

Lightweight Arrhythmia Detection Based on Momentum Contrast Learning

Zhongyu Wang, Caiyun Ma, Shuo Zhang, Yuwei Zhang, Jianqing Li*, Chengyu Liu*

State Key Laboratory of Digital Medical Engineering, School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China

Abstract

Arrhythmia disease can be extremely damaging to the heart, and in severe cases can even lead to death. The ECG smart monitoring device is an effective way for detecting arrhythmia disease, and as wearable devices spreads, it also places certain requirement on lightweight arrhythmia detection algorithms. It is of great importance to implement an efficient arrhythmia detection algorithm with strong generalization performance. This work trains an arrhythmia detection model on the Georgia 12-lead ECG Challenge (G12EC) database and the China Physiological Signaling Challenge 2018 (CPSC2018) database using xResNet18 as the backbone network and momentum contrast learning as the framework, which allows contrast learning of positive samples and a large number of negative samples by introducing queue and momentum update encoder parameters to obtain a more comprehensive information representation. The model was pre-trained using the Georgia 12-lead ECG Challenge (G12EC) Database to obtain better characterization of initialization information and fine-tuned using the China Physiological Signal Challenge 2018 (CPSC2018) dataset. The experimental results showed that the model was effective with an AUC of 0.861, an Acc of 77.04% on the CPSC2018 database.

1. Introduction

In recent years, with the development of society, arrhythmia diseases have become more and more widely concerned. The number of deaths due to arrhythmia disease is increasing every year, making it critical to monitor, screen, and diagnose cardiovascular disease in real time through wearable ECG devices so that patients can detect the condition and receive treatment earlier. At this stage, arrhythmia recognition is mainly diagnosed by forming 12-lead ECGs from the body surface potential difference collected by 12-lead electrode wires placed at different locations on the body. Many methods for automatic arrhythmia diagnosis using DNNs already exist, which has led to considerable development of deep

learning in the field of ECG [1-3].

With the development of Natural Language Processing (NLP) and Computer Vision (CV), the data feature mining capabilities and information characterisation capabilities of Deep Neural Networks (DNNs) have been further developed. Deep neural network models have been made more powerful in terms of data encoding and feature extraction by designing deeper network structures and using large training data sets [4]. However, the large amount of labelled data will consume huge amounts of manual labelling costs, making the training of models difficult.

Past years, self-supervision has received much attention in Deep Neural Network. Self-supervised learning aims to mine unlabelled data for its own representational properties by designing auxiliary tasks that can be used as supervised information [5-7]. It generates supervised information from proxy tasks to pre-train large-scale unlabelled data, obtains a representational encoding of the data itself, and uses the pre-trained model for downstream tasks. Effectively, only a small number of labelled samples are needed for supervised training in the downstream task to achieve the performance of a strongly supervised training model. It is worth mentioning that contrast learning, as a type of self-supervised learning, has excellent performance in optimizing encoder feature extraction performance. It generates different views of the same data by means of data augmentation. Multiple views of the same data are considered as positive classes among themselves and negative classes among different data. Contrast learning maximizes the similarity between negative classes and minimizes the similarity between positive classes for the purpose of enhancing encoder representation. The representative methods in contrast learning are SimCSE [8], SimCLR [9], CMC [10], CPC [11], BYOL [12] and momentum contrast (MoCo) [13]. MoCo transforms the contrast similarity problem into a query and key query problem, and proposes the use of queue to store key values combined with momentum to update the encoder, which solves the problem of large amount of data required for contrast learning that is difficult to train, and fills the gap between unsupervised learning and supervised learning. For 12-lead ECG, the cost of manually annotating ECG data is much higher than

for other types of data. It requires more specialized medical knowledge to support, and some types of arrhythmia disease data are very scarce, which is too expensive for high-volume labelling of ECG data. Therefore, the use of self-supervised learning methods to pre-train large-scale unlabelled ECGs and then fine-tune them using small batches of labelled data in downstream tasks to achieve automatic diagnosis of arrhythmia diseases is highly preferred.

In this paper, a novel self-supervised model for 12-lead ECG arrhythmia classification is proposed. The model is pre-trained with MoCo-based contrast learning for 12-lead ECG, and the pre-trained model is applied to the downstream task representation. Specifically, the main contributions of this paper are as follows: (1) During data enhancement, Gaussian noise was randomly added to the 12-lead ECG, and good characterization was achieved in contrast learning, (2) The lightweight model xResNet18 which is less computationally was used as an encoder for MoCo. The pre-trained xResNet18 have a good performance in downstream tasks, (3) A suitable Gaussian noise intensity for the data enhancement process was found.

2. Methods

2.1. Database

All the databases used in the paper, including the Georgia 12-lead ECG Challenge (G12EC) Database and China Physiological Signal Challenge 2018 (CPSC2018) database. The Georgia 12-lead ECG Challenge (G12EC) Database contains 10344 12-lead ECGs with a length between 5 and 10 seconds and a sampling frequency of 500 Hz. The CPSC2018 database contains 6877 12-lead ECG arrhythmia records from 11 hospitals with a balanced male-to-female ratio. The duration of each record is between several seconds and tens of seconds, and the sampling rate is 500Hz. The types of records are: Normal, Atrial fibrillation (AF), First-degree atrioentric block (I-AVB), Left bundle branch block (LBBB), Right bundle branch block (RBBB), Premature atrial contract (PAC), Premature ventricular contraction (PVC), ST-segment depression (STD) and ST-segment elevated (STE). We use the Georgia 12-lead ECG Challenge (G12EC) Database as a pre-trained dataset and the CPSC2018 database as a dataset to test the performance of the model for downstream tasks. For each database, sliding windows of 10s with a 5s overlap is used to intercept the data segment. The pre-trained database is a self-supervised process, so its data label is not required. The label of the arrhythmia data segment from CPSC2018 database is consistent with the label of the record it originally belonged to. Further details regarding the data are provided in Table 1.

Table 1. The details of the database.

G12EC	Class	Train size	Test size	Total
	-		8233	2059
CPSC2018	Class	Train size	Test size	Total
	Normal	1335	373	1708
	AF	1585	385	1970
	I-AVB	909	236	1145
	LBBB	299	64	363
	RBBB	2329	524	2853
	PAC	1201	276	1477
	PVC	1491	391	1882
	STD	1132	312	1444
STE	347	96	443	

2.2. Preprocessing

To better validate the performance of the model, all ECG records were filtered using a bandpass filter with passband frequencies from 0.5 to 45 Hz, and the filtered signals were normalized to have a mean of 0 and a variance of 1.

2.3. xResNet18

The basic architecture of the xResNet18 model [14] is shown in Figure 1. The model consists of Input stem, output and 4 stage modules. The Input stem module consists of three convolutional layers, each followed by a BN layer and a ReLu activation layer. For the 4 stage modules, each stage consists of 1 Down sampling and 2 Residual blocks.

In this paper, we use the model as a feature extractor for MoCo contrast representation learning and use the pre-trained model parameters for downstream tasks.

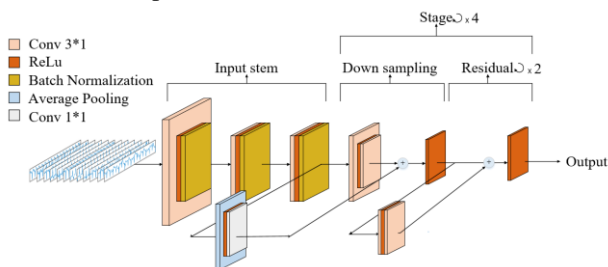


Figure 1. Basic architecture of the adopted xResNet18 model.

2.4. Contrastive Learning Framework

In this paper, we used the MoCo contrast learning representation framework with queue and momentum update parameters at its core to enhance the explanatory power of the model in the pre-training phase. The framework introduces queues in the process of learning representations so that the comparison representations contain more negative samples and ensure stronger

generalization performance of the model. At the same time, the framework updates the parameters of the encoder in a momentum update manner, so that the negative sample representations in the queue are more consistent with the features of the positive samples. The specific pre-training-fine-tuning process is shown in Figure 2. The framework consists of Data Augmentation, Encoder, Momentum Encoder, MLP layer and queue, where the Momentum Encoder is momentum updated as the parameters of the Encoder change, which is calculated as shown as follows:

$$\theta_m \leftarrow m\theta_m + (1 - m)\theta_k$$

where, θ_m is the parameter of the Momentum Encoder, θ_k is the parameter of the Encoder, m is the momentum update parameter and is set to 0.999.

In the pre-training phase, we initialized a queue to store negative samples according to the FIFO principle based on the MoCo setup, and the length of the queue was set to 6400, which is a subset of the training set. For an ECG recording of 10s in length, each second data segment was treated as a computational unit. We add Gaussian noise segments randomly in units of computational units on different leads of the same 12-lead ECG x_i for data enhancement to obtain positive sample pairs x'_i , where the percentage of noise addition was 80% for each lead. The formula for Gaussian noise is shown as follows:

$$G(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where, μ and σ are the mean and variance of the generated Gaussian noise and are set to 0 and 0.1, respectively.

The original data x_i and the augmented data x'_i are used as inputs to the Encoder and the Momentum Encoder, respectively, to obtain the characterization features H and H' . H and H' are then further enhanced by the MLP layer for contrast learning interpretation, as described in the literature[9].

The essence of contrast learning is to improve the feature extraction ability of the model by increasing the representational similarity between positive samples (G and G') and decreasing the representational similarity with negative samples (G and queue). Therefore, InfoNCE was chosen as the loss function to calculate the similarity between positive and negative samples. The loss function is calculated as follows:

$$L_q = -\log \frac{\exp((q \cdot k_+)/\tau)}{\sum_{i=0}^K \exp((q \cdot k_i)/\tau)}$$

where, q and k_+ are the representations of positive sample pairs in the mini-batch, and k_i is the representation of negative samples in the queue. τ denotes the temperature parameter, which is set to 0.07 in this paper. K denotes the size of the queue.

After completing the forward computation of a mini-batch, the Momentum Encoder is updated with the Encoder parameters, G' is sent to the queue, and the earliest data entering the queue is sent out.

In the downstream task, the pre-trained encoder

parameters are used directly and the classifier head is connected for arrhythmia disease classification. During both pre-training and downstream fine-tuning training, the Adadelta optimizer was used and the learning rate was set to 0.005 and 0.001 for 100 and 200 training epochs with batch size of 16 and 32, respectively.

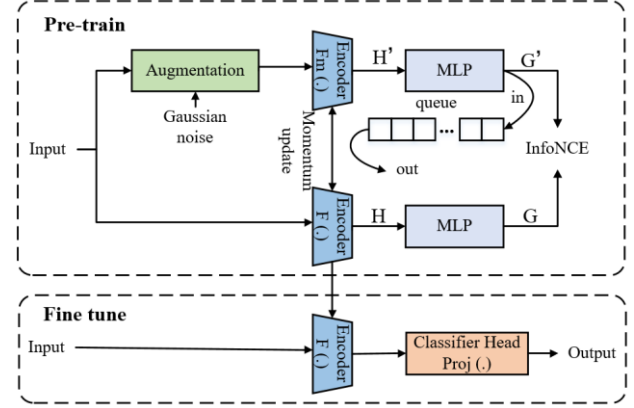


Figure 2. MoCo Comparative Learning Framework.

2.5. Evaluation methods

For classification results, Sensitivity (Se), Specificity (Sp), Accuracy (Acc), and AUC -score are used as evaluation indicators. According to the positive or negative of the label, two indexes were used: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Where,

$$Acc = (TP + TN)/(TP + FN + TN + FP).$$

The AUC -score refers to the area enclosed with the coordinate axis under the ROC curve. The curve is plotted according to a series of different cut-off values, with Se as the vertical coordinate and Sp as the horizontal coordinate. The higher the AUC -score, the better the classifier effect.

3. Results

We added different intensities of noise intensity during the contrast learning pre-training, with σ being set to [0.1, 0.2, 0.3, 0.4, 0.6, 0.8], all with a noise addition ratio of 80%, and after the same number of training epochs. The downstream task performance of the model under different Gaussian noise intensities is shown in Figure 3. As seen in the figure, the AUC is maximum at σ 0.1, which is 0.861. More detailed performance metrics for the model at different σ values are shown in Table 2. When the σ was 0.1, the model achieved an Acc of 77.04% in the CPSC2018 arrhythmia database.

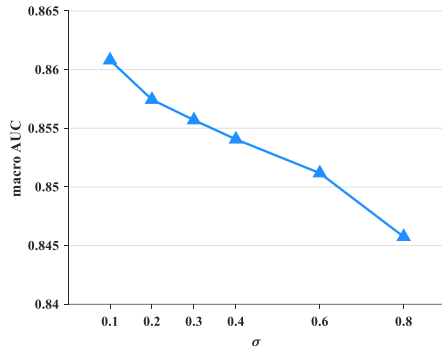


Figure 3. Comparison of macro AUC-ROC score at different Gaussian noise intensities after training 200 epochs.

Table 2. The details of performance metrics for the model.

σ	Acc(%)	AUC
0.1	77.04	0.861
0.2	76.28	0.857
0.3	76.13	0.855
0.4	76.40	0.854
0.6	76.43	0.851
0.8	74.82	0.845

4. Discussion

As shown in Figure 3 and Table 2, different intensities of noise had different degrees of influence on the performance of the model, and the highest Acc and AUC were 77.04% and 0.861 when the variance of Gaussian noise was 0.1. In summary, the method is an effective lightweight arrhythmia detection algorithm, and the model is more amenable to deployment to wearable devices than complex models.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (62171123, 62071241, 62201144 and 62211530112).

References

[1] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Medicine*, vol. 25, no. 1, pp. 65-69, Jan 2019, doi: 10.1038/s41591-018-0268-3.

[2] K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, Jun 27-30 2016, in IEEE Conference on

Computer Vision and Pattern Recognition, 2016, pp. 770-778, doi: 10.1109/cvpr.2016.90.

[3] S. Zhang, Y. W. Li, X. Y. Wang, H. X. Gao, J. Q. Li, and C. Y. Liu, "Label decoupling strategy for 12-lead ECG classification," *Knowledge-Based Systems*, vol. 263, Mar 2023, doi: 10.1016/j.knosys.2023.110298.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, May 2015, doi: 10.1038/nature14539.

[5] R. Krishnan, P. Rajpurkar, and E. J. Topol, "Self-supervised learning in medicine and healthcare," *Nature Biomedical Engineering*, vol. 6, no. 12, pp. 1346-1352, Dec 2022, doi: 10.1038/s41551-022-00914-1.

[6] L. Schmarje, M. Santarossa, S. M. Schroder, and R. Koch, "A Survey on Semi-, Self- and Unsupervised Learning for Image Classification," *IEEE Access*, vol. 9, pp. 82146-82168, 2021, doi: 10.1109/access.2021.3084358.

[7] L. L. Jing and Y. L. Tian, "Self-Supervised Visual Feature Learning with Deep Neural Networks: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 11, pp. 4037-4058, Nov 2021, doi: 10.1109/tpami.2020.2992393.

[8] T. Y. Gao, X. C. Yao, D. Q. Chen, and L. Assoc Computat, "SimCSE: Simple Contrastive Learning of Sentence Embeddings," in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Punta Cana, DOMINICAN REP, Nov 07-11 2021, 2021, pp. 6894-6910.

[9] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," in *International Conference on Machine Learning (ICML)*, Electr Network, Jul 13-18 2020, vol. 119, in Proceedings of Machine Learning Research, 2020.

[10] Y. Tian, D. Krishnan, and P. Isola, "Contrastive Multiview Coding," in *European Conference on Computer Vision (ECCV)*, 2020, pp. 776-794.

[11] Aaron, Y. Li, and O. Vinyals, "Representation Learning with Contrastive Predictive Coding," *arXiv pre-print server*, 2019-01-22 2019, doi: None. arxiv:1807.03748.

[12] J.B. Grill, S. Florian, A. Florent, T. Corentin, H. R. Pierre, B. Elena, D. Carl, A. P. Bernardo, D. G. Zhaohan, G. A. Mohammad, P. Bilal, K. Koray, M. Rémi, V. Michal, "Bootstrap your own latent: A new approach to self-supervised Learning," *arXiv pre-print server*, 2020-09-10 2020, doi: None. arxiv:2006.07733.

[13] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum Contrast for Unsupervised Visual Representation Learning," *arXiv pre-print server*, 2020-03-23 2020, doi: None. arxiv:1911.05722.

[14] T. He, Z. Zhang, H. Zhang, Z. Y. Zhang, J. Y. Xie, M. Li, "Bag of Tricks for Image Classification with Convolutional Neural Networks," in *32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, Jun 16-20 2019, in IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 558-567, doi: 10.1109/cvpr.2019.00065.

Address for correspondence:

Jianqing Li and Chengyu Liu.
Sipailou Road 2, Southeast University, Nanjing, China.
E-mail: lj@seu.edu.cn and chengyu@seu.edu.cn