2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A One-Dimensional Siamese Few-Shot Learning Approach for ECG Classification under Limited Data

Zongjin Li, Huan Wang, Xinwen Liu

Abstract—**Electrocardiogram (ECG) is mainly used by medical domain to diagnose arrhythmia. With the development of deep learning algorithms in the ECG classification field, related algorithms have achieved very high accuracy. However, the training of deep learning algorithms always requires large amounts of samples, while the labeled samples are often lacked in the field of medical signals. Therefore, the performance of deep learning algorithms will be greatly restricted. To overcome the sample scarcity problem, we propose a few-shot ECG classification approach based on the Siamese network. This network architecture first uses two one-dimensional convolutional neural network (CNN) that share weights to extract feature vectors of the paired input signals. Then, L1-distance between the two feature vectors is calculated and inputted into the fully connected layer with an activation function sigmoid to determine whether the input pairs belong to same category. We validated our method on the MIT-BIH arrhythmia database. By experiments, our method performs better than existing networks under the circumstance of extremely few amounts of data.**

Clinical Relevance—**The proposed algorithm can be used to classify arrhythmia types with a small amount of data.**

I. Introduction

Arrhythmia is a very common cardiovascular disease. Though most of arrhythmias are innocuous to human body, some kinds can lead to further impact such as rapid atrial fibrillation and persistent ventricular [1] tachycardia. These can threaten people's life health if not diagnosed timely. The most commonly used clinical diagnosis method for arrhythmia is electrocardiogram (ECG), because it is non-invasive for human body and cheap in price.[2] ECG can record electric potential change of heart to show its health condition.

It is both labor-intensive and time-consuming for experts to diagnose ECG manually, since inter-patient variations are significantly varying with different temporal and physical conditions [3]. Therefore, efficient automatic arrhythmia diagnosis is required. In recent few years, deep learning meet a rapid development in ECG diagnosis. Compared with traditional method, it has great advantage in extracting signal features automatically without human design.

This work was supported by Sichuan Science and Technology Program (2020108) and Glasgow College, University of Electronic Science and Technology of China. The experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board. (Corresponding author: Huan Wang).

Zongjin Li and Xinwen Liu are with the Glasgow College, University of Electronic Science and Technology of China, Chengdu, China.

Huan Wang is with the Department of Industrial Engineering, Tsinghua University and the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, China (email: wh.huanwang@gmail.com).

Among deep learning algorithm, CNN is widely applied in ECG diagnosis and achieved good performance [4]. Yildirim et al. [5] used a 16-layer CNN model and got 95.2% accuracy on MIT-BIH arrhythmia dataset. A Deform-CNN architecture for ECG diagnosis was supposed by Lang et al.[6], the overall diagnostic accuracy rate of which in CPSC-2018 database can reach 86.3%. Gao et al. [7, 8] constructed a network that focus sematic information by combining CNN with temporal attention module, and get 82.32% accuracy in The PhysioNet Computing in Cardiology Challenge 2017. Gao et al.[9] also proposed a new exponential nonlinear loss (EN-Loss) base on CNN to solve the atrial fibrillation classification problem. Considering that the convolutional neural network only pays attention to the spatial characteristics of the signal and ignores the sequential features. A composite neural network structure composed of CNN and long-term short-term memory network (LSTM) [10] is developed. Shu et al. [11] combined LSTM and CNN to consider both the sequential and spatial features of the signal to improve the classification performance of ECG signals. By constructing a hybrid network that can handle variable-length ECG signals, an accuracy of 98.10% is obtained. Qiao et al. [12] get an accuracy of 99.32% on the MIT-BIH Arrhythmia database by using a model of CNN and Bidirectional long-short time memory network (BLSTM).

Although recent neural network models have achieved excellent results in ECG diagnosis, they are quite dependent on large amount of training data. However, large amounts of labeled data are often lacked. Therefore, we adopt the view that a diagnostic classification method for arrhythmia based on extremely few data should be established.

Along development of deep learning, aiming at solving the problem of data scarcity, deep neural networks based on small amount of training sets have made important progress [13]. In the field of computer vision, Qiao et al. [14] proposed an image recognition method to complete the classification task by predicting activation parameters in the matrix space. Zhang et al. [15] proposed Meta-GAN that can be semi-supervised in consideration of sample level and task level. Although few shot learning has extensive research computer vision field, the exploration in the field of ECG diagnosis is still not enough. Thus, inspirated by Siamese neural network proposed by Koch et al [16]. We propose a one-dimensional Siamese few-shot learning approach for ECG diagnosis, which can realize the few-shot task of ECG classification through a feature comparison method. The performance of our method is verified on the MIT-BIH Arrhythmia Database and accuracies of 82.36%, 84.98%, 89.16%, 91.73% and 92.42% are gained respectively for 1 shot, 5 shot, 10 shot, 30 shot and 50 shot cases. The performance is significantly improved compared to existing deep learning methods.

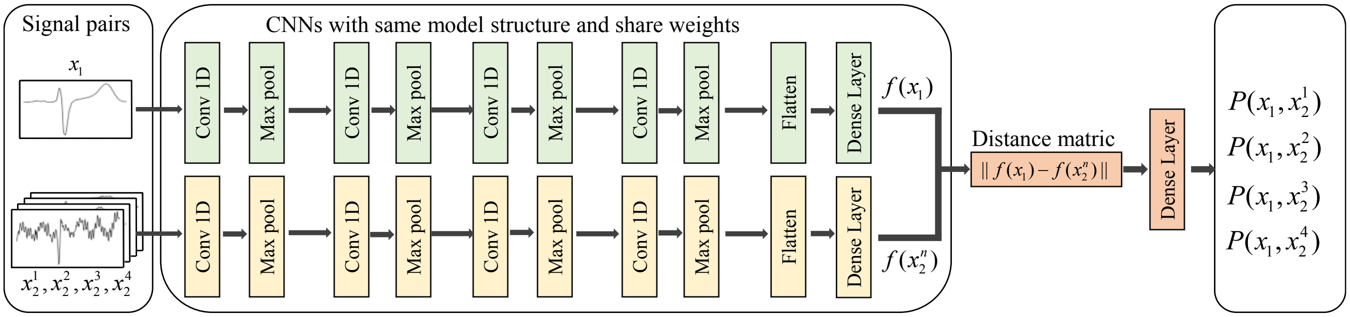

Fig.1. Few-shot-learning model for ECG classification. x_1 and x_2^n denote the paired input signals, $f(x_1)$ and $f(x_2^n)$ denote the feature vectors, $P(x_1, x_2^n)$ denote the probabilities that the paired input signals are of the same category, n=1,2,3,4.

II. METHODOLOGY

In this section, the principle and implementation steps of our model used for ECG signal diagnosis will be described in details.

A. *Siamese Few-shot Learning Algorithm*

In the application scenario of few shot learning, we have a small labeled sample set *D* with a size of 4*N*, that is, there are *N* training samples for each 4 classification categories, as shown in Eq. (1). Each training sample *X* has an independent label *Y*.

$$D = \{(X_1^C, Y_1^C), (X_2^C, Y_2^C), ..., (X_N^C, Y_N^C)\}, c = 0,1,2,3 \quad (1)$$

In the training process of our model, the paired signals will be extracted from *D* and used as the training set input (X_1^i, X_2^j). Among all signal pairs, half of the pairs are composed of signals of the same type, that is, $i = j$. The other half of the signal pairs are composed of different types of signals, that is, $i \neq j$. The output of the model is the probability $P(X_1^i, X_2^j)$ of which the two signals in the input signal pair are of the same category.

B. *The Proposed Few-shot ECG Diagnosis Method*

The framework used for arrhythmia diagnosis contains three main steps, data preparation, model training and model testing. In the data preparation stage, signal pairs will be input into two 4-layer one-dimensional CNNs with same model structure for feature extraction. The two CNNs have exactly the same parameters and share weights. Outputs of the two CNNs is two feature vectors of the same length. The structure of our model is shown in **Fig. 1** and the details of the CNNs are shown in **TABLE I**. In the model training stage, we calculate the distance of all the feature vectors generated by the pairs composed of the training set signals to obtain the probability matrix. By using the binary cross-entropy loss function, the parameters update. In the model testing stage, we will use the parameters saved during model training to classify the test signals and get the accuracy. A judgement method named *K* nearest similarity judgement will be used in the model testing, and the details of which will be explained below.

1) The Input of ECG Diagnosis Method

Before running the algorithm, we preprocessed the data in the dataset. The original ECG signals are segmented into beats and then the heartbeats are down-sampling to a certain length that contain 250 sample points. In addition, Normalization processing will be used for each data as shown in Eq. (2), where $\overline{x}$ and σ^2 denote the mean value and the variance of the sample attribute.

$$x = \frac{x - \overline{x}}{\sigma^2} \quad (2)$$

Every input of our framework is in pairs because this can generate more different pairs to train the model through permutation and combination of data. So that even though the amount of training data is small, it is not easy to overfit. For example, if we assume that the training set has *C* categories, and each category has *N* samples, the number of different pairs that may be generated is shown in Eq. (3). Considering case that *C*=4 and *N*=5, 190 different pairs are generated compared to the original 20 signals.

$$N_{pairs} = \binom{C \times N}{2} = \frac{(C \times N)!}{2!(C \times N - 2)!} \quad (3)$$

2) Distance Matrix

Each element in the distance matrix is actually the L1 distance (also called Manhattan distance) of the paired feature vectors as expressed in Eq. (4). Among the equation, *i* represents the index of minibatch, and *f* represents the feature vectors gained as the outputs of the two convolutional neural networks.

$$d_f^2(x_1^i, x_2^i) = \| f(x_1^i) - f(x_2^i) \| \quad (4)$$

TABLE I. Details of the layers of our convolutional neural network. The Layer details contain the size and dimensions of the convolution kernels.

No.	Layers	Layer details	Padding	Output shape
1	Conv1	1×10,64	valid	1×241×64
2	MaxPool1	1×2	same	1×121×64
3	Conv2	1×7,128	valid	1×115×128
4	MaxPool2	1×2	same	1×58×128
5	Conv3	1×4,128	valid	1×55×128
6	MaxPool3	1×2	same	1×28×128
7	Conv4	1×4,256	valid	1×25×256
8	MaxPool4	1×2	same	1×13×256

9	Dense	2048	1×2048

3) The Output of ECG Diagnosis Method

As we have described in the section A, the output of our network is the probability that the two signals in the input signal pair are of the same category. Through a fully connected layer with an activation function of sigmoid, the probability is obtained by the distance matrix shown in Eq. (4). The calculation of probability is shown in Eq. (5). Among them, i represents the index of minibatch, sigm represents the sigmoid activation function, which is shown in Eq. (6) and *FC* represents the fully connected layer.

$$P(x_1^i, x_2^i) = sigm(FC(d_f^2(x_1^i, x_2^i))) \quad (5)$$

$$S(x) = 1/(1+e^{-x}) \quad (6)$$

C. K-Nearest Similarity Judgement Testing

In order to improve the robustness and reduce the contingency, we introduce the *K* nearest similarity judgment to both testing and validation. Here we take testing process as example. In the testing process of Siamese few shot learning, we have a test set *T* and a support set *S* whose involved signals are all derived from the labelled training set. The value of *K* depends on *N*, which is the number of samples of a single category in the training set. In our test process, if $N \leq 10$, $K=N$, if $N>10$, $K=10$. In the *N*-shot learning test process, we randomly select a signal from the test set T, and randomly select four labeled signals of different types from the training set to form the support set S and this process is repeated *K* times (S_1, ..., S_K). Then the four signals in each support sets and the test signal are formed into pairs respectively and then inputted into the network to obtain similar probabilities. After that, if $K=1$, the category of the test sample is predicted as the largest probable type of signals. In general cases, the type of the test signal is predicted by Eq. (7).

$$C(x,(S_1,...,S_K)) = \arg\max_s(\sum_{k=1}^{K} P(x, x_{sk})), x_{sk} \in S_k \quad (7)$$

III. Experiment And Results

In this section, we will provide comparative experiments and results to prove the effectiveness and superiority of our proposed method.

A. Data Description

The experiments are all implemented on the MIT-BIH Arrhythmia Database. The records in this Database are obtained from 47 subjects studied by the BIH Arrhythmia Laboratory from 1975 to 1979. Each record is sampled at 360Hz in the range of 10 mV. All the recorded signals are annotated by multiple cardiologists. We divide the signals into five categories according to the Association for the Advancement of medical Instrumentation (AAMI), which are F (Fusion of ventricular and normal beat), N (Normal beat), S (Premature or ectopic supraventricular beat), V (Premature ventricular contraction) and Q (unclassified beat). The four categories considered in our experiments are F, N, S and V, because category Q is ignored since it accounted for very little proportion in the database. The details of the database are shown in **TABLE II**.

B. Experimental Setup

In order to verify our proposed method, we compared it with VGG12 [17], LSTM, CNN-LSTM and the CNN baseline described in TABLE I. For the experiment of VGG12, in order to enable the network to process one-dimensional ECG signals, we transform the original network architecture so that it can process one-dimensional data. In the LSTM experiment, we used two layers of LSTM with an output dimension of 256. And in the experiment of CNN-LSTM, two layers of LSTM with output dimension of 256 is added after the max pooling in step No.8 of the CNN structure shown in **TABLE I**.

Considering that the number of training sets for few shot learning is too small, the composition of the training set has a greater impact on the experimental results. Our experimental results are the average of multiple independent experiments, and the training set of each independent experiment is different and random. In the experimental process of our Siamese few-shot learning method, the average value of results is obtained by calculating 10 independent testing accuracies. In all other comparative experiments, the final accuracy is the mean of 10 independent experiment results. In all experiments, we considered five cases, 1 shot, 5 shot, 10 shot, 30 shot and 50 shot. As for *N* shot case, the number of samples of each category in the training set is *N*. Therefore, the size of training set of the five cases we considered above is 4, 20, 40, 120, and 200 respectively. The size of the test sets used in all experiments are all 400.

Our proposed architecture is implemented on Keras using an Nvidia 2070 GPU. To train the Siamese few-shot learning model, Adam optimizer with a learning rate of 0.00006 is used and the batch size is selected to be 32. The loss function used for network training is binary cross-entropy and the number of training epoch is set to 2500 in order to make sure the loss value to be stable.

TABLE II. Quantities of each category in the MIT-BIH Arrhythmia Database.

Category	F	N	S	V	Total
Quantity	802	89841	2927	7008	100578

C. Effectiveness of Siamese Few-Shot Learning Algorithm

Our proposed method, one-dimensional Siamese few-shot learning approach, is denoted as 1D-SIAMESE in the following part. And the CNN architecture used in 1D-SIAMESE for features extraction is denoted as CNN-BASE. It is worth mentioning that the training of the CNN-BASE network and that of 1D-SIAMESE are exactly the same in terms of the selection of activation function and convolution configurations such as the number of channels, kernel size and length of stride. The experimental results of 1D-SIAMESE and CNN-BASE in the case of 1 shot, 5 shot, 10 shot, 30 shot and 50 shot are shown in **TABLE III**.

It can be seen from the results that, due to the function of Siamese network architecture, the performance of 1D-SIAMESE is much better than that of CNN-BASE in all cases, leading by 19.13%, 14.78%, 17.36%, 11.20%, and 7.17% respectively. Especially in 1 shot case, the accuracy of 1D-SIAMESE is 82.36% while CNN-BASE only gains 63.23%. This illustrates that the paired input method of Siamese architecture can indeed enrich the effective features of the input samples, so that the network can obtain more classification judgments than normal networks when there are few training samples, so as to obtain better classification performances. In addition, as the training set becomes larger, the accuracies of 1D-SIAMESE and CNN-BASE both gradually increases. This is consistent with the law of deep learning.

TABLE III. The results of CNN-BASE and 1D-SIAMESE within 5 cases.

Accuracy (%)	1 shot	5 shot	10 shot	30 shot	50 shot
CNN-BASE	63.23	70.18	71.80	80.53	85.25
1D-SIAMESE	**82.36**	**84.98**	**89.16**	**91.73**	**92.42**

D. Comparison with Existing Methods

In order to prove the superiority of Siamese few shot learning algorithm, we will compare the experimental results of VGG 12, LSTM, CNN-LSTM and 1D-SIAMESE in the case of 1 shot, 5 shot, 10 shot, 30 shot and 50 shot. The results are shown in **TABLE IV** and **Fig. 2**.

The results show that in all cases, the accuracy of 1D-SIAMESE is the best among all methods, which are 82.36%, 84.98%, 89.16%, 91.73% and 92.42% respectively. For convenience, we only take 1 shot and 30 shot as examples. Under the extreme conditions of 1 shot, 1D-SIAMESE got an accuracy of 82.36%, while VGG 12, LSTM and CNN- LSTM gains accuracies of only 52.90%, 62.05% and 67.35%, leading by 29.46%, 20.31% and 15.01% respectively. In the case of 30 shots, the accuracy of our method is 91.73%, which means that raises of 17.25%, 11.9% and 17.3% happen since the accuracies of VGG 12, LSTM and CNN-LSTM are only 74.48%, 79.83% and 74.43% respectively. It is worth mentioning that, with the raise of data in the training set, our proposed method still has a good performance. According to the experimental results, our method still gains the highest accuracy in the case of 50 shot.

TABLE IV. The results of VGG 12, LSTM, CNN-LSTM and 1D-SIAMESE within 5 cases.

Accuracy (%)	1 shot	5 shot	10 shot	30 shot	50 shot
VGG 12	52.90	70.40	70.78	74.48	79.08
LSTM	62.05	69.15	74.53	79.83	80.40
CNN-LSTM	67.35	73.95	74.35	74.43	78.56
1D-SIAMESE	**82.36**	**84.98**	**89.16**	**91.73**	**92.42**

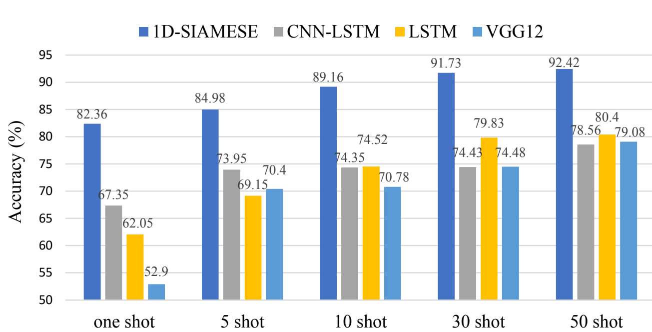

Fig.2. Results of VGG 12, LSTM, CNN-LSTM and 1D-SIAMESE within 5 cases.

IV. Conclusion

In this paper, we propose a few shot learning method based on the Siamese network in the ECG classification field to solve the problem of sample scarcity. The Siamese network reduces the requirements for the network ability of feature extraction through the combination of feature extraction and comparison methods. Also, we use paired inputs instead of normal single input. By constructing image pairs, the number of training samples is increased. Thus, good classification results can be obtained even when the training set is small. In addition, the comparative experiments implemented on the MIT-BIH Arrhythmia Database show that when the training set is small, the one-dimensional Siamese network can perform better than CNN-BASE, VGG12, LSTM and CNN-LSTM. Besides, as the amount of data in the training set increases, the one-dimensional Siamese network still has the best performance, as shown by 50 shot case. These illustrate the effectiveness and superiority of our method.

References:

[1]. Taylor, G.J., 150 Practice ECGs. 2008: Blackwell Publishing Ltd.
[2]. Liu, X., et al., Deep learning in ECG diagnosis: A review. Knowledge-Based Systems, 2021. 227: p. 107187.
[3]. Hoekema, et al., Geometrical Aspects of the Interindividual Variability of Multilead ECG Recordings. IEEE Transactions on Biomedical Engineering, 2001.
[4]. Faust, O., et al., Deep learning for healthcare applications based on physiological signals: A review. Computer Methods and Programs in Biomedicine, 2018. 161: p. 1-13.
[5]. Yıldırım, Ö., et al., Arrhythmia detection using deep convolutional neural network with long duration ECG signals. Computers in Biology and Medicine, 2018. 102: p. 411-420.
[6]. Qin, L., et al., An End-to-End 12-Leading Electrocardiogram Diagnosis System Based on Deformable Convolutional Neural Network With Good Antinoise Ability. IEEE Transactions on Instrumentation and Measurement, 2021. 70: p. 1-13.
[7]. Gao, Y., H. Wang and Z. Liu, A Novel Approach for Atrial Fibrillation Signal Identification Based on Temporal Attention Mechanism. Vol. 2020. 2020. 316-319.
[8]. Gao, Y., H. Wang and Z. Liu, An end-to-end atrial fibrillation detection by a novel residual-based temporal attention convolutional neural network with exponential nonlinearity loss. Knowledge-Based Systems, 2020. 212: p. 106589.
[9]. Gao, Y., H. Wang and Z. Liu, An end-to-end atrial fibrillation detection by a novel residual-based temporal attention convolutional neural network with exponential nonlinearity loss. Knowledge-Based Systems, 2021. 212: p. 106589.
[10]. Qi, L.A.J.Z., Revisit Long Short-Term Memory: An Optimization Perspective. 2015.
[11]. Oh, S.L., et al., Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. Computers in Biology and Medicine, 2018. 102: p. 278-287.
[12]. Qiao, F., et al., A Fast and Accurate Recognition of ECG Signals Based on ELM-LRF and BLSTM Algorithm. IEEE Access, 2020. PP(99): p. 1-1.
[13]. L., W., H. T. and T. Y. The One-Shot similarity kernel. in 2009 IEEE 12th International Conference on Computer Vision. 2009.
[14]. Qiao, S., et al., Few-Shot Image Recognition by Predicting Parameters from Activations. 2018. 7229-7238.
[15]. Che, T., Metagan: An adversarial approach to few-shot learning. 2019.
[16]. G Koch, R.Z.R.S., Siamese neural networks for one-shot image recognition. ICML deep learning workshop, 2015.
[17]. Simonyan, K. and A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 1409.1556, 2014.