ECG Generation Based on Denoising Diffusion Probabilistic Models

Zhongyu Wang[1], Caiyun Ma[1], Minghui Zhao[1], Shuo Zhang[1], Jianqing Li[1*], Chengyu Liu[1*]

[1] State Key Laboratory of Digital Medical Engineering, School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China

Abstract

Arrhythmia diseases seriously damage people's life and health, and identifying abnormal points in ECG signals by deep neural networks is an effective method for detecting arrhythmias. However, their accuracy is often limited by the biased data distribution of the training set, and a large number of labeled ECG signals are usually harder to obtain. Therefore, this paper proposes to synthesize virtual heart beat data by denoising diffusion probability model (DDPM) based on the MIT-BIH arrhythmia database to complement the real data. Three different methods for generating heartbeat signals are also used, which are (i) generating heartbeat signals directly, (ii) generating time-frequency maps of heartbeats and transforming them into heartbeat signals, and (iii) generating sub-signals of heartbeats and fusing them into complete heartbeat signals. Regarding the evaluation of the synthesized signals, we compare the advantages and disadvantages of the three heartbeat generation methods by four metrics: DTW, PCC, ED and KLD. The experimental results showed that the optimum values of 4.37, 17.09, 0.972 and 0.0094 were obtained for ED, DTW of method (i) and PCC, KLD of method (iii), respectively.

1. Introduction

With the changes in lifestyle and the influence of environmental factors, cardiovascular disease has gradually become one of the major health challenges in today's society. Traditionally, the diagnosis and treatment of cardiovascular diseases have mainly relied on doctors' experience and analysis methods based on traditional medical knowledge. However, this method is often limited by subjective factors and limitations and cannot meet the needs for rapid and accurate diagnosis. In recent years, with the continuous development and application of deep learning technology, the application of artificial intelligence in the medical field has shown increasingly broad prospects [1, 2]. Deep learning can learn feature representations from large amounts of medical data and plays an important role in medical image analysis, disease diagnosis and prognosis prediction [3, 4]. Nonetheless, there are still some challenges with deep learning models in cardiovascular diseases.

In the diagnosis of arrhythmia diseases, the model learns the morphological characteristics of time series signals from ECG data and converts them into latent representations for diagnostic classification. Robust disease diagnosis models rely on large-scale class-balanced training data to extract generalized knowledge representations. However, the data sets used for training are usually highly imbalanced, that is, there are far more normal heart beats than abnormal heart beats. Data augmentation enriches the training data distribution by synthesizing new samples, which is an effective way to solve data imbalance. Common data generation methods include Generative Adversarial Network (GAN) [5] and variational autoencoder (VAE) [6].

At present, there have been some studies using GAN to synthesize electrocardiograms. Wulan et al. [7] proposed SpectroGAN and WaveletGAN. SpectroGAN captures the detailed frequency domain characteristics of the ECG signal by synthesizing the time-frequency diagram of the ECG signal and using the inverse Fourier transform to convert the time-frequency diagram into a signal. WaveletGAN uses wavelets to decompose signals into multiple frequency domain sub-bands, allowing the generator to synthesize sub-signals of different frequency bands, and finally fuse the sub-signals to obtain a complete ECG. Golany et al. [8] used a convolutional network as a discriminator to capture the spatial location information of the ECG signal, and used a deconvolution network as a generator to synthesize the heartbeat signal. However, there are two important problems in the training of GAN. One is that GAN easily falls into a fixed pattern of training data, that is, it can only generate fixed types of samples. The other is that the dynamic adversarial training process of GAN is unstable and prone to gradient disappearance and gradient explosion [9].

As a new paradigm of generative modeling, the diffusion model effectively copes with the above problems. The diffusion model [10] destroys the real samples to obtain a priori Gaussian distribution by continuously adding Gaussian noise to the data, and predicts the posterior distribu-

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.027

tion by predicting the noise in the samples in order to make the results generated by the model approximate to the real data, so as to achieve the purpose of learning the distribution of the target data.

In this work, we synthesize heartbeat signals using a diffusion probability model. Specifically, this paper has the following contributions: (1) different types of heartbeat signals are synthesized using denoising diffusion models, (2) heartbeats are generated in three ways: by generating the signal directly, by generating a time-frequency map and transforming it into a signal, and by generating subsignals and fusing them into a complete signal, and (3) several methods are used to evaluate the authenticity of the synthesized signals.

2. Method

2.1. Database

This work used the MIT-BIH Arrhythmia Database (MIT-BIHA). The MIT-BIHA contains 48 dual-lead arrhythmia ECG recordings of half an hour's duration, sampled at a frequency of 360 Hz. Each of the recordings in the database is annotated with the location of the R-peak, the type of heartbeat and an expert annotation for the diagnosis of the type of arrhythmia. The MLII lead of the MIT-BIHA was used to acquire individual heart beat signals. Of all the records, four (#102, #104, #107, and #217) were excluded due to the inclusion of paced heartbeats. In this paper, we synthesize the types of heart beats including normal heartbeats, left bundle branch block (LBBB), right bundle branch block (RBBB) and premature ventricular contractions (PVC) heartbeats (hereafter referred to as N, L, R, V).

2.2. Preprocessing

To ensure that the training data were not disturbed by noise, we filtered the ECG signals using a 0.5-45 Hz bandpass filter. Subsequently, Second, in order to prevent overlapping of the sampled heartbeats, we selected 100 sampling points forward and 150 sampling points backward of the R peak as a complete heartbeat based on the labeling information of the R peak position in the recordings, and selected four types (N, L, R, V) of heartbeats as the training data. of the model, we up-sampled the heartbeat signal to 256 samples by nearest neighbor interpolation. At the same time, we found that the number of different types of heart beats in the training set was highly unbalanced, i.e., more normal heart beats and fewer abnormal heart beats. We randomly intercepted some of the normal heartbeats in the training set and doubled the number of abnormal heartbeats to balance the dataset. The details of the training dataset are shown in Table 1.

Table 1. Training dataset detials.

Type of heartbeat	Number
Normal (N)	20000
Left bundle branch block (L)	16144
Right bundle branch block (R)	14510
Premature ventricular contractions (V)	13804
Total	64458

2.3. Diffusion Probabilistic Heartbeat Generation

Diffusion modeling is a paradigm of generative modeling. Its training consists of two stages: forward diffusion process and backward reverse diffusion process. In this case, the forward diffusion process corrupts the training data by continuously adding Gaussian noise to the data, and then, the target data distribution is learned by removing the noise from the data in the backward diffusion process. More specifically, the diffusion model approximates the model distribution $p_\theta(x)$ to the real data distribution $q(x)$ by predicting Gaussian noise in the data. In the inference stage, the sample data can be sampled from the Gaussian noise by iterative denoising.

Forward diffusion gradually adds Gaussian noise to the target distribution $q(x_0)$ by means of a Markov chain to obtain the noise distribution $q(x_T)$. Then the distribution $q(x_t)$ at each time step can be calculated by the following equation:

$$q(x_{1:T} \mid x_0) := \prod_{t=1}^{T} q(x_t|x_{t-1}) \tag{1}$$

$$q(x_t \mid x_{t-1}) := \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbf{I}) \tag{2}$$

where β_t is a fixed variance schedule whose value is linearly related to time step. When T is large enough, $q(x_T)$ is an isotropic Gaussian distribution. Let $\alpha_t := 1-\beta_t$, and $\overline{\alpha}_t := \prod_{s=1}^{t}\alpha_s$, then the prior distribution $q(x_t \mid x_0)$ for an arbitrary time step can be expressed as follows:

$$q(x_t \mid x_0) := \mathcal{N}(x_t; \sqrt{\overline{\alpha}_t}x_0, (1-\overline{\alpha}_t)\mathbf{I}) \tag{3}$$

then x_t can be expressed as:

$$x_t = \sqrt{\overline{\alpha}_t}x_0 + (1-\overline{\alpha}_t)\epsilon \tag{4}$$

where $\epsilon \sim \mathcal{N}(0, \mathbf{I})$.

The *reverse diffusion* process samples x_0 from the Gaussian noise data x_T in a Markov chain fashion. Let $x_T \sim \mathcal{N}(0, \mathbf{I})$, then the reverse diffusion process can be expressed as:

$$p_\theta(x_{0:T}) := p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1} \mid x_t) \tag{5}$$

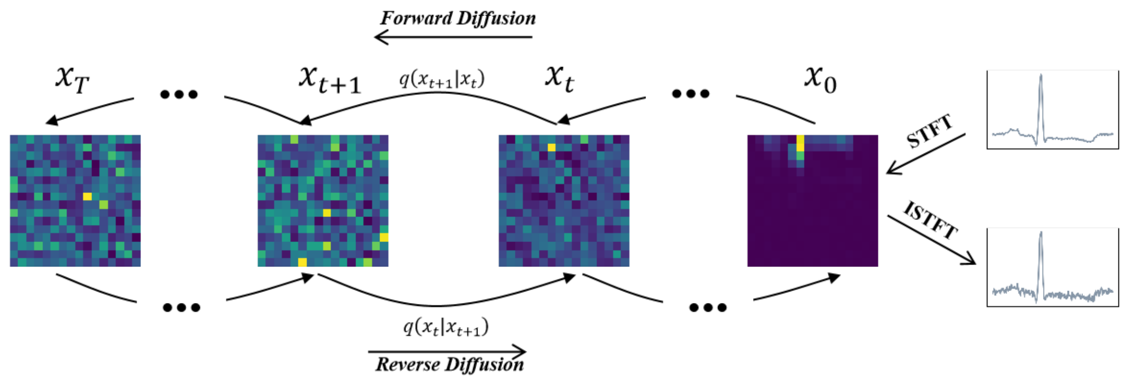

Figure 1. *Forward* and *reverse* diffusion procedures for generating heartbeat signals via time-frequency map diffusion.

$$p_\theta(x_{t-1} \mid x_t) := \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sigma_\theta(x_t, t)\mathbf{I}) \quad (6)$$

where,

$$\sigma_\theta(x_t, t) = \frac{1-\overline{\alpha}_{t-1}}{1-\overline{\alpha}_t}\beta_t \quad (7)$$

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(x_t, t)) \quad (8)$$

where $\epsilon_\theta(x_t, t)$ is the noise at time step t predicted by the model. With this Markov chain approach, it is possible to sample an approximate target data distribution from a pure Gaussian distribution by predicting the noise. Then the objective loss function can be expressed as:

$$\mathcal{L} = \|\epsilon - \epsilon_\theta(\sqrt{\overline{\alpha}_t}x_0 + \sqrt{1-\overline{\alpha}_t}\epsilon, t)\|^2 \quad (9)$$

In this work, Attention U-Net [11] was used as the backbone network and three ECG diffusion generation methods were employed. (i) **Signal-diffusion**: direct generation of the heartbeat timing signal. (ii) **Spectro-diffusion**: conversion of the heartbeat signal into a time-frequency map by short-time Fourier transform (STFT), generation of the time-frequency map of the heartbeat signal by using diffusion model, and finally reconstruction of the heartbeat signal by using the Griffin-Lim algorithm [12]. (iii) **Wavelet-diffusion**: decomposition of the signal into four sub-signals according to the frequency bands by wavelet transform to generate the four sub-signals individually and fusing the sub-signals into the heartbeat signal.

2.4. Evaluation method

In this paper, we visually evaluate the model's ability to learn the target data distribution by showing synthetic heartbeats. Meanwhile, we measure the data fidelity of synthetic heartbeats by calculating the dynamic template matching distance (DTW), euclidean distance (ED), pearson correlation coefficient (PCC), and KL divergence (KLD) for the average synthetic heartbeats and the average original heartbeats.

2.5. Experimental settings

In the three training approaches of the diffusion model, the learning rate was fixed at 0.0001, T was designated as 1000, and β_t was assigned a linear uniform distribution ranging from 0.0001 to 0.02. Regarding Spectro-diffusion, the window length of the STFT and the number of points of the FFT were both set to 30, and the amount of overlap between neighboring windows was set to 12. Regarding Wavelet-diffusion, we use a 'db2' wavelet basis to perform a 2-level decomposition of the heartbeat signal.

3. Results

Figure 2 demonstrates the comparison of the synthesized heartbeat signals of different modes with the original heartbeat signals using Signal-diffusion. The blue line in the figure indicates the original signal and the green line indicates the synthesized signal. The results show that the synthesized heartbeat is basically consistent with the real heartbeat in terms of important waveform morphology and rhythm location information.

In addition to this, we compare the synthesis effects of three different heartbeat synthesis methods by four evaluation metrics, and the results are shown in Table 2. As can be seen from the table, Signal-diffusion achieved optimal values of 4.37 and 17.09 for ED and DTW, while Wavelet-diffusion achieved optimal values of 0.972 and 0.0094 for PCC and KLD by capturing the multi-scale waveform morphology of the signal and the data distribution. Spectro-diffusion gives poor results due to the fact that the STFT and ISTFT processes suffer from information loss.

4. Discussion

The experimental results show that the diffusion model is able to learn the data distribution of the heartbeat signal by continuously removing the noise. Direct synthesis of heart beats is better than synthesizing time-frequency

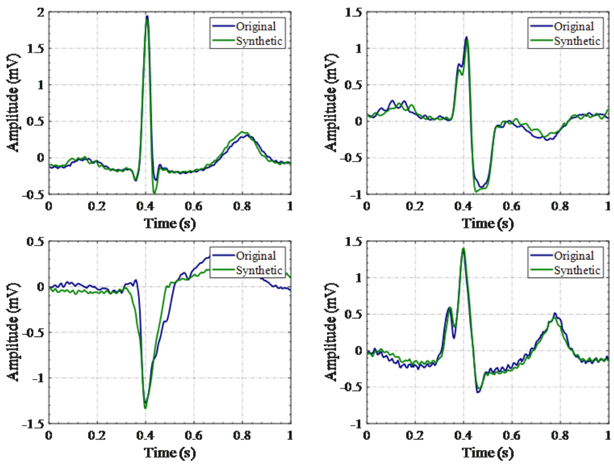

Figure 2. Comparison of various patterns of heartbeats in original data and data synthesized using **Signal-diffusion**. The blue lines indicate the original heartbeats, the green lines indicate the synthetic heartbeats.

Table 2. Comparison of the fidelity of heartbeat synthesis.

Method	**ED↓**	**DTW↓**	**PCC↑**	**KLD↓**
Signal-diffusion	**4.37**	**17.09**	0.969	0.0234
Spectro-diffusion	9.08	74.83	0.643	0.0795
Wavelet-diffusion	4.39	20.93	**0.972**	**0.0094**

maps and transforming them, which may be due to the loss of information in the time-frequency transformation process. Meanwhile, synthesizing multi-scale sub-signals and fusing them also achieves better results, and this effect may be enhanced with the increase of signal scale. Synthesizing ECG signals through diffusion modeling may become a new direction to enrich the distribution of ECG data and solve the data privacy problem.

Acknowledgments

This work was funded by the National Natural Science Foundation of China (62171123, 62201144), and the National Key Research and Development Program of China (2023YFC3603600, 2022YFC2405600).

References

[1] Hammad M, Iliyasu AM, Subasi A, Ho ESL, Abd El-Latif AA. A multitier deep learning model for arrhythmia detection. IEEE Transactions on Instrumentation and Measurement 2021;70. ISSN 0018-9456.

[2] Ma C, Wei S, Chen T, Zhong J, Liu Z, Liu C. Integration of results from convolutional neural network in a support vector machine for the detection of atrial fibrillation. IEEE Transactions on Instrumentation and Measurement 2021; 70. ISSN 0018-9456.

[3] Chen X, Wang X, Zhang K, Fung KM, Thai TC, Moore K, Mannel RS, Liu H, Zheng B, Qiu Y. Recent advances and clinical applications of deep learning in medical image analysis. Medical Image Analysis JUL 2022;79. ISSN 1361-8415.

[4] Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson V J, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Medicine SEP 27 2021;13(1). ISSN 1756-994X.

[5] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. In Ghahramani Z, Welling M, Cortes C, Lawrence N, Weinberger K (eds.), Advances in Neural Information Processing Systems 27 (NIPS 2014), volume 27 of Advances in Neural Information Processing Systems. ISSN 1049-5258, 2014; 2672–2680. 28th Conference on Neural Information Processing Systems (NIPS), Montreal, CANADA, DEC 08-13, 2014.

[6] Kingma D, Welling M. Auto-encoding variational bayes. arXiv Machine Learning Dec 2013;.

[7] Wulan N, Wang W, Sun P, Wang K, Xia Y, Zhang H. Generating electrocardiogram signals by deep learning. Neurocomputing SEP 3 2020;404:122–136. ISSN 0925-2312.

[8] Golany T, Lavee G, Yarden ST, Radinsky K. Improving ecg classification using generative adversarial networks. volume 34 of AAAI Conference on Artificial Intelligence. Assoc Advancement Artificial Intelligence. ISBN 978-1-57735-835-0. ISSN 2159-5399, 2020; 13280–13285. 34th AAAI Conference on Artificial Intelligence / 32nd Innovative Applications of Artificial Intelligence Conference / 10th AAAI Symposium on Educational Advances in Artificial Intelligence, New York, NY, FEB 07-12, 2020.

[9] Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X. Improved techniques for training gans. Cornell University arXiv Jun 2016;.

[10] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. Neural Information Processing Systems Jan 2020;.

[11] Oktay O, Schlemper J, Folgoc L, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla N, Kainz B, Glocker B, Rueckert D. Attention u-net: Learning where to look for the pancreas. arXiv Computer Vision and Pattern Recognition Apr 2018;.

[12] Griffin D, Lim J. Signal estimation from modified short-time fourier transform. IEEE Transactions on Acoustics Speech and Signal Processing Apr 1984;236–243.

Address for correspondence:

Jianqing Li and Chengyu Liu
Sipailou Road 2, Southeast University, Nanjing, China
Email: ljq@seu.edu.cn and chengyu@seu.edu.cn