Contrastive Pretraining Scaling and Multimodal Alignment in ECG-Text Foundation Models

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: How does contrastive pretraining scaling affect multimodal alignment scores between ECG signals and clinical text descriptions in foundation models. 16 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Pretraining Strategies and Scaling for ECG Foundation Models: A Systematic Study. Research question: How does contrastive pretraining scaling affect multimodal alignment scores between ECG signals and clinical text descriptions in foundation models?.

2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.5/10.

3 Results

16 papers retrieved. 16 claims extracted; 1 independently verified. Quality review score: 4.5/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
The study covers five different pretraining methodologies.	×	0.11
The pretraining corpus used comprises over 11 million samples.	×	0.04
State space models (SSM) are confirmed as the superior architecture choice across all pretraining paradigms compared to	×	0.10
CPC (Contrastive Predictive Coding) shows the strongest and most transferable representations across diverse clinical ta	✓	0.16
data2vec consistently lags behind other methodologies across all evaluation modes and scaling regimes.	×	0.05
Scaling behavior is most clearly identified for CPC and JEPA methodologies.	×	0.10
Lower pretraining loss correlates with small residual errors in downstream tasks.	×	0.04
Recent ECG foundation models have been pretrained on tens of millions of recordings.	×	0.11
Structured state space models have shown superior performance on long sequences in supervised ECG settings in prior stud	×	0.13
The CNN stem used in the study consists of four convolutional layers with batch normalization.	×	0.02
The study evaluates three backbone variants: S4-based, Transformer with RoPE and GELU, and CNN-based (Net1D).	×	0.04
All models in the study operate at 240 Hz on 12-lead ECG inputs.	×	0.04
The S4 backbone with model dimension 512 consistently outperforms larger dimensions (768, 1024) and alternative configur	×	0.05
The default backbone adopted for the study is a 4-layer S4 with dimension 512.	×	0.01
The study investigates five self-supervised pretraining objectives spanning contrastive, predictive, and clustering-base	×	0.11
data2vec trains the model to predict the EMA teacher's averaged top-k contextualized layer representations at masked pos	×	0.03

References

- http://arxiv.org/abs/2303.00915v3
- http://arxiv.org/abs/2510.21551v1
- http://arxiv.org/abs/2605.12241v1