Motor Imagery EEG Classification with Convolutional Neural Networks: impact of reduced trials, channels and samples across different datasets

1st Giorgio Scavalli Veccia[1,2]
scavalliveccia.1858129@studenti.uniroma1.it

2nd Neil Mehta[1,3]
neilashm@andrew.cmu.edu

3rd Shay Englander Bendor[1]
shay.englander@tecnico.ulisboa.pt

4th Pulkit Grover[3]
pgrover@andrew.cmu.edu

5th Athanasios Vourvopoulos[1]
athanasios.vourvopoulos@tecnico.ulisboa.pt

6th Patrícia Figueiredo[1]
patricia.figueiredo@tecnico.ulisboa.pt

[1]Institute for Systems and Robotics and Department of Bioengineering, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal.
[2]Department of Computer Science, La Sapienza, Rome, Italy.
[3]Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, USA.

***Abstract*—The classification of EEG signals during motor imagery (MI) tasks is a key element of several brain-computer interfaces (BCIs), especially those used for motor rehabilitation of stroke patients. Despite a large body of literature, the classification of MI EEG signals remains challenging, with deep learning approaches showing improved performance but only limited success in real-life applications. This study focuses on two convolutional neural networks (CNNs) developed for classifying EEG signals, EEGNet and EDPNet, and aims to investigate the impact of different aspects of data degradation in real-life acquisitions on the accuracy of left vs. right MI classification. For this purpose, we consider a large public dataset with a well-defined MI task, as well as a private dataset with fewer trials across different types of MI tasks. We further test the impact of reducing the number of channels and time samples used, and compare the performance of subject-specific and group models. High accuracy was achieved using both CNNs in both datasets, with slightly better performance for subject-specific models compared to group models, highlighting high inter-subject variability. Interestingly, reducing the number of trials in the public dataset to match those in the private dataset yielded only a small decrease in performance, consistently with the similar performance obtained between datasets despite the difference in the number of trials. Both reducing the number of channels (from the total of 18 or 32 to only C3, C4 and Cz) and time samples, or trial duration (from 5.5 to 1.0 seconds) decreased performance, with the latter producing the greatest degradation. Nevertheless, even when using only 1.0 seconds of data from C3, C4 and Cz channels, model accuracy remained always above chance level, with a median of approximately 0.8 across subjects. In conclusion, we show that EEGNet and EDPNet are appropriate for MI EEG BCIs, yielding high left vs. right classification accuracies within a 1-second interval and with a limited number of channels.**

I. INTRODUCTION

Brain-Computer Interfaces (BCIs) are technologies that enable direct communication between the brain and external devices by bypassing conventional neuromuscular pathways [1]. Electroencephalography (EEG)-based BCIs offer a cost-effective, non-invasive approach with high temporal resolution, making them particularly suitable for applications such as neurorehabilitation and assistive technologies. Among various BCI paradigms, Motor Imagery (MI)-based BCIs hold particular promise, as they allow users to control external devices by mentally simulating movements without actual muscle activation [2]. The mental rehearsal movement modulates EEG signals primarily in the Alpha (8–13 Hz) and Beta (13–30 Hz) frequency bands. This capability is especially beneficial for individuals with motor impairments, enabling them to interact with their environment using brain activity alone [3].

Despite substantial progress in EEG-based BCIs, achieving high classification accuracy and robustness remains challenging due to inter-subject variability, signal non-stationarity, and noise contamination [4]–[6]. Traditionally, MI-BCI classification relied on conventional machine learning (ML) techniques, with Common Spatial Pattern (CSP) and Linear Discriminant Analysis (LDA) being among the most widely used approaches [7]. CSP enhances discriminability by computing spatial filters that maximize variance differences between MI classes, while LDA provides a computationally efficient linear separation of features. However, these methods are sensitive to noise and require extensive feature engineering.

Deep learning (DL) approaches have more recently been explored for this purpose [8], [9]. In particular, convolutional neural networks (CNNs) have been proposed that incorporate a temporal layer before a spatial one, enabling the extraction of discriminative temporal and spatial EEG features by training filters specific to each layer. Additionally, CNN models for EEG classification often include a summarization layer to efficiently represent information, keeping the model lightweight, followed by a final classification layer [10]. Although DL models have shown improved accuracy compared with conventional machine learning, their success in real-life BCI applications is still limited.

In this study, we focus on two CNNs that have been successfully employed for the classification of left vs. right MI from EEG signals, EEGNet [11] and EDPNet [12].

EEGNet is a well-known and widely reviewed network for EEG classification, while EDPNet is a new network featuring an interesting structure with attention mechanisms and dual prototype learning. Our aim is to investigate the impact of different aspects of data degradation in real-life acquisitions on model performance. For this purpose, we consider a large public dataset with a well-defined MI task, as well as a private dataset with fewer trials across different types of MI tasks. We further test the impact of reducing the number of channels and time samples used, and compare the performance of subject-specific and group models.

II. Methods

A. Datasets

1) Public Dataset: The BCI Competition IV dataset 2A was used, including EEG data recorded from 9 subjects while performing MI of the left hand, right hand, feet, and tongue [13]. EEG data were acquired from 20 channels with a sampling rate of 250 Hz, over a total of 576 MI trials. After removal of eye movement artifacts, the data were epoched around the onset of each MI trial, with a trial duration of 4.0 seconds immediately after the cue. Here, we focus only on the left and right hand MI trials, totaling 288 trials (144 for each hand / class) and the binary classification of left vs. right hand MI.

2) Private Dataset: EEG data were collected from 14 healthy participants (7 male, 7 female; mean age: 26 ± 6.25 years) during the execution of an MI grasping task [14]. The protocol was approved by the Ethics Committee of CHULN and CAML (Faculty of Medicine, University of Lisbon) and each participant provided written informed consent. The task involved imagination of right or left hand grasping over a total of 120 trials (60 trials per hand / class), across different conditions (including a VR environment or not). EEG data were acquired using a wearable wireless amplifier (LiveAmp; Brain Products GmBH, Germany) with 32 active electrodes at a sampling rate of 250 Hz. Electrodes were arranged according to the 10-20 EEG montage. EEG data were epoched around trial onsets to extract left and right hand MI trials with two different durations: 5.5 seconds, between 1.5 seconds before and 4 seconds after stimulus onset; and 1.0 seconds, from half a second before to half a second after the stimulus onset. The effect on classification of bandpass filtering the data between 1 and 40 Hz was also investigated. No other preprocessing was performed.

B. CNN architectures

1) EEGNet: The EEGNet architecture [11] was designed for EEG-based BCIs, structured in two main blocks. The first is composed by a temporal 2D convolutional filter that takes as input the downsampled EEG signal. This filter returns feature maps representing the EEG signal at different band-pass frequencies. It is composed also by a spatial filter, a Depthwise convolution that helps to understand how the signal varies across EEG channels. A depth parameter controls the number of spatial filters to learn for each feature map. After the first block, EEGNet applies Batch Normalization, followed by the ELU (Exponential Linear Unit) activation function, average pooling to reduce the signal dimension, and dropout. The second block consists of a separable convolution: a Depthwise Convolution that summarizes the temporal information in the feature maps to reduce the amount of data in the model. A Pointwise Convolution with size (1, 1) merges the feature maps obtained from the previous filter. At this point, Batch Normalization, activation, average pooling, and dropout are applied again. Finally, on the newly obtained features, EEGNet uses a Softmax function to assign a probability score to each feature for classification.

2) EDPNet: The EDPNET architecture [12] was designed for EEG MI classification, performing well with small datasets, low computational cost, and fast training time. EDPNET consists of four modules: spatial-spectral embedding, adaptive spatial-spectral fusion, multi-scale variance pooling, and dual prototype learning. The first module takes a vector of size channels * times as input, divides the data into h groups along the channel dimension and applies a Light Convolution to extract temporal and spectral features. The adaptive spatial-spectral fusion module is based on a lightweight attention mechanism consisting of three parts: global context embedding, channel normalization, and gating adaptation.

A mean-variance operation aggregates temporal information from each channel, α controls the weight of each channel. Channel normalization is applied. The attention mechanism, uses attention weights γ and bias β and it determines which features are most relevant for each channel. The adaptive spatial-spectral fusion also employs a pointwise 1D convolution filter of size 1x1 across all channels to extract spatial features. The multi-scale variance pooling module provides a more compact representation of the features. It first splits the input along the first dimension into three groups, applying variance pooling with three kernel sizes to summarize the features for each group. The final step of the network is Dual Prototype Learning (DPL), a technique that builds two prototypes: the inter-class separation prototype. This prototype minimizes the distance to the correct class and maximizes the distance to other classes, thereby increasing the margin between classes.

C. Model training and validation

For each CNN architecture, dataset, and scenario (described below), two types of models were trained. Subject-specific models were obtained by training on all train trials of each individual subject and testing on test trials. Group models were obtained by training on all trials of all subjects except one used for the test, using a leave-one-out scheme.

For each model, the training involves two phases: an initial phase with a 50/50 training-validation split for N_1 epochs using early stopping after N_e epochs without improvement, and a second phase training on the combined data for N_2 epochs, following the strategy used by the authors of EDPNet [12]. The hyperparameters of the EEGNet and EDPNet architectures and the respective model training and validation are presented in Tables I and II.

EEGNet uses a single learning rate and weight decay, as it does not rely on prototype losses. Its hyperparameters include the dropout rate to prevent overfitting, and the number of temporal, spatial, and pointwise filters, as well as the window lengths for temporal and spatial filters. In the public data set for individual subject technique, the learning rate used to train the networks is increased to 0.001.* The private dataset requires a lower learning rate to achieve good test accuracy results. Although it contains fewer trials, it provides more channels, enabling both networks to classify its signals more effectively compared to the other datasets.

The hyperparameters of EDPNet include the number of convolutional filters in the first (F_1) and second layers (F_2), the temporal filter window size, the pooling window size in the multi-scale variance pooling module, and the coefficients α and λ for ISP and ICP prototypes. Three learning rates and weight decay parameters are used for feature extraction, ICP prototype learning, and ISP prototype learning.

TABLE I: Hyperparameters used for the public dataset.

EEGNet	**EDPNet**
General Parameters	
Dropout rate = 0.2	1st filter F_1 = 9
Kernel length = 64	2nd filter F_2 = 48
Temporal filters F_1 = 8	Time kernel 1 = 75
Spatial filter D = 2	Pool kernels = [50, 100, 200]
Pointwise filters F_2 ($F_1 * D$)=16	α = 0.00001
Kernel length2 = 16	λ = 0.001
Learning Rates and Decay	
Learning rate = 0.0001 *	Learning rate = 0.0001 *
Weight decay = 0.01	Weight decay = 0.01
	Learning rate ISP = 0.0001
	Learning rate ICP = 0.0001
	Weight decay ISP = 0
	Weight decay ICP = 0
Training Parameters	
N_1 = 1000	N_1 = 1000
N_e = 200	N_e = 200
N_2 = 300	N_2 = 300
Batch size = 32	Batch size = 32
Validation Parameters	
Validation division size = 0.5	Validation division size = 0.5

TABLE II: Hyperparameters changed for the private dataset.

EEGNet	**EDPNet**
Learning Rates and Decay	
Learning rate = 0.0001	Learning rate = 0.0001
Training Parameters	
N_1 = 300	N_1 = 300
N_e = 150	N_e = 150
N_2 = 200	N_2 = 200
Batch size = 32	Batch size = 32

D. Model scenarios

For each model architecture, the following modeling scenarios were considered. In each case, model performance was assessed as the accuracy of the binary classification between left and right hand MI trials.

For the public dataset, which contained a larger number of MI trials (288), we investigated the effect of reducing the number of trials by considering only a subset of 120 trials (60 trials per class), to match the number of trials in the private dataset. This test was important given the typically low numbers of trials of real-life datasets, which may severely impair the performance of DL models.

For the private dataset, we investigated the effect of reducing the number of time samples in each trial used for classification, by considering a longer trial duration (5.5 seconds) as well as a shorter trial duration (1.0 seconds). This is particularly important for real-life neurorehabilitation BCI applications, in which the timing of the neurofeedback received by the patient is a critical parameter.

For both the public and private datasets, we investigated the effect of using only 3 EEG channels close to the motor cortex: C3, C4 and Cz, where there is strong directional connectivity from Cz to C3/C4 during left- and right-hand MIs [15]. A reduced number of channels would be highly beneficial for the practical implementation of BCIs, since it would considerably reduced preparation time.

Finally, we also considered a preprocessing the EEG data with a bandpass filter focused on the frequency bands of interest for MI brain activity, i.e., 8 - 40 Hz. Restricting the EEG frequencies analyzed may help minimize artifact contributions, occurring mostly at lower and higher frequencies.

III. Results

A. Model performance in the default scenario

The test accuracies achieved for the subject and group models, obtained using the EEGNet and EDPNet architectures in the default scenario are presented in Tables III and IV, for the public and private datasets, respectively. High accuracy was achieved using both CNNs in all cases (between 73 and 100% across all subjects of both datasets). For the public dataset, EDPNet yielded a higher accuracy than EEGNet in subject-specific models (92 vs. 85 %). As expected due to the known inter-subject variability of MI EEG signals, the performance was lower for group models. However, it was not significantly different between CNNs in this case (83 vs. 84 %). When moving to the private dataset, somewhat unexpectedly, we obtained slightly better performance in general. Group models performed only slightly worse than subject-specific models, probably due to the greater number of subjects in this data set (14 vs. 9). A small improvement was achieved with EDPNet relative to EEGNet, for both subject models (95 vs. 94 %) and group models (93 vs. 92 %). The model interpretability reveals that the temporal filters of both networks exhibit a sinusoidal weight distribution over time, with key frequencies in the 4–8 Hz and 14–18 Hz ranges. In contrast, the spatial patterns are less consistent, as the weight distribution across channels varies significantly between filters and subjects.

TABLE III: Model test performance for the public dataset (left vs. right hand classification): individual subject model (top) and group model (bottom), in the default scenario, using all trials and channels, and no bandpass filtering.

Subject	EDPNet (%)	EEGNet (%)
1	95.83	77.08
2	85.41	77.08
3	97.91	81.94
4	90.97	79.16
5	96.52	88.19
6	77.08	83.33
7	84.72	88.19
8	97.22	88.19
9	95.83	88.88
AVG ± Dev std	91.28 ± 6.89	83.56 ± 4.69

Subject	EDPNet (%)	EEGNet (%)
1	82.63	74.65
2	76.04	76.73
3	82.98	88.19
4	82.29	78.81
5	84.02	90.62
6	80.55	82.63
7	90.62	93.05
8	91.66	96.18
9	78.12	78.81
AVG ± Dev std	83.21 ± 4.86	84.41 ± 7.34

TABLE IV: Model test performance for the private dataset (left vs. right hand classification): individual subject model (top) and group model (bottom), in the default scenario, using all time samples (5.5 seconds trial duration) and channels, and no bandpass filtering.

Subject	EDPNet (%)	EEGNet (%)
1	93.33	85.0
2	100.0	100.0
3	71.66	71.66
4	98.33	98.33
5	91.66	96.66
6	100.0	98.33
7	100.0	98.33
8	96.66	96.66
9	96.66	93.33
10	90.0	86.66
11	96.66	96.66
12	100.0	100.0
13	100.0	100.0
14	88.33	91.66
AVG ± Dev std	94.52 ± 7.41	93.8 ± 7.67

Subject	EDPNet (%)	EEGNet (%)
1	86.66	84.16
2	100.0	99.16
3	82.5	84.16
4	99.16	100.0
5	89.76	90.55
6	100.0	100.0
7	97.5	97.5
8	98.31	98.31
9	94.16	90.0
10	82.5	78.33
11	83.33	89.16
12	100.0	100.0
13	98.33	100.0
14	84.16	81.66
AVG ± Dev std	92.60 ± 7.08	92.35 ± 7.58

B. Effect of deteriorating scenarios on model performance

The results of the models obtained considering the scenarios corresponding to real-life deteriorating conditions are presented in Figures 1 and 2, for the public and private datasets, respectively.

In the public dataset, we verified that reducing the number of trials by more than one half (from 288 to 120) to match the private dataset yielded only a small decrease in performance, noticeable only for EEGNet with all channels. This is consistent with the high performance obtained with the private dataset, indicating that the limited number of trials was not detrimental to model performance in this case. An appreciable impairment of performance was observed when considering only 3 out of the total 18 channels, with the accuracy dropping to around 70 and 60 % on average across subjects for the subject and group models, respectively. This effect was particularly pronounced in the group models. Interestingly, bandpass filtering produced only a negligible decrease in accuracy.

In the private dataset, reducing the number of channels from 32 to 3 impaired performance slightly less than in the public dataset, while bandpass filtering impaired it negligibly similarly to the public dataset. The most signifcant impairment was caused by the reduction in time samples used, with 1.0 sec trials yields significantly lower accuracies than 5.5 sec trials. Nevertheless, even when using only 1.0 seconds of data from only C3, C4 and Cz channels, model accuracy remained always above chance level, with a median of approximately 0.8 across subjects for both EDPNet and EEGNet.

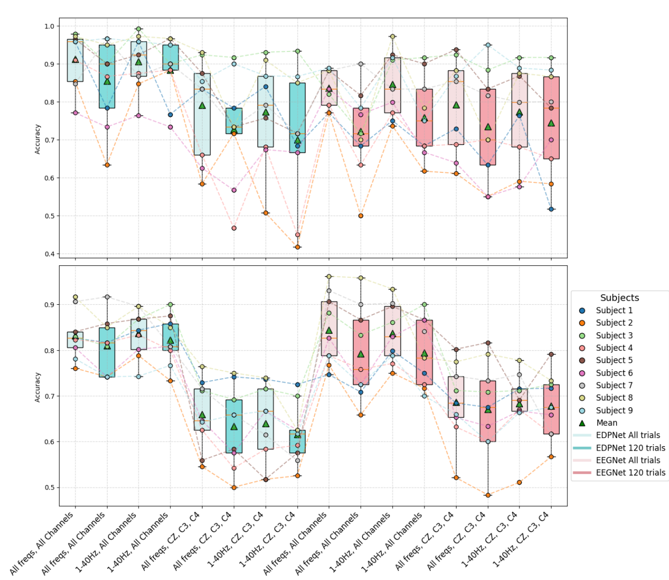

Fig. 1: Model test performance for the public dataset (left vs. right hand classification): individual subject model (top) and group model (bottom), in the different scenarios tested.

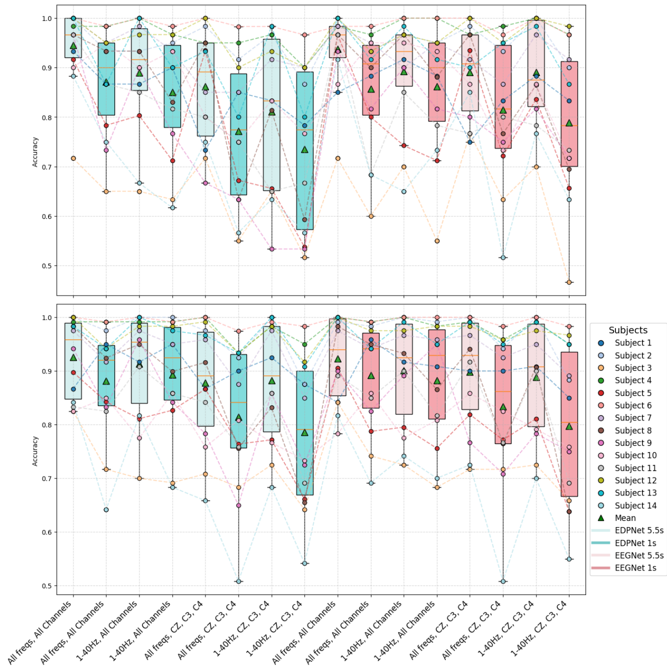

Fig. 2: Model test performance for the private dataset (left vs. right hand classification): individual subject model (top) and group model (bottom), in the different scenarios tested.

IV. Discussion

In this study, we explored DL models for classifying MI EEG signals in BCIs to help individuals with motor disabilities interact with devices through brain signals. This study evaluated generalization across datasets recorded with varying criteria using two reviewed architectures, EDPNet and EEGNet. Both CNNs achieved high accuracy, surpassing conventional machine learning methods like CSP combined with LDA. While EDPNet outperformed EEGNet in subject-specific models of the public dataset, as noted in [12], this advantage was not observed in the private dataset or group models in either dataset. Future improvements could involve Bayesian hyperparameter optimization.

Models trained on individual subjects generally performed better than those trained on groups, due to significant variability in EEG patterns during MI tasks across subjects. However, this gap was smaller in the private dataset, which had more subjects, suggesting that accurate group models might be achievable with sufficiently large cohorts.

Reducing the number of trials by more than one half yielded only a small if any decrease in performance. This is consistent with the high performance obtained with the private dataset, with fewer trials, indicating that this was not detrimental to model performance in this case.

Bandpass filtering to a specific frequency band did not notably improve performance, showing that EEGNet and EDPNet temporal filters can effectively identify key frequencies from raw EEG data. However, reducing the number of channels did affect performance, suggesting these models extract important spatial information beyond motor cortex areas (C3, C4, Cz). Notably, shortening EEG segments to 1 second degraded performance, highlighting the challenge of reducing feedback delay in MI EEG BCIs.

In conclusion, we show that EEGNet and EDPNet are appropriate for MI EEG BCIs, yielding high left vs. right classification accuracies, which are still well above chance level within a 1-second interval and with a limited number of channels.

Acknowledgments

This work is supported by FCT projects LARSyS (DOI: 10.54499/LA/P/0083/2020, 10.54499/UIDP/50009/2020, and 10.54499/UIDB/50009/2020) and NOISyS (DOI: 10.54499/2022.02283.PTDC), and PRR Center for Responsible AI (C645008882-00000055).

References

[1] G. Pfurtscheller, D. Flotzinger, and J. Kalcher, "Brain-computer interface—a new communication device for handicapped persons," *Journal of microcomputer applications*, vol. 16, no. 3, pp. 293–299, 1993.

[2] A. Vourvopoulos, C. Jorge, R. Abreu, P. Figueiredo, J.-C. Fernandes, and S. Bermúdez i Badia, "Efficacy and brain imaging correlates of an immersive motor imagery bci-driven vr system for upper limb motor rehabilitation: A clinical case report," *Frontiers in Human Neuroscience*, vol. 13, 2019.

[3] T. Cattai, S. Colonnese, M.-C. Corsi, D. S. Bassett, G. Scarano, and F. De Vico Fallani, "Phase/amplitude synchronization of brain signals during motor imagery bci tasks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 1168–1177, 2021.

[4] W. Samek, C. Vidaurre, K.-R. Müller, and M. Kawanabe, "Stationary common spatial patterns for brain–computer interfacing," *Journal of neural engineering*, vol. 9, no. 2, 2012.

[5] S. Saha and M. Baumert, "Intra- and inter-subject variability in eeg-based sensorimotor brain computer interface: A review," *Frontiers in Computational Neuroscience*, vol. 13, 2020.

[6] M. C. Meyer, E. S. V. Oort, and M. Barth, "Electrophysiological correlation patterns of resting state networks in single subjects: a combined eeg–fmri study," *Brain Topography*, vol. 26, no. 1, pp. 98–109, 2012.

[7] L. Fabien, B. Laurent, and C. Maureen, "Electroencephalography (eeg)-based brain-computer interfaces," *Wiley encyclopedia of electrical and electronics engineering*, p. 44, 2015.

[8] A. Vavoulis, P. Figueiredo, and A. Vourvopoulos, "A review of online classification performance in motor imagery-based brain–computer interfaces for stroke neurorehabilitation," *Signals*, vol. 4, no. 1, pp. 73–86, 2023.

[9] Q. Xue, Y. Song, H. Wu, Y. Cheng, and H. Pan, "Graph neural network based on brain inspired forward-forward mechanism for motor imagery classification in brain-computer interfaces," *Frontiers in Neuroscience*, vol. 18, p. 1309594, Mar. 2024.

[10] A. Al-Saegh, "Deep learning for motor imagery eeg-based classification: A review," *Artificial Intelligence in Medicine*, 2021.

[11] V. J. Lawhern, "Eegnet: A compact convolutional neural network for eeg-based brain-computer interfaces," *Journal of Neural Engineering*, 2018.

[12] C. Han, "A spatial-spectral and temporal dual prototype network for motor imagery brain-computer interface," *IEEE Transactions on Biomedical Engineering*, 2024.

[13] C. Brunner, "Bci competition 2008 graz data set a," 2008.

[14] S. Englander, D. Esteves, M. Valente, P. Figueiredo, and A. Vourvopoulos, "Understanding the impact of true vs. positive vr feedback on eeg features and bci performance," in *2025 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR)*, pp. 90–97, 2025.

[15] S. Hu, H. Wang, J. Zhang, W. Kong, Y. Cao, and R. Kozma, "Comparison analysis: Granger causality and new causality and their applications to motor imagery," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 7, pp. 1429–1444, 2016.