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

Towards calibration-free online EEG motor imagery decoding using Deep Learning

Martin Wimpff[1], Jan Zerfowski[2] and Bin Yang[1]

[1]Institute of Signal Processing and System Theory, University of Stuttgart, Germany
[2]Clinical Neurotechnology Laboratory, Department of Psychiatry and Neurosciences, Charité Campus Mitte (CCM), Charité - Universitätsmedizin Berlin, Germany

Abstract. The prevalence of stroke-induced disability drives research in motor imagery Brain-Computer Interfaces (BCIs) for rehabilitation. Closed-loop systems using traditional decoding models prevail but deep learning advances in single-trial offline decoding offer promises. However, transferring methods from offline to online decoding poses challenges. To address this, we propose a new approach to tune existing offline deep learning models towards online decoding, outperforming traditional pipelines without the need for subject-specific calibration data. Our proposed method is a step towards calibration-free BCIs that enable immediate feedback and user learning.

1 Introduction

Globally, stroke stands as a primary contributor to enduring disability among adults. An increasing body of evidence indicates the retention of brain plasticity in chronic stroke patients, suggesting the possibility of recovery for affected limbs [1, 2]. Henceforth, different rehabilitation strategies have been investigated, including motor imagery Brain-Computer Interfaces (BCIs) [3, 4]. Motor imagery, the process of movement imagination without execution, shares neural mechanisms with actual movements [5]. This type of BCI is suitable for use with both healthy subjects and severely paralyzed patients.

One major aspect in BCIs for clinical use is the feedback provided to the user for the purpose of self-regulation [2, 6]. BCIs with feedback are also referred to as closed-loop or *online* systems, whereas systems without feedback are labeled *offline*. Online systems typically decode multiple short and overlapping windows within a single trial to provide continuous feedback, while offline decoders usually classify entire trials at once. To date, feedback is mostly delivered by traditional methods such as Common Spatial Patterns (CSP) combined with Linear Discriminant Analysis (LDA) or Support Vector Machine (SVM) classifiers [1]. Only a very limited number of approaches [7, 8] employ deep learning models for online classification with moderate success compared to traditional methods [1]. This is contrary to the developments in single-trial decoding, where deep learning has mostly overtaken traditional methods [9]. We argue that this is due to the naive transfer of deep learning methods from offline single-trial decoding to online decoding as evidenced in [7, 8].

In this work, we propose a new method to tune existing deep learning models towards online decoding to enable efficient training and deployment. We show that

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

a simple deep learning model can outperform a traditional CSP+LDA combination without the need for any subject-specific calibration data by leveraging data from other subjects. Being able to immediately provide patients with feedback facilitates user learning and opens up new possibilities for rehabilitation.

2 Method

2.1 Dataset

We employ the large EEG database [10] recently published. This dataset contains EEG data from 87 subjects performing a binary (left hand vs. right hand) motor imagery task recorded with 27 electrodes. We exclude 8 subjects due to artifacts and missing data, leading to effectively 79 subjects. For each subject there are 120 trials per class, recorded in 6 runs. Each trial lasts 8 seconds featuring a cue appearing after 3 seconds, preceded by a fixation cross and an auditory signal. The cue is presented for 1.25 seconds, followed by 3.75 seconds of visual feedback. The first two calibration runs were recorded with sham feedback and the last four with real feedback, provided by a decoder consisting of CSP and LDA trained on the data from the first two calibration runs.
The feedback is given based on the the last second of EEG data and updated with a frequency of 16 Hz. Since user feedback is provided from 4.25 s after trial onset, only the data from 3.25 s to 8 s is used. This results in $(4.75\,\mathrm{s} - 1\,\mathrm{s}) \cdot 16\,\mathrm{Hz} + 1 = 61$ windows per trial with a $\frac{15}{16} = 93,75\%$ overlap between consecutive windows.
For preprocessing, we employ the same 5 - 35 Hz bandpass filter for all subjects and downsample the data from 512 Hz to $f_s = 256\,\mathrm{Hz}$.

2.2 Model adjustments

As discussed in [1], current deep learning models do not outperform traditional methods in online decoding despite their advancements in single-trial decoding. We argue that this is at least in parts due to the naive transfer of methods from single-trial decoding to online decoding. We propose a new method to tweak existing deep learning models towards online decoding which is applicable to any convolutional architecture. To showcase our method, we employ the simple yet powerful BaseNet architecture [11] which is a modern evolution from ShallowNet [12] and EEGNet [13].
As online decoding needs a high update frequency (e.g., $f_u = 16\,\mathrm{Hz}$), there is a high overlap between consecutive windows (e.g., $\frac{15}{16}$). This overlap is also present in intermediate layers of a deep learning model as previously stated in [12]. In [12], a 'cropped training' strategy is used to stabilize and regularize the training of a model for single-trial decoding. Its idea is to decode multiple smaller sliding windows within one trial to increase the number of training samples. As this results in additional computational load, groups of neighboring windows are decoded together and the intermediate convolution outputs are reused.
We employ a similar strategy, but match it with the online decoding requirements. Specifically, we tune the kernel k_i and stride lengths s_i of the pooling

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

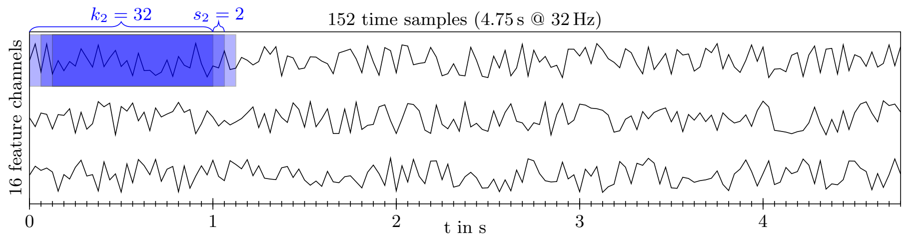

Fig. 1: Second pooling layer of BaseNet.

layers in the model. We propose the following new strategy for 2 pooling layers: The first layer is used to downsample the original input from a sampling frequency f_s to an intermediate frequency $f_{\text{inter}} = \frac{f_s}{k_1}$ with $k_1 = s_1$. For models with only one pooling layer, this stage is dropped and $f_{\text{inter}} = f_s$. For BaseNet (2 pooling layers), we use $f_s = 256\,\text{Hz}, k_1 = s_1 = 8, f_{inter} = 32\,\text{Hz}$. The second pooling layer is then used to extract overlapping windows which fulfill the requirements of the application (window length w and update frequency f_u). Its procedure for BaseNet is visualized in Figure 1. The kernel size $k_2 = f_{\text{inter}} \cdot w$ is chosen based on the window length w. The stride $s_2 = \frac{f_{\text{inter}}}{f_u}$ depends on the update frequency f_u, which determines the overlap between consecutive windows. Our approach is applicable to any number of pooling layers P, where the first $P - 1$ layers are used for downsampling ($k_i = s_i$) and the last one is used to extract the windows. By re-parameterizing the pooling layers in this manner, any model acquires the ability to decode both singular windows and sequences of consecutive windows. Crucially, the prediction for a single window is solely contingent upon that window and does not rely on any accompanying windows. Decoding all windows of one trial jointly is computationally very efficient. Passing every window independently would need $\frac{w \cdot (f_u \cdot (t-w)+1)}{t}$ more operations. For our dataset and setting, this corresponds to a factor of $\frac{1\,\text{s} \cdot 61}{4.75\,\text{s}} \approx 12.84$.
To stabilize training, we average the predictions of all windows of one trial before backpropagation. This averaging minimizes the effect of possible outliers (e.g., due to artifacts) within one trial.

2.3 Training

2.3.1 Data split

Within-subject: We first train one model per subject on the two calibration runs and test the model on the four online runs as done in [10]. Additionally, we investigate how the amount of training data affects the model performance. To do so, we use the first 1 - 5 runs for training and test only on the last run.
Cross-subject: We train one model on the data from 78 subjects and test on the four online runs of the remaining unseen subject (leave-one-subject-out) and repeat this for every subject (cross-validation). We either use the two calibration runs or all 6 runs of the training subjects. Additionally, we vary the number of training subjects by selecting a random subset of the 78 training subjects for each test subject.

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

2.3.2 Training procedure

We train all models with the training procedure described in [11]. We use an Adam optimizer with a learning rate of 10^{-3} and train each model for 100 epochs using a learning rate scheduler with 20 warmup epochs. As the training process is stochastic (e.g., subject selection, data shuffling, weight initialization and dropout), we train each model for five different random seeds and report the average of these five runs. The complete source code is available at https://github.com/martinwimpff/eeg-online.

3 Results and Discussion

Table 1: Average trial-wise accuracy [10] over all subjects, standard deviation calculated between subjects.

	Method	training subjects	training runs	test runs	test accuracy (%)
within - subject	BaseNet	1	1 - 2	6	57.36 ± 12.40
	CSP+LDA[10]	1	1 - 2	6	64.62 ± 17.27
	BaseNet	1	1 - 5	6	65.82 ± 15.82
	BaseNet	1	1 - 2	3 - 6	57.60 ± 11.60
	CSP+LDA[10]	**1**	**1 - 2**	**3 - 6**	$\mathbf{63.29 \pm 15.82}$
cross - subject	BaseNet	10	1 - 2	3 - 6	62.19 ± 11.62
	BaseNet	10	1 - 6	3 - 6	63.68 ± 12.41
	BaseNet	20	1 - 2	3 - 6	64.27 ± 12.59
	BaseNet	20	1 - 6	3 - 6	66.49 ± 13.25
	BaseNet	78	1 - 2	3 - 6	67.80 ± 13.80
	BaseNet	**78**	**1 - 6**	**3 - 6**	$\mathbf{69.29 \pm 13.70}$

The reported accuracies are trial-wise accuracies as in [10] and describe the percentage of correctly classified trials. Since there are no other works employing online classification with this dataset, we only compare our method against the original results presented in [10].

Within-subject: The within-subject results are shown in the first part of Table 1 and in Figure 2a. Using the setting of [10] (i.e., only using the two calibration runs for training), their method is superior as the number of training samples is too low for deep learning models. This observation is supported by Figure 2a which demonstrates a consistent improvement for BaseNet with the addition of more training data.

Cross-subject: The cross-subject results are shown in the second part of Table 1 and in Figure 2b. As the algorithm in [10] selects a subject-specific frequency band, it cannot be applied to the cross-subject settings. Surprisingly, our cross-subject performance (using all subjects) is significantly above the within-subject performance without using any subject-specific data. Using more training data (i.e., using all runs of the training subjects) improves the trial-wise accuracy by around 1.7%. The number of subjects used for training also influences the performance as shown in Figure 2b. By only using the calibration runs of 20

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

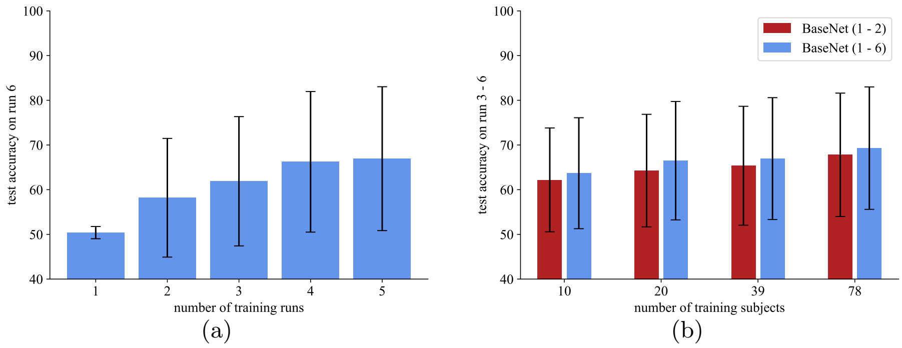

Fig. 2: Within-subject results (a) and cross-subject results (b) for different amounts of training data. Black bars indicate the standard deviation between subjects.

randomly selected subjects other than the test subject, our cross-subject setting (64.27 ± 12.59) already performs at the same level as the within-subject setting in [10], a notable achievement given the scarcity of large BCI databases.

Apart from looking at the average performance, we also investigated the number of subjects who reached a performance level significantly ($p < 0.05$) above the individual chance level ($> 56.25\%$ for 160 test trials, see [14]). For [10], 40 subjects were above this threshold, whereas for BaseNet (cross-subject, 78 training subjects) 56 subjects (trained on the calibration runs) and 64 subjects (trained on all runs) were above this threshold. This improved robustness between subjects is further evidenced by the lower standard deviation of our method compared to [10] (see Table 1). The inference time of BaseNet for one window is 2.15 ms (Intel i7-1195G7, 4 cores), making it computationally suitable for online decoding.

These results showcase not only the effectiveness of our proposed method, but also provide insights on how EEG data can be decoded in real-time. Ultimately, the choice of decoding algorithm depends on the data availability. With enough data, deep learning is able to outperform traditional methods. Importantly, deep learning might be better suited for cross-subject decoding which is promising as it does not need any subject-specific calibration data. Apart from saving time, such calibration-free decoders can be more efficient than their subject-dependent counterparts. If a subject-specific decoder is built on bad data, the feedback does not help the subject to elicit the proper brain activity [6]. As user learning is an essential part of BCI deployment [1], we recommend calibration-free cross-subject decoders over subject-dependent decoders. Future studies should investigate how to transfer knowledge across subjects effectively and how to adapt a given subject-independent decoder towards the target subject over time [6, 15] to allow mutual learning of user and decoder.

4 Conclusion

Our work introduces a novel approach to tailor BCI deep learning models towards a specific decoding task by adjusting the existing pooling layers. Our method outperforms conventional approaches without using subject-specific data by leveraging data from other subjects. This allows calibration-free decoding and enables immediate user learning, a crucial part of BCI usage.

References

[1] Vavoulis, A., Figueiredo, P. & Vourvopoulos, A. A Review of Online Classification Performance in Motor Imagery-Based Brain-Computer Interfaces for Stroke Neurorehabilitation. *Signals*. **4**, 73-86 (2023)

[2] Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., Weiskopf, N., Blefari, M., Rana, M., Oblak, E. & Others Closed-loop brain training: the science of neurofeedback. *Nature Reviews Neuroscience*. **18**, 86-100 (2017)

[3] Cervera, M., Soekadar, S., Ushiba, J., Millán, J., Liu, M., Birbaumer, N. & Garipelli, G. Brain-computer interfaces for post-stroke motor rehabilitation: a meta-analysis. *Annals Of Clinical And Translational Neurology*. **5**, 651-663 (2018)

[4] Soekadar, S., Witkowski, M., Mellinger, J., Ramos, A., Birbaumer, N. & Cohen, L. ERD-based online brain-machine interfaces (BMI) in the context of neurorehabilitation: optimizing BMI learning and performance. *IEEE Transactions On Neural Systems And Rehabilitation Engineering*. **19**, 542-549 (2011)

[5] Decety, J. The neurophysiological basis of motor imagery. *Behavioural Brain Research*. **77**, 45-52 (1996)

[6] Kumar, S., Alawieh, H., Racz, F., Fakhreddine, R. & Millán, J. Transfer learning promotes acquisition of individual BCI skills. *PNAS Nexus*. **3**, pgae076 (2024)

[7] Tayeb, Z., Fedjaev, J., Ghaboosi, N., Richter, C., Everding, L., Qu, X., Wu, Y., Cheng, G. & Conradt, J. Validating deep neural networks for online decoding of motor imagery movements from EEG signals. *Sensors*. **19**, 210 (2019)

[8] Karácsony, T., Hansen, J., Iversen, H. & Puthusserypady, S. Brain computer interface for neuro-rehabilitation with deep learning classification and virtual reality feedback. *Proceedings Of The 10th Augmented Human International Conference 2019*. pp. 1-8 (2019)

[9] Craik, A., He, Y. & Contreras-Vidal, J. Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal Of Neural Engineering*. **16**, 031001 (2019)

[10] Dreyer, P., Roc, A., Pillette, L., Rimbert, S. & Lotte, F. A large EEG database with users profile information for motor imagery brain-computer interface research. *Scientific Data*. **10**, 580 (2023)

[11] Wimpff, M., Gizzi, L., Zerfowski, J. & Yang, B. EEG motor imagery decoding: A framework for comparative analysis with channel attention mechanisms. *ArXiv Preprint ArXiv:2310.11198*. (2023)

[12] Schirrmeister, R., Springenberg, J., Fiederer, L., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W. & Ball, T. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*. **38**, 5391-5420 (2017)

[13] Lawhern, V., Solon, A., Waytowich, N., Gordon, S., Hung, C. & Lance, B. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *Journal Of Neural Engineering*. **15**, 056013 (2018)

[14] Combrisson, E. & Jerbi, K. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal Of Neuroscience Methods*. **250** pp. 126-136 (2015)

[15] Wimpff, M., Döbler, M. & Yang, B. Calibration-free online test-time adaptation for electroencephalography motor imagery decoding. *2024 12th International Winter Conference On Brain-Computer Interface (BCI)*. pp. 1-6 (2024)