



# Multi-classification of arrhythmias using ResNet with CBAM on CWGAN-GP augmented ECG Gramian Angular Summation Field

Ke Ma <sup>a</sup>, Chang'an A. Zhan <sup>a,b,c,\*</sup>, Feng Yang <sup>a,c,\*</sup>

<sup>a</sup> School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China

<sup>b</sup> Department of Rehabilitation Medicine, Zhujia Hospital, Southern Medical University, Guangzhou 510280, China

<sup>c</sup> Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou, China

## ARTICLE INFO

### Keywords:

Cardiac arrhythmias  
Gramian Angular Summation Field  
Generative Adversarial Network  
Convolutional Block Attention Modules  
Residual network

## ABSTRACT

Cardiovascular diseases are the leading cause of death globally. Arrhythmias are the most common symptoms and can cause sudden cardiac death. Accurate and reliable detection of arrhythmias from large amount of ECG signals remains a challenge. We here propose to use ResNet with convolutional block attention modules (CBAM-ResNet) to classify the major types of cardiac arrhythmias. To facilitate the classifier in extracting the rich information in the ECG signals, we transform the time series into Gramian angular summation field (GASF) images. In order to overcome the imbalanced data problem, we employ the conditional Wasserstein generative adversarial network with gradient penalty (CWGAN-GP) model to augment the minor categories. Tested using the MIT-BIH arrhythmia database, our method shows classification accuracy of 99.23%, average precision of 99.13%, sensitivity of 97.50%, specificity of 99.81% and the average F1 score of 98.29%. Compared with the performance of the state-of-the-art algorithms in the extant literature, our method is highest accuracy and specificity, comparable in precision, sensitivity and F1 score. These results suggest that transforming the ECG time series into GASF images is a valid approach to representing the rich ECG features for arrhythmia classification, and that CWGAN-GP based data augmentation provides effective solution to the imbalanced data problem and helps CBAM-ResNet to achieve excellent classification performance.

## 1. Introduction

Cardiovascular diseases (CVDs) pose great threats to human health. According to the 2021 report of the World Health Organization (WHO), CVDs are the leading cause of death globally and it is important to detect cardiovascular disease as early as possible for timely proper patient management [1]. Arrhythmias, characterized by abnormal heart rhythm, are the most common symptom of CVDs and can cause sudden cardiac death (SCD) [2]. Arrhythmias include atrial fibrillation, bradycardia, conduction disorders, premature beats, tachycardia, ventricular fibrillation, and other types of rhythm disorders [3]. According to ANSI/AAMI EC57, arrhythmias can be further divided into many sub-classes [4]. For example, the commonly used MIT-BIH arrhythmia database approximately included 40 types of rhythms [5], among which the top five heartbeat types include normal beat (NOR), left bundle branch block (LBBB), right bundle branch block, atrial premature contraction (APC), and premature ventricular contraction (PVC) [6]. Accurate identification of arrhythmia types is conducive to the early diagnosis

and early treatment of cardiovascular diseases. Electrocardiogram (ECG) is one of the most commonly used non-invasive diagnostic tools. Monitoring abnormal rhythm requires extended period of ECG recording. The detection and classification of arrhythmia depends on the in-depth interpretation of ECG signals by experienced doctors. This process is time-consuming and prone to human subjectivity. Therefore, computer-assisted arrhythmia classification becomes critical to improve accuracy and scalability [7].

Existing heart rhythm classification can be divided into two categories. One of them relies on classic machine learning, implementable with three steps: ECG signal preprocessing, ECG feature extraction and selection, and classification on the selected features [8]. Since the raw ECG records are usually contaminated by different types of noises and artifacts, the goal of preprocessing is to obtain the 'clean' ECG signal fiducially representing the condition of the heart. As a key step for ECG classification, the feature extraction step aims to extract the signal waveform morphological, statistical, wavelet and others features [8]. Different combinations of features and classification methods form

\* Corresponding authors.

E-mail addresses: [czenlab@smu.edu.cn](mailto:czenlab@smu.edu.cn) (C.A. Zhan), [yangf@smu.edu.cn](mailto:yangf@smu.edu.cn) (F. Yang).

classification algorithms with various performance. Existing literature show that k-nearest neighbor (k-NN) [9], support vector machine (SVM) [10], linear discriminant analysis (LDA) [11] and artificial feature selection shallow artificial neural network (ANN) [12] have been successfully applied to the detection of arrhythmia. These algorithms rely on supervised learning, which makes full use of the correlation between the ECG features and the class labels to find the optimal solution for the classification task. The performance of these methods depends on explicit extraction and choice of the proper ECG features, and the appropriate combination of classifiers, limiting their scalability.

As a second category, deep learning method can implement end-to-end training of classifiers, automatically learning the proper features from the sample data, without the need of hand-crafted features and thus avoiding the subjectivity and limitations of the traditional machine learning methods [13]. Chauhan *et al.* constructed a long short-term memory network (LSTM) to detect normal and abnormal signals in the ECG [14]. Shaker *et al.* proposed two convolutional neural network (CNN) classification approaches, namely the end-to-end and the two-stage hierarchical approaches, to achieve automatic classification of 15 types of arrhythmias [6]. Existing research has shown that the deep learning method can even surpass the experienced cardiologists in ECG classification [7]. Nevertheless, how to provide valuable input information appropriate to deep learning models and how to maintain the stability of network training warrant further research [15].

First, it is particularly important to provide enough and appropriate data for the deep network in order for which to learn the deep level features. The ECG signal belongs to nonlinear nonstationary time series. Most studies use one-dimensional (1D) ECG data [6] or transform it into two-dimensional (2D) space (e.g., time-frequency transformation) as network input [16,17]. Studies comparing the 1D and 2D representations of the ECG signals for deep learning suggest that transforming the ECG signals into images can facilitate the arrhythmias classification [47,50,51]. Wu *et al.* (2018) printed the ECG waveforms into a 2D plane to form images as input to the CNN model, and achieved an accuracy of 98%, higher than that for the 1D-CNN model (96%) using the ECG waveforms directly [50]. Alternatively, Huang *et al.* (2019) used short-time Fourier transform (STFT) [47] and Ullah *et al.* (2021) used continuous wavelet transform (CWT) [51] to convert the 1D ECG signals into time-frequency images as input to the learning algorithms. Although these approaches can expand the local frequency details, it also is visible that the large areas in the time-frequency space contain little information (see also the second column in Fig. 5 of this article). In addition, these transforms do not make full use of the temporal correlation structures in the original ECG signals. Gramian Angular Summation Field (GASF) [18] can fully take advantage of the correlation information in the time series and GASF has been proposed to transform EEG signals into images in combination with CNN for epilepsy prediction [19]. Tang *et al.* used GASF to retain the time correlation of the ECG RR series and significantly improved the performance of obstructive sleep apnea detection [20]. Liu *et al.* utilized GASF to highlight the periodicity of the PPG signal, thereby overcoming the classification error caused by individual differences and improving the generalizability of their classifier [21]. Above findings inspire us to explore the utility of GASF images in effective representing the 1D ECG waveforms for our learning model.

Second, the arrhythmia data is highly imbalanced [5] and the imbalance of data necessarily increases false positive to the majority and false negative to the minority, leads to instability in the network training, over-fitting and poor generalization [6,22]. Augmentation of the minority in the sample is usually employed to overcome the problem. Jun *et al.* used cropping methods to augment samples of multiple arrhythmia types to improve class balance [23]. Ukil *et al.* combined the over-sampling method and the semi-supervised feedback control approach to achieve the intelligent augmentation algorithm [24]. These resampling methods rely too much on the original samples and can achieve limited performance improvement [25]. Generative Adversarial

Networks (GAN) can estimate and adapt to the distribution of training data and generate highly similar samples. It consists of a pair of competing generator and discriminator, where the former generates an estimated image and the later tells to what extent the generated differs from the original and guide the generator to update its output. After iterations of the process, the GAN model reach the Nash equilibrium state [23]. Shaker *et al.* used GAN to augment ECG data in the time domain and offered evidence that the GAN-based approach was superior to other conventional data augmentation algorithms (e.g., SMOTE and ADASYN) [6]. However, the vanilla GAN generally suffers such problems as mode collapse, discriminator winning, instability and slow convergence speed [26,27]. Recently developed conditional Wasserstein generative adversarial network with gradient penalty (CWGAN-GP) can not only overcome these problems, but also train multi-class data at the same time and greatly reduce the training time. [28-30].

Third, deep learning can realize end-to-end training and automatically learn the proper features, avoiding the subjectivity and limitations of the traditional machine learning methods. Existing research reveals that, compared with the shallow counterparts, deep networks have better performance [31]. For example, deep convolutional neural network (DCNN) can learn more high-level features than LeNet and AlexNet (the two typical shallow networks) [17]. However, during deep network training, as more layers were stacked, the vanishing / exploding gradients phenomenon is more likely to occur, causing a reduction of the performance [32]. In order to overcome this problem, He *et al.* proposed a skip connection scheme [33]. They found that, embedded with the skip connection structure, the DCNN can speed up convergence without increasing the computational complexity. In addition, to enable the network to focus mainly on the useful features, Hu *et al.* proposed the channel attention module highlighting the correlation between feature channels and achieved high classification performance [34]. In the image domain, the spatial relationships in addition to the channels are also informative. The convolutional block attention module (CBAM) offers a good way to combine the channel and spatial information to further improve the performance [35].

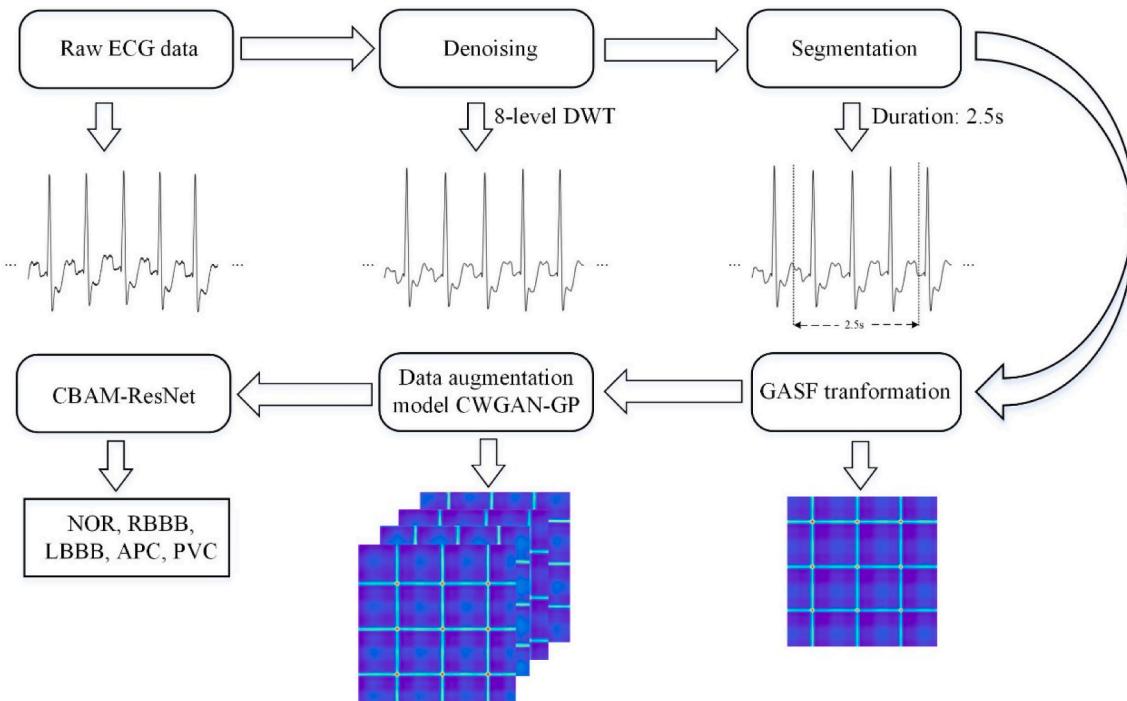
Taking advantages of the GASF in representing the rich characteristics of ECG signals, the excellent properties of CWGAN-GP in learning stability and efficiency for imbalanced data augmentation, we propose an algorithm that integrates these components for arrhythmias classification and examine the algorithm performance based on the widely used MIT-BIH arrhythmia database. Specifically, we first converted the pre-processed ECG signals into GASF images, then fed them into the CWGAN-GP network to obtain the augmented dataset. Thirdly, we introduced the CBAM mechanism able to utilize both the spatial and channel information in training the ResNet for arrhythmia classification. In the sections to follow, we first describe in detail our proposed algorithm in methods, then the key intermediate results from each module and the final performance in results. Finally, we summarize our key findings and discuss the implications of the results and the limitations that warrant further study.

## 2. Methods

The block diagram of the ECG-based arrhythmia classification algorithm proposed in this paper is shown in Fig. 1. The main steps include signal preprocessing (denoising and segmentation), transformation of 1D ECG segments into 2D GASF images, data augmentation using CWGAN-GP model, and finally, classification of five arrhythmia types using ResNet incorporated with an attention module CBAM.

### 2.1. Data

The ECG data in this paper comes from the MIT-BIH arrhythmia standard database established by the Massachusetts Institute of Technology and Beth Israel Hospital [5,36]. The database contains 48 ECG records from 47 patients between 1975 and 1979. Each record contains



**Fig. 1.** The block diagram of our proposed ECG-based arrhythmia classification algorithm.

two-channel data sampled at 360 Hz for 30 min. At least two cardiologists annotated each heartbeat in the record independently. The MIT-BIH database includes approximately 110,000 heartbeats and dozens of different types of arrhythmia. We select the modified limb lead II (MLII) channel data as the experimental data and select top four common arrhythmia and normal types for classification, namely Right Bundle Branch Block (RBBB), Left Bundle Branch Block (LBBB), Atrial Premature Contraction (APC), Premature Ventricular Contraction (PVC) and Normal (NOR). These choices follow the common practice in many previous studies in the literature [42-44,46,47].

## 2.2. Preprocessing

The preprocessing included denoising and segmentation. To remove the high frequency noise and baseline drifting, the raw ECG signals were decomposed using discrete wavelet transform (DWT) with the db5 wavelet. Given the sampling frequency of 360 Hz, eight-level of wavelet transform yielded nine coefficients. The detail coefficients for the first, second and third levels spanning frequency range [22.5 180] Hz, and the eighth approximation coefficients spanning frequency range [0 0.703] Hz were set to 0. The remaining wavelet coefficients were processed with soft thresholding. All the processed coefficients were used to reconstruct the denoised ECG signals through inverse wavelet transform. To allow the GASF images containing correlational structures of neighboring heartbeats, the ECG segments were set to have a duration of 2.5 s (900 data points due to a sampling frequency of 360 Hz). The label of the ECG segments were marked according to the annotation file accompanied with the dataset. For those marked as normal segments, the heartbeats within were all normal. Those segments marked as one of four arrhythmia types were determined by the majority label (R, L, A, V), while the total of other types appear at most once.

## 2.3. GASF transformation

The core idea of the GASF algorithm is to map the 1D time series in the Cartesian coordinate system to the polar coordinate system, and then use the trigonometric function to generate the 2D GASF matrix [37]. First, each 2.5 s ECG time series  $X = \{x_1, x_2, x_3, \dots, x_n\}$  ( $n = 900$ ) was

normalized into the interval [0, 1]:

$$\tilde{X} = (x_i - \min(X)) / (\max(X) - \min(X)) \quad (1)$$

Then the normalized time series  $\tilde{X}$  was encoded in polar coordinates, the value the angular cosine and the timestamps as the radius:

$$\begin{cases} \phi_i = \arccos(\tilde{x}_i) & 0 \leq \tilde{x}_i \leq 1, \tilde{x}_i \in \tilde{X} \\ r_i = i/N & i \in N \end{cases} \quad (2)$$

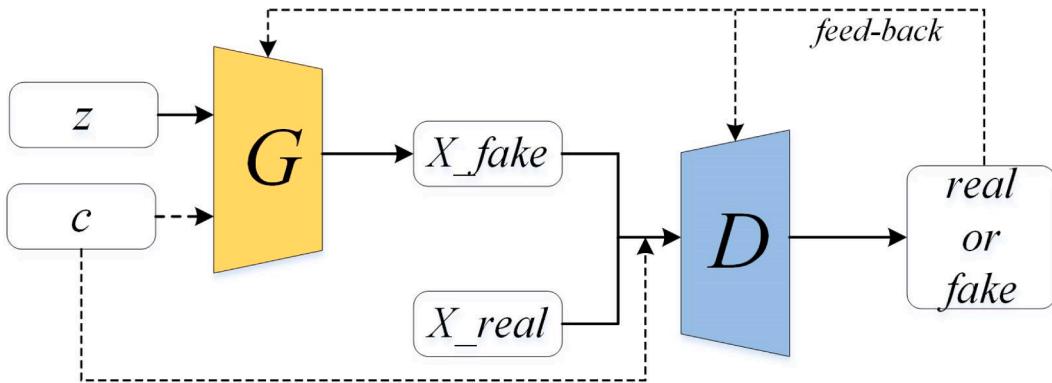
where  $i$  was the time stamp, and  $N$  was the constant factor (i.e., 900 in the current study). Because  $\tilde{x}_i \in [0, 1]$  and  $\phi_i \in [0, \pi/2]$ , the above conversion has a unique mapping result and a unique reverse mapping. 3). For each ECG segment, a new operation was defined:  $x_i \oplus x_j = \cos(\phi_i + \phi_j)$ , where  $\phi_i, \phi_j$  respectively represents the corresponding angle of  $x_i, x_j$  in the polar coordinates. The GASF matrix was finally defined:

$$GASF = \begin{bmatrix} \cos(\phi_1 + \phi_1) & \cos(\phi_1 + \phi_2) & \dots & \cos(\phi_1 + \phi_n) \\ \cos(\phi_2 + \phi_1) & \cos(\phi_2 + \phi_2) & \dots & \cos(\phi_2 + \phi_n) \\ \vdots & \vdots & \ddots & \vdots \\ \cos(\phi_n + \phi_1) & \cos(\phi_n + \phi_2) & \dots & \cos(\phi_n + \phi_n) \end{bmatrix} \quad (3)$$

Each element in the GASF matrix corresponds the cosine function of angular cosine sum in different time points.

## 2.4. Data augmentation model CWGAN-GP

The imbalance of training data hinders the network from learning how to recognize minority types of samples and affects the performance of the classification network. In the context of medical applications, the minority usually corresponds to abnormality (i.e. the markers of disease) and misclassifying the minority is costly. Therefore, it is essential to augment the sample of minority categories and alleviate the imbalanced data problem. This study employed the CWGAN-GP model for GASF data augmentation (Fig. 2). Like the vanilla GAN, it contains a generator (G) and a discriminator (D). The goal of the generator is to learn the distribution of real data and generate as real data as possible to fool the discriminator, and the goal of the discriminator is to distinguish the data



**Fig. 2.** The Conditional Wasserstein Generative Adversarial Network with Gradient Penalty (CWGAN-GP) model for data augmentation.

generated by the generator from the real data as much as possible. In addition to the vanilla GAN, CWGAN-GP introduces category information ( $c$ ) as auxiliary information. Generator  $G$  receives both Gaussian distribution noise  $z$  and class label  $c$  as inputs, generates fake GASF data  $x_{fake}$ . Discriminator  $D$  receives fake GASF data  $x_{fake}$ , real GASF data  $x_{real}$  and class label  $c$ , and determines whether the input GASF data is 'real' and matches the class label  $c$ .

The advantage of this CWGAN-GP method is that the multiple categories of data can be trained at the same time, greatly reducing the training time [28]. In comparison, the vanilla GAN is less efficient since it is trained to generate one category of data at a time. Second, CWGAN-GP uses the Wasserstein distance value in constructing the loss function in place of the JS divergence value and overcome the vanishing gradient problem [29]. Furthermore, CWGAN-GP adds the Gradient Penalty to enhance the convergence of training [30]. The specific process of CWGAN-GP is shown in Algorithm 1.

#### Algorithm 1 CWGAN-GP.

**Input:**  $\Theta$ : GASF images dataset;  $\lambda$ : the gradient penalty coefficient;  $B$ : the batch size;  $c$ : the class labels;  $\alpha$ : learning rate.  
**Output:** optimal generator parameters  $\theta$ .  
01: initial discriminator ( $D$ ) and generator ( $G$ ) parameters  $\omega_0$ ,  $\theta_0$ .  
02: **repeat**  
03:   **for** each  $b \in [1, B]$  **do**  
04:     Sample real data  $x$  from dataset  $\Theta$   
05:     Noise vector  $z$  follow standard normal distribution  $N(0, 1)$   
06:     Random number  $\mu$  follow uniform distribution  $U[0, 1]$   
07:      $\tilde{x} \leftarrow G_\theta(z|c)$   
08:      $\hat{x} \leftarrow \mu x + (1-\mu)\tilde{x}$   
09:      $L_b \leftarrow D_\omega(\hat{x}|c) - D_\omega(x|c) + \lambda(||\nabla_x D_\omega(\hat{x}|c)||_2 - 1)^2$   
10:   **end for**  
11:   Update the weights  $\omega$  of discriminator by descending:  
12:      $\omega \leftarrow \text{optimizer}(\nabla_\omega \frac{1}{B} \sum_{b=1}^B L_b, \omega, \alpha)$   
13:   Update the weights  $\theta$  of generator by descending:  
14:      $\theta \leftarrow \text{optimizer}(\nabla_\theta \frac{1}{B} \sum_{b=1}^B -D_\omega(G_\theta(z|c)|c), \theta, \alpha)$   
14: **until**  $\theta$  converge

## 2.5. Network framework

The main components of CWGAN-GP model are shown in Table 1.

**Table 1**

The main components of CWGAN-GP model.

Layer	Generator	Discriminator
1	FC (512)	Conv2d (64)
2	Deconv2d (512)	Conv2d (128)
3	Deconv2d (256)	Conv2d (256)
4	Deconv2d (128)	Conv2d (512)
5	Deconv2d (64)	Conv2d (512)
6	Conv2d (1)	FC (1)

The generator network included one fully connected layer, four deconvolution layers and one convolution layer. The number of neurons in the fully connected layer was 512. The size of the convolution kernel for each layer of deconvolution was  $5 \times 5$ , the step size was 2, and the number of convolution kernels were in order 512, 256, 128, and 64. The size of the convolution kernel of the convolution layer was  $5 \times 5$ , and the step size was 1. The discriminator network included five convolutional layers and one fully connected layer. The size of the convolution kernel of each convolution layer was  $5 \times 5$ , the step size was 2, and the number of convolution kernels were in order 64, 128, 256, 512, and 512. The number of neurons in the fully connected layer was 1.

## 2.6. Network training

The loss function  $D_{loss}$  of the discriminator and the loss function  $G_{loss}$  of the generator are defined as follows:

$$D_{loss} = E_{x \sim P_g}[D(\tilde{x}|c)] - E_{x \sim P_r}[D(x|c)] + \lambda E_{\tilde{x} \sim \tilde{x}}[(||\nabla_{\tilde{x}} D(\tilde{x}|c)||_2 - 1)^2] \quad (4)$$

$$G_{loss} = -E_{x \sim P_g}[D(\tilde{x}|c)] \quad (5)$$

where  $\tilde{x}$  is the generated GASF images,  $x$  is the real GASF images,  $c$  is the class label,  $\tilde{x}$  represents the sample randomly sampled in the real data and the generated data space,  $\lambda$  is the gradient penalty coefficient (10 in this experiment). We used the Adam optimizer with a learning rate of 0.0002 to update the parameters iteratively. Due to the large difference in the number distribution of different types of ECG signals, it was difficult to set a uniform batch size to make the network converge. We set RBBB and LBBB type samples to be trained together with a batch size of 64. The APC and PVC samples were trained together and the batch size was 32. The proposed CWGAN-GP was trained for 30,000 iterations. In each train iterations,  $G$  was updated once and  $D$  five times.

The quality of the augmented data was evaluated. Two metrics, structural similarity (SSIM) and peak signal-to-noise ratio (PSNR) [38], were introduced to quantify the quality of CWGAN-GP generated GASF images. SSIM (equ. 6) is a full-reference quality evaluation metrics that compared the similarity between the brightness, contrast, and structure of two images. Its value is between 0 and 1. The larger the value, the higher the similarity. The PSNR is an inverse-like function of MSE (mean squared error) between the two images (equ. 7). The larger the PSNR value, the smaller MSE and the degree of distortion.

$$\begin{cases} SSIM(f, g) = l(f, g) \cdot c(f, g) \cdot s(f, g) \\ l(f, g) = (2\mu_f\mu_g + C_1)/(\mu_f^2 + \mu_g^2 + C_1) \\ c(f, g) = (2\sigma_f\sigma_g + C_2)/(\sigma_f^2 + \sigma_g^2 + C_2) \\ s(f, g) = (\sigma_{fg} + C_3)/(\sigma_f\sigma_g + C_3) \end{cases} \quad (6)$$

$$\begin{cases} PSNR = 10 \cdot \log_{10}(255^2 / MSE) \\ MSE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [f(i,j) - g(i,j)]^2 \end{cases} \quad (7)$$

where  $f, g$  represent two images of size  $M \times N$ ,  $l(f,g)$ ,  $c(f,g)$ ,  $s(f,g)$  are used to measure the brightness, contrast and structure of two images.  $\mu$  refers to the mean of the image,  $\sigma$  the covariance. The small positive constants  $C_1, C_2$  and  $C_3$  are used to avoid a null denominator.

To gain qualitative knowledge about the augmented GASF images in relation the original ECG segments, the generated GASF images were inverse-transformed into 1D ECG time series (Eq. (8)).

$$\tilde{x}_i = \sqrt{(\cos(2\phi_i) + 1)/2} \quad \phi_i \in [0, \pi/2] \quad (8)$$

In order to improve the quality of the augmented sample set, a cleaning process was implemented in the data augmentation and only some of generated samples were included into the training dataset. First, the trained generator was instructed to generate a set of samples for a specific class (e.g., APC type), then these samples were compared with the real samples in the specific class one by one to quantify their quality, and finally the generated sample with the highest SSIM values were added into the training dataset. After necessary number of iterations, the training sample size for arrhythmia data reached the same number of samples in the NOR class training set.

## 2.7. CBAM-ResNet classification model

The CBAM-ResNet used in this study to classify the GASF images and thus cardiac arrhythmias is shown in Fig. 3. The network contained four residual blocks, each residual block consisted of three convolutional layers, the size of the convolution kernel was  $3 \times 3$ , and the step size was 1. Following the convolutional layer, there were a batch normalization layer and a ReLU activation function layer, which can alleviate the overfitting and ensure the gradual convergence of the network. The model also included two attention residual blocks. Each of them consisted of three convolutions and one attention residual operation. The model included five global maximum pooling layers, the pooling window was  $2 \times 2$ , and the step size was 2. Finally, the number of neurons in the two fully connected layers was set to 1024 and 1, respectively.

To improve the training efficiency by focusing on the channel and spatial information, we introduced into the ResNet two CBAM modules, whose detailed structure is shown in Fig. 4. CBAM was divided into Channel Attention Module (CAM) and Spatial Attention Module (SAM). CAM allowed the network to focus on useful feature channels and ignore the rest. SAM let the network focus on the local regions of interest in the feature map. For any given intermediate feature map  $F \in \mathbb{R}^{H \times W \times C}$ , CAM was used to compress along the spatial dimension to obtain a channel feature map  $Mc(F) \in \mathbb{R}^{H \times W \times C}$ .  $Mc(F)$  and  $F$  were merged by using element-wise multiplication to obtain  $F' \in \mathbb{R}^{H \times W \times C}$ .  $F'$  was then compressed along the channel dimension by SAM to generate a spatial feature map  $M_s(F') \in \mathbb{R}^{H \times W \times 1}$ . Finally, with the element-wise

multiplication between the spatial feature Map  $M_s(F')$  and  $F'$ , a refined feature map  $F'' \in \mathbb{R}^{H \times W \times C}$  was generated. The overall attention process can be summarized as:

$$\begin{aligned} F' &= Mc(F) \otimes F \\ F'' &= M_s(F') \otimes F' \end{aligned} \quad (9)$$

where  $\otimes$  denotes the element-wise multiplication.

The CAM module aggregated the spatial information of the input feature map  $F$  through max-pooling and average-pooling, generating two different types of features: max-pooled features and average-pooled features. Then, they were forwarded to a weight-sharing network composed of a multi-layer perceptron (MLP) with one hidden layer, followed by further Max- and Average-pooling and a nonlinear activation unit, to produce a channel attention map  $Mc(F)$ . In short, the channel attention was computed as:

$$Mc(F) = \sigma(MLP(\text{AvgPool}(F)) + MLP(\text{MaxPool}(F))) \quad (10)$$

where  $\sigma$  denotes the sigmoid function.

The SAM module conducted the max-pooling and average-pooling operations on the input feature map  $F$  along the channel dimension to obtain two feature maps with size  $oH \times W \times 1$ . These two pooling operations was shown to effectively highlight the informative regions [39]. The two feature maps were concatenated and convolved through a convolution layer (convolution kernel was  $3 \times 3$ ) to produce the final spatial attention map  $M_s(F')$  with size  $oH \times W \times 1$ . In short, the spatial attention was computed as:

$$M_s(F') = \sigma(f([\text{AvgPool}(F'); \text{MaxPool}(F')])) \quad (11)$$

where  $\sigma$  denotes the sigmoid function and  $f$  represents a convolution operation with the filter size.

## 2.8. Evaluation metrics

This paper evaluated the classification performance using five metrics: accuracy (Acc), precision (Pr), sensitivity (Se), specificity (Sp), and F1 score (F1). The equations of the metrics are as follows:

$$Acc = (TN + TP) / (TN + TP + FN + FP) \quad (12)$$

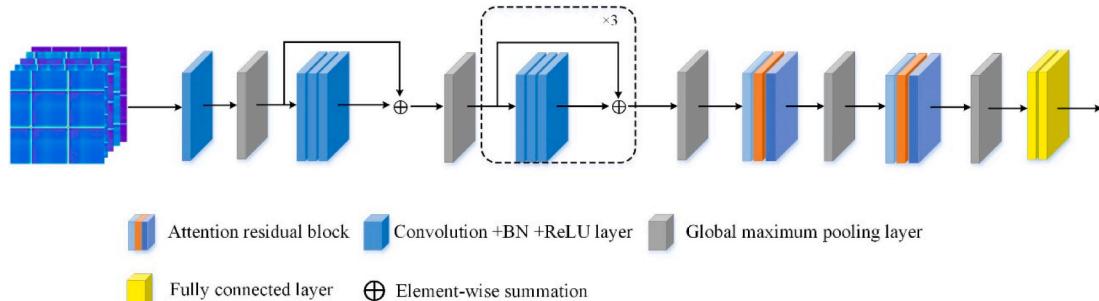
$$Pr = TP / (TP + FP) \quad (13)$$

$$Se = TP / (TP + FN) \quad (14)$$

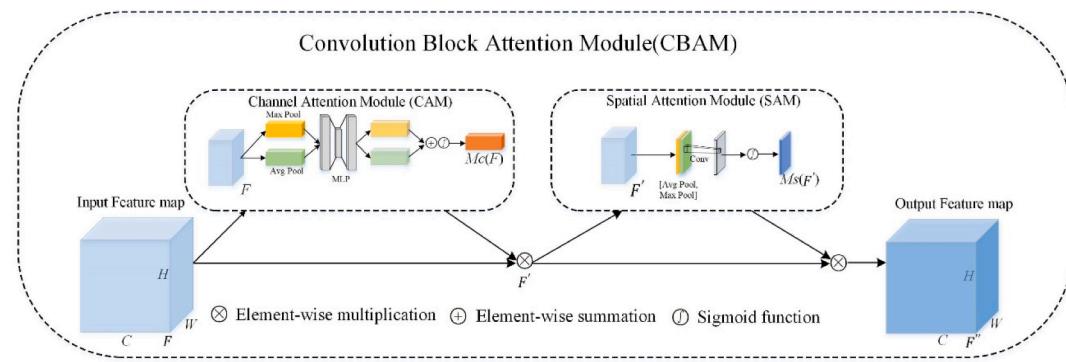
$$Sp = TN / (TN + FP) \quad (15)$$

$$F1 = 2 \times Se \times Pr / (Se + Pr) \quad (16)$$

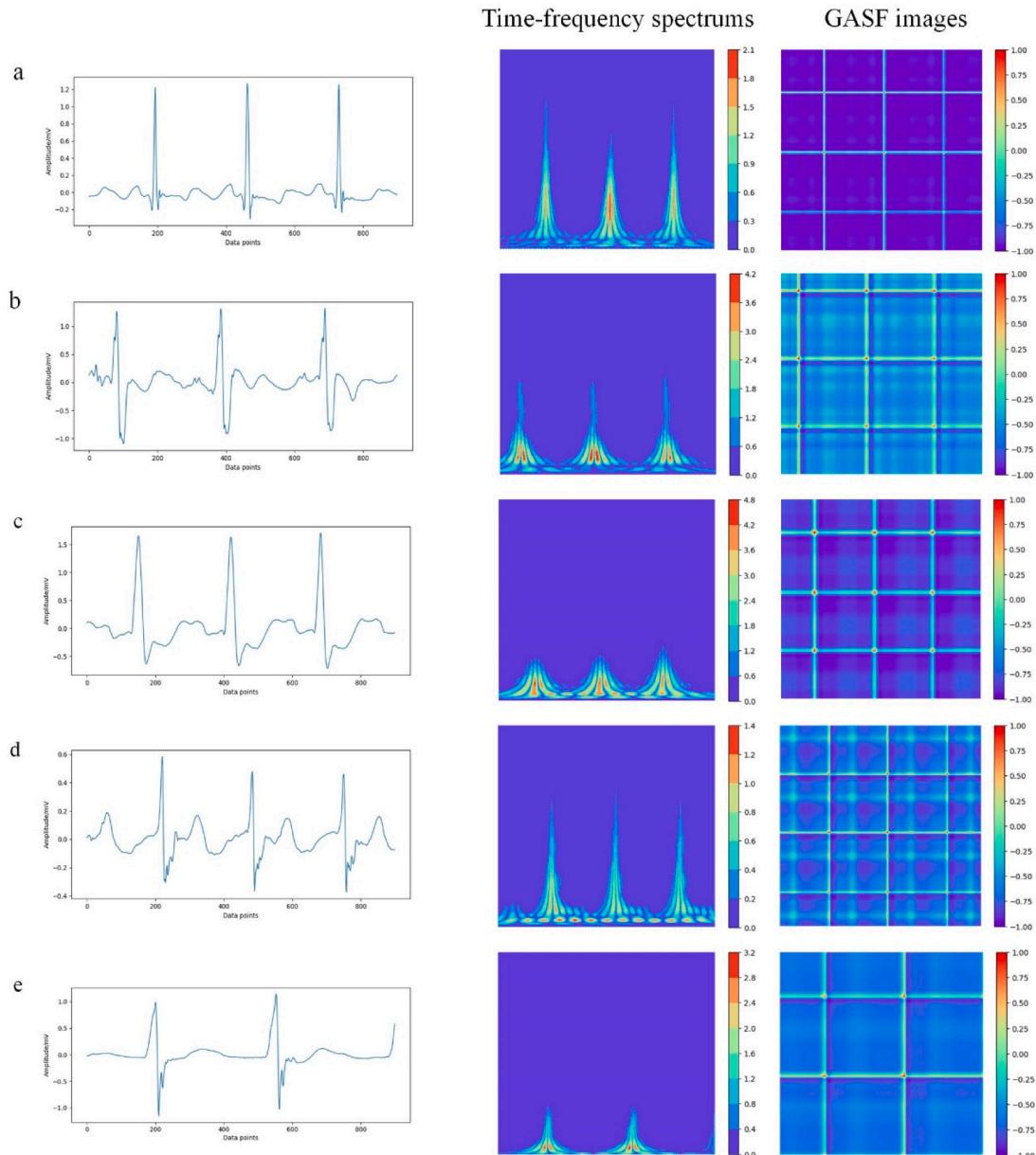
where TP refers true positive, FP refers false positive, TN refers true negative and FN refers false negative.



**Fig. 3.** The ResNet with Convolutional Block Attention Modules for arrhythmias classification on the ECG Gramian Angular Summation Field images.



**Fig. 4.** The Convolutional Block Attention Modules (CBAM).



**Fig. 5.** The different types of ECG signals (left column) and their time-frequency spectrums (middle column) and GASF images (right column). (a) Normal (NOR) type, (b) Right Bundle Branch Block (RBBB) type, (c) Left Bundle Branch Block (LBBB) type, (d) Atrial Premature Contraction (APC) type, (e) Premature Ventricular Contraction (PVC) type.

### 3. Results

#### 3.1. GASF transform of 1-D ECG signals into 2-D images

By segmenting the raw ECG signal into 2.5 s segments and labeling each segment with an arrhythmia type according to the criteria described in Methods section. As the sampling frequency was 360 Hz, each 2.5 s ECG segment was converted into GASF images with size of 900 × 900 pixels. In order to reduce the amount of parameters and computing cost, and reduce the risk of overfitting, the preprocessing step extracted ECG signals in the frequency range of [0.7–22.5] Hz and the raw GASF images (900 × 900) were down-sampled to 128 × 128 pixels. Fig. 5 shows the representative ECG segments (900 data points for 2.5 s) (the first column) and corresponding GASF images (128 × 128 pixels) (the right column). It can be seen that different ECG segment types (NOR, RBBB, LBBB, APC, PVC, from top to bottom row) correspond to strikingly different image textures. As a comparison, the results of the commonly used continuous wavelet transform (CWT) approach to 1D ECG time series transform to 2D images are shown in the middle column. It is also obvious that these spectra images are visually distinctive. However, they aggregate information in limited time-frequency space, leaving large areas almost blank. In addition, the CWT-based time-frequency spectrums are mother wavelet and parameter dependent. The GASF transform images and the ECG segments have one-to-one mapping relationship, with each element in the matrix represents the temporal correlation strength of the signal data point pair through the superposition of the direction of the time interval [37]. The bijection relationship between the GASF images and the signals is helpful for quality evaluation of the GAN-based dataset augmentation, which will be discussed in next sub-section.

#### 3.2. Augmentation of imbalanced data using CWGAN-GP

The samples of 2.5 s from the raw dataset are described in Table 2. The number of samples of the five heart rhythm types is severely imbalanced. For example, the PVC type has only 453 samples whereas the NOR type has 19,005 samples, with almost 50 times difference. In order to alleviate the imbalance in the sample sizes for model training, we randomly selected 3000 samples for the NOR type, while the maximum number of other four arrhythmia types were used to form the experimental dataset (DS), with a total of 9086 samples. Then dataset was divided at a ratio of 4:1 to form training set (DS1, 7268 samples) and test set (TS, 1818 samples). To further solve the imbalance problem, we augmented the RBBB, LBBB, APC and PVC samples using CWGAN-GP model to generate GASF images. With the augmented data and the original training set data, each of the four types also reached 2400 samples, to form the new training set (DS2).

Fig. 6 visualizes four typical examples, where the left column shows the input images to the CWGAN-GP model and the second column the generated images. Through visual inspection, the generated images (the second column) are highly similar to the original (the first column) (Fig. 6a-d). Fig. 6e-f show the ECG waveforms (blue color) inverse-

transformed from the generated GASF images. Compared to the original ECG (orange color), the generated ECG waveforms capture the key features (e.g. rhythm, peak location, QRS and ST waves) of corresponding categories. The quality of the generated samples can further be quantified using SSIM and PSNR. Table 3 shows that the SSIM values are all above 0.88, indicating that the generated samples can well capture the properties of the original samples. The PSNR values of the four arrhythmia types are all between 20 and 30 dB. Although there is no standard PSNR range to evaluate the quality of the generated image, our results were comparable to those typically reported in the literature [48,49], indicating that the degree of distortion of our generated samples is acceptable.

#### 3.3. Classification result

The whole dataset (DS) contained 9086 samples, of which 80% (DS1, 7268 samples, before data augmentation; DS2, 12,000 samples, after data augmentation) were used as the training set for model development, and 20% (TS, 1818 samples) for test. Five-fold cross-validation was used for model development. The batch size of the model was 40 and the number of epochs was set to 80. Fig. 7(a) and (b) respectively show the accuracy and loss curves for training (in blue) and validation (in orange) using DS2. It can be seen from both curves that the model converges at around 25 epochs and does not show overfitting.

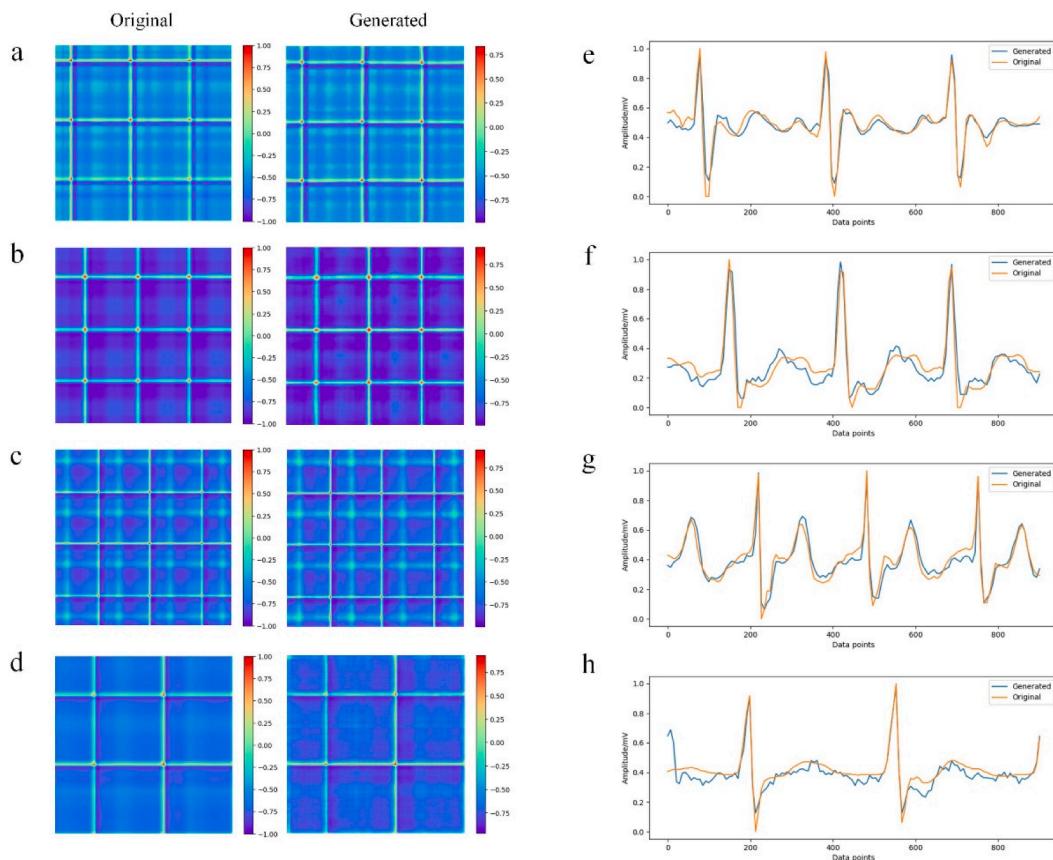
To examine to what extent data augmentation using CWGAN-GP improves the performance of the classification model, the original training set (DS1, 7268 samples, the imbalanced data) and the augmented training set (DS2, 12,000 samples with each type of 2400 samples, the balanced data) were used to train the model, respectively. The two trained models were tested on the common test set (TS, 1818 samples) and their performance compared. The confusion matrices before (left) and after (right) data augmentation are presented in Table 4. It is observed that, after data augmentation, the numbers of false positives for NOR, RBBB, LBBB, and APC are respectively reduced from 22, 16, 20 and 4 to 2, 3, 4 and 3. More importantly, the false negatives from the model trained in DS1 for APC and PVC are radically reduced from 28 and 32 to 6 and 6. The false positives and false negatives are summarized by the receiver operating characteristic (ROC) curves and the area under curve (AUC) is used to evaluate the classification performance [40]. ROC curve can reflect the relationship between true positive rate and false positive rate. As shown in Fig. 8, the AUCs for APC and PVC are significantly improved after data augmentation (from 0.9407 and 0.9951 to 0.9996 and 0.9997). Table 5 shows all the commonly used classification performance metrics before and after the data augmentation. The overall accuracy has increased from 96.48% for the model trained using the non-augmented dataset DS1, to 99.23% for that using the augmented dataset DS2. More importantly, the biggest improvement by using the augmented dataset comes for the APC and PVC classes in sensitivity and F1 measures. Specifically, the classification sensitivity values of APC and PVC types increase from 74.07% and 64.84%, much lower than other three types, to 94.44% and 93.41%, respectively, more comparable to the other classes. Similarly, the F1 scores increase from 83.33% and 77.63% to 96.79% and 95.51%, respectively for APC and PVC. The results in Tables 4 and 5 and Fig. 8 offer evidence consistent to the idea that the learning algorithms trained using imbalanced data sacrifice the performance for the minority class [6,17]. Note that APC and PVC only accounted for 5.94% and 4.98% in the entire training dataset before augmentation. The results also show that our approach using CWGAN-GP to data augmentation not only enhances the classification performance for the minority classes, but also likely improve the overall classification performance.

To understand the impact of the attention mechanism modules on the classification performance, CBAM-ResNet and CAM-ResNet were compared with ResNet. The results in the first three rows in Table 6 show that the integration of attention modules (CAM and CBAM can improve the classification performance of ResNet to varying degrees.

**Table 2**  
Distribution of data segments in the experimental dataset.

Category	# Seg.s	DS	DS1	DS1 (%)	DS2	DS2 (%)	TS
NOR	19,005	3000	2400	33.02	2400	20.00	600
RBBB	2454	2454	1963	27.01	2400	20.00	491
LBBB	2639	2639	2111	29.05	2400	20.00	528
APC	540	540	432	5.94	2400	20.00	108
PVC	453	453	362	4.98	2400	20.00	91
Total	25,091	9086	7268	100.00	12,000	100.00	1818

DS1: training set before augmentation; DS2: training set after augmentation; TS: test set; NOR: normal; RBBB: right bundle branch block; LBBB: left bundle branch block; APC: atrial premature contraction; PVC: premature ventricular contraction.



**Fig. 6.** Evaluation of the generated samples. (a) Example of the original GASF image of RBBB type and its generated counterpart. (b-d) same as (a), but respectively for LBBB, APC and PVC type. (e) The 1D ECG signals reconstructed from the original GASF image of RBBB type and its generated counterpart. (f-h) same as (e), but for LBBB, APC and PVC type, respectively.

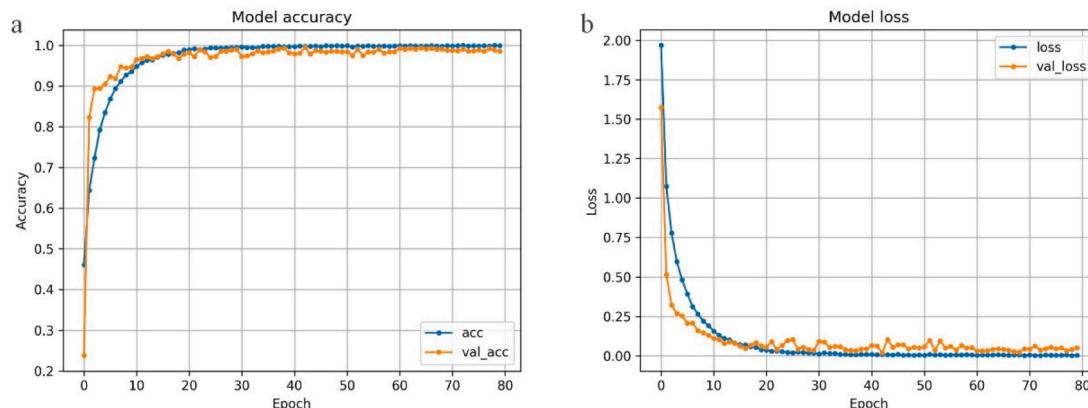
**Table 3**  
Quality evaluation of the augmented GASF samples.

Class	SSIM	PSNR(dB)
RBBB	$0.88 \pm 0.02$	$25.93 \pm 2.72$
LBBB	$0.88 \pm 0.02$	$25.18 \pm 1.58$
APC	$0.90 \pm 0.03$	$27.97 \pm 2.33$
PVC	$0.89 \pm 0.03$	$27.08 \pm 1.56$

Data are expressed as the mean  $\pm$  standard deviation.

The introduction of the CBAM module leads to the best network performance, with an accuracy of 99.23%, precision of 99.13%, specificity of 99.81%, and the F1 score of 98.29%. This shows that the introduction of the spatial attention module can effectively supplement the channel attention module (CAM), and the hybrid attention module (CBAM) can effectively extract more distinguishing features and improve the classification performance of the baseline model.

To verify the generalization ability of the trained model across datasets, we have tested our final model on the China Physiological Signal Challenge (CPSC) 2018 dataset, which includes 9831 records of



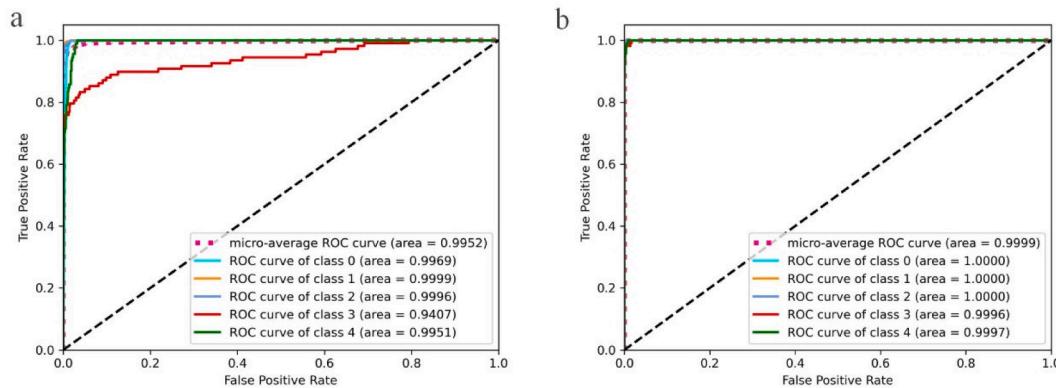
**Fig. 7.** Training and validation curves. (a) The accuracy curves for training (blue) and validation (orange). (b) The loss curves for training (blue) and validation (orange). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 4**

Classification confusion matrix before and after data augmentation.

Before data augmentation						After data augmentation					
		Predicted Labels					Predicted Labels				
Ground Truth Labels		NOR	RBBB	LBBB	APC	PVC	NOR	RBBB	LBBB	APC	PVC
Summary	NOR	599	0	1	0	0	NOR	599	0	1	0
	RBBB	1	489	1	0	0	RBBB	0	491	0	0
	LBBB	0	0	527	0	1	LBBB	0	0	527	1
	APC	21	0	6	80	1	APC	2	1	2	102
	PVC	0	16	12	4	59	PVC	0	2	2	85
	Actual	600	491	528	108	91	Actual	600	491	528	108
FP		22	16	20	4	2	FP	2	3	4	2
FN		1	2	1	28	32	FN	1	0	1	6

FP: false positive; FN: false negative.

**Fig. 8.** Receiver operating characteristic (ROC) curves. (a) The ROC curve for each arrhythmia type and the micro-average (overall) ROC curve for all types without data augmentation. (b) The ROC curve for each arrhythmia type and the overall ROC curve for all types with data augmentation. Classes 0–4 are respectively for NOR, RBBB, LBBB, APC and PVC types. Area represents the area under curve (AUC) value.**Table 5**

Classification performance before and after data augmentation.

Class	DS1					DS2				
	Acc(%)	Pr(%)	Se(%)	Sp(%)	F1(%)	Acc(%)	Pr(%)	Se(%)	Sp(%)	F1(%)
NOR	96.48	96.46	99.83	98.13	98.17	99.23	99.67	99.83	99.83	99.75
RBBB	96.83	99.59	98.75	98.19	98.19	99.39	100.00	99.82	99.69	99.69
LBBB	96.34	99.81	98.40	98.04	98.04	99.62	99.81	99.77	99.71	99.71
APC	95.24	74.07	99.76	83.33	83.33	99.25	94.44	99.76	96.79	96.79
PVC	96.72	64.84	99.88	77.63	77.63	97.70	93.41	99.88	95.51	95.51

**Table 6**

Performance of classification models for the single-channel ECG data.\*

Datasets	Model	Acc (%)	Pr (%)	Se (%)	Sp (%)	F1 (%)
MIT-BIH (ambulatory ECG)	ResNet	99.01	97.75	97.28	99.76	97.51
	CAM- ResNet	99.12	98.54	97.58	99.77	98.06
	CBAM- ResNet	99.23	99.13	97.50	99.81	98.29
CSPC (12-lead ECG)	CBAM- ResNet	86.95	87.21	86.95	96.39	87.08

\* The models were all trained using 80% data from the MIT-BIH two-channel ambulatory ECG database; they were tested on 20% data from the MIT-BIH database; and the final model was tested on channel II data from the CPSC (China Physiological Signal Challenge) database of the standard 12-lead ECG.

12-lead ECG signals sampled at 500 Hz from 9458 patients in 11 hospitals [41]. This is the only database containing all the five ECG categories studied in our model and available to us. Because the model was trained using one channel of ECG from the 2-channel ambulatory ECG in

the MIT-BIH database, we randomly selected 400 samples of channel II of the 12-lead ECG in the CPSC database for each category for a total of 2000 samples, comparable to the sample size of the test set (TS, 1818 samples) from the MIT-BIH dataset. The last row in Table 6 shows the performance of the final trained model CBAM-ResNet on the new dataset: accuracy (86.95%), precision (87.21), sensitivity (86.95%), specificity (96.39), and F1 (87.08%). It is clear that the performance of the model trained using MIT-BIH database drops dramatically. The most probable reason is that the channel II in the standard 12-lead ECG in the CPSC database is not comparable to the single channel II from the 2-channel ambulatory ECG in the MIT-BIH database. In addition, the subjects in the CPSC database are totally new to the model.

The performance of our proposed model is also compared to that of the extant literature also using the MIT-BIH database in Table 7. Most previous studies were heartbeats-based [42–45] and two segment-based [46,47]. Overall, our proposed algorithm has achieved the best performance. Among the 7 algorithms, ours ranks the 1st in accuracy and in specificity, and 3rd in precision and the 5th in sensitivity at 97.5%. Among the segment-based classification, our proposed method ranks the 1st in every performance metrics.

**Table 7**

Performance comparison between the proposed algorithm and those in the literature.

Method	class	Length	Acc (%)	Pr (%)	Se (%)	Sp (%)	F1 (%)
Li et al. [42]	5	0.69 s (beat)	98.80	98.91	98.50	99.69	—
Martis et al. [43]	5	0.56 s (beat)	93.48	99.33	99.27	98.31	—
Martis et al. [44]	5	0.56 s (beat)	98.11	99.61	99.90	99.10	—
Ji et al. [45]	5	0.3 s (beat)	99.21	—	98.06	99.45	—
Oh et al. [46]	5	2.78 s (segment)	98.10	98.69	97.50	98.70	—
Huang et al. [47]	5	10 s (segment)	99.00	—	—	—	—
Proposed	5	2.5 s (segment)	99.23	99.13	97.50	99.81	98.29
Rank	—	—	1	3	5	1	—

## 4. Discussion

In this article, we proposed to use ResNet with Convolutional Block Attention Modules (CBAM) to classify the major types of cardiac arrhythmias. To facilitate the classifier in extracting the rich information in the ECG signals, we transformed the time series into Gramian Angular Summation Field (GASF) as images. In order to overcome the imbalance problem of training dataset, we employed the Wasserstein Generative Adversarial Network with Gradient Penalty (CWGAN-GP) model to augment the minor categories. Compared with the state-of-the-art algorithms in the extant literature, our algorithm has achieved the highest performance in two out of five commonly used evaluation metrics. These algorithms worked on the ECG data in two modes: the heartbeats, and the ECG segments. Performance of algorithms based on heartbeat is generally higher than on segments, but using segments can remove the need to extract heartbeat and simplify the algorithm complexity and increase its applicability. Designed to classify fixed-length segments, our algorithm has achieved both high performance and simplicity in signal preprocessing at same time. In the category of using ECG segments as units for classification, our algorithm outperforms Oh *et al.* [46] may imply the importance of training set balancing and proper data segmenting. We used CWGAN-GP to augment the minority classes whereas Oh *et al.* used weighted loss for training. Our own experiment also showed the great improvement after augmentation (Table 5). We segmented the data into equal length of 2.5 s, and Oh *et al.* used a variable-length segmentation method but later either zero-padding or cutting to obtain 2.78 s segments to train the CNN-LSTM model [46]. The cutting and zero-padding may introduce artifacts into the training set. Huang *et al.* also used ECG segments of 10 s for classification and achieved an accuracy of 99%, but no other performance was reported [47]. One of the reason may lie in converting the ECG time series into images using Short-Time Fourier Transform (STFT) to take advantage of two-dimensional convolutional neural network (2D-CNN) to achieve high performance classification. With the inspiration of related studies, our algorithm converted the time series into GASF images to use the ResNet with CBAM for classification and achieved even higher performance.

Although our algorithm has shown good performance, some of our choices may have negatively influence its performance. First, we down-sampled the GASF images in order to remove the high frequency noises and reduce the computation cost in model training, necessarily also causing detail lost and missing the high frequency components in the ECG signals. Second, data augmentation in image space was relatively time-consuming and the quality of the generated class samples was uneven. Therefore, optimization of the generative adversarial network becomes necessary. Finally, there are many cardiac rhythm categories in the MIT-BIH dataset, future efforts should aim to accurately classify more categories with a good strategy in dealing with the more severe imbalance problem.

## 5. Conclusion

Based on the performance evaluation results and comparison with recent studies, we may conclude that transforming ECG time series into Gramian Angular Summation Field images is a valid approach to representing the rich ECG features for arrhythmia classification. The results also provide solid evidence that data augmentation using Conditional Wasserstein Generative Adversarial Network with Gradient Penalty (CWGAN-GP) model can effectively solve the data imbalance problem for robust classification. Thirdly, the introduction of Convolutional Block Attention Modules (CBAM) into ResNet offers value improving the classification performance. Given the existence of dozens of arrhythmia types and many of them occur at a very probability, accurate and reliable detection of all types of arrhythmias requires much efforts in future studies.

### CRediT authorship contribution statement

**Ke Ma:** Methodology, Investigation, Software, Writing – original draft, Writing – review & editing. **Chang'an A. Zhan:** Project administration, Methodology, Writing – original draft, Writing – review & editing, Supervision. **Feng Yang:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgement

This study was funded by the National Natural Science Foundation of China (grant number 61771233). The sponsors had no involvement in the study design, in the collection, analysis and interpretation of data; in the writing of the manuscript; or in the decision to submit the manuscript for publication. We appreciate the anonymous reviewers for their insightful and constructive comments, which have greatly helped us to improve this manuscript.

### References

- [1] World Health Organization. Cardiovascular diseases (CVDs). <http://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-cvds> (accessed on Jul. 13, 2021).
- [2] H.J. Arevalo, et al., Arrhythmia risk stratification of patients after myocardial infarction using personalized heart models, *Nat. Commun.* 7 (2016) 11437.
- [3] American Heart Association. Types of Arrhythmias. <https://www.heart.org/en/health-topics/arrhythmia/about-arrhythmia> (accessed on Oct. 6, 2021).
- [4] Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms, Standard ANSI/AAMI EC57:1998/(R)2008, Association for the Advancement of Medical Instrumentation, 1998.
- [5] G.B. Moody, R.G. Mark, The impact of the MIT-BIH arrhythmia database, *IEEE Eng. Med. Biol. Mag.* 20 (3) (2001) 45–50, <https://doi.org/10.1109/51.932724>.

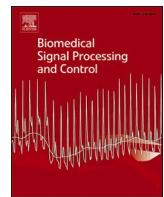
- [6] A.M. Shaker, M. Tantawi, H.A. Shedeed, M.F. Tolba, Generalization of convolutional neural networks for ECG classification using generative adversarial networks, *IEEE Access* 8 (2020) 35592–35605.
- [7] A.Y. Hannun, P. Rajpurkar, M. Haghpanahi, G.H. Tison, C. Bourn, M.P. Turakhia, A.Y. Ng, Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network, *Nat. Med.* 25 (1) (2019) 65–69.
- [8] S. Kaplan Berkaya, A.K. Uysal, E. Sora Gunal, S. Ergin, S. Gunal, M.B. Gulmezoglu, A survey on ECG analysis, *Biomed. Signal Process. Control* 43 (2018) 216–235.
- [9] Y. Kutlu, D. Kuntalp, Feature extraction for ECG heartbeats using higher order statistics of WPD coefficients, *Comput. Methods Programs Biomed.* 105 (2012) 257–267, <https://doi.org/10.1016/j.cmpb.2011.10.002>.
- [10] S. Raj, K.C. Ray, ECG signal analysis using DCT-based DOST and PSO optimized SVM, *IEEE Trans. Instrum. Meas.* 66 (3) (2017) 470–478.
- [11] R.J. Martis, U.R. Acharya, L.C. Min, ECG beat classification using PCA, LDA, ICA and discrete wavelet transform, *Biomed. Signal Process. Control* 8 (5) (2013) 437–448.
- [12] Y. Ozbay, A new approach to detection of ECG arrhythmias: complex discrete wavelet transform based complex valued artificial neural network, *J. Med. Syst.* 33 (2009) 435–445, <https://doi.org/10.1007/s10916-008-9205-1>.
- [13] X. Liu, H. Wang, Z. Li, L. Qin, Deep learning in ECG diagnosis: a review, *Knowl.-Based Syst.* 227 (2021) 107187.
- [14] S. Chauhan, L. Vig, Anomaly detection in ECG time signals via deep long short-term memory networks, in: Proceedings of the IEEE International Conference on Data Science and Advanced Analytics, 2015, <https://doi.org/10.1109/DSAA.2015.7344872>.
- [15] P. Saman, et al., Electrocardiogram monitoring and interpretation: from traditional machine learning to deep learning, and their combination, *Proceedings of the Computing in Cardiology*, 2018.
- [16] T. Wang, C. Lu, Y. Sun, M. Yang, C. Liu, C. Ou, Automatic ECG classification using continuous wavelet transform and convolutional neural network, *Entropy* 23 (1) (2021) 119.
- [17] T.J. Lan, et al., Arrhythmias classification using Short-Time Fourier Transform and GAN based data augmentation. *Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, 10.1109/EMBC44109.2020.9176733.
- [18] Z. Wang, T. Oates, Imaging time-series to improve classification and imputation, *Proceedings of the 24th International Conference on Artificial Intelligence*, 2015.
- [19] A. Shankar, et al., Epileptic seizure classification based on Gramian angular field transformation and deep learning. *Proceedings of 2020 IEEE Applied Signal Processing Conference (ASP CON)*, 2020, 10.1109/ASPCON49795.2020.9276717.
- [20] L. Tang, G. Liu, The novel approach of temporal dependency complexity analysis of heart rate variability in obstructive sleep apnea, *Comput. Biol. Med.* 135 (2021) 104632, [10.1016/j.combiom.2021.104632](https://doi.org/10.1016/j.combiom.2021.104632).
- [21] X. Liu, et al., Motion artifact detection in PPG signals based on Gramian angular field and 2-D-CNN. *Proceedings of the 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2020, 10.1109/CISP-BMEI51763.2020.9263630.
- [22] P.u. Wang, B. Hou, S. Shao, R. Yan, ECG arrhythmias detection using auxiliary classifier generative adversarial network and residual network, *IEEE Access* 7 (2019) 100910–100922.
- [23] T.J. Jun, et al., ECG arrhythmia classification using a 2-D convolutional neural network, 2018. arXiv preprint arXiv: 1804.06812.
- [24] A. Ukil, et al., Class augmented semi-supervised learning for practical clinical analytics on physiological signals, *Proceedings of the Machine Learning for Health Workshop at NeurIPS*, 2018. arXiv preprint arXiv: 1812.07498.
- [25] I.J. Goodfellow, et al., Generative adversarial nets, *Proceeding of the 27th International Conference on Neural Information Processing Systems*. 2 (2014) 2672–2680. <https://dl.acm.org/doi/10.5555/2969033.2969125>.
- [26] M. Arjovsky, L. Bottou, Towards principled methods for training generative adversarial networks, 2017. arXiv preprint arXiv: 1701.04862.
- [27] B. Zhu, J. Jiao, D. Tse, Deconstructing generative adversarial networks, *IEEE Trans. Inf. Theory* 66 (11) (2020) 7155–7179.
- [28] M. Mirza, S. Osindero, Conditional Generative Adversarial Nets, 2014. arXiv preprint arXiv: 1411.1784.
- [29] M. Arjovsky, et al., Wasserstein GAN, 2017. arXiv preprint arXiv: 1701.07875v3.
- [30] I. Gulrajani, et al., Improved training of wasserstein GANS, in: *Proceedings of the 31th International Conference on Neural Information Processing Systems*, 2017, pp. 5769–5779.
- [31] K. Simonyan, A. Zisserman, Very deep convolution networks for large-scale image recognition, *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015. arXiv preprint arXiv: 1409.1556.
- [32] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, *Proceedings of the 13th international conference on artificial intelligence and statistics*. 9 (2010) 249–256.
- [33] K. He, et al., Deep residual learning for image recognition. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, 10.1109/CVPR.2016.90.
- [34] J. Hu, et al., Squeeze-and-excitation network, *IEEE Trans. Pattern Anal. Mach. Intell.* 42 (8) (2017) 2011–2013, <https://doi.org/10.1109/TPAMI.2019.2913372>.
- [35] S. Woo, et al., CBAM: Convolutional Block Attention Module, *Proceedings of the European Conference on Computer Vision*. (2018) 3–19. arXiv preprint arXiv: 1807.06521.
- [36] A.L. Goldberger, et al., PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, *Circulation*. 101 (23) (2000) 215–220. <http://doi.org/10.1161/01.cir.101.23.e215>.
- [37] Z. Wang, T. Oates, Ending time series as images for visual inspection and classification using tiled convolutional neural networks. *Proceedings of the Workshops at the 29th AAAI on Artificial Intelligence*, 2015.
- [38] A. Horé, D. Ziou, Image quality metrics: PSNR vs, in: *SSIM*, *Proceedings of the 20th International Conference on Pattern Recognition*, 2010, pp. 2366–2369, 10.1109/ICPR.2010.579.
- [39] S. Zagoruyko, N. Komodakis, Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. *Proceedings of the International Conference on Learning Representations*, 2017.
- [40] A.P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, *Pattern Recogn.* 30 (17) (1997) 1145–1159, [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2).
- [41] F. Liu, C. Liu, L. Zhao, X. Zhang, X. Wu, X. Xu, Y. Liu, C. Ma, S. Wei, Z. He, J. Li, E. N. Yin Kwee, An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection, *J. Med. Imaging Hlth Inform.* 8 (7) (2018) 1368–1373.
- [42] H. Li, et al., Arrhythmia classification based on multi-domain feature extraction for an ECG recognition system, *Sensors*. 16 (10) (2016) 1744. <http://doi.org/10.3390/s16101744>.
- [43] R.J. Martis, U.R. Acharya, K.M. Mandana, A.K. Ray, C. Chakraborty, Cardiac decision making using higher order spectra, *Biomed. Signal Process. Control* 8 (2) (2013) 193–203.
- [44] R.J. Martis, U.R. Acharya, K.M. Mandana, A.K. Ray, C. Chakraborty, Application of principal component analysis to ECG signals for automated diagnosis of cardiac health, *Expert Syst. Appl.* 39 (14) (2012) 11792–11800.
- [45] Y. Ji, S. Zhang, W. Xiao, *Electrocardiogram classification based on faster regions with convolutional neural network*, *Sensors*. 19 (11) (2019) 2558.
- [46] S.L. Oh, E.Y.K. Ng, R.S. Tan, U.R. Acharya, Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats, *Comput. Biol. Med.* 102 (2018) 278–287.
- [47] J. Huang, B. Chen, B. Yao, W. He, ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network, *IEEE Access* 7 (2019) 92871–92880.
- [48] X. Zhang, H. Song, K. Zhang, J. Qiao, Q. Liu, Single image super-resolution with enhanced Laplacian pyramid network via conditional generative adversarial learning, *Neurocomputing*. 398 (2020) 531–538.
- [49] Y.Q. Wang, et al., Image super-resolution reconstruction based on generative adversarial network model feedback and attention mechanisms, *Multimedia Tools Appl.* 81 (2022) 6633–6652, <https://doi.org/10.1007/s11042-021-11679-1>.
- [50] Y. Wu, et al., A comparison of 1-D and 2-D deep convolutional neural networks in ECG classification. *Proceedings of the IEEE Engineering in Medicine and Biology Society*, 2018.
- [51] A. Ullah, S.u. Rehman, S. Tu, R.M. Mahmood, Fawad, M. Ehatisham-ul-haq, A hybrid deep CNN model for abnormal arrhythmia detection based on cardiac ECG signal, *Sensors*. 21 (3) (2021) 951.

## Update

# **Biomedical Signal Processing and Control**

Volume 77, Issue , August 2022, Page

DOI: <https://doi.org/10.1016/j.bspc.2022.103834>



## Corrigendum

### Corrigendum to “Multi-classification of arrhythmias using ResNet with CBAM on CWGAN-GP augmented ECG Angular Summation Field” [Biomed. Signal Process. Control 77 (2022) 103684]



Ke Ma<sup>a</sup>, Chang'an A. Zhan<sup>a,b,c,\*</sup>, Feng Yang<sup>a,c,\*</sup>

<sup>a</sup> School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China

<sup>b</sup> Department of Rehabilitation Medicine, Zhujian Hospital, Southern Medical University, Guangzhou 510280, China

<sup>c</sup> Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou, China

The authors regret to inform that several typos need to be corrected in the article.

In the first paragraph of Section 2.7, the sentence “The model included five global maximum pooling layers, ...” should be read as “The model included five maximum pooling layers, ...”. In the same paragraph, the number “1” should be “5” in the sentence “Finally, the

number of neurons in the two fully connected layers was set to 1024 and 1, respectively.”

In Fig. 3, “Global maximum pooling layer” should be read as “Maximum pooling layer”.

The authors would like to apologize for any inconvenience caused.

DOI of original article: <https://doi.org/10.1016/j.bspc.2022.103684>.

\* Corresponding authors at: School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China.

E-mail addresses: [czenlab@smu.edu.cn](mailto:czenlab@smu.edu.cn) (C.A. Zhan), [yangf@smu.edu.cn](mailto:yangf@smu.edu.cn) (F. Yang).