IMPROVING GENERALIZATION OF MOTOR-IMAGERY BRAINWAVE DECODING VIA DYNAMIC CONVOLUTIONS

Anonymous authors
Paper under double-blind review

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 across subjects in various MI tasks for four well-established deep architectures.

1 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. One of the first and most popular BCI paradigm 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 (Decety & Ingvar (1990)). Developing MI-BCI systems 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. Mane et al. (2020), Robinson et al. (2021), Sebastián-Romagosa et al. (2020)).

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 (Chai et al. (2021)). Due to their massive progress, CNN-based feature extractors have been introduced in various paradigms in the field of BCIs (e.g. Antoniades et al. (2016), Rezaeitabar & Halici (2017), Längkvist et al. (2012), Wulsin et al. (2011)), in an effort to become generic EEG signal processing tools compared to classical feature extraction techniques (McFarland et al. (2006), Blankertz et al. (2008), Ang et al. (2008)), in which the exact design of the spatio-temporal filtering pipeline is accomplished in a principled manner so as to ensure the reliability of the subsequent brain activity decoding. One of the core challenges that a BCI - or more generally the decoding of EEG signals - faces is to cope with changes 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. Therefore, modern DL-based BCIs tend to fail to generalize well in unseen subjects due to this type of data distribution shift (Saha & Baumert (2020)).

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 Schölkopf et al. (2021), 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 Chen et al. (2020), 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.

2 Background

2.1 Deep Learning in MI

DeepConvNet and ShallowConvNet (Schirrmeister et al. (2017)) are among the first deep learning architectures employed in MI-BCIs and are inspired by common spatial pattern (CSP) filters (Blankertz et al. (2008)) since they include convolutions across time followed by convolutions across EEG channels. EEGNet (Lawhern et al. (2018)) 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 (Ang et al. (2008)). EEG-Inception (Santamaría-Vázquez et al. (2020)) 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 Szegedy et al. (2015). These branches consist of trainable convolutional temporal filters of different scales, capturing several temporal modulations of the EEG signals.

2.2 Inter-subject variability and Transfer Learning

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, preventing the successful deployment of a previously trained MI classifier to new unseen subjects. 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 Marcel & R. Millan (2007), Valsaraj et al. (2020), Yang et al. (2021)). For many years, normalization techniques (e.g Barachant et al. (2012), Kang et al. (2009)) - 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 Zhao et al. (2019), Olesen et al. (2020), Zhang et al. (2021a), Zhang et al. (2021b)). 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 Zhang et al. (2021b) only the last fully-connected layers are fine-tuned while the previous layers are frozen. While in Zhao et al. (2019) 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, Özdenizci et al. (2020) 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.

3 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 Schölkopf et al. (2021), 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.

3.1 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 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.

3.2 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 (Castro et al. (2020)):

- 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 Castro et al. (2020), we can consider X as a sequence of imperfect observed measurements (in sensor-space) 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. 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. Subjects are required to spend hours in a laboratory facility performing successive motor-imagery tasks. 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 publically available EEG datasets extremely difficult.
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. 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.

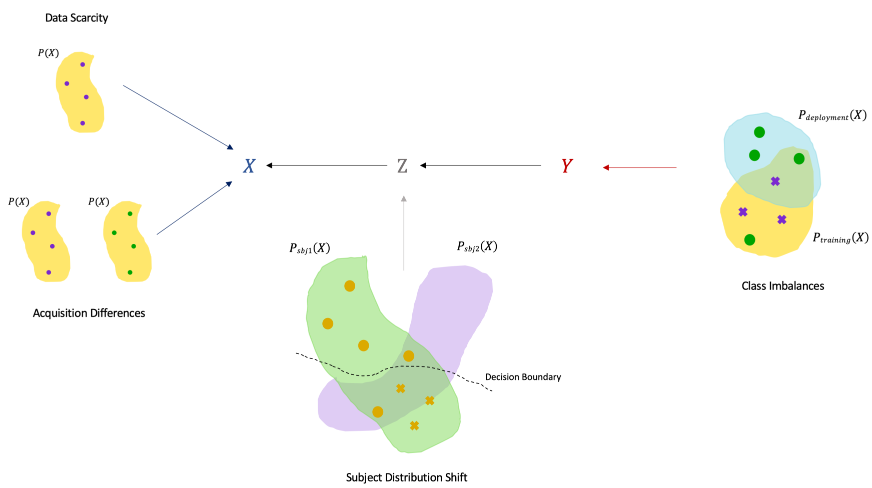

Figure 1: Key challenges in machine learning for a MI EEG classification task. X represents input EEG signals, Y the associated MI labels. Big circles and crosses represent EEG signals of different labels. Dots represent data points of any label and their color represent different EEG acquisition devices.

4 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 3, 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 (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 a statistically significant performance increase. Inspired by Chen et al. (2020), 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 kernels (corresponding to the k available training subjects) for each convolutional block

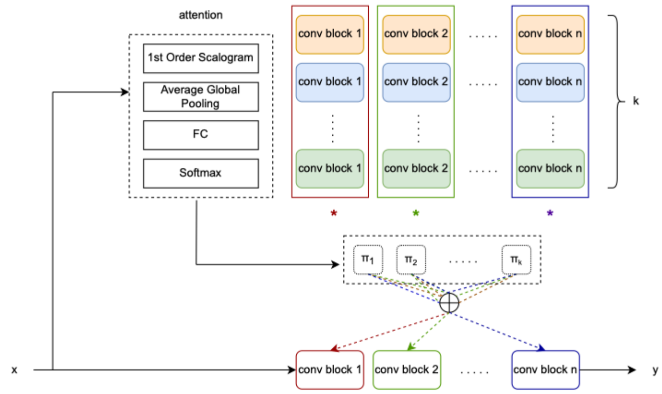

Figure 2: Dynamic convolution framework for BCI architectures. X represents input EEG signals, Y the associated MI labels. The K different subjects in the training set are represented by different colors in the convolutional blocks. Colored rectangles and arrows (namely green, red and dark blue) demonstrate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

of a CNN-based BCI network. Using a subject attention network that learns to distinguish between the available individuals, we decouple the subjects and essentially train simultaneously k parallel personalized models of the same BCI architecture, as illustrated in Figure 2.

Our proposed framework is inspired by the work of Chen et al. (2020) 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 Sections 4.1 and 4.2, these differences can be summarized as follows:

- Instead of fully trainable attention mechanisms, it utilizes our novel subject attention network (described in 4.1) 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 Chen et al. (2020) 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 kernels in our proposed framework is not a tunable hyperparameter (like in Chen et al. (2020)) but coincides with the number of available subjects in the training set.
- Instead of using the output vector of the attention mechanism as Chen et al. (2020), our framework utilizes the proposed “uniformly attended” vector **A*** (described in 4.2) in order to be more robust to the low Signal-to-Noise Ratio (SNR) of the EEG signal.

4.1 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 $\psi_\lambda(t)$ be a wavelet. The 1st order scalogram is defined as $\mathbf{X}(\lambda, t) = |\mathbf{x}(t) * \psi_\lambda(t)|$. 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:

$$\psi_\lambda(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} cos(2\pi\lambda t) \quad (4)$$

with t $\in [-\frac{W}{2}, ..., \frac{W}{2}]$ and $\frac{1}{\sigma}$ denotes the bandwidth and λ the normalized frequency of the Gabor wavelet and these two properties are the only trainable parameters of this layer. During training, λ is restricted ($\lambda \in [0, \frac{1}{2}]$) to satisfy the Nyquist theorem. The first order wavelet scalogram $\mathbf{X}(\lambda, t) \in \mathbb{R}^{TxFxC}$ (where F is the number of Gabor filters and C the number of EEG channels) 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 π.

4.2 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 kernels are trained simultaneously using the following loss function:

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

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 kernels for the different MI tasks. As also suggested in Chen et al. (2020), 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 from a new unseen subject S_x is processed, it passes firstly through the attention network and the subject attention vector $\boldsymbol{\pi}$ is computed. We empirically observed that this vector is quite sparse, and if it was used during inference, only a handful of parallel convolutional kernels would be utilized during the kernel 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" **A***. If there was no attention network, the k parallel convolutional kernel would be mixed with a uniform factor $A_i = \frac{1}{k}$. To compute the "uniformly attended vector", the uniform attention vector **A** is combined with the subject attention vector $\boldsymbol{\pi}$ 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 B for a performance comparison between using $\boldsymbol{\pi}$ and **A*** as attention). Mathematically, this operation can be described as:

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

where σ denotes the softmax operation, $\mathbf{A}$ the uniform attention vector with $A_i = \frac{1}{k}$, $\boldsymbol{\pi}$ the subject attention vector with $\sum_i \pi_i = 1$ and **A*** the "uniformly attended" vector where $\sum_i A_i^* = 1$.

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:

$$P_{S_x}(Z|Y) = A_1^* \times P_{S_1}(Z|Y) + A_2^* \times P_{S_2}(Z|Y) + ... + A_k^* \times P_{S_k}(Z|Y) \tag{7}$$

5 Experiments

To validate our proposed framework based on our causal breakdown in Section 3, two publically available MI datasets are used namely:

1. **PhysioNet** (Goldberger et al. (2000)): 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** (Kwon et al. (2020)): 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 powerline noise. We also applied a notch filter at 460Hz due to a spurious artifact (consistent across all trials).

5.1 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 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 to quickly adapt to the data) and 0.001 for the remainder 20 epochs. As shown in Table 1, 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.

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

Table 1: 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

5.2 MI Classification

We tested our proposed framework in four well-established BCI architectures, namely DeepConvNet (Schirrmeister et al. (2017)), ShallowConvNet (Schirrmeister et al. (2017)), EEGNet (Lawhern et al. (2018)) and EEG-Inception (Santamaría-Vázquez et al. (2020)) in the following MI tasks: for the publically available MI dataset Physionet (Goldberger et al. (2000)) 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 (Kwon et al. (2020)) one MI binary classification task (MI Left vs Right Hand).

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 (Table 2).

[1] $\pm$% refers to the rounded standard deviation across 10 runs of 10-fold cross-validation experiments

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

Table 2: 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.

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 3). Furthermore, we compared our framework with two other commonly used transfer learning EEG techniques: 1) an adversarial approach, namely Özdenizci et al. (2020), that (similarly to our approach) does not use a calibration set and 2) Euclidean alignment (He & Wu (2020)) 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 $\lambda = 0.005$ (hyperparameters taken from the original paper Özdenizci et al. (2020)). As it can be seen from Table 3, 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.

Model	5-Fold CV ($K \approx 83$)	10-Fold CV ($K \approx 93$)	20-Fold CV ($K \approx 98$)	Trials of Unseen Subject Used
Vanilla DeepConvNet	$81.4 \pm 1.3\%$	$81.7 \pm 2.9\%$	$82.4 \pm 5.5\%$	-
DeepConvNet with Euclidean Alignment	$82.7 \pm 1.2\%$	$83.03 \pm 3.2\%$	$\mathbf{83.2} \pm \mathbf{5.0\%}$	All Trials
Adversarial DeepConvNet ($\lambda = 0.05$)[2]	$81.4 \pm 1.35\%$	$81.8 \pm 3.7\%$	$81.9 \pm 4.6\%$	-
Dynamic DeepConvNet (Ours)	$\mathbf{82.8} \pm \mathbf{2.03\%}$	$\mathbf{83.14} \pm \mathbf{3.9\%}$	$83.2 \pm 5.3\%$	1
Vanilla ShallowConvNet	$79.7 \pm 1.75\%$	$80.4 \pm 3.3\%$	$80.95 \pm 4.6\%$	-
ShallowConvNet with Euclidean Alignment	$79.8 \pm 2.0\%$	$80.3 \pm 3.5\%$	$81.1 \pm 4.3\%$	All Trials
Adversarial ShallowConvNet ($\lambda = 0.05$)[2]	$81.0 \pm 1.26\%$	$81.3 \pm 3.7\%$	$82.35 \pm 4.65\%$	-
Dynamic ShallowConvNet (Ours)	$\mathbf{81.0} \pm \mathbf{1.06\%}$	$\mathbf{82.32} \pm \mathbf{2.86\%}$	$\mathbf{83.10} \pm \mathbf{4.7\%}$	1

Table 3: Performance of generic (trained and evaluated in a leave-M-subjects-out fashion) models of MI-classification (Left / Right hand) tasks in Physionet. CV stands for Cross-Validation across subjects

Finally, 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 4, the calibrated dynamic models also outperform their equivalent vanilla calibrated networks.

[2]$*$ early stopping has been applied to some folds during the fine-tuning phase since these particular subjects presented signs of overfitting prior to the epoch's hard boundary.

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

Table 4: 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 Shallow-ConvNet and their Calibrated Dynamic equivalent networks. CV stands for Cross-Validation across subjects

6 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 5 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 $\times$ 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 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 C, there is a less than $\times$ K increase in terms of inference time cost. Inspired by related works (Wu et al. (2019)), 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 5 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.

7 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.

REFERENCES

Kai Keng Ang, Zheng Yang Chin, Haihong Zhang, and Cuntai Guan. Filter bank common spatial pattern (fbcsp) in brain-computer interface. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 2390–2397, 2008. doi: 10.1109/IJCNN.2008.4634130.

Andreas Antoniades, Loukianos Spyrou, Clive Cheong Took, and Saeid Sanei. Deep learning for epileptic intracranial eeg data. In *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, 2016. doi: 10.1109/MLSP.2016.7738824.

Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Multiclass brain–computer interface classification by riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, 2012. doi: 10.1109/TBME.2011.2172210.

Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and Klaus-robert Muller. Optimizing spatial filters for robust eeg single-trial analysis. *IEEE Signal Processing Magazine*, 25(1):41–56, 2008. doi: 10.1109/MSP.2008.4408441.

D.C. Castro, I. Walker, and B. Glocker. Causality matters in medical imaging. Commun 11, 3673. 2020. doi: doi.org/10.1038/s41467-020-17478-w.

Junyi Chai, Hao Zeng, Anming Li, and Eric W.T. Ngai. Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications*, 6: 100134, 2021. ISSN 2666-8270. doi: https://doi.org/10.1016/j.mlwa.2021.100134.

Yinpeng Chen, Xiyang Dai, Mengchen Liu, Dongdong Chen, Lu Yuan, and Zicheng Liu. Dynamic convolution: Attention over convolution kernels. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11027–11036, 2020. doi: 10.1109/CVPR42600.2020.01104.

Jean Decety and David H. Ingvar. Brain structures participating in mental simulation of motor behavior: A neuropsychological interpretation. *Acta Psychologica*, 73(1):13–34, 1990. ISSN 0001-6918. doi: 10.1016/0001-6918(90)90056-L.

A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P.C. Ivanov, R. Mark, J.E. Mietus, G.B. Moody, C.K. Peng, and H.E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220. *Publicly Accessible Dataset*, 2000. URL `https://physionet.org/content/eegmmidb/1.0.0/`.

He He and Dongrui Wu. Transfer learning for brain–computer interfaces: A euclidean space data alignment approach. *IEEE Transactions on Biomedical Engineering*, 67(2):399–410, 2020. doi: 10.1109/TBME.2019.2913914.

Hyohyeong Kang, Yunjun Nam, and Seungjin Choi. Composite common spatial pattern for subject-to-subject transfer. *IEEE Signal Processing Letters*, 16:683–686, 2009.

O-Yeon Kwon, Min-Ho Lee, Cuntai Guan, and Seong-Whan Lee. Subject-independent brain–computer interfaces based on deep convolutional neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 31(10):3839–3852, 2020. doi: 10.1109/TNNLS.2019.2946869. URL `https://github.com/PatternRecognition/OpenBMI`.

Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 15(5):056013, jul 2018. doi: 10.1088/1741-2552/aace8c.

Martin Längkvist, Lars Karlsson, and Amy Loutfi. Sleep stage classification using unsupervised feature learning. *Advances in Artificial Neural Systems*, 2012, 07 2012. doi: 10.1155/2012/107046.

Ravikiran Mane, Tushar Chouhan, and Cuntai Guan. BCI for stroke rehabilitation: motor and beyond. *Journal of Neural Engineering*, 17(4):041001, aug 2020. doi: 10.1088/1741-2552/aba162.

Sebastien Marcel and Jose Del R. Millan. Person authentication using brainwaves (eeg) and maximum a posteriori model adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):743–752, 2007. doi: 10.1109/TPAMI.2007.1012.

D.J. McFarland, C.W. Anderson, K.-R. Muller, A. Schlogl, and D.J. Krusienski. Bci meeting 2005-workshop on bci signal processing: feature extraction and translation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):135–138, 2006. doi: 10.1109/TNSRE.2006.875637.

Alexander Neergaard Olesen, Poul Jennum, Emmanuel Mignot, and Helge B. D. Sorensen. Deep transfer learning for improving single-eeg arousal detection. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pp. 99–103, 2020. doi: 10.1109/EMBC44109.2020.9176723.

Yousef Rezaeitabar and Ugur Halici. A novel deep learning approach for classification of eeg motor imagery signals. *Journal of Neural Engineering*, 14:016003, 02 2017. doi: 10.1088/1741-2560/14/1/016003.

Neethu Robinson, Ravikiran Mane, Tushar Chouhan, and Cuntai Guan. Emerging trends in bci-robotics for motor control and rehabilitation. *Current Opinion in Biomedical Engineering*, 20: 100354, 2021. ISSN 2468-4511. doi: 10.1016/j.cobme.2021.100354.

Simanto Saha and Mathias Baumert. Intra- and inter-subject variability in eeg-based sensorimotor brain computer interface: A review. *Frontiers in Computational Neuroscience*, 13, 2020. ISSN 1662-5188. doi: 10.3389/fncom.2019.00087.

Eduardo Santamaría-Vázquez, Víctor Martínez-Cagigal, Fernando Vaquerizo-Villar, and Roberto Hornero. Eeg-inception: A novel deep convolutional neural network for assistive erp-based brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28 (12):2773–2782, 2020. doi: 10.1109/TNSRE.2020.3048106.

Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human Brain Mapping*, 38(11):5391–5420, 2017. doi: https://doi.org/10.1002/hbm.23730.

Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, 2021. doi: 10.1109/JPROC.2021.3058954.

Marc Sebastián-Romagosa, Woosang Cho, Rupert Ortner, Nensi Murovec, Tim Von Oertzen, Kyousuke Kamada, Brendan Z. Allison, and Christoph Guger. Brain computer interface treatment for motor rehabilitation of upper extremity of stroke patients—a feasibility study. *Frontiers in Neuroscience*, 14:1056, 2020. ISSN 1662-453X. doi: 10.3389/fnins.2020.591435.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, 2015. doi: 10.1109/CVPR.2015.7298594.

Akshay Valsaraj, Ithihas Madala, Nikhil Garg, Mohit Patil, and Veeky Baths. Motor imagery based multimodal biometric user authentication system using eeg. In *2020 International Conference on Cyberworlds (CW)*, pp. 272–279, 2020. doi: 10.1109/CW49994.2020.00050.

Felix Wu, Angela Fan, Alexei Baevski, Yann Dauphin, and Michael Auli. Pay less attention with lightweight and dynamic convolutions. In *International Conference on Learning Representations*, 2019.

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. 8(3):036015, apr 2011. doi: 10.1088/1741-2560/8/3/036015.

Liuyin Yang, Arno Libert, and Marc M. Van Hulle. Chronic study on brainwave authentication in a real-life setting: An lstm-based bagging approach. *Biosensors*, 11(10), 2021. ISSN 2079-6374. doi: 10.3390/bios11100404.

Kaishuo Zhang, Neethu Robinson, Seong-Whan Lee, and Cuntai Guan. Adaptive transfer learning for eeg motor imagery classification with deep convolutional neural network. *Neural Networks*, 136:1–10, 2021a. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2020.12.013.

Ruilong Zhang, Qun Zong, Liqian Dou, Xinyi Zhao, Yifan Tang, and Zhiyu Li. Hybrid deep neural network using transfer learning for eeg motor imagery decoding. *Biomedical Signal Processing and Control*, 63:102144, 2021b. ISSN 1746-8094. doi: https://doi.org/10.1016/j.bspc.2020.102144.

Dongye Zhao, Fengzhen Tang, Bailu Si, and Xisheng Feng. Learning joint space–time–frequency features for eeg decoding on small labeled data. *Neural Networks*, 114:67–77, 2019. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2019.02.009.

Ozan Özdenizci, Ye Wang, Toshiaki Koike-Akino, and Deniz Erdoğmuş. Learning invariant representations from eeg via adversarial inference. *IEEE Access*, 8:27074–27085, 2020. doi: 10.1109/ACCESS.2020.2971600.

A Investigation of Negative Transfer Learning

Although our proposed framework showed increased cross-subject performance as experimentally demonstrated in Section 5, 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, the vast majority is either marginally or significantly better compared to the vanilla architectures.

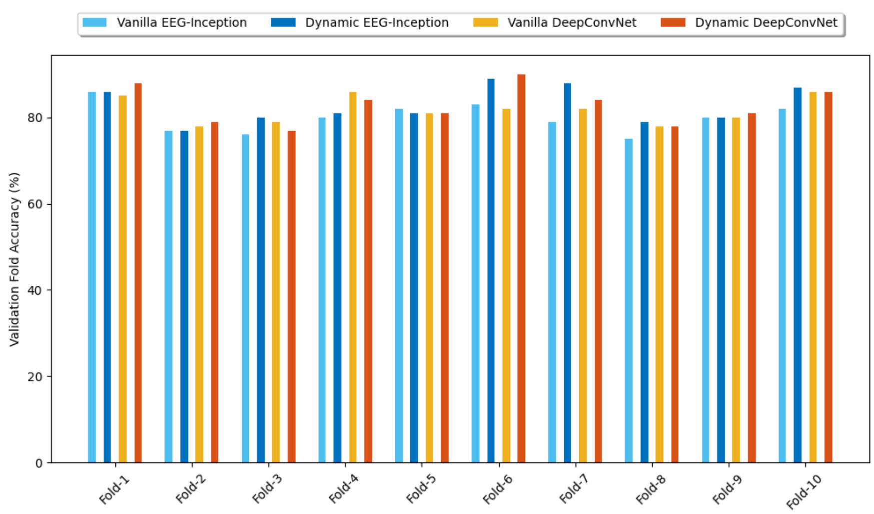

Figure 3: Per-fold comparison of the performance of vanilla architectures versus their equivalent dynamic networks (ours)

B Comparison between vectors π and $A*$

As described in Section 4, 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 $\boldsymbol{\pi}$ 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" **A*** that uses knowledge from all k individuals and "shift" the attention more to the most relevant subjects. A comparison between using the vector $\boldsymbol{\pi}$ versus our proposed uniformly attended vector $\boldsymbol{A}$***** as attention in our proposed dynamic framework can be seen in the following Figure 4.

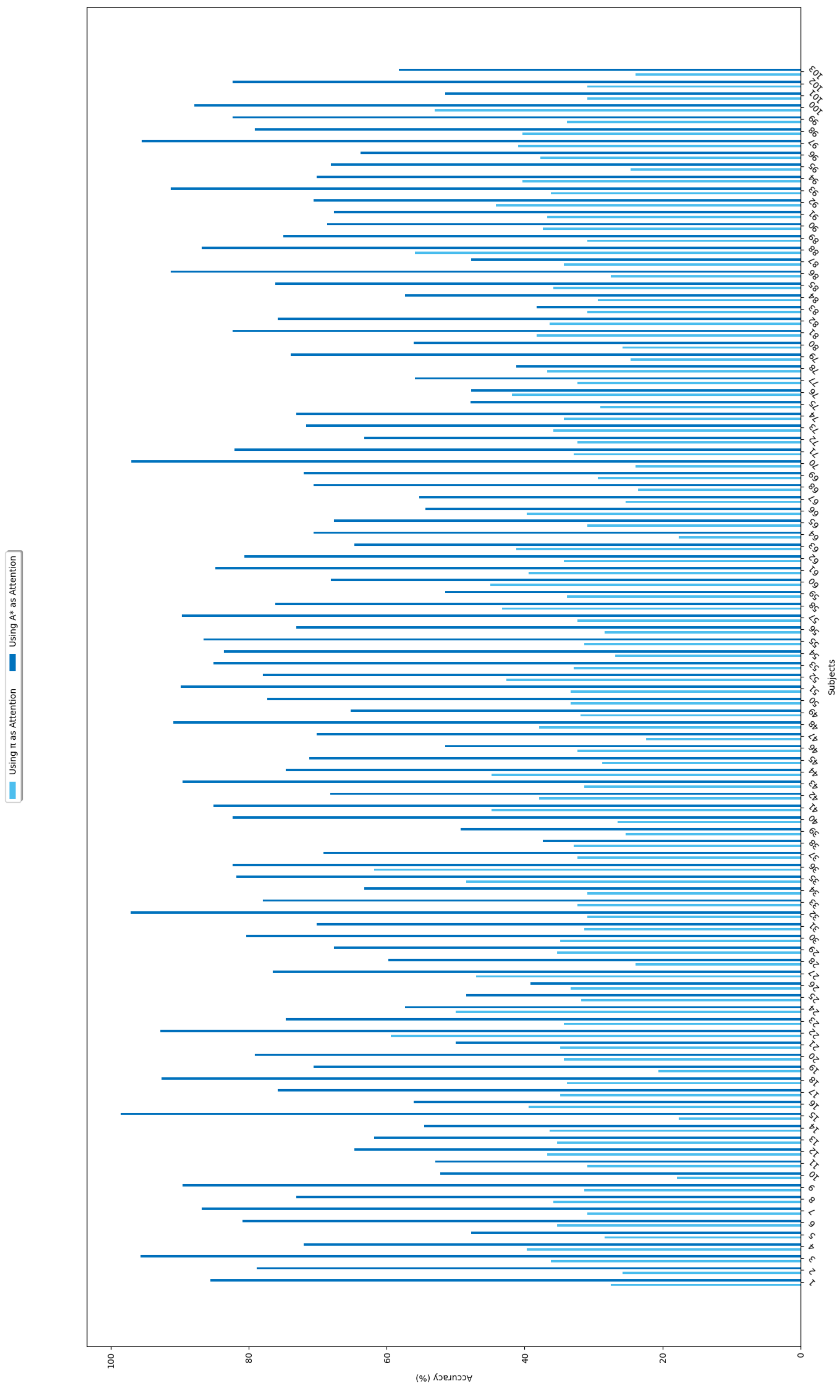

Figure 4: Comparison between using the vector π versus our proposed uniformly attended vector A^* as attention in our proposed dynamic framework. The performances correspond to generic (trained and evaluated in a leave-one-subject-out fashion) Dynamic EEG-Inception models in the MI-classification (Left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

C INFERENCE TIMINGS COMPARISONS

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 tables, 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.

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
Subject 11	544.8242	898.6418
Subject 12	545.5861	899.7143
Subject 13	543.9782	1137.1084
Subject 14	524.6781	880.7658
Subject 15	546.7437	892.9803
Subject 16	539.1946	881.5262
Subject 17	551.6413	876.1219
Subject 18	555.4468	895.4225
Subject 19	553.101	1142.7909
Subject 20	536.0449	1031.1428
Subject 21	535.9892	870.7953
Subject 22	548.059	1156.7307
Subject 23	535.9896	886.2561
Subject 24	552.912	911.1381
Subject 25	526.7927	1124.0788
Subject 26	551.9285	910.0632
Subject 27	536.4123	891.3182
Subject 28	539.1051	1117.4125
Subject 29	539.3462	905.8769
Subject 30	529.274	872.0497
Subject 31	542.4147	1116.2725
Subject 32	542.7398	894.614
Subject 33	543.7969	900.8855
Subject 34	546.0942	915.9275
Subject 35	522.5048	868.7032
Subject 36	553.5467	1147.1854
Subject 37	544.6818	1151.4641
Subject 38	533.6177	895.8355
Subject 39	548.0503	877.6697
Subject 40	544.3172	894.4753

Subject	Standard (ms)	Dynamic (ms)
Subject 41	544.4666	887.1042
Subject 42	530.282	876.1537
Subject 43	545.9131	890.2133
Subject 44	541.8491	893.9466
Subject 45	525.7591	1099.8516
Subject 46	546.3638	894.8608
Subject 47	544.9518	1131.7571
Subject 48	540.1192	901.2862
Subject 49	552.4357	903.1747
Subject 50	526.1646	1104.5922
Subject 51	572.2273	909.8803
Subject 52	543.0584	907.7985
Subject 53	543.0958	891.8089
Subject 54	539.5095	1073.8354
Subject 55	536.8121	1125.4846
Subject 56	540.3361	1102.8461
Subject 57	549.2922	894.1009
Subject 58	537.4204	1137.1086
Subject 59	546.1475	1134.5581
Subject 60	560.6594	1152.4353
Subject 61	548.8822	948.3445
Subject 62	540.4975	1119.9238
Subject 63	546.902	1152.3956
Subject 64	541.2633	1152.5192
Subject 65	545.0002	927.2082
Subject 66	544.3959	1130.6614
Subject 67	542.9522	1133.8398
Subject 68	551.0722	913.637
Subject 69	549.351	1136.2486
Subject 70	543.6982	1129.6688
Subject 71	541.5467	1118.4973
Subject 72	560.7404	1137.3549
Subject 73	538.0839	885.301
Subject 74	539.8588	901.1848
Subject 75	551.6779	1142.0281
Subject 76	535.3296	889.4693
Subject 77	545.8409	1137.8778
Subject 78	539.4993	905.6306
Subject 79	549.3963	1150.3135
Subject 80	537.3423	1010.6793

Table 5: Inference timings for 80 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	293.1013	592.0185
Subject 2	261.3001	567.4535
Subject 3	277.6414	762.4132
Subject 4	332.2469	714.878
Subject 5	307.2616	579.3735
Subject 6	323.8485	1040.4726
Subject 7	419.3959	817.5334
Subject 8	287.3716	792.4063
Subject 9	268.1497	770.9701
Subject 10	269.5422	814.7743
Subject 11	363.2819	877.2561
Subject 12	339.4558	924.5848
Subject 13	409.5779	932.7032
Subject 14	297.9082	583.1962
Subject 15	271.8576	851.6035
Subject 16	276.3279	917.3601
Subject 17	274.9757	617.1784
Subject 18	271.5192	666.8381
Subject 19	266.7319	817.1083
Subject 20	352.0864	939.2702
Subject 21	269.7336	567.0578
Subject 22	270.1747	587.6435
Subject 23	263.6932	812.134
Subject 24	325.0687	652.048
Subject 25	269.4698	621.5127
Subject 26	295.0868	634.1263
Subject 27	406.8556	1244.4247
Subject 28	526.1524	767.5288
Subject 29	346.6407	666.8721
Subject 30	267.8148	585.2428
Subject 31	274.5486	676.3858
Subject 32	277.2021	862.4104
Subject 33	266.4326	576.9763
Subject 34	370.4928	649.1246
Subject 35	376.0236	833.6711
Subject 36	320.1059	942.6238
Subject 37	409.8353	1095.825
Subject 38	269.5554	577.8968
Subject 39	262.9665	576.0119
Subject 40	266.718	632.9843
Subject 41	272.6256	582.0228
Subject 42	262.7292	565.8136
Subject 43	267.493	567.319
Subject 44	284.8254	621.0434
Subject 45	304.4975	843.2986
Subject 46	340.2609	653.5012
Subject 47	263.6114	572.389
Subject 48	268.7476	578.8455
Subject 49	272.6755	598.1583
Subject 50	264.9827	582.6363
Subject 51	290.8646	872.6904
Subject 52	309.326	678.8556
Subject 53	272.7245	861.544
Subject 54	303.32	597.5879
Subject 55	277.3168	630.2398
Subject 56	271.6752	991.2502
Subject 57	318.5117	972.6068
Subject 58	294.7329	1045.9693
Subject 59	275.9514	815.0113
Subject 60	349.2239	810.4091
Subject 61	263.9984	577.374
Subject 62	269.4928	569.9094
Subject 63	290.156	587.2844
Subject 64	268.5258	831.6206
Subject 65	270.2256	823.4725
Subject 66	335.8103	735.2804
Subject 67	265.9914	826.7936
Subject 68	271.2678	588.4226
Subject 69	269.1232	580.561
Subject 70	263.2801	568.7038
Subject 71	263.3383	567.7978
Subject 72	265.307	828.3046
Subject 73	264.4986	596.5381
Subject 74	262.5872	824.8643
Subject 75	352.4282	706.9853
Subject 76	319.8665	717.6244
Subject 77	267.0403	585.0451
Subject 78	268.896	822.8302
Subject 79	302.2626	702.9593
Subject 80	267.0334	585.4106
Subject 81	266.3547	586.8337
Subject 82	260.3958	587.4193
Subject 83	363.4541	773.0845
Subject 84	283.8774	773.0845
Subject 85	289.3468	610.5704
Subject 86	281.237	847.6461
Subject 87	274.8509	1019.282
Subject 90	315.6247	884.607
Subject 91	273.0971	616.6718
Subject 93	266.9891	821.2341
Subject 94	266.7666	579.734
Subject 95	267.361	588.8292
Subject 96	277.6225	688.1226
Subject 97	267.322	603.2024
Subject 98	272.2102	844.2862
Subject 99	444.3941	669.1419
Subject 101	288.218	720.4358
Subject 102	367.7827	957.0791
Subject 103	272.9728	618.8917
Subject 105	262.0691	799.0899
Subject 107	271.324	582.4897
Subject 108	351.929	579.6055
Subject 109	269.9653	591.113

Table 6: Inference timings for all trained models of MI-classification (Left / Right hand and Left / Right hand / Feet) from the leave-one-subject-out cross-validation for EEG-Net in Physionet. Measured with $torch.autograd.profiler$ in 2.9 GHz 6-core CPU Intel Core i9.