Parameter-Efficient Self-Supervised Adaptation for EEG-FM under Fixed Computational Budgets

Meghal Dani[1,3][0009−0007−1249−3630] and Stefanie Liebe[2,3,4][0000−0003−2873−2943]

[1] University of Tübingen, Germany
[2] Dept. of Neurology and Epileptology, University Clinic Tübingen, Germany
[3] Hertie Institute for AI in Brain Health (Hertie AI), Tübingen, Germany
[4] Hertie Institute for Clinical Brain Research, Tübingen, Germany
meghal.dani@uni-tuebingen.de

Abstract. EEG foundation models pretrained via self-supervised learning promise transferable representations, but their generalization remains limited, especially across diverse clinical datasets. Full fine-tuning is impractical for resource-constrained clinical settings due to high computational requirements. In this work, we investigate whether parameter-efficient self-supervised adaptation, updating only 9% of parameters suffices to align representations to target tasks. We evaluate our method on two state-of-the-art models with different pretraining objectives: BIOT (contrastive) and CBraMod (masked reconstruction), and evaluate on three clinical EEG datasets for abnormality detection (TUAB), event classification (TUEV), and seizure detection (CHB-MIT) under both in-distribution and out-of-distribution conditions. SSL adaptation yields consistent gains over linear probing, up to 20× AUCPR. Under a fixed compute budget, peak performance requires only 20–50% of available unlabeled data. Critically, when total window count is fixed, performance remains invariant to patient count, suggesting that performance is dependent on overall temporal window diversity only. Our findings demonstrate that parameter-efficient adaptation enables effective deployment of EEG Foundation models (EEG-FM) with minimal computational overhead and data collection burden.
Code available at: https://github.com/c3n-group/efficient-eeg-adapt

Keywords: EEG · Self Supervised Learning · Foundation Models.

1 Introduction

Electroencephalography (EEG) is the gold standard for non-invasive neurological monitoring and still heavily relies on labor-intensive manual expert inspection [14,5]. Deep learning algorithms using supervised training have addressed automating this process [20,3,2,7]. Although unlabeled EEG recordings are abundant at clinical sites, annotated data sets are rare, limiting these approaches.

In contrast, Foundation models (FMs), neural networks pretrained on massive unlabeled corpora via self-supervised learning (SSL), hold promise for not

Accepted at MICCAI 2026 (Strasbourg, France). Preprint of the submitted version.

requiring any expert annotations. A first generation of EEG foundation models (EEG-FMs) has emerged [11,23,10,9,22,21], demonstrating that broad pretraining on unlabeled EEG can yield transferable representations. Most models segment raw multichannel EEG into fixed-length time-series patches per channel, tokenized into embeddings that preserve both spatial and temporal structure, and processed by transformers. Their SSL objectives fall into two families: contrastive learning (e.g. BIOT [23]) and masked reconstruction (e.g. CBraMod [22]).

Despite this progress, deploying a pretrained EEG-FM to a clinical site requires bridging distribution shifts caused by differences in recording hardware, channel montages, and sampling rates. Retraining from scratch is impractical given the annotation costs, motivating parameter-efficient adaptation strategies. REMEDIS [4] established that SSL-based adaptation of pretrained models outperforms both direct transfer and full fine-tuning in medical imaging, particularly under limited labeled data. For EEG, concurrent work has been explored using graph-based adapters [19], while recent EEG-FMs such as LaBraM [10] and EEGPT [21] have scaled pretraining data, though with diminishing returns. Specifically, NeuroLM [9], trained on roughly ten times more data than LaBraM, achieves comparable downstream performance. This raises a fundamental question: *how much unlabeled target-domain data is actually needed for reliable EEG-FM adaptation?*

We address this by studying parameter-efficient SSL adaptation in a setting that mirrors clinical deployment: adapt a pretrained EEG-FM using only unlabeled target-domain EEG, then evaluate representation quality via frozen-encoder linear probing. We probe two EEG-FMs with different pretraining objectives, BIOT (contrastive) and CBraMod (masked reconstruction), updating only the final encoder layer while preserving each model's original SSL objective. To make data-efficiency conclusions actionable, we evaluate our adaptation under a normalized fixed-compute protocol [1] that holds total computation constant. Across three clinical EEG datasets covering both in-distribution (ID) and out-of-distribution (OOD) conditions, we find that: (1) lightweight SSL adaptation substantially improves linear probe representations with gains up to $20\times$ in AUCPR on rare-event tasks, (2) near-peak performance is achieved using only a fraction ($20-50\%$) of available unlabeled data, (3) the number of total temporal windows seen, rather than the number of unique patients, impacts adaptation performance under a fixed compute budget.

2 Method

2.1 Parameter Efficient SSL Adaptation

Our pipeline consists of two stages: (i) SSL-based representation learning for parameter-efficient adaptation, and (ii) linear probing for downstream evaluation.

Self Supervised Learning Pretraining: Let f_θ denote an EEG foundation model encoder that maps an input EEG segment $\mathbf{x} \in \mathbb{R}^{C\times T}$ (with C channels

and T time steps) to a representation $\mathbf{h} = f_\theta(\mathbf{x}) \in \mathbb{R}^d$. To effectively adapt the foundation model under resource-constrained settings, we adopt a parameter-efficient fine-tuning strategy: we partition the full parameter set into frozen lower layers θ_{frozen} and a trainable subset θ_{adapt} corresponding to the final encoder layer, such that $\theta = \{\theta_{\text{frozen}}, \theta_{\text{adapt}}\}$. We preserve the original SSL objective of each model, i.e., the masked reconstruction for CBraMod (with masking ratio 0.5 [22]), contrastive learning for BIOT and optimize the loss with respect to θ_{adapt} on an unlabeled target dataset $\mathcal{D}$, keeping θ_{frozen} fixed:

$$\min_{\theta_{\text{adapt}}} \mathcal{L}_{\text{SSL}}(\theta_{\text{frozen}}, \theta_{\text{adapt}}; \mathcal{D}) \tag{1}$$

This strategy retains the generic signal representations learned during pretraining in lower layers while enabling domain-specific calibration in the final layer.

Downstream Evaluation via Linear Probing: To evaluate representation quality, we employ linear probing [1,6,17] on a labeled downstream dataset $\mathcal{D}_{\text{task}}$ containing EEG segment-class pairs $(\mathbf{x}, \mathbf{y})$. For each EEG window $\mathbf{x}$, we extract representations from our pretrained encoder f_{θ^*} and train only a linear classifier g_ϕ to obtain $g_\phi(f_{\theta^*}(\mathbf{x}))$. Overall, the linear head is adapted by optimizing only for ϕ, keeping θ^* frozen as defined below:

$$\phi^* = \arg\min_{\phi} \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{D}_{\text{task}}}\Big[\mathcal{L}_{\text{task}}\big(g_\phi(f_{\theta^*}(\mathbf{x})), \mathbf{y}\big)\Big], \tag{2}$$

where $\mathcal{L}_{\text{task}}$ is the task-specific cross entropy loss.

2.2 Normalized Evaluation

To correctly assess the impact of data diversity, we must isolate its effects from variations in computational cost and hyperparameter optimization. Following established protocols for fair comparison under varying data regimes [1], we use: (1) a fixed computational budget to normalize computation across experiments, and (2) data diversity decomposed with respect to the number of unique EEG windows and (3) unique patients to characterize what the model observes or learns from, during the SSL adaptation.

Computational Budget and Data Diversity: We standardize adaptation compute using the number of optimizer update steps. For an unlabeled training set of data with size N_{total}, batch size B, and a chosen reference epoch-equivalent $\mathcal{E}$ for SSL, we define the steps per epoch as:

$$S_{\text{epoch}} = \left\lceil \frac{N_{\text{total}}}{B} \right\rceil, \qquad S = \mathcal{E} \cdot S_{\text{epoch}}, \qquad C = S \cdot B, \tag{3}$$

where S is the total number of optimizer updates and C is the computational budget, representing the total number of EEG windows "seen" during training. Equivalently, $C = N \times \mathcal{E}$, where N is the number of unique EEG windows in the adaptation set $\mathcal{D}$. The **repetition factor** r, defined as C/N or $\mathcal{E}$ quantifies how many times, on average, each unique window is processed during training.

Table 1. Performance comparison between linear probing (LP) versus our parameter-efficient SSL adaptation (SA). Results shown as mean ± SEM across 5 seeds. Blue subscripts indicate absolute improvement; best results in **bold**. Metrics: Balanced Accuracy / Cohen's Kappa / Weighted F1 (TUEV); Balanced Accuracy / AUCPR / AUCROC (TUAB, CHB-MIT). θ_{adapt}: adapted parameters (of total θ in EEG-FM).

Dataset	Method	Bal. ACC	AUCPR / Kappa	AUCROC / F1
BIOT ($\|\theta\|$=3.19M, $\|\theta_{\text{adapt}}\|$=0.92M)				
TUEV	LP	30.52 ± 1.21	33.86 ± 2.84	65.87 ± 1.64
	SA	**31.78 ± 0.46** $_{+1.26}$	**40.45 ± 1.41** $_{+6.59}$	**69.43 ± 0.63** $_{+3.56}$
TUAB	LP	63.93 ± 0.15	74.97 ± 0.13	75.08 ± 0.11
	SA	**70.86 ± 0.23** $_{+6.93}$	**77.39 ± 0.24** $_{+2.42}$	**78.34 ± 0.26** $_{+3.26}$
CHB-MIT	LP	50.74 ± 0.21	1.53 ± 0.03	46.93 ± 0.30
	SA	**52.71 ± 0.72** $_{+1.97}$	**31.52 ± 1.21** $_{+30.0}$	**83.08 ± 1.89** $_{+36.2}$
CBraMod ($\|\theta\|$=4.92M, $\|\theta_{\text{adapt}}\|$=0.44M)				
TUEV	LP	28.97 ± 0.30	35.82 ± 0.94	67.27 ± 0.42
	SA	**44.36 ± 1.28** $_{+15.4}$	**40.91 ± 1.02** $_{+5.09}$	**69.06 ± 0.62** $_{+1.79}$
TUAB	LP	59.36 ± 0.04	56.44 ± 0.05	62.72 ± 0.07
	SA	**66.37 ± 1.70** $_{+7.01}$	**77.16 ± 0.40** $_{+20.7}$	**77.43 ± 0.52** $_{+14.7}$
CHB-MIT	LP	57.87 ± 0.20	17.35 ± 0.24	54.14 ± 1.39
	SA	**61.37 ± 3.50** $_{+3.50}$	**24.62 ± 2.31** $_{+7.27}$	**83.67 ± 1.36** $_{+29.5}$

A model adapted with large r (small N, many epochs) sees limited diversity with high repetition, and vice versa.

For each dataset, we vary sample sizes $N \leq N_{\text{total}}$ (selecting roughly log-spaced fractions of the total data, rounded to multiples of batch size for stability), while adjusting $\mathcal{E}$ to maintain a fixed C. This protocol enables us to observe *how much unique data is needed to achieve peak performance under resource constraints*, allowing us to disentangle whether performance gains arise from increased repetition r (more epochs on limited data) or increased diversity (exposure to more unique samples).

Patient vs. Window Diversity. EEG datasets exhibit hierarchical structure: each patient contributes multiple temporal windows. To determine whether adaptation benefits more from increasing the number of patients or collecting longer recordings, we fix both total windows N and compute budget C while varying the number of patients P. For each value of P, we randomly sample P patients and extract windows from their available recordings such that the total number of windows $\sum_i W_i = N$ (where W_i is the number of windows from patient i, constrained by that patient's available data). Since N is fixed, smaller P implies more windows per patient on average (longer recordings from fewer patients), while larger P implies fewer windows per patient (brief recordings from many patients). We repeat each configuration with 3 data sampling seeds.

3 Experiments and Results

3.1 Dataset and Setup

Data and Preprocessing: We evaluate on three clinical EEG tasks: (i) abnormality detection using TUH Abnormal EEG Corpus (TUAB) [12], (ii) event

type classification using TUH EEG Events (TUEV) [8], and (iii) seizure detection in pediatric epilepsy patients using CHB-MIT [18]. TUAB and TUEV originate from Temple University Hospital and share recording infrastructure with CBraMod's pretraining corpus (TUEG), making them *in-distribution* (ID) for CBraMod and *out-of-distribution* (OOD) for BIOT. CHB-MIT is a pediatric dataset recorded at a different institution with different hardware, making it OOD for both models. For TUAB and TUEV we use the official train/test splits; for CHB-MIT (24 patients) we use patients 1–20 for training, 21–22 for validation, and 23–24 for testing.

We use the common 16 bipolar montage channels in the international 10-20 system following BIOT [23] to obtain clean and uniformly formatted data. Signals are band-pass filtered (0.3–75 Hz), notch-filtered (60 Hz), and resampled to 200 Hz. For SSL adaptation we segment the EEG signal into 30-second non-overlapping windows for all three datasets used in this study; while for downstream evaluation we use 10-second windows (TUAB, CHB-MIT) and 5-second windows (TUEV), following CBraMod [22].

Foundation Models and SSL Adaptation: We evaluate two EEG foundation models with distinct pretraining objectives: BIOT [23], a contrastive learning based model pretrained on a sleep EEG corpus (SHHS [16]) and a proprietary resting EEG dataset (PREST), and CBraMod [22], a masked reconstruction model pretrained on Temple University EEG data (TUEG) [15]. We establish baselines through linear probing (LP), where all encoder layers remain frozen and only a linear classification head is trained on labeled downstream data, evaluating off-the-shelf representation quality [17].

Training Details: For SSL Adaptation, we use only the *training set* of each target dataset using AdamW optimizer [13] with learning rate 3×10^{-5}, weight decay 5×10^{-2}, and $(\beta_1, \beta_2) = (0.9, 0.999)$, with a cosine annealing schedule (minimum learning rate 10^{-6}). For linear probe evaluation, we train the classifier for 10 epochs, with learning rate 4×10^{-4}. We use a batch size of 128 for TUAB and CHB-MIT, and 256 for TUEV.

Evaluation Metrics: For binary tasks (TUAB, CHB-MIT) we report Balanced Accuracy, AUCPR, and AUCROC, with AUCROC as the monitor score. For multi-class classification (TUEV, 6 classes) we report Balanced Accuracy, Cohen's Kappa, and Weighted F1, with Kappa as the monitor score. All results are reported as mean ± standard error of the mean (SEM) over 5 independent random seeds, on the test set, if not mentioned otherwise.

3.2 Parameter-Efficient SSL Adaptation on Clinical Tasks

Table 1 summarizes the comparison between EEG-FM linear probing (LP) and our SSL adaptation. Across all three clinical datasets, SSL adaptation (SA) improves performance ranging from +1.79 weighted F1 score (CBraMod on TUEV) to +36.2 AUCROC (BIOT on CHB-MIT). Importantly, SA updates only 0.92M of 3.19M parameters for BIOT (28.8%) and 0.44M of 4.92M for CBraMod (9.0%). Thus, we obtain better aligned feature representations using parameter efficient adaptation regardless of pretraining objective. For BIOT, which is OOD on all

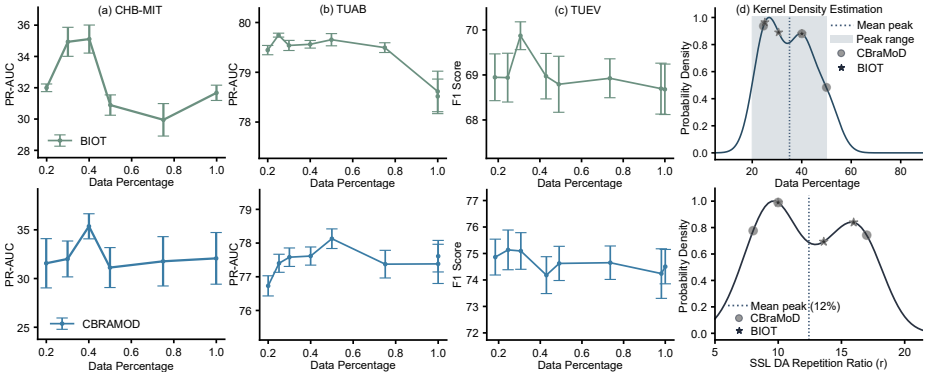

Fig. 1. SSL adaptation exhibits data efficiency under normalized compute $C{=}N{\times}\mathcal{E}$: performance peaks at $20-50\%$ of data then plateaus. (a) CHB-MIT seizure detection (AUCPR), N_{total}=4,174; (b) TUAB abnormality detection (AUCPR), N_{total}=57,614; (c) TUEV event classification (F1), N_{total}=86,587. (d) Kernel density estimation confirms saturation at N/N_{total}=0.2–0.5 (repetition factor $r{=}\mathcal{E}{=}12$). Green: BIOT; Blue: CBraMod. Error bars denote SEM (3 data shuffle $\times$ 3 model seeds).

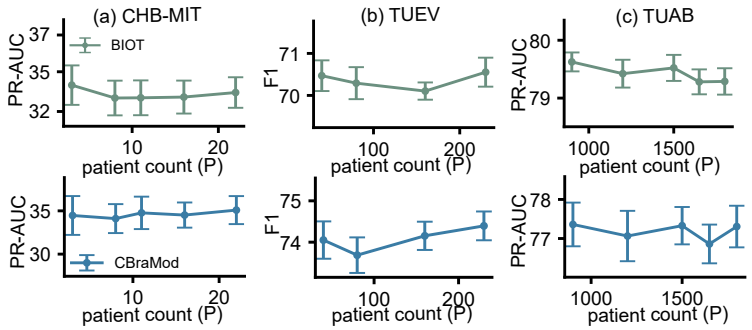

Fig. 2. SSL adaptation performance is patient-agnostic under fixed computational budget C. At constant $N{=}P{\times}W$, varying patient count P yields flat performance curve: (a) CHB-MIT seizure detection (AUCPR), $P{\in}[3, 8, 11, 16, 22]$; (b) TUEV event classification (F1), $P{\in}[40, 80, 160, 230]$; (c) TUAB abnormality detection (AUCPR), $P{\in}[900, 1200, 1500, 1650, 1800]$. Green: BIOT; Blue: CBraMod. Error bars denote SEM over 3 data shuffle $\times$ 3 model initialization seeds.

three datasets, adaptation yields +6.59 Kappa on TUEV, +3.26 AUCROC on TUAB (63.93 $\rightarrow$ 70.86% Balanced Accuracy), and +36.2 AUCROC on CHB-MIT seizure detection (46.93 $\rightarrow$ 83.08% AUCROC). For CBraMod, gains are observed on both ID tasks (TUAB: +14.7 AUCROC; TUEV: +5.09 Kappa) and the OOD task (CHB-MIT: +29.5 AUCROC: (54.14 $\rightarrow$ 83.67%)). These improvements indicate that pretrained representations are not sufficiently aligned to target-domain signal characteristics for direct deployment. The gains are particularly pronounced on OOD tasks, similar to findings in medical imaging where

SSL-based adaptation of pretrained models yields the largest improvements under distribution shifts [4].

CHB-MIT seizure detection task, a classical rare event with prevalence 1.4%, exposes a critical limitation of frozen foundation model features: EEG-FM linear probing achieves near-random AUCPR performance (BIOT: 1.53%, CBraMod: 17.35%), barely exceeding random classifier baseline precision (1.48%). We verified this result across all 5 seeds, confirming that frozen features fail to capture discriminative patterns for rare clinical events. SSL adaptation recovers clinically meaningful performance (BIOT: 31.52%, CBraMod: 24.62% AUCPR), demonstrating that even modest unlabeled target-domain training can align representations when frozen features provide no discriminative signal.

To investigate the trade-off between adaptation capacity and parameter efficiency, we selectively update different encoder layer subsets during SSL adaptation (final layer only, final two layers, or all layers). As shown for the CHB-MIT dataset in Table 2, we observe only marginal gains (+1.17% AUCPR) for BIOT and even slightly degraded performance for CBraMod (31.25% AUCPR $\rightarrow$ 29.41%), at the cost of significantly more parameters. Final-layer adaptation, thus, provides an optimal cost-performance balance. This is consistent with transfer learning principles that selective updates preserve pretrained knowledge while enabling domain-specific calibration [24].

Takeaway: *Across all datasets, models and pretraining objectives evaluated, last-layer SSL adaptation yields consistent gains (up to* $20\times$ *AUCPR), making it a highly efficient and lightweight adaptation method especially in cases where distribution shifts impact performance.*

3.3 Data Efficiency under Normalized Compute Budget:

We systematically vary the amount of unique unlabeled data (N) while holding total compute ($C{=}N{\times}\mathcal{E}$) constant across all adaptation runs. Figure 1 shows downstream performance as a function of data percentage for both models across all three datasets. Interestingly, both BIOT and CBraMod reach peak performance using only $20-50\%$ of the available unlabeled data across all tasks. On CHB-MIT, both models achieve $\approx$35% AUCPR at 40% of the data (N=86K, repetition factor r=12). On TUEV, both peak at $\approx$30% with r=14, then plateau. The kernel density plot (Fig. 1d), aggregated over all models, datasets, and seeds, confirms that the distribution of peak performance concentrates at N/N_{total}=0.2–0.5. Beyond this point, additional unique data provides no measurable benefit under fixed compute. The model sees each window fewer times without gaining sufficient new information to offset the reduced repetition. This is consistent with recent findings on pretraining data diversity for SSL in vision, where beyond a saturation threshold, increasing unique samples under fixed compute yields diminishing returns as models require sufficient repetition to consolidate learned features [1]. We note that this result characterizes the optimal diversity–repetition balance at a given compute budget.

Table 2. Layer unfreezing ablation on CHBMIT dataset with one seed. Unfreezing only the final layer (L) provides optimal parameter efficiency.

Model	Layers	Param.	AUCROC	AUCPR	Bal. Acc
BIOT	L	0.92M	88.67	35.62	60.62
	$L, L-1$	1.71M	88.44	35.05	60.82
	All	3.19M	88.60	36.79	**62.18**
CBRAMOD	L	0.44M	87.16	31.25	**67.60**
	$L, L-1$	0.84M	86.62	21.90	58.69
	All	4.4M	88.20	29.41	50.98

Takeaway: *Under fixed compute, peak SSL adaptation requires only* $20-50\,\%$ *of available unlabeled data with sufficient repetition of approx.* 12*, enabling rapid deployment with minimal data collection.*

3.4 Total EEG Window Count Dominates Over unique patient ID:

To disentangle whether adaptation benefits more from patient diversity or temporal coverage, we fix total windows $N{=}P{\times}W$ and compute C, then systematically vary the patient-window composition. Figure 2 shows downstream performance as a function of patient count P across all three datasets. Across the tested ranges (CHB-MIT: $P{\in}\{3, 8, 11, 16, 22\}$; TUEV: $P{\in}\{40, 80, 160, 230\}$; TUAB: $P{\in}\{900, 1200, 1500, 1650, 1800\}$), increasing number of unique patients, produces no significant change in downstream performance, suggesting that SSL objectives capture transferable temporal structure rather than patient-specific features.

Takeaway: *Under FM-initialized SSL adaptation with fixed samples and compute, total windows seen, not patient count, affects the performance, suggesting clinical sites can adapt effectively using longer recordings from smaller cohorts.*

4 Conclusion

We present a parameter-efficient SSL adaptation strategy for EEG foundation models that updates 9% of parameters, yet achieves substantial performance gains. Across three clinical tasks and two models with distinct architectures and pretraining objectives, adaptation consistently outperforms linear probing (AUCROC gains up to +36.2 points). On rare-event seizure detection (CHB-MIT), adaptation recovers clinically meaningful performance (31.52% AUCPR) from near-random baselines (1.53%), demonstrating that domain adaptation is essential before clinical deployment. We systematically characterize data requirements under fixed compute, finding that peak performance requires only $20-50\%$ of available data. Additionally, with fixed compute budget and total samples while varying patient-window composition ($N{=}P{\times}W$), no significant performance difference is observed, suggesting that for FM-initialized SSL adaptation, temporal

window coverage may matter more than patient diversity. We note that for one of the datasets (CHB-MIT), this observation is over limited patient range (3–22). Future work should compare alternative parameter-efficient pretraining methods, vary the compute budget C within our fixed-compute design to test the robustness of findings in this work across compute regimes, and test whether these data-efficiency patterns hold for general-purpose time-series foundation models and other medical time-series data.

Acknowledgments. This work was supported by the Else Kröner Fresenius Foundation, German Research Foundation (DFG): 493665037, SPP 2241 - PN 520287829, the Machine Learning Cluster of Excellence EXC number 2064/1 PN 390727645, Tübingen AI Center and BMFTR (01GQ2502). M.D. is a member of the International Max Planck Research School for Intelligent Systems Tübingen (IMPRS-IS). The authors thank C3N lab members, Prof. Dr. Jakob H. Macke and Julius Vetter for discussion and feedback.

Disclosure of Interests. The authors declare no competing interests.

References

1. Al Kader Hammoud, H.A., Das, T., Pizzati, F., Torr, P.H., Bibi, A., Ghanem, B.: On pretraining data diversity for self-supervised learning. In: European Conference on Computer Vision. pp. 54–71. Springer (2024)
2. Alhussein, M., Muhammad, G., Hossain, M.S.: Eeg pathology detection based on deep learning. IEEE Access **7**, 27781–27788 (2019)
3. Amin, S.U., Hossain, M.S., Muhammad, G., Alhussein, M., Rahman, M.A.: Cognitive smart healthcare for pathology detection and monitoring. IEEE Access **7**, 10745–10753 (2019)
4. Azizi, S., Culp, L., Freyberg, J., Mustafa, B., Baur, S., Kornblith, S., Chen, T., Tomasev, N., Mitrović, J., Strachan, P., et al.: Robust and data-efficient generalization of self-supervised machine learning for diagnostic imaging. Nature Biomedical Engineering **7**(6), 756–779 (2023)
5. Bitar, R., Khan, U.M., Rosenthal, E.S.: Utility and rationale for continuous eeg monitoring: a primer for the general intensivist. Critical Care **28**(1), 244 (2024)
6. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PMLR (2020)
7. Gemein, L.A., Schirrmeister, R.T., Chrabąszcz, P., Wilson, D., Boedecker, J., Schulze-Bonhage, A., Hutter, F., Ball, T.: Machine-learning-based diagnostics of eeg pathology. NeuroImage **220**, 117021 (2020)
8. Harati, A., Golmohammadi, M., Lopez, S., Obeid, I., Picone, J.: Improved eeg event classification using differential energy. In: 2015 IEEE Signal Processing in Medicine and Biology Symposium (SPMB). pp. 1–4. IEEE (2015)
9. Jiang, W.B., Wang, Y., Lu, B.L., Li, D.: NeuroLM: A universal multi-task foundation model for bridging the gap between language and EEG signals. In: The Thirteenth International Conference on Learning Representations (2025), https://openreview.net/forum?id=Io9yFt7XH7

10. Jiang, W.B., Zhao, L., Lu, B.L.: Large brain model for learning generic representations with tremendous eeg data in bci. In: International Conference on Learning Representations. vol. 2024, pp. 16405–16426 (2024)
11. Kuruppu, G., Wagh, N., Kremen, V., Varatharajah, Y.: Eeg foundation models: a critical review of current progress and future directions. Journal of neural engineering **23**(2), 021001 (2026)
12. Lopez, S., Suarez, G., Jungreis, D., Obeid, I., Picone, J.: Automated identification of abnormal adult eegs. In: 2015 IEEE signal processing in medicine and biology symposium (SPMB). pp. 1–5. IEEE (2015)
13. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2019)
14. Mushtaq, F., Welke, D., Gallagher, A., Pavlov, Y.G., Kouara, L., Bosch-Bayard, J., Van Den Bosch, J.J., Arvaneh, M., Bland, A.R., Chaumon, M., et al.: One hundred years of eeg for brain and behaviour research. Nature human behaviour **8**(8), 1437–1443 (2024)
15. Obeid, I., Picone, J.: The temple university hospital eeg data corpus. Frontiers in neuroscience **10**, 196 (2016)
16. Quan, S.F., Howard, B.V., Iber, C., Kiley, J.P., Nieto, F.J., O'Connor, G.T., Rapoport, D.M., Redline, S., Robbins, J., Samet, J.M., et al.: The sleep heart health study: design, rationale, and methods. Sleep **20**(12), 1077–1085 (1997)
17. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PMLR (2021)
18. Shoeb, A., Guttag, J.: Application of machine learning to epileptic seizure detection. In: International Conference on Machine Learning. pp. 975–982 (2010)
19. Suzumura, T., Kanezashi, H., Akahori, S.: Graph adapter of eeg foundation models for parameter efficient fine tuning. arXiv preprint arXiv:2411.16155 (2024)
20. Van Leeuwen, K., Sun, H., Tabaeizadeh, M., Struck, A., Van Putten, M., Westover, M.: Detecting abnormal electroencephalograms using deep convolutional networks. Clinical neurophysiology **130**(1), 77–84 (2019)
21. Wang, G., Liu, W., He, Y., Xu, C., Ma, L., Li, H.: Eegpt: Pretrained transformer for universal and reliable representation of eeg signals. Advances in Neural Information Processing Systems **37**, 39249–39280 (2024)
22. Wang, J., Zhao, S., Luo, Z., Zhou, Y., Jiang, H., Li, S., Li, T., Pan, G.: Cbramod: A criss-cross brain foundation model for eeg decoding. In: International conference on learning representations. vol. 2025, pp. 75310–75346 (2025)
23. Yang, C., Westover, M.B., Sun, J.: Biot: Biosignal transformer for cross-data learning in the wild. In: Thirty-seventh Conference on Neural Information Processing Systems (2023), https://openreview.net/forum?id=c2LZyTyddi
24. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? Advances in neural information processing systems **27** (2014)