Digital Signal Processing 116 (2021) 103094

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

# Few-shot learning for cardiac arrhythmia detection based on electrocardiogram data from wearable devices

Tianyu Liu, Yukang Yang, Wenhui Fan *, Cheng Wu

*Department of Automation, Tsinghua University, Beijing, China*

ARTICLE INFO

*Article history:*
Available online 18 May 2021

*Keywords:*
Cardiac arrhythmia classification
Wearable devices
Few-shot learning
Electrocardiogram data

ABSTRACT

Wearable devices have dramatically developed over the past decade as their functions extended from the simple posture analysis to non-invasive condition monitoring for early warning and proactive healthcare, which are especially significant for the dangerous disease such as cardiac arrhythmia. However, it is difficult for the wearable devices to collect plentiful and high-quality training samples so as to meet the fundamental requirements for the learning-based methods. To address this challenge, we propose a meta-transfer based few-shot learning method to handle arrhythmia classification with the ECG signal from the wearable devices. First, the original ECG signals are converted into spectrograms applicable to the 2D-CNN models. Second, we propose the special large-training scheme to pre-train the feature extractor to emphasize the meaningful information for classification, and the feature output dimension is reshaped to reduce the influence of irrelevant and redundant information. Then, the meta-transfer scheme is developed to avoid the training from scratch, which is prone to overfitting without the adequate samples. Finally, we conduct the extensive experiments to assess the performance of our method. The experimental results illustrate that the proposed method outperforms in accuracy than other comparative methods when handling the various few-shot tasks under the same training samples.


## 1. Introduction

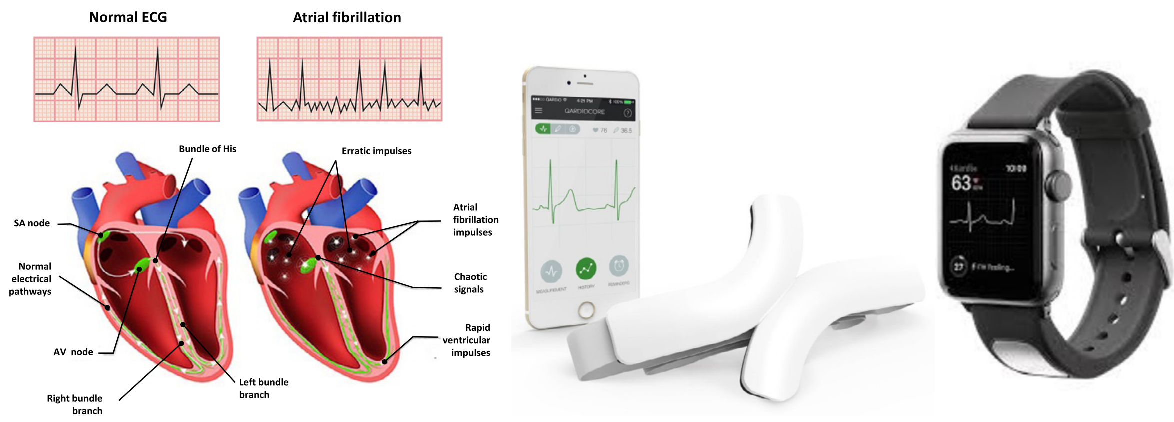
Cardiac arrhythmia, a typical and important group of the clinical cardiovascular domain, extremely requires early diagnosis and proactive treatment due to its severe mortality rates [1]. The arrhythmia diagnosis highly depends on electrocardiogram (ECG), which can record the signal data of heart pumping over time with specific devices. Wearable devices have attracted considerable attention since the advent of portable smart sensors [2], thanks to which physiological information can be obtained to tackle the limitations of proactive healthcare in a dynamic and non-invasive manner [3–5]. Manual arrhythmia detection based on ECG is a laborious and time-consuming task, while automatic method can improve the efficiency of cardiologists and even provide prioritizing diagnoses to alert potential patients to prevent irreversible disease and reactive treatment via wearable devices [6], as shown in Fig. 1. The learning-based methods have made tremendous progress in the various types of arrhythmia detection, and they play an irreplaceable role in computer-aided diagnosis of heart disease. The successful application of learning-based method in arrhythmia detection principally benefits from the construction and selection of suitable databases, which usually contain plentiful training samples with high quality collected by the professional medical devices in clinic.

However, the wearable devices are usually served by edge computing, whose limited computing capacity makes it difficult to deal with the large-scale dataset for accurate detection [7]. The collected ECG data from wearable devices inevitably contain meaningless information and noises due to different physical states [8]. Moreover, some certain arrhythmia types can only collect a few amount of similar and relevant samples for training models as a result of the individual variation, rare disease case and limitation of current medical knowledge. The suitable training dataset is the primary requirement for standard supervised learning. Therefore, automatic arrhythmia detection with wearable devices under limited-size training datasets remains a challenging task.

To tackle the aforementioned challenges, we propose a meta-transfer based few-shot learning method to achieve automatic arrhythmia detection with wearable devices. The original ECG signals are transformed into the time-frequency spectrograms as the input of 2D-CNN models, which can automatically extract the meaningful feature for target category recognition. This step is used to avoid the noise filtering and hand-crafted feature extraction [9]. Then, we pre-train the feature extractor based on auxiliary dataset before features dimension reshaping, and adopt meta-transfer learn-

* Corresponding author.
*E-mail address:* fanwenhui@tsinghua.edu.cn (W. Fan).

https://doi.org/10.1016/j.dsp.2021.103094




(a) Arrhythmia detection with ECG (b) Various types of wearable devices

**Fig. 1.** The illustration of the wearable devices for arrhythmia detection.[2]

ing strategy to recognize the unseen target samples. According to the experiments, the results show that the proposed method is able to generalize to new arrhythmia types, even that with the limited-size datesets. To the best of our knowledge, the previous studies have not attempted to deal with arrhythmia detection in a few-shot learning manner. The main contributions of this paper are summarized as follows:

- We propose a trainable few-shot learning method to deal with arrhythmia detection with limited-size ECG samples, which is more practical for the collected data of wearable devices.
- The base learner is pre-trained on an auxiliary dataset to improve the efficiency of feature extraction and alleviate the requirements on the training- samples scale, and feature dimension reshape can ignore the irrelevant features and improve the performance of classification.
- We propose the meta-transfer scheme to relieve the gap between the source and the target domain for unseen category recognition, and the comparison experiments are designed to demonstrate the advantage of our method.

The remainder parts are organized as follows: Section 2 briefly introduces the related work on automatic arrhythmia detection and few-shot learning. Section 3 describes the details of the proposed method, mainly containing large-scale training, feature reshaper and meta-transfer scheme. The experimental setting and results are presented in Section 4, and we conclude the paper and discuss the future research direction in Section 5.

## 2. Related work

Manual arrhythmia detection by clinicians is usually laborious, inefficient, and prone to clinical experience variability. Recently, a growing body of learning-based methods has made remarkable process in automatic arrhythmia detection, such as DNN [10,11], RCNN [12,13] and LSTM [14,15]. In the following, we roughly review the representative learning-based methods in two branches: one-dimensional and two-dimensional convolutional neural network (CNN) based methods. The ECG data contains rich information of heartbeat rate and rhythm [16], which can simply adopt the one-dimensional CNN methods to achieve arrhythmia detection as a multi-classification problem. For instance, a 9-layer CNN was developed to address the multi-classification problem of arrhythmia detection [17]. To further improve accuracy, batch normalization and dropout module were integrated into the basic CNN framework to relieve the overfitting during the training process [18]. The CNN optimization would be more difficult as the more layers were added to improve performance. Rajpurkar et al. [13] adopted the shortcut mechanism based on residual network to relieve the optimization problem of arrhythmia detection. The one-dimensional based methods can easily and efficiently deal with arrhythmia detection, since the raw ECG data can be directly utilized without any transformation. However, one-dimensional methods generally neglect the temporal properties of ECG data and cause performance reduction and undesired detection results [19]. The two-dimensional based methods are suitable to handle the 2D time-frequency spectrograms, which are robust to the noise of low-quality signal and can extract meaningful feature without manual setting and complicated framework designing [20]. Lu et al. [21] investigated a 2D-CNN network to extract the special feature as the input of random forest classifier and their experiment demonstrated the adaptability and robustness of proposed methods. Huang et al. [9] trained the 2D-CNN network on the large-size data for better performance than traditional methods. Meanwhile, the transfer learning was applied to identify four ECG patterns based on 2D time-frequency spectrograms [22]. The above 2D-CNN based studies successfully improved the classification performance, and these methods obviously outperform one-dimensional representations in the same tasks. However, the existing 2D-CNN based methods still depend on the computing capacity and the plentiful training samples, which are still hardly feasible for wearable devices.

Few-shot learning (FSL) is proposed to recognize new concepts with scarce training samples, and attempts to bridge the enormous gap between model generalization performance and training samples size [23,24]. In this paper, we have summarized mainstream FSL methods and divided them into two major categories: metric learning and meta learning based discriminative model [25]. Metric learning advocates to measure a similarity score between the extracted features and the target samples based on a learnable metric function. Koch et al. [26] proposed a unique Siamese Net to extract generic feature for similarity ranking, which utilized the neural network for FSL classification for the first time. To improve the metric performance, SPRN [27] was developed to replace the basic framework of Siamese Net as a skip ResNet [28] for efficient computation of pairwise similarity. The histogram loss was a simple measurement of pairwise similarity based on cosine distance, which was utilized to efficiently extract the potential feature space of the source domain [29]. Meta learning is a meaningful and

[2] The figures are respectively cited from https://southdenver.com/atrial-fibrillation-afib, uncrate.com/qardio-core-ecg-monitor and www.apple.com/watch.

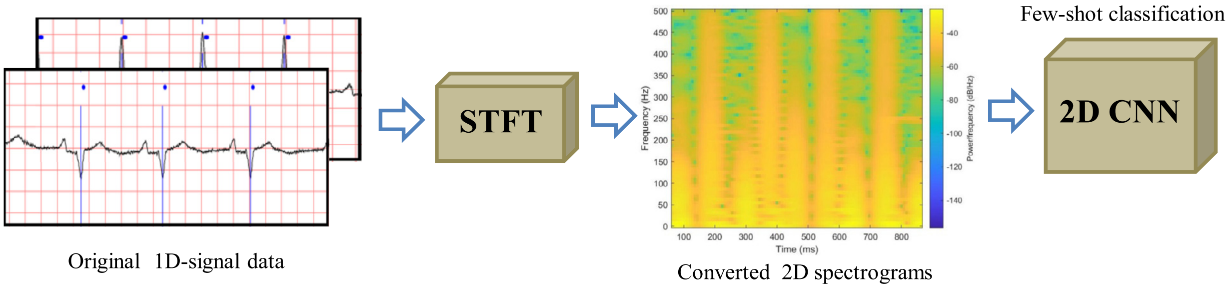



**Fig. 2.** The brief block representation of data preprocessing. The STFT module can convert the original data into the 2D spectrograms to describe the time-varying information of heart pumping, which meets the input requirement of 2D-CNN models for few-shot classification.

fundamental paradigm that can learn knowledge from across tasks and then be generalized to unseen tasks, and it is also called learning to learn [30]. Matching network defined the framework and the training standard for few-shot learning, which obtained remarkable success and was generally used as the baseline of other approaches [31]. Few-shot learning method has been applied to deal with the automatic recognition of medical data. Roy et al. [32] introduced a few-shot network to segment the volumetric medical images based on limited-size annotated samples. The one-shot learning was presented for the drug-resistant epilepsy classification and detection with only a few intracranial electroencephalography (iEEG) samples [33]. The CNNs was extended to train frequent conditions for the rare condition detection with a few fundus photographs [34]. Despite noticeable progress in various domains, the few-shot learning is still unable to propose a suitable and general paradigm in the medical field, especially for the arrhythmia detection with wearable devices.

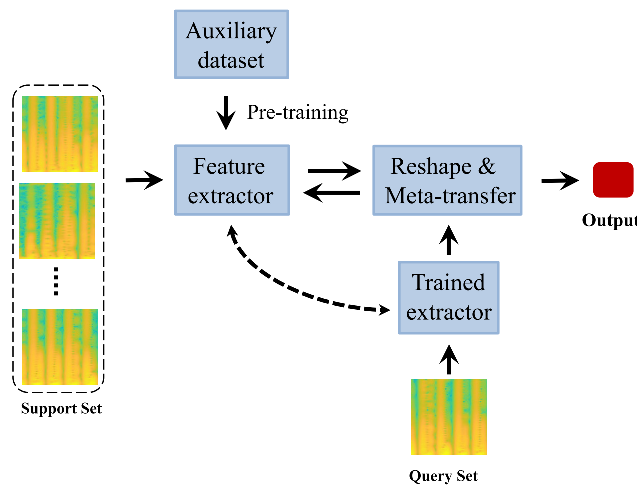



**Fig. 3.** The overall framework of proposed method. The auxiliary dataset is utilized to pre-train the CNN-based feature extractor for effective feature embedding, and reshape and meta-transfer module will emphasize the meaningful feature to adapt to the unseen task in the fine-tuning stage.

## 3. The proposed method

### 3.1. Problem formulation

In this study, we formulate arrhythmia detection based ECG data as a few-shot learning problem (FSL), which aims to effectively classify the unannotated (new) ECG data with minimal training samples. FSL datasets usually consist of three parts: a support set $\mathcal{D}_S$, a query set $\mathcal{D}_Q$ and an auxiliary set $\mathcal{D}_A$, among which the set $\mathcal{D}_S$ contains $K$ labeled samples for each of $C$ unique category. This task is usually called $C$-way $K$-shot classification problem. The set $\mathcal{D}_Q$ is composed of unlabeled samples, and FSL aims to predict the label of each sample in set $\mathcal{D}_Q$ based on the set $\mathcal{D}_S$, while the support set and query set share the same label space. Unlike the support set, the auxiliary set $\mathcal{D}_A$ is used to extract transferable knowledge for better performance of classifier [35], which contains abundant training samples and categories. It is noted that the label spaces of set $\mathcal{D}_A$ and set $\mathcal{D}_S$ are orthogonal, $\mathcal{D}_A \cap \mathcal{D}_S = \emptyset$. The goal of FSL is to correctly recognize the samples of query set $\mathcal{D}_Q$ only based on the support set $\mathcal{D}_S$ and the prior knowledge learned from the auxiliary set $\mathcal{D}_A$.

### 3.2. Data processing

The original ECG data is a typical nonstationary signal that records the variation of frequencies and amplitudes over time. The traditional 1D-CNN models are prone to ignore the temporal information of heart pumping, and they depend on complicated designing of feature extractor [20].

To deal with above problems, we convert the original one-dimensional ECG data into the time-frequency spectrograms as the input of 2D-CNN models, as shown in Fig. 2. The Fourier transform can represent the spectral changes with a time function in the fixed temporal window, and it has been widely used to process the stationary signal. The short-time Fourier transform (STFT) is an extended method of the discrete time fourier transform, which can efficiently describe the time-frequency representations. STFT uses an approximate method to represent the nonstationary signals, and 2D spectrograms are defined as:

$$S[X(n)] = X(m, \omega) = \sum_{n=-\infty}^{\infty} x(n)\omega(n-m)e^{-j\omega n}, \tag{1}$$

where $x(n)$ is ECG signal at time $n$, and $\omega(n)$ describes the Hanning window function [36]. In this paper, we use the spectrogram images of ECG signal to meet the input requirement of general 2D-CNN models, which can indirectly relieve the influence of noise and interruption in the original data to some degree, since the hidden features can be automatically extracted by 2D-CNN models.

### 3.3. Few-shot learning for arrhythmia detection

In this section, we propose a meta-transfer learning method to address the arrhythmia detection problem with the few labeled ECG spectrograms. The overview of proposed framework is shown in Fig. 3, which mainly consists of three phases: large-scale training (Section 3.3.2), feature dimensions reshaping (Section 3.3.3) and meta-transfer learning (Section 3.3.4).

#### 3.3.1. Meta learning for few-shot classification

Meta learning generally utilizes a learnable similarity metric to extract transferable feature for diverse domain adaptation, and measure-based method is a crucial branch of the meta learning

[25]. In few-shot task, the measure-based learning mainly consists of feature extractor and metric module. The extracted features are fed into the metric module $\mathcal{M}(\cdot,\cdot)$ to measure the relationship among the different samples for unseen class recognition. The final probability is measured by neighborhood components analysis [37] with special metric module, which is given by:

$$p(y=s|\mathbf{x}) = \frac{\exp(-\mathcal{M}(f(\mathbf{x},\theta),\mathbf{c}_s))}{\sum_{s'}\exp(-\mathcal{M}(f(\mathbf{x},\theta),\mathbf{c}_{s'}))}, \tag{2}$$

where $y$ is label of the samples $\mathbf{x}$, and $\mathbf{c}_s$ is the center of congener class in support set. The training objective is to minimize the averaged loss of all query-support pairs under the special metric module. In this paper, we adopt the Mahalanobis distance as the similar metric module $\mathcal{M}$ instead of Euclidean and cosine distance [31]. The proposed metric module is able to efficiently discriminate the different distributions with a unified scale, since it is able to emphasize the cluster variance to describe the classification boundaries. The inherent geometric feature of the samples in various domains can be preserved and represented, and the traditional Euclidean distance only considers that each class is distributed according to a unit normal. The Mahalanobis distance metric is more suitable to measure the pairwise similarity for the limited-size training samples, which has been well proved by [38]. Moreover, the learnable Mahalanobis distance has been formulated as a plug-and-play module to improve the performance of most meta-learning methods [39].

#### 3.3.2. Large-scale training

The mainstream few-shot learning methods usually focus on the valid metric setting and the development of optimization algorithms. However, it is difficult to effectively extract the underlying information for the recognition task due to their low-complexity architecture, and the shallow neural network tends to cause the model overfitting under the limited training samples when it is trained from scratch.

To address this problem, we leverage the modified EfficientNet to train a base learner for feature extraction with the abundant samples, which is inspired by the supervised fine-tuning on large-scale ImageNet for medical images classification [40]. EfficientNet [41] is a novel baseline network that adopts the compound coefficient to effectively scale each dimension of ConvNets for better performance. In particular, this method can dramatically reduce the amount of parameters in the network, which is more suitable for the transfer learning and few-shot learning. In this paper, we use random-weight method to train the basic network from scratch with the auxiliary dataset $D_A$, and aim to minimize the cross-entropy loss function $\mathcal{L}(\cdot)$, which is given by:

$$\mathcal{L}(\theta_l) = \frac{1}{|\mathcal{D}_A|}\sum_{(\mathbf{x}_a,y)\in\mathcal{D}_A} \ell(f(\mathbf{x}_a,\theta_l),y) + \mu(\theta_l), \tag{3}$$

where $f(\cdot)$ denotes the feature extractor, $\mathbf{x}_a$ is one sample of $\mathcal{D}_A$, $y$ represents its corresponding label and $\mu(\cdot)$ is the regularizer term. To further improve the performance of feature extractor, we consider the class imbalance problem of the ECG classification task, in which the healthy cases are usually much more than the ones with illness. Therefore, we propose a modified loss function to alleviate the impact of data distribution skews. The loss function can be rewritten as:

$$\mathcal{L}(\theta_l) = \frac{1}{|\mathcal{D}_A|}\sum_{(\mathbf{x}_a,y)\in\mathcal{D}_A} \big(\ell(f(\mathbf{x}_a,\theta_l),y_h) + \lambda\cdot\ell(f(\mathbf{x}_a,\theta_l),y_i\big) + \mu(\theta_l), \tag{4}$$

where $\lambda$ denotes the weight coefficient of binary classes, $y_h$ and $y_i$ denote the labels of healthy and ill cases, respectively. Then, we train the efficient feature extractor on large-scale dataset [42] to reduce the implicit gap between the source and the target domain, which helps the subsequent learning easily be trained.

#### 3.3.3. Feature dimensions reshaping

In a $C$-way $K$-shot classification task, the samples of support set $\mathcal{D}_S$ and query set $\mathcal{D}_Q$ are taken as the input of $f(\cdot,\theta_l)$ to extract feature for metric learning. The extracted feature inevitably contains redundancy and meaningless information since the differences between source domain and target domain.

To reduce the redundancy information, we add a simple CNN layer for the output reshaping so as to emphasize the meaningful feature. The output shape of feature extractor can be defined as $\{CK, m_1, d_1, d_1\}$, the $m_1$ and $d_1$ are the channel number and spatial size, respectively. The feature map can be simply reshaped as follows:

$$f(\cdot,\theta_l): \{CK, m_1, d_1, d_1\} \xrightarrow{reshape} \mathbf{o}: \{C, m_2, d_2, d_2\}, \tag{5}$$

where $m_2$ and $d_2$ are the reshaped channel number and spatial size [43]. This simple method aims to relieve the discrepancies of intra-class redundancy among the samples in one class, and the commonality among samples is extracted to form the decision boundaries. The reshaper module is considered more efficient than the averaging operation. Unlike CTM [43], this paper does not utilize *projector* module to mask out the certain feature for inter-class uniqueness, since the extractable and discriminable feature of ECG spectrogram is significantly less than the natural image as usual.

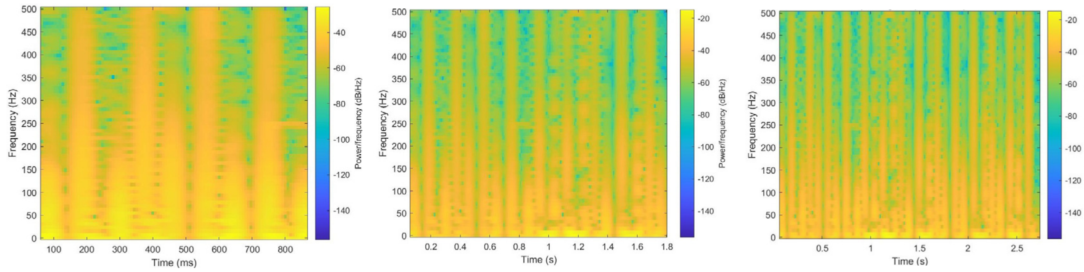
#### 3.3.4. Meta-transfer learning

The distribution of target domain is merely similar to the source domain, which is unable to directly adopt a metric function to discriminate the gap between target domain and source domain. However, we advocate that some useful discriminative information can be extracted and used to adapt to the target domain after fine-tuning the pre-trained network parameter of the source domain, and it is meaningful to learn a metric function for new class recognition.

To improve the recognition accuracy of target domain, we leverage the learned network parameter $\theta_l$ to feed into the feature extractor as the initial weights in the fine-tuning stage, and the output dimension of feature map has been reshaped to ignore irrelevant feature. The support set samples (few annotated data) are used to retrain the model by parameter optimization and loss minimization. Unlike non-parametric scheme of the ProtoNet [37], we fine-tune the final layer of new classifier by minimizing the cross-entropy loss function on the support set. The prototypic $\mathbf{c}_s$ are the means of each class of support set samples, which is given by:

$$\mathbf{c}_s = \frac{1}{|\mathcal{D}_s|}\sum_{(\mathbf{x}_s,y)\in\mathcal{D}_s} f(\mathbf{x}_s,\theta), \tag{6}$$

the parameter $\theta$ needs to be further fine-tuned based on $\theta_l$. Inspired by [44], the initial weight and the bias of final layer are respectively set as $\mathbf{W}_s = 2\cdot\mathbf{c}_s$ and $b_s = -\|\mathbf{c}_s\|^2$. The weights and bias are differentiable with parameter $\theta$ of the embedding function, which can be calculated for parameter $\theta$ updating on the support set by the universal optimization-based learning algorithms (i.e., Adam and SGD). It is noted that all parameters of model will be fine-tuned to improve the recognition performance in this stage.




(a) 5 records merged (b) 10 records merged (c) 15 records merged

**Fig. 4.** The illustration of various types of spectrogram images. The converted images can merge with unequal length records to describe the information of different sampling periods for arrhythmia detection, which are more suitable to the practical conditions of wearable devices. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

**Table 1**
The details of class information, and each class represents its corresponding arrhythmia type.

| Class | Annotations |
|---|---|
| **N** | Normal signals |
| **V** | Ventricular escape |
| **Q** | Paced / Unidentification |
| **S** | Supra-ventricular premature |
| **F** | Fusion of ventricular and normal |

## 4. Experiments

### 4.1. Datesets collection

In this paper, we collect the original ECG data from the renowned public datasets, MIT-BIH [45] and PTB diagnostic database [46], which are also used as the benchmark data source of the Kaggle tournament.[3] The MIT-BIH database consists of five class records, among which four are with arrhythmia and one is for the normal cases, and the sampling rate of ECG is fixed as 360 Hz. The PTB database is annotated seven types of arrhythmia, which samples the ECG signals at the frequency of 1000 Hz. To obtain the plentiful amounts of samples from two databases, we resample and preprocess the original signals with the sampling rate uniformly set as 125 Hz.[4] The resampled PTB dataset contains 14 K ECG records with two categories (ill and healthy cases), which is adopted to pre-train the feature extractor of the proposed method. Each record is used to form the training samples, and each sample records varying signals with the duration of 10 seconds. Then, we randomly down-sample the processed MIT-BIH database to form a few-shot learning dataset, which includes 2 K ECG records with five types of the beat rhythm, the details of labeled class information are listed in Table 1. The window size of STFT is set as 256 to achieve the spectrogram conversion [9]. To further consider the practical conditions of wearable devices, we linearly merge the multi-records (i.e., 5, 10 and 15 records) into one record as the input of the spectrogram conversion, since the wearable devices discontinuously sample the signal data several times a day for comprehensive analysis of the physical state. Then, the converted 2D spectrogram images are respectively fed into large-scale network and meta-transfer learning for pre-training and fine-tuning, the different types of spectrograms are shown in Fig. 4. The time-frequency responses of the ECG data are shown in Fig. 5. The implementation details of the experiment and its corresponding results will be elaborated in the subsequent sections.

[3] https://www.kaggle.com/shayanfazeli/heartbeat.

[4] This paper does not focus on the preprocessing operation, interested readers can refer to [47] for more details.

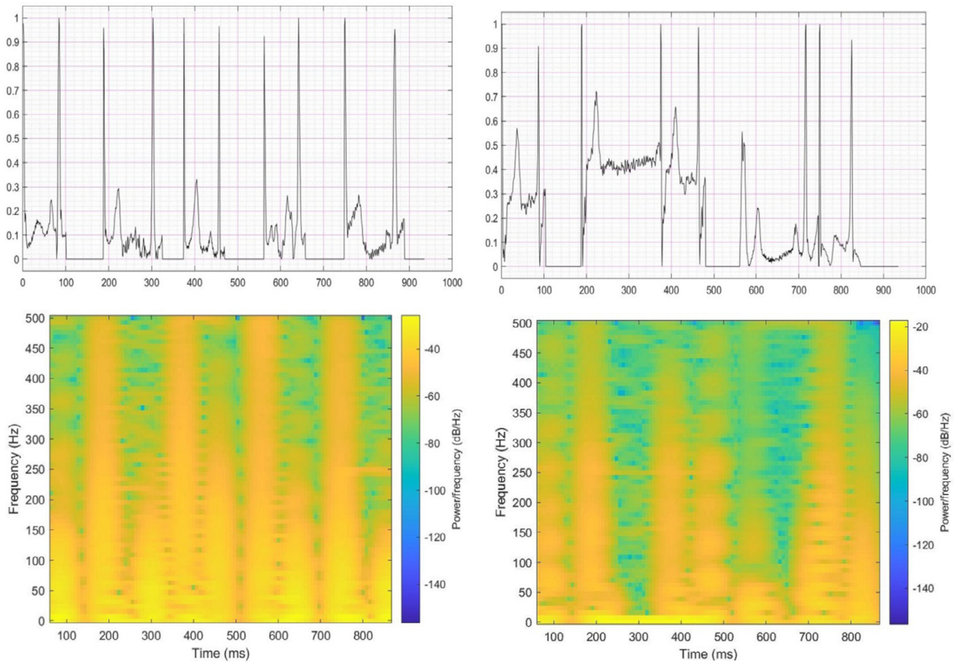
### 4.2. Implementation details

In the pre-training phase, we utilize the EfficientNet-B0 [41] as the basic network to train the feature extractor on the collected dataset. The amount of learnable parameter is 5.3M, which is much less than other famous methods, and it includes 16 MB Conv-modules, 2 Conv-layer with $3\times3$ kernels and a $2\times2$ pooling layer. The compound coefficients of basic network are set as $\alpha$(width)=1.2, $\beta$(depth)=1.1, $\gamma$(resolution)=1.15, respectively. The RMSProp optimizer is adopted to optimize the learnable parameter, the momentum and decay are 0.9, weight decay is 1e-5, initial learning rate is set as 0.3, and the decay rate set as 0.2 after every 100 epochs, which is similar to [41]. The collected dataset consists of 14 K spectrogram images with $128 \times 128$ pixel size, and we randomly select 10000 spectrogram images as the training dataset, and the remaining data is divided into two groups, among which 3000 for validation and 1000 for testing. In this paper, we set the weight coefficient $\lambda$ as 5 to balance the binary class, according to the approximate ratio of the major and minor-class samples. The stride of the CNN layer is set as 2 to reshape the feature map, and $m_2$ and $d_2$ of the reshaped feature map are 32 and 10 [43], respectively.

In the fine-tuning phase, the collected spectrogram images are uniformly resized to $84\times84$. We randomly select 2 categories from the dataset to form 2-way classification problem, and the similar operation is exerted to form 4-way one. Then, k annotated samples are picked to constitute the support set for k-shot learning, and corresponding query set is constituted by 5 non-annotated ones. To roundly demonstrate the model performance, we have trained the 1-shot, 5-shot, and 10-shot with 5 K, 3 K and 1 K episodes, respectively. The support set is used to optimize a classifier for 50 iterations, and its batch size is 10. In this phase, each experiment is repeated over 3 times to obtain the averaged results. Moreover, we adopt the Adam optimizer with the initial learning rate 0.001 for meta-transfer leaning. This paper employs the PyTorch to implement above experiments on two NVIDIA 2080Ti GPUs. Moreover, the wearable device can directly use the well-trained parameters of network to deal with the arrhythmia detection without training the parameters from scratch.

### 4.3. Experiment results and analysis

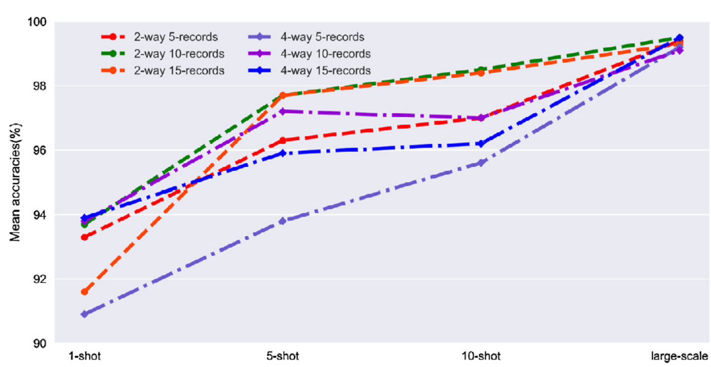
In this paper, we select the accuracy as our assessment metric, which evaluates the amount of correct classification in the test dataset. To verify the proposed the few-shot learning for the arrhythmia detection, we conduct the experiments to provide the evaluation and comparisons with two classification tasks. The 2-way k-shot task aims to detect the abnormal ECG signals, and the 4-way one is to classify the arrhythmia types. In 2-way task, we respectively set the k as 1, 5 and 10 to evaluate the perfor-



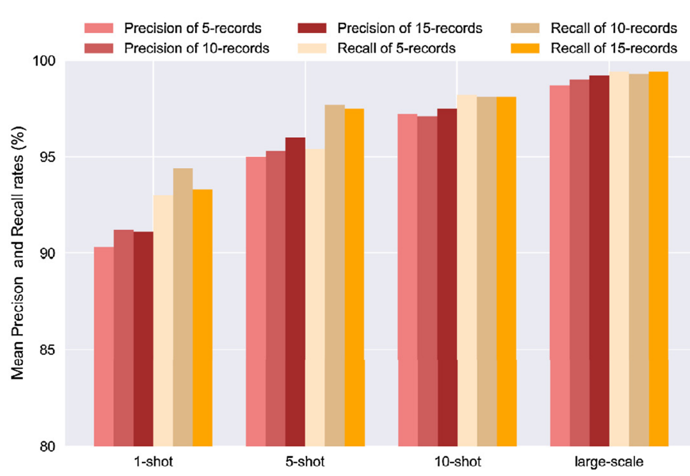

(a) The normal ECG data

(b) The arrhythmia-ECG data

**Fig. 5.** The illustration of time-frequency responses of various ECG data. There are some differences between the normal and the arrhythmia-ECG signal data, and our method is used to recognize such differences for arrhythmia detection.




(a) The mean accuracies of various tasks.

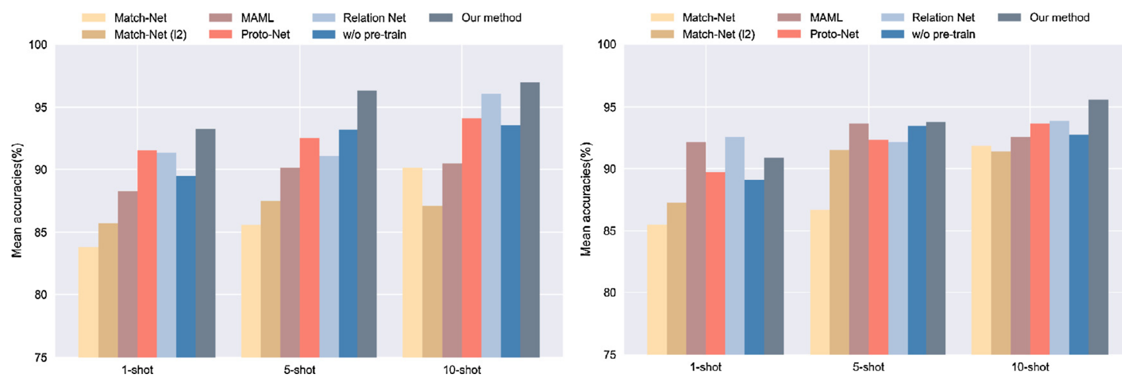



(b) The precision and recall results of 2-way tasks.

**Fig. 6.** (a) shows the mean accuracies of different few-shot learning tasks with different types of the training samples. (b) presents the mean precision and recall rates of 2-way k-shot tasks, which randomly selects the fixed class of query set for the fair evaluation of precision and recall rates.

**Table 2**
The mean accuracies (%) of various arrhythmia detection tasks on the various training samples, with 95% confidence intervals. It is noted that the mean accuracies of various few-shot learning tasks are not significantly less than those of the large-scale training, especially for the 15-records samples.

| Sample-size | 2-way task | | | 4-way task | | |
|---|---|---|---|---|---|---|
| | 5-records | 10-records | 15-records | 5-records | 10-records | 15-records |
| 1-shot | 93.3±0.3 | 93.7±2.6 | 91.6±2.4 | 90.9±1.8 | 93.8±1.2 | 93.9±1.8 |
| 5-shot | 96.3±0.5 | 97.7±1.0 | 97.7±0.2 | 93.8±0.6 | 97.2±0.9 | 95.9±1.1 |
| 10-shot | 97.0±0.4 | 98.5±0.2 | 98.4±0.4 | 95.6±0.2 | 97.0±0.7 | 96.2±0.7 |
| large-scale | 99.4±0.2 | 99.5±0.1 | 99.3±0.1 | 99.2±0.1 | 99.1±0.1 | 99.5±0.2 |




(a) 2-way learning tasks (b) 4-way learning tasks

**Fig. 7.** The experiment results of compared methods. All experiments adopt the same setting for a fair comparison, and our method outperforms in accuracy than others.

**Table 3**
Comparative results of various famous methods for C-way K-shot classifications on the merged 5-record samples. The performance is evaluated using mean accuracies (%) with 95% confidence intervals. The results show that the proposed method obtains the best accuracies in various classification tasks.

| Method | | 2-way task | | | 4-way task | | |
|---|---|---|---|---|---|---|---|
| | backbone | 1-shot | 5-shot | 10-shot | 1-shot | 5-shot | 10-shot |
| PrototyNet [37] | Conv-4 | 91.5±0.4 | 94.4±0.8 | 96.2±0.4 | 89.7±0.5 | 92.3±0.2 | 93.7±0.4 |
| Relation-Net [48] | Conv-4 | 91.3±0.2 | 91.1±1.3 | 96.1±1.9 | 92.5±1.3 | 92.1±1.2 | 93.9±1.8 |
| MatchingNet [31] | Conv-4 | 85.7±2.3 | 87.5±3.3 | 87.1±4.1 | 85.5±1.9 | 86.7±2.1 | 91.8±2.7 |
| MatchingNet($\ell$-2) | Conv-4 | 83.8±5.1 | 85.6±3.3 | 90.0±3.5 | 87.3±2.1 | 91.5±1.7 | 91.4±2.0 |
| Meta-learning [30] | Conv-4 | 88.3±1.6 | 90.1±2.3 | 90.5±1.3 | 92.1±1.5 | 93.7±1.2 | 92.5±1.0 |
| Without pre-train | EffiNet-b0 | 90.9±2.7 | 91.8±3.2 | 94.7±1.8 | 89.1±0.7 | 93.5±0.7 | 92.7±0.6 |
| Pre-training ($\lambda = 3$) | EffiNet-b0 | 90.4±1.0 | 90.1±1.9 | 94.1±1.7 | 90.0±2.1 | 93.0±1.3 | 91.9±1.6 |
| **Our method** | EffiNet-b0 | **93.3±0.3** | **96.3±0.8** | **97.0±0.4** | **90.9±1.8** | **93.8±0.6** | **95.6±0.3** |

mance of the proposed method, which is similar to the 4-way ones. Moreover, the different types of training samples are also used to further assess the performance. The experiment results are listed in Table 2 and Fig. 6. They show the performances of the different training-sample sizes with their corresponding classification task, which demonstrate the proposed method can handle the arrhythmia detection problem with the few training samples. The pre-trained feature extractor can effectively extract the useful information for alleviating the requirement of plentiful training samples. According to the results of various spectrogram types, more recorders can be merged into one spectrogram to further improve the model performance if a better accuracy is required for the clinical application. The mean accuracies of large-scale learning are only slightly better than the few-shot learning with the numerical experiment results, it means that the proposed method has great potential to handle the arrhythmia detection based on wearable devices without plentiful samples. Moreover, we present the metrics of *Precision* and *Recall* to comprehensively evaluate the performance of 2-way k-shot tasks, which can avoid the unfair evaluation due to the long-tailed distributions of query dataset, as shown in Fig. 6(b).

To further illustrate the outperformance of the proposed method, four widely applied few-shot learning networks with the same dataset and task are also tested as baseline methods. In this phase, we select the merged 5-record as the basic samples to perform the numerical experiments since the amount of 5-record is larger than other types, and it can demonstrate the performance of model generality for a fair comparison.

The comparison methods are Matching Net [31], Prototypical Net [37], Relation Net [48] and MAML [30]. Without loss of generality, we repeat all the experiments 3 times to obtain the averaged results, and their experiment results are shown in Table 3 and Fig. 7. It can be shown that the proposed model suppresses the compared methods under the same training dataset and task. The probable reason is that the Matching and Prototypical Net select the unsuitable metric function to handle the arrhythmia detection with the spectrogram samples, and their shallow neural network may focus on the meaningless and irrelevant feature to impact the classification performance without the feature dimensions reshape. The meta-transfer learning scheme can avoid optimizing a large amount of parameters with few training samples, which is inevitable to cause overfitting when training from scratch. Besides, the suitable backbone can improve the performance of the feature extraction for a better accuracy of the arrhythmia detection. The performance of prediction has no obvious variance when the hyperparameter $\lambda$ is slightly adjusted. The above numerical experiments illustrate that our method really outperforms the baseline methods in dealing with the arrhythmia detection problems under a few training samples.

## 5. Conclusions

In this paper, we propose a meta-transfer based few-shot learning method to automatically detect the arrhythmia type with few samples, which can be handled by the wearable devices. The proposed method can help alleviate the requirements of high-quality and abundant training samples for the standard learning-based methods. Furthermore, we introduce a modified pre-training method to improve the performance of feature extractor. The meta-transfer strategy can use the learned knowledge to avoid the overfitting to some degree, which is hard to deal with the traditional few-shot learning methods. The comprehensive experiments demonstrate that our approach has obtained the outstanding result compared with other common methods. In the future, we will further improve the performance of the feature extractor and generalize to other fields of medical applications by wearable devices, such as early diagnosis of cerebral stroke and long-term management of chronic diseases.

## Declaration of competing interest

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

## References

[1] P. Li, Y. Wang, J. He, L. Wang, Y. Tian, T. Zhou, T. Li, J. Li, High-performance personalized heartbeat classification model for long-term ecg signal, IEEE Trans. Biomed. Eng. 64 (1) (2017) 78–86, https://doi.org/10.1109/TBME.2016.2539421.
[2] X. Chen, Y. Wang, J. He, S. Pan, Y. Li, P. Zhang, Cap: context-aware app usage prediction with heterogeneous graph embedding, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 3 (1) (2019) 1–25.
[3] T. Strain, K. Wijndaele, P.C. Dempsey, S.J. Sharp, M. Pearce, J. Jeon, T. Lindsay, N. Wareham, S. Brage, Wearable-device-measured physical activity and future health risk, Nat. Med. (2020) 1–7.
[4] A. Pantelopoulos, N.G. Bourbakis, A survey on wearable sensor-based systems for health monitoring and prognosis, IEEE Trans. Syst. Man Cybern., Part C, Appl. Rev. 40 (1) (2009) 1–12.
[5] X. Chen, Z. Zhu, M. Chen, Y. Li, Large-scale mobile fitness app usage analysis for smart health, IEEE Commun. Mag. 56 (4) (2018) 46–52.
[6] M. Bariya, H.Y.Y. Nyein, A. Javey, Wearable sweat sensors, Nat. Electron. 1 (3) (2018) 160–171.
[7] M.Z. Uddin, A wearable sensor-based activity prediction system to facilitate edge computing in smart healthcare system, J. Parallel Distrib. Comput. 123 (2019) 46–53.
[8] M. Choi, G. Koo, M. Seo, S.W. Kim, Wearable device-based system to monitor a driver's stress, fatigue, and drowsiness, IEEE Trans. Instrum. Meas. 67 (3) (2017) 634–645.
[9] J. Huang, B. Chen, B. Yao, W. He, Ecg arrhythmia classification using stft-based spectrogram and convolutional neural network, IEEE Access 7 (2019) 92871–92880.
[10] K. Luo, J. Li, Z. Wang, A. Cuschieri, Patient-specific deep architectural model for ecg classification, J. Healthc. Eng. (2017).
[11] Y. Xia, H. Zhang, L. Xu, Z. Gao, H. Zhang, H. Liu, S. Li, An automatic cardiac arrhythmia classification system with wearable electrocardiogram, IEEE Access 6 (2018) 16529–16538.
[12] U.R. Acharya, H. Fujita, O.S. Lih, Y. Hagiwara, J.H. Tan, M. Adam, Automated detection of arrhythmias using different intervals of tachycardia ecg segments with convolutional neural network, Inf. Sci. 405 (2017) 81–90.
[13] 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.
[14] B. Hou, J. Yang, P. Wang, R. Yan, Lstm-based auto-encoder model for ecg arrhythmias classification, IEEE Trans. Instrum. Meas. 69 (4) (2019) 1232–1240.
[15] R.S. Andersen, A. Peimankar, S. Puthusserypady, A deep learning approach for real-time detection of atrial fibrillation, Expert Syst. Appl. 115 (2019) 465–473.
[16] A. Gacek, W. Pedrycz, ECG Signal Processing, Classification and Interpretation: a Comprehensive Framework of Computational Intelligence, Springer Science & Business Media, 2011.
[17] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, M. Adam, A. Gertych, R. San Tan, A deep convolutional neural network model to classify heartbeats, Comput. Biol. Med. 89 (2017) 389–396.
[18] M.M. Al Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, R.R. Yager, Deep learning approach for active classification of electrocardiogram signals, Inf. Sci. 345 (2016) 340–354.
[19] Q. Yao, R. Wang, X. Fan, J. Liu, Y. Li, Multi-class arrhythmia detection from 12-lead varied-length ecg using attention-based time-incremental convolutional neural network, Inf. Fusion 53 (2020) 174–182.
[20] F. Murat, O. Yildirim, M. Talo, U.B. Baloglu, Y. Demir, U.R. Acharya, Application of deep learning techniques for heartbeats detection using ecg signals-analysis and review, Comput. Biol. Med. (2020) 103726.
[21] W. Lu, H. Hou, J. Chu, Feature fusion for imbalanced ecg data analysis, Biomed. Signal Process. Control 41 (2018) 152–160.
[22] M. Salem, S. Taheri, J.-S. Yuan, Ecg arrhythmia classification using transfer learning from 2-dimensional deep cnn features, in: 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS), IEEE, 2018, pp. 1–4.
[23] E. Triantafillou, R. Zemel, R. Urtasun, Few-shot learning through an information retrieval lens, in: Advances in Neural Information Processing Systems, 2017, pp. 2255–2265.
[24] B. Oreshkin, P.R. López, A. Lacoste, Tadam: task dependent adaptive metric for improved few-shot learning, in: Advances in Neural Information Processing Systems, 2018, pp. 721–731.
[25] J. Lu, P. Gong, J. Ye, C. Zhang, Learning from very few samples: a survey, arXiv preprint, arXiv:2009.02653.
[26] G. Koch, R. Zemel, R. Salakhutdinov, Siamese neural networks for one-shot image recognition, in: ICML Deep Learning Workshop, vol. 2, Lille, 2015.
[27] A. Mehrotra, A. Dukkipati, Generative adversarial residual pairwise networks for one shot learning, arXiv preprint, arXiv:1703.08033.
[28] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[29] T. Scott, K. Ridgeway, M.C. Mozer, Adapted deep embeddings: a synthesis of methods for k-shot inductive transfer learning, in: Advances in Neural Information Processing Systems, 2018, pp. 76–85.
[30] M. Andrychowicz, M. Denil, S. Gomez, M.W. Hoffman, D. Pfau, T. Schaul, B. Shillingford, N. De Freitas, Learning to learn by gradient descent by gradient descent, Adv. Neural Inf. Process. Syst. 29 (2016) 3981–3989.
[31] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, D. Wierstra, Matching networks for one shot learning, in: Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016, pp. 3637–3645.
[32] A.G. Roy, S. Siddiqui, S. Pölsterl, N. Navab, C. Wachinger, 'Squeeze & excite' guided few-shot segmentation of volumetric images, Med. Image Anal. 59 (2020) 101587.
[33] A. Burrello, K. Schindler, L. Benini, A. Rahimi, One-shot learning for ieeg seizure detection using end-to-end binary operations: local binary patterns with hyperdimensional computing, in: 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS), 2018, pp. 1–4.
[34] G. Quellec, M. Lamard, P.-H. Conze, P. Massin, B. Cochener, Automatic detection of rare pathologies in fundus photographs using few-shot learning, Med. Image Anal. 61 (2020) 101660.
[35] X. Chen, S. Xu, X. Liu, X. Xu, H.Y. Noh, L. Zhang, P. Zhang, Adaptive hybrid model-enabled sensing system (hmss) for mobile fine-grained air pollution estimation, IEEE Trans. Mob. Comput. (October 2020) (Early Access).
[36] D. Griffin, J. Lim, Signal estimation from modified short-time Fourier transform, IEEE Trans. Acoust. Speech Signal Process. 32 (2) (1984) 236–243.
[37] J. Snell, K. Swersky, R. Zemel, Prototypical networks for few-shot learning, in: Advances in Neural Information Processing Systems, 2017, pp. 4077–4087.
[38] D. Das, C.G. Lee, A two-stage approach to few-shot learning for image recognition, IEEE Trans. Image Process. 29 (2019) 3336–3350.
[39] P. Bateni, R. Goyal, V. Masrani, F. Wood, L. Sigal, Improved few-shot visual classification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 14493–14502.
[40] H.-C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R.M. Summers, Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imaging 35 (5) (2016) 1285–1298.
[41] M. Tan, Q.V. Le, Efficientnet: rethinking model scaling for convolutional neural networks, arXiv preprint, arXiv:1905.11946.
[42] X. Chen, X. Xu, X. Liu, S. Pan, J. He, H.Y. Noh, L. Zhang, P. Zhang, Pga: physics guided and adaptive approach for mobile fine-grained air pollution estimation, in: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, 2018, pp. 1321–1330.
[43] H. Li, D. Eigen, S. Dodge, M. Zeiler, X. Wang, Finding task-relevant features for few-shot learning by category traversal, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 1–10.
[44] E. Triantafillou, T. Zhu, V. Dumoulin, P. Lamblin, U. Evci, K. Xu, R. Goroshin, C. Gelada, K. Swersky, P.-A. Manzagol, et al., Meta-dataset: a dataset of datasets for learning to learn from few examples, arXiv preprint, arXiv:1903.03096.

[45] G.B. Moody, R.G. Mark, The impact of the mit-bih arrhythmia database, IEEE Eng. Med. Biol. Mag. 20 (3) (2001) 45–50.
[46] R. Bousseljot, D. Kreiseler, A. Schnabel, Nutzung der ekg-signaldatenbank cardiodat der ptb über das internet, Biomed. Tech. 40 (s1) (1995) 317–318.
[47] M. Kachuee, S. Fazeli, M. Sarrafzadeh, Ecg heartbeat classification: a deep transferable representation, arXiv preprint, arXiv:1805.00794, 2018.
[48] F. Sung, Y. Yang, L. Zhang, T. Xiang, P.H. Torr, T.M. Hospedales, Learning to compare: relation network for few-shot learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1199–1208.

**Tinayu Liu** received the M.S degree in mechanics institute from Jilin University, Changchun, China, in 2016. He is currently pursuing the Ph.D. degree with the Department of Automation, Tsinghua University, Beijing, China. His current research interests include machine learning, deep learning and computer-aided diagnosis based on artificial intelligence methods.

**Yukang Yang** is currently a master student of Control Science and Engineering in the Department of Automation, Tsinghua University. Before that, he received his B.S. degree in Automation from ShenYuan Honors College, Beihang University. His main research interests focus on deep learning, computer vision, especially for medical image analysis.

**Wenhui Fan** received the PhD degree in mechanical engineering from Zhejiang University, Hangzhou, China in 1998. He obtained postdoctoral certificate from Tsinghua University, Beijing, China in 2000. He is a Vice President of China Simulation Federation. He is currently a professor in Tsinghua University, Beijing, China. His current research interests include machine learning, intelligent healthcare, multi-agent modeling and simulation, large scale agent modeling and simulation.

**Cheng Wu** received the B.S. and M.S. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1962 and 1966, respectively. Since 1967, he has been with Tsinghua University, where he is currently a Professor with the Department of Automation. He has also been an Academician of the Chinese Academy of Engineering since 1995. His current research interests include computer integrated manufacturing systems, manufacturing systems scheduling, optimization of supply chains, and system identification. Prof. Wu was a recipient of the National Science and Technology Progress Award of China in 1995, 1999, and 2006, and the First Scientific Development Award from the State Educational Committee of China in 1994.