Variational autoencoder–based neural electrocardiogram synthesis trained by FEM-based heart simulator

Ryo Nishikimi, PhD,* Masahiro Nakano, MS,*† Kunio Kashino, PhD,*† Shingo Tsukada, MD, PhD†

*From the *NTT Communication Science Laboratories, Atsugi, Japan, and †NTT Basic Research Laboratories, Atsugi, Japan.*

BACKGROUND For comprehensive electrocardiogram (ECG) synthesis, a recent promising approach has been based on a heart model with physical and chemical cardiac parameters. However, the problem is that such approach requires a high-cost and limited environment using supercomputers owing to the massive computation.

OBJECTIVE The purpose of this study is to develop an efficient method for synthesizing 12-lead ECG signals from cardiac parameters.

METHODS The proposed method is based on a variational autoencoder (VAE). The encoder and decoder of the VAE are conditioned by the cardiac parameters so that it can model the relationship between the ECG signals and the cardiac parameters. The training data are produced by a comprehensive, finite element method (FEM)-based heart simulator. New ECG signals can then be synthesized by inputting the cardiac parameters into the trained VAE decoder without relying on enormous computational resources. We used 2 metrics to evaluate the quality of ECG signals synthesized by the proposed model.

RESULTS Experimental results showed that the proposed model synthesized adequate ECG signals while preserving empirically important feature points and the overall signal shapes. We also explored the optimal model by varying the number of layers and the size of latent variables in the proposed model that balances the model complexity and the simulation accuracy.

CONCLUSION The proposed method has the potential to become an alternative to computationally expensive FEM-based heart simulators. It is able to synthesize ECGs from various cardiac parameters within seconds on a personal laptop computer.

KEYWORDS Variational autoencoder; Electrocardiogram; Synthesis; Cardiac parameters; Finite element method

(Cardiovascular Digital Health Journal 2024;5:19–28)

Introduction

Electrocardiograms (ECG) are graphical representations of the electric potential changes generated by the excitation of the myocardial cells of the heart detected using electrodes placed at specific locations on the body surface. ECG is a common tool for monitoring cardiac health and detecting heart disease in medical practice.[1–8] Various tools and techniques to analyze ECG signals and quickly detect abnormalities have been devised over many decades in order to improve medical care. In recent years, advances in machine learning technologies such as deep neural networks (DNNs)[4,9,10] have led to rapid development in research that seeks to find useful clinical clues from real ECGs. Efforts to date using even limited amounts of properly annotated ECG data suggest that this could be a productive approach for early detection of abnormalities and diagnosis.[10–17] However, it is not easy for experts to manually add clinical insights to real ECGs from multiple perspectives, and this is a serious obstacle facing this AI-based research strategy. The acquisition, analysis, and labeling of electrocardiograms is extremely time consuming, requiring both specialized expertise and specialized equipment. In addition, using actual ECGs from medical patients requires strict compliance with applicable laws and regulations covering data privacy.

In this context, there is a need for new approaches that do not rely solely on the analysis of real ECGs by human experts with specialized knowledge about these ECGs.

Over the years, attempts have been made to synthesize and simulate ECGs on a computer as an alternative to relying solely on actual ECGs from patients interpreted by clinical experts. Conventional approaches to synthesizing ECG signals are based on the dynamic models described by ordinary

Address reprint requests and correspondence: Dr Ryo Nishikimi, NTT Communication Science Laboratories, 3-1, Morinosato Wakamiya, Atsugi-shi, Kanagawa Pref., 243-0198, Japan. E-mail address: ryo.nishikimi@ntt.com.

2666-6936/

https://doi.org/10.1016/j.cvdhj.2023.12.002

differential equations.[18,19] Newer DNN-based approaches to synthesizing ECGs have also emerged more recently which focus on the waveform shapes (morphological patterns) of ECG signals. As such, EGG signals synthesized by these approaches are linked only to shape features representing morphological patterns (eg, timing and voltage of R positions).[9,18,20]

There are also methods that can synthesize ECG signals from physical and chemical parameters, including electrochemical parameters at the cellular level.[21] In this paper, we refer to such parameters as *cardiac parameters* and use the cardiac parameters listed in Table 1. ECG synthesis from cardiac parameters is expected not only to more accurately reflect the relationship between an ECG and the internal state of the human heart, but also to help better understand clinical states, because it is more consistent with the underlying mechanisms of the ECG signal generation and propagation process than ECG synthesis derived only from ECG appearances.

The ability to easily synthesize ECG signals from cardiac parameters is expected to provide important clues for early detection of heart disease and screening treatment strategies for individual patients. A straightforward approach, for example, is the compilation of an encyclopedia of electrocardiograms for each patient using synthetic ECGs for a variety of cardiac parameters in advance with reference to the case histories. Such an encyclopedia would allow an actual ECG from the patient to be systematically searched for similar ECGs.

There have been promising efforts to use advanced heart simulators based on the finite element method (FEM) to computationally synthesize ECGs. While capable of principled synthesis, the FEM-based heart simulators are too computationally expensive to be a practical solution for applications, such as creating a comprehensive encyclopedia of ECGs. For example, the UT-Heart[21] simulator takes about 350 node-hours[1] on the supercomputer Fugaku[2] to synthesize a single-beat 12-lead ECG. In addition, when the cardiac simulation is used to analyze complex multifactorial pathologies such as heart failure, it is difficult to use computationally expensive methods owing to the combinatorial explosion of factors.

Against this background, the goal of this study is to construct a fast method that simulates the mechanism of the FEM-based method and synthesizes ECGs having the same quality as those obtained by the FEM-based method. That is, this method is intended not to improve the accuracy of the simulation of the original FEM-based method, but to significantly increase its speed while maintaining as much accuracy as possible. A fast and high-quality ECG synthesis method has potential to discover new clinical knowledge and insight. For example, a fast ECG simulator may generate many reference ECGs by interpolating and extrapolating the cardiac parameters representing the various heart conditions. These generated ECGs may be used, for example, in medical practice, for training other machine learning models, and for creating an ECG encyclopedia by accumulating them. The potential applications and challenges to use of a fast, parametric ECG synthesis method are further discussed in the Discussion section.

To achieve this goal, we propose a DNN-based method to accelerate the speed of ECG generation by simulating the process of the FEM-based method. In this paper, we adopt a conditional variational autoencoder (VAE) architecture for representing the cardiac parameters and the other factors as conditional inputs and latent variables in the VAE. This study is the first report to point to a method for replicating an elaborate physical/electrical heart simulator for ECG synthesis using a general-purpose machine learning model, greatly reducing the computational cost. Using the accuracy measures in terms of the peak signal-to-noise ratios and the mean absolute errors of Q, R, S, and T positions, we confirmed that the proposed method is equivalently accurate to reproduce the ECG generation compared with the FEM-based heart simulator using the same cardiac parameters. ECG synthesis, which currently requires slow and expensive computation of FEM-based heart simulators on a supercomputer, can therefore potentially be replaced by general-purpose machine learning tools running on a standard laptop.

Methods

Dataset

We used an open dataset of 12-lead ECG signals synthesized by the multi-scale multi-physics heart simulator "UT-Heart"[21] to train and evaluate the proposed model. In the dataset, UT-Heart takes as input 16 cardiac parameters, named SCon, LCon, I_{Na}, SERCA, I_{CaL}, CaRC, NCX, I_{Ks}, I_{Kr}, I_{to}, I_{K1}, LV, EX, CELL, HR, and CIR, major known factors associated with heart failure as listed in Table 1. Other factors that may affect ECG, such as body shape, sex, environment, and personality, are not considered. The detailed description of the cardiac parameters can be found on their official web page.[22] The dataset includes 2880 ECG variations, corresponding to the number of combinations of the above cardiac parameters.

In the information provided with the current UT-Heart data, the continuous value parameter is set at 100% after careful selection by the authors of the UT-Heart simulator of situations that are assumed to be healthy, and candidate values that may cause potential adverse effects such as disease are evaluated relative to each other. Such a setup is very reasonable, since UT-Heart has a vast number of parameters other than cardiac parameters that govern body shape, sex, etc; and since the absolute values themselves, which can be representative of the situation of a healthy person, are heavily influenced by factors other than those cardiac parameters. Thus, it is not easy to give a clinical interpretation to absolute values of cardiac parameters. Therefore, continuous values of

[1] *Node* means a management unit of a supercomputer, and *node-hour* is a unit of computation time equivalent to 1 hour of computation on 1 node.

[2] Fugaku is a supercomputer developed in Japan, and each node of the supercomputer Fugaku has 1 Fujitsu A64FX CPU with 48 cores.

Table 1 Cardiac parameters used as inputs for the proposed machine learning-based method and for the conventional finite element–based method

Name	Description of individual cardiac parameters[†]
SCon, LCon	Cardiac conduction velocities in the short- and long-axis directions of cardiomyocytes. The tuple of them takes 3 continuous values: (100%, 100%), (120%, 80%), or (100%, 80%).
I_{Na}	Conductance of sodium current. It takes 2 continuous values: 100% or 70%.
SERCA, I_{CaL}, CaRC	Amount of sarcoplasmic reticulum Ca^{2+} ATPace, the scale of L-type Ca^{2+} current, and the degree of phosphorylation. The tuple of them takes 2 continuous values: (100%, 100%, 100%) or (50%, 70%, 150%).
NCX	Activity of Na^{+}-Ca^{2+} exchanger current. It takes 2 continuous values: 100% or 150%.
I_{Ks}, I_{Kr}, I_{to}, I_{K1}	Slow and fast components of delayed rectifier outward potassium current, transient outward potassium current, and inward rectifier potassium current. The tuple of them takes 3 continuous values: (100%, 100%, 100%, 100%), (50%, 80%, 80%, 80%), or (30%, 50%, 30%, 50%).
LV	Sphericity of the left ventricular shape. It takes 2 discrete patterns corresponding to the cases SI = 0 · 53 and SI = 0· 68, where SI stands for sphericity index, defined by (sphericity index) = (heart long-axis length) / (heart short-axis length).
EX	Ventricular activation pattern. It takes 4 discrete patterns: normal, left axis deviation, right axis deviation, and mild delay in right ventricular activation.
CELL	Arrangement pattern of the 3 ventricular cell models: endocardial cell, M-cell, and epicardial cell. It takes 5 discrete patterns.
HR	Heart rate. It takes 1 discrete pattern indicating the heart rate of 60 beats per minute.
CIR	Circulation parameter that indicates either normal, heart failure, or diastolic dysfunction.[26] The current data contains only electrocardiograms of a normal case, and therefore, the circulation parameter was fixed to the normal pattern.

[†]A value of 100% for the continuous parameters means normal.

cardiac parameters in this study will use relative assessed values.

We preprocessed the ECGs and cardiac parameters for use in the proposed method. The amplitude of ECG signals was normalized to be between 0 and 1 by dividing it by 1024 before inputting them into the proposed model. The parameters that take continuous values were also normalized to be between 0 and 1 by dividing them by 200. Those 11 continuous parameters were represented as an 11-dimensional vector. Each discrete parameter was converted to a 1-hot vector, a common format for representing discrete inputs and outputs in DNNs, and inputted into the proposed model. That is, the LV, EX, CELL, HR, and CIR parameters were respectively converted to 2-, 4-, 5-, 1-, and 1-dimensional 1-hot vectors, depending on the number of discrete patterns that each discrete parameter can take. Therefore, an input vector for the proposed method comprises the 11 continuous parameters and the 1-hot vectors, and the dimension of the parameter vector given to the proposed model was 11 + 2 + 4 + 5 + 1 + 1 = 24. Note that our model can be trained even when the number of possible values that a specific parameter (eg, HR) can take increases for future extensibility.

Model

Our model synthesizes each of the 12-lead signals used for ECG from the cardiac parameters. For synthesis, we adopted a conditional VAE architecture (Figure 1). The conditional VAE represents variations inherent in the process of ECG generation using a Gaussian distribution, while the changes in the ECG owing to differences in the cardiac parameters are represented by the conditional vector. The former corresponds to the factors not included in the cardiac parameters, such as environment, physique, and personality. Note that DNNs that convert the cardiac parameters directly into ECGs cannot represent different ECGs from the same cardiac parameters, since their architecture assumes a 1-to-1 correspondence between the cardiac parameters and ECGs. In contrast, conditional VAE can represent the factors other than a given set of parameters as the Gaussian distribution in the latent space.

The VAE assumes that the ECG signal is stochastically generated from a D-dimensional latent variable $\mathbf{z}\in\mathbb{R}^D$, representing ECG characteristics, where $D\in\{2, 4, 8, 16, 32, 64, 128\}$. We vary the value of D to optimize the VAE model. Our model also assumes that ECG signals are generated from the cardiac parameters $Y\in\mathbb{R}^{24}$ to simulate the synthesis process of ECG signals by the UT-Heart. Letting $\mathbf{x}=\{x_1,\ldots,x_T\}\in\mathbb{R}^T$ be an ECG signal, where T is the number of samples per ECG cycle, we formulate the hierarchical generative model of the ECG signal $\mathbf{x}$ as follows:

$$p_\theta(\mathbf{x},\mathbf{z}|Y)=p_\theta(\mathbf{x}|\mathbf{z},Y)p(\mathbf{z}),$$

where $p(\mathbf{z})$ is the prior distribution of $\mathbf{z}$. $p_\theta(\mathbf{x}|\mathbf{z},Y)$ is a likelihood function of $\mathbf{x}$ and is represented by an arbitrary deep neural network called a decoder with a parameter set θ. In this paper, we formulate the generative model so that both $\mathbf{x}$ and $\mathbf{z}$ follow a standard Gaussian distribution, as follows:

$$p(\mathbf{z})=\prod_{d=1}^{D}p(z_d)=\prod_{d=1}^{D}\mathcal{N}(z_d|0,1),$$

$$p_\theta(\mathbf{x}|\mathbf{z},Y)=\prod_{t=1}^{T}p_\theta(x_t|\mathbf{z},Y)$$

$$=\prod_{t=1}^{T}\mathcal{N}\left(x_t\middle|\mu_{\theta,t}(\mathbf{z},Y),\sigma^2_{\theta,t}(\mathbf{z},Y)\right),$$

where $\mathcal{N}(\cdot|\mu,\sigma^2)$ represents a Gaussian distribution with a mean parameter μ and a variance parameter σ^2.

$\mu_{\theta,t}(\mathbf{z}, Y) \in \mathbb{R}$ and $\sigma^2_{\theta,t}(\mathbf{z}, Y) \in \mathbb{R}_+$ are the output of the decoder DNN with the parameter set θ.

We use a stack of L_{dec} fully connected (FC) layers as the decoder network, where $L_{\text{dec}} \in \{4, 5, 6, 7, 8, 9, 10\}$ is the number of the FC layers. A rectified linear unit layer follows each FC layer except the last one, and a hyperbolic tangent (tanh) layer follows the last FC layer. The input dimension of the decoder network is $D + 24$. The output dimension of the decoder network is $T \times 2$. The output dimension of each layer except the last layer is set so that the input and output dimensions of the decoder network are exponentially interpolated as follows:

$$D_i^{\text{dec}} = T \times \left(\frac{D}{T}\right)^{1-\frac{i}{L_{\text{dec}}}} (i = 1, \ldots, L_{\text{dec}} - 1),$$

where D_i^{dec} is the output dimension of i-th FC layers in the decoder network.

Training

The objective of VAE training is to find the parameter set θ by maximizing the log-marginal likelihood $\log p_\theta(\mathbf{x}|Y)$. Since calculating the log-marginal likelihood is intractable, we introduce an encoder network with a parameter set ϕ to approximately represent the variational posterior $q_\phi(\mathbf{z}|\mathbf{x}, Y)$. Although we can use $q_\phi(\mathbf{z}|\mathbf{x})$ as a variational posterior, we actively use Y as auxiliary information since Y is available. Then, we train the parameter sets θ and ϕ by maximizing the following evidence lower bound $\mathscr{L}(\theta, \phi)$:

$$\begin{aligned}
\log p_\theta(\mathbf{x}|Y) &= \log \int p_\theta(\mathbf{x}|\mathbf{z}, Y) p(\mathbf{z}) d\mathbf{z} \\
&= \log \int \frac{q_\phi(\mathbf{z}|\mathbf{x}, Y)}{q_\phi(\mathbf{z}|\mathbf{x}, Y)} p_\theta(\mathbf{x}|\mathbf{z}, Y) p(\mathbf{z}) d\mathbf{z} \\
&\geq \int q_\phi(\mathbf{z}|\mathbf{x}, Y) \log \frac{p_\phi(\mathbf{x}|\mathbf{z}, Y) p(\mathbf{z})}{q_\phi(\mathbf{z}|\mathbf{x}, Y)} dz \\
&= -D_{\text{KL}}\left(q_\phi(\mathbf{z}|\mathbf{x}, Y) \| p(\mathbf{z})\right) + E_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z}, Y)] \\
&= \mathrm{L}(\theta, \phi),
\end{aligned}$$

where $\mathscr{D}_{\text{KL}}(\cdot \| \cdot)$ is the Kullback-Leibler (KL) divergence and $\mathbb{E}[\cdot]$ is the expectation by the variational posterior $q_\phi(\mathbf{z}|\mathbf{x}, Y)$ given by

$$\begin{aligned}
q_\phi(\mathbf{z}|\mathbf{x}, Y) &= \prod_{d=1}^{D} q_\phi(z_d|\mathbf{x}, Y) \\
&= \prod_{d=1}^{D} \mathscr{N}\left(z_d \middle| \mu_{\phi,d}(\mathbf{x}, Y), \sigma^2_{\phi,d}(\mathbf{x}, Y)\right),
\end{aligned}$$

where $\mu_{\phi,d}(\mathbf{x}, Y) \in \mathbb{R}$ and $\sigma^2_{\phi,d}(\mathbf{x}, Y) \in \mathbb{R}_+$ are a mean parameter and a variance parameter that are the outputs of the encoder network. Since we provide the decoder with the latent variable $\mathbf{z}$ and the cardiac parameters Y, $\mathbf{z}$ should represent the ECG features that the parameters Y cannot describe. Therefore, we provide the encoder with the cardiac parameters Y as a hint for training the encoder.

The encoder network is a stack of L_{enc} FC layers like the decoder network, where $L_{\text{enc}} \in \{4, 5, 6, 7, 8, 9, 10\}$ is the number of the FC layers and equals L_{dec}. A rectified linear unit layer follows each FC layer except the last one. The input dimension of the encoder network is the summation of the length of an ECG signal and the size of the parameter vector, ie, $T + 24$. The output dimension of the decoder network is $D \times 2$, which corresponds to the number of the mean and variance variables of $\mathbf{z}$. The output dimension of each FC layer, except the last layer, is set so that the input and output dimensions of the encoder network are exponentially interpolated as follows:

$$D_i^{\text{enc}} = T \times \left(\frac{D}{T}\right)^{\frac{i}{L_{\text{enc}}}} (i = 1, \ldots, L_{\text{enc}} - 1),$$

where D_i^{enc} is the output dimension of the i-th FC layer in the encoder network.

Inference

A new ECG signal is obtained by inputting a latent variable $\mathbf{z} = \mathbf{0}$ and an unseen set of cardiac parameters $\tilde{Y}$ into the decoder of the trained VAE. Since the proposed method assumes that the latent variables represent the generative factors of ECG, including the cardiac parameters and the others, we can also use $\mathbf{z} \neq \mathbf{0}$ for synthesizing ECGs by the trained decoder. However, we simply use $\mathbf{z} = \mathbf{0}$ to obtain the ECG that is most likely to be synthesized from the given cardiac parameters. The mean vector $\tilde{\mathbf{x}} = \mu_\theta(\mathbf{0}, \tilde{Y}) = \{\mu_{\theta,1}(\mathbf{0}, \tilde{Y}), \ldots, \mu_{\theta,T}(\mathbf{0}, \tilde{Y})\}$ is considered as the synthesized ECG.

Experimental condition

We randomly split 2880 synthesized 12-lead ECG signals into a training dataset (1728 ECG signals), a validation dataset (576 ECG signals), and a test dataset (576 ECG signals). The lengths of individual synthesized signals were all $T = 500$. To search for the appropriate VAE model for the ECG signal generation, we varied the number of layers of the encoder and decoder networks $L_{\text{enc}} = L_{\text{dec}} \in \{4, 5, 6, 7, 8, 9, 10\}$ and the size of the latent variable $D \in \{2, 4, 8, 16, 32, 64, 128\}$. We optimized the proposed VAE using the Adam optimizer[23] with the parameters $\alpha = 0.001$ (learning rate), $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$. The weight and bias parameters of the FC layers of the encoder and decoder networks were initialized to random values between $-\sqrt{1/k}$ and $\sqrt{1/k}$, where k represents the input dimension of each FC layer. The batch size and the number of epochs were 1024 and 2000, respectively.

We evaluated the degree to which the artificial ECG synthesized by the proposed method can reproduce the original ECG signal using objective evaluation metrics. The

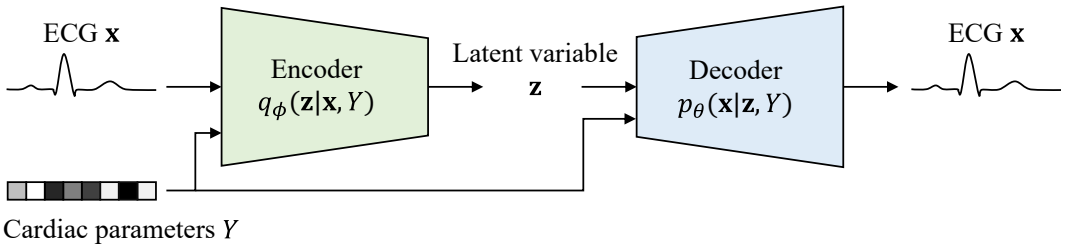

Figure 1 Overview of the proposed conditional variational autoencoder (VAE). The proposed method is based on a conditional VAE, which is composed of encoder and decoder neural networks. The encoder takes the electrocardiogram (ECG) **x** and cardiac parameters Y and outputs the distribution parameter of the latent variable **z**. The decoder takes the latent variable **z** and the cardiac parameters Y and outputs the ECG **x**.

appropriate evaluation criteria depend on the purpose. For example, in the context of synthesizing ECGs for arrhythmias, a metric that emphasizes the periodicity of the signal may be appropriate. In order to evaluate the general reproducibility as a waveform and the degree to which clinically important features are maintained, here we used the following 2 metrics: position mean absolute error (PMAE) and peak signal-to-noise ratio (PSNR).

PMAE represents the gaps of Q, R, S, and T positions of the ground-truth and generated ECG signals and is calculated as follows: $\mathrm{PMAE}=\frac{1}{N}\sum_{n=1}^{N}\left|t_n^{\mathrm{gt},*}-t_n^{\mathrm{es},*}\right|$, where N is the number of test data; "$*$" represents Q, R, S, or T;

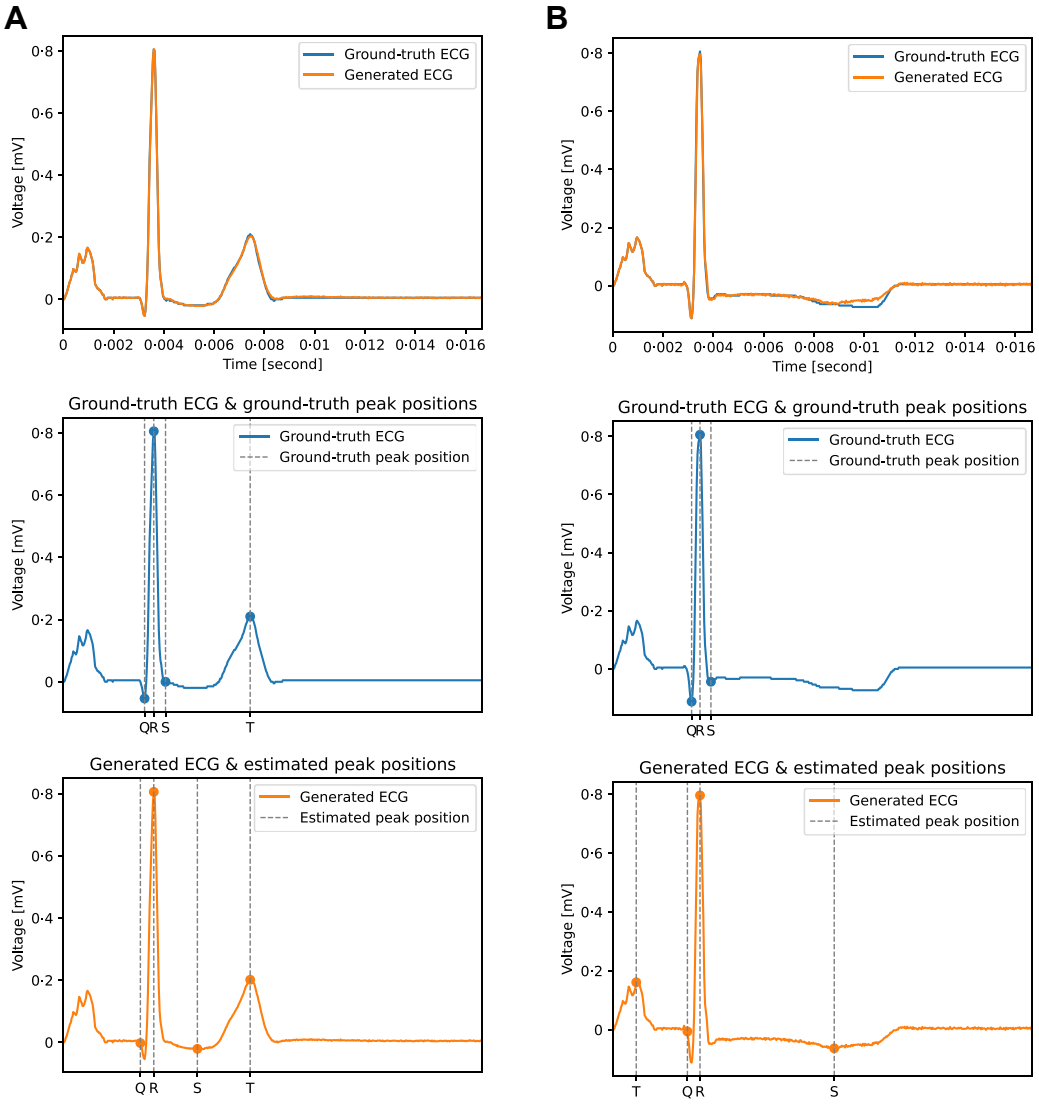

Figure 2 Examples of the ground-truth and generated electrocardiogram (ECG) signals. Blue and orange lines represent the ground-truth and generated ECG signals, respectively. Gray vertical dotted lines represent the ground-truth or estimated positions. **A:** Example including the error of the S positions estimated by the NeuroKit2. **B:** Example with the lowest peak signal-to-noise ratio of lead I.

Table 2 Performance of the NeuroKit2 measured by the mean absolute errors of individual electrocardiogram positions

	Q	R	S	T
PMAE [sample][†]	7.58 ± 8.30	0.55 ± 0.61	19.15 ±14.28	0.51 ± 0.73

PMAE = position mean absolute error.
[†]Results show the mean absolute errors and their standard deviations between the positions manually annotated and those estimated by Neurokit2 for the ground-truth electrocardiograms.

$t_n^{\mathrm{gt},*} \in \{1, \ldots T\}$ is the position in sample unit of the n-th ground-truth ECG signal; and $t_n^{\mathrm{es},*} \in \{1, \ldots, T\}$ is the position in sample unit of the n-th generated ECG signal. The ground-truth positions $t_n^{\mathrm{gt},*}$ were manually annotated by a non-author worker under the supervision of a non-author medical specialist. On the other hand, the positions $t_n^{\mathrm{es},*}$ for the large numbers of ECGs generated by the system were automatically estimated by NeuroKit2.[24] We found that NeuroKit2 sometimes failed to output meaningful positions. For example, NeuroKit2 estimated an incorrect S position in Figure 2A and an incorrect T position in Figure 2B. To exclude the effect of these obvious errors of NeuroKit2, the results for the Q, S, and T positions were excluded from the PMAE calculation in the following cases:

- Q position more than 100 milliseconds (ie, 50 samples) away from the corresponding R position
- S position more than 150 milliseconds (ie, 100 samples) away from the corresponding R position
- T position appearing prior to the corresponding Q position

Furthermore, we did not measure PMAE for the P positions because all the P waves in the UT-Heart dataset used for training and tests are actually copies of an identical template P wave. The HR parameter is fixed to 60 beats per minute, so the duration of 1 beat is 1 second. Since the number of samples of all ECG signals is $T = 500$ per second, the interval between adjacent samples is 1/500 of a second. Therefore, PMAE representing the position gap in samples can be converted to a gap in seconds by PMAE/500.

PSNR is a metric used to evaluate the quality of synthesized ECG signals[25] and is calculated as follows: $\mathrm{PSNR} = \frac{1}{N}\sum_{n=1}^{N} 10\log_{10}\frac{\mathrm{MAX}^2}{\mathrm{MSE}_n}$, where MSE_n represents the mean squared error calculated as: $\mathrm{MSE}_n = \frac{1}{T}\sum_{t=1}^{T}\left(x_{n,t}^{\mathrm{gt}} - x_{n,t}^{\mathrm{es}}\right)^2$, where $\mathbf{x}_n^{\mathrm{gt}} = \{x_{n,1}^{\mathrm{gt}}, \ldots, x_{n,T}^{\mathrm{gt}}\}$ and $\mathbf{x}_n^{\mathrm{es}} = \{x_{n,1}^{\mathrm{es}}, \ldots, x_{n,T}^{\mathrm{es}}\}$ are n-th ground-truth and generated ECG signals in the test dataset, respectively. MAX represents the maximum amplitude that ECG signals can take, and we set MAX = 1 because all ECG signals were normalized before inputting them into the proposed model.

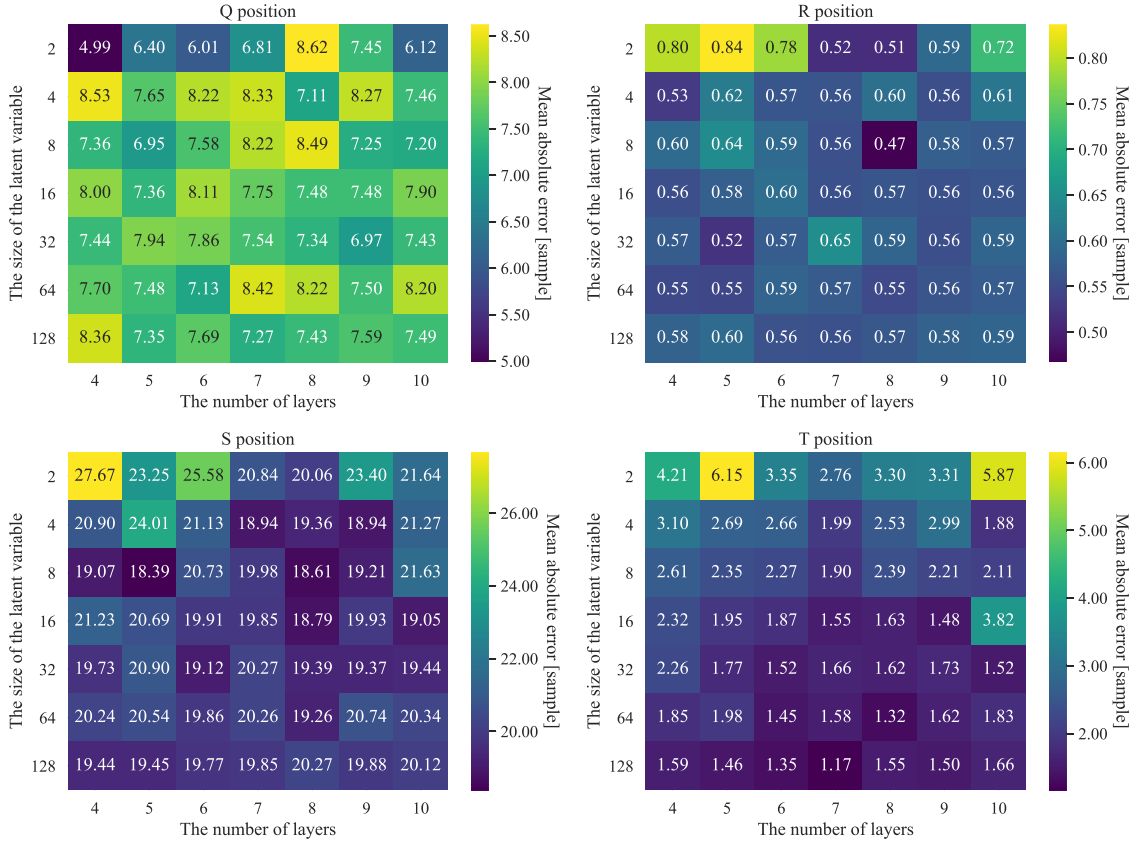

Figure 3 Mean absolute errors of individual electrocardiogram positions. Each heatmap corresponds to the position mean absolute error (PMAE) of each position. The vertical and horizontal axes of each heatmap represent the latent variable size and the number of layers. The numbers and colors of individual elements in the heatmaps represent the calculated PMAEs.

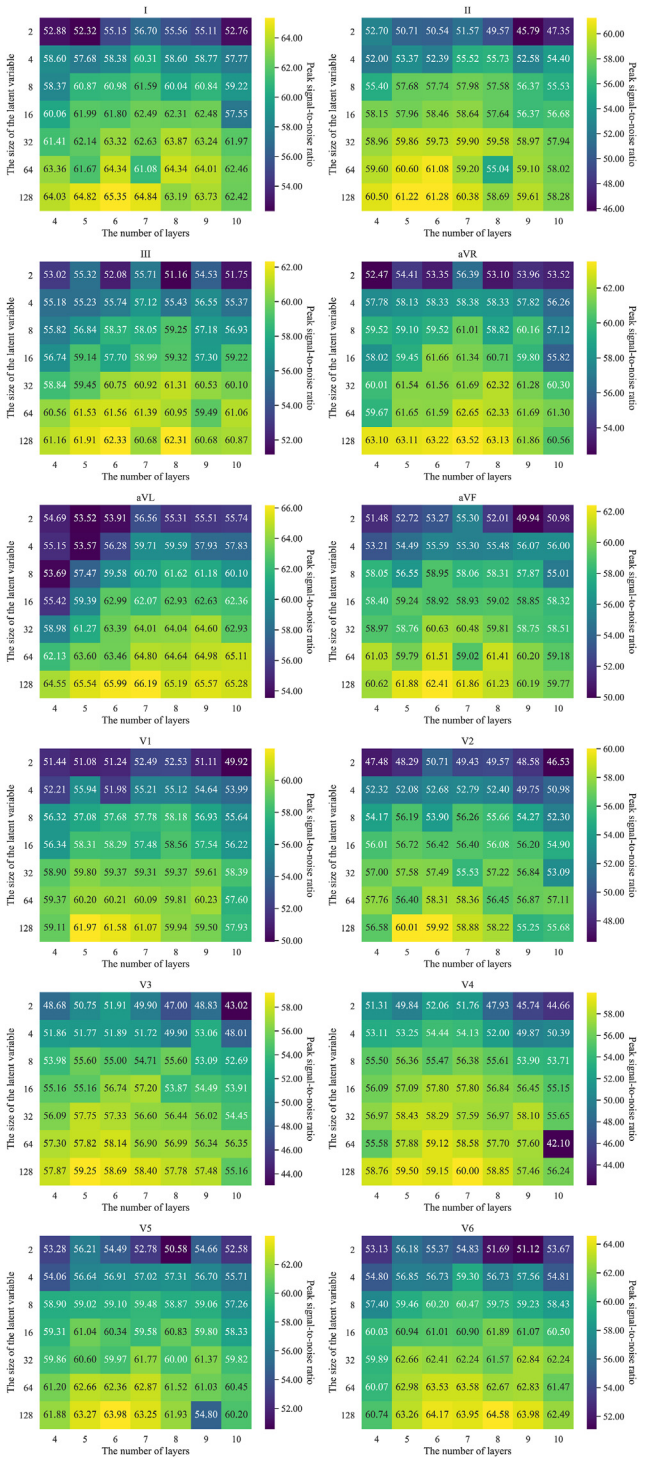

Figure 4 Peak signal-to-noise ratios (PSNRs) of individual leads. Each heatmap corresponds to the PSNRs of each position. The vertical and horizontal axes of each heatmap represent the latent variable size and the number of layers. The numbers and colors of individual elements in the heatmaps represent the calculated PSNRs.

Results

First, we examined the accuracy of NeuroKit2 by measuring PMAE for Q, R, S, and T positions. This was done by applying NeuroKit2 to the manually annotated ground truth waveforms and comparing the results. The results are shown in Table 2. A lower PMAE indicates better accuracy. While the R and T positions were very accurately recognized, the errors of the Q and S positions were larger than those of the R and T positions. We consider that this is because NeuroKit2 tends to recognize the Q and S positions at the lowest local minimum of the entire Q and S waves, respectively, which is not always the case clinically.

The PMAEs calculated on the test data for the combination of the number of layers and the size of latent variables in the proposed network are illustrated in Figure 3. We found that differences in the size of the latent variable and the number of layers did not cause differences in PMAE performance for all positions. Furthermore, the PMAEs of the R and T positions were sufficiently small. The major cause of larger PMAEs for the Q and S positions than the other positions was considered to be the detection error of the Q and S positions by NeuroKit2[25] shown in Table 2.

The PSNRs calculated on the test data for the combination of the number of layers in the proposed network and the size of latent variable are illustrated in Figure 4. The higher PSNR represents the better performance of the proposed method. We found that the differences of the latent variable sizes and the number of layers affected the PSNR performance of the proposed model. The overall tendency of the PSNR heatmaps is similar for all leads and increasing the latent variable size of the proposed network improved the PSNR performance. In fact, the mean PSNRs of the individual leads are 60.60, 56.61, 58.23, 59.44, 60.69, 57.80, 57.03, 54.69, 54.38, 55.00, 58.87, and 59.76, and the standard deviation calculated from these mean PSNRs is 2.15, which is sufficiently smaller than the mean PSNRs. The best performance is obtained for all leads when the size of the latent variable is 128. The number of layers with the best PSNR performance depends on each lead and ranges from 5 to 7. Figure 2B shows the example ECG with the lowest PSNR of lead I among the ECGs generated the proposed VAE with the size of the latent variable of 128 and the number of layers of 7. The lowest PSNR of 55.53 has a gap of more than $3\sigma = 6.15$, where σ is the standard deviation of the PSNRs, from the mean PSNR of 64.84 in Figure 4I, and this gap would be due to the estimation error around 0.01 seconds. The other parts of the estimated ECG waveform are close to the ground-truth waveform.

The PMAE and PSNR calculated for the different cardiac parameters are illustrated in Figures 5 and 6. As shown in these figures, the value of the parameter does not seem to affect the error significantly in most cases. The most notable exception was the effect of I_{Na} on PMAE for Q and S positions, where there was a significant difference in PMAE ($P = .05$) between $I_{Na} = 100\%$ and $I_{Na} = 70\%$.

Discussion

This study examines the effectiveness of a computationally inexpensive machine learning model for reproducing an accurate but expensive process of ECG synthesis using an

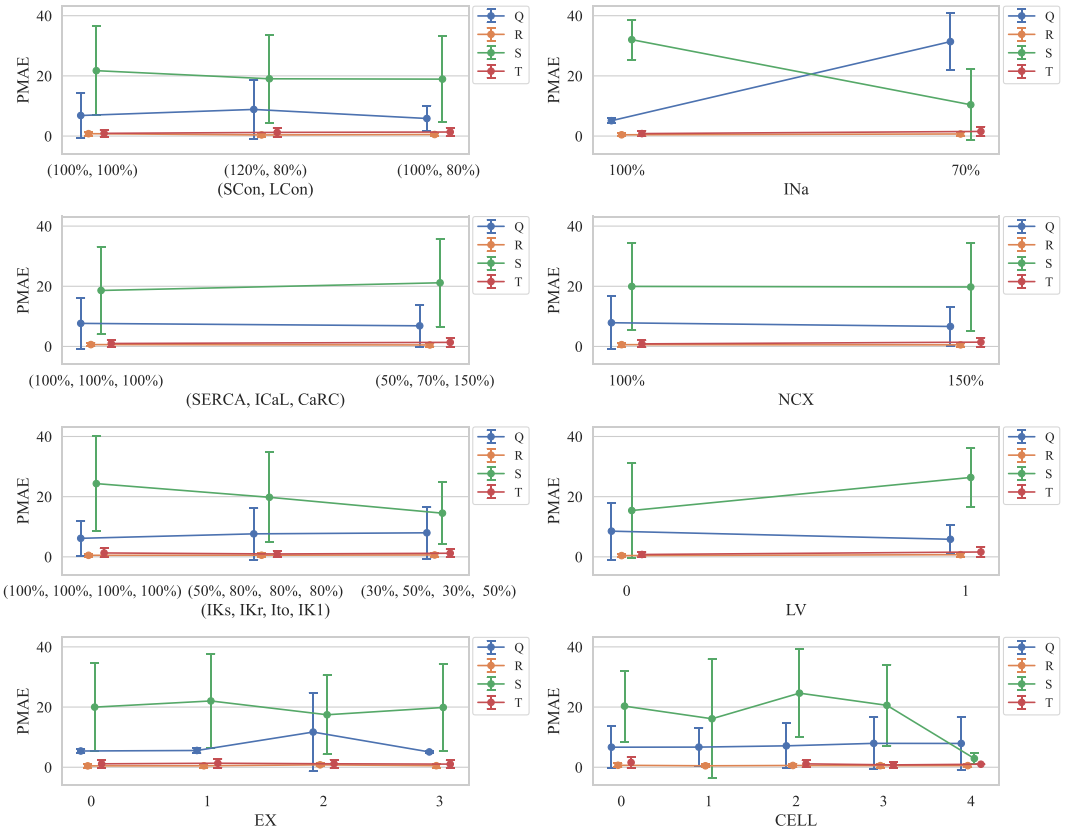

Figure 5 Position mean absolute errors (PMAE) obtained for the different cardiac parameter values. The 8 graphs correspond to the first 8 rows of Table 1, respectively. The vertical and horizontal axes of each graph represent the PMAE and the values that the cardiac parameter can take. The blue, orange, green, and red lines show the PMAE differences of the Q, R, S, and T positions, respectively. The vertical lines on the dots of the line plots represent the standard deviations of PMAEs.

elaborate physical heart model based on the FEM. We confirmed that the VAE model achieved sufficient reproducibility in the objective measure of PSNR for 576 test data when the number of dimensions and layers were set to 128 and 5–8, respectively. Furthermore, whereas UT-Heart[21] takes approximately 350 node-hours to synthesize 1 12-lead ECG on the supercomputer Fugaku, the proposed VAE method takes only a few seconds on a standard personal computer with Intel® Core™ i7-1065G7 CPU and 16 GB memory. This result suggests that it is feasible, in terms of computational cost, to create an encyclopedia of ECGs for given cardiac parameters by exhaustively applying the proposed method to the cardiac parameters in advance. An added benefit of the proposed method is sharp reduction in the electric power required to synthesize ECGs, which in turn reduces the cost of this approach relative to FEM-based ECG synthesis and reduces associated carbon dioxide (CO_2) emissions.

One important byproduct obtained from the proposed VAE machine learning process is the knowledge that can be gleaned about hidden structures in the ECG. When the latent variable dimension is increased, the accuracy of reproducing the ECG is also increased. We also confirmed that good reproduction accuracy could be achieved, especially when the number of dimensions is set to about 128. This result indicates that reasonably rich model representation capability and model complexity are required to reproduce reliable synthetic ECGs, taking into consideration the full range of factors affecting ECG signals owing to real-world patient variations in heart condition and body shape. This implies that it may be feasible in the future to find significant new clues within the ECG to suggest heart condition and body type.

A remaining challenge for this research is further analysis of the extrapolation capability. The present study confirmed that the proposed lightweight model has sufficient extrapolation capability to replicate the ECG data synthesized by the elaborate FEM-based UT-Heart physical/electrical heart model. However, we still need to verify the validity of the ECGs synthesized by the general-purpose machine learning model from parameters that deviate significantly from those used in this study. To improve such capability, we consider that it is important to efficiently select a small amount of training data to best train general-purpose machine learning models.

Another challenge is to improve the proposed method, which increased the speed of ECG synthesis, so that it can support the discovery of new clinical knowledge and insights and be used in actual medical practice. We dealt with only synthetic ECGs in this paper, but the real ECGs may contain deviations that do not appear in the synthetic data and are caused by factors not represented by the cardiac parameters.

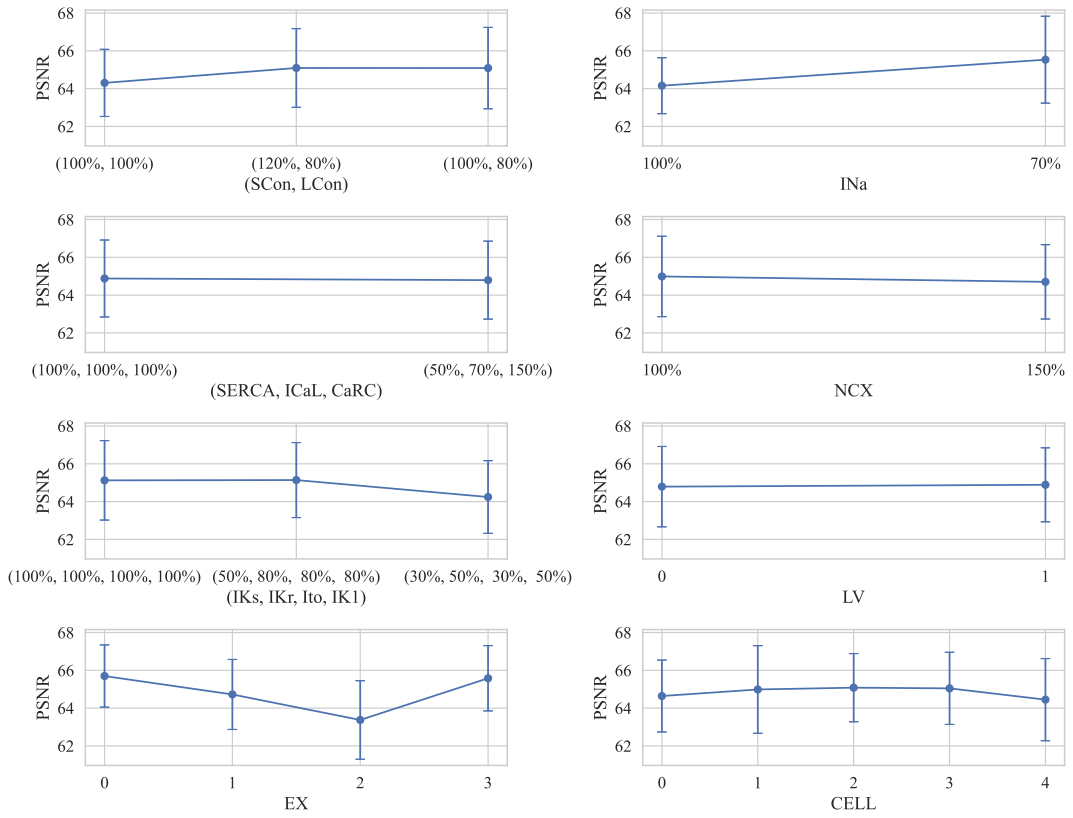

Figure 6 Peak signal-to-noise ratios (PSNR) of lead I obtained for the different cardiac parameter values. The 8 graphs correspond to the first 8 rows of Table 1, respectively. The vertical and horizontal axes of each graph represent the PSNR and the values that the cardiac parameter can take.

Therefore, to address the challenge, we will extend our framework for extracting clinical knowledge and insights from real ECGs by integrating the framework with an additional estimator of labels useful for other ECG analysis. We anticipate that the dramatic increase presented here in simulation speed will facilitate analysis and comparison to real-world ECGs.

We expect that this study will suggest new research perspectives in ECG analysis beyond just creating an ECG encyclopedia for diagnostic reference—for example, research into ways of estimating a patient's cardiac parameters backwards from their actual ECG signals. Because this study has confirmed that ECGs can be adequately represented by general-purpose machine learning models such as VAE, the direct mapping from ECGs to cardiac parameters could also potentially be learned as inputs and outputs of general-purpose machine learning models. If such is confirmed to be practical, it could help automate and expand the search for clinically useful cues to cardiovascular health from real-world ECGs.

Conclusion

This paper proposed the conditional VAE-based method that synthesizes 12-lead ECGs from the cardiac parameters. The experimental results showed that the proposed method could efficiently synthesize accurate ECGs and can be used as an alternative to the FEM-based heart simulator, which requires much time and computational resources. This research opens the door to practical implementation of a comprehensive ECG encyclopedia, needed to bridge the model-based simulation and the real ECG observations, by exhaustively synthesizing ECGs for a wide range of cardiac parameter sets within a reasonable time and cost. Further investigation is needed to determine whether our experimental model maintains its accuracy on a wider range of cardiac parameters beyond the specific parameters used to train our VAE machine learning model so far. In the future, we hope to develop a scheme in which the FEM model and the machine learning model can improve each other, using active learning and self-supervised learning frameworks to improve the accuracy of our VAE-based ECG synthesis tool. We will also investigate the potential of our VAE approach to extrapolate salient synthetic ECG training data from sparse source data of various types.

Acknowledgments

The authors would like to thank Drs Jun-ichi Okada, Seiryo Sugiura, and Toshiaki Hisada at UT-Heart Inc. for valuable discussion and sharing their synthetic ECG dataset with us.

Funding Sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Disclosures
The authors have no conflicts to disclose.

Authorship
All authors attest they meet the current ICMJE criteria for authorship.

Patient Consent
This work does not include any experiments with patients.

Ethics Statement
The research reported in this paper adhered to TRIPOD guidelines.

Data Availability
For reproducibility of this work, we will publish the source code used for our study, including model training, ECG synthesis, experiments, and evaluation. In addition, we will attach a readme file describing how to use the source code. Right after the acceptance of the manuscript, the source code will be published on GitHub of NTT Communication Science Laboratories (https://github.com/nttcslab). Subsequent studies are permitted to use the source code without restriction only if they agree to a license to reproduce the method.

References

1. Attallah O. ECG-BiCoNet: an ECG-based pipeline for COVID-19 diagnosis using bi-layers of deep features integration. Comput Biol Med 2022;142:105210.
2. Choi J, Hong D, Jung C, Hwang E, Park S, Roh S. A multi-view learning approach to enhance automatic 12-lead ECG diagnosis performance. International Workshop on Data Mining in Bioinformatics 2022;1–9.
3. Revathi J, Anitha J, Hemanth DJ. An intelligent medical decision support system for diagnosis of heart abnormalities in ECG signals. Intelligent Decision Technologies 2021;15:19–31.
4. Liu X, Wang H, Li Z, Qin L. Deep learning in ECG diagnosis: a review. Knowledge-Based Systems 2021;227:107187.
5. Sundararajan P, Moses K, Potes C, Parvaneh S. Automatic diagnosis of cardiac disease from twelve-lead and reduced-lead ECGs using multilabel classification. Comput Cardiol 2021;1–4.
6. Nonaka N, Seita J. In-depth benchmarking of deep neural network architectures for ECG diagnosis. Proceedings of the 6th Machine Learning for Healthcare Conference 2021;414–439.
7. Oliveira DM, Ribeiro AH, Pedrosa JAO, Paixão GMM, Ribeiro ALP, Meira W Jr. Explaining end-to-end ECG automated diagnosis using contextual features. In: Dong Y, Ifrim G, Mladenić D, Saunders C, Van Hoecke S, eds. Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track; 2020. p. 204–219.
8. Alickovic E, Subasi A. Effect of multiscale PCA denoising in ECG beat classification for diagnosis of cardiovascular diseases. Circuits, Systems, and Signal Processing 2015;34:513–533.
9. Golany T, Radinsky K, Freedman D. SimGANs: simulator-based generative adversarial networks for ECG synthesis to improve deep ECG classification. ICML'20. Proceedings of the 37th International Conference on Machine Learning 2020;3597–3606.
10. Hagiwara Y, Fujita H, Oh SL, et al. Computer-aided diagnosis of atrial fibrillation based on ECG signals: a review. Information Sciences 2018;467:99–114.
11. Karthik S, Santhosh M, Kavitha MS, Paul CA. Automated deep learning based cardiovascular disease diagnosis using ECG signals. Computer Systems Science and Engineering 2022;42:183–199.
12. Zhu J, Lv J, Kong D. CNN-FWS: a model for the diagnosis of normal and abnormal ECG with feature adaptive. Entropy 2022;24:471.
13. Prabhakararao E, Dandapat S. Attentive RNN-based network to fuse 12-lead ECG and clinical features for improved myocardial infarction diagnosis. IEEE Signal Processing Letters 2020;27:2029–2033.
14. Kirodiwal A, Srivastava A, Dash A, et al. A bio-toolkit for multi-cardiac abnormality diagnosis using ECG signal and deep learning. Comput Cardiol 2020;1–4.
15. Acharya UR, Fujita H, Oh SL, et al. Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals. Applied Intelligence 2019;49:16–27.
16. Wang K, Zhang X, Zhong H, Chen T. Automatic diagnosis with 12-lead ECG signals. Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting; 2019. p. 28–35.
17. Kumar SU, Inbarani HH. Neighborhood rough set based ECG signal classification for diagnosis of cardiac diseases. Soft Computing 2017; 21:4721–4733.
18. Golany T, Freedman D, Radinsky K. ECG ODE-GAN: learning ordinary differential equations of ECG dynamics via generative adversarial learning. AAAI Conference on Artificial Intelligence 2021;134–141.
19. Habiba M, Borphy E, Pearlmutter BA, Ward T. ECG synthesis with neural ODE and GAN models. International Conference on Electrical, Computer and Energy Technologies; 2021. p. 1–6.
20. Golany T, Radinsky K. PGANs: personalized generative adversarial networks for ECG synthesis to improve patient-specific deep ECG classification. AAAI Conference on Artificial Intelligence 2019;557–564.
21. Sugiura S, Washio T, Hatano A, Okada J, Watanabe H, Hisada T. Multi-scale simulations of cardiac electrophysiology and mechanics using the University of Tokyo heart simulator. Prog Biophys Mol Biol 2012;110:380–389.
22. [dataset] UT-Heart Inc. Database labeled with subcellular pathologies developed by multi-scale heart simulator, UT-Heart. http://ut-heart.com/ECGdata_fugaku/database_download.html. Accessed November 1, 2022.
23. Kingma DP, Ba JL. Adam: a method for stochastic optimization. International Conference on Learning Representations 2015;1–15.
24. Makowski D, Pham T, Lau ZJ, et al. NeuroKit2: a Python toolbox for neurophysiological signal processing. Behav Res Methods 2021;53:1689–1696.
25. Chen J, Zheng X, Yu H, Chen DZ, Wu J. Electrocardio panorama: synthesizing new ECG views with self-supervision. International Joint Conference on Artificial Intelligence 2021;3597–3605.
26. Kerckhoffs RC, Neal ML, Gu Q, Bassingthwaighte JB, Omens JH, McCulloch AD. Coupling of a 3D finite element model of cardiac ventricular mechanics to lumped systems models of the systemic and pulmonic circulation. Ann Biomed Eng 2007;35:1–18.