Biomedical Signal Processing and Control 56 (2020) 101675

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

A novel data augmentation method to enhance deep neural networks for detection of atrial fibrillation

Ping Cao [a,d], Xinyi Li [a], Kedong Mao [a], Fei Lu [a], Gangmin Ning [b,c], Luping Fang [a], Qing Pan [a,*]

[a] College of Information Engineering, Zhejiang University of Technology, Liuhe Rd. 288, Hangzhou 310023, China
[b] Department of Biomedical Engineering, MOE Key Laboratory of Biomedical Engineering, Zhejiang University, 38 Zheda Road, Hangzhou 310027, China
[c] Zhejiang Lab, 1818 Wenyi Road, 310000, China
[d] Zhijiang College, Zhejiang University of Technology, Shaoxing 312030, China

ARTICLE INFO

Article history:
Received 7 January 2019
Received in revised form 6 July 2019
Accepted 25 September 2019

Keywords:
Atrial fibrillation
Electrocardiogram
Deep neural networks
Data augmentation

ABSTRACT

Automated detection of atrial fibrillation (AF) from electrocardiogram (ECG) recordings remains challenging in real clinical settings. Deep neural networks (DNN) emerge as a promising tool for the task of AF detection. However, the success of DNN for AF detection is hampered by limited size and imbalanced number of samples in datasets. We propose a novel data augmentation strategy based on duplication, concatenation and resampling of ECG episodes to balance the number of samples among different categories as well as to increase the diversity of samples. The performance of the data augmentation method was examined on an AF database from Computing in Cardiology (CinC) challenge 2017. A 2-layer long short-term memory (LSTM) network was trained with the augmented dataset. Its ability of AF detection was evaluated using a 10-fold cross validation approach. And F1 score was adopted as the metrics. The AF detection results show that the proposed method was superior to two conventional data augmentation methods: window slicing and permutation. The network was also submitted to the evaluation system of the CinC challenge 2017. The F1 score obtained by the network using the proposed data augmentation method was close to the winner (0.82 vs. 0.83). In summary, the proposed data augmentation method provides an effective solution to enhance the dataset for improving the performance of DNN in ECG analysis. Such a method promotes the application of deep learning in the analysis of ECG, particularly when the dataset is small and imbalanced.

1. Introduction

Atrial fibrillation (AF) is one of the most common arrhythmias with a prevalence of 0.4–1% [1–3]. Though AF itself does not represent a lethal condition, it is associated with high risks of morbidity and mortality due to AF-related complications [4], such as stroke, heart failure and other cardiac related diseases [5,6]. As AF happens unpredictably and requires timely treatment, more and more wearable devices have been developed for long-term monitoring the electrocardiogram (ECG) of the subjects [7]. The large number of recordings obtained with these wearable devices can hardly be inspected by cardiologists. Therefore, algorithms for automated detection of AF in real time are highly necessary to reduce the AF-related mortality and morbidity.

The presence of AF is characterized by two remarkable features in the ECG signals [8]: (1) P waves disappear and are replaced by rapid irregular atrial pulsating waves called fibrillatory waves; (2) R-R intervals are irregular. Various conventional algorithms for automated detection of AF have been developed relying on detecting the absence of P-waves [9–11] or irregularity of R-R intervals [12–15], or the combination of both [16–18]. In recent years, deep learning emerges as a novel approach for AF detection. It allows recognizing the presence of AF without extracting the morphological features in the ECG signal. Xu et al. [19] proposed a framework which combines modified frequency slice wavelet transform (MFSWT) and convolutional neural network (CNN) to identify AF from short-term episode of ECG and got a high performance on the Massachusetts Institute of Technology - Beth Israel Hospital (MIT-BIH) AF dataset. Using a CNN, Xia et al. [20] obtained a superior performance in AF detection compared with the conventional approaches on the MIT-BIH AF dataset. Salloum et al. [21] evaluated different recurrent neural network (RNN) architectures on two public datasets: ECG-Identification and MIT-BIH Arrhyth-

* Corresponding author at: College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China.
E-mail address: pqpq@zjut.edu.cn (Q. Pan).

https://doi.org/10.1016/j.bspc.2019.101675
1746-8094/

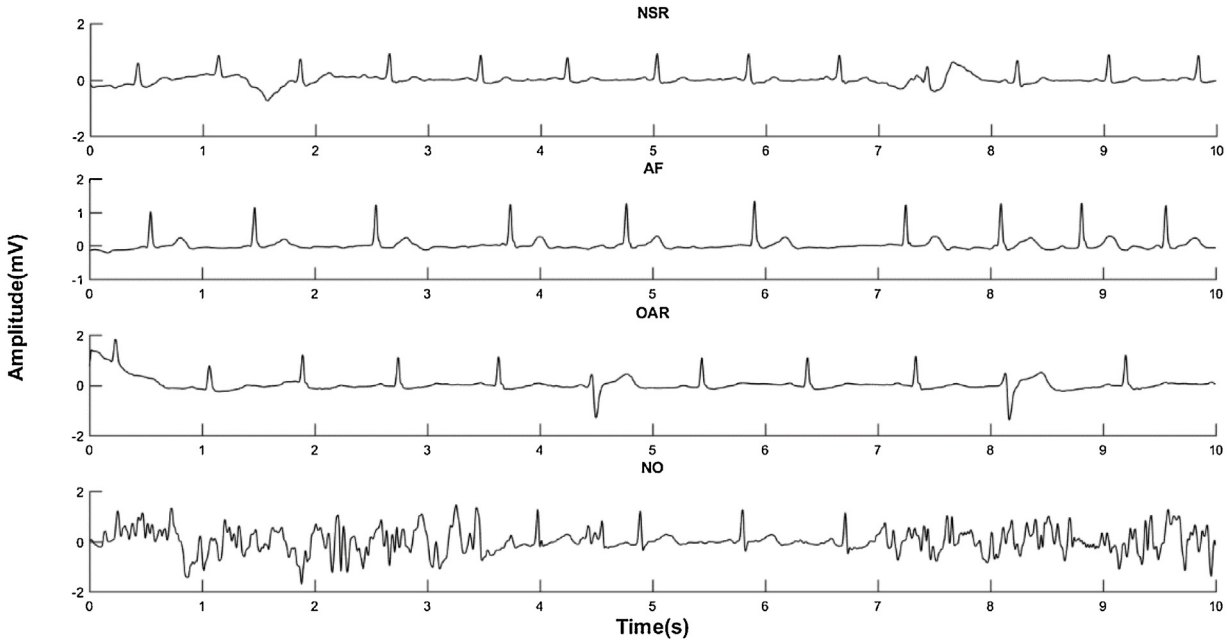

Fig. 1. Representative ECG waveforms of four classes in the CinC challenge 2017 dataset. From top to bottom, they are normal sinus rhythm (NSR), atrial fibrillation (AF), other abnormal rhythm (OAR) and noise (NO).

mia, and concluded that RNN is a suitable architecture for the task of AF detection from ECG recordings. In an AF classification challenge held by Computing in Cardiology (CinC) conference, the algorithm designed by Teijeiro et al., using a deep learning model and a gradient boost decision tree, outperforms many conventional approaches and ranks first with an F1 score of 0.83 [22]. These results indicate that deep learning may be promising for enhancing the performance of AF detection.

The success of a deep learning model is highly dependent on the availability of rich dataset [23]. On one hand, training a deep neural network with insufficient data may result in non-convergence of the network or poor performance. On the other hand, imbalanced dataset will lead to biased classification results. The deep learning based ECG analysis suffers from the problems in these two aspects. The available ECG recordings annotated by cardiologists are limited because only experienced professionals can annotate the recordings accurately and the task is very demanding. In addition, the abnormal recordings are usually much fewer than the normal ones due to the low incidence rate of the abnormal cardiac events. To guarantee the performance, the development of a deep learning model is often accompanied with a data augmentation procedure, which generates extra and potentially non-redundant training data by applying one or several deformations on the training set [24]. Plenty of data augmentation algorithms have been proposed for enhancing the performance of deep learning models. For instance, principal component analysis (PCA) color jittering [25] and radial transform [26] have achieved remarkable effects on the two-dimensional (2D) images. For one-dimensional (1D) time series, effective data augmentation strategies include window slicing (WS), window warping (WW), permutation and scaling [27,28]. WS extracts multiple subsegments from the original time series to generate new samples [27]. Each subsegment is assigned the same label as the original sample. WW deforms a randomly selected segment in a time series by lengthening or shortening it to generate a new sample [27]. Permutation divides a time series into several segments and permutes them to form a new sample [28]. Scaling modifies the amplitude of the time series randomly to produce new samples [28]. However, to our knowledge, a data augmentation particularly suitable for the 1D ECG signal is lacking.

Table 1
Detailed information of the dataset. NSR: Normal Sinus Rhythm. AF: atrial fibrillation. OAR: other abnormal rhythms. NO: too noisy to classify.

Type	Recording#	Time length (s)				
		Mean	SD	Max	Median	Min
NSR	5154	31.9	10.0	61.0	30	9.0
AF	771	31.6	12.5	60	30	10.0
OAR	2557	34.1	11.8	60.9	30	9.1
NO	46	27.1	9.0	60	30	10.2
Total	8528	32.5	10.9	61.0	30	9.0

The main contribution of this study is the proposal of a novel data augmentation strategy specifically designed for deep learning analysis of ECG. It is based on duplication, concatenation and resampling of ECG episodes. The strategy can increase the diversity of samples as well as balance the number of samples in each category, which facilitates the deep learning models in extracting the features from the dataset. We examined the effect of the data augmentation method on AF classification based on a public dataset from the CinC challenge 2017 by training and testing a 2-layer long short-term memory (LSTM) network, which is a widely used RNN architecture.

2. Materials and methods

2.1. Dataset

The ECG dataset used in the study was obtained from the CinC challenge 2017, which focuses on the classification of the ECG signals into four types: (1) Normal Sinus Rhythm (NSR), (2) AF, (3) Other Abnormal Rhythms (OAR), and (4) Noise (NO, too noisy to classify). The OAR class contains episodes with premature ventricular contraction (PVC), premature atrial contraction (PAC) or other abnormal beats. This dataset contains 8528 single-lead ECG recordings with different lengths (9 s – 61 s). The labels were given for each recording instead of for each beat. All the recordings, collected using the AliveCor device (AliveCor Inc, United States), were sampled at 300 Hz. Details of the dataset is summarized in Table 1.

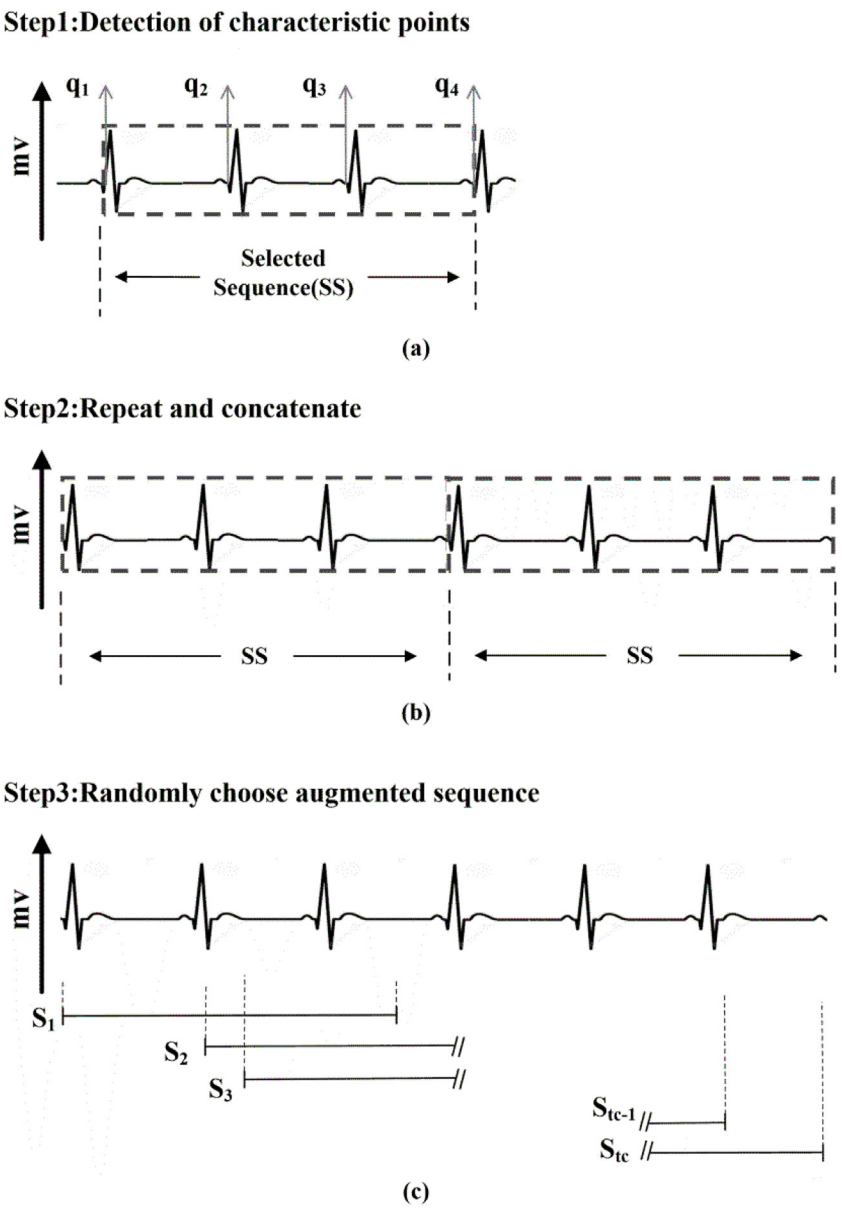

Fig. 2. Schematic diagram of the data augmentation strategy. (a) Step 1. Starting points of QRS complexes are detected and denoted as $q_1, q_2, q_3, \ldots q_n$. The sequence between q_1 and q_n are chosen as the selected sequence (SS). (b) Step 2. SS are concatenated with itself. (c) Step 3. The segment obtained in Step 2 was resampled with a random sliding window of size l_w for repeating t_c times and the resampled segments were identified as $S_1, S_2, S_3, \ldots S_{t_c}$.

The challenge also prepares a hidden test dataset with 3658 recordings for evaluating the performance of the network. The test dataset is not open to public and the evaluation can only be conducted by submitting a testing script to the system. Examples of 10 s excerpts of the ECG recordings representing each class respectively are shown in Fig. 1.

2.2. Pre-processing

First, all the ECG recordings were denoised by a wavelet-based denoising algorithm according to previously reported literatures [29,30]. A 7-level wavelet decomposition based on Coiflets wavelet basis was applied to the ECG signal. For each level, a soft threshold was selected to filter the detail coefficient. The signal was reconstructed from the filtered detail coefficients [29]. Second, in order to remove the baseline interference, the ECG signal was decomposed with a 6-level Daubechies wavelet basis and reconstructed after removing the coefficient of the 6^{th} level [31]. Finally, to address the problem of amplitude scaling, Z-score normalization was applied to each recording according to

$$z(t) = \frac{x(t) - \mu}{\sigma} \quad (1)$$

where $x(t)$ is the raw signal amplitude at the sampling time t, μ is the mean amplitude of the signal and σ is the standard deviation of the amplitude of the signal.

2.3. Data augmentation strategies

We proposed a data augmentation strategy aiming to increase the diversity of the samples and balance the number of samples among the classes. The strategy is based on concatenation of ECG episodes according to the characteristic points. The diagram of the strategy is shown in Fig. 2. First, for each heartbeat, the Pan-Tompkins QRS detector [32] was used to locate the QRS complex. Then the initial downward deflection in the QRS complex (starting

point of the QRS) was determined as the characteristic point. All the starting points detected in the episode were denoted as q_1, q_2, q_3, ... q_n. The sequence between q_1 and q_n was chosen as the selected sequence (SS), as shown in Fig. 2(a). This step was only applied to the NSR, AF and OAR classes but not to the NO class because its samples are too noisy for the QRS detection. Instead, each entire sample in the NO class was chosen as the SS. For the OAR class, we visually inspected all the samples and found that most of the beats at the beginning or end of the episodes are normal. As a result, the data augmentation will not be compromised by the abnormal beats in the OAR class since the concatenation is mostly based on the joint of two normal beats.

Second, the SS was duplicated and the original SS was concatenated to the duplicated one (Fig. 2(b)). Finally, we used a sliding window of size l_w to resample the concatenated ECG sequence for t_c times at random locations (Fig. 2(c)).

The number of samples among different categories can be balanced by setting t_c appropriately. Suppose that there are N_c samples per category (c corresponds to NSR, AF, OAR and NO). $N_{max} = \max(N_{NSR}, N_{AF}, N_{OAR}, N_{NO})$. To balance the number of samples for the four categories, t_c for each recording of class C can be calculated as:

$$t_c = \left\lceil \frac{N_{max}}{N_c} \right\rceil \tag{2}$$

where [x] indicates the rounding value of x. After this operation, the sample numbers for each class can be balanced to about N_{max}.

In addition, the number of samples in the dataset can be increased. Assuming that the target number of samples for each class is M ($M = \lambda N_{max}$), the Eq. (2) can be modified to:

$$t_c = \left\lceil \frac{M}{N_c} \right\rceil = \left\lceil \frac{\lambda N_{max}}{N_c} \right\rceil \tag{3}$$

The resampled segments were identified as $S_1, S_2, S_3, \ldots S_{t_c}$. To ensure that S_i ($i = 1, 2, \ldots, t_c$) have the same label with the original recording, l_w should be greater than or equal to the length of the original ECG recording. In this study, we set l_w to the same length with the original ECG recording. We examined the effects of different M values (5000, 10,000, and 15,000) on network training. The data augmentation processing was only applied to the training dataset rather than to the testing dataset.

Two conventional data augmentation methods—WS and permutation were involved for comparison, as they show superior performance to other approaches in the reported applications [27,28]. In the present study, WS was applied by randomly slicing each sample with a window and giving it the same label as the original one. The length of the window was defined as 90% of the original sample following the literature [27]. The time of slicing for each category was the same as t_c according to Eq. (3) in the proposed method. Permutation was carried out by dividing each sample into N_p subsegments with the same length, randomly permute them t_c times, and assemble the permuted subsegments to generate the new samples with the same label. N_p was defined to satisfy

$$N_p! > t_c > (N_p - 1)! \tag{4}$$

in order to guarantee that the permutation does not produce repeating samples.

2.4. Recurrent neural networks

RNN is a class of deep neural networks designed for classifying, processing and making predictions based on sequences [33]. With the help of its internal memory, it is able to explore the temporal dynamics in a sequence, thus particularly suitable for the applications such as traffic forecast [34,35], natural language processing [36,37], and biomedical signal processing [38,39]. We tested the proposed data augmentation strategy based on an LSTM network, which is a variant of traditional RNN for solving its gradient explosion and vanishing problem [40]. The structure of an LSTM is shown in Fig. 3.

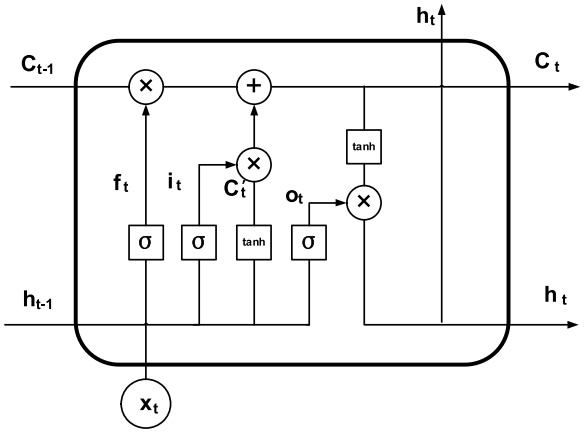

Fig. 3. The structure of an LSTM network. It has a forget gate, an input gate and an output gate within a memory cell.

An LSTM network contains structures of memory cell and gate unit [40]. Each gate unit includes a forget gate f_t, an input gate i_t and an output gate o_t. Their activation at time t can be calculated as

$$f_t = \sigma(W[x_t, h_{t-1}] + b_f) \tag{5}$$

$$i_t = \sigma(W[x_t, h_{t-1}] + b_i) \tag{6}$$

$$o_t = \sigma(W[x_t, h_{t-1}] + b_o) \tag{7}$$

where x_t is the input sequence at time t, h_{t-1} is the activation of previous time step, C_{t-1} represents the memory of the memory cell at time $t-1$, W is the weight vectors and b is the bias vector. σ of the forget gate is a sigmoid function which decides how much of the C_{t-1} will be retained in LSTM cell at time t. The input gate uses a tanh as the activation function. It decides how much of the memory cell C_t updates its activation and can be calculated by the following equation

$$C_t = f_t.C_{t-1} + i_t.\tanh(W[x_t, h_{t-1}] + b_c) \tag{8}$$

Finally, the activated value of the output gate can be calculated as

$$h_t = \tanh(C_t).o_t \tag{9}$$

In this study, a 2-layer LSTM network was developed for the task of AF classification. The structure and parameters of the proposed network are shown in Fig. 4.

2.5. Network training and optimization

A 10-fold cross-validation strategy was used to evaluate the performance of the network. Ninety percent of the whole dataset were used for training and the rest for testing. Thirty percent of the training data was used as a validation set to verify the performance of the network after every training epoch. Fig. 5 shows the number of ECG segments used for training, validation and testing procedures.

As required by the LSTM network, zeros were padded at the rear of each segment to ensure that input data has the same dimension. Accordingly, a masking layer was used as the first layer to mask the padded zeros and the time steps of these zeros would be skipped at all downstream layers of the network. In this way, the ECG segments with variable lengths could be processed by the network.

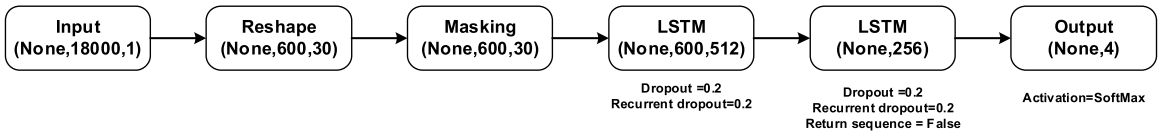

Fig. 4. The structure and parameters of the 2-layer LSTM network.

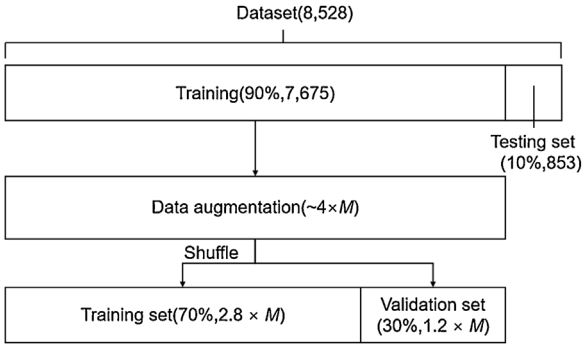

Fig. 5. The number of ECG segments used for training and testing procedures. The dataset contains 8528 single-lead ECG recordings. We used 10% of dataset as testing set. The rest of dataset were used to train the network after data augmentation. There were M samples in each category after data augmentation. 70% of these samples were used as training set and the other were used as validation set.

Categorical cross-entropy was selected as the loss function. A training optimization strategy and an early stopping criterion were applied during the training phase. We used both the Adam [41] and stochastic gradient descent (SGD) [42] optimizers for the training process. The training was started by using the Adam optimizer with a learning rate of 0.001. If the validation loss did not decease for 5 consecutive epochs, the SGD replaced the Adam for further optimization. The training would be terminated if the loss of validation did not decrease for 10 consecutive epochs. During the training procedure, any network that generated a decreased loss of validation would be saved for testing.

Dropout [43] and recurrent dropout [44] strategies were applied during the training process to avoid overfitting. Neurons were randomly dropped out at the specified rates (shown in Fig. 4) in each LSTM layer to make the network less prone to overfitting in the training phase.

2.6. Performance evaluation

We used the F1 score as the evaluation index following the rules of the CinC challenge 2017 [45]. F1 score of each class is calculated by counting the total number of true positives (TP), false negatives (FN) and false positives (FP) to evaluate the performance of the algorithms. The F1 score for class C is given as:

$$F_{1c} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{10}$$

The multiclass F1 score is the weighted average of the F1 score for the three classes "NSR", "AF" and "OAR" following the rule of the CinC challenge 2017.

$$F_1 = \frac{F_{1N} + F_{1A} + F_{1O}}{3} \tag{11}$$

It has to be noticed that even though the F1 score of the noisy class is not included in the calculation, misclassification of noisy samples into other classes will still influence the final score.

We also calculated the accuracy of the network. The accuracy (Acc) for each class is calculated as:

$$Acc = \frac{TP_{all}}{TP_{all} + FN_{all}} \tag{12}$$

Table 2
The F1 scores and accuracies of the 10-fold cross validation using the 2-layer LSTM network with and without data augmentation. In the test without data augmentation, the data were balanced by randomly removing samples from the prominent categories. The NO class was excluded in the balancing. Acc: accuracy.

	With Data Augmentation	Without Data ugmentation
F_{1N}	0.860 ± 0.028	0.380 ± 0.075
F_{1A}	0.754 ± 0.029	0.308 ± 0.120
F_{1O}	0.677 ± 0.035	0.380 ± 0.066
F_1	0.764 ± 0.028	0.356 ± 0.046
Acc (%)	78.347 ± 1.457	36.356 ± 3.095

The performance of the proposed data augmentation was first validated by comparing with the results obtained without data augmentation. Because the original dataset is highly imbalanced, which may lead to biased classification results, the number of samples was balanced to 771 (size of AF class) for each category by randomly removing samples from the prominent categories. As the NO class has too few samples and is excluded in the F1 score calculation, we ignored this type in balancing the dataset. The proposed augmentation was then compared with the two conventional data augmentation methods—WS and permutation as introduced in Section 2.3.

3. Results

3.1. Performance without data augmentation

The performance comparison of the networks with and without data augmentation is shown in Table 2. The result with data augmentation was obtained by augmenting the number of samples to 10,000 for each category. The change of loss and accuracy in the training and validation phase is shown in Fig. 6 (one fold). It is observed that the augmented dataset assists the network to obtain a satisfactory F1 score and classification accuracy. By contrast, the network using the balanced original dataset shows an inferior result. Its loss and accuracy curves indicate that the trained network cannot extrapolate to the validation dataset, probably due to the insufficient number of samples.

3.2. Performance with data augmentation

The performance comparison of the network using different data augmentation strategies is shown in Table 3. For each method, the performance increases with the elevation of M. Our proposed method gives the best F1 score for all M values. We submitted the best network ($M = 15{,}000$) to the CinC challenge 2017 during the unofficial stage. The network obtained an F1 score of 0.82 on the hidden test set with F1 scores of 0.91, 0.84 and 0.70 for NSR, AF and OAR classes, respectively. The comparison of the proposed method with other state-of-the-art methods is given in Table 4. It is observed that our 2-layer LSTM using the proposed data augmentation method achieved a competitive F1 score close to the winner.4

4. Discussion

In spite of the rapid development of deep learning models in ECG analysis, the availability of a rich and balanced ECG dataset remains

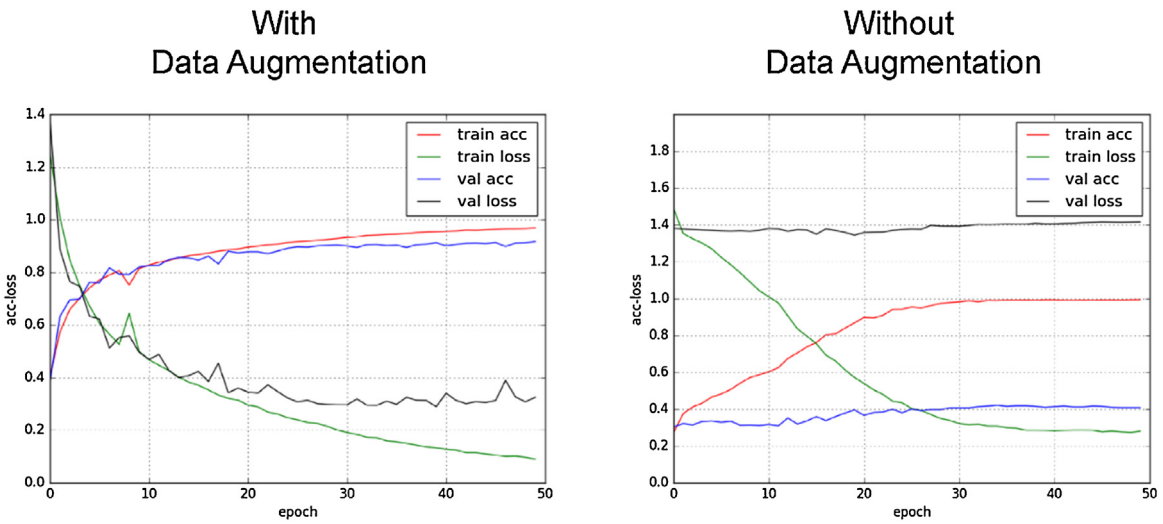

Fig. 6. Loss and accuracy curves for the networks trained with augmented dataset and the balanced original dataset (one fold in the cross validation). Because the validation loss did not decrease using the balanced original dataset, we excluded the early stopping criterion to allow observation of the training process till it reaches a stable state.

Table 3
F1 scores and accuracies of the networks using the three data augmentation methods. *M* indicates the approximate number of samples in each category. The column with the optimal *M* for each method is given in bold. Acc: accuracy.

	WS			Permutation			Our Method		
M	5000	10,000	15,000	5000	10,000	15,000	5000	10,000	15,000
F_{1N}	0.806 ± 0.015	0.835 ± 0.016	**0.846 ± 0.039**	0.799 ± 0.016	0.870 ± 0.011	**0.871 ± 0.034**	0.821 ± 0.029	0.860 ± 0.028	**0.894 ± 0.014**
F_{1A}	0.604 ± 0.042	0.658 ± 0.027	**0.691 ± 0.041**	0.581 ± 0.033	0.664 ± 0.044	**0.699 ± 0.091**	0.697 ± 0.030	0.754 ± 0.029	**0.798 ± 0.031**
F_{1O}	0.580 ± 0.051	0.628 ± 0.023	**0.651 ± 0.048**	0.500 ± 0.020	0.663 ± 0.013	**0.649 ± 0.115**	0.567 ± 0.043	0.677 ± 0.035	**0.736 ± 0.018**
F_1	0.663 ± 0.031	0.707 ± 0.019	**0.729 ± 0.039**	0.627 ± 0.018	0.732 ± 0.121	**0.740 ± 0.079**	0.695 ± 0.028	0.764 ± 0.028	**0.810 ± 0.017**
Acc (%)	70.563 ± 2.202	74.771 ± 1.811	**76.589 ± 4.155**	68.664 ± 1.728	78.652 ± 1.407	**79.086 ± 5.864**	72.165 ± 3.135	78.347 ± 1.457	**82.949 ± 1.437**

Table 4
Comparison of the proposed methods with other state-of-the-art methods. F_{1ht}: F_1 on the hidden test dataset. F_{1cv}: F_1 of cross validation.

	Data augmentation	Feature extraction	Classifier	F_{1ht}	Mean F_{1cv}
Official stage					
Teijeiro et al. [22]	L^2-regularization	rhythm features, morphological features, signal quality features	RNN, Gradient boost decision tree	0.83	0.85[a]
Hong et al. [46]	dynamic oversample	expert features, center wave features, deep features	Ensemble Classifier	0.83	–
Zabihi et al. [47]	N/A	morphological features, frequency features, nonlinear features, time-frequency features	Random Forest	0.83	0.819 ± 0.026[b]
Datta et al. [48]	binary classifiers	HRV features, frequency features,	Adaptive Boosting	0.83	0.826 ± 0.003[c]
Xiong et al. [49]	5s-segments as input and produced a prediction	deep features	CNN	0.82	0.82[c]
Zihlmann et al. [50]	dropout bursts and random resampling	deep features	CNN + LSTM	0.82	0.792[c]
Kropf et al. [51]	N/A	physiologically motivated feature	Random Forest	0.81	0.83[b]
Smole et al. [52]	N/A	inter-beat timing features, statistical features, frequency features, deep features	Gradient Boosting Machine, RNN	0.81	–
Unofficial stage					
Kamaleswaran et al. [53]	L^1↘L^2-regularization	deep features	CNN	0.83	0.842 ± 0.009[c]
Goodfellow et al. [54]	N/A	template features, RRI features, full waveform features	Xtreme Gradient Boosting	0.82	0.824 ± 0.007[d]
Liu et al. [55]	length-adaptive entropy classifier	statistical features, P wave absence, spectrum features	Support Vector Machine	0.80	–
Hannun et al. [56]	N/A	deep features	CNN	0.83	–
Ours	proposed method	deep features	2-layer LSTM	0.82	0.810 ± 0.017[b]

[a] 8-fold cross validation.
[b] 10-fold cross validation.
[c] 5-fold cross validation.
[d] 4-fold cross validation.

a challenge. Many ECG datasets suffer from insufficient samples and imbalanced number of samples among different categories. Therefore, an efficient data augmentation for ECG signal is highly necessary. We proposed a novel data augmentation strategy based on duplication, concatenation and resampling of ECG episodes to balance the number of samples among different categories as well

as to increase the diversity of the dataset. The comparison with the network using balanced original dataset verifies the necessity of data augmentation. Table 2 shows that the network trained with the augmented dataset obtained a significantly higher F1 score and accuracy compared with that using balanced but not augmented dataset. Fig. 6 shows that the network trained with balanced original dataset cannot extrapolate to the validation dataset, probably because the training dataset is too small to include the critical features for respective categories, so that the generalization ability of the network is compromised.

The two conventional data augmentation approaches obtained results inferior to the proposed one. Using the WS approach, each augmented sample is a part of the corresponding original sample, which does not introduce diversity to the augmented dataset, leading to weak ability of the network in extrapolating to the testing dataset. Also, the slicing may exclude the determinant features for the label of the sample. The permutation approach, although increases diversity of the dataset by randomly permuting subsegments, is likely to disrupt the morphology of ECG episodes so that the label for the sample may no longer be valid. By contrast, the proposed method first extended the signal by concatenating the ECG episodes based on characteristic points, ensuring that no significant artefact will intrude the signal thus the label for the resampled signal is unlikely to change. Then, the method resampled the extended signal to increase diversity to the dataset by introducing new non-redundant ECG episodes. In addition to increasing the number and diversity of samples, the data augmentation strategy also balances the number of samples among different classes. For those classes which have fewer data than the others, we may perform non-uniform resampling by controlling the parameter λ in Eq. (3). With the help of the proposed data augmentation, the 2-layer LSTM network gets an F1 score close to the winner of the CinC challenge 2017, which verifies its effectiveness.

The participants of the CinC challenge 2017 also made efforts on data augmentation (Table 4). For example, dropout bursts were applied to mimic bad contact of ECG leads [50]. It allows augmenting the NO class, which is difficult to be balanced by the conventional data augmentation approaches. Random resampling was used to emulate the heart rates unobserved in the raw dataset, which can increase the data diversity with respect to cardiac rhythm [50]. Dynamic oversampling used various strides for different classes to achieve data balance when applying WS method [46]. The groups who incorporated these data augmentation methods in suitable classifiers obtained admirable performance in AF detection. Four teams were tied for first place, with an overall score of 0.83. As listed in Table 4, Teijeiro et al. [22] used an RNN network and a gradient boost decision tree with feature extraction. Hong et al. [46] used the expert features and an ensemble classifier. While Zabihi et al. [47] used a random forest algorithm with feature extraction. Datta et al. [48] used a cascade of adaptive boosting learners. All of these algorithms take a number of handcrafted features obtained by either rule-based feature extraction algorithms or expert knowledge. The quality of the handcrafted features may influence the ability of the algorithms significantly, especially in real clinical settings with high noises and complex interferences. Our network is able to learn the suitable features from the raw data without complex extra operations for classification and the performance is comparable to their algorithms. In this sense, with the help of the proposed data augmentation, a general DNN network has a better potential to deal with the complex noise and interference in real clinical settings.

This study has several limitations. First, although the concatenation based on the characteristic points avoids the artefacts at the joint, it may disrupt the patterns of RR interval of the raw signal thus generate new samples with non-physiological rhythms. In the future study, we aim to evaluate this effect quantitatively and determine the location of concatenating point following the pattern of RR intervals. Second, the data augmentation was not tested in other classical DNN architectures, such as the CNN. In theory, the method could be applied to CNN by keeping the resampling length (l_w) uniform for every sample. Future attempts will be made to verify this hypothesis to extend the application scope of the proposed data augmentation strategy.

5. Conclusions

This study presents a novel data augmentation algorithm for detecting AF from ECG recordings using deep learning models. The algorithm concatenates the duplicated ECG episode to the original one based on characteristic points and resamples randomly from the concatenated signal. It is capable of balancing the number of samples among different categories, as well as increasing the diversity of the dataset, rendering a considerably improved performance of the RNN. Such an algorithm may facilitate the application of deep learning in the analysis of ECG as well as other 1D physiological signals, particularly when the dataset is small and imbalanced.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (Grant 31870938, 81871454), Major Scientific Project of Zhejiang Lab (Grant 2018DG0ZX01), and Zhejiang Provincial Key Laboratory of Communication Networks and Applications.

References

[1] V. Fuster, L.E. Ryden, D.S. Cannom, H.J. Crijns, A.B. Curtis, K.A. Ellenbogen, J.L. Halperin, G.N. Kay, J.Y. Le Huezey, J.E. Lowe, S.B. Olsson, E.N. Prystowsky, J.L. Tamargo, L.S. Wann, 2011 ACCF/AHA/HRS Focused Updates Incorporated into the ACC/AHA/ESC 2006 Guidelines for the Management of Patients with Atrial Fibrillation, Circulation 123 (10) (2011) E269–E367.
[2] W.M. Feinberg, J.L. Blackshear, A. Laupacis, R. Kronmal, R. Hart, Prevalence, age distribution, and gender of patients with atrial fibrillation. Analysis and implications, Arch Intern Med 155 (1995) 469–473.
[3] S. Stewart, C.L. Hart, D.J. Hole, J.J. Mcmurray, Population prevalence, incidence, and predictors of atrial fibrillation in the Renfrew/Paisley study, Heart 86 (5) (2001) 516–521.
[4] C.D. Furberg, B.M. Psaty, T.A. Manolio, J.M. Gardin, V.E. Smith, P.M. Rautaharju, Prevalence of atrial fibrillation in elderly subjects (the Cardiovascular Health Study), Am. J. Cardiol. 74 (3) (1994) 236–241.
[5] P.A. Wolf, R.D. Abbott, W.B. Kannel, Atrial fibrillation as an independent risk factor for stroke: the Framingham Study, Stroke 22 (8) (1991) 983–988.
[6] E.J. Benjamin, P.A. Wolf, R.B. D'Agostino, H. Silbershatz, W.B. Kannel, D. Levy, Impact of atrial fibrillation on the risk of death: the Framingham Heart Study, Circulation 98 (10) (1998) 946–952.
[7] R.S. Andersen, A. Peimankar, S. Puthusserypady, A deep learning approach for real-time detection of atrial fibrillation, Expert Syst. Appl. 115 (2019) 465–473.
[8] I. Savelieva, A.J. Camm, Silent atrial fibrillation—another Pandora's box, Pacing Clin. Electrophysiol. 23 (2) (2000) 145–148.
[9] G.K. Andrikopoulos, P.E. Dilaveris, D.J. Richter, E.J. Gialafos, A.G. Synetos, J.E. Gialafos, Increased variance of P wave duration on the electrocardiogram distinguishes patients with idiopathic paroxysmal atrial fibrillation, Pacing Clin. Electrophysiol. Pace 23 (7) (2000) 1127–1132.
[10] K. Aytemir, N. Ozer, E. Atalar, E. Sade, S. Aksöyek, K. Ovünç, A. Oto, F. Ozmen, S. Kes, P wave dispersion on 12-lead electrocardiography in patients with paroxysmal atrial fibrillation, Pacing Clin. Electrophysiol. 23 (7) (2000) 1109–1112.
[11] S. Ladavich, B. Ghoraani, Rate-independent detection of atrial fibrillation by statistical modeling of atrial activity, Biomed. Signal Process. Control 18 (2015) 274–281.
[12] J. Park, S. Lee, M. Jeon, Atrial fibrillation detection by heart rate variability in Poincare plot, Biomed. Eng. Online 8 (1) (2009) 38.
[13] J. Lian, L. Wang, D. Muessig, A simple method to detect atrial fibrillation using RR intervals, Am. J. Cardiol. 107 (10) (2011) 1494–1497.

[14] X. Zhou, H. Ding, B. Ung, E. Pickwell-MacPherson, Y. Zhang, Automatic online detection of atrial fibrillation based on symbolic dynamics and Shannon entropy, Biomed. Eng. Online 13 (1) (2014) 18.
[15] C. Huang, S. Ye, H. Chen, D. Li, F. He, Y. Tu, A novel method for detection of the transition between atrial fibrillation and sinus rhythm, IEEE Trans. Biomed. Eng. 58 (4) (2011) 1113–1119.
[16] R. Couceiro, P. Carvalho, J. Henriques, M. Antunes, Detection of atrial fibrillation using model-based ECG analysis, in: International Conference on Pattern Recognition, IEEE, 2008, pp. 1–5.
[17] S. Babaeizadeh, R.E. Gregg, E.D. Helfenbein, J.M. Lindauer, S.H. Zhou, Improvements in atrial fibrillation detection for real-time monitoring, J. Electrocardiol. 42 (6) (2009) 522–526.
[18] S. Asgari, A. Mehrnia, M. Moussavi, Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine, Comput. Biol. Med. 60 (2015) 132–142.
[19] X. Xu, S. Wei, C. Ma, K. Luo, L. Zhang, C. Liu, Atrial fibrillation beat identification using the combination of modified frequency slice wavelet transform and convolutional neural networks, J. Healthc. Eng. 2018 (2018), 2102918.
[20] Y. Xia, N. Wulan, K. Wang, H. Zhang, Detecting atrial fibrillation by deep convolutional neural networks, Comput. Biol. Med. 93 (2018) 84–92.
[21] R. Salloum, C.-C.J. Kuo, ECG-based biometrics using recurrent neural networks, in: IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2017, pp. 2062–2066.
[22] T. Teijeiro, C.A. García, D. Castro, P. Félix, Arrhythmia classification from the abductive interpretation of short single-lead ECG records, in: Computing in Cardiology, IEEE, 2017, pp. 1–4.
[23] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436.
[24] J. Salamon, J.P. Bello, Deep convolutional neural networks and data augmentation for environmental sound classification, IEEE Signal Process. Lett. 24 (3) (2017) 279–283.
[25] T.-Y. Lin, S. Maji, Visualizing and understanding deep texture representations, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2016, pp. 2791–2799.
[26] H. Salehinejad, S. Valaee, T. Dowdell, J. Barfett, Image augmentation using radial transform for training deep neural networks, in: International Conference on Acoustics, Speech and Signal Processing, IEEE, 2018, pp. 3016–3020.
[27] A.L. Guennec, S. Malinowski, R. Tavenard, Data augmentation for time series classification using convolutional neural networks, in: ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data, 2016.
[28] T.T. Um, F.M.J. Pfister, D. Pichler, S. Endo, M. Lang, S. Hirche, U. Fietzek, D. Kuli, Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks, in: 19th ACM International Conference on Multimodal Interaction, ACM, 2017, pp. 216–220.
[29] D.L. Donoho, De-noising by soft-thresholding, IEEE Trans. Inf. Theory 41 (3) (1995) 613–627.
[30] D.L. Donoho, J.M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika 81 (3) (1994) 425–455.
[31] B.N. Singh, A.K. Tiwari, Optimal selection of wavelet basis function applied to ECG signal denoising, Digit. Signal Process. 16 (3) (2006) 275–287.
[32] J. Pan, W.J. Tompkins, A real-time QRS detection algorithm, IEEE Trans. Biomed. Eng. BME-32 (3) (1985) 230–236.
[33] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, G.Z. Yang, Deep learning for health informatics, IEEE J. Biomed. Health Inform. 21 (1) (2017) 4–21.
[34] Z. Zhao, W.H. Chen, X.M. Wu, P.C.Y. Chen, J.M. Liu, LSTM network: a deep learning approach for short-term traffic forecast, Iet Intell. Transp. Syst. 11 (2) (2017) 68–75.
[35] X.L. Ma, Z.M. Tao, Y.H. Wang, H.Y. Yu, Y.P. Wang, Long short-term memory neural network for traffic speed prediction using remote microwave sensor data, Transp. Res. C: Emerging Technol. 54 (2015) 187–197.
[36] W. De Mulder, S. Bethard, M.F. Moens, A survey on the application of recurrent neural networks to statistical language modeling, Comput. Speech Lang. 30 (1) (2015) 61–98.
[37] H. Palangi, L. Deng, Y.L. Shen, J.F. Gao, X.D. He, J.S. Chen, X.Y. Song, R. Ward, Deep sentence embedding using long short-term memory networks: analysis and application to information retrieval, IEEE-ACM Trans. Audio Speech Lang. Process. 24 (4) (2016) 694–707.
[38] Y.L. Hsu, Y.T. Yang, J.S. Wang, C.Y. Hsu, Automatic sleep stage recurrent neural classifier using energy features of EEG signals, Neurocomputing 104 (2013) 105–114.
[39] O. Faust, A. Shenfield, M. Kareem, T.R. San, H. Fujita, U.R. Acharya, Automated detection of atrial fibrillation using long short-term memory network with RR interval signals, Comput. Biol. Med. 102 (2018) 327–335.
[40] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[41] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, in: International Conference on Learning Representations, IEEE, 2015, pp. 1–15.
[42] W. Sun, D.H.J. Poot, I. Smal, X. Yang, W.J. Niessen, S. Klein, Stochastic optimization with randomized smoothing for image registration, Med. Image Anal. 35 (2017) 146–158.
[43] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.
[44] Y. Gal, Z. Ghahramani, A theoretically grounded application of dropout in recurrent neural networks, in: Advances in Neural Information Processing Systems, 2016, pp. 1019–1027.
[45] G.D. Clifford, C. Liu, B. Moody, L.-w.H. Lehman, I. Silva, Q. Li, A. Johnson, R.G. Mark, AF classification from a short single lead ECG recording: the physionet computing in cardiology challenge 2017, in: Computing in Cardiology, IEEE, 2017, pp. 1–4.
[46] S. Hong, M. Wu, Y. Zhou, Q. Wang, J. Shang, H. Li, J. Xie, ENCASE: an ENsemble ClASsifiEr for ECG classification using expert features and deep neural networks, in: Computing in Cardiology, IEEE, 2017, pp. 1–4.
[47] M. Zabihi, A.B. Rad, A.K. Katsaggelos, S. Kiranyaz, S. Narkilahti, M. Gabbouj, Detection of atrial fibrillation in ECG hand-held devices using a random forest classifier, in: Computing in Cardiology, IEEE, 2017, pp. 1–4.
[48] S. Datta, C. Puri, A. Mukherjee, R. Banerjee, A.D. Choudhury, R. Singh, A. Ukil, S. Bandyopadhyay, A. Pal, S. Khandelwal, Identifying Normal, AF and other abnormal ECG rhythms using a cascaded binary classifier, in: Computing in Cardiology, IEEE, 2017, pp. 1–4.
[49] Z. Xiong, M.K. Stiles, J. Zhao, Robust ECG signal classification for detection of atrial fibrillation using a novel neural network, in: Computing in Cardiology, IEEE, 2017, pp. 1–4.
[50] M. Zihlmann, D. Perekrestenko, M. Tschannen, Convolutional recurrent neural networks for electrocardiogram classification, in: Computing in Cardiology, IEEE, 2017, pp. 1–4.
[51] M. Kropf, D. Hayn, G. Schreier, ECG classification based on time and frequency domain features using random forests, in: Computing in Cardiology, IEEE, 2017, pp. 1–4.
[52] D. Smole, Atrial fibrillation detection using boosting and stacking ensemble, in: Computing in Cardiology, IEEE, 2017, pp. 1–4.
[53] R. Kamaleswaran, R. Mahajan, O. Akbilgic, A robust deep convolutional neural network for the classification of abnormal cardiac rhythm using single lead electrocardiograms of variable length, Physiol. Meas. 39 (3) (2018), 035006.
[54] S.D. Goodfellow, A. Goodwin, R. Greer, P.C. Laussen, M. Mazwi, D. Eytan, Atrial fibrillation classification using step-by-step machine learning, Biomed. Phys. Eng. Express 4 (4) (2018), 045005.
[55] N. Liu, M. Sun, L. Wang, W. Zhou, H. Dang, X. Zhou, A support vector machine approach for AF classification from a short single lead ECG recording, Physiol. Meas. 39 (6) (2018), 064004.
[56] A.Y. Hannun, P. Rajpurkar, M. Haghpanahi, G.H. Tison, C. Bourn, M.P. Turakhia, A.Y. Ng, Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network, Nat. Med. 25 (1) (2019) 65–69.