

An End-to-end Deep Learning Scheme for Atrial Fibrillation Detection

Yingjie Jia*, Haoyu Jiang, Ping Yang, Xianliang He

Shenzhen Mindray Bio-Medical Electronics Co Ltd, Shenzhen, China

Abstract

The aim of this study was the detection of atrial fibrillation (AF) from continuous ECG analysis. In this study, an end-to-end deep learning scheme was proposed. When the scheme was applied, 30-second multi-lead ECG data segments with an overlapping window of 5 seconds were preprocessed and sequentially fed into a multi-layer residual convolutional neural network (CNN) to extract ECG's multi-scale local morphological (spatial) features. The generated local spatial features were then processed by the following two bidirectional long short-term memory (LSTM) layers, and the output sequences of the LSTM layers were then weighted by an attention module and processed by a following dense network to complete AF detection. Finally, the sequential detection results were further processed to improve accuracy. To demonstrate its effectiveness, the proposed scheme was trained and tested on multiple ECG databases annotated by cardiologists. Episode and duration accuracies were calculated according to the performance evaluation method of atrial fibrillation detection defined in the EC57 standard [1]. An episode F1 score of 85.7% and a duration F1 score of 95.5% were achieved on the independent testing dataset.

1. Introduction

Atrial fibrillation (AF) is a common cardiac arrhythmia with a prevalence of 2% in the adult population [2], and is considered to be associated with ischemic strokes, heart failure and many other serious complications [3]. Obviously, detection of AF is crucial for early treatment. As a no-invasive method, electrocardiography (ECG) is widely applied in AF detection. In practice, AF detection is done by detecting the absence of P wave or the presence of atrial fibrillation waves (f-waves) and the irregular variability of RR intervals in the ECG trace. However, significant feature engineering efforts are always needed to generate handcraft features or empirical thresholds due to the environment interferences, low amplitudes of the P waves and the intra-patient differences. Recently, deep learning models have been popularly applied for AF detection, including convolutional network (CNN), long short-term memory

network (LSTM). Besides, attention mechanism, which is effective in nature language processing (NLP) has also been introduced. For example, a 34-layer CNN with residual network architecture was proposed for detection of 14 cardiac arrhythmias including AF [4], but only deep residual CNN module was utilized in this model, and thus temporal relationships between different parts of each input ECG segment was not utilized. Another model composed of CNN module, two LSTM layers and attention module was reported to classify 8 types of arrhythmias and sinus rhythm, and an average F1-score of 81.2% was achieved [5], but in this model, the CNN module was similar to VGGNet with only 13 CNN layers and the LSTM layers were unidirectional, and thus could only extract temporal features from forward direction. In [6], an attention based deep learning scheme for detection of paroxysmal AF episodes was proposed. In the scheme of [6], non-overlapping 30-second windows split from each 10-minute ECG data were sequentially processed by the wavelet transform to generate their spectrograms, and by a CNN to further extract features from the spectrograms. The features extracted by the CNN were sequentially fed into a bidirectional recurrent neural network (BRNN), whose output results were weighted by a soft attention mechanism to generate the deep learning features for each 10-minute ECG data. Finally, the deep learning features were concatenated with other handcraft features and processed by a softmax regression layer to get the AF detection result. One disadvantage of the scheme in [6] was that it needed too long ECG data to analysis, besides, both wavelets transform and feature engineering were needed.

In this article, an end-to-end deep learning scheme for AF detection was proposed. The proposed scheme composed of a 20-layer residual CNN to automatically extract spatial features, and a following bidirectional LSTM layers to extract temporal features from both forward and backward directions. The features extracted by the residual CNN and the LSTM were concatenated together at every time step to generate spatial-temporal features, which were weighted by an attention module to get the final deep learning feature, and the feature were fed into a softmax regression layer to complete the AF detection. The proposed scheme only needed 30 seconds ECG data without any handcraft features for every AF detection, but it could also be applied for continuous ECG

analysis due to the following simple post-processing module. The scheme was trained on multi-lead 30-second ECG segments, but evaluated on the long-time records from a testing dataset.

2. Method

An overview of the proposed scheme for AF detection was illustrated in Figure 1, and the continuous ECG data was processed by the following stages.

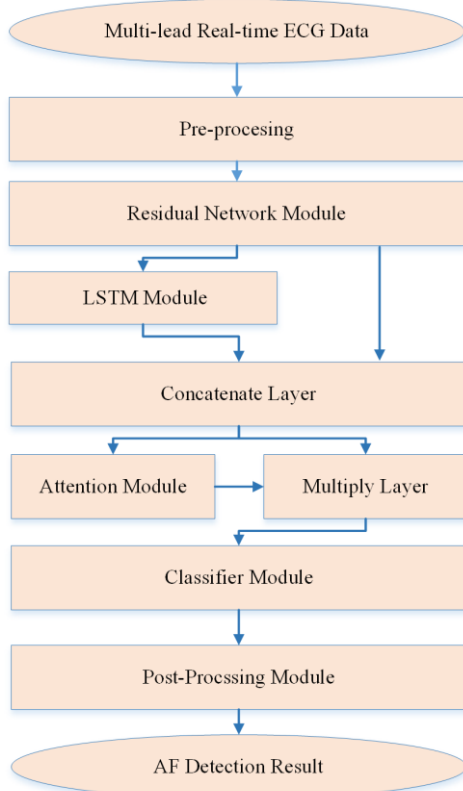


Figure 1. Proposed scheme for atrial fibrillation detection

2.1. Data pre-processing

Multi-lead continuous ECG data was sampled at 250Hz, and filtered by IIR filters with a bandwidth of 0.5-40 Hz. Then the continuous ECG data was sliced into 30-second ECG segments with a step of 5 seconds, and standardized segment-wise, so that each segment had a mean of zero and a standard deviation of one.

2.2. Local spatial features extraction

Constructed by 20 CNN layers with shortcuts, the residual network module was developed to automatically extract spatial features from each 30-second ECG data segment. Residual network was chosen for its ability to generate complex features with deep CNN layers while

avoiding vanishing or exploding of gradients. All of the referred 20 CNN layers had a kernel size of 5 with a stride step of 3 or a kernel size of 3 with a stride step of 1, and were initialized by 'He-Initializer' method [7] and padded with the padding strategy of 'same'. All pooling size and stride step were 3. After processing by the residual network module, the sequence size was reduced by 243 times. Batch normalization (BN) layers and dropout layers with a rate of 0.2 were used to avoid over fitting. Further details were illustrated in Figure 2.

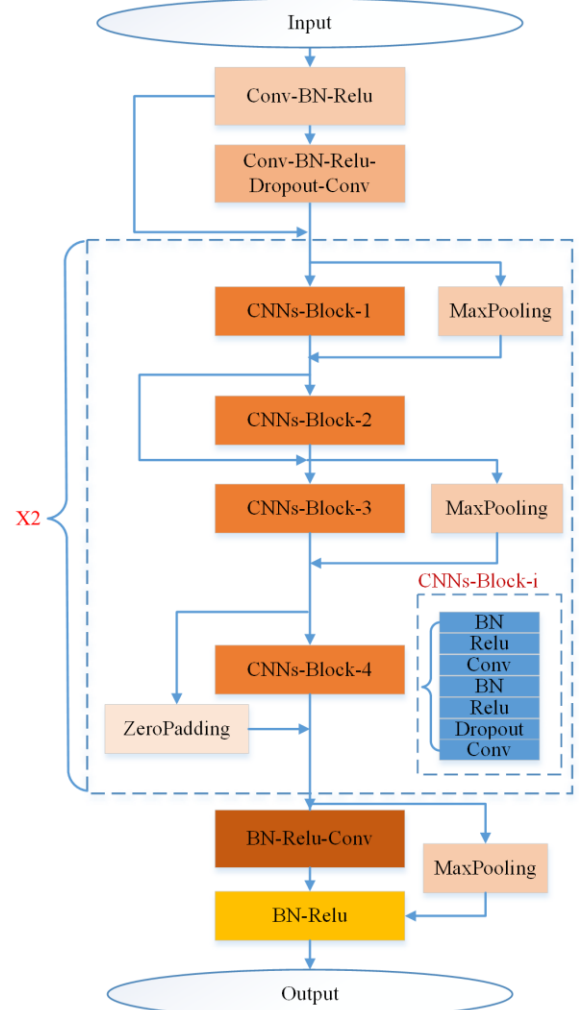


Figure 2. Details of residual network module

2.3. Temporal features extraction

The LSTM module consisted of 2 sequential bidirectional LSTM layers, in which the numbers of the hidden units and output units were both 32. This module was used to extract the temporal information from both forward and backward directions.

2.4. Combining spatial and temporal features

For every time-step of the LSTM module, the input and output were simply concatenated together along the feature dimension, and thus the sequences consisting of spatial features extracted by residual CNN and temporal features extracted by bidirectional LSTM layers were generated.

2.5. Calculation of the representative feature vectors

The attention module, as illustrated in Figure 3, was mainly constructed by 2 sequential dense layers each followed by a nonlinear activation function. Through the attention module, the sequences of spatial-temporal features were weighted to form a representative feature vector for each ECG segment.

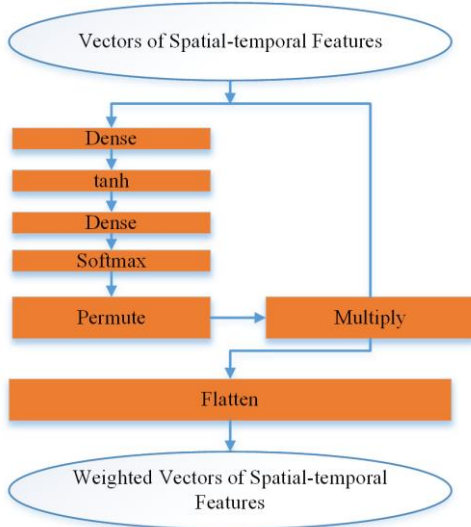


Figure 3 Details of attention module

2.6. Classification

The classifier module was constructed by a dense network followed by the softmax to calculate the likelihood of the AF for each 30-second ECG data segment.

2.7. Post-processing

As in our proposed scheme, segments extracted from continuous ECG data were processed sequentially and separately, and thus the AF detection results of all segments were theoretically independent with each other. But in practice, the AF detections may not be stably correct during long AF/Non-AF episodes and isolating

false negative and positive AF detections may occur due to environment interferences. In order to further improve the accuracy, the post-processing part was needed.

In the post-processing part, AF episodes shorter than 30 seconds were ignored, and AF episodes separated by intervals shorter than 5 seconds were merged to a single one

3. Results

3.1. Database and pre-processing

Four expert annotated databases including MIT-BIH Arrhythmia Database (MIT-BIH), MIT-BIH Atrial Fibrillation Database (MIT-AF), Long-term AF Database (LTAF) and AHA Database (AHA) were used to train and test the model in the proposed scheme.

Since AHA database only contains 80 records (each with a duration of 3 hours) of 8 arrhythmia types excluding AF, so it was used as database of negative instances. Besides, 4 paced records from MIT-BIH were excluded. A brief summary of the used records was listed in Table 1.

All the used records were resampled at the frequency of 250Hz and filtered by 1-order low-pass Butterworth filter with a cut-off frequency of 40 Hz, 1-order high-pass Butterworth filter with a cut-off frequency of 0.5 Hz and a notch filter of 50Hz.

Table 1. Summary of used records

	Rhythm	MIT-BIH	MIT-AF	LTAF	AHA
Records	AF	7	23	83	0
	Non-AF	37	0	1	80
Hours	AF	2.2	93.4	1030.9	0
	Non-AF	19.9	140.9	929.7	46.7

3.2. Experimental setup

Records of each database were divided into training (2/3 records) and testing part (1/3 records), then the four training and testing parts were separately mixed to construct two datasets named as DB1 for training and DB2 for testing. 30-second segments were extracted with a stride of 5 seconds from the records of DB1 to construct the final training dataset. Each extracted segment was labeled as AF only if it contained AF for more than 15 seconds, and otherwise Non-AF. Since the lengths of records varied from 30 minutes to more than 24 hours, in order to balance the segments extracted from each record, data augmentation was performed. Meanwhile, DB2 was used as an independent testing database of continuous ECG data.

The weights of CNN and dense were initialized using the 'He initializer' while the LSTM cells were initialized with orthogonal initializer. Cross entropy loss and Adam

optimizer with default parameters was applied. The initial learning rate was 0.001 and would be multiplied by a factor of 0.1 every 10 epochs while the maximum training epoch was set as 100.

3.3 Evaluation results

The performance of the proposed scheme for AF detection was evaluated on the testing dataset DB2 according to EC57 standard [1].

Firstly, annotations for each record were derived from the sequential results of AF detection, and then were compared with reference annotations to calculate three metrics of precision, recall and F1-score both in terms of episode and duration. When recall was calculated, reference AF episodes or durations overlapped with algorithm ones were counted as true positives (TP) while others were counted as false negatives (FN), and the recall was the ratio of true positives to the sum of true positives and false negatives ($TP / (TP + FN)$). When precision was calculated, algorithm AF episodes or duration overlapped with reference ones were counted as true positives (TP) while others were counted as false positives (FP), and the result was the ratio of true positives to the sum of true and false positives ($TP / (TP + FP)$). The F1-score was calculated as the harmonic average of recall and precision. The final evaluation results were listed in Table 2.

Although the proposed scheme achieved a promising performance on the whole testing datasets, there were still AF misdetection cases. Some misdetection cases were probably due to questionable annotations. For example, as illustrated in Figure 4 (around 63164 second of the record '10.dat', LTAF), considering the presence of little but regular P waves, the scheme possibly made a true negative AF detection.

Besides, the proposed scheme didn't work well on several other ECG segments. For example, as illustrated in Figure 5 (around 2159 second of the record '23.dat', LTAF), a Non-AF was incorrectly detected as AF. This misdetection was probably due to the artifacts similar to f waves, though the RR intervals were almost regular. This false positive AF detection could be improved by incorporating extra artifacts handling strategies.

Table 2. Performance of AF detection on records of testing databases

	Precision	Recall	F1-score
Episode	79.9%	92.5%	85.7%
Duration	94.9%	96.0%	95.5%

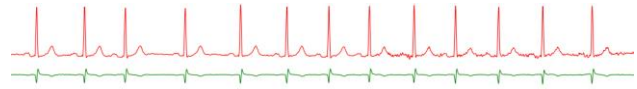


Figure 4 Segment annotated as AF but detected as non-AF

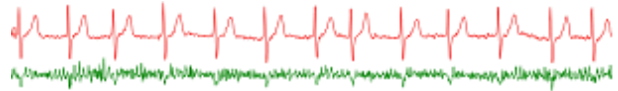


Figure 5 Segment annotated as no-AF but detected as AF

4. Conclusion

An end-to-end deep learning scheme composed of a residual network module, a LSTM module, an attention module and a post-processing module was proposed for AF detection by continuous ECG analysis, and the proposed scheme achieved an episode F1 score of 85.7% and a duration F1 score of 95.5% on the testing dataset of continuous ECG records. The experiment results showed that the proposed scheme could be a valuable method for AF detection.

References

- [1] ANSI/AAMI EC57: "Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms", 2012
- [2] G. Y. H. Lip et al., "Atrial fibrillation", *Nature Reviews Disease Primers*, vol.2, 2016
- [3] A. J. Camm et al., "Guidelines for the management of atrial fibrillation: The Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC)", *European Heart Journal* vol.31, no. 19, p.2369-2429, 2010
- [4] A.Y. Hannun et al., "Cardiologist-level arrhythmia detection with convolutional neural networks". *Nature Medicine*, vol.25, pp.65-69, 2019
- [5] Q. Yao et al., "Multi-class arrhythmia detection from 12-lead varied-length ECG using attention-based time-incremental convolutional neural network". *Information Fusion*, vol. 53, pp.174-182, 2020
- [6] S. P. Shashikumar et al., "Detection of paroxysmal atrial fibrillation using attention-based bidirectional recurrent neural networks", *the 24th ACM SIGKDD International Conference*, pp.715-723, 2018
- [7] K. He et al., "Delving deep into rectifiers: surpassing human-level performance on ImageNet classification", *CVPR IEEE Computer Society*, 2015

Address for correspondence:

Yingjie Jia
Mindray Building
Keji 12th Road South, High-Tech Industrial Park, Nanshan,
Shenzhen, China
jiayingjie@mindray.com