Domain Adaptation and Transfer Learning methods enhance Deep Learning Models used in Inner Speech Based Brain Computer Interfaces

Luciano Ivan Zablocki[1,#], Agustín Nicolás Mendoza[1,#], Nicolás Nieto[2,3,*]

[1] Facultad de Ingeniería y Ciencias Hídricas, FICH, UNL # Equally contributing authors.
[2] Instituto de Matemática Aplicada del Litoral, IMAL, UNL-CONICET.
[3] Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, sinc(i), UNL-CONICET *nnieto@sinc.unl.edu.ar.

Abstract. Brain Computer Interfaces are useful devices that can partially restore the communication from severe compromised patients. Although the advances in deep learning have significantly improved brain pattern recognition, a large amount of data is required for training these deep architectures. In the last years, the inner speech paradigm has drew much attention, as it can potentially allow a natural control of different devices. However, as of the date of this publication, there is only a small amount of data available in this paradigm. In this work we show that it is possible, by means of transfer learning and domain adaptation methods, to make the most of the scarce data, enhancing the training process of a deep learning architecture used in brain computer interfaces.

Keywords: Deep Learning · Domain Adaptation · Transfer Learning · Convolutional Neural Network

1 Introduction

Brain Computer Interfaces (BCIs) are useful tools that allow users to control external devices only by using their cerebral activity [18]. These technologies are fundamental for patients who suffer from strokes, amyotrophic lateral sclerosis, or any other accident that may interrupt the normal communication between the brain and the muscles. By means of a BCI, these patients can partially restore their capability to communicate and interact with the environment, significantly improving their life quality [15].

The surface electroencephalography (EEG) is widely used for measuring the brain activity in a BCI, as it is a standard and noninvasive technique [10]. EEG provides signals with good time resolution, allowing real time applications, but with a low signal-noise ratio and with a high inter and intra-subject variability [11]. One of the existing BCI paradigms is the inner speech, which refers to the mental process of imagining one's own voice, and possibly enabling a natural method for controlling different devices, just by thinking of a certain command [11].

In the last decades, the rise of deep learning has been of great benefit for the BCIs, mainly with the use of Convolutional Neural Networks (CNNs) as brain pattern classifiers [17,3,9]. An important downside of these deep structures is that they require a big amount of data to train its many parameters. As inner speech is a relatively recent paradigm there is only a limited amount of data available for training deep models.

As mentioned before, the high variability in the EEG data makes the BCIs an instrument that needs to be highly personalized [16]. Several transfer learning and domain adaptation techniques are paving the way to take the most of the already scarce data [14,6,19,4,16], and to solve these potential differences in the data.

In this work we combine a domain adaptation technique [6]. that allows us to use data from different subjects (domains). We also combine this method with a transfer learning method and pre-train a deep structure on several different subjects and then fine-tune the network with the target subject. We show that this process enhances the training process of a CNN used in an BCI.

2 Materials and Methods

2.1 Data description

Tasks and participants. The dataset presented in [12], which contains EEG signals from ten healthy participants, was used for the experiments. The participants performed three different paradigms: pronounced speech, inner speech and visualized condition. In this study, we focus on distinguishing between the signals produced in the inner speech and the visualized condition paradigms, as it was shown that is a more challenging task [13].

The EEG data were acquired with a BioSemi ActiveTwo acquisition system of 128+8 channels at 1024 Hz. It was later down-sampled 4 times and a standard EEG signal processing was applied to each subject. The number of the available trials varied among subjects. A more detailed description of the acquisition procedures, number of trials for each subject and preprocessing can be found in [12].

EEG processing. Each trial has a duration of 2.5 seconds but only the last 2 seconds were used, preventing any protocol-evoked EEG potential described in [12]. The trials were later split into ten smaller trials of 200ms (51 samples) with the main goal of data augmentation. This procedure was performed after separating the trials between train, validation and test folds, avoiding data leakage. Finally, the data was scaled between 0 and 1 with a minmax scaler that finds its parameters in the train set.

2.2 Classification algorithm

Convolutional Neural Networks. These architectures are one of the most used in deep learning, as it allows end-to-end learning, working as both feature extractor and classifier. They were originally introduced in [7] and had a huge impact on artificial intelligence in general and in the BCIs application in particular ever since ever since [17,3,9,5].

The convolutional architecture used in this work is summarize in Table 1.1. The kernel size was 3x3, with a stride and a padding equal to 1. The dropouts probabilities were set to 0.2 for the first two layers, and 0.3 for the last one. The activation function used was the Exponential Linear Unit (ELU). The network had 13,818 trainable parameters that were in their majority parameters that correspond to the linear layers. This architecture was obtained as the network with the minimum number of parameters that converge in the validation set.

Table 1.1: Model Architecture

Layer	Output Shape	Param #
INet	[128, 2]	--
Conv2d	[128, 4, 128, 51]	40
MaxPool2d	[128, 4, 64, 25]	--
Dropout	[128, 4, 64, 25]	--
Conv2d	[128, 8, 64, 25]	296
MaxPool2d	[128, 8, 32, 12]	--
Dropout	[128, 8, 32, 12]	--
Conv2d	[128, 16, 32, 12]	1,168
MaxPool2d	[128, 16, 16, 6]	--
Dropout	[128, 16, 16, 6]	--
Linear	[128, 8]	12,296
Linear	[128, 2]	18

Domain Adaptation and Transfer Learning. As noted before, the EEG data has a high inter and intra-subject variability, requiring highly personalized BCIs [16]. In a traditional approach, the classifiers are trained only on the data from one subject, discarding the data from the others subjects: we call this approach "Simple Model". Aiming to make the most out of the available data, we use the Euclidean Alignment method [6]. We apply this alignment between the data from 9 subjects that were not the target subject. Then, we pre-train a CNN with the aligned data of the 9 subjects. Pre-training was performed using a learning rate of 0.001 and 200 training epochs. Early stopping was also used, with 20 epochs of tolerance, but with a minimum of 40 epochs of training. The used batch size was 1280. Finally we transfer this pretrained model's weights and fine-tune the whole network parameters with the data from the subject of interest, as in [2]. The data of the target subject is aligned using only the train fold, emulating a real-world application, in which this data will be available just at the fine-tuning phase. We call this approach the "Pretrain-finetune Model". Finally, aiming to demonstrate the relevance of the fine-tuning phase, we test the pretrained model over the aligned data but without fine tuning the network with the training fold of the subject of interest. We call this approach the "Pretrain Model". Fine tuning phase was done using a learning rate of 0.001 and 200 training epochs. The early stopping tolerance and minimum epochs was 100 and 40, respectively. Chosen batch size was 128. AdamW [8] was used as optimizer for both parts of the training process.

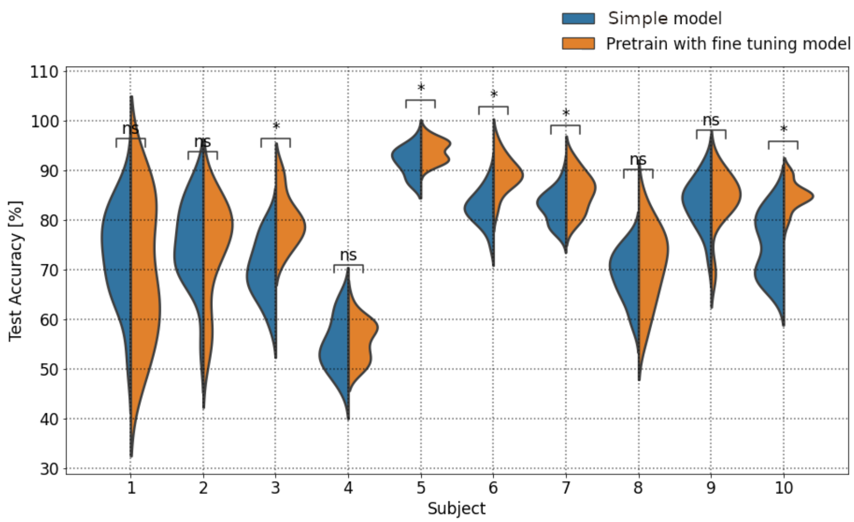

Fig. 1: Violin plots for test accuracy distribution obtained with Simple and Pretrain-finetune models. For each subject the statistical significance according to a T-Test-Paired is marked with " * " ($p \leq 0.01$).

3 Results

Aiming to demonstrate the difference between the three methods, a K-fold cross validation was used, with K = 20, splitting the data in train, validation and test folds, with 70%, 10%, 20%, respectively. For all models, the test folds were the same, aiming to make a paired and more fair comparison. We obtain a performance distribution for each subject and each method, as depicted in Figure 1 and Figure 2. For five out of ten subjects, the accuracy distribution was significantly improved by the "Pretrain-finetune Model" with respect to the "Simple Model", according to a T-Test-Paired with a significance threshold of 0.01. In the rest of the subjects, the "Pretrain-finetune Model" accuracy was similar or not significantly worse that the one obtained from the "Simple Model". Figure 2 clearly shows that the fine tuning step is crucial for the network to work properly. This is mainly due to the fact that the distributions are highly variable and the test data is only aligned with the data of the same subject and not with the data used in the pretraining step. We did not include the target's train fold in

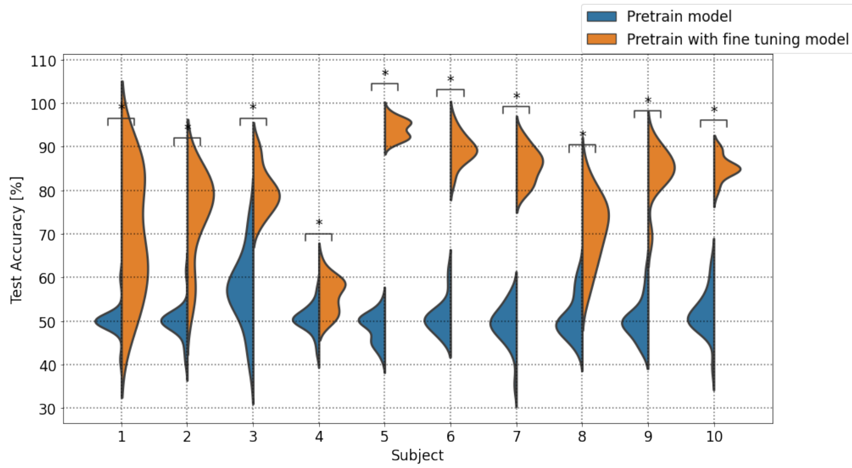

Fig. 2: Violin plots for test accuracy distribution obtained with Pretrain and Pretrain-finetune models. For each subject the statistical significance according to a T-Test-Paired is marked with " * " ($p \leq 0.01$).

the alignment process because it is how the network will be used in real time applications.

A more detailed view of the results is also shown in Table 1.2.

Table 1.2: Test accuracy comparison

	PRETRAIN MODEL		SIMPLE MODEL		PRETRAIN-FINTUNE MODEL	
Sub	Acc	Std	Acc	Std	Acc	Std
1	0.5006	0.0303	0.7495	0.1114	0.6963	0.1336
2	0.4974	0.0375	0.7968	0.0842	0.7417	0.0964
3	0.5713	0.0800	0.7907	0.0576	0.8006	0.0494
4	0.5036	0.0297	0.5525	0.0585	0.5583	0.0388
5	0.4854	0.0314	0.9401	0.0307	0.9406	0.0205
6	0.5092	0.0380	0.8232	0.0814	0.8943	0.0339
7	0.4906	0.0450	0.8442	0.0428	0.8578	0.0380
8	0.5088	0.0465	0.6249	0.0591	0.7094	0.0741
9	0.5000	0.0421	0.8370	0.0795	0.8406	0.0563
10	0.5177	0.0483	0.7817	0.1416	0.8516	0.0266

4 Conclusions and Discussion

We demonstrated that the proposed approach can incorporate new information to the classifier, using the whole available data. Needless to say, much further work has to be done, but this is without doubt a path that could be followed

in order to make the most out of the scarce available data when training deep learning models in BCIs applications.
A future path to explore could be to analyze if it is beneficial to exclude subjects from the pretraining step, as some of them are more difficult to classify, and therefore can be left out of the pretraining process if the information provided is not significant. This could improve the test accuracy obtained in the fine tuning phase, and reduce the pretraining time. Also, the whole 128 available sensors were used as an input to CNN. Another approach, used in [1], is to use just a subset of channels, giving more importance to the ones that are related to the left-hemisphere of the brain, where more speech related brain activity should be present.
For the fine-tuning training, the weights of the last linear layer are reinitialized, and the whole model is trained again with the subject that will be tested. We can freeze some of the layers, leaving the parameters unchanged, or we can reinitialize more layers than just the last one.
Finally, other performance indicators, as training speed, convergence or loss, could also be improved by the proposed approach. We did not analyze them in this work but could also be important in the future, as these values are crucial for real applications.

References

1. van den Berg, B., van Donkelaar, S., Alimardani, M.: Inner speech classification using EEG signals: A deep learning approach. 2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS) pp. 1–4 (2021)
2. Chu, B., Madhavan, V., Beijbom, O., Hoffman, J., Darrell, T.: Best practices for fine-tuning visual classifiers to new domains. European conference on computer vision (2016)
3. Cooney, C., Korik, A., Raffaella, F., Coyle, D.: Classification of imagined spoken word-pairs using convolutional neural networks. In: The 8th Graz BCI Conference, 2019. pp. 338–343 (2019)
4. Courty, N., Flamary, R., Tuia, D., Rakotomamonjy, A.: Optimal transport for domain adaptation. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(9), 1853–1865 (2017)
5. Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. MIT press (2016)
6. He, H., Wu, D.: Transfer learning for brain–computer interfaces: A euclidean space data alignment approach. IEEE Transactions on Biomedical Engineering **67**(2), 399–410 (2019)
7. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998)
8. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
9. Mao, W., Fathurrahman, H., Lee, Y., Chang, T.: EEG dataset classification using CNN method. In: Journal of physics: conference series. vol. 1456, p. 012017. IOP Publishing (2020)
10. Nicolas-Alonso, L.F., Gomez-Gil, J.: Brain computer interfaces, a review. sensors **12**(2), 1211–1279 (2012)

11. Nieto, N.: Algoritmos para interfaces cerebro-computadora en paradigmas relacionados con el habla (2022)
12. Nieto, N., Peterson, V., Rufiner, H.L., Kamienkowski, J.E., Spies, R.: Thinking out loud, an open-access EEG-based BCI dataset for inner speech recognition. Scientific Data **9**(1), 1–17 (2022)
13. Nieto, N., Runer, H.L., Spies, R.: Preliminary feasibility analysis of inner speech as a control paradigm for brain-computer interfaces. In: XXII Simposio Argentino de Inteligencia artificial (ASSAI 2021)-JAIIO 50 (Modalidad virtual) (2021)
14. Peterson, V., Nieto, N., Wyser, D., Lambercy, O., Gassert, R., Milone, D.H., Spies, R.D.: Transfer learning based on optimal transport for motor imagery brain-computer interfaces. IEEE Transactions on Biomedical Engineering **69**(2), 807–817 (2021)
15. Rousseau, M.C., Baumstarck, K., Alessandrini, M., Blandin, V., De Villemeur, T.B., Auquier, P.: Quality of life in patients with locked-in syndrome: Evolution over a 6-year period. Orphanet journal of rare diseases **10**(1), 1–8 (2015)
16. Samek, W., Meinecke, F.C., Müller, K.R.: Transferring subspaces between subjects in brain–computer interfacing. IEEE Transactions on Biomedical Engineering **60**(8), 2289–2298 (2013)
17. Tang, Z., Li, C., Sun, S.: Single-trial EEG classification of motor imagery using deep convolutional neural networks. Optik **130**, 11–18 (2017)
18. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain-computer interfaces for communication and control. Clinical Neurophysiology **113**(6), 767–791 (2002)
19. Wu, D., Xu, Y., Lu, B.: Transfer learning for EEG-Based Brain-Computer Interfaces: A review of progresses since 2016. arXiv preprint arXiv:2004.06286 (2020)