IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 1, 2021, accepted March 21, 2021, date of publication March 30, 2021, date of current version April 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3069827

ProEGAN-MS: A Progressive Growing Generative Adversarial Networks for Electrocardiogram Generation

HAIXU YANG[1], JIHONG LIU[1], LVHENG ZHANG[1], YAN LI[1], AND HENGGUI ZHANG[2,3,4]

[1]College of Information Science and Engineering, Northeastern University, Shenyang 110819, China
[2]International Laboratory for Smart Systems and Key Laboratory of Intelligent of Computing in Medical Image, Ministry of Education, Northeastern University, Shenyang 110004, China
[3]School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, U.K.
[4]Peng Cheng Laboratory, Shenzhen 518055, China

Corresponding author: Jihong Liu (liujihong@mail.neu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61572152 (to HZ), and in part by the Science Technology and Innovation Commission of Shenzhen Municipality under Grant JSGG20160229125049615 (to HZ).

ABSTRACT Electrocardiogram (ECG) is a physiological signal widely used in monitoring heart health, which is of great significance to the detection and diagnosis of heart diseases. Because abnormal heart rhythms are very rare, most ECG datasets have data imbalance problems. At present, many algorithms for ECG anomaly automatic recognition are affected by data imbalance. Conventional data augmentation methods are not suitable for the augmentation of the ECG signal, because the ECG signal is one-dimensional and their morphology has physiological significances. In this paper, we propose a ProGAN based ECG sample generation model, called ProEGAN-MS, to solve the problem of data imbalance. The model can stably generate realistic ECG samples. We evaluate the fidelity and diversity of the data generated by the model and compare the data distribution of the original and generated data. In addition, in order to show the diversity of the generated ECG data more intuitively, we manually checked the diversity and calculate the statistics of the data. The results show that compared with other ECG augmentation methods based on GANs, the ECG data generated by our model has higher fidelity and diversity, and the distribution of generated samples is closer to the distribution of original data. Finally, we established neural network models for arrhythmia classification, and used them to evaluate the improvement of the classification model performance by ProEGAN-MS. The results show that augmented data by ProEGAN-MS can effectively improve the insufficient sensitivity and precision of the classification model.

INDEX TERMS Generative adversarial networks, data augmentation, electrocardiogram signals, ECG generation.

I. INTRODUCTION

Electrocardiogram (ECG) is a physiological signal widely used in heart health monitoring. It contains a large amount of pathological information related to heart activity that is very important for the detection and diagnosis of heart diseases. The amount of ECG data is usually very large since the duration of symptoms is one day or longer. It is very time-consuming and difficult for cardiologists to diagnose arrhythmia only by manually analyzing a large amount of ECG data, as some important information may be overlooked [1]. Therefore, some computer-aided diagnosis (CAD) algorithms have been proposed and used to automatically identify arrhythmias in the past few decades. Traditional ECG automatic classification methods [2]–[10] have achieved acceptable ECG classification performance, but they also have disadvantages. Conventional methods need to design a feature extractor, and then input the extracted features into the appropriate classifier for ECG classification.The feature extractors are usually designed and selected through multiple trials or experience, and the extracted features are difficult to express all the features of the signal. The methods use hand-crafted features for classification, which tend to generate more false positives, which may lead to misdiagnosis [11].

Deep learning methods have the capability to automatically learn features from input signals and can extract more abstract and advanced features [12]. Therefore, recently many

The associate editor coordinating the review of this manuscript and approving it for publication was Bin Liu.

studies use deep learning methods to design ECG classifiers [13]–[20] and achieved better performance than traditional methods. Although deep learning methods have shown promising results in ECG classification, they are seriously affected by data imbalance. The imbalance of the datasets will make training models (especially deep learning models) tend to be biased towards classes containing a large number of samples [21]. For ECG datasets, most abnormal heart rhythms are extremely rare, which limits the establishment of a machine learning model. There are many methods to overcome data imbalance in computer vision tasks, such as random flipping, random translation, random cropping, affine transformation, etc. But these methods are not suitable for the augmentation of the ECG signal, because the ECG signal is one-dimensional and their morphology has physiological significance. Oversampling is the most commonly used ECG augmentation method, but it is easy to cause overfitting [22]. There are also some ECG generation methods based on mathematical models [23]–[27]. The methods based on traditional mathematical models can generate very realistic heartbeats, but these heartbeats are too standard and similar. In addition, each type of heartbeat corresponds to a set of equations. In order to generate different types of heartbeats, the equations must be modified. This modification process is usually complicated and troublesome [28].

Generative adversarial networks (GANs) were proposed by Goodfellow *et al.* [29] in 2014 as a tool for generating data. The trained GAN maps the latent input noise to the real data distribution to generate synthetic data. GANs and improved GANs algorithms are widely used in image generation, style conversion, speech signals generation, and other fields in recent years, and have excellent performance. GAN is also widely used in medical image processing. Madani *et al.* [30] used GAN to generate chest X-ray images and finally used it for the classification of cardiovascular abnormalities. Tim *et al.* [31] proposed a GAN-based method to reduce noise in CT images to ensure the quality of the CT image. Han *et al.* [32] used GAN to generate multi-sequence brain magnetic resonance (MR) images and successfully generated 128×128 brain MR images avoiding artifacts. Quan *et al.* [33] used GAN to quickly and accurately reconstruct CS-MRI (Compressed Sensing Magnetic Resonance Imaging).

However, there are very few studies [21], [28], [34], [35] on ECG signal generation using GAN, and their quality assessment of the generated ECG signal is not comprehensive. The above GAN-based ECG generation studies have different evaluation indicators. Some studies only show the accuracy of the classifier after data augmentation, but do not evaluate the quality of the production data. Some studies only compare the distribution of the generated data with the original data and did not evaluate the fidelity or diversity of the generated data. In addition, these studies do not consider the problem of mode collapse. GANs tend to generate only a subset of the types found in the training data. This tendency will cause the model to generate only a certain type or a few types of samples, which is what we call mode collapse. Mode collapse is collectively referred to the lack of diversity of the generated samples [36]. Each type of ECG signal usually presents more than one forms, but after testing, the above model can only generate samples of one or several shapes. In other words, although the generated samples are realistic, the distribution of generated samples is different from the original samples.

In this work, we further explored GAN-based methods to solve ECG data imbalance. The proposed method is based on ProGAN [37]. Different from the traditional GAN structure, the structure of the generator and discriminator progressively grow with the training process. The model does not directly generate an ECG signal with a target length but starts with a low-resolution ECG signal with a length of 32 sample points, and then generates a higher resolution ECG signal by progressively increases the convolutional layer used for feature extraction in the model structure. In addition, the across-minibatch standard deviation [36] is applied, which alleviates the problem of model collapse, making the generated ECG distribution closer to the real data distribution.

The contributions of this paper are:

1) ProEGAN-MS, an ECG signal generation model, is proposed, which can generate ECG signals with higher fidelity and the training process is more stable. In addition, it can generate ECG signals of any desired resolution without changing the model.
2) The problem of mode collapse in ECG signal generation by the GAN method is solved. There is better diversity in the generated ECG data. The distribution of ECG signals generated by the proposed model is more similar to the real data distribution. There will not be a situation where the generated signal of one type of heartbeat has only one form, or the proportion of one or more forms is very high.
3) Some of the GAN-based ECG generation methods are reproduced and a comprehensive comparison of various methods under the unified evaluation indicators are conducted. In addition, in order to show the diversity of the generated ECG data more intuitively, we manually checked the diversity of the data and calculated the statistical results.

The structure of the paper is as follows. Section II reviews some studies on ECG signal generation based on GAN. Section III expatiates on the proposed method including network architecture and loss function in detail. Section IV describes the experimental platform, database, evaluation index, evaluation method. Model training details and results are also described in Section IV. Section V concludes the contributions of this paper.

II. RELATED WORKS

In this section, we review some ECG augmentation methods based on GAN proposed in the past.

There are few studies on using GAN to generate ECG signals. Wang *et al.* [34] were the first to use GAN to

generate ECG signals to solve data imbalance. They construct a data augmentation model based on an auxiliary classifier generative adversarial network (ACGAN) [38]. The statistical indicators Euclidean Distance (ED), Pearson Correlation Coefficient(PCC), and Kullback Leibler Divergence (KLD) were used to compare the distribution of the generated signal and the original signal. Finally, classification experiments on MIT-BIH [39] datasets was conducted to show that the augmentation model can improve the classification accuracy. Shaker *et al.* [21] used generative adversarial networks (GANs) to restore the balance of the dataset. The generator network in their model consists of four fully connected layers; the discriminator network consists of five fully connected layers. This model is used to balance the fifteen types of samples in the MIT-BIH dataset. Wulan *et al.* [28] proposed two GAN-based ECG generation models the SpectroGAN model and WaveletGAN model. SpectroGAN and WaveletGAN used short-term Fourier transform (STFT) and stationary wavelet transform (SWT) respectively to convert a 1-D time-domain ECG signal into a 2-D time-frequency representation as input for the deep convolutional generative adversarial networks (DCGAN) [40]. The GAN-Train and GAN-Test score obtained by the SVM classifier is used to evaluate the fidelity and diversity of the generated ECG samples. Golany *et al.* [35] also used DCGAN for ECG data augmentation to improve classification model performance. But they only showed the accuracy of the classification model after augmentation and did not evaluate the quality of the generated ECG signal.

III. METHODOLOGY

In this chapter, we describe in detail the structure of the proposed ECG generation model and the loss function used.

A. GENERATIVE ADVERSARIAL NETWORKS

Generative adversarial networks (GANs) were proposed by Goodfellow *et al.* in 2014 as a tool for generating data. GANs consist of a generator network and a discriminator network. The generator captures the distribution of real data and produces data similar to the original data. The discriminator is used to determine whether the input data is real (original data) or fake (generated data). The training process of this framework is defined as a game between two competing networks. The discriminator learns to discriminate between the real and fake samples and the generator learns to fool the discriminator. The loss function of the original GAN is defined as follows:

$$\min_G \max_D \mathbb{E}_{x\sim\mathbb{P}_r}[\log(D(x))] + \mathbb{E}_{\tilde{x}\sim\mathbb{P}_g}[\log(1-D(\tilde{x}))] \quad (1)$$

where $\tilde{x} = G(z)$, z represents the input of the generator and follows noise distribution p, such as Gaussian distribution. G and D represent generator and discriminator respectively. $\mathbb{E}_r$ is the distribution of real data and $\mathbb{E}_g$ is the distribution of the data generated by the model.

In order to make the training more stable, we use WGAN-GP [41] loss function instead of the classic GAN loss function (formula 1). The loss function of the WGAN-GP is:

$$L = \mathbb{E}_{\tilde{x}\sim\mathbb{P}_g}[D(\tilde{x})] - \mathbb{E}_{x\sim\mathbb{P}_r}[D(x)] + \lambda\mathbb{E}_{\hat{x}\sim\mathbb{P}_{\hat{x}}}\left[\left(\left\|\nabla_{\hat{x}}D(\hat{x})\right\|_2 - 1\right)^2\right] \quad (2)$$

where $\lambda\mathbb{E}_{\hat{x}\sim\mathbb{P}_{\hat{x}}}\left[\left(\left\|\nabla_{\hat{x}}D(\hat{x})\right\|_2 - 1\right)^2\right]$ is the gradient penalty, which is an alternative way to enforce the Lipschitz constraint. $\hat{x}$ is random samples obtained by random interpolation sampling between x and $\tilde{x}$. $\mathbb{P}_{\hat{x}}$ is distribution satisfied by $\hat{x}$. This method interpolates each real data with a generated one, with a random weight. The penalty is calculated from the gradient of the input with respect to the discriminator's score for that interpolated data. If the gradient of the discriminator is far away from 1, the penalty term will prevent the training from wandering to unstable regions. In addition, an additional drift penalty [36] was added to the discriminator loss. It can prevent the discriminator output from deviating too far from zero. The final loss function we used is as follows:

$$\begin{aligned} Loss = {} & \mathbb{E}_{\tilde{x}\sim\mathbb{P}_g}[D(\tilde{x})] - \mathbb{E}_{x\sim\mathbb{P}_r}[D(x)] \\ & + \lambda\mathbb{E}_{\hat{x}\sim\mathbb{P}_{\hat{x}}}\left[\left(\left\|\nabla_{\hat{x}}D(\hat{x})\right\|_2 - 1\right)^2\right] \\ & + \epsilon_{drift}\mathbb{E}_{x\sim\mathbb{P}_r}\left[D(x)^2\right] \end{aligned} \quad (3)$$

where ϵ_{drift} is a weight of drift penalty.

B. MODEL STRUCTURE

The structure of the ECG generation model we proposed is shown in Fig. 1. The model does not directly generate an ECG signal with a length of 256 sample pointsbut starts with a low-resolution ECG signal with a length of 32 sample points, and then generates a higher resolution ECG signal by progressively increases the convolutional layer used for feature extraction in the model structure. When a new convolution structure is added to the model, there will be a fade-in process. This is to avoid the influence of the sudden increase of the gradient on the parameters of the trained low-resolution convolutional layer. The advantage of progressive training is that it can achieve stable synthesis ECG signals in high-resolution, and it can also greatly speed up training. Because the generation of shorter signals was more stable in the early stage, and it is easier to supplement details progressively than to generate all details directly. The training process is divided into four stages. Each stage generates ECG signals with lengths of 32, 64, 128, and 256 sample points. The real samples used for training in each stage will also be downsampled to the corresponding length. Taking the training process of type N as an example, the signals generated at each stage are shown in Fig. 2. (for the convenience of comparison, we upsampled the signals with a length less than 256 to 256). Table 1 describes the structure of the generator and discriminator in detail. In particular, we merely referred to ProGAN to design our own generator and discriminator rather than directly use the whole structure of ProGAN.

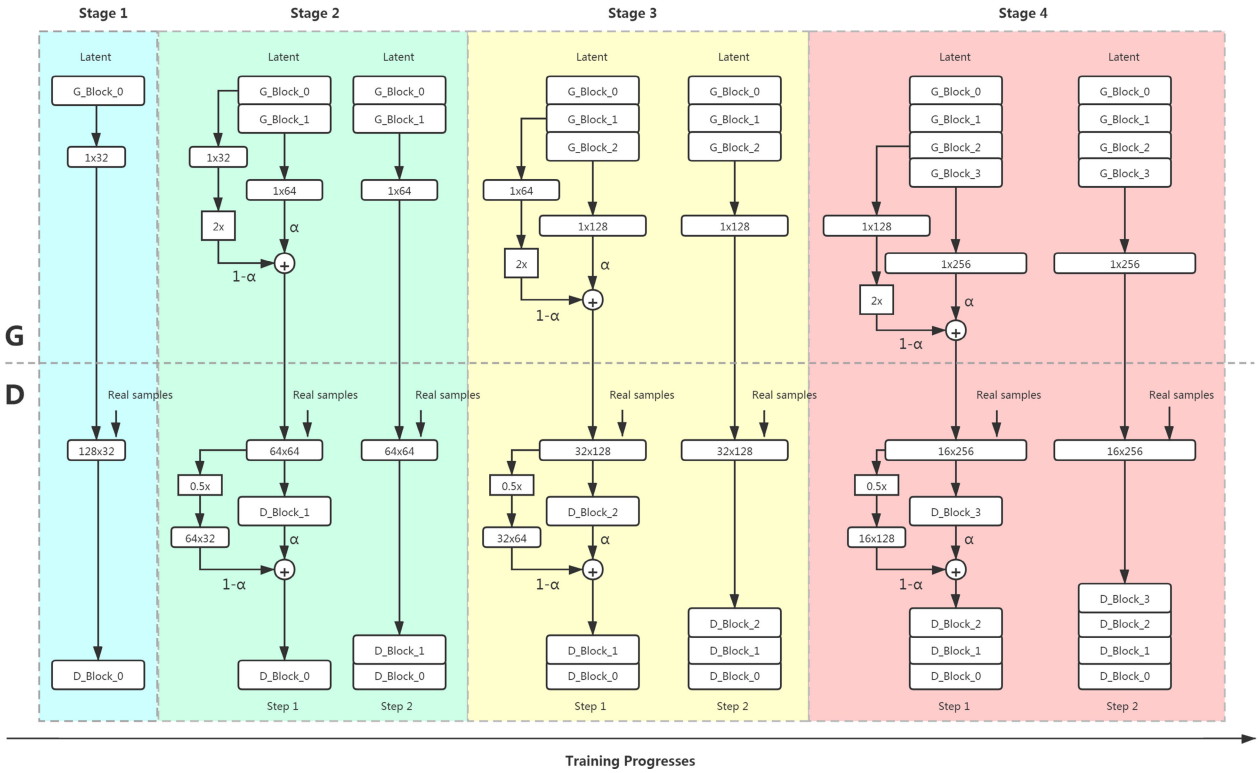
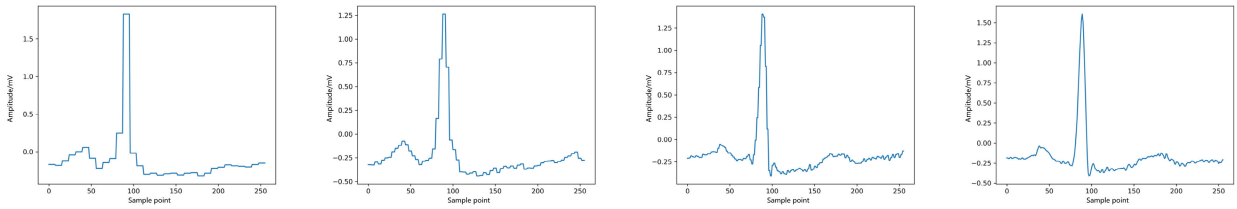


FIGURE 1. The upper part represents the generator, and the lower part represents the discriminator. The training process is divided into four stages, starting with a length of 32-sampling points ECG. After the training of the previous stage is completed, the new convolution structure will be added to the current network. Except for the first stage, each stage of training is divided into two steps. In step 1, the newly high-resolution block will fade-in the network with a weight (linearly increasing from 0 to 1). Then in step 2, the entire network parameters will be optimized. The model finally generates a heartbeat signal with a length of 256 sampling points. Here, 2x and 0.5x represent doubling or halving the signal length. In the generator, 1×32, 1×64, 1×128, and 1×256 represent the process of converting the feature vector into an ECG signal of corresponding length. In the discriminator, 128×32, 64×64, etc. represent the process of converting ECG signals of different lengths into feature vectors. G_block and D_ block represent convolution structure described in TABLE 1.



(a) Length: 32 (b) Length: 64 (c) Length: 128 (d) Length: 256

FIGURE 2. Generated samples for each stage. To facilitate comparison, the generated samples with a length of less than 256 are all up-sampled to 256.

C. INCREASING DATA DIVERSITY USING MINIBATCH STANDARD DEVIATION

GANs tend to generate only a subset of the types found in the training data. This tendency will cause the model to generate only a certain type or a few types of samples, which is what we call mode collapse. This phenomenon is collectively referred to as the lack of diversity of generated samples [42]. Each type of ECG signal usually presents more than one forms, but the model with mode collapse can only generate samples of one or several shapes. In order to improve the diversity of the generated data and make the distribution of the generated data more similar to the original data, we add the across-minibatch standard deviation [36] as an extra feature channel to the last layer of the discriminator. This gives the discriminator a clue to detect generated batches if the generator only learns how to generate one type of sample, which forces the generator to learn a distribution of samples to generate. The calculation process of across-minibatch standard deviation is as follows:

$$F_{B,L} = f_{B,L} - \frac{1}{B}\sum_{i=1}^{B} f_{i,L} \tag{4}$$

which was used to calculate the difference from the average of each location, where $f_{B,L}$ are feature of a batch. B and L represent the number of features in the batch and the length of a feature, respectively.

$$S_L = \sqrt{\frac{\sum_{i=1}^{B}\left(F_{B,L} - \overline{F_{B,L}}\right)^2}{B}} \tag{5}$$

TABLE 1. Structure of the generator and discriminator.

Generator	Act	Output shape
Latent vector	-	128x1
Conv 1x41	LReLU	128x32
Conv1x3	LReLU	128x32
UpSampling	-	128x64
Conv1x3	LReLU	64x64
Conv1x3	LReLU	64x64
UpSampling	-	64x128
Conv1x3	LReLU	32x128
Conv1x3	LReLU	32x128
UpSampling	-	32x256
Conv1x3	LReLU	16x256
Conv1x3	LReLU	16x256
Conv1x1	Linear	1x256

Discriminator	Act	Output shape
Input sample	-	1x256
Conv 1x3	LReLU	16x256
Conv 1x3	LReLU	16x256
DownSampling	-	16x128
Conv 1x3	LReLU	32x128
Conv 1x3	LReLU	32x128
DownSampling	-	32x64
Conv 1x3	LReLU	64x64
Conv 1x3	LReLU	64x64
DownSampling	-	64x32
Conv 1x3	LReLU	128x32
Conv 1x32	LReLU	128x32
Conv 1x	LReLU	128x1
Fully-connect	Linear	1x1

Equation 5 was used to calculate the standard deviation of each feature over the minibatch.

$$\mathrm{M} = \frac{\sum_{i=1}^{L} s_i}{L} \tag{6}$$

Equation 6 was used to calculate the mean standard deviation across the entire feature map. Finally, these calculations are repeated and the obtained results are concatenated to all signal locations over the minibatch.

IV. EXPERIMENTAL AND ANALYSIS

In this section, we evaluate the effectiveness of the proposed method. We first introduce the database and experimental platform and then describe the details of the model training. GAN-train and GAN-test scores were used to evaluate the fidelity and diversity of the generated data. Statistical indices of index Euclidean Distance (ED),Dynamic Time Warping (DTW), Pearson Correlation Coefficient(PCC), and Kullback Leibler Divergence (KLD) were used to compare the distribution of the generated data and the distribution of the original data. Because the evaluation indexes of previous studies are not uniform, we have reproduced some methods and compared them with our methods under the above evaluation criteria. In addition, in order to visually show the improvement of generated data distribution. We performed artificial diversity statistics on ECG samples generated by different methods and showed the statistical results. Finally, we established a neural network classification model for arrhythmia and used the evaluation index of the classification task to evaluate the improvement of the classification model performance by ProEGAN-MS.

A. EXPERIMENTAL PLATFORM

The experiments were executed on a computer with Windows 10 operating system, Intel (R) Core (TM) i7-8700MCPU, GeForce RTX 2080 SUPER, 16GB RAM, and by programming in Python. All the models are run over GPU using the PyTorch deep learning framework.

B. MIT-BIH DATABASE AND SIGNAL PREPROCESSING

In this study, we use the MIT-BIH arrhythmia database [38]. The database contains 48 half-hour records, each record contains two 30-minute ECG leads signals with a sampling frequency of 360 Hz. Four of the records (102, 104, 107, 217) are generated by pacemakers and are excluded in this study. The database contains 16 types of heartbeats. Five majority types: normal beat (N), left bundle branch block beat (L), right bundle branch block beat (R), premature ventricular contraction (V), and atrial premature contraction (A) are involved in this study.

There are two mainstream data category selection methods in papers that use MIT-BIH data: one uses five categories data of NOR, LBBB, RBBB, PVC and APC (which is the category selected in this article), and the other uses five categories data of N, S, V, F and Q. (where N includes NOR, LBBB, RBBB, AE and NE; S includes APC, AP, BAP and NP; V includes PVC, VE and VF, F only includes VFN, and Q includes FPN and UN.) We chose the first data selection method for the same reason as many other studies because these five types of heartbeat data are abundant in each category. Four records (102, 104, 107, and 217) of MIT-BIH arrhythmia database are generated by pacemakers and only contain paced heartbeats. They were excluded from this study, as has been done in many studies using the five data categories, because it does not contain the five major heartbeats we need. If the second data classification rule is adopted, the number of heartbeats in the heartbeat categories other than the above five types is very small. For example, the number of AE and UN is only 16 and 15. The number of NP, VE, and AP is also less than 100. Lack of data will seriously affect the performance of machine learning algorithm. Although we can also use the second data category selection rule, the neural network model trained with such a small amount of data is not reliable. There is no special standard in this experiment for the selected data, which is commonly used in ECG research, so reproducibility can be guaranteed.

Lead MLII ECG data were sliced to single heart-beats which were used for sample generation in this study. On the basis of the position of R-peak provided by Annotation Files, we intercept 0.24s (88 points) forward and 0.47s (168 points) backward (256 points in total) to form a single heartbeat sample because a heartbeat usually lasts 0.6 to 0.8 seconds. After segmentation, some samples with a length of fewer than 256 points are excluded. In addition, samples containing two R-peak are also excluded. The separated ECG samples are shown in Fig. 3. The number of heartbeats in each type

TABLE 2. Overview of the data used in this work.

Type	Number of samples		
	Original samples	Generated samples	Augmented samples
N	70879	-	10000
L	7908	2000	10000
R	7223	2000	10000
V	6282	2000	10000
A	2168	2000	10000

It is obvious that there is an imbalance in the number of five beats in the original data.

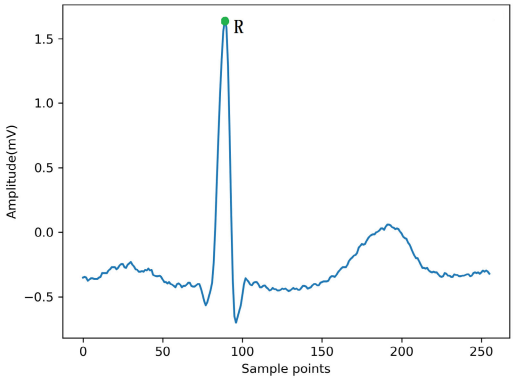


FIGURE 3. A heartbeat sample of type N.

is shown in Table 2. Original samples represent the data of original data, generated samples represent the data used for diversity and fidelity evaluation and distribution evaluation, and augmented samples represent the augmentation data used in classification experiments. In order to get the generated data closer to the original sample, but also to make the GAN model more robust, we did not perform any filtering on ECG samples used for generator training.

C. MODEL TRAINING AND GENERATED ECG

1) TRAINING DETAILS OF THE MODEL

The input z of the generator is a noise vector of length 128, and z follows a normal distribution with a mean of 0 and a variance of 1. The ProEGAN-MS is trained by Adam [43] with $\beta_1 = 0.0$, $\beta_2 = 0.99$. The learning rates of generator and discriminator are 0.001. The hyperparameters λ and ϵ_{drift} of the gradient penalty and drift penalty are 10 and 0.001, respectively. Since the number of N categories in the original data is sufficient, we only generate data for abnormal rhythms (L, R, V, and A). The models of type L, R, and V are trained for 30 epochs each stage, the model of type A is trained for 50 epochs each stage. We call the model after removing the across-minibatch standard deviation feature in ProEGAN-MS as ProEGAN. We did the same training on the ProEGAN model and compared the following parts to verify the effect of across-minibatch standard deviation on improving the diversity of the generated data. All experiments are performed five times and the middle value is taken.

2) GENERATED ECG SAMPLES

Fig. 4 (a) to (d) shows four types of original data, and (e) to (h) shows the generated data. The goal of the GAN model is to generate data similar to the original data but not in the training set. As we can see from Fig. 4, the generated sample is very similar to the original sample but slightly different. Different types of ECG samples have different characteristics, ProEGAN-MS can automatically learn and generate samples with high fidelity. In addition, even the same type of heartbeat, there will be several different main forms, as shown in Fig. 5, for example, there are four main forms in the type of LBBB. ProEGAN-MS can also generate realistic samples of the same type of heartbeat in different forms.

D. EVALUATION OF FIDELITY AND DIVERSITY

1) EVALUATION INDICATORS

It is important to evaluate the generated data in two aspects: the fidelity and the diversity of the data. The fidelity of the generated data is not enough, which clearly indicates that the model is performing poorly. If the fidelity of the generated data is enough but the diversity is insufficient, it indicates that the model has mode collapse. This phenomenon is also what we do not want. There are several indicators, such as Inception Score (IS) [44], Frĺechet Inception Distance (FID) [45], Mode Score [46], GAN-train, GAN-test [47], and so on, are widely used to evaluate the performance of a GAN. Although IS and FID are widely used, they are not suitable for evaluating the ECG data generation model because they rely too much on Inception V3. In this study, we utilize GAN-train and GAN-test to evaluate the fidelity and the diversity of the ECG data generation model.

2) EVALUATION METHODS

In order to get the two scores GAN-train and GAN-test, we need a basic multi-category heartbeat classifier. We removed the feature normalized part of the SVM heartbeat classifier described in the article [28] as the basic classifier used in this study. The classification method is as follows: Firstly, Daubechies wavelet 6 is utilized to decompose each heartbeat at 4 levels. Then, approximation coefficients at level 4 as a feature vector were input to SVM. Finally, the trained SVM is used as a classifier to classify the heartbeats based on the feature vector.

For getting the GAN-train score, SVM needs to be trained on the generated ECG sample and tested on the original ECG sample. If the SVM can classify the original data correctly, it means that the generated data is similar to the original data. Under the premise that the generated data is realistic, the higher the GAN-train score, the better the diversity of the generated data. For getting the GAN-train score, SVM needs to be trained on the original ECG sample and tested on the generated ECG sample. The higher the GAN-test score, the better the fidelity of the generated data.

3) RESULTS

We utilize GAN-train and GAN-test to evaluate the diversity and the fidelity of the ECG data generated by different models. The results are shown in Table 3. The higher the GAN-train score, the better the diversity of the generated

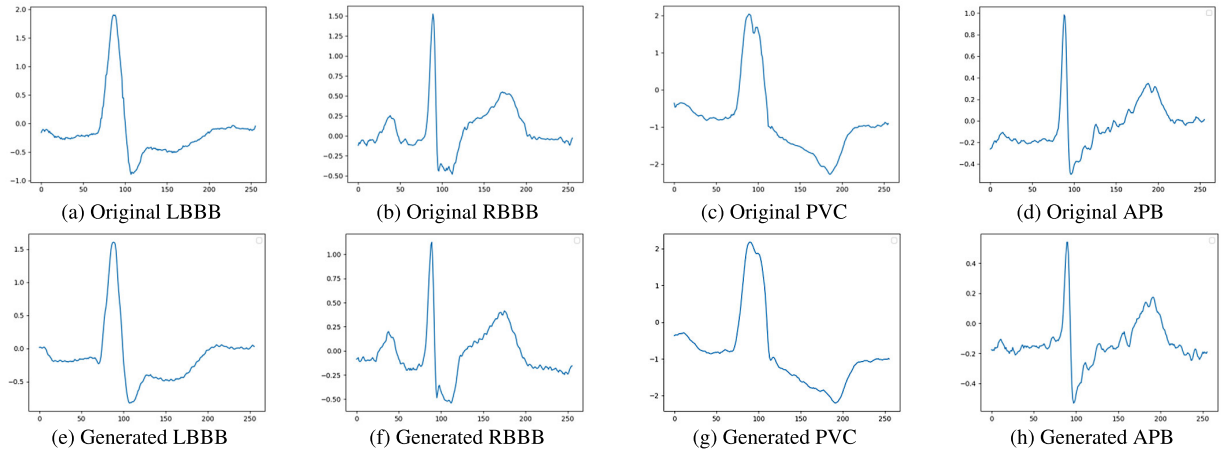


FIGURE 4. Different types of original heartbeat samples and generated heartbeat samples. (a) to (d) show the original samples of type LBBB(L), type RBBB(R), type PVC(V), and type APB(A). (e) to (h) show the generated samples of type LBBB(L), type RBBB(R), type PVC(V), and type APB(A). The generated heartbeat has the characteristics of the original heartbeat and is slightly different from the original sample.

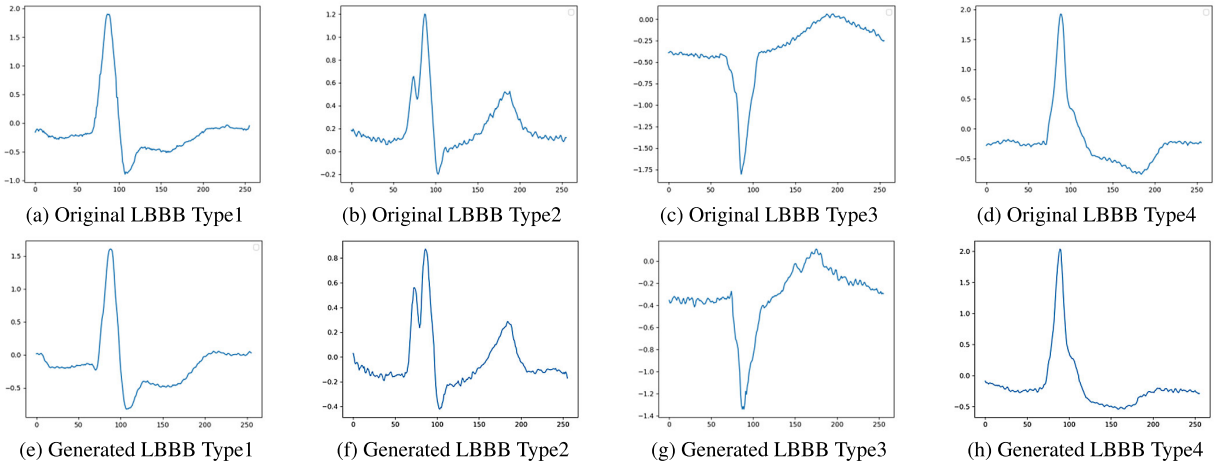


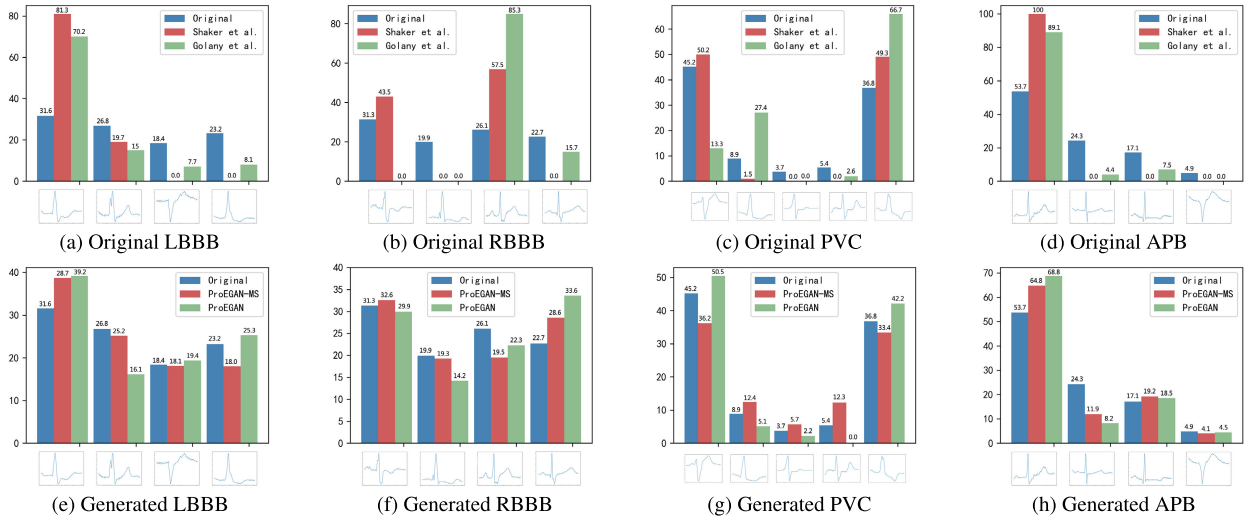
FIGURE 5. Four main forms of LBBB in MIT-BIH arrhythmia database. (a) to (d) show the four main forms of the LBBB-type heartbeat in the original data. (e) to (h) show the four main forms of the LBBB-type heartbeat in the original data. The proposed augmentation model can effectively generate different forms of heartbeats of the same type.

data. The higher the GAN-test score, the better the fidelity of the generated data. We used GAN-train and GAN-test score from real data as benchmarks. We can see that the GAN-train score of the model proposed by Shaker *et al.* [21] and Golany *et al.* [35] are about 30 and 40 lower than the benchmark, respectively. The GAN-train score of ProEGAN is 91.73, which is much higher than the scores of other methods and is also the closest to the benchmark. Compared with ProEGAN, ProEGAN-MS has further improved diversity and fidelity due to the addition of MSS features, with GAN-train score of 92.64. The GAN-Test score of the model proposed by Shaker *et al.* [21] is 98.04 or even higher than the benchmark. Through the analysis of the generated samples, this is due to the poor diversity of generated samples. The model only generates a few types of samples with extremely obvious characteristics, so it is easier to identify. The GAN-Test scores of ProEGAN and ProEGAN-MS were 87.31 and 89.84, respectively. The result shows that compared with other methods, ProEGAN and ProEGAN-MS can not only ensure high fidelity, but also ensure the diversity of the generated data. However, other methods can guarantee the fidelity of the generated data, but the diversity is poor. This result is also reflected in subsequent experiments.

TABLE 3. Evaluation of generated data fidelity and diversity.

Model	GAN-Train (%)	GAN-Test (%)
Benchmark	97.96	97.96
Shaker et al.[21]	68.39	**98.04**
Golany et al.[35]	58.17	87.54
ProEGAN	91.73	87.31
ProEGAN-MS	**92.64**	89.84

In order to visually show the improvement of generated data diversity and distribution, we divided the main forms of heartbeats in each type in the original data and made manual statistics on their proportions. From Fig. 6 (a) to (d), we can see that the ECG data generated by the models proposed by Shaker and Golany has a particularly different distribution from the original data. In addition, there are some forms that cannot be generated in each type of heartbeat. From Fig. 6 (e) to (h), we can see that the ECG data generated by ProEGAN and ProEGAN-MS has a highly similar distribution to the original data, and there is almost no ECG form that cannot be generated.



(a) Original LBBB (b) Original RBBB (c) Original PVC (d) Original APB

(e) Generated LBBB (f) Generated RBBB (g) Generated PVC (h) Generated APB

FIGURE 6. The forms distribution statistics of the original data and the generated data. (a) to (d) show the distribution of the form of the four type LBBB(L), RBBB(R), PVC(V), and APB(A) in the original data and the data generated by the compared method. From the red and green bars in (a) to (d), it is obvious that the form of generated samples have obvious biases, and some forms cannot be generated. The distribution of the data generated by these two models is very different from the original data distribution. (e) to (h) show the distribution of the form of the four types of the original data and the data generated by our method. The blue, red and green bars indicate the proportions of forms in the original data and the data generated by the two methods we proposed, respectively. The form distribution of the generated data is highly consistent with that of the original data.

TABLE 4. Quantitative evaluation of the generated data.

Model	Indicators	ED	DTW	PCC	KLD
Template	CIs	8.00	7.09	1.0	0.011
Shaker et al. [21]	FIs	6.62	6.18	0.971	0.050
Golany et al. [35]	FIs	5.56	5.09	0.925	0.167
ProEGAN	FIs	**7.43**	**6.54**	**0.999**	0.023
ProEGAN-MS	FIs	7.38	6.44	**0.999**	**0.018**

E. EVALUATION OF SAMPLE DISTRIBUTION

1) EVALUATION INDICATORS

We evaluate the data distribution difference between the generated data and the original data by some indicators based on statistical characteristics, such as Euclidean Distance (ED), Dynamic Time Warping (DTW), Pearson Correlation Coefficient (PCC), and Kullback Leibler Divergence (KLD). ED quantifies the distance between the generated data and the original data, DTW reflects dissimilarity between two different time series signals [48], PCC reflects the linear correlation between distributions, and KLD measures the closeness of two distributions.

2) EVALUATION METHODS

If there is no comparison standard, it is difficult to measure the quality of generated data by indicators based on statistics. Therefore, we use the method proposed by Wang *et al.* [34] to conduct a comparative evaluation of the generated data. The calculation steps are as follows:

Step 1: Calculating the average of the original data as the template;

Step 2: Calculating averaged ED, DTW, PCC, and KLD between the template and original data for indicators of CIs;

Step 3: Calculating averaged ED, DTW, PCC, and KLD between the template and original data for indicators of FIs;

Step 4: Comparing FIs with CIs.

TABLE 5. Quantitative evaluation of the generated data for each category.

Category	Model	Indicators	ED	DTW	PCC	KLD
L	Template	CIs	5.66	5.10	1.0	0.006
	ProEGAN	FIs	5.58	5.02	0.999	0.033
	ProEGAN-MS	FIs	5.51	5.00	0.999	0.027
R	Template	CIs	7.09	6.83	1.0	0.009
	ProEGAN	FIs	6.85	6.65	0.999	0.032
	ProEGAN-MS	FIs	6.91	6.66	0.999	0.034
V	Template	CIs	10.47	9.57	1.0	0.028
	ProEGAN	FIs	10.14	9.25	0.999	0.074
	ProEGAN-MS	FIs	9.95	9.08	0.999	0.065
A	Template	CIs	4.26	3.14	1.0	0.003
	ProEGAN	FIs	3.89	2.95	0.999	0.022
	ProEGAN-MS	FIs	3.69	2.62	0.995	0.021

TABLE 6. F_1 score for each type.

Type	F1 Mathematical Definition
F_{11}	$2*N_{11}/(N_{1x}+N_{x1})$
F_{12}	$2*N_{22}/(N_{2x}+N_{x2})$
F_{13}	$2*N_{33}/(N_{3x}+N_{x3})$
F_{14}	$2*N_{44}/(N_{4x}+N_{x4})$
F_{15}	$2*N_{55}/(N_{5x}+N_{x5})$

The smaller the difference between CIs and FIs, the higher the distribution similarity between the generated data and the original data.

3) RESULT

The results are listed in Table 4 and Table 5. Table 4 shows the distribution difference between the original data and the generated data by different methods. We can see that the ED and DTW of the data generated by ProEGAN have the smallest difference to the template, and the PCC and KLD indicators of the data generated by ProEGAN-MS have the smallest gap with the template. This shows that the distribution of the data generated by our proposed method is most similar to the original data. This result is consistent with the result of the sample distribution of manual statistics in the

TABLE 7. Overview of the data used in classification experiment.

Type	Number of train set				Number of test set	Data length
	Original	Golany et al.	Shaker et al.	ProEGAN-MS		
N	63791	10000	10000	10000	7088	256
L	7117	10000	10000	10000	791	256
R	6501	10000	10000	10000	722	256
V	5654	10000	10000	10000	628	256
A	1951	10000	10000	10000	217	256

TABLE 8. F_1 score with different data augmentation strategies on the simple network.

Score	Augmentation strategies			
	Original	Golany et al.	Shaker et al.	ProEGAN-MS
F_{11}	0.963	0.981	0.988	0.991
F_{12}	0.980	0.982	0.994	0.994
F_{13}	0.933	0.988	0.994	0.989
F_{14}	0.874	0.945	0.980	0.975
F_{15}	0.009	0.590	0.706	0.837
F_1	0.752	0.897	0.932	**0.957**

TABLE 9. Comparison of the effects of different GAN-based methods on the performance of classifiers.

Author	Augmentation methods	Performance		
		ACC	SEN	PRE
Template	-	96.08%	75.44%	85.88%
Golany et al.	DCGAN	97.10%	90.10%	89.52%
Shaker et al.	GAN	98.09%	94.86%	91.68%
Our method	ProEGAN	97.84%	94.15%	88.33%
	ProEGAN-MS	**98.55%**	**99.36%**	**92.89%**

previous part. For the ECG data generated by ProEGAN and ProEGAN-MS, Table 5 shows the distribution differences between the generated data and the original data in each category.

F. EVALUATION OF PERFORMANCE

1) EVALUATION INDICATORS

We established neural network models for arrhythmia classification, and use evaluation indexes of the classification task to illustrate the improvement of the performance of the model by ProEGAN-MS. The evaluation indexes used in this part include accuracy (ACC), sensitive (SEN), precision (PRE) and F_1 score. ACC shows the proportion of correct prediction examples to the total data. SEN indicates the proportion of all positive examples that are divided into pairs, which measures the ability of the classifier to find positive examples in an all-round way. PRE indicates the proportion of positive examples that are actually positive examples, measuring the query accuracy of the classifier to positive examples. Their calculation method is shown in formula 7-9.

$$ACC = \frac{TN + TP}{TN + TP + FN + FP} \times 100\% \tag{7}$$

$$SEN = \frac{TP}{TP + FN} \times 100\% \tag{8}$$

$$PRE = \frac{TP}{TP + FP} \times 100\% \tag{9}$$

where TP means the number of true positive samples, TN means the number of true negative samples, FP means the number of false positive samples, and FN means the number of false negative samples.

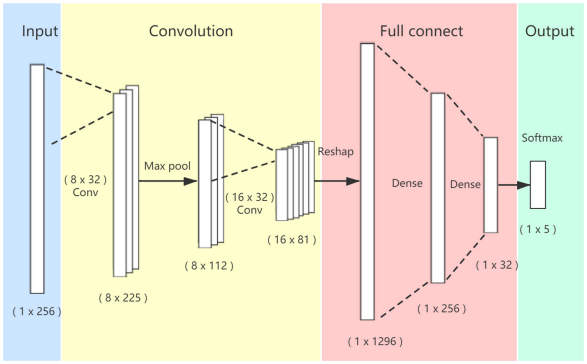


FIGURE 7. The simple neural network consisting of two convolutional layers and two fully connected layers.

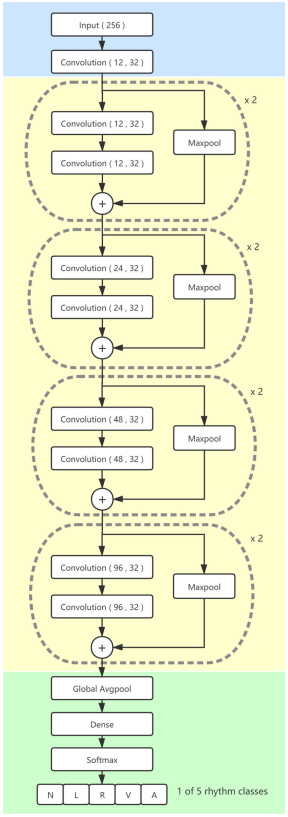
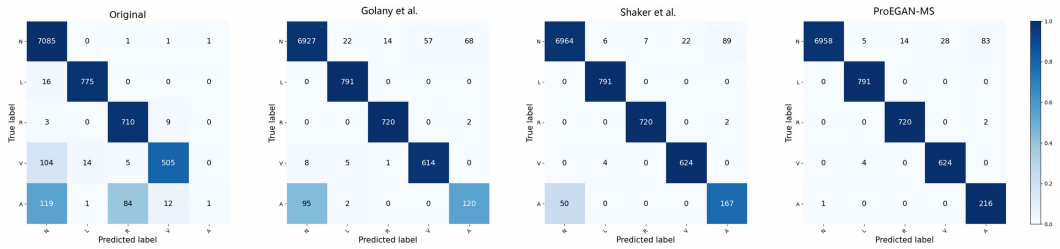


FIGURE 8. The network consists of 19 layers, including 17 convolution layers, an average pooling layer and a fully connected layer. Residual blocks are used to ensure that the depth of the network increases without reducing performance. A large convolution kernel with a length of 32 is used to ensure that the features of ECG signals are extracted.

F_1 is the harmonic average of sensitive and precision, with maximum of 1 and a minimum of 0, the calculation method is shown in formula (10). F_1 score for each type are obtained based on the confusion matrix, as shown in Table 6, in which



(a) Original (b) Golany et al. (c) Shaker et al. (d) ProEGAN-MS

FIGURE 9. Confusion matrix comparison.

TABLE 10. Comparison of our proposed recognition model with related work.

Author	Approach	Performance		
		ACC	SEN	PRE
Yang et al. [49]	SVM	97.77%	-	86.34%
Yazdanian et al.[50]	SVM	96.97%	94.16%	96.97%
Acharya et al.[51]	9-layer CNN	94.03%	96.71%	97.86%
Kiranyaz et al. [52]	5-layer CNN	99.00%	93.90%	90.60%
Wang et al. [53]	7-layer CNN	98.64%	87.91%	93.94%
Our method	19-layer Resnet	99.52%	99.10%	96.00%
	19-layer Resnet (augmentation by ProEGAN)	**99.59%**	97.68%	**99.89%**
	19-layer Resnet (augmentation by ProEGAN-MS)	99.58%	**99.88%**	97.90%

N_{1x} is the sum of a row, N_{x1} is the sum of a column.

$$F_1 = \frac{2 \times PRE \times SEN}{PRE + SEN} \tag{10}$$

2) EVALUATION METHODS

We established two neural network classification models for arrhythmia, one of which is a simple neural network consisting of two convolutional layers and two fully connected layers, as shown in Fig. 7. The other network is a 19-layer residual network, and its structure is shown in Fig. 8. The purpose of designing a simple network is to clearly reflect the performance difference between different GAN-based ECG augmentation methods. ResNet19 was established to compare the classification performance with other excellent ECG recognition algorithms. For simple network, we set the L2 regularization coefficient to 0.0005, the learning rate to 0.0001, and training for 150 epochs. For 19-layer residual network, we set the L2 regularization coefficient to 0.001, the learning rate to 0.01, and training for 50 epochs. The data used in this part is shown in Table 7. We randomly select 10% of the original sample as the test set, and the remaining data with less than 10,000 are augmented to 10,000 by three GAN-based ECG augmentation methods.

3) RESULT ON SIMPLE NETWORK

In order to illustrate the advantages of ProEGAN-MS, we compared the F_1 scores of three GAN-based ECG augmentation methods. The purpose of using data augmentation is to improve the recognition performance of the classifier on minor classes. From the first column of Table 8, we can see that due to the imbalance of the sample, the neural network classifier trained with the original data can hardly recognize the minor category A heartbeat, with a F_{15} score of 0.009. Augmentation by ProEGAN-MS, F_{15} increased from 0.009 to 0.837, which shows that the performance of small category recognition has been significantly improved. This score is 0.247 and 0.131 higher than the other two GAN-based methods, respectively. In addition, the overall F_1 score also increased from 0.752 to 0.957. This result is also higher than the two compared methods. It can be seen from the comparison of the confusion matrix in Fig. 9. that the number of correct recognition of type A by the classifier has increased significantly after data augmentation.

As shown in Table 9, we listed ACC, SEN, PRE of the simple ECG recognition network as a reference for classification performance. The accuracy of the model is acceptable, but due to the imbalance of data, sensitive and precision are relatively low. After data augmentation by ProEGAN-MS, the ACC, SEN and PRE of the classifier are increased to 98.55%, 99.36% and 92.89% respectively, which is the highest score among all comparison methods. This shows that using ProEGAN-MS to augment the ECG signal is more helpful to improve the performance of classifier than using other GAN-based methods to augment the ECG signal.

4) RESULT ON RESIDUAL NETWORK

As shown in Table 10, we listed ACC, SEN, PRE of different related works as a reference for classification performance. The accuracy of different classification methods is very high, but due to the imbalance of data, sensitive and precision will be relatively low. Before data augmentation, although the ACC and SEN of the 19-layer convolutional network we built were the highest among all methods, the PRE was still not satisfactory. After data augmentation by ProEGAN-MS, our classifier not only improved further in ACC and SEN, but also increased by 1% in PRE, making it the highest 97.90% of all methods. This shows that augmented data by ProEGAN-MS can effectively improve insufficient sensitive and precision. It also illustrates the role of ProEGAN-MS in improving data imbalance.

V. CONCLUSION

In this paper, we proposed a progressive ECG generation model called ProEGAN-MS to solve the problem of ECG data imbalance. The model does not directly generate an ECG signal with a target length but starts with a low-resolution ECG signal with a length of 32 sampling points, and then generates a higher resolution ECG signal by progressively increases the convolutional layer used for feature extraction in the model structure. In addition, the across-minibatch standard deviation is used to enrich the diversity of generated data. In the experimental part, we comprehensively evaluated the proposed method from three parts: the fidelity and diversity of generated data, the distribution of generated data and the performance improvement of classification model. The experimental results show that the ECG data generated by ProEGAN-MS has outstanding performance under the above three evaluation standards, which shows that our proposed method can effectively solve the problem of ECG data imbalance and promote the development of ECG classification technology.

ACKNOWLEDGMENT

(Haixu Yang and Jihong Liu are co-first authors.)

REFERENCES

[1] S. L. Oh, E. Y. K. Ng, R. S. Tan, and U. R. Acharya, "Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats," *Comput. Biol. Med.*, vol. 102, pp. 278–287, Nov. 2018.

[2] S. Sahoo, B. Kanungo, S. Behera, and S. Sabut, "Multiresolution wavelet transform based feature extraction and ECG classification to detect cardiac abnormalities," *Measurement*, vol. 108, pp. 55–66, Oct. 2017.

[3] T. Li and M. Zhou, "ECG classification using wavelet packet entropy and random forests," *Entropy*, vol. 18, no. 8, p. 285, Aug. 2016.

[4] F. A. Elhaj, N. Salim, A. R. Harris, T. T. Swee, and T. Ahmed, "Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals," *Comput. Methods Programs Biomed.*, vol. 127, pp. 52–63, Apr. 2016.

[5] H. Li, D. Yuan, Y. Wang, D. Cui, and L. Cao, "Arrhythmia classification based on multi-domain feature extraction for an ECG recognition system," *Sensors*, vol. 16, no. 10, p. 1744, Oct. 2016.

[6] R. J. Martis, U. R. Acharya, K. M. Mandana, A. K. Ray, and C. Chakraborty, "Application of principal component analysis to ECG signals for automated diagnosis of cardiac health," *Expert Syst. Appl.*, vol. 39, no. 14, pp. 11792–11800, Oct. 2012.

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

[8] M. H. Vafaie, M. Ataei, and H. R. Koofigar, "Heart diseases prediction based on ECG signals' classification using a genetic-fuzzy system and dynamical model of ECG signals," *Biomed. Signal Process. Control*, vol. 14, pp. 291–296, Nov. 2014.

[9] M. C. De Jongh, C. C. T. Haar, S. Man, R. W. Treskes, A. C. Maan, M. J. Schalij, and C. A. Swenne, "Intra-individual ECG changes over 25 years: How long can elective ECGs be used as reference for acute ischemia detection?" *J. Electrocardiol.*, vol. 48, no. 4, pp. 490–497, Jul. 2015.

[10] B. Liu, J. Liu, G. Wang, K. Huang, F. Li, Y. Zheng, Y. Luo, and F. Zhou, "A novel electrocardiogram parameterization algorithm and its application in myocardial infarction detection," *Comput. Biol. Med.*, vol. 61, pp. 178–184, Jun. 2015.

[11] J. G. Lee, S. Jun, Y.-W. Cho, H. Lee, G. B. Kim, J. B. Seo, and N. Kim, "Deep learning in medical imaging: General overview," *Korean J. Radiol.*, vol. 18, no. 4, pp. 570–584, 2017.

[12] Z. Wu, T. Lan, C. Yang, and Z. Nie, "A novel method to detect multiple arrhythmias based on time-frequency analysis and convolutional neural networks," *IEEE Access*, vol. 7, pp. 170820–170830, 2019.

[13] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Med.*, vol. 25, no. 1, pp. 65–69, Jan. 2019.

[14] R. He, Y. Liu, K. Wang, N. Zhao, Y. Yuan, Q. Li, and H. Zhang, "Automatic cardiac arrhythmia classification using combination of deep residual network and bidirectional LSTM," *IEEE Access*, vol. 7, pp. 102119–102135, 2019.

[15] M.-G. Kim and S. B. Pan, "Deep learning based on 1-D ensemble networks using ECG for real-time user recognition," *IEEE Trans. Ind. Informat.*, vol. 15, no. 10, pp. 5656–5663, Oct. 2019.

[16] R. He, K. Wang, N. Zhao, Y. Liu, Y. Yuan, Q. Li, and H. Zhang, "Automatic detection of atrial fibrillation based on continuous wavelet transform and 2D convolutional neural networks," *Frontiers Physiol.*, vol. 9, p. 1206, Aug. 2018.

[17] B. Hou, J. Yang, P. Wang, and R. Yan, "LSTM-based auto-encoder model for ECG arrhythmias classification," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1232–1240, Apr. 2020.

[18] A. H. Kashou, W.-Y. Ko, Z. I. Attia, M. S. Cohen, P. A. Friedman, and P. A. Noseworthy, "A comprehensive artificial intelligence–enabled electrocardiogram interpretation program," *Cardiovascular Digit. Health J.*, vol. 1, no. 2, pp. 62–70, Sep. 2020.

[19] A. S. Eltrass, M. B. Tayel, and A. I. Ammar, "A new automated CNN deep learning approach for identification of ECG congestive heart failure and arrhythmia using constant-Q non-stationary Gabor transform," *Biomed. Signal Process. Control*, vol. 65, Mar. 2021, Art. no. 102326.

[20] J. P. Allam, S. Samantray, and S. Ari, "SpEC: A system for patient specific ECG beat classification using deep residual network," *Biocybern. Biomed. Eng.*, vol. 40, no. 4, pp. 1446–1457, Oct. 2020.

[21] A. M. Shaker, M. Tantawi, H. A. Shedeed, and M. F. Tolba, "Generalization of convolutional neural networks for ECG classification using generative adversarial networks," *IEEE Access*, vol. 8, pp. 35592–35605, 2020.

[22] P. Branco, L. Torgo, and R. P. Ribeiro, "A survey of predictive modeling on imbalanced domains," *ACM Comput. Surv.*, vol. 49, no. 2, pp. 1–50, Nov. 2016.

[23] N. Jafarnia-Dabanloo, D. C. McLernon, H. Zhang, A. Ayatollahi, and V. Johari-Majd, "A modified zeeman model for producing HRV signals and its application to ECG signal generation," *J. Theor. Biol.*, vol. 244, no. 2, pp. 180–189, Jan. 2007.

[24] L. Zheying, L. Shuo, and M. Minjie, "ECG signal generator design based on DFG model," in *Proc. 8th Int. Conf. Signal Process.*, vol. 4, 2006, p. 1.

[25] G. D. Clifford, S. Nemati, and R. Sameni, "An artificial vector model for generating abnormal electrocardiographic rhythms," *Physiol. Meas.*, vol. 31, no. 5, pp. 595–609, May 2010.

[26] E. K. Roonizi and R. Sameni, "Morphological modeling of cardiac signals based on signal decomposition," *Comput. Biol. Med.*, vol. 43, pp. 1453–1461, Oct. 2013.

[27] S. Das and K. Maharatna, "Fractional dynamical model for the generation of ECG like signals from filtered coupled van-der pol oscillators," *Comput. Methods Programs Biomed.*, vol. 112, no. 3, pp. 490–507, Dec. 2013.

[28] N. Wulan, W. Wang, P. Sun, K. Wang, Y. Xia, and H. Zhang, "Generating electrocardiogram signals by deep learning," *Neurocomputing*, vol. 404, pp. 122–136, Sep. 2020.

[29] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 3, 2014, pp. 2672–2680.

[30] A. Madani, M. Moradi, A. Karargyris, and T. Syeda-Mahmood, "Chest X-ray generation and data augmentation for cardiovascular abnormality classification," *Proc. SPIE*, vol. 10574, Mar. 2018, Art. no. 105741M.

[31] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Isgum, "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2536–2545, Dec. 2017.

[32] C. Han, H. Hayashi, L. Rundo, R. Araki, W. Shimoda, S. Muramatsu, Y. Furukawa, G. Mauri, and H. Nakayama, "GAN-based synthetic brain MR image generation," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 734–738.

[33] T. M. Quan, T. Nguyen-Duc, and W.-K. Jeong, "Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1488–1497, Jun. 2018.

[34] P. Wang, B. Hou, S. Shao, and R. Yan, "ECG arrhythmias detection using auxiliary classifier generative adversarial network and residual network," *IEEE Access*, vol. 7, pp. 100910–100922, 2019.

[35] T. Golany, G. Lavee, S. T. Yarden, and K. Radinsky, "Improving ECG classification using generative adversarial networks," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 8, pp. 13280–13285.
[36] Z. Lin, A. Khetan, G. Fanti, and S. Oh, "PacGAN: The power of two samples in generative adversarial networks," presented at the 32nd Int. Conf. Neural Inf. Process. Syst., Montreal, QC, Canada, 2018.
[37] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," 2017, *arXiv:1710.10196*. [Online]. Available: http://arxiv.org/abs/1710.10196
[38] A. Odena, C. Olah, and J. Shlens. (2017). *Conditional Image Synthesis With Auxiliary Classifier GANs*. [Online]. Available: http://proceedings.mlr.press/v70/odena17a.html
[39] R. Mark and G. Moody, "MIT-BIH arrhythmia database directory," Massachusetts Inst. Technol., Cambridge, MA, USA, 1988.
[40] F. Liu, C. Liu, L. Zhao, X. Zhang, X. Wu, X. Xu, Y. Liu, C. Ma, S. Wei, Z. He, J. Li, and E. N. Y. Kwee, "An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection," *J. Med. Imag. Health Informat.*, vol. 8, no. 7, pp. 1368–1373, Sep. 2018.
[41] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: http://arxiv.org/abs/1511.06434
[42] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5767–5777.
[43] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–15.
[44] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 2234–2242.
[45] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6626–6637.
[46] T. Che, Y. Li, A. P. Jacob, Y. Bengio, and W. Li, "Mode regularized generative adversarial networks," 2016, *arXiv:1612.02136*. [Online]. Available: http://arxiv.org/abs/1612.02136
[47] K. Shmelkov, C. Schmid, and K. Alahari, "How good is my GAN?" in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. 11206. Munich, Germany: Springer, 2018, pp. 218–234. [Online]. Available: https://hal.inria.fr/hal-01850447/document and https://hal.inria.fr/hal-01850447/file/ganeval.pdf
[48] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-26, no. 1, pp. 43–49, Feb. 1978.
[49] W. Yang, Y. Si, D. Wang, and B. Guo, "Automatic recognition of arrhythmia based on principal component analysis network and linear support vector machine," *Comput. Biol. Med.*, vol. 101, pp. 22–32, Oct. 2018.
[50] H. Yazdanian, A. Nomani, and M. R. Yazdchi, "Autonomous detection of heartbeats and categorizing them by using support vector machines," in *Proc. 20th Iranian Conf. Biomed. Eng. (ICBME)*, Dec. 2013, pp. 223–227.
[51] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. S. Tan, "A deep convolutional neural network model to classify heartbeats," *Comput. Biol. Med.*, vol. 89, pp. 389–396, Oct. 2017.
[52] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016.
[53] H. Wang, H. Shi, X. Chen, L. Zhao, Y. Huang, and C. Liu, "An improved convolutional neural network based approach for automated heartbeat classification," *J. Med. Syst.*, vol. 44, no. 2, 2020, Art. no. 35.

HAIXU YANG received the B.S. degree from the Taiyuan University of Science and Technology, Shanxi, China, in 2017. He is currently pursuing the M.S. degree with the College of Information Science and Engineering, Northeastern University, China. His research interests include image processing, pattern recognition, and deep learning.

JIHONG LIU received the Ph.D. degree in pattern recognition and intelligent system from Northeastern University, China, in 2003. She is currently an Associate Professor with the College of Information Science and Engineering, Northeastern University. Her research interests include computational cardiology, intelligent information processing, and biomedical signal acquisition.

LVHENG ZHANG received the B.S. degree from the Hubei University of Technology, China, in 2019. He is currently pursuing the M.S. degree with the College of Information Science and Engineering, Northeastern University, China. His current research interests include deep learning in medical images and the application of generation adversarial networks.

YAN LI received the B.S. degree in material forming and control engineering from the Wuhan University of Technology University, China, in 2013. He is currently pursuing the M.S. degree with the College of Information Science and Engineering, Northeastern University, China. His research interests include deep learning, pattern recognition, and signal processing of ECG.

HENGGUI ZHANG received the Ph.D. degree in mathematical cardiology from the University of Leeds, in 1994. He worked as a Postdoctoral Research Fellow with the School of Medicine, Johns Hopkins University, from 1994 to 1995, and the University of Leeds, from 1996 to 2000, where he also a Senior Research Fellow, from 2000 to 2001. In October 2001, he moved to UMIST to take up the lectureship. He also worked with The University of Manchester, as a Lecturer with UMIST, from 2001 to 2004; a Senior Lecturer, from 2004 to 2006; and a Reader, from 2006 to 2009. He is currently the Chair of the Biological Physics Group, School of Physics and Astronomy, The University of Manchester. He is also a Professor of biological physics. He has published more than 400 scientific articles, among them over 200 articles were published in prestigious peer-reviewed journals in his field. Related works have attracted wide public interests, and been covered by many prestigious media, such as BBC. He has been elected as a fellow of world renowned societies as recognition of distinctions.