Deep Multi-Scale Fusion Neural Network for Multi-Class Arrhythmia Detection

Ruxin Wang, Jianping Fan, and Ye Li, *Senior Member, IEEE*

***Abstract*—Automated electrocardiogram (ECG) analysis for arrhythmia detection plays a critical role in early prevention and diagnosis of cardiovascular diseases. Extracting powerful features from raw ECG signals for fine-grained diseases classification is still a challenging problem today due to variable abnormal rhythms and noise distribution. For ECG analysis, the previous research works depend mostly on heartbeat or single scale signal segments, which ignores underlying complementary information of different scales. In this paper, we formulate a novel end-to-end Deep Multi-Scale Fusion convolutional neural network (DMSFNet) architecture for multi-class arrhythmia detection. Our proposed approach can effectively capture abnormal patterns of diseases and suppress noise interference by multi-scale feature extraction and cross-scale information complementarity of ECG signals. The proposed method implements feature extraction for signal segments with different sizes by integrating multiple convolution kernels with different receptive fields. Meanwhile, joint optimization strategy with multiple losses of different scales is designed, which not only learns scale-specific features, but also realizes cumulatively multi-scale complementary feature learning during the learning process. In our work, we demonstrate our DMSFNet on two open datasets (CPSC_2018 and PhysioNet/CinC_2017) and deliver the state-of-art performance on them. Among them, CPSC_2018 is a 12-lead ECG dataset and CinC_2017 is a single-lead dataset. For these two datasets, we achieve the F1 score 82.8% and 84.1% which are higher than previous state-of-art approaches respectively. The results demonstrate that our end-to-end DMSFNet has outstanding performance for feature extraction from a broad range of distinct arrhythmias and elegant generalization ability for effectively handling ECG signals with different leads.**

***Index Terms*—Deep learning, ECG, multi-scale fusion, convolutional neural network.**

I. Introduction

CARDIOVASCULAR diseases are the leading cause of death and disability on a global scale. It is an important cause of death and disability, which seriously affects people's health. Recently, the World Health Organization (WHO) announces the top ten health threats in the world in 2019. Heart disease is as a typical non-infectious disease on the list. Because of the difficulty in curing, early screening and treatment are particularly important. Electrocardiogram (ECG) is an essential tool which can record the electrical activity of the heart over a period of time (Fig. 1). Every year there are more than 300 million clinical ECG records in global hospitals. ECG is the most basic, convenient and economical routine examination approach. It is very commonly performed for clinical medical screening of many cardiac diseases, such as judging arrhythmia, diagnosing myocardial ischemia, reflecting the structure of the heart, and provides important reference information for clinicians [1], [2]. With the emerging of Healthcare 4.0 and the development of Artificial Intelligence (AI), the importance of automatic diagnosis has become increasingly prominent. Automated analysis of ECG not only provides auxiliary diagnostic information, but also can monitor hearts situation for 24 hours, which is also beneficial for mobile medical and remote diagnosis.

Over the past decade, a large number of automatic analysis algorithms have been introduced [4]–[8]. Although these methods improve the accuracy of ECG signal classification through reasonably combining feature extraction and classifier, they still have some common defects: 1) They must rely on experts to design and extract the characteristics of ECG signals, other potential information in the original signal is neglected. 2) The artificial definition of different diseases characteristics may be slightly different, therefore the generalization ability of the model is restricted. 3) At the same time, as the feature dimension increases, the choice of model parameters has also become more difficult.

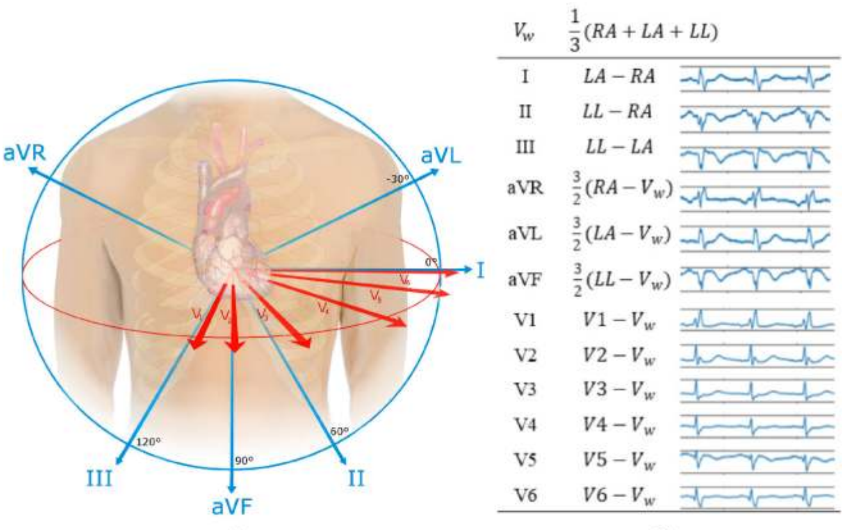
With the development of deep learning, it has achieved outstanding performance in ECG processing [9]–[14]. Deep neural network realizes the effective combination of feature extraction and disease classification through end-to-end learning. However, to automatically detect variable heart arrhythmias from ECG signals, an algorithm must implicitly recognize the distinct wave

Manuscript received July 17, 2019; revised January 29, 2020 and March 5, 2020; accepted March 11, 2020. Date of publication April 13, 2020; date of current version September 3, 2020. This work was supported in part by National Natural Science Foundation of China under Grants U1913210 and 61771465, in part by the Strategic Priority CAS Project under Grant XDB38000000, in part by the Major Special Project of Guangdong Province under Grant 2017B030308007, in part by the Major Projects from General Logistics Department of People's Liberation Army under Grant AWS13C008, in part by the Shenzhen Basic Research Projects under Grant JCYJ20180703145202065, in part by the Shenzhen Science and Technology Innovation Project under Grant JSGG20170823144843046, and in part by the Special Fund Project for Innovation of High-level Overseas Talents under Grant KQJSCX20170731165939298. *(Corresponding authors: Jianping Fan; Ye Li.)*

Ruxin Wang and Ye Li are with the Joint Engineering Research Center for Health Big Data Intelligent Analysis Technology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China (e-mail: rx.wang@siat.ac.cn; ye.li@siat.ac.cn).

Jianping Fan is with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China (e-mail: jp.fan@siat.ac.cn).

Digital Object Identifier 10.1109/JBHI.2020.2981526



(a) (b)

Fig. 1. Illustration for 12-lead ECG system. (a) Spatial orientation of ECG leads [3]. (b) Examples of ECG signals in 12 leads.

types and discern the complex relationships between them. It is difficult due to the variability in wave morphology of different diseases as well as the presence of noise. At present, most of these models based on deep learning mainly use single-scale convolution filters for feature extraction, ignoring other potentially useful information of different scales. Furthermore, they are hard to utilize the implicit correlated complementary advantages across scales for enhancing the recognition performance.

Specifically, small-scale convolution filters are suitable for extracting amplitude and local statistical information from signals, such as the amplitude of P, R, S and T waves of ECG. And the large-scale convolution filters with larger receptive fields are more better at encoding the interval information between different waves and some morphological features, such as P-R interval, R-R interval and QRS duration etc. All these features are crucial for analyzing the ECG signal. For example, the typical performances of atrial fibrillation are the disappearance of P wave at small scale and irregular RR interval at large scale.

In this paper, we formulate a novel Deep Multi-Scale Fusion convolutional neural network architecture for ECG classification by multi-scale features optimized simultaneously integrating multi-loss learning. This is significantly different from existing ECG detection approaches considering only heartbeat or single-scale signal information. The main contributions of this paper are summarized as follows:

1) For ECG analysis, we investigate the multi-scale feature learning problem for multi-class arrhythmia detection. Cross-scale features of segments with different sizes are extracted by multiple convolution kernels with different receptive fields. In addition, spatial attention is utilized for further mining the discriminative information.
2) Joint optimization strategy with multiple losses of different scales is designed, which not only learns scale-specific features, but also realizes cumulatively multi-scale complementary feature learning during the learning process.
3) Finally, we evaluate the proposed method for ECG classification on two public ECG datasets (CPSC_2018 [15] and PhysioNet/CinC_2017 [16]) and compare it with state-of-the-art methods. The experimental results convince the effectiveness and efficiency of the proposed method.

The rest of this paper is organized as follows: Section II is the related work in ECG processing. Section III we propose a new end-to-end deep multi-scale fusion CNN architecture for ECG classification. Section IV presents experimental results with different methods on the two ECG datasets. Section V gives the discussion of our proposed method. Finally, we conclude the paper in Section VI.

II. Related Work

Traditional ECG analysis methods, such as ECG-based disease classification, mainly consist of two parts: feature extraction and classifier training. The first and the most important step is feature extraction, which need be manually designed and extracted from raw signal. Early approaches mostly rely on classical waveform features, such as amplitudes, hermite coefficients [4], morphological features [5], heartbeat interval features [6] etc. After that, some new features such as time-frequency, wavelet, high-order statistics and other factors based on the detection of waveform features are employed. In order to further mine the effective information, some commonly extraction algorithms are used including wavelet decomposition [7], principal component analysis (PCA) [8], Kalman filter [17] and some statistical methods [18]. In terms of classification, different learning algorithms have been well studied, containing support vector machines (SVM) [19], artificial neural networks (ANN) [20], and Hidden Markov Models (HMM) [21] and so on [22], [23].

In recent years, deep learning have achieved remarkable performance in various fields of medicine, such as medical image processing [24], genomic analysis [25], electronic health records analysis [26] and physiological signal analysis [27]. Deep neural network can form more abstract high-level features by reasonably designing multi-layer and non-linear network structure. Meanwhile, some novel multi-scale-based methods have been designed and proposed in many computer vision task and achieved outstanding results compared to single-scale methods [28]–[31], which illustrates the advantages of the multi-scale approach.

For ECG signal processing, many achievements have been made based on deep learning over these few years. Most of the studies focused on ECG-based auxiliary diagnosis and signal analysis of heart disease [32]–[34]. Kiranyaz *et al.* [9] proposed an adaptive 1-D Convolutional Neural Networks (CNN) model which is used for both feature extraction and classification of the raw ECG data from each individual patient. Rahhal *et al.* [10] used stacked denoising autoencoder with sparsity constraint for active classification of ECG signals which provided significant accuracy improvements with less expert interaction. Li *et al.* [11] implemented a parallel general regression neural network to classify the heartbeat for long-term ECG signal. Baloglu *et al.* [12] proposed an end-to-end deep learning model using the standard 12-lead ECG signal for the diagnosis of myocardial infarction and achieved high performance on myocardial infarction detection. Pranav Rajpurkar *et al.* [13] adopted a 34-layers CNN model to classify 12 rhythm categories using 91232 ECG signals recorded by a single-lead Holter monitoring device from 53549 patients. The classification results were compared with human experts, which displayed the similar diagnostic performance with human. Bahareh Taji *et al.* [14] used deep belief networks to reduce the false alarm rate caused by poor-quality ECG signal measurement during atrial fibrillation recognition. In addition, ECG-based assisted diagnosis of other diseases also has some research results [35], [36]. For example, Hirotaka Kaji *et al.* [35] employed a multi-task learning technique to predict the degree of concentration with heart-rate features and significantly improved the accuracy of concentration prediction in small samples situations.

III. Methodology

A. Problem Formulation

ECG-based disease detection belongs to the time series classification problem. Given a set of ECG signals and their corresponding disease labels, the target is that judging the ECG records belong to what kind of cardiac diseases. In this paper, we aim to learn a deep representation for ECG records and use them for end-to-end disease classification. For simplicity, we define $D = \{(x_i, y_i)|i = 1, 2, \ldots, N\}$ as the ECG data set. where x_i indicates one ECG signal with length l_i. $y_i \in \{1, \ldots, C\}$ denotes the corresponding category of x_i, and C is the number of disease categories. N refers to the total number of samples. In order to get meaningful feature representation of records, we formulate a Deep Multi-Scale Fusion (DMSF) CNN architecture for capturing discriminative signal features from multiple scales. Then, this powerful features are directly used for classification. Mathematically, it can be described by minimizing the cross-entropy between the reference labels and outputs.

TABLE I
The Network Structure for Proposed Method

Configurations of DMSFNet	
branch-1	**branch-2**
input (raw ECG signal)	
backbone CNN	
conv3_64_1 conv3_64_1	
max-pooling	
conv3_128_1 conv3_128_1	
max-pooling	
conv3_256_1 conv3_256_1 conv3_256_1	
max-pooling	
conv3_512_1 conv3_512_1 conv3_512_1	conv3_512_3 conv3_512_3 conv3_512_3
max-pooling	
conv3_512_1 conv3_256_1 conv3_128_1	conv3_512_3 conv3_256_3 conv3_128_3
fusion	
global-pooling	

note: The convolutional parameters are denotes as ”conv(kernel size)_(number of filters)_(dilation rate)”. Padding operation is adopted for maintaining the previous size in all convolutional layers. And the pooling window is set as 3 with stride 3.

B. Model Overview

The proposed DMSFNet is composed of three main components: 1) Backbone network for learning shared low-level features; 2) multiple sub-networks to learn the high-level scale-specific signal features using different scales convolution kernels collaboratively; 3) multi-scale features fusion for integrating features from sub-networks and further discovering correlated complementary informations from different scales by using attention. Meanwhile, the joint multi-loss optimization strategy is adopted for simultaneously optimizing multi-branch feature representation and realizes cumulatively multi-scale complementary feature learning during the learning process. Specifically, the backbone network is built based on the VGG net due to its powerful data representation ability. Table I shows the configuration of our proposed network. We use the first seven convolution layers for shared learning and obtain the feature maps f_b. Then the features f_b are fed into two sub-branches with six convolution layers in each branch to extract scale-specific feature maps f_{b_1} and f_{b_2}. Concatenate different feature maps f_{b_1}, f_{b_2} and adopt attention to obtain the fusion cross-scale features F. At last, all the learning features including the fusion and single branch features are employed for multi-loss optimization. The overall framework design is illustrated in Fig. 2.

In particular, for obtain the multi-scale receptive fields, dilated convolution is adopted in this paper, which has been demonstrated in solving many computer vision task with significant performance [37], [38]. To make it clearer, we use a 3×3 2-D convolution kernel as an example to describe the operation (Fig. 3). Mathematically, a 2-D dilated convolution can be

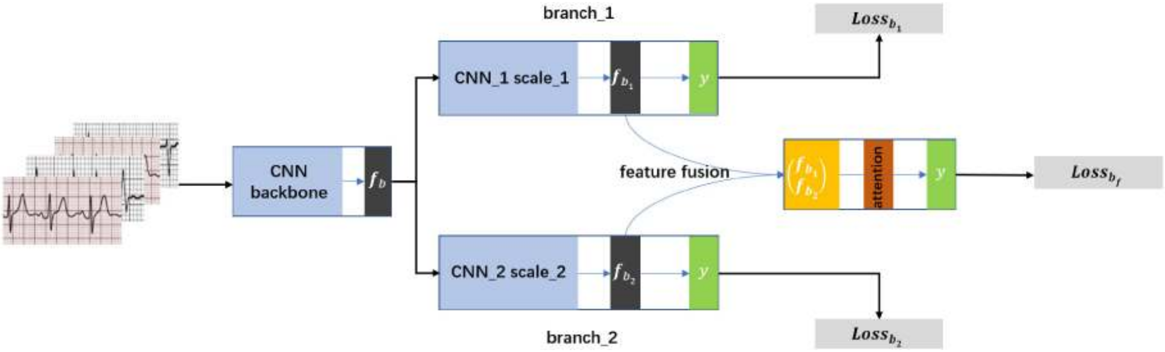


Fig. 2. Overview of the proposed model architecture, which consists a backbone network, two different scale-specific networks and one multi-scale feature fusion branch.

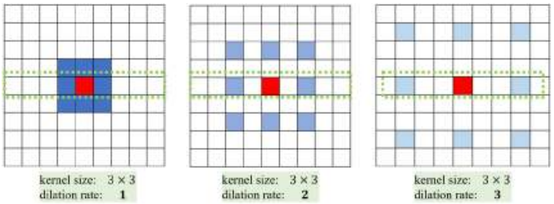


Fig. 3. 3×3 convolution kernels with different dilation rate. The green dotted frame indicates 1-D convolution situation.

defined as follows:

$$y[i,j] = \sum_p \sum_q x[i + r \cdot p, j + r \cdot q] w[p,q] \tag{1}$$

where r denotes the dilation rate, $w[p,q]$ is the convolution parameters, $[i,j]$ is the center of convolution, $y[i,j]$ indicates the output of convolution from input $x[i,j]$. Classic convolution kernels with different sizes would overlap to a certain extent at the same output position and produce redundant parameters. The dilated convolutions dramatically reduce the redundant repeated information using filters with holes. For different dilation rates 1, 2, 3, the 3×3 dilated convolutions instead of classic convolution kernels with size 3×3, 5×5 and 7×7. As shown in Fig. 3, the red patch denotes the center position, and blue patch refers to the convolution area. We can find that the convolution have larger receptive field with larger dilation rate. And the number of parameters has not been increased, which reduces the duplicate convolution for overlapped areas.

C. Single Scale Feature Learning

The shared feature maps by the backbone are fed into the different scale branch. We construct the single-scale branch using six layers CNN framework. Specifically, the first three convolution layers use the same number of convolution kernels for effectively extracting the signal features. And the number of convolution kernels in the latter three layers decreases step by step. It can continue to extract high-level features and reduce features dimension very well. Specific configuration details refer to Table I.

For the input $x_i, i \in \{1, 2, \ldots, N\}$, the branch output $f_{b_j}, j \in \{1, 2\}$ can be defined as:

$$f_{b_j} = N_{b_j}(N_b(x_i; \theta_b); \theta_{b_j}) \tag{2}$$

where f_{b_j} denotes the branch feature of raw input x_i, N_b and $N_{b_j}, j \in \{1, 2\}$ denote the backbone network and scale-specific sub-network. θ_b and θ_{b_j} are the network parameters, respectively.

ECG classification is a multi-class classification problem. In this paper, the softmax loss is employed for single branch model training. According to the top output feature map f_{b_j} of b_jth branch, the global max-pooling is first adopted for squeezing the features dimension, which produces a reduced dimension feature embedding for each sample. Then the posterior probability of each class is calculated:

$$z_{b_j} = g_m(f_{b_j}) \tag{3}$$

$$p(z_{b_j}) = \frac{\exp(w_{y_i}^\top z_{b_j})}{\sum_{k=1}^{C} \exp(w_k^\top z_{b_j})} \tag{4}$$

where g_m denotes the global max-pooling operation, $p^{(b_j)}$ is the probability that model assigns the label y_i to the input x_i, and w_k is the parameter of class k. Therefore, for all the observable instances in the training set, the objective lost function can be defined as:

$$L_{b_j} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{C} I\{y_i = k\} \log p(z_{b_j}) \tag{5}$$

where $I(\cdot)$ is the indicator function, so that $I(true) = 1$, and otherwise is 0.

D. Multi-Scale Feature Fusion Learning

For achieving cross-scale information complementation and obtaining robust features for classification, we first obtain the fusion feature maps F with c channels by concatenating the multiple scale-specific features f_{b_j}:

$$F = Cat(f_{b_1}, f_{b_2}) \tag{6}$$

where Cat is the concatenation operation. Then a spatial attention module is adopted for further mining the discriminative features and improving the performance. In this work, a global

TABLE II
CLASSIFICATION PERFORMANCE ON CPSC_2018 DATASET

Type	**F1 score**								
	Resnet [39]	***VGG*** [40]	***LSTM*** [41]	***Acharya et al.*** [42]	***Fan et al.*** [43]	***Yao et al.*** [44]	***Liu et al.*** [1] [47]	***Liu et al.***[2] [47]	***Ours***
N	0.76	0.78	0.75	0.70	0.79	0.75	0.80	**0.82**	**0.82**
AF	**0.92**	0.89	0.91	0.91	**0.92**	0.90	0.89	0.91	0.90
I-AVB	0.82	0.81	0.77	0.75	0.81	0.81	**0.87**	**0.87**	0.86
LBBB	0.84	0.85	0.86	0.85	**0.87**	**0.87**	0.77	**0.87**	**0.87**
RBBB	0.92	0.91	0.92	0.91	0.92	0.92	0.90	0.91	**0.93**
PAC	0.62	0.70	0.57	0.72	0.76	0.64	0.65	0.63	**0.78**
PVC	0.84	0.83	0.79	0.84	0.83	0.83	0.79	0.82	**0.88**
STD	0.75	0.75	0.73	0.75	0.77	0.76	**0.80**	**0.80**	**0.80**
STE	0.52	0.54	0.53	0.44	0.50	0.46	0.56	0.60	**0.62**
Average									
Precision	0.789	0.815	0.784	0.773	0.831	0.799	-	-	**0.838**
Recall	0.767	0.768	0.743	0.771	0.781	0.758	-	-	**0.822**
F1 score	0.776	0.785	0.758	0.761	0.797	0.772	0.780	0.810	**0.828**

note: *Liu et al.*[1] indicates their proposed network model without expert features. *Liu et al.*[2] refers to their method with expert features and deep features.

feature map S is first obtained by global average pooling operation at each spatial location u of F:

$$S_u = \frac{1}{c} \sum_{k=1}^{c} F_{u,k} \tag{7}$$

Then we use a 1×1 convolution and sigmoid function to S and produce the spatial attention map f_{att}. Thus, the new fusion features f_{b_f} can be obtained by summing with the weighted features. The details can be expressed as follows:

$$f_{att} = \sigma(W * S + b) \tag{8}$$

$$F = F + f_{att} \otimes F \tag{9}$$

where $\sigma(\cdot)$ denotes the sigmoid function and $\otimes$ indicates the channel-wise product operation. At last, a global pooling layer is adopted for integrating features from different convolutional channels and squeezing the features dimension. In this paper, both the global max-pooling and average-pooling are used for the fusion feature maps F. Max-pooling can effectively extract the specific and discriminative information of signals by extracting the maximum value in each region. And average-pooling is more conducive to extracting global information of the signal by average operation:

$$z_{b_f} = g_m(F) + g_a(F) \tag{10}$$

where g_a refers to the global average-pooling operation. Then features z_{b_f} are adopted for prediction. In our work, we similarly utilize the softmax classification loss as the objective function. The details can be expressed as follows:

$$L_{b_f} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{C} I\{y_i = k\} \log p(z_{b_f}) \tag{11}$$

E. Joint Optimization With Multiple Losses

Critically, the scale-specific branches are not independent but related to each other. In order to learn effective and discriminative classification features, we train the whole model by jointly optimizing the losses of multiple branches.

Optimization for each branch aims to maximize scale-specific feature discriminative capability by supervision, whilst optimization for scale-fusion branch is designed to concurrently optimize the potential complementary information across scales.

Based on the above considerations, we use the joint optimization for model training, which balances between individual learning and correlation learning. Compared with the scale-specific learning, it can optimize multiple classification loss on same ECG label information concurrently. Importantly, the model parameters are optimized by back propagation to all individual branches, which not only learns scale-specific features, but also realizes cumulatively multi-scale complementary feature learning during the learning process. Through joint learning in an end-to-end fashion, the model maximizes the scale-specific feature learning and discriminative selection from multi-scale representations for arrhythmia detection. Thus the robustness of the model is improved effectively in training stage and achieves a better classification performance. For the overall network training, The final objective function is as follows:

$$L = L_{b_f} + \lambda_1 L_{b_1} + \lambda_2 L_{b_2} \tag{12}$$

where λ_1, λ_2 are the balance parameters which are set to 1.0 in our experiments.

IV. EXPERIMENT

In this section, implementation details and experiments are given. We choose two ECG data set for validating the proposed method. The final results are shown in Table II–III.

TABLE III
CLASSIFICATION PERFORMANCE ON CINC_2017 DATASET

Type	F1 score						
	Resnet [39]	*VGG* [40]	*LSTM* [41]	*Acharya et al.* [42]	*Fan et al.* [43]	*Datta et al.* [45]	*Ours*
N	0.89	0.88	0.90	0.89	0.90	0.90	**0.92**
AF	0.72	0.75	0.68	0.72	0.77	0.79	**0.83**
Other	0.71	0.69	0.71	0.69	0.74	0.77	**0.78**
Average							
Precision	0.798	0.800	0.813	0.795	0.826	-	**0.856**
Recall	0.761	0.764	0.734	0.752	0.793	-	**0.829**
F1 score	0.777	0.779	0.762	0.770	0.807	0.826	**0.841**

TABLE IV
DATA DETAILS FOR CPSC_2018 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

TABLE V
DATA DETAILS FOR CINC_2017 DATASET

Type	Records	Time length(s)				
		Mean	*SD*	*Min*	*Median*	*Max*
N	5154	31.90	10.00	61.00	30.00	9.00
AF	771	31.60	12.50	60.00	30.00	9.10
O	2557	31.40	11.80	60.00	30.00	10.20
noise	46	27.10	9.00	60.00	30.00	9.00
Total	8528	32.50	10.90	61.00	30.00	9.00

A. Data Description

1) CPSC_2018 Dataset: This dataset was from the China Physiological Signal Challenge (CPSC 2018). The data is collected from 11 hospitals containing 6,877 12-lead ECG records (female: 3,178; male: 3,699) for training and 2,954 records for testing. The ECG records are sampled as 500 Hz, and the signal length of the data is from 6 s to 60 s. The labels of these records include one normal type and eight abnormal types, which are detailed as: Normal (N), Atrial fibrillation (AF), First-degree atrioventricular block (I-AVB), Left bundle brunch block (LBBB), Right bundle brunch block (RBBB), Premature atrial contraction (PAC), Premature ventricular contraction (PVC), ST-segment depression (STD) and ST-segment elevated (STE). Table IV shows the details of the data.

2) PhysioNet/CinC_2017 Dataset: This dataset contains 8,528 single lead ECG records lasting from 9 s to just over 60 s, and ECG records were sampled as 300 Hz. All the signals were manually labeled by ECG experts into Normal rhythm, Atrial fibrillation rhythm, Other rhythm and noisy recordings. In this paper, only Normal (N), Atrial fibrillation (AF) and Other rhythm (O) are used for classification. Table V shows the details of the data.

B. Reference Model

To evaluate the proposed model's performance, we choose some common network structures and state-of-art ECG classification algorithms for comparison.

Baselines: In our experiments, three common deep neural network frameworks, Resnet [39], VGG [40] and LSTM [41] are adopted for performance comparison. Both the Resnet and VGG net are classical convolutional neural network for processing images and signals. The Resnet designs a residual learning framework by shortcut identity connections to ease the training of very deep networks and make feature maps from shallower layers available at later stages. The VGG is also a classical CNN containing multiple convolutional and fully connected layers. The LSTM is a variant of recurrent neural network, which is designed for time series processing. In addition, two state-of-the-art ECG analysis methods are also used for testing. Acharya *et al.* [42] implemented a 11-layers convolutional neural network algorithm for the automated detection of a normal and myocardial infarction ECG signals. And Fan *et al.* [43] proposed a multi-scale CNN (MS-CNN) for screening out AF recordings from ECG records. Both methods have achieved excellent classification results at present for ECG classification task.

CPSC_2018: For the CPSC_2018 dataset, several latest reported algorithms are also compared. Yao *et al.* [44] proposed a time-incremental convolutional neural network (TI-CNN) using the spatial-temporal network framework which consists of multiple convolutional layers for feature extracting and a Long Short-Term Memory (LSTM) layer for time-series processing and classification. Liu *et al.* [47] used extracted expert features and deep features by a modified Resnet framework, CL3, containing 17 layers of convolution and a fully connected layer for 12-lead ECGs classification.

CinC_2017: We also choose related algorithm for classifying normal, AF and other signals on the Cinc_2017 dataset. Datta *et al.* [45] introduced a multi-layer cascaded binary classifier instead of a single multi-class classifier with about

150 features (including morphological ECG features, prior art AF features, HRV features Frequency features and statistical features etc.).

C. Evaluation Criteria

In this paper, the average precision, recall rate and F1 score are adopted for measuring the classification performances. The details is as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{13}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{14}$$

$$F1 = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \tag{15}$$

For a certain class in multi-classification problem, TP is the true positives which indicates the number of correctly classified samples in this class, FN is the false negatives which refers to the number of samples belonging to this class which are misclassified into other classes, and FP denotes the false positives which indicates the number of samples misclassified in this class. The average of three metrics among classes were calculated to give a final evaluation. Among these metrics, $F1$ score mainly assesses the recognition effect, which is the most important evaluation metric in these two datasets.

D. Implementation Details

Training setting: All the experiments in this paper are implemented based on Pytorch (http://pytorch.org/). In the model, we use the stochastic gradient descent optimizer (SGD) with 0.9 momentum for training in each mini-batch and update the parameters. ReLU is chose as the default activation function. For fair comparison, we initialize our model using the Kaiming initializer [39] and set the initial learning rate as 10^{-3}. The learning rate is divided by 10 at 150 and 200 epochs, eventually terminated at 250 epochs. All the training data is divided into mini-batches for network training, the mini-batch size is set as 128 during the training stage.

Data preprocessing: As shown in Table IV–V, the different kinds of diseases' ECG signal show imbalanced sample distribution. In addition, the length of the signal varies from few seconds to 60 seconds. So several data augmentation and padding/sampling strategies are used before training the model. For data augmentation, we first use horizontal and vertical flip operation to expand small sample data, such as the samples of LBBB, PAC, PVC and STE in CPSC_2018 dataset, the samples of AF and other signals in CinC_2017 dataset. Secondly, to further expand the sample number and increase sample diversity, we add random noise and use random erasure strategy [46] to the original samples for data augmentation. These method have been proved to be effective in expanding sample data and improving robustness of model.

Input with same length is necessary for model training. In that the length of the data varies from few seconds to 60 seconds, padding operation is applied to fix input length. Firstly, we fill all the data into 60 s using replication strategy, and then cut out 50 s length data from the padded signal as training data randomly.

Besides, ECG signals among different individuals as well as different lead positions tend to have large variation of amplitudes, which affects model performance greatly. In this work, all padded records are normalized to zero mean and unit standard deviation in training stage, which would help the model to converge faster.

E. Results

1) Evaluation on CPSC_2018 Dataset: Table II compares the class-level F1 score, average precision, average recall and average F1 score of eight reference models and our work in identifying cardiac arrhythmias. As can be observed, the proposed DMSFNet performs favorably against other counterparts in all evaluation metrics (Precision, Recall and F1 score). Specifically, the proposed method reaches an overall classification F1 score of 82.8%. Compared with the Resnet and plain VGG network, about 5.2% (0.828–0.776) and 4.3% (0.828–0.785) improvements are obtained by the proposed approach, respectively. And our method obtains 7.0% (0.828–0.758) gain compared to LSTM. The Acharya *et al.* and MS-CNN methods are surpassed by our approach in F1 score by 6.7% (0.828–0.761) and 3.1% (0.828–0.797). We also compare the proposed method with two latest reported algorithm on this dataset. Compared with TI-CNN (Yao et al.), our method increases by about 5.6% (0.828–0.772) in average F1 score. Compared with CL3 (Liu *et al.*[2]) combining the expert features and learning features, our DMSFNet has a 1.8% (0.828–0.810) improvement.

Furthermore, for each individual class of N, AF, I-AVB, LBBB, RBBB, PAC, PVC, STD and STE, the gains on F1 score are almost the highest than others. In particular, for single disease, accuracy increases 16.0%, 21.0% in detecting paroxysmal arrhythmias (PAC) compared with Resnet and LSTM. And our method is 18.0% and 16.0% higher than Acharya *et al.* and TI-CNN in detecting ST-segment elevated, respectively. And in [47], only using the expert features and deep features, the F1 score of CL3 (Liu et al.[1]) are 58.0% and 78.0% respectively. Compared with the best competitor CL3, our method achieves F1 score of 82.8% only using the learning features by the designed end-to-end neural network, which implies the effectiveness of our model.

2) Evaluation on CinC_2017 Dataset: Table III compares the related metrics of six reference models and DMSFNet in detecting Normal rhythm, AF rhythm, Other rhythm. In the experiment, we evaluate the classification performance on the training dataset using 5-fold cross validation. As shown in the table, we can find that the DMSFNet has the best performance than other methods in all average precision, recall rate and F1 score. Specifically, the proposed method reaches an overall classification F1 score of 84.1% in average. In comparison with Resnet, VGG and Acharya *et al.* that are based on convolutional network, the average F1 score increases by about 6.4% (0.841–0.777), 6.2% (0.841–0.779) and 7.1% (0.841–0.770). Compared to the recurrent neural network model, LSTM, our model has a 7.9% (0.841–0.762) improvement. And Compared

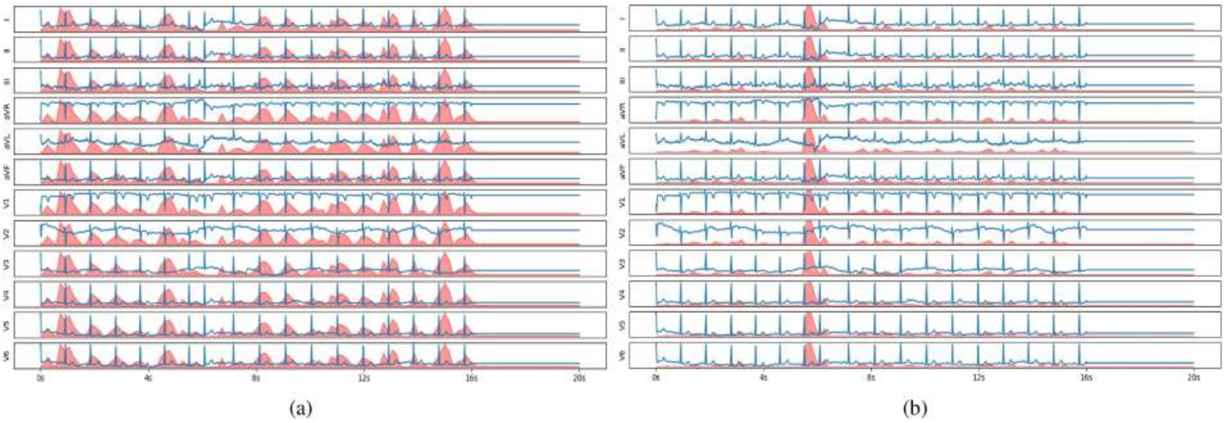


Fig. 4. Visualization of the responses assigned for different segments in ECG records, where (a) is the result of single-scale model and (b) indicates our model.

TABLE VI
EVALUATING MODEL WITH DIFFERENT SCALES

dataset	CPSC_2018		
metric	precision	recall	F1 score
scale-1	0.828	0.785	0.803
scale-2	0.820	0.788	0.801
DMSFNet	**0.838**	**0.822**	**0.828**

TABLE VII
EVALUATING MODEL WITH ATTENTION

dataset	CPSC_2018		
metric	precision	recall	F1 score
w/o attention	0.832	0.807	0.817
DMSFNet	**0.838**	**0.822**	**0.828**

with the MS-CNN, about 3.4% (0.841–0.807) improvement are obtained by our method.

In the classification of single disease, our method achieves a F1 value of 0.83 in the diagnosis of AF, which exceeds 11% of Resnet and Acharya *et al.*, 8% of VGG and 6% of MS-CNN. And for other rhythmic diseases, the proposed method has also improved to some extent. Compared with Datta *et al.* method, it uses more than 150 extracted features for ECG classification by a multi-layer cascaded binary classifier. In order to get more accurate results, more expert features need to be designed and extracted. DMSFNet can solve this problem effectively. Compared with the designed features of ECG, the above results indicate that our method can effectively extract abnormal features of ECG signals by using the multi-scale features in an end-to-end mechanism.

3) Ablation Studies: To analyze the relative contributions of different components of our model, we evaluate some variants of the proposed method with different settings.

Single-scale Framework vs. Multi-scale Framework: The classification performance of the model is effectively improved by complementary information between cross-scales. As shown in Table VI, we evaluate the performance using single-scale and multi-scale framework respectively. In this experiment, we use two single-scale CNNs with dilation rate 1 and 3 for classification task. The DMSFNet also adopts these two scales convolutional kernels for constructing the neural network. By comparison, we can find that the multi-scale model has better performance than single-scale model. The average recall rate increases by about 4.0% and the multi-scale model beats the single-scale model with about 3.0% rise in average F1 score. Also, Fig. 4 shows a visual comparison of features with DMSFNet and single-scale model. From the figure, we can find that the single-scale model does not capture the abnormal pattern of the signal and misidentifies this PVC signal as STD in the decision. Our proposed method gives high weight to the anomalous signal segment and accurately identifies it. In addition, a significant test is conducted with multiple sampling, the p value is less than 0.05. Overall, it suggests that the multi-scale features are consistently better than the results of single-scale model.

Fusion without Attention vs. Fusion with Attention: To evaluate the effectiveness of fusion method using spatial attention, we conduct additional experiments by comparing with the model without attention module. In this experiment, we delete attention module before global pooling and keep other network configurations unchanged in the training stage. Experimental results are shown in Table VII. As we can see, the average recall rate increases by 1.5%, the F1 sore increases by about 1.1% with the attention module. This shows that the attention module for fusion multi-scale features is helpful for mining the discriminative features and improving the classification performance of the model.

Single-loss Optimization vs. Multi-loss Optimization: The hyper parameter λ_1 and λ_2 dominates the participation level of scale-specific subnetworks in our model. Both of them are essential to our model. So we conduct experiments to investigate the sensitiveness of the two parameters. In the first experiment, to evaluate the effectiveness of the joint multi-losses optimization, we remove all branch losses and only keep the loss for the last fusion features in the training stage by fixing λ_1 and λ_2 to 0. Experimental results are shown in Table VIII. As we can see,

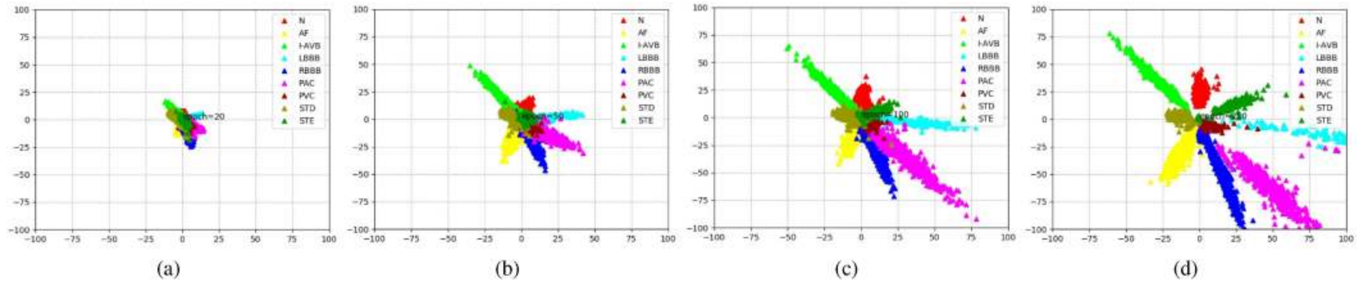


Fig. 5. Visualization of the learning features at 20, 50, 100 and 250 epochs on CPSC_2018 dataset.

TABLE VIII
Evaluating Model With Different Loss

dataset	CPSC_2018		
metric	precision	recall	F1 score
$\lambda_1 = 0.0,\ \lambda_2 = 0.0$	0.827	0.791	0.805
$\lambda_1 = 1.0,\ \lambda_2 = 0.0$	0.833	0.804	0.819
$\lambda_1 = 0.0,\ \lambda_2 = 1.0$	0.820	0.810	0.813
$\lambda_1 = 0.1,\ \lambda_2 = 0.1$	0.832	0.816	0.821
$\lambda_1 = 0.5,\ \lambda_2 = 0.5$	0.836	0.811	0.823
$\lambda_1 = 1.0,\ \lambda_2 = 1.0$	**0.838**	**0.822**	**0.828**

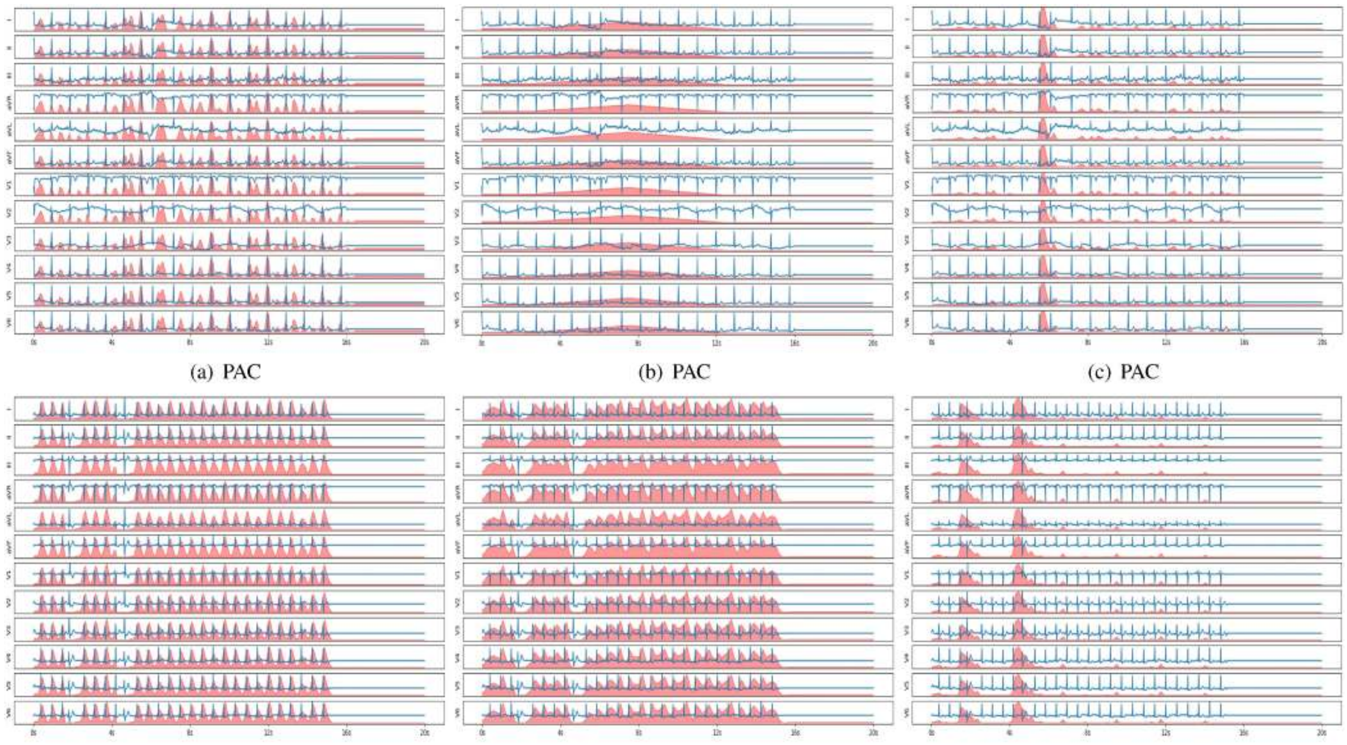
the F1 sore increases by about 2.3% (0.828–0.805) with the joint branch losses. Further, we add one scale-specific branch loss and observe the performance of classification. After adding a single scale branch loss, the F1 scores have 1.4% (0.819–0.805) and 0.8% (0.813–0.805) improvements. When all scale branches are considered for joint optimization, the accuracy gradually increases. We carry λ_1 and λ_2 from 0.1 to 1.0, it is very clear that simply using the scale-fusion branch loss (in this case λ_1 and λ_2 are 0) is not an optimal result. And the recognition accuracy of the deeply learned features is improved as scale-specific branches participation increases. This shows that joint multi-loss optimization can improve the classification performance of the model.

4) Effectiveness of Learning Features: The quantitative metrics show the effectiveness of our proposed multi-scale features fusion approach. Taking CPSC_2018 dataset as an example, to further intuitively evaluate the proposed method, Fig. 5 shows the visualization of the learning features on different training stages. To facilitate visualization, we reduce the dimensions of the network features from 256 to 2 before output. Then the results of these features for all the categories can be visualized into a two-dimensional plane. We intercept the results of 20, 50, 100 and 250 epochs. As shown in Fig. 5, the degree of feature discrimination becomes more and more obvious as the number of iterations increases on the whole. Specifically, in the first 20 epochs, the features of all categories overlap and are difficult to identify. Before 50 epochs, the distinction of learning features for all categories is not obvious. After 50 epochs, the discrimination is obviously enhanced, and the distance between classes is gradually enlarged. By the end of training, classes are basically separated from each other.

To illustrate the effectiveness of the proposed method, two samples with PAC and PVC are drawn according to the response degree of the category to the feature, as shown in Fig. 6. We use Resnet and MS-CNN as examples. From the figure, we can find that the response performances of arrhythmias with obvious occasional patterns get more attention. Specifically, our DMSFNet clearly assigns larger weights for abnormal segments and gives lower weight to normal signal segments, which achieves the right judgment. However, it is difficult for the MS-CNN (Fig. 6(b)(e)) and Resnet (Fig. 6(a)(d)) to catch the accurate abnormal pattern in the ECG segment, so they get the wrong recognition.

5) MS-CNN vs. DMSFNet: Both the DMSFNet and MS-CNN are based on multi-scale learning methods for ECG analysis. But there are still some differences between the two methods: (1) In DMSFNet, a backbone CNN sharing the same parameters is adopted for extracting shared features firstly. And unlike MS-CNN which uses general convolution kernels, the dilated convolution is used for decreasing the correlations among the different kernels and the number of network parameters. (2) DMSFNet uses spatial attention for further mining the discriminative features and improving the representations of the network. And using global pooling replaces full connection layer, which effectively extracts features from different channels and reduces feature dimension. (3) Joint optimization with multiple losses of different scales is adopted, which improves the discriminant ability of learning features.

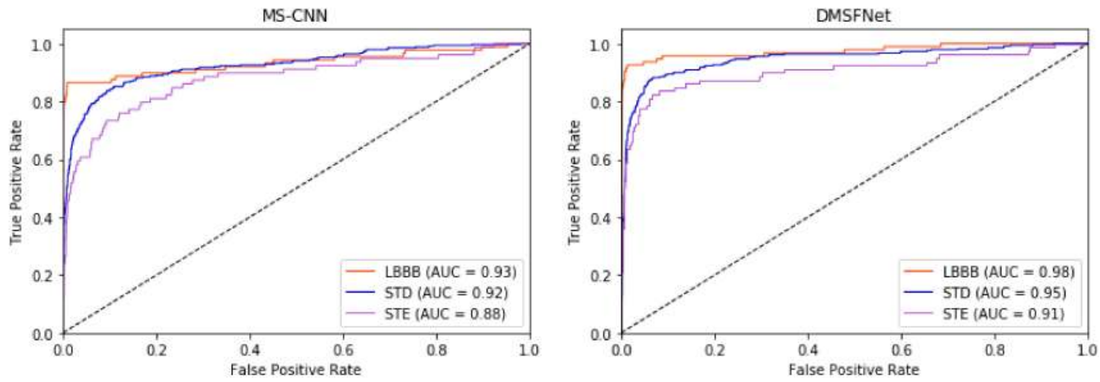
Take LBBB, STD and STE for example, Fig. 7 shows the receiver operating characteristic (ROC) curve and area under curve (AUC) value between these two method for these three disease. From the figure, the AUC values of DMSFNet are 0.98, 0.95 and 0.91 respectively, which are higher than MS-CNN in all kinds of listed diseases. Further, we compare the classification performance of the two methods. As shown in Table II, the DMSFNet outperforms the MS-CNN in almost all disease classification performances. Specially, F1 value increases by 12% in detecting STE and 5.0% for PVC. The average recall rate increases by about 4.0% and F1 score increases by about 3.0% compared with MS-CNN. It demonstrates that the DMSFNet has a higher classification performance by learning multi-scale features than MS-CNN. Compared with the MS-CNN, the proposed method is more competitive in mining the implicit correlated complementary advantages across scales and improving the processing ability of the model.



(a) PAC (b) PAC (c) PAC

(d) PVC (e) PVC (f) PVC

Fig. 6. Visualization of the responses assigned for different segments in ECG records on CPSC_2018 dataset, where the images from the first to third columns are the results using Resnet [39], MS-CNN [43] and the proposed DMFSNet respectively.



(a) (b)

Fig. 7. The ROC curve of MS-CNN and DMSFNet for LBBB, STD and STE.

V. Discussion

A. Diagnostic Effect

Different arrhythmias show various rhythmic characteristics. For example, AF patients are usually associated with characteristics of P-wave absence or irregular variability of R-R intervals in the ECG signal. Other abnormal rhythms also show different abnormal patterns in a single ECG signal. Traditional methods need to design and extract different features for distinct diseases, which largely depends on expert experience and extraction accuracy. In traditional methods of detecting PAC and PVC, it is necessary to detect QRS position and calculate R-R interval and QRS width, so that the accuracy of detecting premature beats depends heavily on the location of QRS and the accuracy of QRS starting and ending points. In this work, we focus on developing a novel deep learning method for automated multi-class arrhythmia detection. First, the proposed method has great potential to reduce the dependence of hand-crafted features by end-to-end neural network. From Fig. 6, we can see that the proposed model can identify premature beat without calculating parameters such as R-R interval, and locate the position of premature beat accurately and clearly. In addition, with the help of multi-scale feature fusion, the proposed method highlights the related irregular area compared with other approaches. The segment location of abnormal pattern is more accurate, which can help doctors to locate and diagnose abnormal patterns faster and better, and then improve the efficiency and accuracy of diagnosis. Results show that the multi-scale informations bring great benefit to identify the arrhythmias, the proposed DMSFNet effectively employs the underlying correlations among features of distinct scales and obtains more abundant feature embeddings for arrhythmias detection.

B. Model Applicability

The recent works on ECG processing with deep network improve the recognition accuracy and generalization capacity. Different network frameworks illustrate different performances. From the results, we can find that the models based convolutional network generally exceeds plain LSTM. This is mainly because if the input time step is very long, it is difficult for the network to capture long-term memory information. Compared to using LSTM alone, TI-CNN (Yao et al.) integrates LSTM cell layers after multi-layer convolution, which improves the above problem. But it also only utilizes the single scale information that limited the classification performance without complementary cross-scale information. Compared with above methods, our work effectively extracts the cross-scale features of segments by multiple convolution filters with different receptive fields and spatial attention mechanism. Meanwhile, joint optimization strategy optimizes multiple classification loss on same ECG label information concurrently, which further promotes the learning of different scale features. The above results suggest that the scheme design of multi-scale features fusion effectively improves the performance of multi-class arrhythmia detection.

VI. Conclusion

In this paper, we present a novel end-to-end deep learning method (DMSFNet) for ECG signal classification by utilizing the multi-scale ECG signal features. At same time, we integrate joint optimization with multiple losses of different scales into an unified convolutional neural network. Compared with the existing deep learning methods for ECG analysis using single scale, our proposed approach can effectively achieve multi-scale feature extraction and cross-scale information complementarity of ECG signals. We demonstrate outstanding performance for ECG classification on two public datasets comparing with some state-of-the-art methods. The experimental results convince the effectiveness of the proposed method. In the future, we will apply the DMSFNet to other physiological signal analysis and processing requirements.

Acknowledgment

We would like to appreciate Prof. Chengyu Liu from Southwestern University (China) for the experiment database he and his group provided for this research.

References

[1] A. Galassi, K. Reynolds, and J. He, "Metabolic syndrome and risk of cardiovascular disease: A meta-analysis," *Amer. J. Medicine*, vol. 119, no. 10, pp. 812–819, 2006.

[2] J. Schläpfer, and H. J. Wellens, "Computer-interpreted electrocardiograms: Benefits and limitations," *J. Amer. Coll. Cardiol.*, vol. 70, no. 9, pp. 1183–1192, 2017.

[3] 2015. [Online]. Available: https://en.wikipedia.org/wiki/Electrocardiography

[4] S. Osowski, L. T. Hoa, and T. Markiewic, "Support vector machine based expert system for reliable heartbeat recognition," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 582–589, Apr. 2004.

[5] P. D. Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, Jul. 2004.

[6] T. J. Jun, H. J. Park, and Y. H. Kim, "Premature ventricular contraction beat detection with deep neural networks," in *Proc. IEEE Int. Conf. Mach. Learn Appl.*, 2016, pp. 859–864.

[7] J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala, and R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1204–1211, May 1999.

[8] V. Monasterio, P. Laguna, and J. P. Martinez, "Multilead analysis of t-wave alternans in the ecg using principal component analysis," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 7, pp. 1880–1890, Jul. 2009.

[9] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016.

[10] M. M. Al. Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Inf. Sci.*, vol. 345, pp. 340–354, 2016.

[11] P. Li *et al.*, "High-performance personalized heartbeat classification model for long-term ECG signal," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 78–86, Jan. 2017.

[12] U. B. Baloglu, M. Talo, O. Yildirim, R. S. Tan, and U. R. Acharya, "Classification of myocardial infarction with multi-lead ECG signals and deep CNN," *Pattern Recognit. Lett.*, vol. 122, pp. 23–30, 2019.

[13] A. Y. Hannun *et al.*, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nat. Med.*, vol. 25, no. 1, pp. 65–69, 2019.

[14] B. Taji, A. D. C. Chan, and S. Shirmohammadi, "False alarm reduction in atrial fibrillation detection using deep belief networks," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 5, pp. 1124–1131, May 2018.

[15] F. Liu *et al.*, "An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection," *J. Med. Imag. Health Inform.*, vol. 8, no. 7, pp. 1368–1373, 2018.

[16] G. D. Clifford *et al.*, "AF classification from a short single lead ECG recording: The physionet/computing in cardiology challenge 2017," in *Proc. Comput. Cardiol. (CinC)*, 2017, pp. 1–4.

[17] N. Zeng, Z. Wang, and H. Zhang, "Inferring nonlinear lateral flow immunoassay state-space models via an unscented Kalman filter," *Sci. China-Inf. Sci.*, vol. 59, no. 11, 2016, Art. no. 112204.

[18] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "ECG analysis: A new approach in human identification," *IEEE Trans. Instrum. Meas.*, vol. 50, no. 3, pp. 808–812, Jun. 2001.

[19] M. R. Homaeinezhad, S. A. Atyabi, E. Tavakkoli, H. N. Toosiab, A. Ghaffariabc, and R. Ebrahimpour, "ECG arrhythmia recognition via a neuro-SVM–KNN hybrid classifier with virtual QRS image-based geometrical features," *Expert Syst. Appl.*, vol. 39, no. 2, pp. 2047–2058, 2012.

[20] R. Silipo, and C. Marchesi, "Artificial neural networks for automatic ECG analysis," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1417–1425, May 1998.

[21] R. V. Andreao, B. Dorizzi, and J. Boudy, "ECG signal analysis through hidden Markov models," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 8, pp. 1541–1549, Aug. 2006.

[22] S. Raj and K. C. Ray, "Sparse representation of ECG signals for automated recognition of cardiac arrhythmias," *Expert Syst. Appl.*, vol. 105, pp. 49–64, 2018.

[23] T. Ince, S. Kiranyaz, and M. Gabbouj, "A generic and robust system for automated patient-specific classification of ECG signals," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 5, pp. 1415–1426, May 2009.

[24] Q. Yang *et al.*, "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018.

[25] Y. Gurovich *et al.*, "Identifying facial phenotypes of genetic disorders using deep learning," *Nat. Med.*, vol. 25, no. 1, pp. 60–64, 2019.

[26] A. Rajkomar *et al.*, "Scalable and accurate deep learning with electronic health records," *NPJ Digit. Med.*, vol. 1, no. 1, pp. 1–10, 2018.

[27] P. Zanini, M. Congedo, C. Jutten, S. Said, and Y. Berthoumieu, "Transfer learning: A Riemannian geometry framework with applications to brain–Computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 5, pp. 1107–1116, May 2018.

[28] S. H. Gao, M. M. Cheng, K. Zhao, X. Zhang, M. Yang, and P. Torret, "Res2Net: A new multi-scale backbone architecture," 2019. [Online]. Available: https://arxiv.org/abs/1904.01169

[29] Q. Zhou *et al.*, "Multi-scale deep context convolutional neural networks for semantic segmentation," *World Wide Web*, vol. 22, no. 2, pp. 555–570, 2019.

[30] Q. Zhou, B. Zheng, W. Zhu, and L. J. Latecki, "Multi-scale context for scene labeling via flexible segmentation graph," *Pattern Recognit.*, vol. 59, pp. 312–324, 2016.

[31] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 510–519.

[32] T. Yang, L. Yu, Q. Jin, L. Wu, and B. He, "Localization of origins of premature ventricular contraction by means of convolutional neural network from 12-lead ECG," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 7, pp. 1662–1671, Jul. 2018.

[33] B. Pourbabaee, M. J. Roshtkhari, and K. Khorasani, "Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients," *IEEE Trans. Syst. Man Cybern.: Syst.*, vol. 48, no. 12, pp. 2095–2104, Dec. 2018.

[34] Y. Shahriari, R. Fidler, M. M. Pelter, Y. Bai, A. Villaroman, and X. Hu, "Electrocardiogram signal quality assessment based on structural image similarity metric," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 4, pp. 745–753, Apr. 2018.

[35] H. Kaji, H. Iizuka, and M. Sugiyama, "ECG-based concentration recognition with multi-task regression," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 1, pp. 101–110, Jan. 2019.

[36] A. Zarei, and B. M. Asl, "Automatic detection of obstructive sleep apnea using wavelet transform and entropy based features from single-lead ECG signal," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 3, pp. 1011–1021, May 2018.

[37] F. Yu, and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015. [Online]. Available: https://arxiv.org/abs/1511.07122

[38] N. Lessmann *et al.*, "Automatic calcium scoring in low-dose chest CT using deep neural networks with dilated convolutions," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 615–625, Feb. 2018.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014. [Online]. Available: https://arxiv.org/abs/1409.1556

[41] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[42] U. R Acharya, H. Fujita, O. S. Lih, Y. Hagiwaraa, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *Inf. Sci.*, vol. 405, pp. 81–90, 2017.

[43] X. Fan, Q. Yao, Y. Cai, F. Miao, F. Sun, and Y. Li, "Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ECG recordings," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 6, pp. 1744–1753, Nov. 2018.

[44] Q. Yao, X. Fan, Y. Cai, R. Wang, L. Yin, and Y. Li, "Time-incremental convolutional neural network for arrhythmia detection in varied-length electrocardiogram," in *Proc. IEEE Int. Conf. DASC/PiCom/DataCom/CyberSciTech*, 2018, pp. 754–761.

[45] S. Datta, C. Puri, A. Mukherjee, R. Banerjee, A. D. Choudhury, and R. Singh, "Identifying normal, AF and other abnormal ECG rhythms using a cascaded binary classifier," in *Proc. Comput. Cardiol. (CinC)*, 2017, pp. 1–4.

[46] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," 2017. [Online]. Available: https://arxiv.org/abs/1708.04896

[47] Z. Liu, X. Meng, J. Cui, Z. Huang, and J. Wu, "Automatic identification of abnormalities in 12-lead ECGs using expert features and convolutional neural networks," in *Proc. Int. Conf. Sens. Netw. Signal Process.*, 2018, pp. 163–167.