PVC Recognition for Wearable ECGs using Modified Frequency Slice Wavelet Transform and Convolutional Neural Network

Zhongyao Zhao¹,², Xingyao Wang¹, Zhipeng Cai¹, Jianqing Li¹, Chengyu Liu¹

¹School of Instrument Science and Engineering, Southeast University, Nanjing, China
²School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

Abstract

Progress in wearable techniques makes the long-term daily electrocardiogram (ECG) monitoring possible. It allows the physician to diagnose heart diseases and risks earlier and thus more accurately. Premature ventricular contraction (PVC) is one of the most common cardiac arrhythmias. This study proposed a method by combining the modified frequency slice wavelet transform (MFSWT) and convolutional neural network (CNN). Training data are from the 2018 China physiological signal challenge (934 PVC and 906 non-PVC recordings). The first 10-s ECG waveforms in each recording were transformed into 2-D time-frequency images (frequency range of 0-50 Hz and size of 300×100) using MFSWT. A 25-layer CNN structure was constructed, which includes five convolution layers with kernel size of 3×3, five dropout layers, five ReLU layers, five maximum pooling layers with kernel size of 2×2, a flatten layer, two fully connected layers, as well as the input and output layers. Test data were recorded from 12-lead Smart ECG vests, including 775 PVC and 742 non-PVC recordings. Results showed that, the proposed method achieved a high accuracy of 97.89% for PVC/non-PVC episodes classification, indicating that the combination of MFSWT and CNN provides new insight to accurately identify PVC from the wearable ECG recordings.

1. Introduction

As a comprehensive reflection of cardiac activity, electrocardiogram (ECG) analysis has proven to be the archetypal method for detection of dangerous cardiac conditions. ECG effectively presents valuable clinical information regarding the rate, morphology, and regularity of the heart while being a low-cost and non-invasive test [1,2]. Advancement of wearable technology has enabled the recording of long-term dynamic ECGs. Specifically, wearable ECG analysis can be used for real-time detection of cardiac arrhythmias, such as premature ventricular contractions (PVCs).

PVCs result from irritated ectopic foci in the heart’s ventricles, and are independent of the pace set by the sinoatrial node. Recent studies have shown that the occurrence of PVCs is indicative of increased risk of sudden cardiac death, and is linked to mortality when associated with myocardial infarction [3]. Consequently, their immediate detection and treatment is essential for patients with heart disease.

Different methods have been proposed for heart beat classification, such as support vector machine (SVM) [4,5], neural network [6,7], fuzzy mathematics [8], disease rule [9] and so on. Most PVC detection algorithms can obtain good classification performance on the standard ECG database [10]. However, the traditional neural network algorithm has a large amount of computation and a long training time, and it is hard to realize real-time detection. Fuzzy mathematics methods need to simplify many clinical discriminating rules and it is difficult to adapt to complex ECG records.

This study aims to explore the possibility of PVC identification method based on the combination of modified frequency slice wavelet transform (MFSWT) and convolutional neural network (CNN). The idea is that with the highlight of PVC characteristic in 2-D MFSWT images, the employed CNN model can accurately identify the PVC incidents in ECGs.

2. Methods

2.1 Data

Training data were from the 2018 China physiological signal challenge (CSPC-2018) [11] and test data were recorded from a newly developed 12-lead Lenovo Smart ECG vest [12]. All recordings were intercepted the only first 10-s segments. The database consists of 1,840 training 10-s segments and 1,517 test segments. ECGs have a sampling rate of 400 Hz and a resolution of 16 bits. A large part of PVC signals only has a heart beat’s abnormality, often lasting only about 1 second. Thus the 10-s ECG segments have been manually labeled by clinical experts and technicians as two types: PVC and non-PVC. The data profile was given in Table 1.

<table>
<thead>
<tr>
<th>Database</th>
<th>Type</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>PVC</td>
<td>934</td>
</tr>
<tr>
<td></td>
<td>non-PVC</td>
<td>906</td>
</tr>
<tr>
<td>Test</td>
<td>PVC</td>
<td>775</td>
</tr>
<tr>
<td></td>
<td>non-PVC</td>
<td>742</td>
</tr>
</tbody>
</table>
2.2 Signal pre-processing

Smooth denoising and normalization were used as signal pre-processing. Smooth denoising removes high frequency components from the signal and normalization can eliminate the dimensions of the sample and summarize the statistical distribution of the sample. This operation can improve the performance of learning speed rate of the neural network since the singular sample can not only increase the network training time, but also may cause the model fail to converge.

2.3 Modified frequency slice wavelet transform (MFSWT)

MFSWT can efficiently contain the time-frequency information of ECG in the transformed 2-D images, such as P-wave, QRS complex and T-wave, and were successfully applied in the previous studies [13,14]. A bound signal-adaptive frequency slice function (FSF) was introduced in MFSWT, which can realize the adaptive measurement of signal energy distribution at different observation frequencies. The reconstruction is readily accepted by clinicians. The model of MFSWT was expressed as follows:

\[
W_f(t, \omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(k) \hat{\rho}^* \left( \frac{\omega - \mu}{q(\omega)} \right) e^{-ikt} dk
\]  

where \( \hat{f}(k) \) is the Fourier transform of \( f(t) \), \( t \) and \( \omega \) are observed time and frequency, respectively, * represents conjugation operator, \( \hat{\rho} \) is the FSF defined as \( \hat{\rho}(x) = e^{-x^2/2} \), \( q \) is a scale function of \( \hat{f}(k) \). It uses the function form of Eq. (2):

\[
q = \delta + \text{sign}(\nabla \hat{f}(\mu))
\]  

where \( \delta \) is the frequency position of the signal main component. It can be estimated from the frequency position corresponding to maximum \( \hat{f}(\mu) \). \( \nabla(\cdot) \) is a differential operator, and \( \text{sign}(\cdot) \) means signum function.

FSF here adopts a Gaussian function form and \( \hat{\rho}(0) = 1 \) is always true. Then the original signal can be reconstructed as follows:

\[
f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W_f(t, \omega)e^{i\omega(t-\tau)} d\tau d\omega
\]  

Figure 1 shows the examples of preprocessed 10-s normal and PVC segments and their corresponding MFSWT images with different frequency ranges: 0-90 Hz and 0-50 Hz with a fixed pixel size of 300×100.

2.4 Convolutional neural network (CNN)

CNNs are now commonly used for deep learning tasks, voiding the need for any manual feature extraction and postprocessing. In the current study, a 25-layer CNN structure was constructed. Except the input and output layer, it includes five convolution layers, five dropout layers, five ReLU layers, five maximum pooling layers, a flatten layer and two fully connected layers. Figure 2 illustrates the architecture of the implemented network. Table 2 gives the specific parameter settings (after optimization) for the CNN architecture used in this study.

2.5 Model evaluation

Six widely used metrics, i.e., sensitivity (Se), specificity (Sp), accuracy (Acc), F-measure, area under the receiver operating characteristic curve (ROC), i.e., AUC, and Kappa coefficient, were used for evaluation. According to the labelled references, the result can generate four basic parameters: true positive (TP), false positive (FP), true negative (TN) and false negative (FN). In this case, \( \text{Acc} = (TP + TN)/(TP + TN + FP + FN) \). Se is the true positive rate, is probability of incorrectly
diagnosing into positive among all positive patients, so \( Se = TP/(TP + FN) \). \( Sp \) is proportion of incorrectly diagnosing into negative among all negative patients, so \( Sp = TN/(TN + FP) \).

F-measure is defined as:
\[
F_F = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}
\]
where \( \beta \) is a parameter, \( P \) is precision rate and it is defined as \( P = \frac{TP}{TP + FP} \), \( R \) is recall rate, and its value equals to \( Se \). Take \( \beta = 1 \), and we get \( F_1 = \frac{2PR}{P + R} \). Kappa coefficient can verify consistency and can be used to measure classification accuracy. It is defined as:
\[
K = \frac{p_0 - p_e}{1 - p_e}
\]
where \( p_0 \) is Acc. Suppose that the number of real samples in each class is \( a_1, a_2 \) respectively, and the number of samples predicted in each class is \( b_1, b_2 \). The total number of samples is \( n \), and \( p_e = \frac{a_1b_1 + a_2b_2}{n^2} \).

ROC curve has the true positive rate (\( Se \)) as the ordinate and the false positive rate (1-\( Sp \)) as the abscissa. AUC is defined as the area under the ROC curve.

3. Results
We equally split the 1,517 test ECG episodes into ten groups. The recording numbers for the ten groups are 151, 151, 151, 152, 152, 152, 152, 152, 152, 152, and an overall AUC of 97.88% was achieved, with less than 1% inter-group variability.

<table>
<thead>
<tr>
<th>Fold</th>
<th>TP</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
<th>( Se ) (%)</th>
<th>( Sp ) (%)</th>
<th>( K ) (%)</th>
<th>( F_1 ) (%)</th>
<th>Acc (%)</th>
<th>AUC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>72</td>
<td>76</td>
<td>2</td>
<td>1</td>
<td>98.63</td>
<td>97.44</td>
<td>96.02</td>
<td>97.96</td>
<td>98.01</td>
<td>98.03</td>
</tr>
<tr>
<td>2</td>
<td>77</td>
<td>72</td>
<td>1</td>
<td>1</td>
<td>98.72</td>
<td>98.63</td>
<td>97.36</td>
<td>98.72</td>
<td>98.68</td>
<td>98.67</td>
</tr>
<tr>
<td>3</td>
<td>73</td>
<td>73</td>
<td>2</td>
<td>3</td>
<td>96.05</td>
<td>97.33</td>
<td>93.38</td>
<td>96.69</td>
<td>96.69</td>
<td>96.69</td>
</tr>
<tr>
<td>4</td>
<td>81</td>
<td>67</td>
<td>3</td>
<td>1</td>
<td>98.78</td>
<td>95.71</td>
<td>94.70</td>
<td>97.59</td>
<td>97.37</td>
<td>97.25</td>
</tr>
<tr>
<td>5</td>
<td>85</td>
<td>66</td>
<td>1</td>
<td>0</td>
<td>100.0</td>
<td>98.51</td>
<td>98.66</td>
<td>99.42</td>
<td>99.34</td>
<td>99.25</td>
</tr>
<tr>
<td>6</td>
<td>74</td>
<td>73</td>
<td>3</td>
<td>2</td>
<td>97.37</td>
<td>96.05</td>
<td>93.42</td>
<td>96.73</td>
<td>96.71</td>
<td>96.71</td>
</tr>
<tr>
<td>7</td>
<td>81</td>
<td>68</td>
<td>3</td>
<td>0</td>
<td>100.0</td>
<td>95.77</td>
<td>96.03</td>
<td>98.18</td>
<td>98.03</td>
<td>97.89</td>
</tr>
<tr>
<td>8</td>
<td>72</td>
<td>77</td>
<td>3</td>
<td>0</td>
<td>100.0</td>
<td>96.25</td>
<td>96.06</td>
<td>97.96</td>
<td>98.03</td>
<td>98.13</td>
</tr>
<tr>
<td>9</td>
<td>73</td>
<td>76</td>
<td>3</td>
<td>0</td>
<td>100.0</td>
<td>96.20</td>
<td>96.06</td>
<td>97.99</td>
<td>98.03</td>
<td>98.10</td>
</tr>
<tr>
<td>10</td>
<td>76</td>
<td>73</td>
<td>0</td>
<td>3</td>
<td>96.20</td>
<td>100</td>
<td>96.06</td>
<td>98.06</td>
<td>98.03</td>
<td>98.10</td>
</tr>
</tbody>
</table>

**Total** 764 721 21 11 98.58± 97.17± 95.78± 97.95± 97.89± 97.88± 1.47 1.38 1.54 0.78 0.77 0.76

Table 2. CNN specifications designed for the ECG classification problem.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning rate</td>
<td>0.005</td>
</tr>
<tr>
<td>Convolutional layer kernel size</td>
<td>3×3</td>
</tr>
<tr>
<td>No. of feature maps in the first convolutional</td>
<td>16</td>
</tr>
<tr>
<td>No. of feature maps in the second convolutional</td>
<td>32</td>
</tr>
<tr>
<td>No. of feature maps in the third convolutional</td>
<td>64</td>
</tr>
<tr>
<td>No. of feature maps in the fourth convolutional</td>
<td>128</td>
</tr>
<tr>
<td>No. of feature maps in the fifth convolutional</td>
<td>256</td>
</tr>
<tr>
<td>Leave probability of dropout layer</td>
<td>0.5</td>
</tr>
<tr>
<td>Max pooling layer kernel size</td>
<td>2×2</td>
</tr>
<tr>
<td>No. of neurons in the first fully connected layer</td>
<td>256</td>
</tr>
<tr>
<td>No. of neurons in the second fully connected layer</td>
<td>2</td>
</tr>
<tr>
<td>No. of epoch</td>
<td>50</td>
</tr>
<tr>
<td>Size of mini-batch</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 3. Results from the evaluation metrics

[Diagram of the network architecture]

Figure 2. The architecture of the network.
4. Discussion

This study proposes an innovative framework for PVC detection based on MFSWT time-frequency representation and CNN classifier, which can accurately identify PVC ECG segments from the wearable big data ECGs. We build a unique CNN architecture to train the classification model that uses the time-frequency image generated by MFSWT method as input. The test data recorded from 12-lead Lenovo Smart ECG vest is evaluated by the trained model and obtain an accuracy of 97.89%, indicating that the method has clinical significance. Xu et al. [15] applied this MFSWT method in the detection of atrial fibrillation, and enhanced the AF detection accuracy up to 84.85%.

Yang et al. combined clinical diagnostic criteria with image processing methods, extracted heart rhythm parameters and QRS complex morphological features, and then used SVM to perform heart beat classification. The MIT-BIH database was used for training and verification and the PVC recognition rate was 95.31% [16]. It is noted that performance of the classifier relied too much on the quality of feature extraction. Since the morphologies of PVC beats can vary enormously from person to person, the model may suffer from overfitting and non-universality. A CNN method can eliminate the feature design and extraction process required in this kind of approaches.

In addition, considering that PVC is a dynamic process, real-time monitoring is important. Our work also has obvious advantage in the running time of the algorithm. Li et al. proposed a method to automatically discriminate PVC beats from other beats and artifacts with the use of wavelet transform and CNN. Ten-fold cross validation results on MIT-BIH showed that this algorithm achieved a highest overall accuracy of 97.96% using Paul wavelet [17]. However, it has been confirmed that MFSWT can better capture the tiny changes in the frequency domain than CWT [13]. Moreover, the ECG signal data used in our study were all from wearable ECG monitoring equipment so as to ensure clinical applicability. The next step of research should focus on expanding the types of diseases detected based on ensuring accuracy.

Acknowledgement

The study was partly supported by the National Natural Science Foundation of China (81871444), The authors thank the support from the Southeast-Lenovo Wearable Intelligent Monitoring Lab.

References


Address for correspondence.

Chengyu Liu, PhD
School of Instrument Science and Engineering
Southeast University, China
E-mail: chengyu@seu.edu.cn