Deep Learning on M/EEG Signals : Adapt Your Model, Not Your Preprocessing

Jarod LÉVY[1,2] Hubert Jacob BANVILLE[1] Jean-Rémi KING[1] Svetlana PINET[3,4] Jérémy RAPIN[1]
Stéphane D'ASCOLI[†, 1] Thomas MOREAU[†,2]

[1]Meta AI, Paris, France

[2]Université Paris-Saclay, Inria, Palaiseau, France

[3]Basque Center on Cognition, Brain and Language, San Sebastian

[4]Ikerbasque, Basque Foundation for Science, Bilbao

[†]These authors share last authorship.

Résumé – Cette étude examine l'impact du prétraitement des signaux EEG (électroencéphalographie) et MEG (magnétoencéphalographie) sur les performances de modèles de deep learning. Nos résultats montrent qu'un prétraitement minimal permet de réduire significativement le coût computationnel, tout en maintenant des performances comparables aux approches plus complexes, à travers divers modèles et jeux de données. Nos observations suggèrent que le choix du modèle a un impact plus déterminant sur les résultats que le degré de prétraitement appliqué.

Abstract – This study investigates the impact of preprocessing EEG (electroencephalography) and MEG (magnetoencephalography) signals on the performance of deep learning models. Our results show that minimal preprocessing significantly reduces computational cost while maintaining performance comparable to more complex approaches, across datasets and models. Our observations suggest that model choice has a more decisive influence on the outcome than the complexity of the applied preprocessing.

1 Introduction

The analysis of EEG and MEG signals typically involves a preprocessing stage, where methodological choices vary considerably between research groups. Traditionally, manual preprocessing has been favored [21], relying on expert-driven selection and cleaning of signals. However, this approach is strongly dependent on human expertise, leading to issues of reproducibility and efficiency. The field has yet to reach consensus : some argue that extensive preprocessing is indispensable [7], while others favor minimalist methods to preserve information contained in the raw signal [9]. Automated procedures such as PyPREP [3] and AutoReject [15] have been developed to standardize preprocessing.

These solutions nonetheless present limitations. For example, PyPREP is only partially generalizable to MEG signals. AutoReject, while effective in some cases, does not consistently eliminate artifacts and often needs to be combined with additional steps (e.g., high-pass filtering, ICA), thereby increasing pipeline complexity and the number of hyperparameters to tune. Furthermore, these methods induce substantial computational costs, making them unsuitable for training on large-scale datasets. In particular, when channel density is very low or very high, they may become either inapplicable or computationally prohibitive. PyPREP, for instance, relies on the RANSAC algorithm [10], whose complexity may scale quadratically with the number of channels.

In the context of deep learning, several studies [26] have shown that convolutional models trained on EEG signals with minimal filtering and basic normalization can achieve performance comparable to more complex approaches. Delpup et al. [24] similarly demonstrated that while minimal preprocessing remains necessary, excessive cleaning may even become counterproductive. However, their study is limited to classification tasks on pathological or motor datasets and compares only four handcrafted preprocessing strategies. Moreover, although they highlight the benefits of minimalist pipelines, the results were not systematically benchmarked against established baselines in the literature.

In this study, we introduce **SimplePrep**, a minimal preprocessing pipeline combining three widely used steps in EEG and MEG analyses [2, 6, 20, 26] : band-pass filtering, robust scaling, and clamping of extreme values. SimplePrep reduces the influence of artifacts without discarding data segments and is directly applicable to both EEG and MEG recordings. We compare its performance against PyPREP and AutoReject across three lightweight deep learning models applied to diverse tasks (cognitive, clinical, sleep, BCI) and datasets with between 2 and 306 sensors.

2 Methods

We evaluate the impact of preprocessing and model choice by comparing three methods and three architectures across nine datasets.

2.1 Datasets and Performance Metrics

Table 1 presents the nine datasets considered (2 MEG, 7 EEG), covering a range of tasks : image classification, motor decoding, sleep stage classification, keystroke decoding, and pathological signal detection. These datasets span 4 to 2329

TABLE 1 : **Summary of the datasets used in this study.**

Dataset	Task	Category	Signal	Nb. Channels	Avg. Session Duration (min)	Nb. Hours	Nb. Participants	Window Length (s)	Metric	Reference Method
THINGS-MEG [14]	Image	Cognitive	MEG	272	5.8	46.4	4	1	Pearson	[2]
PINET-MEG [20]	Typing	Cognitive	MEG	306	15.8	24.7	20	0.5	CER	[20]
PINET-EEG [20]	Typing	Cognitive	EEG	61	17	18	20	0.5	CER	[20]
THINGS-EEG2 [11]	Image	Cognitive	EEG	63	60	80	10	1	Pearson	[2]
TUAB (Lopez, 2017)	Pathology	Clinical	EEG	23	23	1141	2329	10	Bal. Accuracy	[12]
SleepEDF [16]	Sleep	Clinical	EEG	2	1361	3470	78	30	Accuracy	[19]
MMI [25]	Motor	BCI	EEG	64	2	49	109	4	Accuracy	[4]
BNCI2014_001 [27]	Motor	BCI	EEG	22	7	12	9	6	Accuracy	[5]
BNCI2014_002 [27]	Motor	BCI	EEG	15	4	7	14	6	AUC-ROC	[5]

participants and represent more than 5000 cumulative hours of recordings. For each task, signals were segmented into temporal windows, whose lengths were chosen according to prior literature. Windows were centered on events of interest, or uniformly distributed for clinical datasets.

Model performance was compared against state-of-the-art results reported on Papers With Code[1] or, where unavailable, against the best published results for each dataset. Our scores are expressed as a percentage of this reference performance, defined as the best performance obtained within our pipeline averaged across three distinct splits. Results are aggregated across datasets, with medians and interquartile ranges reported to reduce the influence of any single dataset. Preprocessing time was evaluated as the duration required to process one hour of data. All methods were run with default parameters, applied session by session.

Dataset partitioning followed the original authors' recommendations, typically subject-level splits, but occasionally more restrictive rules (e.g., by file path in image datasets).

2.2 Preprocessing Methods

1. **SimplePrep :** This method applies a band-pass filter (0.1–40 Hz) from the MNE-Python library [13], followed by session-wise scaling using the *RobustScaler* [23], and finally clamps extreme values (±16). SimplePrep is not a methodological innovation per se, but rather a formalization of a minimal pipeline combining three essential steps frequently observed in the literature.

2. **PyPREP :** The PyPREP pipeline consists of three stages : frequency filtering, identification of noisy channels, and robust rereferencing. For MEG, specific adaptations were required, including separate treatment of gradiometers and magnetometers and removal of the rereferencing step, which is unsuitable for MEG signals.

3. **AutoReject :** AutoReject relies on Bayesian optimization to determine thresholds for rejecting or correcting noisy segments. It combines trial rejection with automatic sensor repair. This method was applied only to the training set segments to avoid bias.

A scaler was added after PyPREP and AutoReject to ensure stable model training. When the number of channels was too low, only applicable steps were executed. Conversely, for very long sessions (over 30 minutes), recordings were divided into 10-minute segments to keep preprocessing times reasonable.

[1] https://paperswithcode.com/

2.3 Models

1. **EEGNet** [18] : a compact convolutional model tailored for EEG and MEG signals, widely used in BCI applications. We use the default configuration.

2. **BrainModule** [8] : a deeper convolutional architecture with residual connections and dilated convolutions, designed for cognitive decoding tasks.

3. **Green** [22] : a hybrid model combining wavelet-based convolutional kernels with Riemannian projections. Default parameters are used, except for BCI tasks where the original recommended configuration is applied. This model performs particularly well on clinical datasets.

3 Results

3.1 Overall Comparison

The results, summarized in Figure 1, highlight the efficiency of SimplePrep, which combines speed with high performance across datasets. SimplePrep is significantly faster than PyPREP (p = 0.039) and AutoReject (p = 0.031), while achieving superior performance in several model–preprocessing combinations. In particular, paired with the Green model, SimplePrep yields statistically significant improvements compared to PyPREP (p = 0.027) and AutoReject (p = 0.0078). With BrainModule, SimplePrep also outperforms PyPREP (p = 0.039). In all other comparisons, SimplePrep performs at least as well as PyPREP and AutoReject.

Using EEGNet with SimplePrep, the best median score was obtained, corresponding to 96% of reference performance, with reduced interquartile variability, reflecting consistently strong performance across datasets. Although slightly higher, this score is not significantly different from BrainModule (p = 0.054) or Green (p = 0.16). Importantly, preprocessing with SimplePrep is remarkably fast (median : 44.05 seconds), compared to AutoReject, which required 20,006 seconds with EEGNet to achieve 95% of reference performance. PyPREP achieved its best results with Green (87.08%) but at a cost of 1,532 seconds per hour of data. Overall, SimplePrep emerges as a fast, effective, and robust alternative to existing pipelines.

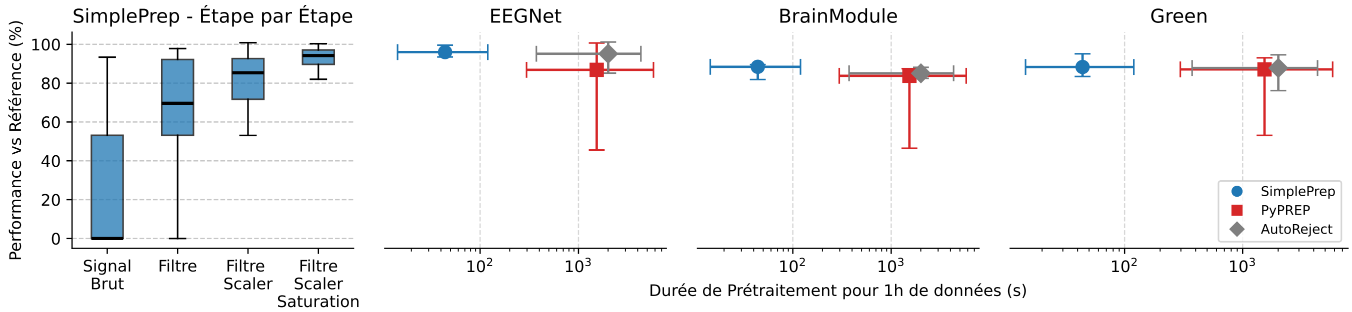

FIGURE 1 : **Step-by-step SimplePrep, performance and preprocessing time.**
Ablation of SimplePrep (left). Median performance and preprocessing time for EEGNet, BrainModule and Green with SimplePrep, PyPREP and AutoReject (right).

3.2 Ablation Study

We conducted an ablation study using EEGNet to evaluate the contribution of each component of SimplePrep. Results show progressive improvement with each added step. Raw signals proved unsuitable for direct modeling due to dominant low-frequency drifts, instrumental noise, and amplitude ranges varying by several orders of magnitude across subjects. Applying a band-pass filter (0.1–40 Hz) made the data more usable, but the most impactful improvement came from scaling : median performance increased from 69.65% ± 35.43 to 85.33% ± 16.86, a gain of 23%. This is largely explained by MEG signals, expressed in femto-tesla units, which without scaling operate near floating-point precision limits, destabilizing training. Adding clamping provided an additional 11% gain by mitigating high-amplitude artifacts, often caused by movement or electrical interference.

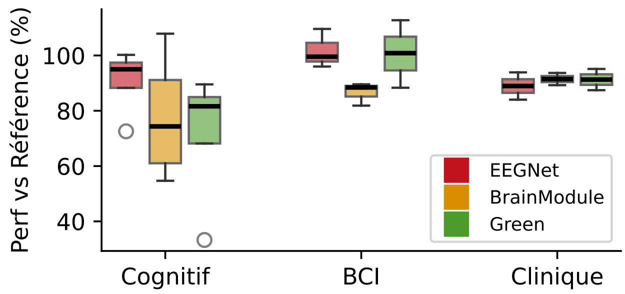

FIGURE 2 : **Performance (%) by model and dataset category.**

3.3 Performance by Task Category

Given dataset diversity, we further analyzed performance by task type, using SimplePrep as preprocessing (Figure 2). With only nine datasets, statistical comparisons are limited, but grouping into three categories is informative : **BCI** (BNCI2014_001, BNCI2014_002, MMI), **cognitive** (THINGS-MEG, THINGS-EEG2, Pinet-EEG, Pinet-MEG), and **clinical** (TUAB, Sleep-EDF).

EEGNet achieved the best performance on cognitive tasks, averaging 95% $\pm$ 10.73. For BCI tasks, Green performed best (100.84% $\pm$ 9.97). For clinical tasks, BrainModule outperformed others (91.47% $\pm$ 2.21). These results suggest emerging specialization across models, warranting further study with more datasets to identify features underlying these differences.

TABLE 2 : **Performance (%) by preprocessing method and number of sensors.** Best score in bold, second-best underlined.

Nb. sensors	**SimplePrep**	AutoReject [15]	PyPREP [3]
< 60	$\mathbf{94.09 \pm 1.35}$	91.82 ± 1.78	$\underline{93.19 \pm 0.77}$
≥ 60	$\mathbf{90.53 \pm 16.90}$	$\underline{84.61 \pm 15.22}$	55.78 ± 30.10

As shown in Table 2, preprocessing methods behave differently depending on the number of sensors. PyPREP performs well with fewer than 60 sensors but drops sharply (by more than 40%) with larger sensor arrays, partly due to limitations in handling MEG signals. AutoReject shows a similar trend, performing slightly better with fewer channels. In contrast, SimplePrep maintains consistent performance across both groups, with only a modest 4% median decrease.

4 Conclusion

This study highlights the effectiveness of minimal preprocessing for EEG and MEG in deep learning applications. It demonstrates the potential for lightweight pipelines better suited to large-scale deep learning. Several open questions remain. Future work should evaluate whether these findings generalize to other architectures such as ShallowConvNet [17] and ATCNet [1]. Another avenue is to study how model size influences preprocessing needs : EEGNet, a shallow model with only 1.4K parameters, already achieves strong results with minimal preprocessing. Further analysis of correlations between linear models and deep learning performance under different pipelines could also be insightful. Expanding the benchmark to include a broader set of tasks (BCI, clinical, cognitive), architectures, and datasets would provide stronger evidence for the robustness of these observations, and contribute to establishing standardized practices for the community.

Références

[1] H. Altaheri, G. Muhammad, and M. Alsulaiman. Physics-informed attention temporal convolutional network for eeg-based motor imagery classification. *IEEE transactions on industrial informatics*, 19(2) :2249–2258, 2022.

[2] H. Banville, Y. Benchetrit, S. d'Ascoli, J. Rapin, and J.-R. King. Scaling laws for decoding images from brain activity. *arXiv preprint arXiv :2501.15322*, 2025.

[3] N. Bigdely-Shamlo, T. Mullen, C. Kothe, K.-M. Su, and K. A. Robbins. The prep pipeline : standardized preprocessing for large-scale eeg analysis. *Frontiers in neuroinformatics*, 9 :16, 2015.

[4] J. Y. Cheng, H. Goh, K. Dogrusoz, O. Tuzel, and E. Azemi. Subject-aware contrastive learning for biosignals. *arXiv preprint arXiv :2007.04871*, 2020.

[5] S. Chevallier, I. Carrara, B. Aristimunha, P. Guetschel, S. Sedlar, B. Lopes, S. Velut, S. Khazem, and T. Moreau. The largest eeg-based bci reproducibility study for open science : the moabb benchmark. *arXiv preprint arXiv :2404.15319*, 2024.

[6] S. d'Ascoli, C. Bel, J. Rapin, H. Banville, Y. Benchetrit, C. Pallier, and J.-R. King. Decoding individual words from non-invasive brain recordings across 723 participants, 2024.

[7] A. de Cheveigné. Is eeg is better left alone ? *bioRxiv*, pages 2023–06, 2023.

[8] A. Défossez, C. Caucheteux, J. Rapin, O. Kabeli, and J.-R. King. Decoding speech perception from non-invasive brain recordings. *Nature Machine Intelligence*, 5(10) :1097–1107, 2023.

[9] A. Delorme. Eeg is better left alone. *Scientific reports*, 13(1) :2372, 2023.

[10] M. A. Fischler and R. C. Bolles. Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6) :381–395, 1981.

[11] A. T. Gifford, K. Dwivedi, G. Roig, and R. M. Cichy. A large and rich eeg dataset for modeling human visual object recognition. *NeuroImage*, 264 :119754, 2022.

[12] S. Gijsen and K. Ritter. Eeg-language modeling for pathology detection. *arXiv preprint arXiv :2409.07480*, 2024.

[13] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. S. Hämäläinen. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7(267) :1–13, 2013.

[14] M. N. Hebart, O. Contier, L. Teichmann, A. H. Rockter, C. Zheng, A. Kidder, A. Corriveau, M. Vaziri-Pashkam, and C. I. Baker. "things-meg", 2023.

[15] M. Jas, D. A. Engemann, Y. Bekhti, F. Raimondo, and A. Gramfort. Autoreject : Automated artifact rejection for meg and eeg data. *NeuroImage*, 159 :417–429, 2017.

[16] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. Kamphuisen, and J. J. Oberye. Analysis of a sleep-dependent neuronal feedback loop : the slow-wave microcontinuity of the eeg. *IEEE Transactions on Biomedical Engineering*, 47(9) :1185–1194, 2000.

[17] S.-J. Kim, D.-H. Lee, and S.-W. Lee. Rethinking cnn architecture for enhancing decoding performance of motor imagery-based eeg signals. *IEEE Access*, 2022.

[18] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance. Eegnet : a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of Neural Engineering*, 15(5) :056013, July 2018.

[19] S. Lee, Y. Yu, S. Back, H. Seo, and K. Lee. Sleepyco : Automatic sleep scoring with feature pyramid and contrastive learning. *Expert Systems with Applications*, 240 :122551, 2024.

[20] J. Lévy, M. Zhang, S. Pinet, J. Rapin, H. Banville, S. d'Ascoli, and J.-R. King. Brain-to-text decoding : A non-invasive approach via typing, 2025.

[21] R. Näätänen, A. Lehtokoski, M. Lennes, M. Cheour, M. Huotilainen, A. Iivonen, M. Vainio, P. Alku, R. J. Ilmoniemi, A. Luuk, et al. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385(6615) :432–434, 1997.

[22] J. Paillard, J. F. Hipp, and D. A. Engemann. Green : A lightweight architecture using learnable wavelets and riemannian geometry for biomarker exploration with eeg signals. *Patterns*, 2025.

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn : Machine learning in Python. *Journal of Machine Learning Research*, 12 :2825–2830, 2011.

[24] F. D. Pup, A. Zanola, L. F. Tshimanga, A. Bertoldo, and M. Atzori. The more, the better ? evaluating the role of eeg preprocessing for deep learning applications, 2024.

[25] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw. Bci2000 : a general-purpose brain-computer interface (bci) system. *IEEE Transactions on biomedical engineering*, 51(6) :1034–1043, 2004.

[26] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human brain mapping*, 38(11) :5391–5420, 2017.

[27] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. R. Müller-Putz, et al. Review of the bci competition iv. *Frontiers in neuroscience*, 6 :55, 2012.