ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

Machine Learning Methods for BCI: challenges, pitfalls and promises

Jaime Riascos Salas[1 2] and Marta Molina[3] and Fabien Lotte[4] *

1- PECoG, Postdam University, Germany
2- GITESI, Institucion Universitaria de Envigado (IUE), Colombia.
3- Department of Engineering Cybernetics, NTNU, Norway
4- LaBRI, Inria Center at the University of Bordeaux, France

Abstract. The development of Brain-Computer Interfaces (BCIs) has been constrained by a predominant focus on signal classification. This paper rather emphasizes the integration of neurophysiological principles, BCI paradigm selection, and rigorous experimental design. By addressing common pitfalls in Machine Learning implementation, we provide researchers with a tutorial and robust framework for BCI development, promoting reproducibility and rigor. Furthermore, by tackling challenges at the intersection of BCI and Machine Learning, this work contributes to the advancement of practical, real-time BCI applications.

1 Introduction

Brain-computer interface (BCI) proposes direct communication with machines using brain signals, typically electroencephalography (EEG). Since its first appearance [1], BCI has been widely used in applications such as communication and control for people with severe motor disabilities, neurorehabilitation, prosthesis and robotics, entertainment, etc [2]. Traditionally, creating an EEG-based BCI has been focused on the data pipeline: data acquisition, preprocessing, feature extraction, classification, and feedback. Each stage has its challenges and highly depends on the paradigm and neurophysiological patterns to be used. In recent years, a vast amount of research and advances have been made in the accurate decoding of EEG, ranging from exhaustive evaluation of traditional Machine Learning (ML) methods [3], deep learning, and generative models to explore different feature extraction techniques, data modeling and representation. While ML is an angular piece of the BCI pipeline, before delving into ML methods, several aspects of experimental BCI design require our attention.

Nevertheless, the efficacy of machine learning in BCI is intrinsically linked to a well-designed experimental protocol, incorporating carefully chosen paradigms and tasks that minimize confounding factors. Such a robust foundation is indispensable for any ML method, and common pitfalls encountered during model training must be diligently addressed. While the conventional BCI pipeline often prioritizes classification over data processing and feature extraction, this represents a gap from the original vision of BCIs [1], which proposed the integration of neurophysiological considerations at every stage (BCI Framework). We offer

*J.R.S. thanks Colegio Colombiano de Neurociencias (COLNE) for supporting him in this research. F.L. was supported by ANR project PROTEUS (grant ANR-22-CE33-0015-01).

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

a discussion around the impact of experimental factors on BCI development, addressing common pitfalls and suboptimal practices. Furthermore, we explore the advancements and open challenges in machine learning for BCI research, providing insights into the future of reliable BCI applications.

2 Traditional BCI pipeline

Research in EEG-based BCI has followed the traditional pipeline: signal acquisition, signal processing, feature extraction, classification, and feedback.

2.1 Signal acquisition

EEG involves recording the brain's electrical activity using an amplifier and an electrode cap, typically positioned according to the International 10-20 system or its high-density extensions. The choice of electrode type and number of channels is contingent upon available resources and the specific research question. While high-density systems with sintered Ag/AgCl electrodes ($\geq$32 channels) offer superior spatial resolution [4], studies targeting well-localized brain regions associated with specific paradigms may employ fewer channels.

2.2 Signal processing

EEG data frequently contains artifacts from various sources, including ocular and muscle movement, heartbeat, and power line noise. Digital processing techniques are commonly employed to improve the signal-to-noise ratio (SNR). These techniques include digital filtering using Finite/Infinite Impulse Response (FIR/IIR) filters to attenuate specific frequency bands. Furthermore, subsampling reduces the number of samples in the recorded signal, often to mitigate computational demands or enhance signal quality [5]. Spatial filtering methods, such as re-referencing and normalization, can also be applied to remove common-mode noise and standardize signal amplitudes. Statistical approaches such as Independent Component Analysis (ICA) and Principal Component Analysis (PCA) are helpful for the removal of artifacts by separating independent sources of activity within the EEG signal [5]. Although these statistical techniques are also used for feature extraction and dimensionality reduction, their application in artifact removal significantly improves the quality of EEG data.

2.3 Feature Extraction

Feature extraction transforms high-dimensional EEG signals into a concise representation of the user's mental state or intent. This process involves identifying and quantifying the significant patterns in the EEG data, such as spatial distributions, spectral power, and temporal dynamics [5]. By reducing dimensionality and focusing on relevant information, feature extraction enhances the efficiency and accuracy of subsequent classification algorithms, enabling BCIs to decode user intentions and translate them into actionable commands effectively.

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

2.4 Machine Learning classifier

The feature vector, derived from the previous feature extraction stage, serves as input to a classifier. This classifier assigns a class label to the set of features, effectively identifying the user's mental state. A classification model is initially trained on a collection of labeled observations (supervised learning), establishing a relationship between feature vectors and their corresponding mental states. Subsequently, this model is employed to classify new, unseen data. Historically, commonly used classification algorithms in BCIs include linear classifiers (e.g., Linear Discriminant Analysis or Support Vector Machines), neural networks, Bayesian classifiers, and nearest-neighbor classifiers [3]. Recently, more advanced and modern techniques such as Deep Learning or Riemannian Geometry are being used (see Section 4). The accuracy of these classifiers significantly impacts the overall success of a BCI system. Therefore, selecting an appropriate classification technique should be guided by the nature of the extracted features and the number of mental states (classes) to be distinguished.

2.5 Feedback

In a closed-loop BCI system, the feedback application provides real-time information to the user of their performance in executing a mental task. This feedback could be a speller program, a robotic prosthesis, a game, or a simple progress bar. This enables users to learn how to use and improve their BCI control abilities with practice [6]. The application utilizes the output from the classifier to execute actions or events corresponding to the user's inferred mental state. In contrast, in an offline BCI, users do not directly control external devices. Instead, the brain activity is recorded for subsequent analysis, often while receiving stimuli designed to elicit specific brain responses and improve classification performance. Importantly, feedback in online BCI and stimuli in offline BCI can be time-stamped and integrated into recorded data as markers, allowing precise labeling of events for subsequent analysis [7].

3 Common pitfalls in BCI design and evaluation

BCI design is a highly multi-disciplinary and complex effort, there are many ways to make honest mistakes and design a BCI that is either suboptimal or even scientifically invalid and biased. Unfortunately, such mistakes are regularly seen, even in published papers. Here, we would thus like to raise awareness about the most common pitfalls and mistakes that can be made during BCI design, so that our readers can avoid, many of such pitfalls involving the machine learning components of BCIs. Note that we report only the most common ones, as a more comprehensive overview would require a dedicated full paper. Interested readers can refer to [4] and [8] for a more in-depth treatment of some of them.

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

3.1 Experiment design

Effective BCI development requires careful experimental design, balancing data quality with user comfort. Understanding neurophysiological patterns evoked by cognitive processes and stimuli is crucial. These patterns, classified as exogenous (externally-triggered, simpler control) or endogenous (internally-generated, nuanced control), significantly influence BCI use due to their unique temporal and spatial characteristics, e.g., latency, amplitude or distribution.

Two primary BCI paradigms leverage distinct neurophysiological patterns, each with specific strengths and limitations, influencing their suitability for different user tasks. Event-related potentials (ERPs), such as P300 or Error Related Negativity, offer insights into cognitive processes but may be susceptible to noise. Steady-state Visual/Auditory Evoked Potentials (SSVEPs/SSAEPs) provide robust control options but require external stimulation. Sensorimotor Rhythms (SMRs), particularly Mu and Beta rhythms, enable intuitive motor-based control but can be challenging to detect and modulate reliably.

Effective BCI design requires a deep understanding of neurophysiology, EEG patterns, and BCI paradigms, as these directly influence feature extraction and classification. Critical consideration, often overlooked, include balancing session duration, trial lengths, and epoch numbers to ensure data quality and user engagement. Task selection should align with research goals and user capabilities. Incorporating pilot testing, real-time feedback, and participant feedback enables adaptive, user-centric protocols, enhancing the translational potential of BCI research.

3.2 Training and evaluating machine learning models

As mentioned earlier, ML models are often at the heart of BCI designs and are regularly used to infer BCI users' mental states from their EEG. Thus, there is a need to ensure that the training and evaluation of these ML algorithms are done soundly [8, 4]. Unfortunately, with many BCI designers lacking a strong ML background, this has led to various invalid results being published [9]. More recently, with the steep increase in Deep Learning (DL) use, the issue has become even more severe, with many DL BCI results proving to be non-reproducible and/or invalid [9, 10]. To avoid such issues, the main critical point that should be satisfied - but is unfortunately not always so - is that any BCI ML component should be trained on a given training set only, and evaluated on a testing set that is completely different and independent from the training set [8, 4]. This testing set should only be used for evaluating the trained models and for nothing else. While the vast majority of BCI designers know that in theory, some do not seem to realize the full practical implications.

A common mistake is to consider that only the classifier parameters should be trained on the training set. In contrast, there are often other ML components in a BCI than the classifier, e.g., data-driven spatial filters (ex: Common Spatial Patterns), which also have be trained on the training set only. Moreover, ML algorithm hyperparameters (e.g., regularization parameters or DL network

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

architecture) should also be selected on the training set only. Overall, all ML steps should performed on the training set only, which notably includes:

- Hyperparameters selection (including DL architecture selection): hyperparameters should be selected according to the training set only (and/or possibly a validation set that is independent of the testing set), but NOT according to, e.g., the performance obtained on the testing set.
- Feature & channel selection: features and channels should be selected according to the training set only (not according to the testing set nor according to all the data available, which includes the testing set).
- Classifier and feature choice: the choice of the classifier type (e.g., Support Vector Machine or DL) and/or features (e.g., band power or EEG time points) should also be selected based on the training set only, and not based on which ones lead to the highest test set classification performance.

Another common mistake is a misunderstanding of what having the training set and the testing being independent means. In particular, while having these two sets being different is indeed required for them to be independent, this may not be enough: they also need to be statistically independent. This means, e.g., that the training and testing samples should not come from the same experimental block or overlap with each other. For this reason, one should be careful when using Cross-Validation (CV), and refrain from using fully random CV, to avoid having training and testing epochs overlapping or coming from the same block by chance. Rather, a full block should be used either only in the CV training fold or in its testing fold, but not be split across both.

A good way to avoid such mistakes is to try to simulate online BCI use as much as possible when training and assessing a BCI ML model offline. Indeed, in an online BCI use, the training set is the previously collected EEG data, while the test set is the future data, that will be acquired during BCI use. The test set is thus not available for training the models. Ideally, BCI ML algorithms should thus be trained and assessed in a way that could be performed in an actual online BCI use. If this is not possible, the ML training or evaluation is probably biased and unsound, or at least not relevant for practical BCI use.

3.3 Reporting measures and results

When assessing BCI classification/regression performances, and reporting the results, there is a need to 1) report appropriate performance metrics, and ensure that the reported results are 2) reproducible, 3) free of confounding factors, and 4) statistically relevant. For reporting appropriate performance metrics, one should consider the number of classes and the balance between classes, as not all performance metrics can deal with more than two classes or with unbalanced classes (with different numbers of testing epochs per class). In particular, while classification accuracy is fine for balanced classes, it is not for unbalanced classes. Metrics like Area Under the ROC curve (for two classes) or balanced

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

accuracy should be used instead [11]. Please also note that while CV is very popularly used, it tends to overestimate BCI performances [11]. To ensure reproducible results, the code and data used should ideally be shared [10]. If not feasible, all the parameters and hyperparameters of the algorithms used have to be described, along with how they were selected, and based on which data. To ensure results that are free of confounding factors, it is recommended to check which neurophysiological information is used by the ML models, to make sure they are using the cortical EEG features targeted, and not, e.g., muscle artifacts, effects of fatigue [4, 8] or brain responses to cues [12]. Finally, one should ensure the reported results are statistically relevant. This means 1) verifying that the reported performance is actually statistically better than chance, such chance level varying according to the amount of testing data available [13]; and 2) if comparing classifiers, using a statistical test to ensure any apparent difference in performance is real and not due to chance - the variability of BCI data being high, especially due to the small amount of testing data usually available.

4 Machine learning promises, new direction and open challenges for machine learning in BCI

Recently, new promising directions for ML in BCI have emerged, in particular with the advent of DL and Riemannian classifiers [3], as well as several open challenges that are still to be solved. We describe them below.

4.1 Deep Learning

Deep learning has rapidly emerged as a powerful paradigm in BCI research. Unlike traditional methods, it autonomously discerns intricate and complex features within brain signals, skipping the need for manual features. Likewise, these models handle effectively the inter- and intra-subject variability in neural patterns, creating suitable decoders for BCI, for a more direct and efficient translation of neural signals into actionable outputs [14]. DL has notably enhanced the precision and intuitiveness of motor imagery decoding, enabling more natural and effective control of assistive devices like prosthetic limbs. Furthermore, its ability to recognize emotional states from EEG signals promises to transform affective computing and mental health monitoring [15]. While DL offers numerous advantages, it is not without challenges. The demand for large, labeled datasets, inherent model complexity leading to interpretability difficulties, and the need for real-time processing in practical BCI applications pose considerable limitations.

4.2 Riemannian geometry

The last decade also witnessed the rise of Riemannian Geometry-based Classifiers (RGCs) in BCI [16]. RGCs are EEG classifiers that represent EEG signals as Symmetric Positive Definite (SPD) matrices - typically Spatial Covariance Matrices (SCM) measuring the covariance between each pair of EEG channels

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

- and manipulate and classify such matrices using a dedicated geometry: the Riemannian geometry. These last few years, RGCs have won numerous international competitions dedicated to brain signal classification - notably EEG-based BCIs [16, 17], and have proven the best in various meta-analyses and large-scale benchmarks, at least for offline subject-specific BCI designs [18]. At the time of writing this article, they are thus considered by many to be the state-of-the-art classifiers for EEG-based BCIs. Note that this may change in the future when larger EEG databases are available, which DL could exploit better.

4.3 Explainable BCI

Explainable BCI refers to BCI systems designed with the capability to provide understandable and interpretable insights into how they process brain signals to make decisions. The goal is to ensure that the users, developers, and other stakeholders can understand the underlying mechanisms, the reasoning behind outputs, and the reliability of the system's operations. The use of transparent algorithms, identifying and explaining which features are most influential in making decisions, notably using visualization tools (e.g., saliency maps to show how different brain regions contribute to outputs [12]), and model validation, have demonstrated that explainable BCI is tangible and useful for BCIs [19].

4.4 Open challenges for machine learning in BCI

Despite recent and promising directions such as DL or RGCs, BCIs, are still far from being fully reliable and usable everywhere, by anyone, at any time, and without calibration. To make this happen, we have identified several open challenges that the ML community interested in BCIs may want to tackle:

The need for machine learning algorithms able to learn from small data: Typically, BCI experiments and applications can only afford to collect a few (typically a few dozen) training examples from a new user before training the BCI ML components. Moreover, this training data calibration is tedious and time-consuming for the BCI user. Thus, algorithms that can be trained from small data could make this data collection and calibration as short as possible [20].

The need for cross-subjects / cross-data sets algorithms: To go one step further, completely suppress calibration time, and address variability across and within users, we would ideally need algorithms that could learn classifiers from data of multiple users or even from multiple data sets so that they could classify EEG signals for new users and/or new data sets without recalibration [21].

The need to understand and tackle noise and variabilities in BCI: To make the points above feasible, a better understanding of the noise and variabilities affecting BCI is needed, notably about which sources of variability affect EEG and BCI performance, and how they affect them. Such knowledge should enable us to design informed algorithms that are more robust to these variabilities.

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

The need for invariant representations: The knowledge above could hopefully enable BCI designs to come up with invariant EEG representations, that are not affected by BCI sources of noise and variabilities, and thus which would be robust and reliable, even across users, within users, at all time.

The need for online adaptation: Alternatively, if designing invariant representations appears too challenging, the community would need to explore more advanced online adaptive BCI ML algorithms, i.e., the continuous update of the ML model parameters, based on brain data being collected online [3], to track and accommodate variabilities.

Need for self-paced decoding: In terms of user experience, BCIs would be more convenient and usable if the users could issue BCI commands at any time, whenever they want, i.e., if the BCI allowed self-paced use [22]. Unfortunately, most BCIs are not self-paced, but rather synchronous, enabling users to issue commands in specific, system-paced periods only. Research on self-paced BCIs seemed to have slowed down this past decade, but would probably need additional efforts, as current self-paced BCIs are still far from being reliable.

The need for new representations of EEG signals: Most BCI studies use only two main EEG representations: band power features (a.k.a. oscillatory activity) and ERPs. However, neuroscience research has identified additional neurophysiological patterns that could/should be better explored in BCI, including functional connectivity [23] or cross-frequency-coupling [24]. EEG source imaging (ESI) also offers an alternative representation allowing for the identification of the exact neural sources contributing to the recorded signals. By isolating these sources, the BCI can make more accurate predictions about the user's intentions, leading to more precise control over external devices [25]. Additional research to identify such alternative EEG representations may improve BCI reliability when combined with current representations.

The need for algorithms with low computational demand for use in wearable systems: If we aim for BCIs to leave the laboratories, and be used in everyday life, such BCIs will need to be as small as possible, to be embedded, e.g., in headphones, baseball caps or earplugs, which will limit their computational capabilities. There is thus a need for lightweight ML algorithms, that can learn and run quickly and with a low computational demand.

The need for algorithms able to work with few EEG sensors: For the same reason as above, in everyday life, we need BCIs that can work with as few EEG sensors as possible, so that they are as wearable and transparent (or even invisible) as possible. This in turn requires dedicated/optimized BCI ML algorithms that can decode EEG signals from as few EEG sensors as possible.

Overly complex machine learning models: Some overly complex machine learning models have demonstrated potential for high performance. However, they often come with significant drawbacks that can diminish their practical utility in BCI systems. It is often better to start with simpler models, validate their performance, and only introduce complexity when it is clear that it offers a benefit without compromising the usability of interpretability.

The need to consider the human user when designing algorithms: In the end, it is also essential to remember that the core of the BCI interaction loop is not the ML algorithm but rather the BCI user. Thus, while this is unfortunately rarely done, there is a need to consider the user during algorithm design. This could include, e.g., considering the user profile to personalize BCI ML algorithms, to ensure that classifier-based feedback enables user learning or to model users to know when/how to adapt algorithms parameters optimally [26].

5 Conclusions

We urge the BCI community to foster ongoing discussions aimed at establishing robust guidelines and standardized workflows. By doing so, we can effectively bridge the gap between Vidal's foundational BCI framework (1973) and contemporary research. This integration would ensure that experimental factors and neurophysiological insights are seamlessly incorporated into the established data pipeline and latest machine learning advances, ultimately leading to more reliable, impactful, and reproducible BCI research and development.

References

[1] Jacques J Vidal. Toward direct brain-computer communication. *Annual review of Biophysics and Bioengineering*, 2(1):157–180, 1973.

[2] Kaido Värbu, Naveed Muhammad, and Yar Muhammad. Past, present, and future of eeg-based bci applications. *Sensors*, 22(9):3331, 2022.

[3] Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rakotomamonjy, and Florian Yger. A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update. *J. Neur. Eng.*, 15(3):031005, 2018.

[4] Camille Jeunet, Stefan Debener, Fabien Lotte, Jeremie Mattout, Reinhold Scherer, and Catharina Zich. Mind the traps! design guidelines for rigorous bci experiments. In *Brain–computer interfaces handbook*, pages 613–634. CRC Press, 2018.

[5] Ali Bashashati, Mehrdad Fatourechi, Rabab K Ward, and Gary E Birch. A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals. *Journal of Neural engineering*, 4(2):R32, 2007.

[6] Aline Roc, Lea Pillette, Jelena Mladenovic, Camille Benaroch, Bernard NâKaoua, Camille Jeunet, and Fabien Lotte. A review of user training methods in brain computer interfaces based on mental tasks. *Journal of Neural Engineering*, 18(1):011002, 2021.

[7] Siuly Siuly, Yan Li, and Yanchun Zhang. Eeg signal analysis and classification. *IEEE Trans Neural Syst Rehabilit Eng*, 11:141–144, 2016.

[8] Anne-Marie Brouwer, Thorsten O Zander, Jan BF Van Erp, Johannes E Korteling, and Adelbert W Bronkhorst. Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Front. Neuro.*, 9:136, 2015.

[9] Dean F Salisbury. Clinical EEG and neuroscience editorial board announces new criteria for studies reporting on data-driven classifiers, and on novel interventions and treatments. *Clinical EEG and Neuroscience*, 53(6):471–471, 2022.

[10] Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering*, 16(5):051001, 2019.

[11] Eoin Thomas, Matthew Dyson, and Maureen Clerc. An analysis of performance evaluation for motor-imagery based BCI. *Journal of neural engineering*, 10(3):031001, 2013.

[12] David Trocellier, Bernard N'Kaoua, and Fabien Lotte. Visual cues can bias EEG deep learning models. In *Neuroergonomics 2024-The 5th International Conference*, 2024.

[13] Gernot Müller-Putz, Reinhold Scherer, Clemens Brunner, Robert Leeb, and Gert Pfurtscheller. Better than random: a closer look on BCI results. *International journal of bioelectromagnetism*, 10(1):52–55, 2008.

[14] Rafael T Gonzalez, Jaime A Riascos, and Dante AC Barone. How artificial intelligence is supporting neuroscience research: a discussion about foundations, methods and applications. In *Proc. LAWCN*, pages 63–77, 2017.

[15] Khondoker Murad Hossain, Md Ariful Islam, Shahera Hossain, Anton Nijholt, and Md Atiqur Rahman Ahad. Status of deep learning for eeg-based brain–computer interface applications. *Frontiers in computational neuroscience*, 16:1006763, 2023.

[16] Florian Yger, Maxime Berar, and Fabien Lotte. Riemannian approaches in brain-computer interfaces: a review. *IEEE Trans. Neur. Syst. Rehab.*, 25(10):1753–1762, 2017.

[17] Raphaëlle N Roy, Marcel F Hinss, Ludovic Darmet, Simon Ladouce, Emilie S Jahanpour, Bertille Somon, Xiaoqi Xu, Nicolas Drougard, Frédéric Dehais, and Fabien Lotte. Retrospective on the first passive brain-computer interface competition on cross-session workload estimation. *Frontiers in Neuroergonomics*, 3:838342, 2022.

[18] Sylvain Chevallier, Igor Carrara, Bruno Aristimunha, Pierre Guetschel, Sara Sedlar, Bruna Lopes, and other. The largest EEG-based BCI reproducibility study for open science: the MOABB benchmark. *arXiv:2404.15319*, 2024.

[19] Irene Sturm, Sebastian Lapuschkin, Wojciech Samek, and Klaus-Robert Müller. Interpretable deep neural networks for single-trial EEG classification. *Journal of neuroscience methods*, 274:141–145, 2016.

[20] Fabien Lotte. Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces. *Proc. IEEE*, 103(6), 2015.

[21] Xiaoxi Wei, A Aldo Faisal, Moritz Grosse-Wentrup, Alexandre Gramfort, Sylvain Chevallier, Vinay Jayaram, et al. 2021 BEETL competition: Advancing transfer learning for subject independence and heterogenous EEG data sets. In *PMLR*, 2022.

[22] Steven Mason, Julien Kronegg, Jane Huggins, Mehrdad Fatourechi, and Alois Schlögl. Evaluating the performance of self-paced brain-computer interface technology. *Neil Squire Soc., Vancouver, BC, Canada, Tech. Rep*, 2006.

[23] Juliana Gonzalez-Astudillo, Tiziana Cattai, Giulia Bassignana, Marie-Constance Corsi, and Fabrizio De Vico Fallani. Network-based brain–computer interfaces: principles and applications. *Journal of neural engineering*, 18(1):011001, 2021.

[24] Maria Sayu Yamamoto, Apolline Mellot, Sylvain Chevallier, and Fabien Lotte. Novel SPD matrix representations considering cross-frequency coupling for EEG classification using riemannian geometry. In *Proc. EUSIPCO*, pages 960–964. IEEE, 2023.

[25] Jussi T Lindgren. As above, so below? towards understanding inverse models in bci. *Journal of neural engineering*, 15(1):012001, 2017.

[26] Fabien Lotte, Camille Jeunet, Jelena Mladenović, Bernard N'Kaoua, and Léa Pillette. A BCI challenge for the signal processing community: considering the user in the loop. *Signal Processing and Machine Learning for Brain-Machine Interfaces*, pages 1–33, 2018.