# ECG Arrhythmias Detection Using Auxiliary Classifier Generative Adversarial Network and Residual Network

**PU WANG**[1], **BORUI HOU**[1], **(Member, IEEE), SIYU SHAO**[1],
**AND RUQIANG YAN**[1,2]**, (Senior Member, IEEE)**

[1]School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China
[2]School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710000, China

Corresponding author: Ruqiang Yan (ruqiang@seu.edu.cn)

**ABSTRACT** This paper aims at proposing an abnormality detection framework for electrocardiogram (ECG) signals, which owns unbalance distribution among different classes and gaining high accuracy in rhythm/morphology abnormalities classification. The proposed framework is composed of two models: data augmentation model and classification model. In this framework, data augmentation model is designed to recast a class-balanced training dataset by generating artificial data of minor class. The outputs of augmentation model are transferred into classification model. The classification model is designed to identify abnormalities accurately after training using both the experimental and generated datasets. Data augmentation model is supported by auxiliary classifier generative adversarial network (ACGAN). We construct Generator and Discriminator of the ACGAN by stacking multiple 1-dimensional convolutional layers with small kernel size. Dropout function and batch normalization are added to prevent gradients vanish and speed up convergence. In order to evaluate the performance of augmentation model, a set of quantitative indicators are introduced to verify the quality of generated ECG signals. We establish classification model based on stacked residual network parallel connected with long short-term memory (LSTM) network. The experimental study is conducted for single heartbeat detection and consecutive heartbeat detection. The results based on standard benchmark, MIT-BIH, and competition database provided by 2018 China physiological signal challenge (CPSC) have verified the proposed framework can achieve high performance in robustness and accuracy for class-imbalanced dataset.

**INDEX TERMS** Electrocardiogram signals, heartbeat arrhythmias detection, auxiliary classifier generative adversarial network, data augmentation, long short-term memory network, residual network.

## I. INTRODUCTION

In recent years, the incidence of cardiovascular diseases (CVDs) has exploded due to multiple factors such as population ageing, chronic cardiovascular disease and increasing living pressures. With high mortality, heart disease has become a major threat to human life [1], [2]. Therefore, the task about monitoring and preventing it in advance is quite important. One intrinsic presentation of heart diseases is heart's rhythm/morphology abnormal activity, and electrocardiogram (ECG) which records such electrical activity of heart in visible way provides abundant information for abnormality

diagnosis. Thus, detecting abnormalities of ECG signals has often been applied in clinical CVDs diagnosis [3], [4].

The majority of extant models for rhythm/morphology abnormalities detection [5], [6] are comprised by four independent steps: 1) ECG signals acquisition; 2) data processing; 3) features extraction; 4) identification. Each procedure can introduce errors and lead to inaccurate detection. Recently, deep learning-based approach which ensembles feature extraction and classification into one process has been successfully applied for ECG signal analysis to overcome this challenge. Deep learning-based ECG signal processing framework has powerful feature extraction ability which can learn deep features from given signals and optimize model automatically to achieve high accuracy in classification.

---

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

Typically, there are different deep learning architectures, such as Deep Belief Network (DBN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM), and so on. For example, Acharya *et al.* [7] proposed a deep CNN structure to automatically identify 5 different categories of ECGs arrhythmia. Salloum and Kuo [8] applied RNN to build an effective CVDs identification system. Tan *et al.* [9] proposed a stacked LSTM network with CNN to classify normal versus CVD ECGs, which achieve high accuracy.

For deep learning-based approaches, each model is constructed by stacked multi-hidden layers. Meanwhile, each hidden layer contains considerable parameters. Therefore, the number of parameters in deep learning model which needs to be trained is enormous. To achieve a high accuracy, the deep learning model needs adequate training by plentiful balanced training data. However, in practice, the occurrence rate of different abnormalities is diverse. It often leads to an imbalanced distribution between minor and major cases of multiple abnormalities in collected ECG signals. Such class-imbalance prevents the deep network to learn how to identify minority class.

Several researches have been conducted to deal with this data imbalanced issue. Rajesh and Dhuli [10] implemented re-sampling techniques for imbalanced ECG beats classification. Ukil *et al.* [11] combined over-sampling methods with semi-supervised feed-back controlled approach to achieve an intelligent class augmentation algorithm. Gogna *et al.* [12] introduced semi-supervised deep learning approach, stacked auto-encoder (SAE) model for ECG signal reconstruction. However, previous work related to imbalanced data augmentation mainly based on sample from original signal, and the improvements are limited. Generative Adversarial Network (GAN) which was firstly introduced to generate artificial convincing image samples [13] provides a new approach for imbalanced data learning. Madani *et al.* [14] investigated its capability for learning from both labeled and unlabeled medical images, and realized data-efficient cardiac disease diagnosis. Chen *et al.* [15] modified original GAN structure to boost risk prediction performance with limited electronic health records (EHRs). Shao *et al.* [16] utilized auxiliary classifier GAN (ACGAN) with label condition information for imbalanced mechanical signal augmentation. However, limited work has been conducted in ECG signal augmentation using GAN structure.

Therefore, this work further investigate the GAN's potential in dealing with imbalanced ECG signal issue, and constructs a deep-learning based ECG rhythm/morphology abnormalities detection framework. The framework is composed of two models: data augmentation model and classification model. The data augmentation model is supported by ACGAN which is used to generate artificial ECG signal and form a new class-balanced training dataset. It also proposes a quantitative assessment criterion to evaluate the quality of generated signal. The classification model is based on stacked residual block and LSTM network. It is used for ECG abnormalities classification after data augmentation. The novelties and contributions of this work mainly includes: 1) the ACGAN is firstly applied for ECG signal generation to solve the data imbalanced issue; 2) a set of evaluation indicators are proposed to assess the performance of data augmentation model; 3) the proposed detection framework can achieve high performance in robustness, and accuracy for class-imbalanced dataset.

The reminder of the paper is organized as follows: Section II introduces the basic theories of ACGANs, residual network, and LSTM network. Then the proposed detection framework is illustrated in section III with detailed structure of both ACGANs, and classification model. In section IV, experiments are conducted to test the performance of proposed detection framework for single heartbeat detection and consecutive heartbeat detection. After that, results and some discussions are presented. Finally, conclusion is drawn in section V.

## II. BASIC THEORY

In this paper, three deep-learning approaches are utilized to form the detection model. The basic theories of GAN and ACGAN, which are used for data augmentation, are briefly introduced in part A. Residual network and LSTM, which are applied in classification model, are represented in part B and C, respectively.

### A. AUXILIARY CLASSIFIER GENERATIVE ADVERSARIAL NETWORKS

As shown in Fig.1(a), GANs are composed of two parts: the Generator G and the Discriminator D. The principle of GANs is to generate fake data which make the Discriminator hard to classify fake or real, that is the Generator is trained to generate fake data which can fool the Discriminator. The input of the Generator is random noise vector *z*, and the Generator force *z* to model the distribution of real data vector *x* and then output fake data. The input of the Discriminator is from both real data and generated fake data, and the Discriminator
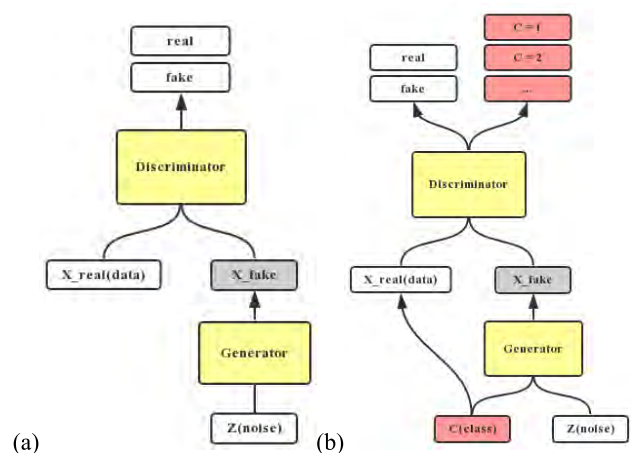


**FIGURE 1.** Typical structure of (a) regular GAN, (b) ACGAN.

is trained to classify real and fake data correctly. In a word, the training process of GANs is a two-player minimax game. The Generator is trained to generate fake data by making the Discriminator recognize fake data as real; on the contrary, the Discriminator is trained to classify real and fake data as accurate as it can. The target of GANs training can be described with value function $V(D, G)$ followed by (1) [13]:

$$
\min_{G} \max_{D} V(D, G) = E_{x \sim P_{data}(x)} \left[ \log D(x) \right]
$$
$$
+ E_{z \sim P_z(z)} \left[ \log (1 - D(G(z))) \right] \quad (1)
$$

where, $D(x)$ represents the probability that $x$ from the real data distribution $P_{data}$ rather than the Generator $P_g$; $G(z)$ represents a mapping from noise vector to generated vector. In training procedure, the Generator is trained after the Discriminator has been well trained. We update parameters in Discriminator to maximum the function $V(G, D)$, while updating parameters in Generator to minimum the function $V(G, D)$.

ACGAN is a type of variant of the GAN. As shown in Fig.1 (b), it introduces category information as auxiliary term to improve the performance of the Generator [17]. And the output of the Discriminator includes categories as well. Therefore, the $G(z)$ existed in (1) transform to $G(c, z)$ in ACGAN, where $c$ represents corresponding class label. The loss functions of ACGAN contains two parts: $L_S$ records the probability of the correct source (the same as GAN) as shown in (2), and $L_C$ records the probability of correct label as shown in (3).

$$
L_S = E_{x \sim P_{data}(x)} \left[ \log D(x) \right] + E_{z \sim P_z(z)} \left[ \log (1 - D(G(z))) \right]
$$
$$
(2)
$$
$$
L_C = E_{c \sim P_{data}(c)} \left[ \log D(c) \right] + E_{c \sim P_z(c)} \left[ \log (1 - D(G(c))) \right]
$$
$$
(3)
$$

Since the Discriminator should be able to correctly classify fake and real data with accurate corresponding label, the Discriminator is trained to maximum $L_S + L_C$. Meanwhile, we hope the output of the Generator can fool the Discriminator to recognize it as real one with correctly corresponding label. Therefore, the Generator is trained to maximum $L_C - L_S$.

### B. RESIDUAL NETWORK

After successful implementation in image classification [18], [19], deep convolutional neural networks have been extended to various fields, including physiological signal analysis. Evidence [20], [21] reveals that networks with deep structure have better performance. For example, outstanding results [20], [21] on ImageNet challenging [22] all exploit "very deep" [20] models. But [19] also refers that with more layers to stack, the phenomenon called vanishing/exploding gradients more likely to occur during model training [23]. This problem can hamper the model converge to correct direction. To overcome such problem, [24] proposed residual block to reduce gradient vanish or explode, as shown in Fig.2.
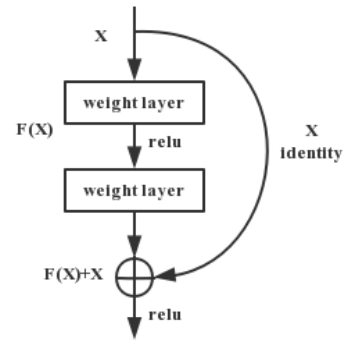


**FIGURE 2.** Structure of residual block [24].

The process of deep model training is actually to optimize parameters in stacked layer to fit one desired mapping $H(x)$. As illustrated in [24], when the optimization of $H(x)$ is hard to conducted, an alternative residual mapping: $F(x) = H(x) - x$ can be established firstly, and then the desired $H(x)$ can be recast by adding $F(x)$ and $x$. The shortcut connection which skips one or more layers in Fig.2 simply performs identity mapping: $x$, and its output is added to the output of stacked layers. Such operation does not introduce extra parameters and complexity. Moreover, it resolves the vanishing gradient problem by preventing partial derivative to zero in chain rule. Besides that, residual block accelerates the convergence speed for shallow network.

### C. LONG SHORT-TERM MEMORY NETWORK

LSTM is derived from RNN, which has capability of learning long-term dependencies. It is consisted of an input layer, memory units and an output layer. Its memory unit has a three-gate structure named input, forget and output gate [25]. A typical structure of a LSTM unit is illustrated in Fig.3. The input gate learns what information is stored in the memory unit. The forget gate is used to learn how much information to be retained or forgotten by generating decision vectors ranged in [0, 1]. Output gate learns when stored information can be used.



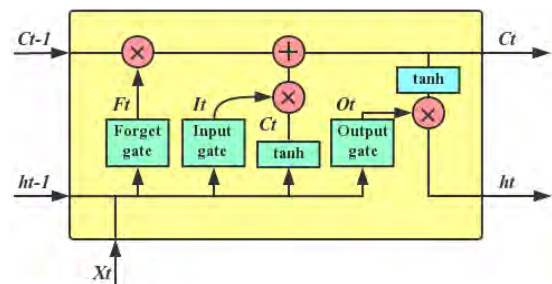**FIGURE 3.** Typical structure of a LSTM unit.

### III. SYSTEM FRAMEWORK

The framework proposed in this study contains two parts: data augmentation model and classification model. The original training data are firstly transferred into data augmentation
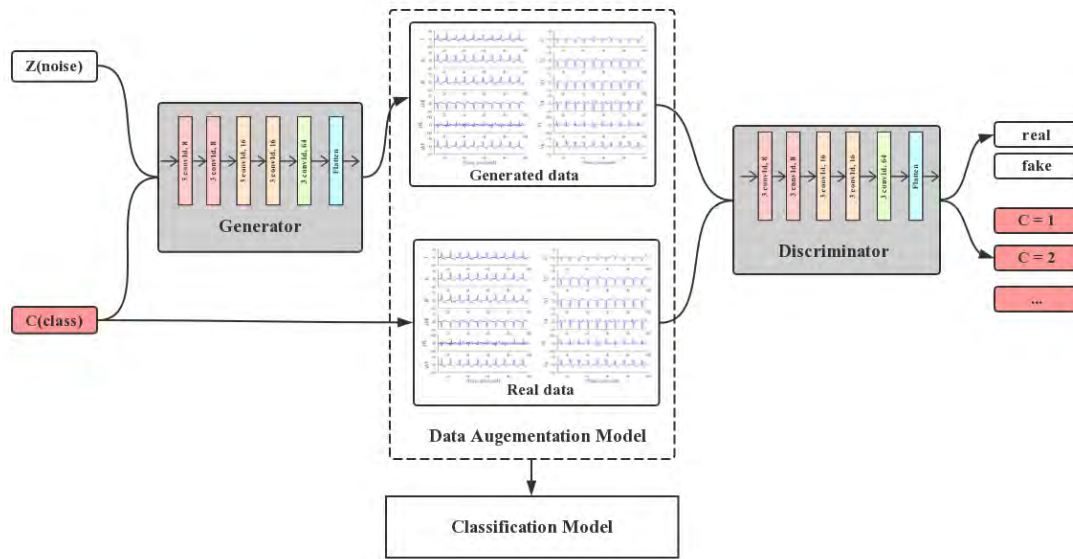
**FIGURE 4.** Proposed ECG arrhythmias detection framework with data augmentation model and classification model.

model to achieve a class-balanced distribution. Then the generated training data is used to train the classification model. Data augmentation model is supported by ACGAN, and classification model is based on residual network and LSTM.

As shown in Fig.4, the procedure of ECG arrhythmias detection proposed in this study consists of four major steps. The random noise $z$ and corresponding label $c$ is put into Generator (in data augmentation model) first to output the generated data. Then, the generated data are mixed with the real data together, and put into Discriminator (in data augmentation model) to train parameters in Discriminator. After that, with the well trained Discriminator, high quality generated data can be obtained in Generator. Finally, a balanced dataset can be formed to train the classification model.

Detailed structures of Generator and Discriminator are discussed in part A. Meanwhile a set of statistical indicators are introduced to evaluate the signal generated by ACGAN. Then the frame of classification model is presented in part B.

## A. ACGAN BASED DATA AUGMENTATION MODEL

### 1) MODEL STRUCTURE

Generator is designed to establish a mapping from latent space*(z, c)* to artificial ECG signal, where $z$ is normal distribution noise and $c$ is random corresponding label. Discriminator is designed to train Generator towards better performance. Since ECG signal from each individual possess different morphology, to better learn hierarchical features of input signal, 14 1-dimensional (1D) convolutional layers including 2 up-sampling operations are utilized with small kernel size to construct the Generator, and 16 1D convolutional layers are applied to form the Discriminator.*Sigmoid* function and *softmax* function are adopted as the output layers of Discriminator to generate predicted sample source and specific label respectively. Batch normalization and dropout are added to

prevent overfitting. The specific structure of ACGAN is listed in Table 1.

**TABLE 1.** Structure of data augmentation model.

| Layer | Generator | Discriminator |
|---|---|---|
| 0 | Input (latent space *(z, c)*) | Input (generated and original signal) |
| 1 | up sampling | Cov3-16 |
| 2 | Cov3-16 | Cov3-16 |
| 3 | Cov3-16 | dropout |
| 4 | batch normalization | Cov3-16 |
| 5 | Cov3-16 | Cov3-16 |
| 6 | Cov3-16 | dropout |
| 7 | batch normalization | Cov3-16 |
| 8 | Cov3-32 | Cov3-16 |
| 9 | Cov3-32 | dropout |
| 10 | batch normalization | Cov3-16 |
| 11 | up sampling | Cov3-16 |
| 12 | Cov3-32 | dropout |
| 13 | Cov3-32 | Cov3-16 |
| 14 | batch normalization | Cov3-16 |
| 15 | Cov3-64 | dropout |
| 16 | Cov3-64 | Cov3-32 |
| 17 | batch normalization | Cov3-32 |
| 18 | Cov3-64 | dropout |
| 19 | Cov3-64 | Cov3-32 |
| 20 | Cov3-1 | Cov3-32 |
| 21 | | dropout |
| 22 | | Cov5-32 |
| 23 | | Cov5-32 |

For model training, based on loss function mentioned in section II, parameters are updated iteratively using ADAM optimizer with learning rate 0.0001 for Generator and 0.0002 for Discriminator. The procedure can be divided into 3 steps during each training epoch:

a): Generating artificial data through Generator by inputting random noise and corresponding labels.

b): Mixing artificial data with real data and transferring them into Discriminator. Parameters in Discriminator are able to update according inputted data and label.

c): After well training Discriminator, fixing parameters in Discriminator into constant status and updating parameters in Generator.

Through adequate iterations, losses of Generator and Discriminator are forced to achieve a balance status called Nash Equilibrium.

### 2) MODEL EVALUATION

The purpose of applying ACGAN architecture is to generate convincing ECG signal for data augmentation, and then transfer them into classification model. Therefore, it is important to evaluate the similarity between the generated signal and the original signal. However, the evaluation criterion is still an open issue. For image generation task, it can be conducted by visual evaluation or statistical measures, such as peak signal to noise ratio (PSNR) and structural similarity (SSIM). For one-dimensional time series, Euclidean distance (ED), Pearson correlation coefficient (PCC) and Kullback–Leibler (K–L) divergence are introduced as evaluation indicators [16].

Considering ECG signal is periodic time series which is composed by several fixed waveforms, including P-wave, QRS complex, and T-wave. The generated signal can be evaluated both by vision and statistical indicators.

In this study, ED, PCC, and K-L divergence are chosen as quantitative indicators. ED represents the distance between the generated signal and original one, PCC measures the linear correlation between the two distributions, and K-L divergence evaluates the difference between the two signals. For each category, specific evaluation procedure can be performed as follows:

a) Calculating average original signal as the template signal;

b) Calculating the ED, PCC, and K-L divergence between the template signal and original data, and calculating the average ED, PCC, and K-L divergence as the compared indicators CIs;

c) Calculating the ED, PCC, and K-L divergence between the template signal and generated data, and calculating the average ED, PCC, and K-L divergence as the final indicators FIs;

d) Comparing FIs with CIs, the smaller difference between FIs and CIs represents high similarity.

### B. CLASSIFICATION MODEL

In classification procedure, multiple stacked 1D convolutional layers with several residual blocks are used to extract deep features of ECG signals. Considering ECG signals are time-series signals, we combine LSTM network with residual network to achieve a better performance in features learning.

The input of the model is parallel processed by 1D-convolutional residual network and LSTM respectively as shown in Fig.5. To achieve concatenate operation of two outputs, global average pooling is added after residual network. The pooling operation translates output of residual network to 1D vector. Then it is concatenated with features extracted by LSTM, and connected to three fully-connected layers. *Softmax* function is adopted as final output layer to determine the class of input ECG segment. The detailed structure of each neural network is illustrated in Table 2.

**TABLE 2.** Structure of classification model.

| Layer | Type | Kernel(filter) size | Stride |
|-------|------|---------------------|--------|
| 0 | Input | - | - |
| 1 | 1D Convolution | (32,3) | 2 |
| 2 | 1D Convolution | (64,3) | 2 |
| 3 | 1D Convolution | (128,3) | 2 |
| 4 | 1D Convolution | (128,3) | 2 |
| 5 | 1D Maxpooling | 3 | 2 |
| 6 | Residual connection | (128,1) | 2 |
| 7 | 1D Convolution | (256,3) | 2 |
| 8 | 1D Convolution | (256,3) | 2 |
| 9 | 1D Maxpooling | 3 | 2 |
| 10 | Residual connection | (256,1) | 2 |
| 11 | 1D Convolution | (512,3) | 2 |
| 12 | 1D Convolution | (512,3) | 2 |
| 13 | 1D Maxpooling | 3 | 2 |
| 14 | Residual connection | (512,1) | 2 |
| | Repeat 6 times | | |
| 34 | 1D Convolution | (728,3) | 2 |
| 35 | 1D Convolution | (1024,3) | 2 |
| 36 | Global Average Pooling | - | - |
| 2 | LSTM | Unit(1024) | - |
| 37 | Concatenate with layer 16 | - | - |
| 38 | Fully-connected(drop out) | - | - |
| 39 | Fully-connected(drop out) | - | - |
| 40 | Fully-connected (softmax) | - | - |

## IV. EXPERIMENT AND RESULTS

In this study, we test the proposed detection model's performance in both single heartbeat detection and consecutive heartbeats detection. For single-beat detection, the standard benchmark: MIT-BIH database is introduced. Comparative experiments are also carried out to compare the classification results with existing methods. Then, the competition database provided by 2018 China Physiological Signal Challenge (CPSC) is applied for consecutive-beats detection. We also compare the classification results with the Top Three Results in the competition.

The proposed model is trained on a workstation with Intel Core i7-7700, CPU 3.6 GHz, RAM 8 GB, and GPU NVIDIA GeForce GTX 1060 8GB.

### A. MIT-BIH DATABASE

MIT-BIH arrhythmia database is developed by Massachusetts Institute of Technology (MIT) [26]. It contains 48 half-hour recordings sampled at 360Hz. Four of them (recording 102, 104, 107, and 217) are generated by pacemaker, and are excluded in this study. MIT also provided Annotation Files with this database. They recorded the location of R-peak
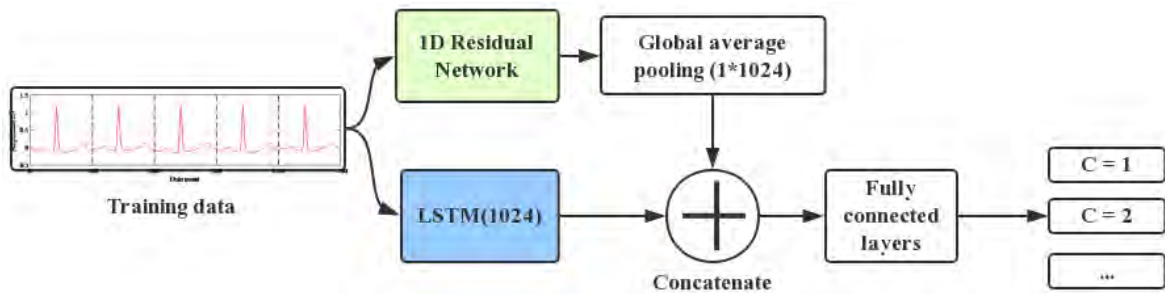
**FIGURE 5.** Structure of proposed classification model.

and type of the each heartbeat. The arrhythmia types of this database are up to 16 including the normal type (N). Five majority types including N, left bundle branch block (L), right bundle branch block (R), premature ventricular contraction (V), and atrial premature contraction (A) obtained from modified limb lead II are used to evaluate the proposed detection model.

As shown in Fig.6, the typical ECG signal contains QRS complex, P-wave, and T-wave. Among them, R-peak of QRS complex divides a heartbeat into two parts. Therefore, the location of R-peak can help to segment consecutive ECG signal into single heartbeat segmentation. Based on the R-peak location provided by Annotation Files, since one heartbeat usually last 0.6s to 0.8s, 0.24s (88 points) offset before R-peak and 0.44s (156 points) offset after R-peak are adopted in this study as single heartbeat duration.
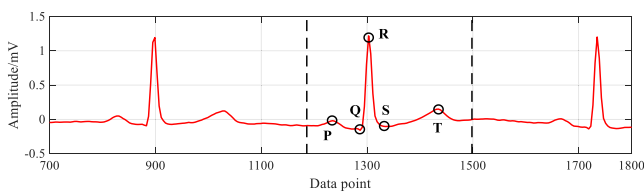


**FIGURE 6.** Typical consecutive ECG signal.

After signal segmentation, total of 99244 single heartbeat segments are used for this work. The number of samples within each type is listed in Table 3. Daubechies wavelet 8 is utilized to remove noise and baseline drift. For each type, 90% of samples are formed as database DS1-1 which is used for classification model training, and the rest 10% are formed as database DS1-2 for testing. Among the DS1-1, 2000 samples of it are formed as the database DS1-3 which is used for ACGAN training, and 200 samples of it are formed as the database DS1-4 for ACGAN testing. Table 3 records the number and data length of each dataset samples.

### 1) DATA AUGMENTATION

As shown in Table 3, the sample distribution of each type is unbalanced, the number of type N accounts for 75%, while the minor type A only represents 3%. Therefore, database DS1-3 is firstly utilized to train ACGAN model for data augmentation.

**TABLE 3.** Data profile of MIT-BIH before data augmentation.

| Type | The number of samples | | | | | |
| | DS1-1 | DS1-2 | DS1-3 | DS1-4 | Data length | Total |
|---|---|---|---|---|---|---|
| N | 67032 | 7448 | 2000 | 200 | 1×244 | 74480 |
| L | 7262 | 807 | 2000 | 200 | 1×244 | 8069 |
| R | 6525 | 725 | 2000 | 200 | 1×244 | 7250 |
| V | 6211 | 690 | 2000 | 200 | 1×244 | 6901 |
| A | 2290 | 254 | 2000 | 200 | 1×244 | 2544 |

The training epoch of ACGAN model is set to be 150, and we generate different numbers of artificial samples for each category to make sure each class contains 10000 recordings for classification model training. Model training process is recorded in Fig. 7. Generation loss in Fig.7 (a) reflects the model's ability in predicting the correct sample source (from real data or generated data); classification accuracy (b) and loss (c) present model's performance in predicting correct sample class.

As shown in Fig.7 (a), Generator and Discriminator go towards Nash Equilibrium after approximately 20 epochs, and there is an obvious improvement in classification accuracy presented in Fig.7 (b).

After data augmentation, the original training dataset DS1-1 is enriched to database DS1-5 as recorded in Table 4.

**TABLE 4.** Data profile of MIT-BIH after data augmentation.

| Type | Before data augmentation | | After data augmentation | |
| | DS1-1 | DS1-3 | DS1-5 | Data length |
|---|---|---|---|---|
| N | 67032 | 2000 | 10000 | 1×244 |
| L | 7262 | 2000 | 10000 | 1×244 |
| R | 6525 | 2000 | 10000 | 1×244 |
| V | 6211 | 2000 | 10000 | 1×244 |
| A | 2290 | 2000 | 10000 | 1×244 |

After adequate training, generated samples of each type is listed in Fig.8 (a) to (e). The red line shown in the figure represents the original signal, and blue line remarks the generated signal. It is clear that although ECG signal within different category has mutiple morphologies in both amplitude and time duration, ACGAN can learn deeper features automatically and generate alike samples.
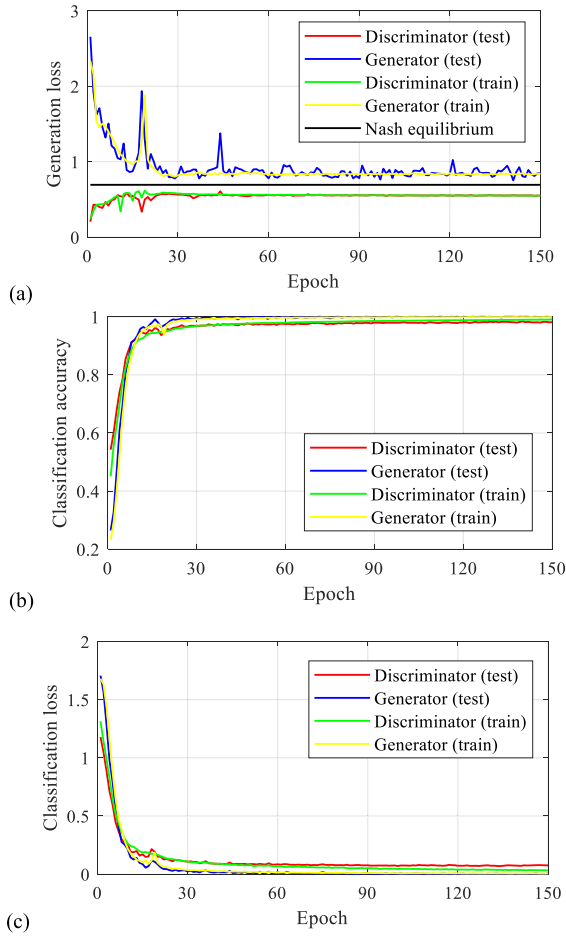
(a)

(b)

(c)

**FIGURE 7.** ACGAN performance in (a) generative loss, (b) classification accuracy, and (c) classification loss.



(a)

(b)

(c)

(d)

(e)

**FIGURE 8.** Generated data and original data of (a) type N; (b) type L; (c) type R; (d) type V; (e) type A.

As illustrated in Section III. Part A, quantitative indicators are conducted to evaluate the similarity between generated samples and original samples. The results are listed in the Table 5. ED and K-L divergence indicates the distance and divergence between the two distributions, and the lower value represents more similar. PCC represents the correlation between two samples, and high value over 0.8 represents strong correlation. As shown in Table 5, the differences between CIs and FIs values are small which indicate the

generated data and original data have similar distribution. Besides that, the PCC values of FIs are all over 0.8 which represents strong similarity among the generated data and original one.

### 2) CLASSIFICATION RESULTS

After data augmentation, generated database DS1-5 is used for classification model training. To test the capability of data augmentation model, the original imbalanced database DS1-1 is applied to training the same classification model

**TABLE 5.** Quantitative evaluation of generated signal.

| Type | | ED | PCC | K-L divergence |
|------|------|-------|--------|----------------|
| N | CIs | 2.237 | 1.000 | 2.482 |
| | FIs | 2.2413 | 0.9106 | 2.633 |
| L | CIs | 4.047 | 1.000 | 3.605 |
| | FIs | 3.817 | 0.8672 | 3.512 |
| R | CIs | 2.926 | 1.000 | 5.203 |
| | FIs | 3.140 | 0.8543 | 5.507 |
| V | CIs | 7.308 | 1.000 | 3.416 |
| | FIs | 6.887 | 0.8332 | 2.981 |
| A | CIs | 2.044 | 1.000 | 2.440 |
| | FIs | 2.323 | 0.8908 | 2.342 |

**TABLE 6.** Classification performance of DS1-1 and DS1-5.

|  | Type | ACC (%) | SEN (%) | SPE (%) |
|---|---|---|---|---|
| | N | 99.36 | 99.79 | 98.06 |
| | L | 99.93 | 99.38 | 99.98 |
| | R | 99.91 | 99.31 | 99.96 |
| **DS1-1** | V | 99.76 | 98.55 | 99.86 |
| | A | 99.59 | 85.83 | 99.95 |
| | Total | 99.71 | 99.27 | 99.82 |
| | N | 99.64 | 99.85 | 98.99 |
| | L | 99.90 | 99.50 | 99.95 |
| | R | 99.98 | 99.86. | 99.99 |
| **DS1-5** | V | 99.77 | 97.83 | 99.91 |
| | A | 99.76 | 93.70 | 99.92 |
| | Total | 99.81 | 99.53 | 99.88 |

as well. The classification performance is evaluated by three measures, sensitivity (SEN), specificity (SPE) and classification accuracy (ACC). SEN is the true positive rate, and it represents the proportion of all positive samples classified to positive. SPE is the true negative rate, and it represents the proportion of all negative samples classified to negative. ACC is the classification accuracy. Higher values of those measures indicate better performance. Mathematically, they are

calculated by (4) to (6)

$$SEN = TP/(TP + FN) \times 100\% \tag{4}$$

$$SPE = TN/(TN + FP) \times 100\% \tag{5}$$

$$ACC = (TN + TP)/(TN + TP + FN + FP) \times 100\% \tag{6}$$

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

The results of classification performance trained by DS1-1 and DS1-5 are listed in Table 6. It shows that the SEN of type A in DS1-1 is relatively low. That means the classification model is not well trained since A class is the minority class, and the model has difficulty in dealing with it. It may wrongly classify other types' samples into type A. After resolving the data imbalance issue by data augmentation, SEN of type A in DS1-5 has a significant improvement, from 85.83% to 93.7% . Besides that, compared with training by imbalance database, the total ACC, SEN, and SPE all have improvement.

For comparison, other related works including supervised, and unsupervised/semi-supervised approaches based on this standard benchmark are listed in Table 7. Considering we

**TABLE 7.** Comparison between the related work and proposed detection model.

| | Data augmentation | Method & Classifier | Training /test dataset | Beat types | SEN (%) | SPE (%) | ACC (%) |
|---|---|---|---|---|---|---|---|
| | No augmentation | Discrete wavelet combined with kernel-independent component analysis & Deep bidirectional LSTM Networks [27] | Imbalanced training dataset/ 7326 samples for testing | 5 | - | - | 99.39 |
| | Random oversampling and under-sampling (ROU) | Higher-Order Statistics (HOS) & AdaBoost ensemble classifier [10] | Balanced training dataset/ 8877 samples for testing | 5 | 95.70 | 98.80 | 98.20 |
| Supervised approaches | Synthetic Minority Over-sampling Technique (SMOTE) | Higher-Order Statistics (HOS) & AdaBoost ensemble classifier [10] | Balanced training dataset/ 8877 samples for testing | 5 | 91.70 | 97.80 | 98.50 |
| | Distribution based data sampling | Higher-Order Statistics (HOS) & AdaBoost ensemble classifier [10] | Balanced training dataset/ 8877 samples for testing | 5 | 97.90 | 99.40 | 99.10 |
| | Generating synthetic data based on the standard deviation and Z-score calculated from the original signals | 9-layer deep CNN [7] | Balanced training dataset/10945 samples for testing | 5 | 96.71 | 91.54 | 94.03 |
| | No augmentation | Unsupervised affinity propagation (AP) clustering algorithm [28] | Using recording #200, 201, 203, 205, 207, 208, 210, 213, 219, 221, and 233 for testing | Average 3.1 | - | - | 91.90 |
| Unsupervised or semi-supervised approaches | No augmentation | Semi-supervised AP clustering algorithm & independent component analysis (ICA) [28] | Using recording #200, 201, 203, 205, 207, 208, 210, 213, 219, 221, and 233 for testing | Average 3.1 | - | - | 98.40 |
| | SAE | SAE [12] | Using European ST-T dataset to augment MIT-BIH dataset /49690 samples for testing (signal compressed by 50%) | 5 | - | - | 96.90 |
| | ACGAN [This work] | Residual network & LSTM | Balanced training dataset /9924 samples for testing | 5 | 99.53 | 99.88 | 99.81 |

introduce ACGAN model to deal with data imbalanced issue, some other results using multiple data augmentation algorithms are also summarized in Table 7. It can be seen that the detection model proposed in this study has better performance over all the comparison methods. For data augmentation, ACGAN has stronger capability than conventional sampling methods and deep-learning based SAE model.

### B. COMPETITION DATABASE

The competition database used for consecutive heartbeats detection is provided by CPSC2018 [29]. It contains 6,877 (female: 3178; male: 3699) 12 leads ECG recordings lasting from 6s to 60s (sampled with 500 Hz). As shown in Table 8, the distribution between minor and major cases is imbalanced. For example, STE and LBBB only have 202 and 207 recordings, respectively, while RBBB has 1695 recordings. Since the most minor case (STE) only occupies 3% of the total dataset, even if the network structure learns nothing about this case and gives out totally wrong detection results about it, we still can get the accuracy as high as 97%. Such class-imbalance prevents the network to learn how to identify minor class.

**TABLE 8.** Data profile of competition database.

| Type | #recording | Time length | | | | |
|------|-----------|------|------|------|--------|-------|
| | | Mean | SD | Min | Median | Max |
| Normal | 918 | 15.43 | 7.61 | 10.00 | 13.00 | 60.00 |
| AF | 1098 | 15.01 | 8.39 | 9.00 | 11.00 | 60.00 |
| I-AVB | 704 | 14.32 | 7.21 | 10.00 | 11.27 | 60.00 |
| LBBB | 207 | 14.92 | 8.09 | 9.00 | 12.00 | 60.00 |
| RBBB | 1695 | 14.42 | 7.60 | 10.00 | 11.19 | 60.00 |
| PAC | 556 | 19.46 | 12.36 | 9.00 | 14.00 | 60.00 |
| PVC | 672 | 20.21 | 12.85 | 6.00 | 15.00 | 60.00 |
| STD | 825 | 15.13 | 6.82 | 8.00 | 12.78 | 60.00 |
| STE | 202 | 17.15 | 10.72 | 10.00 | 11,89 | 60.00 |
| Total | 6877 | 15.79 | 9.04 | 6.00 | 12.00 | 60.00 |

Since data length in each recording is different, original signals need to be segmented into the same date length. As demonstrated in study 1, the location of R-peak can help to segment the heartbeat, so, finding the R-peak location which is also known as QRS detection is conducted firstly. For each 12-lead ECG sample, the lead-I signal is used to determine the specific location of R-peak. Dyadic splines 4-level wavelet is apply to realizing QRS detection. After getting the location of R-peak, 0.25s (125 points) offset before R-peak and 0.35s (175points) offset after R-peak are adopted as one heartbeat duration. Five consecutive heartbeats are integrated as one sample. Hence, each sample in this study is $12 \times 1500$ time-series vector. Fig.6 shows the segment of the original lead-I signal, and locations of R-peak are illustrated in Fig.9 (a) with black circles. The final lead-I sample after preprocessing is listed in Fig.9 (b).

Since the shortest length of one original recording contains 3000 points, after preprocessing, the capacity of the database is up to 13754 samples. Followed by the stipulation of CPSC2018, 50 samples of each type are formed as the testing database (DS2-2) for model verification, and the remainders are formed as database DS2-1 which is used for model
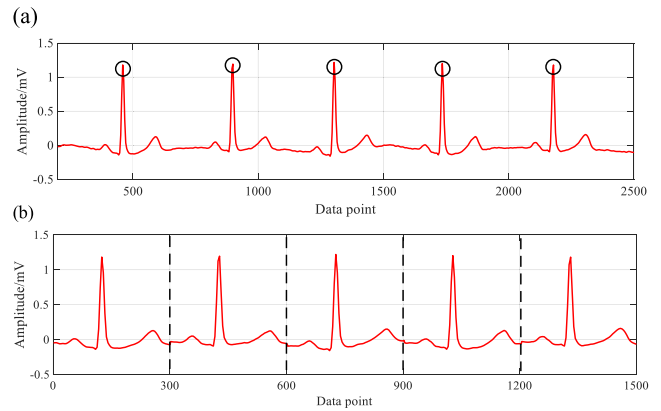


**FIGURE 9.** (a) Detection the location of R-peak; (b) signal segmentation based on R-peak.

training. Among the DS2-1, 300 samples of it are formed as the database DS2-3 which is used for ACGAN training, and 50 samples of it are formed as the database DS2-4 which is for ACGAN testing. Table 9 records the number and data length of each dataset samples. Numbers 1-9 listed in the 'Type' column represent type Normal, AF, I-AVB, LBBB, RBBB, PAC, PVC, STD, and STE, respectively.

**TABLE 9.** Data profile of competition database before data augmentation.

| Type | The number of samples | | | | | |
|------|-------|-------|-------|-------|-------------|-------|
| | DS2-1 | DS2-2 | DS2-3 | DS2-4 | Data length | Total |
| 1 | 1786 | 50 | 300 | 50 | 12×1500 | 1836 |
| 2 | 2146 | 50 | 300 | 50 | 12×1500 | 2196 |
| 3 | 1358 | 50 | 300 | 50 | 12×1500 | 1408 |
| 4 | 364 | 50 | 300 | 50 | 12×1500 | 414 |
| 5 | 3340 | 50 | 300 | 50 | 12×1500 | 3390 |
| 6 | 1062 | 50 | 300 | 50 | 12×1500 | 1112 |
| 7 | 1294 | 50 | 300 | 50 | 12×1500 | 1344 |
| 8 | 1600 | 50 | 300 | 50 | 12×1500 | 1650 |
| 9 | 354 | 50 | 300 | 50 | 12×1500 | 404 |

#### 1) DATA AUGMENTATION

Each sample contains 12-lead signal, since the scale of database for ACGAN training is limited (only 300 samples), training the ACGAN to learn variations among 12-lead signal is difficult. Therefore, we utilize lead-I to lead-XII signal to train ACGAN model respectively, and generating each lead signal one by one. Finally 12 leads generated signal are integrated as one sample.

The training epoch of ACGAN model is set to be 150, and model training process is recorded in Fig. 10. As shown in Fig.10 (a), Generator and Discriminator go towards Nash Equilibrium after approximately 90 epochs, and classification accuracy finally up around 100% as presented in Fig.10 (b).

We generate different numbers of artificial samples for each category to make sure each class contains 4000 recordings for classification model training. After data augmentation, the original training dataset DS2-1 is enriched to database DS2-5 as recorded in Table 10.

Since the scale of one generated sample in this study is large, we do not intend to list the waveform here, and only
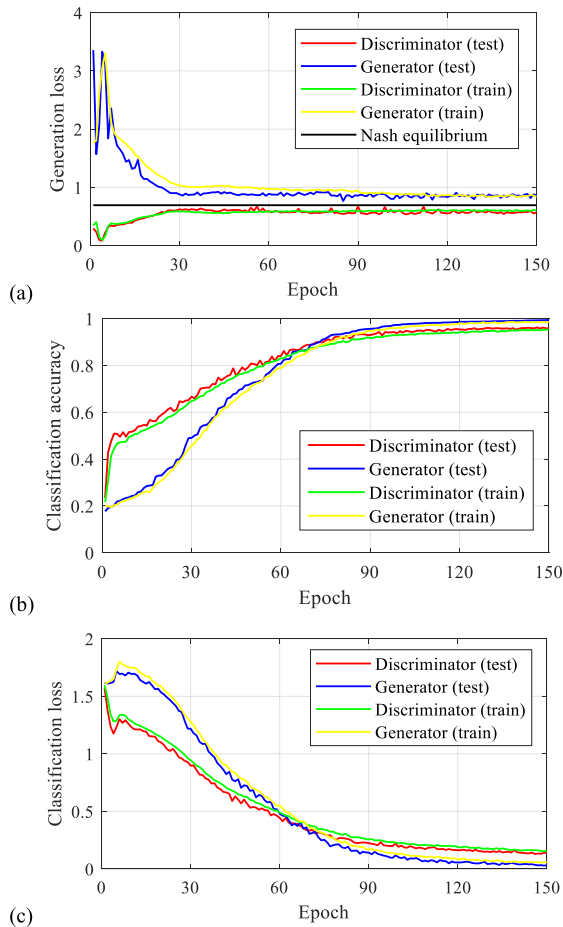
**FIGURE 10.** ACGAN performance in (a) generative loss, (b) classification accuracy, and (c) classification loss.

**TABLE 10.** Data profile of competition database after data augmentation.

| Type | Before data augmentation | | After data augmentation | |
|------|------|------|------|------|
| | DS2-1 | DS2-3 | DS2-5 | Data length |
| 1 | 1786 | 300 | 4000 | 12×1500 |
| 2 | 2146 | 300 | 4000 | 12×1500 |
| 3 | 1358 | 300 | 4000 | 12×1500 |
| 4 | 364 | 300 | 4000 | 12×1500 |
| 5 | 3340 | 300 | 4000 | 12×1500 |
| 6 | 1062 | 300 | 4000 | 12×1500 |
| 7 | 1294 | 300 | 4000 | 12×1500 |
| 8 | 1600 | 300 | 4000 | 12×1500 |
| 9 | 354 | 300 | 4000 | 12×1500 |

evaluating the generated samples based on quantitative indicators. The results are listed in the Table 11.

As shown in Table 11, the differences between value CIs and FIs are small, and the PCC value of FIs are all over 0.8. Therefore, the generated data and original data have similar distribution and strong correlation.

### 2) CLASSIFICATION RESULTS

Considering the database provided by CPSC 2018 is not the standard benchmark, there is few work conducted with

**TABLE 11.** Quantitative evaluation of generated signal.

| Type | | ED | PCC | K-L divergence |
|------|------|------|------|------|
| 1 | CIs | 3.172 | 1.000 | 2.472 |
| | FIs | 3.511 | 0.8120 | 2.531 |
| 2 | CIs | 2.019 | 1.000 | 2.340 |
| | FIs | 2.705 | 0.8113 | 2.274 |
| 3 | CIs | 2.135 | 1.000 | 2.744 |
| | FIs | 2.501 | 0.8237 | 3.013 |
| 4 | CIs | 2.543 | 1.000 | 2.167 |
| | FIs | 2.970 | 0.8317 | 2.574 |
| 5 | CIs | 2.424 | 1.000 | 3.021 |
| | FIs | 2.511 | 0.8246 | 3.317 |
| 6 | CIs | 4.144 | 1.000 | 2.988 |
| | FIs | 4.545 | 0.8131 | 3.009 |
| 7 | CIs | 4.097 | 1.000 | 3.214 |
| | FIs | 4.600 | 0.8293 | 3.512 |
| 8 | CIs | 4.213 | 1.000 | 2.914 |
| | FIs | 4.278 | 0.8204 | 3.327 |
| 9 | CIs | 5.832 | 1.000 | 3.211 |
| | FIs | 5.602 | 0.8107 | 3.518 |

this database. This study is only compared with the Top Three Results outcome in this physiological signal challenge. The evaluation criterion is proposed by the CPSC 2018 as well. As shown in Table 12, the number of classification results is counted by each label firstly.

**TABLE 12.** Counting rules for the numbers of the variables.

| | Predicted | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | total |
| 1 | $N_{11}$ | $N_{12}$ | $N_{13}$ | $N_{14}$ | $N_{15}$ | $N_{16}$ | $N_{17}$ | $N_{18}$ | $N_{19}$ | $N_{1X}$ |
| 2 | $N_{21}$ | $N_{22}$ | $N_{23}$ | $N_{24}$ | $N_{25}$ | $N_{26}$ | $N_{27}$ | $N_{28}$ | $N_{29}$ | $N_{2X}$ |
| 3 | $N_{31}$ | $N_{32}$ | $N_{33}$ | $N_{34}$ | $N_{35}$ | $N_{36}$ | $N_{37}$ | $N_{38}$ | $N_{39}$ | $N_{3X}$ |
| 4 | $N_{41}$ | $N_{42}$ | $N_{43}$ | $N_{44}$ | $N_{45}$ | $N_{46}$ | $N_{47}$ | $N_{48}$ | $N_{49}$ | $N_{4X}$ |
| 5 | $N_{51}$ | $N_{52}$ | $N_{53}$ | $N_{54}$ | $N_{55}$ | $N_{56}$ | $N_{57}$ | $N_{58}$ | $N_{59}$ | $N_{5X}$ |
| 6 | $N_{61}$ | $N_{62}$ | $N_{63}$ | $N_{64}$ | $N_{65}$ | $N_{66}$ | $N_{67}$ | $N_{68}$ | $N_{69}$ | $N_{6X}$ |
| 7 | $N_{71}$ | $N_{72}$ | $N_{73}$ | $N_{74}$ | $N_{75}$ | $N_{76}$ | $N_{77}$ | $N_{78}$ | $N_{79}$ | $N_{7X}$ |
| 8 | $N_{81}$ | $N_{82}$ | $N_{83}$ | $N_{84}$ | $N_{85}$ | $N_{86}$ | $N_{87}$ | $N_{88}$ | $N_{89}$ | $N_{8X}$ |
| 9 | $N_{91}$ | $N_{92}$ | $N_{93}$ | $N_{94}$ | $N_{95}$ | $N_{96}$ | $N_{97}$ | $N_{98}$ | $N_{99}$ | $N_{9X}$ |
| total | $N_{X1}$ | $N_{X2}$ | $N_{X3}$ | $N_{X4}$ | $N_{X5}$ | $N_{X6}$ | $N_{X7}$ | $N_{X8}$ | $N_{X9}$ | |

Then for each type, F1 score is defined in Table 13.

Based on mathematical definition of $F_1$ score, the final scores are defined as (7)-(11):

$$F_1 = (F_{11} + F_{12} + F_{13} + F_{14} + F_{15} + F_{17} + F_{17} + F_{18} + F_{19})/9 \tag{7}$$

$$F_{AF} = 2 \times N_{22}/(N_{2\,X} + N_{X2}) \tag{8}$$

$$F_{\text{Block}} = 2 \times (N_{33} + N_{44} + N_{55})/(N_{3\,X} + N_{X3} + N_{4\,X} + N_{X4} + N_{5\,X} + N_{X5}) \tag{9}$$

$$F_{PC} = 2 \times (N_{66} + N_{\eta})/(N_{6\,x} + N_{X6} + N_{7\,x} + N_{X7}) \tag{10}$$

$$F_{sT} = 2 \times (N_{gs} + N_{gg})/(N_{BX} + N_{X8} + N_{0X} + N_{X9}) \tag{11}$$

**TABLE 13.** Definition for each nine types.

| Type | $F_1$ mathematical definition |
|------|-------------------------------|
| 1 | $F_{11}=2\times N_{11}/(N_{1X}+N_{X1})$ |
| 2 | $F_{12}=2\times N_{22}/(N_{2X}+N_{X2})$ |
| 3 | $F_{13}=2\times N_{33}/(N_{3X}+N_{X3})$ |
| 4 | $F_{14}=2\times N_{44}/(N_{4X}+N_{X4})$ |
| 5 | $F_{15}=2\times N_{55}/(N_{5X}+N_{X5})$ |
| 6 | $F_{16}=2\times N_{66}/(N_{6X}+N_{X6})$ |
| 7 | $F_{17}=2\times N_{77}/(N_{7X}+N_{X7})$ |
| 8 | $F_{18}=2\times N_{88}/(N_{8X}+N_{X8})$ |
| 9 | $F_{19}=2\times N_{99}/(N_{9X}+N_{X9})$ |

**TABLE 14.** Performances with different data augmentation strategies.

| Score | Framework | | |
|-------|-----------|-----------|-------|
| | DS2-1 | Using SMOTE | DS2-5 |
| $F_{11}$ | 0.535 | 0.632 | 0.880 |
| $F_{12}$ | 0.705 | 0.891 | 0.922 |
| $F_{13}$ | 0.658 | 0.882 | 0.902 |
| $F_{14}$ | 0.750 | 0.863 | 0.880 |
| $F_{15}$ | 0.839 | 0.847 | 0.911 |
| $F_{16}$ | 0.209 | 0.548 | 0.891 |
| $F_{17}$ | 0.348 | 0.632 | 0.878 |
| $F_{18}$ | 0.519 | 0.645 | 0.857 |
| $F_{19}$ | 0.157 | 0.525 | 0.816 |
| $F_1$ | 0.524 | 0.718 | 0.883 |

To test the advantage of constructing an ACGAN based data augmentation model before training classification model. We conduct comparison experiment in three frameworks, and the results are listed in Table 14. Those three frameworks have the same classification model but different in data augmentation part. The first one is barely a classification model with no data augmentation procedure. The second one uses SMOTE to augment minor classes [30]. The third one is the proposed framework in this study.

As shown in Table 14, with no data augmentation procedure, classification model performs poorly in minor classes, such as PAC, PVC, and STE. Since those classes occupy small proportion in the training set, the losses generated by those classes are small. Therefore, classification model ignores the limited loss decline.

After applying SMOTE to augment data, the performance of the same classification model significantly improves. But SMOTE's capacity in class-imbalanced resolving is weaker than ACGAN. The ACGAN model performs better in each type.

To evaluate the performance of proposed detection framework, we compare the scores of proposed framework with The Top Three Results whose Entry Numbers are CPSC0236, CPSC0223, and CPSC0183, respectively in 2018 CPSC. The models with final codes they applied are listed in [29]. To deal with data imbalanced issue, CPSC0236 introduced Attention Mechanism to adjust weight, CPSC0223 applied over-sampling strategy to augment dataset. However, none of them attempt to utilize unsupervised or semi-supervised approaches to generated artificial data for augmentation. We summary the detailed model structure of each work, and record classification performance of them in Table 15. We highlight the highest score in each item by red mark.

**TABLE 15.** Comparison between the related work and proposed detection model.

| | Data Augmentation | QRS detection | Classifier | $F_1$ | $F_{AF}$ | $F_{Block}$ | $F_{PC}$ | $F_{ST}$ |
|---|---|---|---|---|---|---|---|---|
| CPSC0236 (Rank 1) [29] | No | No | 15-layer CNN & bi-direction gated recurrent unit (GRU) & Attention Mechanism | 0.837 | 0.933 | 0.899 | 0.847 | 0.779 |
| CPSC0223 (Rank 2) [29] | Concatenating original signal to form the new signal | No | Similar to Inception V3 Concatenating outputs from five same sub modules, and transferring outputs into 1-layer LSTM followed by 2 fully connected layers. Each sub module contains 3-layer CNN & 2-layer bi-direction GRU | 0.83 | 0.931 | 0.912 | 0.817 | 0.761 |
| CPSC0183 (Rank 3) [29] | No | Using Pan-Tomkins algorithm | 12 residual blocks &1-layer bi-direction GRU followed by 2 fully connected layers Each residual block contains 3-layer CNN with max-pooling layer | 0.806 | 0.914 | 0.879 | 0.801 | 0.742 |
| Proposed model | Using ACGAN | Using Dyadic splines 4-level wavelet | 8 residual blocks & 1-layer LSTM | 0.883 | 0.922 | 0.900 | 0.884 | 0.837 |

It is clear that the proposed model has more balanced performance in each category, and gets the highest scores in $F_1$, $F_{PC}$, and $F_{ST}$. Since $F_1$ score has the similar definition with ACC, the proposed framework significantly improves the classification accuracy. Although $F_{AF}$ and $F_{Block}$ calculated based on proposed framework are not the best, the differences between them with the best values are little. Furthermore, one notable point is that the accuracy of minor types ($F_{ST}$) has a great promotion by dealing with data imbalanced issue. Although CPSC0236 and CPSC0223 did not investigate semi-supervised approaches, they adopted other strategies to deal with the imbalanced data and achieve relatively high results. However, the ACGAN applied in this work further improves the performance, and verifies its efficiency in solving data imbalanced issue.

## V. CONCLUSION

In this paper, we propose a heartbeat arrhythmias detection framework which has high performance in robustness and accuracy. The proposed framework contains two parts: ACGAN based data augmentation model and residual network-LSTM based classification model.

Data augmentation model is designed to enrich data of minor class and recast new training dataset which has class-balanced distribution. Generator and Discriminator of ACGAN proposed in this paper are construed by stacked 1D convolutional layers with small size kernel. Dropout and Batch Normalization are utilized to avoid overfitting and gradients vanish. Classification model is designed to extract deep features from the ECGs. We adopt stacked residual network parallel connected with LSTM network as main frame to construct the classification model. Experiments on standard benchmark, MIT-BIH and competition ECG signals provided by 2018 CPSC have verified that proposed framework achieve high accuracy in both single heartbeat abnormalities detection and consecutive heartbeat detection.

However, there are still some limitations in the proposed architecture: 1) the classification accuracy of consecutive heartbeat detection is not as high as the performances of standard benchmark. It may be caused by un-accurate ECG signal segmentation. The location of R-peak in standard benchmark is accurately annotated by MIT, while the R-peak in competition database needs to be calculated by ourselves. In further development, we plan to investigate more accurately QRS detection approach to increase the classification accuracy. 2) The proposed framework is composed of two separated model, and these models are trained separately. However, the performance can be improved if joint training is feasible to achieve the global optimization. Therefore, in future work, we plan to achieve an end-to-end framework to improve the classification performance.

## REFERENCES

[1] J. Shi, B. D. Sekar, M. C. Dong, and X. Y. Hu, "Extract knowledge from site-sampled data sets and fused hierarchical neural networks for detecting cardiovascular diseases," in *Proc. Int. Conf. Biomed. Eng. Biotechnol. (ICBEB)*, Macao, China, May 2012, pp. 275–279.

[2] A. R. Naghsh-Nilchi and A. R. Kadkhodamohammadi, "Cardiac arrhythmias classification method based on MUSIC, morphological descriptors, and neural network," *EURASIP J. Adv. Signal Process.*, vol. 2008, no. 1, 2009, Art. no. 202.

[3] E. J. D. S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "ECG-based heartbeat classification for arrhythmia detection: A survey," *Comput. Methods Programs Biomed.*, vol. 127, pp. 144–164, Apr. 2016.

[4] P. de Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, Jul. 2004.

[5] C. Ye, B. V. K. V. Kumar, and M. T. Coimbra, "Heartbeat classification using morphological and dynamic features of ECG signals," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 10, pp. 2930–2941, Oct. 2012.

[6] P. Li, Y. Wang, J. He, L. Wang, Y. Tian, T.-S. Zhou, T. Li, and J.-S. Li, "High-performance personalized heartbeat classification model for long-term ECG signal," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 78–86, Jan. 2017.

[7] 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.

[8] R. Salloum and C.-C. J. Kuo, "ECG-based biometrics using recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 2062–2066.

[9] J. H. Tan, Y. Hagiwara, W. Pang, I. Lim, S. L. Oh, M. Adam, R. S. Tan, M. Chen, and U. R. Acharya, "Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals," *Comput. Biol. Med.*, vol. 94, pp. 19–26, Mar. 2018.

[10] Kandala N. V. P. S. Rajesh and R. Dhuli, "Classification of imbalanced ECG beats using re-sampling techniques and AdaBoost ensemble classifier," *Biomed. Signal Process. Control*, vol. 41, pp. 242–254, Mar. 2018.

[11] A. Ukil, S. Bandyopadhyay, C. Puri, R. Singh, and A. Pal, "Class augmented semi-supervised learning for practical clinical analytics on physiological signals," in *Proc. Mach. Learn. Health (ML4H) Workshop NeurIPS*, Montreal, QC, Canada, Dec. 2018, pp. 1–5.

[12] A. Gogna, A. Majumdar, and R. Ward, "Semi-supervised stacked label consistent autoencoder for reconstruction and analysis of biomedical signals," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2196–2205, Sep. 2017.

[13] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Int. Conf. Neural Inf. Process. Syst. (NIPS)*, Jan. 2014, pp. 2672–2680.

[14] A. Madani, J. R. Ong, A. Tibrewal, and M. R. K. Mofrad, "Deep echocardiography: Data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease," *npj Digit. Med.*, vol. 1, Oct. 2018, Art. no. 58. doi: 10.1038/s41746-018-0065-x.

[15] Z. Chen, Y. Cheng, S. Zhai, Z. Sun, and Y. Liu, "Boosting deep learning risk prediction with generative adversarial networks for electronic health records," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, New Orleans, LA, USA, Nov. 2017, pp. 787–792.

[16] S. Shao, P. Wang, and R. Yan, "Generative adversarial networks for data augmentation in machine fault diagnosis," *Comput. Ind.*, vol. 106, pp. 85–93, Apr. 2019.

[17] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *Proc. Int. Conf. Mach. Learn.*, Sydney, NSW, Australia, 2017, pp. 2642–2651.

[18] A. Krizhevsky, "Learning multiple layers of features from tiny images," Univ. Toronto, Toronto, ON, Canada, Tech. Rep. TR-2009, 2009.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 1–14.

[21] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[22] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[23] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *J. Mach. Learn. Res.*, vol. 9, pp. 249–256, May 2010.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[25] C. Olah. *Understanding LSTM Networks*. Accessed: Aug. 27, 2015. [Online]. Available: http://colah.github.io/posts/2015-08-Understanding-LSTMs/

[26] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol.*, vol. 20, no. 3, pp. 45–50, May/Jun. 2001.

[27] Z. Wu, X. Ding, and G. Zhang, "A Novel method for classification of ECG arrhythmias using deep belief networks," *Int. J. Comput. Intell. Appl.*, vol. 15, no. 4, Dec. 2016, Art. no. 1650021.

[28] L. Wang, X. Zhou, Y. Xing, M. Yang, and C. Zhang, "Clustering ECG heartbeat using improved semi-supervised affinity propagation," *IET Softw.*, vol. 11, no. 5, pp. 207–213, Oct. 2017.

[29] F. F. Liu, C. Y. Liu, L. N. Zhao, X. Y. Zhang, X. L. Wu, X. Y. Xu, Y. L. Liu, C. Y. Ma, S. S. Wei, Z. Q. He, J. Q. Li, and N. Y. Kwee, "An open access database for evaluating the algorithms of ECG rhythm and morphology abnormal detection," *J. Med. Imag. Health Inform.*, vol. 8, no. 7, pp. 1368–1373, 2018.

[30] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.

**PU WANG** was born in Anhui, China, in 1994. She received the B.S. degree from the School of Instrument Science and Engineering, Southeast University, Nanjing, China, in 2016, where she is currently pursuing the master's degree.

Her research interests include mechanical condition monitoring, biomedical monitoring, and physiological signal processing.

**BORUI HOU** (M'17) was born in Yichang, Hubei, China, in 1992. He received the B.S. degree in the measurement and control technology and instrument from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in June 2015. He is currently pursuing the Ph.D. degree with the School of Instrument Science and Engineering, Southeast University, Nanjing, China.

His academic interests include fault diagnosis, signal analysis, including physiological signal analysis, pattern identification, and computer vision.

**SIYU SHAO** received the bachelor's degree in electrical engineering and automation from Soochow University, Suzhou, China, in 2013. She is currently pursuing the Ph.D. degree in instrument science and technology with the School of Instrument Science and Engineering, Southeast University, Nanjing, China.

Her research interests include deep learning models and mechanical fault diagnosis.

**RUQIANG YAN** (M'07–SM'11) received the M.S. degree from the University of Science and Technology of China, in 2002, and the Ph.D. degree from the University of Massachusetts Amherst, in 2007.

He was a Professor with the School of Instrument Science and Engineering, Southeast University, China, from 2009 to 2018, and joined the School of Mechanical Engineering, Xi'an Jiaotong University, in 2018. His research interests include data analytics, machine learning, energy-efficient sensing, sensor networks for the condition monitoring, health diagnosis of large-scale, complex, and dynamical systems.

Dr. Yan is also a member of the ASME. He received the New Century Excellent Talents in University Award from the Ministry of Education in China, in 2009. He is also an Associate Editor of the IEEE Transactions on Instrumentation and Measurement.

. . .