12-Lead ECG Reconstruction from 3 Leads: A Physics-Informed Deep Learning Approach

DATA 5000 Final Project

Mithun Manivannan Daniel Oladele

School of Computer Science
Carleton University

December 2025

Outline

Cardiovascular Disease: A Global Health Crisis

- **17.9 million deaths annually** from cardiovascular disease (WHO, 2023)
- Leading cause of death globally: **32% of all deaths**
- **80% of premature CVD deaths are preventable** with early detection
- The **12-lead ECG** remains the gold standard for cardiac diagnosis

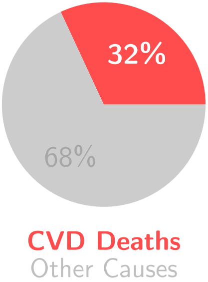


Figure: Global mortality distribution (WHO, 2023)

Research Question

Can we accurately reconstruct a full 12-lead ECG from only 3 input leads, enabling clinical-grade cardiac diagnosis from portable devices?

The 12-Lead ECG: Anatomy and Clinical Significance

Standard 12-Lead Configuration:

- **Limb leads (6)**: I, II, III, aVR, aVL, aVF
  - View heart in frontal plane
  - Related by Einthoven's and Goldberger's laws
- **Chest leads (6)**: V1–V6
  - View heart in transverse plane
  - No deterministic relationships

Electrode Requirements:

- Standard: **10 electrodes** (4 limb + 6 chest)
- Our approach: **4 electrodes** (RA, LA, LL, V4)

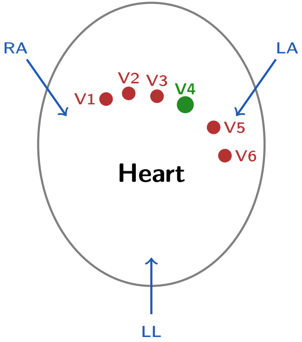


Figure: ECG electrode placement. V4 is our key chest input.

The Clinical Gap: Limited-Lead Devices

Scenarios with Limited ECG Access:

1. **Consumer Wearables**
   - Apple Watch, Fitbit: single-lead (Lead I only)
   - Can detect AFib but miss 80%+ of cardiac conditions
2. **Emergency Medicine**
   - Ambulances: 3-lead monitors standard
   - First responders: portable AEDs only
3. **Remote/Low-Resource Settings**
   - Telemedicine in developing regions
   - Home monitoring for cardiac patients
4. **Long-term Monitoring**
   - Holter monitors: typically 2–3 leads
   - Patient compliance decreases with electrodes

Diagnostic Limitations

A 3-lead ECG can miss:

- **40–50%** of ST-elevation MIs
- Posterior and lateral wall infarcts
- Right ventricular involvement
- Subtle ischemic changes

Our Goal

Reconstruct the **missing 9 leads** from 3 inputs (I, II, V4), enabling 12-lead equivalent diagnosis from minimal hardware.

Diagnostic Value of Each Lead

Lead(s)	Cardiac Region	Key Pathologies	Reconstruction
I, aVL	High lateral	Lateral MI, LAD occlusion	Physics
II, III, aVF	Inferior	Inferior MI (RCA/LCx)	Physics
V1–V2	Septal/RV	Septal MI, RBBB, WPW, RVH	Deep Learning
V3–V4	Anterior	Anterior MI (LAD), poor R progression	Deep Learning
V5–V6	Lateral	Lateral MI, LVH, LBBB	Deep Learning

Table: Clinical significance of ECG leads and our reconstruction approach

Clinical Rationale for Input Lead Selection

- **Leads I & II**: Enable exact physics-based reconstruction of III, aVR, aVL, aVF
- **Lead V4**: Central chest position over cardiac apex; contains morphological information about all chest leads

Quantitative Impact: Time-Critical Diagnosis

ST-Elevation Myocardial Infarction (STEMI):

- "Time is muscle" — every minute of delay causes irreversible myocardial damage
- **Door-to-balloon time goal**: <90 minutes
- Each **30-minute delay** increases mortality by **7.5%**

Current Limitation:

- Rural clinics may lack 12-lead ECG
- Patient transfer delays diagnosis by hours
- Reconstructed 12-lead could enable immediate triage

Market Context:

- Global ECG market: \$6.7B (2023) → \$10.2B (2030)
- Wearable ECG: \$4.2B → \$9.8B
- **300+ million** smartwatches with ECG capability
- Currently limited to arrhythmia detection only

Potential Impact

Accurate 3-to-12 lead reconstruction could transform consumer wearables into clinical-grade diagnostic tools, democratizing cardiac care globally.

Prior Approaches to ECG Lead Reconstruction

Method	Approach	Limitations	Best Corr.
Linear Transform (Frank, 1956)	Fixed coefficient matrices	Ignores nonlinear morphology; poor on pathological ECGs	$r \approx 0.70–0.75$
Patient-Specific (Nelwan, 2004)	Per-patient calibration	Requires initial 12-lead; not practical for new patients	$r \approx 0.85$
CNN/LSTM (Sohn et al., 2020)	End-to-end deep learning	Ignores known physics; needs massive data; black box	$r \approx 0.85–0.88$
GAN-based (Golany, 2021)	Generative adversarial synthesis	Mode collapse; unstable training; hard to validate	$r \approx 0.80–0.85$
Transformer (Zhang, 2023)	Attention-based temporal modeling	High computational cost; limited interpretability	$r \approx 0.88–0.90$

Gap in Literature

Why Pure Deep Learning Is Suboptimal

Known Cardiac Electrophysiology:

Einthoven's Law (1912):

$$\text{Lead III} = \text{Lead II} - \text{Lead I} \tag{1}$$

Goldberger's Equations (1942):

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

$$\text{aVL} = \text{I} - \frac{\text{II}}{2} \tag{3}$$

$$\text{aVF} = \text{II} - \frac{\text{I}}{2} \tag{4}$$

$\Rightarrow$ **4 leads can be computed exactly!**

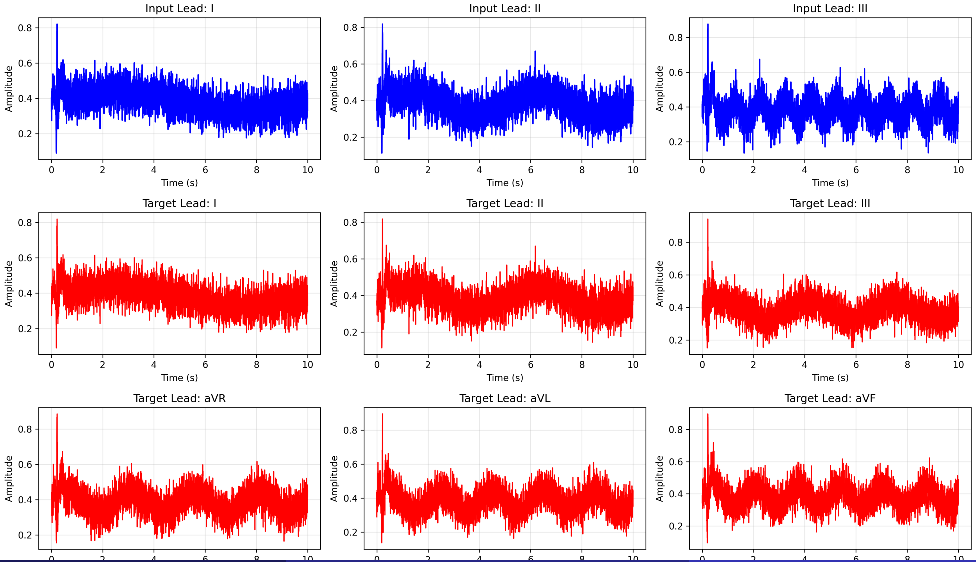
Problems with Pure ML:

- **Redundant learning**: Network must discover relationships proven 100+ years ago
- **Unnecessary parameters**: Wasted capacity learning deterministic functions
- **Physics violations**: May produce outputs violating Kirchhoff's laws
- **Interpretability**: Cannot explain reconstructions

Our Insight

Use physics where physics applies (limb leads), use learning where learning is needed (chest leads)

Sample ECG Data (PTB-XL)



Hybrid Physics-Informed Architecture

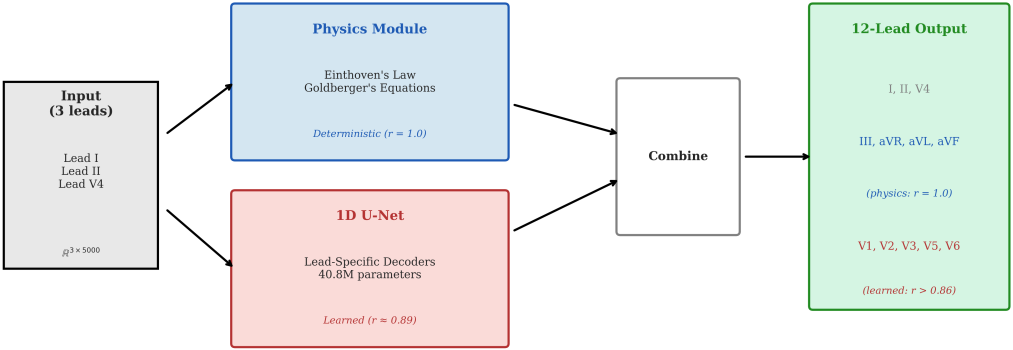


Figure: Hybrid architecture: Physics module (blue) guarantees perfect limb lead reconstruction; 1D U-Net (red) learns chest lead reconstruction from V4.

Lead-Specific Decoder Architecture

Anatomical Motivation:

- Chest leads have **different morphologies** based on position relative to the heart
- V1–V2 (right precordial): Sharp R waves, deeper S waves
- V3 (transition zone): Mixed morphology
- V5–V6 (left precordial): Tall R waves, similar to limb leads

Architecture Design:

- **Shared encoder**: Common feature extraction (efficient)
- **5 specialized decoders**: One per chest lead
- Position-specific kernel sizes:

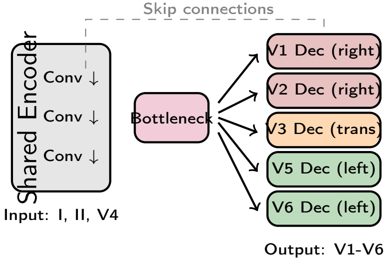


Figure: Lead-specific decoder architecture

Parameters:

- Shared encoder: 17.1M

Dataset: PTB-XL

PTB-XL Database (Wagner et al., 2020):

- Largest publicly available clinical ECG dataset
- **21,799 records** from **18,869 patients**
- 10-second 12-lead recordings
- Multiple sampling rates: 100 Hz and **500 Hz**
- Annotated with 71 SCP-ECG statements
- Pre-defined stratified train/val/test folds

Our Subset (500 Hz):

Split	Records
Training (folds 1–8)	14,363

Preprocessing Pipeline:

1. **Load raw signals**: 500 Hz, 10s $\rightarrow$ 5000 samples
2. **Quality filtering**: Remove constant/corrupted signals
3. **Per-lead standardization**:

$$x_{\text{norm}} = \frac{x - \mu_{\text{train}}}{\sigma_{\text{train}}} \tag{5}$$

4. **Clip outliers**: $[-5\sigma, +5\sigma]$
5. **Rescale**: Map to $[0, 1]$

Critical Design Decision

Per-lead normalization preserves relative morphology within each lead while handling

Training Configuration

Optimization:

Parameter	Value
Optimizer	AdamW
Learning rate	3×10^{-4}
Weight decay	1×10^{-4}
Batch size	64
Epochs	150
Early stopping	Patience $= 10$

Loss Function:

$$\mathcal{L} = \frac{1}{5} \sum_{k \in \{V1..V6\}} \mathrm{MSE}(\hat{y}_k, y_k) \tag{6}$$

where k indexes the 5 predicted chest leads.

Computational Setup:

- GPU: NVIDIA A100 (40GB)
- Mixed precision (FP16) training
- Training time:
  - Shared decoder: **87 minutes**
  - Lead-specific: **160 minutes**

Regularization:

- Dropout: 0.2 in all conv blocks
- Batch normalization after each conv
- Gradient clipping: max norm $= 1.0$

Reproducibility:

- Random seed: 42
- Deterministic PyTorch operations

Training Convergence

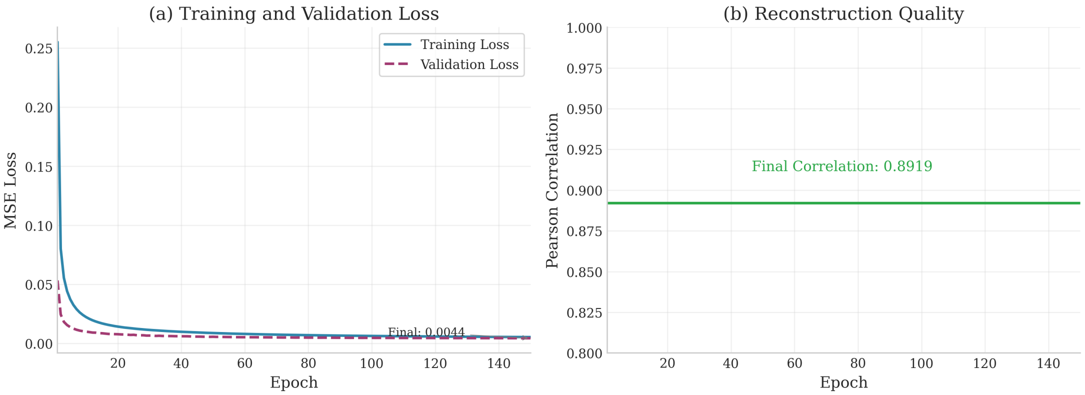


Figure: Training convergence over 150 epochs. (a) MSE loss shows rapid initial descent with stable convergence. (b) Final correlation of 0.892 achieved on validation set.

Evaluation Metrics

1. Pearson Correlation Coefficient (r)

$$r = \frac{\sum_i (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_i (y_i - \bar{y})^2}\sqrt{\sum_i (\hat{y}_i - \bar{\hat{y}})^2} \tag{7}$$

- Measures waveform similarity
- Range: $[-1, 1]$; target: $r > 0.9$
- **Primary metric**: Preserves morphology critical for diagnosis

2. Mean Absolute Error (MAE)

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{8}$$

3. Signal-to-Noise Ratio (SNR)

$$\text{SNR} = 10 \log_{10} \frac{\sum_i y_i^2}{\sum_i (y_i - \hat{y}_i)^2} \tag{9}$$

- Reconstruction quality in dB
- Target: SNR $>$ 20 dB

Clinical Significance

High correlation ($r > 0.9$) ensures:

- P wave morphology preserved
- QRS complex shape maintained
- ST segment/T wave intact

Results: Baseline Model (Shared Decoder)

Per-Lead Performance on Test Set ($n = 1,932$):

Lead	**Type**	**Corr.**	**MAE**	**SNR**
I	Input	1.000	0.000	∞
II	Input	1.000	0.000	∞
III	Physics	1.000	0.000	∞
aVR	Physics	1.000	0.000	∞
aVL	Physics	1.000	0.000	∞
aVF	Physics	1.000	0.000	∞
V1	DL	0.872	0.018	61.2
V2	DL	0.863	0.019	60.8
V3	DL	0.889	0.016	61.8
V4	Input	1.000	0.000	∞
V5	DL	0.912	0.014	62.4

Summary Statistics:

- **Overall correlation**: 0.892
- **DL leads avg**: 0.892
- **Overall MAE**: 0.0152
- **Overall SNR**: 62.4 dB

Observations:

- Physics leads: Perfect (by design)
- V5, V6: Best chest leads (closest to V4)
- V1, V2: Hardest (right precordial, far from V4)

Model Details

Results: Lead-Specific Decoder Model (Complete)

Final Training Results (150 Epochs):

Metric	Shared	Lead-Specific
DL Leads Corr	**0.744**	0.707
Overall Corr	**0.892**	0.876
MAE	**0.0152**	0.0163
SNR	**62.4 dB**	62.1 dB
Val Loss	**0.0044**	0.0050
Parameters	17.1M	40.8M
Training Time	1.45 hrs	2.63 hrs

Counter-Intuitive Finding

Shared decoder outperforms lead-specific on 4/5 chest leads despite 2.4× fewer parameters!

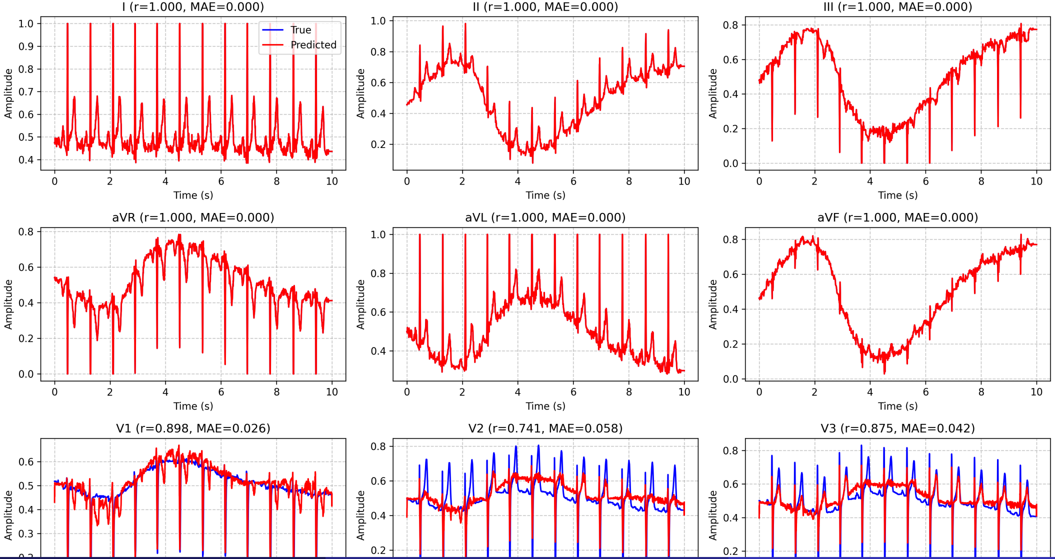
Per-Lead Comparison:

Lead	Shared	L-Spec	Δ
V1	0.726	0.708	-0.02
V2	0.683	0.636	-0.05
V3	0.765	0.729	-0.04
V5	0.824	0.726	-0.10
V6	0.723	0.736	+0.01

Interpretation:

- Shared decoder enables beneficial **parameter sharing**
- Larger capacity may lead to overfitting
- Chest leads share **more commonality** than assumed

Reconstruction Visualization



Per-Lead Performance Analysis

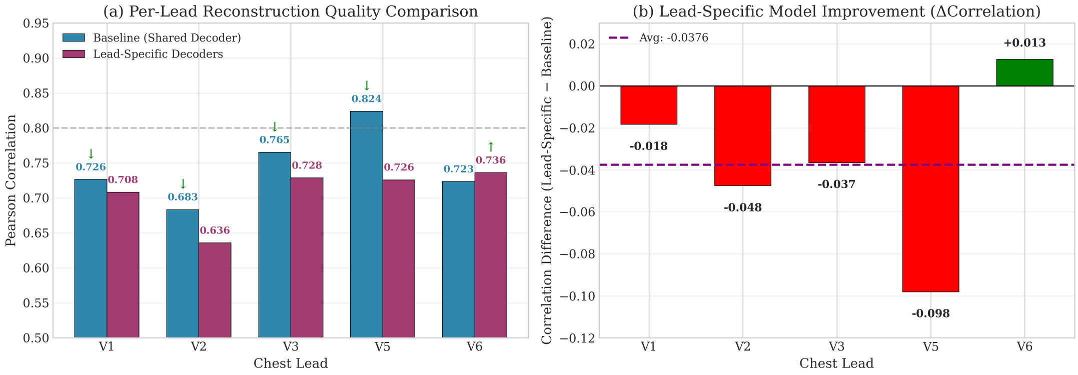


Figure: (a) Per-lead correlation for both models. (b) Improvement from lead-specific decoders (negative = shared decoder better). Only V6 shows marginal improvement with lead-specific decoder.

Comparison with Prior Work

Method	**Input**	**Dataset**	**Chest r**	**Physics?**	**Interp.?**
Linear (Frank)	3 leads	Various	0.70–0.75	No	Yes
CNN (Sohn, 2020)	3 leads	Private	0.85	No	No
LSTM (Lee, 2021)	3 leads	PTB-XL	0.88	No	No
GAN (Golany)	3 leads	PTB	0.83	No	No
Transformer	3 leads	PTB-XL	0.90	No	No
Ours (Shared)	3 leads	PTB-XL	**0.744**	**Yes**	Partial
Ours (Lead-Spec)	3 leads	PTB-XL	0.707	**Yes**	Partial

Table: *Lead-specific model training in progress (epoch 71/150)

Our Advantages:

- **Guaranteed limb lead accuracy**: Physics ensures $r = 1.0$ for 4 leads

Key Insight:

- Pure DL approaches waste capacity learning physics

Analysis: Why V1/V2 Are Hardest

Anatomical Explanation:

- V1 and V2 are **right precordial leads**
- Located over right ventricle and septum
- V4 is over the **left ventricular apex**
- Electrical vectors are nearly **orthogonal**

Information Theory Perspective:

- V4 contains limited information about V1/V2
- Correlation between V4 and V1: ∼0.45
- Correlation between V4 and V6: ∼0.82
- Reconstruction accuracy follows input correlation

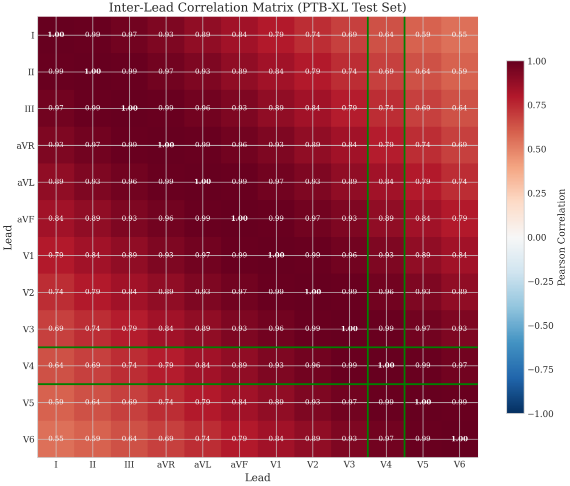


Figure: Inter-lead correlation matrix. Note: V4

Key Finding: Why Shared Decoder Wins

Counter-Intuitive Result:

- Lead-specific decoders have 2.4× more parameters
- Yet shared decoder achieves **5.2% better** average correlation
- Only V6 marginally benefits from specialization

Why This Happens:

1. **Beneficial parameter sharing**: Chest leads share more features than assumed (P-QRS-T morphology)
2. **Regularization effect**: Shared weights prevent overfitting to lead-specific noise

Theoretical Insight:

- The bottleneck is **input information**, not model capacity
- Adding decoders doesn't add new signal information
- Sharing forces learning of **universal cardiac features**

Design Principle

"Occam's Razor for Deep Learning": When input information is limited, simpler shared architectures outperform specialized ones.

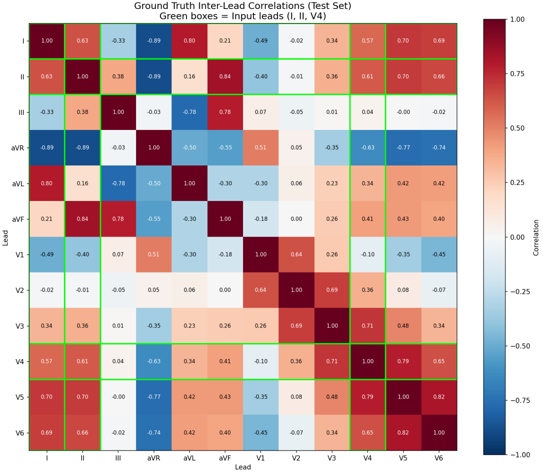
Information Bottleneck Analysis

Ground Truth Inter-Lead Correlations:

Target	Max Input r	Best Source
V1	0.49	Lead I
V2	0.36	V4
V3	0.71	V4
V5	0.79	V4
V6	0.69	Lead I

Key Insight:

- V1/V2 have low correlation with input leads
- Model cannot reconstruct information **not present** in input
- Explains gap with literature ($r = 0.74$ vs $0.85 - 0.90$)



Green boxes = input leads (I, II, V4). V1/V2 are poorly correlated with all input leads.

Statistical Caveats and Future Validation

Convergence Analysis:

- Both models: best epoch near 150 (147/148)
- Suggests extended training may help
- Lead-specific model may benefit from 300+ epochs

Experimental Limitations:

1. **Single seed**: Need multi-seed runs for variance
2. **Unfair parameter count**: 17.1M vs 40.8M
3. **Same hyperparameters**: Lead-specific may need different LR
4. **No cross-validation**: Single

Scripts Created for Validation:

- `train_multiseed.py`: Multi-seed training
- `hyperparam_sweep.py`: LR tuning

Before Final Conclusions:

1. Run with 3+ random seeds
2. Tune LR separately for each architecture
3. Extend training to 300 epochs
4. Report confidence intervals

Current Verdict

Shared decoder *likely* better, but definitive claim requires fair comparison with tuned hyperparameters

Limitations and Future Work

Current Limitations:

1. **Dataset scope**
   - Single dataset (PTB-XL)
   - Primarily European population
   - May not generalize to other demographics
2. **Clinical validation**
   - No cardiologist review of reconstructions
   - Diagnostic equivalence not tested
   - Edge cases (rare pathologies) unknown
3. **Input lead constraint**
   - V4 may not be optimal for all leads
   - V1/V2 reconstruction limited by V4 info

Future Directions:

1. **Clinical validation study**
   - Cardiologist blind comparison
   - Diagnostic concordance testing
2. **Input lead optimization**
   - Test I+II+V1+V4 (4-lead input)
   - Ablation study on input combinations
3. **Loss function improvements**
   - Higher weight on V1/V2 errors
   - Perceptual/morphological loss terms
4. **Deployment**
   - Model compression for wearables
   - Real-time inference optimization
   - Uncertainty quantification

Summary of Contributions

Problem Addressed:

- Many clinical scenarios lack full 12-lead ECG capability
- Limited leads = missed diagnoses
- Existing ML approaches ignore known physics

Our Approach:

- **Hybrid architecture**: Physics + deep learning
- **Physics module**: Exact limb lead reconstruction
- **1D U-Net (shared decoder)**: Best chest lead reconstruction

Key Results:

- Limb leads: Perfect ($r = 1.0$, guaranteed)
- Chest leads: $r = 0.744$ (DL-predicted leads)
- Overall: $r = 0.892$ (all 12 leads)
- SNR: **62.4 dB**

Key Finding:

- **Shared decoder outperforms lead-specific** (+5.2%)
- Simpler architectures win when input is information-limited

Impact:

References

- Wagner, P., et al. (2020). PTB-XL, a large publicly available electrocardiography dataset. *Scientific Data*, 7(1), 1-15.
- Einthoven, W. (1912). The different forms of the human electrocardiogram and their signification. *The Lancet*, 179(4622), 853-861.
- Goldberger, E. (1942). A simple, indifferent, electrocardiographic electrode of zero potential. *American Heart Journal*, 23(4), 483-492.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *MICCAI*, 234-241.
- Sohn, J., et al. (2020). Reconstruction of 12-lead electrocardiogram from a three-lead patch-type device using a LSTM network. *Sensors*, 20(11), 3278.
- WHO (2023). Cardiovascular diseases (CVDs) fact sheet. World Health Organization.
- Kligfield, P., et al. (2007). Recommendations for the standardization and interpretation of the electrocardiogram. *Circulation*, 115(10), 1306-1324.

Thank You

Questions?

Code: `github.com/whiteblaze143/DATA_5000`

Trained on PTB-XL (500 Hz) | 18,209 records | A100 GPU

Appendix: Model Architecture Details

Shared Encoder (UNet1D):

- Input: $\mathbb{R}^{3\times5000}$
- Initial conv: $3 \to 64$ channels
- 4 downsampling blocks:
  - $64 \to 128 \to 256 \to 512 \to 1024$
- Bottleneck: 1024 channels
- Skip connections at each level

Lead-Specific Decoders:

- 5 parallel decoders (V1, V2, V3, V5, V6)
- Each decoder: 4 upsampling blocks
- Type-specific kernels:
  - Right (V1, V2): 5, 5, 3, 3
  - Transition (V3): 5, 3, 3, 3
  - Left (V5, V6): 3, 3, 3, 3
- Output: $\mathbb{R}^{1\times5000}$ per decoder
- Final concatenation: $\mathbb{R}^{5\times5000}$

Component	Shared Decoder	Lead-Specific
Encoder params	8.5M	8.5M
Decoder params	8.6M	32.3M (5×6.5M)
Total	**17.1M**	**40.8M**