



HAL
open science

Unsupervised Adaptive Deep Learning Method For BCI Motor Imagery Decoding

Yassine El Ouahidi, Giulia Lioi, Nicolas Farrugia, Bastien Padeloup, Vincent Gripon

► **To cite this version:**

Yassine El Ouahidi, Giulia Lioi, Nicolas Farrugia, Bastien Padeloup, Vincent Gripon. Unsupervised Adaptive Deep Learning Method For BCI Motor Imagery Decoding. European Signal Processing Conference (EUSIPCO), Aug 2024, Lyon, France. hal-04645175

HAL Id: hal-04645175

<https://imt-atlantique.hal.science/hal-04645175v1>

Submitted on 11 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Unsupervised Adaptive Deep Learning Method For BCI Motor Imagery Decoding

Yassine El Ouahidi, Giulia Lioi, Nicolas Farrugia, Bastien Padeloup and Vincent Gripon

IMT Atlantique

Lab-STICC, UMR CNRS 6285

F-29238 Brest, France

name.surname@imt-atlantique.fr

Abstract—In the context of Brain-Computer Interfaces, we propose an adaptive method that reach offline performance level while being usable online without requiring supervision. Interestingly, our method does not require retraining the model, as it consists in using a frozen efficient deep learning backbone while continuously realigning data, both at input and latent spaces, based on streaming observations. We demonstrate its efficiency for Motor Imagery brain decoding from electroencephalography data, considering challenging cross-subject scenarios. For reproducibility, we share the code of our experiments.

Index Terms—Electroencephalography, brain-computer interface, motor imagery, deep learning, online learning.

I. INTRODUCTION

Brain-Computer Interfaces (BCIs) have gained considerable attention in neuroscience research due to their potential for various applications, from assisting people with disabilities to enhancing human-computer interaction [1]. Among BCI paradigms, Motor Imagery (MI) has attracted interest for its medical and societal benefits. MI-based BCIs translate mental imagery of motor movements into actionable commands or in a feedback for neuromotor rehabilitation. As such, it opens up new ways of helping people with disabilities in their daily life.

Interpretation and classification of ElectroEncephaloGraphy (EEG) signals are central to many BCI systems. Unfortunately, EEG exhibit low signal-to-noise ratio, non-stationarity and high variability between sessions and subjects [2]. This make learning and classification challenging, especially when applying models trained offline to real-time BCI applications.

A common way to report performance in the literature consists in considering *offline* settings where one can benefit from all experimental data for calibration and optimization. On the contrary, *online* systems acquire data sequentially in real time, making it challenging to reach offline level performance. To overcome these difficulties, adaptive classifiers emerged as a promising solution. These classifiers update their parameters based on incoming EEG data, adapting to the evolving signal characteristics. One can distinguish two main types of adaptive classifiers: **supervised** or **unsupervised** [2]. Supervised adaptive classifiers rely on labeled sequential data, while unsupervised classifiers do not require explicit labels, making them more suitable for realistic real-time BCI applications.

Recent advances in Deep Learning have improved the effectiveness of BCI systems. Indeed, Deep Learning models are particularly more efficient than classical Machine Learning

alternatives for transferring knowledge between subjects. However, there are very few methods for adaptive Deep Learning classifiers, especially in the unsupervised setting.

In this work, we propose an unsupervised deep learning adaptive method for MI decoding and demonstrate its effectiveness using off-the-shelf benchmarks. Our method aims to bridge the gap between offline and online performance by dynamically updating normalization statistics throughout the considered Deep Learning architecture with each incoming data segment. The main contributions of this paper are:

- A deep learning method for unsupervised adaptive classification of MI signals, suitable for online setups;
- Experiments showing that our proposed method is competitive and achieves offline level performance, considering open source MI datasets and cross-subject scenarios;
- The code¹ to reproduce the experiments.

II. RELATED WORK

Machine Learning has significantly improved BCI systems, offering models that can effectively decode particular tasks. The MI paradigm is no exception, and remarkable progress has been made with the introduction of various EEG-specific BCI algorithms. In particular, methods using Riemannian geometry and Common Spatial Pattern (CSP) with automated feature selection have proven to be the most effective ones [2].

More recently, deep learning has introduced a promising way to exploit and transfer knowledge across sessions and subjects, and has emerged as a viable candidate solution to overcome the limitations of conventional Machine Learning methods. Multiple architectures have been proposed, including convolutional architectures with models such as Deep ConvNet, Shallow ConvNet [3], or EEGNet [4]. More recently EEG-SimpleConv [5], a simple 1D-CNN, was proposed to serve as a robust baseline for MI decoding. In our study we use the EEG-SimpleConv architecture and training procedure, as it shows state-of-the-art performance.

The development of classifiers adapting to the dynamic nature of brain signals has been driven by the challenge of non-stationarity. In short, adaptive classifiers aim to maintain or improve classification performance over time despite variations in signal characteristics. Both supervised and unsupervised

¹https://github.com/elouayas/eeg_adaptive

approaches to classifier adaptation have been proposed, each addressing various scenarios of BCI use. Supervised methods, which rely on labeled data to update the classifier, have shown promises but are limited by the availability of accurately labeled instances in real-time applications. On the contrary, unsupervised adaptive classifiers, which do not require labeled data, focus on updating the considered classifier to the evolving characteristics of the incoming data stream.

In the context of MI, unsupervised adaptation algorithms have been particularly highlighted for their potential in real-world BCI applications where the true intent of the user is not always explicitly known. Several methods applied to Linear Discriminant Analysis (LDA) and Gaussian mixture [6]–[8] propose to estimate the class labels of new incoming samples before adapting the classifier. A simpler alternative [9] updating the bias only has also been applied to LDA. In Riemannian approaches, an unsupervised adaptive Minimum Distance to Mean (MDM) classifier [10] has been proposed to retrain the classifier after each prediction. Another adaptive Riemannian classifier [11] does not require retraining but only recentering. As in [9], [11], our method needs no retraining of the model.

Not many adaptive approaches were proposed for Deep Learning classifiers. To the best of our knowledge, only a recent work [12] addresses this gap using various normalization techniques. We use this work as a comparison point.

III. METHODOLOGY

Machine learning models, especially Deep Learning ones, perform best when the training and test sets belong to the same distribution. However, this premise faces significant challenges while working with EEG data, particularly in BCI applications. Indeed, due to the inherent variability of EEG signals across subjects and sessions, a mismatch in data distributions is common, leading to suboptimal performance of BCI models. To address this issue, our study introduces a novel approach aimed at minimizing distributional discrepancies, thereby enhancing the utility of pre-trained offline neural network for new, unseen data and facilitating their application in an online context.

Our proposed methodology consists in using an unsupervised adaptive Deep Learning classifier, characterized by its ability to adapt to new data without the need for retraining. This is achieved by dynamically updating the test data statistics when a new EEG trial arrives. The process involves two key steps: first, the incoming trial is aligned using updated Euclidean Alignment (EA) [13] statistics; Then, batch normalization layers statistics within the neural network’s latent space are updated. This dual strategy of alignment and normalization is inspired by previous research demonstrating its effectiveness in MI decoding tasks [5]. By incorporating these adaptive elements into our model, we aim to bridge the distributional gaps between training and test data sets, thereby facilitating the use of Deep Learning models in real-time BCI applications.

A. Euclidean Alignment (EA)

EA enhances BCI models generalization by standardizing EEG data across different subjects and sessions into a

domain-invariant space. For a given subject, it aligns each session’s EEG trial data using the arithmetic mean of the covariance matrices. Through EA, data are transformed such that the mean covariance matrix of the trials is the identity matrix, significantly reducing dissimilarities across subjects and sessions, resulting in better homogeneity [13]. EA is a data transformation step that has proven its worth in BCI and especially in MI decoding, and could be considered as commonly accepted in the offline literature [14]. In our adaptive setup, EA plays a crucial role to recalibrate the data we feed into our classifier in real time, without the need for labeled data.

Formally, let us consider a subject with n trials, where each trial $\mathbf{X}_i \in \mathbb{R}^{C \times T}$ consists of $C \times T$ samples with C the number of channels and T the number of time samples. Let us define:

$$\bar{\mathbf{R}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T. \quad (1)$$

Alignment is performed by using the matrix square root of the arithmetic mean:

$$\tilde{\mathbf{X}}_i = \bar{\mathbf{R}}^{-\frac{1}{2}} \mathbf{X}_i. \quad (2)$$

EA is generally applied offline, as we often wait for all the trials from a given subject or session to become available to perform the alignment with the estimated matrix $\bar{\mathbf{R}}$. To adapt it to the online setting, we propose to recompute an updated version of the matrix $\bar{\mathbf{R}}$ each time a new trial arrives, performing the alignment $\tilde{\mathbf{X}}_i$ on the way.

B. Batch normalization (BN) “trick”

BN is a widely used and effective technique in Deep Learning. It consists of normalizing the activation within the neural network per batch. However, its behaviour during training and testing is different. During training, the running mean and variance are calculated and updated per batch. During evaluation, the learned and stored statistics from training are used to normalize the input. To adapt BN to online testing, we compute statistics directly from the test sets, rather than relying on stored training statistics. More specifically, we adopt an adaptive approach and, for each new incoming trial, we recompute the statistics with that trial and all previous ones from the same session. This allows session-specific statistics to be estimated and improves the ability of the model to adapt to individual users and non-stationarity.

C. Warm-up buffer of data

Our method relies on updating incoming test trials via normalization. However, to get good estimates of normalization statistics it is necessary to have several trials. This means that normalization of the first few trials may be inaccurate, due to imprecise estimation, resulting in a poor initial performance of our adaptive classifier. To overcome this, we propose to initially approximate those statistics using a buffer of data randomly sampled from the calibration set (if Fine-Tuning) or from other subjects (if Cross-Subject), depending on the evaluation settings (see Section III-D). This buffer and incoming

trials are used to compute the statistics up to a certain point. After acquiring a large enough number of trials, we switch to using only session-specific trials to get reliable statistics.

We choose to evaluate our method in the Cross-Subject and Cross-Subject with Fine-Tuning settings, and compare it to offline and online alternatives.

D. Evaluation Settings

In this study, we evaluate the capabilities of our adaptive approach to transfer knowledge across domains. One of the main interests of Deep Learning approaches over Machine Learning approaches lies in their ability to enable transfer learning [5]. More specifically, it allows the use of data from multiple subjects simultaneously to transfer the learned knowledge to a new subject. We consider the following scenarios:

- **Cross-Subject.** In this scenario, the model is trained on data from all but one subject, and its performance is evaluated on data from the omitted subject;
- **Cross-Subject with Fine-Tuning.** Here we pre-train the model on data from all but one of the subjects, then we fine-tune on a part of the data from the omitted subject. We then evaluate the performance of the model on the rest of the data from the omitted subject.

E. Offline, Online and Adaptive Setups

In the offline evaluation setup for MI classification, we consider a scenario in which we want to classify an ensemble of recorded trials, with access to all of them at any time, in an unsupervised way (without having access to their labels). In contrast, in the online evaluation setup, we make predictions on individual trials, one at a time. Consequently, we cannot use the rest of data from the test set for normalization or alignment. The online setup simulates a strict real-life application of a classifier. Due to the varying nature of EEG signals, using more trials from the same session greatly improves classification performance. In [5], we observe a gap of about 15% on a 4-class classification task on a MI benchmark dataset. The adaptive setup, which lies between the online and offline setups, simulates a realistic but less strict use of a classifier, where we have access to the incoming trial and the previous ones. In this setup, we make predictions on individual trials using also previously classified trials to update normalization statistics. Recall that we do not exploit trial labels, which makes the method unsupervised.

IV. EXPERIMENTS & RESULTS

A. Datasets and preprocessing

To validate our method, we consider two open access datasets. A small-scale dataset, BNCI [15] (the IIa dataset from the BCI competition) and a Large dataset [16], which contains 9 times more subjects than BNCI. In our experiments we also consider a subset of the BNCI dataset, BNCI2, which consists in EEG data belonging to only two classes (Right and Left MI) in order to allow a fair comparison to the Large dataset.

Each dataset has a unique recording setup with different devices and configurations, as detailed in Table I. In our experiments, we consider signals starting from the cue indicating the task to be imagined.

TABLE I: MI datasets considered. L = Left hand, R = Right hand, F = Feet, T = Tongue.

Dataset	BNCI	BNCI2	Large
Considered subjects	9	9	85
Sessions per subject	2	2	5
Trials per session	288	144	40
Total trials	5184	2592	17000
Classes	L/R/F/T	L/R	L/R
EEG electrodes	22	22	27
Sampling frequency (Hz)	250	250	512
Trial duration (s)	4	4	4.5

We used the very minimal preprocessing steps suggested in [5], [17]. The idea is to take advantage of the ability of Deep Learning models to work with raw data and leave them as much latitude as possible to learn: we apply a high-pass filter at 0.5 Hz, as recommended in [18], and resample to a lower sampling frequency to speed up the model’s inference, making sure to keep frequencies up to 40-50 Hz.

In the Cross-Subject with Fine-Tuning scenario, we use the sessions as follow: for BNCI, the first session in the omitted subject is used for fine-tuning and the second for evaluation; for the Large database, we use the first two sessions for fine-tuning and the last three for evaluation.

We follow the same training procedure as described in [5].

B. Main Results

Results are shown in Table II. We observe that our adaptive method is as effective as the offline baseline and highly outperforms the online baseline. This has been observed for every dataset and for every evaluation paradigm considered. A notable point is that we can observe that the results in the two two-class datasets, BNCI2 and Large, are similar, which is unsurprising considering the similarity of tasks and setups.

The line termed “Soft K-mean” in Table II refers to an alternative method of performing the classification in our adaptive method. The normalisation steps are the same as described previously. However, we perform a Soft K-mean clustering on the predicted class probabilities at the output of the model. The idea is to exploit the incoming trials during the evaluation session to refine and obtain better decision frontiers for the classification compared with the fixed classification layer. However, we can see that the result is similar and does not add significant value. This could be explained by the fact that the decision boundary is already adjusted by the normalisations, in particular the internal normalisation using the BN-trick.

Our method is able to achieve near-offline performance with only 10 to 20 trials (depending on the number of classes and the scoring paradigm). This corresponds to a EEG acquisition of about 30 to 60 seconds in a MI setup. Figure 1 provides more details on the behaviour of our method over time. In this Figure, we depict the cumulative accuracy over the incoming

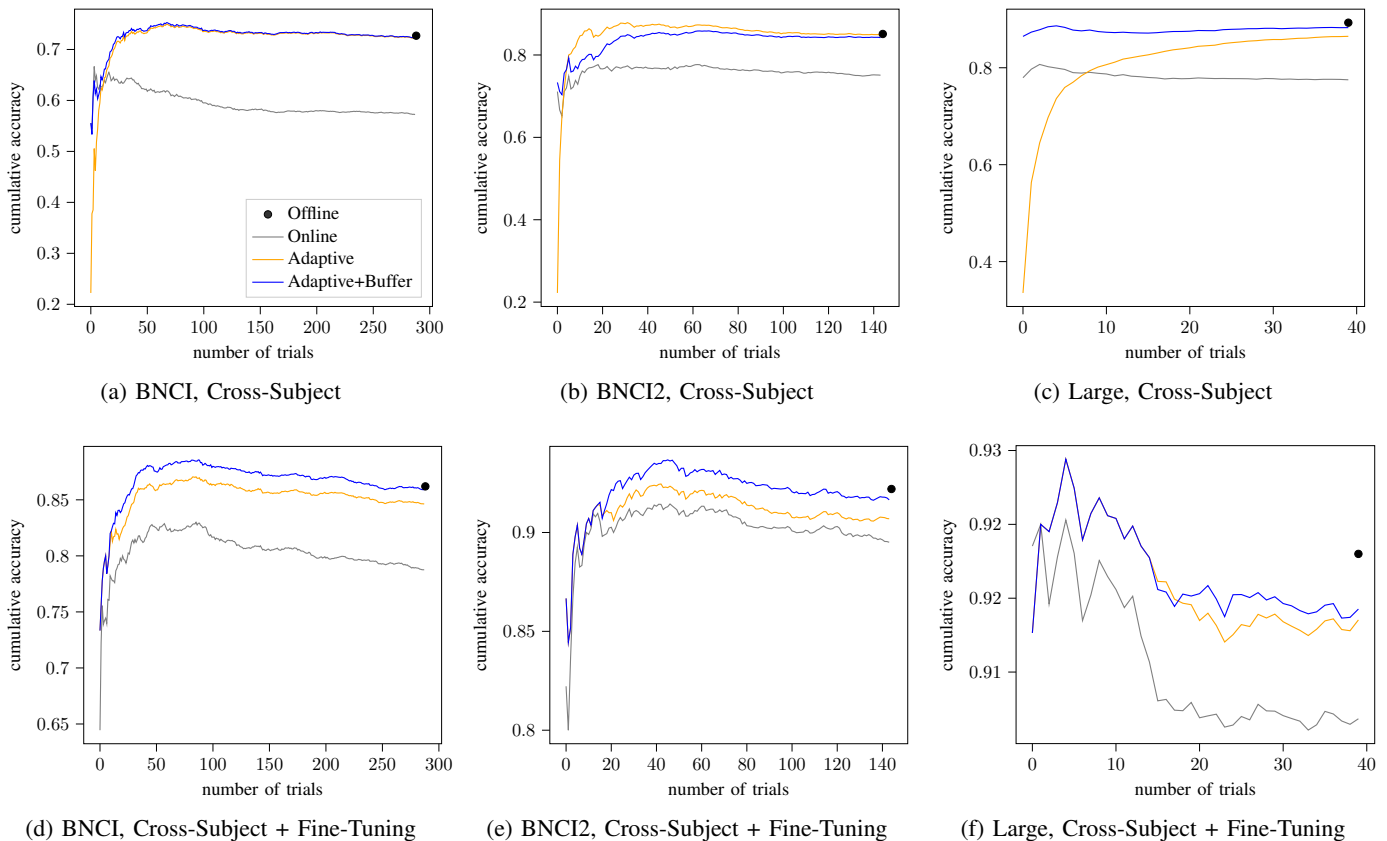


Fig. 1: Adaptive classifier performance compared to online and offline baselines.

trials, *i.e.*, the average accuracy up to this trial. We can observe the rapid convergence of the method to the offline baseline performance, together with its superiority over the online baseline at all given time steps.

TABLE II: Comparison of accuracies (%) of various methods.

	BNCI	BNCI2	Large
	Cross-Subject		
Online	56.9	75.1	77.5
Adaptive	72.1	84.9	86.5
Adaptive + Buffer	72.2	84.2	88.3
Offline	72.7	85.1	89.3
	Cross-Subject + Fine-Tuning		
Online	78.8	89.5	91.2
Adaptive	84.6	90.7	91.9
Adaptive + Buffer	85.9	91.7	91.9
Adaptive + Soft K-mean	84.6	90.6	91.8
Offline	86.2	92.2	92.3

TABLE III: Comparison to other unsupervised adaptive methods on BNCI Cross-Subject evaluation.

	RCT [11] ¹	Wimpff et al. [12]	Ours
Accuracy (%)	44.6	67.3	72.2

¹ Results reproduced by us

In Table III we can see that in the Cross-Subject setting, our method outperforms the other Deep Learning method [12],

by 6%. Since they have many similarities, this shows the importance of using a good backbone with the right training routine. Our method outperforms RCT even more by almost 30%. However, it should be pointed out that RCT was not designed to allow transfer between subjects, and was not meant to be used in a Cross-Subject scenario. We can assume that it would be way more efficient in a Within-Subject scenario.

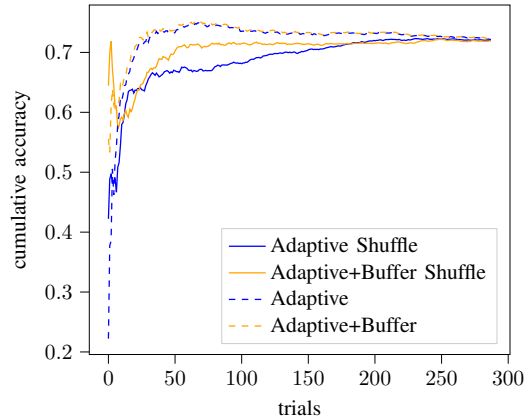
C. Additional experiments

1) *Warm-up buffer of trials*: We investigated the number of early trials for which using a warm-up buffer is beneficial. We observed that this had a marginal impact on the overall score, and found an optimum at around 10 trials. We also investigated the optimal buffer size. We found that 40 trials is enough, and we did not notice benefits while further increasing the size.

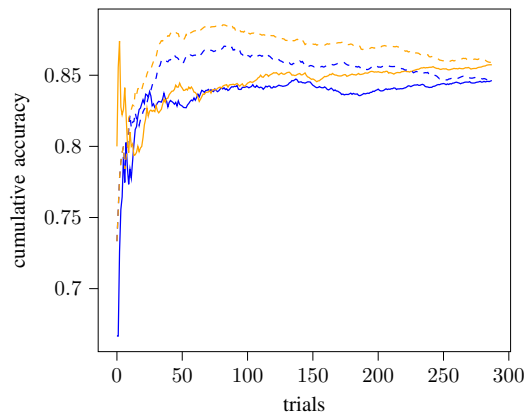
The conclusions from these experiments are that in all the setups and datasets, the addition of a buffer is harmless to performance and always provides a large gain at the beginning of the evaluation, when the first trials are classified. This is shown in Figure 1 where Adaptive+Buffer (blue) always outperforms Adaptive (orange) for the left part with few trials, and at the same performance level or above in the long run. We therefore recommend using a buffer to maximize performance and ensure reliable classification from the start.

Other sorts of buffers have been tried. For example, a fixed size buffer slid temporally in order to use only the last trials

to normalize, with the rationale that there might be variations in the signals within long sessions. However, this did not show significantly better results, at the cost of more complexity.



(a) Cross-Subject



(b) Cross-Subject + Fine-Tuning

Fig. 2: Adaptive classifier performance on BNCI compared to online and offline baselines.

2) *Hypothesis on user fatigue*: We can observe that classification performance of our adaptive method seems to decline over time. This can be seen in the dotted curves of Figure 2, which decreases after a certain point. We hypothesize that subjects experience fatigue at the end of the session, with poorer concentration during the final trials. MI is potentially mentally exhausting as BNCI sessions are long (288 trials).

To evaluate this hypothesis, we shuffled each session trials so that our adaptive method was applied to the trials in a random order rather than a chronological order. By doing that, we can see that the curves are strictly increasing. Also, the shuffled and unshuffled trials achieve the same final performance, but with different trends. This experience supports our hypothesis on user fatigue affecting performance.

V. CONCLUSIONS

We have proposed a method for the adaptive use of BCI deep learning backbones. Our method can be used in real time and does not require the model to be re-trained. It consists of

updating normalizations at both input and latent spaces. We have demonstrated its effectiveness on several MI datasets. The method can be particularly useful for faster calibration of a system in a few-shot setup. The effectiveness of the method remains to be demonstrated in a real online acquisition study.

REFERENCES

- [1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain–computer interfaces for communication and control,” *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.
- [2] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, “A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update,” *Journal of neural engineering*, vol. 15, no. 3, p. 031005, 2018.
- [3] R. T. Schirmer, 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*, vol. 38, no. 11, 2017.
- [4] 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*, vol. 15, no. 5, p. 056013, 2018.
- [5] Y. E. Ouahidi, V. Gripon, B. Pasdeloup, G. Bouallegue, N. Farrugia, and G. Lioi, “A strong and simple deep learning baseline for bci mi decoding,” *arXiv preprint arXiv:2309.07159*, 2023.
- [6] J. Blumberg, J. Rickert, S. Waldert, A. Schulze-Bonhage, A. Aertsen, and C. Mehring, “Adaptive classification for brain computer interfaces,” in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2007, pp. 2536–2539.
- [7] B. A. S. Hasan and J. Q. Gan, “Hangman bci: An unsupervised adaptive self-paced brain–computer interface for playing games,” *Computers in biology and medicine*, vol. 42, no. 5, pp. 598–606, 2012.
- [8] G. Liu, G. Huang, J. Meng, D. Zhang, and X. Zhu, “Improved gmm with parameter initialization for unsupervised adaptation of brain–computer interface,” *International Journal for Numerical Methods in Biomedical Engineering*, vol. 26, no. 6, pp. 681–691, 2010.
- [9] C. Vidaurre, M. Kawanabe, P. von Büna, B. Blankertz, and K.-R. Müller, “Toward unsupervised adaptation of lda for brain–computer interfaces,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 587–597, 2010.
- [10] S. Kumar, F. Yger, and F. Lotte, “Towards adaptive classification using riemannian geometry approaches in brain-computer interfaces,” in *2019 7th International Winter Conference on Brain-Computer Interface (BCI)*. IEEE, 2019, pp. 1–6.
- [11] L. Bougrain, S. Rimbart, P. L. Rodrigues, G. Canron, and F. Lotte, “Guidelines to use transfer learning for motor imagery detection: an experimental study,” in *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2021, pp. 5–8.
- [12] M. Wimpff, M. Döbler, and B. Yang, “Calibration-free online test-time adaptation for electroencephalography motor imagery decoding,” *arXiv preprint arXiv:2311.18520*, 2023.
- [13] H. He and D. Wu, “Transfer learning for brain–computer interfaces: A euclidean space data alignment approach,” *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, 2019.
- [14] B. Junqueira, B. Aristimunha, S. Chevallier, and R. Y. de Camargo, “A systematic evaluation of euclidean alignment with deep learning for eeg decoding,” *arXiv preprint arXiv:2401.10746*, 2024.
- [15] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. Mueller-Putz *et al.*, “Review of the bci competition iv,” *Frontiers in neuroscience*, 2012.
- [16] P. Dreyer, A. Roc, L. Pillette, S. Rimbart, and F. Lotte, “A large eeg database with users’ profile information for motor imagery brain-computer interface research,” *Scientific Data*, vol. 10, no. 1, p. 580, 2023.
- [17] Y. El Ouahidi, L. Drumetz, G. Lioi, N. Farrugia, B. Pasdeloup, and V. Gripon, “Spatial graph signal interpolation with an application for merging bci datasets with various dimensionalities,” in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.
- [18] A. Delorme, “Eeg is better left alone,” *Scientific Reports*, vol. 13, 2023.