Mathematical Biosciences and Engineering

MBE, 20(1): 1–17.
DOI: 10.3934/mbe.2023001
Received: 15 July 2022
Revised: 04 September 2022
Accepted: 12 September 2022
Published: 29 September 2022

http://www.aimspress.com/journal/MBE

Theory article

Deep arrhythmia classification based on SENet and lightweight context transform

Yuni Zeng[1], Hang Lv[1], Mingfeng Jiang[1,*], Jucheng Zhang[2], Ling Xia[3], Yaming Wang[4] and Zhikang Wang[2]

[1] School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China
[2] Department of Clinical Engineering, The Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310019, China
[3] Department of Biomedical Engineering, Zhejiang University, Hangzhou 310027, China
[4] Lishui University, Lishui 323000, China

* **Correspondence:** Email: m.jiang@zstu.edu.cn.

Abstract: Arrhythmia is one of the common cardiovascular diseases. Nowadays, many methods identify arrhythmias from electrocardiograms (ECGs) by computer-aided systems. However, computer-aided systems could not identify arrhythmias effectively due to various the morphological change of abnormal ECG data. This paper proposes a deep method to classify ECG samples. Firstly, ECG features are extracted through continuous wavelet transform. Then, our method realizes the arrhythmia classification based on the new lightweight context transform blocks. The block is proposed by improving the linear content transform block by squeeze-and-excitation network and linear transformation. Finally, the proposed method is validated on the MIT-BIH arrhythmia database. The experimental results show that the proposed method can achieve a high accuracy on arrhythmia classification.

Keywords: continuous wavelet transform; Squeeze-and-Excitation network; lightweight context transform; arrhythmia classification

1. Introduction

In recent years, the global incidence of cardiovascular diseases (CVDs) has increased significantly,

and arrhythmia is one of the main causes of CVDs [1]. Parts of arrhythmias are harmful and even life-threatening. Therefore, the early detection of arrhythmias is necessary, which is also considered as an effective intervention for CVDs. In clinical practice, arrhythmias are usually diagnosed by analyzing the heartbeat of electrocardiogram (ECG) signals [2]. When arrhythmias occur, the ECGs will show abnormal electrical activity, mainly manifested as changes in the morphology of the P wave, QRS wave, and T wave [3]. However, the ECG diagnosis is time-consuming and laborious. With the help of computer-aided intelligent diagnosis [4,5], the identification of ECG can be greatly improved, which is helpful for the treatment of cardiovascular diseases. Therefore, computer-aided tools have great potential to help doctors provide better and faster diagnosis of arrhythmias.

Recent, many studies focus on arrhythmia classification by automatic ways. Study [6] used short-time Fourier transform to convert ECG recordings into two-dimensional spectrograms to obtain frequency and energy information in ECG recordings. Study [7] by calculating RR interval features, morphological features and statistical features based on wavelet packet decomposition. With the development of deep learning, deep models were beginning to be used in arrhythmia classification. Study [8] proposed a deep ventricular tachycardia and ventricular fibrillation classification scheme based on mixed time-frequency features, using wavelet transform, empirical mode decomposition and variable mode decomposition methods to decompose the signal.

In order to capture the multiple morphological change of ECG waves, attention mechanism is used in this paper. The attention mechanism has made great progress in computer vision and natural language processing [9,10]. It has the ability to focus on local patterns, and learn the global information of data, which could help models to learn the small local changes of ECG waves. For example, the Squeeze-and-Excitation network (SENet) [11,12] based on channel attention and aims to selectively emphasize informative channels and suppress trivial channels by explicitly modeling dependencies across channels.

In this paper, we proposed a deep model for arrhythmia classification based on SENet and lightweight context transform (LCT). Firstly, continuous wavelet transform (CWT) are used to preprocess the input ECG sequence as the feature extraction. Then, a new LCT block is proposed to context information of ECG data with Squeeze-and-Excitation (SE) blocks. Here, the SE blocks are improved from SENet by the feature transformation module [13], changing the fusion mode [14], and integrating the spatial attention mechanism [15]. Finally, a classification model with convolution, batch normalization (BN), max pooling layers is used to recognize the arrhythmia. The results on benchmark dataset shows our proposed model has strong ability to classify arrhythmia on ECG.

2. Materials and methods

2.1. Overview

This paper proposes an arrhythmia classification method, which is based on CWT for feature extraction, and then passes the extracted features through the improved SENet method based on LCT, and finally outputs the results. An overview of the method is shown in Figure 1 and will be described in detail below.

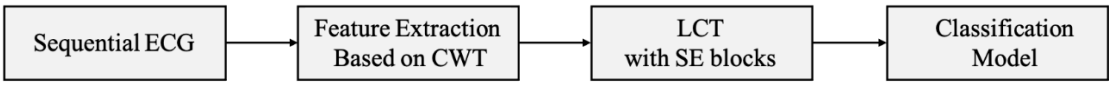

Figure 1. The overall flow of the method.

2.2. Feature extraction

CWT is a widely used method for time-frequency analysis of signals, which overcomes the limitations of Fourier transform (FT), which uses a set of wavelet functions to decompose signals in the time-frequency domain. It differs from (Short-time Fourier Transform, STFT) in that CWT separates the signal components by adjusting the scale and translation parameters [16], resulting in a time-frequency representation of the signal that is more accurate than traditional STFT using a fixed-length sliding time window. CWT is a suitable choice for non-stationary signals such as ECG signals and can replace STFT.

The prototype wavelet used for wavelet transform of ECG signal in this paper is defined as [17]:

$$C_a(b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \varphi\left(\frac{t-b}{a}\right) dt \tag{1}$$

where a is the scale parameter, b is the translation parameter, and φ(t) is the wavelet function (call the mother wavelet). The scale can be converted to frequency by

$$F = \frac{F_c \cdot f_s}{a} \tag{2}$$

F_c is the center frequency of the mother wavelet, f_s is the sampling frequency of the sample $x(t)$.

The choice of mother wavelet is often the fundamental to time-frequency analysis. The study [18] chooses the wavelet Mexican hat (mexh) as the wavelet base, and we will compare the case of Morlet (morl) and Gaussian wavelet bases in experiments. Using different scale factors of wavelet transform, the wavelet coefficients of the signal at different scales are obtained. These wavelet coefficients can be viewed as two-dimensional scales of the ECG signal in the time-frequency domain.

2.3. Squeeze-and-Excitation blocks

For convolutional neural network, the core computation is the convolution operator, which learns new feature maps from the input feature map through the convolution kernel. Essentially, convolution is a feature fusion of a local region, which includes spatial (H × W) and inter-channel (C) feature fusion. A large part of the convolution operation is to improve the receptive field, that is, to integrate more feature fusion in space, or to extract multi-scale spatial information. Feature fusion of channel dimension, the convolution operation basically fuses all channels of the input feature map by default. The SENet [11] is a feature fusion in the channel dimension. It mainly focuses on the relationship between channels, and can automatically learn the importance of features between different channels in the network layer of the model, as shown in Figure 2.

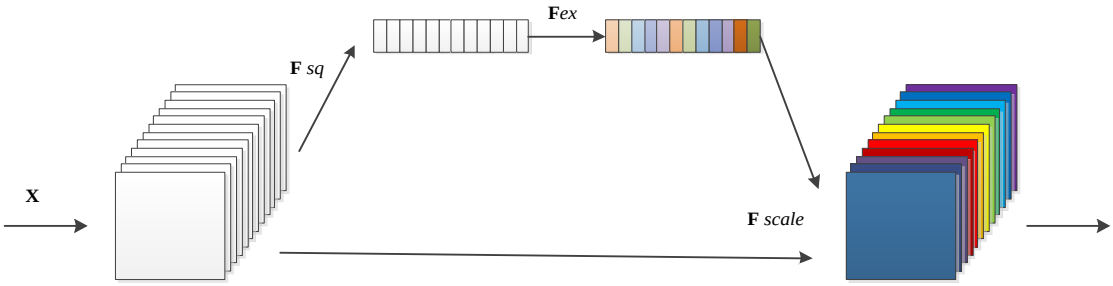

Figure 2. A Squeeze-and-Excitation block.

Squeeze-and-Excitation (SE) is an improved presentation power for the network by establishing interdependencies between convolutional feature channels. By subjecting the network to feature recalibration, it can learn to use global information, selectively emphasizing informative features and suppressing less useful ones. The SE module mainly includes two operations, Squeeze and Excitation, which can be applied to any mapping. For any given transformation,

$$F_{tr}: X \to U,\ X \in R^{H'\times W'\times C'},\ U \in R^{H\times W\times C} \tag{3}$$

Taking convolution as an example, the convolution kernel is $V = [v_1, v_2, \cdots, v_c]$, where v_c represents the c-th convolution kernel. Then the output is $U = [u_1, u_2, \cdots, u_c]$.

$$u_c = v_c * X = \sum_{s=1}^{C'} v_c^s * x^s \tag{4}$$

where * represents the convolution operation, $V = [v_c^1, v_c^2, \cdots, v_c^{C'}]$ and $X = [X^1, X^2, \cdots, X^{C'}]$, while v_c^s represents a 2-D convolution kernel of a s channel, which inputs a spatial feature on a channel, and it learns the feature space relationship. However, since the convolution results of each channel are summed, the feature relationships of the channel are mixed with the spatial relationships learned by the convolution kernels. In order to extract this confusion, the SE module enables the model to directly learn the channel feature relationship.

2.3.1. Squeeze operation

Since convolution only operates in a local space, it is difficult for U to obtain enough information to extract the relationship between channels. Therefore, SE encodes the entire spatial feature on a channel into a global feature through the Squeeze operation, which is implemented by global average pooling (in principle, a more complex aggregation strategy can also be used) as follows:

$$z_c = F_{sq}(u_c) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} u_c(i,j), z \in R^C \tag{5}$$

2.3.2. Excitation operation

The squeeze operation obtains the global description feature, and another operation is required to

capture the relationship between channels. This operation needs to meet two criteria: first, it must be flexible, and it must be able to learn the nonlinear relationship between each channel; the second point is that the learned relationship is not mutually exclusive, because multi-channel features are allowed here. Based on this, the gating mechanism in the form of sigmoid is as follows, according to [11]:

$$s = F_{ex}(z, W) = \sigma\left(g(z, W)\right) = \sigma\left(W_2 \operatorname{Re} LU\left(W_1 z\right)\right) \tag{6}$$

where $W_1 \in R^{\frac{C}{r} \times C}, W_2 \in R^{C \times \frac{C}{r}}$. In order to reduce the complexity of the model and improve the generalization ability, a structure consisting of two fully connected layers is adopted, in which the first FC layer plays the role of dimensionality reduction. The second FC layer restores the original dimension. Finally, the learned activation value of each channel is multiplied by the original feature on U:

$$\widetilde{x_c} = F_{scale}(u_c, s_c) = u_c \cdot s_c \tag{7}$$

The entire operation process can be seen as learning the weight coefficients of each channel, so that the model has more ability to distinguish the characteristics of each channel, which is also an attention mechanism.

2.4. Lightweight context transform blocks

It can be known from reference [11], SE performs nonlinear transformations to learn the negative correlation between global context and attention values by explicitly capturing dependencies across channels. However, this negative correlation is learned from all channels, which may introduce irrelevant information for each channel from other channels and destabilize global context modeling, leading to incorrect mappings. To solve this problem, a new LCT block is proposed based on SE blocks and the linear context transform. According to [13], the LCT also mainly includes two operation modules, context aggregation and context feature transform, its frame is shown in Figure 3.

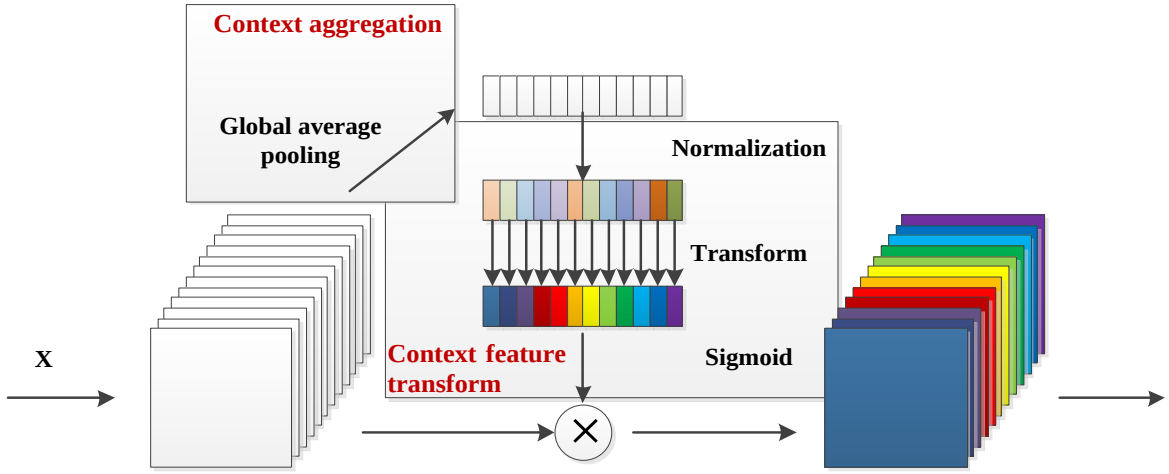

Figure 3. Architecture of our proposed LCT block. Where × denotes broadcast element-wise multiplication.

2.4.1. Context aggregation

The purpose of context aggregation is to help the network capture long-term dependencies by leveraging information beyond each filter's local receptive fields. The method adopted by context aggregation is the same as the squeeze operation in SE, which adopts global average pooling.

2.4.2. Context feature transform

LCT introduces a pair of lightweight operators in context feature transform: a normalization operator and a transformation operator. The normalization operator normalizes the global context features in each group, and the transformation operator takes the normalized results as inputs to output the importance scores. That is, the context aggregation is first divided into groups and then normalized along the channel dimension. We define $v^i = \{z_{mi+1}, \cdots, z_{m(i+1)}\}$ as the i-th local context group, where $i \in \{0, \cdots, G-1\}$ and G are the index and the number of groups, respectively. $m = \frac{C}{G}$ is the number of channels per group. The normalization operator φ formula is as follows:

$$\hat{v}^i = \varphi\left(v^i\right) = \frac{1}{\sigma^i}\left(v^i - \mu^i\right) \tag{8}$$

where μ^i and σ^i are as follows:

$$\mu^i = \frac{1}{m}\sum_{n \in S_i} z_n, \sigma^i = \sqrt{\frac{1}{m}\sum_{n \in S_i}\left(z_n - \mu^i\right)^2 + \varepsilon} \tag{9}$$

and ε is a small constant, S_i is the set of the i-th group of channel index.

Next, a transformation operator ϕ is used, which is defined as follows:

$$a = \phi\left(\hat{z}\right) = w \cdot \hat{z} + b \tag{10}$$

where $\hat{z} = [\hat{v}^0, \hat{v}^1, \cdots, \hat{v}^{G-1}]$, w and b are trainable gain and bias parameters in the same dimension as $\hat{z}$. The transformation operator ϕ is a linear transformation of one channel, which does not consider information from other channels during context transformation. Furthermore, it only introduces the parameters of w and b, which are almost negligible compared to the entire network. Finally, the feature fusion module adjusts the input features according to the transformed context conditions. Specifically, the output $Y \in C \times H \times W$ of the LCT block is obtained by scaling the original response X according to the attention activation $\sigma(a)$ and can be expressed as:

$$Y = X \cdot \sigma\left(a\right) \tag{11}$$

Both LCT and SE use the same context aggregation module and feature fusion module. However, SE uses the global information of other channels to model global features. In contrast, LCT is more lightweight and simplifies global feature modeling by independently transforming the information of each channel.

2.5. Classification model

Convolutional neural network (CNN) is the most common type of deep nonlinear network structure, which can obtain transformation-invariant features when applying neurons with the same parameters to different positions in the previous layer [19]. The two main features of CNN are: sparse connection and parameter sharing. Sparse connections are achieved by making the kernel size and filter smaller than the input, which reduces the computational complexity of the model. Parameter sharing refers to using the same parameters in the multiplication operation, that is, the parameters of each convolution kernel are the same when processing different positions of the input. Other layers commonly used in CNNs are ReLU, BN, and pooling layers. The ReLU acts as an activation function to implement nonlinear functions. The BN is usually located between the convolutional layer and the ReLU layer, and normalizes the feature maps of each channel, reducing the sensitivity of training time and network initialization. The pooling layer, also known as the subsampling layer, is used to reduce the feature dimension and speed up the training process, the role of this layer is to calculate the average or maximum convolutional features of adjacent neurons in the previous convolutional layer [17]. The last layer of the CNN is usually connected to a fully connected layer (FC), which is used to integrate the feature representations extracted by the CNN together.

Study [20] has already demonstrated the effectiveness of Focal Loss in dealing with the imbalance of ECG data, so the loss function of model training adopts the Focal Loss function, and the hyperparameters of the Focal Loss function also refer to [20]. To directly compare the performance of SENet and LCT, we refer to paper [18] to build a simple convolutional neural network classification model. As shown in Figure 4, we combine the convolutional layer, batch normalization layer, ReLU layer and pooling layer into a convolutional unit, and continuously use three convolutional unit operations to extract features, then input to the fully connected layer, and finally softmax output result.

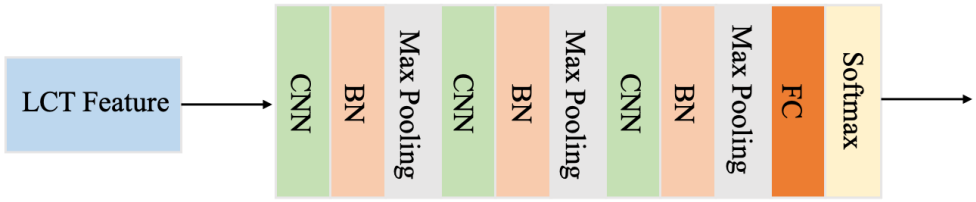

Figure 4. The structure of classification model.

3. Experimental setting

3.1. Dataset

The ECG data used to train the model in this study were obtained from the MIT-BIH arrhythmia database, which randomly selected 23 representative samples of clinical routine and 25 records containing rare but clinically significant arrhythmias from 4000 24-hour ambulate records at a sampling rate of 360 Hz. The dataset consisted of 48 annotated half-hour dual-lead ECG recordings collected by 47 subjects [21]. Each record consists of two leads (lead I and lead II). Lead I (MLII) is modified and is common in the above records, so the ECG data of lead I is used for the evaluation of this method.

MIT-BIH Arrhythmia Database records also contain 15 arrhythmia types labeled by two or more professional physicians. We further classify these arrhythmias into five categories according to ANSI/AAMI EC57-2012 criteria [22]. Four records (102, 104, 107 and 217) were also removed, as shown in Table 1. In addition, since class Q has no application value, it is ignored in the experimental process and does not participate in the evaluation of the method in this paper [23,24].

In order to make a direct comparison with the existing work, the data partition method proposed by [25] is used to segment the database. MIT-BIH arrhythmia database was divided into DS1 and DS2 datasets [25]. Both DS1 and DS2 were composed of 22 records with similar proportions of heart beat types. DS1 is used to train the model, and DS2 is used to test the performance of the method. The specific sample numbers of DS1 and DS2 are shown in Table 2.

Before the ECG is input into the model, we first segment the temporal sequence into the form of individual beats, which are input as samples. In this paper, the position of the R peak that has been marked in the MIT-BIH arrhythmia database is used as the reference point, and the ECG signal is divided into a series of heart beats, which is also convenient for directly comparing the performance of other methods. For each heartbeat, we obtained an ECG signal with a fixed size of 300 sample points by collecting 100 sample points before the R peak and 200 sample points after the R peak, and these sample points contained enough informative features of the waveform.

Table 1. MIT/BIH data set 5 heart beats according to ANSI/AAMI EC57-2012 standard.

N	S	V	F	Q
Normal beat (N)	Atrial premature beat (A)	Premature ventricular contraction (V)	Fusion of ventricular and normal beat (F)	Paced beat (/)
Left bundle branch block beat (L)	Aberrated atrial premature beat (a)	Ventricular escape beat (E)		fusion of paced and normal beat (f)
Right bundle branch block beat (R)	Nodal (junctional) premature beat (J)			Unclassifiable beat (Q)
Atrial escape beat (e)	Premature or ectopic supraventricular beat (S)			
Nodal (junctional) escape beat (j)				

Table 2. ECG samples in train set (DS1) and test set (DS2).

Data Set	N	S	V	F	Sum
DS1	45,824	3788	943	414	50,969
DS2	44,218	3219	1836	388	49,661

3.2. Experimental environment and hyper-parameter setting

In this paper, the G of the LCT block is set as 32. Moreover, the CNN classification model adopts Keras as the deep learning platform, TensorFlow as the backend, NVIDIA GeForce RTX 3090 as the hardware platform, and Ubuntu 18.04.5 as the operating system. For the network parameter of the classification model, the filter number is 16 and kernel size is 7×7 at the first CNN; In the second CNN layer, the filter number is 32 while the kernel size is 3×3. At the third CNN layer, the filter number is 64 and kernel size is 3×3. Among them, cross entropy and focal loss were used to compare the loss function respectively, and both optimizers were Adam. The data was input into the network in batches, and the batch size was set to 32. The weights were initialized randomly before the training began. The learning rate was set to 0.001, decreasing by 0.1 times every 2 epochs. The maximum epoch is set to 30. If the training effect is not significantly improved after training 10 epochs, the training will be terminated in advance.

3.3. Evaluation criterion

This paper uses multiple evaluation metrics to measure the classification performance of the model. Accuracy, precision, recall and F-Score (F1) were used to evaluate the performance of the proposed method. The four indicators (accuracy, precision, recall, F1) are defined as follows:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{12}$$

$$precision = \frac{TP}{TP+FP} \tag{13}$$

$$recall = \frac{TP}{TP+FN} \tag{14}$$

$$F_1 = 2 \times \frac{precision \times recall}{precision+recall} \tag{15}$$

where TP is the correct prediction number of positive samples, TN is the correct prediction number of negative samples, FN is the wrong prediction number of positive samples, FP is the wrong prediction number of negative samples.

4. Experimental results and discussion

4.1. Feature extraction

In this paper, Mexican hat (Mexh), Morlet (Morl) and Gaussian (Gaus) wavelets could be used as the wavelet base of continuous transformation to extract CWT features of original ECG signal and LCT features. For data visualization, we employ Mexh wavelet to show the signals as samples in this section. Figure 6 shows the original ECG signals of N, S, V and F, respectively. Figure 5 shows the CWT features extracted from ECG signals, and Figure 7 shows the features of CWT features processed by LCT. It can be seen from the figure that the time-frequency feature map of the N-type samples has the highest energy at the R peak, and then gradually decreases, and the energy on both sides of the R

peak will suddenly become zero; the time-frequency energy distribution of the S-type samples is similar to the N-type samples, and their biggest difference is that the high frequency of the N-type samples is larger than that of the S-type samples; the time-frequency energy of the V-type samples is more average than that of the N-type samples, and the energy of the R-peak is less, but the energy on both sides of the R-peak of the heart beat is the most, while the F-type samples have more energy. There is more low-frequency energy in the samples, and it can be seen that the time-frequency energy maps of different types of samples are obviously different.

As shown in Figures 5–7, the S-type sample features extracted by continuous wavelet transform are similar to those of N-type samples, and the number of N-type samples is much higher than that of S-type samples. Therefore, most S-type samples are wrongly classified into N labels for the model.

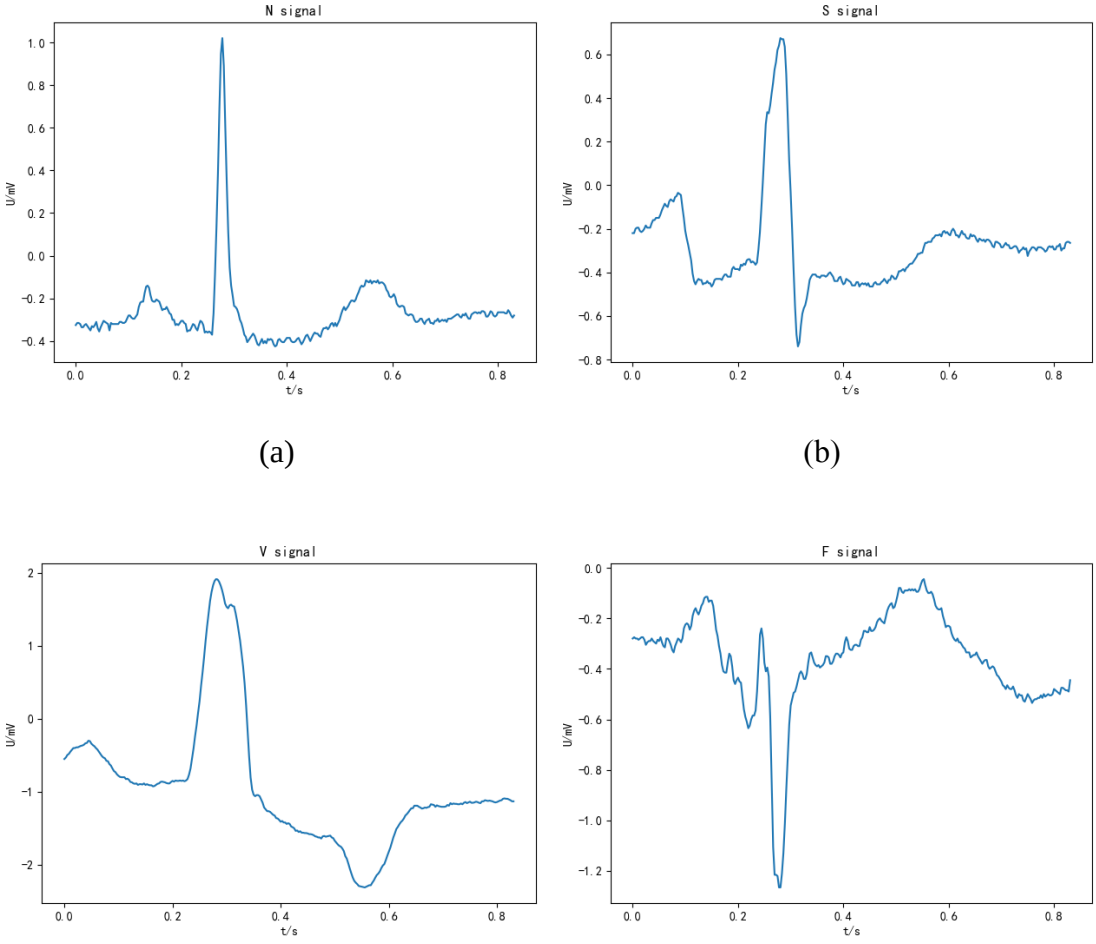

Figure 5. (a)–(d) are the original ECG signals of N, S, V and F, respectively.

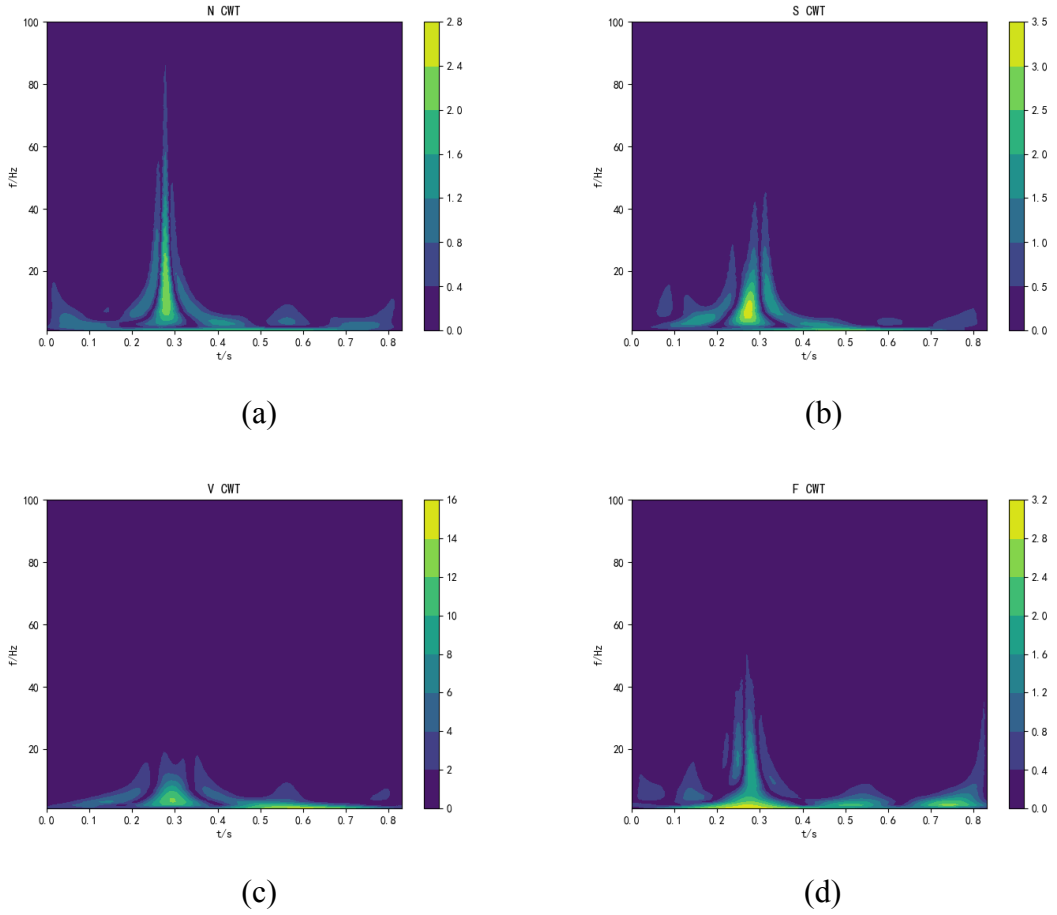

Figure 6. (a)–(d) are the CWT Time-frequency features of N, S, V and F, respectively.

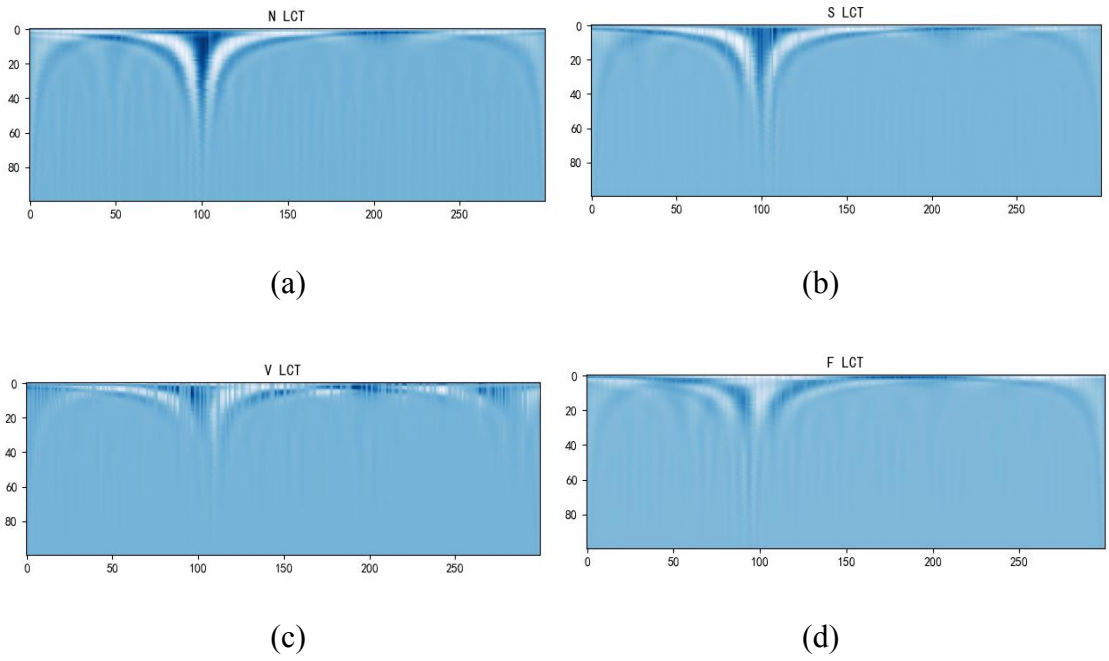

Figure 7. (a)–(d) are the features of N, S, V and F by LCT method, respectively.

4.2. Arrhythmia classification

To verify the validity of the experiments, we separately verify the performance of the following three methods. These three methods are not using any method to process features, only use CNN classification model (denoted as CNN), use SE method to process features and then input CNN classification model (denoted as SE + CNN) and use LCT method to process features and then input CNN Classification model (denoted as LCT + CNN).

Table 3. Comparison of evaluation indexes of different wavelet families using different methods.

METHOD	WAVELET	PRECISION	RECALL	ACCURACY	F1
CNN	Morl	0.8634	0.8153	0.8379	0.8351
	Mexh	0.8695	0.8349	0.8505	0.8414
	Gaus4	0.8636	0.7966	0.8194	0.8302
	Gaus8	0.8690	0.8417	0.8670	0.8491
SENET + CNN	Morl	0.8617	0.8228	0.8481	0.8379
	Mexh	0.8700	0.8316	0.8533	0.8487
	Gaus4	0.8647	0.8003	0.8722	0.8296
	Gaus8	0.8653	0.8491	0.8732	0.8544
LCT + CNN	Morl	0.8690	0.8319	0.8553	0.8490
	Mexh	0.8735	0.8523	0.8757	0.8596
	Gaus4	0.8633	0.8228	0.8751	0.8412
	Gaus8	0.8751	0.8558	0.8770	0.8633

We first use continuous wavelet transform to extract features, and we choose several commonly used wavelet families (Morl, Mexh, Gaus4 and Gaus8) to do experiments. The influence of features extracted from different wavelets on model classification was verified by CNN, SENET + CNN and LCT + CNN respectively, as shown in Table 3. It can be found that Mexh has better precision in extracting features from samples, but recall and F1 have better performance by Gaus8. In the case of gaus8 wavelet, compared with CNN and CNN + SENET, CNN + LCT used in this paper has improved precision by 0.0061 and 0.0098, recall by 0.0141 and 0.0067, accuracy by 0.01 and 0.0038. F1 increased by 0.0142 and 0.0089.

The confusion matrix illustrates the inconsistency between predicted labels and ground truth labels, row labels indicate the ground truth labels in each row, column labels indicate the predicted labels in each column, and colors indicate the proportion of the above records to all records in the same row.

Figures 8–10 show the confusion matrices of the three classification methods. Due to the small number of F samples, the results of model classification are easily affected by other types of samples, but the LCT method proposed in this paper is less affected. On the whole, the CNN + SENET method has no significant improvement compared with the method directly adopted by CNN, and the effect is even worse in the V, S and F classes with fewer data samples. However, the CNN + LCT proposed in this paper achieves the best average performance compared with the previous two methods. Meanwhile, although the number of samples of class V is also small, the confusion matrix shows that the classification effect of class V of the three methods is very good, almost reaching the classification accuracy of class N.

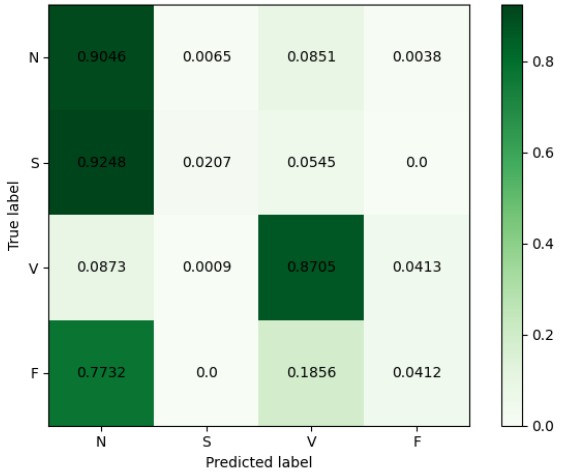

Figure 8. CNN confusion matrix of Gaus8 wavelet families.

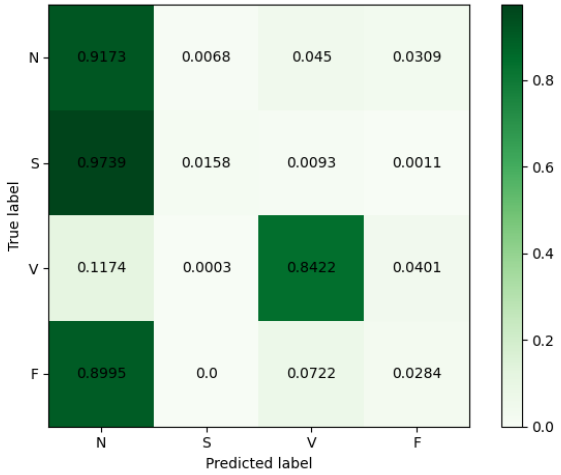

Figure 9. CNN + SENET confusion matrix of Gaus8 wavelet families.

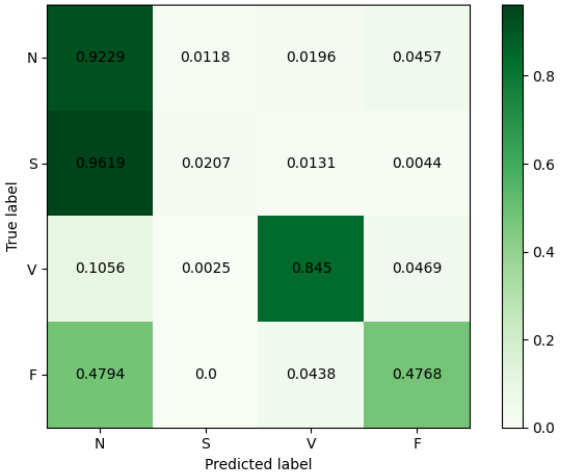

Figure 10. CNN + LCT confusion matrix of Gaus8 wavelet families.

Figure 11 shows the change value of training Loss of the three methods. The convergence speed is slowest when features are directly input into CNN without any processing while the speed of 'SENet + CNN' model is the second. The LCT used in this paper is the fastest, and the training loss value after LCT pretreatment is also the smallest.

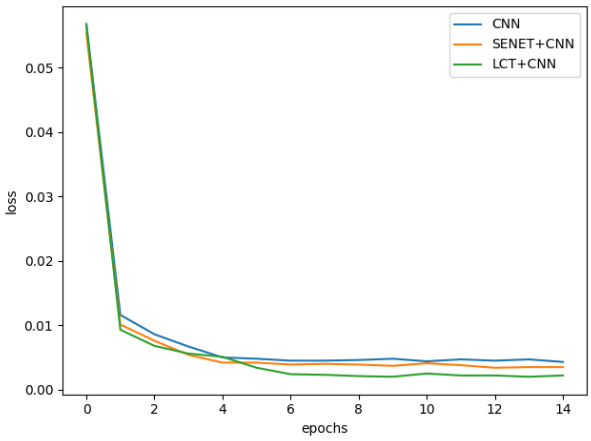

Figure 11. Loss for training.

5. Conclusions

In order to capture the different morphological change of abnormal ECG data, this paper focus on deep arrhythmia classification methods. The proposed method extracts ECG features through CWT firstly, and then utilize the LCT block which combine normalization, linear transformation and SENet to classify ECG data. Especially, the CWT is one kind of time-frequency analysis, thus the feature has the advantages of time-domain and frequency-domain analysis. Finally, the experimental results show the proposed model in this paper has strong ability to recognize arrhythmia in ECG with contribution of the proposed LCT block and classification model. Besides, the performance of classification on N and V samples is significantly improved while the classification performance of classification on F samples has relative less improvement. The reason may be the relative less amount of F samples.

The limitation of this work includes:

1) We do not compare our work and the first proposed linear context transform [13]. The motivation of this work is from work [13], but that work focus on image data and our work is for ECGs data.

2) We do not add RR interval characteristics although it may contribute to the classification. This is because the RR interval may contribute to all methods and the point of this work focus on the improvement and application for LCT on ECGs. As for future works, firstly, we will consider the comparation on complex neural network of LCT + CNN model. Secondly, nowadays, the large amount of available ECG data is always unbalance, and the ECG data with abnormal heart rate are often small in amount. Thus, a large amount of clinical ECG data will be collected from the hospital in the future, hoping to continue to improve the current method, further try to solve the problem of data imbalance and improve the accuracy of the algorithm.

Acknowledgments

This work is supported in part by the Key Research and Development Program of Zhejiang Province (2020C03060), the National Natural Science Foundation of China (61672466 62011530130 and 61671405), Joint Fund of Zhejiang Provincial Natural Science Foundation (LSZ19F010001, LZ20F020003), and this work is also supported by the 521 Talents project of Zhejiang Sci-Tech University.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. H. Chen, L. Shi, M. Xue, N. Wang, X. Dong, Y. Cai, et al., Geographic variations in in-hospital mortality and use of percutaneous coronary intervention following acute myocardial infarction in China: A nationwide cross-sectional analysis, *J. Am. Heart Assoc.*, **7** (2018), e008131. https://doi.org/10.1161/JAHA.117.008131
2. S. Yang, H. Shen, Heartbeat classification using discrete wavelet transform and kernel principal component analysis, in *IEEE 2013 Tencon-Spring*., Sydney, Australia, (2013), 34–38. https://doi.org/10.1109/TENCONSpring.2013.6584412
3. J. Park, S. M. Hwang, J. W. Baek, Y. N. Kim, J. H. Lee, Cardiac arrhythmias auto detection in an electrocardiogram using computer-aided diagnosis algorithm, in *Applied Mechanics and Materials.*, (2014), 2728–2731. https://doi.org/10.4028/www.scientific.net/AMM.556-562.2728
4. T. Xia, M. Shu, H. Fan, L. Ma, Y. Sun, The development and trend of ECG diagnosis assisted by artificial intelligence, in *Proceedings of the 2019 2nd International Conference on Signal Processing and Machine Learning*, ACM, New York, USA, (2019), 103–107. https://doi.org/10.1145/3372806.3372807
5. S. Kaplan Berkaya, A. K. Uysal, E. S. Gunal, S. Ergin, S. Gunal, M. B. Gulmezoglu, A survey on ECG analysis, *Biomed. Signal Process. Control*, **43** (2018), 216–235. https://doi.org/10.1016/j.bspc.2018.03.003
6. A. Çınar, S. A. Tuncer, Classification of normal sinus rhythm, abnormal arrhythmia and congestive heart failure ECG signals using LSTM and hybrid CNN-SVM deep neural networks, *Comput. Methods Biomech. Biomed. Eng.*, **24** (2021), 203–214. https://doi.org/10.1080/10255842.2020.1821192
7. Z. Wang, H. Li, C. Han, S. Wang, L. Shi, Arrhythmia classification based on multiple features fusion and random forest using ECG, *J. Med. Imaging Health Inf.*, **9** (2019), 1645–1654. https://doi.org/10.1166/jmihi.2019.2798
8. S. Sabut, O. Pandey, B. S. P. Mishra, M. Mohanty, Detection of ventricular arrhythmia using hybrid time–frequency-based features and deep neural network, *Phys. Eng. Sci. Med.*, **44** (2021), 135–145. https://doi.org/10.1007/s13246-020-00964-2
9. H. Zhang, K. Dana, J. Shi, Z. Zhang, X. Wang, A. Tyagi, et al., Context encoding for semantic segmentation, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, (2018), 7151–7160. https://doi.org/10.1109/CVPR.2018.00747

10. K. Xu, J. Ba, R. Kiros, K. Cho, A. Couville, R. Salakhutdinov, et al., Show, attend and tell: Neural image caption generation with visual attention, in *International Conference on Machine Learning*, PMLR, (2015), 2048–2057.
11. J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, **42** (2018), 7132–7141. https://doi.org/10.1109/TPAMI.2019.2913372
12. X. Chu, B. Zhang, R. Xu, MoGA: Searching beyond mobilenetv3, in *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Barcelona, Spain (2020), 4042–4046. https://doi.org/ 10.1109/ICASSP40776.2020.9054428
13. D. Ruan, J. Wen, N. Zheng, M. Zheng, Linear context transform block, in *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI, New York, USA **34** (2020), 5553–5560. https://doi.org/10.1609/aaai.v34i04.6007
14. M. S. Moustafa, S. A. Sayed, Satellite imagery super-resolution using squeeze-and-excitation-based GAN, *Int. J. Aeronaut. Space Sci.*, **22** (2021), 1481–1492. https://doi.org/10.1007/s42405-021-00396-6
15. S. Woo, J. Park, J. Y. 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
16. M. F. Guo, X. D. Zeng, D. Y. Chen, N. C. Yang, Deep-learning-based earth fault detection using continuous wavelet transform and convolutional neural network in resonant grounding distribution systems, *IEEE Sens. J.*, **18** (2017), 1291–1300. https://doi.org/10.1109/JSEN.2017.2776238
17. Z. Wu, T. Lan, C. Yang, Z. Nie, A novel method to detect multiple arrhythmias based on time-frequency analysis and convolutional neural networks, *IEEE Access*, **7** (2019), 170820–170830. https://doi.org/10.1109/ACCESS.2019.2956050
18. T. Wang, C. Lu, Y. Sun, M. Yang, C. Liu, C. Ou, Automatic ECG classification using continuous wavelet transform and convolutional neural network, *Entropy*, **23** (2021), 119. https://doi.org/10.3390/e23010119
19. A. Ajit, K. Acharya, A. Samanta, A review of convolutional neural networks, in *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, IEEE, Vellore, India, (2020), 1–5. https://doi.org/10.1109/ic-ETITE47903.2 Vellore, India 020.049
20. Y. Lu, M. Jiang, L. Wei, J. Zhang, Z. Wang, B. Wei, et al., Automated arrhythmia classification using depthwise separable convolutional neural network with focal loss, *Biomed. Signal Process. Control*, **69** (2021), 102843. https://doi.org/10.1016/j.bspc.2021.102843
21. 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
22. ANSI/AAMI EC57:2012/(R)2020, Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms, 2012. Available from: https://array.aami.org/doi/10.2345/9781570204784.ch1
23. T. Mar, S. Zaunseder, J. P. Martínez, M. Llamedo, R. Poll, Optimization of ECG classification by means of feature election, *IEEE Trans. Biomed. Eng.*, **58** (2011), 2168–2177. https://doi.org/10.1109/TBME.2011.2113395

24. V. Mondéjar-Guerra, J. Novo, J. Rouco, M. G. Penedo, M. Ortega, Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers, *Biomed. Signal Process. Control*, **47** (2019), 41–48. https://doi.org/10.1016/j.bspc.2018.08.007
25. P. D. Chazal, M. O'Dwyer, R. B. Reilly, Automatic classification of heartbeats using ECG morphology and heartbeat interval features, *IEEE Trans. Biomed. Eng.*, **51** (2004), 1196–1206. https://doi.org/10.1109/TBME.2004.827359

AIMS Press