Physiology-Informed Diffusion for 12-Lead ECG Generation

Anonymous authors
Paper under double-blind review

Abstract

Large-scale 12-lead ECG data are critical for training reliable cardiac machine learning systems, yet their availability is limited by privacy constraints, annotation cost, and severe class imbalance. Generative models offer a promising solution, but standard diffusion models typically treat ECGs as generic multivariate time series and do not explicitly exploit known physiological structure.

We propose PhysDiff-ECG, a physiology-guided diffusion framework that integrates cardiac ordinary differential equation (ODE) prior into the diffusion trajectory. Our central idea is to make ECG physiology tractable during training by deriving differentiable regularizers from a dynamical model of cardiac activity together with a differentiable 12-lead observation model. Given a denoised reconstruction along the reverse process, PhysDiff-ECG fits a latent physiological explanation via an unrolled inner optimization and penalizes violations of both the simulator dynamics and the induced ECG reconstruction.

This training-time regularization biases the learned denoising trajectories toward physiologically realizable ECGs while preserving the flexibility of latent diffusion. Experiments on standard 12-lead ECG benchmarks show that PhysDiff-ECG improves physiological fidelity, representation-space realism, class-conditional diagnostic consistency, and downstream classification performance relative to strong GAN and diffusion baselines.

1 Introduction

Electrocardiography (ECG) is a fundamental diagnostic modality for assessing cardiac function, with the standard 12-lead ECG providing a rich, multi-view measurement of the heart's electrophysiological activity. Large-scale, diverse ECG datasets are critical for training reliable machine learning systems for tasks such as arrhythmia detection, risk stratification, and clinical decision support. However, access to such data is often severely bottlenecked by privacy concerns, regulatory constraints, and demographic or pathological imbalances (Voigt & Bussche, 2017). These challenges have motivated a growing interest in generative models for synthesizing realistic 12-lead ECG signals that can augment training data and support simulation studies without compromising patient privacy (de Melo et al., 2022; Giuffrè & Shung, 2023).

Generative models, particularly Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) and, more recently, Denoising Diffusion Probabilistic Models (DDPMs) (Ho et al., 2020), have achieved strong results in realistic data synthesis. Diffusion models in particular provide stable training, improved mode coverage, and high-quality samples by reversing a gradual noising process, and have become competitive across images, audio, and sequential data. However, applying standard diffusion models directly to physiological signals exposes a key mismatch. Unlike natural images, whose generative manifold is largely defined by surface statistics, ECG recordings are observations of a structured physical system: cardiac electrophysiology and its body-surface projection (McSharry et al., 2003; Potse, 2018). As a result, treating ECGs as generic multivariate waveforms can yield samples that match local morphology yet violate global constraints, since waveform timing and inter-lead relationships are governed by well-characterized biophysical dynamics and measurement geometry. In this work, by global physiological constraints we refer to properties that cannot be verified from local waveform realism in a single lead alone. These include coordinated beat timing across leads, physiologically plausible waveform morphology, and consistency among the 12 leads as correlated observations of shared cardiac electrical activity. Thus, a generated ECG may appear realistic lead-by-lead

while still being implausible as a 12-lead recording, for example if R-peaks are temporally misaligned across leads or if different leads imply incompatible cardiac activation patterns. Appendix Fig. 3 illustrates typical unconstrained generation failures with physiologically consistent 12-lead structure.

Standard data-driven generative models are largely agnostic to these governing laws. As a result, they may reproduce local ECG morphology while failing to respect the latent dynamical structure and cross-lead relationships induced by cardiac electrophysiology and body-surface projection. In the ECG domain, however, we do have access to tractable physiological models that relate latent cardiac dynamics to observed 12-lead signals. A natural question, then, is how to incorporate such known structure into modern generative models in a way that is differentiable, trainable, and useful for improving the realism and downstream utility of generated ECGs.

Physiological structure in ECG generation is not itself new, and prior work has explored constraints based on lead relationships, simulator-based priors, and physiology-aware diffusion models. Our contribution is therefore not the generic use of physiology, but a specific training-time formulation that combines latent diffusion, differentiable fitting of a low-dimensional physiological explanation, and regularization through both dynamical and observation-model consistency.

In this work, we propose *PhysDiff-ECG*, a diffusion-based framework that bridges data-driven generative modeling and dynamic-based cardiac electrophysiology. Rather than treating physiology as a post-processing constraint or a separate simulator, we incorporate it directly into diffusion training through a differentiable regularization mechanism. Specifically, given a denoised reconstruction along the reverse process, we fit a low-dimensional physiological explanation using a reduced dynamical model together with a differentiable 12-lead observation model, and penalize violations of both the latent dynamics and the reconstructed ECG. In this way, physiology acts as a structured inductive bias on the learned denoising dynamics while preserving the flexibility and scalability of latent diffusion.

Crucially, this physiology-informed regularization provides a structured training signal during denoising. By encouraging denoised reconstructions to admit a plausible physiological explanation, our approach improves biological validity while also improving representation-space fidelity, diagnostic consistency, and downstream usefulness of the generated samples.

Our contributions are threefold:

- **Physiology-Informed Latent Diffusion for 12-Lead ECGs.** We introduce PhysDiff-ECG, a physiology-informed latent diffusion framework that incorporates tractable physiological structure through differentiable regularization derived from a reduced cardiac dynamical model and a differentiable 12-lead observation model.
- **Improved Fidelity and Downstream Utility.** We show that enforcing physiological structure improves physiological fidelity, representation-space realism, class-conditional diagnostic alignment, and downstream ECG classification performance relative to strong GAN and diffusion baselines.
- **Training Stability.** We show that physiology-informed regularization improves sample quality and reduces metric variability across training iterations, relative to the same diffusion backbone without the physiological regularizer.

2 Related Work

The emergence of deep generative models shifted ECG synthesis toward data-driven approaches, with Generative Adversarial Networks (GANs) playing a central role in early work. Methods such as WaveGAN (Donahue et al., 2018) and Pulse2Pulse (Thambawita et al., 2021) demonstrated the feasibility of generating realistic ECG waveforms using adversarial training.

Subsequent work has explored a range of approaches for multi-lead ECG generation, spanning both single-beat synthesis and full 12-lead signal modeling. Vector-quantized variational autoencoders (VQ-VAEs) have been used to learn compact ECG representations (Liu et al., 2020), while 3KG (Gopal et al., 2021) performs data augmentation in vectorcardiogram (VCG) space using three-dimensional geometric transformations. ME-GAN (Chen et al., 2022) extends GAN-based architectures to multi-view ECG synthesis, and (Huang

et al., 2023) incorporates unsupervised noise modeling to improve robustness. More recently, MultiODE-GAN (Yehuda & Radinsky, 2024) combines mechanistic ODE formulations with adversarial learning for multi-lead ECG generation, building on classical dynamical models (McSharry et al., 2003). These approaches improve realism and inter-lead consistency, but many remain tied to single-beat generation.

More recently, diffusion-based generative models have emerged as a strong alternative. Denoising Diffusion Probabilistic Models (DDPMs) (Ho et al., 2020) have achieved strong results in time-series modeling (Rasul et al., 2021; Kollovieh et al., 2023) and have been adapted to biomedical signals, including ECGs. Methods such as SSSD-ECG (Alcaraz & Strodthoff, 2023) apply diffusion modeling to multi-lead ECG generation, while DiffuSETS (Lai et al., 2025) conditions ECG synthesis on auxiliary information such as diagnostic labels, clinical text, or patient-specific metadata.

Despite their empirical success, many existing GAN and diffusion based ECG generators still treat ECGs primarily as generic multivariate time series and rely mainly on statistical structure learned from data. In contrast, earlier biophysical approaches based on dipole models or reaction-diffusion systems (Potse, 2018; Quiroz-Juárez et al., 2019) explicitly encode cardiac electrophysiology, but are not designed to capture the full variability, noise characteristics, and pathological diversity of real-world ECG recordings. This leaves a gap between flexible data-driven generators and physiologically structured simulators.

Incorporating domain knowledge into generative models has also been explored through physics-informed learning and constrained deep generative approaches. Within physics-informed deep learning, Physics-Informed Neural Networks (PINNs) embed differential-equation residuals directly into training objectives, primarily for forward simulation and inverse parameter estimation rather than generative modeling (Raissi et al., 2019; Karniadakis et al., 2021). In cardiac electrophysiology, differentiable ODE solvers have been used to infer patient-specific parameters for ODE/PDE-based models such as FitzHugh-Nagumo (Boulakia et al., 2010), enabling physiologically interpretable reconstruction and personalization (Sayadi et al., 2010; Cantwell et al., 2019). However, these methods are not designed to generate diverse, high-dimensional signals such as full 12-lead ECGs.

Our work is most closely related to physiology-aware diffusion for ECG generation, but differs in how physiological structure is incorporated. Rather than treating physiology as a post-processing constraint or a sampling-time correction, we integrate it into diffusion training through a differentiable regularization mechanism. Given a denoised reconstruction, we fit a low-dimensional physiological explanation using a reduced cardiac dynamical model together with a differentiable 12-lead observation model, and penalize violations of both the latent dynamics and the induced ECG reconstruction.

3 Method

We propose PhysDiff-ECG, a latent diffusion framework for class-conditioned 12-lead electrocardiogram (ECG) generation regularized by a differentiable physiological prior. Our central idea is to make ECG physiology tractable during training by combining a low-dimensional dynamical model of cardiac activity with a differentiable observation model that maps latent physiological states to 12-lead surface ECGs.

Given a full ECG signal, a pretrained encoder maps the signal to a latent representation on which diffusion is performed. The denoiser is trained with the standard diffusion objective together with a physiology-informed regularization term evaluated on denoised reconstructions along the reverse process. This regularizer encourages intermediate reconstructions to admit a plausible physiological explanation, thereby steering the learned denoising dynamics toward physiologically realizable ECGs.

3.1 Problem Setup

Let $S \in \mathbb{R}^{12\times T}$ denote a 12-lead ECG segment with T samples and class label y. In diffusion notation, we write S_0 for the clean signal, where the subscript refers to diffusion time rather than ECG sample index.

We assume that S admits an approximate low-dimensional physiological explanation through a latent trajectory $X(\tau) \in \mathbb{R}^d$, evolving in simulator time τ, and a differentiable observation model $G(\cdot;\theta_{\text{obs}})$ that

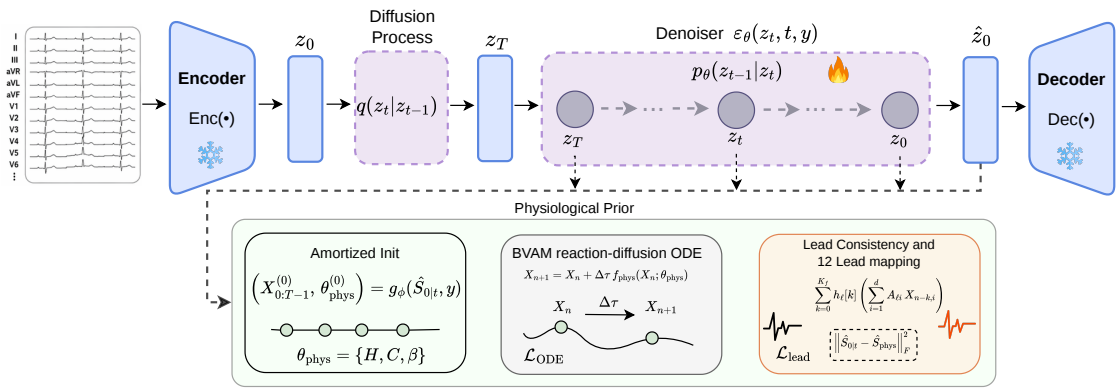


Figure 1: Overview of the physiology-informed latent diffusion framework for 12-lead ECG generation.

maps latent physiological states to the 12-lead surface ECG:

$$\frac{dX(\tau)}{d\tau} = f_{\text{phys}}(X(\tau); \theta_{\text{phys}}),$$

$$S^{\text{phys}}(\tau) = G(X(\tau); \theta_{\text{obs}}).$$

Here, f_{phys} is the reduced cardiac dynamical model, θ_{phys} denotes simulator parameters inferred per reconstruction, and θ_{obs} denotes shared observation-model parameters.

We discretize the simulator on N steps of size $\Delta\tau$, writing $X_n \approx X(n\Delta\tau)$ for $n = 0, \ldots, N-1$. To compare simulator outputs with the observed ECG, we align the simulator grid with the signal length by setting (or resampling to) $N = T$. Throughout, simulator time τ is distinct from diffusion step t: τ indexes the latent physiological trajectory within a single ECG, whereas t indexes the denoising process.

Throughout the paper, t denotes diffusion time, τ denotes simulator time, and n denotes the discrete ECG/sample index. Thus, z_t is the noisy latent variable at diffusion step t, whereas $X_n \approx X(n\Delta\tau)$ is the physiological latent state at simulator step n. The clean ECG is denoted by S_0, and the denoised reconstruction estimated from z_t is denoted by $\hat{S}_{0|t}$.

3.2 Latent Diffusion Preliminaries

We perform diffusion in the latent space of a pretrained VAE. Given a clean ECG signal S_0, the encoder produces a latent code $z_0 = \text{Enc}(S_0)$, and the decoder reconstructs the signal as $\tilde{S}_0 = \text{Dec}(z_0)$. Unless otherwise stated, the VAE is pretrained and kept fixed while training the diffusion model.

For a variance-preserving diffusion process with schedule $\{\alpha_t\}_{t=1}^{K}$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$, the forward process is

$$q(z_t \mid z_0) = \mathcal{N}\big(\sqrt{\bar{\alpha}_t}\, z_0,\, (1 - \bar{\alpha}_t)\mathbf{I}\big). \tag{1}$$

We train a class-conditioned denoiser $\epsilon_\theta(z_t, t, y)$ using the standard noise-prediction objective

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{z_0, t, \varepsilon \sim \mathcal{N}(0,I)} \left[\|\varepsilon - \epsilon_\theta(z_t, t, y)\|_2^2 \right]. \tag{2}$$

Given a noisy latent state z_t, we form the corresponding denoised estimate

$$\hat{z}_{0|t} = \frac{1}{\sqrt{\bar{\alpha}_t}} \left(z_t - \sqrt{1 - \bar{\alpha}_t}\, \epsilon_\theta(z_t, t, y) \right), \qquad \hat{S}_{0|t} = \text{Dec}(\hat{z}_{0|t}). \tag{3}$$

The decoded signal $\hat{S}_{0|t}$ serves as the input to the physiology-informed regularizer introduced in the next subsection.

3.3 Physiological Prior: Reaction–Diffusion System

The first component of the physiological regularizer is a dynamical constraint on the latent cardiac trajectory. Given a denoised reconstruction $\hat{S}_{0|t}$, we infer a discrete latent trajectory $X_{0:T-1}$ that should both explain the reconstructed ECG and remain close to a reduced cardiac dynamical model. We use the BVAM reaction–diffusion ODE as this reduced model, and its vector field f_{phys} is later used to define the ODE-consistency penalty in Section 3.5. As the physiological prior, we use the low-dimensional ODE obtained by spatial discretization of the BVAM reaction–diffusion system of Quiroz-Juárez et al. (2019). This reduced nonlinear system serves as a compact surrogate model of cardiac conduction dynamics. With $X = [x_1, \dots, x_d]^\top \in \mathbb{R}^d$ (specifically $d = 6$ for this system), the dynamics are

$$\dot{x}_1 = x_1 - x_2 - Cx_1x_2 - x_1x_2^2, \tag{4}$$

$$\dot{x}_2 = Hx_1 - 3x_2 + Cx_1x_2 + x_1x_2^2 + \beta\,(x_4 - x_2), \tag{5}$$

$$\dot{x}_3 = x_3 - x_4 - Cx_3x_4 - x_3x_4^2, \tag{6}$$

$$\dot{x}_4 = Hx_3 - 3x_4 + Cx_3x_4 + x_3x_4^2 + \beta\,(x_6 - 2x_4 + x_2), \tag{7}$$

$$\dot{x}_5 = x_5 - x_6 - Cx_5x_6 - x_5x_6^2, \tag{8}$$

$$\dot{x}_6 = Hx_5 - 3x_6 + Cx_5x_6 + x_5x_6^2 + \beta\,(x_4 - x_6), \tag{9}$$

where $\theta_{\text{phys}} = \{H, C, \beta\}$. Following Quiroz-Juárez et al. (2019), H controls the oscillatory regime and effective heart rate, C controls the nonlinearity of the restitution dynamics, and β determines the coupling strength induced by the discrete Laplacian.

The BVAM prior comes from a spatially discretized reaction-diffusion model of cardiac electrical activity and provides a compact system of coupled nonlinear oscillators for generating ECG-like dynamics. We do not use it as a full cellular, anatomical, or patient-specific electrophysiology simulator. In particular, it does not model detailed cellular ionic dynamics, tissue anisotropy, torso geometry, electrode placement variability, or patient-specific conduction.

We use BVAM as a differentiable surrogate that captures coarse dynamical properties relevant to ECG generation, including rhythmic excitation, nonlinear restitution-like behavior, and coupling between latent physiological components. The ODE-consistency term is therefore applied as a soft regularizer rather than a hard constraint, allowing the diffusion model to retain data-driven flexibility when the reduced prior is imperfect.

We treat θ_{phys} as latent variables inferred from the denoised reconstruction. Class information enters the overall model through the class-conditioned denoiser $\epsilon_\theta(z_t, t, y)$ and may also be provided to the initializer g_ϕ.

Numerical integration. To evaluate the ODE-based physiology regularizer, we approximate the continuous latent dynamics with an explicit Euler discretization. Given the current latent state X_n, the next state is approximated as

$$X_{n+1} = X_n + \Delta\tau\, f_{\text{phys}}(X_n; \theta_{\text{phys}}), \quad n = 0, \dots, N-2. \tag{10}$$

This first-order solver provides a simple and differentiable approximation of the ODE trajectory, allowing the regularizer to penalize deviations from the prescribed physiological dynamics. Although higher-order solvers could be used, Euler integration keeps the computation lightweight and the residual easy to optimize.

3.4 Differentiable 12-Lead Mapping

A limitation of the original BVAM formulation is that it produces only a single ECG lead through a scalar linear combination of the latent states, $\text{ECG}(t) = \sum_{i=1}^{d} \alpha_i x_i$. To generate full 12-lead ECGs, we replace this scalar observation model with a differentiable multi-lead mapping.

In the simplest case, we use a shared linear mixing matrix $A \in \mathbb{R}^{12\times d}$:

$$\hat{S}_{\text{phys}}[\ell, n] = \sum_{i=1}^{d} A_{\ell i}\, X_{n,i}, \tag{11}$$

where $\ell \in \{1, \ldots, 12\}$ and $n \in \{0, \ldots, T-1\}$. This model reduces to the original single-lead formulation when selecting one row of A. The matrix A is shared across samples and learned jointly with the rest of the model.

This linear observation model serves as a flexible approximation to the ECG lead-field operator: each lead is modeled as a distinct projection of a shared low-dimensional latent trajectory. Accordingly, the 12 leads are treated as correlated views of common underlying cardiac dynamics rather than as independent channels.

To capture small lead-specific temporal delays and smoothing effects, we optionally augment the observation model with a short causal FIR filter $h_\ell \in \mathbb{R}^{K_f+1}$ for each lead:

$$\hat{S}_{\text{phys}}[\ell, n] = \sum_{k=0}^{K_f} h_\ell[k] \left(\sum_{i=1}^{d} A_{\ell i} \, X_{n-k,i} \right). \tag{12}$$

For out-of-range indices with $n - k < 0$, we use boundary handling via $X_{n-k} = X_0$ (or zero-padding). With this extension, the full observation parameters are $\theta_{\text{obs}} = \{A, \{h_\ell\}_{\ell=1}^{12}\}$, shared across samples and learned jointly with the rest of the model. Setting $h_\ell[k] = \mathbb{1}[k = 0]$ recovers the purely linear model in equation 11.

3.5 Physiology-Informed Regularization

Given a denoised reconstruction $\hat{S}_{0|t} \in \mathbb{R}^{12 \times T}$, we define a physiology-informed regularizer by fitting a latent physiological trajectory that explains the reconstruction through the simulator and observation model.

We initialize this inner fitting problem using an amortized network g_ϕ:

$$\left(X^{(0)}_{0:T-1}, \theta^{(0)}_{\text{phys}} \right) = g_\phi(\hat{S}_{0|t}, y).$$

The network predicts an initial latent trajectory $X^{(0)}_{0:T-1}$ and simulator parameters $\theta^{(0)}_{\text{phys}}$, where (0) denotes initialization before refinement. We then apply M unrolled gradient steps to obtain $(\hat{X}_{0:T-1}, \hat{\theta}_{\text{phys}})$, yielding a differentiable simulator-consistent explanation of $\hat{S}_{0|t}$.

ODE consistency. We encourage the inferred latent trajectory to follow the simulator dynamics through the Euler residual

$$\mathcal{L}_{\text{ODE}} = \sum_{n=0}^{T-2} \left\| \hat{X}_{n+1} - \left(\hat{X}_n + \Delta\tau \, f_{\text{phys}}(\hat{X}_n; \hat{\theta}_{\text{phys}}) \right) \right\|_2^2. \tag{13}$$

Lead consistency. Let $\hat{S}_{\text{phys}} = \mathcal{G}(X; \theta_{\text{obs}})$ denote the ECG reconstructed from the inferred latent trajectory using equation 11 or equation 12. We define the signal-consistency term over all 12 leads as

$$\mathcal{L}_{\text{lead}} = \left\| \hat{S}_{0|t} - \hat{S}_{\text{phys}} \right\|_F^2. \tag{14}$$

This encourages the denoised reconstruction to be globally consistent with the physiological observation model across all leads.

Inner physiology objective. We combine the two terms into

$$\mathcal{L}_{\text{inner}} = \lambda_{\text{ODE}} \, \mathcal{L}_{\text{ODE}} + \lambda_{\text{lead}} \, \mathcal{L}_{\text{lead}}. \tag{15}$$

In practice, $(\hat{X}_{0:T-1}, \hat{\theta}_{\text{phys}})$ are obtained by approximately solving

$$(\hat{X}_{0:T-1}, \hat{\theta}_{\text{phys}}) \approx \arg \min_{X_{0:T-1}, \, \theta_{\text{phys}}} \mathcal{L}_{\text{inner}}(\hat{S}_{0|t}; X_{0:T-1}, \theta_{\text{phys}}, y) \tag{16}$$

This inner optimization induces a reconstruction-dependent physiological explanation

$$(\hat{X}_{0:T-1}(\hat{S}_{0|t}), \hat{\theta}_{\text{phys}}(\hat{S}_{0|t})).$$

The outer physiology regularizer used during diffusion training is then defined as

$$\mathcal{L}_{\text{phys}}(\hat{S}_{0|t}) := \mathcal{L}_{\text{inner}}\big(\hat{S}_{0|t}; \hat{X}_{0:T-1}(\hat{S}_{0|t}), \hat{\theta}_{\text{phys}}(\hat{S}_{0|t}), y\big).$$

Because the inner optimization is unrolled, $\mathcal{L}_{\text{phys}}(\hat{S}_{0|t})$ is differentiable with respect to $\hat{S}_{0|t}$ and can be backpropagated through the denoising network during training.

In practice, we use a small fixed number of unrolled refinement steps ($M \ll T$), which keeps the additional computational overhead moderate relative to the diffusion backbone.

3.6 Training Objective and Reverse-Process Coupling

For each sampled diffusion timestep, the denoiser is trained to predict the diffusion noise while its implied clean reconstruction is regularized to admit a plausible physiological explanation under the simulator and observation model (see Fig. 1).

Given a noisy latent z_t, the denoiser estimates $\hat{z}_{0|t}$ via equation 3, which is decoded into $\hat{S}_{0|t} = \text{Dec}(\hat{z}_{0|t})$. We then infer a latent physiological explanation $(\hat{X}, \hat{\theta}_{\text{phys}})$ for $\hat{S}_{0|t}$ using the unrolled optimization in equation 16, and evaluate the regularization loss $\mathcal{L}_{\text{phys}}(\hat{S}_{0|t})$. This loss is added to the diffusion objective and backpropagated through the denoiser, thereby coupling latent denoising with physiological consistency.

The regularized denoising objective at timestep t is

$$\ell_t(\theta) = \|\varepsilon - \epsilon_\theta(z_t, t, y)\|_2^2 + \lambda_{\text{phys}}\, \mathcal{L}_{\text{phys}}(\hat{S}_{0|t}), \tag{17}$$

where $\hat{S}_{0|t} = \text{Dec}(\hat{z}_{0|t})$ and $\hat{z}_{0|t}$ is defined by equation 3; the dependence of $\mathcal{L}_{\text{phys}}$ on the inner unrolled refinement is implicit.

Averaging over training samples, diffusion timesteps, and Gaussian noise yields the full objective

$$\mathcal{L}_{\text{total}} = \mathbb{E}_{S_0,y,t,\varepsilon}\left[\|\varepsilon - \epsilon_\theta(z_t, t, y)\|_2^2 + \lambda_{\text{phys}}\, \mathcal{L}_{\text{phys}}(\hat{S}_{0|t})\right], \quad z_0 = \text{Enc}(S_0),\ z_t = \sqrt{\bar{\alpha}_t}\, z_0 + \sqrt{1-\bar{\alpha}_t}\, \varepsilon. \tag{18}$$

This objective is backpropagated through the chain

$$z_t \mapsto \hat{z}_{0|t} \mapsto \hat{S}_{0|t} \mapsto (\hat{X}, \hat{\theta}_{\text{phys}}),$$

thereby encouraging the denoiser to produce reverse-process trajectories whose denoised reconstructions are consistent with both the latent cardiac dynamics and the differentiable 12-lead observation model.

We optimize over the full latent trajectory $X_{0:T-1}$ rather than only an initial condition in order to keep the inner problem flexible and numerically stable when fitting noisy intermediate reconstructions. The ODE residual therefore acts as a soft dynamical constraint, encouraging—but not strictly forcing—the refined trajectory to remain close to the simulator manifold.

Importantly, $\mathcal{L}_{\text{phys}}$ is not introduced as a post-hoc sampling correction, but as a training-time regularizer on the learned denoising dynamics. At inference time, sampling proceeds using the learned denoiser, which has already been biased toward physiologically realizable trajectories during training.

Unlike label-conditioning terms that inject class information into the reverse process, our physiological term does not encode semantic class preferences. Instead, it acts as a structured model-based regularizer derived from a low-dimensional cardiac dynamics prior and a differentiable multi-lead observation model, penalizing denoised reconstructions that fail to admit a plausible physiological explanation.

In each training iteration, we sample an ECG S_0 and label y, encode it to z_0, sample a diffusion timestep t and noise ε, construct z_t via equation 1, predict $\hat{z}_{0|t}$ and decode $\hat{S}_{0|t}$, run M unrolled refinement steps to obtain $(\hat{X}, \hat{\theta}_{\text{phys}})$, and optimize equation 18.

4 Experimental Setup

4.1 Model Architecture

PhysDiff-ECG builds on the latent diffusion architecture of DiffuSETS (Lai et al., 2025) and adapts it to class-conditional 12-lead ECG generation. The model consists of a pretrained VAE and a class-conditioned denoising diffusion model operating in the VAE latent space. Given an ECG $\mathbf{x} \in \mathbb{R}^{12\times T}$, the encoder maps it to a latent code $\mathbf{z}_0 = \mathrm{Enc}(\mathbf{x}) \in \mathbb{R}^{C\times L}$, and the decoder reconstructs $\hat{\mathbf{x}} = \mathrm{Dec}(\mathbf{z}_0)$. In our implementation, we use $C = 6$ latent channels and latent length $L = 128$. Additional architectural details are provided in Appendix B.

4.2 ECG Datasets

We evaluate PhysDiff-ECG on two standard 12-lead ECG benchmarks, using one as the primary development dataset and the other for external validation. Our primary experiments are conducted on PTB-XL (Wagner et al., 2020), a large-scale benchmark for 12-lead ECG analysis. It contains 21,799 recordings from 18,869 patients, with each recording spanning 10 seconds and annotated with diagnostic labels. We use PTB-XL as the main dataset for training the generative model and for downstream synthetic-data evaluation. For additional validation, we also evaluate on the Georgia 12-Lead ECG Challenge dataset (G12EC) (Alday et al., 2020). It contains 10,344 recordings from 7,871 patients, each 10 seconds long and sampled at 500 Hz. We use G12EC to test whether the conclusions observed on PTB-XL generalize to a second large-scale 12-lead ECG cohort.

4.3 Evaluation Protocol

Data Splits and Cross-Validation. To prevent data leakage, we use patient-wise splits throughout. For each dataset, 20% of patients are held out for final testing (Xu & Goodacre, 2018), and the remaining 80% form the development set, within which we perform 5-fold cross-validation. Additional preprocessing and split details are provided in Appendix B.

4.4 Downstream Classifiers for Synthetic-Data Evaluation

To assess the utility of synthetic ECGs generated by PhysDiff-ECG, we evaluate downstream classification performance using two strong 12-lead ECG classifiers: a ResNet-based model (Ribeiro et al., 2020) and an attention-augmented ResNet (Nejedly et al., 2021). Using two distinct classifier families reduces the chance that improvements from synthetic augmentation are specific to a single downstream architecture. Architectural details are provided in Appendix B.

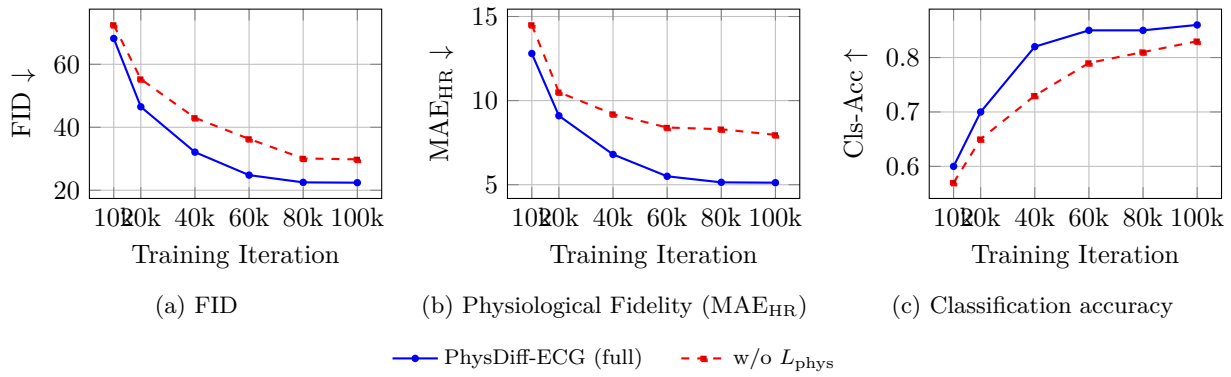
5 Experimental Results

We evaluate PhysDiff-ECG along three complementary axes: (i) signal- and feature-level fidelity, (ii) class-conditional diagnostic alignment, and (iii) downstream classification performance under synthetic-data augmentation.

5.1 Compared Generative Models

We compare PhysDiff-ECG against the following generative baselines:

- **WaveGAN** (Donahue et al., 2018): Originally proposed for raw audio synthesis, WaveGAN models temporal dependencies in one-dimensional signals and is adapted here for multi-lead ECG waveform generation.
- **Pulse2Pulse** (Thambawita et al., 2021): A conditional GAN employing a U-Net style encoder-decoder with one-dimensional convolutions for ECG waveform generation.
- **SSSD-ECG** (Alcaraz & Strodthoff, 2023): A diffusion-based ECG generator that leverages structured state-space (S4) layers within a DDPM framework to model long-range temporal dependencies.



(a) FID (b) Physiological Fidelity (MAE$_{\text{HR}}$) (c) Classification accuracy

Figure 2: Training-Iterations evaluation of ECG generation quality. We compare the full PhysDiff-ECG framework with the same diffusion backbone trained without the physiology-informed regularizer. We report FID ,heart-rate mean absolute error MAE$_{\text{HR}}$ and classifier accuracy.

- **DiffuSETS** (Lai et al., 2025): A diffusion framework originally designed for text-to-ECG synthesis using LLM–derived semantic embeddings, adapted here to a class-conditioned generation.
- **Diffusion-TS** (Yuan & Qiao, 2024): A diffusion model for multivariate time-series generation based on encoder-decoder transformer with disentangled temporal representations. Supports both conditional and unconditional generation.
- **PhysDiff-ECG**: Our class-conditioned latent diffusion model with physiology-informed regularization, which incorporates a cardiac ODE prior.

5.2 Signal and Feature-Level Evaluation

Physiological Fidelity (Heart-Rate Preservation). To assess whether generated ECGs preserve basic physiological characteristics, we compare the heart rate (HR) extracted from generated samples $\hat{S}$ to that of matched real samples from the same class. We report the mean absolute error MAE$_{\text{HR}}$ in beats per minute (bpm), where lower values indicate better HR preservation. Heart rate is computed using a standard R-peak detector and converting the median RR interval to bpm.

Signal- and Feature-Level Distribution Similarity. We assess distributional realism using Fréchet distance in a learned ECG feature space. Real and generated ECGs are embedded with a fixed pretrained Net1D encoder, and the Fréchet distance between the resulting Gaussian fits is computed as

$$\text{FID} = \|\mu_r - \mu_g\|_2^2 + \text{Tr}\Big(\Sigma_r + \Sigma_g - 2\,(\Sigma_r \Sigma_g)^{1/2}\Big)\,,$$

where (μ_r, Σ_r) and (μ_g, Σ_g) are the empirical mean and covariance of the embedded real and generated samples. Lower FID is better.

Class-Conditional Diagnostic Alignment. Because PhysDiff-ECG is class-conditioned, we also test whether generated ECGs are diagnostically consistent with their conditioning label y. To do so, we apply a fixed pretrained ECG classifier to generated samples and report accuracy.

Table 1 summarizes MAE$_{\text{HR}}$, representation-space FID, and classifier accuracy across all generative models.

5.3 Downstream Classification Performance

To evaluate downstream utility, we train classifiers on either real ECGs only or the same real ECGs augmented with synthetic samples generated by each baseline and by PhysDiff-ECG, and test exclusively on unseen real ECGs (Alcaraz & Strodthoff, 2023; Yehuda & Radinsky, 2024). We report per-abnormality sensitivity,

Table 1: Evaluation of ECG generation quality across methods. We report heart-rate mean absolute error MAE_{HR} in bpm, representation-space Fréchet distance (FID), and classifier accuracy (Cls-Acc). Lower is better for MAE_{HR} and FID, while higher is better for Cls-Acc.

Model	$\textbf{MAE}_{\textbf{HR}}$ (↓)	**FID** (↓)	**Cls-Acc** (↑)
WaveGAN	15.70	50.4	0.71
Pulse2Pulse	12.01	41.2	0.76
SSSD-ECG	13.82	39.4	0.77
DiffuSETS (class-cond.)	9.22	29.8	0.82
Diffusion-TS	10.18	34.3	0.80
PhysDiff-ECG	**5.14**	**22.4**	**0.86**

specificity, and AUC in a one-vs-rest setting (Wang et al., 2018; Bressman et al., 2020; Golany et al., 2022). For sensitivity and specificity, thresholds are selected on the validation split and applied to the test set.

Table 2 reports results for the ECG classifier (Ribeiro et al., 2020). The Baseline CLS column uses only real training data, whereas all other columns augment the same real data with synthetic samples generated by the corresponding model. For each abnormality, all methods use the same synthetic augmentation size, set to N, where N is the number of real training samples for that abnormality. Augmenting with PhysDiff-ECG improves performance over the Real Only setting and is competitive with or better than alternative generative baselines on most abnormalities. Gains are typically larger for lower-prevalence conditions, where limited real data constrains generalization. Results with an additional downstream classifier are reported in Sec. C.4.

Table 2: Downstream classification performance (evaluated on a held-out **real** test set). We report per-abnormality sensitivity (Sens.), specificity (Spec.), and AUC for the baseline classifier and for classifiers trained with synthetic augmentation from each Model. Boldface indicates the best value per row within each metric.

Abnormality	Baseline CLS*			**Pulse2Pulse**			**SSSD-ECG**			**DiffuSETS**			**PhysDiff-ECG**		
	Sens.	Spec.	AUC	Sens.	Spec.	AUC	Sens.	Spec.	AUC	Sens.	Spec.	AUC	Sens.	Spec.	AUC
AFL	0.80	0.86	0.84	0.81	0.86	0.85	0.83	0.88	0.87	0.83	0.87	0.87	**0.85**	**0.90**	**0.89**
TAb	0.89	0.70	0.88	0.90	0.71	0.89	0.92	0.72	0.91	0.93	0.71	0.91	**0.94**	**0.73**	**0.93**
QAb	0.84	0.73	0.87	0.84	0.74	0.87	0.86	0.74	0.88	0.85	0.75	0.89	0.86	**0.77**	**0.90**
SA	0.67	0.53	0.65	0.68	0.55	0.67	0.68	0.58	0.68	0.70	0.57	0.69	0.70	**0.61**	**0.71**
LAD	0.90	0.87	0.95	0.90	0.88	0.95	0.91	0.88	0.96	0.90	0.90	0.96	**0.93**	0.90	**0.97**
LBBB	0.98	0.97	0.99	0.98	0.96	0.99	0.98	0.97	0.99	0.98	0.97	0.99	0.98	0.97	0.99
PAC	0.88	0.62	0.79	0.88	0.64	0.80	0.89	0.64	0.81	0.88	0.65	0.81	**0.90**	**0.67**	**0.84**
NSIVCB	0.81	0.67	0.82	0.82	0.67	0.83	0.84	0.68	0.84	0.84	0.67	0.84	0.84	**0.71**	**0.85**

* Baseline classifier follows Ribeiro et al. (2020)

Convergence During Training

We track FID, MAE_{HR}, and class-conditional diagnostic accuracy across training epochs in order to understand how the physiology-informed regularizer affects optimization. Specifically, we compare the full model against the variant trained without physiological regularization.

Fig. 2 shows that physiology-informed regularization improves these metrics throughout training and often reduces performance variance across steps. This suggests that the physiological prior provides a stabilizing inductive bias during training, improving intermediate sample quality in addition to final performance.

Additional ablations, including loss-component analysis, evaluation on the G12EC dataset, and robustness across downstream classifiers, are reported in Appendix C. Qualitative results are provided in Appendix A.

6 Conclusion

We presented PhysDiff-ECG, a physiology-guided diffusion framework for class-conditioned 12-lead ECG synthesis. Rather than treating physiology as a post-processing constraint, PhysDiff-ECG incorporates a low-dimensional cardiac dynamical model and a differentiable 12-lead observation model directly into training through unrolled physiological fitting and consistency-based regularization. Empirically, this formulation improves representation-space realism, physiological fidelity, class-conditional diagnostic alignment, and downstream classification performance under synthetic-data augmentation relative to strong GAN and diffusion baselines. Our analysis further indicates that physiology-informed diffusion improves sample quality and reduces metric variability across training checkpoints, which is consistent with faster convergence.

A limitation of PhysDiff-ECG is its reliance on the informativeness of the chosen physiological prior. If the reduced ODE or observation model is severely misspecified, the regularizer may provide a weaker signal or bias generation toward an incomplete physiological manifold. Future work should explore richer and more adaptive electrophysiological priors, including patient-specific parameterizations and uncertainty-aware physiological constraints.We discuss broader impact, clinical-use limitations, privacy, bias, and responsible-use considerations in Appendix D.

References

Juan Miguel Lopez Alcaraz and Nils Strodthoff. Diffusion-based conditional ecg generation with structured state space models. *Computers in Biology and Medicine*, 163:107115, 2023. ISSN 0010-4825. doi: https://doi.org/10.1016/j.compbiomed.2023.107115. URL https://www.sciencedirect.com/science/article/pii/S0010482523005802.

Erick A Perez Alday, Annie Gu, Amit J Shah, Chad Robichaux, An-Kwok Ian Wong, Chengyu Liu, Feifei Liu, Ali Bahrami Rad, Andoni Elola, Salman Seyedi, et al. Classification of 12-lead ecgs: the physionet/computing in cardiology challenge 2020. *Physiological measurement*, 41(12):124003, 2020.

Muriel Boulakia, Serge Cazeau, Miguel Angel Fernández, Jean-Frédéric Gerbeau, and Nejib Zemzemi. Mathematical modeling of electrocardiograms: A numerical study. *Annals of Biomedical Engineering*, 38 (3):1071–1097, 2010. doi: 10.1007/s10439-009-9873-0. INRIA preprint inria-00400490v2.

Maxwell Bressman, Alon Y. Mazori, Eric Shulman, Jay J. Chudow, Ythan Goldberg, John D. Fisher, Kevin J. Ferrick, Mario Garcia, Luigi Di Biase, and Andrew Krumerman. Determination of sensitivity and specificity of electrocardiography for left ventricular hypertrophy in a large, diverse patient population. *The American Journal of Medicine*, 133(9):e495–e500, 2020. ISSN 0002-9343. doi: https://doi.org/10.1016/j.amjmed.2020.01.042. URL https://www.sciencedirect.com/science/article/pii/S0002934320302023.

Chris D. Cantwell, Yumnah Mohamied, Konstantinos N. Tzortzis, Stef Garasto, Charles Houston, Rasheda A. Chowdhury, Fu Siong Ng, Anil A. Bharath, and Nicholas S. Peters. Rethinking multiscale cardiac electrophysiology with machine learning and predictive modelling. *Computers in Biology and Medicine*, 104:339–351, 2019. ISSN 0010-4825. doi: https://doi.org/10.1016/j.compbiomed.2018.10.015. URL https://www.sciencedirect.com/science/article/pii/S0010482518303147.

Jintai Chen, Kuanlun Liao, Kun Wei, Haochao Ying, Danny Z Chen, and Jian Wu. ME-GAN: Learning panoptic electrocardio representations for multi-view ECG synthesis conditioned on heart diseases. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 3360–3370. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/chen22n.html.

Celso M. de Melo, Antonio Torralba, Leonidas Guibas, James DiCarlo, Rama Chellappa, and Jessica Hodgins. Next-generation deep learning based on simulators and synthetic data. *Trends in Cognitive Sciences*, 26(2):174–187, 2022. ISSN 1364-6613. doi: https://doi.org/10.1016/j.tics.2021.11.008. URL https://www.sciencedirect.com/science/article/pii/S136466132100293X.

Chris Donahue, Julian J. McAuley, and Miller S. Puckette. Synthesizing audio with generative adversarial networks. *CoRR*, abs/1802.04208, 2018. URL http://arxiv.org/abs/1802.04208.

Mauro Giuffrè and Dennis L. Shung. Harnessing the power of synthetic data in healthcare: innovation, application, and privacy. *npj Digital Medicine*, 6(1):186, October 9 2023. ISSN 2398-6352. doi: 10.1038/s41746-023-00927-3. URL https://doi.org/10.1038/s41746-023-00927-3.

T. Golany, K. Radinsky, N. Kofman, I. Litovchik, R. Young, A. Monayer, I. Love, F. Tziporin, I. Minha, Y. Yehuda, T. Ziv-Baran, S. Fuchs, and S. Minha. Physicians and machine-learning algorithm performance in predicting left-ventricular systolic dysfunction from a standard 12-lead electrocardiogram. *Journal of Clinical Medicine*, 11(22):6767, 2022. doi: 10.3390/jcm11226767.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger (eds.), *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.

Bryan Gopal, Ryan Han, Gautham Raghupathi, Andrew Ng, Geoff Tison, and Pranav Rajpurkar. 3kg: Contrastive learning of 12-lead electrocardiograms using physiologically-inspired augmentations. In Subhrajit Roy, Stephen Pfohl, Emma Rocheteau, Girmaw Abebe Tadesse, Luis Oala, Fabian Falck, Yuyin Zhou, Liyue Shen, Ghada Zamzmi, Purity Mugambi, Ayah Zirikly, Matthew B. A. McDermott, and Emily Alsentzer (eds.), *Proceedings of Machine Learning for Health*, volume 158 of *Proceedings of Machine Learning Research*, pp. 156–167. PMLR, 04 Dec 2021. URL https://proceedings.mlr.press/v158/gopal21a.html.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf.

Shaobin Huang, Peng Wang, and Rongsheng Li. Noise ecg generation method based on generative adversarial network. *Biomedical Signal Processing and Control*, 81:104444, 2023. ISSN 1746-8094. doi: https://doi.org/10.1016/j.bspc.2022.104444. URL https://www.sciencedirect.com/science/article/pii/S1746809422008989.

George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021. ISSN 2522-5820. doi: 10.1038/s42254-021-00314-5. URL https://doi.org/10.1038/s42254-021-00314-5.

Marcel Kollovieh, Abdul Fatir Ansari, Michael Bohlke-Schneider, Jasper Zschiegner, Hao Wang, and Yuyang (Bernie) Wang. Predict, refine, synthesize: Self-guiding diffusion models for probabilistic time series forecasting. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 28341–28364. Curran Associates, Inc., 2023.

Yongfan Lai, Jiabo Chen, Qinghao Zhao, Deyun Zhang, Yue Wang, Shijia Geng, Hongyan Li, and Shenda Hong. Diffusets: 12-lead ecg generation conditioned on clinical text reports and patient-specific information. *Patterns*, 6(10):101291, October 2025. ISSN 2666-3899. doi: 10.1016/j.patter.2025.101291. URL http://dx.doi.org/10.1016/j.patter.2025.101291.

Han Liu, Zhengbo Zhao, Xiao Chen, Rong Yu, and Qiang She. Using the vq-vae to improve the recognition of abnormalities in short-duration 12-lead electrocardiogram records. *Computer Methods and Programs in Biomedicine*, 196:105639, 2020. ISSN 0169-2607. doi: https://doi.org/10.1016/j.cmpb.2020.105639.

Patrick E McSharry, Gari D Clifford, Lionel Tarassenko, and Leonard A Smith. A dynamical model for generating synthetic electrocardiogram signals. *IEEE transactions on biomedical engineering*, 50(3):289–294, 2003.

Petr Nejedly, Adam Ivora, Radovan Smisek, Ivo Viscor, Zuzana Koscova, Pavel Jurak, and Filip Plesinger. Classification of ecg using ensemble of residual cnns with attention mechanism. In *2021 Computing in Cardiology (CinC)*, volume 48, pp. 1–4, 2021.

Mark Potse. Scalable and accurate ecg simulation for reaction-diffusion models of the human heart. *Frontiers in Physiology*, 9:370, April 2018. doi: 10.3389/fphys.2018.00370.

M. A. Quiroz-Juárez, O. Jiménez-Ramírez, R. Vázquez-Medina, V. Breña-Medina, J. L. Aragón, and R. A. Barrio. Generation of ecg signals from a reaction-diffusion model spatially discretized. *Scientific Reports*, 9(1):19000, 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-55448-5. URL https://doi.org/10.1038/s41598-019-55448-5.

M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2018.10.045. URL https://www.sciencedirect.com/science/article/pii/S0021999118307125.

Kashif Rasul, Calvin Seward, Ingmar Schuster, and Roland Vollgraf. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8857–8868. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/rasul21a.html.

Antônio H Ribeiro, Manoel Horta Ribeiro, Gabriela MM Paixão, Derick M Oliveira, Paulo R Gomes, Jéssica A Canazart, Milton PS Ferreira, Carl R Andersson, Peter W Macfarlane, Wagner Meira Jr, et al. Automatic diagnosis of the 12-lead ecg using a deep neural network. *Nature communications*, 11(1):1–9, 2020.

Omid Sayadi, Mohammad B. Shamsollahi, and Gari D. Clifford. Synthetic ecg generation and bayesian filtering using a gaussian wave-based dynamical model. *Physiological Measurement*, 31(10):1309–1329, 2010. doi: 10.1088/0967-3334/31/10/002.

Vajira Thambawita, Jonas L. Isaksen, Steven A. Hicks, Jonas Ghouse, Gustav Ahlberg, Allan Linneberg, Niels Grarup, Christina Ellervik, Morten Salling Olesen, Torben Hansen, Claus Graff, Niels-Henrik Holstein-Rathlou, Inga Strümke, Hugo L. Hammer, Mary M. Maleckar, Pål Halvorsen, Michael A. Riegler, and Jørgen K. Kanters. Deepfake electrocardiograms using generative adversarial networks are the beginning of the end for privacy issues in medicine. *Scientific Reports*, 11(1):21896, 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-01295-2. URL https://doi.org/10.1038/s41598-021-01295-2.

Paul Voigt and Axel von dem Bussche. *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Springer Publishing Company, Incorporated, 1st edition, 2017. ISBN 3319579584.

Patrick Wagner, Nils Strodthoff, Ralf-Dieter Bousseljot, Dieter Kreiseler, Fatima I. Lunze, Wojciech Samek, and Tobias Schaeffter. Ptb-xl, a large publicly available electrocardiography dataset. *Scientific Data*, 7 (1):154, May 2020. ISSN 2052-4463. doi: 10.1038/s41597-020-0495-6. URL https://doi.org/10.1038/s41597-020-0495-6.

John J. Wang, Olle Pahlm, James W. Warren, John L. Sapp, and B. Milan Horáček. Criteria for ecg detection of acute myocardial ischemia: Sensitivity versus specificity. *Journal of Electrocardiology*, 51(6, Supplement):S12–S17, 2018. ISSN 0022-0736. doi: https://doi.org/10.1016/j.jelectrocard.2018.08.018. URL https://www.sciencedirect.com/science/article/pii/S002207361830339X.

Yun Xu and Royston Goodacre. On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *Journal of Analysis and Testing*, 2(3):249–262, July 1 2018. ISSN 2509-4696. doi: 10.1007/s41664-018-0068-2. URL https://doi.org/10.1007/s41664-018-0068-2.

Yakir Yehuda and Kira Radinsky. Ordinary differential equations for enhanced 12-lead ecg generation, 2024. URL https://arxiv.org/abs/2409.17833.

Xinyu Yuan and Yan Qiao. Diffusion-TS: Interpretable diffusion for general time series generation. In *The Twelfth International Conference on Learning Representations*, 2024. URL `https://openreview.net/forum?id=4h1apFjO99`.

A Qualitative Results

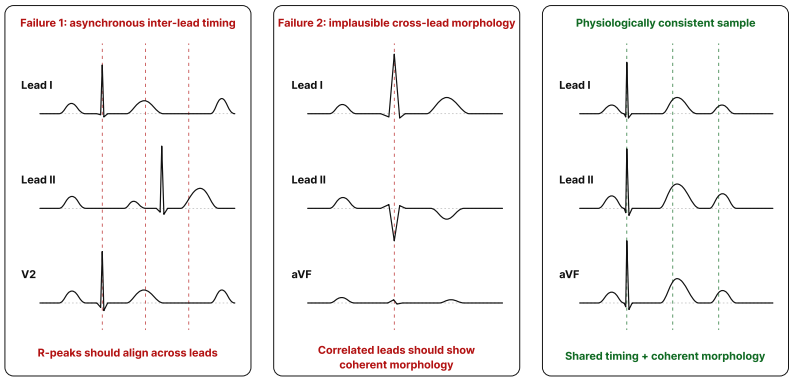


Figure 3: Illustration of global physiological constraints in 12-lead ECG generation. Unconstrained generation may produce locally plausible leads while violating synchronized inter-lead timing or coherent cross-lead morphology. PhysDiff-ECG encourages generated signals to share a consistent latent physiological trajectory.

Figure 3 illustrates two common failure modes: asynchronous R-peaks across leads and implausible morphology among correlated leads, both of which can occur even when individual leads appear locally realistic.

Figures 4–5 present two representative 12-lead ECGs generated by PhysDiff-ECG. Each figure shows a complete 12-lead recording in standard clinical order (I, II, III, aVR, aVL, aVF, V1–V6). The samples display coherent temporal rhythms across leads, realistic waveform morphology, and consistent inter-lead structure.

These qualitative examples complement the quantitative results in the main paper, demonstrating that PhysDiff-ECG yields globally coherent ECGs rather than independently plausible lead-wise signals.

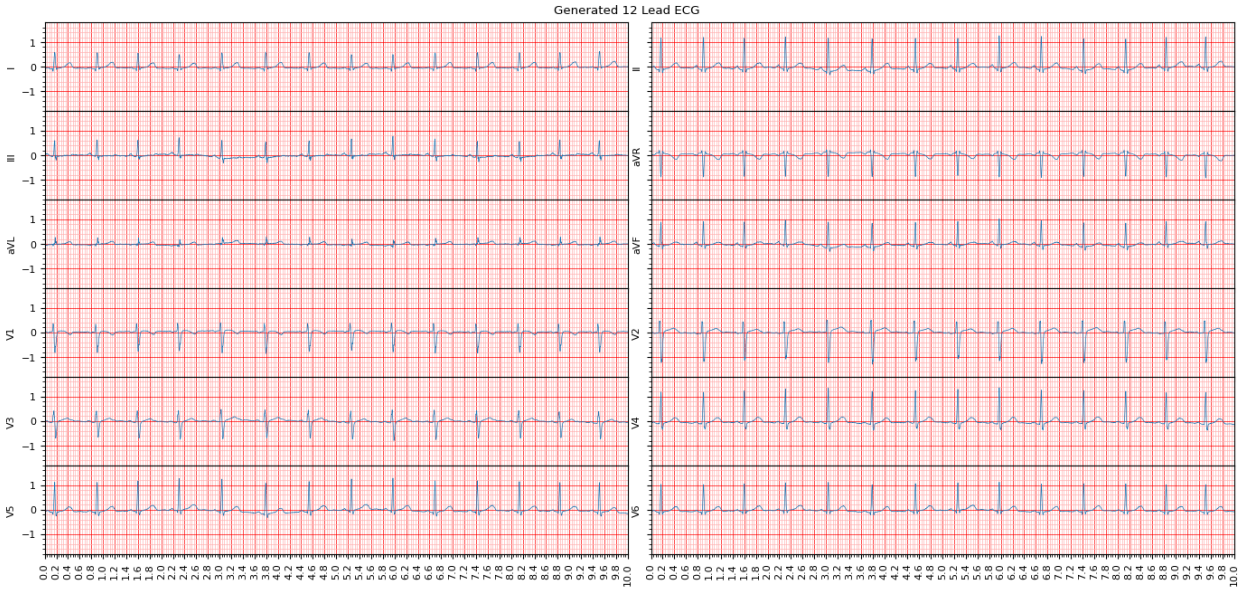


Figure 4: Generated 12-lead ECG example #1 by PhysDiff-ECG.

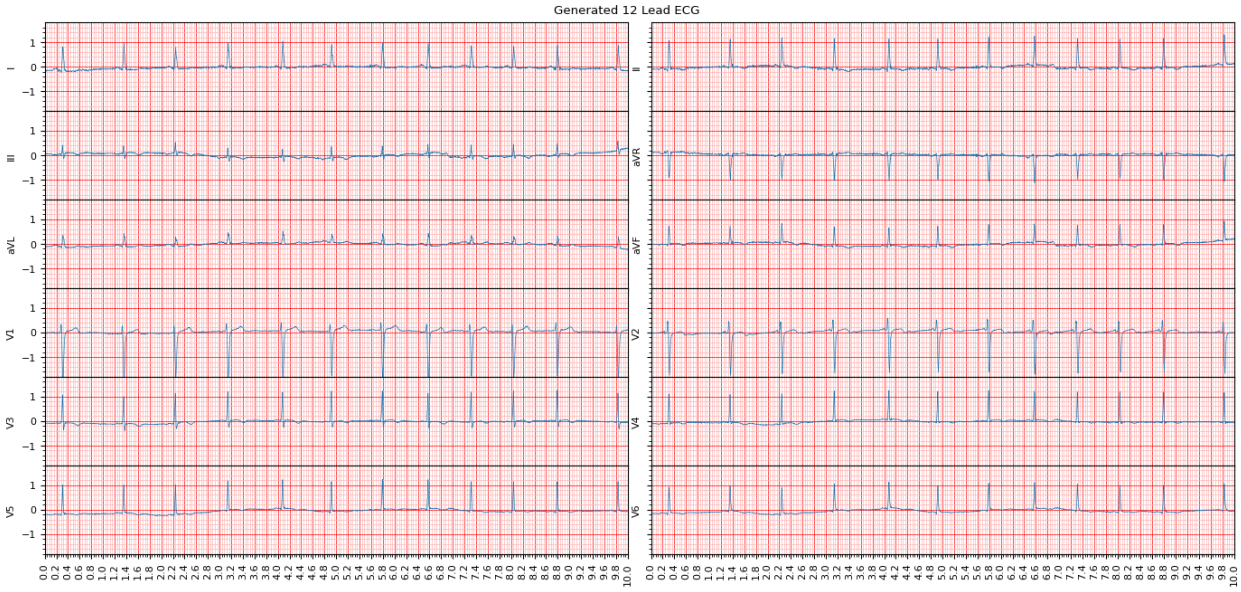


Figure 5: Generated 12-lead ECG example #2 by PhysDiff-ECG .

B Additional Implementation Details

Generative model architecture. The denoiser $\epsilon_\theta(\mathbf{z}_t, t, y)$ is a 1D U-Net with seven resolution levels and kernel size 7, organized as a symmetric downsampling/upsampling hierarchy with skip connections. Self-attention blocks are inserted at selected resolutions to capture long-range temporal dependencies. Class conditioning is implemented through learned label embeddings injected into the denoiser by cross-attention at each resolution level. Unlike text-conditioned diffusion architectures, the conditioning signal here consists of discrete diagnostic labels.

Data splits and cross-validation. For each dataset, 20% of patients are held out as an independent test set and used exclusively for final evaluation (Xu & Goodacre, 2018). The remaining 80% form the development set, within which we perform 5-fold cross-validation for model selection and variability estimation. All splits are constructed at the patient level, so that recordings from the same individual never appear in both development and test sets.

Downstream classifier architectures.

Residual Network (ResNet). Our first downstream classifier is a ResNet-based architecture adapted to ECG signals (Ribeiro et al., 2020). It consists of an initial convolutional layer followed by five residual blocks with batch normalization, ReLU activations, dropout, and skip connections. Temporal downsampling is performed via strided convolutions, and a global average pooling layer feeds a final classification head.

Attention-Enhanced ResNet. Our second downstream classifier augments the ResNet backbone with multi-head attention (Nejedly et al., 2021), enabling stronger modeling of long-range temporal dependencies and inter-lead interactions. This architecture has been shown to perform strongly on large-scale ECG classification benchmarks, making it a useful complementary evaluation backbone.

C Ablation Study

We conduct ablations to isolate the contribution of the physiology-informed regularization in PhysDiff-ECG and its constituent components. Unless otherwise stated, all ablations use the same training data, diffusion backbone, and class conditioning, and are evaluated using the same signal-/feature-level metrics (Sec. 5.2) and downstream classification protocol (Sec. 5.3).

C.1 Ablation on Physiology-Informed Regularization

We first ablate the role of the physiology-informed regularizer during training, and then analyze the contribution of its two components: ODE consistency and lead-space consistency.

With vs. without physiology-informed regularization. We compare the full PhysDiff-ECG against a variant trained without the physiological regularizer, i.e., with $\lambda_{\text{phys}} = 0$ in Eq. equation 18. Both variants use the same pretrained VAE, diffusion backbone, training data, and class conditioning; the only difference is whether the denoiser is trained with the additional physiological regularization term.

As shown in Table 3, removing the physiological regularizer degrades representation-space fidelity (FID), heart-rate preservation (MAE_{HR}), and class-conditional diagnostic accuracy, indicating that the physiological prior improves both realism and label consistency.

Regularizer components. Our inner physiology objective combines latent ODE consistency $\mathcal{L}_{\text{ODE}}$ and lead-space reconstruction $\mathcal{L}_{\text{lead}}$. To isolate their roles, we train variants in which each term is removed by setting its weight to zero. Specifically, we compare the full model against variants without $\mathcal{L}_{\text{ODE}}$ and without $\mathcal{L}_{\text{lead}}$.

Table 3 shows that both terms contribute. Removing $\mathcal{L}_{\text{ODE}}$ primarily harms heart-rate preservation, consistent with weaker enforcement of latent physiological dynamics, whereas removing $\mathcal{L}_{\text{lead}}$ degrades FID and class-

conditional diagnostic accuracy, consistent with weaker alignment to the observed ECG space. The full objective yields the best overall trade-off across metrics.

Table 3: Ablation of physiology-informed regularization and its components. We compare the full model against variants trained without the full physiological regularizer, without $\mathcal{L}_{\text{lead}}$, and without $\mathcal{L}_{\text{ODE}}$. Lower is better for FID and MAE_{HR}; higher is better for Cls-Acc.

Variant	**MAE_{HR} (↓)**	**FID (↓)**	**Cls-Acc (↑)**
Full PhysDiff-ECG	**5.14**	**22.4**	**0.86**
w/o $\mathcal{L}_{\text{phys}}$ (λ_{phys}=0)	7.96	27.1	0.84
w/o $\mathcal{L}_{\text{lead}}$ (λ_{lead}=0)	6.88	26.2	0.83
w/o $\mathcal{L}_{\text{ODE}}$ (λ_{ODE}=0)	9.10	28.9	0.81

These ablations also address prior dependence: removing either the dynamical residual or the observation-model consistency degrades performance, indicating that the gains are not due only to additional loss weighting but to the structured physiological signal.

C.2 Sensitivity to BVAM Parameters

To evaluate robustness to the simplified physiological prior, we perturb each BVAM parameter H, C, and β independently while keeping the diffusion backbone, data split, sampling procedure, and evaluation protocol fixed. We use the same metrics as in the main experiments: heart-rate error, representation-space FID, and classifier accuracy.

Table 4 shows that the perturbations affect the expected aspects of the generated ECGs. Perturbing H causes the largest degradation in heart-rate preservation, consistent with its role in controlling the oscillatory regime. Perturbing C mainly degrades distributional realism, reflecting changes in waveform morphology through the nonlinear restitution terms. Perturbing β reduces class-conditional accuracy, consistent with its effect on latent coupling and cross-lead consistency.

Table 4: Sensitivity to BVAM parameters. Each parameter is perturbed independently while the diffusion backbone and evaluation protocol are fixed.

Setting	MAE_{HR} ↓	FID ↓	Cls-Acc ↑
Default BVAM prior	5.14	22.4	0.86
H perturbed	7.42	24.8	0.85
C perturbed	6.18	25.9	0.85
β perturbed	6.56	26.1	0.84
w/o L_{phys}	7.96	27.1	0.84

C.3 External Augmentation on G12EC

We study whether external data improves downstream classification on G12EC, and whether synthetic PTB-XL samples can substitute for adding real PTB-XL recordings. All classifiers are trained under identical protocols and evaluated on the same held-out *real* G12EC test set.

We compare three training settings:

1. **G12EC (Real Only):** train using only G12EC recordings.
2. **G12EC + PTB-XL (Real External):** augment G12EC with real PTB-XL recordings after mapping diagnoses to a shared label space.
3. **G12EC + Synthetic PTB-XL:** augment G12EC with class-conditional synthetic samples generated by PhysDiff-ECG on PTB-XL, using the same shared label space.

For each class y, we add the same number of external samples in the real-external and synthetic-external settings:

$$N_y^{\text{PTB-XL,added}} = N_y^{\text{syn,added}} = N_y^{\text{added}}.$$

Here, N_y^{added} is a fixed per-class augmentation count shared across settings.

Table 5 reports sensitivity, specificity, and AUC on the held-out real G12EC test set. Adding *real* PTB-XL improves performance across all three metrics, indicating that external data provides useful signal despite dataset shift. Replacing real external recordings with synthetic PTB-XL generated by PhysDiff-ECG yields comparable gains, suggesting that the proposed generator can transfer label-specific diversity to the target dataset without requiring additional real external recordings.

Table 5: External augmentation ablation on G12EC, evaluated on the held-out *real* G12EC test set.

Training Data	**Sens.** ↑	**Spec.** ↑	**AUC** ↑
G12EC (Real Only)	0.85	0.81	0.83
G12EC + PTB-XL	0.89	0.84	0.87
G12EC + Synthetic (PhysDiff-ECG)	0.88	0.84	0.86

C.4 Robustness Across Downstream Classifiers

Because downstream utility is assessed through classifier performance, it is important to verify that the observed gains are not specific to a single evaluation backbone. We therefore repeat the augmentation analysis using two widely adopted 12-lead ECG classifiers: (i) a standard ResNet-based classifier (Ribeiro et al., 2020), and (ii) an attention-augmented ResNet (Nejedly et al., 2021). All training hyperparameters and the test protocol are kept fixed; only the classifier architecture is changed.

Table 6 shows that augmenting with PhysDiff-ECG improves sensitivity, specificity, and AUC for both classifiers, indicating that the benefits of synthetic augmentation are not tied to a particular downstream architecture.

Table 6: Sensitivity analysis across downstream classifier architectures. We compare ResNet (Ribeiro et al., 2020) and attention-augmented ResNet (Nejedly et al., 2021) under the same augmentation protocol. Augmenting with PhysDiff-ECG improves sensitivity, specificity, and AUC for both classifiers.

	ResNet			**Attn-ResNet**		
Training Data	**Sens.**↑	**Spec.**↑	**AUC**↑	**Sens.**↑	**Spec.**↑	**AUC**↑
Real Only	0.86	0.82	0.85	0.84	0.81	0.83
Real + PhysDiff-ECG	**0.89**	**0.84**	**0.88**	**0.88**	**0.84**	**0.87**

D Broader Impact and Responsible Use

Synthetic 12-lead ECGs can support methodological research and data augmentation when real data are limited by privacy, annotation cost, or class imbalance. However, they should be used only as auxiliary training data, not as substitutes for real clinical data or as evidence for patient-level decisions. Generated ECGs may contain subtle artifacts, non-physiological morphology, or inter-lead inconsistencies that are not captured by current automated metrics; clinical use would therefore require clinician review and external validation on independent real-world cohorts. Since PhysDiff-ECG is trained on PTB-XL and G12EC, synthetic samples may inherit demographic, institutional, or diagnostic biases from these public benchmarks, making subgroup evaluation and continued collection of diverse real ECG data essential. Although our experiments do not require access to private clinical records, generated ECGs may still reflect sensitive population-level attributes in the source data or be misused if released without safeguards. Responsible release should therefore include provenance tracking, privacy auditing when appropriate, and clear restrictions against clinical or fraudulent use.