Personalized Heart Disease Detection via ECG Digital Twin Generation

Yaojun Hu, Jintai Chen, Lianting Hu, Dantong Li, Jiahuan Yan, Haochao Ying, Huiying Liang, Jian Wu

1College of Computer Science and Technology, Zhejiang University
2Computer Science Department, University of Illinois at Urbana-Champaign
3Medical Big Data Center, Guangdong Provincial People’s Hospital, Southern Medical University
4Guangdong Cardiovascular Institute, Guangdong Provincial People’s Hospital, Guangdong Academy of Medical Sciences
5Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application, Guangdong Provincial People’s Hospital (Guangdong Academy of Medical Sciences) 
6School of Public Health, Zhejiang University
7State Key Laboratory of Transvascular Implantation Devices of The Second Affiliated Hospital, Zhejiang University School of Medicine

Abstract
Heart diseases rank among the leading causes of global mortality, demonstrating a crucial need for early diagnosis and intervention. Most traditional electrocardiogram (ECG) based automated diagnosis methods are trained at population level, neglecting the customization of personalized ECGs to enhance individual healthcare management. A potential solution to address this limitation is to employ digital twins to simulate symptoms of diseases in real patients. In this paper, we present an innovative prospective learning approach for personalized heart disease detection, which generates digital twins of healthy individuals’ anomalous ECGs and enhances the model sensitivity to the personalized symptoms. In our approach, a vector quantized feature separator is proposed to locate and isolate the disease symptom and normal segments in ECG signals with ECG report guidance. Thus, the ECG digital twins can simulate specific heart diseases used to train a personalized heart disease detection model. Experiments demonstrate that our approach not only excels in generating high-fidelity ECG signals but also improves personalized heart disease detection. Moreover, our approach ensures robust privacy protection, safeguarding patient data in model development. The code can be found at https://github.com/huyjj/LAVQ-Editor.

1 Introduction
Heart disease stands as a leading cause of global mortality [Roth et al., 2018]. In clinical practice, the electrocardiogram (ECG) is the most routinely-used tool for heart disease detection due to its affordability, convenience, and non-invasiveness. However, the lengthy ECG signals (e.g., records that may span several hours from Holter monitors) pose challenges for manual inspection, and cutting-edge deep learning techniques [Li et al., 2023, Chen et al., 2020, Chen et al., 2024] are widely employed to automatically find anomalous signals. Nevertheless, the application of ECG signals in deep learning is constrained by high annotation costs and data privacy concerns. In addition, given the variability of disease symptoms among individual patients, current data-driven deep learning approaches face challenges in ensuring the reliability across diverse patient profiles.

An obvious solution to the data dilemma is to generate more data by deep generative networks. Deep generative networks such as generative adversarial networks (GANs) [Goodfellow et al., 2014] and variational autoencoders (VAEs) [Kingma and Welling, 2013] have been used to synthetic ECGs [Hossain et al., 2021, Chen et al., 2022]. Nevertheless, most prior researches [Galony et al., 2020, Chen et al., 2022, Chen et al., 2021] have predominantly concentrated on generating ECGs following population distributions, largely neglecting the creation of tailored ECGs for individual patients. Such limitation suggests that the generated ECGs are more aptly utilized as supplementary data for training models rather than for direct application in personalized patient care within clinical practice. Digital twins are virtual constructs designed to mirror real-world objects, allowing for rapid and non-invasive simulation of disease progression and treatment trials [Das et al., 2023]. It also involves preserving unique patient information to ensure that subsequent simulations and predictive modeling are meaningful.

To specify, the primary goal of this paper is to generate personalized ECG digital twins, thereby facilitating the individualized management of heart conditions.

We create personalized ECG digital twins by editing an individual’s ECG signals to simulate the heart disease symptoms prospectively for heart disease detection improvement. Our purpose is to ensure that the created ECG digital twins
can assist clinicians and automated diagnosis machine learning approaches in gaining prospective insight into individual symptoms, thereby enhancing the quality of personalized diagnosis. The representations of most heart diseases in ECG signals typically exhibit locality, attributed to the different signal wave segments reflecting conditions at specific parts on the heart [Chen et al., 2022]. Hence, when simulating a target disease, our approach separates the ECG features into normal segments and disease-indicative segments guided by the textual description of the disease. The disease-indicative components are then edited to align with the features of the target disease. In this paper, we define “disease-indicative feature” as the ECG features wherein abnormalities may manifest, serving as the basis upon which clinicians rely for diagnosing the heart disease. Notably, distinct heart diseases may exhibit on different disease-indicative segments.

Our proposed model comprises three components: the VQ-Separator for text-guided separation of different segments of the ECG, a generator employing a feature mapper to incorporate disease information into the ECG signals, and a discriminator to discern authenticity. In the training process, we employ ECG-text pairs from a pre-diagnosis patient whose ECG digital twins are to be constructed (potentially healthy or otherwise) and a reference patient already diagnosed with a specific disease. According to the guidance of the heart disease descriptive text, VQ-Separator picks out the disease-indicative features from the personal normal features. Subsequently, we leverage the normal feature of the pre-diagnosis patient and reference the disease-indicative feature from the reference patient to adjust the ECGs of the pre-diagnosis patient, aligning the disease condition with that of the reference patient. In addition, to preserve personalized characteristics in the ECG signals, the generator not only learns to generate edited ECGs but also possesses the capability to reconstruct ECGs with their own disease-indicative features. The discriminator is trained to discern both the validity and whether the edited ECGs have integrated the desired disease features. Our main contributions are summarized as follows:

- We introduce a novel prospective learning concept for personalized heart detection by creating personalized ECG digital twins that simulate heart disease symptoms. This method provides prospective information to enhance the cognition of heart diseases, thereby improving subsequent detection performance.
- We propose a location-aware model called LA VQ-Editor for personalized ECG digital twins generation, which localizes and separates personal normal features and disease-indicative features guided by textual disease descriptions. This is achieved through the novel VQ-Separator that operates with a Disease Embedding Codebook to precisely model the disease features.
- Empirically, our controlled experiments have demonstrated that patients who provided prospective cognition to the heart disease detection model via ECG digital twins achieved significantly better diagnostic accuracy than those who did not. Furthermore, our experiments also validate that our methodology not only efficiently leverages patient information but also enhances patient privacy to a greater extent.

## 2 Related Work

### 2.1 Generative Adversarial Networks

Currently, generative adversarial networks have made great achievements in the areas of image synthesis [Karras et al., 2019], text generation [Yu et al., 2017] and scientific discovery [Repecka et al., 2021]. In addition, GANs have been instrumental in enabling interactive image editing where users can provide input to guide the generation process. For instance, GauGAN [Park et al., 2019] allowed users to create complex scenes by painting a segmentation map that the network converts into a photorealistic image, effectively turning sketches into detailed landscapes. Patashnik et al. [Patashnik et al., 2021] introduced StyleCLIP, which utilizes the capabilities of StyleGAN and CLIP to manipulate images based on textual descriptions. The method is known for its intuitive nature, allowing non-experts to make significant, detailed changes to images using simple text prompts. Hertz et al. [Hertz et al., 2022] presented a novel approach that refines the interaction between text prompts and the generated images using a cross-attention within transformer [Yaswani et al., 2017]. This advancement allows for more precise and localized control over the image editing process, facilitating detailed adjustments without extensive collateral impacts on the image. Similarly, our method separates disease-indicative and normal features in ECGs through text-guidance and cross-attention, and then edits the ECG by fusing the target disease-indicative features.

### 2.2 Vector Quantization Based Methods

The Vector Quantized Variational Autoencoder (VQ-VAE) [Van Den Oord et al., 2017] is a significant advancement in the field of generative models, blending the concepts of auto encoding and vector quantization to achieve impressive results in data compression and generation. In VQ-VAE, the encoder maps the input data to a discrete latent representation using vector quantization, and the decoder reconstructs the input data from this discrete representation. Its ability to learn meaningful discrete representations has made it a popular choice for tasks that require high-quality, diverse sample generation [Esser et al., 2021, Razavi et al., 2019]. In our work, we build a learnable Disease Embedding Codebook for preserving disease latent space and mapping the disease-indicative features to it by vector quantization. However, we use only the discrete representations of the disease and the output of the encoder for the rest to reconstruct in decoder.

### 2.3 Generation Methods for Electrocardiograms

Electrocardiograms (ECGs) are a cornerstone in diagnosing heart diseases, yet their interpretation hinges on the discernment of clinical physicians, placing a considerable demand on their expertise and time. The imperative for more streamlined and effective automated systems for ECG analysis is driven by the need for rapid and precise heart health assessments. Efforts have been made to create automated ECG classifiers [Baloglu et al., 2019, Li et al., 2023, Chen et al., 2020], but they are often hampered by the
The synthesis of ECG signals emerges as a vital solution, enhancing the variety and volume of training samples while also mitigating privacy issues. Recent advancements, such as SimGAN [Golany et al., 2020], leverage Ordinary Differential Equations (ODEs) to capture cardiac dynamics and generate ECGs, whereas methodologies like Nef-Net [Chen et al., 2021] and MEGAN [Chen et al., 2022] focus on the synthesis of multi-view ECG signals grounded in cardiac electrical activity. Besides, Chung et al. [Chung et al., 2023] introduced Auto-TTE, an autoregressive generative model informed by clinical text reports. Golany et al. [Golany and Radinsky, 2019] generated personalized heartbeats by adding constraints on patient-relevant wave values. While existing studies often generate entirely new ECGs using disease-related information, our approach diverges by directly editing a patient’s normal ECGs to incorporate specific disease features instead of constructing ECGs from noise.

3 Methods

The manifestation of most heart diseases in ECG signals typically exhibits localized patterns, and such information is commonly included in ECG reports. Our objective is to generate ECGs with disease-specific characteristics while preserving the patient’s individual traits according to the guidance provided in ECG reports. Refer to Figure 1, our model, LAVQ-Editor, comprises three components: a Vector Quantized Feature Separator (VQ-Separator) with a Disease Embedding Codebook, a generator G consisting of a feature mapper and a decoder G_{dec}, and a discriminator D. Subsequently, we will introduce each of these three components individually.

3.1 Vector Quantized Feature Separator

The role of the Vector Quantized Feature Separator is to separate the ECG features into normal features and disease-indicative features, guided by descriptive text. As shown in Figure 2, we first encode the ECGs with an encoder G_{enc} based on 1D ResNet-34. This process can be represented as

$$f_{ecg} = G_{enc}(X_{ecg}), f_{ecg} \in \mathbb{R}^{C \times L_e}$$

where $X_{ecg}$ is the input of normalized ECG signals. Here, $C$ and $L_e$ indicate the channel and the embedding dimensions, respectively. In practice, clinicians detect heart diseases by recognizing the disease-specific characteristics and properties of ECG signals, and document a rich source of information in the report. Inspired by [Li et al., 2023], we encode the ECG report texts with a pre-trained ClinicalBERT $G_{text}$ [Alsentzer et al., 2019], which is frozen in the training process. We further use a linear layer to compress $f_{text}$ to one dimension. Formally, the process can be formulated by

$$f_{text} = \text{Linear}(G_{text}(X_{text})), f_{text} \in \mathbb{R}^{L_t}$$

where $X_{text}$ is the descriptive text of the input ECG signals and $L_t$ indicate the embedding dimensions.

Here we utilize the textual feature to guide the separation of features in ECG signals through cross-attention [Hou et al., 2019]. Specifically, full ECG feature $f_{ecg}$ is refined into a query, while $f_{text}$ serves as both key and value. Then we define a criterion to determine whether a feature segment exhibits disease symptoms, simply implemented by a threshold $f$ for the cross-attention elements. Conversely, element values below this threshold are recognized as...
the normal features. This process can be defined as

\[ A = \mathbf{1}[\text{MHCA}(f_{\text{ecg}}, f_{\text{text}}) \geq l], \quad (3) \]

where \( \mathbf{1}[\cdot \geq l] \) binarizes the matrix elements based on whether they are larger than the threshold \( l \), and MHCA means Multi-Head Cross-Attention module. The outcome \( A \in \mathbb{R}^{C \times L_e} \) is a binary mask indicating the segments presenting the disease symptoms.

Utilizing the vector quantization technique, we compress and quantize the disease-indicative embedding within \( f_{\text{ecg}} \) to remove personal information while accentuating disease symptoms. This process involves Disease Embedding Codebook (DEC), comprising a series of embedding vectors \( v_k \in \mathbb{R}^{K \times L_e} \) for \( k \in \{1, 2, \ldots, K\} \), where \( K \) represents the size of the embedding book. To isolate disease-indicative features, we select the nearest quantized discrete feature \( Q(e) \) from the DEC. This procedure can be represented as:

\[ Q(e) := (\text{argmin} \| e_i - v_k \|^2_2 \text{ for } i \text{ in } C). \quad (4) \]

Given that many ECG segments may not exhibit disease symptoms, we specifically choose segments corresponding to disease symptoms as the disease-indicative feature from \( Q(e) \) calculated as \( f_d = Q(e) \ast A \), where “\( \ast \)” denotes point-wise multiplication. Conversely, segments devoid of disease symptoms are selected as personal normal features from \( e \), computed as \( f_p = e \ast (I - A) \), with \( I \) being an all-one matrix matching \( A \)’s shape. In this way, we effectively preserve normal features to retain individual information of pre-diagnosis patients, while leveraging feature quantization to eliminate personal features from the reference patient to maximize the retention of disease-related features.

Different from existing methods, our objective focuses on retaining the quantized latent disease embeddings within the DEC. Thus, there is no necessity to use the entirety of \( Q(e) \). Instead, we selectively utilize only those segments identified as disease-indicative features. This is done with a loss function defined by:

\[ \mathcal{L}_{\text{vq}} = ||(sg(Q(e)) - Q(e)) \ast A||^2_2 + ||(sg(Q(e)) - e) \ast A||^2_2 \quad (5) \]

where \( sg[\cdot] \) means stop-gradient and “\( \ast \)” indicates the point-wise multiplication.

3.2 Generator

After obtaining disease-indicative features and personal normal features, we aim to fuse the target predicted disease-indicative features with personal normal features in the generator. In the training process, we use the reference patient’s disease-indicative feature \( f_d^t \) as the target disease, which is extracted in the same way as \( f_d \). Before fusion, the disease-indicative features \( f_d \) and \( f_d^t \) are transformed into \( \hat{f}_d \) and \( f_d^t \) by a mapper with eight sequential linear layers. Following [Karras et al., 2019, Chen et al., 2022], we introduce adaptive instance normalization (AdaIN) to fuse \( f_d^t \) and \( f_p \) in the decoder. We first add the learnable noise embedding to the personal normal features \( f_p \) and then transform the disease-indicative features \( f_d^t \) to the control vectors, and finally perform AdaIN to obtain \( \hat{f}_p \). Subsequently, we upsample \( \hat{f}_p \) and further fuse it with \( \hat{f}_d \) by nine decoder blocks with AdaIN and convolution layers, as shown in Figure 1. In each decoder block, the features are firstly upsampled with the nearest neighbor interpolation approach, and we inject \( \hat{f}_d^t \) by two AdaIN operations.

Unlike conventional generative models, our generator is tasked with producing not merely realistic ECGs but also with ensuring that the synthesized ECG digital twins accurately represent both the specific disease symptoms and the individual patient characteristics. To achieve this, we employ both reconstruction loss and disease similarity loss to optimize the generator. In order to accurately represent disease-indicative features, we try to ensure ECGs associated with the same symptoms to present similar features from DEC after processing by the VQ-Separator. Hence, we constrain the generated ECG digital twin to have target disease-indicative features by comparing the disease-indicative features separated from the reference ECG. Here we use the cosine similarity between the two to design the disease similarity loss, which is formally defined by:

\[ f_d^G, f_p^G = \text{VQ-Separator}(\hat{X}_{\text{ecg}}^t, X_{\text{text}}^t), \quad (6) \]

\[ \mathcal{L}_{\text{sim}}^G = 1 - \frac{f_d^G \cdot f_d^t}{\| f_d^G \| \| f_d^t \|} \quad (7) \]

where \( X_{\text{text}}^t \) represents the reports of the reference ECG.

Additionally, to ensure that a maximum amount of characteristics about the pre-diagnosis patient is retained during the ECG editing process, we formulate a reconstruction loss. We use the disease-indicative feature \( f_d \) of the ECG from the pre-diagnosis patient instead of the target disease-indicative feature \( f_d^t \) from the reference to reconstruct ECG by the generator. This process is defined as:

\[ \mathcal{L}_{\text{rec}} = \| \hat{X}_{\text{ecg}}^t - X_{\text{ecg}}^t \|^2 \quad (8) \]

where \( \hat{X}_{\text{ecg}}^t \) means the reconstructed ECG by the generator. In line with established generative adversarial networks, we also employ an adversarial loss to optimize the generator, by:

\[ \mathcal{L}_{\text{adv}}^G = \mathbb{E}_{z \sim p_z} \log(1 + e^{\exp(-D(G(z)))}) \quad (9) \]

where \( p_z \) represents the distribution of the pre-diagnosis patients’ ECGs. Thus, the total optimization objective of the generator can be represented as:

\[ \mathcal{L}^G = \mathcal{L}_{\text{adv}}^G + \mathcal{L}_{\text{rec}} + \mathcal{L}_{\text{sim}}^G + \mathcal{L}_{\text{vq}}. \quad (10) \]
3.3 Discriminator
Our discriminator architecture, featuring ten blocks and two heads, is designed for authenticity verification. In the architecture, each block contains two convolution layers with a 3-sized kernel, followed by a ReLU activation function for non-linearity, and a Gaussian blur layer with a kernel size of 2 for feature smoothing.

The output head, responsible for verifying the authenticity of the ECG, incorporates a layer of Standard Deviation [Karras et al., 2019], a convolution layer and two linear layers that finally determines the authenticity of the ECG. The optimization objective for the discriminator also includes the conventional discriminator loss, by:

$$\mathcal{L}_{adv}^{D} = \mathbb{E}_{x \sim p_{data}(x)}[(\log(1 + \exp(-D(x)))) + \mathbb{E}_{x \sim p_{g}(z)}[(\log(1 + \exp(D(z))))]$$

(11)

where \( p_{data} \) represents the distribution of the real ECGs and \( p_{g} \) represents the distribution of the generated ECGs.

In our study, generating realistic ECGs that precisely capture the symptoms of specific heart diseases is also paramount. To achieve this, we designed the disease-indicative feature extraction head that has the similar structure with the output head, but with a key distinction in the output size of the final linear layer. To equip the discriminator with the ability to assess disease-indicative features in ECGs, we optimize it by maximizing the similarity between \( f_{d} \) extracted from the pre-diagnosis patient and \( f_{d}^{D} \) extracted from the reconstructed ECG by the disease-indicative feature extraction head, as:

$$\mathcal{L}_{sim}^{D} = 1 - \frac{f_{d}^{D} \cdot f_{d}}{\| f_{d}^{D} \| \cdot \| f_{d} \|}.$$  

(12)

The optimization objective of the discriminator is:

$$\mathcal{L}^{D} = \mathcal{L}_{adv}^{D} + \mathcal{L}_{sim}^{D}.$$  

(13)

4 Experiments
Our evaluation of the proposed method is organized on three aspects: effectiveness, fidelity, and privacy security. To assess effectiveness, we incorporate our prospective learning method into a data augmentation model for heart disease detection. We then introduce two specialized evaluation metrics to measure the fidelity of our synthesized ECG digital twins, drawing parallels with established benchmarks in image generation. Additionally, we compute the “membership inference risk” metric to assess the privacy security of the generated results. Our study includes comparisons with conventional generative methods condition GAN [Mirza and Osindero, 2014], VQ-VAE2 [Razavi et al., 2019], WGAN [Arjovsky et al., 2017], StyleGAN [Karras et al., 2019], and personalized generative method PGANs [Golany and Radinsky, 2019].

4.1 Implementation Details
In our experiments, we utilize the PTB-XL dataset, a comprehensive and publicly accessible collection of electrocardiograms, which contains 21,837 clinical 12-lead ECGs from 18,885 individuals. Accompanying each ECG is an extensive report and a set of diagnostic labels. These labels are categorized into 5 overarching superclasses and an additional 24 subclasses. Our research strictly focuses on the superclasses, using them as the basis for all analyses presented in this study. Given the importance of clear disease differentiation, our experiments are conducted exclusively with single-label data. To ensure a robust test of our model’s capabilities, we selected 291 patients who have both normal and disease ECG records for testing, while allocating the remaining patients’ data for training and validation purposes.

For the training process of LAQV-Editor, we employ the Adam optimization algorithm, with a learning rate set at 0.00001 and a weight decay of 0.001. The model undergoes a rigorous training regime spanning 500 epochs and is processed with a batch size of 128. The threshold \( t \) is set as 0.5 in the following experiments, and we provide ablation study of the threshold in the supplements. All subsequent experiments are executed using the PyTorch 1.9 on an NVIDIA GeForce RTX-3090 GPU. We benchmark our method against the existing generative models tailored to accommodate the data and maintain consistency in the training configuration to ensure fairness in comparison.

4.2 Personalized Heart Disease Detection
To verify the utility of our method, we test whether there is a benefit of prospective learning for personalized heart disease detection models. We use 1D ResNet-34 and 1D GoogleNet as the base model and compare the performance of the base model enhanced with and without the results of different generative models. In this study, we randomly divide the patients in the test set into two even groups for a comprehensive evaluation of our prospective learning methodology. The first group, designed as the experimental group, is subjected to our prospective learning approach, providing prospective cognition into the heart disease detection model. The second group is assigned as the control group, serving as a baseline against which the effectiveness of our methodology could be compared. For every normal ECG in the experimental group of the test set, we generate 50 ECGs with diseases and add them to the training set. We optimize the base model by the Adam optimizer with a learning rate of 0.001 and weight decay of 0.001, and train 256 epochs with a batch size of 64.

Evaluation at the patient level. To evaluate the performance of our methodology for personalized heart disease detection, we implement metrics of personalized accuracy, F1-score, and Average Treatment Effect (ATE). Personalized accuracy and F1-score represent the aggregate of individual patient accuracies and F1-scores, respectively. The ATE metric is particularly insightful, quantifying the differential in expected accuracy outcomes when a specific treatment is applied across a patient cohort, as opposed to when it is with-
Table 1: Evaluation of the Classification Performance of Several Base Models Trained on Augmented Training Sets Synthesized via Various Techniques. The best performances are marked in bold.

<table>
<thead>
<tr>
<th>Method</th>
<th>Patient Wise</th>
<th>ECG Wise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy(↑)</td>
<td>F1-score(↑)</td>
</tr>
<tr>
<td>ResNet-34 w/o generator</td>
<td>69.13%</td>
<td>0.6358</td>
</tr>
<tr>
<td>ResNet-34+CGAN</td>
<td>70.13%</td>
<td>0.6309</td>
</tr>
<tr>
<td>ResNet-34+WGAN</td>
<td>70.94%</td>
<td>0.6370</td>
</tr>
<tr>
<td>ResNet-34+StyleGAN</td>
<td>70.62%</td>
<td>0.6409</td>
</tr>
<tr>
<td>ResNet-34+VQ-VAE2</td>
<td>69.54%</td>
<td>0.6124</td>
</tr>
<tr>
<td>ResNet-34+PGANs</td>
<td>69.67%</td>
<td>0.6296</td>
</tr>
<tr>
<td>ResNet-34+LA VQ-Editor(Ours)</td>
<td>72.21%</td>
<td>0.6717</td>
</tr>
<tr>
<td>GoogleNet w/o generator</td>
<td>67.19%</td>
<td>0.6088</td>
</tr>
<tr>
<td>GoogleNet+CGAN</td>
<td>67.90%</td>
<td>0.6198</td>
</tr>
<tr>
<td>GoogleNet+WGAN</td>
<td>67.43%</td>
<td>0.6217</td>
</tr>
<tr>
<td>GoogleNet+StyleGAN</td>
<td>67.60%</td>
<td>0.6136</td>
</tr>
<tr>
<td>GoogleNet+VQ-VAE2</td>
<td>69.73%</td>
<td>0.6371</td>
</tr>
<tr>
<td>GoogleNet+PGANs</td>
<td>69.31%</td>
<td>0.6234</td>
</tr>
<tr>
<td>GoogleNet+LA VQ-Editor(Ours)</td>
<td>70.73%</td>
<td>0.6500</td>
</tr>
</tbody>
</table>

Impacts of the quantity of generated ECG digital twins.
Here we conducted an experiment to validate the impact of the quantity of generated ECG digital twins on the results. We trained a 1D ResNet model with varying amounts of digital twins and examined the performance discrepancies. As shown in Table 2, it is evident that increasing the quantity of digital twins improves accuracy at both the patient and ECG levels. However, on ATE, it appears that having more generated digital twins does not yield better results. This could be attributed to duplicated cases introducing redundant information, thus compromising the model’s effectiveness in capturing disease-related patterns.

Table 2: The impact of the quantity of generated ECG digital twins on performance.

<table>
<thead>
<tr>
<th>Data Num</th>
<th>Patient Wise</th>
<th>ECG Wise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acc(↑)</td>
<td>F1(↑)</td>
</tr>
<tr>
<td>15553 8600</td>
<td>72.21%</td>
<td>0.6617</td>
</tr>
<tr>
<td>15553 17200</td>
<td>73.55%</td>
<td>0.6690</td>
</tr>
<tr>
<td>15553 34400</td>
<td>73.81%</td>
<td>0.6759</td>
</tr>
</tbody>
</table>

4.3 Generated ECG Fidelity Evaluation
To evaluate the authenticity of the ECGs, we adapt metrics from image generation tasks, specifically the Fréchet Inception Distance (FID) [Heusel et al., 2017] and the Inception Score (IS) [Salimans et al., 2016], traditionally used for assessing the quality of generated images. Diverging from the conventional approach, we utilize a pre-trained 1D ResNet-101 model instead of the Inception V3 [Szegedy et al., 2016] as the backbone for feature extraction from ECG data. Consequently, we introduce the “Fréchet ResNet Distance” (FRD) and “ResNet Score” (RS) as our novel metrics for quantifying the fidelity of ECG synthesis. Lower FRD scores are indicative of a closer resemblance between the distribution of generated and real ECGs, signifying a higher level of authenticity in the synthesized data. As outlined in Table 3, our method distinguishes itself among various generative approaches by achieving the lowest FRD score, markedly surpassing its counterparts. In terms of the RS, where higher
4.4 Privacy Security Performance

In clinical practice and machine learning development, safeguarding patient privacy is also important. We rigorously assess the synthetic ECG digital twins for any risk of presence disclosure, ensuring that the information within these digital twins is non-sensitive and cannot be linked back to the original patient data. To assess the risk of presence disclosure, a scenario where an adversary could deduce the inclusion of specific samples in the model’s training set from synthetic ECGs, we employ the concept of membership inference risk [Shokri et al., 2017]. This type of disclosure becomes a potential issue when an attacker, possessing a subset of target ECGs from the training set, attempts to analyze synthetic ECGs to confirm the presence of these target samples in the training data. Given a distance threshold $\epsilon$, the adversary claims that a target ECG is in the real training set if there exists at least one ECG with a distance less than the threshold. In our experiments, we set the threshold $\epsilon$ based on the mean value of the distance between all ECGs, which can be expressed as $\epsilon = \tau \times mean$. To compute the distance between ECGs, we first extract the ECG features using 1D ResNet-101 pretrained on training set and then compute the Euclidean distance between the features. Finally, we use F1-score as membership inference risk. The higher the F1-score, the higher the risk. As shown in Figure 4, our method demonstrates a lower risk of presence disclosure compared to the other methods across most of thresholds. Notably, although our model uses samples from the training set when generating ECGs, our risk is still lower than other models that do not use samples from the training set, indicating that our model is effective in preventing presence disclosure. Additionally, the line LAVQ-Editor (w/o VQ) shows the performance of the model without VQ-Separator. When compared with the full LAVQ-Editor, there is a heightened risk of presence disclosure. This suggests that the VQ-Separator effectively separates the personal normal feature of the reference patient from their disease-indicative feature, thereby enhancing privacy security.

4.5 Personal Information Retention Performance

To further demonstrate the retention of personal information in the digital twins, we utilized t-SNE to visualize differences between ECG digital twins created using the same reference but for different pre-diagnosis patients. As depicted in Figure 5, “Patient1-MI” represents the reference ECG with Myocardial Infarction derived from Patient1, while “Patient1-MI Digital Twin” refers to an ECG generated for Patient1’s nor-

<table>
<thead>
<tr>
<th>Method</th>
<th>FRD(%)</th>
<th>RS(%)</th>
<th>Precision(%)</th>
<th>Recall(%)</th>
<th>F1-Score(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real</td>
<td>0.3014</td>
<td>16.8195</td>
<td>0.7865</td>
<td>0.9668</td>
<td>0.9287</td>
</tr>
<tr>
<td>CGAN</td>
<td>2.2869</td>
<td>17.8568</td>
<td>0.9668</td>
<td>0.6688</td>
<td>0.7287</td>
</tr>
<tr>
<td>WGAN</td>
<td>3.9171</td>
<td>1.8853</td>
<td>0.9000</td>
<td>0.2250</td>
<td>0.3600</td>
</tr>
<tr>
<td>StyleGAN</td>
<td>5.4443</td>
<td>1.8412</td>
<td>0.8969</td>
<td>0.0781</td>
<td>0.1437</td>
</tr>
<tr>
<td>VQ-VAE2</td>
<td>4.2539</td>
<td>3.5462</td>
<td>0.9094</td>
<td>0.4187</td>
<td>0.6240</td>
</tr>
<tr>
<td>PGANs</td>
<td>3.8477</td>
<td>2.0853</td>
<td>0.7781</td>
<td>0.4750</td>
<td>0.5899</td>
</tr>
<tr>
<td>Ours</td>
<td>1.6286</td>
<td>9.9771</td>
<td>0.8719</td>
<td>0.6438</td>
<td>0.7407</td>
</tr>
</tbody>
</table>

Table 3: Evaluation of the fidelity of ECGs generated by diverse generative models. The best performances are marked in bold.
Figure 4: The F1-score for the membership inference risk. The horizontal coordinate indicates $\tau$, and the vertical coordinate indicates the F1-score. The higher the F1-score, the higher the risk.

Figure 5: t-SNE visualization demonstrating the distinction between ECG digital twins created using the same reference but for different pre-diagnosis patients. Normal ECG using “Patient1-MI” ECG as a reference. Similarly, “Patient2-MI Digital Twin” indicates that Patient2’s normal ECG is used as the pre-diagnosis patient’s ECG. Figure 5 clearly shows greater overlap between the “Patient1-MI” and “Patient1-MI Digital Twin” distributions, whereas “Patient2-MI Digital Twin” is distinctly separated from both, highlighting that the “Patient1-MI Digital Twin” successfully retains personal information of Patient1. Remarkably, this section prioritizes retaining personal information in pre-diagnosis patients’ ECGs. In contrast, Section 4.4 is focused on removing identifiable information from the references. The marked similarity between “Patient1-MI” and “Patient1-MI Digital Twin” and their pronounced distinction from “Patient2-MI Digital Twin” exemplify this strategy effectively.

### 4.6 Ablation Study

In addition, we also perform ablation study to validate the rationale of our framework. By employing Dynamic Time Warping (DTW) [Keogh and Pazzani, 2001] which is acclaimed for its precision in aligning varying temporal sequences, we evaluate the distance between generated ECG digital twin and the real ECGs from pre-diagnosis patients with the same disease which indicates similarity to the pre-diagnosis patients. Given that ECGs naturally differ in length and rhythm, DTW is particularly adept at adapting these time series to closely align their patterns, enhancing the reliability of the comparison. As detailed in Table 4, our ablation study results suggest that vector quantization is pivotal for the retention of personalized disease characteristics in the heart disease detection model. Without it, the model may produce ECGs that closely mirror the original but fail to enhance the detection model effectively. The implementation of both generator’s and discriminator’s disease similarity losses is instrumental in imbuing the generated ECGs with disease traits. However, this narrowly focused approach comes at the expense of individualized patient information, as reflected by the diminished ATE and high DTW observed in the third line of Table 4. As shown in the second and third lines, reconstruction loss ensures the synthesized ECG digital twins retain patient-specific characteristics and are not disproportionately influenced by disease features, which is crucial for achieving lower DTW scores. In addition, the absence of either similarity loss in the generator or discriminator results in unstable adversarial learning dynamics, preventing the model from reaching its optimal performance.

### 5 Conclusions

In this paper, we introduce a novel prospective learning framework for creating personalized ECG digital twins that simulate the heart conditions of diagnosed individuals. For the purpose of realistic ECG generation, we propose a location-aware ECG digital twin generation model based on vector quantization, termed LA VQ-Editor. The model skillfully segregates and edits normal and disease-indicative features in ECGs, preserving individual characteristics while injecting target disease information. This enhances understanding of personalized heart diseases prospectively while safeguarding privacy in model development. To the best of our knowledge, our LA VQ-Editor is the first framework for personalized ECG digital twin generation in contrast to previous work that only generated data at population level. The versatility of our approach extends beyond diagnostics, holding potential for application in ethically-conducted scientific research, such as clinical trial simulation.

<table>
<thead>
<tr>
<th>VQ</th>
<th>$L^G_{sym}$</th>
<th>$L^D_{sym}$</th>
<th>$L_{rec}$</th>
<th>DTW ($\times10^3$)</th>
<th>ATE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>6.664</td>
<td>1.24%</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>44.096</td>
<td>2.76%</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>12.264</td>
<td>1.98%</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>5.772</td>
<td>1.52%</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>5.519</td>
<td>1.35%</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>5.674</td>
<td>2.55%</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>5.322</td>
<td>5.13%</td>
</tr>
</tbody>
</table>

Table 4: Evaluation of the significance of different components of the proposed framework. The best performances are marked in bold. VQ: Vector Quantized Feature Separator; DTW: Dynamic Time Warping; ATE: Average Treatment Effect.
Acknowledgments
This research was supported by Zhejiang Provincial Natural Science Foundation of China (LY23F020019), National Natural Science Foundation of China (62106218 & 82200558), Excellent Young Scientists Fund (82122036), Basic and Applied Basic Research Foundation of Guangdong Province (2022A1515110722), and Guangdong Provincial Medical Science & Technology Research Fund Project (A2023011).

References


[Hou et al., 2019] Ruibing Hou, Hong Chang, Bingpeng Ma, Shiguang Shan, and Xilin Chen. Cross attention network


