Assessing the Impact of Population Data Domain Differences on Transfer Learning in P300-based Brain-Computer Interfaces (Student Abstract)

Rally Lin[1*], Christina Mo[1*], Reyan Shariff[1*], Darrick Zhang[1*], Abdullah Alumar[2], Kaleb Kassaw[1], Leslie M. Collins[1], Boyla O. Mainsah[1]

[1]Duke University, Durham, NC, USA
[2]Manchester Metropolitan University, Manchester, UK
boyla.mainsah@duke.edu

Abstract

Brain-computer interfaces (BCIs) can provide a means of communication for individuals with severe neuromuscular diseases, the target end-users. While personalized BCI machine learning models are the current standard, models trained on data from other users could reduce BCI calibration time. We use a novel dataset with BCI users with and without amyotrophic lateral sclerosis (ALS) and a popular BCI deep learning model, EEGNet, to assess the impact of population domain data on transfer learning of a P300 speller task in the ALS cohort. Results show that training on source data from the non-ALS cohort was detrimental to transfer learning. In contrast, generic EEGNet models trained on source data from the ALS cohort performed comparably as user-specific models. Our findings highlight the need for more data from target end-users populations in publicly available BCI datasets.

Introduction

Brain-computer interfaces (BCIs) have wide ranging applications to replace impaired neural output or provide an alternative means of communication by interpreting brain signals to control an external system. Due to inter-individual variabilities, the BCI machine learning (ML) model used for brain signal decoding is trained on user-specific data collected during a calibration session. To minimize BCI calibration time, there has been an increased interest in using data from other BCI users for transfer learning to develop generic deep learning models for BCI tasks.

Beyond inter-individual variabilities, differences in brain signal characteristics and BCI experiment conditions create a data domain mismatch between the general population and target BCI end-users (Wan et al. 2021). There is a lack of data from target BCI end-users in publicly available BCI datasets (Bianchi et al. 2021). Thus, transfer learning in the BCI field focuses primarily on analyzing BCI data from the general population (Wan et al. 2021). There is a lack of exploration of how differences across BCI populations impact the utility of transfer learning in target BCI end-users. The main contribution of this work is a first investigation of the impact of population data on transfer learning of a BCI task in a target BCI end-user population.

*These authors contributed equally.

Study	Number of Subjects	Grid Size	Inter-stimulus Interval	ALS Study?
A	13	9×8	125 ms	No
B	18	6×6	250 ms	Yes
C	19	9×8	125 ms	No
L	11	6×6	250 ms	Yes

Table 1: Summary of bigP3BCI studies used in this work.

Methodology

This work will focus on the P300-based BCI speller (Farwell and Donchin 1988) that relies on detecting event-related potentials embedded in electroencephalogram (EEG) data elicited in response to target stimuli. In this work, the target end-user population is individuals with amyotrophic lateral sclerosis (ALS). ALS is a neuro-degenerative disease that causes a progressive loss of voluntary muscle control.

Dataset. We used *bigP3BCI*, a novel BCI dataset with data from P300 speller studies with individuals with and without ALS (Mainsah et al. 2024). To minimize confounds across experiment conditions, we used data from the checkerboard paradigm condition from Studies A and C for the non-ALS cohort and Studies B and L for the ALS cohort. Additional study details are summarized in Table 1.

BCI Simulations. P300 speller use was simulated with EEG data from electrodes {Fz, Cz, P3, Pz, P4, PO7, PO8, Oz} and Bayesian inference for target character estimation.

Stopping Criteria. Various stopping criteria were implemented to vary the amount of data collection prior to target character estimation. In static stopping, the amount of data collection is fixed. In dynamic stopping (DS), the stopping rule was set based on attaining a character probability threshold, $P_{th} = 0.9$. DS with a language model (DSLM) is the same as DS but character probabilities are initialized based on a bigram language model (Mainsah et al. 2015). The data limit was the same across all stopping criteria.

Classifier Models. We trained a traditional ML model and a deep learning model. The traditional ML model was stepwise linear discriminant analysis (SWLDA), which was used in the original P300 speller (Farwell and Donchin 1988) and is still popular in the P300 BCI field (Kalra et al. 2023). Since model inference times need to be within real-time P300 signal processing limits, a deep learning model with low computational complexity is desirable. EEGNet (Lawhern et al. 2018b), a compact convolutional neural network for generic EEG signal classification, has been tested in a real-time P300 speller study. We used the EEGNet implementation in Python from the model developers (Lawhern et al. 2018a). Raw features extracted 800 ms from stimulus onset were used to train SWLDA and EEGNet models.

Classifier Learning. Classifier models were trained either only on user-specific data (SWLDA, EEGNet) or used data from other BCI users for transfer learning (EEGNet). Domain definitions include: *within-domain* (WD) source data from individuals with ALS (Study B) excluding data from a novel BCI user with ALS (Study L); and *cross-domain* (CD) source data from individuals without ALS (Studies A and C). Generic EEGNet models were either: i) pretrained on CD data and fine-tuned (FT) on WD data or on novel user-specific data; or ii) pre-trained on WD data and FT on novel user-specific data. EEGNet models were trained with a cross-validation test split of 0.2 for 300 epochs using the Adam optimizer and a cross-entropy loss. The trained models were evaluated on the test data partition of study L.

Statistical Tests. Performance measures include character accuracy and expected stopping time. A linear mixed effects model was fit with fixed effects of classifier model and stopping criterion and random effect of subject. Post-hoc pairwise comparisons with Bonferroni adjustment were conducted with estimated marginal means.

Results and Discussion

Subject-specific performance measures for DSLM are shown in Figure 1; performance trends are similar across data stopping criteria. Statistical results for all conditions are summarized in Figure 2. Overall, generic EEGNet models trained on CD source data performed poorly relative to conventional user-specific models. In contrast, generic EEGNet models trained on WD source data performed comparably or better than user-specific models. With the user-specific SWLDA model trained on raw features as the baseline, the degree of benefit conferred with transfer learning was dependent not only on the source data domain, but also on the baseline performance, Figure 1. Most mid- to high-level BCI performers experienced a sharp performance drop with generic EEGNet models trained on CD source data, even after fine-tuning on WD or user-specific data. Low-level performers mostly benefitted from generic EEGNet models.

Differences between EEGNet models trained on WD and CD data might be due to different experiment conditions across population cohorts. Physical limitations in target BCI end users usually define the range of BCI system parameters that are used during experiments. People with neuromuscular limitations usually find it easier to navigate a simple visual interface and have a lower tolerance for high stimulus presentation rates. The ALS studies used a 6×6 (number of rows $\times$ number of columns) speller grid and an inter-stimulus interval (ISI) of 250 ms while the non-ALS studies used a 9×8 grid and an ISI of 125 ms (Table 1). Since the feature extraction window is from stimulus onset, it is more likely that the ISI has a greater impact on the feature distribution than the speller grid size during CD transfer learning.

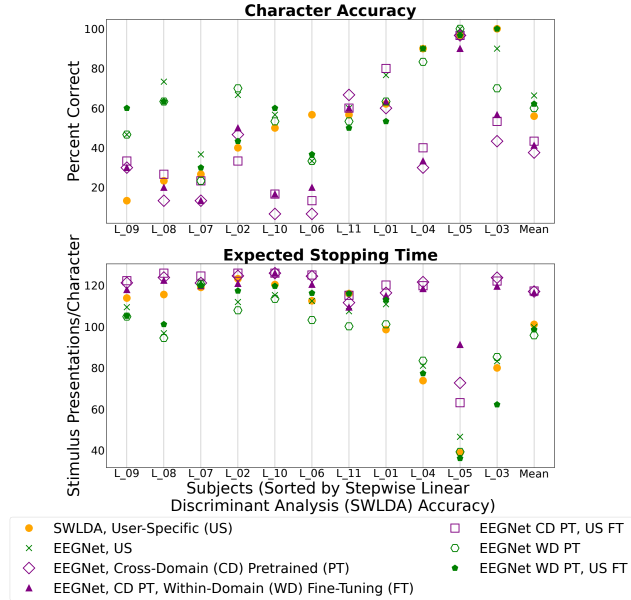

Figure 1: Character accuracy and expected stopping time from P300 speller simulations with dynamic stopping with a language model using EEG data from BCI users with ALS.

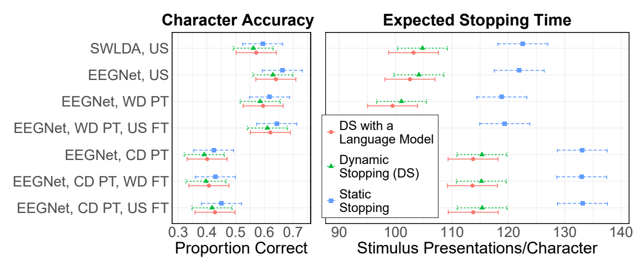

Figure 2: Estimated marginal means and 95% confidence intervals (error bars) of performance measures from simulations of P300 speller use. No overlap of error bars indicates a statistically significant difference ($p \leq 0.05$). Abbreviations: US, User-specific; WD, Within-Domain; CD, Cross-Domain; PT, Pretrained; FT, Fine-tuning.

Conclusion

The use of generic deep learning models could potentially minimize BCI calibration time while performing comparably as user-specific models. Transfer learning of the P300 BCI speller task in the test ALS cohort was more effective with source data from the same user population and was more beneficial to lower level BCI performers. Our work highlights the importance of including data from target BCI end-users and a wide range of user performance levels in publicly-available BCI datasets for BCI transfer learning analysis. Potential confounds from different experiment conditions across and within study populations should also be considered during BCI transfer learning. Future work includes a comprehensive analysis with other real-time compatible deep learning models and enhanced domain adaptation on a larger dataset partition.

Acknowledgements

This work was funded by the Data+ 2024 summer research program at Duke University.

References

Bianchi, L.; Antonietti, A.; Bajwa, G.; Ferrante, R.; Mahmud, M.; and Balachandran, P. 2021. A functional BCI model by the IEEE P2731 working group: data storage and sharing. *Brain-Computer Interfaces*, 8(3): 108–116.

Farwell, L. A.; and Donchin, E. 1988. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70(6): 510–523.

Kalra, J.; Mittal, P.; Mittal, N.; Arora, A.; Tewari, U.; Chharia, A.; Upadhyay, R.; Kumar, V.; and Longo, L. 2023. How Visual Stimuli evoked P300 is transforming the Brain-Computer Interface Landscape: A PRISMA Compliant Systematic Review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31: 1429–1439.

Lawhern, V. J.; Solon, A. J.; Waytowich, N. R.; Gordon, S. M.; Hung, C. P.; and Lance, B. J. 2018a. The Army Research Laboratory (ARL) EEGModels project: A Collection of Convolutional Neural Network (CNN) models for EEG signal processing and classification, using Keras and Tensorflow. https://github.com/vlawhern/arl-eegmodels.

Lawhern, V. J.; Solon, A. J.; Waytowich, N. R.; Gordon, S. M.; Hung, C. P.; and Lance, B. J. 2018b. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 15(5): 056013.

Mainsah, B. O.; Collins, L. M.; Colwell, K. A.; Sellers, E. W.; Ryan, D. B.; Caves, K.; and Throckmorton, C. S. 2015. Increasing BCI communication rates with dynamic stopping towards more practical use: an ALS study. *Journal of Neural Engineering*, 12(1): 016013.

Mainsah, B. O.; Fleeting, C.; Balmat, T.; Sellers, E. W.; and Collins, L. M. 2024. bigP3BCI: An Open, Diverse and Machine Learning Ready P300-based Brain-Computer Interface Dataset. *PhysioNet*.

Wan, Z.; Yang, R.; Huang, M.; Zeng, N.; and Liu, X. 2021. A review on transfer learning in EEG signal analysis. *Neurocomputing*, 421: 1–14.