

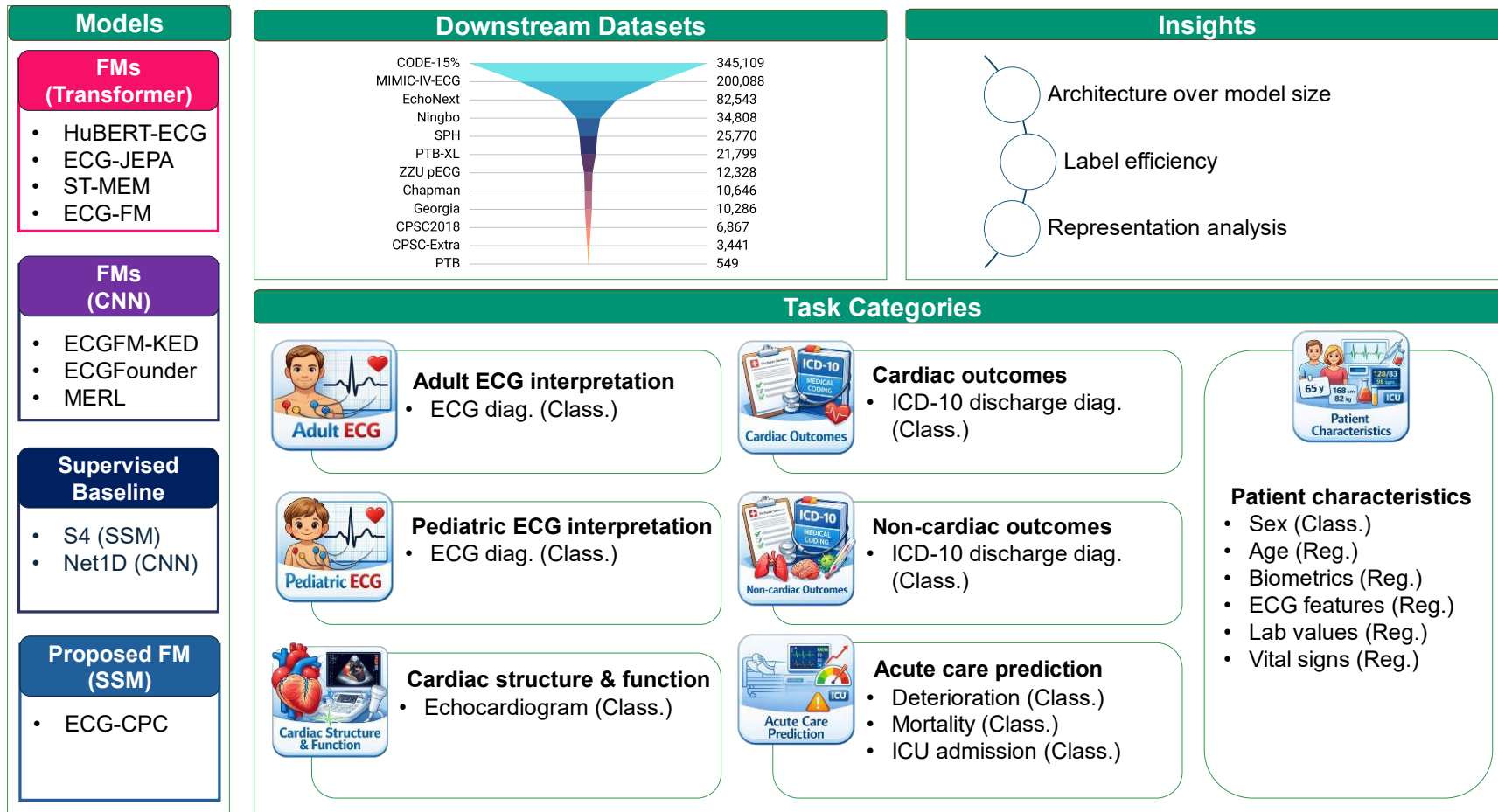
Benchmarking ECG FMs:

A Reality Check Across Clinical Tasks

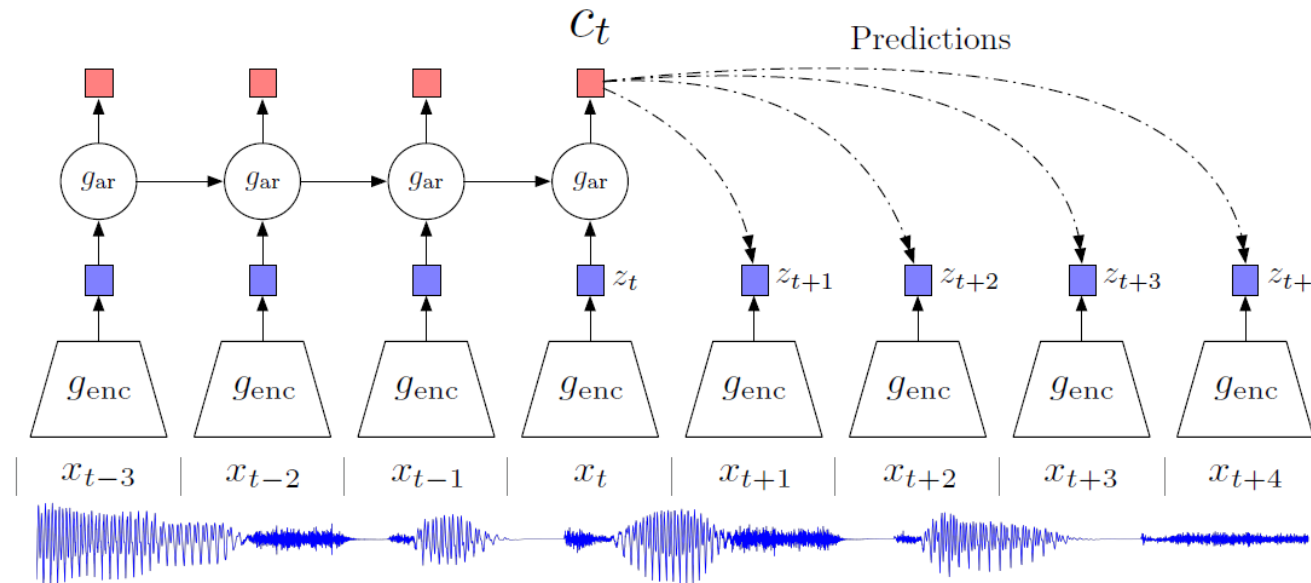
M A Al-Masud, Juan Miguel Lopez Alcaraz &
Nils Strodthoff



Methods



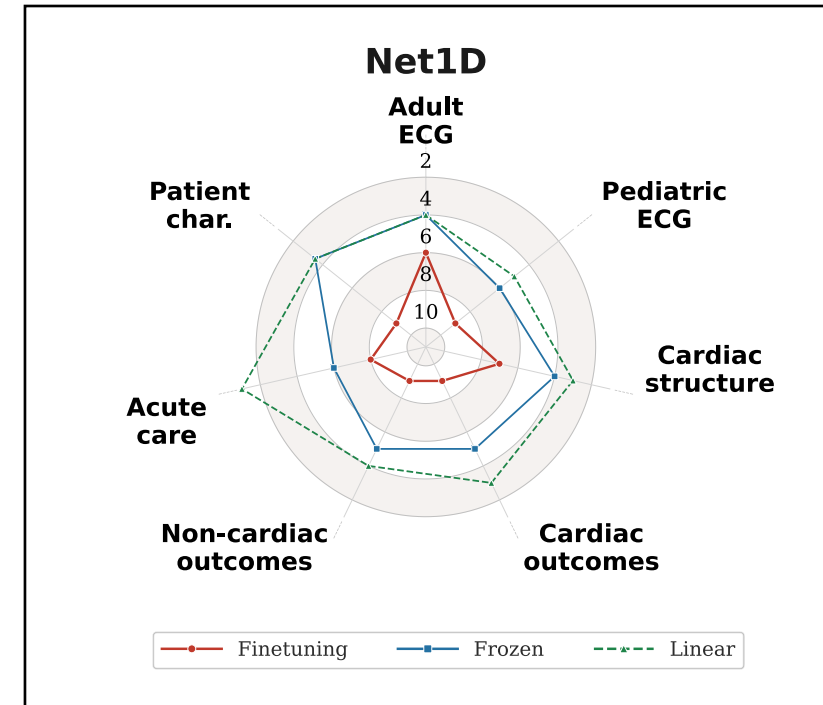
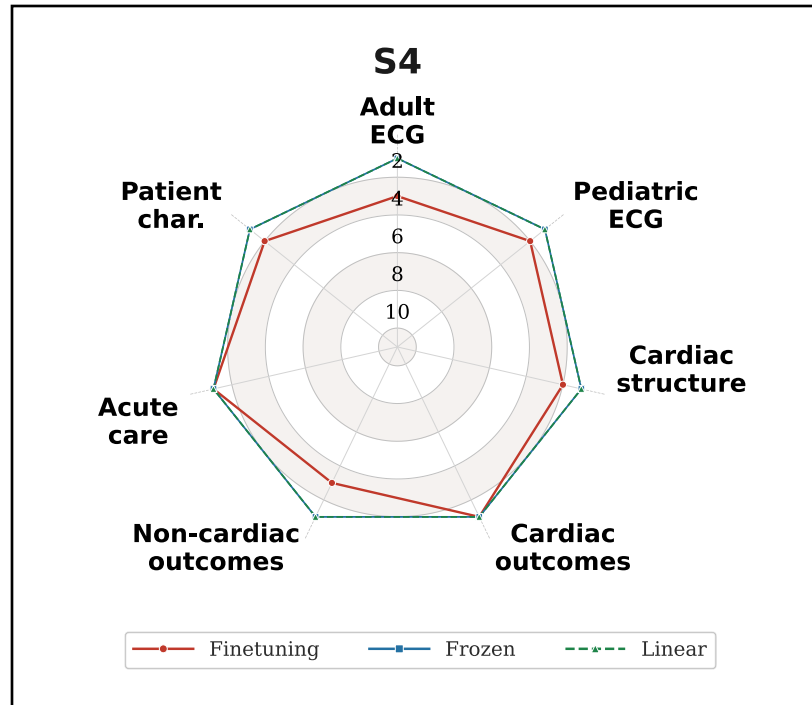
Proposed FM (ECG-CPC)



Overview of Contrastive Predictive Coding (Oord et al., 2018)

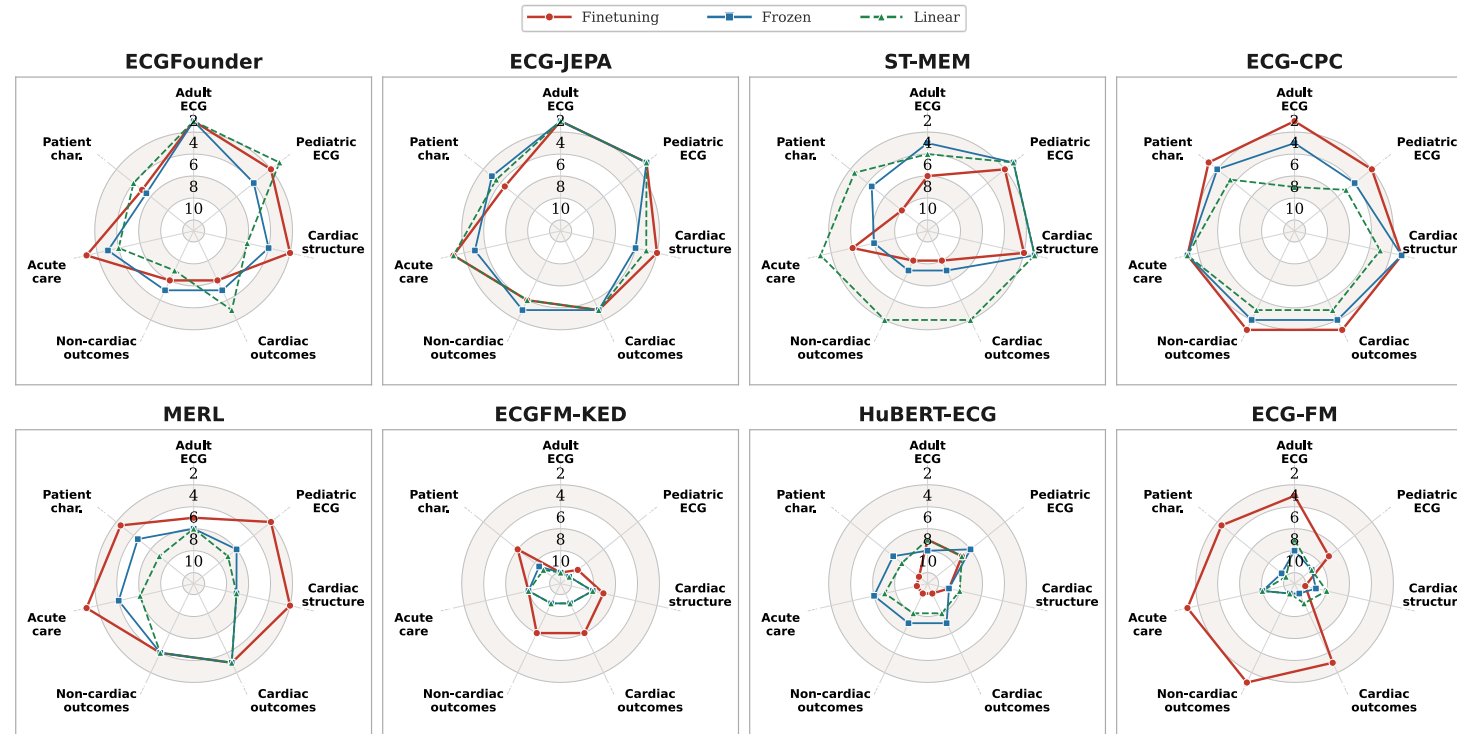
- **Backbone:** Structured State Space Sequence (S4) Model
- **Pretraining Dataset:** HEEDB (10M samples).
- **Pretraining Method:** Contrastive Predictive Coding (CPC)
- **Model Complexity:** 3.8M parameters, 1.741 GFLOPs

Supervised Baselines



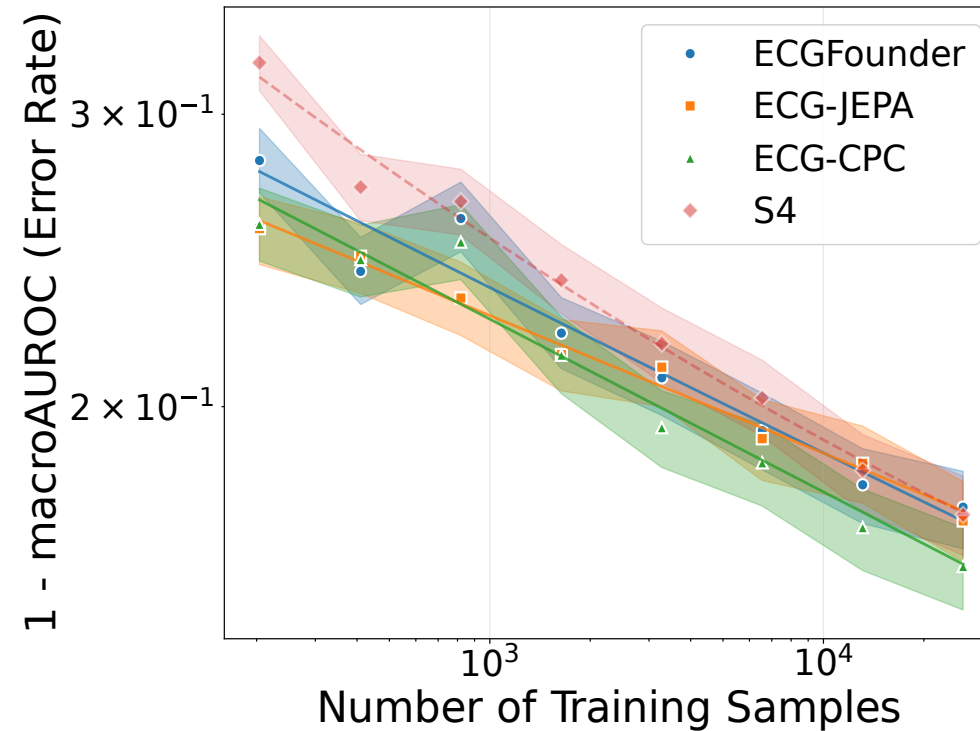
- **S4 vs Net1D:** S4 superior to Net1D (CNN) when trained from scratch
- **Frozen/Linear Eval:** No FM beats the supervised baseline (S4)
- **Finetuning Evaluation:** Only a few FMs outperform the supervised baseline (S4)

FM Rankings



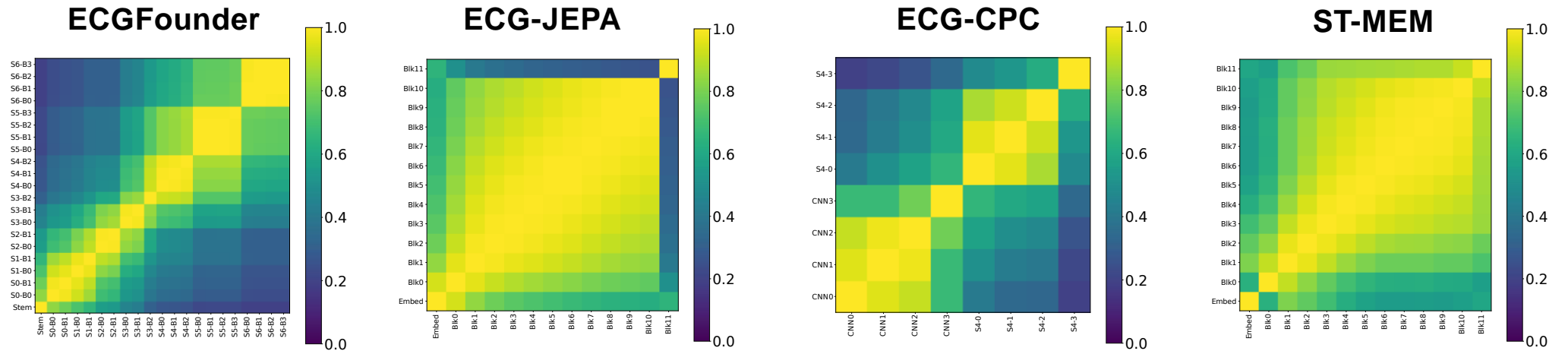
- **Finetuning:** ECGFounder, ECG-JEPA & ECG-CPC achieve top performance for adult ECG interpretation, surpassing supervised baseline (S4). ECG-CPC excels in cardiac structure, outcomes, acute care predictions, and patient characteristics.
- **Frozen:** ECGFounder and ECG-JEPA serve as effective frozen feature extractors for adult ECG interpretation.

Label Efficiency



- Subsampled EchoNext (powers of 2) using multi-label stratification.
- Fitted error rates to power-law curves ($CN^{-\alpha} + L_0$)
- **FMs improve label efficiency:** ECGFounder: 3.3, ECG-JEPA: 9, ECG-CPC: 4.8
- **Low-sample:** ECG-JEPA leads, ECG-CPC catches up gradually.

Representation Similarity



- CKA heatmaps of internal layers within each frozen FM.
- Gaussian RBF kernel ($\sigma = 1.0$); 2,500 PTB-XL samples.

Summary

- **Architecture > Scale:** The lightweight S4-based ECG-CPC outperforms larger Transformer models across most tasks, showing design beats size.
- **Label Efficiency:** ECG FMs improve label efficiency by 3.3–9x vs. supervised baselines.
- **Representation Similarity:** Models with similar accuracy learn distinct internal patterns, revealing multiple paths to effective ECG understanding.