



## OPEN ACCESS

## EDITED BY

Jianjun Meng,  
Shanghai Jiao Tong University, China

## REVIEWED BY

Sung-Phil Kim,  
Ulsan National Institute of Science and  
Technology, Republic of Korea  
Vassily Tsytarev,  
University of Maryland, United States

## \*CORRESPONDENCE

Tetiana Aksenova  
✉ tetiana.aksenova@cea.fr

## SPECIALTY SECTION

This article was submitted to  
Brain-Computer Interfaces,  
a section of the journal  
Frontiers in Human Neuroscience

RECEIVED 29 November 2022

ACCEPTED 27 February 2023

PUBLISHED 16 March 2023

## CITATION

Śliwowski M, Martin M, Souloumiac A,  
Blanchart P and Aksenova T (2023) Impact of  
dataset size and long-term ECoG-based BCI  
usage on deep learning decoders performance.  
*Front. Hum. Neurosci.* 17:1111645.  
doi: 10.3389/fnhum.2023.1111645

## COPYRIGHT

© 2023 Śliwowski, Martin, Souloumiac,  
Blanchart and Aksenova. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).  
The use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in this  
journal is cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# Impact of dataset size and long-term ECoG-based BCI usage on deep learning decoders performance

Maciej Śliwowski<sup>1,2</sup>, Matthieu Martin<sup>1</sup>, Antoine Souloumiac<sup>2</sup>,  
Pierre Blanchart<sup>2</sup> and Tetiana Aksenova<sup>1\*</sup>

<sup>1</sup>Université Grenoble Alpes, CEA, LETI, Clnatec, Grenoble, France, <sup>2</sup>Université Paris-Saclay, CEA, List, Palaiseau, France

**Introduction:** In brain-computer interfaces (BCI) research, recording data is time-consuming and expensive, which limits access to big datasets. This may influence the BCI system performance as machine learning methods depend strongly on the training dataset size. Important questions arise: taking into account neuronal signal characteristics (e.g., non-stationarity), can we achieve higher decoding performance with more data to train decoders? What is the perspective for further improvement with time in the case of long-term BCI studies? In this study, we investigated the impact of long-term recordings on motor imagery decoding from two main perspectives: model requirements regarding dataset size and potential for patient adaptation.

**Methods:** We evaluated the multilinear model and two deep learning (DL) models on a long-term BCI & Tetraplegia ([ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT02550522) identifier: NCT02550522) clinical trial dataset containing 43 sessions of ECoG recordings performed with a tetraplegic patient. In the experiment, a participant executed 3D virtual hand translation using motor imagery patterns. We designed multiple computational experiments in which training datasets were increased or translated to investigate the relationship between models' performance and different factors influencing recordings.

**Results:** Our results showed that DL decoders showed similar requirements regarding the dataset size compared to the multilinear model while demonstrating higher decoding performance. Moreover, high decoding performance was obtained with relatively small datasets recorded later in the experiment, suggesting motor imagery patterns improvement and patient adaptation during the long-term experiment. Finally, we proposed UMAP embeddings and local intrinsic dimensionality as a way to visualize the data and potentially evaluate data quality.

**Discussion:** DL-based decoding is a prospective approach in BCI which may be efficiently applied with real-life dataset size. Patient-decoder co-adaptation is an important factor to consider in long-term clinical BCI.

## KEYWORDS

ECoG, motor imagery, deep learning, tetraplegia, adaptation, dataset size, learning curve

## 1. Introduction

Permanent motor deficits as a result of a spinal cord injury (SCI) affect hundreds of thousands of people worldwide each year (12,000 people each year just in the United States [Hachem et al., 2017](#)). In this case, the motor cortex is preserved, but neuronal signals can no longer be transmitted to the muscles. Then, the use of a brain-computer interface (BCI),

which enables interaction with an effector by thought, could enable these patients to regain a certain autonomy in everyday life. For example, motor imagery based BCI has been used for the control of prostheses or exoskeletons of upper limbs (Hochberg et al., 2012; Collinger et al., 2013; Wodlinger et al., 2014; Edelman et al., 2019), lower limbs (López-Larraz et al., 2016; He et al., 2018) and four limbs (Benabid et al., 2019) in subjects with paraplegia or tetraplegia following an SCI. In this study, we focus on electrocorticography (ECoG)-based motor BCIs, promising tools that may enable continuous 3D hand trajectory decoding for neuroprosthesis control while reducing the risk of implantation compared to more invasive approaches (Volkova et al., 2019).

BCIs record neuronal activity and decode it into control commands for effectors. Decoders are generally trained using machine learning algorithms in a supervised manner. In the vast majority of studies, the training dataset is strongly restricted due to limited access to recordings. At the same time, dataset size is an important factor in machine learning analysis and can influence overall system performance drastically. In contrast to recent computer vision and natural language processing studies (Kaplan et al., 2020; Rosenfeld et al., 2020; Hoiem et al., 2021), the optimal quantity of training data, i.e., the quantity at which decoder's performance reaches a plateau for a given application, is rarely studied for BCI (Perdikis and Millan, 2020). Especially, learning curves, providing insight into the relationship between model performance and training set size, are rarely presented. Learning curves can be used for model selection, decreasing the computational load of model training, or estimating the theoretical influence of adding more data to training datasets (Viering and Loog, 2021). The last point is particularly important in BCI, considering the limited access to datasets recorded with humans. Without knowing the relationship between system performance and dataset size, it is hard to determine the strategy to improve the accuracy of decoders: increase the amount of training data or increase the capacity of the models. In the case of ECoG-based motor BCI, most models have a limited capacity. The decoders used are Kalman filters (Pistohl et al., 2012; Silversmith et al., 2020) and mostly variants of linear models (Flamary and Rakotomamonjy, 2012; Liang and Bougrain, 2012; Nakanishi et al., 2013, 2017; Chen et al., 2014; Bundy et al., 2016; Eliseyev et al., 2017). In most of these studies, decoder optimization has been carried out on databases containing a few minutes or tens of minutes of the signal. This results in usable models but does not provide any information on the performance gain that could be achieved with more data, nor does it compare the data quantity/performance relationship between several decoders.

Model characteristics and learning curves are not the only factors influencing decoders' performance in the case of BCI. The human ability to generate distinct brain signal patterns is crucial for a BCI system. Research in recent years has focused mainly on the development of increasingly efficient decoders, for example, deep learning (DL) (Bashivan et al., 2015; Elango et al., 2017; Schirrmeyer et al., 2017; Du et al., 2018; Lawhern et al., 2018; Pan et al., 2018; Xie et al., 2018; Zhang et al., 2019; Rashid et al., 2020; Śliwowski et al., 2022) rather than on patient learning or co-adaptation (Wolpaw et al., 2002; Millan, 2004), even though several studies have shown the crucial importance of

patient learning (Carmena, 2013; Lotte et al., 2013; Orsborn et al., 2014; Chavarriaga et al., 2017; Benaroch et al., 2021). Thanks to recording device developments and clinical trial advances, long-term studies of chronic BCI enable recording of bigger datasets than ever before. Current techniques for recording brain activity, such as the ElectroCorticoGram (ECoG), provide stable recordings for at least 2 years (Nurse et al., 2017). It offers the possibility to train and test a decoder over several months. It also enables studying potential patient learning and provides insight into the optimal quantity of data necessary to get the best out of a decoder. These questions have largely been put aside (Perdikis and Millan, 2020).

Closed-loop learning allows for short-term patient-model co-adaptation through the visual feedback received by the patient. This feedback leads to a modification of the brain activity and has shown capabilities for improving the control of neuroprostheses (Cunningham et al., 2011; Jarosiewicz et al., 2013; Orsborn et al., 2014; Shanechi et al., 2017; Sitaram et al., 2017). Nevertheless, motor learning is a process that takes place in the short term and in the long term (Dayan and Cohen, 2011; Krakauer et al., 2019). This long-term learning is little studied in BCI, and most studies in humans are limited to a few sessions (<15) (Holz et al., 2013; Höhne et al., 2014; Leeb et al., 2015; Meng et al., 2016) to show that a fast and efficient calibration of the proposed decoders is possible. Several studies with a larger number of sessions (>20) were nevertheless carried out: Neuper et al. (2003), Wolpaw and McFarland (2004), Hochberg et al. (2006), McFarland et al. (2010), Collinger et al. (2013), Wodlinger et al. (2014), Ajiboye et al. (2017), Perdikis et al. (2018), Benaroch et al. (2021), and Moly et al. (2022). Some have focused on patient learning (Neuper et al., 2003; McFarland et al., 2010; Collinger et al., 2013; Leeb et al., 2015; Perdikis et al., 2018; Benaroch et al., 2021) by seeking an improvement in performance coming from changes in the signal or the characteristics extracted from it. The last point is required to distinguish between performance improvement due to patient learning, increased data available for decoder optimization, or changes in the experimental environment (Lotte and Jeunet, 2018; Perdikis and Millan, 2020).

As dataset size is an important limitation influencing BCI research, we investigated the relationship between BCI decoders' (predicting 3D upper-limb movements from ECoG signals) performance and the training dataset size used to optimize model parameters. Learning curves obtained in different offline computational experiments showed that DL models could provide similar or better performance without requiring more training data than a multilinear model. Moreover, learning curves revealed characteristics that were unlikely caused by just the dataset increase. Extended analysis using unsupervised ML methods showed dataset characteristic (e.g., intrinsic dimensionality, states separability) changes with time, suggesting that long-term patient learning may play an important role in achieving higher BCI performance. This kind of analysis was possible thanks to the access to a rare database of ECoG signals (Moly et al., 2022) containing imagined hand movements performed by a tetraplegic patient to control upper-limb 3D translation in a virtual environment. However, our study was limited to only one patient and a specific task. Despite that, our results may be a reference for other BCI researchers showing that deep

learning could be advantageous also in the case of limited clinical datasets.

## 2. Materials and methods

### 2.1. Clinical trial and patient

The data was recorded and analyzed as a part of the “BCI and Tetraplegia” (ClinicalTrials.gov identifier: NCT02550522) clinical trial, which was approved by the Agency for the Safety of Medicines and Health Products (Agence nationale de sécurité du médicament et des produits de santé—ANSM) with the registration number: 2015-A00650-49 and the ethical Committee for the Protection of Individuals (Comité de Protection des Personnes—CPP) with the registration number: 15-CHUG-19.

The participant was a 28-year-old right-handed man following tetraplegia after a C4-C5 spinal cord injury. He had residual control over upper limbs with American Spinal Injury Association Impairment (ASIA) scores of 4 (right hand), 5 (left hand) at the level of the elbow, and 0 (right hand), 3 (left hand) at the extensors of the wrist. All motor functions below were completely lost (ASIA score of 0) (Benabid et al., 2019). Two WIMAGINE implants (Mestais et al., 2015), recording ECoG signal at 586 Hz sampling rate, were implanted above the left and right primary motor and sensory cortex responsible for upper limb movements. The implants consisted of an 8 × 8 electrodes grid. Due to the data transfer limit, only 32 electrodes organized on a chessboard-like grid were used for recording at each implant, totaling 64 electrodes.

The data recordings used in this study started after 463 days post-implantation. The subject was already experienced in the BCI setup as the clinical trial experiments began shortly after the surgery. During the clinical trial, the participant gradually learned how to control the BCI, starting by using discrete/1D effectors and finally achieving control of up to 8D movements in one experimental session.

### 2.2. Data and experimental paradigm

The dataset analyzed in this study contains 43 experimental sessions recorded over more than 200 days. In the experiments, the tetraplegic patient was asked to perform motor imagery tasks in order to move virtual exoskeleton effectors (see the virtual environment in Figure 1). In particular, the patient used an MI strategy in which he repeatedly imagined/attempted fingers, hands, and arm movements to control 8 dimensions (3D left and right hand translation and 1D left and right wrist rotation). In every trial, the patient’s goal was to reach the target displayed on the screen, one after another, without returning to the center position (Moly et al., 2022). The dataset consisted of 300 and 284 min of the recorded signal, comprising 19 and 18 trials per session on average, 811 and 756 trials in common, for the left hand and right hand, respectively. During the experiment, target localization was defined by the experimenter and was recorded together with the hand position. Based on these data points, the desired hand direction, i.e., the direct path connecting hand and target, was estimated. The resulting vector was used to train and evaluate all the



FIGURE 1

Screenshot from the virtual environment. The patient was asked to reach the blue sphere with his right hand.

models. For evaluation, we used cosine similarity (CS), measuring the cosine of the angle between two vectors, which is equal to 1 for perfect prediction, 0 for orthogonal vectors, and  $-1$  in case of opposite direction. For cosine similarity equal to 1, the target reach would be as fast as possible (direct path) according to exoskeleton parameters. Unfortunately, we could not use the time to reach the target as evaluation criteria directly. It may better reflect the target reach performance, however, it would require online experiments that we could not perform due to experimental constraints.

During the experimental sessions, 1 out of 5 states (idle state, left hand translation, right hand translation, left wrist rotation, right wrist rotation) was decoded from the recorded ECoG signal with a multilinear gate model. Accordingly to the gate predictions, an appropriate multilinear expert was selected to provide a trajectory of hand movement or direction of wrist rotation. For further analysis, we selected only left and right hand translation datasets.

Multilinear model parameters were optimized online during the recordings using recursive exponentially weighted n-way partial least squares (REW-NPLS) (Eliseyev et al., 2017). Models were trained on the first six sessions, further referred to as the calibration dataset. For the next 37 sessions, models’ weights were fixed and used for the performance evaluation. In our computational experiments, we concatenated calibration and test sessions to perform offline simulations in different scenarios, studying the dataset and model characteristics in-depth.

### 2.3. Preprocessing and feature extraction

ECoG signal was referenced to 5 electrodes (on the edge of recording grids) and on-chip filtered with a high pass (0.5 Hz) and low pass (300 Hz) analog filters, followed by a digital low pass FIR filter with a cutoff frequency of 292.8 Hz (Moly et al., 2022). After the recording, no additional reference or filtering methods were applied. Raw ECoG signal was processed with feature extraction pipeline creating time-frequency representation that is a popular way of describing motor imagery brain signals (Lotte et al., 2018). Continuous complex wavelet transform was used with 15 Morlet wavelets with central frequencies in the range of 10–150 Hz (10 Hz interval). Wavelet parameters were selected according to previous

analysis (Chao et al., 2010; Chen et al., 2013; Benabid et al., 2019) to match the informative frequency range of ECoG signal. Every 100 ms, 1 s of signal (90% overlap) was selected and convolved with the set of wavelets coefficients. Then modulus of the convolved complex signal was averaged over 0.1 s fragments. Finally, every  $i$ -th window of the signal was represented with time-frequency representation in the form of a tensor  $\mathbf{X}_i \in \mathbb{R}^{64 \times 15 \times 10}$  with dimensions corresponding to ECoG channels, frequency bands, and time steps.

In this study, samples for which predicted and desired states did not match each other were removed. By removing the gate errors, we minimize the influence of low gate model performance on the visual feedback and thus on the patient imagination patterns. In addition, one session was removed from the dataset as during the online experiment patient reached a highly negative cosine similarity (outliers compared to other sessions) which may as well influence recorded signals by providing erroneous visual feedback to the patient.

## 2.4. UMAP embeddings and artifacts identification

High-dimensional datasets are almost not possible to visualize without any dimensionality reduction before. What can be trivial to observe in low-dimensional space may easily stay hidden in the noise in high-dimensional representations. Due to the curse of dimensionality, understanding the topology of distributions or even noticing outliers is challenging. The main goal of the visualization was to see the evolution of data distributions between sessions. To map time-frequency representation into lower-dimensional space, an unsupervised learning algorithm, namely Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018) was used. We decided to apply UMAP as it preserves the global manifold structure similarly to t-SNE (Kobak and Linderman, 2021) but has a lower computational load according to McInnes et al. (2018) and UMAP (2018). Thanks to that, we could avoid additional dimensionality reduction (e.g., PCA), which is usually done before feeding high-dimensional datasets into t-SNE (van der Maaten and Hinton, 2008). We used flattened time-frequency features  $\mathbf{X}_i \in \mathbb{R}^{64 \times 15 \times 10} \rightarrow \mathbb{R}^{9600}$  (the same as for motor imagery decoding) as the input to UMAP. Every tenth observation from the dataset was selected for UMAP to avoid redundancy in the data (90% overlap between samples) and decrease the computational load. UMAP was fitted on three datasets to all the sessions together, i.e., one UMAP for both hands optimized together and one per hand trained individually. The first scenario lets us better see the data distributions within the state classification framework, with samples being colored due to the state they belong to. This gave us a global overview of the dataset. In the per hand scenario, we focused more locally on the structure of each dataset. This may have a bigger influence on the decoding performance while being harder to analyze due to the lack of explicit labels for visualization (like states in the previous case).

In the case of UMAP optimized together for both hands, we proposed an indirect indicator of data quality reflecting the separability of the left and right hand clusters. This was assessed

using a linear support-vector machine (SVM). SVM was fitted to every session separately. Then every sample in the session was classified into two categories, i.e., left hand or right hand movement. The accuracy of the state classification was further used as a state separability indicator. We did not perform any cross-validation as we focused on the separability of the clusters and not on the state classification performance itself. On the UMAP embeddings, we also visualized the SVM decision boundary dividing the space between categories of movements.

UMAP as a dataset visualization method may also be used for an overall sanity check of the dataset, especially for artifacts that are easy to spot when the dataset is small, but it is impossible to review every sample individually when analyzing thousands of observations. In our case, UMAP helped us to observe artifacts coming from connection loss resulting in singular outliers samples that were not caught during recording. Those, on the UMAP plots, created suspicious clusters of observations (Figure 2). The clusters of artifacts after recognition on the UMAP plots and further manual review were fixed by interpolation of points in the raw signal domain.

## 2.5. Evaluated models

A multilinear model optimized with REW-NPLS algorithm (Eliseyev et al., 2017) was used as a “traditional” ML benchmark to predict 3D hand translation. The same algorithm was also used for providing online control to the patient during recordings. PLS models embed both high-dimensional input features and output variables into lower-dimensional latent space, aiming to extract latent variables with the highest correlation between input and output. REW-NPLS model can be updated online thanks to low-computational cost, recursive validation of the number of latent factors, and model parameters being updated with only chunks of the dataset. Online training eases performing the experiments and makes it possible to use ECoG decoders almost from the beginning of the first recording session. Even if decoders may show unstable performance at the beginning of the experiment due to the small amount of data used for training, it provides visual feedback to the patient. For our offline computational experiments, multilinear models were trained in pseudo-online mode, simulating real-life experiments with updates based on 15 s-long chunks of data.

The second group of models used deep learning to predict the desired hand translation. In particular, methods proposed and described in detail in Śliwowski et al. (2022) were evaluated—i.e., multilayer perceptron (MLP—simple approach) and mix of CNN and LSTM [with multiple trajectories (MT) modification, explained in detail in Śliwowski et al., 2022] (CNN+LSTM+MT) providing the best performance for a given dataset (Śliwowski et al., 2022). MLP was built from two fully-connected layers with 50 neurons with dropout and batch normalization in-between (see Table 1). CNN-based method exploited the spatial correlation between electrodes by analyzing data organized on a grid reflecting the electrodes’ arrangement with convolutional layers. As the CNN-based method utilizes data structure, it has fewer parameters while maintaining similar capabilities to MLP. In CNN+LSTM+MT, LSTMs were used to aggregate temporal information extracted

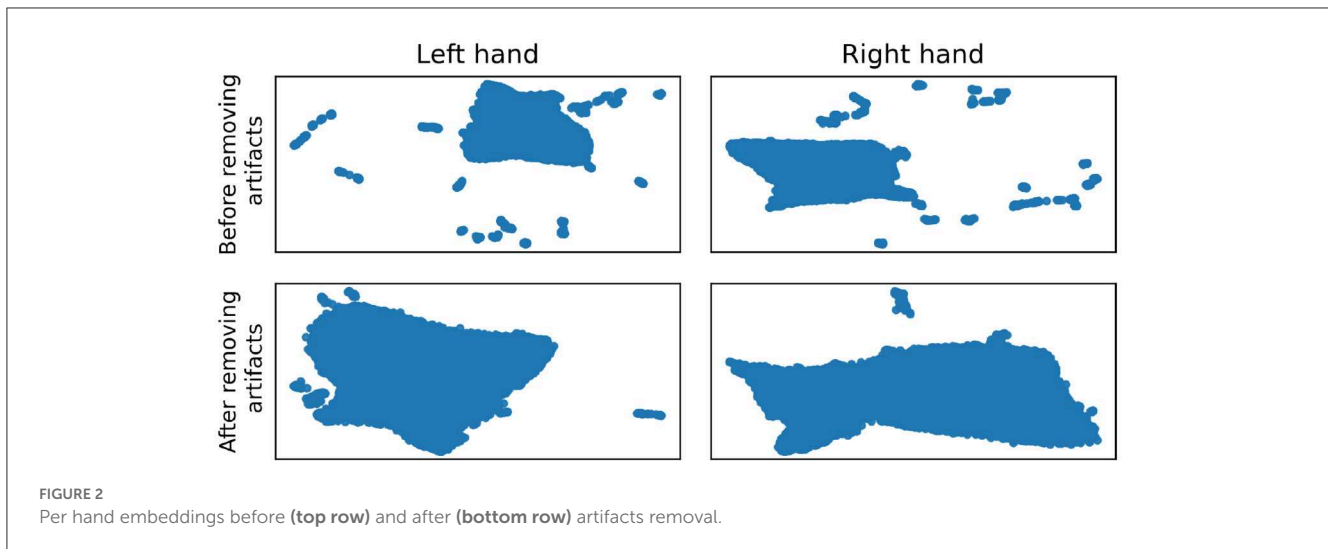


TABLE 1 MLP architecture from Śliwowski et al. (2022).

| Layer           | Kernel shape | Output shape |
|-----------------|--------------|--------------|
| Flatten         | -            | [200, 9,600] |
| Fully connected | [9,600, 50]  | [200, 50]    |
| BatchNorm       | [50]         | [200, 50]    |
| ReLU            | -            | [200, 50]    |
| Dropout         | -            | [200, 50]    |
| Fully connected | [50, 50]     | [200, 50]    |
| BatchNorm       | [50]         | [200, 50]    |
| ReLU            | -            | [200, 50]    |
| Dropout         | -            | [200, 50]    |
| Fully connected | [50, 3]      | [200, 3]     |

TABLE 2 CNN+LSTM+MT architecture from Śliwowski et al. (2022).

| Layer             | Kernel shape      | Output shape        |
|-------------------|-------------------|---------------------|
| Input             |                   | [200, 15, 8, 8, 10] |
| Input per implant |                   | [200, 15, 8, 4, 10] |
| Conv space        | [15, 32, 3, 3, 1] | [200, 32, 6, 4, 10] |
| ReLU              | -                 | [200, 32, 6, 4, 10] |
| BatchNorm         | [32]              | [200, 32, 6, 4, 10] |
| Dropout           | -                 | [200, 32, 6, 4, 10] |
| Conv space        | [32, 64, 3, 3, 1] | [200, 64, 4, 2, 10] |
| ReLU              | -                 | [200, 64, 4, 2, 10] |
| Dropout           | -                 | [200, 64, 4, 2, 10] |
| LSTM              | -                 | [200, 10, 50]       |
| LSTM              | -                 | [200, 10, 3]        |

by convolutional layers into desired translation trajectory (see Table 2). The DL models were trained to maximize cosine similarity (CS) between predicted and optimal trajectories. We used early stopping to limit the overfitting with a validation dataset consisting of the last 10% of the calibration dataset. The best model on the validation dataset was used for further evaluations. The procedure was repeated five times for every scenario and model to limit the influence of the stochasticity of the training process on our results. To train DL models, we used a fixed set of hyperparameters, i.e., learning rate equals 0.001, weight decay (L2 regularization) equals 0.01, and batch size equals 200.

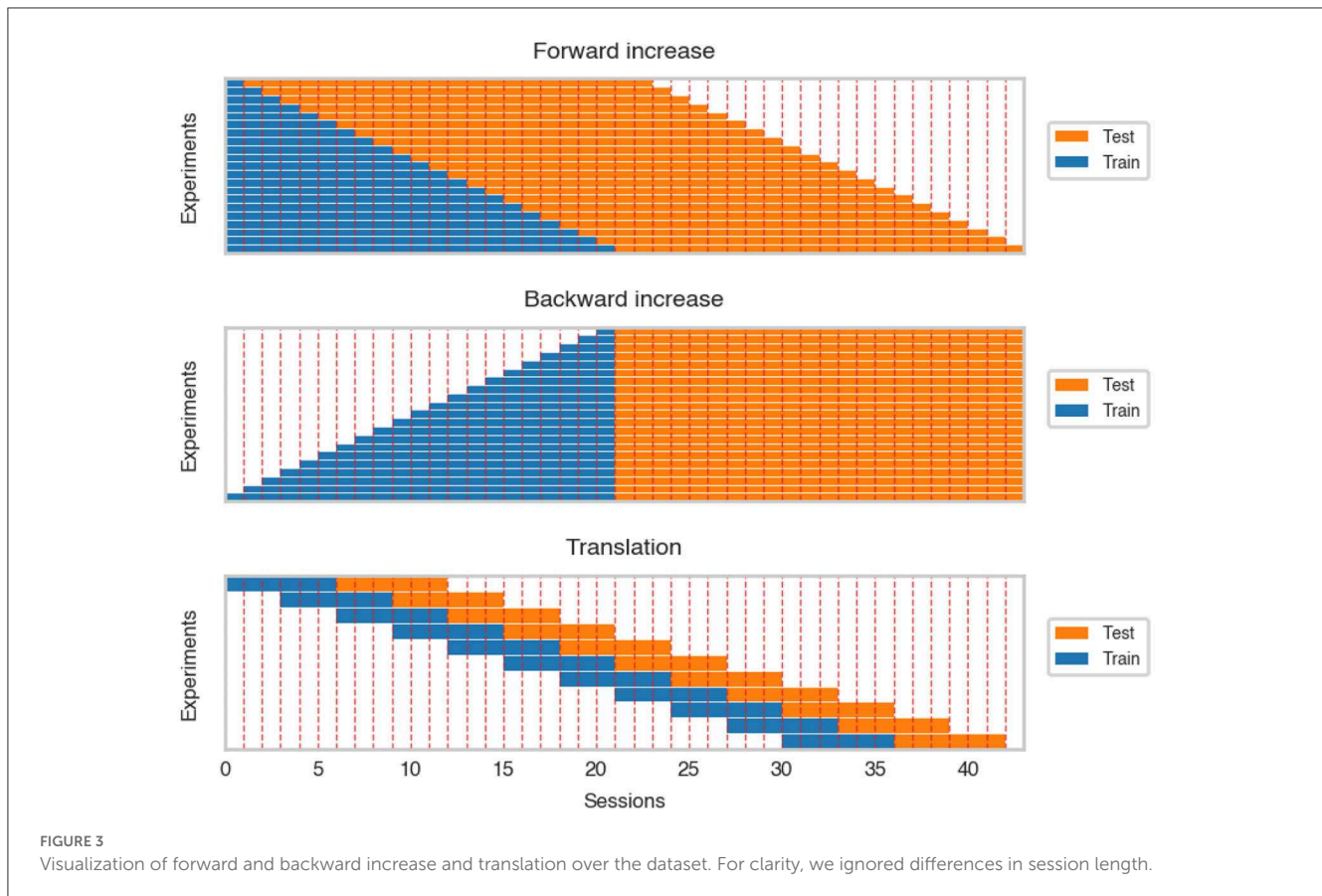
## 2.6. Computational experiments

Multiple offline computational experiments were performed on the prerecorded ECoG BCI dataset to assess the impact of training dataset size on decoding performance. The results computed on a real-life dataset may be impacted by multiple factors that cannot be observed directly. Thus, we

proposed several ways of increasing the training dataset as well as iterating over it. By modifying the training datasets in different manners, we aimed to isolate different factors that can potentially influence learning curves. In every scenario, all the models were trained on a different subset of the database and then evaluated on test datasets accordingly to the experiment.

### 2.6.1. Forward increase

The forward increase (FI) scenario measured the change of cosine similarity when adding more recording sessions to the dataset. This experiment corresponds to a real-life situation where more data is collected during the experiment. The sessions were incrementally added (session by session) to the training dataset. After every step, all the decoders were trained from scratch and evaluated on the following 22 sessions (see Figure 3).



### 2.6.2. Backward increase

An important factor influencing model training may be the nonstationarity of signal in time originating from the plasticity of the brain as well as the patient's adaptation. To assess the influence of this factor, an inverse of forward increase was performed, further referred to as backward increase (BI). Similar to the FI simulation, the training dataset was increased session by session. However, the increase was started from the 21st session and the previous sessions were added until including the first calibration session. After every training, models were evaluated on a fixed test set consisting of 22 last recordings (see Figure 3).

### 2.6.3. Random increase

An alternative way of assessing the influence of training dataset size on the decoder performance is random dataset increase (RI). Instead of maintaining the temporal order of recorded samples, we artificially removed the connection between neighboring observations, i.e., for every dataset size, a respective number of observations was selected from the first 22 sessions, and then the model was trained. This may reduce the effects of neuronal signal nonstationarity and/or patient adaptation and provide results closer to theoretical learning curves when assumptions about the stationarity of observations are fulfilled. Evaluations were performed on the same test set as in BI.

### 2.6.4. Dataset translation

As data may change over time, we trained models on approximately the same amount of data but recorded in different periods of the experiment. This enabled us to rule out the effect of the increased dataset and focus on data shift and potential patient adaptation that may modify the data representation and influence the performance of trained decoders. The training dataset was translated over the whole dataset and evaluated on the test dataset consisting of the following six sessions (see Figure 3).

## 2.7. Learning curve

The learning curve<sup>1</sup> describes the relationship between model performance and the training dataset size  $l$  (Cortes et al., 1993). It can be used, for example, to infer a potential change in the performance from adding more data to the system. This can be particularly efficient in application to BCI because we can estimate the hypothetical performance of decoders when recording more data without actually performing the experiments. Learning curves can also be used to select an appropriate model for a specific dataset size. For example, Strang et al. (2018) showed that non-linear models are more likely to outperform linear models for

<sup>1</sup> In this context, the learning curve does not refer to the relationship between the number of training epochs and model performance which the name learning curve is also commonly used for.

bigger datasets. On the other hand, Hoiem et al. (2021) showed that models with more parameters can be more efficient in the case of small datasets despite the higher potential for overfitting.

The learning curve may be formulated with power law (Cortes et al., 1993; Gu et al., 2001). In our case, the relationship between cosine similarity and training dataset size  $l$  may be expressed as:

$$CS(l; a, b, c) = a - b \cdot l^{-c} \quad (1)$$

Where  $b$  and  $c$  can be interpreted as learning rate and decay rate, respectively.  $a$  corresponds to theoretical asymptotic performance when  $l \rightarrow \infty$ . Parameters  $a$ ,  $b$ , and  $c$  were fitted to the results obtained in RI experiment with non-linear least squares using Trust Region Reflective algorithm with bounds  $a \in [-1, 1]$ ,  $b > 0$ , and  $c > 0$ .

## 2.8. Intrinsic dimensionality estimation

The idea of patient adaptation and improving BCI skills using visual feedback is based on the assumption that the patient can modify/adjust motor imagery patterns to solve the task better. As a result, the data distribution and the shape of the data manifold may change. To estimate the data distribution changes, intrinsic dimensionality (ID) estimation methods may be used. ID reflects the minimum number of variables needed to represent the dataset without a significant information loss. Thus, the ID indicator is strictly connected to a dataset's true dimensionality, which is an important factor in data analysis, influencing the performance and changing the number of samples needed to train models. Intuitively, in a typical case, higher-dimensional manifolds are harder to learn due to the 'curse of dimensionality.' ID is better studied for images that, although have thousands of pixels, lie on a lower-dimensional manifold (e.g., less than 50 for ImageNet Pope et al., 2021). We use ID as a potential data quality indicator, which may vary with different recording sessions. ID estimates were computed for every session, and values from the respective sessions were averaged to obtain training dataset estimates for the dataset translation experiment. To compute ID, we used current state-of-the-art methods, namely expected simplex skewness (ESS) (Johnsson et al., 2015) estimating local ID in data neighborhoods (in our case 100 points) and TwoNN (Facco et al., 2017) estimating global dataset ID. ESS, according to Tempczyk et al. (2022) provides better estimates for high ID values, while most of the methods tend to underestimate the ID (e.g., TwoNN Facco et al., 2017). It is especially important because our preliminary analysis showed that ECoG data is high dimensional, with ECoG features' mean local ID being significantly higher than the mean local ID for images (around 300 for ECoG, below 15 for MNIST, EMNIST, and FMNIST Bac and Zinovyev, 2020). For ID computations we used scikit-dimensions package (Bac et al., 2021).

## 3. Results

### 3.1. UMAP

Data distributions for every session were shown in Figure 4 with colors indicating left and right hand states. With time, clusters

of states get better separated from each other. We quantified the separability of different states with SVM classification accuracy (Figure 5). A statistically significant (slope = 0.175, intercept = 83.37,  $R = 0.539$ ,  $p = 0.0002$ ) increase in accuracy can be observed for sessions recorded later in the experiment, with a maximum accuracy of 95% for session 37. Note that UMAP, similarly to t-SNE, does not preserve the density of points when mapping to the lower dimensional space and may, in some cases, create sub-clusters that originally may not exist in the input space.

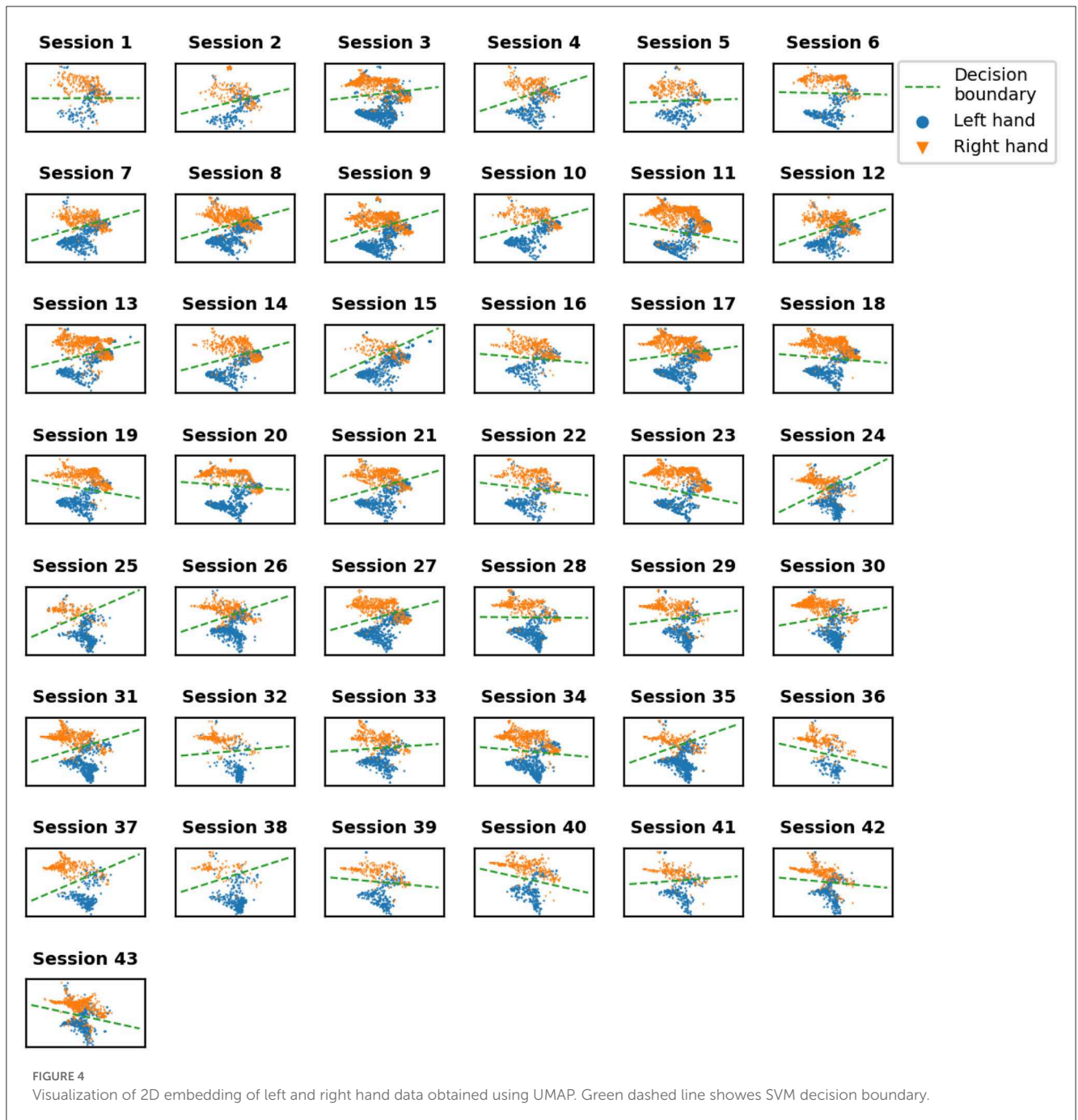
### 3.2. Forward and backward increase

Forward increase results (Figure 6) show learning curves in a situation close to a real-life scenario when more recordings are performed in the experiment. For all the models, a sharp increase in performance can be observed for small datasets. After 30–40 min of data, the curves become flat, reaching 70–80% of maximum FI performance (except 100% for the multilinear right-hand model) until 100–120 min of the signal. For datasets with more data than 100–120 min, a slow performance increase can be noticed. In the case of the left hand dataset, it starts earlier and is also visible for the multilinear model, while for the right hand, REW-NPLS performance stays stable. Overall, multilinear and DL models have similar learning curves and reach a performance plateau after including the same amount of data. However, multilinear models usually perform worse than DL models for the same amount of data.

Extending the dataset backward, starting from the middle of the recorded dataset, does not correspond to any real-life scenario. However, by doing this, we were able to assess the potential influence of data quality change on the results computed in the FI computational experiment. In the case of backward increase (Figure 7), high performance can be observed for relatively small datasets—with just 3 (left hand) and 2 (right hand) sessions. For bigger datasets, the performance stabilizes or decreases slightly. The curves for all the models behave similarly. Performance of DL models starts to increase for >130 min of signal for the right hand and achieves the best cosine similarity. When comparing FI and BI, in the case of the left hand, the best performance can be observed for BI and only 3 sessions of data in the training dataset. In the case of the right hand, the highest performance is achieved for the biggest dataset, suggesting that recording more data may improve the cosine similarity. The small amount of data needed to achieve high performance (2–3 sessions) in the BI experiment may suggest brain activity improvement resulting in dataset quality increase (the amount of data required to reach a given performance).

### 3.3. Random increase

In the RI experiment, the influence of patient adaptation and signal nonstationarity is reduced as all the links between neighboring samples are destroyed when selecting data for the training dataset. Results for RI are more similar to theoretical learning curves of DL models, with a sharp increase in performance in the beginning and saturation when the model's maximum



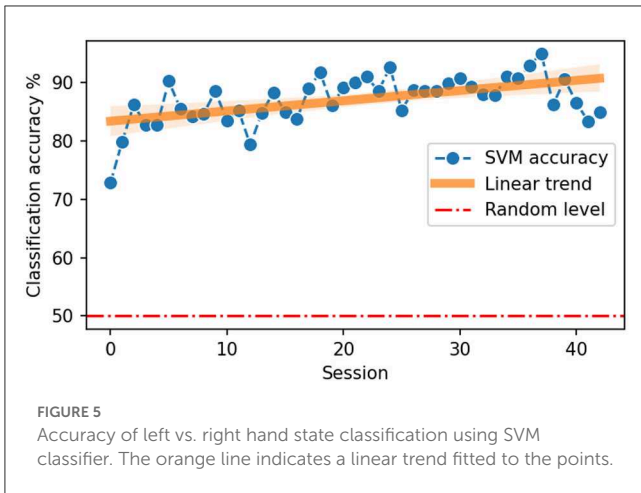
capacity is achieved. The performance is saturated after adding approximately 60–90 min of data to the training dataset at 95% of maximum cosine similarity for RI experiment. Only a small improvement can be observed from using more data. For the multilinear model, we can observe that saturated best performance is lower than in the case of DL models. DL methods are able to learn more complex functions and thus can reach higher performance. Fitted learning curves show the relationship between cosine similarity and dataset size within a theoretical framework, emphasizing the bigger capabilities of DL methods. The best models trained in the RI experiment showed lower performance compared to the best models from

other experiments (dataset translation for both hands and BI for the left hand). However, in every experiment except BI and RI, models were evaluated on different test datasets (see Figure 3).

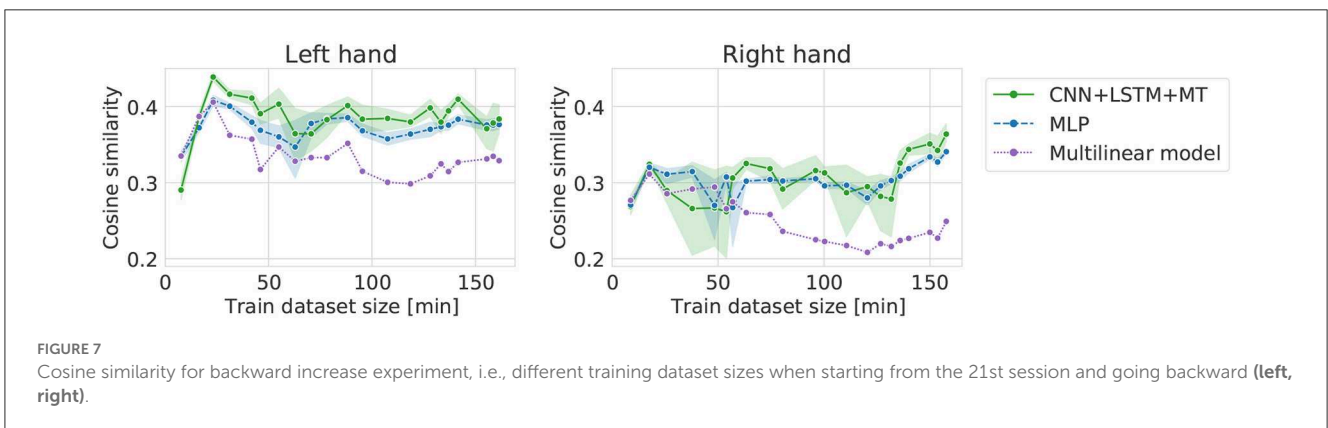
### 3.4. Dataset translation

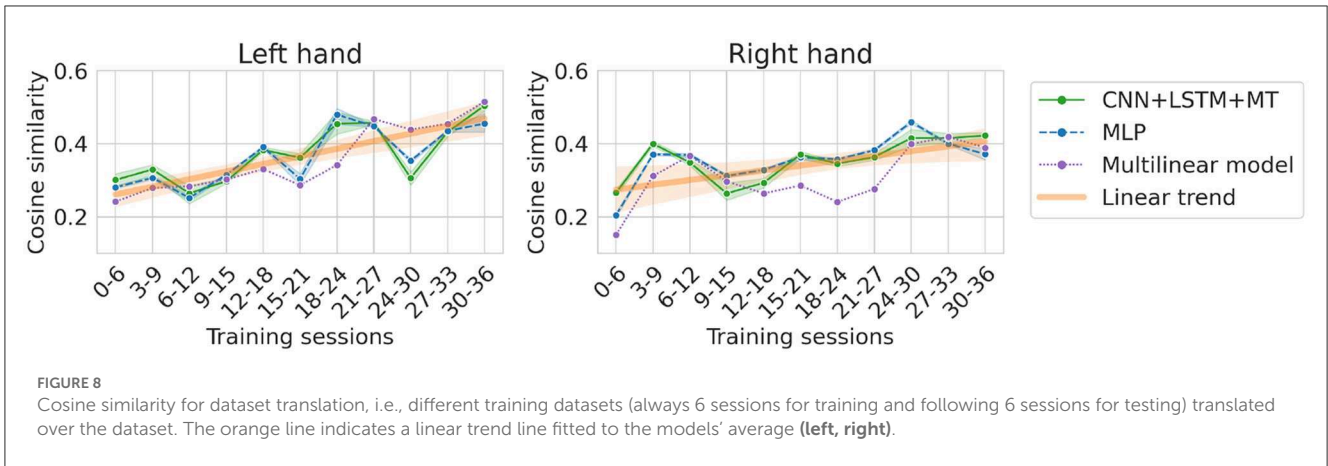
The dataset translation experiment shows the change in performance while maintaining approximately the same amount of data (six sessions) in the training dataset. Generally, all



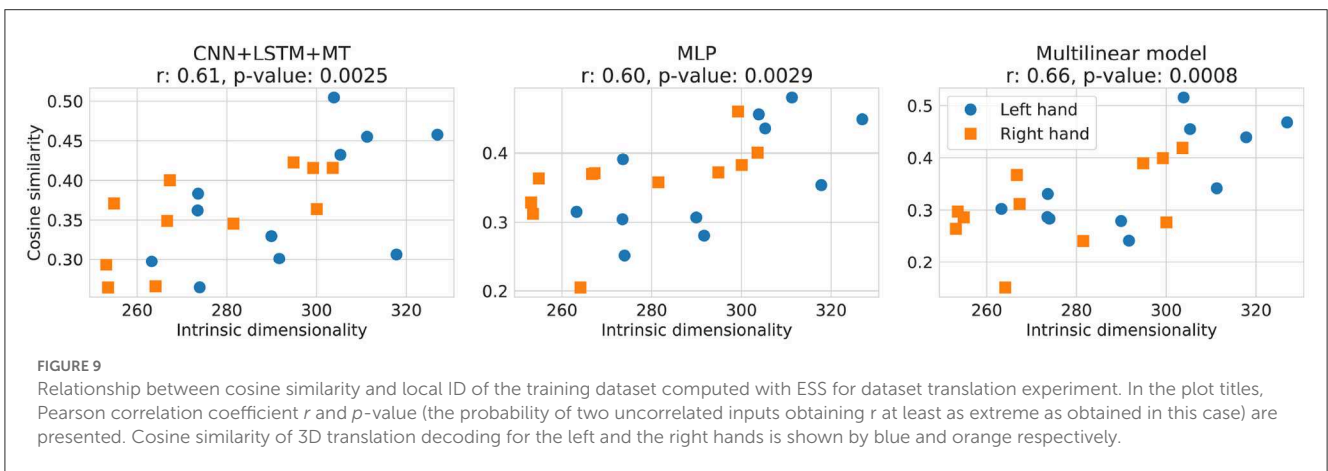


models show similar trends (see Figure 8). For the left hand, we can observe an increase in cosine similarity for datasets recorded later in the experiment suggesting an improvement in data quality. The increase is less visible for the right hand dataset. This is confirmed by the slope of the linear trend fitted to the average performance of all the models (Table 3). Expected cosine similarity improvement from training a model on the dataset recorded later was equal to 0.0069 per session and 0.0044 per session for left and right hand datasets, respectively. For both datasets, the most significant performance increase between the first and last evaluation can be observed for the multilinear model (Table 4). It may suggest that the patient, to some extent, adapted specifically to the linear model family. The multilinear model does not follow the same fluctuations as the DL methods. The difference could be caused by the way of validating models (10% validation set for DL,





**FIGURE 8** Cosine similarity for dataset translation, i.e., different training datasets (always 6 sessions for training and following 6 sessions for testing) translated over the dataset. The orange line indicates a linear trend line fitted to the models' average (left, right).



**FIGURE 9** Relationship between cosine similarity and local ID of the training dataset computed with ESS for dataset translation experiment. In the plot titles, Pearson correlation coefficient  $r$  and  $p$ -value (the probability of two uncorrelated inputs obtaining  $r$  at least as extreme as obtained in this case) are presented. Cosine similarity of 3D translation decoding for the left and the right hands is shown by blue and orange respectively.

recursive validation on the last 15 s of data at every step for pseudo-online REW-NPLS).

In **Figure 9**, the relationship between the local ID of the training dataset computed with ESS and the cosine similarity of different models for the translation experiment is presented. A statistically significant ( $\alpha < 0.05$ ) correlation between local ID and models' performance was observed for all the methods, reaching up to 0.66 of the  $r$  correlation coefficient for the multilinear model. An overall trend of achieving higher cosine similarity can be observed for training datasets with a higher ID. ID for the analyzed datasets varies between 250 and 330, which is much more compared to  $<15$  reported for MNIST, EMNIST, and FMNIST (Bac and Zinovyev, 2020).

### 4. Discussion

Our results showed that DL-based methods provide similar or higher performance in almost all cases, enabling achieving higher performance than the multilinear model while using the same amount of data. Given the limited evaluation possibilities, including more data in the training dataset for this patient and task may not be immediately visible on the performance metrics if already having access to 40 min of the signal. Indeed, a drastic increase in performance can be noticed for datasets smaller

**TABLE 3** Parameters of trend lines fitted to the dataset translation results.

|            | Slope  | Intercept | R      | $p$ -value |
|------------|--------|-----------|--------|------------|
| Left hand  | 0.0069 | 0.2612    | 0.8816 | 0.0003*    |
| Right hand | 0.0044 | 0.2744    | 0.6999 | 0.0165*    |

Statistically significant  $p$ -values for the correlation coefficient are marked with asterisks.

**TABLE 4** Differences between models trained on sessions 0–6 and 30–36 in the dataset translation experiment.

|                   | Left hand | Right hand |
|-------------------|-----------|------------|
| CNN+LSTM+MT       | 0.203     | 0.157      |
| MLP               | 0.175     | 0.167      |
| Multilinear model | 0.274     | 0.239      |

than 40 min. This justifies the current experimental paradigms in which 40–50 min of the signal is collected (corresponding to achieving approximately 70–80% of maximum performance achieved with datasets up to 160 min of data) for training 3D hand translation models.

Theoretically, models with bigger capacity can benefit stronger from having access to more data. One of the indicators of model capacity can be the number of trainable parameters. In our

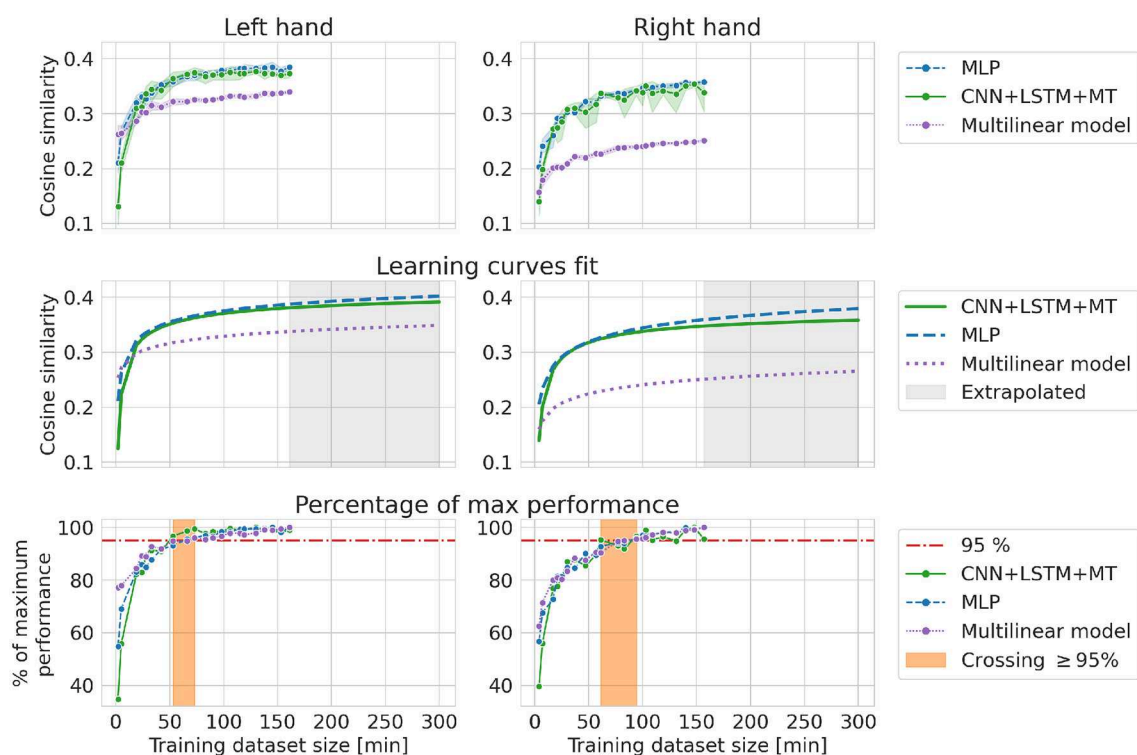


FIGURE 10

Cosine similarity for random increase experiment, i.e., different training dataset sizes when randomly selecting a subset of observations from the first 22 sessions. Every evaluation was performed 10 times (left, right).

case, MLP had the biggest number of trainable parameters (482 953), followed by CNN+LSTM+MT (238 772) and the multilinear model (28 800). The difference in potential performance gains can be visible in Figure 10. For small datasets, a multilinear model outperforms DL-based approaches (Figure 10 left) or provides approximately the same cosine similarity. However, the multilinear model saturates at a lower level of cosine similarity, resulting in a performance gap that could be explained by the difference in model capacity. Multilinear models are more likely to provide high performance compared to DL for small datasets, which is consistent with the ML theory of less complex functions being less prone to overfitting. RI results and fitted theoretical learning curves revealed the models' characteristics while limiting the influence of other factors like distribution shifts or patient adaptation on the decoding performance. Finally, all models saturate for relatively small training datasets (50–90 min for RI, 50 min for FI, 30 min for BI) with only slight improvement from adding more data (~ 5%). This amount of data is similar to the usual amount of data used in BCI studies.

While this result validates previously developed data processing and experimental pipelines, a question arises whether it is an actual property/characteristic of brain signals or the shape of the curve is influenced by the previous years of research in which a relatively small amount of data was usually used to develop pipelines. There are hundreds of hyperparameters influencing data processing characteristics, starting from recording devices (e.g., number of electrodes, mental task design), signal processing pipelines (e.g., electrodes montage, filtering, standardization),

ending on hyperparameters of machine learning models of all kinds (e.g., models' capacity, regularization weight, the architecture of models). The lack of huge improvement from increasing the dataset may be caused just because we reached the level of decoding close to maximum due to a lack of information in the data needed for prediction. However, from another perspective, one can hypothesize that the observed lack of huge improvement from increasing the dataset is an effect of researchers overfitting to the specific conditions observed so far, especially years of analysis of small datasets.

#### 4.1. Models optimization for big datasets

All offline experiments were performed with a fixed set of hyperparameters. At the same time, different dataset sizes may require a change in the hyperparameters. For example, regularization limits overfitting, which should be less severe when the training dataset is big. Similar logic applies to dropout, which limits overfitting but on the other hand, it decreases models' capacity by introducing redundancy in the network representation. In the BI experiment, we observed a decrease in performance when adding more data for the left hand dataset. Hypothetically, increasing models' capacity may solve this problem (assuming it is caused by adding samples from different distributions to the dataset) because models with bigger capacity might not have to "choose" on which motor imagery patterns they should focus.

However, hyperparameters search is time and resource-consuming, so performing hyperparameters search for every dataset size may not be reasonable. In the future, DL architectures with bigger capacities in terms of the number of layers, number of neurons, etc., should be evaluated.

Datasets can also be artificially extended by using data augmentation methods. A variety of beneficial data augmentation methods exist for brain signals, especially EEG (Rommel et al., 2022b), that might improve decoding accuracy for the 3D hand movement control. Hoiem et al. (2021) showed, for computer vision datasets, that data augmentation may act as a multiplier of the number of examples used for training. In the light of recent advancements in EEG data augmentation, i.e., class-wise automatic differentiable data augmentation (Rommel et al., 2022a), it can be interesting to investigate how the reported results generalize to ECoG signals and influence, presented here, learning curves.

## 4.2. Patient training

UMAP embeddings may reveal interesting data manifold structures. In our case, we observed signs of distribution change on the embedding visualization and the separability of left/right hand observations. Points start to be distributed denser in some regions of the plots and align along lines (see for examples sessions 35, 42, 43 for the left hand in [Supplementary Figure 1](#) or sessions 31, 41, 42 in [Supplementary Figure 2](#)). Additionally, in the dataset translation experiment, we can see an increase in cosine similarity, stronger for the left hand. Moreover, the overall best performance for the left hand was achieved with only 3 sessions (~25 min of signal), outperforming models trained on much bigger datasets. This suggests improvements in patient BCI skills by adapting motor imagery patterns to the ML pipeline used in the study but non-specific to the multilinear model because trends are visible for all the evaluated approaches. At the same time, adding more data with noisy and changing patterns may not be profitable for the predictions. Thus, more focus should be placed on obtaining high-quality and well-separable motor imagery patterns in the signal. Patient adaptation is possible thanks to the visual feedback provided to the participant during recordings. The potential for patient adaptation creates a perspective for further improvements of BCI performance with the experience gained by the patient in long-term usage of the system. However, the reason adaptation is visible only for the left hand remains unknown. We hypothesize that the motor imagery patterns are easier to adapt for the left hand thanks to the remaining residual control resulting in better cortex preservation. Differences between hands in residual control level can also affect the shape of presented learning curves and be the reason why we observed significantly similar but distinct characteristics. It is only a hypothesis that would require extended experiments and analysis.

Our results showed a correlation between the local ID of the training dataset and the models' performance. This may indicate that models achieve better results when trained on more complicated manifolds. However, this hypothesis is counterintuitive and contradictory to research in computer vision. Thus, we hypothesize that higher ID may also indicate more diverse motor imagery patterns, better representing those found in the

test set. Diversity of patterns may be harmful to models with a too-small capacity to learn them all. However, to some extent, it may be helpful as it creates a more diverse dataset that better reflects/covers the real manifold of all motor imagery patterns. Finally, we hypothesize that another hidden factor affects both the local ID and the amount of information needed for prediction, like the diversity of motor imagery patterns, so a change in local ID may not cause an increase in the performance itself. For example, local ID can also be increased by adding Gaussian noise to the signal, decreasing cosine similarity instead. Investigation of this kind of relationship is especially challenging in the case of brain signals due to a lack of data understanding with "the naked eye," which would significantly ease finding a correct interpretation of observed phenomena. As a next step, more ID estimation methods could be evaluated as statistically significant correlations for DL models were observed only for local ID computed with ESS. In the case of TwoNN, global ID did not show a significant correlation for DL approaches (see [Supplementary Figure 3](#)). This could be caused by worse TwoNN precision for high ID values as well as a lack of local per-sample ID information in the global ID dataset estimate. The relationship between local ID and performance should be further analyzed on different brain signal datasets.

## 4.3. Interpretation limitations

All the computational experiments analyzed in this study were obtained offline using data recorded with only one patient. Thus, the learning curves and potential of patient adaptation should be further investigated in a bigger population with online experiments verifying our conclusions. Specifically, an online experimental protocol aiming to isolate patient training (with or without visual feedback) and the decoder's update influence on performance should be designed. Despite the limitations of the analysis, our study could be a reference point for future work. It may also influence experimental paradigm design and model selection, especially considering difficult access to datasets allowing for this kind of analysis.

Our results were computed on a real-life dataset recorded with a tetraplegic patient. Analyzing this kind of dataset allows us to draw conclusions about the population in real need of assistive technology. However, interpretation of results is even more challenging than in the case of healthy subjects because we do not have access to solid ground-truth labels to train ML models. This increases the already long list of factors that can affect the performance of the decoders and may not be easily noticed when analyzing ECoG signals. For example, in the ideal ML world, one could analyze the learning curve and draw conclusions about the required dataset size to effectively train ML models. In our case, other factors like the nonstationarity of the signal play an important role in the process. In some cases, we may add more data to the dataset (e.g., BI experiment) and decrease the performance. Part of the aforementioned issues limiting our interpretation capabilities might be addressed with generative models (Goodfellow et al., 2014) that are a popular tool in computer vision. In the case of brain signals, the ability to produce signals with the given parameters and characteristics may be used to verify and understand phenomena observed in real-life experiments. First attempts to train GANs

for EEG (Hartmann et al., 2018) data analysis were made, but a significant amount of work still has to be done to create a consistent framework for easier hypothesis evaluation.

## 4.4. Conclusions

Deep learning models performed better than a multilinear model for almost all dataset sizes without requiring extended training datasets, indicating DL models' compatibility with BCI dataset size restrictions. Furthermore, we showed the importance of patient adaptation in the human-in-the-loop system that enabled obtaining high-performance models with relatively small training datasets. Finally, we propose UMAP embeddings and local intrinsic dimensionality as a way to visualize the data and potentially evaluate data quality. While our analysis was limited to only one patient and a specific experimental paradigm, considering difficult access to clinical data and the lack of similar studies for this problem, computed results can be a reference for future ECoG BCI research.

## Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The data analyzed during the current study are not publicly available for legal/ethical reasons. Part of the dataset may be provided upon reasonable request. Requests to access these datasets should be directed to TA, [tetiana.aksenova@cea.fr](mailto:tetiana.aksenova@cea.fr).

## Ethics statement

The studies involving human participants were reviewed and approved by National Agency for the Safety of Medicines and Health Products (Agence nationale de sécurité du médicament et des produits de santé: ANSM), registration number 2015-A00650-49, and the Committee for the Protection of Individuals (Comité de Protection des Personnes—CPP), registration number 15-CHUG-19. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

MŚ and MM contributed to the conception and design of the study and were responsible for software, experiments

## References

- Ajiboye, A. B., Willett, F. R., Young, D. R., Memberg, W. D., Murphy, B. A., Miller, J. P., et al. (2017). Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: a proof-of-concept demonstration. *Lancet* 389, 1821–1830. doi: 10.1016/S0140-6736(17)30601-3
- Bac, J., Mirkes, E. M., Gorban, A. N., Tyukin, I., and Zinovyev, A. (2021). Scikit-dimension: a python package for intrinsic dimension estimation. *Entropy* 23, 1368. doi: 10.3390/e23101368
- Bac, J., and Zinovyev, A. (2020). "Local intrinsic dimensionality estimators based on concentration of measure," in *2020 International Joint Conference on Neural Networks (IJCNN)* (Glasgow), 1–8. doi: 10.1109/IJCNN48605.2020.9207096
- Bashivan, P., Rish, I., Yeasin, M., and Codella, N. (2015). Learning representations from eeg with deep recurrent-convolutional neural networks. *arXiv preprint arXiv:1511.06448*. doi: 10.48550/arXiv.1511.06448
- Benabid, A. L., Costecalde, T., Eliseyev, A., Charvet, G., Verney, A., Karakas, S., et al. (2019). An exoskeleton controlled by an epidural wireless brain-machine interface in a tetraplegic patient: a proof-of-concept demonstration. *Lancet Neurol.* 18, 1112–1122. doi: 10.1016/S1474-4422(19)30321-7
- Benaroch, C., Sadatnejad, K., Roc, A., Appriou, A., Monseigne, T., Pramij, S., et al. (2021). Long-term bci training of a tetraplegic user: adaptive riemannian classifiers and user training. *Front. Hum. Neurosci.* 15, 635653. doi: 10.3389/fnhum.2021.635653

validation, investigation, formal analysis, data curation, results visualization, and writing the initial draft and final manuscript version. TA, AS, and PB contributed to the conception and design of the study and reviewed/corrected the final manuscript, supervised the research, and participated in the investigation process. All authors contributed to the article and approved the submitted version.

## Funding

Clinatec is a Laboratory of CEA-Leti at Grenoble and has statutory links with the University Hospital of Grenoble (CHUGA) and with University Grenoble Alpes (UGA). This study was funded by CEA (recurrent funding) and the French Ministry of Health (Grant PHRC-15-15-0124), Institut Carnot, Fonds de Dotation Clinatec. MM was supported by the cross-disciplinary program on Numerical Simulation of CEA. MŚ was supported by the CEA NUMERICS program, which has received funding from European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 800945—NUMERICS—H2020-MSCA-COFUND-2017.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2023.1111645/full#supplementary-material>

- Bundy, D. T., Pahwa, M., Szrama, N., and Leuthardt, E. C. (2016). Decoding three-dimensional reaching movements using electrocorticographic signals in humans. *J. Neural Eng.* 13, 026021. doi: 10.1088/1741-2560/13/2/026021
- Carmena, J. M. (2013). Advances in neuroprosthetic learning and control. *PLoS Biol.* 11, e1001561. doi: 10.1371/journal.pbio.1001561
- Chao, Z., Nagasaka, Y., and Fujii, N. (2010). Long-term asynchronous decoding of arm motion using electrocorticographic signals in monkey. *Front. Neuroeng.* 3, 3. doi: 10.3389/fneng.2010.00003
- Chavarriga, R., Fried-Oken, M., Kleih, S., Lotte, F., and Scherer, R. (2017). Heading for new shores! overcoming pitfalls in bci design. *Brain Comput. Interfaces* 4, 60–73. doi: 10.1080/2326263X.2016.1263916
- Chen, C., Shin, D., Watanabe, H., Nakanishi, Y., Kambara, H., Yoshimura, N., et al. (2013). Prediction of hand trajectory from electrocorticography signals in primary motor cortex. *PLoS ONE* 8, e83534. doi: 10.1371/journal.pone.0083534
- Chen, W., Liu, X., and Litt, B. (2014). “Logistic-weighted regression improves decoding of finger flexion from electrocorticographic signals,” in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Chicago, IL: IEEE)*, 2629–2632.
- Collinger, J. L., Wodlinger, B., Downey, J. E., Wang, W., Tyler-Kabara, E. C., Weber, D. J., et al. (2013). High-performance neuroprosthetic control by an individual with tetraplegia. *Lancet* 381, 557–564. doi: 10.1016/S0140-6736(12)61816-9
- Cortes, C., Jackel, L. D., Solla, S., Vapnik, V., and Denker, J. (1993). “Learning curves: asymptotic values and rate of convergence,” in *Advances in Neural Information Processing Systems, Vol. 6*, eds J. Cowan, G. Tesauro, and J. Alsppector (Denver, CO: Morgan-Kaufmann).
- Cunningham, J. P., Nuyujukian, P., Gilja, V., Chestek, C. A., Ryu, S. I., and Shenoy, K. V. (2011). A closed-loop human simulator for investigating the role of feedback control in brain-machine interfaces. *J. Neurophysiol.* 105, 1932–1949. doi: 10.1152/jn.00503.2010
- Dayan, E., and Cohen, L. G. (2011). Neuroplasticity subserving motor skill learning. *Neuron* 72, 443–454. doi: 10.1016/j.neuron.2011.10.008
- Du, A., Yang, S., Liu, W., and Huang, H. (2018). “Decoding ECoG signal with deep learning model based on LSTM,” in *TENCON 2018-2018 IEEE Region 10 Conference (Jeju: IEEE)*, 0430–0435.
- Edelman, B. J., Meng, J., Suma, D., Zurn, C., Nagarajan, E., Baxter, B., et al. (2019). Noninvasive neuroimaging enhances continuous neural tracking for robotic device control. *Sci. Robot.* 4, 31. doi: 10.1126/scirobotics.aaw6844
- Elango, V., Patel, A. N., Miller, K. J., and Gilja, V. (2017). Sequence transfer learning for neural decoding. *bioRxiv*, 210732. doi: 10.1101/210732
- Eliseyev, A., Auboiroux, V., Costecalde, T., Langar, L., Charvet, G., Mestais, C., et al. (2017). Recursive exponentially weighted n-way partial least squares regression with recursive-validation of hyper-parameters in brain-computer interface applications. *Sci. Rep.* 7, 16281. doi: 10.1038/s41598-017-16579-9
- Facco, E., d’Errico, M., Rodriguez, A., and Laio, A. (2017). Estimating the intrinsic dimension of datasets by a minimal neighborhood information. *Sci. Rep.* 7, 11873. doi: 10.1038/s41598-017-11873-y
- Flamary, R., and Rakotomamonjy, A. (2012). Decoding finger movements from ECoG signals using switching linear models. *Front. Neurosci.* 6, 29. doi: 10.3389/fnins.2012.00029
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). “Generative adversarial nets,” in *Advances in Neural Information Processing Systems, Vol. 27*, eds Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger (Montreal, QC: Curran Associates, Inc.).
- Gu, B., Hu, F., and Liu, H. (2001). “Modelling classification performance for large data sets,” in *Advances in Web-Age Information Management*, eds X. S. Wang, G. Yu, and H. Lu (Berlin; Heidelberg: Springer Berlin Heidelberg), 317–328.
- Hachem, L. D., Ahuja, C. S., and Fehlings, M. G. (2017). Assessment and management of acute spinal cord injury: from point of injury to rehabilitation. *J. Spinal Cord Med.* 40, 665–675. doi: 10.1080/10790268.2017.1329076
- Hartmann, K. G., Schirrmeyer, R. T., and Ball, T. (2018). EEG-gan: generative adversarial networks for electroencephalographic (EEG) brain signals. *arXiv:1806.01875 [eess.SP]*. doi: 10.48550/arXiv.1806.01875
- He, Y., Eguren, D., Azorin, J. M., Grossman, R. G., Luu, T. P., and Contreras-Vidal, J. L. (2018). Brain-machine interfaces for controlling lower-limb powered robotic systems. *J. Neural Eng.* 15, 021004. doi: 10.1088/1741-2552/aaa8c0
- Hochberg, L. R., Bacher, D., Jarosiewicz, B., Masse, N. Y., Simeral, J. D., Vogel, J., et al. (2012). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature* 485, 372–375. doi: 10.1038/nature11076
- Hochberg, L. R., Serruya, M. D., Friehe, G. M., Mukand, J. A., Saleh, M., Caplan, A. H., et al. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* 442, 164–171. doi: 10.1038/nature04970
- Höhne, J., Holz, E., Staiger-Sälzer, P., Müller, K.-R., Kübler, A., and Tangermann, M. (2014). Motor imagery for severely motor-impaired patients: evidence for brain-computer interfacing as superior control solution. *PLoS ONE* 9, e104854. doi: 10.1371/journal.pone.0104854
- Hoiem, D., Gupta, T., Li, Z., and Shlapentokh-Rothman, M. (2021). “Learning curves for analysis of deep networks,” in *Proceedings of the 38th International Conference on Machine Learning, Vol. 139 of Proceedings of Machine Learning Research*. p. 4287–4296.
- Holz, E. M., Höhne, J., Staiger-Sälzer, P., Tangermann, M., and Kübler, A. (2013). Brain-computer interface controlled gaming: evaluation of usability by severely motor restricted end-users. *Artif. Intell. Med.* 59, 111–120. doi: 10.1016/j.artmed.2013.08.001
- Jarosiewicz, B., Masse, N. Y., Bacher, D., Cash, S. S., Eskandar, E., Friehe, G., et al. (2013). Advantages of closed-loop calibration in intracortical brain-computer interfaces for people with tetraplegia. *J. Neural Eng.* 10, 046012. doi: 10.1088/1741-2560/10/4/046012
- Johnsson, K., Soneson, C., and Fontes, M. (2015). Low bias local intrinsic dimension estimation from expected simplex skewness. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 196–202. doi: 10.1109/TPAMI.2014.2343220
- Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., et al. (2020). Scaling laws for neural language models. *CoRR*, abs/2001.08361. doi: 10.48550/arXiv.2001.08361
- Kobak, D., and Linderman, G. C. (2021). Initialization is critical for preserving global data structure in both t-SNE and UMAP. *Nat. Biotechnol.* 39, 156–157. doi: 10.1038/s41587-020-00809-z
- Krakauer, J. W., Hadjiosif, A. M., Xu, J., Wong, A. L., and Haith, A. M. (2019). Motor learning. *Compr. Physiol.* 9, 613–663. doi: 10.1002/cphy.c170043
- Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., and Lance, B. J. (2018). Egnet: a compact convolutional neural network for EEG-based brain-computer interfaces. *J. Neural Eng.* 15, 056013. doi: 10.1088/1741-2552/aac8c8
- Leeb, R., Tonin, L., Rohm, M., Desideri, L., Carlson, T., Millan, J., et al. (2015). Towards independence: a bci telepresence robot for people with severe motor disabilities. *Proc. IEEE* 103, 969–982. doi: 10.1109/PROC.2015.2419736
- Liang, N., and Bougrain, L. (2012). Decoding finger flexion from band-specific ECoG signals in humans. *Front. Neurosci.* 6, 91. doi: 10.3389/fnins.2012.00091
- López-Larraz, E., Trincado-Alonso, F., Rajasekaran, V., Pérez-Nombela, S., Del-Ama, A. J., Aranda, J., et al. (2016). Control of an ambulatory exoskeleton with a brain-machine interface for spinal cord injury gait rehabilitation. *Front. Neurosci.* 10, 359. doi: 10.3389/fnins.2016.00359
- Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., et al. (2018). A review of classification algorithms for eeg-based brain-computer interfaces: a 10 year update. *J. Neural Eng.* 15, 031005. doi: 10.1088/1741-2552/aab2f2
- Lotte, F., and Jeunet, C. (2018). Defining and quantifying users’ mental imagery-based bci skills: a first step. *J. Neural Eng.* 15, 046030. doi: 10.1088/1741-2552/aac577
- Lotte, F., Larrue, F., and Mühl, C. (2013). Flaws in current human training protocols for spontaneous brain-computer interfaces: lessons learned from instructional design. *Front. Hum. Neurosci.* 7, 568. doi: 10.3389/fnhum.2013.00568
- McFarland, D. J., Sarnacki, W. A., and Wolpaw, J. R. (2010). Electroencephalographic (eeg) control of three-dimensional movement. *J. Neural Eng.* 7, 036007. doi: 10.1088/1741-2560/7/3/036007
- McInnes, L., Healy, J., and Melville, J. (2018). UMAP: uniform manifold approximation and projection for dimension reduction. *arXiv e-prints*, arXiv:1802.03426. doi: 10.21105/joss.00861
- Meng, J., Zhang, S., Bekyo, A., Olsoe, J., Baxter, B., and He, B. (2016). Noninvasive electroencephalogram based control of a robotic arm for reach and grasp tasks. *Sci. Rep.* 6, 1–15. doi: 10.1038/srep38565
- Mestais, C. S., Charvet, G., Sauter-Starace, F., Foerster, M., Ratel, D., and Benabid, A. L. (2015). Wimage: Wireless 64-channel ecog recording implant for long term clinical applications. *IEEE Trans. Neural Syst. Rehabil. Eng.* 23, 10–21. doi: 10.1109/TNSRE.2014.2333541
- Millan, J. R. (2004). “On the need for on-line learning in brain-computer interfaces,” in *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)*, vol. 4 (Budapest: IEEE), 2877–2882.
- Moly, A., Costecalde, T., Martel, F., Martin, M., Larzabal, C., Karakas, S., et al. (2022). An adaptive closed-loop ecog decoder for long-term and stable bimanual control of an exoskeleton by a tetraplegic. *J. Neural Eng.* 19, 026021. doi: 10.1088/1741-2552/ac59a0
- Nakanishi, Y., Yanagisawa, T., Shin, D., Fukuma, R., Chen, C., Kambara, H., et al. (2013). Prediction of three-dimensional arm trajectories based on ECoG signals recorded from human sensorimotor cortex. *PLoS ONE* 8, e72085. doi: 10.1371/journal.pone.0072085
- Nakanishi, Y., Yanagisawa, T., Shin, D., Kambara, H., Yoshimura, N., Tanaka, M., et al. (2017). Mapping ECoG channel contributions to trajectory and muscle activity prediction in human sensorimotor cortex. *Sci. Rep.* 7, 45486. doi: 10.1038/srep45486
- Neuper, C., Müller, G. R., Kübler, A., Birbaumer, N., and Pfurtscheller, G. (2003). Clinical application of an eeg-based brain-computer interface: a case study

- in a patient with severe motor impairment. *Clin. Neurophysiol.* 114, 399–409. doi: 10.1016/S1388-2457(02)00387-5
- Nurse, E. S., John, S. E., Freestone, D. R., Oxley, T. J., Ung, H., Berkovic, S. F., et al. (2017). Consistency of long-term subdural electrocorticography in humans. *IEEE Trans. Biomed. Eng.* 65, 344–352. doi: 10.1109/TBME.2017.2768442
- Orsborn, A. L., Moorman, H. G., Overduin, S. A., Shanechi, M. M., Dimitrov, D. F., and Carmena, J. M. (2014). Closed-loop decoder adaptation shapes neural plasticity for skillful neuroprosthetic control. *Neuron* 82, 1380–1393. doi: 10.1016/j.neuron.2014.04.048
- Pan, G., Li, J.-J., Qi, Y., Yu, H., Zhu, J.-M., Zheng, X.-X., et al. (2018). Rapid decoding of hand gestures in electrocorticography using recurrent neural networks. *Front. Neurosci.* 12, 555. doi: 10.3389/fnins.2018.00555
- Perdikis, S., and Millan, J., d. R. (2020). Brain-machine interfaces: A tale of two learners. *IEEE Syst. Man Cybern. Mag.* 6, 12–19. doi: 10.1109/MSMC.2019.2958200
- Perdikis, S., Tonin, L., Saeedi, S., Schneider, C., and Millán, J., d. R. (2018). The cybathlon bci race: successful longitudinal mutual learning with two tetraplegic users. *PLoS Biol.* 16, e2003787. doi: 10.1371/journal.pbio.2003787
- Pistohl, T., Schulze-Bonhage, A., Aertsen, A., Mehring, C., and Ball, T. (2012). Decoding natural grasp types from human ECoG. *Neuroimage* 59, 248–260. doi: 10.1016/j.neuroimage.2011.06.084
- Pope, P., Zhu, C., Abdelkader, A., Goldblum, M., and Goldstein, T. (2021). “The intrinsic dimension of images and its impact on learning,” in *International Conference on Learning Representations* (Vienna).
- Rashid, M., Islam, M., Sulaiman, N., Bari, B. S., Saha, R. K., and Hasan, M. J. (2020). Electrocorticography based motor imagery movements classification using long short-term memory (LSTM) based on deep learning approach. *SN Appl. Sci.* 2, 211. doi: 10.1007/s42452-020-2023-x
- Rommel, C., Moreau, T., Paillard, J., and Gramfort, A. (2022a). “CADDA: class-wise automatic differentiable data augmentation for EEG signals,” in *International Conference on Learning Representations*.
- Rommel, C., Paillard, J., Moreau, T., and Gramfort, A. (2022b). Data augmentation for learning predictive models on eeg: a systematic comparison. *ArXiv*, abs/2206.14483. doi: 10.1088/1741-2552/aca220
- Rosenfeld, J. S., Rosenfeld, A., Belinkov, Y., and Shavit, N. (2020). “A constructive prediction of the generalization error across scales,” in *International Conference on Learning Representations* (Addis Ababa).
- Schirrmeyer, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangemann, M., et al. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Hum. Brain Mapp.* 38, 5391–5420. doi: 10.1002/hbm.23730
- Shanechi, M. M., Orsborn, A. L., Moorman, H. G., Gowda, S., Dangi, S., and Carmena, J. M. (2017). Rapid control and feedback rates enhance neuroprosthetic control. *Nat. Commun.* 8, 1–10. doi: 10.1038/ncomms13825
- Silversmith, D. B., Abiri, R., Hardy, N. F., Natraj, N., Tu-Chan, A., Chang, E. F., et al. (2020). Plug-and-play control of a brain-computer interface through neural map stabilization. *Nat. Biotechnol.* 39, 326–335. doi: 10.1038/s41587-020-0662-5
- Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., et al. (2017). Closed-loop brain training: the science of neurofeedback. *Nat. Rev. Neurosci.* 18, 86–100. doi: 10.1038/nrn.2016.164
- Śliwowski, M., Martin, M., Souloumiac, A., Blanchart, P., and AksenoVA, T. (2022). Decoding ECoG signal into 3d hand translation using deep learning. *J. Neural Eng.* 19, 026023. doi: 10.1088/1741-2552/ac5d69
- Strang, B., Putten, P., v. d., Rijn, J. N., v., and Hutter, F. (2018). “Don’t rule out simple models prematurely: a large scale benchmark comparing linear and non-linear classifiers in openml,” in *Advances in Intelligent Data Analysis XVII*, eds W. Duivestijn, A. Siebes, and A. Ukkonen (Cham: Springer International Publishing), 303–315.
- Tempczyk, P., Michaluk, R., Garncarek, L., Spurek, P., Tabor, J., and Golinski, A. (2022). “LIDL: local intrinsic dimension estimation using approximate likelihood,” in *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research* (Baltimore, MD: PMLR), 21205–21231.
- UMAP. (2018). *Performance Comparison of Dimension Reduction Implementations*. Available online at: <https://umap-learn.readthedocs.io/en/latest/performance.html#> (accessed July 7, 2022).
- van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.
- Viering, T. J., and Loog, M. (2021). The shape of learning curves: a review. *CoRR*, abs/2103.10948. doi: 10.48550/arXiv.2103.10948
- Volkova, K., Lebedev, M. A., Kaplan, A., and Ossadtchi, A. (2019). Decoding movement from electrocorticographic activity: a review. *Front. Neuroinform.* 13, 74. doi: 10.3389/fninf.2019.00074
- Wodlinger, B., Downey, J., Tyler-Kabara, E., Schwartz, A., Boninger, M., and Collinger, J. (2014). Ten-dimensional anthropomorphic arm control in a human brain-machine interface: difficulties, solutions, and limitations. *J. Neural Eng.* 12, 016011. doi: 10.1088/1741-2560/12/1/016011
- Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3
- Wolpaw, J. R., and McFarland, D. J. (2004). Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proc. Natl. Acad. Sci. U.S.A.* 101, 17849–17854. doi: 10.1073/pnas.0403504101
- Xie, Z., Schwartz, O., and Prasad, A. (2018). Decoding of finger trajectory from ECoG using deep learning. *J. Neural Eng.* 15, 036009. doi: 10.1088/1741-2552/aa9d8e
- Zhang, R., Zong, Q., Dou, L., and Zhao, X. (2019). A novel hybrid deep learning scheme for four-class motor imagery classification. *J. Neural Eng.* 16, 066004. doi: 10.1088/1741-2552/ab3471