# Correlation Between Pre-training Dataset Diversity and Noise Robustness in EEG Foundation Models on OmniEEG-Bench

Assignee Research

June 11, 2026

**Abstract**

Electroencephalography (EEG) supports a variety of brain-computer interface (BCI) tasks ranging from brain-state monitoring to human-LLM interactions. EEG foundation models are emerging, but evaluation remains fragmented due to heterogeneous datasets and nconsistent task protocols. Here, we introduce OmniEEG-Bench, a unified benchmark and downstream task roadmap for EEG foundation models (FMs). It organizes evaluation of EEG FMs into six task families spanning (i) signal reliability, (ii) biometrics and disease, (iii) consciousness and state, (iv) cognition and emotion, (v) naturalistic stimul

## 1 Introduction

This paper examines: OmniEEG-Bench: A Standardized Evaluation Benchmark for EEG Foundation Models. Research question: What is the correlation between pre-training dataset diversity and robustness to signal noise in EEG foundation models evaluated on the signal reliability task family of OmniEEG-Bench?.

## 2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.3/10.

## 3 Results

13 papers retrieved. 10 claims extracted; 7 independently verified. Quality review score: 7.3/10.

## 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 perfo | ✓ | 0.23 |
| EEG foundation models with a larger number of parameters tend to achieve lower average ranks (i.e., better performance). | ✓ | 0.24 |
| 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-s | ✓ | 0.27 |
| The benchmark preprocesses 54 EEG datasets through a standardized pipeline of downsampling, band-pass and notch filterin | ✓ | 0.25 |
| The benchmark randomly selects up to 40 samples per subject per class for linear probing. | × | 0.15 |
| The benchmark evaluates ten representative EEG foundation models: BENDR, BIOT, LaBraM, CBraMod, BrainOmni, FEMBA, Neuro- | ✓ | 0.21 |
| The benchmark uses linear probing as the primary evaluation method, freezing the pretrained backbone and training only t | ✓ | 0.21 |
| The benchmark performs full fine-tuning on all the datasets from each task category under the cross-subject setting to c | ✓ | 0.23 |
| Models pretrained on multiple datasets outperform those pretrained on a single dataset overall. | × | 0.06 |

## References


- http://arxiv.org/abs/1801.02477v1

- http://arxiv.org/abs/2512.03307v1
- http://arxiv.org/abs/2606.00815v1