PHYSICS-INFORMED REPRESENTATION LEARNING FOR ELECTROCARDIOGRAMS

Anonymous authors
Paper under double-blind review

ABSTRACT

Electrocardiograms (ECGs) are central to cardiac screening and diagnosis. Recent advances in self-supervised learning have enabled the development of generalizable ECG representations from large-scale unlabeled recordings. However, incorporating the physical structure of multi-lead ECG acquisition into self-supervised representation learning remains relatively underexplored. Each lead observes cardiac electrical activity from a distinct viewpoint determined by its positioning. The resulting signals exhibit structured dependencies across both lead and time dimensions. Building on these physical priors, we propose Φ-ECG, a **PH**ysics-**I**nformed **ECG** self-supervised representation learning framework. Φ-ECG encodes three-dimensional lead directions using real spherical harmonic basis functions, explicitly incorporating acquisition into lead representations. We further introduce structured lead–time masking that samples visible leads separately from the limb and precordial groups and applies a shared temporal mask across them. By limiting reconstruction shortcuts arising from inter-lead redundancy and withholding simultaneous observations at masked time positions, this design encourages the model to jointly exploit complementary lead observations and temporal context for reconstruction. We evaluate Φ-ECG across ECG disease classification, echocardiographic phenotype prediction, reduced-lead transfer, and multi-lead reconstruction. Across these settings, Φ-ECG consistently improves over strong self-supervised baselines, demonstrating robust generalization across diverse downstream tasks and evaluation settings. These results highlight the value of incorporating physically grounded priors into self-supervised ECG representation learning. The code will be released upon acceptance.

1 INTRODUCTION

Electrocardiograms (ECGs) are central to cardiac screening, clinical diagnosis, and risk assessment. Deep learning has enabled automated ECG analysis for applications such as arrhythmia detection and cardiac dysfunction screening (Hannun et al., 2019; Attia et al., 2019; Ribeiro et al., 2020; Poterucha et al., 2025). However, developing task-specific models typically requires large annotated datasets, whose acquisition depends on costly clinical interpretation. Self-supervised learning offers a scalable alternative by learning transferable representations from large collections of unlabeled ECG recordings (Liu et al., 2024; Na et al., 2024; Jin et al., 2025; Lin et al., 2026; Gu et al., 2026; Huang et al., 2026). In addition to data scale, effective ECG pretraining also depends on how acquisition-specific structure informs both signal representations and pretraining objectives.

Recent ECG self-supervised methods have increasingly incorporated ECG-specific structures into the pretraining stage. ST-MEM (Na et al., 2024) captures spatio-temporal relationships through masked reconstruction, ECG-FM (McKeen et al., 2025) combines masked signal modeling and contrastive learning with random lead masking, and HeartLang (Jin et al., 2025) models ECG morphology and rhythm at different semantic scales. In parallel, Nef-Net (Chen et al., 2021a) and LVCG (Huang et al., 2026) incorporate lead observation directions for ECG view synthesis or modeling in vectorcardiogram space. These advances motivate further investigation into how lead geometry and structured signal dependencies can jointly inform the design of representations and pretraining objectives of self-supervised ECG models.

To address this gap, we propose Φ-ECG, a **PH**ysics-**I**nformed **ECG** self-supervised representation learning framework that incorporates the physical structure of ECG acquisition. First, we build on the directional nature of multi-lead ECG acquisition, as illustrated in Fig. 1(a). Lead-vector theory

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

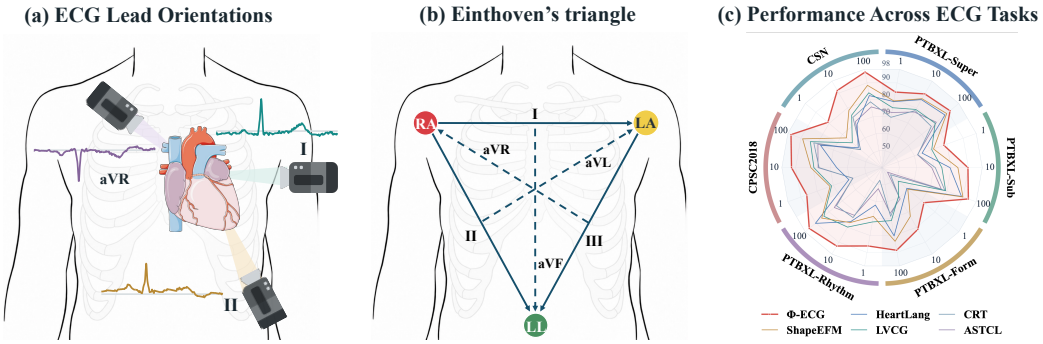

Figure 1: **Physical priors and performance of Φ-ECG. (a)** Each lead observes cardiac electrical activity from a characteristic viewpoint determined by its positioning. **(b)** Electrode configuration and relationships among limb leads. A standard 12-lead ECG uses ten electrodes, one on each limb and six on the chest. Potentials at the right arm (RA), left arm (LA), and left leg (LL) determine the six limb leads, while each chest electrode is referenced to their mean potential to form leads V1–V6. The RA, LA, and LL sites form Einthoven's triangle, and any two independent limb leads (e.g., I and II) can algebraically determine the remaining four. **(c)** AUROC (%) comparison of Φ-ECG and baseline methods on six ECG classification tasks at 1%, 10%, and 100% label availability.

in vectorcardiography (VCG) (Burger & van Milaan, 1948) models each ECG lead as a directional measurement of cardiac electrical activity characterized by a three-dimensional orientation, which can be parameterized in the Frank coordinate system (Frank, 1956; Dower, 1968). Motivated by this physical prior, we introduce a physics-informed lead direction embedding that maps lead directions through real spherical harmonic basis functions. This shared, smooth mapping leads to representations of their effective observation directions, explicitly incorporating acquisition information.

In addition, we exploit the structured dependencies of ECG signals across leads and time. Specifically, limb and precordial leads provide complementary observations of cardiac activity, while limb leads satisfy established algebraic relations, as illustrated in Fig. 1(b) (Clifford et al., 2006; Macfarlane et al., 2010). Meanwhile, the temporal continuity of ECG signals makes missing intervals inferable from their surrounding context (Na et al., 2024). Based on these structural priors, we develop a structured lead–time masking strategy that samples visible leads separately from the limb and precordial groups and applies a shared temporal mask across the retained leads. Keeping only one limb lead avoids exact algebraic reconstruction, while shared temporal masking removes same-time observations across leads. Together, these choices encourage reconstruction through complementary lead views and surrounding temporal context. Our main contributions are summarized as follows:

- **Lead–time position encoding.** We introduce a physics-informed lead direction embedding that maps effective ECG lead directions through real spherical harmonics, providing a shared, smooth parameterization of lead geometry. We combine this embedding with conditional temporal encoding to represent both lead direction and local temporal context.
- **Structured lead–time masking.** We sample limb and precordial leads separately and apply a shared temporal mask across them, limiting reconstruction shortcuts and encouraging the model to integrate complementary lead views with temporal context during pretraining.
- **Comprehensive downstream evaluation.** We evaluate Φ-ECG on ECG disease classification, echocardiographic phenotype prediction, reduced-lead transfer, and multi-lead view reconstruction, including full- and low-label evaluation for classification and phenotype prediction.

2 Related Work

Self-Supervised ECG Representation Learning. Self-supervised ECG methods learn transferable representations through contrastive learning and signal reconstruction. CLOCS (Kiyasseh et al., 2021) constructs positive pairs across time, leads, and patients, while ASTCL (Wang et al., 2023) learns invariances to spatio-temporal transformations. Other approaches explore temporal–contextual contrasting (Eldele et al., 2021) and cross reconstruction (Zhang et al., 2024). More

recent methods incorporate ECG-specific structures into pretraining. ST-MEM (Na et al., 2024) captures spatio-temporal relationships through masked reconstruction, while HeartLang (Jin et al., 2025) models morphology and rhythm at different semantic scales. To accommodate varying lead availability, lead-agnostic pretraining (Oh et al., 2022) exposes models to different lead subsets through random lead masking. ECG-FM (McKeen et al., 2025) further combines random lead masking with masked signal modeling and contrastive learning. Building on these efforts, Φ-ECG uses limb–precordial group sampling and shared temporal masks to limit reconstruction shortcuts and encourage the joint use of complementary lead observations and temporal context.

ECG Lead Structure and Vectorcardiography. The relationship between ECG leads and cardiac electrical activity has long been studied through vectorcardiography (VCG). Classical linear mappings, including the Kors and Dower transformations (Kors et al., 1990; Dower, 1968), relate standard ECG leads to a three-dimensional representation of cardiac electrical activity. Neural approaches also exploit the directional nature of ECG acquisition. Nef-Net (Chen et al., 2021a) uses viewpoint information to synthesize ECGs at queried directions, while ME-GAN (Chen et al., 2022) generates multiple lead views conditioned on cardiac abnormalities. More recently, LVCG (Huang et al., 2026) learns cardiac representations in a latent VCG space. These studies motivate the use of acquisition geometry in ECG modeling. Φ-ECG incorporates this geometry into self-supervised learning through a shared lead-direction embedding based on real spherical harmonics, providing a smooth mapping from effective observation directions to lead representations.

3 Background

Lead Orientation and Algebraic Structure. VCG provides a useful first-order abstraction of multi-lead ECG acquisition (Frank, 1956; Dower, 1968). Let $\mathbf{q}(t) \in \mathbb{R}^3$ denote the instantaneous cardiac electrical vector in the Frank coordinate system. Under a first-order directional measurement model, the signal recorded by lead c is approximated as

$$x_c(t) = \mathbf{a}_c^\top \mathbf{q}(t) + \epsilon_c(t), \quad \mathbf{v}_c = \frac{\mathbf{a}_c}{\|\mathbf{a}_c\|_2} \in \mathbb{S}^2, \tag{1}$$

where $\mathbf{a}_c \in \mathbb{R}^3$ is the effective lead vector that characterizes the sensitivity of lead c to the three Frank axes, $\mathbf{v}_c$ is its normalized orientation, and $\epsilon_c(t)$ captures effects beyond this first-order approximation. This formulation associates each lead with an effective observation direction in a common three-dimensional coordinate system (Macfarlane et al., 2010). The standard limb leads also satisfy the Einthoven and Goldberger relations (Clifford et al., 2006; Macfarlane et al., 2010):

$$x_{\mathrm{III}} = x_{\mathrm{II}} - x_{\mathrm{I}}, \quad x_{\mathrm{aVR}} = -\tfrac{1}{2}(x_{\mathrm{I}} + x_{\mathrm{II}}), \quad x_{\mathrm{aVL}} = x_{\mathrm{I}} - \tfrac{1}{2}x_{\mathrm{II}}, \quad x_{\mathrm{aVF}} = x_{\mathrm{II}} - \tfrac{1}{2}x_{\mathrm{I}}. \tag{2}$$

These relations determine the other four limb leads from leads I and II. More generally, any two distinct standard limb leads suffice to recover the remaining four.

Spherical Harmonics. Real spherical harmonics form an orthonormal basis of $L^2(\mathbb{S}^2)$, providing a basis for representing functions of normalized lead directions on the unit sphere (Atkinson & Han, 2012). Let $Y_{\ell m}^{\mathbb{R}}(\mathbf{v})$ denote the real spherical harmonic of degree ℓ and order m, where $\ell \geq 0$ and $-\ell \leq m \leq \ell$. Truncating the basis at degree L gives the directional feature map

$$\boldsymbol{\phi}_L : \mathbb{S}^2 \to \mathbb{R}^{(L+1)^2}, \quad \boldsymbol{\phi}_L(\mathbf{v}) \left[Y_{\ell m}^{\mathbb{R}}(\mathbf{v})\right]_{0 \leq \ell \leq L,\ -\ell \leq m \leq \ell}. \tag{3}$$

Degree-ℓ subspaces contain $2\ell + 1$ components, with higher degrees resolving finer angular detail.

4 Method

This section showcases Φ-ECG, a physics-informed self-supervised framework for ECG representation learning. As illustrated in Fig. 2, structured lead–time masking selects visible waveform patches, which are embedded and augmented with lead-direction and temporal information before an encoder. A lightweight decoder then reconstructs the masked waveforms.

4.1 Structured Lead–Time Masking

Let $\mathbf{X} \in \mathbb{R}^{C \times T}$ denote a multi-lead ECG recording, where C is the number of leads and T is the number of time points per lead. We divide each lead signal into $N = T/P$ non-overlapping temporal patches, each containing P consecutive samples. Let $\mathbf{x}_{c,n} \in \mathbb{R}^P$ denote the n-th patch of lead c, where $c \in \{1, \ldots, C\}$ and $n \in \{1, \ldots, N\}$. We then construct a structured mask over the patch grid through group-wise lead sampling and a temporal mask shared across the sampled leads.

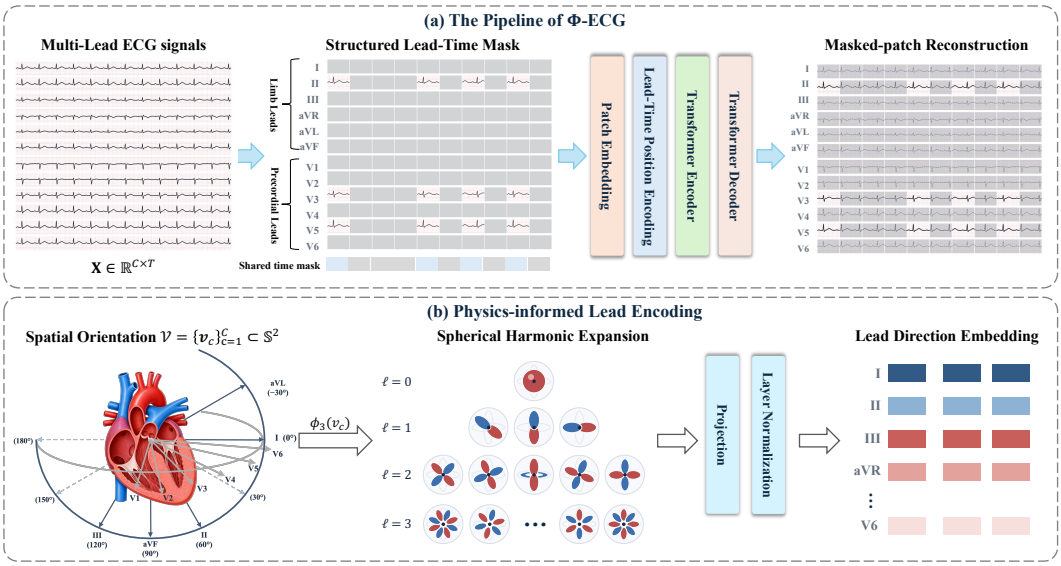

Figure 2: **Overview of Φ-ECG. (a)** Structured lead–time masking retains one limb and two precordial leads with a shared temporal mask, with gray cells indicating masked patches. The visible patches are embedded and augmented with lead-direction and temporal information before encoding. A lightweight decoder reconstructs the masked waveforms. **(b)** Each lead direction is expanded using real spherical harmonics, then projected and normalized into a lead direction embedding. This mapping is shared across leads, and each embedding is reused across its lead's temporal patches.

Group-Wise Lead Sampling. We partition the standard leads into limb and precordial groups:

$$\mathcal{G}_{\text{limb}} = \{\text{I}, \text{II}, \text{III}, \text{aVR}, \text{aVL}, \text{aVF}\}, \quad \mathcal{G}_{\text{pre}} = \{\text{V1}, \text{V2}, \text{V3}, \text{V4}, \text{V5}, \text{V6}\}. \tag{4}$$

The limb leads satisfy relations as shown in Eq. (2), whereas the precordial leads provide complementary views across the chest. For each recording, we uniformly sample $k_{\text{limb}} = 1$ limb lead and $k_{\text{pre}} = 2$ precordial leads, yielding subsets $\mathcal{S}_{\text{limb}}$ and $\mathcal{S}_{\text{pre}}$. Let $\mathcal{S} = \mathcal{S}_{\text{limb}} \cup \mathcal{S}_{\text{pre}}$ be the selected lead set, and define $m_c^{\text{lead}} \in \{0, 1\}$ to be 1 if lead c belongs to $\mathcal{S}$ and 0 otherwise. We retain only one limb lead to prevent exact recovery of the remaining limb leads solely through the Einthoven and Goldberger relations, and two precordial leads to preserve complementary chest views. We randomly resample these subsets for each recording to expose the model to different lead configurations.

Shared Temporal Masking. We construct a temporal mask $\mathbf{m}^{\text{time}} = [m_n^{\text{time}}]_{n=1}^{N} \in \{0, 1\}^N$ by uniformly selecting a proportion r_{time} of patch positions to mask. Masked positions are assigned 0 and retained positions 1. The same temporal mask is shared across all selected leads. Combining it with the lead-selection mask gives

$$m_{c,n} = m_c^{\text{lead}} m_n^{\text{time}}, \quad \mathbf{M} = [m_{c,n}] \in \{0, 1\}^{C\times N}. \tag{5}$$

Here, $m_{c,n} = 1$ indicates a visible patch, while $m_{c,n} = 0$ indicates a masked patch to be reconstructed. The corresponding sets of visible and masked patch indices are defined as

$$\Omega = \{(c, n) \mid m_{c,n} = 1\}, \quad \mathcal{M} = \{(c, n) \mid m_{c,n} = 0\}. \tag{6}$$

An unselected lead is masked at every temporal position, while a masked temporal position is hidden across all leads. This removes same-time observations from other leads at temporally masked positions, requiring reconstruction of these patches from surrounding temporal context.

4.2 Lead–Time Position Encoding

Patch Embedding. Given the visible set Ω, the shared patch embedding module f_{patch} independently maps each retained waveform patch to a D-dimensional token:

$$\mathbf{z}_{c,n} = f_{\text{patch}}(\mathbf{x}_{c,n}) \in \mathbb{R}^D, \quad (c, n) \in \Omega. \tag{7}$$

The module uses one-dimensional convolutions along the P waveform samples, and its architecture is provided in Table 11. The resulting tokens represent local ECG morphology and are subsequently augmented with lead-direction and temporal information.

Lead Direction Embedding. We associate each ECG lead with an effective direction $\mathbf{v}_c \in \mathbb{S}^2$, as defined in Eq. (1). For leads I, II, and V1–V6, we use the corresponding rows of the canonical forward Dower matrix. Meanwhile, the remaining limb-lead vectors are derived from the Einthoven and Goldberger relations. All vectors are then normalized onto the unit sphere. As illustrated in Fig. 2(b), we expand each direction using real spherical harmonics through degree three and project the resulting 16-dimensional features into the model space:

$$\mathbf{e}_c^{\mathrm{lead}} = \mathrm{LayerNorm}(\mathbf{W}_{\mathrm{sh}}\boldsymbol{\phi}_3(\mathbf{v}_c)) \in \mathbb{R}^D, \quad \mathbf{W}_{\mathrm{sh}} \in \mathbb{R}^{D\times 16}. \tag{8}$$

Here, $\mathbf{W}_{\mathrm{sh}}$ learns task-relevant combinations of the angular modes. The mapping is shared across leads, while each lead embedding is reused across its temporal patches. This provides a smooth parameterization of lead representations by their observation directions.

In practice, electrodes may not be placed exactly at their standard locations. To account for the uncertainty in lead directions, we add independent Gaussian noise with standard deviation 0.1 to each normalized direction during pretraining and renormalize it to obtain $\tilde{\mathbf{v}}_c$. This encourages the shared mapping to tolerate directional deviations. Downstream evaluation uses the canonical directions without perturbation. This exposes the shared mapping to local neighborhoods of the directions.

Conditional Temporal Encoding. We incorporate local temporal context using a depthwise one-dimensional convolution g_{time} spanning five consecutive patch positions along each lead (Chu et al., 2023). For this convolution, we form an auxiliary sequence $\bar{\mathbf{z}}_{c,:}$ by retaining the visible patch tokens and filling masked positions with a learned placeholder. This sequence preserves the original temporal positions, and the placeholders are excluded from the encoder input. For each visible position, the patch, lead, and temporal components are combined as

$$\mathbf{h}_{c,n}^0 = \mathrm{LN}\big(\mathbf{z}_{c,n} + \mathbf{e}_c^{\mathrm{lead}} + g_{\mathrm{time}}(\bar{\mathbf{z}}_{c,:})_n\big), \quad (c,n) \in \Omega. \tag{9}$$

The resulting visible token $\mathbf{h}_{c,n}^0$ combines local ECG morphology, the lead observation direction, and neighboring temporal context, forming the input representation for the encoder.

4.3 Masked Reconstruction and Downstream Evaluation

Masked Waveform Reconstruction. Let E_Θ denote the Transformer encoder, which processes only the visible tokens indexed by Ω:

$$\mathbf{H}_\Omega = E_\Theta\big([\mathbf{h}_{c,s}^0]_{(c,s)\in\Omega}\big) = [\mathbf{h}_{c,s}^{\mathrm{enc}}]_{(c,s)\in\Omega}. \tag{10}$$

We adopt an asymmetric masked-autoencoder architecture (He et al., 2022). The encoder outputs are projected to the decoder dimension and combined with learned decoder mask tokens to restore the masked tokens. After adding positional embeddings, a lightweight Transformer decoder predicts a waveform patch $\hat{\mathbf{x}}_{c,s} \in \mathbb{R}^P$. We optimize the mean absolute error (MAE) over the masked positions:

$$\mathcal{L}_{\mathrm{rec}} = \frac{1}{|\mathcal{M}|P} \sum_{(c,s)\in\mathcal{M}} \|\hat{\mathbf{x}}_{c,s} - \mathbf{x}_{c,s}\|_1 . \tag{11}$$

The loss covers both the withheld leads and the temporally masked patches of the selected leads.

Downstream Linear Probing. For downstream tasks, we mean pooling the encoder outputs over all patches to obtain an ECG representation. We freeze the pretrained backbone and train only a task-specific linear classifier, evaluating the transferability of the learned representations without updating the encoder. Baseline-specific input and feature-extraction paths are described in Sec. F.

5 Experiments

5.1 Experimental Setup

Datasets and Preprocessing. We pretrain Φ-ECG on approximately 3.6 million unlabeled ECGs from MIMIC-IV-ECG (Gow et al., 2023) and CODE (Ribeiro et al., 2020). Downstream evaluation covers six disease-classification tasks: PTB-XL superclass, subclass, form, and rhythm (Wagner et al., 2020; Strodthoff et al., 2021), CPSC2018 (Liu et al., 2018), and Chapman–Shaoxing–Ningbo (CSN) (Zheng et al., 2020; 2022). We further evaluate on EchoNext-Mini (Hughes et al., 2026; Elias & Finer, 2026), which pairs ECG recordings with labels derived from echocardiography. We predict 12 echocardiographic endpoints from ECG waveforms to assess whether the learned representations capture information about cardiac structure and function. Missing-lead reconstruction is

Table 1: Linear-probing AUROC (%) across six ECG classification tasks. Best and second-best results are shown in **bold** and gray. The best result among comparison methods is underlined.

Method	PTBXL-Super			PTBXL-Sub			PTBXL-Form			PTBXL-Rhythm			CPSC2018			CSN		
	1%	10%	100%	1%	10%	100%	1%	10%	100%	1%	10%	100%	1%	10%	100%	1%	10%	100%
SimCLR (Chen et al., 2020)	63.41	69.77	73.53	60.84	68.27	73.39	54.98	56.97	62.52	51.41	69.44	77.73	59.78	68.52	76.54	59.02	67.26	73.20
BYOL (Grill et al., 2020)	71.70	73.83	76.45	57.16	67.44	71.64	48.73	61.63	70.82	41.99	74.40	77.17	60.88	74.42	78.75	54.20	71.92	74.69
BarlowTwins (Zbontar et al., 2021)	72.87	75.96	78.41	62.57	70.84	74.34	52.12	60.39	66.14	50.12	73.54	77.62	55.12	72.75	78.39	60.72	71.64	77.43
MoCo-v3 (Chen et al., 2021b)	73.19	76.65	78.26	55.88	69.21	76.69	50.32	63.71	71.31	51.38	71.66	74.33	62.13	76.74	75.29	54.61	74.26	77.68
SimSiam (Chen & He, 2021)	73.15	72.70	75.63	62.52	69.31	76.38	55.16	62.91	71.31	49.30	69.47	75.92	58.35	72.89	75.31	58.25	68.61	77.41
TS-TCC (Eldele et al., 2021)	70.73	75.88	78.91	53.54	66.98	77.87	48.04	61.79	71.18	43.34	69.48	78.23	57.07	73.62	78.72	55.26	68.48	76.79
CLOCS (Kiyasseh et al., 2021)	68.94	73.36	76.31	57.94	72.55	76.24	51.97	57.96	72.65	47.19	71.88	76.31	59.59	77.78	77.49	54.38	71.93	76.13
ASTCL (Wang et al., 2023)	72.51	77.31	81.02	61.86	68.77	76.51	44.14	60.93	66.99	52.38	71.98	76.05	57.90	77.01	79.51	56.40	70.87	75.79
CRT (Zhang et al., 2024)	69.68	78.24	77.24	61.98	70.82	78.67	46.41	59.49	68.73	47.44	73.52	74.41	58.01	76.43	82.03	56.21	73.70	78.80
ST-MEM (Na et al., 2024)	61.12	66.87	71.36	54.12	57.86	63.59	55.71	59.99	66.07	51.12	65.44	74.85	56.69	63.32	70.39	59.77	66.87	71.36
HeartLang (Jin et al., 2025)	78.94	85.59	87.52	64.68	79.34	88.91	58.70	63.99	80.23	62.08	76.22	90.34	60.44	66.26	77.87	57.94	68.93	82.49
LVCG (Huang et al., 2026)	75.33	79.03	80.13	70.61	74.62	79.19	52.28	59.12	71.24	72.03	79.87	83.94	71.09	79.44	84.15	62.47	75.17	84.14
ShapeEFM (Liu et al., 2026)	79.32	86.23	88.86	71.70	83.91	89.79	59.45	72.72	83.44	68.75	76.82	85.59	72.03	81.19	88.90	66.88	78.38	88.52
Φ-ECG (MIMIC)	83.98	87.03	89.75	**78.99**	88.43	92.08	63.63	73.97	86.31	83.35	89.00	94.28	82.94	90.65	94.82	77.81	89.56	95.12
Φ-ECG	**84.56**	**89.45**	**91.38**	78.05	**89.72**	**93.17**	**71.66**	**81.39**	**88.73**	**86.20**	**92.57**	**94.98**	**86.34**	**92.39**	**96.06**	**80.93**	**91.77**	**96.37**

evaluated on CPSC2018 and PTB Diagnostic (Bousseljot et al., 1995). For Φ-ECG, recordings are standardized to 10-second, 12-lead waveforms at 500 Hz through signal repair, resampling, filtering, and amplitude clipping. Derived limb leads in CODE follow Eq. (2). Dataset partitions and preprocessing details are provided in Sec. E and Tab. 13.

Baselines and Pretraining Data. We compare against general self-supervised methods, including SimCLR (Chen et al., 2020), BYOL (Grill et al., 2020), Barlow Twins (Zbontar et al., 2021), MoCo-v3 (Chen et al., 2021b), SimSiam (Chen & He, 2021), TS-TCC (Eldele et al., 2021), and ECG methods, including CLOCS (Kiyasseh et al., 2021), ASTCL (Wang et al., 2023), CRT (Zhang et al., 2024), ST-MEM (Na et al., 2024), HeartLang (Jin et al., 2025), LVCG (Huang et al., 2026), and ShapeEFM (Liu et al., 2026). The baseline results adopted from the benchmark reported by (Liu et al., 2024; Jin et al., 2025) use MIMIC-IV-ECG pretraining. We report both a MIMIC-only variant, denoted Φ-ECG (MIMIC), and the default model pretrained on MIMIC-IV-ECG and CODE.

Model and Optimization. We pretrain Φ-ECG on eight NVIDIA A800 GPUs. The model uses a 12-layer Transformer encoder with width 512 and eight attention heads, together with a four-layer decoder. The pretraining model contains 42.6M parameters, including the decoder. Each token represents a one-second waveform patch. Structured masking retains one limb lead, two precordial leads, and 60% of temporal positions, resulting in 18 visible tokens per recording. Both the MIMIC+CODE and MIMIC-only variants are pretrained for 200 epochs using AdamW with cosine learning-rate decay. Architecture and optimization settings are provided in Tables 8 and 9.

Evaluation Protocol. The classification and phenotype prediction use frozen-encoder linear probing with 1%, 10%, and 100% of the training labels. A linear multi-label classifier is optimized using binary cross-entropy, and the checkpoint with the highest validation AUROC is evaluated on the test set. For EchoNext-Mini and reduced-lead evaluation, we retain each model's native inference pipeline and tune the linear-probing hyperparameters for each model starting from its original repository configuration. Both evaluations use shared downstream splits and five probe seeds. Implementation and hyperparameter-tuning details are provided in Sec. F.

Evaluation Metrics. We report AUROC for disease classification and macro-AUROC, macro-AUPRC, and macro-balanced accuracy (BACC) for EchoNext-Mini. BACC uses a probability threshold of 0.5. These metrics are reported as percentages, with differences expressed in percentage points. Following the evaluation protocol of LVCG (Huang et al., 2026), we report mean squared error (MSE) and mean absolute error (MAE) over the missing leads.

5.2 Main Results

Disease Classification. As shown in Table 1, both Φ-ECG variants outperform the baselines across all six classification tasks and three label fractions. The gains are particularly pronounced with limited annotations: with 1% labels, the default model exceeds the strongest baseline by 14.31 AUROC points on CPSC2018 and 14.05 points on CSN. The consistently strong performance of the MIMIC-only variant indicates that these gains do not depend solely on additional CODE data.

Table 2: EchoNext linear-probing performance (mean ± std, %) over five seeds. Best and second-best results are shown in **bold** and gray. The best result among comparison methods is underlined.

Method	1%			10%			100%		
	AUROC	AUPRC	BACC	AUROC	AUPRC	BACC	AUROC	AUPRC	BACC
ST-MEM (Na et al., 2024)	56.10 ± 1.85	15.74 ± 0.22	55.97 ± 0.48	69.92 ± 0.78	23.64 ± 0.12	64.95 ± 0.56	72.61 ± 0.43	24.30 ± 0.07	66.68 ± 0.30
HeartLang (Jin et al., 2025)	66.96 ± 1.78	21.05 ± 0.20	62.87 ± 1.32	73.58 ± 0.93	24.00 ± 0.22	67.23 ± 1.11	75.48 ± 0.31	26.05 ± 0.10	68.69 ± 0.39
LVCG (Huang et al., 2026)	69.51 ± 1.06	21.57 ± 0.24	64.42 ± 0.98	74.99 ± 0.79	25.37 ± 0.21	68.12 ± 0.67	75.49 ± 0.36	25.89 ± 0.60	68.30 ± 0.54
ShapeEFM (Liu et al., 2026)	69.02 ± 1.05	22.14 ± 0.46	63.60 ± 1.06	74.71 ± 0.49	25.49 ± 0.24	67.79 ± 0.88	77.37 ± 0.08	27.46 ± 0.07	70.18 ± 0.30
Φ-ECG (MIMIC)	72.57 ± 0.85	24.82 ± 0.60	66.15 ± 0.97	78.30 ± 0.44	29.91 ± 0.31	70.61 ± 0.09	80.72 ± 0.09	32.45 ± 0.10	72.44 ± 0.77
Φ-ECG	**73.60 ± 1.50**	**25.57 ± 0.87**	**66.41 ± 1.07**	**79.02 ± 0.47**	**30.94 ± 0.36**	**71.71 ± 0.59**	**81.17 ± 0.17**	**32.61 ± 0.11**	**73.87 ± 0.40**

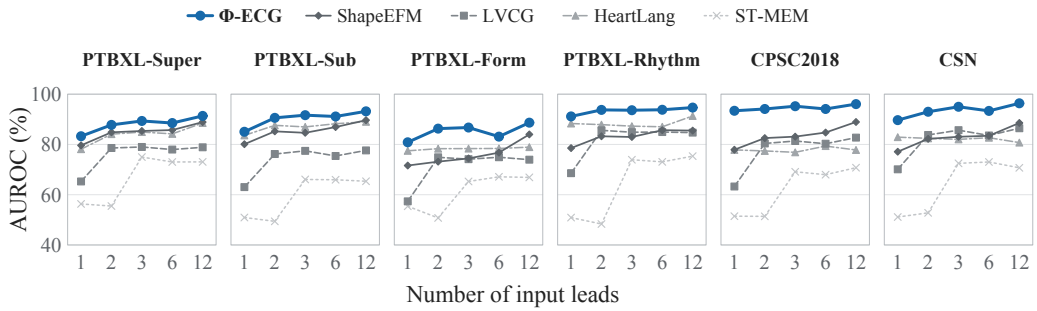

Figure 3: Classification with fewer input leads. Points show mean AUROC over five seeds.

Pretraining on MIMIC+CODE generally further improves performance, although the MIMIC-only variant achieves the highest AUROC on PTBXL-Sub with 1% labels.

Echocardiographic Phenotype Prediction. Φ-ECG achieves the highest macro-AUROC, macro-AUPRC, and macro-BACC at all three label fractions on EchoNext-Mini (Table 2). With full labels, it improves over ShapeEFM by 3.80 AUROC points and 5.15 AUPRC points. With only 10% labels, its AUROC of 79.02% exceeds the strongest baseline using full labels, highlighting effective transfer to echocardiographic phenotype prediction under limited supervision. We further provide per-label results and statistical comparisons in Tables 17 and 18 of Sec. F.

Classification with Fewer Leads. Following HeartLang (Jin et al., 2025), we evaluate five lead configurations with $K \in \{1, 2, 3, 6, 12\}$ using full training labels and five seeds. These configurations correspond to I, I/II, I/II/V2, all six limb leads, and all 12 standard leads, respectively. Further evaluation details are provided in Sec. F. Φ-ECG achieves the highest mean AUROC across all 30 tasks, outperforming the strongest baseline by 15 points on CPSC2018 with lead I alone. Performance also depends on lead composition: I/II/V2 outperforms all six limb leads on PTBXL-Form (86.72% vs. 83.11%) and CSN (94.98% vs. 93.34%). These results suggest that complementary observations across lead groups can support better classification with fewer channels, consistent with the motivation for limb–precordial sampling during pretraining.

View Reconstruction. Following LVCG (Huang et al., 2026), we evaluate missing-lead reconstruction on CPSC2018 and PTB from three or five observed leads. MSE and MAE are computed over the remaining leads. As shown in Table 3, Φ-ECG achieves the lowest MSE and MAE across both datasets and lead settings. With three observed leads, it reduces MAE relative to LVCG from 0.37 to 0.27 on CPSC2018 and from 0.36 to 0.24 on PTB. The MIMIC-only variant also achieves lower MAE in all four settings, indicating that the gains do not rely solely on additional CODE data. These improvements may reflect greater reconstruction flexibility. While LVCG reconstructs ECGs through a fixed linear projection from three-dimensional VCG space, Φ-ECG incorporates lead geometry into its representations and reconstructs ECG waveforms, potentially preserving details.

5.3 Ablation Studies

We examine structured lead–time masking, position encoding, lead-direction assignment, and spherical-harmonic degree. Within each ablation family, all variants share the pretraining data, backbone, 200-epoch training schedule, and downstream probing protocol. We average results over

Table 3: Multi-lead reconstruction on CPSC2018 and PTB.

Method	CPSC2018				PTB			
	$3 \rightarrow 12$		$5 \rightarrow 12$		$3 \rightarrow 12$		$5 \rightarrow 12$	
	MSE	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**
KIM (Kors et al., 1990)	0.85	0.52	1.00	0.47	0.74	0.43	1.02	0.75
VAE-CNN (Matyschik et al., 2020)	0.80	0.49	1.02	0.46	1.02	0.56	0.99	0.70
E-LSTM (Sohn et al., 2020)	0.72	0.48	1.02	0.46	0.54	0.38	0.94	0.71
Nef-Net (Chen et al., 2021a)	0.68	0.40	0.44	0.35	1.56	0.52	1.00	0.58
LVCG (Huang et al., 2026)	0.49	0.37	0.40	0.35	0.47	0.36	0.78	0.34
Φ-ECG (MIMIC)	0.50	0.30	0.40	0.25	0.35	0.25	0.25	0.21
Φ-ECG	**0.48**	**0.27**	**0.39**	**0.24**	**0.31**	**0.24**	**0.23**	**0.20**

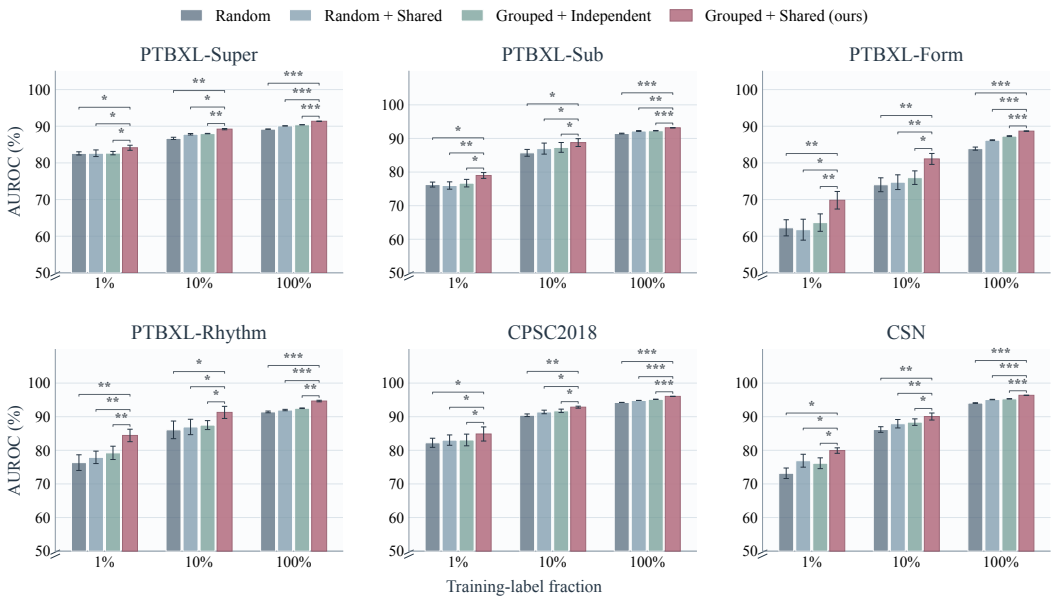

Figure 4: Ablation of masking strategies with bars and error bars representing mean AUROC and std, respectively. Holm-adjusted significance is indicated by $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$.

five probe seeds. The masking experiment covers all three label fractions, while the tables report results with 1% labels. Complete results are provided in Tables 19 to 21.

Structured Lead–Time Masking. We compare random token masking with random lead selection and shared temporal masks, grouped selection and independent temporal masks, and grouped selection and shared temporal masks. Grouped selection retains one limb lead and two precordial leads, and all strategies retain 18 of 120 tokens. As shown in Fig. 4, grouped selection with shared temporal masks achieves the highest AUROC across all 18 settings. With 1% labels on PTBXL-Rhythm, grouping improves AUROC by 6.51 points under shared masks, while sharing improves AUROC by 5.18 points under grouped selection. These gains are consistent with our hypothesis that redundant same-time observations may enable reconstruction shortcuts. Restricting these shortcuts encourages the model to exploit complementary lead observations and temporal context.

Lead–Time Position Encoding. As shown in Table 4, removing temporal or lead-direction encoding reduces average AUROC by 8.49 or 5.46 points, respectively. We further compare learnable lead embeddings with a parameter-matched coordinate multilayer perceptron (MLP), consisting of two linear layers with an intervening GELU activation and a final LayerNorm. The MLP maps the same normalized Dower directions to D-dimensional embeddings. Learnable embeddings, the coordinate MLP, and spherical harmonics achieve average AUROCs of 76.04%, 78.41%, and 80.36%, respectively. The MLP improvement supports incorporating observation geometry. The additional

Table 4: Position-encoding ablation (AUROC, mean ± std, %) with 1% labels.

Variant	PTBXL-Super	PTBXL-Sub	PTBXL-Form	PTBXL-Rhythm	CPSC2018	CSN	Avg.
w/o temporal encoding	82.93 ± 0.75	72.46 ± 1.55	58.29 ± 2.28	68.27 ± 3.37	81.82 ± 0.81	67.42 ± 0.88	71.87
w/o lead encoding	82.00 ± 0.68	72.55 ± 0.92	62.73 ± 1.11	79.21 ± 2.48	80.91 ± 2.85	72.00 ± 1.12	74.90
w/o both	80.55 ± 0.19	69.73 ± 1.79	54.84 ± 2.85	65.63 ± 2.74	78.69 ± 0.59	65.72 ± 1.44	69.19
Learnable lead embedding	83.27 ± 0.35	74.66 ± 1.88	62.67 ± 1.56	76.45 ± 3.35	82.43 ± 1.21	76.74 ± 1.09	76.04
Dower-coordinate MLP	84.06 ± 0.35	76.03 ± 1.24	66.99 ± 1.67	78.85 ± 4.05	**85.69 ± 1.55**	78.84 ± 0.83	78.41
Φ-ECG (full)	**84.13 ± 0.69**	**78.98 ± 0.86**	**69.81 ± 2.40**	**84.42 ± 1.85**	84.86 ± 2.10	**79.92 ± 0.82**	**80.36**

1.95-point gain from spherical harmonics under a matched parameter budget suggests that predefined angular modes provide a useful inductive bias for learning directional representations.

Table 5: Lead-direction assignment and noise ablation (AUROC, mean ± std, %) with 1% labels.

Variant	PTBXL-Super	PTBXL-Sub	PTBXL-Form	PTBXL-Rhythm	CPSC2018	CSN	Avg.
Shuffled Dower	82.92 ± 0.48	76.08 ± 1.02	67.33 ± 1.45	84.19 ± 2.17	84.16 ± 2.76	77.43 ± 0.72	78.68
Random sphere	82.36 ± 0.74	74.63 ± 0.92	66.52 ± 1.66	80.29 ± 2.91	82.63 ± 2.18	77.20 ± 0.70	77.27
Dower w/o noise	83.41 ± 0.52	76.34 ± 1.17	67.46 ± 3.33	**84.47 ± 1.52**	84.63 ± 2.23	77.28 ± 0.93	78.93
Φ-ECG (Canonical Dower)	**84.13 ± 0.69**	**78.98 ± 0.86**	**69.81 ± 2.40**	84.42 ± 1.85	**84.86 ± 2.10**	**79.92 ± 0.82**	**80.36**

Lead–Direction Correspondence. We compare canonical, shuffled, and random directions using the same spherical-harmonic mapping and perturbation level $\sigma_v = 0.1$. As shown in Table 5, canonical directions outperform shuffled and random assignments by 1.68 and 3.09 AUROC points on average. Shuffling preserves the direction set but breaks its correspondence with the waveforms, supporting the value of physically meaningful assignments beyond distinct lead identifiers. Direction perturbations further improve average AUROC from 78.93% to 80.36%, with gains on five of six tasks and nearly unchanged performance on PTBXL-Rhythm. These results suggest that preserving canonical directions while allowing small directional deviations benefits representation learning.

Table 6: Spherical-harmonic degree ablation (AUROC, mean ± std, %) with 1% labels.

Maximum degree	PTBXL-Super	PTBXL-Sub	PTBXL-Form	PTBXL-Rhythm	CPSC2018	CSN	Avg.
$L = 0$ (1 basis)	81.62 ± 0.78	76.25 ± 0.41	63.62 ± 1.82	80.54 ± 1.83	83.65 ± 1.44	71.41 ± 0.95	76.18
$L = 1$ (4 bases)	82.59 ± 0.65	76.47 ± 0.72	66.20 ± 3.38	82.44 ± 1.91	84.39 ± 1.69	71.74 ± 1.05	77.31
$L = 2$ (9 bases)	83.48 ± 0.76	78.68 ± 0.74	67.92 ± 1.95	83.77 ± 1.80	83.96 ± 2.34	74.14 ± 0.61	78.66
$L = 3$ (16 bases; full)	**84.13 ± 0.69**	**78.98 ± 0.86**	**69.81 ± 2.40**	**84.42 ± 1.85**	**84.86 ± 2.10**	**79.92 ± 0.82**	**80.36**

Spherical-Harmonic Degree. We investigate the effect of the maximum spherical-harmonic degree $L \in \{0, 1, 2, 3\}$ on downstream tasks. Degree zero supplies a constant feature, while degree one introduces features proportional to the Cartesian coordinates. Higher degrees provide finer angular modes and greater flexibility in representing lead directions. As shown in Table 6, average AUROC increases from 76.18% at $L = 0$ to 77.31, 78.66%, and 80.36% at $L = 1, 2, 3$, respectively. The gains beyond $L = 1$ suggest that a richer angular basis provides a more effective parameterization for learning lead embeddings than first-degree coordinate features alone. Overall, $L = 3$ achieves the highest average performance within the tested range, although the gains vary across tasks.

6 Conclusion

We presented Φ-ECG, a physics-informed self-supervised framework that incorporates lead geometry and lead–time structure into ECG representation learning. Spherical-harmonic encoding represents leads through their observation directions, while structured masking encourages reconstruction from complementary lead observations and temporal context. Experiments demonstrated improved disease classification and echocardiographic phenotype prediction across label budgets, effective transfer to reduced-lead inputs, and missing-lead reconstruction without additional training. Ablation studies further supported the contributions of physically meaningful direction assignments, higher-degree angular features, and structured masking. These findings highlight the value of ECG acquisition priors for learning generalizable representations from unlabeled recordings.

AI USE STATEMENT

Generative AI tools are used for manuscript drafting and language editing. The authors retain full responsibility for the final manuscript, including its claims, references, and experimental results.

ETHICS STATEMENT

This work studies ECG representation learning using existing research datasets. Access to and reuse of these datasets remain subject to their respective licenses and data-use agreements. Any release associated with this work will respect restrictions on patient-level data redistribution.

REPRODUCIBILITY STATEMENT

We provide architecture and optimization settings in Sec. C, lead geometry and spherical-harmonic conventions in Sec. D, and dataset partitions and preprocessing procedures in Sec. E. Checkpoint sources, baseline configurations, evaluation protocols, and additional results are documented in Sec. F. Ablation protocols and full results across label fractions are provided in Sec. G. The code and models will be released upon acceptance.

REFERENCES

Kendall Atkinson and Weimin Han. *Spherical Harmonics and Approximations on the Unit Sphere: An Introduction*, volume 2044 of *Lecture Notes in Mathematics*. Springer Berlin, Heidelberg, 2012.

Zachi I. Attia, Suraj Kapa, Francisco Lopez-Jimenez, Paul M. McKie, Dorothy J. Ladewig, Gaurav Satam, Patricia A. Pellikka, Maurice Enriquez-Sarano, Peter A. Noseworthy, Thomas M. Munger, Samuel J. Asirvatham, Christopher G. Scott, Rickey E. Carter, and Paul A. Friedman. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nature Medicine*, 25(1):70–74, 2019.

Ralf Bousseljot, Dieter Kreiseler, and Allard Schnabel. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. *Biomedizinische Technik/Biomedical Engineering*, 40(s1): 317–318, 1995.

H. C. Burger and J. B. van Milaan. Heart-vector and leads. Part III: Geometrical representation. *British Heart Journal*, 10(4):229–233, 1948.

Jintai Chen, Xiangshang Zheng, Hongyun Yu, Danny Z Chen, and Jian Wu. Electrocardio panorama: synthesizing new ECG views with self-supervision. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 3597–3605, 2021a.

Jintai Chen, Kuanlun Liao, Kun Wei, Haochao Ying, Danny Z Chen, and Jian Wu. ME-GAN: Learning panoptic electrocardio representations for multi-view ECG synthesis conditioned on heart diseases. In *Proceedings of the International Conference on Machine Learning*, volume 162, pp. 3360–3370. PMLR, 2022.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Proceedings of the International Conference on Machine Learning*, volume 119, pp. 1597–1607. PMLR, 2020.

Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15750–15758, 2021.

Xinlei Chen, Saining Xie, and Kaiming He. An empirical study of training self-supervised vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9640–9649, 2021b.

Xiangxiang Chu, Zhi Tian, Bo Zhang, Xinlong Wang, and Chunhua Shen. Conditional positional encodings for vision transformers. In *Proceedings of the International Conference on Learning Representations*, 2023.

Gari D Clifford, Francisco Azuaje, Patrick McSharry, et al. *Advanced Methods and Tools for ECG Data Analysis*, volume 10. Artech House, 2006.

Gordon E Dower. A lead synthesizer for the Frank system to simulate the standard 12-lead electrocardiogram. *Journal of Electrocardiology*, 1(1):101–116, 1968.

Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and Cuntai Guan. Time-series representation learning via temporal and contextual contrasting. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 2352–2359, 2021.

Pierre Elias and Joshua Finer. EchoNext: A dataset for detecting echocardiogram-confirmed structural heart disease from ECGs. *PhysioNet*, 2026. Version 1.1.1.

Ernest Frank. An accurate, clinically practical system for spatial vectorcardiography. *Circulation*, 13(5):737–749, 1956.

Brian Gow, Tom Pollard, Larry A. Nathanson, Alistair Johnson, Benjamin Moody, Chrystinne Fernandes, Nathaniel Greenbaum, Jonathan W. Waks, Parastou Eslami, Tanner Carbonati, Ashish Chaudhari, Elizabeth Herbst, Dana Moukheiber, Seth Berkowitz, Roger Mark, and Steven Horng. MIMIC-IV-ECG: Diagnostic electrocardiogram matched subset. *PhysioNet*, 2023. Version 1.0.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent: A new approach to self-supervised learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 21271–21284, 2020.

Xiao Gu, Wei Tang, Jinpei Han, Veer Sangha, Fenglin Liu, Shreyank N Gowda, Antonio H Ribeiro, Patrick Schwab, Kim Branson, Lei Clifton, et al. Cardiac health assessment across scenarios and devices using a multimodal foundation model pretrained on data from 1.7 million individuals. *Nature Machine Intelligence*, 8(2):220–233, 2026.

Awni Y. Hannun, Pranav Rajpurkar, Masoumeh Haghpanahi, Geoffrey H. Tison, Codie Bourn, Mintu P. Turakhia, and Andrew Y. Ng. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, 25(1): 65–69, 2019.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16000–16009, 2022.

Bosong Huang, Panzhen Zhao, Zengxiang Li, Patricia Lee, Wei Jin, Alan Wee-Chung Liew, Ming Jin, and Shirui Pan. Learning cardiac latent representations in vectorcardiogram space. In *Proceedings of the International Conference on Machine Learning*, 2026.

John Weston Hughes, Linyuan Jing, Joshua Finer, Dustin Hartzel, Christopher Kelsey, Aaron Long, Daniel Rocha, Jeffrey Ruhl, Timothy Poterucha, and Pierre Elias. EchoNext-mini: A dataset and baseline AI model for detecting structural heart disease from electrocardiograms. *NEJM AI*, 3(5), 2026.

Jiarui Jin, Haoyu Wang, Hongyan Li, Jun Li, Jiahui Pan, and Shenda Hong. Reading your heart: Learning ECG words and sentences via pre-training ECG language model. In *Proceedings of the International Conference on Learning Representations*, 2025.

Dani Kiyasseh, Tingting Zhu, and David A Clifton. CLOCS: Contrastive learning of cardiac signals across space, time, and patients. In *Proceedings of the International Conference on Machine Learning*, volume 139, pp. 5606–5615. PMLR, 2021.

JA Kors, G Van Herpen, AC Sittig, and JH Van Bemmel. Reconstruction of the Frank vectorcardiogram from standard electrocardiographic leads: diagnostic comparison of different methods. *European Heart Journal*, 11(12):1083–1092, 1990.

Siying Lin, Zhaoqi Li, Qifan Wu, Yifeng Chen, Yuedong Yang, and Huiying Zhao. A self-supervised electrocardiogram foundation model for empowering cardiovascular disease prediction and genetic factor discovery. *Nature Communications*, 2026.

Che Liu, Zhongwei Wan, Cheng Ouyang, Anand Shah, Wenjia Bai, and Rossella Arcucci. Zero-shot ECG classification with multimodal learning and test-time clinical knowledge enhancement. In *Proceedings of the International Conference on Machine Learning*, volume 235, pp. 31949–31963. PMLR, 2024.

Feifei Liu, Chengyu Liu, Lina Zhao, Xiangyu Zhang, Xiaoling Wu, Xiaoyan Xu, Yulin Liu, Caiyun Ma, Shoushui Wei, Zhiqiang He, et al. An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. *Journal of Medical Imaging and Health Informatics*, 8(7):1368–1373, 2018.

Huan Liu, Junyi Wang, Jiaqi Fan, et al. ShapeEFM: A shape-based pretrained foundation model on ECG morphology comprehension. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2026.

Peter W Macfarlane, Adriaan Van Oosterom, Olle Pahlm, Paul Kligfield, Michiel Janse, and John Camm. *Comprehensive Electrocardiology*. Springer Science & Business Media, 2010.

Maksymilian Matyschik, Henry Mauranen, Pietro Bonizzi, and Joël Karel. Feasibility of ECG reconstruction from minimal lead sets using convolutional neural networks. In *Proceedings of Computing in Cardiology*, pp. 1–4. IEEE, 2020.

Kaden McKeen, Laura Oliva, Sameer Masood, Augustin Toma, Barry Rubin, and Bo Wang. ECG-FM: An open electrocardiogram foundation model. *JAMIA Open*, 8(5):ooaf122, 2025.

Yeongyeon Na, Minje Park, Yunwon Tae, and Sunghoon Joo. Guiding masked representation learning to capture spatio-temporal relationship of electrocardiogram. In *Proceedings of the International Conference on Learning Representations*, 2024.

Jungwoo Oh, Hyunseung Chung, Joon-myoung Kwon, Dong-gyun Hong, and Edward Choi. Lead-agnostic self-supervised learning for local and global representations of electrocardiogram. In *Proceedings of the Conference on Health, Inference, and Learning*, volume 174, pp. 338–353. PMLR, 2022.

Timothy J Poterucha, Linyuan Jing, Ramon Pimentel Ricart, Michael Adjei-Mosi, Joshua Finer, Dustin Hartzel, Christopher Kelsey, Aaron Long, Daniel Rocha, Jeffrey A Ruhl, et al. Detecting structural heart disease from electrocardiograms using AI. *Nature*, 644(8075):221–230, 2025.

Antonio H. Ribeiro, Manoel Horta Ribeiro, Gabriela M. M. Paixao, Derick M. Oliveira, Paulo R. Gomes, Jessica A. Canazart, Milton P. S. Ferreira, Carl R. Andersson, Peter W. Macfarlane, Wagner Meira Jr., Thomas B. Schon, and Antonio Luiz P. Ribeiro. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nature Communications*, 11:1760, 2020.

Jangjay Sohn, Seungman Yang, Joonnyong Lee, Yunseo Ku, and Hee Chan Kim. Reconstruction of 12-lead electrocardiogram from a three-lead patch-type device using a LSTM network. *Sensors*, 20(11):3278, 2020.

Nils Strodthoff, Patrick Wagner, Tobias Schaeffter, and Wojciech Samek. Deep learning for ECG analysis: Benchmarks and insights from PTB-XL. *IEEE Journal of Selected Topics in Signal Processing*, 15(1):151–167, 2021.

Patrick Wagner, Nils Strodthoff, Ralf-Dieter Bousseljot, Dieter Kreiseler, Fatima I Lunze, Wojciech Samek, and Tobias Schaeffter. PTB-XL, a large publicly available electrocardiography dataset. *Scientific Data*, 7(1):1–15, 2020.

Ning Wang, Panpan Feng, Zhaoyang Ge, Yanjie Zhou, Bing Zhou, and Zongmin Wang. Adversarial spatiotemporal contrastive learning for electrocardiogram signals. *IEEE Transactions on Neural Networks and Learning Systems*, 35(10):13845–13859, 2023.

Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *Proceedings of the International Conference on Machine Learning*, volume 139, pp. 12310–12320. PMLR, 2021.

Wenrui Zhang, Ling Yang, Shijia Geng, and Shenda Hong. Self-supervised time series representation learning via cross reconstruction transformer. *IEEE Transactions on Neural Networks and Learning Systems*, 35(11):16129–16138, 2024.

Jianwei Zheng, Huimin Chu, Daniele Struppa, Jianming Zhang, Sir Magdi Yacoub, Hesham El-Askary, Anthony Chang, Louis Ehwerhemuepha, Islam Abudayyeh, Alexander Barrett, et al. Optimal multi-stage arrhythmia classification approach. *Scientific Reports*, 10(1):2898, 2020.

Jianwei Zheng, Hangyuan Guo, and Huimin Chu. A large scale 12-lead electrocardiogram database for arrhythmia study. *PhysioNet*, 2022. Version 1.0.0.

APPENDIX CONTENTS

A Limitations **15**

B Notation **15**

C Implementation Details **15**

C.1 Model and Optimization Settings 15

C.2 Patch Embedding 17

D Lead Structure and Spherical-Harmonic **17**

D.1 Canonical Lead Vectors 17

D.2 Real Spherical-Harmonic Features 17

E Dataset and Preprocessing Details **19**

E.1 Pretraining Datasets 19

E.2 Downstream Evaluation Datasets 19

E.3 Signal Preprocessing 20

E.4 Label Subsets 20

F Evaluation Protocols and Additional Results **20**

F.1 Model Sources and Probing Setup 20

F.2 Disease Classification 21

F.3 Echocardiographic Phenotype Prediction 21

F.4 Classification with Fewer Leads 22

F.5 Missing-Lead Reconstruction 22

G Ablation Protocols and Full Results **23**

G.1 Lead-Direction Assignment and Perturbation 23

G.2 Full Results Across Label Fractions 24

A LIMITATIONS

Our canonical lead geometry provides a shared approximation rather than patient-specific acquisition directions. Although direction perturbations allow small deviations during pretraining, they do not explicitly model individual anatomy or actual electrode displacement. Future work will explore adapting the directional representation to alternative acquisition geometries and evaluating its robustness on wearable recordings under varying placement and recording conditions.

B NOTATION

Table 7: Summary of notations.

Notations	Explanation
C, T, P, N	Numbers of leads, temporal samples, samples per patch, and patches per lead, respectively ($N = T/P$).
c, n, s	Lead, temporal-patch, and retained-position indices, respectively.
$\mathbf{X} \in \mathbb{R}^{C\times T}$	Multi-lead ECG signal with C leads and T temporal samples.
$\mathbf{x}_{c,n} \in \mathbb{R}^{P}$	Waveform patch n from lead c.
$\mathbf{X}^{\text{patch}} \in \mathbb{R}^{C\times N\times P}$	Tensor collecting waveform patches over all leads and temporal positions.
g	Index of a lead group, either limb or precordial.
$\mathcal{G}_{\text{limb}}$, $\mathcal{G}_{\text{pre}}$	Sets containing the six limb leads and six precordial leads, respectively.
k_g	Number of leads sampled from lead group g.
$\mathcal{S}_g$	Subset of leads sampled from lead group g.
$\mathcal{S}$	Union of the sampled limb and precordial lead subsets.
m_c^{lead}, m_n^{time}, $m_{c,n}$	Lead-selection, shared temporal-retention, and patch-retention indicators, respectively.
$\mathbf{M}$	Complete mask over all lead–time patch positions.
Ω, $\mathcal{M}$	Sets of retained and masked lead–time patch indices, respectively.
D	Dimensionality of each latent token.
f_{patch}	Shared Patch Embedding module for waveform patches.
$\mathbf{z}_{c,s} \in \mathbb{R}^{D}$	Visible waveform token at (c, s).
$\mathbf{a}_c$	Forward-Dower coefficient vector for lead c.
$\mathbf{v}_c$, $\widetilde{\mathbf{v}}_c$	Original and perturbed unit directions of lead c, respectively.
L	Maximum spherical-harmonic degree ($L = 3$).
$\boldsymbol{\phi}_L(\cdot)$	Real spherical-harmonic feature map truncated at degree L.
$\mathbf{W}_{\text{sh}}$	Learned projection from spherical-harmonic features to token space.
$\mathbf{e}_c^{\text{lead}}$	Lead Direction Embedding associated with lead c.
g_{time}	Conditional encoder for temporal position information.
$\bar{\mathbf{z}}_{c,:}$	Token sequence completed with placeholders at masked positions.
$\mathbf{h}_{c,s}^{0}$	Visible token combining waveform, lead-direction, and temporal context.
E_Θ	Transformer encoder operating on the retained tokens.
$\mathbf{H}_\Omega$	Collection of encoder outputs over all retained positions.
$\mathbf{h}_{c,s}^{\text{enc}}$	Encoder output at position (c, s).
$\widehat{\mathbf{x}}_{c,s}$	Waveform patch predicted by the reconstruction decoder.
$\mathcal{L}_{\text{rec}}$	Mean absolute reconstruction loss evaluated over masked patches.
$\mathbf{r}$	Mean-pooled ECG representation used for downstream linear probing.
$Y_{\ell m}^{\mathbb{R}}$	Real spherical harmonic of degree ℓ and order m.
P_ℓ	Legendre polynomial of degree ℓ.
$\mathbf{u}$, $\mathbf{v}$, γ	Two unit directions and their angular separation, respectively.
K_L, $K_{L,\mathbf{W}}$	Directional kernels before and after the learned linear projection, respectively.
$\mathbf{p}(\mathbf{u})$, $\mathbf{e}(\mathbf{u})$	Projected directional features before and after applying LayerNorm, respectively.
K	Number of observed leads provided during reconstruction.

C IMPLEMENTATION DETAILS

C.1 MODEL AND OPTIMIZATION SETTINGS

Tabs. 8 to 10 summarize the model architecture, pretraining configuration, and linear-probing settings. The 42.6M parameter count includes the reconstruction decoder, which is discarded for downstream linear probing.

Table 8: Encoder and decoder hyperparameters of Φ-ECG.

Hyperparameters	Encoder	Decoder
Transformer blocks	12	4
Embedding dimension	512	256
FFN expansion ratio	4	4
FFN hidden dimension	2,048	1,024
Attention heads	8	4
Dropout	0.1	0.1

Table 9: Pretraining hyperparameters of Φ-ECG.

Hyperparameters	Values
Batch size	1,024 per GPU
Peak learning rate	1×10^{-3}
Minimum learning rate	5×10^{-5}
Learning-rate scheduler	Linear warmup + cosine decay
Optimizer	AdamW
Adam $\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$	$(0.9, 0.999)$ and 10^{-8}
Weight decay	$0.05 \rightarrow 0.01$ (cosine)
Warmup epochs	10
Mixed precision	bfloat16
Hardware	$8\times$ NVIDIA A800-SXM4 (80 GB)
Gradient clipping	Global norm 1.0
Visible lead sampling	1 limb + 2 precordial leads
Temporal mask ratio	0.4 (6 of 10 positions visible)
Visible encoder tokens	18 of 120 per recording
Direction noise	Gaussian ($\sigma = 0.1$), then unit normalization
MIMIC+CODE pretraining MIMIC-only pretraining	200 epochs

Table 10: Downstream linear-probing hyperparameters.

Hyperparameters	Values
Batch size	256
Peak learning rate	5×10^{-3}
Minimum learning rate	10^{-5}
Learning-rate scheduler	10-epoch warmup + cosine decay
Optimizer	AdamW
Weight decay	0.05
Total epochs	100
Trainable module	Task-specific linear head; encoder frozen
Objective	Binary cross-entropy
Checkpoint selection	Highest validation AUROC
Probe seeds	$\{0, 1, 2, 3, 42\}$ (five-seed)

Weight decay is applied to weights other than biases, normalization parameters, embeddings, and mask tokens. Distributed pretraining uses seed $42 + r$ on rank r, and evaluation uses the fixed canonical lead directions. Five-seed results report the mean and sample standard deviation across the seeds listed in Tab. 10.

C.2 PATCH EMBEDDING

Patch Embedding maps each one-second lead segment to one token. Four bias-free convolutional stages reduce the 500-sample input to a D-dimensional vector, with eight-group normalization and GELU after every convolution. Kernels operate along the sample axis within each lead–patch pair. The layer configuration is given in Tab. 11.

Table 11: Patch Embedding architecture for 500-sample waveform patches.

Stage	Channels	Kernel	Stride	Padding	Length
Conv 1	$1 \to 64$	31	4	15	125
Conv 2	$64 \to 64$	15	2	7	63
Conv 3	$64 \to 64$	7	2	3	32
Conv 4	$64 \to D$	31	31	0	1

D LEAD STRUCTURE AND SPHERICAL-HARMONIC

This section provides the background and construction details of the lead-direction embedding used in Φ-ECG. We first describe how canonical lead vectors are obtained from a linear VCG-to-ECG synthesis model. We then explain how real spherical harmonics map these directions to features that are shared across leads.

D.1 CANONICAL LEAD VECTORS

Vectorcardiography (VCG) was developed to represent the changing magnitude and spatial direction of the heart's electrical activity over the cardiac cycle. It builds on the approximation of the cardiac electrical source as a time-varying equivalent dipole, represented by a three-dimensional vector. In an orthogonal VCG system, this vector is described by three components, denoted by $X(t)$, $Y(t)$, and $Z(t)$ (Frank, 1956; Macfarlane et al., 2010). Under a linear lead-vector approximation, an ECG lead is represented as a weighted combination of these components. The corresponding coefficients form a lead vector that characterizes the lead's sensitivity to the three coordinate directions (Burger & van Milaan, 1948; Dower, 1968).

We obtain canonical lead vectors from the 8×3 forward-Dower VCG-to-ECG synthesis matrix (Dower, 1968). Under this linear approximation, the waveform of lead c is expressed as

$$s_c(t) \approx a_{c,X} X(t) + a_{c,Y} Y(t) + a_{c,Z} Z(t) = \mathbf{a}_c^\top \mathbf{q}(t), \tag{12}$$

where $\mathbf{q}(t) = (X(t), Y(t), Z(t))^\top$ and $\mathbf{a}_c = (a_{c,X}, a_{c,Y}, a_{c,Z})^\top$ contains the corresponding synthesis coefficients. We use the Frank coordinate convention, with $+X$ pointing from the patient's right to left, $+Y$ from head to feet, and $+Z$ from anterior to posterior (Frank, 1956; Dower, 1968).

The eight rows of the synthesis matrix correspond to leads I, II, and V1–V6. The remaining four limb leads are determined by the Einthoven and Goldberger relations (Clifford et al., 2006; Macfarlane et al., 2010). Because these relations are linear, they also apply to the synthesis coefficient vectors:

$$\mathbf{a}_{\mathrm{III}} = \mathbf{a}_{\mathrm{II}} - \mathbf{a}_{\mathrm{I}}, \quad \mathbf{a}_{\mathrm{aVR}} = -\frac{\mathbf{a}_{\mathrm{I}} + \mathbf{a}_{\mathrm{II}}}{2}, \quad \mathbf{a}_{\mathrm{aVL}} = \mathbf{a}_{\mathrm{I}} - \frac{\mathbf{a}_{\mathrm{II}}}{2} \quad \mathbf{a}_{\mathrm{aVF}} = \mathbf{a}_{\mathrm{II}} - \frac{\mathbf{a}_{\mathrm{I}}}{2}.$$

Thus, the twelve lead vectors follow from a common synthesis model that preserves the algebraic relationships among the limb leads. The direction of lead c is represented by the unit vector $\mathbf{v}_c = \frac{\mathbf{a}_c}{\|\mathbf{a}_c\|_2} \in \mathbb{S}^2$, where $\mathbb{S}^2$ denotes the unit sphere in $\mathbb{R}^3$. These canonical directions provide a shared approximation of lead geometry across recordings. The coefficients and corresponding unit directions are listed in Tab. 12.

D.2 REAL SPHERICAL-HARMONIC FEATURES

Spherical harmonics are basis functions for representing angular variation on the sphere (Atkinson & Han, 2012). They play a role analogous to sine and cosine functions in Fourier representations, but their domain is a three-dimensional direction rather than a one-dimensional periodic coordinate. Here, they are evaluated at lead directions to construct positional features.

Table 12: Forward-Dower coefficients and normalized lead directions.

Lead	Forward-Dower coefficients a_X	a_Y	a_Z	Normalized direction $\mathbf{v}_c$ v_X	v_Y	v_Z
I	0.6320	-0.2350	0.0590	0.933733	-0.347195	0.087168
II	0.2350	1.0660	-0.1320	0.213724	0.969490	-0.120049
III	-0.3970	1.3010	-0.1910	-0.289028	0.947168	-0.139054
aVR	-0.4335	-0.4155	0.0365	-0.720606	-0.690685	0.060674
aVL	0.5145	-0.7680	0.1250	0.551551	-0.823307	0.134002
aVF	-0.0810	1.1835	-0.1615	-0.067657	0.988547	-0.134897
V1	-0.5150	0.1570	-0.9170	-0.484308	0.147643	-0.862350
V2	0.0440	0.1640	-1.3870	0.031488	0.117365	-0.992590
V3	0.8820	0.0980	-1.2770	0.567175	0.063019	-0.821182
V4	1.2130	0.1270	-0.6010	0.892129	0.093405	-0.442019
V5	1.1250	0.1270	-0.0860	0.990834	0.111854	-0.075744
V6	0.8310	0.0760	0.2300	0.960045	0.087802	0.265716

Each real spherical harmonic $Y_{\ell,m}$ is indexed by a degree $\ell \geq 0$ and an order $m \in \{-\ell, \ldots, \ell\}$. The degree determines the angular complexity of the basis functions, while the order indexes the $2\ell + 1$ functions within that degree. Degree zero is constant over the sphere, degree one represents linear dependence on the direction coordinates, and higher degrees introduce finer angular variation (Atkinson & Han, 2012).

For a unit direction $\mathbf{v} = (x, y, z)^\top$, we concatenate real orthonormal spherical harmonics by increasing degree ℓ and, within each degree, increasing order m:

$$\boldsymbol{\phi}_L(\mathbf{v}) = \left[Y_{0,0}(\mathbf{v}), Y_{1,-1}(\mathbf{v}), Y_{1,0}(\mathbf{v}), Y_{1,1}(\mathbf{v}), \ldots, Y_{L,L}(\mathbf{v})\right]^\top \in \mathbb{R}^{(L+1)^2}. \tag{13}$$

The feature dimension follows from $\sum_{\ell=0}^{L}(2\ell + 1) = (L + 1)^2$. Under our real-basis convention, the first four components are

$$Y_{0,0}(\mathbf{v}) = \frac{1}{\sqrt{4\pi}}, \qquad \begin{bmatrix} Y_{1,-1}(\mathbf{v}) \\ Y_{1,0}(\mathbf{v}) \\ Y_{1,1}(\mathbf{v}) \end{bmatrix} = \sqrt{\frac{3}{4\pi}} \begin{bmatrix} y \\ z \\ x \end{bmatrix}. \tag{14}$$

These examples show how the basis converts a direction into numerical features: the constant component is shared by all leads, while the degree-one components already distinguish their directions. The maximum degree L controls the angular expressiveness of the shared mapping. Increasing L adds higher-degree basis functions and allows more flexible dependence on lead direction. Our default setting uses $L = 3$, yielding $1 + 3 + 5 + 7 = 16$ basis features. A shared learned projection maps these fixed features to the embedding space, followed by LayerNorm:

$$\mathbf{e}_c^{\text{lead}} = \mathrm{LN}(\mathbf{W}_{\text{sh}} \boldsymbol{\phi}_3(\mathbf{v}_c)), \qquad \mathbf{W}_{\text{sh}} \in \mathbb{R}^{D \times 16}. \tag{15}$$

Before LayerNorm, each embedding coordinate is a learned linear combination of the basis functions. The same projection is used for all leads, so their embeddings are generated by a common function of direction. Each resulting lead embedding is reused across the temporal patches of that lead. During pretraining, we perturb each canonical unit vector with independent Gaussian noise and normalize it again:

$$\widetilde{\mathbf{v}}_c = \frac{\mathbf{v}_c + \boldsymbol{\epsilon}_c}{\|\mathbf{v}_c + \boldsymbol{\epsilon}_c\|_2}, \qquad \boldsymbol{\epsilon}_c \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_3), \qquad \sigma = 0.1. \tag{16}$$

The perturbed direction replaces $\mathbf{v}_c$ in the embedding computation. This extends training to neighborhoods around the canonical directions, where features are generated by the same spherical-harmonic basis and learned projection. Downstream evaluation uses the canonical directions without perturbation. Together, the basis expansion and shared projection provide a continuous mapping from lead geometry to learned lead representations.

E Dataset and Preprocessing Details

E.1 Pretraining Datasets

We use MIMIC-IV-ECG and CODE for self-supervised pretraining. These datasets provide large collections of clinical ECGs acquired in different healthcare settings. Pretraining uses ECG waveforms without diagnostic labels.

MIMIC-IV-ECG. MIMIC-IV-ECG (Gow et al., 2023) contains approximately 800,000 diagnostic ECGs from nearly 160,000 patients at Beth Israel Deaconess Medical Center. Each recording contains twelve leads acquired over ten seconds at 500 Hz. The dataset includes ECGs collected during routine clinical care and can be linked to the broader MIMIC-IV clinical database. In this study, signal-level quality control retains 773,000 recordings for the pretraining and self-supervised validation partitions. MIMIC-IV-ECG also serves as the pretraining source for the MIMIC-only comparison involving HeartLang, LVCG, ShapeEFM, and Φ-ECG (MIMIC).

CODE. CODE (Clinical Outcomes in Digital Electrocardiology) (Ribeiro et al., 2020) is a large-scale ECG resource collected through the Telehealth Network of Minas Gerais in Brazil. It provides recordings acquired in a different clinical setting from MIMIC-IV-ECG and constitutes the larger source in our combined pretraining corpus. From the available CODE archives, we retain files whose identifiers end in `N1`, yielding 2,800,000 recordings. The retained files contain eight independent leads, I, II, and V1–V6, sampled at 300 Hz. The remaining limb leads are derived during preprocessing.

Pretraining partitions. Within each source, every retained recording is independently assigned to self-supervised validation with probability 2% and the remaining recordings form the training partition. We concatenate the two training partitions, with MIMIC-IV-ECG and CODE contributing approximately 21.6% and 78.4% of the combined training records, respectively.

E.2 Downstream Evaluation Datasets

Table 13: Downstream datasets, task sizes, and data splits.

Task	Labels	Train	Validation	Test	Split unit / provenance
PTBXL-Super (Wagner et al., 2020)	5	17,084	2,146	2,158	Official patient folds
PTBXL-Sub (Wagner et al., 2020)	23	17,084	2,146	2,158	Official patient folds
PTBXL-Form (Wagner et al., 2020)	19	7,197	901	880	Official patient folds
PTBXL-Rhythm (Wagner et al., 2020)	12	16,832	2,100	2,098	Official patient folds
CPSC2018 (Liu et al., 2018)	9	4,950	551	1,376	Released participant lists
CSN (Zheng et al., 2020; 2022)	38	16,546	1,860	4,620	Released record split
EchoNext-Mini (Hughes et al., 2026; Elias & Finer, 2026)	12	72,475	4,626	5,442	Official split

Downstream evaluation covers six ECG disease-classification tasks, echocardiographic phenotype prediction, and multi-lead waveform reconstruction. The disease-classification tasks comprise four annotation groups from PTB-XL and one task each from CPSC2018 and CSN. EchoNext-Mini evaluates prediction of echocardiogram-derived targets, while CPSC2018 and the PTB Diagnostic ECG Database are used for reconstruction. Task-specific sample counts and split provenance for disease classification are summarized in Tab. 13.

PTB-XL. PTB-XL (Wagner et al., 2020; Strodthoff et al., 2021) is a clinical 12-lead ECG dataset with annotations describing diagnostic findings, waveform morphology, and cardiac rhythm. The version used in our experiments contains 21,837 ten-second recordings. We evaluate 4 tasks: diagnostic superclass classification with 5 labels, diagnostic subclass classification with 23 labels, form classification with 19 labels, and rhythm classification with 12 labels. These tasks assess representations at different levels of diagnostic detail and across different aspects of ECG interpretation. We follow the recommended patient-level folds, and task-specific sample counts are reported in Tab. 13.

CPSC2018. CPSC2018 (Liu et al., 2018) was introduced for evaluating the detection of ECG rhythm and morphology abnormalities. It contains 6,877 ECG recordings from distinct participants with nine diagnostic labels. We use the released participant lists to form training, validation, and test partitions containing 4,950, 551, and 1,376 recordings, respectively. Each recording is represented by one ten-second input after the temporal standardization described below. The same 1,376-record test partition is also used for multi-lead reconstruction evaluation.

CSN. CSN (Zheng et al., 2020; 2022) provides clinical 12-lead ECG recordings for studying arrhythmias and other ECG abnormalities. In our evaluation, its 38-label primary task complements the smaller label sets in PTB-XL and CPSC2018. After removing records with unknown annotations, we retain 23,026 recordings and follow the released split of 16,546 training, 1,860 validation, and 4,620 test recordings.

EchoNext-Mini. EchoNext-Mini (Hughes et al., 2026; Elias & Finer, 2026) links ECG recordings to echocardiogram-derived labels for structural heart disease. It therefore evaluates the transfer of ECG representations to cardiac phenotypes assessed through another clinical modality. The dataset contains 100,000 ten-second, 12-lead ECGs sampled at 250 Hz. After excluding records marked `no_split`, we use the official training, validation, and test partitions containing 72,475, 4,626, and 5,442 recordings, respectively. The twelve prediction targets comprise eleven component conditions and their composite structural-heart-disease endpoint. The validation and test partitions each contain one ECG per patient.

PTB Diagnostic ECG Database. The PTB Diagnostic ECG Database (Bousseljot et al., 1995) contains 549 recordings from 290 subjects, including patients with cardiac conditions and healthy controls. Each recording includes the twelve standard ECG leads and three Frank leads. We use the standard ECG leads for multi-lead reconstruction evaluation following the protocol of LVCG (Huang et al., 2026). Together with the CPSC2018 test partition, this dataset provides a second collection for assessing recovery of unobserved lead waveforms.

E.3 Signal Preprocessing

Signal repair and resampling. We repair non-finite samples by linear interpolation within each lead, convert signal amplitudes to millivolts, and resample the waveforms to 500 Hz. This establishes a common amplitude unit and sampling frequency across the data sources.

Filtering and amplitude clipping. Signals are processed using a fifth-order Butterworth bandpass filter with cutoffs of 0.5 and 40 Hz, applied with zero-phase filtering. We also apply a notch filter at the source-specific power-line frequency. The resulting amplitudes are clipped to ± 3 mV.

Recording length and lead order. Each recording is center-cropped or zero-padded to 5,000 samples, corresponding to ten seconds at 500 Hz. The standard lead order is $(\mathrm{I}, \mathrm{II}, \mathrm{III}, \mathrm{aVR}, \mathrm{aVL}, \mathrm{aVF}, \mathrm{V1}, \mathrm{V2}, \mathrm{V3}, \mathrm{V4}, \mathrm{V5}, \mathrm{V6})$. For CODE, the four derived limb leads are computed from I and II after resampling using the Einthoven and Goldberger relations. Pretraining and reconstruction use the processed waveforms in millivolts.

E.4 Label Subsets

We evaluate disease classification using 1%, 10%, and 100% of the training labels. For a training partition containing N recordings and a label fraction $r \in \{0.01, 0.1\}$, we randomly sample $\max(1, \lfloor rN \rfloor)$ recordings without replacement and retain their complete label vectors. For each task and repeated run, all models use the same selected training records and the same validation and test partitions. Full-label experiments use the entire training partition.

F Evaluation Protocols and Additional Results

F.1 Model Sources and Probing Setup

Pretrained Checkpoints. For the five-seed disease-classification comparison in Tab. 16, the EchoNext-Mini evaluation in Tab. 2, and the reduced-lead evaluation in Fig. 3, we use official publicly released checkpoints for ST-MEM and HeartLang, and pretrain LVCG and ShapeEFM on MIMIC-IV-ECG using their official implementations. The pretraining datasets, schedules, and checkpoint sources are summarized in Tab. 14. HeartLang, LVCG, ShapeEFM, and Φ-ECG (MIMIC) form the matched-pretraining-data comparison where evaluated together. The default Φ-ECG model additionally uses CODE recordings.

Input Adaptation and Probe Configuration. Each encoder retains its native lead order, sampling rate, amplitude transformation, tokenization, and recording-level feature extraction. For the within-study classification, EchoNext-Mini, and reduced-lead evaluations, we freeze the pretrained encoders and tune linear-probing hyperparameters separately for each model, starting from its original repository configuration. Model-specific input adaptations and probe configurations are summarized in Tab. 15.

Table 14: Pretraining data, optimization budgets, model sizes, and checkpoint sources.

Model	Pretraining data	ECGs	Budget	Parameters	Checkpoint source
ST-MEM (Na et al., 2024)	Chapman + Ningbo + CODE-15	–	–	85M	Official released checkpoint
HeartLang (Jin et al., 2025)	MIMIC-IV-ECG	–	–	39.08M	Official released checkpoint
LVCG (Huang et al., 2026)	MIMIC-IV-ECG	773k	500k updates	8.29M	Our implementation
ShapeEFM (Liu et al., 2026)	MIMIC-IV-ECG	773k	200 epochs	31.48M	Our implementation
Φ-ECG (MIMIC)	MIMIC-IV-ECG	773k	200 epochs	42.6M	Our implementation
Φ-ECG	MIMIC-IV-ECG + CODE	3.57M	200 epochs	42.6M	Our implementation

Table 15: Model-specific input preprocessing and recording-level feature extraction for linear probing. Input dimensions refer to full-lead recordings.

Model	Input preprocessing	Recording representation
ST-MEM (Na et al., 2024)	12 leads at 250 Hz, 2,250 samples. Filtering and standardization follow the official implementation.	Mean-pooled final-layer tokens.
HeartLang (Jin et al., 2025)	QRS-centered tokenization with the official lead order.	Final-layer recording token.
LVCG (Huang et al., 2026)	12 leads at 100 Hz with the official z-score normalization.	640-dimensional `ext_ecg_emb`.
ShapeEFM (Liu et al., 2026)	Morphology preprocessing at 500 Hz with the official lead order.	Mean-pooled final-layer tokens.
Φ-ECG	12 leads at 500 Hz, 5,000 samples. Normalization uses training-partition statistics.	Mean-pooled final-layer tokens.

F.2 Disease Classification

Baselines. The general self-supervised baselines represent complementary learning objectives. SimCLR (Chen et al., 2020) contrasts augmented views, BYOL (Grill et al., 2020) aligns online and target networks, and Barlow Twins (Zbontar et al., 2021) reduces cross-view redundancy. MoCo-v3 (Chen et al., 2021b) uses a momentum target encoder, while SimSiam (Chen & He, 2021) employs stop-gradient Siamese prediction. TS-TCC (Eldele et al., 2021) combines temporal and contextual contrastive learning. ECG-specific baselines exploit signal structure at different scales. CLOCS (Kiyasseh et al., 2021) constructs contrastive views using temporal segments, leads, and patient identity. ASTCL (Wang et al., 2023) employs adversarial spatio-temporal contrastive learning, while CRT (Zhang et al., 2024) connects temporal and spectral representations through reconstruction. ST-MEM (Na et al., 2024) reconstructs masked ECG regions, and HeartLang (Jin et al., 2025) models QRS-centered beat tokens. LVCG (Huang et al., 2026) learns representations in latent vectorcardiogram space, while ShapeEFM (Liu et al., 2026) learns transferable morphology representations.

Literature Comparison. Tab. 1 combines previously reported baseline results, obtained under their original evaluation settings, with results from both Φ-ECG variants. Table 16 reports the within-study comparison using the downstream partitions described above.

Within-Study Evaluation. We evaluate frozen checkpoints of ST-MEM, HeartLang, LVCG, ShapeEFM, and both Φ-ECG variants on all six disease-classification tasks using the dataset partitions and label definitions in Tab. 13. For each fixed pretrained checkpoint, we train five linear probes with seeds $\{0, 1, 2, 3, 42\}$ and report their mean AUROC and sample standard deviation in Tab. 16.

F.3 Echocardiographic Phenotype Prediction

Evaluation Protocol. We evaluate ST-MEM, HeartLang, LVCG, ShapeEFM, and both Φ-ECG variants on the same 12 EchoNext-Mini endpoints using waveform-only linear probing. Models share training, validation, and test partitions and are evaluated with 1%, 10%, and 100% of the training labels. For each label fraction, all models use identical labeled subsets indexed by probe seeds $\{0, 1, 2, 3, 42\}$.

Endpoint-Level Results. Tab. 17 reports five-seed means for individual endpoints. AS, AR, MR, TR, and PR denote moderate-or-greater aortic stenosis, aortic regurgitation, mitral regurgitation,

Table 16: Linear-probing AUROC (mean ± std, %) over five seeds.

Method	PTBXL-Super	PTBXL-Sub	PTBXL-Form	PTBXL-Rhythm	CPSC2018	CSN
			1% labels			
ST-MEM	59.83 ± 2.81	53.10 ± 2.26	56.81 ± 2.73	51.36 ± 7.20	56.60 ± 4.43	57.96 ± 1.01
HeartLang	77.90 ± 0.97	62.89 ± 1.93	59.69 ± 1.62	63.18 ± 4.89	60.85 ± 0.91	58.02 ± 2.57
LVCG	75.48 ± 0.65	70.95 ± 0.87	52.75 ± 2.13	73.90 ± 2.48	68.07 ± 1.61	64.91 ± 1.43
ShapeEFM	79.63 ± 0.78	72.89 ± 1.73	58.36 ± 1.55	70.33 ± 2.07	71.62 ± 1.02	67.94 ± 0.91
Φ-ECG (MIMIC)	83.41 ± 0.58	78.18 ± 1.04	59.76 ± 2.49	80.16 ± 2.28	83.23 ± 0.78	75.87 ± 1.55
Φ-ECG	**84.13 ± 0.69**	**78.98 ± 0.86**	**69.81 ± 2.40**	**84.42 ± 1.85**	**84.86 ± 2.10**	**79.92 ± 0.82**
			10% labels			
ST-MEM	67.93 ± 1.38	59.22 ± 1.80	60.16 ± 1.65	63.61 ± 6.09	63.73 ± 4.02	66.53 ± 0.63
HeartLang	85.39 ± 0.21	78.02 ± 0.74	63.89 ± 1.46	77.70 ± 2.95	68.02 ± 0.94	70.70 ± 1.74
LVCG	78.77 ± 0.47	72.99 ± 0.89	60.32 ± 1.71	82.29 ± 1.77	80.59 ± 0.49	75.83 ± 1.30
ShapeEFM	86.28 ± 0.17	83.32 ± 0.73	72.05 ± 1.18	79.46 ± 1.63	81.48 ± 0.41	78.19 ± 1.09
Φ-ECG (MIMIC)	87.44 ± 0.30	86.76 ± 1.13	74.13 ± 2.03	87.71 ± 1.49	91.05 ± 0.39	87.79 ± 1.29
Φ-ECG	**89.21 ± 0.18**	**88.77 ± 1.16**	**81.08 ± 1.49**	**91.28 ± 1.80**	**92.84 ± 0.29**	**90.05 ± 1.04**
			100% labels			
ST-MEM	73.06 ± 0.93	65.35 ± 1.29	66.89 ± 0.63	75.34 ± 4.14	70.74 ± 2.34	70.70 ± 0.55
HeartLang	86.52 ± 0.07	88.01 ± 0.92	79.93 ± 0.16	91.32 ± 0.62	77.80 ± 0.08	83.73 ± 0.09
LVCG	78.86 ± 0.47	77.63 ± 0.60	73.94 ± 1.08	84.70 ± 1.93	82.71 ± 0.53	86.44 ± 0.35
ShapeEFM	88.89 ± 0.02	89.65 ± 0.19	83.99 ± 0.43	85.51 ± 0.12	88.93 ± 0.04	88.54 ± 0.02
Φ-ECG (MIMIC)	89.75 ± 0.01	92.06 ± 0.09	85.73 ± 0.33	94.20 ± 0.13	94.83 ± 0.01	95.09 ± 0.03
Φ-ECG	**91.35 ± 0.03**	**93.17 ± 0.06**	**88.66 ± 0.07**	**94.66 ± 0.23**	**96.05 ± 0.02**	**96.37 ± 0.01**

tricuspid regurgitation, and pulmonary regurgitation. RVSD denotes moderate-or-greater right-ventricular systolic dysfunction, PE denotes moderate or large pericardial effusion, and SHD denotes moderate-or-greater structural heart disease.

Statistical Comparisons. We use two-sided paired Student's t-tests. Holm correction is applied across the 36 EchoNext-Mini comparisons at $\alpha = 0.05$, separately from the disease-classification comparisons. Mean differences and adjusted p-values are reported in Tab. 18.

F.4 Classification with Fewer Leads

Lead Configurations and Input Adaptation. We evaluate five configurations comprising lead I, leads I/II, leads I/II/V2, all six limb leads, and all 12 standard leads. Tokens corresponding to unavailable leads are removed for each evaluated model. The remaining inputs follow the encoder's native inference pipeline.

Linear-Probing Protocol. A linear probe is trained for each task and lead configuration using full training labels and probe seeds $\{0, 1, 2, 3, 42\}$. The pretrained backbone remains frozen, and each lead configuration is held fixed throughout probe training, validation, and testing. The Φ-ECG results use the default checkpoint pretrained on MIMIC-IV-ECG and CODE.

F.5 Missing-Lead Reconstruction

Baselines and Result Sources. KIM (Kors et al., 1990) uses a linear transformation between ECG and Frank vectorcardiogram signals. VAE-CNN (Matyschik et al., 2020) uses a variational convolutional model, while E-LSTM (Sohn et al., 2020) models lead reconstruction with recurrent units. Nef-Net (Chen et al., 2021a) predicts ECG waveforms at queried observation views. LVCG (Huang et al., 2026) projects decoded latent VCG signals onto target leads. Baseline values in Tab. 3 are taken from the reconstruction benchmark compiled by Huang et al. (2026).

Lead Sampling and Inference. Following the three- and five-lead settings of LVCG (Huang et al., 2026), we evaluate missing-lead reconstruction on CPSC2018 and PTB. For each recording, we uniformly sample subsets of three or five distinct leads without replacement using seed 42. The remaining nine or seven leads serve as reconstruction targets. Φ-ECG retains its pretraining preprocessing and reconstructs waveforms in millivolts at 500 Hz without additional training.

Normalization and Metrics. Predictions and reference waveforms are resampled to 100 Hz with anti-aliasing filtering and aligned on the same temporal grid. For scoring, we apply the same

Table 17: Per-label EchoNext performance (mean, %) over five seeds across label fractions.

Label	1%					10%					100%				
	ST-MEM	HeartLang	LVCG	ShapeEFM	Φ-ECG	ST-MEM	HeartLang	LVCG	ShapeEFM	Φ-ECG	ST-MEM	HeartLang	LVCG	ShapeEFM	Φ-ECG
AUROC															
LVEF $\leq 45\%$	67.02	80.37	81.13	81.14	**85.80**	84.20	83.55	85.16	84.42	**89.09**	84.52	84.81	85.48	85.66	**89.98**
LVWT ≥ 13 mm	57.99	67.03	67.78	68.13	**72.30**	69.63	70.41	70.21	72.16	**75.34**	69.92	72.86	71.44	73.59	**76.53**
AS	56.17	62.46	60.16	66.49	**73.55**	68.78	71.12	75.39	71.54	**81.17**	71.57	74.73	74.79	75.15	**83.94**
AR	50.60	55.25	59.61	**62.29**	61.69	56.18	68.46	61.80	**70.62**	69.93	62.54	67.09	63.34	**74.28**	73.90
MR	55.04	73.41	70.59	72.41	**76.48**	75.52	77.44	78.09	75.55	**80.03**	76.99	79.06	77.41	77.24	**81.77**
TR	56.57	70.55	73.31	69.29	**76.40**	75.41	74.37	76.93	74.65	**80.82**	76.74	77.17	78.64	76.98	**82.48**
PR	44.10	54.21	**68.75**	60.60	66.79	47.85	74.62	80.00	76.29	**80.61**	59.12	77.01	81.12	85.64	**85.98**
RVSD	60.85	77.96	79.35	78.27	**84.31**	82.46	82.57	84.03	83.14	**87.20**	83.13	84.11	83.77	84.46	**88.36**
PE	47.87	56.85	62.91	63.93	**65.94**	60.71	64.58	68.35	68.56	**72.37**	66.09	66.82	68.66	71.35	**73.59**
PASP ≥ 45 mmHg	58.00	66.60	68.50	67.29	**71.93**	72.50	69.85	71.65	71.55	**75.83**	73.03	71.71	71.33	73.05	**77.83**
TR $V_{\max} \geq 3.2$ m/s	45.90	63.50	65.20	63.21	**68.52**	66.48	67.81	69.46	69.72	**73.83**	68.19	71.04	70.49	71.57	**76.68**
SHD	73.05	75.36	76.86	75.18	**79.44**	79.34	78.16	78.84	78.37	**82.04**	79.43	79.32	79.37	79.53	**83.02**
AUPRC															
LVEF $\leq 45\%$	31.40	50.38	49.94	53.69	**60.12**	55.80	56.52	58.75	59.96	**69.99**	56.41	59.79	60.00	61.84	**72.13**
LVWT ≥ 13 mm	25.87	33.14	31.88	34.36	**38.28**	33.88	36.67	34.08	38.34	**42.53**	34.15	39.48	35.14	40.34	**43.94**
AS	6.76	8.83	7.76	10.56	**15.15**	10.26	11.68	15.41	13.16	**21.49**	12.05	13.04	14.73	16.27	**25.32**
AR	1.37	1.54	2.08	**2.64**	2.08	1.61	2.54	1.83	**3.39**	2.90	1.89	2.46	2.53	**3.80**	3.66
MR	8.34	14.74	14.60	18.09	**18.73**	17.99	18.88	19.30	20.39	**22.06**	19.00	21.41	18.87	22.89	**23.82**
TR	8.94	14.35	17.11	12.68	**18.46**	17.53	16.95	19.72	16.56	**25.91**	18.50	20.14	22.66	19.02	**29.24**
PR	0.39	0.66	0.93	0.69	**1.19**	0.44	1.23	3.56	1.72	**4.48**	0.61	3.25	3.74	**4.32**	2.83
RVSD	11.55	23.91	24.81	26.66	**34.38**	28.96	30.83	34.50	34.09	**44.49**	30.11	34.67	34.62	37.31	**48.47**
PE	1.52	2.61	2.56	**2.76**	2.70	1.97	2.35	2.92	2.91	**3.94**	2.58	2.86	3.35	2.89	**3.72**
PASP ≥ 45 mmHg	17.75	22.10	24.31	22.20	**27.16**	27.78	24.12	26.54	26.62	**33.62**	28.33	25.90	26.28	28.85	**36.10**
TR $V_{\max} \geq 3.2$ m/s	6.74	10.40	11.18	10.50	**13.27**	12.49	12.72	13.71	14.19	**20.53**	12.88	14.88	13.89	15.84	**21.66**
SHD	68.28	69.98	71.66	70.78	**75.36**	74.96	73.56	74.14	74.56	**79.35**	75.04	74.71	74.81	76.11	**80.47**
BACC															
LVEF $\leq 45\%$	62.90	73.02	73.69	73.78	**78.20**	76.74	75.58	78.05	76.67	**81.31**	76.90	76.53	77.91	77.54	**82.15**
LVWT ≥ 13 mm	55.93	62.27	63.04	63.22	**66.34**	64.77	65.27	65.10	66.50	**68.89**	65.09	67.25	65.72	67.42	**70.07**
AS	55.19	59.17	57.67	61.87	**67.14**	63.93	65.39	68.67	65.72	**72.83**	65.85	69.26	68.71	69.58	**75.70**
AR	52.15	52.12	55.49	**55.59**	50.43	54.76	61.92	56.20	**65.01**	63.91	55.60	62.72	55.27	**67.54**	67.31
MR	55.78	67.88	65.75	66.57	**69.90**	69.15	70.38	72.46	69.37	**73.48**	70.01	71.24	71.14	70.86	**75.27**
TR	56.16	65.83	67.81	64.10	**70.32**	67.55	68.60	70.42	67.91	**73.84**	70.05	69.83	71.62	69.57	**75.45**
PR	50.35	54.91	**62.77**	55.83	57.44	50.25	63.82	62.73	61.48	**67.46**	58.77	65.12	66.18	70.95	**76.23**
RVSD	58.81	71.40	72.71	72.01	**77.04**	74.68	75.59	76.53	75.80	**79.53**	74.93	76.36	76.03	77.60	**80.46**
PE	50.29	55.94	59.27	**59.76**	58.95	57.07	60.89	64.32	63.05	**67.40**	60.50	62.28	64.71	65.29	**67.95**
PASP ≥ 45 mmHg	56.34	62.51	63.22	62.17	**66.05**	66.66	64.99	66.07	66.15	**69.48**	67.32	65.95	65.30	67.29	**70.73**
TR $V_{\max} \geq 3.2$ m/s	50.71	60.51	61.35	59.97	**62.86**	61.81	63.00	64.64	64.85	**68.05**	62.86	65.79	64.81	66.45	**69.78**
SHD	67.00	68.88	70.31	68.33	**72.18**	72.06	71.26	72.18	70.96	**74.33**	72.31	71.92	72.25	72.00	**75.33**

reference-derived normalization to both waveforms

$$z_{i,c,t} = \frac{x_{i,c,t} - \mu_{i,c}}{s_{i,c}}, \qquad \hat{z}_{i,c,t} = \frac{\hat{x}_{i,c,t} - \mu_{i,c}}{s_{i,c}}, \qquad s_{i,c} = \max(\sigma_{i,c}, \epsilon), \tag{17}$$

where $x_{i,c,t}$ and $\hat{x}_{i,c,t}$ are the aligned reference and predicted values. The temporal mean $\mu_{i,c}$ and standard deviation $\sigma_{i,c}$ are computed from the reference lead and used only for scoring. MSE and MAE are computed in standardized units and averaged over missing leads and time samples within each subset, then over sampled subsets within each recording, and finally over recordings.

G Ablation Protocols and Full Results

G.1 Lead-Direction Assignment and Perturbation

We compare three lead-direction assignments, with full results reported in Tab. 20. The canonical assignment uses the normalized Dower direction $\mathbf{u}_c = \mathbf{v}_c$ for each lead c. The shuffled assignment uses $\mathbf{u}_c = \mathbf{v}_{\pi(c)}$, where π is a fixed random permutation of the 12 lead indices. The random assignment associates each lead with an independently sampled fixed direction on the unit sphere. The underlying assignments remain fixed throughout pretraining and evaluation, while waveform channels and reconstruction targets retain their original lead identities.

All three variants are pretrained separately using the same spherical-harmonic mapping and direction perturbation

$$\tilde{\mathbf{u}}_c = \frac{\mathbf{u}_c + \boldsymbol{\epsilon}_c}{\|\mathbf{u}_c + \boldsymbol{\epsilon}_c\|_2}, \qquad \boldsymbol{\epsilon}_c \sim \mathcal{N}(\mathbf{0}, \sigma_v^2 \mathbf{I}_3), \qquad \sigma_v = 0.1. \tag{18}$$

Evaluation uses the assigned directions without perturbation. The Dower w/o noise variant retains the canonical assignment and sets $\sigma_v = 0$ during pretraining, with all other settings unchanged.

Table 18: Paired significance tests for EchoNext-Mini. Each cell reports the mean metric difference in points and, in parentheses, its Holm-adjusted p-value for Φ-ECG (MIMIC) minus the indicated comparator. Bold adjusted p-values are below 0.05.

Comparator	AUROC	AUPRC	BACC
1% labels			
ST-MEM	16.48 (**0.0001**)	9.08 (**4.66e-5**)	10.18 (**0.0002**)
HeartLang	5.61 (**0.0037**)	3.77 (**0.0015**)	3.28 (**0.0171**)
LVCG	3.06 (**0.0228**)	3.25 (**0.0036**)	1.73 (0.0729)
ShapeEFM	3.56 (**0.0171**)	2.68 (**0.0022**)	2.55 (**0.0171**)
10% labels			
ST-MEM	8.38 (**0.0009**)	6.27 (**3.02e-5**)	5.66 (**0.0007**)
HeartLang	4.72 (**0.0073**)	5.91 (**8.57e-5**)	3.39 (**0.0171**)
LVCG	3.31 (**0.0022**)	4.54 (**4.57e-5**)	2.50 (**0.0098**)
ShapeEFM	3.59 (**0.0042**)	4.42 (**8.03e-5**)	2.83 (**0.0145**)
100% labels			
ST-MEM	8.11 (**7.13e-5**)	8.16 (**5.29e-7**)	5.76 (**0.0022**)
HeartLang	5.24 (**0.0001**)	6.40 (**5.43e-6**)	3.75 (**0.0073**)
LVCG	5.23 (**0.0002**)	6.57 (**0.0003**)	4.13 (**0.0098**)
ShapeEFM	3.35 (**3.47e-5**)	5.00 (**4.89e-6**)	2.26 (**0.0170**)

Table 19: Lead–time position-encoding ablation (AUROC, mean ± std, %) over five seeds.

Variant	PTBXL-Super	PTBXL-Sub	PTBXL-Form	PTBXL-Rhythm	CPSC2018	CSN
1% labels						
w/o temporal encoding	82.93 ± 0.75	72.46 ± 1.55	58.29 ± 2.28	68.27 ± 3.37	81.82 ± 0.81	67.42 ± 0.88
w/o lead encoding	82.00 ± 0.68	72.55 ± 0.92	62.73 ± 1.11	79.21 ± 2.48	80.91 ± 2.85	72.00 ± 1.12
w/o both	80.55 ± 0.19	69.73 ± 1.79	54.84 ± 2.85	65.63 ± 2.74	78.69 ± 0.59	65.72 ± 1.44
Learnable lead emb.	83.27 ± 0.35	74.66 ± 1.88	62.67 ± 1.56	76.45 ± 3.35	82.43 ± 1.21	76.74 ± 1.09
Dower-coordinate MLP	84.06 ± 0.35	76.03 ± 1.24	66.99 ± 1.67	78.85 ± 4.05	**85.69 ± 1.55**	78.84 ± 0.83
Spherical harmonics (full)	**84.13 ± 0.69**	**78.98 ± 0.86**	**69.81 ± 2.40**	**84.42 ± 1.85**	84.86 ± 2.10	**79.92 ± 0.82**
10% labels						
w/o temporal encoding	88.19 ± 0.12	85.87 ± 1.20	76.29 ± 0.52	85.03 ± 1.78	88.66 ± 0.58	85.21 ± 1.38
w/o lead encoding	87.43 ± 0.20	85.15 ± 1.23	78.48 ± 1.04	88.67 ± 1.69	91.39 ± 0.12	85.14 ± 1.38
w/o both	85.99 ± 0.20	82.42 ± 1.19	73.28 ± 1.47	83.14 ± 1.08	86.74 ± 0.21	83.07 ± 1.49
Learnable lead emb.	88.32 ± 0.18	86.89 ± 1.04	76.66 ± 1.67	88.09 ± 2.46	**92.85 ± 0.04**	88.88 ± 0.95
Dower-coordinate MLP	88.47 ± 0.19	86.59 ± 1.17	78.69 ± 1.41	87.35 ± 1.44	92.30 ± 0.24	89.61 ± 0.79
Spherical harmonics (full)	**89.21 ± 0.18**	**88.77 ± 1.16**	**81.08 ± 1.49**	**91.28 ± 1.80**	92.84 ± 0.29	**90.05 ± 1.04**
100% labels						
w/o temporal encoding	90.51 ± 0.01	92.07 ± 0.03	83.95 ± 0.19	90.39 ± 0.10	93.48 ± 0.06	93.56 ± 0.03
w/o lead encoding	89.92 ± 0.01	90.77 ± 0.07	87.57 ± 0.21	93.23 ± 0.14	95.13 ± 0.02	94.30 ± 0.03
w/o both	88.73 ± 0.01	89.80 ± 0.04	84.41 ± 0.11	89.23 ± 0.06	92.33 ± 0.02	91.30 ± 0.02
Learnable lead emb.	90.31 ± 0.01	91.98 ± 0.07	86.21 ± 0.33	92.86 ± 0.08	95.45 ± 0.01	95.53 ± 0.03
Dower-coordinate MLP	90.64 ± 0.01	92.40 ± 0.03	87.63 ± 0.16	94.00 ± 0.24	95.13 ± 0.02	95.40 ± 0.03
Spherical harmonics (full)	**91.35 ± 0.03**	**93.17 ± 0.06**	**88.66 ± 0.07**	**94.66 ± 0.23**	**96.05 ± 0.02**	**96.37 ± 0.01**

These comparisons assess lead–direction correspondence under a common perturbation setting and the contribution of perturbations under the canonical assignment. Each variant is evaluated using a fixed pretrained checkpoint and five linear-probing seeds $\{0, 1, 2, 3, 42\}$ following Sec. 5.1. The permutation and random-direction samples remain fixed across probe runs.

G.2 Full Results Across Label Fractions

The main-text ablation tables summarize results with 1% of the training labels. Here, we report the complete results at 1%, 10%, and 100% for all six disease-classification tasks. Values are mean AUROC and sample standard deviation over the five probe seeds.

Lead–Time Position Encoding. Tab. 19 reports the effects of removing temporal encoding, lead-direction encoding, or both, together with comparisons against learnable lead embeddings and the Dower-coordinate MLP.

Table 20: Lead-direction assignment and noise ablation (AUROC, mean ± std, %) over five seeds. All variants use $\sigma_v = 0.1$ except Dower w/o noise ($\sigma_v = 0$).

Variant	PTBXL-Super	PTBXL-Sub	PTBXL-Form	PTBXL-Rhythm	CPSC2018	CSN
			1% labels			
Dower (Ours)	**84.13 ± 0.69**	**78.98 ± 0.86**	**69.81 ± 2.40**	84.42 ± 1.85	**84.86 ± 2.10**	**79.92 ± 0.82**
Shuffled Dower	82.92 ± 0.48	76.08 ± 1.02	67.33 ± 1.45	84.19 ± 2.17	84.16 ± 2.76	77.43 ± 0.72
Random Sphere	82.36 ± 0.74	74.63 ± 0.92	66.52 ± 1.66	80.29 ± 2.91	82.63 ± 2.18	77.20 ± 0.70
Dower w/o noise	83.41 ± 0.52	76.34 ± 1.17	67.46 ± 3.33	**84.47 ± 1.52**	84.63 ± 2.23	77.28 ± 0.93
			10% labels			
Dower (Ours)	**89.21 ± 0.18**	**88.77 ± 1.16**	**81.08 ± 1.49**	**91.28 ± 1.80**	**92.84 ± 0.29**	**90.05 ± 1.04**
Shuffled Dower	88.23 ± 0.09	85.92 ± 1.54	80.28 ± 2.06	90.56 ± 1.42	**92.84 ± 0.18**	88.61 ± 0.84
Random Sphere	87.57 ± 0.21	85.61 ± 1.73	79.98 ± 1.26	89.40 ± 1.15	91.70 ± 0.23	88.59 ± 0.73
Dower w/o noise	88.88 ± 0.15	87.40 ± 2.02	80.65 ± 1.14	90.70 ± 0.89	92.64 ± 0.40	88.88 ± 1.09
			100% labels			
Dower (Ours)	**91.35 ± 0.03**	**93.17 ± 0.06**	88.66 ± 0.07	94.66 ± 0.23	**96.05 ± 0.02**	**96.37 ± 0.01**
Shuffled Dower	90.78 ± 0.02	91.79 ± 0.18	88.64 ± 0.11	94.25 ± 0.07	95.89 ± 0.02	95.59 ± 0.02
Random Sphere	90.11 ± 0.04	91.60 ± 0.04	87.87 ± 0.08	93.94 ± 0.03	95.26 ± 0.01	95.05 ± 0.02
Dower w/o noise	91.20 ± 0.02	92.41 ± 0.09	**89.03 ± 0.08**	**94.88 ± 0.13**	95.72 ± 0.02	95.63 ± 0.02

Lead–Direction Correspondence. Tab. 20 compares canonical, shuffled, and random direction assignments, together with the canonical assignment without direction perturbations.

Spherical-Harmonic Degree. Tab. 21 reports results for maximum degrees $L \in \{0, 1, 2, 3\}$, showing how the angular expressiveness of the lead representation affects performance across label fractions.

Table 21: Spherical-harmonic degree ablation (AUROC, mean ± std, %) over five seeds.

Variant	PTBXL-Super	PTBXL-Sub	PTBXL-Form	PTBXL-Rhythm	CPSC2018	CSN
			1% labels			
$L = 0$ (1 basis)	81.62 ± 0.78	76.25 ± 0.41	63.62 ± 1.82	80.54 ± 1.83	83.65 ± 1.44	71.41 ± 0.95
$L = 1$ (4 bases)	82.59 ± 0.65	76.47 ± 0.72	66.20 ± 3.38	82.44 ± 1.91	84.39 ± 1.69	71.74 ± 1.05
$L = 2$ (9 bases)	83.48 ± 0.76	78.68 ± 0.74	67.92 ± 1.95	83.77 ± 1.80	83.96 ± 2.34	74.14 ± 0.61
$L = 3$ (16 bases; Ours)	**84.13 ± 0.69**	**78.98 ± 0.86**	**69.81 ± 2.40**	**84.42 ± 1.85**	**84.86 ± 2.10**	**79.92 ± 0.82**
			10% labels			
$L = 0$ (1 basis)	86.54 ± 0.28	84.63 ± 1.23	78.08 ± 2.47	87.31 ± 0.72	90.92 ± 0.31	87.89 ± 0.53
$L = 1$ (4 bases)	87.46 ± 0.11	86.63 ± 1.70	79.94 ± 2.04	88.54 ± 0.83	91.18 ± 0.30	87.30 ± 1.24
$L = 2$ (9 bases)	88.68 ± 0.16	88.35 ± 1.12	79.65 ± 1.60	89.92 ± 1.53	92.13 ± 0.29	89.63 ± 0.94
$L = 3$ (16 bases; Ours)	**89.21 ± 0.18**	**88.77 ± 1.16**	**81.08 ± 1.49**	**91.28 ± 1.80**	**92.84 ± 0.29**	**90.05 ± 1.04**
			100% labels			
$L = 0$ (1 basis)	89.10 ± 0.02	90.04 ± 0.04	85.61 ± 0.12	92.32 ± 0.33	94.07 ± 0.02	93.56 ± 0.04
$L = 1$ (4 bases)	89.88 ± 0.02	91.64 ± 0.05	87.86 ± 0.22	92.98 ± 0.11	94.41 ± 0.03	94.79 ± 0.04
$L = 2$ (9 bases)	90.92 ± 0.03	92.78 ± 0.04	88.30 ± 0.07	94.50 ± 0.06	95.50 ± 0.04	95.84 ± 0.01
$L = 3$ (16 bases; Ours)	**91.35 ± 0.03**	**93.17 ± 0.06**	**88.66 ± 0.07**	**94.66 ± 0.23**	**96.05 ± 0.02**	**96.37 ± 0.01**