Information Fusion 53 (2020) 174–182

Contents lists available at ScienceDirect

Information Fusion

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

Full Length Article

# Multi-class Arrhythmia detection from 12-lead varied-length ECG using Attention-based Time-Incremental Convolutional Neural Network

Qihang Yao, Ruxin Wang, Xiaomao Fan, Jikui Liu, Ye Li*

*Joint Engineering Research Center for Health Big Data Intelligent Analysis Technology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China*

ARTICLE INFO

*Keywords:*
Convolutional neural network
Recurrent cells
Attention module
Arrhythmia detection
Spatial temporal fusion

ABSTRACT

Automatic arrhythmia detection from Electrocardiogram (ECG) plays an important role in early prevention and diagnosis of cardiovascular diseases. Convolutional neural network (CNN) is a simpler, more noise-immune solution than traditional methods in multi-class arrhythmia classification. However, suffering from lack of consideration for temporal feature of ECG signal, CNN couldn't accept varied-length ECG signal and had limited performance in detecting paroxysmal arrhythmias. To address these issues, we proposed attention-based time-incremental convolutional neural network (ATI-CNN), a deep neural network model achieving both spatial and temporal fusion of information from ECG signals by integrating CNN, recurrent cells and attention module. Comparing to CNN model, this model features flexible input length, halved parameter amount as well as more than 90% computation reduction in real-time processing. The experiment result shows that, ATI-CNN reached an overall classification accuracy of 81.2%. In comparison with a classical 16-layer CNN named VGGNet, ATI-CNN achieved accuracy increases of 7.7% in average and up to 26.8% in detecting paroxysmal arrhythmias. Combining all these excellent features, ATI-CNN offered an exemplification for all kinds of varied-length signal processing problems.

## 1. Introduction

Cardiovascular disease (CVD) is the most common cause of death, accounting for over 31% of deaths around the world [1]. It was shown by statistics that sudden cardiac deaths, over 80% of which were closely related to cardiac arrhythmias, were responsible for half of deaths caused by heart diseases [2]. The most widely applied solution for arrhythmia detection is electrocardiography (ECG), which records the electrical activity of heart over a period of time using electrodes placed over skin, as illustrated in Fig. 1. Capturing heart's electrical potential from different angles through different leads, ECG shows the morbid status of cardiovascular system by changes in its waveforms or rhythms [3]. Using ECG, doctors can learn about the risk for a patient to have various kinds of cardiac disease, for instance, ischemia [4], heart attack [5] and stroke [6]. Automatic arrhythmia detection based on ECG provides important assistances for doctors, and also helps common people to self-monitor their heart conditions using wearable devices. Accurate automatic arrhythmia detection plays as the foundation of machine-aided diagnosis and treatment of cardiovascular diseases.

In past decades, automatic arrhythmia detection from ECG has been widely investigated. Greatly benefited by the construction and continuous refinement of open-source ECG databases, like MIT-BIH [7], heartbeat-level analysis of ECG became approachable. Many methods were proposed for discrimination between heartbeats in five classes (Normal, Supraventricular ectopic beat (SVEB), Ventricular ectopic beat (VEB), Fusion beat and Unknown beat) specified by the AAMI standard [8], where waveform of heartbeats could be differentiated. Lin et al. proposed to apply weighted linear discriminator on the normalized RR-intervals [9], and gained an overall classification performance of 93%. Huang et al. proposed a method using an ensemble of support vector machine (SVM) to learn from the random projections of ECG signals, which achieved sensitivity of over 90% for both SVEB and VEB [10]. Apart from this coarse-grain classification between heartbeat types, detection and classification of other cardiac arrhythmias which doesn't or seldom show in a single heartbeat was also a major topic to be studied. For instance, Dai et al. proposed a novel method to reduce QRS residual in atrial activity extraction, and achieved great accuracy in identifying atrial fibrillation [11]. Perlman et al. classified five types of supraventricular tachycardia using a clinically oriented decision tree on atrial and ventricular information extracted through atrial wave detection [12]. Noises unavoidably introduced in collection of ECG signals, and individual differences in ECG waveforms, lead to strong difficulties in the design of empirical features. Constrained by these two factors, application of traditional signal processing and machine learning methods were mostly limited to relatively simple problems, when their usage in more complex arrhythmia detection problems, like fine-grain arrhythmia classification, were less concerned.

* Corresponding author.
*E-mail address:* ye.li@siat.ac.cn (Y. Li).

https://doi.org/10.1016/j.inffus.2019.06.024
Received 31 October 2018; Received in revised form 22 February 2019; Accepted 15 June 2019
Available online 17 June 2019

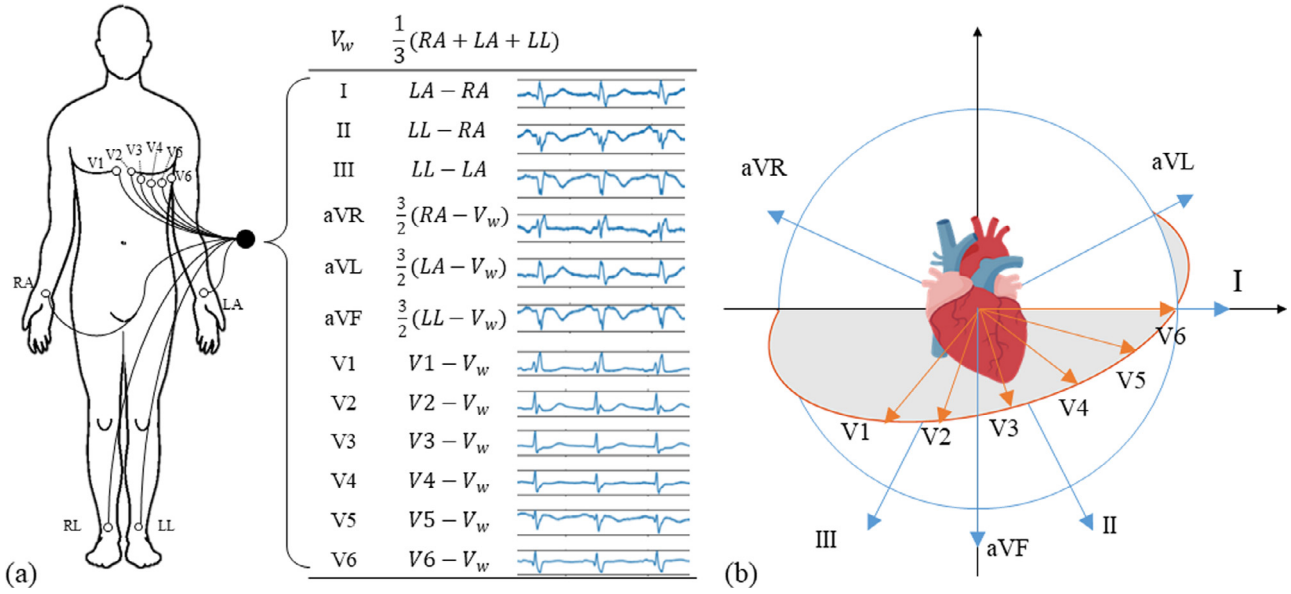


**Fig. 1.** Illustration for 12-lead ECG system. a) Spatial placement of 10 electrodes used by 12-lead ECG system. Three limb leads and three augmented limb leads as well as six precordial(chest) leads were formed between physical electrodes and a virtual electrode known as Wilson's central terminal. b) Representing electrical potential differences between electrodes placed on human skin, ECG signals from 12 leads reflect electrical activity of heart from different spatial angles.

Constricted by limited computing power and amount of data, previous machine learning methods showed less satisfying performance. Recent advances of intelligence, the application of deep neural networks, has delivered a highly data-driven method applicable to a wide range of problems. In medicine, deep learning was demonstrated to be effective in diagnosis of cancers [13,14], brain diseases [15], Alzheimer's disease [16]. In the future, it is anticipated to be able to provide assistances for doctors in disease diagnosis and provide extramural care services for patients. With deep learning, more complex arrhythmia detection problems could also be solved. Rajpurkar et al. proposed a 34-layer convolutional neural network (CNN) for arrhythmia detection, and reported cardiologist-level accuracy in classifying fourteen cardiac arrhythmias [17]. Fan et al. proposed multi-scale fusion of CNNs which was proved to be effective for detection of atrial fibrillation from short ECG signals [18]. Kiranyaz et al. proposed an adaptive CNN model for intra-patient ventricular ectopic beats and supraventricular ectopic beats detection [19], which required only very few patient-specific data to achieve high accuracy. Acharya et al. detected myocardial infarction using a 11-layer CNN and obtained state-of-art performance [20]. In these works, ECG's temporal properties were usually omitted and led to less optimal results concerning both performance and efficiency of solutions. In other research works, ECG signals were treated as time series and deep network structures which were designed to deal with time series were adopted. Li et al. proposed a model based on deep neural network and hidden Markov model for obstructive sleep apnea detection from ECG segments [21]. Chauhan et al. achieved great results using deep long short-term memory network in detecting four types of arrhythmias, without any preprocessing of ECG signals needed [22]. However, without convolutional structures, models used in these works were limited to only have dense connections. Therefore, depths of these models, as well as the ability to generate more complex features, were limited.

Above mentioned works demonstrated the application of many widely-used deep learning network structures in arrhythmia detection from ECG, but deep learning models tailored for physiological signal were less studied. Several characteristics should be taken into consideration. First, periodicity exists in ECG signals. Localized waveform features and global features like heart rate variability (HRV) both contributed to arrhythmia detection, and should therefore be both emphasized [23]. Second, as many paroxysmal rhythms or ectopic beats intermittently show up in ECG records [24], and beat-wise annotations are of great cost, the ideal model should be robust to interference from non-informative part of a record. Third, widely-applied 12-lead ECG recordings provide richer information for diagnosis of arrhythmias. Effective fusion of information from spatially distributed sources of this multi-sensor system, along with troubles it might brought [25], are to be considered in the model design.

Major contributions of this research are as follows: 1) A novel neural network model named attention-based time-incremental convolutional neural network (ATI-CNN) was proposed to fully utilize the temporal and spatial characteristics of ECG. This model divides ECG processing pipeline into two phases: spatial information fusion based on CNN and temporal information fusion based on recurrent neural network (RNN) and attention mechanism. 2) Unwrapping ability of recurrent cells was exploited to extend the proposed model's input to varied-length ECG signals. Comparing to traditional CNN model, the proposed model get rid of cropping/padding signal beforehand when dealing with varied-length signal database and is more robust in the detection of paroxysmal arrhythmias. 3) Attention mechanism was introduced to output the signal segment of interest along with classification result. It was demonstrated through experiments how this mechanism help to locate the abnormality in ECG signals and improve the interpretability of deep learning model. The proposed model was validated on the China Physiological Signal Challenge 2018 database[1], and demonstrated its performance in classifying 9 classes of cardiac arrhythmias using 12-lead ECG signals.

The rest of this paper is orgnaized as follows: Section 2 introduces architecture of ATI-CNN, and the detailed experiment process is described in Section 3. Section 4 presents the result and Section 5 give discussion explaining ATI-CNN's advantages over traditional CNNs in several aspects. Section 6 concludes the paper.

## 2. Methods

### 2.1. Problem formulation

ECG arrhythmia detection could be generalized as a time-series classification problem, in which a model was required to extract useful in-

[1] http://www.icbeb.org/Challenge.html.

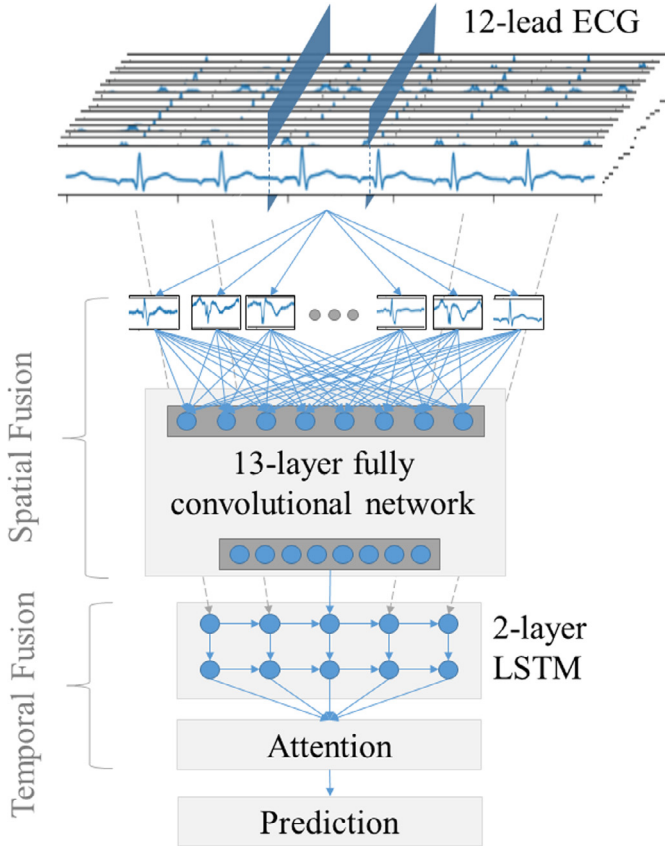


**Fig. 2.** Architecture for the proposed model. 12-lead ECG is fed into fully convolutional layers and therefore generate a preprocessed time series. This time series is further fed into LSTM cells to exchange information among different time points. An attention module accepts output of LSTM cells, assigns weights for different time points and outputs a final result.

formation from varied-length ECG records and predict the correct class for the record. This task requires a model which takes a varied-length signal $X = [x_1, \dots x_k]$ as input, and outputs a label which indicates the class a signal belongs to. The objective of a model is to minimize the cross-entropy between the reference labels and outputs, given by:

$$loss(X, r) = -\log\left(\frac{\exp(p(X, r))}{\sum_j \exp(p(X, j))}\right) \tag{1}$$

where $p(X, j)$ is the probability the model assigns the label $j$ to the input $X$, and $r$ is reference label.

### 2.2. Model architecture

The propsoed model, illustrated in the Fig. 2, was constructed integrating a fully convolutional neural network, LSTM layers and an attention module. The convolutional neural network was designed with reference to the VGGNet developed by Visual Geometry Group [26]. VGGNet was chosen as it features smooth information flow and simple implementation. Nevertheless, it should be noted that the proposed model was designed to be compatible with a variety of convolutional neural network designs, and in other signal processing problems the depth and design of convolutional layers could be adjusted based on the specific requirements of those problems. THe channel-wise information sharing in convolutional layers facilitated the spatial fusion of information from different ECG leads. Following the convolutional layers were two LSTM cells. In the runtime, they would be unwrapped to the length of input feature series. These layers facilitated information exchange between signal segments, and their output were fed to an attention module. The attention module assigned weights for features extracted from different signal segments, and output final classification result based on synthesized feature vector.

| Conv3 x 64 |
|---|
| Conv3 x 64 |
| Pooling |
| Conv3 x 128 |
| Conv3 x 128 |
| Pooling |
| Conv3 x 256 |
| Conv3 x 256 |
| Conv3 x 256 |
| Pooling |
| Conv3 x 256 |
| Conv3 x 256 |
| Conv3 x 256 |
| Pooling |
| Conv3 x 256 |
| Conv3 x 256 |
| Conv3 x 256 |
| Pooling |

**Fig. 3.** Layer configuration for fully convolutional neural network. 'Conv3 x64' represents a convolutional layer with 64 kernels of size 3.

#### 2.2.1. Spatial fusion

With reference to the model used in [18], 13 convolutional layers and 5 pooling layers were used in the fully convolutional neural network, forming an architecture as illustrated in Fig. 3. All the convolutional layers use a kernel size of 3, a boundary padding of 1 and a stride of 1. Therefore, the length of the signals was maintained during convolutions and only controlled by pooling layers. The pooling layers all have a kernel size of 3 and a stride of 3, therefore each of them reduces the length of the signal by 3 times. Starting from 64, numbers of feature maps generated by each convolutional layer scales up by a factor of 2 when it passes through each pooling layer. A batch normalization layer and a rectified linear unit (ReLU) function follows each convolutional layer. Batch normalization layer [27] is a trainable layer designed to re-normalize the distribution of data and mitigate alternation made by convolutional layers. It helps to obtain more stable parameter update in training. ReLU [28] is a widely used activation function recognized for its ability to prevent the vanishing gradient problem in deep neural networks.

#### 2.2.2. Temporal fusion

Two LSTM cells were used in the proposed model. A LSTM cell could be regarded as a cell which replicate itself to form a sequence of cells along time axis. Between these cells there are unidirectional connections to pass down information in the direction of time. In another view, this model is a single cell which read through the series input, and update its state vector and generate output based on what it is reading as well as what it has memorized. The cell's behavior is concluded using following equations:

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_{t-1}\right] + b_f\right) \tag{2}$$

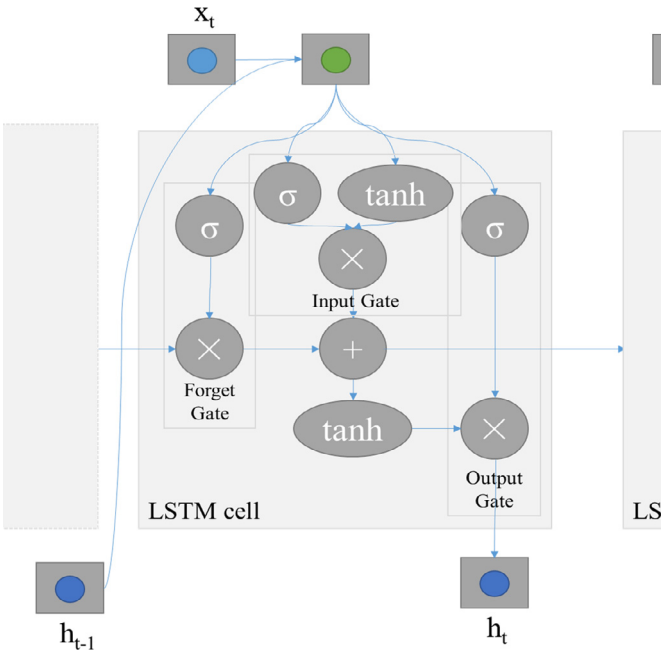


**Fig. 4.** Internal sturcture and information flow during inference for an LSTM cell. $x_t$ and $h_t$ indicates the input size and the output size for a LSTM cell.

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_{t-1}\right] + b_i\right) \tag{3}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot tanh\left(W_C \cdot \left[h_{t-1}, x_{t-1}\right] + b_C\right) \tag{4}$$

$$h_t = \sigma\left(W_o\left[h_{t-1}, x_{t-1}\right] + b_o\right) \odot tanh\left(C_t\right) \tag{5}$$

where $h_t$, $x_t$ and $C_t$ denotes the cell's output, input and state at time t, and $f_t$, $i_t$ denotes the output of forget gate and input gate inside the LSTM cell. $\sigma$ denotes the sigmoid function widely used as activation function in machine learning. The internal structure of LSTM cell is illustrated in Fig. 4. Trainable parameters including $W_f$, $W_i$, $W_C$ and $W_o$, as well as their corresponding biases, making it possible for the cell to decide whether to remember or forget certain part of the information in series. Input size of first LSTM cell is dependent on the number of feature maps generated by convolutional layers and is 512 in this paper. The output size of both LSTM cells are 32. Therefore, LSTM cells are also responsible for reducing feature dimension.

#### 2.2.3. Attention mechanism

The attention module used in ATI-CNN is based on the similar principle as the one used in [29], as illustrated in Fig. 5. In an attention module, two fully connected layers along with a hyperbolic tangent function were applied to every single feature vector in the input feature series, by iterating through the series. The latter fully connected layer output a series in which every element is a vector of length $N$, or $N$ series in which every element in every series is a single real value. Softmax was applied to each of $N$ series and therefore values in each series summed to 1. As a result, for each possible class in $N$ classes, the attention module generated an independent group of values which would be used as weights. For each of $N$ classes, a unique weighted average of input was computed, and was then used to calculate the probability that the input signal belongs to that class.

Benefits in two aspects were expected by introducing attentions module. First, it could help the model to concentrate on the informative and important segments of a signal and therefore improve detection performance. Second, it could help to highlight the location of abnormal signal patterns which need further investigation, and adds to the interpretability of our model.

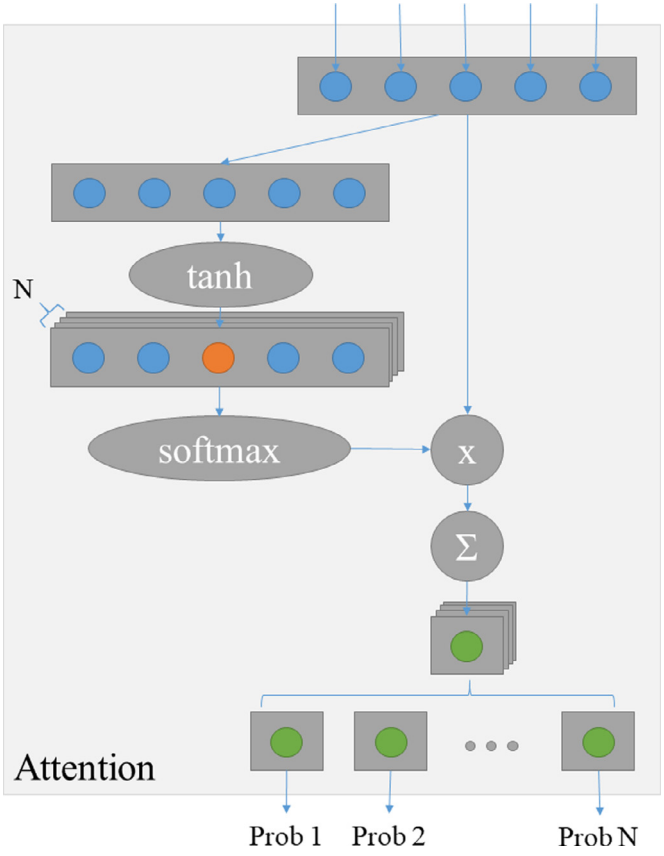


**Fig. 5.** Structure of attention module. The attention module generates weights for different signal segments based on different target class and combines information from different signal segments in the form of weighted average of them. $N$ is the number of classes in this classification problem.

## 3. Experiment

### 3.1. Environment

The proposed model was trained and tested on a server with a Xeon E5 2650 CPU, 128GB memory and four Titan Xp graphic cards. This server runs an Ubuntu 16.04 LTS system, and the model was implemented using the PyTorch 0.4.1 framework.

### 3.2. Data source

The training data used in this study was from the 1st China Physiological Signal Challenge. This dataset contains 6877 12-lead ECG records from 6 s to 60 s. A total of 8 types of arrhythmia and normal sinus rhythms were to be classified in these records. These records were collected from 11 hospitals and sampled at 500 Hz. Table 1 give details concerning this dataset. The independent test dataset contains 2954 records collected in the same way as our training dataset.

### 3.3. Preprocessing

As the dataset used is relatively small compared with common use cases for deep learning like computer vision and natural language processing, the problem of overfitting could easily occur. Several data augmentation strategies were used in the training of models. Training signals were stretched or compressed in the time axis by a random factor which is a real value uniformly sampled from [1, 1.2]. This would introduce noises as some arrhythmias were sensitive to heart rate, but in practice it helped the model in achieving better performance. In [30],

**Table 1**
Data profile for training dataset.

| Type | Records | Time length(s) | | | | |
|---|---|---|---|---|---|---|
| | | *Mean* | SD | Min | Median | Max |
| N | 918 | 15.43 | 7.61 | 10.00 | 13.00 | 60.00 |
| AF | 1098 | 15.01 | 8.39 | 9.00 | 11.00 | 60.00 |
| I-AVB | 704 | 14.32 | 7.21 | 10.00 | 11.27 | 60.00 |
| LBBB | 207 | 14.92 | 8.09 | 9.00 | 12.00 | 60.00 |
| RBBB | 1695 | 14.42 | 7.60 | 10.00 | 11.19 | 60.00 |
| PAC | 556 | 19.46 | 12.36 | 9.00 | 14.00 | 60.00 |
| PVC | 672 | 20.21 | 12.85 | 6.00 | 15.00 | 60.00 |
| STD | 825 | 15.13 | 6.82 | 8.00 | 12.78 | 60.00 |
| STE | 202 | 17.15 | 10.72 | 10.00 | 11.89 | 60.00 |
| Total | 6877 | 15.79 | 9.04 | 6.00 | 12.00 | 60.00 |

[a] Arrhythmia names were abbreviated in the table and the exact names are as follows: (1) Normal (N); (2) Atrial fibrillation (AF); (3) First-degree atrioventricular block (I-AVB); (4) Left bundle branch block (LBBB); (5) Right bundle branch block (RBBB); (6) Premature atrial contraction (PAC); (7) Premature ventricular contraction (PVC); (8) ST-segment depression (STD); (9) ST-segment elevation (STE).

DeVries et al. proposed a data augmentation method for image processing, in which randomly picked squares were cut from the input image. This method was proved to be effective in preventing deep neural networks from memorizing the training data. In this paper, this method was extended to 1-D signal processing and found to be useful, as it encouraged the model to rely less on a single local pattern. For each training sample, a signal segment which is at most 1.5 s long was masked with zero. This approach also worked in the sense that it simulated the cases when leads were accidentally detached from patient's skin.

Besides, all records were normalized to have a mean of zero and a standard deviation of one. Normalization is vital for a variety of machine learning methods, and in deep learning, a normalized distribution of input data would help the model to converge faster [31].

### 3.4. Training setting

The weights of convolutional layers and fully connected layers in ATI-CNN were initialized using the kaiming initializer [32] and the LSTM cells were initialized using orthogonal initializer [33], which were shown to greatly improve the converging speed of model parameters. The entire model was trained using Adam optimizer [34] with default parameters and a learning rate of 0.0001 at the beginning. Adam was applied for its ability to balance gradient update between different classes, and therefore alleviate adverse effects caused by unbalanced data. In a total of 150 epochs of training, learning rate was multiplied by a factor of 0.1 after every 50 epochs.

L2 loss of all the parameters in the model was multiplied by a factor of 0.004 and added to the training loss, which helped to prevent the cases that some parameters in the model become too large and dominate calculation of model. Dropout [35], as a widely used regularization method, was proved to have similar effects as training a group of smaller models, and averaging them in testing, which is beneficial to model's generalizability. In this experiment, dropout with a rate of 0.2 was applied in LSTM cells. Though it was proposed that a dropout rate of 0.5 lead to optimal result, in this experiment our configuration was found to be comparably effective and more efficient. As an extension to dropout which introduce noises to the middle layers, DisturbLabel [36] proposed by Xie et al., counter-intuitively adds incorrect labels to the training data, and thus adds noises in the calculation of loss function. This method implicitly average models trained with different label sets, and therefore effectively prevent overfitting. A disturbing probability of 0.4, as recommended by the author, was used in this paper.

In order to accelerate training process, the most time-consuming calculation in convolutional neural network were batched to achieve better parallelism. Before convolutional layers, records were all padded to 60 s and the original lengths were recorded. They were then grouped into batches of 128 records and fed in to convolutional neural network. When the feature series extracted by convolutional layers were fed into first LSTM layer, the original lengths was provided to the LSTM cell to indicate the length the LSTM cell need to unwrap to. The output feature series of LSTM cells will be cropped based on the lengths recorded. The attention module, as a small and efficient module, iterated among 128 samples in the batch sequentially, and output predictions which would then be grouped again. Cross entropy loss was calculated for the batched output and corresponding labels, and the averaged gradient was back propagated to all the weights in previously mentioned layers.

### 3.5. Reference model

Limited by the size of available datasets, previous research for classification of multiple types of arrhythmias primarily worked on beat-wise classification, and on different groups of arrhythmias, which make it difficult to offer direct comparison. To evaluate the proposed model's performance, we implemented models including three 1D VGGNets with different input size and a time-incremental convolutional neural network (TI-CNN) as references.

VGGNet shares the same convolutional layer configuration with ATI-CNN, while three fully connected layers instead of LSTM layers or attention module were placed behind convolutional layers. Three fully connected layers have 1,024, 1,024, 256 neurons. In the design of three reference VGGNets, different input sizes, including 6 s, 12 s and 60 s were applied, respectively, and therefore determined the number of weights needed for the connections between convolutional layers and fully connected layers. These thmodels are referred to as VGG-6, VGG-12 and VGG-60 in following passages. These specific sizes are used as they are the minimum, median and maximum lengths of ECG records from the training dataset. If a ECG record is too short, it is padded with zero and is cropped if too long. Random padding and cropping were used in training and central padding and cropping were used in validation and testing. Cropping and padding played as naive strategies to make VGGNet compatible with varied-length input and introduced minimal computation overhead or modification to the original CNN models.

Apart from naive VGGNets, we also used TI-CNN as reference. TI-CNN have network architecture very similar to ATI-CNN, while attention module was not included in TI-CNN. The final output of last LSTM layer, instead of weighted average of all outputs, were used to predict the class of an ECG signal. TI-CNN used exact the same input as ATI-CNN without any additional preprocessing, as it also in its design supports varied-length input. Comparison between ATI-CNN and TI-CNN helped to demonstrate how attention mechanism helped to improve not only the interpretability, but also the performance of a model.

Same hyper parameters including learning rate and batch size, etc. were used for the proposed model and all reference models, and same data augmentation and regularization strategies were used, in order to give a fair comparison.

## 4. Result

### 4.1. Performance metrics

In this research, typical classification metrics, including precision, recall, and F1 score were used for each class. They were defined as:

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1 = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \tag{8}$$

where *TP* refers to the number of correctly classified samples in a certain class, *FN* refers to the number of samples belonging to a certain class

**Table 2**
Classification Performance for VGGNet and the proposed model.

| Type | F1 score | | | | | fSupport |
|---|---|---|---|---|---|---|
| | VGG-6 | VGG-12 | VGG-60 | TI-CNN | ATI-CNN | |
| N | 0.717 | 0.733 | 0.743 | 0.753 | **0.789** | 394 |
| AF | 0.889 | 0.890 | 0.889 | 0.900 | **0.920** | 466 |
| I-AVB | 0.774 | 0.776 | 0.776 | 0.809 | **0.850** | 295 |
| LBBB | 0.846 | 0.852 | 0.841 | **0.874** | 0.872 | 97 |
| RBBB | 0.915 | 0.907 | 0.910 | 0.922 | **0.933** | 756 |
| PAC | 0.397 | 0.446 | 0.468 | 0.638 | **0.736** | 250 |
| PVC | 0.689 | 0.739 | 0.780 | 0.832 | **0.861** | 276 |
| STD | 0.704 | 0.716 | 0.721 | 0.762 | **0.789** | 340 |
| STE | 0.455 | 0.484 | 0.484 | 0.462 | **0.556** | 80 |
| **Average** | | | | | | |
| Precision | 0.724 | 0.738 | 0.770 | 0.799 | **0.826** | 2954 |
| Recall | 0.700 | 0.719 | 0.718 | 0.758 | **0.801** | |
| F1 score | 0.710 | 0.727 | 0.735 | 0.772 | **0.812** | |

which were misclassified as in other classes, and *FP* refers to the number of samples misclassified as in a certain class when they belong to other classes. The average of three metrics among classes were calculated, to give a final evaluation of the model's performance.

### 4.2. Classification performance

Table 2 compares the class-level F1 score and average precision, recall and F1 score of four reference models and ATI-CNN in identifying cardiac arrhythmias. It was shown that ATI-CNN outperformed VGGNets in the F1 score of all classes, and almost outperformed TI-CNN in all classes except for one class where two models perform comparably. The performance of VGGNets increased progressively with input signal length, indicating that padding zeros is a comparatively better strategy than cropping for accurate classification. Comparing to the VGG-60, TI-CNN and ATI-CNN achieved about 3.7%, 7.7% average F1-score increase, correspondingly. While two models both outperformed VGGNets in all classes, largest performance gain introduced by recurrent structures and attention module lies in the screening of premature atrial compression (PAC) and premature ventricular compression (PVC). TI-CNN gained about 17.0% and 5.2% F1-score increase in classifying these two arrhythmias, while ATI-CNN further improved these two figures by 7.2% and 2.9%. In identifying first-order atrioventricular block (I-AVB) and ST-segment elevation (STE), ATI-CNN achieved significantly better result compared with TI-CNN.

## 5. Discussion

### 5.1. Performance analysis

As shown in the result, for traditional fixed-length input CNN models, cropping caused significant performance drop, especially in classifying PAC and PVC. It is not surprising, as these two arrhythmias sometimes only appeared for a few times in the entire signal. The cropping strategy could very possibly cropped that informative part out, leaving the input signal very similar to records from normal class, as illustrated in Fig. 6. VGGNet with padding strategy to some degree compensated for this issue, showing their performance improvement in detecting PAC and PVC. However, due to lack of focus on abnormality and noises added through padding zeros, its performance is still not satisfying, indicating that in addition to spatial information fusion, more effective temporal fusion strategy need to be applied. TI-CNN, on the contrary, took advantage of design of recurrent cells, could effectively identify these two arrhythmias. Weighted temporal fusion of information introduced by attention module in ATI-CNN further improved model's performance on these two arrhythmias, indicating the importance of focuses on informative part of ECG signal when dealing with paroxysmal arrhythmias.

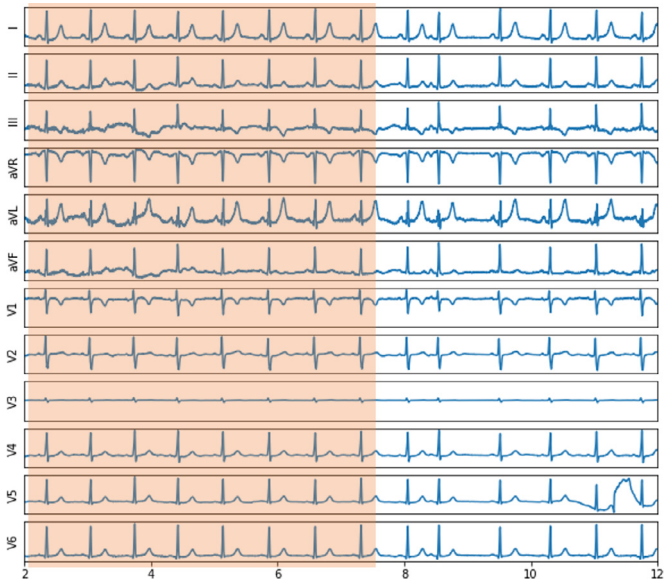


**Fig. 6.** Example for failure of VGGNet with cropping strategy to identify records with PAC. The rectangle indicates a possible cropped signal segment as the input to the model. This is a signal segment which will be misclassified as it doesn't contain the part of signal where PAC happens.

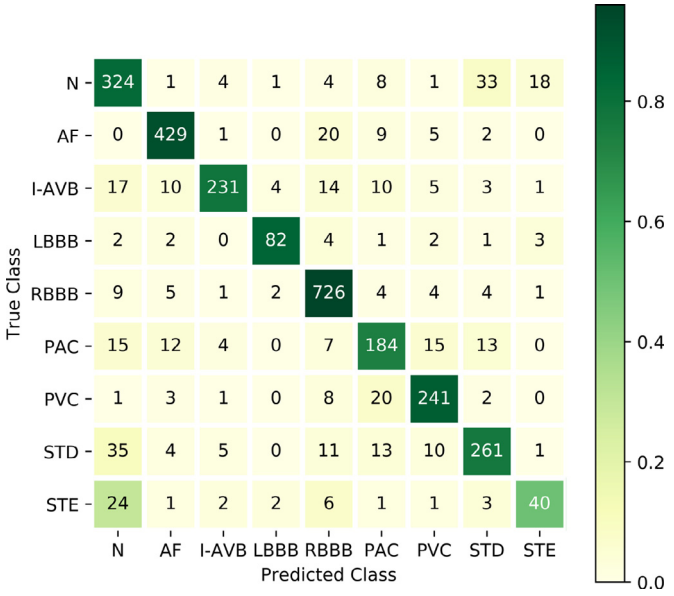


**Fig. 7.** Confusion Matrix of ATI-CNN. The row labels indicate the true class records in each row belong to, and the column labels indicate the class predicted by ATI-CNN for records in each column. Numbers in each grid show the number of records classified as column label when its true class is indicated by row label. The color represents the proportion of aforementioned records to all records in the same row.

In Fig. 7, the confusion matrix of ATI-CNN was drawn. It could be inferenced from this figure that the model made most mistakes in screening records with ST-depression from normal records. Taking the total number records in each class into consideration, it was found that the mistakes which influenced the performance of ATI-CNN the most were from discrimination between normal records and records with ST-elevation. ATI-CNN's lack of sensitivity to changes in ST-segment of an ECG waveform could be attributed to the fact that changes in ST-segment could be easily covered by noises in many cases.

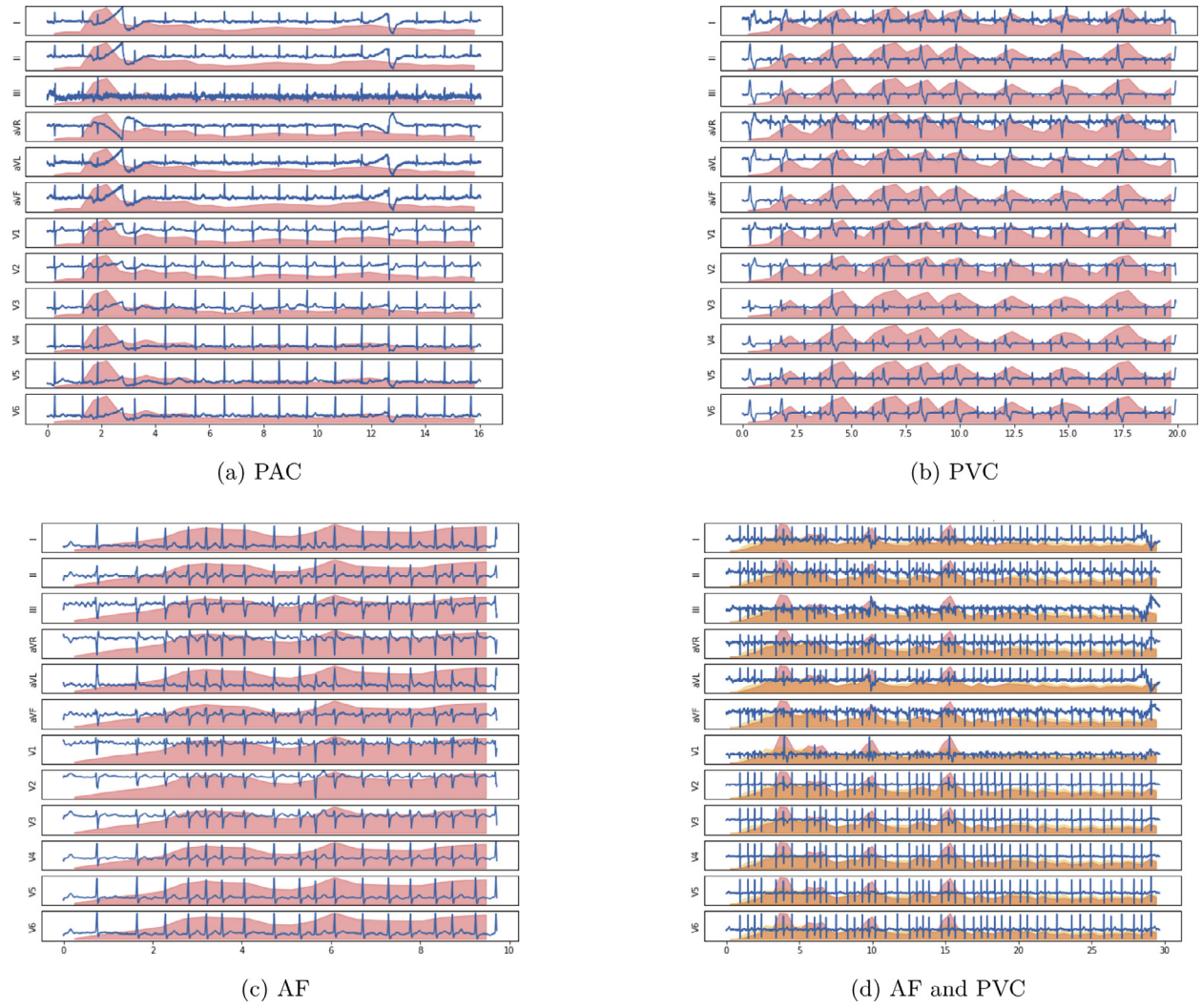


**Fig. 8.** Weights assigned for different segments in ECG records. In (a), (b) and (c), height of red area indicate the weight that signal segment in the same position is assigned with, when their correct class is concerned. In (d), weights for AF was drawn in orange, while weights for PVC was drawn in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

### 5.2. Attention weights

The attention module in ATI-CNN assigned different weights for different ECG segments in an ECG record, which helped the model to focus on informative part of a signal. 9 groups of weights were generated for one ECG record, representing the model's biases for different kind of waveform segments when different kinds of arrhythmia is considered. To illustrate the effectiveness of attention module, three samples with PAC, PVC, and AF was drawn, along with the weights generated by attention module for each record's true class, as shown in Fig. 8.

It could be seen that for arrhythmias with obvious occasional patterns like PAC and PVC, larger weights were clearly assigned for abnormal segments, demonstrating ATI-CNN's ability to locate ECG segment where abnormality appeared. What's more, in Fig. 8a, it is shown that though an irregular signal pattern occurred at about 13s, it was not weighted. This demonstrated that ATI-CNN didn't simply weight signal segments with more information or irregular pattern, but considered a signal segment informative only if it is concerned with the target arrhythmia.

In Fig. 8(c), as AF's pattern consistently appeared in the signal, weights were more uniformly assigned. Nevertheless, correlation between weights assigned and heartrate can be identified. In this case, the actual location of arrhythmia wasn't told using the weights, but attention still added to the interpretability of the model. In Fig. 8(d), a signal with both AF and PVC was drawn, together with two groups of weights for AF and PVC. From the discrepancy of weights, it was clearly shown that attention module has the ability to weight the same signal differently for different target arrhythmias.

### 5.3. Model parameter

In the design of ATI-CNN, the fully connected layers were abandoned to remove the requirement for fixed-length input. Apart from this flexibility, which is important for arrhythmia classification as aforementioned, the proposed model is also benefited by the reduced parameter amount brought by this design. In Table 3, the parameter amount of the VGG-6 network and the proposed model were calculated, to make a thorough comparison. It is shown that ATI-CNN has less than half of the model parameters of VGG-6 network. As two models actually share the same convolutional layer configuration, the cause of parameter reduction lies in the rest part. The connection between convolutional layers and fully connected layers was shown to be the major cause for the VGG16 network to inflate. It could be inferred that high redundancy exist in this connection, as it failed to take advantage of the periodicity of signals. On the contrary, both recurrent layers and attention module achieved heavy reuse of parameters. LSTM cells iterated through input

**Table 3**
Parameter Amount for VGG-6 and ATI-CNN.

| Layer | | Parameters | |
|---|---|---|---|
| VGG-6 | ATI-CNN | VGG-6 | ATI-CNN |
| Conv3-64 | Conv3-64 | 12 × 64 × 3 | 12 × 64 × 3 |
| Conv3-64 | Conv3-64 | 64 × 64 × 3 | 64 × 64 × 3 |
| Conv3-128 | Conv3-128 | 64 × 128 × 3 | 64 × 128 × 3 |
| Conv3-128 | Conv3-128 | 128 × 128 × 3 | 128 × 128 × 3 |
| Conv3-256 | Conv3-256 | 128 × 256 × 3 | 128 × 256 × 3 |
| Conv3-256 | Conv3-256 | 256 × 256 × 3 | 256 × 256 × 3 |
| Conv3-256 | Conv3-256 | 256 × 256 × 3 | 256 × 256 × 3 |
| Conv3-512 | Conv3-512 | 256 × 512 × 3 | 256 × 512 × 3 |
| Conv3-512 | Conv3-512 | 512 × 512 × 3 | 512 × 512 × 3 |
| Conv3-512 | Conv3-512 | 512 × 512 × 3 | 512 × 512 × 3 |
| Conv3-512 | Conv3-512 | 512 × 512 × 3 | 512 × 512 × 3 |
| Conv3-512 | Conv3-512 | 512 × 512 × 3 | 512 × 512 × 3 |
| Conv3-512 | Conv3-512 | 512 × 512 × 3 | 512 × 512 × 3 |
| FC1024 | LSTM32 | 512 × 12 × 1,024 | (512 + 32) × 32 × 4 |
| FC256 | LSTM32 | 1,024 × 256 | (32 + 32) × 32 × 4 |
| FC9 | Attention | 256 × 9 | 32 × 32 + 32 × 9 + 32 × 1 × 9 |
| **Total** | | **11,461,120** | **4,984,640** |

series to update its internal state vector and gave outputs, while fully connected layers in attention module iterated through the series to give weights for each segment. The experiment result demonstrated that parameter reduction based on property of input signal didn't sacrifice the model's ability to learn, but helped the model to achieve better performance. A more parameter-efficient model structure would be less prone to overfit, better in generalizability and also less memory-consuming.

### 5.4. Incremental inference

Apart from parameter reuse, LSTM cells in ATI-CNN made this model in its innate design support incremental inference. In real-time processing of an ECG signal, the LSTM part of ATI-CNN unwrap along the time axis. Previous information concerning a record will be saved in LSTM as its cell state vector, and the update of prediction doesn't require the model to draw its inference from scratch. As a result, ATI-CNN only need to process a signal segment with a length of 243 data points to generate a new prediction, while the VGG-6 needs to inference from scratch with an input of a 6-second-long signal, or 3000 data points, when sampling frequency is 500 Hz.

According to Molchanov et al. [37], we can compute the number of floating-point operations (FLOPs) for convolutional kernels as follows:

$$FLOPs = 2HW(C_{in}K^2 + 1)C_{out}, \tag{9}$$

where $H$, $W$ and $C_{in}$ are height, width and number of channels of the input feature map, $K$ is the kernel width, and $C_{out}$ is the number of output channels. For fully connected layers we have:

$$FLOPs = (2I - 1)O, \tag{10}$$

where $I$ is the input dimensionality and $O$ is the output dimensionality. As gates inside LSTM cells are special forms of fully connected layers, this formula could also be used on LSTM cells. Using formula above, we are able to calculate the computation needed by each layer in our model, and therefore estimate the computation cost of the entire model.

Consistent with the result in [38], the calculation needed by VGG16 model for a single inference is about 30 GFlops. However, for a VGG-6 model modified for signal, the computational complexity is greatly reduced to be approximately 1.1 GFlops. This is because that for deep learning on 1-D signal, only 1-D convolution is involved, resulting in the formula to be modified as:

$$FLOPs = 2L(C_{in}K + 1)C_{out}, \tag{11}$$

where $L$ indicates the length of input signal. As shown in [39], current mobile devices on market need about 100 to 200 ms to finish single image classification using MobileNet, which requires about 1.14 GFlops of calculation. It is thereby demonstrated that deep convolutional neural network for 1-D signal has the potential to run on mobile devices in real time.

What's more, in ATI-CNN the input for each layer is scaled down by a factor of 12, and the calculation needed is correspondingly reduced. It could be expected that ATI-CNN would work very well on mobile devices without consuming too much computing power.

## 6. Conclusion

In this paper, a model named ATI-CNN for arrhythmia classification from varied-length ECG signals was proposed. This model extracted information from ECG signals in two steps: spatial information fusion based on convolutional neural network and temporal information fusion based on LSTM cells and an attention module. These modules with different functions were incorporated into a uniform neural network architecture and formed an end-to-end trainable model. With this novel network architecture, ATI-CNN is in its nature compatible with varied-length input, leading to its superior classification performance, especially in detecting paroxysmal arrhythmias. It achieved an average F1-score of 81.2% in classification of 8 types of arrhythmias and sinus rhythm, which exceeded reference CNN model by 7.7%. Attention mechanism helps the model to locate informative part of signals and improves interpretability. Halved parameter amount makes ATI-CNN less memory-hungry and less prone to overfit, and ability to memorize brought by recurrent cells makes ATI-CNN much efficient in real-time processing. Other types of abnormal ECG patterns, like trigeminy, T-wave alternans, are not included in the evaluation of ATI-CNN. Nevertheless, it is expected that ATI-CNN, as a fully data-driven method, could also perform well in detecting these arrhythmias once corresponding data is collected. All in all, ATI-CNN was an excellent solution to the problem of multi-class arrhythmia classification from varied-length 12-lead ECG signals. It provided an exemplification for other signal processing problems where spatial and temporal fusion of signal was concerned.

## Acknowledgments

This work was supported by Joint Engineering Research Center for Health Big Data Intelligent Analysis Technology, and in part of Major Special Project of Guangdong Province (2017B030308007), Basic Research discipline Planning in Shenzhen (JCYJ20170413161515911), Special Fund Project for Innovation of High-level Overseas Talents (KQJSCX20170731165939298), Ph.D Start-up Fund Project for Natural Science of Guangdong Province (2018A030310006). We also appreciate Prof.Chengyu Liu from Southwestern University (China) for the experiment database he and his group provided for this research.

## Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.inffus.2019.06.024.

## References

[1] S. Mendis, P. Puska, B. Norrving, W.H. Organization, et al., Global Atlas on Cardiovascular Disease Prevention and Control, Geneva: World Health Organization, 2011.
[2] R. Mehra, Global public health problem of sudden cardiac death, J. Electrocardiol. 40 (6) (2007) S118–S122.
[3] C. Van Mieghem, M. Sabbe, D. Knockaert, The clinical value of the ecg in noncardiac conditions, Chest 125 (4) (2004) 1561–1576.
[4] T. Stamkopoulos, K. Diamantaras, N. Maglaveras, M. Strintzis, Ecg analysis using nonlinear pca neural networks for ischemia detection, IEEE Trans. Signal Process. 46 (11) (1998) 3058–3067.
[5] P. Leijdekkers, V. Gay, A self-test to detect a heart attack using a mobile phone and wearable sensors, in: 2008 21st IEEE International Symposium on Computer-Based Medical Systems, IEEE, 2008, pp. 93–98.
[6] D.S. Goldstein, The electrocardiogram in stroke: relationship to pathophysiological type and comparison with prior tracings., Stroke 10 (3) (1979) 253–259.

[7] G.B. Moody, R.G. Mark, The impact of the mit-bih arrhythmia database, IEEE Eng. Med. Biol. Mag. 20 (3) (2001) 45–50.
[8] A.-A. EC57, Testing and reporting performance results of cardiac rhythm and st segment measurement algorithms, Assoc. Adv. Med. Instrum. Arlington, VA (1998).
[9] C.-C. Lin, C.-M. Yang, Heartbeat classification using normalized rr intervals and morphological features, Math. Probl. Eng. 2014 (2014).
[10] H. Huang, J. Liu, Q. Zhu, R. Wang, G. Hu, A new hierarchical method for inter–patient heartbeat classification using random projections and rr intervals, Biomed. Eng. Online 13 (1) (2014) 90.
[11] H. Dai, L. Yin, Y. Li, Qrs residual removal in atrial activity signals extracted from single lead: a new perspective based on signal extrapolation, IET Signal Process. 10 (9) (2016) 1169–1175.
[12] O. Perlman, A. Katz, G. Amit, Y. Zigel, Supraventricular tachycardia classification in the 12-lead ecg using atrial waves detection and a clinically based tree scheme, IEEE J. Biomed. Health Inform. 20 (6) (2016) 1513–1520.
[13] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (7639) (2017) 115.
[14] D. Wang, A. Khosla, R. Gargeya, H. Irshad, A.H. Beck, Deep learning for identifying metastatic breast cancer, arXiv preprint arXiv:1606.05718(2016).
[15] R. Li, W. Zhang, H.-I. Suk, L. Wang, J. Li, D. Shen, S. Ji, Deep learning based imaging data completion for improved brain disease diagnosis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2014, pp. 305–312.
[16] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, D. Feng, Early diagnosis of Alzheimer's disease with deep learning, in: 2014 IEEE 11th international symposium on biomedical imaging (ISBI), IEEE, 2014, pp. 1015–1018.
[17] P. Rajpurkar, A.Y. Hannun, M. Haghpanahi, C. Bourn, A.Y. Ng, Cardiologist-level arrhythmia detection with convolutional neural networks, arXiv preprint arXiv:1707.01836(2017).
[18] X. Fan, Q. Yao, Y. Cai, F. Miao, F. Sun, Y. Li, Multi-scaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ecg recordings, IEEE J. Biomed. Health Inform. (2018). 1–1
[19] S. Kiranyaz, T. Ince, M. Gabbouj, Real-time patient-specific ecg classification by 1-d convolutional neural networks, IEEE Trans. Biomed. Eng. 63 (3) (2016) 664–675.
[20] U.R. Acharya, H. Fujita, S.L. Oh, Y. Hagiwara, J.H. Tan, M. Adam, Application of deep convolutional neural network for automated detection of myocardial infarction using ecg signals, Inf. Sci. 415 (2017) 190–198.
[21] K. Li, W. Pan, Y. Li, Q. Jiang, G. Liu, A method to detect sleep apnea based on deep neural network and hidden Markov model using single-lead ecg signal, Neurocomputing 294 (2018) 94–101.
[22] S. Chauhan, L. Vig, Anomaly detection in ecg time signals via deep long short-term memory networks, in: 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2015, pp. 1–7.
[23] K.-K. Tseng, D. Lee, C. Chen, Ecg identification system using neural network with global and local features., Int. Assoc. Dev. Inf. Soc. (2016).
[24] L. Jordaens, A clinical approach to arrhythmias revisited in 2018, Netherlands heart journal : monthly journal of the Netherlands Society of Cardiology and the Netherlands Heart Foundation, 2018.
[25] R. Gravina, P. Alinia, H. Ghasemzadeh, G. Fortino, Multi-sensor fusion in body sensor networks: state-of-the-art and research challenges, Inf. Fusion 35 (2017) 68–80.
[26] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556(2014).
[27] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167(2015).
[28] V. Nair, G.E. Hinton, Rectified linear units improve restricted Boltzmann machines, in: Proceedings of the 27th international conference on machine learning (ICML-10), 2010, pp. 807–814.
[29] M. Ilse, J.M. Tomczak, M. Welling, Attention-based deep multiple instance learning, CoRR abs/1802.04712 (2018).
[30] T. DeVries, G.W. Taylor, Improved regularization of convolutional neural networks with cutout, arXiv preprint arXiv:1708.04552(2017).
[31] H. Shimodaira, Improving predictive inference under covariate shift by weighting the log-likelihood function, J. Stat. Plan. Inference 90 (2) (2000) 227–244.
[32] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, in: Proceedings of the IEEE international conference on computer vision, 2015, pp. 1026–1034.
[33] A.M. Saxe, J.L. McClelland, S. Ganguli, Exact solutions to the nonlinear dynamics of learning in deep linear neural networks, arXiv preprint arXiv:1312.6120(2013).
[34] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv preprint arXiv:1412.6980(2014).
[35] 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.
[36] L. Xie, J. Wang, Z. Wei, M. Wang, Q. Tian, Disturblabel: Regularizing CNN on the loss layer, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4753–4762.
[37] P. Molchanov, S. Tyree, T. Karras, T. Aila, J. Kautz, Pruning convolutional neural networks for resource efficient inference, arXiv preprint arXiv:1611.06440(2016).
[38] A. Canziani, A. Paszke, E. Culurciello, An analysis of deep neural network models for practical applications, arXiv preprint arXiv:1605.07678(2016).
[39] A. Ignatov, R. Timofte, W. Chou, K. Wang, M. Wu, T. Hartley, L. Van Gool, Ai benchmark: running deep neural networks on android smartphones, in: European Conference on Computer Vision, Springer, 2018, pp. 288–314.