



**HAL**  
open science

## **Learning from small datasets – review of workshop 6 of the 10th International BCI Meeting 2023**

Michael Tangermann, Sylvain Chevallier, Matthias Dold, Pierre Guetschel,  
Reinmar Kobler, Theodore Papadopoulo, Jordy Thielen

### ► **To cite this version:**

Michael Tangermann, Sylvain Chevallier, Matthias Dold, Pierre Guetschel, Reinmar Kobler, et al.. Learning from small datasets – review of workshop 6 of the 10th International BCI Meeting 2023. *Journal of Neural Engineering*, 2025, 22 (3), pp.033001. <10.1088/1741-2552/addf80>. <hal-05462858>

**HAL Id: hal-05462858**

**<https://universite-paris-saclay.hal.science/hal-05462858v1>**

Submitted on 22 Jan 2026

**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.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

**Learning from small datasets - Review of workshop 6 of the 10th International BCI Meeting  
2023**

Tangermann, M.W.; Chevallier, S.; Dold, M.R.; Guetschel, P.J.; Kobler, R.; Papadopoulo, T.; Thielen, J.

2025, Article / Letter to editor (Journal of Neural Engineering, 22, 3, (2025), article 033001)

Doi link to publisher: <https://doi.org/10.1088/1741-2552/addf80>

Version of the following full text: Publisher's version

Downloaded from: <https://hdl.handle.net/2066/320633>

Download date: 2026-01-16

**Note:**

To cite this publication please use the final published version (if applicable).

PERSPECTIVE • OPEN ACCESS

## Learning from small datasets—review of workshop 6 of the 10th International BCI Meeting 2023

To cite this article: Michael Tangermann *et al* 2025 *J. Neural Eng.* **22** 033001

View the [article online](#) for updates and enhancements.

### You may also like

- [Amorphous silicon carbide probe mechanics for insertion in the cerebral cortex of rats, pigs, and macaques](#)  
Mahasty Khajehzadeh, Negar Geramifard, Justin R Abbott *et al.*
- [SSVEP-based BCI using ultra-low-frequency and high-frequency peripheral flickers](#)  
Zexin Pang, Ruoqing Zhang, Meng Li *et al.*
- [P300-based brain–computer interface for communication in assistive technology centres: influence of users' profile on BCI access](#)  
Galiotta Valentina, Caracci Valentina, Toppi Jlenia *et al.*



## PERSPECTIVE

## OPEN ACCESS

RECEIVED  
31 July 2024REVISED  
5 May 2025ACCEPTED FOR PUBLICATION  
2 June 2025PUBLISHED  
24 June 2025

Original Content from  
this work may be used  
under the terms of the  
[Creative Commons  
Attribution 4.0 licence](#).

Any further distribution  
of this work must  
maintain attribution to  
the author(s) and the title  
of the work, journal  
citation and DOI.



## Learning from small datasets—review of workshop 6 of the 10th International BCI Meeting 2023

Michael Tangermann<sup>1,\*</sup> , Sylvain Chevallier<sup>2</sup> , Matthias Dold<sup>1</sup> , Pierre Guetschel<sup>1</sup> , Reinmar Kobler<sup>3,4</sup>, Theodore Papadopoulo<sup>5</sup> and Jordy Thielen<sup>1</sup> <sup>1</sup> Donders Institute, Radboud University, Nijmegen, The Netherlands<sup>2</sup> INRIA TAU, LISN-CNRS, Université Paris-Saclay, Orsay, France<sup>3</sup> Department of Dynamic Brain Imaging, Advanced Telecommunications Research Institute (ATR), Kyoto, Japan<sup>4</sup> RIKEN Center for Advanced Intelligence Project (AIP), Tokyo, Japan<sup>5</sup> INRIA Sophia Antipolis Méditerranée, Sophia Antipolis Cedex, France

\* Author to whom any correspondence should be addressed.

E-mail: [michael.tangermann@donders.ru.nl](mailto:michael.tangermann@donders.ru.nl)**Keywords:** brain–computer interfacing, machine learning, one-shot learning, regularization, sample efficiency, transfer learning, zero-training**Abstract**

In a brain–computer interface (BCI), a primary objective is to reduce calibration time by recording as few as possible novel data points to (re-)train decoder models. *Objective.* Minimizing the calibration can be crucial for enhancing the usability of a BCI application with patients, increasing the acceptance by healthy users, facilitating a fast adaptation during non-stationary recordings, or transferring between sessions. *Approach.* At the 10th International BCI Meeting in 2023, our workshop addressed the latest proposed techniques to train classification or regression machine learning models with small datasets. *Main results.* We explored methodologies from both traditional machine learning and deep learning. In addition to talks and discussions, we discussed Python toolboxes for various presented methods and for the benchmarking of classification models. *Significance.* This review provides a comprehensive overview of the workshop's content and discusses the insights that were obtained.

**1. Introduction**

This review focuses on the latest techniques to train classification or regression machine learning models with small datasets in the field of brain–computer interfaces (BCIs). It covers techniques for supervised and unsupervised decoding of brain signals presented at a workshop on small datasets organized as part of the 10th International BCI Meeting in 2023 by seven researchers from five labs.

**1.1. Comparison to large data domains**

Contrary to widely adopted application scenarios of machine learning which use text or image material for training and evaluation that is available in abundance, the machine learning approaches in the field of brain–computer interfacing (BCI) act upon signals derived from magneto-/electroencephalogram (M/EEG) recordings, and thus cannot (yet) build upon comparably large data collections. Instead,

these brain signal recordings represent a small data domain, because each recording session typically involves a sophisticated experimental setup supported by a domain expert. Thus M/EEG data is less readily available and cannot simply be scraped from websites or social media. The resulting small data problem is recognized by the wider BCI community [1–5] well beyond the scope of the workshop organized by the authors. One possibility to address this problem by the community is to build up collections of BCI datasets, such as the ‘mother of all BCI benchmarks’ (MOABBs) [6].

A BCI session, i.e. a continuous recording of one user, rarely provides more than a few hours of brain signals. This is specifically true for patient studies in clinical contexts, where treatments and studies compete for the available time of a patient. The most valuable data segments of a session consist of labeled data, which has been obtained during an actively executed task. This so-called calibration data typically makes

for only a fraction of the data collected during a session.

In standard machine learning approaches for BCI, the calibration data is used to train a decoding model. Subsequently, this decoding model is applied during the second part of the session to run the actual online application. For longer online applications, frequently a degradation of the decoding performance is observed indicating non-stationarity of the brain signals over time [7]. The non-stationarity can rarely be pinpointed to a single cause only but may reflect changes related to the recording setup as well as user-related factors, such as psychology, morphology, fatigue, or a changed strategy to execute the task [8], but also fluctuating influences of for instance medication.

In situations involving non-stationary data, quickly re-calibrating the decoding model may help [7]. Ideally, such re-calibration should require a very short time only, that is, it must again be realized using a small labeled dataset. Naturally, non-stationarities are a hurdle specifically when the same decoding model shall be transferred to the following sessions of the same user, to sessions of other users, or even across different tasks and protocols.

### 1.2. ‘Small’ depends on the protocol, user group, or time available per session

The experimental protocol used in BCI determines roughly how many M/EEG data points can be collected within a single session. Typically, data points are multiple repetitions of a trial or an epoch, that is, one executed task or one presented stimulus, respectively. For instance, motor imagery (MI) or attempted motor execution tasks, which take one to multiple seconds to cue and execute, may result in a few tens or a few hundreds of trials per class (e.g. left versus right hand) within one session, while evoked responses obtained from event-related potential (ERP), steady-state, or code-modulated stimulation protocols may result in a few thousand epochs per session under ideal conditions. Please note, however, that only a fraction of these are labeled data points when an online application is run during the session. In addition, sessions with patient groups in general require more and longer breaks, are more limited regarding the session duration, have to respect the patient’s attention capacity and stamina, and consequently result in fewer task repetitions than sessions conducted with age-matched healthy users or even young participants [9].

### 1.3. Definitions and clarification of terminology

Realizing, that the notion of general scarcity of data in the BCI field can be further aggravated by the context, we generally will base our definition of ‘small datasets’ on the goal of using a minimal amount of labeled

or unlabeled data from either a novel user, a novel session, or a novel experimental protocol/paradigm. For each of the three scenarios the goals can be to (1) train/calibrate a decoding model from scratch, (2) re-calibrate or update an existing model with comparable but more recent data, or (3) to fine-tune a decoding model for the *downstream task* of the BCI session, which has been *pre-trained* on historic data and potentially on a different *pretext task*. All these scenarios and goals have in common, that in a concrete new session or protocol, only minimal time shall be spent to obtain a decoding model capable of running some BCI application.

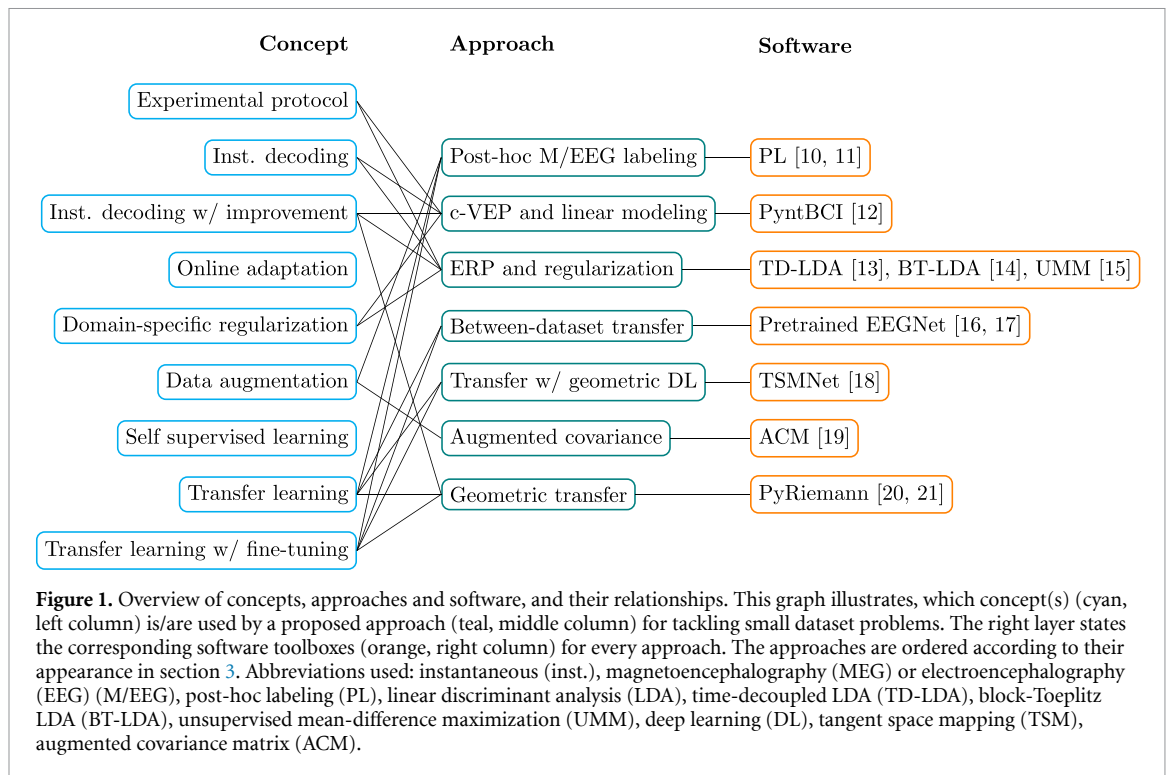
In the following, we introduce general concepts of how to cope with the small dataset challenge. Subsequently, we describe recent approaches in more detail, which have been presented as part of Workshop 6 titled ‘Learning from Small Datasets’ held at the 10th International BCI Meeting themed ‘Balancing Innovation and Translation’, which took place on 6–9 June 2023 in the Sonian Forest in Brussels in Belgium. For brevity, we have limited the author list of this workshop paper to the workshop presenters only. Please note, that every presented approach typically is the result of a collaborative effort, as reflected by the papers cited in the corresponding sections.

## 2. General concepts for coping with the small dataset challenge

In the following, we have refined and explained distinct concepts that have the potential to address the small dataset challenge. We cover (1) the experimental protocol, (2) instantaneous decoding, (3) instantaneous decoding with improvements across trials, (4) online adaptation, (5) regularization, (6) data augmentation, (7) self-supervised learning, (8) transfer learning, and (9) transfer learning with fine-tuning. The novel approaches presented in section 3 make use of these concepts, often in the form of combinations. The relationships between these concepts, the novel approaches and available software tools are illustrated in figure 1.

### 2.1. Experimental protocol

Whether a dataset is considered small or large is significantly influenced by the amount of information it contains per unit of time. This is closely tied to the experimental protocol used to evoke or induce task-specific brain activity, which may accumulate data at varying rates and exhibit differing signal-to-noise ratios (SNRs). In the past, preferable imagery classes were investigated [22], improvements to visual evoked potential protocols [23] and steady-state visual evoked potential (SSVEP) protocols [24] had been proposed. Consequently, one of the key



factors in managing limited datasets is selecting an experimental protocol that efficiently gathers reliable information over time.

There are several strategies to enhance the information content of the experimental protocol. For protocols based on evoked responses, fine-tuning the inter-stimulus interval can accelerate stimulus delivery while avoiding habituation of responses [25, 26]. Moreover, instead of relying solely on stimuli eliciting early sensory responses, introducing more complex stimuli [27] may engage higher-level cortical areas, yielding additional relevant features crucial for precise decoding. In ERP-based and MI-based protocols, selecting intuitive and easily learnable tasks and ensuring participant comfort can mitigate fatigue, thereby reducing the likelihood of encountering non-stationary data.

## 2.2. Instantaneous decoding

An approach classified as *instantaneous* operates solely on the data of the ongoing trial without relying on any prior training data. Essentially starting from scratch for every trial, it processes the current trial's data, for which the label is unknown. Such methods also fall under the categories of *zero-training* or *calibration-free*, as they do not use a prior calibration stage that typically involves labeled training data [28].

Instantaneous methods may also be considered *unsupervised*, given their independence from label information. However, some methods in this category may leverage prior knowledge, blurring the line into *semi-supervised* and *weakly supervised* methods,

especially if they use detailed information about all potential stimuli in the experimental protocol to constrain the search space.

Instantaneous methods exclusively use local information from the ongoing trial, often spanning only several seconds. Consequently, they operate with minimal available information. This characteristic renders instantaneous methods relatively robust against non-stationary data. Essentially, after each trial, a new classifier is used, that takes into account only the local data distribution.

## 2.3. Instantaneous decoding with improvements across trials

As instantaneous approaches operate with minimal data, they may result in lower performance or require longer trials compared to their calibrated, supervised counterparts. To address this limitation, an instantaneous method could leverage previously classified trials as additional information alongside the current trial data. Then a classifier still starts from scratch, but having seen more and more trials it gradually learns an improved representation. Such approaches can potentially converge to their calibrated counterparts [29], but may then share with these models the reduced ability to flexibly react to non-stationarity.

For certain decoding models, incorporating past data can enhance the estimation of certain model parameters. If, for example, a model uses a covariance matrix to estimate the background brain signal activity and noise structure, then this activity is independent of the classes and may be estimated without

relying on label information. In such cases, previous trials can improve the reliability of the covariance estimate and improve the performance and efficiency of the decoding model. Alternatively, in scenarios where label information is essential, such as for estimating class means, an optional optimistic strategy utilizes previously classified trials as labeled data, assuming their classifications were correct. This technique is known as *naive labeling* [30].

#### 2.4. Online adaptation

The preceding categories explored the concept of a calibration-free decoder. In contrast, more conventional approaches involve calibrating a classifier with labeled training data and subsequently applying the classifier to novel testing data. However, since the training and testing data may differ significantly, e.g. the testing data may exhibit evolving distributions over time, using such a fixed classifier under non-stationary conditions can lead to suboptimal performance.

One strategy to address the challenge of changing data distributions is to update the classifier during testing, a technique commonly known as an *adaptive* classifier [31, 32]. In its most extreme form, this involves updating the classifier after each new trial, a practice referred to as *online learning*.

#### 2.5. Domain-specific regularization

A fundamental challenge when dealing with limited data is to prevent overfitting. In this scenario a model fits the training data too closely, leading to a poor generalization for unseen data. Overfitting often occurs when the number of features, and consequently the number of the model's parameters, are large compared to the available data points used for training. During the workshop, we emphasized on domain-specific regularization approaches, assuming that unspecific regularization methods (L1/L2 norm, drop-out, etc) are known.

Regularization aims to address overfitting under small sample conditions by constraining the model's degrees of freedom. This is typically achieved by integrating prior knowledge about the brain signal data domain, leveraging insights from the experimental protocol, making informed assumptions about the geometry of recording electrodes, the data distribution, expected outliers and artifacts, or imposing other domain-specific constraints on the model parameters. As a result, the model typically becomes simpler which enhances its sample efficiency and reduces the risk of overfitting. Regularization thus allows to either obtain better models based on the same amount of data, or to effectively use limited

training data while maintaining a comparable performance. In the context of the small data challenge in BCI, the second use case is interesting.

#### 2.6. Domain-specific data augmentation

Most of the preceding concepts have addressed the challenge of small datasets by devising sophisticated strategies within the decoding approach. In contrast, the concept of *data augmentation* tackles the issue at its core. Essentially, data augmentation aims to expand the available dataset without necessitating the collection of additional labeled data points.

Traditionally, data augmentation enlarges the training data set by modified versions of the original training data points. The modifications are realized by either general transformations, i.e. adding unspecific noise, or by leveraging domain-specific knowledge to determine suitable transformations [5, 33]. The objective is to augment the dataset in a manner that enhances the robustness of the decoding model and ultimately improves its performance.

Alternatively, the advent of generative AI, specifically transformer and diffusion models, enables the creation of synthetic data, providing another avenue for augmenting the dataset [34].

#### 2.7. Self-supervised learning on a pretext task

The primary objective of *self-supervised learning* is to discern underlying patterns or structures within data, and to do so without relying on labeled examples [35]. This approach opens up the possibility of using unlabeled data, which is much more abundant than labeled data, and enables the development of larger and more general models. Such models hold promise for transfer learning applications, where they can be fine-tuned to new contexts, as discussed in the subsequent sections.

#### 2.8. Transfer learning

*Transfer learning* involves using data collected from one or multiple users, sessions, or tasks and applying it to a different yet potentially related context. The underlying concept is to capture any shared information between the source(s) and receiver, thereby reducing the amount of data needed for the receiving task. Transfer learning includes two main types: domain adaptation and rule adaptation [36]. Domain adaptation focuses on learning transformations so that a shared rule generalizes across various contexts [37]. Conversely, rule adaptation aims to discover adjustments necessary for implementing context-specific decoders in novel environments, as shown in [38].

A more and more popular tool for transfer learning is deep learning (DL), or deep transfer-learning [39]. Specifically, a recent comprehensive

review discusses many DL approaches, which aim at forming deep representations of EEG data that are useful for BCI transfer learning and decoding tasks [40].

### 2.9. Transfer learning with fine-tuning

In an ideal scenario, data could seamlessly transition between contexts without any degradation in performance. However, due to inherent variation among users, sessions, tasks, and even within a single recording session, the approach to directly translate data from one context to another is challenging and will probably result in a significant performance loss. One approach to address this issue is to record a limited amount of calibration data within the novel context. This calibration data can then be used to fine-tune a model initially pre-trained on a different task, thereby mitigating the performance degradation and adapting the model to the novel task very sample-efficiently.

## 3. Approaches

This review focuses on the contributions made in Workshop 6 titled ‘Learning from Small Datasets’, which was held at the 10th International BCI Meeting and was themed ‘Balancing Innovation and Translation’. Each contribution presents a recent approach to tackling the small dataset challenge. While these approaches of course are not the only ones to do so, they represent a large spectrum of methods that currently are available to BCI researchers and practitioners. Along the presented approaches, we will give pointers to existing other studies. For every approach, we state which concepts it connects to and provide a link to source code or trained models, see figure 1.

### 3.1. Post-hoc M/EEG labeling

The post-hoc labeling (PL) method [41] augments data (section 2.6) and generates matching labels while preserving the characteristics of M/EEG recordings. It allows for the extraction of large amounts of continuously labeled data from any, potentially short, multi-channel M/EEG recording, i.e. also from non-BCI data. Its main use cases are the generation of benchmark data for regression methods and the pre-training of decoding models and feature embeddings using deep neural networks. As such, it may support transfer learning (section 2.8) and fine-tuning (section 2.9).

PL requires a solution for the inverse problem  $\mathbf{s}(t) = \mathbf{W}^T \cdot \mathbf{x}(t)$ . Solving this problem delivers many estimated sources  $\mathbf{s}(t)$  which underlie an observed M/EEG recording  $\mathbf{x}(t) \in \mathbb{R}^{N_c}$  with  $N_c$  channels. The post-hoc relabeling method defines novel labels for the  $i$ th source  $\tilde{y}_i(t)$  as either the activity of that source

$s_i(t)$ , its (log) band-power or other characteristics. Optionally, the activity of a source can be amplitude-modulated or disrupted by added noise to form  $\tilde{s}_i(t)$ , before applying the associated forward model which generates a modulated M/EEG recording  $\tilde{\mathbf{x}}(t) = \mathbf{A} \cdot \tilde{\mathbf{s}}(t)$ . The modulation allows to influence the SNR of the source and thus the difficulty of the corresponding novel labels  $\tilde{y}_i(t)$ . Repeating this procedure with different modulations and for different sources delivers a multitude of very realistic labeled datasets.

Post-hoc relabeling can be applied either fully data-driven or using biologically motivated and constrained models. For a data-driven approach, blind source separation methods [42, 43] will provide source signal activity estimates  $\mathbf{s}(t)$  as well as a corresponding forward model  $\mathbf{A}_{\text{data}}$ . In the case of a biologically motivated forward model  $\mathbf{A}_{\text{bio}}$ , the solution (sources) can be obtained under, e.g. minimum norm constraints [44], or using domain-specific constraints suitable for expressing spatial priors, for dealing with non-stationarities [45, 46] or label noise. Once sources are obtained and modulated, they can be mixed to deliver virtual M/EEG data analogously to sources delivered by blind source separation methods. Creating such modulated, labeled data sets will largely preserve the characteristics of natural M/EEG recordings, such as narrow-band power features, typical spindle dynamics, and the general  $1/f$  noise floor, which may be an advantage compared to fully synthetic M/EEG generation [41]. A related approach by Zlatov and Blankertz [47] modulates the direction of estimated dipoles to augment MI datasets.

Providing control over the SNR levels and label noise is an interesting feature of the approach. It allows running parameter searches over SNR levels to evaluate the robustness of decoding pipelines in benchmarking scenarios. Additionally, a retrospective estimation of the SNR and label noise characteristic for a given recording can be attempted by approximating the decoding performance  $P_{\text{real}}$  of a pipeline predicting real labels with the performance of the same pipeline using multiple versions of noisy synthetic labels  $\tilde{y}(t)$ .

Repositories are available containing the code for the benchmarking of decoding methods [10] used in [41] and an introductory notebook [11] augmenting MOABB [6] data.

### 3.2. The code-modulated visual evoked potential (c-VEP) paradigm and linear modeling

Despite recent reports about record-breaking communication speed, exceeding 200 bits per minute [48, 49], the experimental BCI protocol called *noise-tagging*, is still relatively unknown to the BCI community.

Noise-tagging evokes the c-VEP [50], which is one of the major classes of evoked responses besides

the ERP and SSVEP. The ERP paradigm commonly presents stimuli at a relatively slow rate, at around 2–5 Hz, like an oddball protocol evoking the P300 [51]. In contrast, the SSVEP paradigm presents stimuli at a faster pace, typically at 8–16 Hz, with each stimulus encoded using a unique narrow-band frequency, a technique known as *frequency tagging* [52], or *joint frequency and phase modeling* [53]. The c-VEP paradigm, similarly to the SSVEP, also presents stimuli at a rapid rate but instead uses broad-band stimuli, hence named *noise tagging* [50].

These three paradigms go well in hand with the time-modulated, frequency-modulated and code-modulated multiple access schemes in telecommunication [54]. The success of the c-VEP paradigm can be attributed to several factors addressing the challenge of small datasets.

Firstly, the advanced experimental protocol (section 2.1) employs rapidly presented pseudo-random noise codes, carefully crafted to ensure dissimilarity in both stimulus and response domains. This design facilitates decoding [49], while also limiting the required training dataset [50, 55].

Secondly, sophisticated decoding approaches extensively leverage domain-specific knowledge (section 2.5). For example, the structure of the stimulus protocol and insights into underlying neural mechanisms were combined to develop an encoding–decoding framework, called *reconvolution*, integrated into a *canonical correlation analysis* (CCA) [56]. It works under the assumption of the *linear superposition hypothesis*, which poses that the response to a sequence of flashes is the summation of the responses to the individual flashes [57]. The framework significantly reduces the number of learnable parameters and, consequently, the amount of required labeled data [58].

Thirdly, the availability of a forward model has led to the development of a calibration-free instantaneous decoding method, which predicts the attended class by identifying the stimulus sequence that best explains the data [58]. This calibration-free approach can operate instantaneously (section 2.2) or integrate information over trials (section 2.3), assuming previously classified trials were predicted correctly [59].

Of course, there are various other recent c-VEP works dealing with the small dataset challenge that involve, but are not limited to, improved stimulus design for faster decoding and better user comfort (section 2.1) with other binary codes [55], burst-codes and grating stimuli [60, 61], non-binary codes [62], narrow-band random sequences [63], and white-noise codes [49]. Additionally, calibration is shortened with transfer learning (section 2.8) using good donor subjects [60], session-to-session [64] and subject-to-subject [65] transfer. Finally, single-trial

decoding is shortened using sophisticated dynamic stopping [66].

Fortunately, more and more shared resources facilitate the improvement of the c-VEP paradigm. For instance, several raw datasets [67–69] are integrated in MOABB [6]. Additionally, software packages to run experimental protocols are shared such as FlickersOnTop [70], Dareplane [71, 72], and the Medusa Platform [73]. Finally, software libraries for analysis are openly available with informative examples in the PyntBCI [12] and Medusa Kernel [73] toolboxes.

### 3.3. Domain-specific regularization and structural information for ERP decoding

BCI protocols that use ERP responses present a sequence of stimuli. The user focuses on one of the stimuli (target) and ignores the others (non-targets) to issue a command within one trial. On average, the resulting target-locked ERP responses differ from non-target responses. Decoding the stimulus type based on its noisy ERP response (epoch) is a binary classification problem. As the randomized stimuli are typically considered independent, the decoding does not account for the actual sequence of stimuli. The target vs. non-target classification problem is tackled by machine learning methods, among which linear discriminant analysis (LDA) is still predominantly used, as LDA is easy to implement, requires limited training data points, shows competitive performance with shrinkage regularization of the feature covariance matrix [74] and has few hyperparameters. LDA also has theoretical advantages when both classes follow the same Gaussian distribution, which is approximately fulfilled by ERP data after outlier removal. But even shrinkage regularization can only partially mitigate the small sample problem, which impedes specifically a good estimate of the covariance matrix  $\Sigma$  with its  $D \times (D + 1)/2$  free parameters, given  $D = N_c \times N_t$  features, comprised of  $N_c$  EEG channels, and  $N_t$  amplitudes per channel.

Recently, the use of additional (protocol-specific) information has been proposed for tackling small-sample ERP decoding problems. Sosulski *et al* used assumptions about the recorded ERP-EEG signals to formulate two improved domain-specific regularizations (section 2.5) for  $\Sigma$ . The first, time-decoupled LDA (TD-LDA) [75] assumes that noise on top of the ERP features is Gaussian and unrelated to the user task, i.e. the same for target stimuli, non-target stimuli or even in the absence of stimulation. These assumptions allow to virtually increase the number of independent datapoints by a factor of  $N_t$  to estimate parts of  $\Sigma$  more robustly. The second approach, block-Toeplitz LDA (BT-LDA) [76], assumes that the covariance between two time samples depends

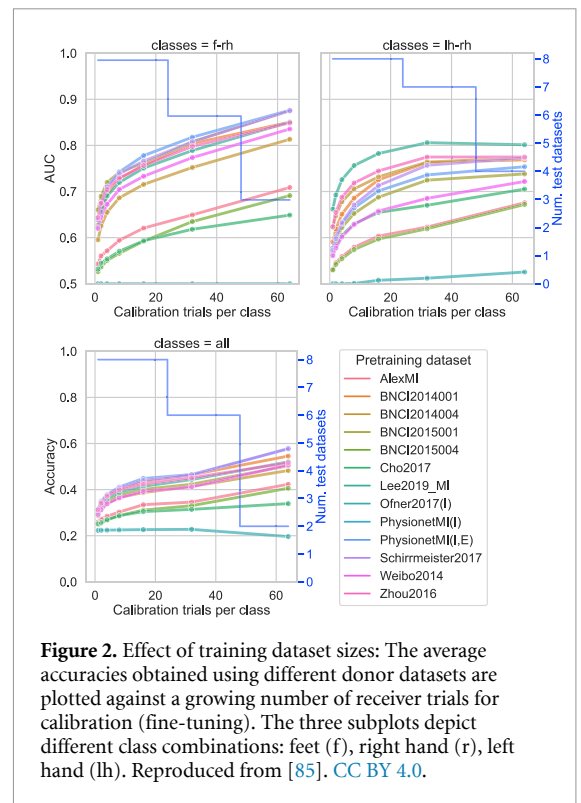
on their time difference  $\delta$  only, not on their position within an epoch, and reaches zero for  $\delta \rightarrow \infty$ . This reduces the free parameters of  $\Sigma$  to  $N_c \times (N_c + 1)/2 \times N_t$ . Both regularizations are applied on top of the non-domain-specific shrinkage regularization. They result in significantly improved classification performances over a range of ERP datasets provided, e.g. by MOABB [6], with strongest benefits for small datasets with large  $N_t$ .

Alternatively, the temporal structure of a stimulus sequence can be exploited by decoding a full trial instead of separately classifying single epochs. This change has allowed a formulation for unsupervised decoding methods like *learning from label proportions* (LLPs) [77, 78], which, however, required small changes of the stimulus protocol (section 2.1). A very recent improvement, unsupervised mean-difference maximization (UMM) [79] overcomes this limitation. Equipped with a confidence metric, UMM can learn across trials (section 2.3), but can also decode trials instantaneously (section 2.2). The latter enables UMM to minimize the performance degradation under non-stationary feature distributions. Specifically for visual ERP data of healthy users, UMM performs almost flawlessly [79]. Recently, UMM was also applied to the fast c-VEP protocol [59]. Code examples of the three ERP classification approaches are provided in [13–15].

### 3.4. Between-dataset transfer

As outlined in the introduction, the problems posed by small datasets can be addressed by transfer learning (section 2.8) and fine-tuning (section 2.9) techniques. While cross-session transfer [80], cross-subject transfer [81, 82], or both [83] have been widely studied, cross-dataset transfer has been relatively underexplored but not completely uncharted [84]. The efficiency of transfer learning in BCI systems is largely dependent on two criteria: a small distance between the data distributions of the pre-existing (donor) and new (receiver) datasets, and the availability of sufficient and diverse data in the donor to enable the model to generalize to the novel data. While choices of donors for cross-session or cross-subject transfer are necessarily limited, possibilities open up when considering cross-dataset transfer. And finding a donor dataset that meets one of these criteria is relatively straightforward while optimizing the balance between the two is a non-trivial task.

In [85], a simple framework was proposed for systematically testing the compatibility of datasets from different studies for MI transfer tasks. For every pair of donor and receiver datasets, first a model was trained on the donor before training merely an additional new linear classification layer based on a few receiver trials. The performance of this transfer approach was then tested on other trials of the receiver



**Figure 2.** Effect of training dataset sizes: The average accuracies obtained using different donor datasets are plotted against a growing number of receiver trials for calibration (fine-tuning). The three subplots depict different class combinations: feet (f), right hand (r), left hand (lh). Reproduced from [85]. CC BY 4.0.

dataset. As this simple framework is computationally efficient and makes very few assumptions about the datasets, it allows running large-scale compatibility benchmarks.

With this framework the transferability between all MI datasets of the MOABB library [6] was evaluated, providing a comprehensive guide for compatibility. Furthermore, the datasets were characterized based on their overall performance as donors (figure 2) and receivers, and a method was provided for applying this framework to a new receiver. The trained models and code thereof are available [16, 17], and may serve as a reference for future BCI research, allowing making informed donor choices and thereby enhancing the effectiveness of transfer learning in mental imagery decoding tasks. Finally, an applied example of such use is presented to strengthen this claim. In this example, the study of Xie and colleagues [84] is reproduced, but replacing the pre-training dataset they used with the one recommended by the new method. This change leads to statistically significant score improvement.

### 3.5. Transfer with geometric deep learning (GDL)

GDL aims to use domain-specific geometric structures in an end-to-end decoding model. Covariance matrices which are symmetric and positive definite (SPD) are one particular geometric structure that is frequently encountered in EEG-based BCIs. Differential geometry and geometric statistics offer principled mathematical frameworks to analyze this

geometric structure and perform statistical inference, giving rise to a family of geometry-aware shallow decoders [86]. Since their introduction about a decade ago [87], they have been highly competitive in all major BCI paradigms [88, 89].

The success in the regime of small data is driven by *transfer learning* (section 2.8) techniques [90, 91] and intrinsic *domain-specific regularization* (section 2.5) properties. First, estimating spatial covariance matrices reduces dimensionality, since the number of spatial channels is typically an order of magnitude smaller than the number of temporal samples. Second, popular Riemannian metrics measure distances on a logarithmic scale, rendering them robust to high-variance outliers. Third, these metrics come with invariances relevant to spatial mixing models of EEG data, i.e. distances are similar in channel and source space. On the downside, the parameters are fitted sequentially, requiring prior knowledge about spectral and temporal task-related effects.

Neural networks acting on SPD manifolds use and retain the underlying geometric structure of the data. Numerous standard layers have recently been generalized to the SPD manifold, realizing projections [92], nonlinear activations [92] and batch normalization [93, 94]. Additionally, layers that implement tangent space mapping (TSM) were introduced for the log-Euclidean [92] and the affine invariant [83] Riemannian metrics. TSM, a key concept in Riemannian geometry, refers to an invertible mapping that globally (for SPD manifolds) projects points between the manifold and the tangent space at any reference point on the manifold [95]. Tangent spaces are Euclidean vector spaces, where standard Euclidean distances locally approximate distances on the manifold. Generally, GDL BCI decoders either operate with SPD matrix-valued input features [94, 96–98] or learn temporal/spectral feature extractors jointly with latent SPD representations [83, 94, 99–101].

To address non-stationarities in cross-session and cross-subject transfer, Kobler *et al* [83] extended shallow geometry-aware unsupervised domain adaptation (UDA) techniques [90, 91] to SPD neural nets. The proposed TSMNet architecture combines a shared convolutional feature extractor with a domain-specific TSM layer and a shared linear classification head, resulting in a parameter-efficient (orders of magnitude fewer parameters than conventional neural nets) model that is robust to scarce data [83]. Unlike shallow geometry-aware alignment methods [90, 91], end-to-end learning extracts task-relevant latent representations, resulting in improved SNR for the UDA problem [83, 96, 102, 103]. The hypothesis class of TSMNet essentially corresponds to traditional TSM models [87, 104]. As such the fitted model can be globally interpreted via

transforming the parameters into spectral and spatial patterns [105].

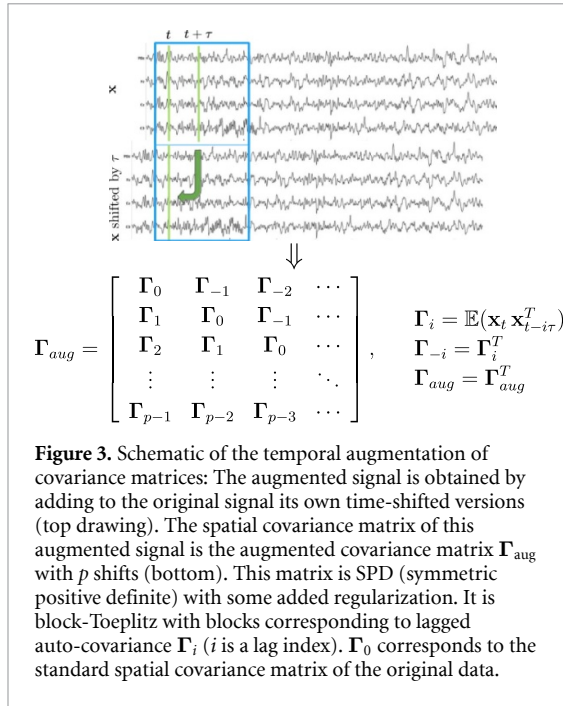
The code of the TSMNet architecture is available online [18].

### 3.6. Augmented covariance matrix (ACM)

The idea of the *augmented covariance matrix* (ACM) stems from autoregressive models [106]. Such models assume that the data at time  $t$  can be explained as a linear combination of the same data at some previous time points  $t-\tau_1, t-\tau_2, \dots, t-\tau_p$  plus some random noise called the innovation term. Here,  $p$  is the order of the autoregressive model. It is assumed that  $\tau_i=i\tau$ , where  $\tau$  is called the lag. Thus, the multichannel EEG signal is modeled as  $\mathbf{x}_t = \sum_{i=1}^p \mathbf{A}_i \mathbf{x}_{t-i\tau} + \varepsilon_t$ , where the  $\mathbf{A}_i, i=1 \dots p$  are the linear coefficients and  $\varepsilon_t$  is the noise. For an EEG signal with  $n$  channels,  $\mathbf{x}_t \in \mathbb{R}^n$ ,  $\varepsilon_t \in \mathbb{R}^n$  and  $\mathbf{A}_i \in \mathbb{R}^{n \times n}$ . Estimating such a model from data typically involves solving a linear system, the so-called Yule–Walker equations. The matrix involved in this linear system is called the ACM  $\Gamma_{\text{aug}}$ . It is a block-Toeplitz symmetric positive (and definite with some care) matrix [76, 106]. Figure 3 illustrates how this matrix is built and summarizes its structure.

Being SPD,  $\Gamma_{\text{aug}}$  can be classified using a Riemannian-based approach using an affine invariant metric of SPD matrices. All the data (training and testing augmented covariance matrices) are projected on the same tangent space of the set of SPD matrices, where an SVM classifier is applied. This method has proven to be extremely efficient in many recent studies for various BCI paradigms (e.g. MI, SSVEP, P300) [86, 107, 108]. Using this approach leads to 2% to 9% of classification improvements over competing methods at the small cost of introducing two hyperparameters ( $p$  and  $\tau$ ). These improvements can either be used to reduce the number of channels or to reduce the number of training samples, while maintaining an accuracy comparable to the non-ACM state of the art.

To conclude, the ACM methodology is a form of domain specific data augmentation (section 2.6) for dynamical systems. It can be considered as a spatio-temporal extension of the usual purely spatial approaches and basically relies on a more precise spatio-temporal modeling of the recorded signals. The increase in classification performance it provides can be used to work with smaller data sets. In this work, either reducing the number of EEG electrodes or reducing the number of training samples was explored. In both cases, ACM has shown its capacity to effectively handle such a small dataset situation. An interesting note is that the time embedding procedure depicted in figure 3 (top) is exactly what appears in mathematical theorems for the characterization of the dynamics of a system given some observations



**Figure 3.** Schematic of the temporal augmentation of covariance matrices: The augmented signal is obtained by adding to the original signal its own time-shifted versions (top drawing). The spatial covariance matrix of this augmented signal is the augmented covariance matrix  $\mathbf{\Gamma}_{aug}$  with  $p$  shifts (bottom). This matrix is SPD (symmetric positive definite) with some added regularization. It is block-Toeplitz with blocks corresponding to lagged auto-covariance  $\mathbf{\Gamma}_i$  ( $i$  is a lag index).  $\mathbf{\Gamma}_0$  corresponds to the standard spatial covariance matrix of the original data.

(e.g. Takens theorem [109]). Code is available on GitHub [19] and in MOABB [6].

### 3.7. Geometric transfer

As delineated in section 2.8, transfer learning aims to capture shared information between the source domains  $\mathcal{D}_S$ —in essence, the BCI users that could serve as training data—and the target domain  $\mathcal{D}_t$ , representing a specific BCI user. Notably, if data from the target user is accessible, it becomes feasible to fine-tune the model (see section 2.9 or [110]). Both domain adaptation and rule adaptation have been explored using geometric methodologies, relying on the estimated covariance of M/EEG data. Within the small datasets regime, transfer learning allows to use as much data as possible without hindering performance by training on samples with a different distribution.

Domain adaptation can be realized through composite mean techniques, employing either Euclidean [111] or Riemannian [112, 113] geometries. A weighted combination of each trial from  $\mathcal{D}_t$  and the (Euclidean or Riemannian) mean is computed from  $\mathcal{D}_S$ . Furthermore, based on the selected weighted combination, one can transition from pure transfer learning, excluding any data from  $\mathcal{D}_t$ , to instantaneous decoding with performance improvements across trials, as detailed in section 2.3. Proper parametrization significantly enhances accuracy across all BCI paradigms [113].

*Riemannian Procrustes analysis* (RPA) [91] serves as a prototypical instance of rule adaptation, wherein all data undergo transformation to facilitate training a classifier within a shared feature space encompassing

both  $\mathcal{D}_S$  and  $\mathcal{D}_t$ . RPA involves three consecutive pre-processing steps: recentering, rescaling, and rotating the data. Consequently, the preprocessed data are normalized across users, enabling the training of a single classifier to handle all domains. A notable limitation of RPA is its dependence on calibration data for estimating the preprocessing steps.

The code for both, domain adaptation and rule adaptation functionalities are incorporated within the pyRiemann [20] toolbox, with illustrative examples accessible in the documentation [21], as depicted in figure 4. Transfer learning makes the most out of all available samples, a useful approach for small datasets.

## 4. Discussion

A large audience attending the workshop acknowledged, that solutions for working with small datasets are desired in the field. Different approaches (see section 3) capable of benchmarking, training or fine-tuning BCI decoding models with minimal labeled or unlabeled training data points were presented, each with code resources that allowed workshop participants to apply the methods to their own data.

While the approaches cover a large range of ideas on how to maximize the information extracted from a few samples, this review could introduce the ideas superficially only, and we had to refer interested readers to extended publications for every method.

Common to all approaches was the observation, that with limited data they can outperform established decoding models and pipelines, which are widespread in standard BCI toolboxes. The approaches achieved this by requiring a few, but not very strict assumptions. Any performance advantage obtained can be used to either derive better decoding models with a larger amount of training data, or to reach state-of-the-art decoding performance with substantially reduced data. Our workshop has mainly focused on the latter.

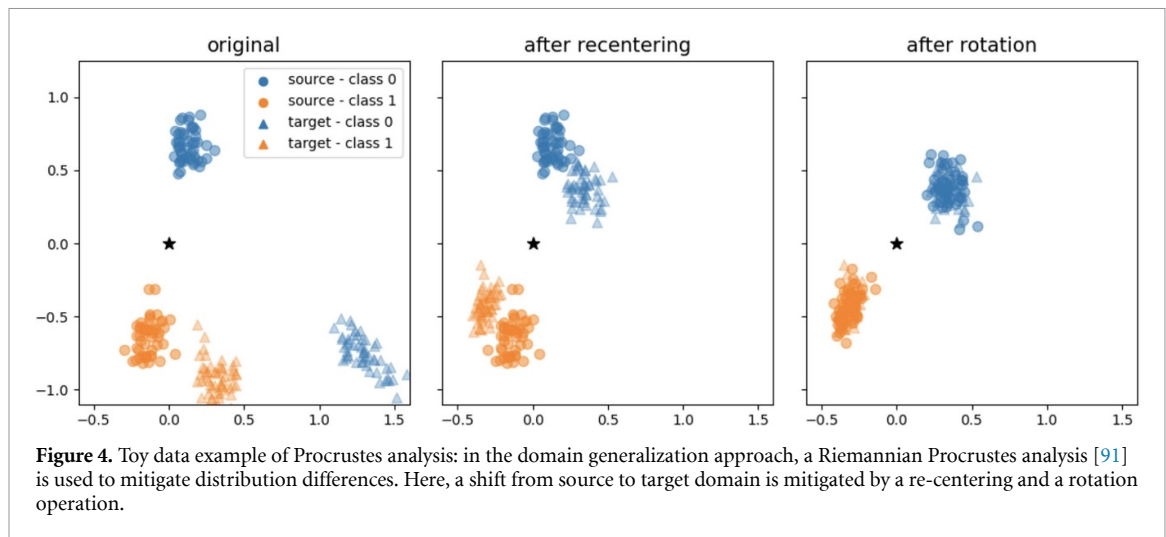
### 4.1. Recurring topics

A thorough discussion between the workshop participants revealed a few recurring concepts that played a key role in the presented methods:

#### 4.1.1. Transfer learning

For MI problems, transfer learning was the dominating approach to deal with the small dataset reality. It was either at the core of every proposed solution presented in the workshop (sections 3.4, 3.7 and 3.5) or implicitly considered (section 3.6).

At the same time, transfer learning was not comparably prominent in algorithms proposed for ERP data. At first glance, this is surprising, as transfer learning is a concept, which generally can be used



**Figure 4.** Toy data example of Procrustes analysis: in the domain generalization approach, a Riemannian Procrustes analysis [91] is used to mitigate distribution differences. Here, a shift from source to target domain is mitigated by a re-centering and a rotation operation.

on top of other algorithmic improvements of decoding models and as this has shown benefits in session-to-session transfer for patient applications [114] and healthy users [115]. A possible reason for this observation is, that ERP-based signals may pose a less difficult decoding problem overall concerning small datasets, as the fast stimulus presentation allows for accumulating a relatively large amount of information within a short amount of time. This is backed by studies with a larger number of participants, which indicate that a higher percentage of healthy BCI users may be able to reach control within a single session using visual ERP protocols, compared to MI protocols [115–118].

#### 4.1.2. Using temporal structure

For evoked data, stimuli are presented repeatedly before a trial decision is derived. The c-VEP reconstruction approach (section 3.2) and UMM decoding method (section 3.3) explicitly exploit the temporal structure of the known flow of events within a trial. They tackle the decoding problem on the granularity of full trials/sequences instead of independently decoding every single evoked potential. Leveraging the temporal/structural information contained in a trial reduces the number of hypotheses about an attended item substantially and thus simplifies the decoding problem. This further mitigates the small dataset reality and even enables—(semi-)unsupervised—instantaneous decoding of single trials.

Unfortunately, most datasets in public repositories do not provide the necessary details about which stimuli belong to which trial/selection step, such that these methods cannot (yet) be benchmarked on a comparably large number of datasets as, for instance, MI decoding.

Furthermore, temporal structure is also used in the LDA improvements TD-LDA and BT-LDA

(section 3.3), which make use of temporal independence assumptions to improve the estimation of the parameters contained in the feature covariance matrix. This can also be interpreted as a form of data augmentation, as the assumptions exploited by those methods allow to estimate the covariance based on multiples of the original number of available epochs.

MI protocols generally do not have a known temporal structure over multiple trials. This makes it more difficult to use the covariance improvements for enhanced decoding. In the workshop the augmentation of spatial covariance matrices using time-shifted signals (section 3.6) was presented as a related approach. Specifically, a temporal embedding was proposed, that captures the within-trial dynamics (but not the temporal structure across trials) to build improved spatial covariance matrices for a Riemannian approach.

#### 4.1.3. Invariances

Becoming invariant to certain transformations of the data, to noise, or to individual differences between subjects was another recurrent concept that generally helps to avoid the recording of novel calibration data points. Affine transformations, for example, do not influence the feature representation when data points are represented in the form of spatial covariance matrices, as in the Riemannian contributions described in sections 3.6 and 3.5. Invariances between datasets can also explicitly be created by the Procrustes transformation (section 3.7), which may enable the transfer between subjects and sessions. However, it requires at least a minimal amount of novel data points from the transfer target domain to estimate mean shifts, shape or rotation of distributions.

Invariances can also be created by suitable forms of data augmentation, for instance by varying the intensity of single sources (section 3.1), or by learning

representations that generalize across sessions, subjects or even datasets (section 3.4). In this context, recently the first approaches for the learning of channel invariance were proposed [119, 120], which may allow to fine-tune models, that had been pre-trained on datasets with different channel sets. In addition, research on generative models of BCI data [121] may provide leverage on tackling invariances, specifically if these models allow for conditioning on varying factors that influence the data distributions.

#### 4.1.4. Approaches inspired by neuroscience

Neuroscientific priors have led to a number of domain-specific methods presented in the workshop. Posthoc labeling (section 3.1) as well as the reconvolution decoding for c-VEP data (section 3.2) made use of models that assume a (linear) interaction between the activity of sources in the brain and the resulting mixed brain signals that can be recorded on the scalp surface after volume conduction. While this linear ERP superposition hypothesis for subsequently presented stimuli is challenged by various observations, it suffices to obtain a highly efficient decoding approach.

Similarly, domain-specific regularization was at the core of the TD-LDA and BT-LDA approaches (section 3.3). They made use of the known autocorrelation structure of measured M/EEG signals over time and could use recorded ERP epochs more efficiently than state-of-the-art methods, which made a substantial difference under small dataset conditions.

## 4.2. Next steps

As part of the workshop, next steps forward in the small dataset challenge were discussed.

Specifically for MI protocols, the workshop participants concluded that the small dataset problem in the BCI domain will probably not be resolved easily in the coming years. The following possible remedies were discussed:

#### 4.2.1. Data collections

First, the potential availability of longer or even 24/7 recordings by future implanted systems could provide new avenues for, e.g. the learning of invariances or other strategies. Second, it was criticized that a lot of existing datasets which partially have been obtained under public funding are not openly accessible. The latter is a goal of data sharing infrastructures such as MOABB [6], OpenNeuro [122], Physionet [123], Open MEG Archive (OMEGA) [124], BNCI Horizon 2020 consortium [125], Radboud Data Repository [126], or GigaDB [127]. However, it was speculated that specifically the labeled datasets contained in those collections will grow slowly only for obvious reasons.

Thus it may be more realistic to expect improvements based on more general DL models/foundation models, which could be pre-trained also on non-BCI datasets before they are fine-tuned for BCI applications. Research on BCI foundation models, however, is starting only [119, 128].

It was proposed that colleagues in the field should support research on decoding models by enlarging the existing dataset collections with their own datasets, ideally organized according to the *brain imaging data structure (BIDS)* format [129], a standard adopted by the field. For data that is not yet recorded, it is helpful to foresee its publication (in anonymized form) right from the beginning of the protocol development and during the preparation of the ethics proposals. If possible, standard paradigms such as resting state recordings should be integrated into protocols, as these allow for the pooling of heterogeneous data over many studies, different user groups and recording conditions, to eventually train foundation models.

#### 4.2.2. Exploiting domain-specific knowledge

The workshop has shown examples on how small dataset problems can be tackled by domain-specific regularization and -decoding approaches. However, these methods are just at their beginning, and maybe a better embracing of domain knowledge can be implemented in novel forms. Two promising directions have been indicated in the workshop, which may be the basis for further improvements, namely (1) making better use of the background signals (i.e. noise) instead of focusing on the foreground signals of interest only, and (2) thinking in sources. The first direction generally allows learning from non-BCI data, or data that is not labeled with user tasks or has not been segmented strictly into epochs or trials. This may provide novel opportunities to learn from inter-trial data, resting state recordings and generally non-BCI data. The second direction describes that insights from neuroscience, specifically different models of data generation in the human brain should be adopted for proposing novel regularization or data augmentation methods, which may have the ability to limit the optimization of model parameters to a reasonably small, neuroscientifically plausible subspace/manifold.

Last but not least it was concluded, that a combination of existing methods might provide advantages for tackling the small dataset problem.

#### 4.2.3. Opportunities for self-supervised learning

In the aftermath to the workshop, we realized that self-supervised learning has not been covered yet substantially by the BCI field, and it was also not contained in the approaches presented during the workshop, see figure 1. We believe that self-supervised

learning could be particularly interesting, as the challenge of working with small datasets extends beyond mere scarcity. Specifically, it may also involve dealing with data that is inherently noisy, potentially rendering portions of the dataset unusable due to artifacts. A general model trained by self-supervised learning might become capable of capturing the intrinsic characteristics of the data. In this case, it could prove valuable in tasks such as repairing faulty segments of a dataset or translating datasets into a standardized format, thereby transcending the limitations imposed by the original hardware setup, like the used electrode layout.

## 5. Conclusion

Finding solutions for small dataset problems is a very active research field. In the workshop, we identified top-down strategies as well as concrete approaches that even now allow to successfully tackle small dataset problems. The underlying research is highly interdisciplinary and profits from a combination of competencies that stem from machine learning, BCI paradigms and neuroscience. For the most pressing remaining challenges, we have identified potential research directions.

## Data availability statement

No new data were created or analyzed in this study.

## Acknowledgment

M T was supported by the project Dutch Brain Interface Initiative (DBI2) with Project Number 024.005.022 of the research programme Gravitation, which is financed by the Dutch Ministry of Education, Culture and Science (OCW) via the Dutch Research Council (NWO). M T and M D received funding from the Dareplane collaboration project, which is co-funded by PPP Allowance awarded by Health Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships as well as by a contribution from the Dutch Brain Foundation. M T and J T were supported by the European Union's research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101118964. R K was supported by the Innovative Science and Technology Initiative for Security Grant Number JPJ004596, ATLA, Japan. S C was supported by DATAIA Convergence Institute as part of the 'Programme d'Investissement d'Avenir', (ANR-17-CONV-0003) operated by LISN-CNRS.

## Author contributions

M T and J T: conceptualization. Writing—original drafts: sections 1, 2 and 4 (M T and J T); section 3.1 (M D); section 3.2 (J T), section 3.3 (M T), section 3.4

(P G), section 3.5 (R K), section 3.6 (T P), section 3.7 (S C). Writing—review & editing: All authors.

## ORCID iDs

Michael Tangermann  <https://orcid.org/0000-0001-6729-0290>  
 Sylvain Chevallier  <https://orcid.org/0000-0003-3027-8241>  
 Matthias Dold  <https://orcid.org/0009-0003-1477-4912>  
 Pierre Guetschel  <https://orcid.org/0000-0002-8933-7640>  
 Theodore Papadopoulou  <https://orcid.org/0000-0002-1643-9988>  
 Jordy Thielen  <https://orcid.org/0000-0002-6264-0367>

## References

- [1] Blankertz B, Müller K-R, Krusienski D J, Schalk G, Wolpaw J R, Schlögl A, Pfurtscheller G, Millán J R, Schröder M and Birbaumer N 2006 The BCI competition III: validating alternative approaches to actual BCI problems *IEEE Trans. Neural Syst. Rehabil. Eng.* **14** 153–9
- [2] Khademi Z, Ebrahimi F and Kordy H M 2023 A review of deep critical challenges in MI-BCI: from conventional to deep learning methods *J. Neurosci. Methods* **383** 109736
- [3] Zhang Z, Duan F, Sole-Casals J, Dinares-Ferran J, Cichocki A, Yang Z and Sun Z 2019 A novel deep learning approach with data augmentation to classify motor imagery signals *IEEE Access* **7** 15945–54
- [4] Aznan N K N, Atapour-Abarghouei A, Bonner S, Connolly J D, Al Moubayed N and Breckon T P 2019 Simulating brain signals: creating synthetic EEG data via neural-based generative models for improved SSVEP classification *2019 Int. Joint Conf. on Neural Networks (IJCNN)* (IEEE) pp 1–8
- [5] Lotte F 2011 Generating artificial EEG signals to reduce BCI calibration time *5th Int. Brain-Computer Interface Workshop* pp 176–9
- [6] Aristimunha B et al 2023 Mother of all BCI Benchmarks *Zenodo* <https://doi.org/10.5281/zenodo.10034223>
- [7] Kindermans P J, Schreuder M, Schrauwen B, Müller K-R and Tangermann M 2014 True zero-training brain-computer interfacing—an online study *PLoS One* **9** e102504
- [8] Trocellier D, N'Kaoua B and Lotte F 2024 Validating neurophysiological predictors of BCI performance on a large open source dataset *9th Graz Brain-Computer Interface Conf. 2024-GBCIC2024*
- [9] Kübler A, Mattia D, Rupp R and Tangermann M 2013 Facing the challenge: bringing brain-computer interfaces to end-users *Artif. Intell. Med.* **59** 55–60
- [10] Castaño-Candamil S Post-HocLabeling (available at: <https://github.com/bsdlab/Post-HocLabeling>)
- [11] Dold M BCIC2023 Post-Hoc (available at: [https://github.com/bsdlab/BCIC2023\\_posthoc](https://github.com/bsdlab/BCIC2023_posthoc))
- [12] Thielen J PyntBCI (available at: <https://github.com/thijor/pyntbci>)
- [13] Sosulski J Time-decoupled-lda (available at: <https://github.com/jsosulski/time-decoupled-lda>)
- [14] Sosulski J Toeplitzlda (available at: <https://github.com/jsosulski/toeplitzlda>)
- [15] Sosulski J umm (available at: <https://github.com/jsosulski/umm>)
- [16] Pre-trained motor-imagery models (available at: [https://neurotechlab.socsci.ru.nl/resources/pretrained\\_imagery\\_models/](https://neurotechlab.socsci.ru.nl/resources/pretrained_imagery_models/))

- [17] Model Card for pre-trained EEGNet models on mental imagery datasets (available at: <https://huggingface.co/PierreGtch/EEGNetv4>)
- [18] Kobler R TSMNet (available at: <https://github.com/rkobler/TSMNet>)
- [19] Carrara I Augmented covariance BCI (available at: [https://github.com/carraraig/Augmented\\_Covariance\\_BCI](https://github.com/carraraig/Augmented_Covariance_BCI))
- [20] Barachant A et al 2023 pyRiemann/pyRiemann: v0.5 Zenodo <https://doi.org/10.5281/zenodo.8059038>
- [21] Barachant A et al pyRiemann (available at: <https://github.com/pyRiemann/pyRiemann>)
- [22] Friedrich E V, Scherer R and Neuper C 2012 The effect of distinct mental strategies on classification performance for brain-computer interfaces *Int. J. Psychophysiol.* **84** 86–94
- [23] Tangermann M, Schreuder M, Dähne S, Höhne J, Regler S, Ramsay A, Quek M, Williamson J and Murray-Smith R 2011 Optimized stimulation events for a visual ERP BCI *Int. J. Bioelectromagn.* **13** 119–20
- [24] Jia C, Gao X, Hong B and Gao S 2010 Frequency and phase mixed coding in SSVEP-based brain-computer interface *IEEE Trans. Biomed. Eng.* **58** 200–6
- [25] Sugi M, Hagimoto Y, Nambu I, Gonzalez A, Takei Y, Yano S, Hokari H and Wada Y 2018 Improving the performance of an auditory brain-computer interface using virtual sound sources by shortening stimulus onset asynchrony *Front. Neurosci.* **12** 108
- [26] Höhne J and Tangermann M 2012 How stimulation speed affects event-related potentials and BCI performance *2012 Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (IEEE)* pp 1802–5
- [27] Kaufmann T, Schulz S M, Grünzinger C and Kübler A 2011 Flashing characters with famous faces improves ERP-based brain-computer interface performance *J. Neural Eng.* **8** 056016
- [28] Kindermans P J, Verschore H and Schrauwen B 2013 A unified probabilistic approach to improve spelling in an event-related potential-based brain-computer interface *IEEE Trans. Biomed. Eng.* **60** 2696–705
- [29] Kindermans P J, Tangermann M, Müller K-R and Schrauwen B 2014 Integrating dynamic stopping, transfer learning and language models in an adaptive zero-training ERP speller *J. Neural Eng.* **11** 035005
- [30] Kuncheva L I, Whitaker C J and Narasimhamurthy A 2008 A case-study on naïve labelling for the nearest mean and the linear discriminant classifiers *Pattern Recognit.* **41** 3010–20
- [31] Shenoy P, Krauledat M, Blankertz B, Rao R P and Müller K-R 2006 Towards adaptive classification for BCI *J. Neural Eng.* **3** R13
- [32] Vidaurre C, Kawanabe M, von Bünau P, Blankertz B and Müller K-R 2010 Toward unsupervised adaptation of LDA for brain-computer interfaces *IEEE Trans. Biomed. Eng.* **58** 587–97
- [33] Rommel C, Paillard J, Moreau T and Gramfort A 2022 Data augmentation for learning predictive models on EEG: a systematic comparison *J. Neural Eng.* **19** 066020
- [34] Fahimi F, Zhang Z, Goh W B, Ang K K and Guan C 2019 Towards EEG generation using GANs for BCI applications *2019 IEEE EMBS Int. Conf. on Biomedical & Health Informatics (BHI)* (IEEE) pp 1–4
- [35] Zhang K et al 2024 Self-supervised learning for time series analysis: taxonomy, progress and prospects *IEEE Trans. Pattern Anal. Mach. Intell.* **46** 6775–94
- [36] Jayaram V, Alamgir M, Altun Y, Schölkopf B and Grosse-Wentrup M 2016 Transfer learning in brain-computer interfaces *IEEE Comput. Intell. Mag.* **11** 20–31
- [37] Mellot A, Collas A, Chevallier S, Engemann D and Gramfort A 2024 Physics-informed and unsupervised Riemannian domain adaptation for machine learning on heterogeneous EEG datasets (arXiv:2403.15415)
- [38] Junqueira B, Aristimunha B, Chevallier S and de Camargo R Y 2024 A systematic evaluation of Euclidean alignment with deep learning for EEG decoding *J. Neural Eng.* **21** 036038
- [39] Wu D, Xu Y and Lu B L 2020 Transfer learning for EEG-based brain-computer interfaces: a review of progress made since 2016 *IEEE Trans. Cogn. Dev. Syst.* **14** 4–19
- [40] Guetschel P, Ahmadi S and Tangermann M 2024 Review of deep representation learning techniques for brain-computer interfaces *J. Neural Eng.* **21** 061002
- [41] Castaño-Candamil S, Meinel A and Tangermann M 2019 Post-hoc labeling of arbitrary M/EEG recordings for data-efficient evaluation of neural decoding methods *Front. Neuroinform.* **13** 55
- [42] Makeig S, Bell A, Jung T P and Sejnowski T J 1995 Independent component analysis of electroencephalographic data *Advances in Neural Information Processing Systems* vol 8, ed D Touretzky, M C Mozer and M Hasselmo (MIT Press)
- [43] Haufe S, Meinecke F, Görgen K, Dähne S, Haynes J-D, Blankertz B and Bießmann F 2014 On the interpretation of weight vectors of linear models in multivariate neuroimaging *NeuroImage* **87** 96–110
- [44] Grech R, Cassar T, Muscat J, Camilleri K P, Fabri S G, Zervakis M, Xanthopoulos P, Sakkalis V and Vanrumste B 2008 Review on solving the inverse problem in EEG source analysis *J. NeuroEng. Rehabil.* **5** 25
- [45] Castaño-Candamil S, Höhne J, Martínez-Vargas J D, An X W, Castellanos-Domínguez G and Haufe S 2015 Solving the EEG inverse problem based on space-time-frequency structured sparsity constraints *NeuroImage* **118** 598–612
- [46] Gramfort A, Strohmeier D, Hauelsen J, Hämäläinen M S and Kowalski M 2013 Time-frequency mixed-norm estimates: sparse M/EEG imaging with non-stationary source activations *NeuroImage* **70** 410–22
- [47] Zlatov O and Blankertz B 2022 Towards physiology-informed data augmentation for EEG-based BCIs (arXiv:2203.14392)
- [48] Shi N, Miao Y, Huang C, Li X, Song Y, Chen X, Wang Y and Gao X 2024 Estimating and approaching the maximum information rate of noninvasive visual brain-computer interface *NeuroImage* **289** 120548
- [49] Miao Y, Shi N, Huang C, Song Y, Chen X, Wang Y and Gao X 2024 High-performance c-VEP-BCI under minimal calibration *Expert Syst. Appl.* **249** 123679
- [50] Martínez-Cagigal V, Thielen J, Santamaría-Vázquez E, Pérez-Velasco S, Desain P and Hornero R 2021 Brain-computer interfaces based on code-modulated visual evoked potentials (c-VEP): a literature review *J. Neural Eng.* **18** 061002
- [51] Fazel-Rezai R, Allison B Z, Guger C, Sellers E W, Kleih S C and Kübler A 2012 P300 brain computer interface: current challenges and emerging trends *Front. Neuroeng.* **5** 28055
- [52] Zhu D, Bieger J, Garcia Molina G and Aarts R M 2010 A survey of stimulation methods used in SSVEP-based BCIs *Comput. Intell. Neurosci.* **2010** 1–12
- [53] Chen X, Wang Y, Nakanishi M, Gao X, Jung T P and Gao S 2015 High-speed spelling with a noninvasive brain-computer interface *Proc. Natl Acad. Sci.* **112** E6058–67
- [54] Gao S, Wang Y, Gao X and Hong B 2014 Visual and auditory brain-computer interfaces *IEEE Trans. Biomed. Eng.* **61** 1436–47
- [55] Thielen J 2023 Effects of stimulus sequences on brain-computer interfaces using code-modulated visual evoked potentials: an offline simulation *Int. Work-Confer. on Artificial Neural Networks* (Springer) pp 555–68
- [56] Thielen J, van den Broek P, Farquhar J and Desain P 2015 Broad-band visually evoked potentials: re(con)volution in brain-computer interfacing *PLoS One* **10** e0133797
- [57] Capilla A, Pazo-Alvarez P, Darriba A, Campo P and Gross J 2011 Steady-state visual evoked potentials can be explained

- by temporal superposition of transient event-related responses *PLoS One* **6** e14543
- [58] Thielen J, Marsman P, Farquhar J and Desain P 2021 From full calibration to zero training for a code-modulated visual evoked potentials for brain–computer interface *J. Neural Eng.* **18** 056007
- [59] Thielen J, Sosulski J and Tangermann M 2024 Exploring new territory: calibration-free decoding for c-VEP BCI (arXiv:2403.15521)
- [60] Castillos K C, Ladouce S, Darmet L and Dehais F 2023 Burst c-VEP based BCI: optimizing stimulus design for enhanced classification with minimal calibration data and improved user experience *NeuroImage* **284** 120446
- [61] Dehais F, Castillos K C, Ladouce S and Clisson P 2024 Leveraging textured flickers: a leap toward practical, visually comfortable and high-performance dry EEG code-VEP BCI *J. Neural Eng.* **21** 066023
- [62] Martínez-Cagigal V, Santamaría-Vázquez E, Pérez-Velasco S, Marcos-Martínez D, Moreno-Calderón S and Hornero R 2023 Non-binary m-sequences for more comfortable brain–computer interfaces based on c-VEPs *Expert Syst. Appl.* **232** 120815
- [63] Zheng L, Dong Y, Tian S, Pei W, Gao X and Wang Y 2024 A calibration-free c-VEP based BCI employing narrow-band random sequences *J. Neural Eng.* **21** 026023
- [64] Stawicki P and Volosyak I 2022 CVEP training data validation—towards optimal training set composition from multi-day data *Brain Sci.* **12** 234
- [65] Huang Z, Zheng W, Wu Y and Wang Y 2020 Ensemble or pool: a comprehensive study on transfer learning for c-VEP BCI during interpersonal interaction *J. Neurosci. Methods* **343** 108855
- [66] Ahmadi S, Desain P and Thielen J 2024 A Bayesian dynamic stopping method for evoked response brain-computer interfacing *Front. Hum. Neurosci.* **18** 1437965
- [67] Thielen J, Farquhar J and Desain P 2024 *Broad-Band Visually Evoked Potentials: Re(con)volution in Brain-Computer Interfacing* (Radboud University)
- [68] Thielen J, Marsman P, Farquhar J and Desain P 2023 *From Full Calibration to Zero Training for a Code-Modulated Visual Evoked Potentials Brain Computer Interface* (Radboud University)
- [69] Castillos K C 2023 4-class code-VEP EEG data *Zenodo* (<https://doi.org/10.5281/zenodo.8255618>)
- [70] Tien D, Darmet L, Gravis V 2022 ludovicdm/FlickersOnTop: release for Zenodo *Zenodo* (<https://doi.org/10.5281/zenodo.7445755>)
- [71] Dold M, Pereira J, Sajonz B, Coenen V A, Thielen J, Janssen M L and Tangermann M 2025 A modular open-source software platform for BCI research with application in closed-loop deep brain stimulation *J. Neural Eng.* **22** 026029
- [72] Thielen J dp-cvep (available at: <https://github.com/thijor/dp-cvep>)
- [73] Santamaría-Vázquez E, Martínez-Cagigal V, Marcos-Martínez D, Rodríguez-González V, Pérez-Velasco S, Moreno-Calderón S and Hornero R 2023 MEDUSA©: a novel Python-based software ecosystem to accelerate brain-computer interface and cognitive neuroscience research *Comput. Methods Programs Biomed.* **230** 107357
- [74] Blankertz B, Lemm S, Treder M, Haufe S and Müller K-R 2011 Single-trial analysis and classification of ERP components—a tutorial *NeuroImage* **56** 814–25
- [75] Sosulski J, Kemmer J P and Tangermann M 2021 Improving covariance matrices derived from tiny training datasets for the classification of event-related potentials with linear discriminant analysis *Neuroinformatics* **19** 461–76
- [76] Sosulski J and Tangermann M 2022 Introducing block-Toeplitz covariance matrices to remaster linear discriminant analysis for event-related potential brain–computer interfaces *J. Neural Eng.* **19** 066001
- [77] Hübner D, Verhoeven T, Schmid K, Müller K-R, Tangermann M and Kindermans P J 2017 Learning from label proportions in brain-computer interfaces: online unsupervised learning with guarantees *PLoS One* **12** e0175856
- [78] Hübner D, Verhoeven T, Müller K-R, Kindermans P J and Tangermann M 2018 Unsupervised learning for brain-computer interfaces based on event-related potentials: review and online comparison *IEEE Comput. Intell. Mag.* **13** 66–77
- [79] Sosulski J and Tangermann M 2023 UMM: unsupervised mean-difference maximization (arXiv:2306.11830)
- [80] Zhu Y, Li Y, Lu J and Li P 2021 EEGNet with ensemble learning to improve the cross-session classification of SSVEP based BCI from Ear-EEG *IEEE Access* **9** 15295–303
- [81] Samek W, Meinecke F C and Müller K R 2013 Transferring subspaces between subjects in brain–computer interfacing *IEEE Trans. Biomed. Eng.* **60** 2289–98
- [82] Jeon E, Ko W, Yoon J S and Suk H I 2021 Mutual information-driven subject-invariant and class-relevant deep representation learning in BCI *IEEE Trans. Neural Netw. Learn. Syst.* **34** 739–49
- [83] Kobler R, Hirayama J, Zhao Q and Kawanabe M 2022 SPD domain-specific batch normalization to crack interpretable unsupervised domain adaptation in EEG *Advances in Neural Information Processing Systems* vol 35, ed S Koyejo, S Mohamed, A Agarwal, D Belgrave, K Cho and A Oh (Curran Associates, Inc.) pp 6219–35
- [84] Xie Y, Wang K, Meng J, Yue J, Meng L, Yi W, Jung T-P, Xu M and Ming D 2023 Cross-dataset transfer learning for motor imagery signal classification via multi-task learning and pre-training *J. Neural Eng.* **20** 056037
- [85] Guetschel P and Tangermann M 2023 Transfer learning between motor imagery datasets using deep learning – validation of framework and comparison of datasets (arXiv:2311.16109)
- [86] Congedo M, Barachant A and Bhatia R 2017 Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review *Brain Comput. Interfaces* **4** 155–74
- [87] Barachant A, Bonnet S, Congedo M and Jutten C 2012 Multiclass brain-computer interface classification by Riemannian geometry *IEEE Trans. Bio Med. Eng.* **59** 920–8
- [88] Roy R N, Hinss M F, Darmet L, Ladouce S, Jahanpour E S, Somon B, Xu X, Drougard N, Dehais F and Lotte F 2022 Retrospective on the first passive brain-computer interface competition on cross-session workload estimation *Front. Neuroergonom.* **3** 838342
- [89] Chevallier S et al 2024 The largest EEG-based BCI reproducibility study for open science: the MOABB benchmark (arXiv:0912.1470)
- [90] Zanini P, Congedo M, Jutten C, Said S and Berthoumieu Y 2018 Transfer learning: a Riemannian geometry framework with applications to brain–computer interfaces *IEEE Trans. Biomed. Eng.* **65** 1107–16
- [91] Rodrigues P L C, Jutten C and Congedo M 2019 Riemannian procrustes analysis: transfer learning for brain–computer interfaces *IEEE Trans. Biomed. Eng.* **66** 2390–401
- [92] Huang Z and Gool L V 2017 A Riemannian network for SPD matrix learning *Proc. 31st AAAI Conf. on Artificial Intelligence. AAAI'17 (AAAI Press)* (Event-place) pp 2036–42
- [93] Brooks D, Schwander O, Barbaresco F, Schneider J Y and Cord M 2019 Riemannian batch normalization for SPD neural networks *Advances in Neural Information Processing Systems* vol 32, ed H Wallach, H Larochelle, A Beygelzimer, A-B Fd, E Fox and R Garnett (Curran Associates, Inc.)
- [94] Kobler R J, Hirayama J and Kawanabe M 2022 Controlling the Fréchet variance improves batch normalization on the symmetric positive definite manifold *ICASSP 2022 - 2022 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE) pp 3863–7

- [95] Pennec X 2020 Manifold-valued image processing with SPD matrices *Riemannian Geometric Statistics in Medical Image Analysis* (Elsevier) pp 75–134
- [96] Wei X et al 2022 2021 BEETL competition: advancing transfer learning for subject independence and heterogeneous EEG data sets *Proc. NeurIPS 2021 Competitions and Demonstrations Track (Proc. Machine Learning Research* vol 176) ed D Kiela, M Ciccone and B Caputo (PMLR) pp 205–19
- [97] Ju C and Guan C 2023 Tensor-CSPNet: a novel geometric deep learning framework for motor imagery classification *IEEE Trans. Neural Netw. Learn. Syst.* **34** 10955–69
- [98] Ju C and Guan C 2023 Graph neural networks on SPD manifolds for motor imagery classification: a perspective from the time–frequency analysis *IEEE Trans. Neural Netw. Learn. Syst.* **35** 17701–15
- [99] Pan Y T, Chou J L and Wei C S 2022 MAtt: a manifold attention network for EEG decoding *Advances in Neural Information Processing Systems* vol 35, ed S Koyejo, S Mohamed, A Agarwal, D Belgrave, K Cho and A Oh (Curran Associates, Inc.) pp 31116–29
- [100] Wilson D, Schirrmeyer R T, Gemein L A W and Ball T 2022 Deep Riemannian networks for EEG decoding (arXiv:2212.10426)
- [101] Paillard J, Hipp J F and Engemann D A 2024 GREEN: a lightweight architecture using learnable wavelets and Riemannian geometry for biomarker exploration *bioRxiv Preprint* (version 03 June 2024) (<https://doi.org/10.1101/2024.05.14.594142>)
- [102] Hehenberger L, Kobler R J, Lopes-Dias C, Srisrisawang N, Tumfart P, Uroko J B, Torke P R and Müller-Putz G R 2021 Long-term mutual training for the CYBATHLON BCI Race with a tetraplegic pilot: a case study on inter-session transfer and intra-session adaptation *Front. Hum. Neurosci.* **15** 635777
- [103] Bakas S, Ludwig S, Adamos D A, Laskaris N, Panagakis Y and Zafeiriou S 2023 Latent alignment with deep set EEG decoders (arXiv:2311.17968)
- [104] Sabbagh D, Ablin P, Varoquaux G, Gramfort A and Engemann D A 2020 Predictive regression modeling with MEG/EEG: from source power to signals and cognitive states *NeuroImage* **222** 116893
- [105] Kobler R J, Hirayama J I, Hehenberger L, Lopes-Dias C, Müller-Putz G R and Kawanabe M 2021 On the interpretation of linear Riemannian tangent space model parameters in M/EEG 2021 43rd Annual Int. Conf. IEEE Engineering in Medicine & Biology Society (EMBC) (IEEE) pp 5909–13
- [106] Carrara I and Papadopoulou T 2024 Classification of BCI-EEG based on the augmented covariance matrix *IEEE Trans. Biomed. Eng.* **71** 2651–62
- [107] Barachant A and Congedo M A Plug&Play P300 BCI using information geometry *CoRR* 2014 (arXiv:1409.0107)
- [108] Chevallier S, Kalunga E K, Barthélemy Q and Monacelli E 2021 Review of Riemannian distances and divergences, applied to SSVEP-based BCI *Neuroinformatics* **19** 93–106
- [109] Takens F 2006 Detecting strange attractors in turbulence *Dynamical Systems and Turbulence, Warwick 1980: Proc. Symp. Held at the University of Warwick 1979/80* (Springer) pp 366–81
- [110] Chevallier S, Kalunga E K, Barthélemy Q and Yger F 2018 *Riemannian Classification for SSVEP-Based BCI: Offline Versus Online Implementations* (CRC Press) pp 371–96
- [111] Kang H, Nam Y and Choi S 2009 Composite common spatial pattern for subject-to-subject transfer *IEEE Signal Process. Lett.* **16** 683–6
- [112] Kalunga E K, Chevallier S and Barthélemy Q 2018 Transfer learning for SSVEP-based BCI using Riemannian similarities between users *European Signal Processing Conf. (EUSIPCO)* pp 1685–9
- [113] Khazem S, Chevallier S, Barthélemy Q, Haroun K and Nouïs C 2021 Minimizing subject-dependent calibration for BCI with Riemannian transfer learning *IEEE/EMBS Conf. on Neural Engineering (NER)* pp 523–6
- [114] Musso M, Hübner D, Schwarzkopf S, Bernodsson M, LeVan P, Weiller C and Tangermann M 2022 Aphasia recovery by language training using a brain–computer interface: a proof-of-concept study *Brain Commun.* **4** fcac008
- [115] Lee M H, Kwon O-Y, Kim Y-J, Kim H-K, Lee Y-E, Williamson J, Fazli S and Lee S-W 2019 EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy *GigaScience* **8** giz002
- [116] Blankertz B, Sannelli C, Müller K-R, Curio G and Blankertz B 2009 Predicting BCI performance to study BCI illiteracy *BMC Neuroscience* **10** 84
- [117] Edlinger G, Allison B Z and Guger C 2015 How many people can use a BCI system? *Clinical Systems Neuroscience* (Springer) pp 33–66
- [118] Volosyak I, Rezeika A, Benda M, Gembler F and Stawicki P 2020 Towards solving of the Illiteracy phenomenon for VEP-based brain-computer interfaces *Biomed. Phys. Eng. Express* **6** 035034
- [119] Guetschel P, Moreau T and Tangermann M 2024 S-JEPA: towards seamless cross-dataset transfer through dynamic spatial attention (arXiv:2403.11772)
- [120] Yang C, Westover M B and Sun J 2023 BIOT: cross-data biosignal learning in the wild (arXiv:2305.10351)
- [121] Klein G, Guetschel P, Silvestri G and Tangermann M 2024 Synthesizing EEG signals from event-related potential paradigms with conditional diffusion models (arXiv:2403.18486)
- [122] Markiewicz C J et al 2021 The OpenNeuro resource for sharing of neuroscience data *eLife* **10** e71774
- [123] Moody G B 2022 PhysioNet *Encyclopedia of Computational Neuroscience* (Springer) pp 2806–8
- [124] Niso G et al 2016 OMEGA: the open MEG archive *NeuroImage* **124** 1182–7
- [125] Brunner C et al 2015 BNCI Horizon 2020: towards a roadmap for the BCI community *Brain Comput. Interfaces* **2** 1–10
- [126] RDR Radboud data repository (available at: <https://data.ru.nl/>)
- [127] Sneddon T P, Li P and Edmunds S C 2012 GigaDB: announcing the GigaScience database *GigaScience* **1** 2047–217X
- [128] Cui W, Jeong W, Thölke P, Medani T, Jerbi K, Joshi A A and Leahy R M 2023 Neuro-GPT: developing a foundation model for EEG (arXiv:231103764)
- [129] Gorgolewski K J et al 2016 The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments *Sci. Data* **3** 1–9