

# Pretraining Dataset Scale and Downstream ECG Classification Accuracy on Chaparral

Assignee Research

June 9, 2026

## Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: What is the correlation between pretraining dataset scale and downstream classification accuracy for ECG foundation models when evaluated on the Chaparral benchmark. 11 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: CLEF: Clinically-Guided Contrastive Learning for Electrocardiogram Foundation Models. Research question: What is the correlation between pretraining dataset scale and downstream classification accuracy for ECG foundation models when evaluated on the Chaparral benchmark?.

## 2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.2/10.

## 3 Results

12 papers retrieved. 11 claims extracted; 0 independently verified. Quality review score: 3.2/10.

## 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.

## 5 Extracted Claims

Claim	Verified	Confidence
CLEF-M is a medium-sized model evaluated across 13 downstream classification tasks.	×	0.09
The CLEF framework includes a negative weighting loss (Lw) and a dissimilarity alignment loss (Ld).	×	0.03
Collecting sufficiently large, human-annotated, single-lead ECGs is impractical due to variability in downstream tasks a	×	0.08
ECGs from healthcare organizations are usually provided in the standard 12-lead format but contain limited annotations.	×	0.07
Prior work incorporated subject and signal attributes in contrastive learning but did not incorporate clinical knowledge	×	0.13
Risk scores estimate a person’s risk of experiencing an adverse health outcome in a specified time period, such as the r	×	0.03
The CLEF pretraining dataset consists of 10-second 12-lead ECGs.	×	0.12
The MOCO (2020) model achieved an AUROC of 0.8112 on the Small (S) configuration in the reported results.	×	0.02
The SimCLR (2020) model achieved an AUROC of 0.8243 on the Medium (M) configuration in the reported results.	×	0.03
The BYOL (2020) model achieved an AUROC of 0.8221 on the Small (S) configuration and 0.8421 on the Medium (M) configurat	×	0.03
Clinical metadata used in the CLEF framework includes age, blood pressure, and clinical risk scores.	×	0.09

## References

- <http://arxiv.org/abs/2512.02180v1>
- <http://arxiv.org/abs/2606.00815v1>
- <http://arxiv.org/abs/2605.12241v1>