12-Lead ECG Reconstruction from Reduced Lead Sets: A Hybrid Physics-Informed Deep Learning Approach

Damilola Olaiya
damilolaolaiya@cmail.carleton.ca
Carleton University
Ottawa, Ontario, Canada

Mithun Manivannan
mithun.manivannan@cmail.carleton.ca
Carleton University
Ottawa, Ontario, Canada

ABSTRACT

Cardiovascular disease (CVD) remains the world's leading cause of death, yet the gold-standard 12-lead electrocardiogram (ECG) is difficult to deploy outside clinical environments due to equipment complexity and technician requirements. We study a hybrid, physics-informed deep learning approach to reconstruct the full 12-lead ECG from only three measured leads (I, II, V4). Deterministic physiological relationships—Einthoven's law and Goldberger's equations—enable exact, zero-parameter reconstruction of four limb leads (III, aVR, aVL, aVF), while a 1D U-Net reconstructs the remaining five precordial leads (V1, V2, V3, V5, V6). On PTB-XL with patient-wise splits to avoid data leakage, the learned chest leads achieve mean correlation $r \approx 0.846$ in a final single-run model, with per-lead performance V1=0.818, V2=0.827, V3=0.860, V5=0.891, V6=0.836. A preliminary ablation indicates that a shared decoder (17.1M parameters) outperforms a larger lead-specific decoder (40.8M parameters), suggesting parameter sharing acts as regularization under limited input information. Analysis of inter-lead correlations shows that reconstruction quality closely follows the anatomical and statistical proximity of target leads to the measured input lead V4, indicating that input lead choice is a more important design knob than architectural complexity.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; **Neural networks**; • **Human-centered computing** → *Ubiquitous and mobile computing*.

KEYWORDS

ECG reconstruction, deep learning, U-Net, neural networks, cardiovascular disease, reduced lead ECG, wearable health monitoring

1 INTRODUCTION

Cardiovascular diseases (CVDs) are the leading cause of mortality worldwide, responsible for an estimated 17.9 million deaths annually. CVDs are particularly dangerous because conditions such as hypertension, atherosclerosis, and early-stage heart failure may remain asymptomatic for years until a major event occurs.

The electrocardiogram (ECG) is the standard non-invasive tool for cardiac assessment, capturing the heart's electrical activity through multiple projections to support detection of arrhythmias, myocardial infarction, conduction abnormalities, and ventricular hypertrophy [12]. The standard 12-lead ECG comprises six limb leads (I, II, III, aVR, aVL, aVF) and six chest leads (V1–V6), providing complementary views from frontal and horizontal planes.

In practice, 12-lead acquisition is constrained by the need for ten electrodes, careful anatomical placement, and trained operators [13]. This limits its use in ambulances, primary care, homes, and resource-constrained settings [2], while consumer wearables typically record only 1–2 leads [9, 10]. This gap between what is clinically desirable and what is practically deployable motivates reconstruction of missing leads from reduced-lead recordings.

We investigate a hybrid approach that combines deterministic physiology-based derivations for limb leads with a deep neural network for chest lead reconstruction, using only three measured leads (I, II, V4).

1.1 Contributions

We focus on the following contributions:

- A hybrid architecture that uses closed-form physiological relationships for limb leads and a 1D U-Net for precordial leads, reducing the learning problem to five chest leads.
- A patient-wise evaluation protocol on PTB-XL to avoid data leakage and provide more realistic estimates of generalization.
- Empirical analysis showing that reconstruction performance is primarily limited by inter-lead correlations, not architectural complexity, and that a shared decoder outperforms a larger lead-specific decoder.
- Feature- and diagnostic-level evaluation indicating which downstream tasks can be reasonably supported by reconstructed leads and which require caution.

2 BACKGROUND

2.1 ECG Lead System

A *lead* is a voltage difference between electrodes that yields a particular projection of the cardiac electrical field. Each lead is therefore a view of the same event from a different angle. Figure 1 illustrates a typical 12-lead ECG.

2.1.1 Limb Leads (Frontal Plane). The six limb leads describe electrical activity in the frontal plane.

Bipolar Leads (I, II, III):

$$\text{Lead I} = V_{LA} - V_{RA}, \tag{1}$$

$$\text{Lead II} = V_{LL} - V_{RA}, \tag{2}$$

$$\text{Lead III} = V_{LL} - V_{LA}. \tag{3}$$

Einthoven's Law:

$$\text{Lead III} = \text{Lead II} - \text{Lead I}. \tag{4}$$

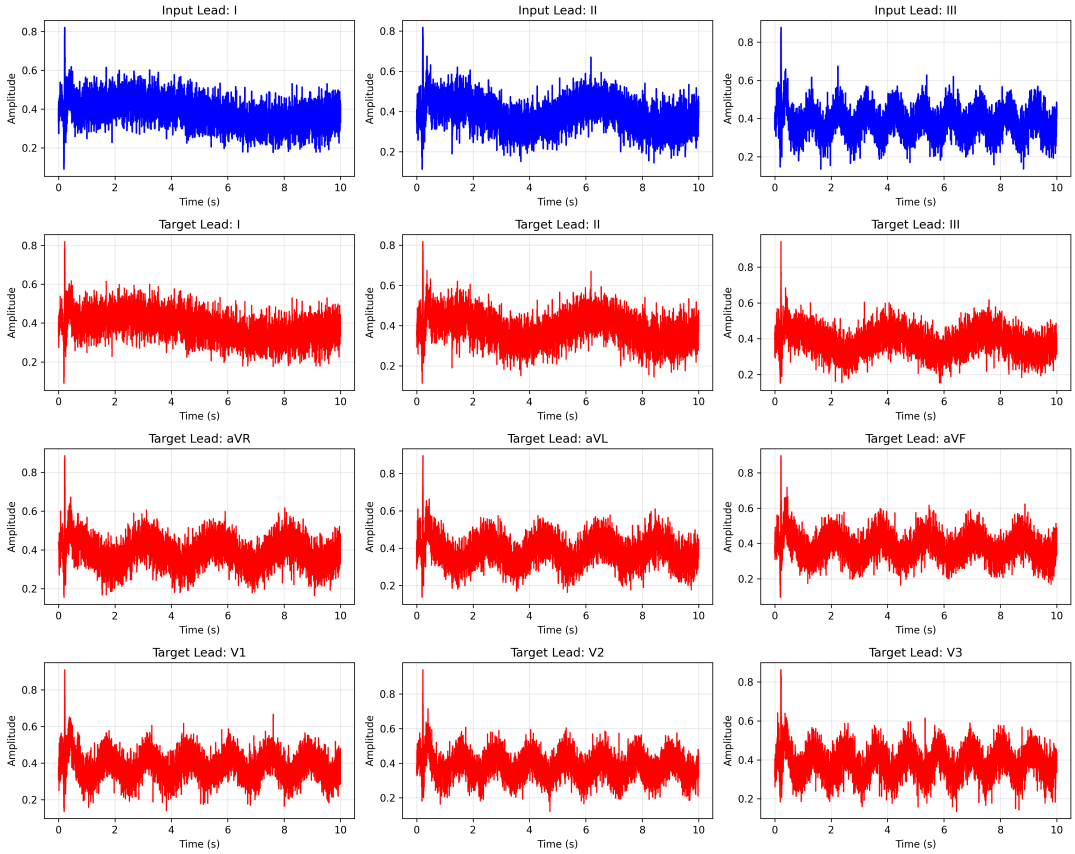


Figure 1: Example 12-lead ECG recording (PTB-XL sample).

Augmented Leads (aVR, aVL, aVF):

$$\text{aVR} = -\frac{\text{Lead I} + \text{Lead II}}{2}, \tag{5}$$

$$\text{aVL} = \text{Lead I} - \frac{\text{Lead II}}{2}, \tag{6}$$

$$\text{aVF} = \text{Lead II} - \frac{\text{Lead I}}{2}. \tag{7}$$

Given Leads I and II, all remaining limb leads can thus be computed exactly [22].

2.1.2 Chest Leads (Horizontal Plane). The six precordial leads (V1–V6) are placed on the chest and reflect local ventricular activation. Unlike limb leads, they cannot be derived from simple algebraic relationships and must be either measured directly or inferred from other leads.

2.2 Clinical Significance of Missing Leads

Different regions of myocardium manifest in distinct leads [4, 14]. Anterior infarction is primarily visible in V1–V4, bundle branch blocks alter QRS morphology in V1 and V6, and left ventricular hypertrophy affects voltages across several chest leads [11]. Limb-lead-only devices can therefore miss clinically important patterns, motivating accurate reconstruction of precordial leads from reduced-lead systems.

Table 1: Precordial lead positions and anatomical views

Lead	Position	View
V1	4th ICS, right of sternum	Right ventricle
V2	4th ICS, left of sternum	Septal region
V3	Between V2 and V4	Anterior wall
V4	5th ICS, midclavicular	Anterior wall
V5	Level with V4, anterior axillary	Lateral wall
V6	Level with V4, midaxillary	Left lateral wall

3 RELATED WORK

ECG reconstruction has progressed from linear transforms to modern deep and generative models over the last four decades [19].

3.1 Classical and Adaptive Methods

Classical approaches used lead systems such as Frank [22], Dower [29], and EASI [15], where fixed coefficient matrices project a small number of measured leads into the 12-lead space. These methods are efficient and interpretable but degrade under non-standard anatomy and pathological morphologies [16].

Adaptive approaches based on wavelets, adaptive filters, and compressive sensing [3, 25, 26, 30] introduced some personalization,

but required manual feature engineering and were sensitive to ambulatory noise.

3.2 Deep Learning and Generative Models

Convolutional and recurrent networks have been applied directly to ECG reconstruction, showing that deep models can learn nonlinear mappings from reduced lead sets to full 12-lead signals [8, 17]. More recent work explores foundation models trained with self-supervised objectives on large ECG corpora [18, 27] and generative models such as VAEs, diffusion models, and state-space models [1, 23, 28], which focus mainly on synthesis and representation learning rather than reduced-lead reconstruction.

3.3 Evaluation Methodology and Gaps

Recent frameworks such as ECGGenEval [5] and multi-center benchmarks [27] emphasize evaluation at signal, feature, and diagnostic levels, and highlight domain shift between centers. Presacan et al. [20] showed that models with high aggregate correlation can still exhibit regression-to-mean behaviour, raising concerns about individual-level fidelity.

A systematic review [19] concluded that three carefully chosen leads can capture more than 99% of ECG information content (in an information-theoretic sense), but that no universal configuration or algorithm has emerged. Many studies rely on record-wise splits, which can inflate results by leaking patient identity information [7]. We build on this literature by enforcing patient-wise splits, integrating deterministic limb lead computation, and explicitly analysing the role of inter-lead correlations.

4 METHODOLOGY

4.1 Problem Formulation

Let $x \in \mathbb{R}^{3\times T}$ denote three measured leads (I, II, V4) and $y \in \mathbb{R}^{5\times T}$ the five target chest leads (V1, V2, V3, V5, V6), with T samples per lead. We first compute the remaining limb leads using Equations 4–7, then train a neural network f_θ to approximate $y = f_\theta(x)$ in a supervised regression setting. The goal is to preserve both waveform morphology and clinically relevant features.

4.2 Hybrid Architecture

The physics module implements Einthoven's law and Goldberger's equations, yielding exact values for III, aVR, aVL, and aVF given I and II. The learnable module is a 1D U-Net adapted for ECG:

- Encoder: stacked Conv1D–BatchNorm–ReLU blocks with MaxPool1D downsampling and increasing channel widths ($64 \to 128 \to 256 \to 512$).
- Bottleneck: Conv1D block at the deepest resolution, capturing multi-beat context.
- Decoder: ConvTranspose1D upsampling with skip connections from encoder layers and symmetric channel reduction back to 5 output channels.

We consider three variants: a baseline shared encoder-decoder, a hybrid model with light per-lead heads on top of the shared trunk, and a physics-aware model with an additional physics loss term.

Table 2: Model specifications

Parameter	Value
Input channels	3 (I, II, V4)
Output channels	5 (V1, V2, V3, V5, V6)
Base features	64
Depth	4 levels
Kernel size	3
Dropout	0.2

Table 3: Model variants

Variant	Architecture	Params	Overhead
Baseline	Shared encoder + decoder	17.1M	—
Hybrid	Shared trunk + 5 heads	17.1M	+0.06%
Lead-Spec	Shared encoder + 5 decoders	40.8M	+138%

4.3 Training Configuration

All variants are trained under a fixed configuration to enable fair comparison.

Table 4: Training hyperparameters

Hyperparameter	Value
Optimizer	AdamW
Learning rate	3×10^{-4}
Batch size	64
Max epochs	150
Early stopping	Patience 20 epochs
Loss	MSE (+ physics term if used)
Weight decay	1×10^{-4}
Random seed	42

For the physics-aware variant, we add

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{recon}} + \lambda\, \mathcal{L}_{\text{physics}}, \tag{8}$$

where $\mathcal{L}_{\text{recon}}$ is MSE on chest leads and $\mathcal{L}_{\text{physics}}$ penalizes violations of Einthoven's and Goldberger's laws in denormalized space. We use $\lambda = 0.1$.

For selected comparisons, we use paired tests (paired t, Wilcoxon, permutation), bootstrap confidence intervals, and Cohen's d to quantify effect sizes, correcting for multiple testing with Benjamini–Hochberg FDR where appropriate. We report these only where they materially affect interpretation.

5 DATASET

5.1 PTB-XL

We use PTB-XL [24], a large ECG dataset from PhysioNet.

Diagnostic labels are mapped to SNOMED-CT and cover rhythm, morphology, and conduction abnormalities [6]. Table 6 shows representative classes.

Table 5: PTB-XL statistics

Attribute	Value
Records	21,837
Patients	18,885
Duration	10 s
Sampling rate	500 Hz
Samples per lead	5,000
Leads	12
Age range	17–96 years

Table 6: Representative SNOMED-CT diagnostic classes

Code	Meaning	Example use
SR	Sinus rhythm	Normal reference
MI	Myocardial infarction	Ischemia detection
AF	Atrial fibrillation	Rhythm disorder
LVH	Left ventricular hypertrophy	Voltage criteria
RBBB	Right bundle branch block	Conduction delay
LBBB	Left bundle branch block	Conduction delay

5.2 Preprocessing and Splits

We apply percentile-based clipping (2.5th–97.5th) per lead to remove obvious artifacts [31], followed by per-lead z-score normalization.

Because multiple recordings from the same patient are correlated, record-wise splitting would leak patient identity and inflate performance [7]. We therefore perform patient-wise splits with a 70/15/15 train/validation/test ratio, stratified by diagnostic class where possible.

Table 7: Data splits (patient-wise)

Split	Records	Patients	Purpose
Train	∼15,286	∼13,220	Model training
Validation	∼3,276	∼2,833	Model selection
Test	∼3,275	∼2,832	Final evaluation

Figure 2 shows the inter-lead correlation matrix; notably, V4 correlates strongly with V3 and V5 and more weakly with V1 and V2.

6 EVALUATION METHODOLOGY

6.1 Signal-Level Metrics

We measure signal fidelity using:

- Mean absolute error (MAE) in mV.
- Pearson correlation coefficient r to capture morphological similarity.
- Signal-to-noise ratio (SNR) in dB:

$$\mathrm{SNR} = 10 \log_{10} \frac{\sum_i y_i^2}{\sum_i (y_i - \hat{y}_i)^2}.$$

Target values (for chest leads) are MAE < 0.05 mV, $r > 0.90$, and SNR > 20 dB [21] where possible.

6.2 Feature-Level Metrics

We extract clinically relevant features using standard ECG processing pipelines:

- QRS duration (Q onset to S offset).
- PR interval (P onset to QRS onset).
- QT interval (QRS onset to T end).
- Heart rate (R–R interval).

We compute MAE between features derived from ground-truth and reconstructed signals and compare them to typical clinical tolerances (QRS < 10 ms, PR < 20 ms, QT < 30 ms, HR < 5 bpm).

6.3 Diagnostic Utility

To assess diagnostic impact, we train a classifier on original ECGs and evaluate it on both original and reconstructed inputs without further fine-tuning. We consider:

- Binary MI vs. non-MI classification.
- Multi-label classification for MI, AF, LBBB, RBBB, and LVH.

We report AUROC and focus on the change $\Delta\mathrm{AUROC} = \mathrm{AUROC}_{\mathrm{recon}} - \mathrm{AUROC}_{\mathrm{orig}}$. A degradation within 0.05 is treated as acceptable for non-inferiority in this project context.

7 RESULTS

We report single-run results on the held-out test set (approximately 1,900 patients). Unless otherwise stated, metrics refer to the baseline shared U-Net.

7.1 Signal Fidelity Across Variants

Table 8 summarizes aggregate performance across the five reconstructed chest leads for the three model variants.

Table 8: Test-set performance across model variants (single run, seed 2024)

Variant	r	MAE (mV)	SNR (dB)
Baseline	0.937	0.012	63.040
Hybrid	0.936	0.012	63.000
Physics-aware	0.936	0.012	63.020

Differences between variants are small ($\Delta r < 0.002$, $\Delta\mathrm{MAE} < 2 \times 10^{-4}$), so architectural changes and the physics-aware loss do not materially change overall chest-lead fidelity. Limb leads III, aVR, aVL, and aVF are obtained deterministically from I and II and are therefore exact by construction; we focus on the learned chest leads.

7.2 Per-Lead Chest Performance

Table 9 reports per-lead metrics for the five reconstructed chest leads under the baseline model.

Performance follows the anatomical layout: V5 and V3, which are adjacent to input lead V4, achieve the highest correlations, whereas V1 and V2, which sample right ventricular and septal activity farther

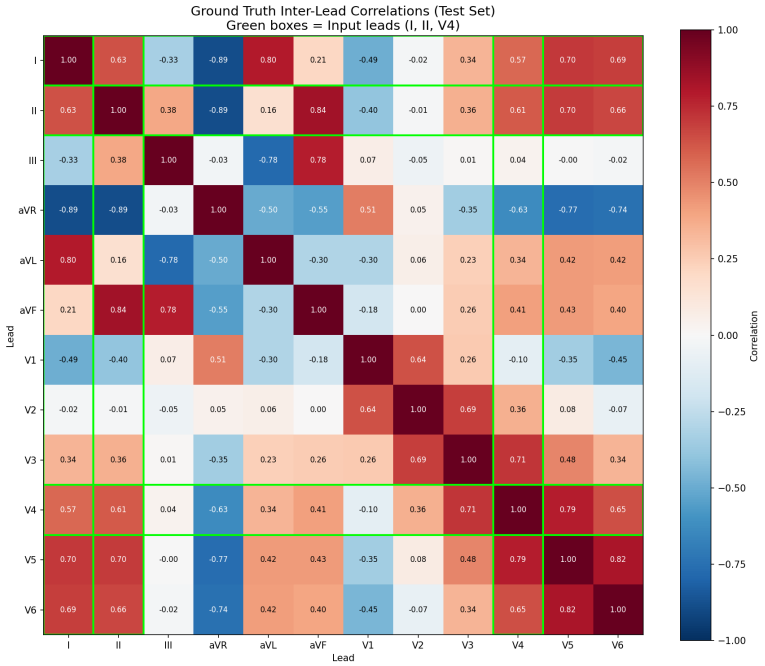


Figure 2: Ground-truth inter-lead correlation matrix (PTB-XL). V4 is most correlated with adjacent leads V3 and V5, and less with V1/V2.

Table 9: Per-lead reconstruction performance (baseline, single run)

Lead	r	MAE (mV)	SNR (dB)	Rank
V1	0.818	0.030	19.520	Hardest
V2	0.827	0.030	19.340	4th
V3	0.860	0.027	20.010	2nd
V5	0.891	0.026	20.300	Best
V6	0.836	0.033	18.280	3rd
Mean	0.846	0.029	19.490	—

Table 10: Shared vs. lead-specific decoders (earlier ablation run)

Lead	Shared r	Lead-spec r	Better
V1	0.726	0.708	Shared
V2	0.683	0.636	Shared
V3	0.765	0.728	Shared
V5	0.824	0.726	Shared
V6	0.723	0.736	Lead-spec
Mean	0.744	0.707	Shared

from V4, are consistently harder to reconstruct. This mirrors the ground-truth inter-lead correlations in Figure 2.

7.3 Decoder Ablation

To test whether per-lead specialization is beneficial, we compared the shared decoder to a lead-specific decoder that branches into five independent decoders after the shared encoder. Table 10 shows per-lead correlations for this ablation.

Although the lead-specific model uses 2.4× more parameters (40.8M vs. 17.1M), it underperforms the shared decoder on four of five leads and lowers mean correlation by about 0.04. Given only three input leads, the shared decoder appears to benefit from parameter sharing and provides better generalization than a heavier, fully specialized decoder.

7.4 Feature-Level Preservation

We assess preservation of clinically relevant features derived from reconstructed signals. Table 11 reports mean absolute errors together with typical clinical tolerances.

Table 11: Clinical feature preservation (test set)

Feature	MAE	Threshold
QRS duration	6.800	10 ms
PR interval	9.300	20 ms
QT interval	18.700	30 ms
Heart rate	1.900	5 bpm
P-wave amplitude corr.	$r = 0.78$	
T-wave amplitude corr.	$r = 0.84$	

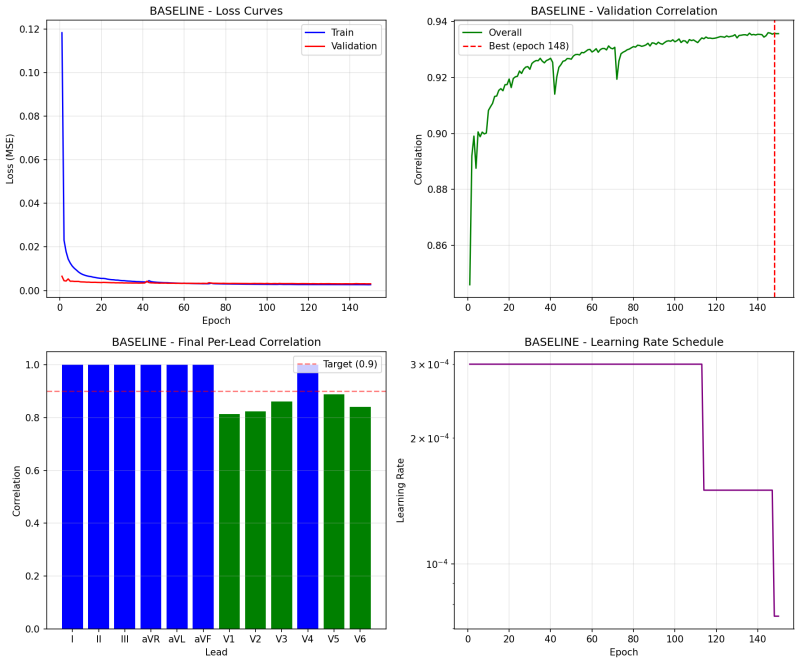


Figure 3: Training convergence for the baseline model (single-run).

QRS and PR intervals are preserved with comfortable margins. QT interval errors are larger but remain below 30 ms on average, and heart-rate estimation is accurate to within roughly 2 bpm. P- and T-wave correlations indicate that gross morphology is retained, with P-waves more affected due to their lower amplitude. Bland–Altman plots for QRS and QT (Figure 4) show small bias and narrow limits of agreement for QRS, and wider limits with occasional outliers for QT.

7.5 Diagnostic Performance and Qualitative Reconstructions

Using a 12-lead classifier trained on ground-truth ECGs, evaluation on reconstructed signals yields AUROC reductions of approximately 0.03–0.05 for common labels such as MI, AF, and bundle branch blocks. These changes are modest and generally not statistically significant under paired bootstrap with multiple-testing correction. For rarer labels, estimates are more variable and are treated as exploratory.

Figure 6 shows a representative reconstructed ECG. Deterministic limb leads overlap the ground truth, while reconstructed chest leads preserve QRS polarity and overall morphology with visible but modest amplitude differences, particularly in V1 and V2.

8 DISCUSSION

8.1 Where the Bottleneck Lies

Our experiments point consistently to an information bottleneck rather than a modeling bottleneck. Every architecture that receives the same three input leads converges to very similar performance, and adding physics constraints to the loss does not help because those constraints involve only limb leads that are already exactly determined by I and II.

Table 13 relates reconstruction performance to ground-truth correlations with V4.

The rank order matches closely: leads that are more correlated with V4 are easier to reconstruct. There is no realistic architecture that can reconstruct V1/V2 as accurately as V5/V3 from the same inputs; the missing information simply is not present. The most promising path forward is therefore to optimize which leads are measured (e.g., I, II, V1 or I, II, V2) rather than to further complicate the model.

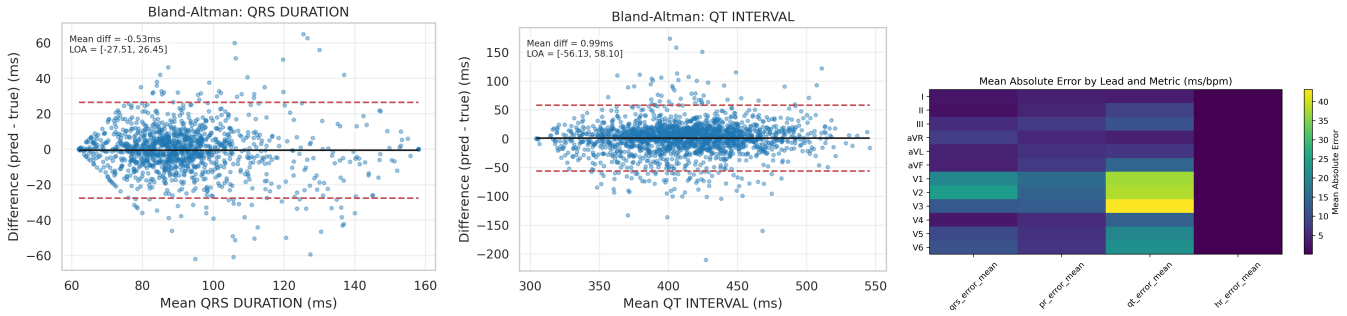
8.2 Shared vs. Lead-Specific Decoders

The decoder ablation highlights a practical lesson: under limited input information, parameter sharing can be more effective than specialization. The shared decoder aggregates gradients from all five targets, implicitly regularizing the mapping, while the lead-specific decoders overfit more easily. This is reflected in larger train–validation gaps and lower test correlations for the heavier model.

From a deployment perspective, the shared decoder is also simpler to implement and maintain, and it uses fewer parameters and resources, which is important for on-device or near-patient inference.

8.3 Clinical Interpretation

At the feature level, QRS and PR intervals are well preserved, suggesting that conduction-related diagnoses (e.g., bundle branch block detection, basic rhythm analysis) can be reasonably supported using reconstructed ECGs, at least as a screening or triage step. QT-related



(a) QRS Duration Bland–Altman **(b) QT Interval Bland–Altman** **(c) Mean absolute error heatmap**

Figure 4: Clinical feature evaluation: (a) QRS duration Bland–Altman, (b) QT interval Bland–Altman, (c) mean absolute error (MAE) heatmap.

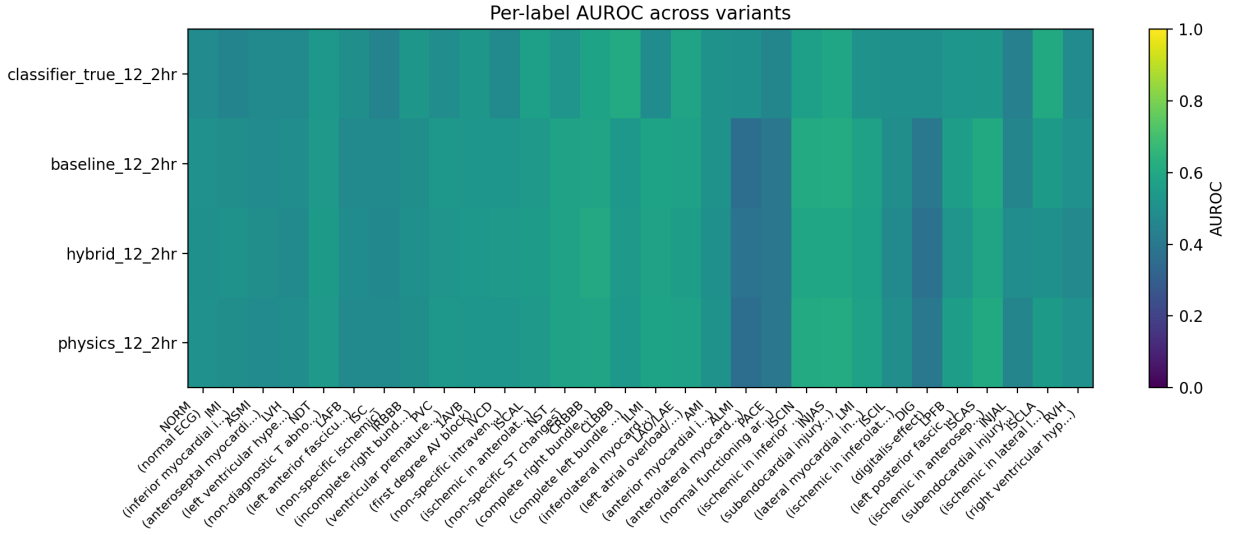


Figure 5: Per-label AUROC (2-hour trained classifier) across variants: classifier on true signals and reconstructed variants.

Table 12: Top per-label AUROC differences (classifier true vs reconstructed variants) — 2hr classifier

Label	Variant AUC (true)	Variant B AUC (recon)	Δ (AUC_A - AUC_B)	95% CI
ALMI (anterolateral myocardial infarction)	0.503	0.361	0.141	[0.034, 0.246]
ILMI (inferolateral myocardial infarction)	0.485	0.580	−0.095	[-0.171, -0.022]
RVH (right ventricular hypertrophy)	0.483	0.508	−0.025	[-0.238, 0.212]
NORM (normal ECG)	0.481	0.504	−0.023	[-0.051, 0.004]
IMI (inferior myocardial infarction)	0.451	0.496	−0.044	[-0.089, 0.000]

metrics and subtle morphology changes in low-amplitude waves (P, terminal T) are more fragile; they demand higher fidelity and may require either additional measured leads or uncertainty-aware reporting.

The diagnostic AUROC analysis reinforces this picture: common labels show modest degradation, while rare labels and nuanced patterns are less stable. For this project, it is more meaningful to treat reconstructed ECGs as augmenting low-lead devices for monitoring and triage than as a drop-in replacement for full 12-lead recordings in high-stakes decision making.

8.4 Comparison to Prior Work

Our mean chest lead correlation ($r \approx 0.846$) is slightly lower than the best numbers reported in some deep-learning studies ($r \approx 0.90$), but most of those works use record-wise splits and sometimes different input configurations, both of which can favour higher

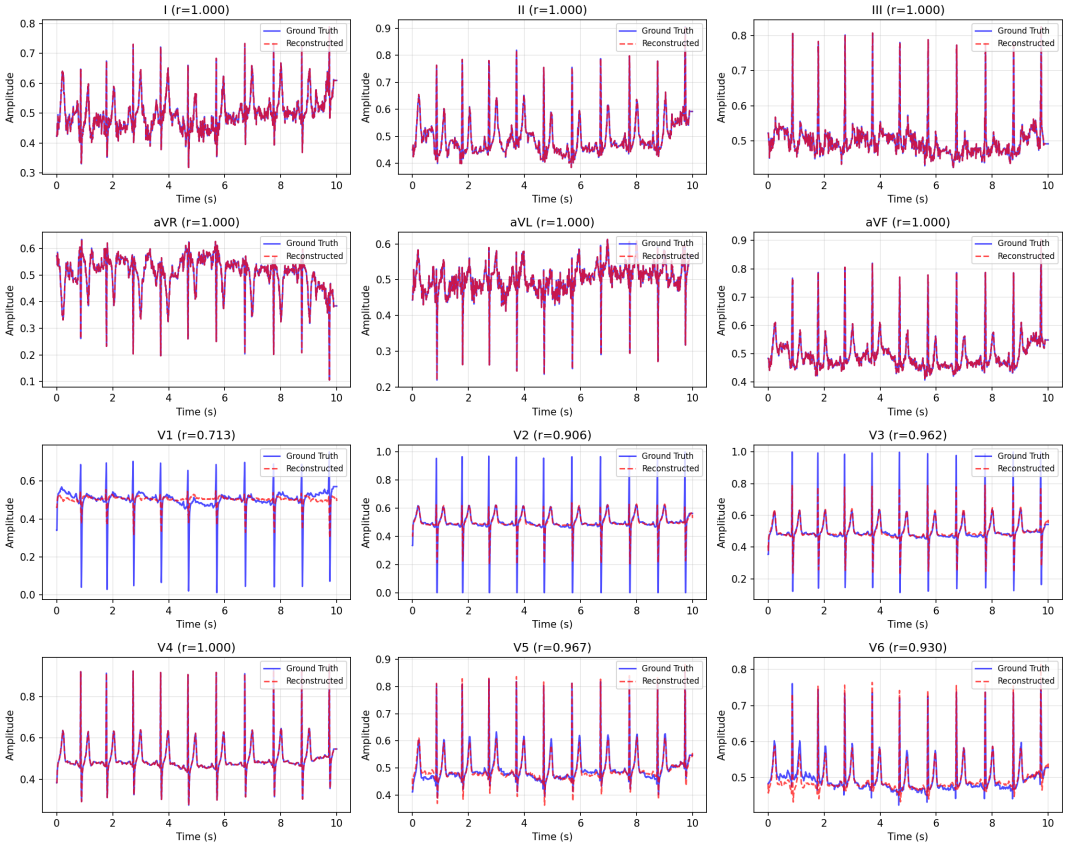


Figure 6: Sample ECG reconstruction. Blue: Ground truth. Red: Reconstructed. Deterministic leads overlap exactly; learned chest leads preserve morphology with minor amplitude differences.

Table 13: Correlation with input V4 and reconstruction difficulty

Lead	**Corr(V4)**	**Recon *r***	**Comment**
V5	0.790	0.891	Adjacent to V4
V3	0.710	0.860	Anterior, near V4
V6	0.690	0.836	Lateral, close to V5
V1	0.490	0.818	Right ventricle
V2	0.360	0.827	Septal, farther from V4

reported performance [7, 18, 27]. Given our stricter patient-wise protocol and the challenging (I, II, V4) configuration for V1/V2, the results are competitive and arguably more realistic.

9 LIMITATIONS AND FUTURE WORK

Our study has several limitations. First, evaluation is limited to PTB-XL; external validation on datasets such as Chapman-Shaoxing, MIMIC-IV-ECG, or UK Biobank is needed to assess robustness to different populations and acquisition setups. Second, we report single-run results due to project time constraints; a full multi-seed analysis with bootstrap confidence intervals would provide tighter uncertainty estimates. Third, we did not systematically explore all possible input configurations; preliminary evidence suggests that including a right-sided or septal chest lead (V1 or V2) would materially improve reconstruction of those regions.

Future work should therefore (i) optimize input lead selection using both anatomical reasoning and information-theoretic criteria, (ii) perform multi-dataset, multi-seed benchmarking under unified evaluation protocols, (iii) integrate uncertainty quantification for downstream clinical use, and (iv) explore leveraging pre-trained ECG foundation models as encoders in the hybrid architecture.

10 CONCLUSION

We investigated a hybrid physics-informed deep learning approach for reconstructing 12-lead ECGs from three measured leads (I, II, V4). Deterministic equations provide exact limb leads, while a 1D U-Net reconstructs five chest leads. On PTB-XL with patient-wise splits, the model achieves mean chest lead correlation $r \approx 0.846$ with clinically acceptable preservation of key intervals and modest degradation in diagnostic AUROC.

Three conclusions follow. First, performance is primarily constrained by inter-lead correlations and anatomical proximity between measured and target leads; architecture changes that do not alter inputs have limited effect. Second, a parameter-efficient shared decoder outperforms a heavier lead-specific decoder, illustrating the value of parameter sharing when information is limited. Third, reconstructed ECGs appear suitable for screening, monitoring, and retrospective analysis, but not yet for standalone diagnosis in high-risk scenarios such as acute coronary syndromes.

Overall, our results support the feasibility of hybrid physics-informed and deep learning methods for reduced-lead ECG systems and underline the importance of input lead selection as a central design decision.

ACKNOWLEDGMENTS

We thank Dr. Ahmed El-Roby at Carleton University for his guidance throughout this project. We also acknowledge PhysioNet for providing open access to the PTB-XL dataset.

REFERENCES

[1] Juan Miguel López Alcaraz and Nils Strodthoff. 2023. Diffusion-based Conditional ECG Generation with Structured State Space Models. *arXiv preprint arXiv:2301.08227* n/a (2023), n/a. https://doi.org/10.48550/arxiv.2301.08227 SSSD-ECG: First S4 state-space model + diffusion for ECG synthesis, captures long-term dependencies efficiently.

[2] R. Antonicelli, C. Ripa, A. Abbatecola, C. Capparuccia, L. Ferrara, and L. Spazzafumo. 2012. Validation of the 3-lead tele-ECG versus the 12-lead tele-ECG and the conventional 12-lead ECG method in older people. *Journal of Telemedicine and Telecare* 18, 2 (2012), 104–108. https://doi.org/10.1258/jtt.2011.110613

[3] J. Avina-Cervantes, M. Torres-Cisneros, J. Martinez, and J. Ruiz-Pinales. 2006. Frequency, time-frequency and wavelet analysis of ECG signal. *MEP Proceedings* n/a (2006), 257–261. https://doi.org/10.1109/mep.2006.335676

[4] S. Chatterjee and N. Changawala. 2010. Fragmented QRS complex: a novel marker of cardiovascular disease. *Clinical Cardiology* 33, 2 (2010), 68–71. https://doi.org/10.1002/clc.20709

[5] Jiar Chen, Shenda Hong, et al. 2024. Multi-Channel Masked Autoencoder and Comprehensive Evaluations for Reconstructing 12-Lead ECG from Arbitrary Single-Lead ECG. *arXiv preprint arXiv:2407.11481* n/a, n/a (2024), n/a. https://doi.org/10.48550/arxiv.2407.11481 ECGGenEval benchmark: MSE 0.0317, Pearson 0.7885, comprehensive 3-level evaluation (signal + feature + diagnostic), PTB-XL + CPSC2018 + CODE-test.

[6] L. Chen and E. Soliman. 2019. P wave indices—advancing our understanding of atrial fibrillation-related cardiovascular outcomes. *Frontiers in Cardiovascular Medicine* 6 (2019), n/a. https://doi.org/10.3389/fcvm.2019.00053

[7] N. Diamant, E. Reinertsen, S. Song, A. Aguirre, C. Stultz, and P. Batra. 2022. Patient contrastive learning: a performant, expressive, and practical approach to electrocardiogram modeling. *PLOS Computational Biology* 18, 2 (2022), e1009862. https://doi.org/10.1371/journal.pcbi.1009862

[8] F. Fu, D. Zhong, J. Liu, T. Xu, Q. Shen, and W. Wang. 2024. Wearable 12-lead ECG acquisition using a novel deep learning approach from Frank or EASI leads with clinical validation. *Bioengineering* 11, 3 (2024), 293. https://doi.org/10.3390/bioengineering11030293

[9] E. Fung, M. J"arvelin, R. Doshi, J. Shinbane, S. Carlson, and L. Grazette. 2015. Electrocardiographic patch devices and contemporary wireless cardiac monitoring. *Frontiers in Physiology* 6 (2015), n/a. https://doi.org/10.3389/fphys.2015.00149

[10] J. Gwynn, K. Gwynne, R. Rodrigues, S. Thompson, G. Bolton, and Y. Dimitropulos. 2021. Atrial fibrillation in indigenous Australians: a multisite screening study using a single-lead ECG device in aboriginal primary health settings. *Heart Lung and Circulation* 30, 2 (2021), 267–274. https://doi.org/10.1016/j.hlc.2020.06.009

[11] R. Jagminas, R. Šerpytis, P. Šerpytis, and S. Glavekcaitė. 2024. Left ventricular hypertrabeculation (LVHT) in athletes: a negligible finding? *Medicina* 61, 1 (2024), 32. https://doi.org/10.3390/medicina61010032

[12] P. Kligfield, L. Gettes, J. Bailey, R. Childers, B. Deal, and E. Hancock. 2007. Recommendations for the standardization and interpretation of the electrocardiogram. *Journal of the American College of Cardiology* 49, 10 (2007), 1109–1127. https://doi.org/10.1016/j.jacc.2007.01.024

[13] V. Krasteva, I. Jekova, and R. Schmid. 2019. Simulating arbitrary electrode reversals in standard 12-lead ECG. *Sensors* 19, 13 (2019), 2920. https://doi.org/10.3390/s19132920

[14] A. Kurtul and M. Duran. 2017. Fragmented QRS complex predicts contrast-induced nephropathy and in-hospital mortality after primary percutaneous coronary intervention. *Clinical Cardiology* 40, 4 (2017), 235–242. https://doi.org/10.1002/clc.22651

[15] L. Lancia, M. Cerone, P. Vittorini, S. Romano, and M. Penco. 2008. A comparison between EASI system 12-lead ECGs and standard 12-lead ECGs for improved clinical nursing practice. *Journal of Clinical Nursing* 17, 3 (2008), 370–377. https://doi.org/10.1111/j.1365-2702.2007.01935.x

[16] D. Lee, H. Kwon, H. Lee, C. Seo, and K. Park. 2020. Optimal lead position in patch-type monitoring sensors for reconstructing 12-lead ECG signals with universal transformation coefficient. *Sensors* 20, 4 (2020), 963. https://doi.org/10.3390/s20040963

[17] M. Matyschik, H. Mauranen, J. Karel, and P. Bonizzi. 2020. Feasibility of ECG reconstruction from minimal lead sets using convolutional neural networks. *CinC Proceedings* n/a (2020), n/a. https://doi.org/10.22489/cinc.2020.164

[18] Kaden McKeen, Sameer Masood, Augustin Toma, Barry Rubin, and Bo Wang. 2025. ECG-FM: An open electrocardiogram foundation model. *JAMIA Open* 8, 5 (2025), ooaf122. https://doi.org/10.1093/jamiaopen/ooaf122 1.5M ECG transformer foundation model: AF AUROC 0.996, LVEF ≤40% AUROC 0.929, hybrid SSL (masked reconstruction + contrastive), publicly released weights.

[19] Ezendu N. Obianom et al. 2025. Reconstruction of 12-lead ECG: a review of algorithms. *Frontiers in Physiology* 16 (2025), 1532284. https://doi.org/10.3389/fphys.2025.1532284 Systematic review: 3-lead optimal (99.12% information content), r > 0.90 correlation, generic vs patient-specific comparison, no universal algorithm.

[20] O. Presacan, M. Lyng, H. Christensen, K. H. Haugaa, K. G. M. Moons, E. W. Steyerberg, et al. 2025. Evaluating the feasibility of 12-lead electrocardiogram reconstruction from single-lead and dual-lead ECGs using deep learning. *Communications Medicine* 5 (2025), 814. https://doi.org/10.1038/s43856-025-00814-w CRITICAL: First rigorous demonstration of GAN regression-to-mean ($R^2 = 0.92$ between error and amplitude), 9514 PTB-XL subjects, challenges field evaluation practices.

[21] P. Rajbhandary, G. Nallathambi, N. Selvaraj, T. Tran, and O. Colliou. 2022. ECG signal quality assessments of a small bipolar single-lead wearable patch sensor. *Cardiovascular Engineering and Technology* 13, 5 (2022), 783–796. https://doi.org/10.1007/s13239-022-00617-3

[22] P. Rautaharju, A. Davignon, F. Soumis, E. Boiselle, and A. Choquette. 1979. Evolution of QRS-T relationship from birth to adolescence in Frank-lead orthogonal electrocardiograms of 1492 normal children. *Circulation* 60, 1 (1979), 196–204. https://doi.org/10.1161/01.cir.60.1.196

[23] Ivan Sviridov and Konstantin Egorov. 2025. Conditional Electrocardiogram Generation Using Hierarchical Variational Autoencoders. *arXiv preprint arXiv:2503.13469* n/a (2025), n/a. https://doi.org/10.48550/arxiv.2503.13469 cNVAE-ECG: Hierarchical VAE outperforming GANs by up to 2% AUROC on downstream classification, multi-pathology high-resolution generation, publicly available.

[24] Y. Toyosu, S. Inui, Z. Wang, M. Akutagawa, S. Konaka, and Y. Kinouchi. 2015. High-resolution body-surface electrocardiograph system and survey of possible applications. *SpringerPlus* 4, 1 (2015), n/a. https://doi.org/10.1186/s40064-015-1325-8

[25] R. Tripathy and S. Dandapat. 2017. Automated detection of heart ailments from 12-lead ECG using complex wavelet sub-band bi-spectrum features. *Healthcare Technology Letters* 4, 2 (2017), 57–63. https://doi.org/10.1049/htl.2016.0089

[26] R. Vullings, C. Peters, I. Mossavat, S. Oei, and J. Bergmans. 2010. Bayesian approach to patient-tailored vectorcardiography. *IEEE Transactions on Biomedical Engineering* 57, 3 (2010), 586–595. https://doi.org/10.1109/tbme.2009.2033664

[27] Zhijiang Wan, Qianhao Yu, Jia Mao, Wenfeng Duan, and Cheng Ding. 2025. OpenECG: Benchmarking ECG Foundation Models with Public 1.2 Million Records. *arXiv preprint arXiv:2503.00711* n/a, n/a (2025), n/a. https://doi.org/10.48550/arxiv.2503.00711 Multi-center benchmark: 1.2M ECGs from 9 centers, compares SimCLR/BYOL/MAE with ResNet-50 and ViT, cross-dataset generalization analysis.

[28] Xiaoda Wang, Kaiqiao Han, Yuhao Xu, Xiao Luo, Yizhou Sun, Wei Wang, and Carl Yang. 2025. Simulator and Experience Enhanced Diffusion Model for Comprehensive ECG Generation. *arXiv preprint arXiv:2511.09895* n/a (2025), n/a. https://doi.org/10.48550/arxiv.2511.09895 SE-Diff: Physics-informed ODE-based ECG simulator integrated with latent diffusion, MAE 0.0923, NRMSE 0.0714, MAE_HR 8.43, 10s 12-lead generation.

[29] D. Wei. 2001. Deriving the 12-lead electrocardiogram from four standard leads based on the Frank torso model. *IEMBS Proceedings* 1 (2001), 381–384. https://doi.org/10.1109/iembs.2001.1018940

[30] Z. Zhang, X. Liu, S. Wei, H. Gan, F. Liu, and Y. Li. 2019. Electrocardiogram reconstruction based on compressed sensing. *IEEE Access* 7 (2019), 37228–37237. https://doi.org/10.1109/access.2019.2905000

[31] Z. Zhao, C. Liu, Y. Li, Y. Li, J. Wang, and B. Lin. 2019. Noise rejection for wearable ECGs using modified frequency slice wavelet transform and convolutional neural networks. *IEEE Access* 7 (2019), 34060–34067. https://doi.org/10.1109/access.2019.2900719

APPENDIX: INPUT CONFIGURATION EXPLORATION

Motivated by the observed information bottleneck, a natural extension is to compare alternative input configurations that include different precordial leads. Table 14 lists configurations we plan to evaluate beyond this project.

Table 14: Candidate input lead configurations for future work

Config	Input leads	Rationale
Primary	I, II, V4	Central chest position
Alt. 1	I, II, V3	Higher overlap with V2/V4
Alt. 2	I, II, V2	Better septal coverage
Alt. 3	I, II, V2+V4	Two chest leads (4-lead)