Robustness of EEG Foundation Models vs. Supervised Baselines on OmniEEG-Bench Under Noise and Adversarial Attacks

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: What is the robustness of EEG foundation models on OmniEEG-Bench when evaluated with adversarial or noisy EEG data, and how does this compare to traditional supervised learning models. 12 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.4/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Utilizing Class Separation Distance for the Evaluation of Corruption Robustness of Machine Learning Classifiers. Research question: What is the robustness of EEG foundation models on OmniEEG-Bench when evaluated with adversarial or noisy EEG data, and how does this compare to traditional supervised learning models?.

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 3.4/10.

3 Results

12 papers retrieved. 12 claims extracted; 0 independently verified. Quality review score: 3.4/10.

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 paper proposes the 'MSCR' metric to evaluate and compare classifier corruption robustness.	×	0.13
The MSCR metric approach is independent of prior knowledge about corruption distances.	×	0.06
For the 2D dataset, the model trained on ε_train=0.007 performs best when testing with lower noise levels and clean test	×	0.07
For the 2D dataset, the maximum overall accuracy is achieved with a model trained on ε_train=0.007 and tested on ε_test=	×	0.04
For the 2D dataset at higher noise levels, optimum robust accuracies are achieved with ε_train $\leq$ ε_test.	×	0.05
For the CIFAR-10 dataset at low noise levels, training with ε_train=0.01 appears to be optimal for clean accuracy.	×	0.06
For the CIFAR-10 dataset, the maximum overall accuracy is achieved with ε_train=0.02 and ε_test=0.01.	×	0.03
For the CIFAR-10 dataset, the optimum ε_train for ε_test=0 is approximately 10 times lower than the ε_min value.	×	0.02
An optimum ε_train of 0.01 for CIFAR-10 translates to a 2.5/255 color grade corruption for every pixel.	×	0.03
For both the 2D and CIFAR-10 datasets, the MSCR value steadily increases with increasing ε_train.	×	0.06
In the 2D dataset experiments, the highest recorded accuracy is 99.785% $\pm$ 0.009, achieved with ε_train=0.01 and ε_test=0	×	0.03
In the CIFAR-10 dataset experiments, the highest recorded accuracy is 91.964% $\pm$ 0.408, achieved with ε_train=0.02 and ε_	×	0.02

References

- http://arxiv.org/abs/2206.13405v2
- http://arxiv.org/abs/2512.03307v1
- http://arxiv.org/abs/2210.16365v1