How do tabular foundation models pretrained on mixed synthetic-real datasets compare to pure synthetic pretrai

Assignee Research

June 10, 2026

Abstract

Specialized foundation models are beginning to emerge in various medical subdomains, but pretraining methodologies and parametric scaling with the size of the pretraining dataset are rarely assessed systematically and in a like-for-like manner. This work focuses on foundation models for electrocardiography (ECG) data, one of the most widely captured physiological time series world-wide. We present a comprehensive assessment of pretraining methodologies, covering five different contrastive and non-contrastive self-supervised learning objectives for ECG foundation models, and investigate their s

1 Introduction

This paper examines: Pretraining Strategies and Scaling for ECG Foundation Models: A Systematic Study. Research question: How do tabular foundation models pretrained on mixed synthetic-real datasets compare to pure synthetic pretraining in terms of alignment evaluation metrics like accuracy and fairness on downstream tasks?.

2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.2/10.

3 Results

11 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 6.2/10.

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.

References

- http://arxiv.org/abs/2504.20900v1
- http://arxiv.org/abs/2605.12241v1
- http://arxiv.org/abs/2403.13430v2