

# Scaling Tabular Foundation Models and Zero-Shot Transfer Accuracy in Heterogeneous Domains

Assignee Research

June 8, 2026

## Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: Does scaling the size of tabular foundation models improve zero-shot transfer accuracy across heterogeneous domains in the TabTime benchmark. 10 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: OmniEEG-Bench: A Standardized Evaluation Benchmark for EEG Foundation Models. Research question: Does scaling the size of tabular foundation models improve zero-shot transfer accuracy across heterogeneous domains in the TabTime benchmark?.

## 2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.8/10.

## 3 Results

14 papers retrieved. 10 claims extracted; 0 independently verified. Quality review score: 3.8/10.

## 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
EEG foundation models pretrained on a greater number of datasets tend to achieve lower average ranks (i.e., better performance).	×	0.11
EEG foundation models with a larger number of parameters tend to achieve lower average ranks (i.e., better performance).	×	0.12
The benchmark includes 58 datasets for evaluating EEG foundation models.	×	0.11
The benchmark evaluates four complementary evaluation paradigms: cross-subject transfer, multi-subject adaptation, few-shot.	×	0.03
The benchmark preprocesses 54 EEG datasets through a standardized pipeline of downsampling, band-pass and notch filtering.	×	0.06
The benchmark randomly selects up to 40 samples per subject per class for linear probing.	×	0.02
The benchmark evaluates ten representative EEG foundation models: BENDR, BIOT, LaBraM, CBraMod, BrainOmni, FEMBA, Neuro-EEG.	×	0.08
The benchmark uses linear probing as the primary evaluation method, freezing the pretrained backbone and training only the probe.	×	0.03
The benchmark performs full fine-tuning on all the datasets from each task category under the cross-subject setting to compare models.	×	0.07
Models pretrained on multiple datasets outperform those pretrained on a single dataset overall.	×	0.06

## References

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

- <http://arxiv.org/abs/2512.03307v1>
- <http://arxiv.org/abs/2502.04573v2>