P2Es: Physiology-Conditioned Diffusion for 12-Lead ECG Generation from PPG

Hui Ji[a], Wei Gao[b], Pengfei Zhou[a]

[a]*Department of Informatics and Networked Systems, University of Pittsburgh, Pittsburgh, PA, USA*
[b]*Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA, USA*

Abstract

The 12-lead electrocardiogram (ECG) is the gold standard for cardiovascular monitoring, offering superior diagnostic granularity and specificity compared to photoplethysmography (PPG). However, existing 12-lead ECG systems rely on cumbersome multi-electrode setups, limiting sustained monitoring in ambulatory settings, while current PPG-based methods fail to reconstruct multi-lead ECG due to the lack of explicit inter-lead constraints and insufficient modeling of spatial-temporal dependencies across leads. To bridge this gap, we introduce P2Es, a physiology-conditioned diffusion framework that integrates several synergistic innovations. First, to resolve the physiological information gap, we propose GroupFinder, a contrastive alignment module that retrieves optimal affinity matrices by aligning the PPG latent space with ECG-derived clusters. Furthermore, to mimic real-world distortions, the forward process applies frequency blurring and temporal noise, while the reverse process employs a hierarchical multi-scale generation module with frequency

Email addresses: huj16@pitt.edu (Hui Ji), weigao@pitt.edu (Wei Gao), pengfeizhou@pitt.edu (Pengfei Zhou)

deblurring to recover fine-grained diagnostic details. Finally, we implement a resource-efficient edge architecture for rapid on-device inference. Extensive experiments demonstrate that P2Es significantly outperforms state-of-the-art baselines. This provides a scalable solution for longitudinal cardiac screening and pre-clinical risk stratification.
Keywords: Conditional Diffusion, 12-lead ECG, PPG

1. Introduction

In current clinical practices, the 12-lead electrocardiogram (ECG) is the gold standard for diagnosing cardiac abnormalities, providing comprehensive insights into cardiac electrical activity across the six limb leads (I, II, III, aVR, aVL, aVF) and six chest leads (V1–V6) (Tash and Al-Bawardy, 2023), as shown in Fig. 1(a). It enables the detection of conditions such as myocardial ischemia, arrhythmia subtypes, and conduction disorders which are often undetectable using other single-lead alternatives.

However, widespread 12-lead ECG use, especially outside clinical settings, faces barriers as shown in Fig. 1(c): Holter monitors are bulky, wired 12-lead devices restrict mobility, and hospital-grade setups are cost-prohibitive in low-resource settings (Lamichhane and Neupane, 2022). Although recent advances in wearable ECG monitors (e.g., Zio patch) partially address portability, they still require users to manually place electrodes, limiting continuous out-of-clinic monitoring. While wearable devices equipped with photoplethysmography (PPG) sensors offer ubiquitous, long-term surveillance, they merely reflect peripheral hemodynamics and lack the multi-dimensional spatial granularity required for clinical-grade diagnostics. Consequently, a

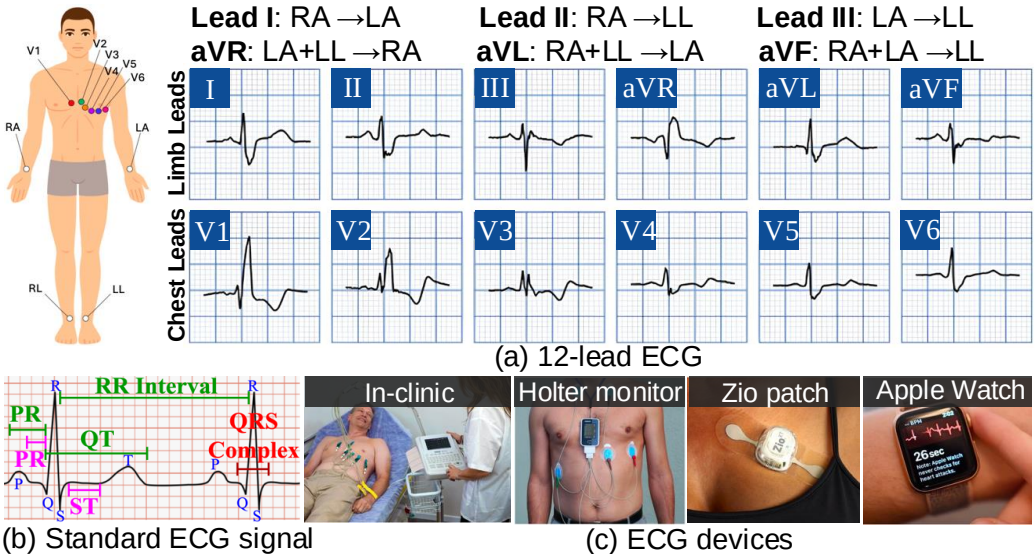


Figure 1: (a) Standard 12-lead ECG electrode placement and waveforms. (b) Components in the standard ECG signal. (c) Different ECG devices.

critical question arises: *Can diagnostically meaningful 12-lead ECGs be directly generated from wearable PPG signals?*

While existing literature has advanced specific sub-tasks—such as translating PPG to single-lead ECG (Ji and Zhou, 2024) or interpolating limited leads to full 12-lead formats (Chehaibi et al., 2024)—establishing a direct, high-fidelity mapping from PPG to 12-lead ECG remains an open challenge. Simply cascading existing models fails to capture the intricate physiological dependency between peripheral blood volume and the heart's global electrical axis. This cross-modality synthesis is fundamentally constrained by three obstacles: ***1)*** Absence of Inter-Lead Constraints: Generative models typically operate in global latent spaces, overlooking the physiology-dependent spatial correlations across chest leads and the electrical axis constraints governing limb leads. ***2)*** Insufficient Modeling of Spatial-Temporal Dependencies:

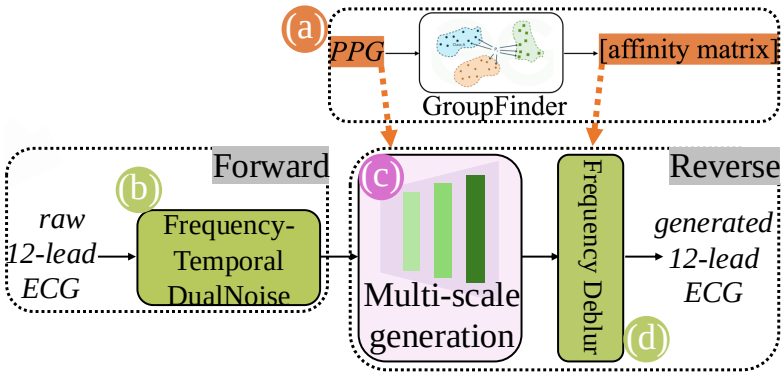


Figure 2: Schematic overview of the proposed P2Es framework.

Clinical ECG demands the joint preservation of time-domain features (e.g., ST elevation) and high-frequency spectral components (e.g., QRS sharpness). *3)* Multi-Scale Morphological Variations: The 12-lead ECG exhibits extreme dynamic range disparities. High-amplitude macro-structures (e.g., QRS complexes) tend to dominate the optimization landscape, often overshadowing clinically critical microvolt-level details (e.g., P-waves and ST-segments) during reconstruction.

To address these challenges, we present *P2Es*, a physiology-conditioned diffusion framework capable of generating full 12-lead ECGs from single-channel PPG. Rather than relying on a naive data-driven mapping, P2Es explicitly bridges the modality gap through three core design insights, as illustrated in Fig. 2.

❶ **Physiology-Aware Spatial Conditioning.** Wearable PPG primarily reflects peripheral hemodynamics and inherently lacks the spatial resolution required to localize cardiac electrophysiology. To bridge this gap, P2Es introduces the **GroupFinder** module to explicitly model the structural con-

straints of the heart. By querying shared population priors, **GroupFinder** dynamically constructs an Affinity Matrix for the chest leads, serving as an explicit spatial condition during the reverse diffusion process. Simultaneously, an Electrical Axis Loss is designed to constrain the geometric projection of limb leads. Together, these constraints ensure that the synthesized multi-lead signals are grounded in authentic cardiac physiology rather than mathematical hallucinations.

❷ **Frequency-Temporal Diffusion for High-Fidelity Morphology.** ECG signals contain both complex temporal trends and critical localized structural features (e.g., QRS notches). To prevent these subtle microvolt-level features from being overshadowed, P2Es incorporates a Frequency-Temporal DualNoise mechanism. By cascading a hierarchical multi-scale temporal generator with a frequency deblurring module, the model progressively and accurately recovers both the global rhythms and the clinically salient morphological details.

❸ **Lightweight Execution for Edge Inference.** To translate this high-fidelity generation into practical utility on mobile platforms, P2Es features a resource-efficient architecture optimized for ubiquitous monitoring. By integrating accelerated sampling (e.g., DDIM (Song et al., 2020)) and structural parameter sharing, we significantly curb model size and computational latency without compromising reconstruction quality.

We extensively evaluate P2Es primarily on the MIMIC-IV dataset, supported by MIMIC-II/III for auxiliary validation. Benchmarked against representative methods, P2Es achieves state-of-the-art performance, reducing reconstruction errors by **16.3%** in MSE and **26.4%** in DTW compared

to current PPG-to-single-lead baselines. More importantly, the synthesized 12-lead ECGs demonstrate strong downstream diagnostic utility, achieving 87.5% specificity for myocardial infarction detection and 83.5% precision for arrhythmia classification.

Our main contributions are summarized as follows:

- **The First PPG to 12-lead ECG Framework.** P2Es is the first end-to-end sensing system to directly synthesize 12-lead ECG from PPG, bridging the modality gap through diffusion-based cross-domain alignment.
- **Three Unique Designs.** We introduce three innovations to realize P2Es. First, a GroupFinder module based on KNN and contrastive learning is designed to retrieve optimal affinity matrices by aligning the PPG with shared anatomical priors, guiding the reverse process. Second, the dual-domain frequency–temporal diffusion process preserves spectral coherence while resisting temporal distortions. Third, the hierarchical temporal multi-scale reverse path uses dilated convolutions and expanding receptive fields to recover global rhythm, local morphology, and high-frequency detail.
- **Edge Deployment and Real-World Validation.** We implement a resource-efficient edge architecture that enables rapid on-device inference. Our extensive evaluation on MIMIC-series datasets demonstrates that P2Es consistently achieves superior reconstruction fidelity compared to state-of-the-art baselines. Furthermore, the practical utility of the proposed framework is validated through a real-world multi-modal case study, confirming its robustness in ambulatory settings.

2. Background

2.1. Multi-Lead ECG and Wearable PPG

The 12-lead ECG captures cardiac electrical activity across multiple anatomical planes and remains the clinical gold standard for cardiovascular diagnosis. It comprises six limb leads (I, II, III, aVR, aVL, aVF) and six chest leads (V1–V6). Limb leads derive from three electrodes (RA, LA, LL), forming Einthoven's triangle (Fig. 3 (a)), with Lead I (RA $\rightarrow$ LA), Lead II (RA$\rightarrow$LL), and Lead III (LA$\rightarrow$LL). Augmented leads (aVR, aVL, aVF) enhance sensitivity to specific axes via Wilson's central terminal. Chest leads (V1–V6) are placed sequentially across the chest, with spatial correlations quantified by an affinity matrix (Fig. 3 (b)). This standardized system enables precise diagnosis: limb leads assess frontal plane electrical axis (θ) for axis deviation, while chest leads localize transverse plane pathologies. As a wearable alternative, PPG measures peripheral blood-volume changes using optical sensors, offering a low-cost, continuous monitoring (Ji and Zhou, 2024). However, since PPG reflects hemodynamic responses (e.g., pulse transit time) rather than direct electrophysiology, its diagnostic role is limited to rhythm screening and heart-rate estimation.

2.2. Physiological Pathway: PPG to 12-Lead ECG

Wearable PPG captures peripheral blood-volume fluctuations that arise as a hemodynamic consequence of cardiac electrical activationVentricular depolarization initiates myocardial contraction, producing a pressure pulse that travels through the arterial tree and manifests as the systolic upstroke in the PPG after a characteristic electromechanical delay and pulse transit time

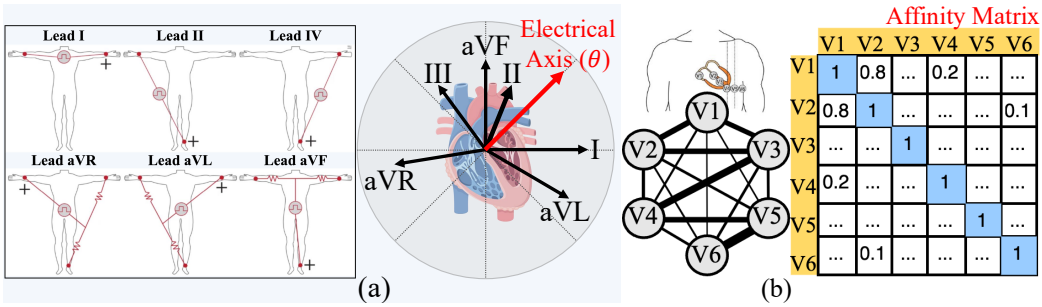


Figure 3: (a) Electrical axis in limb leads. (b) Principle of affinity matrix in chest leads.

(PTT) (Ji and Zhou, 2024). This causal linkage allows PPG to preserve rhythm, beat timing, and contraction-related morphological cues originating from the ECG. Recent studies (Nambu et al., 2025; Lin et al., 2024) demonstrate that these downstream hemodynamic patterns can reconstruct electrically consistent cardiac waveforms from PPG, including ECG-like representations.

However, as a single-site optical measurement, PPG lacks the spatial information required to map cardiac electrophysiology. Therefore, translating PPG into a 12-lead ECG is an ill-posed inverse problem that demands explicit anatomical constraints. To bridge this spatial gap, we introduce two biophysical priors: an electrical axis constraint to govern the frontal-plane geometric projections of the limb leads, and an affinity matrix to model the transverse-plane spatial continuity among the chest leads (V1–V6). By strictly bounding the generative solution space within this physiologically coherent multi-lead manifold, the cross-modal reconstruction becomes mathematically well-regularized and clinically viable.

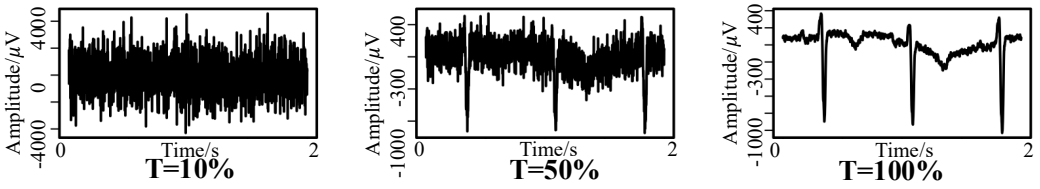


Figure 4: Progressive ECG generation during the reverse process.

2.3. How Diffusion Reduces Uncertainty

Diffusion model (Song et al., 2020) excels at cross-modal time-series generation: their forward process gradually injects noise, driving structured ECG signals toward high-entropy representations, while the reverse process incrementally restores physiological structure (Sivarajkumar et al., 2025). Using preliminary results from the MIMIC-II dataset, Fig. 4 visualizes entropy reduction during the DDPM (Ho et al., 2020) reverse diffusion process at representative steps: 10%, 50%, and 100%. At T=10%, the output remains noise-like. By T=50%, cardiac harmonics emerge, and by T=100%, the model generates a physiologically plausible waveform. This shows that diffusion model does not hallucinate but systematically reduces uncertainty via conditional PPG guidance and learned physiological priors. Compared to GANs (Durgadevi et al., 2021) and VAEs (Bond-Taylor et al., 2021), which suffer from mode collapse or blurry outputs, diffusion models preserve temporal structure and exhibit strong noise robustness—critical for real-world PPG signals. Yet, prior work is limited to unimodal or single-lead ECG generation (Tang et al., 2023; Ji and Zhou, 2024), leaving cross-modal, multi-lead synthesis underexplored.

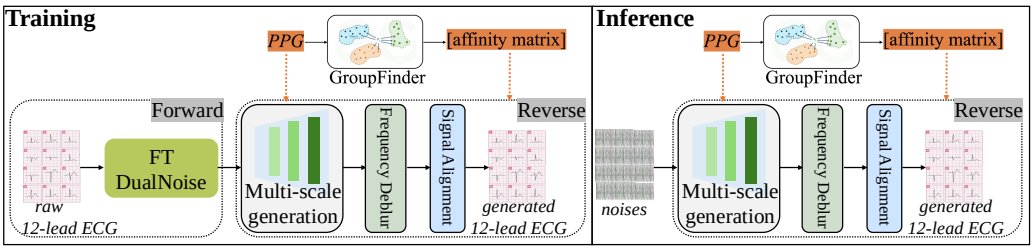


Figure 5: Overview of the P2Es framework. (Left) Training: The GroupFinder module extracts an affinity matrix from the input PPG. A forward process injects FT DualNoise, while the reverse process reconstructs the 12-lead ECG via Multi-scale generation, Frequency Deblur, and Signal Alignment, jointly conditioned on both the raw PPG and the affinity matrix. (Right) Inference: The target 12-lead ECG is synthesized from pure noise, guided solely by the PPG and its derived affinity matrix.

3. P2Es Framework

3.1. Overview

We propose P2Es, a physiology-conditioned diffusion framework illustrated in Fig. 5. Prior to the diffusion process, our framework employs a GroupFinder module to establish explicit anatomical constraints by dynamically retrieving an optimal affinity matrix based on the input PPG. Guided by these spatial constraints, the diffusion model operates in two phases. During the training phase, the forward process injects frequency-time noise into ground-truth 12-lead ECG through the Frequency-Temporal (FT) DualNoise module. The reverse process then reconstructs ECG by progressively performing multi-scale temporal denoising, frequency deblurring, and signal alignment, all conditioned on PPG and the affinity matrix produced by the GroupFinder module. During the inference phase, the model executes exclusively the reverse process, where the PPG signal and its associated affinity

matrix serve as the sole conditional inputs to guide the generation of the final 12-lead ECG.

3.2. GroupFinder

Recent studies have shown that inter-lead correlation structures in multi-lead ECGs are strongly modulated by demographic factors such as age, sex, and cardiac hypertrophy (Tash and Al-Bawardy, 2023; Macfarlane et al., 2010). Motivated by this observation, P2Es groups subjects into multiple clusters so that each cluster shares similar physiological-dependent electrical patterns, enabling physiologically consistent ECG reconstruction (Fig. 6).

3.2.1. Affinity Matrix Construction

The affinity matrix encodes the medically established spatial correlation structure of the six precordial leads (V1-V6)–the key determinant for maintaining physiological plausibility in generated multi-lead ECGs (Macfarlane et al., 2010). For each subject i, we compute the Pearson correlation coefficient $A_i(m, n)$:

$$\mathbf{A}_i(m,n) = \frac{\mathbb{E}[(V_m^{(i)} - \mu_{V_m})(V_n^{(i)} - \mu_{V_n})]}{\sigma_{V_m}\sigma_{V_n}},\ 1 \le m, n \le 6. \tag{1}$$

The normalized signal of lead V_m for the ith subject is represented as $V_m^{(i)}$. Additionally, μ_{V_m} and σ_{V_m} denote the mean and standard deviation of lead V_m, respectively. This formulation extends Kornreich's anatomical correlation analysis to dynamic physiological characteristics. When $m = n$, the numerator is the variance of V_m, and the denominator is $\sigma_{V_m}^2$, thus $A_i(m, m) = 1$ (self-correlation). This is expected because signals from the same lead are perfectly correlated.

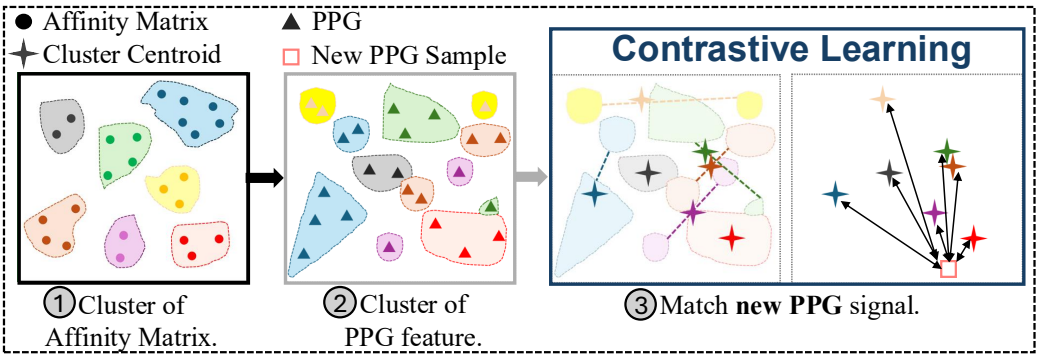


Figure 6: Illustration of the GroupFinder Module.

3.2.2. KNN-Based Clustering

To model physiological similarity and PPG-ECG affinity, we propose a multi-modal KNN clustering process. For each subject, we flatten the upper triangle of $\mathbf{A}_i$ into $\mathbf{a}_i \in \mathbb{R}^{21}$, preserving spatial symmetry:

$$\mathbf{a}_i = [\mathbf{A}_i(1,1), \mathbf{A}_i(1,2), ..., \mathbf{A}_i(5,6), \mathbf{A}_i(6,6)]. \tag{2}$$

Physiological features $\mathbf{d}_i \in \mathbb{R}^p$ (e.g., age, gender, medical flags, and ventricular hypertrophy status) are concatenated to form a joint feature vector:

$$\mathbf{a}_i^{\text{joint}} = [\mathbf{a}_i \parallel \mathbf{d}_i] \in \mathbb{R}^{21+p}, \tag{3}$$

where $\parallel$ denotes concatenation, and p is the dimension of physiological features.

3.2.3. Cluster Optimization

Next, we select the most appropriate number of group through multi-criteria validation. For each sample i, the silhouette score is computed as:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \tag{4}$$

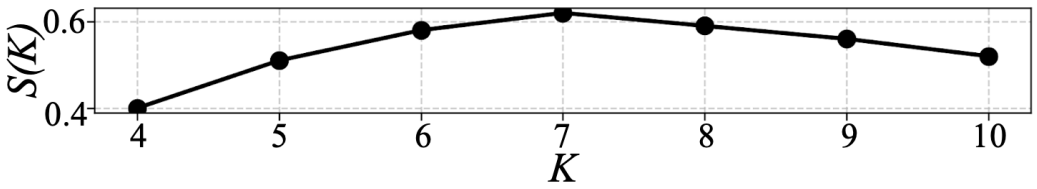


Figure 7: Global score $S(K)$ analysis across varying K.

where $a(i)$ is the average intra-cluster distance, and $b(i)$ is the smallest average inter-cluster distance (separation). The global score $S(K)$ is the mean of all $s(i)$, that is, $S(K) = \frac{1}{N}\sum_{i=1}^{N} s(i)$. In Fig. 7, we observe that $S(7) = 0.62$ (higher than $S(5) = 0.51$ and $S(9) = 0.56$), indicating well-separated clusters with strong intra-group coherence; thus we set $K = 7$.

3.2.4. Centroid Calculation

For each cluster C_k, we compute centroids as follows:

$$\bar{\mathbf{A}}_k = \frac{1}{|\mathcal{C}_k|} \sum_{i \in \mathcal{C}_k} \mathbf{a}_i^{\text{joint}}. \tag{5}$$

3.2.5. Contrastive Assignment for New PPG Inputs

To dynamically map new PPG signals to the predefined ECG clusters, our contrastive assignment framework aligns incoming PPG embeddings with the joint physiological–waveform centroids.

3.2.6. Dual-Branch Encoder Architecture

The PPG encoder $\mathcal{E}_{\text{PPG}}$ maps raw PPG signals x_i and features $f_i = [A_s, T_{PA}, S_d]$ to an embedding space: $z_i^{\text{PPG}} = \mathcal{E}_{\text{PPG}}(x_i) \in \mathbb{R}^{d_e}$. Here, A_s, T_{PA}, S_d represent the peak amplitude of the systolic phase in the PPG signal, the time delay from the R wave to the PPG systolic peak, and the PPG diastolic

Algorithm 1 Real-Time Contrastive Group Assignment

Require: New PPG signal x_{new}, pre-trained encoders $\mathcal{E}_{\text{PPG}}$ and $\mathcal{E}_{\text{A}}$, cluster centroids $\{c_k\}_{k=1}^{7}$.

Ensure: Assigned affinity matrix $\bar{A}_{k^*}$.

1: Extract features from x_{new}: $f_{\text{new}} = [A_s, T_{PA}, S_d]$.

2: Embed features using encoder $\mathcal{E}_{\text{PPG}}$: $z_{\text{new}} = \mathcal{E}_{\text{PPG}}(f_{\text{new}})$.

3: **for** $k = 1$ to 7 **do**

Compute cosine similarity: $s_k = \frac{z_{\text{new}} \cdot c_k}{\|z_{\text{new}}\| \|c_k\|}$

4: **end for**

5: Identify optimal cluster index: $k^* = \arg\max_k s_k$

6: Retrieve affinity matrix $\bar{A}_{k^*}$ for diffusion conditioning.

slope, respectively. The affinity matrix encoder $\mathcal{E}_{\text{A}}$ encodes cluster centroids $\bar{A}_k$ into the same space: $c_k = \mathcal{E}_{\text{A}}(\bar{A}_k) \in \mathbb{R}^{d_e}$.

3.2.7. Contrastive Loss Function

We optimize a contrastive objective that pulls each PPG embedding toward its closest centroid while pushing it away from others:

$$\mathcal{L}_{\text{cont}} = -\log \frac{\exp(z_i^{\text{PPG}} \cdot c_k / \tau)}{\sum_{m=1}^{7} \exp(z_i^{\text{PPG}} \cdot c_m / \tau)}, \tag{6}$$

where τ is a hyperparameter. A lower τ makes the model focus more on the most similar matches, while a higher τ smooths the distribution. For a new PPG signal x_{new}, the cluster assignment proceeds as Algorithm 1.

3.3. Forward Process: Frequency Blurring and Temporal Noise

The forward process models realistic degradations in ECG by gradually increasing uncertainty in both the frequency and temporal domains. This

prepares the diffusion model to reverse these distortions during generation.

Stage 1: Frequency Blurring Module

Given a 12-lead ECG signal $\mathbf{x}_0 \in \mathbb{R}^{N\times 12}$ (N: time steps), we first apply a discrete Fourier transform:

$$\mathbf{X}_0 = \mathcal{F}[\mathbf{x}_0] = U\mathbf{x}_0 \in \mathbb{C}^{N\times 12}, \tag{7}$$

where $U \in \mathbb{C}^{N\times N}$ is the unitary DFT matrix. To simulate spectral distortion, we apply a frequency blur operator $\mathcal{B}$:

$$\mathbf{X}_{\text{blur}} = \mathcal{B}(\mathbf{X}_0) = \mathbf{X}_0 \odot M + \epsilon_f, \tag{8}$$

where $M \in \mathbb{R}^{N\times 12}$ means Gaussian low-pass mask with bandwidth $\omega_c = \frac{\pi t}{T} \cdot f_{\max}$, where $f_{\max}$ is the maximum frequency of the signal, t is the diffusion timestep and T the total steps. $\epsilon_f \sim \mathcal{CN}(0, \sigma_f^2 I)$ is complex Gaussian noise added to high-frequency components. The blurred signal is then mapped back to the time domain:

$$\mathbf{x}_{\text{blur}} = \mathcal{F}^{-1}[\mathbf{X}_{\text{blur}}] = U^*\mathbf{X}_{\text{blur}}. \tag{9}$$

Stage 2: QRS-Adaptive Temporal Noise

In order to further capture the most important QRS area in the ECG signal, the blurred signal $\mathbf{x}_{\text{blur}}$ then undergoes temporal noise injection, modulated by QRS complexes detected via an adaptive search. This QRS-aware strategy (Ji and Zhou, 2024) is preferred over random masking because QRS regions contain the highest diagnostic information and most beat-to-beat variability. Targeted perturbation introduces realistic timing uncertainty without distorting non-QRS morphology.

3.3.1. QRS Mask Generation.

Using a modified Pan-Tompkins algorithm (Pan and Tompkins, 1985), we generate a binary mask $\mathbf{m} \in \{0,1\}^{N\times 12}$:

$$m_{\tau,j} = \begin{cases} 1, & \text{if } \tau \in \text{QRS region of lead } j; \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

3.3.2. Temporal Noise Injection.

At diffusion timestep t, noise is selectively added to non-QRS regions:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_{\text{blur}} + \sqrt{1-\bar{\alpha}_t}(1-\mathbf{m}) \odot \epsilon_t, \tag{11}$$

where $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$. Here $\epsilon_t \sim \mathcal{N}(0, I)$ denotes standard Gaussian noise and $\odot$ is Hadamard product.

3.4. Reverse Process: Multi-Scale Temporal Recovery and Frequency Deblurring

As illustrated in Fig. 5, the reverse process reconstructs 12-lead ECG from noisy latent states using PPG signals and the affinity matrix as conditioning. It proceeds through three stages that progressively refine temporal structure, restore spectral fidelity, and align clinical morphology.

3.4.1. Multi-Scale Temporal Recovery

Letting T be the total number of steps in the diffusion model, and given the noisy signal $\mathbf{x}_T$ from the forward process, the multi-scale temporal reverse iteratively refines the signal across three stages:

Early-Stage Coarse Denoising ($t \in [0, T/3]$). Dilated convolutions with kernel sizes (9×1) are used to capture long-range dependencies.

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_0(\mathbf{x}_t, t, \mathbf{c})\right) + \sigma_t \mathbf{z}, \tag{12}$$

where $\mathbf{c} = \{\text{PPG}, \bar{\mathbf{A}}_k\}$ concatenates PPG features and the assigned affinity matrix, and $\mathbf{z} \sim \mathcal{N}(0, I)$.

Mid-Stage Structural Refinement ($t \in [T/3, 2T/3]$). Hybrid layers combine convolutions (7×1) with transformer attention:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}, \tag{13}$$

where $\mathbf{Q} = \mathbf{x}_t W_Q$, $\mathbf{K} = \mathbf{x}_t W_K$, and $\mathbf{V} = \mathbf{x}_t W_V$ are learnable projections into different feature spaces.

Late-Stage Detail Recovery ($t \in [2T/3, T]$.) Stacked convolutions (5×1) and residual connections reconstruct QRS, P-waves, and T-waves with sub-millisecond precision.

3.4.2. Frequency-Domain Deblurring

To restore spectral sharpness degraded in the forward process, a frequency-domain operator performs controlled enhancement. Here, assuming that the multi-scale restored signal is $\mathbf{x}^{\text{pred}}$:

$$\mathbf{X}^{\text{pred}} = \mathcal{F}[x^{\text{pred}}] \in \mathbb{C}^{N \times 12}, \tag{14}$$

where N is the signal length, and each column corresponds to a lead's spectral representation. To amplify diagnostically critical high-frequency components (> 15 Hz), a frequency-selective mask M_{high} is applied:

$$\mathbf{X}^{\text{deblur}} = \mathbf{X}^{\text{pred}} \odot (1 + \gamma M_{\text{high}}), \tag{15}$$

where $M_{\text{high}} \in \mathbb{R}^{N \times 12}$ is a bandpass mask for high-frequency components. γ is a learnable parameter that controls the strength of the high-frequency

enhancement. Next is the inverse discrete Fourier transform to recover the final signal:

$$\mathbf{x}^{\text{final}} = \mathcal{F}^{-1}[\mathbf{X}^{\text{deblur}}]. \tag{16}$$

3.4.3. Signal Alignment

We further refine the reconstruction using an amplitude–frequency alignment module (Ji and Zhou, 2024). The loss function is calculated as:

$$\mathcal{L}_a = (\frac{1}{N_{\min}} \sum_{i=1}^{N_{\min}} |p_{g_i} - p_{t_i}|) + |E_g - E_t| + |N_g - N_t|\,, \tag{17}$$

where $N_{\min}$ is the number of shared R-peaks between generated and reference ECGs. The R-peck positions p_{g_i} and p_{t_i} denote the i-th R-peak in the generated and ground-truth signals. The amplitude ranges E_g and E_t, calculated as $\max(x) - \min(x)$, quantify peak-to-peak voltage differences. Heart rate consistency is measured by N_g and N_t, derived from R-peak counts per unit time for the synthesized and reference signals.

3.4.4. Unified Loss Function

In our proposed P2Es, the overall training objective is summarized as follows:

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{L}} + \lambda_2 \mathcal{L}_{\text{F}} + \lambda_3 \mathcal{L}_{\mathbf{S}_\theta} + \lambda_4 \mathcal{L}_a. \tag{18}$$

Among this,

$$\mathcal{L}_L = \sum_{i \neq j} \left\| \Delta ECG_{i,j}^{Pred} - \Delta ECG_{i,j}^{GT} \right\|_2, \tag{19}$$

$$\mathcal{L}_F = \left\| \mathcal{F}(ECG^{GT}) - \mathcal{F}(ECG^{Pred}) \right\|_2, \tag{20}$$

$$\mathcal{L}_{\mathbf{S}_\theta} = \left\| \mathbf{S}_\theta^{Pred} - \mathbf{S}_\theta^{GT} \right\|_2, \tag{21}$$

where $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, $\lambda_3 = 0.2$, and $\lambda_4 = 0.3$. Here, the electrical axis reflects the geometric projection of the cardiac dipole onto the limb leads and is an intrinsic ECG property, not an external input. It must therefore be enforced through loss functions that maintain physiologically coherent axis relationships across generated leads.

3.5. Lightweight Model Design

Deploying PPG-to-12-lead ECG generation on mobile devices requires strict control of latency, memory, and energy. Traditional compression methods (e.g., pruning, distillation) improve efficiency but often degrade clinically critical ECG details (Wang et al., 2026), such as sharp QRS transitions and inter-lead spatial relationships. To preserve these diagnostic features, we prioritize architectural and algorithmic optimizations that reduce computation while maintaining full representational fidelity.

- **Efficient Sampling.** We adopt DDIM-based (Song et al., 2020) sampling to reduce the number of reverse steps. Instead of a fixed-length schedule, we employ a step-size controller that allocates more iterations to high-uncertainty phases (early denoising) and fewer iterations once morphological structures stabilize, reducing inference latency by up to 60% while preserving reconstruction quality.

- **Parameter-sharing.** To limit model size, P2Es shares parameters across diffusion steps and within the 12-lead generation network. Only the input splicing index for each lead is uniquely embedded. This design enforces inter-lead consistency and reduces memory consumption without reducing output resolution.

- **Input Length Optimization.** We evaluate multiple PPG input lengths and examine performance–latency trade-offs for mobile deployment. A 2-second input is selected as the optimal balance, enabling real-time inference on consumer-grade devices.

Ultimately, the number of parameters is reduced by 28%, and the inference speed is increased by 1.6 times.

4. Experiments

4.1. Datasets and Experiment Details

Datasets. We use the MIMIC series (II–IV) datasets (Mehrgardt et al., 2022; Johnson et al., 2016, 2023), which include over 65,000 patients and millions of physiological recordings from Beth Israel Deaconess Medical Center (2001–2019). MIMIC II includes signals such as non-invasive blood pressure, PPG, and up to 12-lead ECG. MIMIC III extends this with arterial blood pressure (ABP) and respiratory data. MIMIC IV further incorporates clinical records and continues to provide comprehensive ECG, PPG, ABP, and respiratory signals. PPG is recorded via bedside pulse oximeter probes (mainly Philips IntelliVue) typically attached to the fingertip, similar to sensors now embedded in consumer wearables such as smart rings and watches (Oura Health Ltd., 2024; Ultrahuman Healthcare Pvt. Ltd., 2024). Together, the MIMIC datasets offer the largest open-access, clinically annotated 12-lead ECG repository with paired PPG and diagnostic waveforms, serving as a benchmark for AI-driven cardiac research.

Data pre-processing. All datasets were de-identified using consistent subject IDs across cohorts. Following MIMIC guidelines, subjects appear-

ing in earlier datasets were removed from later ones to ensure independence. Records were retained only if 12-lead ECG and PPG were both available. Preprocessing included: ❶ **Resampling**: ECG and PPG signals were standardized to 125 Hz via cubic interpolation. ❷ **Filtering**: A 4th-order Butterworth highpass filter (cutoff: 0.5 Hz) was applied to ECG signals, and a bandpass filter (0.5–8 Hz) was applied to PPG signals. ❸ **Normalization**: Subject-wise z-score normalization was performed to reduce physiological variability. ❹ **Segmentation**: Signals were divided into 2-second non-overlapping windows (250 samples at 125 Hz).

The final dataset was split by patient (80% training / 20% testing), ensuring no subject overlap. This pipeline ensures *(i)* cross-modal signal consistency, *(ii)* prevention of data leakage, and *(iii)* reproducibility for cross-dataset validation.

Train setup. P2Es is trained on two NVIDIA A100 GPUs with a batch size of 512. The training process employs the AdamW optimizer and runs for 400 epochs. A linear variance scheduler is applied consistently, with the hyperparameter values gradually increasing from 0.0001 to 0.2.

Baselines. Given the absence of established models for 12-lead ECG generation from PPG, we construct three evaluation categories to rigorously benchmark our approach.

• **Category I**: We extend existing PPG-to-single-lead ECG models—PPGG (Ji and Zhou, 2024), RDDM (Shome et al., 2024), and TSDFD (Crabbé et al., 2024)—by training 12 independent instances (one per lead). PPGG leverages non-Markovian noise scheduling, TSDFD integrates spectral-temporal decomposition, and RDDM uses residual-driven diffusion to model ECG dy-

namics.

• **Category II**: We implement a clinically inspired two-stage pipeline: Lead I is first synthesized from PPG via PPGG, then used to generate the remaining 11 leads using four models—CycleGAN (Mehrabadi et al., 2022) (cyclic consistency), TSGAN (Banerjee et al., 2023) (conditional U-Net GAN), DiffWave (Kong et al., 2020) (non-autoregressive diffusion), and TimeGrad (Rasul et al., 2021) (autoregressive temporal gradient diffusion). This reflects practical cascaded setups where synthesized Lead I serves as an intermediate representation.

• **Category III**: To establish performance upper bounds, we generate 11-lead ECG directly from original Lead I signals, eliminating PPG-to-ECG errors.

All models share identical preprocessing and patient-stratified splits (80%/20%) for fair comparison. This tripartite evaluation systematically assesses (i) PPG-to-ECG translation fidelity, (ii) lead interdependence modeling from Lead I, and (iii) cumulative errors in multi-stage workflows, enabling fine-grained analysis of signal reconstruction and clinical waveform preservation across methods.

4.2. Evaluation Metrics

To evaluate the quality of generated 12-lead ECG, we compare synthesized outputs (x_{gen}) using six metrics as formulated in Table 1. Mean Squared Error (MSE) measures overall amplitude fidelity across leads and time. Dynamic Time Warping (DTW) computes the minimal temporal alignment path between generated and reference waveforms. Variance (Var) assesses robustness to PPG-driven variability. KL Divergence (KL) compares distributions

Table 1: Formulations of evaluation metrics.

Metric	**Mathematical Formulation**
MSE	$\frac{1}{N \cdot L} \sum_{l=1}^{L} \sum_{\tau=1}^{N} \left(x_{\text{gen}}^{l}(\tau) - x_{\text{GT}}^{l}(\tau) \right)^2$
DTW	$\min_{\pi} \sum_{(i,j)\in\pi} \|x_{\text{gen}}(i) - x_{\text{GT}}(j)\|_2$
Var	$\frac{1}{L} \sum_{l=1}^{L} \text{Var}(x_{\text{gen}}^{l} - x_{\text{GT}}^{l})$
KL	$\sum_{k} P_{\text{GT}}(k) \log \frac{P_{\text{GT}}(k)}{P_{\text{gen}}(k)}$
MW	$\int_0^1 \left(F_{X_{\text{gen}}}^{-1}(u) - F_{X_{\text{GT}}}^{-1}(u) \right) du$
S_θ	$\cos(\theta_{\text{gen}} - \theta_{\text{GT}}), \theta = \arctan\left(\frac{\sum x_{\text{II}}}{\sum x_{\text{I}}} \right)$

Notations: $L = 12$ (ECG leads), N: time steps, π: optimal warping path, k: clinical features (ST-segment, QRS amplitude), θ: mean electrical axis.

of ECG features (e.g., QRS amplitudes). The Marginal Wasserstein (MW) distance quantifies differences in marginal distributions. Electrical Axis Similarity (S_θ) evaluates the alignment of the mean electrical axis θ, critical for detecting cardiac axis deviations.

5. Results

5.1. Visualization

Figure 8 compares the generated 12-lead ECG from P2Es against original ECG, covering both limb leads (upper row) and chest leads (lower row) to assess transformation fidelity.

The upper row displays the six limb leads, which provide a global view of

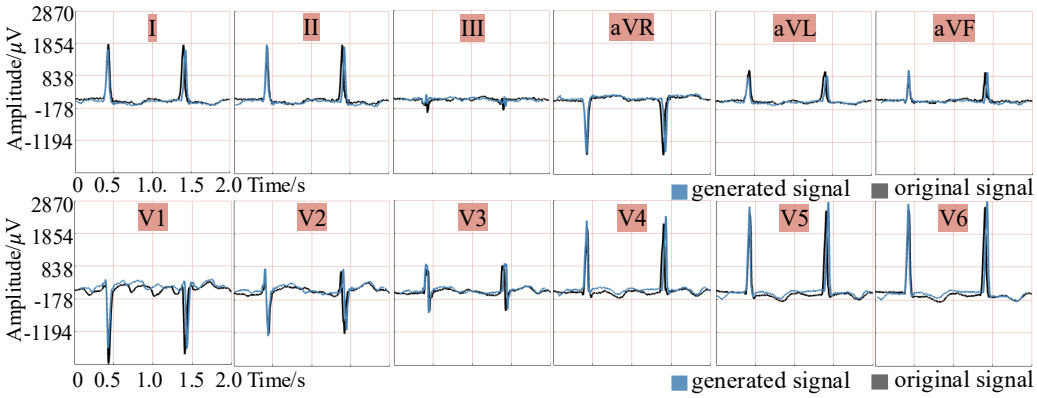


Figure 8: High-fidelity 12-lead ECG reconstruction. Visual comparison of original (black line) and generated (blue line) 12-lead ECG signals. The upper and lower rows are limb and chest leads, respectively. *Note*: 12-lead ECG signals have the same time axis (0-2.0 s).

cardiac electrical activity and are essential for assessing overall rhythm and inter-lead voltage relationships. Specifically, Leads I and II show strong signal consistency, accurately reproducing most of the QRS complex and P wave structure in the original ECG, confirming that the model captures rhythm information and inter-lead correlation. Furthermore, Leads aVR, aVL, and aVF perform well in voltage polarity alignment, exhibiting the expected inverse voltage relationship relative to Lead II. This suggests that our explicitly formulated Electrical Axis Loss effectively preserves the geometric projection and anatomical plausibility of the frontal plane.

The lower row showcases the six chest leads, which provide critical insights into ventricular depolarization and repolarization. These leads are particularly important for detecting ischemic changes and conduction abnormalities. The ST-segment dynamics in V1–V3 are well preserved, including the characteristic ST-slope, demonstrating the effectiveness of the

frequency–temporal diffusion in suppressing high-frequency artifacts while retaining low-frequency diagnostic markers. In V4–V6, the R-wave morphology is accurately reconstructed, demonstrating that our explicitly derived affinity matrix effectively preserves inter-lead spatial correlations. Together, these results confirm that the multi-scale reverse process recovers both low-amplitude features and high-energy QRS components. A slight QRS-onset shift is observed in V2, which we plan to further mitigate by refining diffusion-step scheduling to better capture rapid transitions in chest-lead signals.

5.2. Comparison with Category I

Our proposed method achieves SOTA performance in PPG-to-12-lead ECG translation, outperforming all baseline approaches of category I. As reported in Table 2, P2Es achieves the lowest MSE scores (overall: **0.0902**, chest: **0.0984**, limb: **0.0820**), surpassing the best baseline by **16.3%** and **12.0%** for overall and limb leads, respectively. The smallest DTW score (**0.0583**) further confirms its superior temporal waveform alignment, crucial for arrhythmia detection. Additionally, our model demonstrates enhanced stability, with a variance of **0.4647**, which is **51.5%** lower than the nearest baseline, indicating robustness against PPG variability. The lowest KL divergence values (chest: **0.5113**, limb: **0.4727**) confirm that the generated multi-lead ECGs closely match the original distribution, reducing the risk of pathological misrepresentation. Furthermore, our method improves clinical utility and morphological precision, reducing MW error by **12.4%** (chest) and **8.8%** (limb) compared to the second-best baseline, with the highest similarity score (**0.9542** for limb leads) highlighting its ability to retain diagnostically critical features such as ST-segment morphology details.

Table 2: Performance comparison of our model and category I models on the MIMIC IV dataset. The best models have been shown in bold. Where θ only exists in each set of limb leads.

				PPG-ECG								
	Ours			PPGG			RDDM			TSDFD		
	overall	**chest**	**limb**	**overall**	**chest**	**limb**	**overall**	**chest**	**limb**	**overall**	**chest**	**limb**
MSE (↓)	**0.0902**	**0.0984**	**0.0820**	0.1078	0.1224	0.0932	0.2502	0.2644	0.2360	0.2978	0.3093	0.2863
DTW (↓)	**0.0583**	**0.0615**	**0.0551**	0.0792	0.0794	0.0790	0.1782	0.1802	0.1762	0.2247	0.2269	0.2225
Var (↓)	**0.4647**	**0.4709**	**0.4585**	0.9583	0.9601	0.9565	1.4516	1.4575	1.4457	1.2411	1.2553	1.2269
KL (↓)	**0.4920**	**0.5113**	**0.4727**	0.7664	0.7833	0.7495	2.0173	2.1044	1.9302	1.3265	1.3962	1.2568
MW (↓)	**8.0902**	**9.1417**	**7.0387**	9.2401	10.7548	7.7254	27.3327	28.1034	26.5620	18.9352	22.8534	15.0170
$\mathbf{S}_\theta(\uparrow)$	-	-	**0.9542**	-	-	0.8718	-	-	0.7439	-	-	0.8206

Although chest leads exhibit slightly higher errors (MSE: **0.0984**) due to larger amplitude variations, our KNN-based affinity conditioning effectively mitigates this gap. Additionally, limb leads achieve near-perfect S_θ (**0.9542**), demonstrating the success of our frequency-temporal denoising strategy in preserving subtle ECG features. Overall, our framework sets a new benchmark in PPG to 12-lead ECG translation by integrating demographic-aware clustering, frequency-temporal diffusion process, and multi-scale recovery, addressing key challenges in cross-modal physiological signal synthesis.

5.3. Comparison with Categories II & III

As demonstrated in Table 3, our model achieves superior performance compared to conventional Two-step approaches in category II, with particular advantages in physiological consistency. The cascaded PPGG+DiffWave pipeline (MSE: 0.1492, DTW: 0.0719) underperforms our unified model by 41.7% in waveform fidelity (MW metric) and 12.8% in temporal alignment (DTW), revealing the error accumulation inherent in sequential generation

stages. While DiffWave achieves marginally better MSE (0.0814 vs. 0.0902) when directly generating from original Lead I ECG signals (Table 4), our method maintains competitive performance across all metrics despite requiring only PPG inputs - a clinically critical advantage given PPG's ubiquity in wearable devices.

Table 3: Performance comparison with two-step approaches of category II (PPGG combined with other ECG to 12-lead ECG models). The best is in bold.

		Two-step: PPGG + the following			
	Ours	**CycleGAN**	**TSGAN**	**DiffWave**	**TimeGrad**
MSE(↓)	**0.0902**	0.1540	0.1709	0.1492	0.1639
DTW(↓)	**0.0583**	0.0732	0.0801	0.0719	0.0798
Var(↓)	**0.4647**	0.5917	0.6466	0.5823	0.6013
KL(↓)	**0.4920**	0.5463	0.6035	0.5203	0.5924
MW(↓)	**8.0902**	14.0017	15.4268	14.0749	15.2390
$\mathbf{S}_\theta(\uparrow)$	**0.9542**	0.8200	0.7994	0.8656	0.8175

Table 4: Performance comparison of our model with category III approaches. The best and second-best models are shown in bold and underlined, respectively.

	ECG (Lead I) - 12-lead ECG				
	Ours	**CycleGAN**	**TSGAN**	**DiffWave**	**TimeGrad**
MSE(↓)	<u>0.0902</u>	0.1074	0.1253	**0.0814**	0.0927
DTW(↓)	0.0583	0.0611	0.0792	**0.0525**	<u>0.0564</u>
Var(↓)	<u>0.4647</u>	0.5148	0.5368	**0.4291**	0.4802
KL(↓)	<u>0.4920</u>	0.5393	0.5820	**0.4405**	0.5188
MW(↓)	**8.0902**	12.8394	13.2871	<u>10.2839</u>	11.9368
$\mathbf{S}_\theta(\uparrow)$	**0.9542**	0.9012	0.8937	<u>0.9365</u>	0.9144

Three key differentiators emerge from the analysis: First, our model demonstrates enhanced electrical axis preservation with 18.9% lower variance (0.4647 vs. 0.5823-0.6466 in Table 3) compared to Two-step approaches of category II, demonstrating the anatomically consistent R-wave progression in V1-V6 leads. Second, the S_θ similarity metric (0.9542) surpasses all baseline methods of category II and III by $\geq$4.8%, indicating better preservation of diagnostic waveform features. Third, our KL divergence (0.4920) remains 9.4-18.5% lower than cascaded approaches, confirming superior distributional alignment with clinical ECG patterns.

Though the Lead I-to-11-lead benchmarks establish an upper performance bound reference (best MSE: 0.0814), our PPG-driven approach achieves 89.3% of this bound reference while eliminating the need for initial ECG acquisition–a fundamental paradigm shift in continuous cardiac monitoring.

Table 5: Multi-lead ECG characteristics for different cardiovascular diseases (MIMIC IV dataset). (MI: myocardial infarction. ARR: arrhythmias.)

	Infarct site	ICD-10 code	Primary Leads	Typical ECG features	Abnormal/Normal	Total
MI	Precordial	121.0	V1-V4	ST elevation, Q waves, T wave inversion	15/15	30
	Inferior	121.1	II, III, aVF	ST elevation, Q waves	10/10	20
	Lateral	121.21	I, aVL, V5-V6	ST elevation, Q wave, R wave amplitude	4/4	8
	Posterior	121.22	V1-V2	V1-V2 T wave inversion, ST depression	5/5	10
ARR	SVT	147.1	V1-V6	Narrow QRS complex (< 120 ms)	10/10	20
	VT	147.2	V1-V6	Wide QRS complex (> 120 ms)	10/10	20

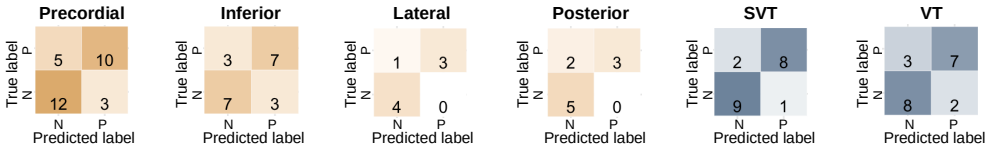


Figure 9: Confusion matrices for classification of cardiovascular diseases using the generated multi-lead ECG. Yellow and blue matrices represent the subdivisions of Myocardial Infarction (MI) and Arrhythmia (ARR), respectively.

5.4. Cardiovascular Disease Diagnosis

To assess downstream utility, we perform an initial clinical validation on the MIMIC-IV dataset using ICD-10-coded myocardial infarction (MI) and arrhythmia (ARR) cases. Despite the limited sample size for specific subtypes, this evaluation effectively demonstrates the model's fundamental capability to synthesize clinically precise 12-lead ECG patterns. The experiment settings for the two cases are detailed in Table 5, and results are reported in Fig. 9. Here we use sensitivity, specificity, accuracy, and precision to measure the performance. Specifically, sensitivity $= \frac{TP}{TP+FN}$, specificity $= \frac{TN}{TN+FP}$, accuracy $= \frac{TP+TN}{TP+TN+FP+FN}$, precision $= \frac{TP}{TP+FP}$. True Positive (TP) refers to cases where the model correctly predicts a positive outcome. True Negative (TN) represents that the model correctly classifies a negative case. False

Positive (FP) occurs when the model incorrectly predicts a negative case as positive. False Negative (FN) happens when the model incorrectly classifies a positive case as negative. For MI subtypes, the overall specificity reached 87.5%, and the lateral and posterior cases reached 100%, proving that model can correctly identify normal ECG and avoid classifying normal variations as diseases. That is, the model detects actual ECG wave amplitude changes, especially V1-V2 T wave inversion, ST elevation, avoiding false positives due to slight normal variations. On the other hand, the model achieves 68% sensitivity in distinguishing pathological features, and the sensitivity of the inferior and lateral cases reached 70% and 75% respectively, which means that the model can correctly identify most of the diseased individuals.

In arrhythmias, the accuracy of SVT and VT cases are 85% and 75% respectively, while the precision is 89% and 78% respectively. High accuracy means that the model correctly distinguishes normal QRS complexes from narrow (<120 ms) or wide (>120 ms) QRS morphology, thus reducing the overall misclassification rate. It reflects the overall reliability of the model in distinguishing normal and abnormal ECG signals under different conditions. Precision is key for correctly distinguishing between the narrow QRS complexes of SVT and the wide QRS complexes of VT. A high-precision model ensures that only actual cases of VT are flagged, preventing false diagnoses based on minor QRS widening. High precision improves clinical trust in the model's decisions, ensuring throughly meaningful and reducing false alerts.

12-lead ECG can comprehensively evaluate cardiac structural and functional abnormalities through multi-angle electrical signal acquisition, and is particularly irreplaceable in locating lesions and identifying complex arrhyth-

mias. Single-lead ECG (such as wearable devices) is suitable for screening heart rate or simple arrhythmias, but it cannot replace the accuracy of 12-lead ECG in clinical diagnosis.

5.5. Ablation Study

To evaluate the contributions of each component in the P2Es framework, we conduct comprehensive ablation studies on the MIMIC IV dataset by progressively removing key modules: the Affinity Matrix (AM) for cross-lead correlation modeling, Frequency Blur and deblur (FB) for spectral regularization, and Multi-scale (MS) reverse for hierarchical feature learning. As shown in Fig. 10, the complete model consistently achieves optimal performance across all evaluation metrics (MSE: 0.0902, DTW: 0.0583), demonstrating clear synergistic effects among architectural components.

Removing the AM module notably degrades temporal alignment by 4.5% (DTW: 0.0609 vs. 0.0583) and increases morphological distortion (MW: 8.5122 vs. 8.0902), underscoring its critical role in preserving inter-lead electrophysiological relationships and ensuring realistic signal morphology. The absence of FB causes severe variance inflation (0.7315 vs. 0.4647), indicating its effectiveness in suppressing high-frequency artifacts while maintaining essential diagnostic waveform features–evidenced by the 6.3% drop in S_θ score (0.9428 vs. 0.9542). Disabling MS generation leads to the most pronounced performance deterioration, with KL divergence increasing by 18.7% (0.5842 vs. 0.4920), confirming the necessity of multi-resolution processing for capturing both global rhythm patterns and local depolarization details.

Notably, the AM-FB-MS triad exhibits significant non-linear complementarity: while removing any single module increases MSE by 1.4–7.4%, exclud-

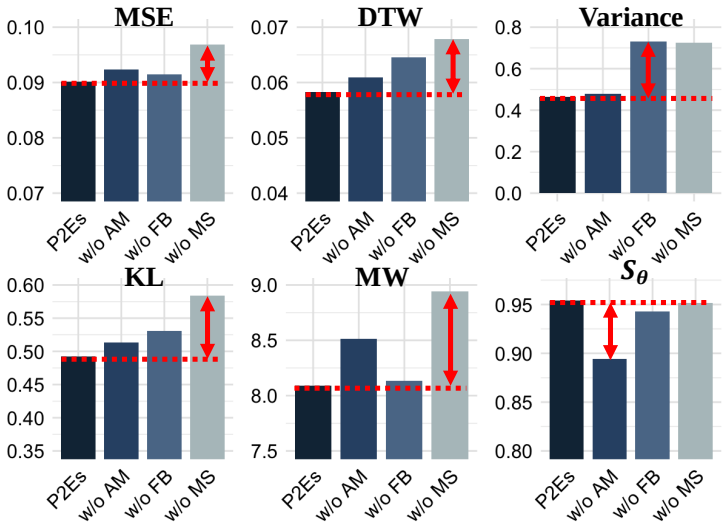


Figure 10: Ablation study for different metrics on the MIMIC IV dataset. (AM: Affinity Matrix; FB: Frequency Blur and Deblur; MS: Multi-scale reverse).

ing any two components degrades MSE by over 12%, highlighting their interdependence. This confirms our architectural design's efficacy in addressing three core challenges.

5.6. Cross-dataset Validation

Cross-dataset validation assesses the generalizability and effectiveness of models across distinct data sources. This experiment aims to confirm that the model does not become overly specialized to characteristics unique to the training dataset. To evaluate this, models were separately trained on the MIMIC III and MIMIC IV datasets, with their performance subsequently tested on the other dataset. Specifically, Fig. 11(a) illustrates the performance of models trained on the MIMIC IV dataset when evaluated against MIMIC III, showing that the overall MSE was 0.0921 for subjects from MIMIC III (purple boxes) and 0.0902 for subjects from MIMIC IV (yellow

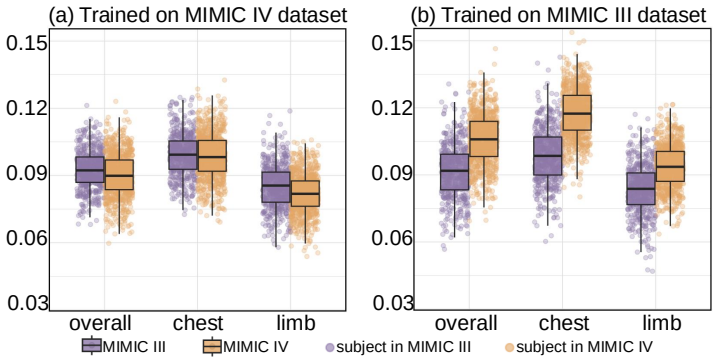


Figure 11: Cross-dataset MSE performance comparison between models trained on MIMIC III and MIMIC IV datasets.

boxes). This indicates a slightly better performance on the training dataset (MIMIC IV), but comparable effectiveness across datasets, suggesting limited overfitting.

Conversely, Fig. 11(b) presents results for the model trained on the MIMIC III dataset, where the overall MSE was 0.0914 for subjects from MIMIC III and increased significantly to 0.1056 for subjects from MIMIC IV. This notable rise in MSE suggests reduced generalizability of the model trained on MIMIC III, highlighting potential dataset-specific features or limitations that impact performance on MIMIC IV. A likely explanation for this phenomenon is that the MIMIC IV dataset includes data not only from ICU and emergency departments but also hospital-wide foundational clinical data, whereas the MIMIC III dataset is limited to ICU and emergency department data. This broader scope of clinical information in MIMIC IV might introduce variability and complexity not captured during training on the narrower MIMIC III dataset, thus impacting generalization.

5.7. Real-world Case Study

To evaluate the practicality and robustness of P2Es in real-world settings, we conduct a case study involving the simultaneous acquisition of multiple physiological signals. This study was approved by our Institutional Review Board (IRB), and all participants are provided informed consent. As illustrated in Fig. 12, our sensing setup includes a fingertip PPG sensor (Max30101) (Zhou et al., 2023) consistent with the finger-clip acquisition used in the MIMIC datasets, a SpO_2 monitor (Nonin Onyx Vantage 9590), a clinical-grade Omron cuff-based blood-pressure device, and a standard 12-lead ECG acquisition system (PCECG-500). All sensors are time-synchronized while the participant remains in a resting supine posture to minimize motion artifacts and ensure high-fidelity multimodal recordings.

In this experiment, the wearable PPG waveform is used as the sole input to the generative model. The concurrently collected 12-lead ECG, blood-pressure readings, and SpO_2 measurements serve as the ground-truth reference. We evaluate the reconstructed ECG quality along two complementary dimensions: **Spatiotemporal coherence**, using the metrics summarized in Table 1; **Downstream diagnostic utility**, quantified through clinically actionable tasks, including cuff-free blood-pressure estimation and SpO_2 prediction.

Fig. 13 summarizes the performance of three representative models (P2Es, PPGG, and RDDM) across two male and two female participants, with all values normalized ($\hat{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min} + \epsilon}$ with $\epsilon = 10^{-8}$) for comparison. Despite using only fingertip PPG as input, P2Es reconstructs 12-lead ECG waveforms that closely match the clinical ground truth and substantially outperform

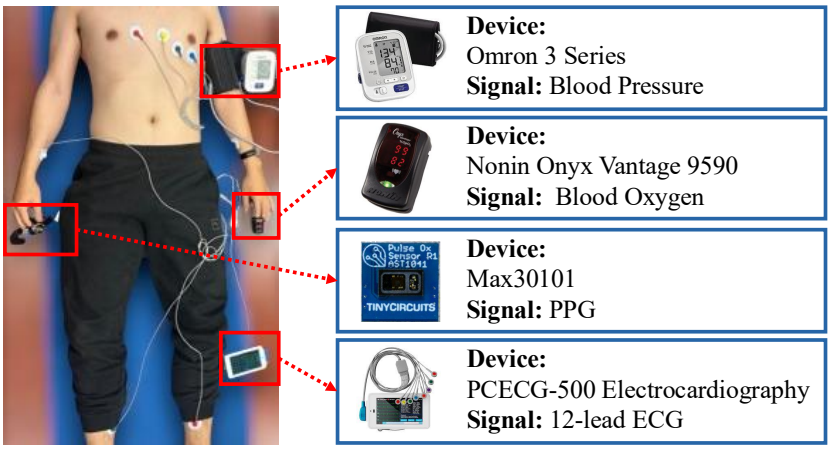


Figure 12: Experimental setup for the real-world case study.

the two comparison methods. Furthermore, we evaluate downstream blood-pressure and SpO_2 estimation using the input PPG and generated ECG. The resulting MAE errors are 2.91 and 1.34, respectively, meeting the AAMI clinically endorsed accuracy thresholds. Collectively, these results demonstrate that P2Es enables continuous, wearable, and lead-free cardiac monitoring, making it feasible to obtain diagnostic-grade 12-lead ECG representations from consumer-grade PPG sensors in everyday environments.

5.8. Model Memory and Latency Overhead

The P2Es model achieves a precision-optimized design for edge deployment, featuring a memory footprint of 3.2GB during training with 57M+ parameters. Evaluations on heterogeneous platforms demonstrate deployment efficiency: On a MacBook Pro 2023 (Apple M2 Pro chip, 16GB unified memory) with GPU acceleration, it achieves 620ms (±45ms) average latency per inference over 100 runs. Mobile benchmarking reveals that on iPhone 15 Pro Max (A17 Bionic, 3nm process, 45 TOPS neural engine), Core

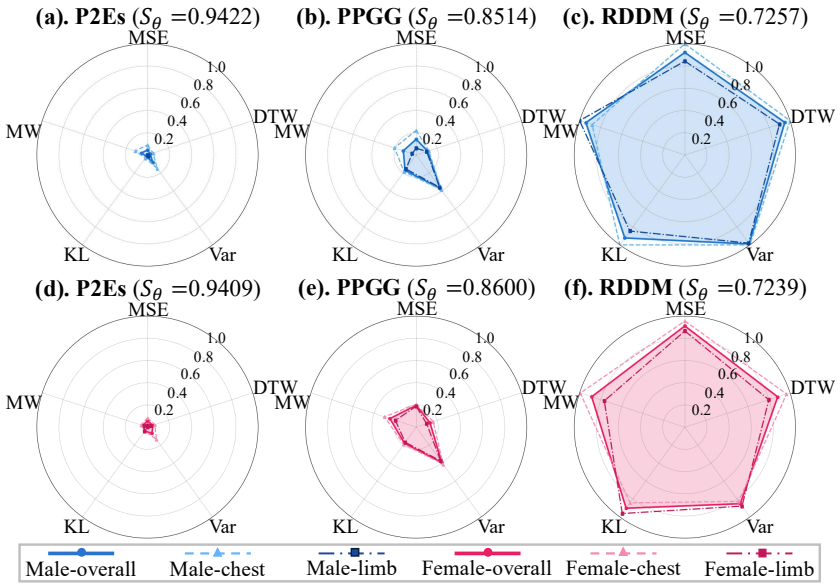


Figure 13: Radar-chart comparison of reconstruction performance across models and participants. Smaller area denotes better performance.

ML-accelerated inference completes 12-lead ECG generation (2s input) in 4.46s (±0.8s) while maintaining sub-900MB peak memory through memory pooling-enabling sustained background operation. This latency is clinically practical: rather than continuous synthesis, generating intermittent background ECG snapshots satisfies ambulatory screening protocols while saving battery.

6. Discussion

6.1. Physiological Interpretability and Mechanism

The clinical utility of P2Es hinges on its ability to recover diagnostic features (e.g., QRS morphology, ST-segment deviations) from a single-point PPG. This feasibility is rooted in the inherent electro-mechanical coupling of

the cardiovascular system: cardiac electrical depolarization drives mechanical ventricular contraction, which propagates as the pulse wave detected by PPG. Unlike "black-box" approaches, P2Es explicitly models this inverse relationship through physiological constraints. The GroupFinder module does not merely cluster data; it acts as a proxy for the cardiac axis, establishing the necessary spatial boundary conditions for the chest leads. Similarly, the Electrical Axis Loss mathematically enforces the geometric projection consistency of the limb leads. By synergistically combining these constraints with the frequency-temporal diffusion, P2Es effectively reverses the physiological degradation, recovering fine-grained electrical details that are latently encoded in the hemodynamic waveform.

6.2. Diagnostic Reliability and Boundaries

Our evaluation demonstrates robust performance across diverse pathologies, yet it is crucial to delineate the diagnostic boundaries of PPG-based synthesis. 1) Arrhythmias (Timing Anomalies): Disorders like SVT and VT primarily manifest as rhythm disturbances. Since electrical rhythm directly dictates mechanical pulse timing, P2Es achieves high accuracy (85% for SVT) by effectively capturing these temporal patterns. 2) Ischemic Events (Morphological Anomalies): Myocardial Infarction (MI) alters the repolarization trajectory (e.g., ST-elevation). Although challenging, P2Es successfully recovers these features (87.5% specificity) because severe ischemia often compromises local ventricular wall motion, subtly influencing the PPG morphology. 3) Limitations: Admittedly, the mapping is not bijective. Certain atypical pathologies (e.g., localized conduction blocks) or minor electrical abnormalities may manifest negligibly in peripheral hemodynamics. In such

"silent" cases, the lack of prominent PPG manifestations inevitably limits reconstruction fidelity. To address this, future iterations will integrate uncertainty estimation to actively flag low-confidence generations where the physiological signal-to-noise ratio is insufficient.

6.3. Future Directions

While currently validated on resting fingertip PPG, future work will systematically evaluate cross-location generalization (e.g., wrist-worn or ear-worn devices) and robustness against motion artifacts. To democratize access, we are optimizing the architecture for low-resource edge inference, aiming to reduce latency for real-time screening. We also plan to explore personalized calibration, utilizing sparse longitudinal data to further enhance diagnostic precision in ambulatory settings.

7. Conclusion

This paper introduces P2Es, an end-to-end framework for synthesizing 12-lead ECGs directly from wearable PPG. To resolve the ill-posed nature of this cross-modality mapping, we propose a physiology-conditioned approach: utilizing GroupFinder to retrieve affinity matrices and an electrical axis penalty to enforce physiology coherence. Furthermore, our frequency-temporal diffusion mechanism ensures signal fidelity by recovering fine-grained morphological details. Validated by comprehensive benchmarks, our resource-efficient edge architecture enables rapid on-device inference, significantly outperforming state-of-the-art baselines and paving the way for ubiquitous, lead-free cardiovascular screening.

CRediT authorship contribution statement

Hui Ji: Conceptualization, Investigation, Methodology, Software, Data curation, Writing - original draft. **Wei Gao**: Funding acquisition, Supervision, Formal analysis, Writing - review & editing. **Pengfei Zhou**: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The MIMIC series (II-IV) dataset used in this study is publicly available. The real-world PPG data will be made available on request.

Acknowledgements

This work was supported by the National Institutes of Health (NIH) under grant number R01HL184168. The authors would like to thank the reviewers for their comments and feedback. Special thanks go to Prof. Gowda and his team for providing the PPG devices and supporting the real-world data collection infrastructure. Furthermore, we would like to express our sincere gratitude to all the volunteers who participated in the study and contributed to the real-world data collection.

References

Banerjee, R., Mazumder, O., Mukherjee, A., Sinhahajari, S., Sinha, A., 2023. Reconstruction of body surface potential from 12-lead ecg: A conditional gan based approach, in: 2023 31st European Signal Processing Conference (EUSIPCO), IEEE. pp. 1180–1184.

Bond-Taylor, S., Leach, A., Long, Y., Willcocks, C.G., 2021. Deep generative modelling: A comparative review of vaes, gans, normalizing flows, energy-based and autoregressive models. IEEE transactions on pattern analysis and machine intelligence 44, 7327–7347.

Chehaibi, O., Majdoub, N., Ladhari, T., Saïd, S.H., 2024. Ecg monitor design: From one to three-advancing ecg monitoring with ai-generated lead synthesis (gans), in: 2024 IEEE International Conference on Artificial Intelligence & Green Energy (ICAIGE), IEEE. pp. 1–6.

Crabbé, J., Huynh, N., Stanczuk, J., Van Der Schaar, M., 2024. Time series diffusion in the frequency domain. arXiv preprint arXiv:2402.05933 .

Durgadevi, M., et al., 2021. Generative adversarial network (gan): A general review on different variants of gan and applications, in: 2021 6th International Conference on Communication and Electronics Systems (ICCES), IEEE. pp. 1–8.

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. Advances in neural information processing systems 33, 6840–6851.

Ji, H., Zhou, P., 2024. Advancing ppg-based continuous blood pressure monitoring from a generative perspective, in: Proceedings of the 22nd ACM Conference on Embedded Networked Sensor Systems, pp. 661–674.

Johnson, A.E., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., Pollard, T.J., Hao, S., Moody, B., Gow, B., et al., 2023. Mimic-iv, a freely accessible electronic health record dataset. Scientific data 10, 1.

Johnson, A.E., Pollard, T.J., Shen, L., Lehman, L.w.H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Anthony Celi, L., Mark, R.G., 2016. Mimic-iii, a freely accessible critical care database. Scientific data 3, 1–9.

Kong, Z., Ping, W., Huang, J., Zhao, K., Catanzaro, B., 2020. Diffwave: A versatile diffusion model for audio synthesis. arXiv preprint arXiv:2009.09761 .

Lamichhane, B., Neupane, N., 2022. Improved healthcare access in low-resource regions: A review of technological solutions. arXiv preprint arXiv:2205.10913 .

Lin, Y., Ma, J., Wang, W., Wu, Z., Dong, S., Luo, G., Wang, K., 2024. Biomedically informed ecg synthesis: Customizing cardiac cycle phases with diffusion model, in: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE. pp. 3505–3508.

Macfarlane, P.W., Van Oosterom, A., Pahlm, O., Kligfield, P., Janse, M., Camm, J., 2010. Comprehensive electrocardiology. Springer Science & Business Media.

Mehrabadi, M.A., Aqajari, S.A.H., Zargari, A.H.A., Dutt, N., Rahmani, A.M., 2022. Novel blood pressure waveform reconstruction from photoplethysmography using cycle generative adversarial networks, in: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society, IEEE. pp. 1906–1909.

Mehrgardt, P., Khushi, M., Poon, S., Withana, A., 2022. Pulse transit time ppg dataset. PhysioNet 10, e215–e220.

Nambu, Y., Kohjima, M., Yamamoto, R., 2025. Cardioflow: Learning to generate ecg from ppg with rectified flow, in: ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE. pp. 1–5.

Oura Health Ltd., 2024. Oura Ring Official Website. `https://ouraring.com`.

Pan, J., Tompkins, W.J., 1985. A real-time qrs detection algorithm. IEEE transactions on biomedical engineering , 230–236.

Rasul, K., Seward, C., Schuster, I., Vollgraf, R., 2021. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting, in: International conference on machine learning, PMLR. pp. 8857–8868.

Shome, D., Sarkar, P., Etemad, A., 2024. Region-disentangled diffusion model for high-fidelity ppg-to-ecg translation, in: Proceedings of the AAAI conference on artificial intelligence, pp. 15009–15019.

Sivarajkumar, S., Zhang, H., Ji, Y., Bilalpur, M., Wu, X., Li, C., Kwak, M.G., Visweswaran, S., Wang, Y., 2025. Generative foundation model

for structured and unstructured electronic health records. arXiv preprint arXiv:2508.16054 .

Song, J., Meng, C., Ermon, S., 2020. Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 .

Tang, Q., Chen, Z., Ward, R., Menon, C., Elgendi, M., 2023. Ppg2ecgps: An end-to-end subject-specific deep neural network model for electrocardiogram reconstruction from photoplethysmography signals without pulse arrival time adjustments. Bioengineering 10, 630.

Tash, A.A., Al-Bawardy, R.F., 2023. Cardiovascular disease in saudi arabia: facts and the way forward. Journal of the Saudi Heart Association 35, 148.

Ultrahuman Healthcare Pvt. Ltd., 2024. Ultrahuman Ring: Metabolic and Sleep Tracking. Available at: `https://ultrahuman.com`.

Wang, P., Song, S., Ji, H., Cao, S., Yu, H., Liu, Z., Yang, H., Lin, Y., Chen, B., Bansal, M., et al., 2026. From models to systems: A comprehensive survey of efficient multimodal learning. TechRxiv .

Zhou, H., Lu, T., Liu, Y., Zhang, S., Liu, R., Gowda, M., 2023. One ring to rule them all: An open source smartring platform for finger motion analytics and healthcare applications, in: Proceedings of the 8th ACM/IEEE Conference on Internet of Things Design and Implementation, pp. 27–38.