Multimodal Transformers vs. Domain-Specific Fine-Tuning in Few-Shot Clinical Waveform Adaptation

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: How does the performance of multimodal transformers (e.g., MM-1) pre-trained on large-scale clinical waveforms compare to domain-specific fine-tuning when evaluated on few-shot adaptation benchmarks. 14 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.0/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Enhancing clinical decision support with physiological waveforms – a multimodal benchmark in emergency care. Research question: How does the performance of multimodal transformers (e.g., MM-1) pre-trained on large-scale clinical waveforms compare to domain-specific fine-tuning when evaluated on few-shot adaptation benchmarks in emergency care using accuracy and calibration metrics?.

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 4.0/10.

3 Results

11 papers retrieved. 14 claims extracted; 0 independently verified. Quality review score: 4.0/10.

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 multimodal model integrating ECG waveforms and clinical routine data achieved a macro-AUROC of 0.8256 (0.8222, 0.828	×	0.07
The multimodal model integrating ECG waveforms and clinical routine data achieved an AUROC of 0.9115 (0.8991, 0.9222) fo	×	0.08
The model predicted 609 out of 1,428 individual ICD diagnoses with high accuracy, defined as conditions where the lower	×	0.05
The model's predictive performance ranges from an AUROC of 0.7405 for the musculoskeletal system and connective tissue (	×	0.03
The inclusion of ECG waveforms alongside clinical routine data consistently improves performance over ECG features and c	×	0.04
The highest gains are observed for XII (Skin, 13.06%), VI (Nervous, 10.75%), and XIX (Injury, Poisoning, 9.95%).	×	0.02
Smaller improvements are seen for III (Blood, 2.36%) and IV (Endocrine, 5.25%).	×	0.02
For clinical deterioration, the model achieves an AUROC of 0.9070.	×	0.11
For ICU admissions, the model reports an overall AUROC of 0.9063.	×	0.05
For mortality predictions, the model exhibits an overall AUROC of 0.9168.	×	0.03
The MDS-ED pipeline involves feature collection encompassing patient demographics, biometrics, vital parameters and tren	×	0.11
The MDS-ED pipeline predicts patient discharge diagnoses out of 1428 cardiac and non-cardiac ICD10-CM codes and predicts	×	0.12
The MDS-ED pipeline collects features from a window of 90 minutes from the patient's arrival at the ED.	×	0.04
The MDS-ED was created by linking ECG waveforms from the MIMIC-IV-ECG dataset to clinical features and outcomes from the	×	0.03

References

- http://arxiv.org/abs/2311.04937v1
- http://arxiv.org/abs/2407.17856v4
- http://arxiv.org/abs/2308.10783v2