mmJEPA-ECG: Cross-Posture Robust Contactless Electrocardiogram Monitoring via Millimeter Wave Radar Sensing

Ziyang Liu*, Siyuan He*, Feng Liang, Chang Huang, Shuxin Zhong†, Kaishun Wu†

The Hong Kong University of Science and Technology (Guangzhou)
{zliu785,she209,fliang302,chuang932}@connect.hkust-gz.edu.cn, {shuxinzhong, wuks}@hkust-gz.edu.cn

Abstract

Continuous cardiac monitoring during sleep is vital for detecting silent arrhythmia and other nocturnal cardiac events. While electrocardiogram (ECG) is the clinical gold standard, its reliance on electrodes and physical contact makes it intrusive for daily long-term use. Millimeter-wave (mmWave) radar offers a compelling non-contact alternative by capturing cardiac-induced chest-wall micro-vibrations. Existing radar-to-ECG methods often rely on direct waveform regression, assuming posture-stable mappings that break under natural sleep movements and obscure true cardiac rhythms. Inspired by the modality-invariant perception observed in speech and vision, we introduce `mmJEPA-ECG`, a physiology-guided framework for reconstructing clinical ECGs by anchoring radar sensing to invariant cardiac dynamics. It addresses two fundamental challenges: (i) disentangling robust cardiac representations from posture-induced artifacts, and (ii) generalizing ECG reconstruction across individuals under signal ambiguity. To address these challenges, *Physiology-Oriented Self-Supervised Pretraining* builds on a Joint Embedding Predictive Architecture (JEPA) with domain-informed masking and heart rate consistency to extract posture-robust cardiac embeddings. *Conditional Diffusion-based ECG Reconstruction* then generates personalized ECG waveforms through a hierarchical conditional diffusion process by spectral fidelity and denoising constraints. Extensive experiments on both public and self-collected multi-subject datasets demonstrate that our method outperforms state-of-the-art across waveform and rhythm metrics, halving R-R peak errors even under posture shifts and arrhythmic conditions.

Introduction

Cardiovascular diseases (CVDs) are the leading cause of death worldwide, claiming over 17.9 million lives annually (WHO 2025). Patients with obstructive sleep apnea (OSA) face heightened risk: nearly 46% of sudden cardiac deaths (SCDs) occur silently during sleep (Gami et al. 2005)(Jaspan et al. 2024), when symptoms go unnoticed in the absence of monitoring. These unobserved nocturnal events underscore an urgent need for *continuous, unobtrusive cardiac monitoring during rest*. Electrocardiogram (ECG) is the clinical gold standard for assessing cardiac health (Siontis et al. 2021). Yet conventional tools (e.g., Holter monitors (DiMarco and Philbrick 1990) and implantable recorders (Brignole et al. 2009)) are intrusive and costly for long-term home use (Walsh III, Topol, and Steinhubl 2014)(Mercuri et al. 2019).

This has sparked growing interest in *non-contact ECG monitoring* using ambient Radio Frequency (RF) signals including ultra-wideband (UWB) (Wang et al. 2023), and millimeter-wave (mmWave) (Chen et al. 2022) which capturing thoracic micro-vibrations induced by heartbeats and use deep-learning model to map the relation between heart vibration and electrical activity. They work well under fixed conditions but degrades with posture shifts, particularly in lateral or reclined sleep positions due to reduced signal quality from altered reflections (Zhang et al. 2024), which limit their usage. A common workaround, inspired by the scaling law (Kaplan et al. 2020), is to collect posture-diverse training datasets for domain transformation. However, this approach is time-consuming, labor-intensive, and challenging to scale in real-world settings, as it requires lots of synchronized data collection and annotation across various postures.

Just as robust perception in speech and vision anchors to *invariant meaning*, recognizing a word across accents (Huang et al. 2022), or identifying an object from any angle (Chen, Ye, and Du 2022), we argue that robust cardiac sensing should anchor to physiology. While mmWave signals dramatically across postures, respiration, and occlusions, the underlying cardiac rhythm remains consistent. This motivates a paradigm shift: rather than directly mapping distorted signals to ECGs, we aim to learn a **physiology-aligned representation** that preserves intrinsic cardiac rhythm while remaining invariant to external distortions. Achieving this vision poses two key challenges:

- **Posture-Induced Representation Drift (C1).** Postural changes reshape radar reflections to such an extent that identical heartbeats yield drastically different signal patterns. Without explicit constraints, models trained on raw geometry produce embeddings that drift with posture, failing to isolate rhythm-consistent features.
- **Physiological and Morphological Variability across Individuals (C2).** Even if a stable rhythm is captured, mapping it to ECG remains ill-posed: the same latent rhythm can manifest as different waveforms across individuals

*These authors contributed equally.
†Corresponding authors

and conditions. This inherent ambiguity poses significant challenges for deep generative models, which must learn to capture the full variability and complexity of physiological signal distributions (Bond-Taylor et al. 2021).

To this end, we introduce `mmJEPA-ECG`, a *physiology-guided framework* that reconstructs clinically reliable ECGs from mmWave radar by anchoring sensing to invariant cardiac dynamics. Figure 1 depicts the practical application, with the device mounted above the bed to contactless monitor ECG signals continuously even as the user changes posture. To address **C1**, we design a *Physiology-Oriented Self-Supervised Pretraining* scheme based on Joint Embedding Predictive Architecture (JEPA) (Assran et al. 2023), incorporating domain knowledge to learn posture-robust cardiac embeddings. **The core idea is to predict masked information directly within the latent space, rather than reconstructing them.** Multi-channel mmWave signal patches are masked across temporal, channel, and spatiotemporal views to capture periodic rhythms and posture-dependent representation. Additionally, a heart rate consistency constraint aligns embeddings with periodic cardiac dynamics. To address **C2**, we implement a *Conditional Diffusion-based ECG Reconstruction* that models reconstruction as a stochastic denoising process, enabling the clinically salient features across diverse plausible outputs. Generation is regularized by two domain-informed objectives:(i) *noise-prediction loss* enforces stable denoising trajectories under ambiguous embeddings. (ii) *frequency alignment loss* promotes spectral coherence with ground-truth ECGs. Our main contributions are listed as follows:

- We propose `mmJEPA-ECG`, the first physiology-guided framework to anchor mmWave cardiac sensing to posture-invariant dynamics, addressing the breakdown of mapping-based methods under sleeping posture shifts.
- We design *Physiology-Oriented Self-Supervised Pretraining* which disentangles posture invariant cardiac representations via semantically multi-granularity masking and heart rate consistency, and utilize *Conditional Diffusion-based ECG Reconstruction* that leverages diffusion models enhanced by Hierarchical Radar Conditioning (HRC) to effectively integrate radar-derived physiological priors, enabling robust ECGs generation.
- Experimentally, we evaluate `mmJEPA-ECG` using `MMECG` public dataset and a nearly 300GB self-collected in-house dataset (10 subjects). `mmJEPA-ECG` generalizes across challenging real-world conditions which achieving $2\times$ lower ECG R-R peak relative error (from 15ms to 8ms) and higher waveform correlation (from 0.91 to 0.95) in supine posture, and uniquely preserving reconstruction fidelity under side-lying distortions and pathological rhythms (e.g., arrhythmia, tachycardia).

Codes — https://github.com/lanyangyang/mmJEPA-ECG

Related Work

ECG Reconstruction by mmWave Signal

Non-contact ECG monitoring with mmWave radar aims to map the relationship between cardiac electrical activity and thoracic vibrations. CardicWave (Xu et al. 2021) pioneers the Cardiac-mmWave Scattering Effect (CaSE) framework, leveraging electromagnetic scattering theory to reconstruct ECG yet overlooking RF characteristics. Subsequently, mmECG (Chen et al. 2022) exploits mmWave RF properties through transformer and Temporal Convolution Network (TCN) to map micro-vibrations to ECG; however, its performance degrades under noisy or complex data. RF-ECG (Wang et al. 2023) extends this paradigm to UWB signals, combining conditional Generative Adversarial Network (GAN) but suffers from mode collapse and heavy data demands. RSSRnet (Wu et al. 2023) uses U-net attention for reconstruction but is high frequency noise-sensitive; AirECG (Zhao et al. 2024) proposes a cross-domain diffusion model with calibration, improving ECG reconstruction yet raising complexity and data needs; RadarODE (Zhang et al. 2025) embeds Ordinary Differential Equation (ODE) decoders for interpretability, yet remains data-specific.

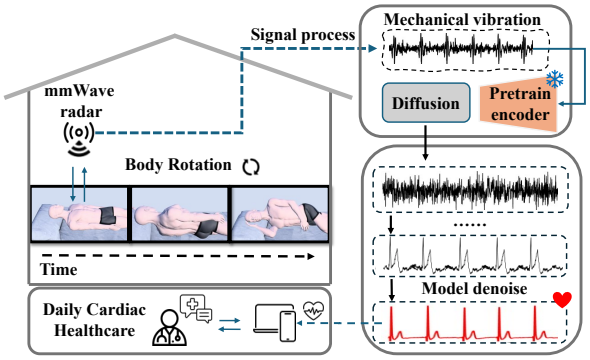


Figure 1: The application scenario of mmJEPA-ECG, which provides contactless ECG monitoring under multi-posture.

Preliminary

Physiological Basis of mmWave radar-to-ECG

mmWave radar provides a high-resolution (Deng, Xue, and Chen 2025), non-contact modality for sensing sub-millimeter thoracic cavity vibrations arising from cardiac activity even in darkness (Xu et al. 2022). Specifically, Frequency Modulated Continuous Wave (FMCW) radar transmits chirp signals and analyzes their phase shifts $\phi(t)$ to track surface displacements (Li et al. 2021) defined as:

$$\phi(t) = \frac{2\pi d(t)}{\lambda} \tag{1}$$

where $d(t)$ is chest displacement and λ is the radar wavelength, encoding cardiac-induced micron-scale motions.

The feasibility of radar-based ECG reconstruction arises from the heart's electromechanical coupling: each ECG segment (P wave, QRS complex, T wave) represents an electrical event that induces mechanical responses (Quinn and Kohl 2021) such as myocardial contraction and blood flow. These mechanical actions generate periodic thoracic vibrations, cause phase shifts in radar signals, which can be extracted and temporally aligned with cardiac activity, providing a biophysically grounded surrogate for ECG signals.

Problem Formulation

We formalize ECG transformation from mmWave radar as structured cross domain generation task that decodes cardiac electrical dynamics from radar phase condition. Specifically, given a mmWave phase signal sequence $X_i = [\phi_i^1, \cdots, \phi_i^T]$ associated with subject i, the goal is to learn a function $\mathcal{D}_\theta$ that reconstructs the corresponding ECG waveform sequence $Y_i = [e_i^1, \cdots, e_i^T]$.

$$\hat{Y}_i = \mathcal{D}_\theta(X_i, \epsilon), \quad \epsilon \sim \mathcal{N}(0, \mathbf{I}) \tag{2}$$

where ϵ is Gaussian noise with identity matrix $\mathcal{N}(0, \mathbf{I})$.

Methodology

In this section, we outline the methodology of mmJEPA-ECG with three components (illustrated in figure 2):

- **Data Pre-processing:** Extracts and amplifies thoracic micro-vibrations from radar data using range-Fast Fourier Transform (FFT) and beamforming, followed by suppression of motion artifacts with filtering.
- **Physiology-Oriented Self-Supervised Pretraining:** Extends JEPA with domain-specific adaptations for **C1**. Specifically, (i) tokenize pre-processed thoracic vibration signal into cardiac cycle aligned temporal patch embeddings; (ii) apply multi-granularity masking across time and spatial views to simulate structured corruption; (iii) enforce heart rate consistency across views to regularize embeddings and align with underlying vital dynamics.
- **Conditional Diffusion-based ECG Reconstruction:** Adopt probabilistic treatment for **C2**, as (i) we utilize Diffusion Transformer (DiT) with HRC, integrating pretrained radar features across layers for fine-grained cross-modal conditioning; (ii) enforce dual-domain supervision by applying a time-domain loss for noise prediction and a frequency-domain loss for the generated signals, ensuring both morphological and spectral fidelity; (iii) apply Denoising Diffusion Implicit Models (DDIM) based fast sampling to enable real-time, high-fidelity ECG synthesis.

Data Pre-processing

To ensure high-fidelity ECG reconstruction, we implement a radar signal processing pipeline that (i) localizes the thoracic wall via maximum vibration energy detection, (ii) uses sum-and-delay beamforming to enhance target signals over the thoracic surface, and (iii) applies second-order differential filtering to suppress respiratory interference while preserving fine-grained cardiac micro-vibrations.

Distance-Domain Projection: We begin with the raw mmWave radar signals and convert them into a range FFT (Brigham and Morrow 1967) data cube to analyze reflections across different distances. To localize the thoracic region primarily driven by cardiac motion, we compute the vibration energy at each range bin and select the one with maximal energy, using round-trip length (RTL) analysis:

$$\text{RTL} = \arg\max_r \sum_t |S(r,t)|^2 \tag{3}$$

where $S(r,t)$ is the radar signal at range r and time t.

Beamforming to Enhance Target Voxel Signal: To further enhance the signal quality at the target thoracic region, we perform *sum-and-delay beamforming* (Matrone et al. 2014) across all antenna channels. This coherently aligns and aggregates signals from different radar elements. The final signal martrix as $S(x,y,z,t)$:

$$S(x,y,z,t) = \sum_{n=1}^{N} \sum_{t=1}^{T} y_{n,t}\, e^{j2\pi \frac{kd(x,y,z,n)}{c} t}\, e^{j2\pi \frac{d(x,y,z,n)}{\lambda}} \tag{4}$$

where $y_{n,t}$ is the signal from the n-th antenna at time t, $d(x,y,z,n)$ is its distance to the scan point (x,y,z), λ is the wavelength and c is the speed of light.

Cardiac Micro-Vibration Extraction and Filtering: We extract the phase signal $\phi(t)$ from the focused voxel, which captures sub-millimeter displacements of the thoracic wall induced by cardiopulmonary activity as Eq. (1) where $d(t)$ is the instantaneous thoracic displacement. Since this phase includes both cardiac and respiratory components, we apply a *second-order differential filter* (Holoborodko 2014) to suppress low-frequency respiratory effects and retain high-frequency cardiac micro-vibrations as ϕ_0'':

$$\phi_0'' = \frac{(\phi_{-3} + \phi_3) + 2\,(\phi_{-2} + \phi_2) - (\phi_{-1} + \phi_1) - 4\phi_0}{16\,(\Delta t)^2} \tag{5}$$

where ϕ_k is the value at k, and Δt is the sampling interval.

Physiology-Oriented Self-Supervised Pretraining

To address **C1**, we introduce *Physiology-Oriented Self-Supervised Pretraining*, which extends the Image-JEPA (Assran et al. 2023) with domain-specific adaptations to mmWave radar signals. While Image-JEPA is effective in visual domains, it falls short in our setting due to two key issues: (i) *structured multi-view corruption* arising from occlusions, and posture variation; and (ii) the *absence of physiological coherence* across masked views.

To this end, we design three coordinated mechanisms: i) *Temporal Patch Tokenization* segments radar inputs into patch tokens; ii) *Multi-Granularity Masked Strategies* occludes time and space simulate posture induced partial input; and iii) *Vital Sign Regularization via Heart Rate Consistency* enforces physiological invariance across views.

Temporal Patch Tokenization: We first partition $X_i \in \mathbb{R}^{T \times C}$ into N non-overlapping temporal patches to preserve local physiological patterns (e.g., waveform morphology within a cardiac cycle) (Nie et al. 2022). Each patch is added with learnable position embedding.

$$\mathcal{P}_i = (X_i[(n-1)W + 1 : nW])_{n=1}^{N}, \quad \mathcal{P}_i \in \mathbb{R}^{N \times W \times C} \tag{6}$$

Here, $\mathcal{P}_i$ is a sequence of N non-overlapping temporal patches of length W, C is the number of channels.

Multi-Granularity Masked Strategies: To enforce robustness under real-world corruptions, we define three distinct mask targets in: *Physiology-Oriented Self-Supervised Pretraining* as Figure 2: (i) temporal masking $\mathcal{M}_{\text{time}}$ (random patch removal on time axis), (ii) channel masking

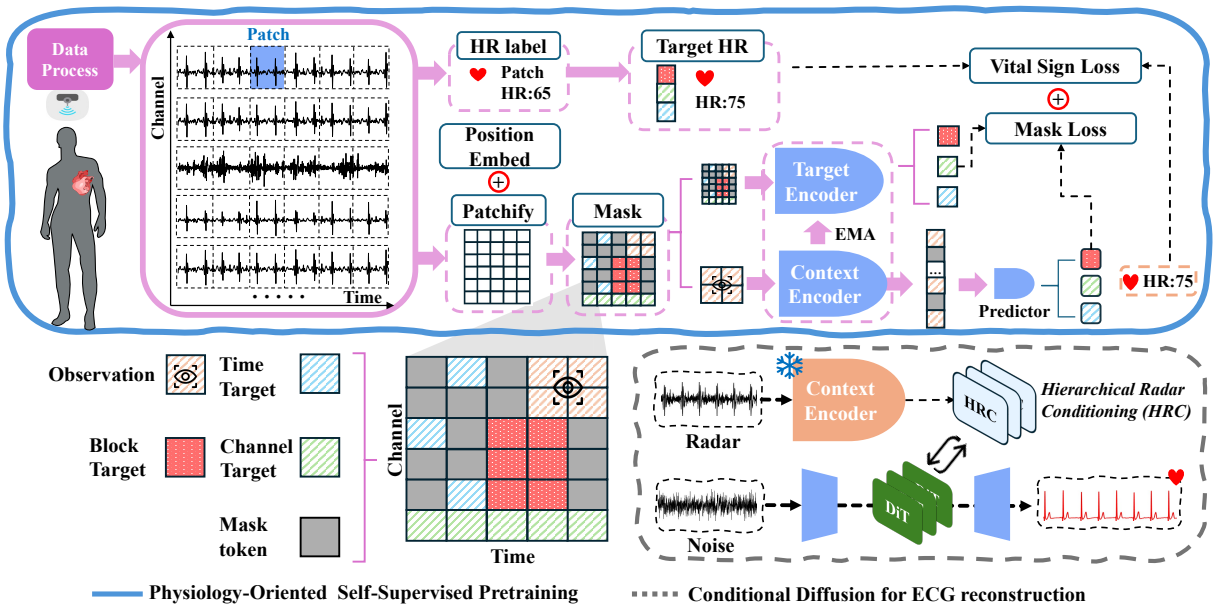


Figure 2: System overview with three strategies: (i) Data processing (ii) Physiology-Oriented Self-Supervised Pretraining (iii) Conditional Diffusion-based ECG Reconstruction.

$\mathcal{M}_{\text{chan}}$ (random patch removal on channel axis), and (iii) joint masking $\mathcal{M}_{\text{joint}}$ (random block patches combined temporal and spatial disruptions).

We first select an observation region non-overlapping with the masked area as the context input $\mathcal{P}_i^{(ob)}$ for the context encoder f_θ, while all patches are fed to the target encoder f'_θ updated via momentum. The context encoder output $f_\theta(\mathcal{P}_i^{(ob)})$ is combined with mask tokens η and passed to the predictor g_φ. This process repeats for each masking strategy m, yielding predictions $\hat{\mathbf{z}}_i^{(m)}$ with shared weights. Predicted features are compared with the corresponding regions of the target encoder output $\mathbf{z}_i^*$ as described in Eq. (7):

$$\hat{\mathbf{z}}_i^{(m)} = g_\varphi(f_\theta(\mathcal{P}_i^{(ob)}) + \eta), \quad \mathbf{z}_i^* = [f'_\theta(\mathcal{P}_i)]_m \tag{7}$$

Vital Sign Regularization via Heart Rate Consistency: Although the multi-mask strategy enhances context-aware feature learning, it does not explicitly enforce physiological consistency across postures. To address this, we design an auxiliary task that employs heart rate, computed from each radar patch via band-pass filtering (0.18–2.5 Hz) and peak detection, as a posture-invariant physiological anchor.

In this task, each patch is assigned a pseudo ground-truth heart rate label h_i. The predictor g_φ not only predicts the mask representation but also predicts the corresponding heart rate $\hat{h}_i$ based on the observation data. We define a weighted mean squared error as the vital sign consistency loss as $\mathcal{L}_{\text{vt}}$.

Loss Function of Pretraining Stage: Our pretraining objective jointly optimizes semantic representation learning and physiological consistency as a total loss ($\mathcal{L}_{\text{mmJEPA}}$):

$$\mathcal{L}_{\text{mmJEPA}} = \mathcal{L}_{\text{mask}} + \lambda_{\text{vt}}\mathcal{L}_{\text{vt}} \tag{8}$$

where the loss terms $\mathcal{L}_{\text{mask}}$ and $\mathcal{L}_{\text{vt}}$ are,

$$\mathcal{L}_{\text{mask}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{m\in\mathcal{M}}\left\|\hat{\mathbf{z}}_i^{(m)} - \mathbf{z}_i^*\right\|_1 \tag{9}$$

$$\mathcal{L}_{\text{vt}} = \frac{1}{N}\sum_{i=1}^{N} w_i\left(\hat{h}_i - h_i\right)^2, w_i = \begin{cases} W_a, & \text{if } |\hat{h}_i - h_i| < \delta \\ W_b, & \text{otherwise} \end{cases} \tag{10}$$

where δ is a physiological tolerance (e.g., 5 BPM), and $W_b \gg W_a$ emphasizes large deviations. $\mathcal{M}$ indexes the masking targets ($\mathcal{M}_{\text{time}}, \mathcal{M}_{\text{chan}}, \mathcal{M}_{\text{joint}}$).

Conditional Diffusion-based ECG Reconstruction

Pretraining on radar data yields generalizable representations, which subsequently guide accurate ECG reconstruction. To address **C2**, we design DiT (Peebles and Xie 2023a) with Hierarchical Radar Conditioning (HRC) to reconstruct ECG with context from *Physiology-Oriented Self-Supervised Pretraining* encoder using three mechanisms:

HRC-DiT for Noise Prediction: The training objective of diffusion-based models is to predict the added noise on data at each step. In inference, it follows a fixed denoising process, starting with pure noise input and iteratively reconstructing the data. Specifically, as illustrated in Figure 3, the HRC-DiT block receives the Gaussian-noise-perturbed ECG E_i^t as input and uses the radar signal X_i^t as condition. The objective is to estimate added noise ϵ_θ at each step.

The HRC-DiT consists of 6 stacked blocks, each integrating two core components: Multi-Head Self-Attention (MHSA) for modeling temporal dependencies within the ECG, and Multi-Layer Cross-Attention (MLCA) (Wang et al. 2024) for radar-conditioned fusion. At each block,

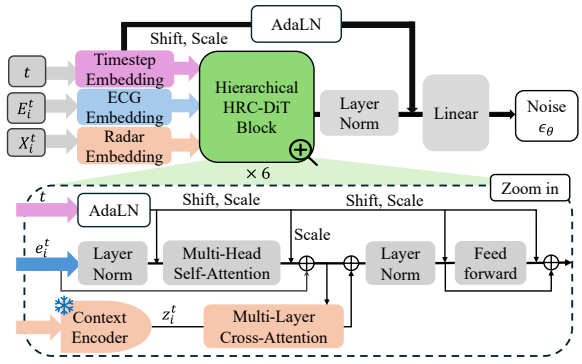


Figure 3: The stacked HRC-DiT diffusion backbone.

MHSA is first applied to the ECG token embeddings $\mathbf{e}_i^t$ to capture intra-beat representation. The resulting features are then fused with radar context $\mathbf{z}_i^t$ via MLCA, enabling joint modeling of ECG dynamics and radar priors. Formally, the radar-conditioned hidden representation is defined as:

$$\mathbf{H} = \mathrm{MLCA}\left(\mathrm{MHSA}(LN(\mathbf{e}_i^t)) + \mathbf{e}_i^t,\ \mathbf{z}_i^t\right) \tag{11}$$

where $\mathrm{LN}(\cdot)$ denotes layer normalization.

To further enhance generative flexibility across diffusion steps, we employ Adaptive LayerNorm (AdaLN) (Peebles and Xie 2023a) with learnable, timestep-dependent scaling and shifting:

$$\mathrm{AdaLN}(h, t) = \gamma(t) \cdot \mathrm{LN}(h) + \beta(t) \tag{12}$$

where h denotes the current hidden state, and $[\gamma(t), \beta(t)] = \mathrm{MLP}(t)$, with learnable Multilayer Perceptron (MLP) that maps diffusion timestep t to scale and shift parameters.

Fast Deterministic Sampling: To meet real-time constraints, we adopt deterministic DDIM sampling (Song, Meng, and Ermon 2020) in place of the standard stochastic approach, enabling efficient generation with significantly fewer steps. DDIM at each timestep is updated by:

$$x_{t-1} = \sqrt{\alpha_{t-1}}\, D_\theta(x_t, t) + \sqrt{1 - \alpha_{t-1}}\, z, \tag{13}$$

where $D_\theta(x_t, t)$ denotes the model prediction at step t, $z = 0$ for deterministic sampling, and α_{t-1} denotes the signal retention factor at step $t-1$ in the diffusion process.

Loss Function of Diffusion: To ensure both morphological accuracy and diagnostic fidelity, we supervise the model with two complementary objectives: (i) a noise prediction loss (Peebles and Xie 2023b), and (ii) a frequency-domain loss between the generated and ground truth signals. The final loss function $\mathcal{L}_{\text{diffusion}}$ as Eq. (14)

$$\mathcal{L}_{\text{diffusion}} = \mathcal{L}_{\text{noise}} + \lambda_{\text{freq}}\, \mathcal{L}_{\text{freq}} \tag{14}$$

where the loss terms are defined as,

$$\mathcal{L}_{\text{noise}} = \mathrm{MSE}\left(\epsilon_\theta(e_i^t, t, \mathbf{z}_i^t), \epsilon\right) \tag{15}$$

$$\mathcal{L}_{\text{freq}} = \mathrm{MAE}\left(\mathcal{F}(\hat{\mathbf{Y}}_i^t),\ \mathcal{F}(\mathbf{Y}_i^t)\right) \tag{16}$$

where ϵ is the true noise, ϵ_θ the predicted noise, e_i^t is the noisy ECG, and $\mathbf{z}_i^t$ the radar condition at time step t. $\hat{\mathbf{Y}}_i^t$ and $\mathbf{Y}_i^t$ are generated and reference ECG signals, and $\mathcal{F}$ denotes the Fourier transform.

Experiments

Evaluation Settings

Datasets. We evaluate our model on two datasets:

- MMECG (Public Benchmark, only supine posture): This dataset contains 91 trials from 11 subjects (age from 18 to 65), each providing 3 minutes of synchronized radar and ECG signals sampled at 200 Hz. Data spans four test conditions: normal/abnormal breathing, post-exercise, and sleep, which apply diverse cardiac dynamics.
- Self-collected (Multi-Posture): Our in-house dataset consists of two parts (i) *Unpaired Radar Data*: 280 GB of overnight mmWave radar data recordings from 3 subjects; (ii) *Paired Radar–ECG Data*: 3 GB synchronized data of 3 hours recording from 10 subjects (6 males, 4 females), evenly collected across three postures (i.e., supine, left-/right lateral). Both radar and ECG data are sampled at 200Hz. This setup highlights `mmJEPA-ECG` can be effectively trained with limited paired data.

Baselines. We compare our model with five baselines: **(i) MMECG** (Chen et al. 2022) uses temporal convolutional and transformer-based model for radar-to-ECG mapping. **(ii) Rssrnet** (Wu et al. 2023) designs a U-Net encoder-decoder with STFT weights for ECG reconstruction. **(iii) RF-ECG** (Wang et al. 2023) utilizes a GAN-based framework for ECG generation. **(iv) AirECG** (Zhao et al. 2024) leverages a diffusion-based model for radar-to-ECG generation. **(v) RadarODE** (Zhang et al. 2025) uses neural ODE solvers to model radar-to-ECG translation.

Evaluation Metrics. We evaluate `mmJEPA-ECG` with three types of metrics that are widely used in prior studies:

- **Event-based Timing Errors Metrics:** We quantify the time accuracy of key cardiac events (Q-Q, R-R, T-T, S-S) by measuring absolute timing errors between corresponding peaks in the generated and reference ECGs, reflecting the precision for clinical interpretation.
- **Morphological Metrics:** We report normalized cosine correlation (CC) and root mean square error (RMSE) to assess the waveform fidelity and the overall performance.
- **Interval-based Clinical Metrics:** We assess clinically relevant intervals (Chazal 2004), which including R-R, QRS, Q-T, and P-R by computing relative differences. This metric are used in downstream task as they are essential for arrhythmia and heart rate variability analysis.

Implementation Details. (i) **Data Acquisition Platform:** We employ a TI IWR6843AOP FMCW mmWave radar (3 Tx, 4 Rx) integrated with DCA1000EVM board, ICBOOST board (Texas Instruments 2024) and a CFDA certified PC-80B ECG monitor (Shenzhen Creative Industry 2024) (both are sampled at 200 Hz). The radar is mounted 50cm above the chest and face toward the target as shown in figure 4. (ii) **Experiments Settings:** Radar data are segmented into 512-sample windows with 50 channels. The mmJEPA encoder is a two-layer Patch Transformer (Nie et al. 2022) (d_{model}=192, d_{ff}=256, 8 heads), with patch size and stride of 32. Multi-granularity masking with (2 time, 10 channel,

Methods	Supine Posture						Multi-Posture					
	Q-Q	R-R	S-S	T-T	RMSE	CC	Q-Q	R-R	S-S	T-T	RMSE	CC
MMECG (2022)*	0.048	0.023	0.029	0.035	0.120	0.855	0.067	0.036	0.047	0.039	0.203	0.697
Rssrnet (2023)*	0.017	0.020	0.018	0.021	0.112	0.916	0.055	0.049	0.050	0.047	0.189	0.652
RF-ECG (2023)*	0.075	0.018	0.035	0.064	0.170	0.803	0.062	0.060	0.057	0.067	0.223	0.565
AirECG (2024)	0.035	0.020	0.023	0.020	0.168	0.823	0.073	0.069	0.074	0.071	0.252	0.441
RadarODE (2025)	0.027	0.015	0.020	0.023	0.097	0.896	–	–	–	–	0.392	0.235
Ours	**0.010**	**0.008**	**0.011**	**0.010**	**0.069**	**0.952**	**0.011**	**0.008**	**0.009**	**0.013**	**0.088**	**0.924**

Table 1: Comparison with current radar-ECG reconstruction methods on supine and multi-Posture datasets. Peaks errors are reported in seconds (s). **Bold** indicates the state-of-the-art result, and underline denotes the second-best. Methods marked with * were built from scratch as no public code. − indicates the method not generate meaningful results for cardiac event evaluation.

8×8 block). The predictor is a single-layer Patch Transformer (d_{model}=192, d_{ff}=64, 4 heads) with an MLP head. Diffusion-based model uses stacked 6 HRC-DiT blocks (d_{model}=384, d_{ff}=1536, 6 heads), with a linear noise schedule (β_{noise}=(0.0001,0.02), steps=1000). The toolkit used for cardiac event metrics is NeuroKit2 (Makowski et al. 2021). The platform is 4 NVIDIA RTX4090 GPU and Intel Xeon Platinum 8352V CPU. Additional optimization details are provided in the Appendix.(iii) **Computation & Energy.** The N attention tokens dominates computational complexity, yielding $\mathcal{O}(N^2)$ for training and $\mathcal{O}(T \times N^2)$ for inference steps. The model has 20.6M parameters, with per-step complexity of 0.69 GFLOPs, with about $10ms$ per step latency. Power draw averages $82\,W$ (8.2 J/sample).

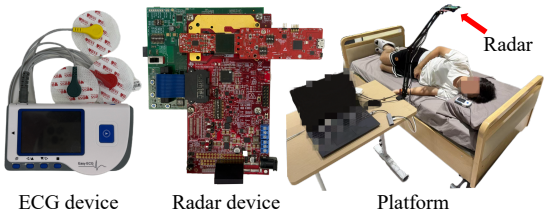


Figure 4: Synchronized radar and ECG acquisition platform.

Overall Performance

Table 1 presents a comprehensive comparison across event-based timing errors and morphological metrics. All results are reported as normalized mean values. Among baselines, *Rssrnet* performs avage well on each peaks error and correlation in supine posture. *RadarODE* excels in R-R estimation and second-best RMSE due to its continuous-time dynamics, but lacks sufficient resolution to accurately capture smooth morphological events such as Q-Q. *RF-ECG* achieves low R-R error but exhibits extremely high errors in other intervals, suggesting it captures rhythm but sacrifices waveform fidelity, likely due to its adversarial objective. *MMECG* and *AirECG* achieve balanced but suboptimal results. However, all existing methods perform poorly when evaluated on different postures, as they fail to extract the underlying physical consistency across postural variations. *RadarODE* even not able to generate meaningful cardiac pattern, making it impossible to compute certain cardiac event errors due to limitation of its ODE structure.

In contrast, `mmJEPA-ECG` achieves state-of-the-art performance across all metrics, with consistent improvements in both cardiac event accuracy (up to 50% error reduction in the supine posture) and waveform similarity. These improvements stem from self-supervised pretraining and conditional diffusion-based reconstruction, which jointly preserve global rhythms and detailed morphological features essential for clinically valid ECG synthesis. Figure 5 visualizes our reconstructed ECG signals, which closely match the reference over multiple cycles and key cardiac events.

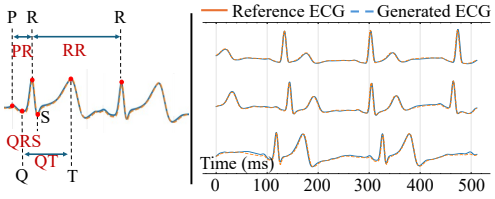
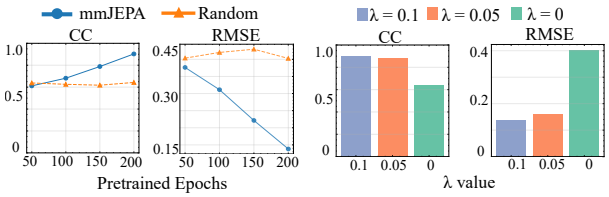


Figure 5: (a) Cardiac events (b) Generated vs reference.



(a) Different mask strategies (b) Different λ weights

Figure 6: Ablation studys of different mask strategies and different weights of vital sign consistency loss.

Key Component Assessment

To evaluate pre-training components, we conduct an ablation study on two aspects: (i) the proposed *Multi-Granularity Masked Strategies* versus random time masking under different pre-training epochs; (ii) the weight of vital sign consistency loss in `mmJEPA-ECG` pre-training. All results are reported after full training, with only encoder configurations differing across variants.

We focus on morphological metrics since competing methods fail to generate meaningful cardiac events. Fig-

ure 6 shows that our multi-granularity masking strategy significantly improves both correlation and RMSE over 50–200 training epochs, while random masking fails to learn posture-invariant features (stagnant performance). Notably, removing the vital sign consistency loss reduces correlation and increases RMSE, whereas introducing even a small weight (e.g., 0.05) yields substantial gains.

Robustness Analysis with User Independence Test

To evaluate generalization to unseen users, we perform 10-fold user-independent cross-validation on the self-collected dataset (6 males, 4 females), holding out one subject per fold for testing and training on the other nine. Relative errors of four cardiac peaks (Q, R, S, T) are computed per subject and normalized by cardiac intervals to yield error ratios.

As shown in Figure 7, the relative errors of all four cardiac peaks remain consistently low across all 10 subjects, generally below 6%. There are individual differences among subjects; for example, Subject 3 exhibits the lowest R-R error, while its S-S error is comparatively higher. The CC and RMSE exhibit little variation among subjects. The low variance across subjects demonstrates the mmJEPA−ECG's robust performance and adaptability to individual cardiac dynamics without prior exposure.

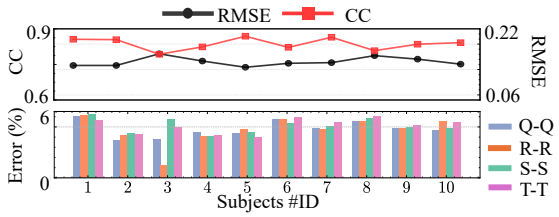


Figure 7: Evaluation results of ten subjects with cardiac events relative peaks error, CC and RMSE.

To further validate the robustness of our model, we performed a supplementary **Statistical Significance Test** using bootstrap resampling on our dataset. The analysis was conducted over 14 groups (20 iterations, 0.8 ratio), with the following results: **CC Mean**: 0.921 ± 0.002 (95% CI [0.918, 0.925]) **RMSE Mean**: 0.090 ± 0.004 (95% CI [0.081, 0.096]) The bootstrap narrow confidence intervals indicate that the results are statistically stable.

Clinical Potential: Arrhythmia Assessment

To evaluate mmJEPA−ECG's clinical robustness, we test it on a public dataset with arrhythmic ECG signals and protocol-induced rhythm disturbances (Chen et al. 2022), which includes realistic pathological challenges (e.g., tachycardia, arrhythmia). We report *Interval-based Clinical Metrics* to assess its clinical potential.

Figure 8 shows our framework achieves low R-R, QRS, and Q-T interval reconstruction errors. The P-R interval has higher error (14.1%) due to inherent arrhythmic variability. The reconstructed signals correlate highly with ground truth (mean: 0.80, median: 0.92). Figure 9 shows example results with tachycardia and arrhythmia. All the results indicate mmJEPA−ECG's reliability under abnormal conditions, as effectively preserving important clinical cardiac events.

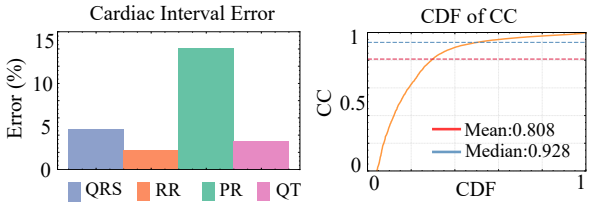


Figure 8: Evaluation of clinical interval and morphology.

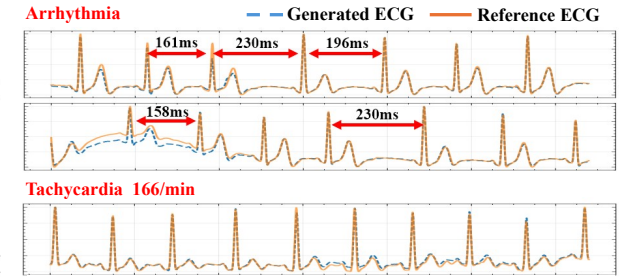


Figure 9: Result samples with tachycardia and arrhythmia.

Discussion

Insight: We found that self-supervised pretraining enables the encoder to extract robust and transferable physiological patterns from unlabeled radar data, significantly improving ECG reconstruction and generalization. In particular, mmJEPA with **cross-view structure, multi-granularity masking, and physiological rhythm-aware designs** effectively captures global context and posture-invariant features essential for accurate contactless ECG monitoring.

Conclusion

This work is the first to address radar-to-ECG reconstruction across varying sleeping postures, rethinking the task from direct signal mapping to physiology-anchored reasoning. We introduce mmJEPA−ECG, a physiology-guided framework that addresses two challenges: disentangling posture-induced artifacts from true heart rhythms, and synthesizing clinically reliable ECG morphologies under ambiguity. It addresses these challenges with two components: a *Physiology-Oriented Self-Supervised Pretraining* that learns posture-invariant cardiac embeddings via structured masking and heart-rate-guided alignment; and a *Conditional Diffusion-based ECG Reconstruction* that stochastically reconstructs clear, personalized waveforms, regularized by spectral coherence and noise prediction. Empirically, mmJEPA−ECG improves waveform fidelity on both datasets, remaining robust under posture variations and arrhythmia. Looking ahead, we hope this work advances accessible, continuous health monitoring, enabling earlier disease detection and improved public health.

Ethics Statement

All data were collected with informed consent, and processed under data minimization and anonymization principles. Millimeter-wave radar captures only non-identifiable physiological motion. No personal appearance or identity information can be inferred from raw millimeter-wave (mmWave) radar signals. Distinct from conventional video or audio sensing modalities, mmWave radar functions purely as a radio-frequency motion sensor, capturing only low-level, non-identifiable physiological displacement patterns.

Acknowledgments

This work was supported in part by the Guangdong Provincial Key Laboratory of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things (No. 2023B1212010007), the National Natural Science Foundation of China (NSFC) under Grant 62472366, the Project of DEGP (No. 2024GCZX003, 2023KCXTD042), the 111 Center (No. D25008), and the Shenzhen Science and Technology Foundation (No. ZDSYS20190902092853047).

References

Assran, M.; Duval, Q.; Misra, I.; Bojanowski, P.; Vincent, P.; Rabbat, M.; LeCun, Y.; and Ballas, N. 2023. Self-supervised learning from images with a joint-embedding predictive architecture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15619–15629.

Bond-Taylor, S.; Leach, A.; Long, Y.; and 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(11): 7327–7347.

Brigham, E. O.; and Morrow, R. E. 1967. The fast Fourier transform. *IEEE Spectrum*, 4(12): 63–70.

Brignole, M.; Vardas, P.; Hoffman, E.; Huikuri, H.; Moya, A.; Ricci, R.; Sulke, N.; Wieling, W.; Auricchio, A.; Lip, G. Y.; et al. 2009. Indications for the use of diagnostic implantable and external ECG loop recorders. *Europace*, 11(5): 671–687.

Chazal, P. D. 2004. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE transactions on biomedical engineering*, 51(7): 1196–1206.

Chen, J.; Zhang, D.; Wu, Z.; Zhou, F.; Sun, Q.; and Chen, Y. 2022. Contactless electrocardiogram monitoring with millimeter wave radar. *IEEE Transactions on Mobile Computing*, 23(1): 270–285.

Chen, S.; Ye, M.; and Du, B. 2022. Rotation invariant transformer for recognizing object in uavs. In *Proceedings of the 30th ACM International Conference on Multimedia*, 2565–2574.

Deng, H.; Xue, T.; and Chen, H. 2025. Fusegrasp: Radar-camera fusion for robotic grasping of transparent objects. *IEEE Transactions on Mobile Computing*.

DiMarco, J. P.; and Philbrick, J. T. 1990. Use of ambulatory electrocardiographic (Holter) monitoring. *Annals of internal medicine*, 113(1): 53–68.

Gami, A. S.; Howard, D. E.; Olson, E. J.; and Somers, V. K. 2005. Day–night pattern of sudden death in obstructive sleep apnea. *New England Journal of Medicine*, 352(12): 1206–1214.

Holoborodko, P. 2014. Noise robust differentiators for second derivative estimation.

Huang, P.-Y.; Xu, H.; Li, J.; Baevski, A.; Auli, M.; Galuba, W.; Metze, F.; and Feichtenhofer, C. 2022. Masked autoencoders that listen. *Advances in Neural Information Processing Systems*, 35: 28708–28720.

Jaspan, V. N.; Greenberg, G. S.; Parihar, S.; Park, C. M.; Somers, V. K.; Shapiro, M. D.; Lavie, C. J.; Virani, S. S.; and Slipczuk, L. 2024. The role of sleep in cardiovascular disease. *Current atherosclerosis reports*, 26(7): 249–262.

Kaplan, J.; McCandlish, S.; Henighan, T.; Brown, T. B.; Chess, B.; Child, R.; Gray, S.; Radford, A.; Wu, J.; and Amodei, D. 2020. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

Li, X.; Wang, X.; Yang, Q.; and Fu, S. 2021. Signal processing for TDM MIMO FMCW millimeter-wave radar sensors. *IEEE Access*, 9: 167959–167971.

Makowski, D.; Pham, T.; Lau, Z. J.; Brammer, J. C.; Lespinasse, F.; Pham, H.; Schölzel, C.; and Chen, S. H. A. 2021. NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods*, 53(4): 1689–1696.

Matrone, G.; Savoia, A. S.; Caliano, G.; and Magenes, G. 2014. The delay multiply and sum beamforming algorithm in ultrasound B-mode medical imaging. *IEEE transactions on medical imaging*, 34(4): 940–949.

Mercuri, M.; Lorato, I. R.; Liu, Y.-H.; Wieringa, F.; Hoof, C. V.; and Torfs, T. 2019. Vital-sign monitoring and spatial tracking of multiple people using a contactless radar-based sensor. *Nature Electronics*, 2(6): 252–262.

Nie, Y.; Nguyen, N. H.; Sinthong, P.; and Kalagnanam, J. 2022. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*.

Peebles, W.; and Xie, S. 2023a. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, 4195–4205.

Peebles, W.; and Xie, S. 2023b. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, 4195–4205.

Quinn, T. A.; and Kohl, P. 2021. Cardiac mechano-electric coupling: acute effects of mechanical stimulation on heart rate and rhythm. *Physiological reviews*, 101(1): 37–92.

Shenzhen Creative Industry. 2024. Lepu PC-80B Portable ECG Monitor. https://www.creative-sz.com/products/lepu-pc-80b-portable-ecg-monitor-easy-ekg-machine-handheld/. Accessed: 2024-07-16.

Siontis, K. C.; Noseworthy, P. A.; Attia, Z. I.; and Friedman, P. A. 2021. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nature Reviews Cardiology*, 18(7): 465–478.

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

Texas Instruments. 2024. MMWAVEICBOOST: mmWave EVM System Development Kit. https://www.ti.com.cn/tool/cn/MMWAVEICBOOST. Accessed: 2024-07-16.

Walsh III, J. A.; Topol, E. J.; and Steinhubl, S. R. 2014. Novel wireless devices for cardiac monitoring. *Circulation*, 130(7): 573–581.

Wang, H.; Guo, P.; Zhou, P.; and Xie, L. 2024. Mlca-avsr: Multi-layer cross attention fusion based audio-visual speech recognition. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 8150–8154. IEEE.

Wang, Z.; Jin, B.; Li, S.; Zhang, F.; and Zhang, W. 2023. ECG-grained cardiac monitoring using uwb signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(4): 1–25.

WHO. 2025. Cardiovascular diseases. https://www.who.int/health-topics/cardiovascular-diseases. Last accessed March 12, 2024.

Wu, Y.; Ni, H.; Mao, C.; and Han, J. 2023. Contactless reconstruction of ECG and respiration signals with mmWave Radar based on RSSRnet. *IEEE Sensors Journal*, 24(5): 6358–6368.

Xu, C.; Li, H.; Li, Z.; Zhang, H.; Rathore, A. S.; Chen, X.; Wang, K.; Huang, M.-c.; and Xu, W. 2021. Cardiacwave: A mmwave-based scheme of non-contact and high-definition heart activity computing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(3): 1–26.

Xu, X.; Yu, J.; Ma, C.; Ren, Y.; Liu, H.; Zhu, Y.; Chen, Y.-C.; and Tang, F. 2022. mmECG: Monitoring Human Cardiac Cycle in Driving Environments Leveraging Millimeter Wave. In *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications*, 90–99.

Zhang, D.; Zhang, X.; Xie, Y.; Zhang, F.; Yang, H.; and Zhang, D. 2024. From single-point to multi-point reflection modeling: Robust vital signs monitoring via mmwave sensing. *IEEE Transactions on Mobile Computing*.

Zhang, Y.; Guan, R.; Li, L.; Yang, R.; Yue, Y.; and Lim, E. G. 2025. radarODE: An ODE-embedded deep learning model for contactless ECG reconstruction from millimeter-wave radar. *IEEE Transactions on Mobile Computing*.

Zhao, L.; Lyu, R.; Lei, H.; Lin, Q.; Zhou, A.; Ma, H.; Wang, J.; Meng, X.; Shao, C.; Tang, Y.; et al. 2024. AirECG: Contactless electrocardiogram for cardiac disease monitoring via mmWave sensing and cross-domain diffusion model. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(3): 1–27.