Scaling Dataset Size and Zero-Shot Reasoning in ECG Foundation Models for Arrhythmia Detection

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 9 peer-reviewed papers addressing the following research question: What is the impact of dataset size scaling on the zero-shot reasoning capabilities of ECG foundation models when evaluated on arrhythmia detection benchmarks. 7 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.6/10. 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: What is the impact of dataset size scaling on the zero-shot reasoning capabilities of ECG foundation models when evaluated on arrhythmia detection benchmarks?.

2 Methodology

Systematic literature search across multiple databases yielded 9 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.6/10.

3 Results

9 papers retrieved. 7 claims extracted; 1 independently verified. Quality review score: 4.6/10.

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 trained on over 11M samples.	✓	0.16
The S4-based backbone is confirmed as the superior architecture choice across all pretraining paradigms.	×	0.02
CPC shows the strongest and most transferable representations across diverse clinical tasks.	×	0.11
Data2vec consistently lags behind across all evaluation modes and scaling regimes.	×	0.03
Lower pretraining loss correlates with small residual errors in downstream tasks.	×	0.04
The S4 backbone with model dimension 512 consistently outperforms larger and alternative configurations.	×	0.01
The study investigates five self-supervised pretraining objectives spanning contrastive, predictive, and clustering-base	×	0.11

References

- http://arxiv.org/abs/2304.06427v2
- http://arxiv.org/abs/2207.07089v2
- http://arxiv.org/abs/2605.12241v1