



Research article

ECG classification efficient modeling with artificial bee colony optimization data augmentation and attention mechanism

Mingming Zhang*, Huiyuan Jin and Ying Yang

School of Mathematics, Statistics and Mechanics, Beijing University of Technology, Beijing, 100124, China

* **Correspondence:** Email: mmzhang@bjut.edu.cn.

Abstract: In addressing the key issues of the data imbalance within ECG signals and modeling optimization, we employed the TimeGAN network and a local attention mechanism based on the artificial bee colony optimization algorithm to enhance the performance and accuracy of ECG modeling. Initially, the TimeGAN network was introduced to rectify data imbalance and create a balanced dataset. Furthermore, the artificial bee colony algorithm autonomously searched hyperparameter configurations by minimizing Wasserstein distance. Control experiments revealed that data augmentation significantly boosted classification accuracy to 99.51%, effectively addressing challenges with unbalanced datasets. Moreover, to overcome bottlenecks in the existing network, the introduction of the Efficient network was adopted to enhance the performance of modeling optimized with attention mechanisms. Experimental results demonstrated that this integrated approach achieved an impressive overall accuracy of 99.70% and an average positive prediction rate of 99.44%, successfully addressing challenges in ECG signal identification, classification, and diagnosis.

Keywords: ECG modeling; data enhancement; TimeGAN network; artificial bee colony optimization; relative position matrix; local attention mechanism; Ca-EfficientNet.

1. Introduction

Cardiovascular disease is one of the major threats to human health, and electrocardiogram (ECG) is an essential tool for detecting and diagnosing cardiovascular disease. However, there are technical challenges in manual diagnosis due to substantial differences in electrocardiograms. In recent years,

with the application of deep learning in the medical field, automatic extraction of essential features and the automatic diagnosis of cardiovascular diseases have become research topics. The extraction of features from ECG signals is a crucial step in the intelligent recognition of cardiovascular diseases. The effectiveness of diagnosis is mainly dependent on the quality of feature extraction.

Kumar proposed an ECG classification method based on R-wave interval using discrete cosine transform (DCT) conversion [1]. However, it is necessary to pay attention to the time problem, the richness of the features as well as the accuracy in the extraction process. This thesis [2] introduced the relative position matrix (RPM) as a novel feature extraction method to address the aforementioned issue. The efficacy of RPM is validated through controlled experiments compared with other methods.

Additionally, the theoretical knowledge in the field of cardiac disease is highly complex in abstraction, making the automatic diagnosis and classification of ECGs an urgent area of research. Currently, the investigation on enhancing classification accuracy and physician efficiency is a hot topic in the medical field. Thus, the exploration on automatic diagnosis optimization algorithms with technologies is of great significance for human life and health.

Support vector machine (SVM) classifiers based on machine learning are used frequently in the recognition and classification modeling of ECG signals. Qin [3] extracted morphological features of heartbeats from the MIT-BIH arrhythmia database using the discrete wavelet transform (DWT) method. An optimized SVM algorithm was proposed for six-class recognition combined with the time domain features of ECG signals. This achieves impressive results in the beat-based training scheme. Kumari [4] used SVM classifier with DWT to classify ECG signals of three different classes with an accuracy of 95.92%. The features were extracted by Monika [5] with DWT from filtered and denoised ECG signals to construct a feature matrix, which was then input into an SVM classifier for training. The XGBoost machine learning classifier has also been used for ECG signal classification with alleviating the overfitting problem. The above proposed methodologies excel in utilizing machine learning on feature extraction and achieving high accuracy in arrhythmia classification. However, further exploration into deep learning with consideration of alternative classifiers is needed, as well as additional details on computational efficiency.

These methods have achieved some progress in the recognition and diagnosis of ECG under specific classifiers by extracting various features of signals. However, the self-designed feature extraction algorithms are complex, resulting in a quantity of cost. Moreover, the quality of extracted features is largely dependent on personal knowledge of the researcher. If the initial feature extraction cannot reflect the essential properties of the ECG accurately, the classifier would not be able to fit an appropriate decision function, causing unsatisfactory classification results. To overcome the limits of manual feature extraction, end-to-end deep learning algorithms have been the focus of attention in recent years. The raw ECG signals can be input into deep learning neural networks for automatic feature extraction. This kind of ability to learn deep features can be better applied for classification, which can compensate for some of the shortcomings of manually extracting signal features.

Initially, convolutional neural networks (CNN) were widely used to enhance the ability to automatically extract features of ECG signals by increasing depth. Yildirim [6] modeled ECG signals using a one-dimensional CNN model. A CNN framework with eleven-layers was proposed for congestive heart failure evaluation by Acharya [7], which was required a minimal ECG data preprocessing with eliminating the need for artificial attribute. A novel approach for detecting atrial fibrillation was proposed on the basis of single-lead short ECG recordings [8]. A multi-scale CNN was utilized for incorporation with the fusion strategy to capture features of varying scales. A dual-stream

convolutional architecture was adopted to enable the extraction of linearly separable ECG features. Acharya [9] employed a 9-layer deep CNN model to manually detect five different types of cardiac cycles in ECG data. Additionally, a diagnostic accuracy of 94.03% in normal ECGs and 93.47% in noise-free ECGs using enhanced data was achieved.

However, due to the one-dimensional nature of ECG signals, the ability of deep network is limited for automatic extraction on original signal features. Therefore, researchers aim to improve neural networks by combining multiple machine learning classifiers into a hybrid classifier. Liu [10] compared three automatic algorithms based on one-dimensional CNN for ECG classification, achieving a high classification accuracy through SVM and stacking methods. The stacking algorithm was found to be the optimal approach in ten-fold cross-validation using the MIT-BIH Arrhythmia database. Mihaela [11] segmented single heartbeats and proposed a CNN-based multi-layer perceptron for identifying congestive heart failure. Based on idea of the manually extracted features and deep learning, a classification approach was used to convert two-dimensional images of signals, such as time-domain enhancement graphs, S-transformation graphs, and Gram-angle field graphs [12]. A fusion classification algorithm was formed of the deep network modeling for atrial fibrillation using machine learning classifiers with SVM, the random forest model, and adaboost model.

Imbalanced sample sizes are a common problem encountered in the medical field. There are often fewer disease samples than that of normal samples, which leads to a high misclassification rate for the smaller size category. To address this issue, Zheng [13] proposed a heart data augmentation method based on Generative Adversarial Networks (GAN). The data enhancement was employed by GAN method to enhance the accuracy of ECG classification effectively as demonstrated through evaluation. Wang [14] employed a generative adversarial network (GAN) to enhance ECG data. Multiple deep convolutional GAN models were constructed using an improved loss function. The models were combined with a classification model for experimental validation, demonstrating effective improvements in classification accuracy. An ACGAN-based framework [15] was introduced for detecting anomalies in ECG signals, addressing imbalanced category distribution through a data augmentation model. Utilizing the ACGAN generator and discriminator, the framework showcased high performance in handling datasets with imbalanced categories. However, they may struggle to capture temporal dependencies in time-series data like ECG. In contrast, TimeGAN excels in efficiently handling such dependencies, making it well-suited for time-series signal generation.

In summary, the field of electrocardiogram signal recognition and diagnosis has benefited greatly from the application of machine learning and deep learning techniques. Nonetheless, in the process of machine learning modeling, the manually extracted features are commonly served as the inputs for the classifiers; and it is required further research to determine the fundamental features of ECG signals. On the other hand, when utilizing deep neural networks, there is a significant issue of data dependence. This becomes especially pronounced when working with limited datasets. Moreover, it possesses a high number of parameters in deep learning modeling, posing additional challenges in practice with a lower optimization efficiency. To address the needs, transfer learning has been proved as a promising approach that makes use of previous learned knowledge to improve present classification performance. Therefore, in this study, we employed a combination of deep learning and transfer learning methods to diagnose arrhythmia, aiming to enhance the accuracy and efficiency of ECG diagnosis while addressing the challenges posed by limited data availability.

For the automatic arrhythmia classification, we address crucial challenges, namely data imbalance in the multi-classification task, classification accuracy enhancement, and the network model

architecture optimization. To tackle these issues, the proposed approach incorporates the following strategies.

1) Data augmentation with TimeGAN network: Employing the TimeGAN network is proved to be effective in enhancing classification accuracy and mitigating the challenges associated with imbalanced datasets.

2) Artificial bee colony algorithm for hyperparameter tuning: An artificial bee colony algorithm is explored automatically to search the hyperparameter configurations suitable for TimeGAN network by minimizing the Wasserstein distance. This process yields optimal parameter settings, enhancing the overall performance of TimeGAN models.

3) Transfer learning for pre-training: The concept of transfer learning is involved in pre-training the model with an existing dataset, effectively capitalizing on previously acquired knowledge in subsequent classification tasks. This significantly improves the efficiency and accuracy of the model.

4) Introduction of the Ca-Efficient network: Proposing the Ca-Efficient_network accelerates the training process while ensuring accuracy, including the local attention mechanism, optimizations, and adjustments that are affiliated in the network architecture, which further enhances the classification accuracy. This approach makes the model more adaptive to the needs of ECG classification.

The major contents of this paper are as follows. We focus on the issue of imbalanced ECG samples, and the incorporated TimeGAN network is constructed for data augmentation. The artificial bee colony algorithm is employed to search for optimal TimeGAN hyperparameters to enhance the robustness and generalization of the model by minimizing Wasserstein distance. The data enhancement method is outlined in detailed in Chapter 3. Chapter 4 underscores the significance of the TimeGAN balanced dataset, demonstrating its superiority over ACGAN data enhancement. The ECG modeling with the Gam-Resnet18 network markedly improves the classification accuracy. Then the optimization of deep learning model is discussed based on the Relative Position Matrix (RPM) by introducing the Efficient network to further improve classification accuracy. Finally, the network architecture is adjusted by incorporating the local attention mechanism. The experiments on automatic arrhythmia classification are presented in combination of the Ca-EfficientNet modeling. The results demonstrate a significant performance improvement and a notable increase in classification accuracy.

2. Methodology of data augmentation

2.1. TimeGAN network

We seek to enhance the training effectiveness and classification accuracy of deep learning models through the segmentation of ECG signals with varying input lengths. However, this process falls short of effectively addressing the challenge of category imbalance, specifically the uneven distribution of samples across different categories. To tackle this issue, we explore the potential of the TimeGAN network [16] in handling the generation and balance of ECG signals. Synthetic data are generated through TimeGAN-enabled data augmentation models to create a new balanced training dataset. This involves synthesizing additional samples of ECG signals using the TimeGAN generator, ensuring a balanced representation across different categories. The primary contributions of this study could be twofold: 1) The pioneering application of TimeGAN for the generation of ECG signal to address data imbalance and 2) the demonstrated ability of TimeGAN to achieve accuracy in class-imbalanced dataset, thereby contributing to overall performance improvement of ECG modeling.

The TimeGAN network architecture is comprised of five major components.

1) Encoder: This component maps time-series data from the original feature space into an embedded representation in the latent space using a GRU recurrent neural network. The embedding representation is then obtained through linear layers and Sigmoid activation functions.

2) Restorer Network: In contrast to the encoder, the restorer network employs a GRU recurrent neural network to restore the embedded representation back to time series data.

3) Generator Function: It generates an embedded representation of time series data in the latent space. The output embedded representation is processed through a linear layer and Sigmoid activation function, incorporating a weight initialization function for the proper weight initialization.

4) Supervisor: Responsible for generating the sequence for the next time step based on the embedding representation generated by the generator. This task is also achieved by the help of a GRU recurrent neural network with an accompanying weight initialization function.

5) Discriminator: The discriminator distinguishes the original data and the synthetic data for classification. Together, these modules form the core architecture of the TimeGAN model for generating and evaluating the performance of synthetic data, as depicted in Figure 1.

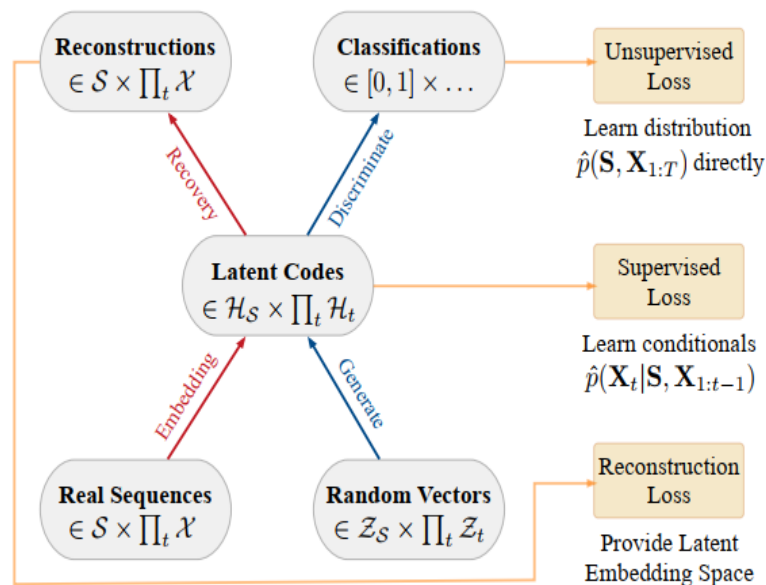


Figure 1. TimeGAN network [16] structure diagram.

2.2. Artificial bee colony optimization

In the process of the optimizing TimeGAN network, the selection of hyperparameters plays a pivotal role in determining both performance and quality of synthetic samples. We employ a biogenetic algorithm for parameter exploration through a global search, drawing inspiration from the principles of biological evolution. The Artificial Bee Colony (ABC) algorithm, functioning as a population intelligence optimization tool, offers the advantages of simplicity, global search capability and adaptability. The application of TimeGAN optimization proves to be instrumental in enhancing modeling performance, expediting the search process, and reducing the complexity associated with manual adaptation.

2.2.1. Artificial bee colony genetic algorithm

1) Initialization of honey sources

First, determine the number of nectar sources, denoted as nPop. The quality of each nectar source is analogous to its ability to attract bees. It is represented by the solution to the function commonly referred to as the fitness value. For the specified nPop nectar sources, their individual locations are determined autonomously according to Eq (1),

$$x_{id} = L_d + rand(0,1)(U_d - L_d) \quad (1)$$

where L_d and U_d represent the lower and upper bounds of the traversal, respectively.

2) Update of nectar source

The leading bee explores a new source in the vicinity of nectar source i , determined by Eq (2),

$$x_{id}^{new} = x_{id} + \varphi(x_{id} - x_{jd}), j \neq i \quad (2)$$

where φ represents uniformly distributed random numbers between $[-1,1]$. When a newly discovered nectar source proves to be adapted, it is selected to replace the original one.

3) Probability of following bee choosing leading bee

The probability that a following bee selects the leading bee is determined through the roulette rule as the following probability formula Eq (3):

$$p_i = \frac{fit_i}{\sum_{i=1}^{nPop} fit_i} \quad (3)$$

4) Generation of scout bees

Throughout the search process, if a nectar source fails to be upgraded to a superior one after reaching the specified threshold L iterations, it is deemed abandoned. In such instances, scout bees are enlisted to locate a new honey source. The location of the honey source discovered by the scout bee is determined as Eq (4):

$$x_i = \begin{cases} L_d + rand(0,1)(U_d - L_d), & trial \geq L \\ x_i, & trial < L \end{cases} \quad (4)$$

2.2.2. Objective function of Wasserstein distance

The objective function of the ABC algorithm is gauged using the Wasserstein distance to optimize the high-quality generation of ECG signals. The Wasserstein distance with its robustness and insensitivity to small noises provides a quantitative measure in evaluating the performance of a generative model. It is considered to be a powerful tool for assessing the similarity between generated samples and real data. The mathematical formulation for Wasserstein distance is expressed as Eq (5):

$$W(P, Q) = \inf \sum \sum d(x, y) \cdot \gamma(x, y) \quad (5)$$

The Wasserstein distance, denoted as $W(P, Q)$, measures the distance between two probability distributions (P, Q) . The function $d(x, y)$ represents the distance between points x and y in the metric space. Additionally, the transportation function $\gamma(x, y)$ illustrates the mass distribution transported from P to Q . The term "inf" denotes the minimum value among all potential transportation plans.

3. Application of data augmentation using the TimeGAN network with ABC optimization

After individual heartbeat segmentation, the mixed heartbeat datasets represent an imbalanced state. Data augmentation with the TimeGAN network is then employed to mitigate the challenge associated with imbalanced datasets. An artificial bee colony algorithm is explored automatically to search the hyperparameter configurations suitable for the TimeGAN network by minimizing the Wasserstein distance. The most compelling one-dimensional spurious samples are generated and incorporated into the original database for subsequent operations. With two-dimensional transformations of the RPM, the same numbers of graph samples are generated to represent the five heart rates. The flowchart is indicated in Figure 2.

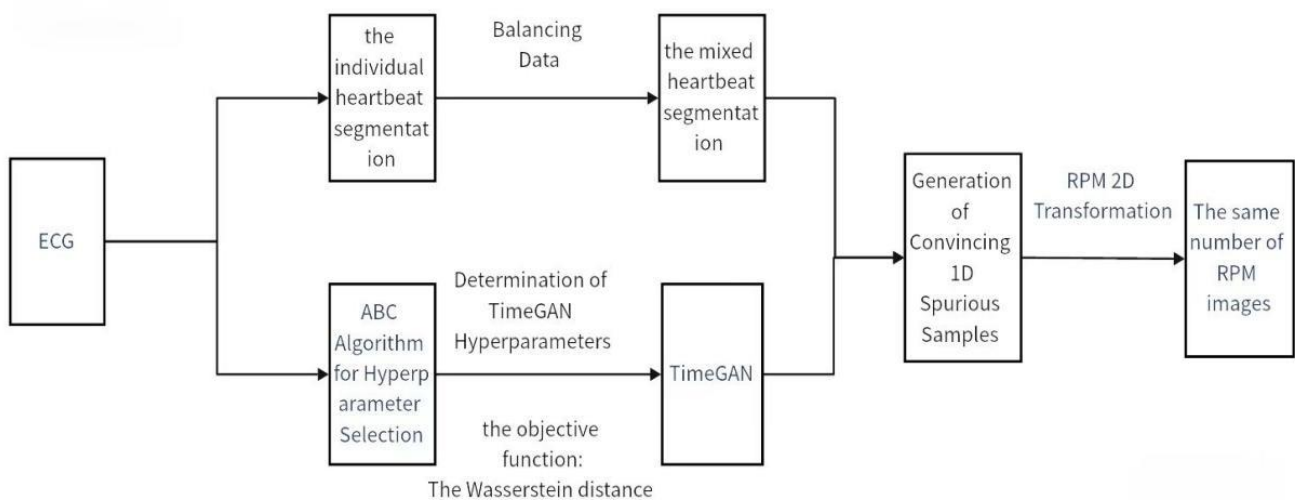


Figure 2. The flowchart of data augmentation.

3.1. Introduction of the dataset

The MIT-BIH arrhythmia database is utilized in this research, comprising of 47 participants including 25 males aged 32 to 89 years and 22 females aged 23 to 89 years. The database is focused on 38 records with MLII lead configurations. The selected rhythms encompass normal ECG and four common arrhythmia types named left bundle branch block (LBBB), right bundle branch block (RBBB), atrial pre-systole (APC), ventricular pre-systole (VPC), and premature ventricular contractions (PVC). In subsequent discussions, the signals are represented with the symbols of N, L, R, A, and V sequentially. Heartbeat segmentation is a key step in ECG signal processing. Commonly, the segmentation relies on the single heartbeat with QRS peak positions. In previous work [2], the single heartbeat segmentation was conducted to obtain the sample sizes. The heartbeat datasets represent an obvious imbalanced state as shown in Table 1.

To address data imbalance in specific normal signals, a method is employed to convert individual normal heartbeats into 5-second data segments. The intervals of 6 or 7s is also taken into account. However, from the consideration of the impact of noise and interfering signals, the interval of 5s is presented as the optimal option. In accordance with the established criteria, the segments will be labeled as a designation of N if all the heartbeats within the segments correspond to the normal category.

By converting the single normal heartbeat into segments of 5s, the imbalanced data can be effectively tackled. This approach effectively mitigates the data imbalance caused by simply segmenting individual heartbeats. The results for the mixed heartbeat samples are detailed in Table 1.

Table 1. Classification results of the mixed heartbeat segmentation.

Heart Rate Types	A	V	L	R	N
Single heartbeat segmentation	1950	6974	6578	4967	71723
Mixed heartbeat segmentation	1950	6974	6578	4967	8996

3.2. Hyperparameter optimization with the ABC algorithm

The artificial bee colony algorithm is employed for modeling optimization with the specific process presented in Figure 3.

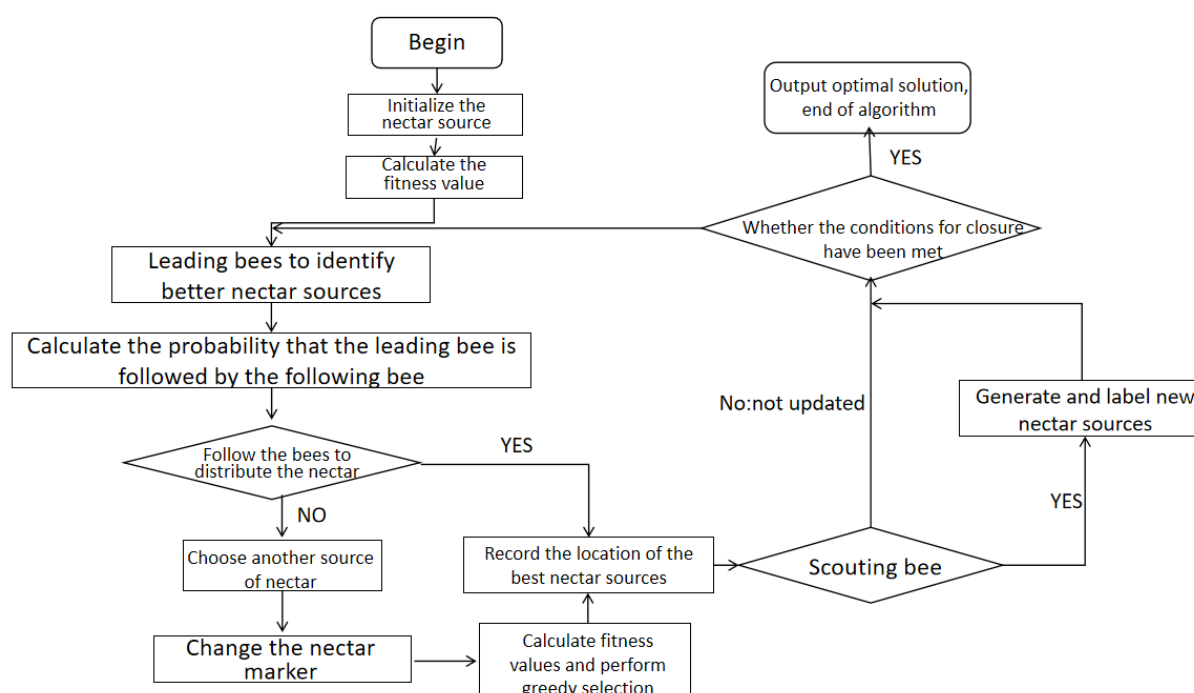


Figure 3. Artificial bee colony algorithm flowchart.

- 1) Generate an initial honey source with initializing the customized parameters.
- 2) The leader bee searches for a new nectar source and replaces it if a better one is found.
- 3) Calculate the probability that the leading bee is followed by the subsequent bee.
- 4) The following bee selects the leading bee according to the roulette rule and searches for a superior nectar source.
- 5) If a bee does not find a better one within the threshold L , it is abandoned. The scout bee takes on the task of searching for a better nectar source.
- 6) Iterate the objective function continuously set by itself.

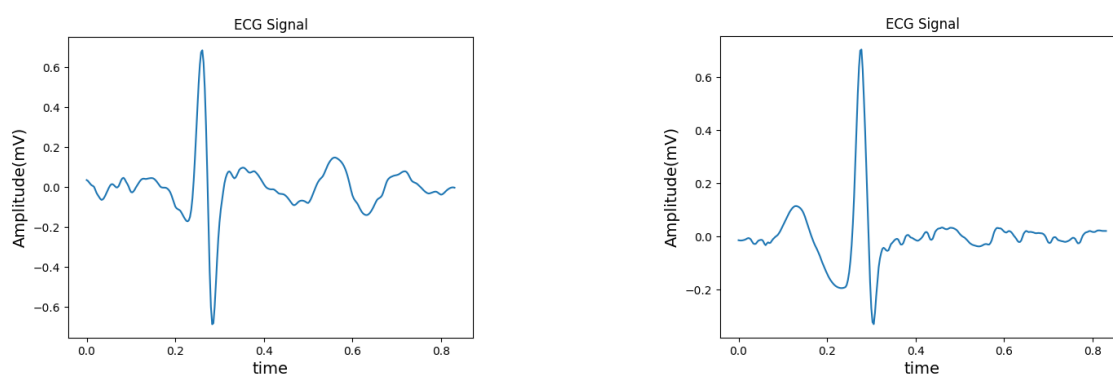
The Wasserstein distance is served as the objective function to assess the similarity between the generated samples and the original samples. ECG samples (A, L, R, and V) are generated with 500 samples for each type. The Wasserstein distances are calculated between each generated sample set and its corresponding original sample set. The sum of these distances are regarded as the optimization objective. To identify the optimal configuration, an artificial bee colony algorithm is employed to explore the hyperparameter space. The optimization objective aims to minimize the sum of Wasserstein distances, ensuring that the distribution difference between the generated and original samples is minimized.

These generated samples can effectively facilitate data expansion, constructing a more diverse dataset with a larger size. This, in turn, enhances generalization ability and classification accuracy of modeling. The optimal configurations for the four hyperparameters are identified as 64, 2, 45, and 256. The key contribution of this section lies in employing the Wasserstein distance as the objective metric for evaluating generative model performance. This metric is optimized by efficiently searching the hyperparameter space through ABC algorithm. This approach has proven highly successful in the task of ECG signal synthesis, providing a robust tool and methodology for future research and applications.

3.3. Augmentation of ECG signal with TimeGAN network

After searching for optimal hyperparameter combinations by ABC algorithm, the TimeGAN network is utilized for training subsequently. Ultimately, mixed samples are obtained with 8996 samples for each type. The results in Figure 4 display both the original (left) and generated (right) samples of the four abnormal heart rates. Subsequently, the generated data is transformed into a 2D image with Relative Position Matrix (RPM) [2] as illustrated in Figure 5.

The Relative Position Matrix (RPM) is a visualization method, which captures the relative positions between different moments in a time series. Furthermore, it enhances data interpretability by reflecting the relative position of each moment within the entire time series. This algorithm provides a more comprehensive characterization of the correlation and trend among different data points. The dependency relationship is reflected among various temporal moments within a time series. Therefore, RPM is proposed to convert ECG signals into two-dimensional images, facilitating a better feature extraction and an accurate classification.



(a)

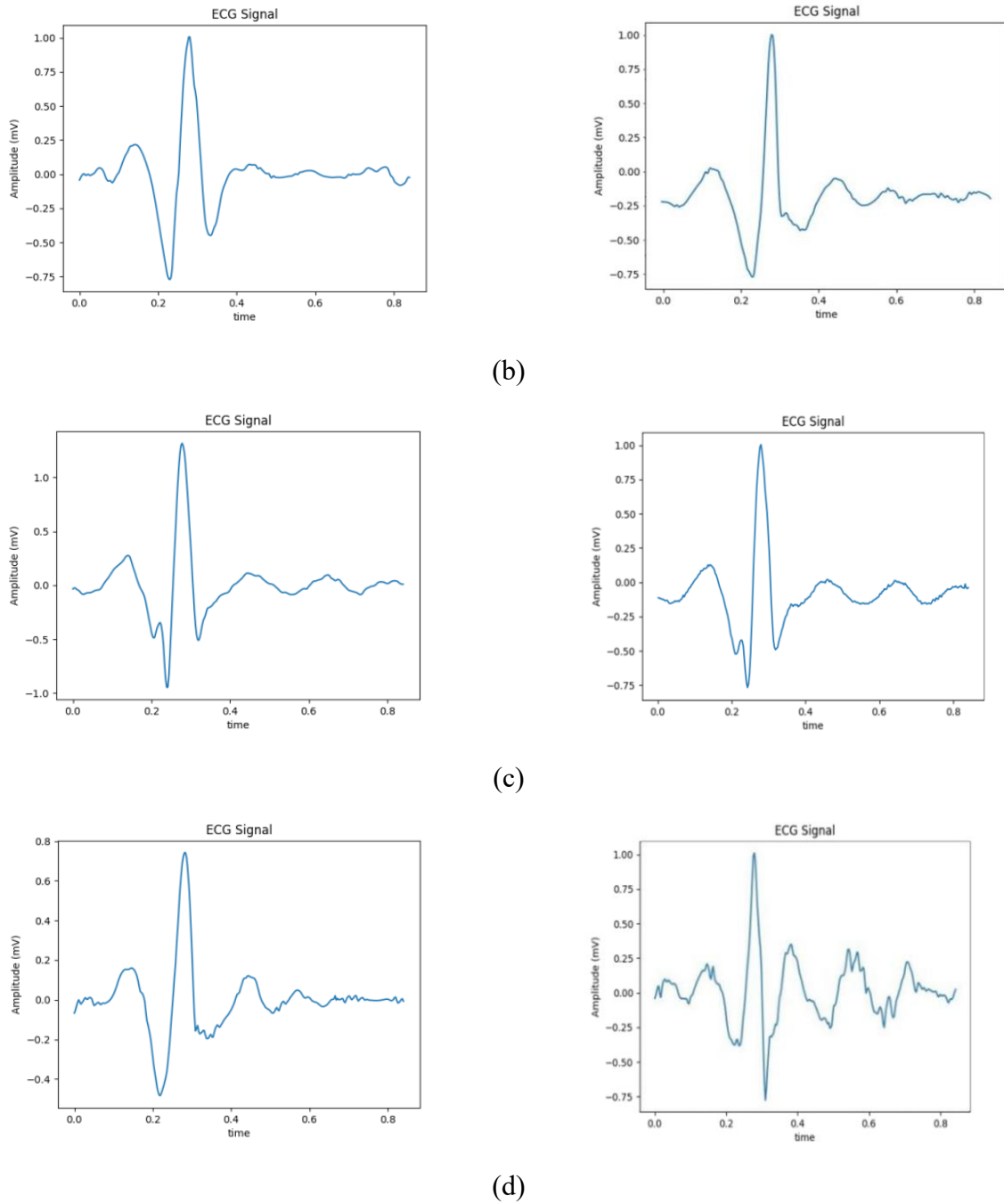


Figure 4. Comparisons of original and generated samples:(a) APC; (b) LBBB;(c) RBBB; and (d) PVC.

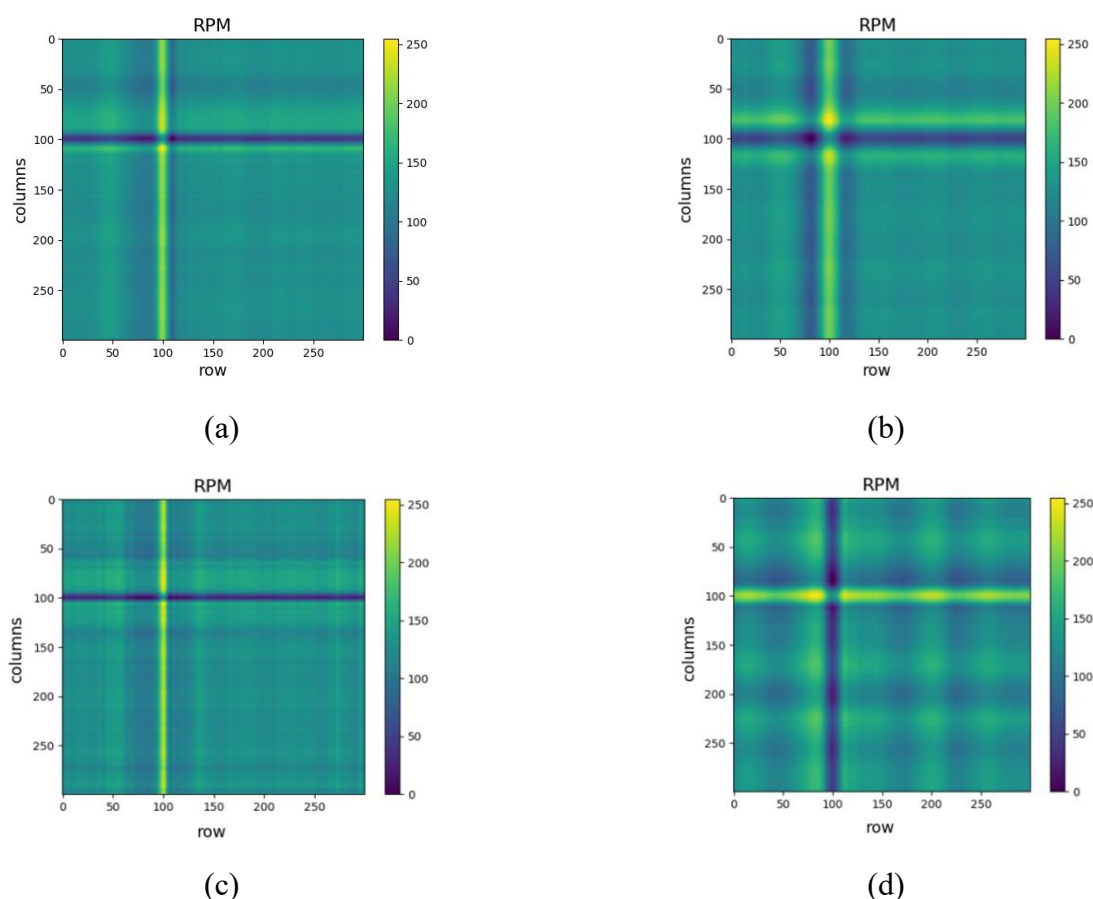


Figure 5. RPM images of ECG signals: (a) APC; (b) LBBB;(c) RBBB; and (d) PVC.

4. Classification modeling of ECG with the Gam-ResNet18 network

In this chapter, the Gam-ResNet18 network is utilized to classify the constructed ECG signal dataset and thoroughly compare its results with those presented in [2]. This comprehensive comparison serves to validate the availability of the TimeGAN network in data expansion. Furthermore, the controlled experiment involving a comparison with the ACGAN network provides an additional confirmation of the TimeGAN network effectiveness.

4.1. Classification with the TimeGAN-Gam-ResNet18 network model

The dataset is divided into training and test sets by an 8:2 ratio. For the training set, the TimeGAN network is applied for data augmentation to balance the five types of heart rates. Subsequently, the Gam-Resnet18 model is employed for training, utilizing the same hyperparameter settings in [2] with a batch size of 32, an image size of 224×224 , and a learning rate of 0.0001. The Adam optimizer is adopted for network optimization. The performance of modeling is evaluated by monitoring the accuracy and loss values on the validation set. Then training is stopped when the validation accuracy no longer improves to avoid overfitting. Visualization results in Figure 6 indicate that the loss value of validation set stabilizes after the 8th epoch, and the accuracy rate closely aligns with the training set after the 9th epoch. This verifies the robust generalization ability of modeling.

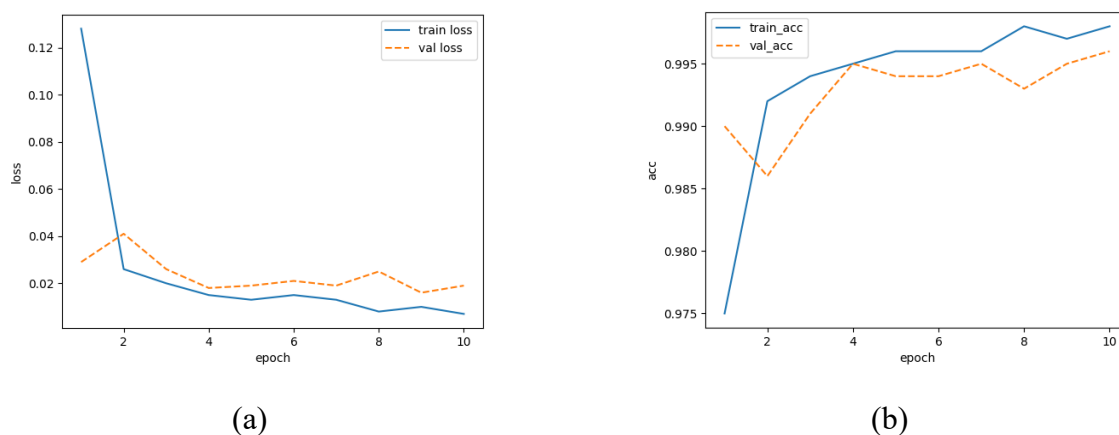


Figure 6. Modeling process of Gam-Resnet training: (a) loss variation; and (b) accuracy variation.

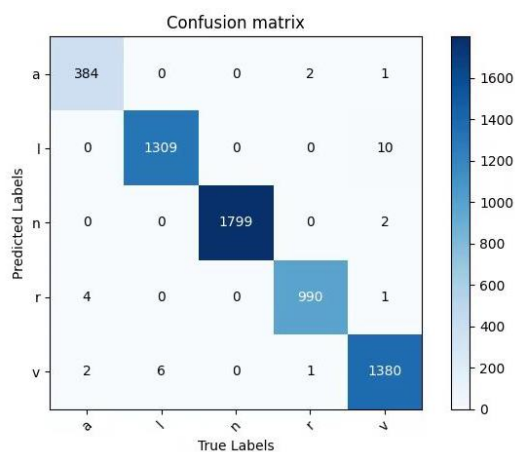


Figure 7. Confusion matrix of 5-class classification with TimeGAN.

Table 2. Performances for 5-class classification.

Method	Type	ACC	PPV	SE	SP	F1 Score	Average PPV	Average SE	Average SP
TimeGAN +RPM+ Gam- Resnet18	A		99.2%	98.5%	100%	98.8%			
	L		99.2%	99.5%	99.8%	99.4%			
	N	99.51%	99.9%	100%	100%	99.9%	99.44%	99.34%	99.88%
	R		99.5%	99.7%	99.9%	99.6%			
	V		99.4%	99.0%	99.8%	99.2%			
RPM+ Gam- Resnet18 [2]	A		95.7%	96.9%	99.7%	96.3%			
	L		99.3%	99.7%	99.8%	99.5%			
	N	99.30%	99.9%	100%	100%	99.9%	98.76%	99.84%	98.90%
	R		99.7%	99.0%	99.9%	99.3%			
	V		99.2%	98.9%	99.8%	99.0%			

The classification on the test set is assessed by applying the trained model to generate the corresponding confusion matrix depicted in Figure 7. Evaluation metrics are calculated based on the confusion matrix including total classification accuracy (ACC), positive predictive value (PPV), sensitivity (SE), specificity (SP), and F1 score as shown in Table 2.

The trained ECG model exhibits an outstanding performance in the five classifications of signals, achieving a classification accuracy up to 99.51%. In comparison with the related findings, the experimental results demonstrate that data enhancement with the TimeGAN network effectively enhances the classification accuracy and significantly addresses the issue of unbalanced datasets.

4.2. Classification with the ACGAN-Gam-Resnet18 network model

The potential of the ACGAN network is also investigated in addressing the signal imbalance problem. This model comprises two parts as a data augmentation model and a classification model. ACGAN network is employed to generate artificial ECG signals for creating the new class-balanced training datasets. Similar to the previous section, these approaches result in mixed samples of 8996 for each class. As a control experiment, these datasets are fed into the Gam-Resnet18 network for training. The corresponding confusion matrix is obtained to evaluate the classifiers' performance on the test set as presented in Figure 8. According to Table 3, the trained model achieves a 99.42% classification accuracy in the five classifications of ECG signals. In comparison, the TimeGAN network has demonstrated a significant advantage in data enhancement.

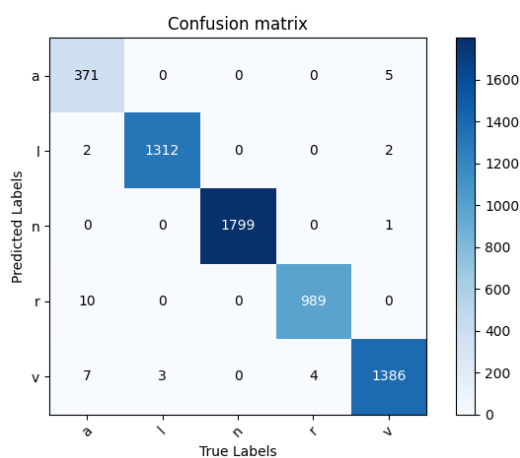


Figure 8. Confusion matrix of 5-class classification with ACGAN.

Table 3. Performances for 5-class classification with ACGAN.

Method	Type	ACC	PPV	SE	SP	F1 Score	Average PPV	Average SE	Average SP
ACGAN +RPM+ Gam- Resnet18	A		98.7%	95.1%	99.9%	96.9%			
	L		99.7%	99.8%	99.9%	99.7%			
	N	99.42%	99.9%	100%	100%	99.9%	99.26%	98.78%	99.86%
	R		99.0%	99.6%	99.8%	99.3%			
	V		99.0%	99.4%	99.7%	99.2%			

5. Classification modeling of ECG with attention mechanism optimized Efficient Network

In this chapter, the focus is shifted to the optimization of the Gam-ResNet18 classification network. Initially, the Gam-ResNet18 framework is replaced with the Efficient network [17]. A superior model of the Efficient network is established through controlled experiments. To further enhance the performance, the CA (Channel Attention) local attention mechanism [18] is introduced based on the Efficient network. Controlled experiments conducted for identical samples aim to substantiate the improved performance of the Ca-Efficient network on the classification and recognition of ECG signals. The flowchart is expressed in Figure 9.

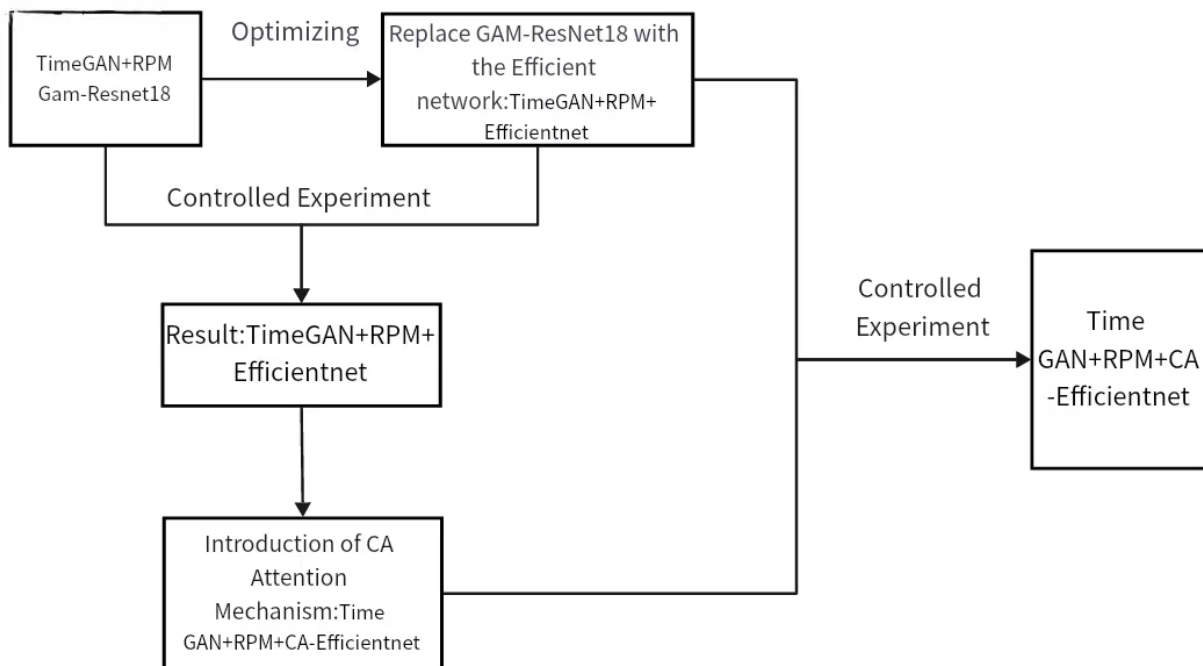


Figure 9. The flowchart of Classification modeling.

5.1. Transfer learning model for arrhythmias classification based on EfficientNet

The Gam-ResNet18 network, serving as an image classification model, faces challenges such as high computation requirement, insensitivity to input image size and susceptibility to overfitting. In contrast, EfficientNet [17] exhibits a superior performance with fewer parameters, signifying enhanced feature learning within the same computational constraints. Its scalability enables adaptability to diverse datasets and computational resources, optimizing the feature extraction for improved data abstraction. It results in heightened generalization performance, rendering EfficientNet widely applicable across practical scenarios. It encompasses multiple models of the same network architecture, adjusting complexity through scaling hyperparameters. The whole structure comprises four parts: Initial processing with the convolution layer, multiple EfficientNet blocks, batch normalization, and an activation function adjusting width and depth. Subsequently, a global average pooling layer converts image feature vectors into fixed-length vector representations, ultimately outputting classification results via a fully connected layer.

Transfer learning involves migrating knowledge from a pre-trained model to a new task. This approach accelerates training process and enhances model accuracy by transferring knowledge structures and feature representations from a trained model. For image classification modeling, the pre-trained EfficientNet is utilized as the base network to apply for the ECG signal classification task. In the training process, a cosine annealing learning rate scheduling function is employed to prevent gradient explosion and overfitting with utilizing the batch_size of 16 and the image size of 224×224 . As depicted in Figure 10, the performance is evaluated by monitoring the accuracy and loss value of the validation by the SGD optimizer. Training is stopped when the validation accuracy ceases to improve.

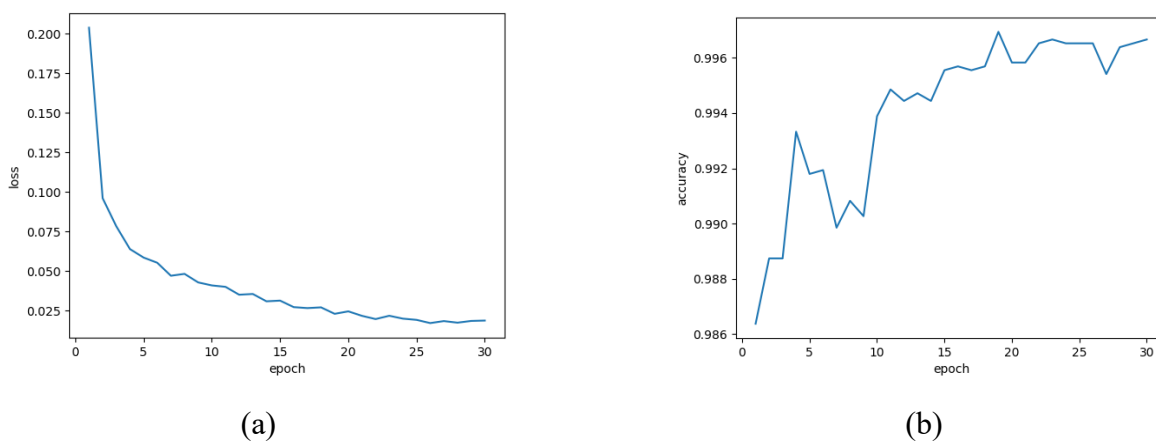


Figure 10. Modeling process of EfficientNet training: (a) Loss variation; and (b) accuracy variation.

After screening by evaluation, the modeling with 18th epoch weights assesses the performance of the 2D image classifier for ECG signals, with the corresponding confusion matrix shown in Figure 11. On the basis of this matrix, the evaluation metrics are calculated to appraise performance for five ECG signal classifications. According to the results in Table 4, the ECG model achieves a classification accuracy of 99.62%, significantly outperforming the Gam-Resnet18 model. Moreover, improvements are observed in all the indexes. This underscores the high effectiveness of the proposed EfficientNet model for the task of image classification of ECG signals.

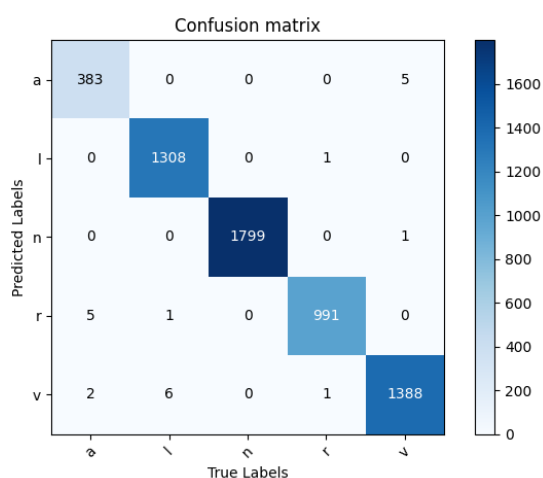


Figure 11. Confusion matrix of 5-class classification results of EfficientNet.

Table 4. Performance metrics for 5-class classification of EfficientNet.

Method	Type	ACC	PPV	SE	SP	F1 Score	Average PPV	Average SE	Average SP
TimeGAN +RPM+ Efficientnet	A		98.7%	98.2%	99.9%	98.5%			
	L		99.9%	99.5%	100%	99.7%			
	N	99.62%	99.9%	100%	100%	100%	99.47%	99.41%	99.44%
	R		99.4%	99.8%	99.9%	99.6%			
	V		99.4%	99.6%	99.8%	99.5%			

5.2. EfficientNet modeling with attention mechanism fusion

5.2.1. Attention mechanism

In this section, the performance of modeling is discussed to be further enhanced by the addition of the channel attention (CA) mechanism [18]. The CA attention mechanism is a crucial technique widely used in convolution neural networks. It achieves effective weighting of the feature map by learning the importance of each channel. It involves two major steps. First, the input feature maps undergo processing through a global average pooling layer to derive the global average of each channel. Subsequently, attention weights for each channel are obtained by nonlinear transformation of the channel averages through two fully connected layers. Second, in the feature fusion stage, the computed weights of channel attention are utilized to sum the original feature maps, generating the final weighted feature maps. This enables the network to intelligently select the importance of each channel, thereby improving overall performance.

In comparison to the squeeze excitation (SE) attention mechanism [2], CA offers advantages such as improved computation efficiency and fewer parameters. CA is designed to acquire channel weights through a small fully-connected layer, significantly reducing the number of parameters. The flexibility of the CA mechanism is evident as it can be conveniently embedded into different locations of the convolution neural network, leading to a significant improvement in performance.

5.2.2. CA-Efficient hybrid network

The fusion of CA-Efficient network combines the Efficient network and the CA mechanism to achieve adaptive learning for each channel and the weighted sum optimization of feature maps. Compared to traditional convolution neural networks, Ca-EfficientNet has the following advantages:

1) It provides a better feature expression due to the strong capability on image classification. By the addition of the CA mechanism, it further enhances the feature representation ability of the network by selectively capturing the importance of each channel.

2) It reveals that Ca-EfficientNet achieves a higher classification accuracy when training with identical data size. This outcome is primarily attributable to the CA mechanism being able to extract plentiful feature information.

3) It utilizes fewer parameters for Ca-EfficientNet than other attention-based networks while maintaining a high performance. This results in faster training, thereby reducing computation cost.

In conclusion, it holds broad applications in various image classifications and object detections by Ca-EfficientNet and exhibits a potential for improving accuracy of modeling with parameter reduction. The topology of Ca-EfficientNet is shown in Figure 12, and the architecture of CA-MBConv in the Ca-EfficientNet is indicated in Figure 13.

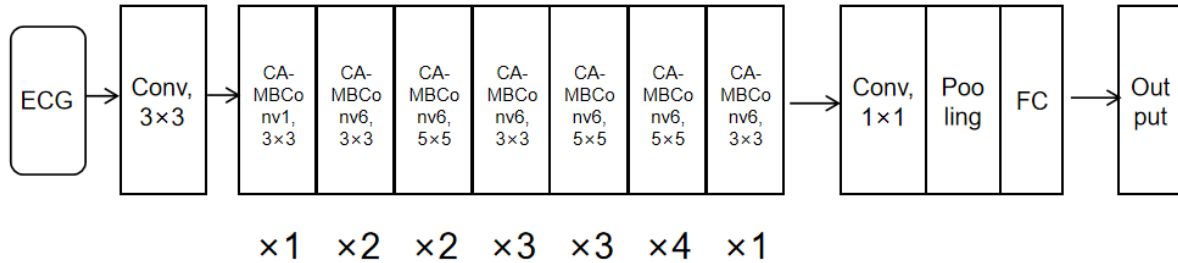


Figure 12. Topology of Ca-EfficientNet.

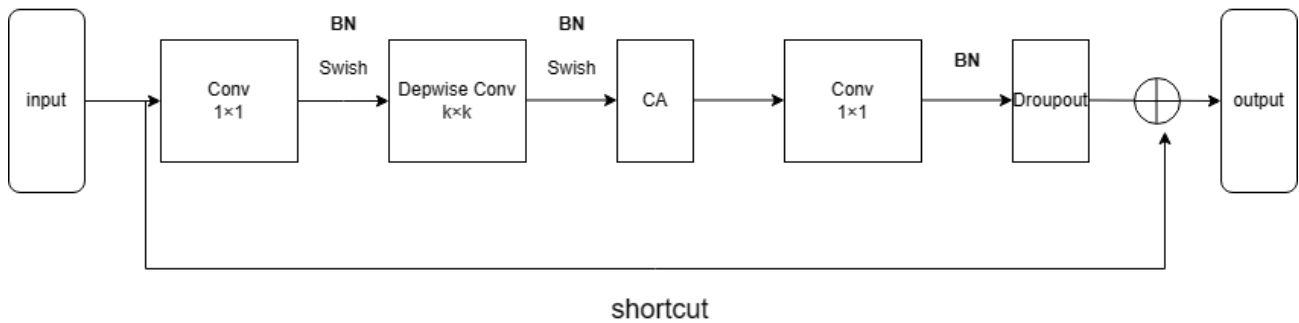


Figure 13. Architecture of CA-MBConv.

After introducing the CA mechanism, the fine-tuning strategy is adopted. Transfer learning approach involves loading the pre-trained model and conducting supervised fine-tuning on a fresh classification task. This approach makes full use of feature learning capability of the pre-trained model, leading to reduced training time and sample requirements. Additionally, it contributes to enhance the generalization performance of modeling, representing the effectiveness and superiority.

5.3. Implementation of classification with Ca-Efficientnet modeling

Choosing Efficientnet with an input size of 224×224 as the preferred network, the model is proceeded with fine-tuning. This involves transforming the MBConv modules with the CA mechanism modules. Moreover, through the concept of transfer learning, the pre-trained weights of EfficientNet is applied in transition to the enhanced ECG signal dataset. Through the visualization results in Figure 14, it is indicated that the loss value tends to stabilize after the 20th epoch. Additionally, the precision rate of the validation is observed to be steady state around the 22th epoch. These findings suggest that the model demonstrates excellent generalization.

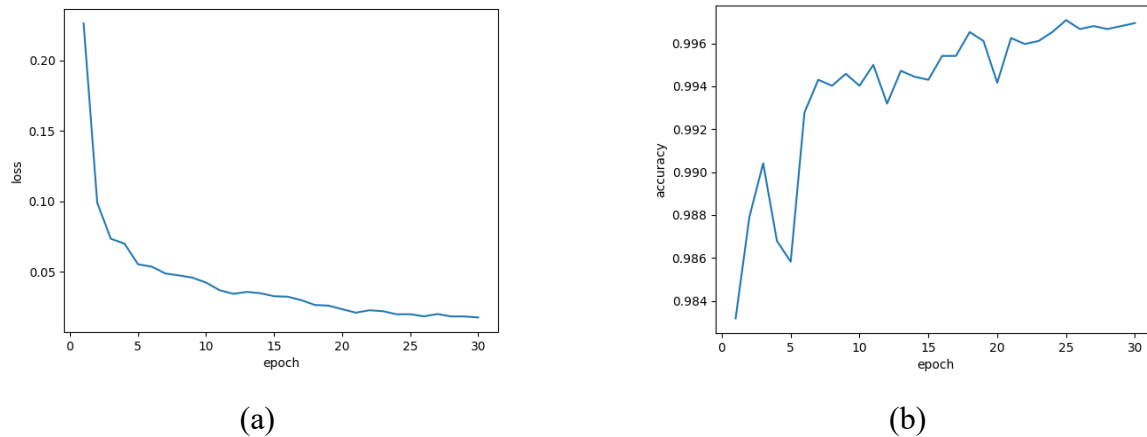


Figure 14. Modeling process of Ca-EfficientNet training: (a) Loss variation; and (b) accuracy variation.

The optimal weights performed on the test set are identified in the parameter selection approach. The weights in the 25th epoch are determined as the optimal parameters. The corresponding confusion matrix is then illustrated in Figure 15. As outlined in Table 5, the Ca-EfficientNet model achieves an excellent accuracy up to 99.70%, surpassing the performance of the Gam-Resnet18 model. The proposed ECG modeling with Ca-EfficientNet is compared with other network techniques. A transfer learning model with AlexNet [19] transformed ECG signals into grayscale images, achieving an accuracy of 94.95%. Similarly, a support vector machine (SVM) model [20] was employed with particle swarm optimization and achieved an accuracy rate of 98.57% for five-class classification. One-dimensional modeling [21] of convolution neural network (CNN) was built with classification accuracy of 98.10% in a combination of long short-term memory (LSTM). Additionally, Kandala [22] proposed a feature extraction method, named ensemble empirical mode decomposition (EEMD). Subsequently, it was integrated with sequential minimal optimization support vector machines (SMO-SVM) for classification. Compared to the mentioned methods, the combination of the proposed Ca-EfficientNet modeling with the relative position matrix algorithm is of the opinion that it produces a satisfied classification performance, confirming its superiority for ECG signal classification modeling.

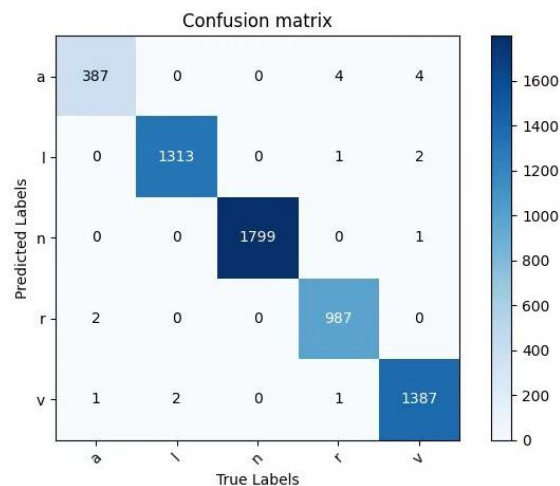


Figure 15. Confusion matrix of 5-class classification results of Ca-EfficientNet.

Table 5. Comparison of results with related literature.

Method	Type	ACC	SE	SP	PPV
AlexNet-like+Grayscale Image[19]	5	94.95%	-	-	-
LibSVM [20]	5	98.57%	-	-	-
CNN+LSTM [21]	5	98.10%	97.50%	98.70%	-
EEMD+SMO-SVM [22]	5	99.20%	98.01%	99.49%	-
RPM+Gam-Resnet18 [2]	5	99.30%	98.90%	99.84%	98.76%
TimeGAN+RPM+Gam-Resnet18	5	99.51%	99.34%	99.88%	99.44%
TimeGAN+RPM+EfficientNet	5	99.62%	99.41%	99.44%	99.47%
TimeGAN+RPM+Ca-EfficientNet	5	99.70%	99.55%	99.92%	99.45%

6. Discussions

To examine the superiority of the proposed approach with the TimeGAN+RPM+Ca-EfficientNet modeling, the Wilcoxon signed-rank test [23] is employed. Specifically, the proposed model is compared with the RPM+Gam-Resnet18 model [2], setting the significance level at $\alpha = 0.05$. The null hypothesis assumes no significant difference between the two models in ECG signal classification, with rejection contingent on a p-value below 0.05.

By the comparison of the proposed model and contrast model [2], the subsequent statistical test yielded a p-value of 0.001414. This result indicates a significant improvement for ECG signal classification, meeting the significance test criteria. Therefore, the TimeGAN+RPM+Ca-EfficientNet model demonstrates its capability to achieve satisfactory results in the classification and recognition of ECG signals.

In artificial intelligence learning, optimization, and enhancement of modeling, there are constant concerns. The limitations of the TimeGAN+RPM+Ca-EfficientNet model primarily stem from the challenge of parameter tuning. Integrating multiple modules may necessitate adjusting numerous hyperparameters, demanding heightened expertise and computational resources. The interpretability of complex models poses a challenge in elucidating their decision-making process, particularly in medical applications. While applying diverse techniques to optimize the performance and generalization ability, it can be explored in the following aspects.

1) Improvement of the model structure. The design of deep learning network has a significant impact on its performance and efficiency. To boost the efficiency and efficacy of modeling, it can explore more advanced and effective model structures, such as Adaptive CNN and Depthwise Separable CNN.

2) Data augmentation and transfer learning. Data augmentation employment is an effective way to improve the generalization ability and accuracy of modeling. In the next step, it can further investigate different data augmentation methods and transfer learning techniques by being applied to image classifications and visual tasks.

3) Knowledge distillation and model compression. There are represented as important ways of optimizing deep learning models by knowledge distillation and model compression. They can be approached with knowledge distillation, network pruning, and weight sharing to compress and refine deep learning models, ultimately enhancing efficiency and generalization ability.

4) Real-time Optimization. This streamlines the network structure for healthcare applications' real-time demands, enhancing computational efficiency for practical, time-sensitive requirements.

5) Explanatory Enhancement. This is used to investigate improvements in model interpretability, incorporating techniques like interpretability or model fusion strategies to elucidate the decision-making process more comprehensively.

In summary, the deep learning technology presents potential applications for areas like language processing, speech recognition and computer vision. However, it also faces challenges and difficulties. In the future, it continues to be developed and improves the performance and generalization ability of modeling, applying to more practical scenarios.

7. Conclusions

The TimeGAN data augmentation with the artificial bee colony optimization and the CA-EfficientNet modeling for the ECG signal image classification is introduced in this research. The goal is to enhance the extraction ability and classification accuracy in ECG signals through data augmentation, deepening and widening the neural network and reducing model parameters. The major conclusions are summarized as follows.

1. To enhance model performance, our results underscore the role of data augmentation in improving generalization and accuracy. The TimeGAN network, a data augmentation technique, is employed for ECG data augmentation to address the data imbalance problem. Furthermore, the artificial bee colony algorithm is utilized for hyperparameter selection, optimizing the quality of the generated data.

2. Trained with the balanced ECG data generated by TimeGAN, a notable accuracy of 99.51% is achieved with the combination of Gam-ResNet18 modeling. This highlights the benefits of data balancing, providing a robust support for model performance improvement, along with the efficacy of the Gam-ResNet18 network.

3. For further optimization on Gam-ResNet18 modeling, the CA-EfficientNet algorithm is applied with a local channel attention mechanism. The optimal weights are determined through the model selection method. An excellent accuracy up to 99.70% is achieved with Ca-EfficientNet modeling. The results demonstrate that applying the CA-EfficientNet significantly enhances the model performance compared to the Gam-ResNet18 network. This validates the superiority and potential of the CA-EfficientNet in image classification, offering novel insights for optimization and improvement of deep learning modeling.

In summary, we investigate and explore the application of deep learning modeling with TimeGAN and CA-EfficientNet in ECG image classification. The results indicate the excellent performance of the CA-EfficientNet modeling, providing new perspectives for model optimization and enhancement.

Use of AI tools declaration

The authors declare that we have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare that there are no conflicts of interest.

References

1. R. G. Kumar, Y. S. Kumaraswamy, Investigating cardiac arrhythmia in ECG using random forest classification, *J. Int. J. Computer Appl.*, **37** (2012), 31–34. <https://doi.org/10.5120/4599-6557>
2. M. Zhang, H. Jin, B. Zheng, W. Luo, Deep learning modeling of cardiac arrhythmia classification on information feature fusion image with attention mechanism, *Entropy*, **25** (2023), 1264. <https://doi.org/10.3390/e25091264>
3. Q. Qin, J. Li, L. Zhang, C.Y. Liu, Combining low-dimensional wavelet features and support vector machine for arrhythmia beat classification, *Sci. Rep.*, **7** (2017), 6067. <https://doi.org/10.1038/s41598-017-06596-z>
4. C. U. Kumari, A. S. D. Murthy, B. L. Prasanna, M. P. P. Reddy, A. K. Panigrahy, An automated detection of heart arrhythmias using machine learning technique: SVM, *Mater. Today Proceed.*, **45** (2021), 1393–1398. <https://doi.org/10.1016/j.matpr.2020.07.088>
5. M. R. Ekta, R. Devi, Arrhythmia discrimination using support vector machine, in *2017 4th International Conference on Signal Processing, Computing and Control (ISPCC)*, 2017, 283–287. <https://doi.org/10.1109/ISPCC.2017.8269690>
6. Ö. Yıldırım, P. Pławiak, R. S. Tan, U. Rajendra Acharya, Arrhythmia detection using deep convolutional neural network with long duration ECG signals, *J. Comput. Biol. Med.*, **102** (2018), 411–420. <https://doi.org/10.1016/j.compbimed.2018.09.009>
7. U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, R. S. Tan, Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals, *Appl. Intell.*, **49** (2019), 16–27.
8. X. Fan, Q. Yao, Y. Cai, F. Miao, F. Sun, Y. Li, Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ECG recordings, *IEEE J. Biomed. Health Inform.*, **22** (2018), 1744–1753. <https://doi.org/10.1007/s10489-018-1179-1>
9. U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, R. San Tan, A deep convolutional neural network model to classify heartbeats, *Comput. Biol. Med.*, **89** (2017), 389–396. <https://doi.org/10.1016/j.compbimed.2017.08.022>
10. J. Liu, M. Fu, S. Zhang, Application of convolutional neural network in automatic classification of arrhythmia, in *Proceedings of the ACM Turing Celebration Conference-China*, (2019), 1–8. <https://doi.org/10.1145/3321408.3326660>
11. M. Porumb, E. Iadanza, S. Massaro, L. Pecchia, A convolutional neural network approach to detect congestive heart failure, *J. Biomed. Signal Process. Control*, **55** (2020), 101–597. <https://doi.org/10.1016/j.bspc.2019.101597>
12. H. Sun, *Research on Automatic Detection Algorithm of Atrial Fibrillation Based on Feature Fusion*, Ph.D thesis, Shandong University in ShanDong, China, 2021.
13. T. Zheng, *Research on ECG Data Augmentation Method Based on Generative Adversarial Networks*, Ph.D thesis, Jiangxi University of Finance and Economics in Jiangxi, China, 2021.
14. Y. Wang, *Research on ECG Data Augmentation Algorithm Based on Generative Adversarial Neural Network*, Ph.D thesis, Beijing University of Posts and Telecommunications in Beijing, China, 2020.
15. P. 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. <https://doi.org/10.1109/ACCESS.2019.2930882>

16. J. Yoon, D. Jarrett, M. Van der Schaar, Time-series generative adversarial networks, *Adv. Neural Inform. Process. Syst.*, **32** (2019). <https://dl.acm.org/doi/abs/10.5555/3454287.3454781>
17. M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in *Proceedings of the International conference on machine learning*, (2019), 6105–6114. <https://doi.org/10.48550/arXiv.1905.11946>
18. Q. B. Hou, D. Q. Zhou, J. S. Feng, Coordinate attention for efficient mobile network design, in *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2021), 13708–13717. <https://doi.org/10.1109/CVPR46437.2021.01350>
19. X. F. Zha, F. Yang, Y. N. Wu, Y. Liu, S. F. Yuan, ECG classification based on transfer learning and deep convolution neural network, *Chin. J. Med. Phys*, **35** (2018), 1307–1312. <https://doi.org/10.3969/j.issn.1005-202X.2018.11.013>
20. J. Wang, M. Shi, X. Zhang, Research on classification of arrhythmia based on EMD and ApEn feature extraction, *J. Instrum. Meas.*, **37** (2016), 168–173.
21. S. L. Oh, E. Y. Ng, S. T. Ru, A. U. R, Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats, *Comput. Biol. Med.*, **102** (2018), 278–287. <https://doi.org/10.1016/j.combiomed.2018.06.002>
22. K. N. V. P. Rajesh, R. Dhuli, Classification of ECG heartbeats using nonlinear decomposition methods and support vector machine, *Comput. Biol. Med.*, **87** (2017), 271–284. <https://doi.org/10.1016/j.combiomed.2017.06.006>
23. D. Li, M. Jiang, M. Li, W. H, R. Xu, A floating offshore platform motion forecasting approach based on EEMD hybrid ConvLSTM and chaotic quantum ALO, *Appl. Soft Comput.*, **144** (2023), 110–487. <https://doi.org/10.1016/j.asoc.2023.110487>



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)