Improving Generalization of CNN-Based Motor-Imagery EEG Decoders via Dynamic Convolutions

Konstantinos Barmpas, Yannis Panagakis, *Member, IEEE*, Stylianos Bakas, Dimitrios A. Adamos, *Member, IEEE*, Nikolaos Laskaris, and Stefanos Zafeiriou, *Member, IEEE*

***Abstract*— Deep Convolutional Neural Networks (CNNs) have recently demonstrated impressive results in electroencephalogram (EEG) decoding for several Brain-Computer Interface (BCI) paradigms, including Motor-Imagery (MI). However, neurophysiological processes underpinning EEG signals vary across subjects causing covariate shifts in data distributions and hence hindering the generalization of deep models across subjects. In this paper, we aim to address the challenge of inter-subject variability in MI. To this end, we employ causal reasoning to characterize all possible distribution shifts in the MI task and propose a dynamic convolution framework to account for shifts caused by the inter-subject variability. Using publicly available MI datasets, we demonstrate improved generalization performance (up to 5%) across subjects in various MI tasks for four well-established deep architectures.**

***Index Terms*— Brain-computer interfaces (BCIs), causality, motor-imagery (MI), electroencephalogram (EEG).**

I. Introduction

BRAIN-COMPUTER Interface (BCI) technology primarily aspires to provide neural communication and control between a user and a machine bypassing the normal neuromuscular pathways. This is feasible by analyzing brainwaves captured by electroencephalogram (EEG) signal recordings using signal processing and Machine Learning (ML) techniques. Nowadays, BCIs find application in various areas, including emotion recognition (e.g. [2], [3]), epileptic seizure detection (e.g. [4], [5]), robotic control [6] as well as video gaming [7].

Manuscript received 28 November 2022; revised 10 February 2023 and 20 March 2023; accepted 28 March 2023. Date of publication 6 April 2023; date of current version 14 April 2023. An earlier version of this paper was presented at the ICLR2022 Workshop on the Elements of Reasoning: Objects, Structure and Causality [1]. *(Corresponding author: Konstantinos Barmpas.)*

Konstantinos Barmpas, Dimitrios A. Adamos, and Stefanos Zafeiriou are with the Department of Computing, Imperial College London, SW7 2RH London, U.K., and also with Cogitat Ltd., W12 0BZ London, U.K. (e-mail: konstantinos.barmpas16@imperial.ac.uk).

Yannis Panagakis is with the Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, 15784 Athens, Greece, and also with Cogitat Ltd., W12 0BZ London, U.K.

Stylianos Bakas and Nikolaos Laskaris are with the School of Informatics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece, and also with Cogitat Ltd., W12 0BZ London, U.K.

Digital Object Identifier 10.1109/TNSRE.2023.3265304

One of the first and most popular BCI paradigms is Motor-Imagery (MI). MI-BCIs are based on a neural process, by which a subject mentally simulates a motor action, for example the movement of a hand or foot, without actually executing it [8]. Developing MI-BCI systems (e.g. [9], [10]) mainly relies on robust decoding of a subject's motor intentions from the recorded EEG signals, under the prior assumption that these signals encode that relevant information, and are mainly used for movement rehabilitation purposes (e.g. [11], [12], [13], [14], [15], [16]) as well as the wheelchair/exoskeleton control [17].

Several works have addressed the problem of EEG-based motor-imagery classification using classical feature extraction techniques [18]. The technique of common spatial pattern (CSP) algorithm [19] and its various extensions, like the Filter-Bank CSP (FBCSP) [20], are among the most popular methods of this category due to their simplicity in design and computational efficiency in implementation. In all these methods, specific band-pass filters are applied to the EEG signals prior to the design of spatial filters, sacrificing flexibility and adaptivity to some extend.

In recent years, Deep Learning (DL) techniques - and most specifically Convolutional Neural Networks (CNNs) - have largely alleviated the need for manual feature extraction, achieving state-of-the-art performance in various areas, most notably computer vision [21]. Due to their massive progress, CNN-based feature extractors have been introduced in various paradigms in the field of BCIs (e.g. [22], [23], [24], [25]), in an effort to become generic EEG signal processing tools compared to classical feature extraction techniques (e.g. [18], [19], [20]). DeepConvNet and ShallowConvNet [26] are among the first deep learning architectures employed in MI-BCIs and are inspired by common spatial pattern (CSP) filters [19] since they include convolutions across time followed by convolutions across EEG channels. EEGNet [27] is a lightweight BCI architecture which consists of a compound of temporal and spatial filtering inspired by the filter bank common spatial pattern (FBCSP) technique [20]. EEG-Inception [28] shares the exact same fundamentals with EEGNet and has strong performance results across different benchmarks. Although it is similar to EEGNet, it includes several Inception branches,

originally introduced in [29]. These branches consist of trainable convolutional temporal filters of different scales, capturing several temporal modulations of the EEG signals.

Although these deep learning architectures are inspired by classical EEG feature extraction techniques and achieve impressive performance in MI classification tasks, they usually fail to tackle the problem of inter-subject variability [30], preventing the successful deployment of a previously trained MI classifier to new unseen subjects. Inter-subject variability is defined as the change in data distributions across different subjects: each individual has a unique brain anatomy and functionality that makes the discovery and exploitation of shared invariant features extremely difficult. In fact, these differences are so distinct that previous works have shown that the identification of a specific subject out-of-many is actually feasible (e.g. [31], [32], [33]). Therefore, modern DL-based BCIs tend to fail to generalize well in unseen subjects due to this type of data distribution shift. For many years, normalization techniques (e.g. [34], [35]) - data scaling using a mean and standard deviation - in conjunction with classical machine learning techniques have been considered the gold standard to solve the problem of inter-subject variability. With the advent of deep learning, methods like transfer learning have emerged in an effort to provide a solution (e.g. [36], [37], [38], [39]). In most of these methods, a small calibration set from the unseen subject is utilized to fine-tune parts of the pre-trained deep network architecture. In [39] only the last fully-connected layers are fine-tuned while the previous layers are frozen. In [36] some identified layers are fine-tuned to maximize knowledge transfer for MI classification. Although transfer learning has been proven to perform well, it still requires a calibration session in order to generalize well to unseen subjects. In the direction of zero-calibration networks, [40] proposes an adversarial inference framework that learns subject invariant features. In this work, we aspire to provide an alternative solution to the problem of inter-subject variability and enhance the above mentioned BCI deep architectures dynamically without the need of a calibration session.

Causal reasoning provides tools to breakdown and analyze important aspects of a BCI task, identify and possibly resolve some of these challenges by employing appropriate ML strategies. The methodical breakdown of a BCI task and the identification of the causal relationships between the various variables of interest take into account the expert's knowledge of the involved biological and neurophysiological processes and can be of vital importance when designing and building ML-based models in the field of BCIs. In this work, we focus mainly on MI-BCI systems, and inspired by the work of [41], we analyze the task of MI EEG signal classification through the lens of causal reasoning. Motivated by this causal analysis, we introduce a framework based on dynamic convolutions that provably tackles the identified problem of data distribution shift across subjects.

Our contributions can be summarized as follows:

1) We employ causal reasoning to breakdown and analyze important challenges / distribution shifts in the task of MI brainwave decoding
2) We propose a subject attention network based on learnable Gabor wavelets that can accurately identify the different available subjects
3) Inspired by [42], we propose a framework based on dynamic convolutions that utilizes our proposed subject attention network and with zero calibration provably tackles the issue of inter-subject variability in the task of MI brainwave decoding according to our proposed causal breakdown. More specifically, our causal analysis allows us to design an evaluation setup which keeps all the identified distribution shifts intact but the inter-subject variability. Therefore, unlike other works in the area which claim improved cross-subject performance and often utilize a mixture of techniques like data augmentation (which can affect also other causal variables of interest), our work is theoretically proven to target the problem of inter-subject variability through this specifically crafted evaluation setup.

The remainder of the paper is organized as follows: Section II describes our causal analysis to breakdown important challenges / distribution shifts in the task of MI brainwave decoding. Section III outlines our proposed framework based on dynamic convolutions that improves the generalization of MI-BCI systems. Section IV consists of the experimental part, where performance results and comparisons are detailedly presented. Section V acts as a discussion part to demonstrate the advantages and disadvantages of our proposed framework. The last section summarizes and concludes our work and briefly outlines future research steps.

II. Characterizing Distribution Shifts in Motor-Imagery (MI) Decoding Using Causal Reasoning

The main goal of this paper is to propose a framework that tackles the issue of inter-subject variability in CNN-based BCI models. To achieve this, we will first investigate the problem of MI brainwave decoding through the lens of causal reasoning. As it has been demonstrated in [41], causal models encode naturally more information which can be vital in the machine learning design process and if appropriately used can lead to models which are more robust to certain types of distribution shifts. But why is this causal analysis important in this work and for the proposed framework? By performing this causal breakdown, we can identify most of the possible distribution shifts that can be met in the task of MI classification. By associating the inter-subject variability to a distribution shift in one of the core variables of interest, we can design an evaluation setup which keeps all the identified challenges intact but the inter-subject variability. Therefore, we can certainly claim that our framework specifically contributes in solving the targeted problem.

A. Preliminaries

Causal reasoning is the analysis of a task / problem in terms of cause-effect relationships between the different variables of interest: if a variable A is a direct cause of variable B, we express it as $A \rightarrow B$ (A causes B or B is the effect

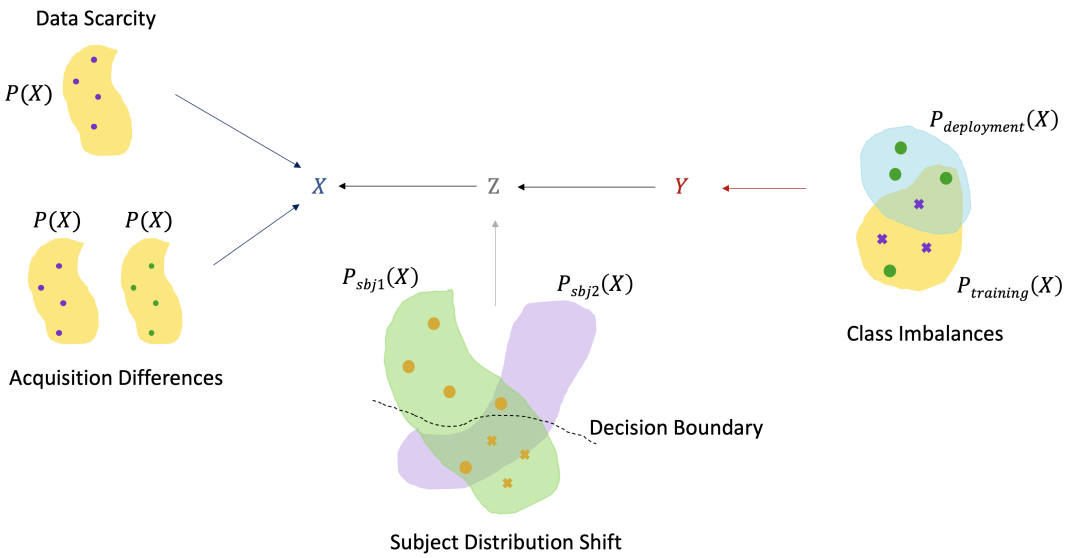

Fig. 1. Key challenges in machine learning for a MI EEG classification task. X represents input EEG signals, Z the true unobserved brain activity, Y the associated MI labels. • and × represent EEG signals of different labels. Dots represent data points of any label and their color represent different EEG acquisition devices.

of A). When designing a machine learning algorithm, it is crucial to understand all the involved factors as well as their causal relationships. A causal breakdown of a system can be represented as a directed acyclic graph (DAG) where the nodes are the variables of interests and the edges represent direct causal relationships. These diagrams can capture vital information for the involved variables of interests such as conditional dependencies as well as independencies.

B. Causality in Motor-Imagery Decoding

In a MI classification problem, we want to accurately predict the mentally performed task from a recorded EEG signal. Mathematically, given an input EEG signal X, we train a statistical model to predict the correct MI task Y, which can be the imagery movement e.g. of a hand or foot. In essence, this statistical model tries to estimate the conditional probability $P(Y|X)$ using an appropriate objective function.

In machine learning tasks, given the input X and the prediction target Y, we can establish that the task to estimate $P(Y|X)$ can be either [43]:

- Causal: when $X \rightarrow Y$, predict effect from cause
- Anti-causal: when $Y \rightarrow X$, predict cause from effect

Using the above basis, we can define an MI EEG classification task as an anti-causal problem, since the true MI intention (observed with the MI label Y) can be considered the cause of the recorded EEG signal X. Additionally, inspired by [43], we can consider X as a sequence of imperfect observed measurements of the true unobserved brain activity Z within, mainly, the cortical areas responsible for the sensorimotor rhythms, i.e. $Z \rightarrow X$. Therefore, using a causal diagram, an MI EEG classification task can be described as:

$$X \leftarrow Z \leftarrow Y \tag{1}$$

As a consequence of the above anti-causal definition and causal diagram, we can explore the problem of MI EEG classification through the following causal factorization:

$$P(X, Y, Z) = P(X|Z)P(Z|Y)P(Y) \tag{2}$$

Through this causal breakdown, we can categorize the major challenges associated with Motor-Imagery (MI) EEG classification tasks into three main categories as illustrated in Figure 1. Challenges related with the:

1) **Training EEG signals - X.** One of the renowned challenges in motor-imagery classification problem - as in any medical-related machine learning problem - is the scarcity of labelled data due to the lengthy acquisition process (e.g. [44], [45], [46]). Subjects are required to spend hours in a laboratory facility performing successive motor-imagery tasks [47]. This process has been reported to cause fatigue and discomfort, even when devices with dry electrodes are utilized. To make things worse, due to the wide variety of available EEG recorders in the market, the data acquisition can be undertaken with various devices (acquisition shift $P(X|Z)$) which have completely different specifications (e.g. number of electrodes, sampling frequency to name just a few), making the combination of publicly available EEG datasets extremely difficult [48].
2) **Anatomical differences of subjects - $P(Z|Y)$.** Each subject has a unique brain anatomy and functionality that results in polymorphous neural activity patterns when appeared in the surface observed EEG signal (e.g. [49], [50]). When designing a generic ML-based MI-BCI, researchers need to take this inter-subject variability (data distribution shift across subjects) into account.
3) **Class Imbalance - $P(Y)$.** Class imbalances can arise between the training and the deployment set of a MI-BCI. It is necessary for the training set to be as closely balanced to the deployment set as possible when training machine learning models.

III. Our Proposed Framework

In this work, we mainly focus on the challenge of subject distribution shift (or inter-subject variability). Using the causal breakdown described in Section II, we will use two publicly available MI datasets - which contain a large number of different subjects, are class balanced, have relatively enough trials per subject and all trials come from a single EEG recorder (within each dataset) - essentially solving all the above identified challenges but the subject distribution shift. In terms of the causal factorization (Eq. 2), the problem of inter-subject variability can be seen as a distribution shift S where:

$$P(X, Y, Z) = P(X|Z)P_{\mathbf{S}}(Z|Y)P(Y) \tag{3}$$

Our framework can be applied to any established CNN-based MI-BCI architecture, resulting in performance increase. Inspired by [42], we utilize dynamic convolutions in the domain of MI brainwave decoding. Instead of having a BCI architecture that tries to discover a common latent space for all K subjects in the training set, we use K parallel trainable convolutional layers (corresponding to the K available training subjects) for each convolutional block of a CNN-based BCI network. Using a subject attention network that learns to distinguish between the available individuals, the subjects are separated from one another and essentially K parallel personalized models of the same BCI architecture are trained simultaneously, as illustrated in Figure 2.

Our proposed framework is inspired by the work of [42] in the field of computer vision, but it includes various modifications to address challenges apparent in the EEG domain. Although the complete framework will be detailedly described in the following subsections III-A and III-B, these differences can be summarized as follows:

- Instead of fully trainable attention mechanisms, it utilizes our novel subject attention network (described in III-A) which uses only trainable Gabor filters making it more lightweight and explainable than a shallow fully trainable neural network and it achieves very high performance in the subject identification task.
- Unlike [42] where there is an attention mechanism for each convolutional layer and these mechanisms are trained in an unsupervised manner, our framework uses only one attention mechanism for all convolutional layers, and with supervised training, it learns to distinguish between the available different subjects.
- The K number of parallel layers in our proposed framework is not a tunable hyperparameter (like in [42]) but coincides with the number of available subjects in the training set.
- Instead of using the output vector of the attention mechanism as [42], our framework utilizes the proposed "uniformly attended" vector $\mathbf{A^*}$ (described in III-B) in order to be more robust to the low Signal-to-Noise Ratio (SNR) of the EEG signal.

A. Attention Network

The first layer of our subject attention network is the first order wavelet scalogram of the input EEG signal X. Mathematically, let $\mathbf{x}(t) \in \mathbb{R}^T$ denote a one-dimensional input EEG signal, where T is the number of initial EEG time points, and $\boldsymbol{\psi}_\epsilon(t)$ be a wavelet. The 1st order scalogram is defined as $\mathbf{X}(\epsilon, t) = |\mathbf{x}(t) * \boldsymbol{\psi}_\epsilon(t)|$, where $*$ stands for the convolution operator. To perform this operation, the raw input signal from each EEG channel is convolved with a wavelet kernel with size $(1, W) = (1, \frac{F_s}{2})$ where F_s is the sampling frequency. This wavelet kernel follows the real Gabor wavelet format:

$$\boldsymbol{\psi}_\epsilon(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} \cos(2\pi\epsilon t) \tag{4}$$

with $t \in [-\frac{W}{2}, \frac{W}{2}]$ and $\frac{1}{\sigma}$ denotes the bandwidth and ϵ the normalized frequency of the Gabor wavelet and these two properties are the only trainable parameters of this layer. During training, ϵ is restricted ($\epsilon \in [0, \frac{1}{2}]$) to satisfy the Nyquist theorem. The three-dimensional (3D) tensor $X(c, \epsilon, t) \in \mathbb{R}^{C \times F \times T}$ (where F is the number of Gabor filters and C the number of EEG channels) containing the first order wavelet scalograms $\mathbf{X}(\epsilon, t)$ for each EEG channel is followed by a global average pooling across time and frequency. Finally, the resulted vector is passed through a fully-connected layer to compute the subject id vector π.

B. Subject-Attended Dynamic Convolutions

The proposed framework takes the EEG signal X as input and tries to learn both the correct MI task Y (estimate the conditional probability $P(Y|X)$) as well as the correct subject id π (estimate the conditional probability $P(\pi|X)$). The subject attention network and the K parallel convolutional layers are trained simultaneously using the following loss function:

$$Loss = (1 - acc) \times \ell_{Attention} + acc \times \ell_{MI} \tag{5}$$

where acc is the training accuracy of the subject attention network and ℓ denotes the cross-entropy function ($\ell_{Attention}$ for the subject attention network and ℓ_{MI} for the MI classification task). This loss function effectively enforces first the training of the subject attention network and, as the attention's accuracy increases, it switches its focus to train the parallel convolutional layers for the different MI tasks. As also suggested in [42], since softmax does not work well due to its near one-hot output, we use a large temperature in the softmax of the attention network during training in order to flatten the framework's attention, allow a broader gradient backpropagation and effectively assist in the subject attention network's training in the early epochs.

During inference, when an input EEG signal (x) from a new unseen subject S_x is processed, it passes firstly through the attention network and the subject attention vector π is computed where $\sum_i \pi_i = 1$. We empirically observed that this vector is quite sparse, and if it was used during inference, only a handful of parallel convolutional layers would be utilized during the mixing. Instead, we would ideally like to use knowledge from all K individuals and "shift" the attention more to the most relevant subjects. To accomplish that, we compute what we call the "uniformly attended vector" $\mathbf{A^*}$. If there was no attention network, the K parallel convolutional layers would be mixed with a uniform factor $A_i = \frac{1}{K}$. To compute

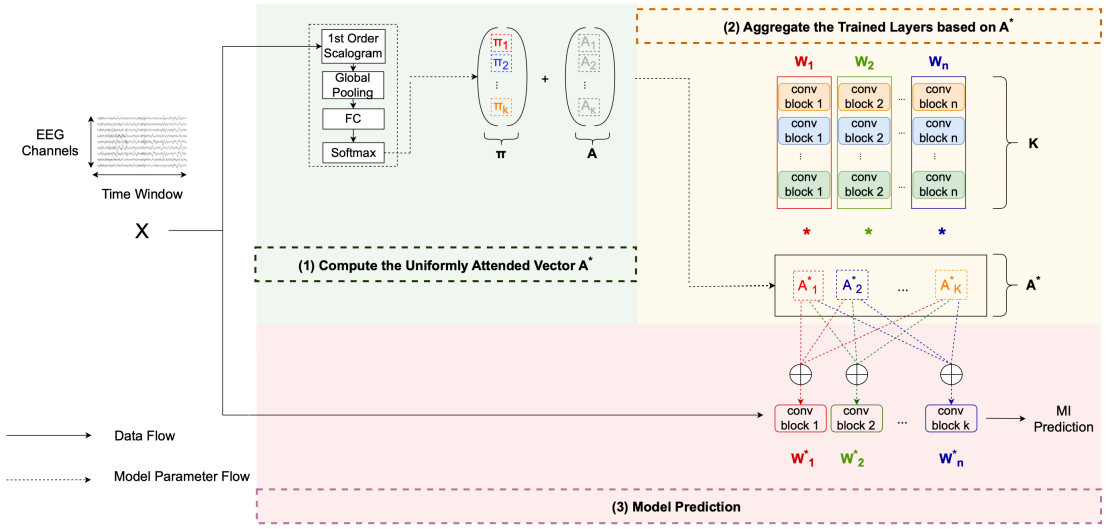

Fig. 2. Dynamic convolution framework for BCI architectures. X represents input EEG signal trial. The K different subjects in the training set are represented by different colors in the convolutional blocks. Colored rectangles and arrows (namely green, red and dark blue) demonstrate the different blocks that are taken into account when computing the final convolutional blocks for the MI classification task.

the "uniformly attended vector", the uniform attention vector **A** is combined with the subject attention vector π and the result is passed through a softmax activation to flatten the attention across all subjects - while maintaining the focus on the most relevant ones (we refer the reader to the Appendix for a performance comparison between using π and **A*** as attention). Mathematically, this operation can be described as:

$$\underbrace{\begin{pmatrix} A_1 \\ A_2 \\ \dots \\ A_k \end{pmatrix}}_{\mathbf{A}} + \underbrace{\begin{pmatrix} \pi_1 \\ \pi_2 \\ \dots \\ \pi_k \end{pmatrix}}_{\pi} \xrightarrow{\sigma} \underbrace{\begin{pmatrix} A_1^* \\ A_2^* \\ \dots \\ A_k^* \end{pmatrix}}_{\mathbf{A^*}} \tag{6}$$

where σ denotes the softmax operation, **A** the uniform attention vector with $A_i = \frac{1}{K}$ and **A*** the "uniformly attended" vector where $\sum_i A_i^* = 1$. Let us denote with $\mathbf{W}_i^j$ the learned convolutional kernel of the network's i^{th} convolutional layer from the j^{th} parallel network and with $\mathbf{W}_i^*$ the dynamic convolutional kernel of the network's i^{th} layer, as illustrated in Figure 2. In our proposed framework, we compute the dynamic convolutional as follows:

$$\mathbf{W}_i^*(x) = \sum_{k=1}^{K} A_k^*(x)\mathbf{W}_i^k \tag{7}$$

In other words, using the causal factorization (3), our proposed framework estimate the probability $P_{S_x}(Z|Y)$ of a new unseen subject S_x as the linear combination of K learned conditional probabilities. More specifically:

$$P_{S_x}(Z|Y) = A_1^* \times P_{S_1}(Z|Y) + A_2^* \times P_{S_2}(Z|Y) + \dots + A_k^* \times P_{S_k}(Z|Y) \tag{8}$$

IV. Experiments

To validate our proposed framework based on our causal breakdown in Section II, two publically available MI datasets are used namely:

1) **PhysioNet** [51]: The original Physionet dataset includes brain recordings from 109 healthy participants, registered via 64 EEG sensors with a sampling frequency of 160 Hz, while performing a series of pseudorandomized cue-triggered MI tasks. In our experiments, we first excluded data from 6 participants (subjects 88, 89, 92, 100, 104 and 106) due to differences in either the sampling frequency or duration of the performed tasks. We extracted trials corresponding to MI hand or feet movements in the form of segments starting with the visual cue and lasting for 4.1 seconds.
2) **OpenBMI - MI** [52]: The original OpenBMI dataset consists of 3 BCI paradigms: ERP-based speller, MI and SSVEP. The MI trials include brain recordings from 54 healthy participants, registered via 62 EEG sensors with a sampling frequency of 1000 Hz. In the MI part of the dataset, the participants performed a series of cue-triggered MI tasks either with or without receiving feedback (cursor moved according to the prediction of a trained classifier). For the purpose of this study, we kept only the MI-trials without feedback, since the neurofeedback was not included as a factor in our initial causal analysis. In particular, we extracted trials corresponding to MI hands in the form of segments starting with the visual cue and lasting for 4 seconds. Furthermore, we applied a notch filter at 60Hz - and its harmonics (120, 180, 240, 300, 360, 420, 480) - to remove power-line noise. We also applied a notch filter at 460Hz due to a spurious artifact (consistent across all trials).

A. Subject Verification

The subject attention mechanism is a vital part in our proposed framework. Therefore, we evaluated its performance separately first in order to ensure its ability to distinguish between the various available subjects in the two datasets. We performed 10-fold trial-wise cross-validation to measure its performance. Adam optimizer was used with learning rate of 0.01 for the first 30 epochs (to allow the Gabor filters

TABLE I
PERFORMANCE OF SUBJECT ATTENTION NETWORK (TRAINED AND EVALUATED USING 10-FOLD CROSS-VALIDATION) IN PREDICTING THE SUBJECT ID IN PHYSIONET AND OPENBMI - MI DATASETS. CV STANDS FOR CROSS-VALIDATION

Dataset	CV Average Accuracy[1]
PhysioNet (103 Subjects)	98.5% ± 0.13%
OpenBMI - MI (54 Subjects)	90.3% ± 0.07%

TABLE II
HYPER-PARAMETER CHOICES FOR THE EXPERIMENTS

Model Setup	Vanilla	Dynamic
Optimizer	Adam	Adam
Training epochs	30	30
Learning Rate	0.001	0.01 (first 20 epochs) and 0.001 (last 10 epochs)
Number of Gabor Filter	-	8
Initialized Center Frequencies (Hz) of Gabor Filters	-	$\{8, 12, 16, 20, 24, 28, 32, 36\}$
Temperature used in the softmax	-	30

TABLE III
PERFORMANCE OF GENERIC (TRAINED AND EVALUATED IN A LEAVE-ONE-SUBJECT-OUT FASHION) MODELS FOR DEEPCONVNET, SHALLOWCONVNET, EEGNET, EEG-INCEPTION AND THEIR DYNAMIC EQUIVALENT NETWORKS (OURS). THE K PARAMETER USED IN DYNAMIC MODELS IS COLOURED WITH VIOLET. THE P-VALUE OF PAIRED T-TESTS BETWEEN PERFORMANCE OF STANDARD AND DYNAMIC IS COLOURED WITH GRAY. THE RATIO OF TRAINABLE PARAMETERS ($\frac{Dynamic}{Standard}$) IS COLOURED WITH BLUE

Dataset **Task**	PhysioNet MI Left / Right	PhysioNet MI Left / Right Hand / FeetHand	OpenBMI - MI MI Left / Right Hand
ShallowConvNet	80.6 ± 11.4%	66.3 ± 16.0%	66.3 ± 11.1%
Dynamic ShallowConvNet	**83.3 ± 12.7%** (102 / ≈0.0055 / 97.5)	**69.0 ± 16.3%** (102 / ≈0.0043 / 95.4)	**70.3 ± 11.1%**[2] (53 / 50.68)
DeepConvNet	82.5 ± 11.5%	67.1 ± 14.6%	71.7 ± 12.0%
Dynamic DeepConvNet	**83.5 ± 13.0%** (102 / ≈0.13 / 100.8)	**71.3 ± 16.1%** (102 / ≈0.0000328 / 100.36)	**73.1 ± 11.6%**[2] (53 / 52.47)
EEGNet	79.5 ± 12.4%	66.5 ± 13.2%	**74.9 ± 11.0%**
Dynamic EEGNet	**80.2 ± 13.5%** (102 / ≈0.26 / 79.8)	**67.5 ± 15.4%** (102 / ≈0.15 / 72.2)	71.9 ± 12.1%[2] (53 / 41.42)
EEG-Inception	81.6 ± 11.8%	67.6 ± 15.1%	76.5 ± 10.8%
Dynamic EEG-Inception	**83.9 ± 11.9%** (102 / ≈0.0068 / 94.5)	**71.4 ± 15.0%** (102 / ≈0.00025 / 92.84)	**77.4 ± 10.0%**[2] (53 / 49.20)

to quickly adapt to the data) and 0.001 for the remainder 20 epochs. As shown in Table I, the subject attention network performs sufficiently well in both datasets which makes it an ideal candidate for the attention mechanism in our proposed dynamic framework.

B. Comparison Between Standard and Dynamic Models

We tested our proposed framework in four well-established BCI architectures, namely DeepConvNet [26], ShallowConvNet [26], EEGNet [27] and EEG-Inception [28] in the following MI tasks: for the publically available MI dataset Physionet [51] one binary classification task (MI Left vs Right Hand) and a 3-class classification problem (MI Left Hand / Right Hand / Feet)) and for OpenBMI - MI [52] one MI binary classification task (MI Left vs Right Hand).

As shown in Table II, we trained the standard networks for 30 epochs with learning rate of 0.001 while their dynamic versions for 30 epochs - in the first 20 epochs with learning rate of 0.01, to assist the attention's Gabor filters to quickly adapt to the data, and 10 epochs with learning rate of 0.001 and frozen attention, to fine-tune to the MI task. In all cases, we used an Adam optimizer. Finally, a temperature of 30 was used during training in the attention mechanism as described in the previous section. We evaluated the performance of the standard networks and their equivalent dynamic networks in a leave-one-subject-out fashion (cross-subject performance) Table III.

C. Comparison With State-of-the-Art Approaches

In this work, we are not only interested in comparing the models trained with our framework versus regularly trained CNN-based BCI architectures but also to compare our framework with other transfer learning approaches in the EEG domain. Therefore, we evaluated the performance of the standard networks and their equivalent dynamic networks in a leave-M-subjects-out fashion Table IV. Furthermore, we compared our framework with two other commonly used transfer learning EEG techniques: 1) an adversarial approach, namely [40], that (similarly to our approach) does not use a calibration set and 2) Euclidean alignment [53] that projects data into a domain-invariant space but it uses all the trials of a subject. We trained the Euclidean alignment networks similar to their vanilla equivalent after performing the data projection for each subject. And we trained the equivalent adversarial networks with early stopping and adversarial regularization weight $\epsilon = 0.005$ (hyperparameters taken from the original paper [40]). As it can be seen from Table IV, our proposed method outperforms adversarial networks (a similar zero-calibration method) while it achieves the same or higher performance when compared with Euclidean alignment. It is worth mentioning though that Euclidean alignment uses all the trials of an unseen subject while our framework is dynamically adapted for each trial during inference.

D. Calibration Methods

We evaluated the performance of the calibrated networks (using a small calibration set of the unseen subjects to fine-tune the final classification layer). For a fair comparison, we also fine-tuned the last layer of the equivalent dynamic networks using the same calibration sets. As it is shown in Table V, the calibrated dynamic models also outperform their equivalent vanilla calibrated networks.

E. Investigation of Negative Transfer Learning

Although our proposed framework showed increased cross-subject performance as experimentally shown above, we wanted to investigate if there are any signs of negative transfer learning during the process. As it is shown in Figure 3, although there are limited cases of negative transfer learning,

[1] ±% Refers to the rounded standard deviation across 10 runs of 10-fold cross-validation experiments.

[2] Early stopping has been applied to some folds during the fine-tuning phase.

[3] Average result of 10 calibrated models. Each model uses 20% of each testing subject's samples for calibration (chosen randomly) and 80% for testing.

TABLE IV

PERFORMANCE OF GENERIC (TRAINED AND EVALUATED IN A LEAVE-M-SUBJECTS-OUT FASHION) MODELS OF MI-CLASSIFICATION (LEFT / RIGHT HAND) TASKS IN PHYSIONET AND OPENBMI-MI. CV STANDS FOR CROSS-VALIDATION ACROSS SUBJECTS

Model **Dataset**	5-Fold CV ($K \approx 83$) PhysioNet	10-Fold CV ($K \approx 93$) PhysioNet	20-Fold CV ($K \approx 98$) PhysioNet	5-Fold CV ($K \approx 43$) OpenBMI - MI	Trials of Unseen Subject Used for Adaptation
Vanilla DeepConvNet	81.4 ± 1.3%	81.7 ± 2.9%	82.4 ± 5.5%	72.1 ± 2.3%	-
DeepConvNet with Euclidean Alignment	82.7 ± 1.2%	83.03 ± 3.2%	**83.2 ± 5.0%**	70.6 ± 2.0%	All Trials
Adversarial DeepConvNet [2]	81.4 ± 1.35%	81.8 ± 3.7%	81.9 ± 4.6%	66.6 ± 2.0%	-
Dynamic DeepConvNet (Ours)	**82.8 ± 2.03%**	**83.14 ± 3.9%**	83.2 ± 5.3%	**72.3 ± 2.3%**	1
Vanilla ShallowConvNet	79.7 ± 1.75%	80.4 ± 3.3%	80.95 ± 4.6%	62.1± 1.3%	-
ShallowConvNet with Euclidean Alignment	79.8 ± 2.0%	80.3 ± 3.5%	81.1 ± 4.3%	63.8 ± 3.8%	All Trials
Adversarial ShallowConvNet [2]	81.0 ± 1.26%	81.3 ± 3.7%	82.35 ± 4.65%	55.1± 0.9%	-
Dynamic ShallowConvNet (Ours)	**81.0 ± 1.06%**	**82.32 ± 2.86%**	**83.10 ± 4.7%**	**68.4 ± 2.9%**	1

TABLE V

PERFORMANCE OF GENERIC (TRAINED AND EVALUATED IN A LEAVE-M-SUBJECTS-OUT FASHION) MODELS OF MI-CLASSIFICATION (LEFT / RIGHT HAND) TASKS IN PHYSIONET FOR CALIBRATED DEEPCONVNET AND SHALLOWCONVNET AND THEIR CALIBRATED DYNAMIC EQUIVALENT NETWORKS. CV STANDS FOR CROSS-VALIDATION ACROSS SUBJECTS

Model	5-Fold CV ($K \approx 83$)	10-Fold CV ($K \approx 93$)	20-Fold CV ($K \approx 98$)
Calibrated DeepConvNet [3]	81.94 ± 1.95%	82.3 ± 3.0%	82.5 ± 5.26%
Calibrated Dynamic DeepConvNet [3]	**83.2 ± 1.5%**	**83.35 ± 3.35%**	**83.43 ± 4.8%**
Calibrated ShallowConvNet[3]	80.6 ± 1.2%	81.5 ± 3.5%	81.45± 4.6%
Calibrated Dynamic ShallowConvNet [3]	**82.65 ± 1.55%**	**83.9 ± 3.85 %**	**84.00 ± 5.2%**

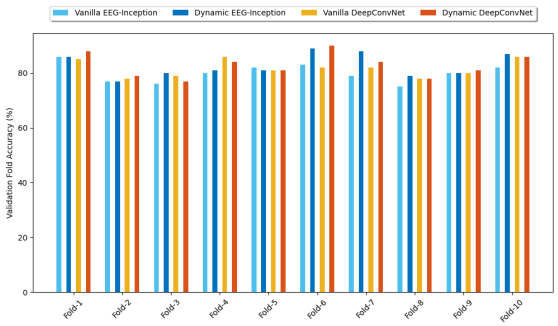

Fig. 3. Per-fold comparison of the performance of vanilla architectures versus their equivalent dynamic networks (ours).

the vast majority is either marginally or significantly better compared to the vanilla architectures.

V. DISCUSSION

The proposed dynamic framework can be used in various CNN-based MI-BCI architectures to increase the cross-subject performance and can take us one step closer in tackling the problem of inter-subject variability as the experimental evaluation in the previous Section IV illustrates. We expect this framework, with certain modifications, to be able to generalize well and get adapted to various BCI paradigms, not only MI. Investigating different BCI paradigms is beyond the scope of this paper where the causal analysis of the MI task is a core factor in ensuring that our proposed framework tackles the targeted problem and there are no misleading performance increases. Extending the framework to different paradigms would require also a causal breakdown for these tasks.

One limitation of our work is the unavoidable increase in the number of trainable parameters (about × K where K is the number of available subjects in the dataset). Although our subject attention mechanism seems to identify well a large number of subject (e.g. 103 on PhysioNet), this increase in the

TABLE VI

INFERENCE TIMINGS FOR 10 TRAINED MODELS OF MI-CLASSIFICATION (LEFT / RIGHT HAND AND LEFT / RIGHT HAND / FEET) FROM THE LEAVE-ONE-SUBJECT-OUT CROSS-VALIDATION FOR EEG-INCEPTION IN PHYSIONET. MEASURED WITH *torch.autograd.profiler* IN 2.9 GHZ 6-CORE CPU INTEL CORE I9

Subject	Standard (ms)	Dynamic (ms)
Subject 1	561.9779	1146.8094
Subject 2	543.0056	1099.2773
Subject 3	563.3565	1139.8957
Subject 4	545.2918	898.6636
Subject 5	545.9808	881.258
Subject 6	550.0676	1141.3383
Subject 7	542.7533	1135.5232
Subject 8	538.7697	1114.9791
Subject 9	541.0152	887.6454
Subject 10	534.8294	1111.1364

number of trainable parameters might be a limiting factor in some cases especially if these models are deployed on real-life applications where devices have limited amount of memory storage. Fortunately, this tremendous increase in number of parameters does not translate to execution time. As it is shown in Appendix, there is a less than × K increase in terms of inference time cost. Inspired by related works [54], we could investigate approaches to mitigate this increase in a future work.

In contrast to other techniques that promise to tackle the issue of inter-subject variability, our framework is dynamically adapted to a new subject during inference without the need of re-training or calibration trials, commonly used in transfer learning methods. Furthermore, an inherent advantage of our framework is the training of K parallel personalized models of the same BCI architecture. During training, these models are not trained using only the samples of one specific subject but also samples from "similar" subjects since the attention mechanism is trained simultaneously. An interesting future step would be to evaluate the performance of these inherent personalized models compared to standard personalized models - trained using strictly the samples of one specific subject. Although the BCI deep architectures used in Section IV are considered state-of-the-art and achieve high performance across different MI-BCI tasks, they are usually comprised of thousands of trainable parameters, making the training of standard personalized models difficult with these publicly available datasets. For that endeavour, we need first to design more lightweight BCI architectures and then perform these comparisons.

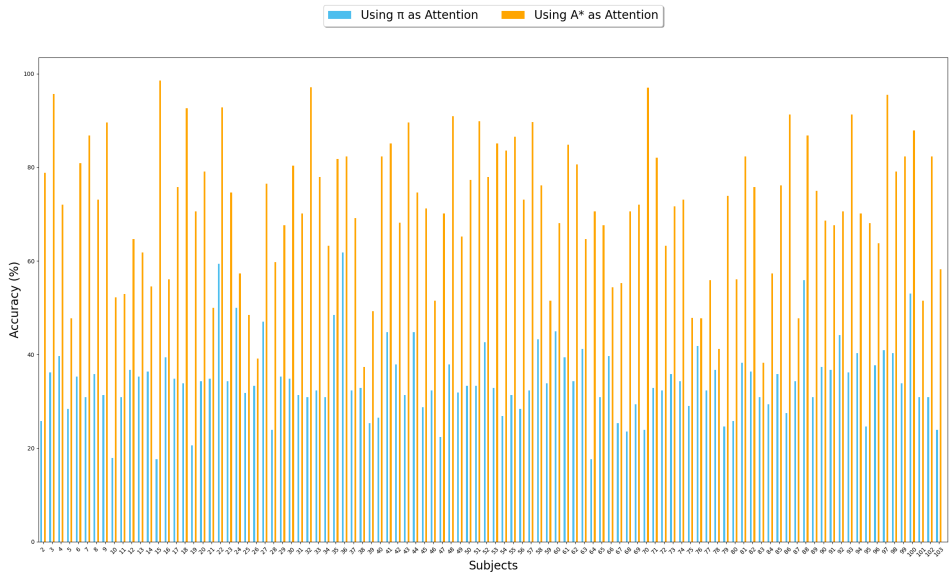

Fig. 4. Comparison between using the vector π versus our proposed uniformly attended vector $\boldsymbol{A}^*$ as attention in our proposed dynamic framework. The performances correspond to generic (trained and evaluated in a leave-one-subject-out fashion) Dynamic EEG-Inception models in the MI-classification (Left / Right hand / Feet) task in Physionet.

VI. Conclusion

In this work, we analyze the task of MI EEG classification through the lens of causal reasoning. To the best of our knowledge, this is the first work that brings machine learning in conjunction with causal reasoning to the domain of EEG. Through this analysis, we identify and analyze some of the major challenges and we introduce a framework based on dynamic convolutions that tackles the problem of subject distribution shift (inter-subject variability). Our proposed subject attention mechanism achieves great performance in identifying subjects and the overall proposed dynamic framework demonstrates increased performance when applied to different BCI architectures while at the same time outperforming other similar methods. In future work, we plan to use it to tackle more, if not all, challenges detailedly described in our causal analysis of MI brainwave decoding.

Appendix

As described in Section III, during inference when an input EEG signal from a new unseen subject S_x is processed, it passes firstly through the attention network and the subject attention vector π is computed. Through investigation, we observed that this vector is quite sparse. Although this is something we would ideally like, the low SNR of the EEG signal makes our framework unstable especially when used in our desired zero-calibration one-trial setup. In order to have a robust network that dynamically adapts to the new trial from an unseen subject, we utilized the "uniformly attended vector" $\mathbf{A}^*$ that uses knowledge from all k individuals and "shift" the attention more to the most relevant subjects. A comparison between using the vector π versus our proposed uniformly attended vector $\mathbf{A}^*$ as attention in our proposed dynamic framework can be seen in the following Figure 4.

A significant drawback of our proposed framework is the unavoidable increase in the number of trainable parameters (about $\times$ K where K is the number of available subjects in the dataset). This factor can have limiting effects when these models are deployed on real-life applications where devices have limited amount of memory storage. As it is shown in the following table, the tremendous increase in number of parameters does not translate to execution time which is less than $\times$ K increase in terms of inference time cost.

References

[1] K. Barmpas, Y. Panagakis, D. A. Adamos, N. Laskaris, and S. Zafeiriou, "A causal viewpoint on motor-imagery brainwave decoding," in *Proc. ICLR Workshop Elements Reasoning, Objects, Struct. Causality*, 2022, pp. 1–7.

[2] E. P. Torres, E. A. Torres, M. Hernández-Álvarez, and S. G. Yoo, "EEG-based BCI emotion recognition: A survey," *Sensors*, vol. 20, no. 18, p. 5083, Sep. 2020.

[3] T. Xu, Y. Zhou, Z. Wang, and Y. Peng, "Learning emotions EEG-based recognition and brain activity: A survey study on BCI for intelligent tutoring system," *Proc. Comput. Sci.*, vol. 130, pp. 376–382, Jan. 2018.

[4] R. Alkawadri, "Brain–computer interface (BCI) applications in mapping of epileptic brain networks based on intracranial-EEG: An update," *Frontiers Neurosci.*, vol. 13, p. 191, Mar. 2019.

[5] L. C. D. Nkengfack, D. Tchiotsop, R. Atangana, V. Louis-Door, and D. Wolf, "Classification of EEG signals for epileptic seizures detection and eye states identification using Jacobi polynomial transforms-based measures of complexity and least-square support vector machine," *Informat. Med. Unlocked*, vol. 23, Jan. 2021, Art. no. 100536.

[6] D. C. Irimia, R. Ortner, G. Krausz, C. Guger, and M. S. Poboroniuc, "BCI application in robotics control," *IFAC Proc. Volumes*, vol. 45, no. 6, pp. 1869–1874, May 2012.

[7] B. Kerous, F. Skola, and F. Liarokapis, "EEG-based BCI and video games: A progress report," *Virtual Reality*, vol. 22, no. 2, pp. 119–135, Jun. 2018.

[8] J. Decety and D. H. Ingvar, "Brain structures participating in mental simulation of motor behavior: A neuropsychological interpretation," *Acta Psychologica*, vol. 73, no. 1, pp. 13–34, Feb. 1990. i

[9] D. J. McFarland and J. R. Wolpaw, "Sensorimotor rhythm-based brain–computer interface (BCI): Feature selection by regression improves performance," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no. 3, pp. 372–379, Sep. 2005.

[10] T. Solis-Escalante, G. Müller-Putz, and G. Pfurtscheller, "Overt foot movement detection in one single Laplacian EEG derivation," *J. Neurosci. Methods*, vol. 175, no. 1, pp. 148–153, Oct. 2008.

[11] R. Mane, T. Chouhan, and C. Guan, "BCI for stroke rehabilitation: Motor and beyond," *J. Neural Eng.*, vol. 17, no. 4, Aug. 2020, Art. no. 041001.

[12] N. Robinson, R. Mane, T. Chouhan, and C. Guan, "Emerging trends in BCI-robotics for motor control and rehabilitation," *Current Opinion Biomed. Eng.*, vol. 20, Dec. 2021, Art. no. 100354.

[13] A. A. Frolov et al., "Post-stroke rehabilitation training with a motor-imagery-based brain–computer interface (BCI)-controlled hand exoskeleton: A randomized controlled multicenter trial," *Frontiers Neurosci.*, vol. 11, pp. 1–10, Jul. 2017.

[14] K. K. Ang and C. Guan, "EEG-based strategies to detect motor imagery for control and rehabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 4, pp. 392–401, Apr. 2016.

[15] M. Lee, J.-H. Jeong, Y.-H. Kim, and S.-W. Lee, "Decoding finger tapping with the affected hand in chronic stroke patients during motor imagery and execution," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1099–1109, 2021.

[16] V. K. Benzy, A. P. Vinod, R. Subasree, S. Alladi, and K. Raghavendra, "Motor imagery hand movement direction decoding using brain computer interface to aid stroke recovery and rehabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 3051–3062, Dec. 2020.

[17] K. Choi and A. Cichocki, "Control of a wheelchair by motor imagery in real time," in *Proc. Int. Conf. Intell. Data Eng. Automated Learn.* Berlin, Germany: Springer, 2008, Art. no. 330b337.

[18] D. J. McFarland, C. W. Anderson, K.-R. Müller, A. Schlogl, and D. J. Krusienski, "BCI meeting 2005-workshop on BCI signal processing: Feature extraction and translation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 135–138, Jun. 2006.

[19] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Jan. 2008.

[20] K. Keng Ang, Z. Yang Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain–computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jun. 2008, pp. 2390–2397.

[21] J. Chai, H. Zeng, A. Li, and E. W. T. Ngai, "Deep learning in computer vision: A critical review of emerging techniques and application scenarios," *Mach. Learn. Appl.*, vol. 6, Dec. 2021, Art. no. 100134.

[22] A. Antoniades, L. Spyrou, C. Cheong Took, and S. Sanei, "Deep learning for epileptic intracranial EEG data," in *Proc. IEEE 26th Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Sep. 2016, pp. 1–6.

[23] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J. Neural Eng.*, vol. 14, no. 1, 2017, Art. no. 016003.

[24] M. Längkvist, L. Karlsson, and A. Loutfi, "Sleep stage classification using unsupervised feature learning," *Adv. Artif. Neural Syst.*, vol. 2012, pp. 1–9, Jul. 2012.

[25] D. F. Wulsin, J. R. Gupta, R. Mani, J. A. Blanco, and B. Litt, "Modeling electroencephalography waveforms with semi-supervised deep belief nets: Fast classification and anomaly measurement," *J. Neural Eng.*, vol. 8, no. 3, 2011, Art. no. 036015.

[26] R. T. Schirrmeister et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Aug. 2017.

[27] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Jul. 2018, Art. no. 056013.

[28] E. Santamaría-Vázquez, V. Martínez-Cagigal, F. Vaquerizo-Villar, and R. Hornero, "EEG-inception: A novel deep convolutional neural network for assistive ERP-based brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 2773–2782, Dec. 2020.

[29] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[30] S. Saha and M. Baumert, "Intra- and inter-subject variability in EEG-based sensorimotor brain computer interface: A review," *Frontiers Comput. Neurosci.*, vol. 13, pp. 1–13, Jan. 2020.

[31] S. Marcel and J. D. R. Millan, "Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 743–752, Apr. 2007.

[32] A. Valsaraj, I. Madala, N. Garg, M. Patil, and V. Baths, "Motor imagery based multimodal biometric user authentication system using EEG," in *Proc. Int. Conf. Cyberworlds*, Oct. 2020, pp. 272–279.

[33] L. Yang, A. Libert, and M. M. Van Hulle, "Chronic study on brainwave authentication in a real-life setting: An LSTM-based bagging approach," *Biosensors*, vol. 11, no. 10, p. 404, Oct. 2021.

[34] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain–computer interface classification by Riemannian geometry," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 4, pp. 920–928, Apr. 2012.

[35] H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," *IEEE Signal Process. Lett.*, vol. 16, no. 8, pp. 683–686, Aug. 2009.

[36] D. Zhao, F. Tang, B. Si, and X. Feng, "Learning joint space–time–frequency features for EEG decoding on small labeled data," *Neural Netw.*, vol. 114, pp. 67–77, Jun. 2019.

[37] A. N. Olesen, P. Jennum, E. Mignot, and H. B. D. Sorensen, "Deep transfer learning for improving single-EEG arousal detection," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 99–103.

[38] K. Zhang, N. Robinson, S.-W. Lee, and C. Guan, "Adaptive transfer learning for EEG motor imagery classification with deep convolutional neural network," *Neural Netw.*, vol. 136, pp. 1–10, Apr. 2021.

[39] R. Zhang, Q. Zong, L. Dou, X. Zhao, Y. Tang, and Z. Li, "Hybrid deep neural network using transfer learning for EEG motor imagery decoding," *Biomed. Signal Process. Control*, vol. 63, Jan. 2021, Art. no. 102144.

[40] O. Ozdenizci, Y. Wang, T. Koike-Akino, and D. Erdogmus, "Learning invariant representations from EEG via adversarial inference," *IEEE Access*, vol. 8, pp. 27074–27085, 2020.

[41] B. Scholkopf et al., "Toward causal representation learning," *Proc. IEEE*, vol. 109, no. 5, pp. 612–634, May 2021.

[42] Y. Chen, X. Dai, M. Liu, D. Chen, L. Yuan, and Z. Liu, "Dynamic convolution: Attention over convolution kernels," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11027–11036.

[43] D. C. Castro, I. Walker, and B. Glocker, "Causality matters in medical imaging," *Nature Commun.*, vol. 11, no. 1, p. 3673, Jul. 2020.

[44] F. Lotte, "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces," *Proc. IEEE*, vol. 103, no. 6, pp. 871–890, Jun. 2015.

[45] Z. Khademi, F. Ebrahimi, and H. M. Kordy, "A review of critical challenges in MI-BCI: From conventional to deep learning methods," *J. Neurosci. Methods*, vol. 383, Jan. 2023, Art. no. 109736.

[46] W. Xiong and Q. Wei, "Reducing calibration time in motor imagery-based BCIs by data alignment and empirical mode decomposition," *PLoS ONE*, vol. 17, no. 2, pp. 1–18, Feb. 2022.

[47] R. Scherer, A. Schloegl, F. Lee, H. Bischof, J. Janša, and G. Pfurtscheller, "The self-paced Graz brain–computer interface: Methods and applications," *Comput. Intell. Neurosci.*, vol. 2007, pp. 1–9, Jan. 2007.

[48] X. Wei et al., "2021 BEETL competition: Advancing transfer learning for subject independence heterogenous EEG data sets," in *Proc. NeurIPS*, 2022, pp. 205–219.

[49] B. Blankertz et al., "Neurophysiological predictor of SMR-based BCI performance," *NeuroImage*, vol. 51, no. 4, pp. 1303–1309, 2010.

[50] R. Zhang, F. Li, T. Zhang, D. Yao, and P. Xu, "Subject inefficiency phenomenon of motor imagery brain–computer interface: Influence factors and potential solutions," *Brain Sci. Adv.*, vol. 6, no. 3, pp. 224–241, Sep. 2020.

[51] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, Jun. 2000, Art. no. e215be220.

[52] O.-Y. Kwon, M.-H. Lee, C. Guan, and S.-W. Lee, "Subject-independent brain–computer interfaces based on deep convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 10, pp. 3839–3852, Oct. 2020.

[53] H. He and D. Wu, "Transfer learning for brain–computer interfaces: A Euclidean space data alignment approach," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 399–410, Feb. 2020.

[54] F. Wu, A. Fan, A. Baevski, Y. Dauphin, and M. Auli, "Pay less attention with lightweight and dynamic convolutions," in *Proc. Int. Conf. Learn. Represent.*, 2019, pp. 1–15.