

# Deep Learning sur les signaux M/EEG : Adaptez votre modèle, pas votre prétraitement

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

<sup>1</sup>Meta AI, Paris, France

<sup>2</sup>Université Paris-Saclay, Inria, Palaiseau, France

<sup>3</sup>Basque Center on Cognition, Brain and Language, San Sebastian

<sup>4</sup>Ikerbasque, Basque Foundation for Science, Bilbao

<sup>†</sup> Ces auteurs partagent la dernière position d'auteur.

**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

L'analyse des signaux EEG et MEG repose sur une étape de prétraitement dont les choix méthodologiques varient considérablement entre les équipes de recherche. Traditionnellement, un prétraitement manuel a été privilégié [21], fondé sur la sélection et le nettoyage des signaux. Cependant, cette approche, très dépendante de l'expertise humaine, présente des problèmes de reproductibilité et d'efficacité. La communauté scientifique peine encore à trouver un consensus : certains pensent qu'un prétraitement approfondi est indispensable [7], tandis que d'autres privilégient des méthodes minimalistes pour préserver les informations contenues dans le signal brut [9]. Des processus automatisés tels que PyPREP [3] et AutoReject [15] ont ainsi été développés pour standardiser le prétraitement.

Cependant, ces solutions présentent certaines limites. Par exemple, PyPREP n'est que partiellement généralisable aux signaux MEG. AutoReject, quant à lui, ne parvient pas toujours à éliminer efficacement certains artefacts et doit être combiné à d'autres approches (filtrage passe-haut, ICA), alourdissant ainsi le pipeline et le nombre de paramètres à choisir. De plus, ces méthodes induisent des coûts computationnels élevés, incompatibles avec un entraînement sur de grands ensembles de données. En particulier, lorsque la densité des canaux est très faible ou très élevée, elles deviennent soit inapplicables, soit exigent un temps d'exécution excessif. PyPREP utilise par exemple l'algorithme RANSAC [10], dont la complexité peut être quadratique en fonction du nombre de canaux.

Dans le contexte du deep learning, plusieurs études [26] ont montré que des modèles convolutionnels entraînés sur des

signaux EEG ayant subi un filtrage minimal et une normalisation (scaler) basique permettent d'obtenir des performances comparables à celles de méthodes plus complexes. Delpup et al. [24] ont également démontré que, même si un prétraitement minimal reste nécessaire, un nettoyage excessif peut devenir contre-productif pour l'entraînement de ces modèles. Cependant, leur étude se limite à des problèmes de classification sur des jeux de données pathologiques ou moteurs, et ne compare que quatre méthodes de prétraitement conçues manuellement. De plus, bien que l'intérêt d'un prétraitement minimaliste soit mis en avant, les auteurs n'ont pas comparé leurs résultats à des performances de référence issues de la littérature.

Dans cette étude, nous définissons SimplePrep, un prétraitement minimal rassemblant trois étapes classiques largement utilisées [2, 6, 20, 26] dans les analyses de signaux EEG et MEG : un filtre passe-bande, un scaler et une saturation des valeurs extrêmes. SimplePrep permet de réduire l'influence des artefacts sans exclure de segments de données, et s'intègre facilement aux traitements EEG comme MEG. Nous comparons ses performances à celles de PyPREP et AutoReject, en nous appuyant sur trois modèles légers, appliqués à des tâches variées (cognition, pathologie, sommeil, BCI) et à des jeux de données comptant entre 2 et 306 capteurs.

## 2 Méthodes

Nous évaluons l'impact du prétraitement et du modèle en comparant trois méthodes et trois architectures sur neuf jeux de données.

TABLE 1 : Résumé des jeux de données utilisés dans cette étude.

Dataset	Tâche	Catégorie	Signal	Nb. Canaux	Durée Moyenne Session (mn)	Nb. Heures	Nb. Participants	Longueur de Fenêtre (s)	Métrique	Méthode de référence
THINGS-MEG [14]	Image	Cognitif	MEG	272	5.8	46.4	4	1	Pearson	[2]
PINET-MEG [20]	Typing	Cognitif	MEG	306	15.8	24.7	20	0.5	CER	[20]
PINET-EEG [20]	Typing	Cognitif	EEG	61	17	18	20	0.5	CER	[20]
THINGS-EEG2 [11]	Image	Cognitif	EEG	63	60	80	10	1	Pearson	[2]
TUAB (Lopez, 2017)	Pathology	Clinique	EEG	23	23	1141	2329	10	Bal. Accuracy	[12]
SleepEDF [16]	Sleep	Clinique	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]

## 2.1 Jeux de Données et Performance

Le tableau 1 présente neuf jeux de données diversifiés (2 MEG, 7 EEG), couvrant plusieurs types de tâches : classification d’images, de gestes moteurs, de phases de sommeil, de touches saisies au clavier et de détection d’enregistrements pathologiques. Ces jeux de données concernent de 4 à 2329 participants et représentent plus de 5000 heures d’enregistrements cumulés. Pour chaque tâche, le signal est segmenté en fenêtres temporelles, dont la durée est fixée selon les recommandations de la littérature : ces fenêtres sont centrées sur les événements d’intérêt, ou uniformément réparties dans le temps pour les jeux de données cliniques.

Les performances des modèles sont comparées aux meilleures performances recensées sur le site Papers With Code<sup>1</sup> ou, à défaut, à la meilleure performance identifiée dans la littérature pour chaque jeu de données. Nos résultats sont exprimés en pourcentage de cette performance de référence, correspondant à la meilleure performance obtenue avec notre pipeline, moyennée sur trois partitions distinctes. Les résultats sont agrégés sur l’ensemble des jeux de données. Afin de réduire l’influence d’une étude spécifique, nous utilisons la médiane ainsi que l’intervalle interquartile pour résumer ces résultats. Le temps de prétraitement est évalué en mesurant la durée nécessaire pour traiter une heure d’enregistrement. Chaque méthode utilise les paramètres par défaut et le prétraitement est effectué session par session.

Le partitionnement des données suit les recommandations des auteurs : généralement par sujet, mais parfois selon des critères plus restrictifs, comme le chemin d’accès des fichiers pour les jeux de données d’images.

## 2.2 Prétraitements

1. **SimplePrep** : Cette méthode applique un filtre passe-bande (0,1–40 Hz) issu de la librairie MNE-Python [13], suivi d’une mise à l’échelle par session via le *RobustScaler* [23], puis d’une saturation des valeurs extrêmes ( $\pm 16$ ). SimplePrep ne constitue pas une contribution méthodologique en soi, mais formalise un prétraitement minimal, fondé sur la combinaison d’étapes essentielles fréquemment observées dans la littérature.
2. **PyPREP** : Le pipeline PyPREP est composé de trois étapes consistant en un filtrage fréquentiel, une identification des canaux bruyants et une recherche robuste du signal de référence (re-référencement). Pour rendre

cette méthode applicable aux signaux MEG, une adaptation spécifique a été réalisée, notamment en traitant séparément les gradiomètres et les magnétomètres, tout en supprimant la re-référenciation inadaptée au MEG.

3. **AutoReject** : AutoReject repose sur une optimisation bayésienne déterminant le seuil optimal pour conserver ou non un segment d’un capteur. Cette approche permet de combiner à la fois le rejet des essais et la correction des capteurs trop bruyants. À noter que cette méthode est appliquée uniquement aux segments de l’ensemble d’entraînement afin d’éviter tout biais.

Un scaler est appliqué en complément des approches PyPREP et AutoReject pour assurer la stabilité de l’entraînement du modèle. Lorsque le nombre de canaux est trop faible, seules les étapes applicables sont réalisées. À l’inverse, lorsque les sessions sont trop longues (plus de 30 minutes), les enregistrements sont découpés en segments de 10 minutes afin de garantir un temps de prétraitement raisonnable avant l’entraînement.

## 2.3 Modèles

1. **EEGNet** [18] : un modèle convolutionnel compact particulièrement adapté aux signaux EEG et MEG, utilisé ici avec ses paramètres par défaut et conçue initialement pour les applications BCI.
2. **BrainModule** [8] : une architecture convolutionnelle plus profonde intégrant des connexions résiduelles et gardant la même taille de représentations latentes grâce à des convolutions dilatées. Elle a été conçue pour du décodage cognitif.
3. **Green** [22] : un modèle hybride combinant des noyaux convolutifs à base d’ondelettes avec une projection dans l’espace riemannien. Les paramètres par défaut sont utilisés, à l’exception des tâches de BCI pour lesquelles une configuration spécifique est choisie selon les recommandations originales des auteurs. Il performe particulièrement bien sur les tâches cliniques.

## 3 Résultats

Les résultats, résumés en Figure 1, mettent en évidence la supériorité de l’approche SimplePrep, qui combine efficacement rapidité et performances élevées sur l’ensemble des jeux de données étudiés. SimplePrep est en effet significativement

<sup>1</sup><https://paperswithcode.com/>

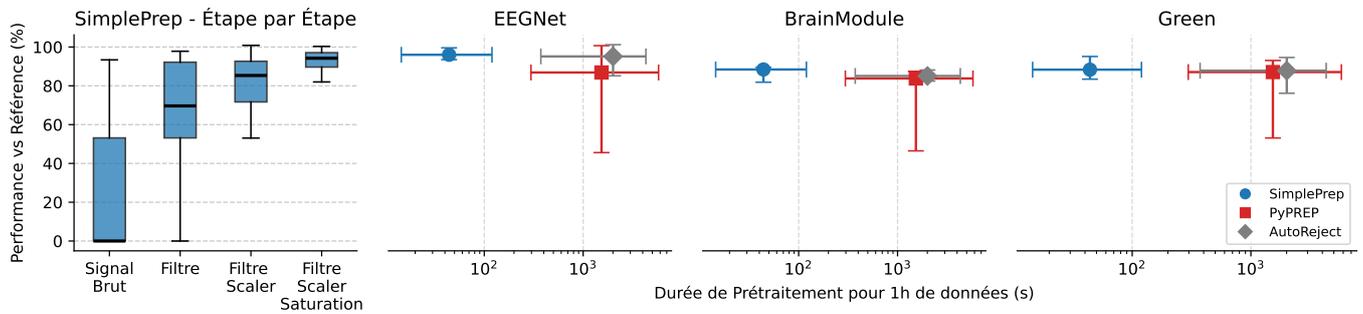


FIGURE 1 : SimplePrep étape par étape, performance et temps de prétraitement.

Ablation de SimplePrep (gauche). Comparaison des performances médianes et des durées de prétraitement pour EEGNet, BrainModule et Green en fonction des méthodes de prétraitement SimplePrep, PyPREP et AutoReject (droite).

plus rapide que PyPREP ( $p = 0,039$ ) et AutoReject ( $p = 0,031$ ), tout en obtenant des performances supérieures sur plusieurs combinaisons modèle-prétraitement. En particulier, associé au modèle Green, SimplePrep présente une amélioration statistiquement significative des performances par rapport à PyPREP ( $p = 0,027$ ) et AutoReject ( $p = 0,0078$ ). De même, pour le modèle BrainModule, SimplePrep permet d'obtenir des résultats significativement supérieurs à ceux obtenus avec PyPREP ( $p = 0,039$ ). Pour les autres combinaisons, SimplePrep affiche des performances comparables aux autres méthodes, sans différence significative, confirmant qu'il est systématiquement au moins aussi performant que PyPREP et AutoReject.

Le modèle EEGNet, utilisé avec SimplePrep, atteint la meilleure performance médiane, correspondant à 96 % des performances de référence, avec un intervalle interquartile réduit, soulignant une compétitivité constante à travers tous les jeux de données évalués. Cette performance, bien que légèrement supérieure, n'est toutefois pas statistiquement différente de celle obtenue avec le BrainModule ( $p = 0,054$ ) ou Green ( $p = 0,16$ ). Par ailleurs, la durée médiane du prétraitement avec SimplePrep est remarquablement courte (44,05 secondes), alors que la meilleure performance obtenue avec AutoReject nécessite un temps de prétraitement médian considérablement plus élevé (20 006,08 secondes pour EEGNet, atteignant 95 % des performances de référence). Enfin, PyPREP présente ses meilleurs résultats avec le modèle Green (87,08 %), mais requiert également un temps de prétraitement nettement supérieur (1 532,54 secondes). SimplePrep se positionne ainsi clairement comme une alternative rapide, efficace et robuste par rapport aux autres méthodes existantes.

Nous menons également une étude d'ablation avec EEGNet afin d'analyser l'impact de chaque étape de SimplePrep sur les performances. Les résultats montrent une amélioration progressive des performances à chaque ajout de composant du pipeline. Le signal brut est généralement inadapté à une analyse directe. En l'absence de traitement, les signaux contiennent des composantes basses fréquences dominantes, du bruit instrumental, et des amplitudes qui varient de plusieurs ordres de grandeur d'un sujet à l'autre. L'ajout d'un filtre passe-bande (0.1–40Hz) permet de rendre le signal exploitable, mais c'est l'ajout d'un scaler qui produit le gain le plus marquant : les performances médianes passent de 69.65% ± 35.43 à 85.33% ± 16.86, soit une amélioration de 23%. Ce gain s'explique en grande partie par le fait que les signaux MEG sont exprimés en unités physiques (femto-teslas), très petites en valeur absolue.

En l'absence de scaler, ces signaux sont traités dans des plages de valeurs qui sont proches de la précision flottante de Python et peuvent nuire à la stabilité de l'entraînement des modèles de deep learning. L'introduction de l'opération de saturation permet un gain additionnel de 11%. Cette étape agit comme une barrière contre les artefacts de forte amplitude, souvent causés par des mouvements ou des interférences électriques.

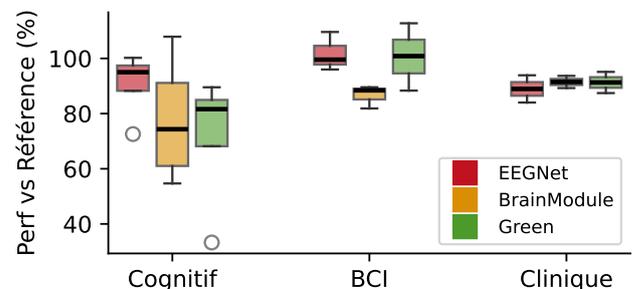


FIGURE 2 : Performance (%) selon le modèle et la catégorie du jeu de données.

Étant donné la diversité des jeux de données, il est pertinent d'examiner comment les performances varient selon le type de tâche considérée avec SimplePrep comme prétraitement. Malgré un nombre limité de jeux de données ( $N=9$ ) qui restreint les comparaisons statistiques, nous avons regroupé en Figure 2 ces jeux en trois catégories : **BCI** (BNCI2014\_001, BNCI2014\_002, MMI), **cognitif** (THINGS-MEG, THINGS-EEG2, Pinet-EEG, Pinet-MEG) et **clinique** (TUAB, Sleep-EDF). En comparant les performances par catégorie de tâche, on observe que EEGNet obtient les meilleurs résultats sur les tâches de décodage, avec une performance moyenne de  $95\% \pm 10.73$ . Pour les tâches de BCI, c'est le modèle Green qui se distingue, avec une performance de  $100.84\% \pm 9.97$ . Enfin, sur les tâches clinique, le BrainModule est le plus performant, atteignant  $91.47\% \pm 2.21$ . Ces résultats suggèrent une certaine spécialisation des modèles, qu'il serait pertinent d'examiner plus en profondeur à l'aide d'autres jeux de données, afin d'identifier les caractéristiques spécifiques qui les rendent plus performants sur certains sous-groupes de tâches.

Le tableau 2 permet également d'observer des comportements contrastés selon les méthodes de prétraitement utilisées et le nombre de capteurs à travers toutes les architectures. PyPREP offre d'excellentes performances lorsque le nombre de capteurs est inférieur à 60 mais devient la méthode la moins performante au-delà de ce seuil, avec une chute marquée

TABLE 2 : **Performances (%) selon le prétraitement et le nombre de capteurs.** Le meilleur score est en gras. Le deuxième meilleur score est souligné.

Nb. capteurs	<b>SimplePrep</b>	AutoReject [15]	PyPREP [3]
< 60	<b>94.09 ± 1.35</b>	91.82 ± 1.78	93.19 ± 0.77
≥ 60	<b>90.53 ± 16.90</b>	<u>84.61 ± 15.22</u>	55.78 ± 30.10

de plus de 40%. Cette chute peut s'expliquer en partie par l'incapacité de PyPREP de produire des résultats satisfaisants pour les signaux MEG. AutoReject suit une tendance similaire, avec une efficacité qui semble légèrement accrue lorsque le nombre de capteurs est réduit. À l'inverse, SimplePrep montre une remarquable constance entre les deux groupes avec une baisse de performance médiane limitée à 4% seulement.

## 4 Conclusion

Cette étude met en évidence l'efficacité d'un prétraitement minimaliste pour les signaux EEG et MEG dans les applications de deep learning. Elle ouvre ainsi des perspectives prometteuses pour le développement de pipelines allégés, mieux adaptés aux contraintes du deep learning à grande échelle. Plusieurs pistes restent toutefois à explorer. Il serait notamment pertinent d'évaluer la généralisation de cette approche à d'autres architectures couramment utilisées, telles que ShallowConvNet [17] et ATCNet [1]. Par ailleurs, l'influence de la taille des modèles sur la nécessité d'un prétraitement mérite d'être approfondie : EEGNet, modèle léger et peu profond (1.4K paramètres, profondeur = 2), montre déjà qu'un prétraitement réduit peut suffire à atteindre de bonnes performances. Une autre direction intéressante consisterait à analyser la corrélation entre les performances des modèles linéaires et celles des modèles de deep learning selon le pipeline appliqué. Un benchmark plus large, intégrant une diversité de tâches (BCI, clinique, cognitive), d'architectures et un plus grand nombre de jeux de données, permettrait de confirmer certaines observations de façon plus robuste. Il constituerait également une avancée utile pour la communauté et fera l'objet de travaux futurs.

## 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. Bencherit, 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. Bencherit, 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. Kabei, 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. Gijssen 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. Obery. 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. Schirrmester, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggenberger, 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.