BIOSerenity

Generalizable EEG-Based Alzheimer's Detection Using Foundation Models: Validation on Independent Clinical Cohorts

Philippe Di Piazza, Aurore Bussalb, Guillaume Jubien, Ulysse Gimenez

Abstract accepted as Poster presentation at the Recent Advances in Machine Learning for Healthcare Workshop
January 27th, Paris

PariSanté campus Inria

Introduction: The early diagnosis of Alzheimer's disease (AD) remains a major clinical challenge limited by the high cost and accessibility of gold-standard neuroimaging tools. Electroencephalography (EEG) offers a low-cost, portable alternative positioning as a promising candidate for AI-driven screening. However, conventional machine learning approaches are constrained by small, single-center datasets and high inter-subject variability (Aviles et al, 2024).

Foundation models, large neural networks pretrained on massive, diverse datasets, can overcome these bottlenecks by learning universal representations. While initial studies show state-of-the-art performance in EEG-based AD detection, their scalability and generalizability remain unproven, limiting their clinical translation (Wang et al., 2025). Here, we finetune a leading EEG foundation model one of the largest EEG-AD cohort to our knowledge and test its generalization on a fully independent, publicly available benchmark.

Methods : The dataset included 1,646 subjects (779 AD, 867 age and comorbidity matched controls) with 20-minute resting-state EEG collected from multiple clinical centers. Subject-independent splits (60% train, 20% validation, 20% test) were used. Final evaluation assessed performance on the held out internal test dataset and critically, on the full independent external dataset (n= 65) assessing real-world generalizability (Miltiadous et al., 2023).

We finetuned the BioSerenity-E1 foundation model (Bettinardi & Rahmouni, 2025), a transformer-based architecture pretrained on over 4,000 hours of clinical EEG to learn generalizable spatiotemporal representations. For binary AD classification, the pretrained encoder was frozen, and a classification head was trained. This classification head processes the encoder's token outputs through intra- and inter-attention mechanisms that model dependencies within and across the temporal and channel dimensions. The features were aggregated to produce record-level predictions specialized in AD pathology while preserving the foundation model's EEG representations.

Results The finetuned foundation model achieved balanced accuracies of 0.74–0.78 on the external dataset, depending on hyperparameter configuration, and 0.69–0.70 on the internal test set. Across all evaluation settings, performance was characterized by consistently high specificity (0.97–1.00) and moderate sensitivity (0.47–0.75). For comparison, a spectro-temporal transformer (Hata et al., 2025) was trained under the same protocol. It exhibited lower and more variable performances specifically on the internal dataset (balanced accuracies : 0.6 - 0.65).

Discussion: This study demonstrates robust generalization of an EEG foundation model for AD detection, validating its use on a large clinical cohort. The strong performance on the external dataset, coupled with higher specificity than sensitivity, confirms the model learns transferable neurophysiological signatures while highlighting the inherent difficulty of the diagnostic task. The advantage of foundation models is further demonstrated when compared to a strong contemporary baseline model trained under identical rigorous conditions.

While higher performances have been reported on the same dataset (accuracy = 91.34, Wang et al., 2025), those studies typically include data from this dataset in the training. In contrast, our methodology uses 100% of the external cohort as a strictly independent benchmark, which yields more moderate but clinically realistic performances.

Our work underscores the need for more rigorous validation protocols in EEG-based AD research. We acknowledge that a primary limitation is the reliance on clinician-derived labels rather than biomarker-confirmed diagnoses. Future work should incorporate biomarker-verified labels, explore advanced fine-tuning strategies to improve sensitivity, and pursue validation in larger multi-center cohorts. This work confirms the potential of foundation models for translating research-grade EEG algorithms into scalable and reliable tools for Alzheimer's disease assessment.

References

Aviles, M., Sánchez-Reyes, L. M., Álvarez-Alvarado, J. M., & Rodríguez-Reséndiz, J. (2024). Machine and Deep Learning Trends in EEG-Based Detection and Diagnosis of Alzheimer's Disease : A Systematic Review. Eng, 5(3), 1464-1484. https://doi.org/10.3390/eng5030078

Bettinardi, R. G., Rahmouni, M., & Gimenez, U. (2025). BioSerenity-E1 : A self-supervised EEG model for medical applications (No. arXiv:2503.10362). arXiv. https://doi.org/10.48550/arXiv.2503.10362

Hata, M., Yanagisawa, T., Miyazaki, Y., Omori, H., Hirashima, A., Nakagawa, Y., Eto, M., Yoshiyama, K., Kanemoto, H., Nyamradnaa, B., Yoshimoto, S., Ezure, K., Takahashi, S., & Ikeda, M. (2025). Accurate deep-learning model to differentiate dementia severity and diagnosis using a portable electroencephalography device. Scientific Reports, 15(1), 26304. https://doi.org/10.1038/s41598-025-12526-1

Miltiadous, A., Tzimourta, K. D., Afrantou, T., Ioannidis, P., Grigoriadis, N., Tsalikakis, D. G., Angelidis, P., Tsipouras, M. G., Glavas, E., Giannakeas, N., & Tzallas, A. T. (2023). A Dataset of Scalp EEG Recordings of Alzheimer's Disease, Frontotemporal Dementia and Healthy Subjects from Routine EEG. Data, 8(6), 95. https://doi.org/10.3390/data8060095

Wang, Y., Huang, N., Mammone, N., Cecchi, M., & Zhang, X. (2025). LEAD : Large Foundation Model for EEG-Based Alzheimer's Disease DetectionarXiv. https://doi.org/10.48550/arXiv.2502.01678