Proceedings of the
9th Graz Brain-Computer Interface Conference 2024

10.3217/978-3-99161-014-4-009

A STUDY OF PERFORMANCE VARIABILITY IN DEEP NEURAL NETWORKS FOR MOTOR IMAGERY CLASSIFICATION: TOWARDS A ZERO-CALIBRATION APPROACH

Pasquale Arpaia[1,2,3], Antonio Esposito[1,2], Fortuna Galdieri[1,2], Angela Natalizio[1,4], Marco Parvis[4], Andrea Pollastro[1,2]

[1] Augmented Reality for Health Monitoring Laboratory (ARHeMLab)
[2]Department of Electrical Engineering and Information Technology (DIETI), Università degli Studi di Napoli Federico II, Naples, Italy
[3]Centro Interdipartimentale di Ricerca in Management Sanitario e Innovazione in Sanità (CIRMIS), Università degli Studi di Napoli Federico II, Naples, Italy
[4]Department of Electronics and Telecommunications (DET), Polytechnic of Turin, Turin, Italy

E-mail: pasquale.arpaia@unina.it

ABSTRACT: This study deals with the adoption of deep learning and transfer learning in motor imagery-based brain-computer interfaces to develop a robust system with a zero-calibration approach. Deep neural networks would be also sought to improve the classification accuracies of these interfaces. However, these approaches are affected by inherent variability in their performance, so that dominating uncertainty sources appears crucial. To assess the performance variability of deep neural networks, the effects of parameter initialisation and pre-processing were studied. EEGNet and Sinc-EEGNet were used for this purpose. The results highlight that network's weight initialisation significantly affect the performance. For instance, classification accuracy can improve from 67 % $\pm$ 3 % to 73 % $\pm$ 3 % by just changing the weight initialisation. Meanwhile, EEG pre-processing does not improve the performance, thus it can be avoided to reduce the computational effort. These results pave the way for real-time application scenarios.

Keyword: brain-computer interface, motor imagery, deep learning, transfer learning, uncertainty.

INTRODUCTION

A motor imagery-based Brain-Computer Interface (BCI) measures voluntarily modulated brain signals generated while imagining a movement [1]. Notably, non-invasive and wearable BCIs based on motor imagery have been investigated more and more [2, 3]. These typically exploit electroencephalographic (EEG) signals acquired through electrodes placed on the scalp [4]. During the execution of a motor imagery task, spectral power changes occur in the μ (7 Hz to 13 Hz) and β (13 Hz to 30 Hz) bands of the signals recorded over the sensorimotor brain area. In particular, event-related desynchronisation and synchronisation can be observed immediately before and after motor imagery, respectively [5]. Therefore, the interface attempts to detect the imagined movement through the analysis of temporal, frequency, and spatial features of the acquired signals.

Despite their potential in several fields [3, 6–8], motor imagery-based BCIs still suffer from many limitations. Firstly, the user must learn how to modulate sensorimotor rhythms. Secondly, motor imagery signals are hindered by noise, which can be either baseline neural activity [9] or artefacts. In addition, a large amount of data are needed for an effective training and testing of processing pipelines, typically relying on machine learning [10–12]. Hence, long calibration periods (20-30 minutes) are needed before properly using the BCI system [10]. Moreover, EEG signals suffer from a significant inter- and intra-subject variability [13]. This implies that calibration data should be acquired for each new user and new session of the same user. The highlighted challenges are exacerbated when the BCI system has to discriminate as the number of motor imagery tasks increases [14].

In this framework, research has recently focused on deep learning methods to improve motor imagery classification, especially in multi-class problems [12, 14, 15]. However, deep learning techniques require a bigger amount of data with respect to classical machine learning approaches [14–16].

EEG data are typically acquired in controlled experimental conditions, but the experimental burden make it difficult to obtain large and significant datasets in practice [16]. Therefore, common strategies to compensate for the lack of data are data augmentation and transfer learning (TL) [10, 16, 17]. In particular, TL is based on training a model by relying on the knowledge gained from another pre-trained model. This approach has the advantage of reducing training time [12, 16, 18] or neglecting it in a "zero calibration" scenario, where the EEG data of a subject are classified by a model identified on independent data from other subjects.

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-009

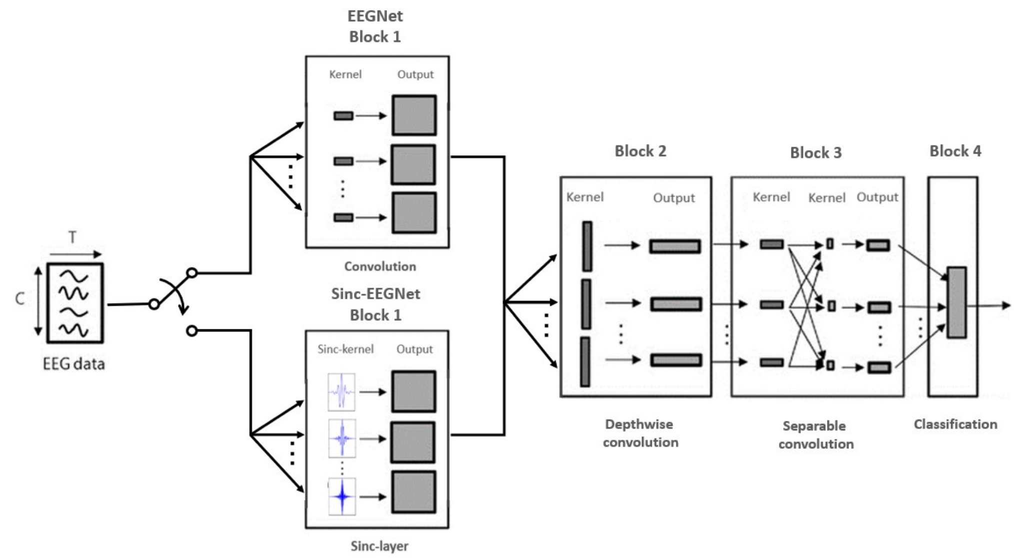

Figure 1: Adopted deep neural network architectures for motor imagery EEG processing, namely EEGNet and Sinc-EEGNet. The only difference between the two is in Block 1, which is a standard convolutional layer for EEGNet and a Sinc-layer for Sinc-EEGNet. The remaining blocks are identical in both architectures.

The combination of TL methods with deep learning-based processing pipelines appears promising in the BCI context [14, 16, 19]. However, it is worth emphasising that deep neural networks performance is very sensitive to the weight initialisation and data pre-processing: the former affects the training process [20–22], while the latter improves the quality of the data [12, 15, 23, 24]. Unfortunately, in context of TL applications for motor imagery-based BCI, there is a lack of studies investigating those performance variations [15, 21, 22], especially regarding the pre-processing [24].

This paper thus focuses on performance variability due to weight initialisation and pre-processing in the context of TL for motor-imagery BCI towards a zero-calibration approach. Notably, EEGNet [25] and Sinc-EEGNet [26] were investigated. As common pre-processing strategies for EEG signals consist of the use of basic filtering techniques, this paper focuses on the use of band-pass filter [27, 28] and Laplacian filter [29].

Therefore, the remainder of the paper is organised as follows. Sec. MATERIALS AND METHOD describes EEGNet and Sinc-EEGNet architectures as well as the analysis conducted on them by involving TL. Sec. RESULTS presents the exploited dataset and discusses inherent results.

MATERIALS AND METHODS

The purpose of this Section is to present the architectures used in the study and the analyses performed on them. It is worth remarking that all the analyses were performed under zero calibration. The Section is structured as follows: in Sec. *Architectures*, EEGNet [25] and Sinc-EEGNet [26] are presented (Fig. 1), while the methodology proposed for comparing different setting is detailed in Sec. *Experimental Setup*.

Architectures: EEGNet is one of the most commonly used deep learning architectures in BCI [25]. It is a low-density convolutional neural network designed to robustly extract information from EEGs. It uses both depthwise and separable convolutions to extract EEG features. The architecture is structured in four blocks. In the first block, two sequential convolutional layers are used as a temporal filter. In the second block, a depth-wise convolutional layer is used. In EEG-specific applications, this type of layer provides a direct way to learn the spatial filters for each temporal filter, allowing for efficient extraction. The third block uses a separable convolution, which reduces the number of parameters to be fitted and explicitly decouples relationships within and between feature maps. Finally, in the classification block, the features are passed directly to a softmax function.

Interestingly, the architecture of EEGNet resembles the steps of the well-known filter bank common spatial pattern algorithm [30], adding flexibility thanks to the end-to-end training procedure of deep learning models. The strengths of this architecture with respects to general-purpose convolutional neural networks include (i) reduced number of trainable parameters due to the use of depthwise and separable convolutions, (ii) applicability to low-dimensional data, and (iii) adaptability across different EEG datasets and tasks [25, 31, 32].

An EEGNet variant called Sinc-EEGNet was also recently proposed [26]. It consists of merging EEGNet [25] with Sinc-Net [33], which is characterised by a convolutional layer having learnable sinc functions as filters. The main strength of Sinc-Net consists in deriving a custom filter bank, specifically tuned for the desired application [33]. Sinc-EEGNet consists of an EEGNet architecture in which the first convolutional layer has been replaced by

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-009

a sinc-layer, resulting in a reduced number of trainable parameters [26]. Different ways for combining the two architectures were proposed in literature [34–36]. This works focused on the version proposed in [26]. It faithfully reproduced the original version of EEGNet except for the first block, in which the first traditional convolutional layer was replaced by a sinc layer [33]. The structures of the two architectures are jointly illustrated in Fig. 1 to stress that the only difference resides in the first block. Such architectures are both suitable for TL, with the Sinc-EEGNet variant that is more prone to explainability [26].

Training hyperparameters	Values
n. of training epochs	1000
learning rate	0.0001
batch size	32
optimizer	Adam
early stopping patience	150
weight decay	0.02

Table 1: Training setup for the conducted analyses.

Experimental Setup: The analyses performed in this work were based on EEGNet and Sinc-EEGNet with either 4 or 32 filters in the first block. In particular, 4 is the optimal number of filters for EEGNet [25] and 32 is the optimal number of filters for Sinc-EEGNet [26]. The name of the architecture followed by the number of filters is used to refer to the specific architectures (e.g. EEGNet-4 refers to EEGNet with 4 filters in the first convolutional layer). Training hyperparameters were optimized in a previous work [26] and they are recalled for clarity in Tab. 1. All the analyses were done in the zero calibration TL scenario. Notably, the leave-one-subject-out [37] technique was used for this purpose. Once the dataset is selected, all samples relating to a single subject are removed and the model is then trained on the remaining samples. The performance of the model is then evaluated on the independent samples from the left-out subject. Such an offline analysis simulates the straightforward usage of a BCI on a previously unseen subject.
Two comparative analyses were carried on by using the above mentioned architectures. The first one consisted of analysing the models' performances with different weight initialisation. The He initialisation [38] was used in this work. To this aim, the seed for pseudo-random generation of initial parameters was firstly varied from 42 to 56 (15 values). Then, by exploiting the optimal seed for each model, a second analysis step was carried on. This consisted of analysing the models' performance as the EEG pre-processing strategy was varied. In particular, the following three cases were tested: (i) no pre-processing, (ii) band-pass filter from 4 Hz to 40 Hz, and (iii) Laplacian filter.
For all the analyses, the metric adopted to assess models' performances was the mean classification accuracy across subjects of the selected dataset and its associated type A uncertainty, i.e. the standard deviation divided by the square-root of the number of averaged accuracies.

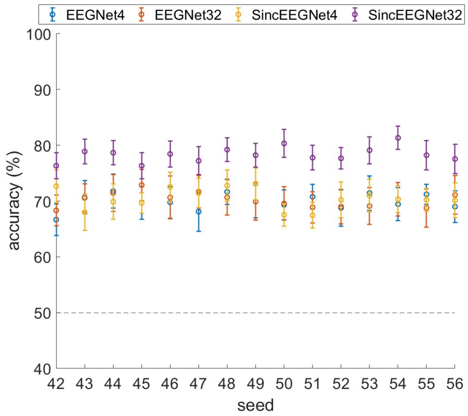

Figure 2: Mean classification accuracies and associated uncertainties obtained as a function of the seed (i.e. weight initialisation) for each architecture. The dotted line refers to the random accuracy.

RESULTS

This Section presents the results of the analyses. In details, Sec. *Dataset* describes the data and its usage, Sec. *Weight initialisation impact* presents the results of the first analysis step, and Sec. *Pre-processing impact* presents the results associated with different pre-processing techniques.

Dataset: the benchmark dataset *BCI competition IV, 2a* was used for the analyses [39]. It includes EEG signals from nine healthy subjects recorded using 22 wet electrodes. The sampling rate was 250 Sa/s. The subjects performed four motor imagery tasks during two sessions recorded on two different days. As the present study adopts an inter-subjective approach, the investigations considered the only first session. Moreover, two classes of motor imagery were used, namely left and right hand motor imagery. Finally, each trial was epoched from 2 s to 6 s, thus including the cue and the motor imagery windows.

Weight initialisation impact: Fig. 2 shows the results obtained for each configuration (i.e., EEGNet-4, EEGNet-32, Sinc-EEGNet-4, and Sinc-EEGNet-32) when varying the seed. In particular, each point represents the mean classification accuracy across the nine subjects of the dataset together with its type A uncertainty. The dotted line shows the random accuracy. As previously found [26], Sinc-EEGNet-32 is the most effective configuration architecture, even as the seed varies. This result was confirmed by the Kruskal-Wallis test (see Tab. 3).
Fig. 2 shows the results obtained for each configuration (i.e., EEGNet-4, EEGNet-32, Sinc-EEGNet-4, and Sinc-EEGNet-32) when varying the seed. In particular, each point represents the mean classification accuracy across the nine subjects of the dataset together with its type A uncertainty. The dotted line shows the random accuracy. As previously found [26], Sinc-EEGNet-32 is the most effective configuration architecture, even as the seed varies. This result was confirmed by the Kruskal-Wallis

10.3217/978-3-99161-014-4-009

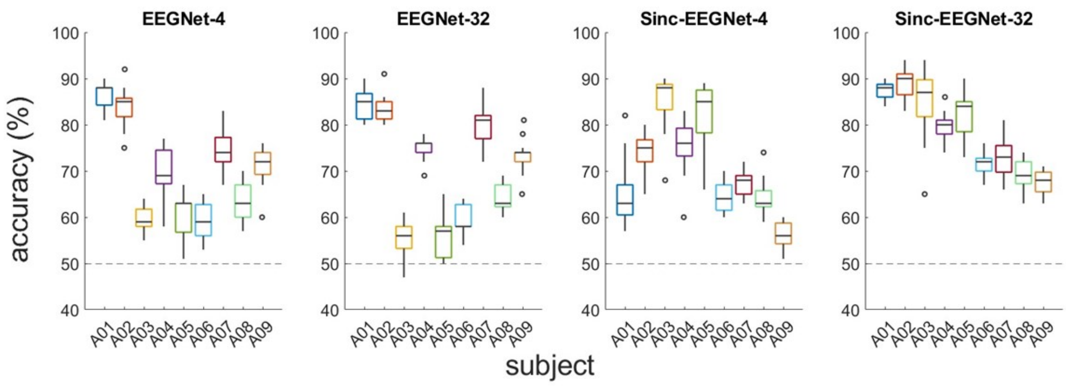

Figure 3: For each configuration, the results obtained for each subject with seed variation are presented in box plots, each containing the 15 associated classification accuracy values for each seed. The dotted line refers to the random accuracy.

	min (%)	max (%)
EEGNet-4	67 ± 4	72 ± 4
EEGNet-32	68 ± 4	73 ± 4
Sinc-EEGNet-4	67 ± 3	73 ± 3
Sinc-EEGNet-32	76 ± 3	81 ± 3

Table 2: Minimum and maximum classification accuracies across seed values and for each configuration. Uncertainties associated with these mean accuracies, estimated by using the standard deviation of the mean, are reported too.

Group A	Group B	p-value
EEGNet-4	EEGNet-32	0.966
EEGNet-4	Sinc-EEGNet-4	0.955
EEGNet-4	Sinc-EEGNet-32	< 0.001
EEGNet-32	Sinc-EEGNet-4	0.100
EEGNet-32	Sinc-EEGNet-32	< 0.001
Sinc-EEGNet-4	Sinc-EEGNet-32	< 0.001

Table 3: Results of post-hoc analysis with Kruskal-Wallis test.

test as extensively reported in Tab. 3.

It is important to stress that, for a specific architecture, the performance can vary significantly with different initialisation. Tab. 2 displays the minimum and maximum results obtained as the seed varies for each configuration. This analysis demonstrates the importance of testing the model's performance with different seeds before effectively using them. For example, in [26] all analyses were performed with a fixed seed of 42, which was found to be the worst case for most of the configurations. However, once the seed was fixed, the accuracy resulted repeatable.

Fig. 3 also shows the results obtained for different seeds, subject by subject. Each box plot represents the 15 classification accuracies, while the dotted line indicates the random classification accuracy reference. Although results are consistent among different architectures for some subjects, there is significant variability in others. For instance, subject A09 displays several outliers when EEGNet is employed, resulting in performance differences of over 15 % when just varying the seed. Next, it can be noted that the zero calibration scenario changes the discrimination between "good" and "bad" subjects with respect to previous literature evidence [40]. For instance, for the notoriously good subject A03, Sinc-EEGNet model leads to a good performance but the EEGNet model does not. In other case, like for A08, performance is relatively low with the proposed approaches, while other literature approaches led to higher performance. In this regard, it is important to recall that the results were obtained by training the models on data from other subjects, which could be associated with a non-compatible probability distribution.

Pre-processing impact: Fig. 4 shows the variations in models' performance by varying the pre-processing for each configuration. The networks were initialised with a random seed (seed = 0), but it is worth noting that compatible results were obtained when the best seed from the previous step was selected. Each box plot contains the results obtained for the nine subjects. The dotted line indicates the random classification accuracy. As usually proposed, one type of pre-processing has been used for each experiment [12]. This also facilitates online classification in terms of computational effort. The models performance resulted reduced by applying a band-pass filter. This is in contrast to what is observed by using classical machine techniques, where filtering the data trial by trial is often recommended [15, 24, 30]. The application of the Laplacian filter, instead, led to compatible performance than the "no pre-processing" case. Hence, this evidence suggests that pre-processing can be avoided. It is worth noting that EEGNet was originally proposed with a band-pass filter [25], whereas Sinc-EEGNet was proposed without any kind of pre-processing [26].

CONCLUSION

Deep learning has attracted more and more attention in the processing of EEG data for motor imagery-based BCIs, and transfer learning promises to improve classification accuracy while reducing the calibration burden. This would disclose a very large use of such BCI technologies in practice. However, literature results are still quite variegated and uncertainty sources are not dominated yet.

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-009

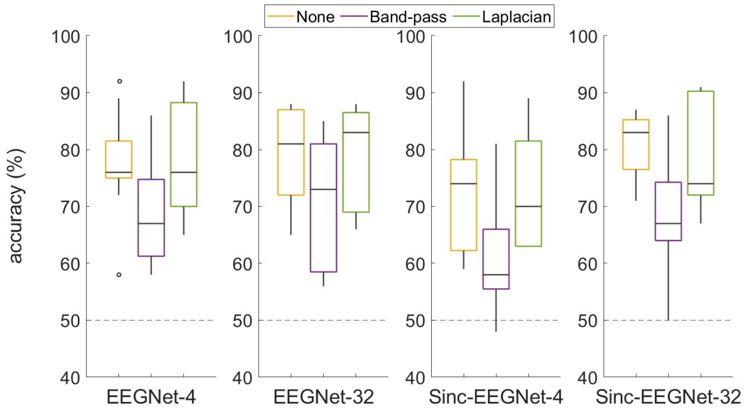

Figure 4: For each configuration, the box plots contains the results obtained using different pre-processing. The dotted line refers to the random accuracy.

To make a further step towards the repeatability and reproducibility of deep neural network results, a comparative analysis of the performance has been proposed for two relevant networks, namely EEGNet and Sinc-EEGNet, when varying fundamental settings. All the analyses were carried out under a zero-calibration approach, hence exploiting the leave-one-subject-out technique on the benchmark dataset *BCI competition IV, 2a* limited to two motor imagery tasks.

It was found that the performance varies significantly with different weights initialisation. Then, by analysing models' performance for varying pre-processing strategies, the performance of the deep network models was unchanged by Laplacian filter and even reduced by band-pass filtering. This finding is in contrast to what is observed by using classical machine learning, where filtering the data is recommended to improve classification results. This outcome suggests the use of pre-trained deep architectures on a new subject without the need for any preliminary data processing. This would reduce the pre-processing time of the data and foster an online classification. Overall, it is worth noting that Sinc-EEGNet-32 resulted the most effective architecture in accordance with previous studies.

This preliminary analysis emphasised the importance of carefully investigating the variability of deep neural networks adopted in BCI. Nonetheless, future works will deal with extending these analyses to more dataset to collect more evidence. Moreover, an experimental plan will be designed for a more comprehensive uncertainty assessment as the main networks settings vary.

ACKNOWLEDGEMENTS

This work was financially supported by Italian Ministry for Universities and Research (MUR) through the project "AGE-IT - Ageing Well in an ageing society: A novel public-private alliance to generate socioeconomic, biomedical and technological solutions for an inclusive Italian ageing society (Spoke: Ageing and clinical practice)", PNRR PE15, CUP E63C22002050006.

REFERENCES

[1] Lotze M, Halsband U. Motor imagery. Journal of Physiology-paris. 2006;99(4-6):386–395.

[2] Lin BS, Pan JS, Chu TY, Lin BS. Development of a wearable motor-imagery-based brain–computer interface. Journal of medical systems. 2016;40:1–8.

[3] Arpaia P *et al.* Paving the way for motor imagery-based tele-rehabilitation through a fully wearable bci system. Sensors. 2023;23(13):5836.

[4] Yadav D, Yadav S, Veer K. A comprehensive assessment of brain computer interfaces: Recent trends and challenges. Journal of Neuroscience Methods. 2020;346:108918.

[5] Pfurtscheller G, Lopes da Silva F. Handbook of Electroencephalography and Clinical Neurophysiology – Event-related desynchronization. Elsevier, Amsterdam, Netherlands (1999).

[6] Paszkiel S. Control based on brain-computer interface technology for video-gaming with virtual reality techniques. Journal of Automation, Mobile Robotics and Intelligent Systems. 2016:3–7.

[7] Chaudhary U, Birbaumer N, Ramos-Murguialday A. Brain–computer interfaces for communication and rehabilitation. Nature Reviews Neurology. 2016;12(9):513–525.

[8] Tong Y *et al.* Motor imagery-based rehabilitation: Potential neural correlates and clinical application for functional recovery of motor deficits after stroke. Aging and disease. 2017;8(3):364.

[9] Singh A, Hussain AA, Lal S, Guesgen HW. A comprehensive review on critical issues and possible solutions of motor imagery based electroencephalography brain-computer interface. Sensors. 2021;21(6):2173.

[10] Huang X *et al.* A review on signal processing approaches to reduce calibration time in eeg-based brain–computer interface. Frontiers in Neuroscience. 2021;15:733546.

[11] Padfield N, Zabalza J, Zhao H, Masero V, Ren J. Eeg-based brain-computer interfaces using

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-009

motor-imagery: Techniques and challenges. Sensors. 2019;19(6):1423.

[12] Arpaia P, Esposito A, Natalizio A, Parvis M. How to successfully classify eeg in motor imagery bci: A metrological analysis of the state of the art. Journal of Neural Engineering. 2022;19(3):031002.

[13] Saha S, Baumert M. Intra-and inter-subject variability in eeg-based sensorimotor brain computer interface: A review. Frontiers in computational neuroscience. 2020;13:87.

[14] Al-Saegh A, Dawwd SA, Abdul-Jabbar JM. Deep learning for motor imagery eeg-based classification: A review. Biomedical Signal Processing and Control. 2021;63:102172.

[15] Craik A, He Y, Contreras-Vidal JL. Deep learning for electroencephalogram (eeg) classification tasks: A review. Journal of neural engineering. 2019;16(3):031001.

[16] Zhang K *et al.* Application of transfer learning in eeg decoding based on brain-computer interfaces: A review. Sensors. 2020;20(21):6321.

[17] Ko W, Jeon E, Jeong S, Phyo J, Suk HI. A survey on deep learning-based short/zero-calibration approaches for eeg-based brain–computer interfaces. Frontiers in Human Neuroscience. 2021;15:643386.

[18] Apicella A, Isgrò F, Pollastro A, Prevete R. On the effects of data normalization for domain adaptation on eeg data. Engineering Applications of Artificial Intelligence. 2023;123:106205.

[19] Lotte F *et al.* A review of classification algorithms for eeg-based brain–computer interfaces: A 10 year update. Journal of neural engineering. 2018;15(3):031005.

[20] Yam JY, Chow TW. A weight initialization method for improving training speed in feedforward neural network. Neurocomputing. 2000;30(1-4):219–232.

[21] Cao W, Wang X, Ming Z, Gao J. A review on neural networks with random weights. Neurocomputing. 2018;275:278–287.

[22] Narkhede MV, Bartakke PP, Sutaone MS. A review on weight initialization strategies for neural networks. Artificial intelligence review. 2022;55(1):291–322.

[23] Pal KK, Sudeep K. Preprocessing for image classification by convolutional neural networks. In: 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). 2016, 1778–1781.

[24] Alzahab NA *et al.* Hybrid deep learning (hdl)-based brain-computer interface (bci) systems: A systematic review. Brain sciences. 2021;11(1):75.

[25] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. Eegnet: A compact convolutional neural network for eeg-based brain–computer interfaces. Journal of neural engineering. 2018;15(5):056013.

[26] Arpaia P *et al.* Sinc-eegnet for improving performance while reducing calibration of a motor imagery-based bci. In: 2023 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE). 2023, 1063–1068.

[27] Togha MM, Salehi MR, Abiri E. An improved version of local activities estimation to enhance motor imagery classification. Biomedical Signal Processing and Control. 2021;66:102485.

[28] Jeon E, Ko W, Yoon JS, Suk HI. Mutual information-driven subject-invariant and class-relevant deep representation learning in bci. IEEE Transactions on Neural Networks and Learning Systems. 2021.

[29] Cohen MX. Analyzing neural time series data: Theory and practice. MIT press (2014).

[30] Ang KK, Chin ZY, Wang C, Guan C, Zhang H. Filter bank common spatial pattern algorithm on bci competition iv datasets 2a and 2b. Frontiers in neuroscience. 2012;6:39.

[31] Özdenizci O, Wang Y, Koike-Akino T, Erdoğmuş D. Learning invariant representations from eeg via adversarial inference. IEEE access. 2020;8:27074–27085.

[32] Autthasan P *et al.* Min2net: End-to-end multi-task learning for subject-independent motor imagery eeg classification. IEEE Transactions on Biomedical Engineering. 2021;69(6):2105–2118.

[33] Ravanelli M, Bengio Y. Speaker recognition from raw waveform with sincnet. In: 2018 IEEE spoken language technology workshop (SLT). 2018, 1021–1028.

[34] Shimizu H, Srinivasan R. Improving classification and reconstruction of imagined images from eeg signals. Plos one. 2022;17(9):e0274847.

[35] Bria A, Marrocco C, Tortorella F. Sinc-based convolutional neural networks for eeg-bci-based motor imagery classification. In: International Conference on Pattern Recognition. 2021, 526–535.

[36] Chen J, Wang D, Yi W, Xu M, Tan X. Filter bank sinc-convolutional network with channel self-attention for high performance motor imagery decoding. Journal of Neural Engineering. 2023;20(2):026001.

[37] Kunjan S *et al.* The necessity of leave one subject out (loso) cross validation for eeg disease diagnosis. In: Brain Informatics: 14th International Conference, BI 2021, Virtual Event, September 17–19, 2021, Proceedings 14. 2021, 558–567.

[38] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision. 2015, 1026–1034.

[39] Brunner C, Leeb R, Müller-Putz G, Schlögl A, Pfurtscheller G. Bci competition 2008–graz data set a. Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology. 2008;16:1–6.

[40] Arpaia P, Esposito A, Moccaldi N, Natalizio A, Parvis M. Online processing for motor imagery-based brain-computer interfaces relying on eeg. In: 2023 IEEE International Instrumentation and Measurement Technology Conference (I2MTC). 2023, 01–06.

Published by
Verlag der Technischen Universität Graz