Embedding neurophysiological signals

Pierre Guetschel, Théodore Papadopoulo, Michael Tangermann

Embedding neurophysiological signals

Pierre Guetschel
Donders Institute for Brain, Cognition and Behaviour
Radboud University
Nijmegen, Netherlands
0000-0002-8933-7640

Thoédore Papadopoulo
INRIA
Université Côte d'Azur
Valbonne, France
0000-0002-1643-9988

Michael Tangermann
Donders Institute for Brain, Cognition and Behaviour
Radboud University
Nijmegen, Netherlands
0000-0001-6729-0290

***Abstract*—Neurophysiological time-series recordings of brain activity like the electroencephalogram (EEG) or local field potentials can be decoded by machine learning models in order to either control an application, e.g., for communication or rehabilitation after stroke, or to passively monitor the ongoing brain state of the subject, e.g., in a demanding work environment. A typical decoding challenge faced by a brain-computer interface (BCI) is the small dataset size compared to other domains of machine learning like computer vision or natural language processing. The possibilities to tackle classification or regression problems in BCI are to either train a regular model on the available small training data sets or through transfer learning, which utilizes data from other sessions, subjects, or even datasets to train a model. Transfer learning is non-trivial because of the non-stationary of EEG signals between subjects but also within subjects. This variability calls for explicit calibration phases at the start of every session, before BCI applications can be used online. In this study, we present arguments to BCI researchers to encourage the use of embeddings for EEG decoding. In particular, we introduce a simple domain adaptation technique involving both deep learning (when learning the embeddings from the source data) and classical machine learning (for fast calibration on the target data). This technique allows us to learn embeddings across subjects, which deliver a generalized data representation. These can then be fed into subject-specific classifiers in order to minimize their need for calibration data. We conducted offline experiments on the 14 subjects of the High Gamma EEG-BCI Dataset [1]. Embedding functions were obtained by training EEGNet [2] using a leave-one-subject-out (LOSO) protocol, and the embedding vectors were classified by the logistic regression algorithm. Our pipeline was compared to two baseline approaches: EEGNet without subject-specific calibration and the standard FBCSP pipeline in a within-subject training. We observed that the representations learned by the embedding functions were indeed non-stationary across subjects, justifying the need for an additional subject-specific calibration. We also observed that the subject-specific calibration indeed improved the score. Finally, our data suggest, that building upon embeddings requires fewer individual calibration data than the FBCSP baseline to reach satisfactory scores.**

***Index Terms*—BCI, EEG, Deep Learning, Embedding, Triplet Loss, Motor Imagery, Calibration**

I. Introduction

Brain-Computer Interfaces (BCI) are neurotechnological systems that rely on machine learning algorithms to decode brain signals and use the extracted information as commands to control applications. In this article, we argue that embeddings can be used in BCI as a way to utilize the benefits of deep learning for signal decoding while still keeping online calibration periods very short. Deep learning is a dynamic field whose advances can benefit the BCI world. It has been applied to nearly all the problems that involve machine learning, and its driving applications, which are computer vision and natural language processing, lead to new progress daily. The neural networks used in deep learning are also known for their good generalization abilities. Thus, they might help to mitigate one of the greatest challenges of BCI which is the transfer of models—between different sessions or different subjects. Moreover, neural networks can be trained in an end-to-end manner which reduces the importance of feature engineering or selection and of priors about the signal features expected for the experimental protocol. Furthermore, we can speculate that there will be more advancement in the deep learning techniques than for the traditional machine learning ones.

The use of deep learning models in BCI is increasing [3] but they typically remain restricted to offline analyses. Possible barriers to their online use are that, first, most BCI systems require calibration data from the ongoing session to train a session-specific machine learning model. This means that the subject has to wait, connected to the recording equipment, while the decoding model is trained on that calibration data. This model training takes place between the recording of the calibration data and the free online phase. For traditional machine learning models, this is not a problem because they can usually be trained within a few seconds only, whereas training deep learning models is substantially longer or even might be infeasible within a session. The model training typically is also harder to automize as it may require a hyper-parameter search or the expertise of the experimenter. Second, deep learning requires specialized hardware. Both the training and inference of deep learning models are computationally expensive, and if not executed on GPU, they can even become prohibitively so, whereas traditional methods do not require such expensive hardware. Finally, deep learning is data-hungry while recording BCI data is time-consuming, and datasets are neither abundant nor necessarily transferable between sessions or subjects due to their large variability.

In this work, we illustrate how a simple framework that

This work was supported by the Donders Institute for Brain, Cognition, and Behaviour. The code to reproduce our results and the supplementary materials can be found at https://github.com/PierreGtch/embedding_eeg_2022.

uses a neural network to extract features from EEG recordings and a linear classification layer can be calibrated with very little data for a new subject or a new session. The feature extractor - or embedding function - is trained in a leave-one-subject-out manner. The individually calibrated classifier is fed the embedding vectors to be trained on data from the test subject. Such deep learning-based embeddings have already been successfully proposed for BCI recordings but not necessarily for the purpose of shortening the calibration time [4]–[6]. The proposed new framework is simple but still demonstrates that deep learning models are not restricted to post-hoc analyses of BCI recordings.

In the analysis of the proposed framework, we will not only focus on performance aspects but also on the comprehension of what is learned by the embedding function, as this may shed light on what is reasonable to expect from the classification layer. We will also compare our method with a well-established classification method. This comparison will cover both raw performance aspects and practical usability in terms of calibration times.

II. Materials and Methods

A. Dataset

In this study, we trained and compared our models on the publicly available High Gamma Dataset [1]. It contains EEG recordings of fourteen healthy subjects executing three different movement classes and a resting condition. The movements were executed by the left hand, the right hand, and both feet. During each trial, the subject was asked to execute one of the movements or to rest for a duration of four seconds. The exact number of trials varies between subjects but is balanced between the four classes. For each subject, approximately 1000 trials were recorded in a single session. We only used 44 out of the original 128 channels, focussing on the central scalp area, specifically the sensori-motor cortex (as did the authors). Aside from this restriction to 44 channels, we used the full published dataset, i. e., we did not remove potential outlier trials or outlier channels. Likewise, we did not apply any artifact correction methods.

Even though the higher gamma frequencies have been reported to contain useful information, the majority of studies conducted with this dataset focused on the alpha and beta bands. In order to limit the network and dataset sizes for fast prototyping, we thus decided to low-pass filter the signal at 40 Hz and downsample it to 128 Hz, which is the same sampling frequency used by the authors of EEGNet [2]. This decision allowed us to use the EEGNet architecture unaltered for our experiments. For all experiments, we corrected for drifts in the signal by high-pass filtering the signal at 0.5 Hz. From each trial, we used the complete four seconds, starting at the appearance of the task cue, as input for the models. This corresponds to the complete time interval of movement execution, including possible delays due to reaction time. This pre-processing led to a representation of each trial in a matrix with dimensionality 512×44. All models were trained using this exact same pre-processing. The dataset was downloaded and pre-processed through the MOABB library [7]. We chose this dataset to conduct our experiments as it provides a relatively large number of trials per class and subject. This allowed us to evaluate how the number of training trials impacts the performance of the classification layer, see II-B.

B. Neural network architecture

EEGNet is a popular architecture that we used in our experiments for multiple purposes. It was originally designed as an end-to-end EEG classifier by Lawhern et al. [2] who successfully trained it on four different BCI classification tasks that include both ERP and oscillatory data. The architecture of EEGNet feeds raw EEG input through a temporal convolution to learn frequency filters and then a depthwise convolutional layer to learn spatial features, followed by a non-linearity and an average pooling layer, followed by a separable convolutional layer with non-linear activation and average pooling, which provides a high-level feature representation for a final linear classification layer. In this article, we refer to EEGNet without its final classification layer as the so-called *embedding function* (cf. Section II-D).

C. Filter Bank Common Spatial Pattern (FBCSP) baseline

FBCSP is a traditional EEG classification approach for imagery data [8], [9]. It does not use a neural network. We used it as a baseline method against which we compared our results. The computation steps of FBCSP are as follows: First, a **bandpass filtering** is applied to obtain 9 non-overlapping frequency bands between 4 Hz and 40 Hz. Second, 4 spatial filters per frequency band are learned in a supervised approach using the multi-class common spatial patterns (CSP) algorithm [10]. Third, these **spatial filters** are applied to the bandpass filtered data to deliver $9 \times 4 = 36$ features per trial. Fourth, the features are classified by a multinomial **logistic regression classifier** with L2 regularisation.

D. Training strategies

In this work we compared three different classification pipelines: **EEGNet**, **EEGNet+LP** and **FBCSP**.

The **EEGNet** pipeline served as a baseline. For this purpose, the vanilla EEGNet network was trained in a leave-one-subject-out (LOSO) validation approach, i.e. training on all the subjects except one and testing on the left-out subject. To train each of the 14 EEGNet models, we used the cross-entropy loss to classify the examples into one of the four classes. Based on our good prior experience with deep learning for BCI data, we decided to train EEGNet using the SGD optimizer [11] and for 1500 iterations with a one-cycle learning rate scheduler [12]. The implementation we used is `EEGNetv4` from Braindecode [1] and we trained our models using PyTorch [13].

The **EEGNet+LP** pipeline is at the core of this work. It consists of two parts. The first part is formed by an EEGNet (obtained for example from the EEGNet pipeline described above) which had been deprived of its final classification layer after training on n-1 subjects. This pre-trained deprived network serves as an embedding function that delivers

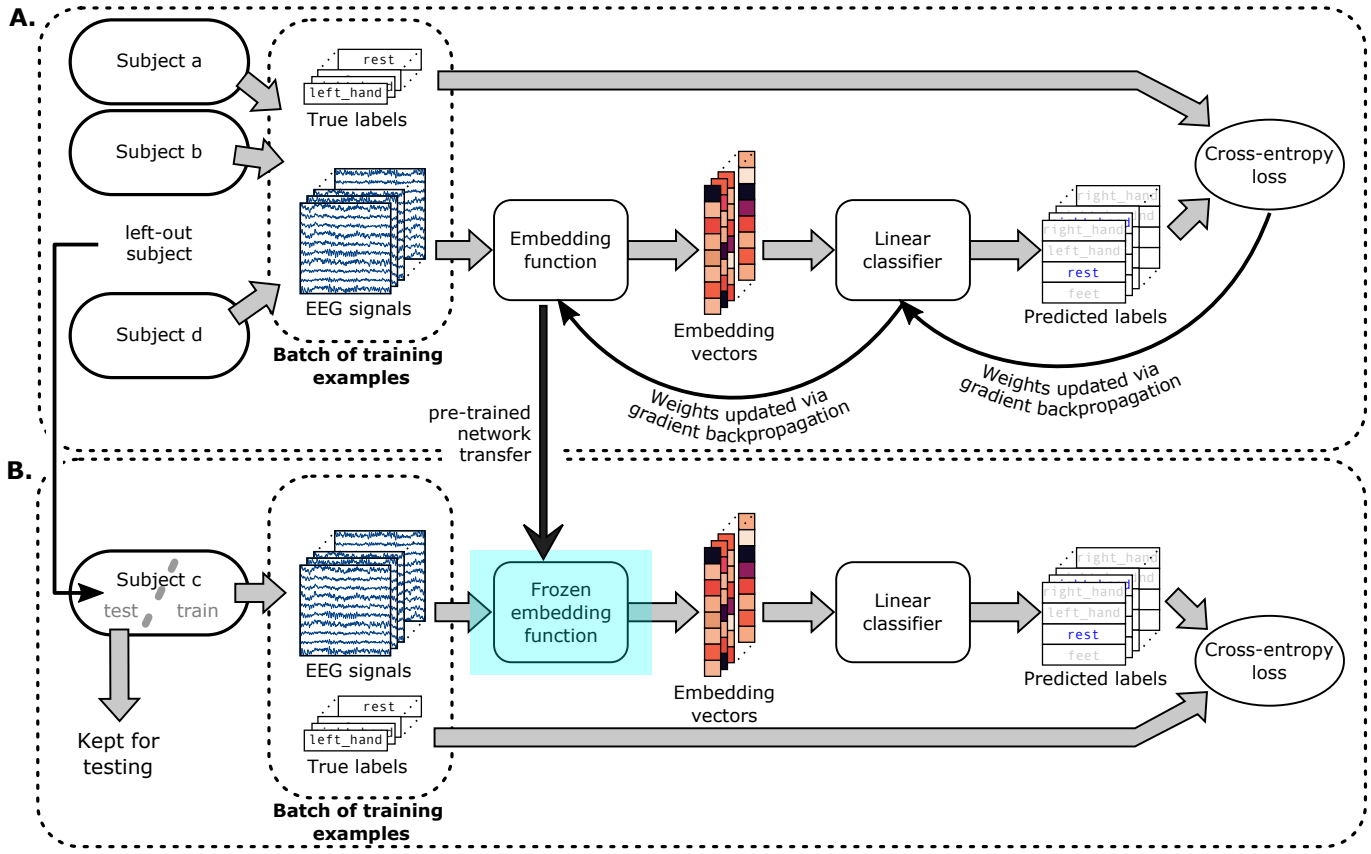

Fig. 1. Diagram **A** describes the training procedure of the **EEGNet** pipeline. Examples from all the subjects except one - the test subject - are used to train the neural network. The weights of the embedding function and the classification layer are updated simultaneously via gradient backpropagation. Diagram **B** describes the training procedure of the **EEGNet+LP** pipeline. First, the pre-trained embedding function is retrieved from the corresponding EEGNet pipeline. Then, a new classification layer is learned by a logistic regression classifier using only some of the recordings from the corresponding left-out subject. In these diagrams, only four subjects are represented for reasons of space, but they are in fact fourteen in all.

256 dimensional embedding vectors for the raw EEG input. Further on, the weights of the embedding function are frozen. The second part is the actual subject-specific training of the EEGNet+LP pipeline: a multinomial logistic regression model with L2 regularisation is trained from scratch to classify the embedding vectors. It is trained on data from the subject that was left out of the embedding function training. A visual description of how the EEGNet and EEGNet+LP pipelines were trained is provided by Figure 1.

A within-subject filter bank common spatial pattern pipeline (**FBCSP**) served as a second baseline. It was trained and tested in the same way as the logistic regression classifier of the EEGNet+LP pipeline. FBCSP thus was used within a single subject only and was not transferred between subjects. We compared to this baseline method as it is widely used for mental imagery decoding, is fast to train (within seconds), known to be trainable from little trials, and has won the BCI competition IV dataset 2a [14]. Overall, it is a good contestant to build a fast to calibrate yet efficient mental imagery-based BCI.

E. Validation Procedures

The validation procedures strictly followed the MOABB benchmarking library [7].

EEGNet pipeline: This pipeline was trained using a leave-one-subject-out cross-validation procedure. The scores reported for this pipeline were computed on the complete data of the left-out test subject.

FBCSP pipeline: FBCSP was trained and evaluated within-subject only. Dependent on the purpose, two different procedures were used: **(1)** For the purpose of evaluating the maximum possible performance, all of a test subject's data was used in a 5-fold cross-validation procedure. Following the MOABB standards, the cross-validation was preceded by a one-time random shuffling of every test subject's data. The resulting performance estimates and the average over the 14 subjects are provided in the subsection III-B and the corresponding Table I. Please note, that the shuffling destroys the chronological order of the trials and thus may lead to over-optimistic generalization estimates compared to true online performance. **(2)** For the purpose of investigating the influence of the calibration data set's size, a so-called *random permutation cross-validation procedure* was applied. Here, the test subject's data set was

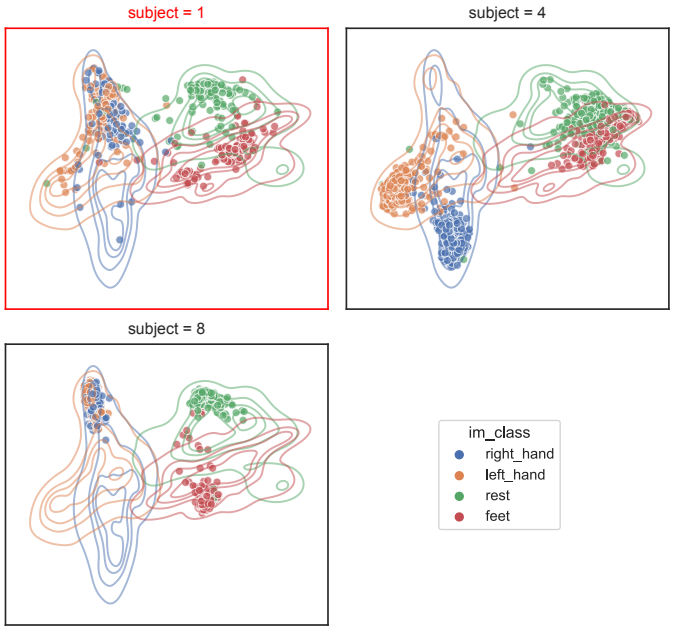

Fig. 2. Projection in 2D of the vectors produced by the embedding function that had been trained on all subjects except subject 1. Each point corresponds to one embedding vector. Its color denotes the class labels. The embedding vectors were projected from 256 to 2 dimensions using UMAP. The sub-plots depict embedding vectors obtained for different subjects, but all vectors were generated by the same embedding function. The sub-plot of test subject 1 is marked by a red frame. The topographic isolines in the background have been plotted after the embedding had been trained and the subject-specific datapoints had been plotted. They indicate the four class distributions as derived from the complete data of all subjects and serve as a visual reference to compare sub-plots. The plots of only three subjects are displayed here for space reasons, but the other subjects can be found in Supplementary Figure 5.

shuffled multiple times. Every time, 20% of data was kept for testing, while from the remaining 80% a limited number of trials were drawn to train the model. The comparison between different calibration set sizes is provided in subsection III-B and the corresponding Figure 3. Please note, that again the chronological structure of the session was not maintained.

EEGNet+LP pipeline: This pipeline always makes use of an embedding obtained from data of n-1 subjects, see Section II-D. The calibration and validation of the logistic regression classifier of this pipeline happened on the data of the test subject and again was done in two versions to fulfill different purposes. Both followed exactly the same procedures as described for the FBCSP pipeline.

Statistical tests: The statistical significance of differences between two processing pipelines was obtained by a Wilcoxon signed-rank test [15]. This score was computed by pairing the scores of each subject from the two pipelines. We choose this test as normality can not be assumed for performance scores due to a range limitation. We considered 0.05 as the maximal threshold for a statistically significant p-value.

III. Results

A. Structure of the Learned Embeddings

In order to evaluate what was learned by the embedding functions, we will first observe the structure of the embedding space. Figure 2 presents the 2D projections of the feature vectors defined by the embedding function which had been trained on the data from all subjects except the test subject 1. Thus it corresponds to the first fold of the LOSO validation. The 2D-projections were obtained by the dimensionality reduction algorithm UMAP [16]. As this is an unsupervised algorithm, the projections from the original 256 dimensions into 2D need to be interpreted cautiously.

The top-left sub-plot of Figure 2 depicts the projection of the features of test subject 1. It overall shows a good class separation, despite that subject 1 had not been seen during the training of the embedding. This indicates a good generalization of the embedding function learned. Figure 2 and Supplementary Figure 5 also allow examining the distributions of embedding vectors of subjects, which—contrary to subject 1—had been used to train the embedding. Across the different sub-figures, we first observe that the blue and orange clusters (representing feature vectors of the right and left hand, respectively) are usually well separated from the remaining two other clusters (rest and feet). Furthermore, the red and green clusters are usually closer together but not overlapping. Lastly, the blue and orange clusters tend to always overlap. These seem to indicate a hierarchy in the capacity of the network at separating the different imagery classes: the right-hand/rest, left-hand/rest, right-hand/feet, and left-hand/feet are the four best-separated class pairs. On the contrary, the right-hand/left-hand pair seem to be the most difficult to separate, while the separability of the rest/feet pair is on a medium level.

On another note, distribution shifts between subjects indicate the inter-subject variability that has frequently been reported for BCI datasets [17]. An example of such a shift is obvious in the sub-plots for subjects 4 and 8. While the clusters of the right and left hand classes of subject 4 are well separated, those of subject 8 overlap in the 2D space. Naturally, a supervised classification layer can be expected to separate the embedding vectors from the four classes more successfully than the un-supervised projector UMAP does. Nevertheless, the distribution shifts in 2D space are an indication that sharing the same classification layer for all the subjects may not be sufficient to obtain a good subject-independent embedding. Thus this observation motivates the use of strategies such as EEGNet-LP and FBCSP which calibrate the classification layer on data from the test subject.

B. Classification Performances

The classification performance of a BCI system measures how well the user's intentions are decoded by the system. It is usually the main point of comparison between models.

In Table I, we can find the average classification accuracies of the three pipelines across all the subjects. As explained in subsection II-E, the scores of the EEGNet pipeline have

TABLE I
CLASSIFICATION ACCURACIES ON THE DIFFERENT SUBJECTS.

Subject Pipeline	1	2	3	4	5	6	7	8	9	10	11	12	13	14	mean	std
EEGNet	0.863	0.734	0.612	0.452	0.741	0.806	0.764	0.308	0.514	0.532	0.414	0.736	0.517	0.790	0.627	0.171
EEGNet+LP	0.931	0.858	0.872	0.908	0.886	0.867	0.881	0.883	0.775	0.796	0.747	0.950	0.816	0.886	0.861	0.058
FBCSP	0.894	0.800	0.947	0.953	0.907	0.930	0.733	0.912	0.743	0.845	0.808	0.918	0.796	0.948	0.867	0.078

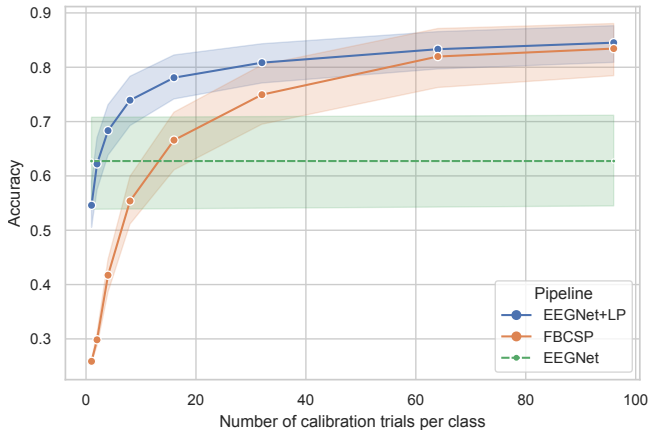

Fig. 3. Development of the classification accuracy for an increasing number of training trials per class to calibrate either the FBCSP baseline method (within-subject scenario, denoted in orange) or the linear classification layer of the EEGNet+LP pipeline (denoted in blue). The performance of the EEGNet baseline model (i.e. without subject-specific calibration of the classification layer) is marked by the green dotted line. The curves express the average accuracies over the fourteen subjects, while the translucent bands represent the standard errors. The score of each subject is estimated by averaging over the different permutations, see subsection II-E. To accurately estimate the scores when few trials are used for calibration, the number of permutations increases in an inverse proportion to the number of calibration trials per class. For 64 and 96 calibration trials per class, we compute 5 permutations, and for one calibration trial per class, 50 permutations are computed. The limitation to 96 trials per class is given by the subject with the lowest trial count contained in the dataset.

been computed on all the data from the test subject, and the scores of the EEGNet+LP and FBCSP pipelines are the results of 5-folds cross-validation procedures over all the available calibration trials. As can be expected from a cross-subject BCI without subject-specific calibration, we observe that the EEGNet pipeline obtains a low average score of 62.7%, which may be problematic for some BCI applications. The standard deviation of this model across subjects is high 17.1%, with high scores for subjects 1, 6, and 14 and performance on chance level for subjects 8 and 11. Moreover, Table I reveals that, for one and the same subject, FBCSP can reach high performances while EEGNet may be lower. This is a clear indication that the EEGNet pipeline performance is not sufficient to make a judgment about the BCI efficiency of a user's dataset. Therefore, the cross-subject generalization performance of this pipeline is highly variable and can not be relied upon.

On the other hand, Table I indicates that the two individually calibrated pipelines FBCSP and EEGNet+LP obtained lower standard deviations of the test performance (7.8% and 5.8% respectively), which suggests that the performance of models using individual calibration data is more consistent than that of un-calibrated models. Moreover, they obtained scores of, respectively, 86.7% and 86.1%, which are relatively high. A Wilcoxon signed-rank test [15] shows that there is no statistically significant difference between the two pipelines over these subjects. However, the scores presented in Table I were computed by 5-folds cross-validation over all the subject's calibration trials, which represents, on average, 241 training trials per subject per class, but when trained with fewer trials, we will see that their performances are no longer equal.

Indeed, in Figure 3, we can observe the scores of these two pipelines for gradually increasing amounts of calibration trials. In these plots, we observe that the EEGNet+LP pipeline (denoted in blue) only requires two trials per class to calibrate its classification layer so that it reaches the same score as the one that was trained by the EEGNet pipeline on the trials from all the other subjects. This shows that the features provided by the embedding are very informative and easy to classify. In contrast, we observe that the FBCSP pipeline requires between 8 and 16 trials per class to reach the EEGNet baseline because the CSP feature extractor also has to be trained from these calibration trials.

If we now assume that a hypothetical BCI application requires a minimal classification accuracy of 70% to function properly, we observe in Figure 3 that this score is reached by the EEGNet+LP pipeline with less than 8 trials per class, whereas between 16 and 32 trials are required by the FBCSP pipeline to reach this threshold. We also note that the low standard deviations of these two pipelines in Table I (i.e. when using all the calibration trials) is still observed with fewer calibration trials in Figure 3. Overall, even if the EEGNet+LP and FBCSP pipelines converge to the same score when many individual calibration trials are available, the EEGNet+LP pipeline reaches satisfactory scores with fewer trials, and without any sacrifice on the uncertainty of its performance.

C. Calibration Time

Reducing the number of calibration samples needed by a system is important because it directly determines the time that a user has to spend before starting to actually use a BCI application productively. If we keep 70% as the minimal classification accuracy threshold for a given BCI application to be usable, we can simulate the time that would be needed to reach it. The trials last four seconds and four additional seconds of rest are spent between trials. The times needed to

optimize the EEGNet+LP and FBCSP pipelines, once all the calibration data is collected, are respectively 1 and 8 seconds, and can be considered negligible. Therefore, the pipelines can respectively be calibrated in 4 min. 15 seconds or 17 min. This difference is non-negligible and matters to the user.

IV. Discussion

In this article, we saw how splitting a neural network into two components (embedding function and linear classification layer) could allow to use it in an online BCI with a calibration. We observed that the trained embedding function was able to reflect the intended motor task's class structure but we also observed non-neglectable distribution shifts between the subjects. Our pipeline obtained a classification accuracy similar to the FBCSP baseline when using a large number of calibration trials, but performed better when fewer trials were available. In particular, less than 8 trials per class are required to obtain a score of 70%. This number of trials can be recorded in around four minutes. In other words, using our pipeline in an online application allows for reducing the duration of the calibration phase. Although the High-Gamma Dataset is well established as a benchmark, it comes with the caveat that it contains only fourteen healthy subjects executing motor tasks. Thus, our results will need to be confirmed with further datasets.

Our data exploitation strategy is a trade-off that lies on a spectrum between two extremes. On one end, the embedding function could be trained using data very close to the test subject's distribution, which should reduce the number of required calibration trials. However, this becomes increasingly impractical as embedding functions would have to be trained for every new subject. On the other end of the spectrum, these functions could be trained on distributions distant from the test subject's one. While it may be easier to find such datasets and while such embeddings could potentially be re-used in different scenarios, they do require more calibration data. The extreme case of self-supervised learning uses unlabeled datasets or even non-BCI EEG recordings. Kostas et al. [5] trained self-supervised embeddings using the extremely large Temple University corpus [18]. Different self-supervised learning techniques have been compared [4]. Further research is required to determine which strategies are optimal in specific scenarios. For example, if a patient regularly uses a BCI, it might be worth investing the time to train a good individual embedding to minimize the calibration per session. Moreover, because of the relative independence between the feature extractor and the classification layer in our pipeline, the classification layer can also be modified with great freedom. We can, for example, imagine optimizing it with an unsupervised algorithm. This would allow starting a BCI application without any session- or subject-specific calibration phase. Finally, testing the proficiency of this setup in online experiments would be interesting in future studies.

To conclude, we proposed to use an established architecture in a slightly unconventional way: we saw that deep learning-based embeddings provide a way to build BCI systems that can be calibrated quickly for new subjects. The pipeline we proposed is simple to implement and optimize, but already brings improvements over state-of-the-art classical analysis pipelines, but also state-of-the-art network architectures. Furthermore, because of its simplicity, it can still be greatly improved.

References

[1] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[2] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, Oct. 2018.

[3] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: A systematic review," *Journal of Neural Engineering*, vol. 16, no. 5, p. 051001, Oct. 2019.

[4] H. Banville, O. Chehab, A. Hyvärinen, D.-A. Engemann, and A. Gramfort, "Uncovering the structure of clinical EEG signals with self-supervised learning," *Journal of Neural Engineering*, vol. 18, no. 4, p. 046020, Aug. 2021.

[5] D. Kostas, S. Aroca-Ouellette, and F. Rudzicz, "BENDR: Using Transformers and a Contrastive Self-Supervised Learning Task to Learn From Massive Amounts of EEG Data," *Frontiers in Human Neuroscience*, vol. 15, p. 653659, Jun. 2021.

[6] H. Alwasiti, M. Z. Yusoff, and K. Raza, "Motor imagery classification for brain computer interface using deep metric learning," *IEEE access : practical innovations, open solutions*, vol. 8, pp. 109 949–109 963, 2020.

[7] V. Jayaram and A. Barachant, "MOABB: Trustworthy algorithm benchmarking for BCIs," *Journal of Neural Engineering*, vol. 15, no. 6, p. 066011, Dec. 2018.

[8] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. IEEE, 2008, pp. 2390–2397.

[9] Z. Y. Chin, K. K. Ang, C. Wang, C. Guan, and H. Zhang, "Multi-class filter bank common spatial pattern for four-class motor imagery BCI," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Sep. 2009, pp. 571–574.

[10] M. Grosse-Wentrup and M. Buss, "Multiclass Common Spatial Patterns and Information Theoretic Feature Extraction," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 8, pp. 1991–2000, Aug. 2008.

[11] S. Ruder, "An overview of gradient descent optimization algorithms," Jun. 2017.

[12] L. N. Smith and N. Topin, "Super-convergence: Very fast training of neural networks using large learning rates," in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, vol. 11006. International Society for Optics and Photonics, 2019, p. 1100612.

[13] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, pp. 8026–8037, 2019.

[14] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. Miller, G. Mueller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the BCI Competition IV," *Frontiers in Neuroscience*, vol. 6, 2012.

[15] F. Wilcoxon, "Individual comparisons by ranking methods," in *Breakthroughs in Statistics*. Springer, 1992, pp. 196–202.

[16] T. Sainburg, L. McInnes, and T. Q. Gentner, "Parametric UMAP Embeddings for Representation and Semisupervised Learning," *Neural Computation*, vol. 33, no. 11, pp. 2881–2907, Oct. 2021.

[17] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update," *Journal of Neural Engineering*, vol. 15, no. 3, p. 031005, Jun. 2018.

[18] I. Obeid and J. Picone, "The temple university hospital EEG data corpus," *Frontiers in neuroscience*, vol. 10, p. 196, 2016.