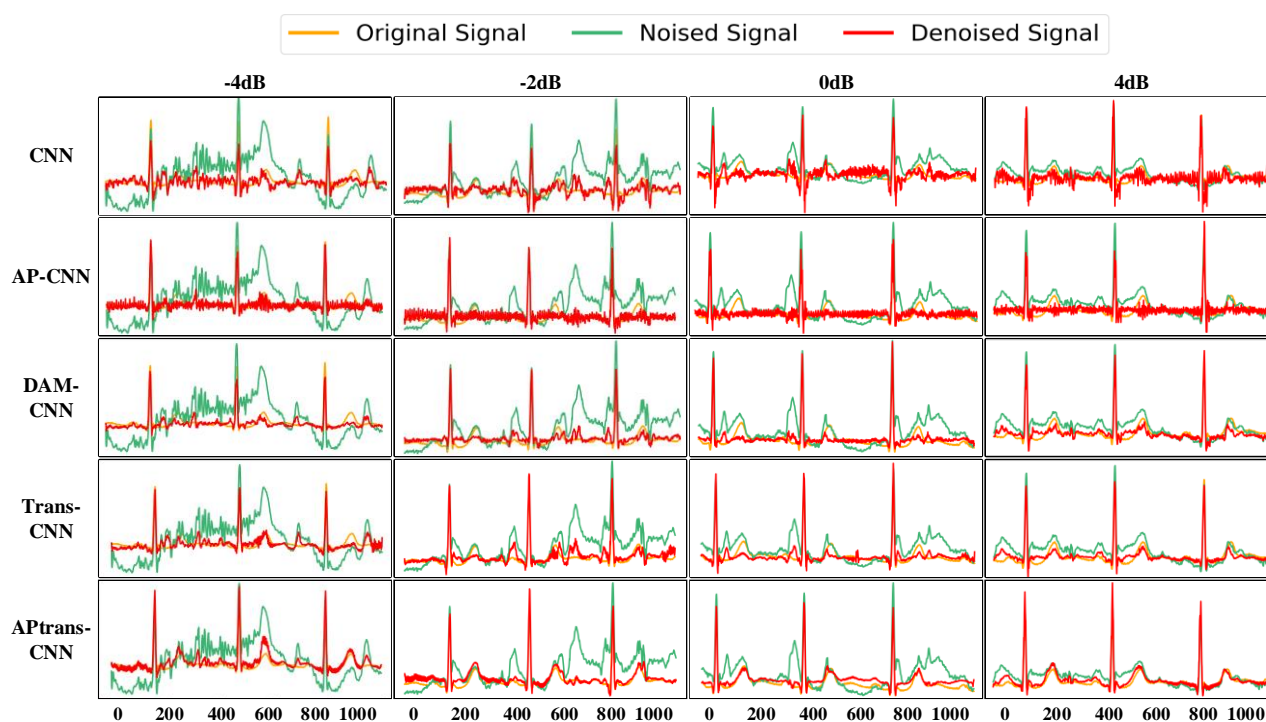

Figure 6. The visualization of the denoising results produced by CNN with the proposed modules.

3.3.3. Comparison with existing denoising methods

To validate the effectiveness of the APtrans-CNN model, we compared it with the discrete wavelet transform (DWT), and we also employed several existing deep learning models, including the LSTM network used by Arsene et al. [13], the FCN proposed by Chiang et al. [39], the DNN proposed by Sannino et al. [40], and the IMUnet proposed by Qiu et al. [41]. The experimental results are presented in Table 2. It is evident that some of the methods only exhibit satisfactory performance in certain noise environments with SNR_{dB} values of -4, -2, 0, and 4 dB. In contrast, the proposed APtrans-CNN consistently exhibits superior denoising performance compared to other benchmarked networks under all four noise conditions.

The proposed APtrans-CNN architecture exhibits significant improvements in SNR compared to other comparison networks. Specifically, when SNR_{dB} = -4 dB, the SNR_{Out} of APtrans-CNN is 6.45 dB, representing an increase of more than 10 dB compared to SNR_{dB}. Among the comparison networks,

the IMUnet has the highest SNR_{Out} of 5.76 dB, while the DNN and FCN only achieve 4.38 and 2.13 dB, respectively. Although LSTM shows improved results compared to the input when $SNR_{dB} = -4dB$, its denoising ability is greatly reduced as the noise level decreases. In contrast, the APtrans-CNN architecture utilizes the Transformer module to obtain more global feature information, making it better suited for processing extremely long feature sequences such as ECG signals. In terms of the evaluation metric MSE, the DWT has the highest MSE of 0.164, followed by FCN with 0.137, while APtrans-CNN has the lowest MSE of 0.057, demonstrating its ability to accurately extract valuable ECG information by retaining the ability to capture important detail features while improving the network's performance in handling long sequence time-series with the help of Transformer.

Table 2. Experiment results under four noise environments (MIT-BIH arrhythmia database).

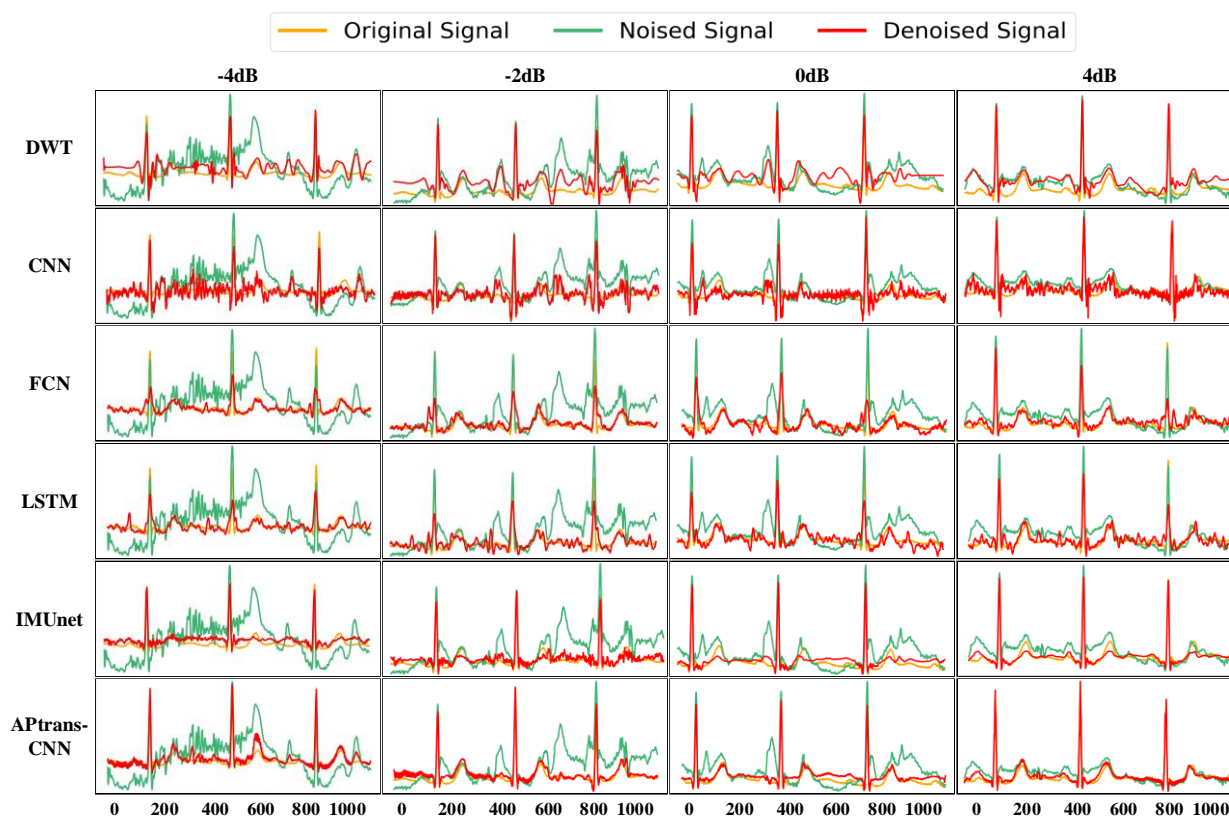| Method	SNR_{Out}				MSE			
Noise	-4 dB	-2 dB	0 dB	4 dB	-4 dB	-2 dB	0 dB	4 dB
DWT	2.25dB	2.51 dB	2.94 dB	4.27 dB	0.164	0.102	0.096	0.074
FCN	2.13 dB	2.43 dB	3.47 dB	5.94 dB	0.137	0.111	0.095	0.052
DNN	4.38 dB	5.19 dB	6.20 dB	6.89 dB	0.072	0.053	0.050	0.041
LSTM	4.46 dB	5.37 dB	5.88 dB	6.52 dB	0.088	0.078	0.052	0.046
IMUnet	5.76 dB	6.63 dB	7.82 dB	9.67 dB	0.059	0.046	0.037	0.032
APtrans-CNN	**6.45 dB**	**7.09 dB**	**8.12 dB**	**10.38 dB**	**0.057**	**0.040**	**0.036**	**0.020**

Figure 7. The visualization of the denoising results produced by APtrans-CNN and other methods: DWT, CNN, FCN, LSTM, IMUnet.

3.3.4. Visualization of denoising results

To further explore the noise reduction effect of each module, we visualized the predicted signal under different noise conditions, and the visualization results are shown in Figure 7. The smoothness of ECG signals processed by discrete wavelet transform is obviously improved. However, its actual performance is not satisfactory because the signal's baseline drift noise is difficult to solve. This problem was more pronounced in the atrial fibrillation dataset. Although the SNR of the signal was improved for FCN, it could be seen that the signal is still greatly affected by the noise and was mixed with many high-frequency noises. DNN performed better than FCN. Most high-frequency noises were removed from the visualized signal, and no baseline drift occurred. However, most of the peak information is not captured and displayed successfully. After noise reduction by LSTM, the noise of the ECG signal is greatly reduced, but the fluctuation of the ECG signal is severe. Moreover, very similar to DNN, the wave peak after processing is far lower than that of real ECG signals. It solves the problem of baseline drifts well. However, its elimination effect on other types of noise is not obvious. For IMUnet, both the smoothness degree and SNR of the ECG signal have been significantly improved. However, there are still some deficiencies in ECG signals after IMUnet processing. The peak of the QRS wave cannot reach the expected peak of the original ECG signals. The APtrans-CNN architecture has an astonishing performance in the field of noise reduction. It can be found that the signals processed by APtrans-CNN are very close to pure ECG signals in both datasets.

3.4. Denoising performance for atrial fibrillation database

3.4.1. Effectiveness of proposed method

The same network model used in the arrhythmia dataset was also applied in the atrial fibrillation dataset. The experimental results are shown in Table 3. It was found that the proposed method also performed well in the MIT-BIH atrial fibrillation database under four SNR_{dB} conditions. The three modules all provide great help to improve the noise reduction performance of the CNN. This shows that the proposed network is capable of noise reduction of different ECG signals. For example, when the SNR_{dB} is -4 dB, the SNR_{Out} of AP-CNN is 3.61 dB higher than that of CNN without AP-ReLU and the MSE of AP-CNN is 0.066 lower than that of the CNN. With the increase of SNR_{dB}, Trans-CNN showed higher performance improvement compared with the other two networks. For example, when the SNR_{dB} is -4dB, the MSE of Trans-CNN is 0.090, which is 0.017 worse than that of DAM-CNN. But, when the SNR_{dB} is 4 dB, Trans-CNN is 0.003 better than DAM-CNN. This is because Transformer can efficiently model relevant features of ECG.

Table 3. Experiment results under four noise environments (MIT-BIH atrial fibrillation database).

Method	SNR_{Out}				MSE			
Noise	-4 dB	-2 dB	0 dB	4 dB	-4 dB	-2 dB	0 dB	4 dB
CNN	2.66dB	3.40 dB	3.84 dB	6.24 dB	0.146	0.133	0.119	0.065
AP-CNN	6.27 dB	6.90 dB	7.53 dB	9.62 dB	0.080	0.059	0.049	0.031
DAM-CNN	5.83 dB	6.81 dB	7.21 dB	9.23 dB	0.073	0.066	0.058	0.036
Trans-CNN	5.29 dB	5.66 dB	6.35 dB	8.67 dB	0.090	0.077	0.067	0.033
APtrans-CNN	**7.48 dB**	**7.94 dB**	**8.49 dB**	**9.95 dB**	**0.049**	**0.046**	**0.035**	**0.027**

By combining the respective advantages of AP-CNN, DAM-CNN, and Trans-CNN, SNR_{Out} can be maintained above 7 dB in the noise environment of -4 to 4 dB. In addition, the tests on two datasets prove that APtrans-CNN does not require any hyperparameter tuning, and can be applied to different databases and achieve good performance with only retraining.
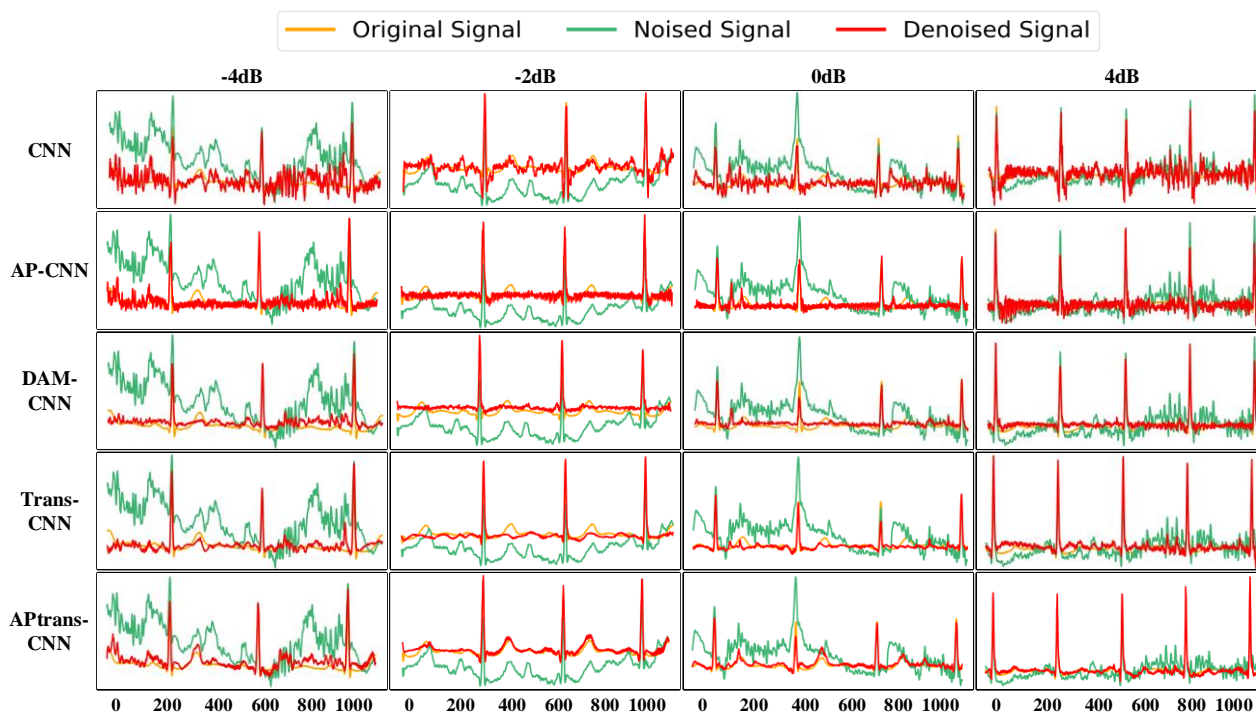

Figure 8. The visualization of the denoising results produced by CNN with the proposed modules.

3.4.2. Visualization of denoising results

This part explores the performance improvement brought by the three modules. Through the method of visualization of noise reduction results, the influence of AP-ReLU, DAM, and Transformer is more directly explained. As shown in Figure 8, the first row is the denoising performance of the original CNN. It is better at processing noise such as baseline drift, but the effect is not obvious on high-frequency noise such as EMG noise. Its actual performance is greatly affected by noise components and its performance is not stable. After introducing AP-ReLU, the waveform after noise reduction is closer to the pure signal, especially in the QRS wave part. However, the signal contains a lot of high-frequency noise that has not been eliminated. This is because AP-ReLU can help extract more feature information, but it lacks effective means to determine which information is valuable. DAM-CNN has successfully removed the interference of most baseline drift and time-frequency noise. The validity of DAM is proved, but some of the peaks did not meet expectations. Combining the CNN and Transformer in a certain order can fully combine the advantages of both. After Trans-CNN processing, the smoothness of the signal is greatly improved, which means that most of the high-frequency noise is eliminated. The trained network can effectively extract ECG signals from noise. For APtrans-CNN, it combines the advantages of the above networks. This allows the denoised signal to be very close to the original ECG signals. Although there are still some deficiencies, the key features

containing disease information are preserved. The smoothness of the processed signal has also been significantly improved, which will facilitate subsequent disease detection tasks.

3.4.3. Comparison with existing denoising methods

The experimental results are shown in Table 4. This study was based on the MIT-BIH atrial fibrillation dataset, SNR remains above 7 dB for SNR_{dB} = -4, -2, 0, and 4 dB. This indicates that the ECG signal processed by APtrans-CNN is very close to real ECG signal in most noisy environments. The proposed network structure has strong universality, which can complete the denoising task well for ECG signals under different noise conditions. For example, when SNR_{dB} = 4 dB, the SNR of IMUnet reached 9.89 dB, which is very close to our proposed method. But, at the high noise condition SNR_{dB} = -4 dB, the SNR of IMUnet is only 6.06 dB. APtrans-CNN has the best performance under the given experimental conditions. Similarly, for another evaluation index, MSE, APtrans-CNN also performs best under the four noise conditions. This proves that the designed architecture has excellent predictive ability and the output signal is very close to the real ECG signal. This is because the Transformer's self-attention mechanism enables the proposed architecture to better model the correlation features of ECG internal information with high feature aggregation and extraction capabilities under any noise conditions. In summary, compared with these comparison networks, the architecture proposed by us has excellent performance in the whole range of SNR_{dB}, which proves the excellence of APtrans-CNN.

Table 4. Experiment results under four noise environments (MIT-BIH atrial fibrillation database).

Method	SNR_Out				MSE			
Noise	-4 dB	-2 dB	0 dB	4 dB	-4 dB	-2 dB	0 dB	4 dB
DWT	1.75 dB	2.21 dB	2.64 dB	4.63 dB	0.211	0.116	0.112	0.094
FCN	2.74 dB	3.22 dB	4.07 dB	6.41 dB	0.142	0.124	0.100	0.066
DNN	4.53 dB	5.23 dB	5.42 dB	7.48 dB	0.090	0.086	0.074	0.052
LSTM	3.82 dB	4.18 dB	4.89 dB	6.13 dB	0.120	0.112	0.096	0.067
IMUnet	6.06 dB	6.52 dB	7.76 dB	9.89 dB	0.062	0.056	0.035	0.029
APtrans-CNN	**7.48 dB**	**7.94 dB**	**8.49 dB**	**9.95 dB**	**0.049**	**0.046**	**0.035**	**0.027**

3.4.4. Visualization of denoising results

In this section, we visualize the experimental results to further explore the performance of the proposed network. Visualization of the atrial fibrillation database shown in Figure 9 demonstrates the superiority of the proposed architecture. DWT has a very significant effect on noise reduction of baseline drift type. However, the noise reduction effect is not obvious with high frequency noise with large amplitude. A similar situation also occurs in the fully convolutional neural network, where the denoised signal contains too much useless noise. The signal processed by FCN can effectively eliminate the high frequency noise, but the wave peak is far from the expected result. This problem also exists in LSTM; the higher the noise content, the greater the gap between the wave peak and the actual wave peak. In addition, it is difficult for the above-mentioned networks to capture useful disease information under the condition of high noise. This can be extremely disruptive to the diagnosis of the medical staff. IMUnet performance has been significantly improved, and the signal smoothness and

noise level have been greatly improved. However, some details are still omitted, which brings greater uncertainty to diagnosis. APtrans-CNN has strong noise reduction ability in the whole noise environment. The signal after noise reduction almost perfectly eliminates the influence of baseline drift electromyography interference and other noises. Meanwhile, the signal also retains information about atrial fibrillation, which is used to help diagnose the patient's condition.
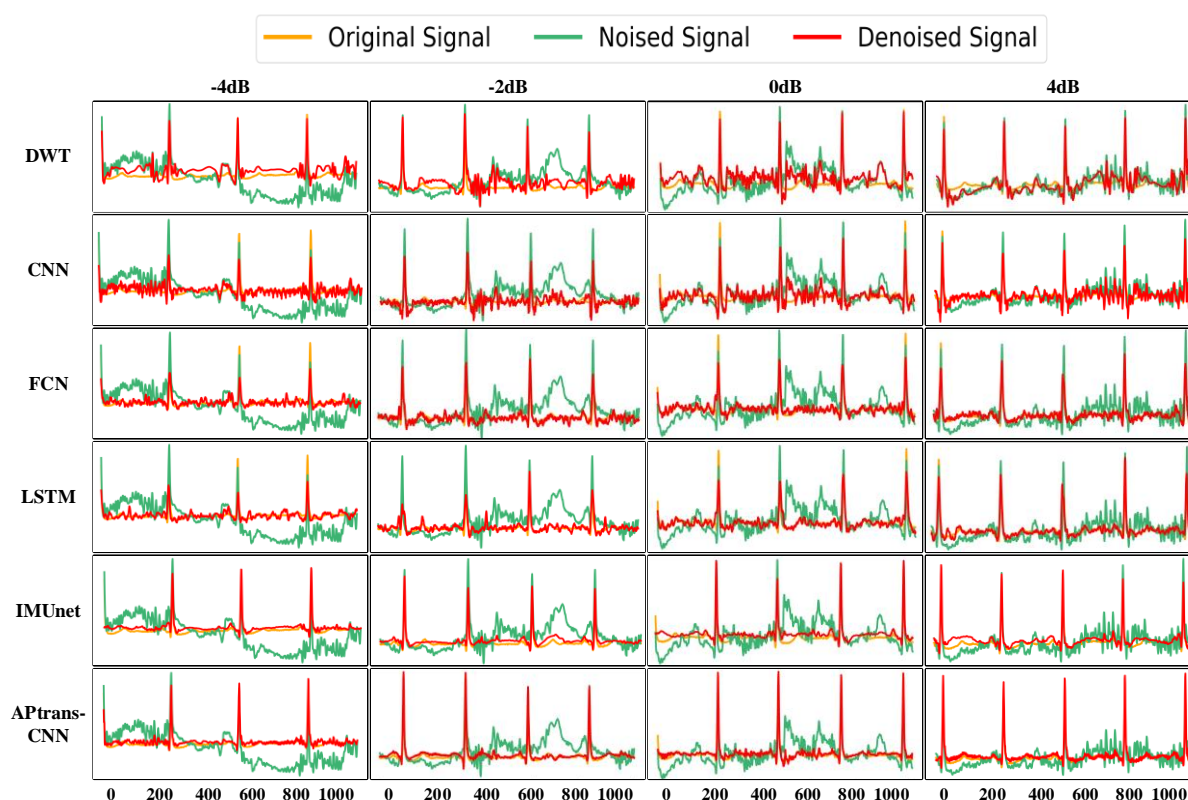

Figure 9. The visualization of the denoising results produced by APtrans-CNN and other methods: DWT, CNN, FCN, LSTM, IMUnet.

3.5.   *Application to diseases diagnosis*

This section introduces the improvement brought by APtrans-CNN for disease diagnosis under noisy conditions. We chose the MIT-BIH Arrhythmia Database as the standard test dataset for evaluating arrhythmia detectors. The MIT-BIH Arrhythmia Database has high confidence and is widely used as a dataset for ECG evaluation. Based on the AAMI criteria, we performed a four-category assessment of this dataset including normal beats (N), supraventricular ectopic beats (S), ventricular ectopic beats (V), and fused beats (F).

We adopted the model proposed by Wang et al. [42] and the model designed by Xu et al. [43]. The model of Wang et al. uses a three-head input, which contains a total of two convolutional layers, two pooling layers, and three fully connected layers. The network proposed by Xu et al. consists of four convolutional layers, two subsampling layers, two fully connected layers and a softmax layer. Each method is tested on the original dataset, the noisy dataset, and the APtrans-CNN denoised dataset. APtrans-CNN is trained in the same way as in the previous section for noise reduction. After denoising

the MIT-BIH Arrhythmia Database with APtrans-CNN and IMUnet, we subject the dataset to 5-fold cross-validation. The average accuracy is used as the final evaluation metric. The experimental results are shown in Table 5.

Table 5. Diagnostic accuracy under different noise levels (MIT-BIH arrhythmia database).

Methods	Noise Level	-4 dB	-2 dB	0 dB	4 dB
Wang et al. [42]	Noise-free	98.47%			
	Noised	42.22%	57.43%	64.11%	76.82%
	IMUnet (Denoised)	77.63%	78.47%	84.71%	95.35%
	APtrans-CNN (Denoised)	**81.76%**	**85.02%**	**91.36%**	**96.21%**
Xu et al. [43]	Noise-free	99.07%			
	Noised	46.03%	60.66%	68.91%	78.61%
	IMUnet (Denoised)	80.28%	82.51%	85.18%	96.36%
	APtrans-CNN (Denoised)	**83.20%**	**86.85%**	**93.36%**	**97.17%**

In the absence of noise, the diagnostic accuracy of the two disease diagnosis methods is very close to 100%. But, with the addition of noise, the diagnostic performance of both methods is greatly limited, and the worst-case accuracy is only around 40%. The addition of APtrans-CNN greatly avoids this situation. For Method I, when the SNR_{dB} = -4 and -2 dB, after APtrans-CNN training, the accuracy was improved by 39.54 and 27.59%. In the test, the denoised network can always maintain a diagnostic accuracy of more than 80%. For Method II, under high noise intensity, APtrans-CNN has a significant improvement in the accuracy of disease diagnosis, and can maintain an accuracy of more than 83%. APtrans-CNN can also play a good role in low noise environments. When SNR_{dB} = 4 dB, the diagnostic accuracy of the two methods after noise reduction reached 93.36 and 97.17%, which is very close to the ECG under pure signal. It can be seen from the test results that APtrans-CNN can effectively help the diagnosis of diseases in a noisy environment, and effectively improve the accuracy of the diagnosis of ECG diseases.

4. Conclusions

ECG signals are an important indicator of human health. Thus, how to separate the pure ECG signal from the noisy ECG signal is an important task. Based on this problem, this paper proposed a novel Transformer-based convolutional neural network framework with adaptively parametric ReLU (APtrans-CNN). APtrans-CNN consists of a Transformer with global feature extraction capabilities and a CNN module which performs well in local feature extraction. This architecture combines the advantages of both and greatly improves the denoising performance of the task. In addition, the AP-ReLU function is introduced to make this model more suitable for the denoising task of ECG signals. The proposed DAM module enables the framework to automatically learn more valuable features from noisy ECG signals.

Through the comparison experiment with the existing ECG signal denoising methods, such as DWT, FCN, DNN, LSTM and IMUnet, we demonstrate the advantages of the proposed method. APtrans-CNN can effectively restore pure ECG signals from noisy ECG signals. It also shows strong robustness at different noise levels and has excellent performance for ECG signals from different

populations. In addition, the experiments also visualize the denoised data and use the denoised results for disease diagnosis tasks. The results showed that ECG signals after denoising retained valuable information which could reflect the patient's health condition well. APtrans-CNN has the potential to become an efficient tool for ECG signal processing, enabling professionals to analyze ECG signals more easily.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

Conflict of interest

The authors declare that there are no conflicts of interest.

References

1. X. Liu, H. Wang, Z. Li, L. Qin, Deep learning in ECG diagnosis: a review, *Knowl.-Based Syst.*, **227** (2021), 107187. https://doi.org/10.1016/j.knosys.2021.107187

2. S. Agrawal, A. Gupta, Fractal and EMD based removal of baseline wander and powerline interference from ECG signals, *Comput. Biol. Med.*, **43** (2013), 1889–1899. https://doi.org/10.1016/j.compbiomed.2013.07.030

3. E. Erçelebi, Electrocardiogram signals de-noising using lifting-based discrete wavelet transform, *Comput. Biol. Med.*, **34** (2004), 479–493. https://doi.org/10.1016/S0010-4825(03)00090-8

4. Z. F. M. Apandi, R. Ikeura, S. Hayakawa, S. Tsutsumi, An analysis of the effects of noisy electrocardiogram signal on heartbeat detection performance, *Bioengineering*, **7** (2020), 53. https://doi.org/10.3390/bioengineering7020053

5. A. O. Boudraa, J. C. Cexus, EMD-based signal filtering, *IEEE Trans. Instrum. Meas.*, **56** (2007), 2196–2202. https://doi.org/10.1109/TIM.2007.907967

6. X. Chen, X. Xu, A. Liu, M. J. McKeown, Z. J. Wang, The use of multivariate EMD and CCA for denoising muscle artifacts from few-channel EEG recordings, *IEEE Trans. Instrum. Meas.*, **67** (2018), 359–370. https://doi.org/ 10.1109/TIM.2017.2759398

7. M. Z. U. Rahman, R. A. Shaik, D. V. R. K. Reddy, Efficient and simplified adaptive noise cancelers for ECG sensor based remote health monitoring, *IEEE Sens. J.*, **12** (2012), 566–573. https://doi.org/10.1109/JSEN.2011.2111453

8. S. Banerjee, M. Mitra, Application of cross wavelet transform for ECG pattern analysis and classification, *IEEE Trans. Instrum. Meas.*, **63** (2014), 326–333. https://doi.org/10.1109/TIM.2013.2278430

9. R. Ranjan, B. C. Sahana, A. K. Bhandari, Cardiac artifact noise removal from sleep EEG signals using hybrid denoising model, *IEEE Trans. Instrum. Meas.*, **71** (2022), 1–10. https://doi.org/10.1109/TIM.2022.3198441

10. B. Weng, M. Blanco-Velasco, K. E. Barner, ECG denoising based on the empirical mode decomposition, in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, (2006), 1–4. https://doi.org/10.1109/IEMBS.2006.260749

11. C. Chandrakar, M. Kowar, Denoising ECG signals using adaptive filter algorithm, *Int. J. Soft Comput. Eng.*, **2** (2012), 120–123.

12. G. Reddy, M. Muralidhar, S. Varadarajan, ECG de-noising using improved thresholding based on wavelet transforms, *Int. J. Comput. Sci. Netw. Secur.*, **9** (2009), 221–225.

13. C. T. C. Arsene, R. Hankins, H. Yin, Deep learning models for denoising ECG signals, in *2019 27th European Signal Processing Conference (EUSIPCO)*, (2019), 1–5. https://doi.org/10.23919/EUSIPCO.2019.8902833

14. P. Singh, G. Pradhan, A new ECG denoising framework using generative adversarial network, *IEEE/ACM Trans. Comput. Biol. Bioinf.*, **18** (2021), 759–764. https://doi.org/10.1109/TCBB.2020.2976981

15. Z. Liu, H. Wang, Y. Gao, S. Shi, Automatic attention learning using neural architecture search for detection of cardiac abnormality in 12-Lead ECG, *IEEE Trans. Instrum. Meas.*, **70** (2021), 1–12. https://doi.org/10.1109/TIM.2021.3109396

16. L. Qin, Y. Xie, X. Liu, X. Yuan, H. Wang, An end-to-end 12-Leading electrocardiogram diagnosis system based on deformable convolutional neural network with good antinoise ability, *IEEE Trans. Instrum. Meas.*, **70** (2021), 1–13. https://doi.org/10.1109/TIM.2021.3073707

17. S. Hong, Y. Zhou, J. Shang, C. Xiao, J. Sun, Opportunities and challenges of deep learning methods for electrocardiogram data: a systematic review, *Comput. Biol. Med.*, **122** (2020), 103801. https://doi.org/10.1016/j.compbiomed.2020.103801

18. K. Antczak, Deep recurrent neural networks for ECG signal denoising, preprint, arXiv:1807.11551. https://doi.org/10.48550/arXiv.1807.11551

19. S. Chatterjee, R. S Thakur, R. N. Yadav, L. Gupta, D. K. Raghuvanshi, Review of noise removal techniques in ECG signals, *IET Signal Process.*, **14** (2020), 569–590. https://doi.org/10.1049/iet-spr.2020.0104

20. H. Chiang, Y. Hsieh, S. Fu, K. Hung, Y. Tsao, S. Chien, Noise reduction in ECG signals using fully convolutional denoising autoencoders, *IEEE Access*, **7** (2019), 60806–60813. https://doi.org/10.1109/ACCESS.2019.2912036

21. P. Singh, A. Sharma, Attention-based convolutional denoising autoencoder for two-lead ECG denoising and arrhythmia classification, *IEEE Trans. Instrum. Meas.*, **71** (2022), 3137710. https://doi.org/10.1109/TIM.2021.3137710

22. F. Samann, T. Schanze, RunDAE model: running denoising autoencoder models for denoising ECG signals, *Comput. Biol. Med.*, **166** (2023), 107553. https://doi.org/10.1016/j.compbiomed.2023.107553

23. P. Xiong, H. Wang, M. Liu, S. Zhou, Z. Hou, X. Liu, ECG signal enhancement based on improved denoising auto-encoder, *Eng. Appl. Artif. Intell.*, **52** (2016), 194–202. https://doi.org/10.1016/j.engappai.2016.02.015

24. E. Fotiadou, T. Konopczyński, J. Hesser, R. Vullings, Deep convolutional encoder-decoder framework for fetal ECG signal denoising, in *2019 Computing in Cardiology (CinC)*, (2019), 1–4. https://doi.org/10.22489/CinC.2019.015

25. R. Hu, J. Chen, L. Zhou, A transformer-based deep neural network for arrhythmia detection using continuous ECG signals, *Comput. Biol. Med.*, **144** (2022), 105325. https://doi.org/10.1016/j.compbiomed.2022.105325

26. L. Meng, W. Tan, J. Ma, R. Wang, X. Yin, Y. Zhang, Enhancing dynamic ECG heartbeat classification with lightweight transformer model, *Artif. Intell. Med.*, **124** (2022), 102236. https://doi.org/10.1016/j.artmed.2021.102236

27. Y. Xia, Y. Xu, P. Chen, J. Zhang, Y. Zhang, Generative adversarial network with transformer generator for boosting ECG classification, *Biomed. Signal Process. Control*, **80** (2023), 104276. https://doi.org/10.1016/j.bspc.2022.104276

28. Y. Xia, Y. Xiong, K. Wang, A transformer model blended with CNN and denoising autoencoder for inter-patient ECG arrhythmia classification, *Biomed. Signal Process. Control*, **86** (2023), 105271. https://doi.org/10.1016/j.bspc.2022.105271

29. J. Yin, A. Liu, C. Li, R. Qian, X. Chen, A GAN guided parallel CNN and transformer network for EEG denoising, *IEEE J. Biomed. Health Inf.*, 2023. https://doi.org/10.1109/JBHI.2023.3146990

30. X. Pu, P. Yi, K. Chen, Z. Ma, D. Zhao, Y. Ren, EEGDnet: Fusing non-local and local self-similarity for EEG signal denoising with transformer, *Comput. Biol. Med.*, **151** (2022), 106248. https://doi.org/10.1016/j.compbiomed.2022.106248

31. J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, *IEEE Trans. Pattern Anal. Mach. Intell.*, **42** (2020), 2011–2023. https://doi.org/10.1109/CVPR.2018.00745

32. S. Woo, J. Park, J. Lee, I. S. Kweon, CBAM: Convolutional block attention module, in *Proceedings of the European conference on computer vision (ECCV)*, (2018), 3–19. https://doi.org/10.1007/978-3-030-01234-2_1

33. M. Zhao, S. Zhong, X. Fu, B. Tang, S. Dong, M. Pecht, Deep residual networks with adaptively parametric rectifier linear units for fault diagnosis, *IEEE Trans. Ind. Electron.*, **68** (2021), 2587–2597. https://doi.org/10.1109/TIE.2020.2972458

34. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., Attention is all you need, *Adv. Neural Inf. Process. Syst.*, (2017), 30. https://doi.org/10.5555/3295222.3295349

35. G. B. Moody, R. G. Mark, The impact of the MIT-BIH arrhythmia database, *IEEE Eng. Med. Biol. Mag.*, **20** (2001), 45–50. https://doi.org/10.1109/51.932724

36. M. B. George, R. G. Mark, A new method for detecting atrial fibrillation using RR intervals, *Comput. Cardiol.*, (1983), 227–230.

37. A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, et al., PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, *Circulation*, **101** (2000), e215–e220. https://doi.org/10.1161/01.CIR.101.23.e215

38. G. B. Moody, W. Muldrow, A noise stress test for arrhythmia detectors, *Comput. Cardiol.*, **11** (1984), 381–384.

39. H. T. Chiang, Y. Hsieh, S. Fu, K. Hung, Y. Tsao, S. Chien, Noise reduction in ECG signals using fully convolutional denoising autoencoders, *IEEE Access*, **7** (2019), 60806–60813. https://doi.org/10.1109/ACCESS.2019.2912036

40. G. Sannino, G. D. Pietro, A deep learning approach for ECG-based heartbeat classification for arrhythmia detection, *Future Gener. Comput. Syst.*, **86** (2018), 446–455. https://doi.org/10.1016/j.future.2018.03.057

41. L. Qiu, W. Cai, Two-stage ECG signal denoising based on deep convolutional network, *Physiol. Meas.*, **42** (2021), 115002. https://doi.org/ 10.1088/1361-6579/ac34ea

42. H. Wang, H. Shi, An improved convolutional neural network based approach for automated heartbeat classification, *J. Med. Syst.*, **44** (2020), 1–9. https://doi.org/ 10.1007/s10916-019-1511-2

43. X. Xu, H. Liu, ECG heartbeat classification using convolutional neural networks, *IEEE Access*, **8** (2020), 8614–8619. https://doi.org/ 10.1109/ACCESS.2020.2964749