ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10095035

MLCGAN: MULTI-LEAD ECG SYNTHESIS WITH MULTI LABEL CONDITIONAL GENERATIVE ADVERSARIAL NETWORK

Jian Wu, Liping Wang⋆*, Hailin Pan, Binyu Wang*

Shanghai Key Laboratory of Trustworthy Computing, East China Normal University, Shanghai, China

ABSTRACT

Electrocardiography(ECG) is a non-invasive tool used to identify the cardiovascular diseases. ECG classification studies have been concerned and made progress well. However, the problems about categories imbalance and absence of labelled clinic data are still dramatically hindered research development. Recently, generative models have been verified as a possible way to handle the data scarcity issues. For ECG synthesis, to the best of our knowledge as the reason of time sequences and multiple labels constraints, no model can generate ECG corresponding to clinic data.

In this paper, we present a novel **m**ulti-**l**abel **c**onditional **g**enerative **a**dversarial **n**etwork, named MLCGAN. To synthesise reasonable long-term multi-lead data, multi-label mixing module is devised to combine with our improved WaveGAN. Moreover, the sampling strategy based on multi-labels distribution is proposed. Comprehensive experiments demonstrate that MLCGAN can generate ECG data satisfied the clinic diagnose requirement and improve the performance of RestNet based ECG classifier.

Index Terms— Class Imbalance, Generative Adversarial Networks(GAN), Multi label classification, Electrocardiography(ECG)

1. INTRODUCTION

Electrocardiography (ECG) is a non-invasive tool frequently used in clinical medicine to identify and monitor heart problems. Multi-lead ECG monitors the status of the heart from different location and has saved many patients' lives. Recently, deep learning based ECG classification studies play an important role for automatic diagnosis and have made progress well [1, 2, 3]. However, training an effective deep learning model requires considerable data. The analysis and labeling of ECG are time-consuming and expensive. Additionally, as the imbalance data distribution among categories, it is very hard to get good performance on the category with minority of ECG data.

Generative adversarial network [4] (GAN) has shown excellent performance in image synthesis [5], which is a deep generative model mainly comprised of a generator and a discriminator. The generator networks are used to generate real data from random noise, and the discriminator network is used to determine whether the input data is real or generated. To solve ECG data problems, GAN is introduced to augment the training set during ECG classification [6, 7]. Many studies [7, 8, 9, 10] have shown that adding the generated ECG data to the training set can improve the classifier's performance. PGANs [6] propose a generative model that learns to synthesize patient-specific ECG signals. Note that as shown in fig.1(a), most existing works use GAN to generate heartbeats-level labeled ECG. However, these efforts did not concentrate on the long-term ECG data which is with multi-label and multi-lead shown in fig.1(b). Here, long-term ECG data classification is a multi-label problem, while heartbeat level ECG data is a multi-class problem. Compare to heartbeats-level ECG synthesis, long-term ECG data can satisfy the clinic diagnose requirement better, meanwhile it need more effort to solve the problems of leads and labels.

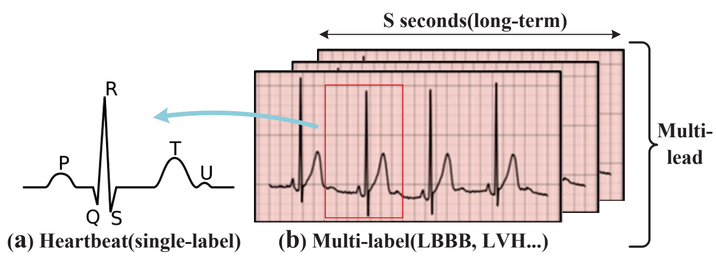

Fig. 1. Comparison of heartbeat level ECG and long-term multi-lead ECG.

In this paper, we focus on improving the classifier's performance by using GAN to generate long-term multi-lead and multi-label ECG. Owing to WaveGAN's outstanding performance in time series data synthesis, WaveGAN is utilized as the central part of the generator. Moreover, the modified WaveGAN [11] is introduced to support multi-lead output. CGAN [12] is improved to mix the multi-label information into the generator.

The main contributions of this paper are summarized as follows:

(1) We examined the shortcomings of existing ECG synthesis methods and proposed a novel multi-label conditional generative adversarial network, MLCGAN. To the best of our knowledge, it is the first time to generate class-specific multi-label ECG data.

(2) To synthesise reasonable long-term multi-lead data, multi-label mixing module is devised and sampling strategy based on actual multi-label distribution is provided.

(3) Comprehensive experiment results demonstrate that our approach achieves synthetic ECG satisfied the clinic requirement, and improves the performance of ECG classifier.

Corresponding author, lipingwang@sei.ecnu.edu.cn.

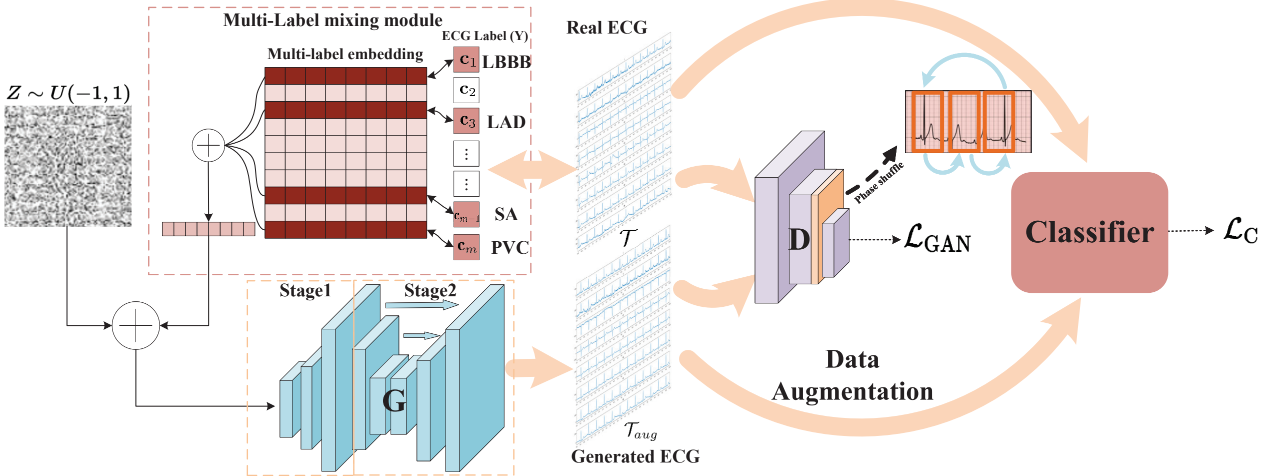

Fig. 2. Illustrating the generation process, in which "G", "D" denote a generator and a discriminator , respectively. "Z" indicates a random noise vector sampled from the uniform distribution. The Label information mixing result is added to the "noise" and entered into the generator.

2. PROPOSED FRAMEWORK

Long-term multi-lead ECG signal is a sequence of data points, each ECG sample can be formulated as $\tau = \langle \mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_l, Y \rangle$. Here, l means the length of sample points, e.g. 10s ECG signal with 500HZ sample rate should be represented as 5000 sample points. Each observation $\mathbf{s}_t \in \mathbb{R}^k$ is a k-dimensional vector. Thus, if $k = 1$, τ is a single lead ECG, and if $k > 1$, τ is a multi lead ECG. Moreover, each signal τ is related to one or more ECG label, denoted as set Y, which means the cardiovascular disease info. In this work, we aimed to construct novel network to generation reasonable τ with $k > 1$ and multi-label in Y. For simplicity, during following discussion we appoint that τ is 8 x 5000. The proposed framework of MLCGAN is shown in Fig.2. The detail about each module is illustrated in the following subsections.

2.1. Multi-Label mixing module

Multi-label Embedding. In clinic, long-term ECG usually contains multiple labels. Specifically, for each ECG τ, let label set $Y = \langle \mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_m \rangle$, where m means the total number of categories and $\mathbf{c}_i = 1$ if τ contains this disease type. In order to generate class-specific data, we need to add some additional information to the generator. In CGAN, a single-label was added as extra information to the generator network. However, we discovered that it is simple to have overfitting conditions during the training process. In order to produce multi-label data, we first employ learnable label embeddings to discover each category's unique vector representation. After that we sum up the vector representation of each category according to the label information. Finally, we combine random noise with embedding information as input to the generator.

Real label sampling. Traditionally, a random number was used to decide the type of data created in the GAN data generation process. However, generating multi-label information will result in an exponential label space. These generated labels may contain contradictory information, such as tachycardia and bradycardia appearing concurrently in labels. To address this issue, we present a sampling approach in actual labels that ensures the labels of the generated data exist in real scenarios. To generate data containing a specific disease, we first acquire all labels in the dataset that contain the disease and then random select a subset of those labels for the generated data. This method assures that all generated data is meaningful.

2.2. Generator

It is essential to monitor data from multiple beats simultaneously because the disease often causes changes in multiple beats over the long-term ECG. As shown in Fig.2, our model architecture is based on WaveGAN, which has achieved excellent results in speech synthesis. The generator is divided into two stages. The first stage primarily generates multi-lead ECG data by combining multi-label information with random noise. It is made up mainly of six identical transposed convolution blocks. Each module includes an up-sampling layer, a 1D convolutional layer with padding, and ReLU activation function. It is worth noting that in the final block, we use Tanh as the activation function with the DCGAN [13] training strategy. The second stage further improves the quality of the generated ECG using a U-Net-like architecture [14] which is widely used in image segmentation. It consists of a contraction path and an expansion path. There are six 1D convolutions along the contraction path, each followed by a leaky rectified linear unit (Leaky ReLU). The extended path is made up of six transposed convolutions that are identical to the first stage. Both stages can be used as generator, and we chose these two stages to splice primarily for the following reasons: a deeper network can be well integrated into the network with label information; if only the first stage is used, it is challenging to generate the desired ECG type; Secondly, the second stage of the network input dimension is 8x5000, adding label information at this stage results in an enormous embedding dimension. Label information is challenging to include in the network. Experiments revealed that combining these two stages of the network produces the most significant outcomes.

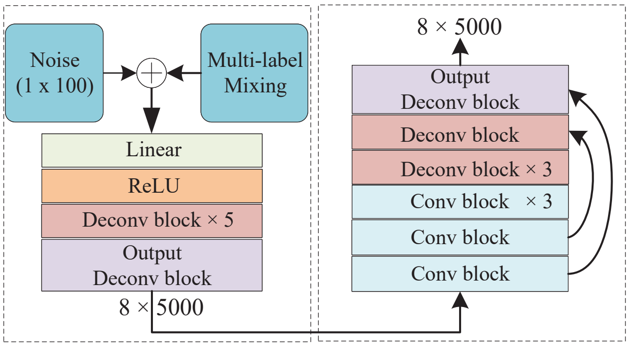

Fig. 3. The framework of two-stage generator.

2.3. Discriminator

The discriminator is used to distinguish whether the input data(8 x 5000) came from the generator or the actual ECG data. It outputs a value indicating the likelihood that the input data is from real data. The discriminator comprises seven modules, each with a one-dimensional convolutional layer, a Leaky ReLU activation layer, and the phase shuffle layer. The discriminator can easily tell whether the ECG data is created data because it has a similar periodicity to the sound signal, which will have an impact on how the GAN network is trained [15]. We follow the phase shuffling process proposed in WaveGan, randomly perturbing the activation layer's phase. It can improve the discriminator's judgment difficulty and prevent it from learning shallow information.

2.4. Objective

GAN can be seen as a two player min-max game. Formally, the objective function of the generator and discriminator can be expressed in the following form:

$$\begin{aligned}\min_G \max_D V(G, D) = \mathbb{E}_{x \sim P_{data}(x)}[\log D(x|y)] \\ + \mathbb{E}_{z \sim P_z(z)}[\log(1 - D(G(z|y)))]\end{aligned} \tag{1}$$

where y is the representation of multi-label information, and D will be trained to assign the correct labels to both real and generated data. We will also train G to minimize the equation $\log(1 - D(G(z|y))$. The discriminator can easily distinguish the generated data during the early stages of training when the generator is weak. In practice, we usually let G maximize $\log D(G(z|y)$, which can provide more efficient gradients. Nevertheless, the GAN training still suffers from gradient instability and mode collapse.

WGAN [16] uses wasserstein distance which is also called Earth Mover's Distance (EMD), to calculate the difference between the two distributions. But the generator must satisfy a Lipschitz constraint. Weight cropping is used in WGAN to meet the Lipschitz constraint by adjusting the model's parameters to a limited range after each training. However, this arbitrary operation can lead to a large number of parameters concentrated in the boundary values, leading to undesired behavior [17].

For the stability of training, We employ the Wasserstein loss with gradient penalty [17] and the objective is:

$$\begin{aligned}\mathcal{L}_{\text{GAN}} = \underset{\tilde{\boldsymbol{x}} \sim \mathbb{P}_g}{\mathbb{E}}[D(\tilde{\boldsymbol{x}})] - \underset{\boldsymbol{x} \sim \mathbb{P}_{data}}{\mathbb{E}}[D(\boldsymbol{x})] \\ + \lambda \underset{\hat{\boldsymbol{x}} \sim \mathbb{P}_{\hat{\boldsymbol{x}}}}{\mathbb{E}}\left[(\|\nabla_{\hat{\boldsymbol{x}}} D(\hat{\boldsymbol{x}})\|_2 - 1)^2\right].\end{aligned} \tag{2}$$

where P_{data} denotes the data distribution, P_g denotes the model distribution implied by $\tilde{x} = G(z|y), z \sim p(z)$ and $P_{\hat{x}}$ comes from sampling uniformly along straight lines between pairs of points sampled from the data distribution P_{data} and the generator distribution P_g, λ is a penalty coefficient.

3. EXPERIMENTAL RESULTS

3.1. Dataset

Tianchi dataset. We evaluate our model on the Tianchi dataset[18]. This dataset contains over 40,000 long-term ECG data divided into 55 categories. It has over 22,000 pieces of anomalous data, including 15 types with >1000 samples, 10 kinds' sample size ≤1000 and ≥100, and 29 types with <100 samples. In our work, we focus on improving the classification performance of the categories with around 1000 samples. Each record contains ten seconds of eight-lead(I, II, V1, V2, V3, V4, V5, V6) information at 500 Hertz. We choose about 8000 data as the test set and the rest of the data as the training set. In the training set, we use 20% of the data as the validation set.

Tianchi dataset subset. Since the total Tianchi data contains more than 50 categories, it is challenging to perform data augmentation for each category, so we choose a small dataset with labels containing left axis deflection(LAD), right axis deflection(RAD) and Right Bundle Branch Block(RBBB) [9]. There are about 2000 data in the training set.

3.2. Evaluation

Base Classify. Since ResNet [19] shows excellent results in ECG classification, we use the modified resnet34[19], which can handle 1D data. We use binary cross-entropy loss with weights to evaluate the accuracy of each category.

$$\mathcal{L}_C = -W\left[Y \cdot \log X + (1 - Y) \cdot \log\left(1 - X\right)\right] \tag{3}$$

W is obtained by calculating the proportion of each category of data in the training set to the overall data, X is the prediction vector and Y is the target vector.

Baseline Models. We compare our model MLCGAN to the following baseline models. We compare to CGAN [12] and

Table 1. Multi-Label classification performance on the data categories used for data generation(T Wave Changes)

Method	Alibaba Dataset					
	P	R	Micro-f1	TWC_P	TWC_R	TWC_Micro-f1
Resnet 1D-34	80.16%	76.33%	78.20%	68.24%	73.52%	70.78%
CGAN	80.32%	76.35%	78.28%	68.27%	**76.71**%	72.25%
Stage1	79.87%	76.37%	78.08%	68.35%	74.33%	71.21%
Stage2	79.24%	75.42%	77.28%	67.95%	75.42%	71.49%
our model	**81.71**%	**77.95**%	**79.79**%	**74.88**%	73.05%	**73.95**%

Table 2. Multi-Label classification performance on the data categories used for data generation(Right Bundle Branch Block)

Method	Alibaba Dataset					
	P	R	Micro-f1	RBBB_P	RBBB_R	RBBB_Micro-f1
Resnet 1D-34	80.16%	76.33%	78.20%	45.94%	57.95%	51.25%
CGAN	79.95%	76.82%	78.35%	45.83%	50.00%	47.82%
Stage1	79.87%	76.37%	78.08%	**50.00**%	46.59%	48.23%
Stage2	80.60%	76.18%	78.33%	47.47%	53.40%	50.26%
our model	**81.23**%	**77.90**%	**79.53**%	48.00%	**59.25**%	**53.03**%

Table 3. Comparison Multi-Label classification performances of 1D ResNet-34 with synthetic ECG produced between our method to other GAN model.

Method	Diseases(f1-score)				
	LAD	RAD	RBBB	micro-f1	macro-f1
1D ResNet-34	0.969	0.981	0.840	0.963	0.930
CGAN	0.966	0.986	0.819	0.962	0.923
ACGAN	0.968	0.984	0.824	0.963	0.926
stage1	0.967	0.985	0.825	0.963	0.926
stage2	0.967	0.986	0.841	0.965	0.931
our model	**0.971**	**0.989**	**0.863**	**0.971**	**0.941**

ACGAN [20] since they can both create ECG data for a specific heart rate type based on label information. We have modified these models so that they can generate eight-lead ECG data. To demonstrate the effectiveness of the two-stage generator, we use only the first and second stages as the main structure of the generator to synthesize the data, respectively.

Experimental Setup. We evaluate our model from two perspectives. First, we used the entire Tianchi dataset, which included 55 categories. We extracted ECGs containing T-wave change(TWC) and Right Bundle Branch Block(RBBB) to be trained using our model separately. After that, the data labels from that particular category are fed into the GAN model to generate the category-specific data. We feed the 500 generated data into the training dataset as supplementary data. We assess the classification metrics for that specific disease type and the general classification metrics. Second, We generate as much data as the original dataset in Tianchi dataset subset and compared them with other GAN-related models.

In the GAN model, we use the Adam [21] optimizer, and we use the WGAN-GP mentioned in Section 2 as the loss function for the model to converge as quickly as possible. We set the penalty coefficient $\lambda = 10$. We continue the training procedure for 2000 epochs. In the classification model, the maximum epoch is 160. We set the learning rate of the initial stage to 1e-3, with a learning rate decay strategy. The decay occurs at the 32, 64, and 128th epochs, and we reduce the learning rate by ten for each decay interval.

We consider precision, recall, and f1-score metrics in our evaluation. We use micro-f1 as an evaluation metric to assess the overall classification effect.

Synthesis Performances. Tables 1 and 2 show the micro average of 55 categories and the overall classification performance after adding the category-specific data, respectively. Our model can successfully improve the categorization of specified categories and the overall classification effect by synthesizing multi-label data.

We assess our model's performance using a subset of the Tianchi dataset. The comparison results are shown in Table 3. We can observe that synthesizing data with GANs can result in better categorization results. The CGAN model is weaker than ours because it can only synthesize single-label data, which does not improve multi-label classification very well and leads to overfitting. Using only stage1 as a generator does not give good results because the model does not extract the correlation between the waveforms. Using stage2 as the generator is less effective because it is challenging to integrate multi-label information into the network.

4. CONCLUSION

In this paper, we propose a GAN model for multi-lead long-term ECG synthesis called MLCGAN. We combine disease information mixing with WaveGAN's ability to process time-series data to generate long-term ECGs. In addition to this, the strategy of sampling from real tags effectively prevents the generation of irrational data. Experiments show that our model can effectively assist in ECG classification.

5. REFERENCES

[1] Mohamad Mahmoud Al Rahhal, Yakoub Bazi, Haikel AlHichri, Naif Alajlan, Farid Melgani, and Ronald R Yager, "Deep learning approach for active classification of electrocardiogram signals," *Information Sciences*, vol. 345, pp. 340–354, 2016.

[2] Mohammad Kachuee, Shayan Fazeli, and Majid Sarrafzadeh, "Ecg heartbeat classification: A deep transferable representation," in *2018 IEEE international conference on healthcare informatics (ICHI)*. IEEE, 2018.

[3] Lahcen El Bouny, Mohammed Khalil, and Abdellah Adib, "ECG heartbeat classification based on multi-scale wavelet convolutional neural networks," in *2020 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2020, Barcelona, Spain, May 4-8, 2020*.

[4] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," *neural information processing systems*, 2014.

[5] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *ICCV 2017*.

[6] Tomer Golany and Kira Radinsky, "Pgans: Personalized generative adversarial networks for ecg synthesis to improve patient-specific deep ecg classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019.

[7] Tomer Golany, Gal Lavee, Shai Tejman Yarden, and Kira Radinsky, "Improving ecg classification using generative adversarial networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020.

[8] Tomer Golany, Kira Radinsky, and Daniel Freedman, "Simgans: Simulator-based generative adversarial networks for ecg synthesis to improve deep ecg classification," in *International Conference on Machine Learning*. PMLR, 2020.

[9] Jintai Chen, Kuanlun Liao, Kun Wei, Haochao Ying, Danny Z. Chen, and Jian Wu, "ME-GAN: learning panoptic electrocardio representations for multi-view ECG synthesis conditioned on heart diseases," in *International Conference on Machine Learning, ICML 2022*.

[10] Faezeh Nejati Hatamian, Nishant Ravikumar, Sulaiman Vesal, Felix P. Kemeth, Matthias Struck, and Andreas K. Maier, "The effect of data augmentation on classification of atrial fibrillation in short single-lead ECG signals using deep neural networks," in *2020 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2020, Barcelona, Spain, May 4-8, 2020*, 2020, pp. 1264–1268.

[11] Vajira Thambawita, Jonas L Isaksen, Steven A Hicks, Jonas Ghouse, Gustav Ahlberg, Allan Linneberg, Niels Grarup, Christina Ellervik, Morten Salling Olesen, Torben Hansen, et al., "Deepfake electrocardiograms using generative adversarial networks are the beginning of the end for privacy issues in medicine," *Scientific reports*, vol. 11, no. 1, pp. 1–8, 2021.

[12] Mehdi Mirza and Simon Osindero, "Conditional generative adversarial nets," *CoRR*, vol. abs/1411.1784, 2014.

[13] Alec Radford, Luke Metz, and Soumith Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun, Eds., 2016.

[14] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015.

[15] Chris Donahue, Julian McAuley, and Miller Puckette, "Adversarial audio synthesis," in *International Conference on Learning Representations*, 2018.

[16] Martin Arjovsky, Soumith Chintala, and Léon Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*. PMLR, 2017.

[17] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville, "Improved training of wasserstein gans," *Advances in neural information processing systems*, vol. 30, 2017.

[18] "Link for tianchi ecg classification competition dataset," [Online], `https://tianchi.aliyun.com/competition/entrance/231754/information`.

[19] Cláudia Brito, Ana Machado, and António Luís Sousa, "Electrocardiogram beat-classification based on a resnet network.," in *MedInfo*, 2019.

[20] Augustus Odena, Christopher Olah, and Jonathon Shlens, "Conditional image synthesis with auxiliary classifier gans," in *International conference on machine learning*. PMLR, 2017.

[21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, 2014.