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

10.3217/978-3-99161-014-4-072

INTER-TASK TRANSFER LEARNING BETWEEN UPPER-LIMB MOTOR EXECUTION AND MOTOR IMAGERY

S. Pérez-Velasco [1, 2], D. Marcos-Martínez[1, 2], E. Santamaría-Vázquez[1, 2], V. Martínez-Cagigal[1, 2], B. Pascual-Roa[1], R. Hornero[1, 2]

[1]Biomedical Engineering Group, E.T.S Ingenieros de Telecomunicación, University of Valladolid, Valladolid, Spain
[2]Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Madrid, Spain

E-mail: sergio.perezv@uva.es

ABSTRACT: This study addresses a key challenge in motor imagery (MI)-based brain-computer interfaces (BCIs): improving the decoding accuracy of electroencephalography (EEG) signals. We investigate the inter-task transfer learning potential between motor execution (ME) and MI to enhance the calibration phase of MI-BCIs. Utilizing the *EEGSym* deep learning network, we demonstrate that ME data can effectively train models for MI classification. Additionally, our analysis identifies a significant positive correlation between performances on ME and MI tasks. These findings support the feasibility of a ME-based calibration approach for MI tasks in BCI systems, leveraging the neural and functional similarities between ME and MI. This approach maintains BCI performance and potentially makes it easier to accommodate new users to the MI task while recording ME data during calibration, which could serve as an indicator of the expected MI accuracy. Furthermore, our results suggest that we can exploit the synergies between ME and MI without significantly reducing decoding accuracy of the user's intentions.

INTRODUCTION

Brain-computer interfaces (BCIs) offer a novel communication channel, directly linking the human brain to external devices [1].These systems are designed as closed-loop systems with three stages: the recording of brain activity, the processing of this data to interpret the user's intent, and providing feedback to the user. Electroencephalography (EEG) is favored in the recording stage for its non-invasive nature, portability, and excellent temporal resolution [2]. Furthermore, it is more affordable than the alternative techniques used for capturing brain dynamics. An EEG-based BCI system captures the brain's electrical activity using electrodes placed on the scalp. In the processing stage, these signals are analyzed to decode the user's intentions [1]. The processed information then translates into feedback, which could be provided as visual cues on a monitor or the manipulation of a prosthetic limb [3]. Despite EEG's advantages, the technique faces significant hurdles, such as its inherently low spatial resolution and the challenge of a poor signal-to-noise ratio (SNR). BCIs, therefore, employ various paradigms to generate recognizable brain patterns in the EEG, facilitating the decoding process. Motor imagery (MI), the voluntary simulation of movement without physical execution, is one paradigm that has gained increased research interest. MI activates the primary motor cortex and associated motor regions, mirroring the neural activation patterns observed during motor execution (ME) [4–6]. This neural overlap between MI and ME has critical implications, particularly in rehabilitative contexts. Research demonstrates that employing MI-based BCIs in a closed-loop system, complemented by functional electrical stimulation as feedback, can significantly bolster brain plasticity. Moreover, such targeted interventions have been crucial in enhancing ME capabilities among stroke patients [6].

Nonetheless, one major drawback of MI-based BCIs lies in the difficulty of achieving high enough decoding accuracy from EEG signals. Conventional machine learning (ML) approaches often struggle with BCI inefficiency [7], a phenomenon where BCI systems cannot reliably interpret and extract distinct features from an individual's EEG signals, impacting an estimated 10-50% of users in MI-based BCI applications [8]. Such users fail to attain effective BCI control, a condition that prior research identifies as exceeding a threshold 70% accuracy in binary MI tasks [9, 10]. This inefficiency has been attributed to the shortcomings in the classification stage [11], recording system limitations, or diminished user motivation over prolonged skill acquisition periods [12]. Moreover, there are elusive additional factors that further contribute to BCI inefficiency. Given that classical ML techniques need a calibration stage at the start of each session to address inter-subject and inter-session variability [13], ensuring this calibration phase captures accurate and relevant information becomes critical for the session's subsequent success. However, verifying whether users have correctly comprehended the instructions or are engaging in the MI task poses a significant challenge. An

CC BY
https://creativecommons.org/licenses/by/4.0/deed.en
This CC license does not apply to third party material and content noted otherwise.

Published by
Verlag der Technischen Universität Graz

 10.3217/978-3-99161-014-4-072

inadequate calibration run can result in confusing feedback, reducing user motivation and potentially leading to BCI inefficiency. This underscores the importance of having a more robust calibration process that can effectively minimize these issues.

Prior research has identified the calibration phase in BCIs as a significant bottleneck, proposing the use of deep learning (DL) models with strong transfer learning capabilities to overcome the inter-subject and inter-session variability [11, 14, 15]. These DL architectures could be employed in a calibration-less scenario by leveraging MI trials from various users. However, this strategy encounters limitations since the data used to train has been recorded providing feedback of a similar flawed calibration, or it is collected without providing any feedback to the user or the observer on the MI being properly performed. In this work, we investigate a less explored strategy: utilizing inter-task transfer learning not just across users but also between ME and MI paradigms [16–18]. Lee *et al.* [16] and Miao *et al.* [18] demonstrate that a model trained in ME data can effectively translate into an MI task with a minimal amount of MI examples. Shuqfa *et al.* [17] employ data from both ME and MI trials to train their classifiers simultaneously, aiming to improve accuracy due to the similarities between the tasks. We will further evaluate the feasibility of using data from users performing ME to classify MI trials, specifically excluding MI trials in the training set. This approach could enable objective verification of ME activity being performed by visual inspection, thus eliciting discernible brain patterns, which will isolate errors to the recording system. Moreover, it could potentially enhance the preparation for rehabilitation-focused MI-based BCI applications through closer alignment with actual ME brain patterns. Furthermore, leveraging ME data from multiple users could be more reliable to train deep learning models that focus on more relevant brain patterns.

Our research investigates the correlation between upper-limb ME and MI, assessing the extent of transfer learning capabilities of *EEGSym* [11], a DL network previously validated in inter-subject MI classification. Additionally, we aim to elucidate the impact of current EEG recording system limitation on BCI inefficiency by examining the performance correlation between ME and MI tasks. To accomplish this goals, we analyze public database of non-invasive EEG recordings from 109 healthy users performing MI and ME tasks without feedback [19].

MATERIALS AND METHODS

Dataset and preprocessing:

The Physionet dataset [19] encompasses recordings from 109 healthy participants during one session. These sessions included one run of 42-46 trials focusing on MI without feedback and another run on ME. In both runs, the duration of the imagination or execution phase for each trial was 3 seconds. The 64-channel EEG signal was recorded using the BCI200 system [20]. The majority of the dataset, covering 105 participants, was recorded at a sampling frequency of 160 Hz, while the recordings from the remaining 4 participants were captured at 128 Hz.

Prior to inputting the dataset into the DL network for classification, we perform a structured preprocessing pipeline, detailed as follows: (1) we apply a notch filter to remove the power line signal, (2) we perform common average reference (CAR) spatial filtering, (3) we do a resampling to 128 Hz to homogenize the dataset across the different sampling rates of the input for the DL model, (4) we extract the trials with a time window length of 3 seconds after the onset, and (5) we apply channel-wise z-score standardization on each trial.

DL architecture and training:

The open implementation of *EEGSym* [11] will be used for classification. *EEGSym* introduces a pioneering convolutional neural network (CNN) architecture designed for the classification of MI across different subjects presented in our prior work [11]. Leveraging cutting-edge DL methodologies, *EEGSym* incorporates residual connections, implements data augmentation strategies, employs inter-subject transfer learning, and features a siamese-network design that capitalizes on the inherent symmetry of the brain along the mid-sagittal plane. This CNN has demonstrated significantly improved accuracy in binary MI inter-subject classification, outperforming the performance of four previously established CNNs developed for EEG classification: ShallowConvNet and DeepConvNet [21], EEGNet [22], and EEG-Inception [23]. T*EEGSym* achieved groundbreaking results, setting a new benchmark for accuracy in inter-subject MI classification.

The selection of this networks is primarily motivated by its tailored design for inter-subject classification scenarios, which was proven by its superior performance in such tasks. It emerges as one of the better choices to discern and emphasize patterns universally present among users engaged in both MI and ME tasks [7]. This property is expected to also boost transfer learning efficiency across these tasks, thus enhancing the robustness of comparative analyses regarding task performance.

This model was trained on a NVIDIA 3080 Ti GPU, with CUDA 11.2 and cuDNN 8.1.0 in Tensorflow 2.10. For each analysis' training iteration, we allocated 10% of the data from each subject present in the training set for validation, to trigger early stopping. This early stopping mechanism halts the training if the validation loss fails to improve for 10 consecutive epochs.

Inter-task transfer learning analysis:

To assess the transfer learning capabilities across MI and ME tasks, we evaluated the following training schemes:

1. Training the DL model on all subjects within the ME dataset, then evaluating the performance on the MI dataset data, treating left-/right- hand movement imagination as if it was the trained left-/right- hand movement execution. In this training scheme, the ME data from every user, whose MI accuracy is assessed, is included in the training data.

CC BY
https://creativecommons.org/licenses/by/4.0/deed.en
This CC license does not apply to third party material and content noted otherwise.

Published by
Verlag der Technischen Universität Graz

 10.3217/978-3-99161-014-4-072

2. Initially, pre-training the DL model on every subjects' trials present in the ME dataset except for one, following a leave one subject out (LOSO) training scheme. Subsequently, the model's accuracy in identifying MI trials for the excluded user is assessed. This process is replicated for every user.

Moreover, we will examine whether including the ME data of the evaluated user significantly impacts the results employing the Wilcoxon signed rank test [24].

Task performance correlation analysis:
We evaluate the correlation between the decoding accuracies for ME and MI data. Accuracies are obtained following a LOSO training scheme. For each task, we train the model on every users' data, except for one subject. The excluded user's trials serve as the test set to determine inter-subject ME or MI prediction accuracy [11]. This correlation is quantified using Spearman's rank correlation coefficient, which will describe the monotonic relationship between these inter-subject performances [25].

RESULTS

Our study yielded several key insights. Firstly, our analysis demonstrated that a DL network, trained on ME trials, is capable of classifying MI trials in the majority of participants with a degree of accuracy ($\geq$70%), which is considered sufficient for BCI control in a binary MI task [9–11]. Secondly, a significant and positive correlation was established between inter-subject performances on ME and MI tasks, evidenced by a highly significant p-value of less than 0.001. Additionally, we found no significant difference in performance between the model trained with ME data, including trials from the target user, and the model trained on MI data from other users.

Inter-task transfer learning analysis:
The efficacy of inter-task transfer learning was examined through two distinct training schemes, the results of which are summarized in Table 1. Our findings highlight that incorporating ME data from the target subject into the model's training signifantly enhances accuracy (p-value$<$0.05), compared to the model that has not been exposed to ME EEG signal from the evaluated user.

Table 1: Inter-task transfer learning accuracies

Training scheme	Accuracy(%)
ME to MI	85.73 ± 10.02
ME to MI without subject's ME trials	85.10 ± 9.93

Task performance correlation analysis:
The accuracies of the inter-subject transfer learning for both ME and MI tasks is presented in Table 2 while the correlation between both tasks performances can be observed in Figure 1. The accuracy for the ME task is significantly superior to the accuracy obtained on the MI task. Nevertheless, there is a positive and significant correlation (i.e., p-value$<$0.001) assessed by Spearman's rank correlation coefficient of 0.6378. Thus, there is a certain expectation of obtaining low or high performances when classifying MI data depending on the accuracy obtained on ME trials. Noteworthy, the accuracy obtained is way above the chance level for the Physionet dataset which is 50% ± 13.86% for individual users and 50% ± 1.40% for the entire dataset, both calculated at a 95% confidence level [26].

Table 2: Accuracies of inter-subject task

Task	Accuracy(%)
Inter-subject ME	87.35 ± 8.40
Inter-subject MI	85.65 ± 10.42

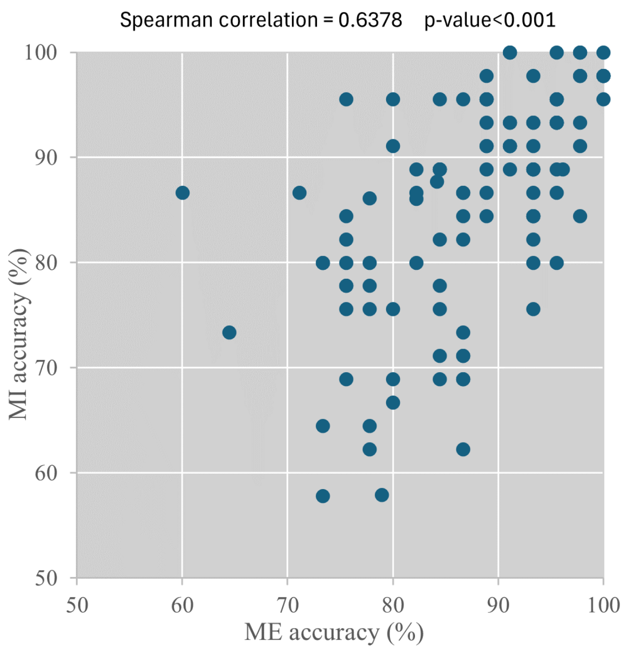

Figure 1: Correlation chart between ME and MI performances

DISCUSSION

ME-based preparation run for MI:
The results obtained in this study have provided a clear picture of the possibilities of transfer learning between the tasks of ME and MI. Of note, we have obtained a comparable performance on decoding MI between a model only trained with ME data, 85.73% ± 10.02%, and the same model trained only on MI data of other participants, 85.65% ± 10.42%. There is a minor, but significant, increase in the MI performance obtained between a model that includes ME data of the final user. Furthermore, the relationship between ME and MI not is only restricted to the possibility of this inter-task transfer learning, but there is also a correlation between ME and MI accuracies as shown in Figure 1.
There have been previous works that have exploited the relationship between ME and MI EEG data [16, 18]. While these works explored this relationship, they have applied it as a previous step to initialize their DL networks without exploring the fully inter-task transfer

CC BY
https://creativecommons.org/licenses/by/4.0/deed.en
This CC license does not apply to third party material and content noted otherwise.

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-072

learning capabilities. The fully transfer-learning possibility explored in this work, in addition to the correlation between ME and MI performances, could be exploited by establishing a preparation run based on ME where the user can be shown how the instruction and feedback will be presented. There are clear benefits from this ME-based preparation run. For new BCI users, understanding how to perform MI tasks can be difficult. Training users with ME tasks, which are more intuitive and easier to perform, can serve as a stepping stone, helping users learn how to modulate their neural signals effectively before transitioning to MI tasks. This can shorten the learning curve and improve overall BCI control. Furthermore, ME tasks can be performed with less mental effort from the user. Furthermore, the accuracy of this ME-based preparation run could be used to indicate the expected accuracy on MI tasks. Additionally, collecting high-quality MI data can be challenging, especially for BCI users who may struggle with performing consistent MI tasks without physical movement. Collecting ME data can provide a more robust dataset for training BCI algorithms, as ME tasks can be more easily performed and monitored for correctness, leading to higher-quality training data. Finally, in rehabilitation settings, a model trained on ME data could lead a more targeted recovery of the lost functions since it will search for the lost patterns common to the users included in the training ME data.

Limitations and future work:
While our study offers promising insights into the relationship between ME and MI, as well as the capabilities of inter-task transfer learning, we recognize certain limitations that future research should address. To begin with, our analysis was focused on binary upper-limb classification tasks, which may not encompass the complexity or challenge of distinguishing among more varied types of ME/MI tasks involving movements with less spatially distinct neural activity. Expanding this research to include multi-class classification tasks that incorporate a wider range of movements could offer a more comprehensive understanding of the applicability of our findings. In addition, the analysis was conducted using data collected in a single session from participants who did not receive feedback, limiting our ability to assess the potential for learning or adaptation over time. Investigating the long-term effects of using ME-based preparation run on MI task performance, as well as user satisfaction, could provide valuable insights for the development of more personalized and effective MI-based BCI systems. Moreover, assessing the impact of MI-based rehabilitation, enhanced with feedback from models trained on ME data, in comparison to those trained solely on MI data, would significantly contribute to our understanding of the most effective strategies for leveraging BCIs in rehabilitation contexts.

CONCLUSION

In this study, we explored the potential of inter-task transfer learning between ME and MI, uncovering that ME data can be effectively utilized to train DL models for MI classification. Additionally, we identified a significant correlation in performance across both tasks. These insights have prompted us to propose an ME-based preparation strategy for MI tasks. By integrating this ME-based preparation run into MI-based BCIs, we introduce a pragmatic solution that leverages the inherent neural and functional similarities between ME and MI. This approach not only maintains BCI performance but also improves accessibility and user experience, making BCIs more intuitive and effective for users. Simultaneously, the ME-based preparation trials offer the opportunity to generate a new corpus of EEG data, which assures the presence of task related information. This enhancement in data quality facilitates the training of deep learning models with improved accuracy.

ACKNOWLEDGEMENTS

This research has been developed under the grants PID2020-115468RB-I00, PDC2021-120775-I00 and TED2021-129915B-I00 funded by MCIN/AEI/10.13039/501100011033/ and European Regional Development Fund (ERDF) A way of making Europe; under project '0124_EUROAGE_MAS_4_E' ('Cooperation Programme Interreg VI-A Spain-Portugal POCTEP 2021–2027') funded by 'European Commission' and ERDF; and by CIBER-BBN through 'Instituto de Salud Carlos III' co-funded with ERDF funds. S. Pérez-Velasco and D. Marcos-Martínez were in receipt of a PIF grant by the 'Consejería de Educación de la Junta de Castilla y León'.

REFERENCES

[1] Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain–computer interfaces for communication and control. Clinical Neurophysiology. 2002;113(6):767–791.
[2] Wolpaw JR, Wolpaw EW. Brain-Computer Interfaces: Principles and Practice. Oxford University Press (2012).
[3] Ramos-Murguialday A *et al.* Proprioceptive Feedback and Brain Computer Interface (BCI) Based Neuroprostheses. PLoS ONE. 2012;7(10):e47048.
[4] Bundy DT *et al.* Contralesional Brain-Computer Interface Control of a Powered Exoskeleton for Motor Recovery in Chronic Stroke Survivors. Stroke. 2017;48(7):1908–1915.
[5] Moldoveanu A *et al.* The TRAVEE system for a multimodal neuromotor rehabilitation. IEEE Access. 2019;7:8151–8171.
[6] Sebastián-Romagosa M *et al.* Brain Computer Interface Treatment for Motor Rehabilitation of Upper Extremity of Stroke Patients—A Feasibility Study. Frontiers in Neuroscience. 2020;14(October):1–12.

CC BY
https://creativecommons.org/licenses/by/4.0/deed.en
This CC license does not apply to third party material and content noted otherwise.

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-072

[7] Pérez-Velasco S, Marcos-Martínez D, Santamaría-Vázquez E, Martínez-Cagigal V, Moreno-Calderón S, Hornero R. Unraveling motor imagery brain patterns using explainable artificial intelligence based on Shapley values. Computer Methods and Programs in Biomedicine. 2024;246(January):108048.
[8] Alkoby O, Abu-Rmileh A, Shriki O, Todder D. Can We Predict Who Will Respond to Neurofeedback? A Review of the Inefficacy Problem and Existing Predictors for Successful EEG Neurofeedback Learning. Neuroscience. 2018;378(January):155–164.
[9] Lee MH *et al.* EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy. GigaScience. 2019;8(5):1–16.
[10] Meng J, He B. Exploring training effect in 42 human subjects using a non-invasive sensorimotor rhythm based online BCI. Frontiers in Human Neuroscience. 2019;13(April):1–19.
[11] Perez-Velasco S, Santamaria-Vazquez E, Martinez-Cagigal V, Marcos-Martinez D, Hornero R. EEGSym: Overcoming Inter-Subject Variability in Motor Imagery Based BCIs With Deep Learning. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2022;30:1766–1775.
[12] Jeunet C, Glize B, McGonigal A, Batail JM, Micoulaud-Franchi JA. Using EEG-based brain computer interface and neurofeedback targeting sensorimotor rhythms to improve motor skills: Theoretical background, applications and prospects. Neurophysiologie Clinique. 2019;49(2):125–136.
[13] Saha S, Baumert M. Intra- and Inter-subject Variability in EEG-Based Sensorimotor Brain Computer Interface: A Review. Frontiers in Computational Neuroscience. 2020;13(January):1–8.
[14] Dose H, Møller JS, Iversen HK, Puthusserypady S. An end-to-end deep learning approach to MI-EEG signal classification for BCIs. Expert Systems with Applications. 2018;114:532–542.
[15] Zhang K, Robinson N, Lee SW, Guan C. Adaptive transfer learning for EEG motor imagery classification with deep Convolutional Neural Network. Neural Networks. 2021;136:1–10.
[16] Lee DY, Jeong JH, Lee BH, Lee SW. Motor Imagery Classification Using Inter-Task Transfer Learning via a Channel-Wise Variational Autoencoder-Based Convolutional Neural Network. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2022;30:226–237.
[17] Shuqfa Z, Belkacem AN, Lakas A. Decoding Multi-Class Motor Imagery and Motor Execution Tasks Using Riemannian Geometry Algorithms on Large EEG Datasets. Sensors. 2023;23(11):5051.
[18] Miao M, Yang Z, Zeng H, Zhang W, Xu B, Hu W. Explainable cross-task adaptive transfer learning for motor imagery EEG classification. Journal of Neural Engineering. 2023;20(6):066021.
[19] Goldberger AL *et al.* PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000;101(23).
[20] Schalk G, McFarland D, Hinterberger T, Birbaumer N, Wolpaw J. BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. IEEE Transactions on Biomedical Engineering. 2004;51(6):1034–1043.
[21] Schirrmeister RT *et al.* Deep learning with convolutional neural networks for EEG decoding and visualization. Human Brain Mapping. 2017;38(11):5391–5420.
[22] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces. Journal of Neural Engineering. 2018;15(5):1–30.
[23] Santamaria-Vazquez E, Martinez-Cagigal V, Vaquerizo-Villar F, Hornero R. EEG-Inception: A Novel Deep Convolutional Neural Network for Assistive ERP-based Brain-Computer Interfaces. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2020;28(12):2773–2782.
[24] Wilcoxon F. Individual Comparisons by Ranking Methods. Biometrics Bulletin. 1945;1(6):80.
[25] Spearman C. The Proof and Measurement of Association between Two Things. The American Journal of Psychology. 1904;15(1):72.
[26] Müller-Putz GR, Scherer R, Brunner C, Leeb R, Pfurtscheller G. Better than random: a closer look on BCI results. 2007 1st COST Neuromath Workgroup Meeting, Rome, Italy. 2007;10(1):95–96.

CC BY
https://creativecommons.org/licenses/by/4.0/deed.en
This CC license does not apply to third party material and content noted otherwise.

Published by
Verlag der Technischen Universität Graz