





Exploring Novel Machine Learning Techniques for Brain Computer Interface (BCI) Applications

Report Version

Group Members:
Riyadh Alghamdi (KAUST)
Arnav Gupta (Duke University)
Albert Liu (Duke University)
Eliza Haines (Clemson University)

Advisors:
Aniketh Shenoy Kota (Duke University)
Dr. Boyla Mainsah (Duke University)
Dr. Leslie Collins (Duke University)

August 27, 2025

Investigating the Impact of Training Data Quality vs. Quantity on P300 Brain-Computer Interface Performance

Riyadh Alghamdi[1], Arnav Gupta[2], Albert Liu[2], Eliza Haines[3]
[1]KAUST, [2]Duke University, [3]Clemson University

Abstract

Brain–computer interfaces (BCIs) translate brain activity into actionable outputs, enabling communication for individuals with severe motor impairments such as amyotrophic lateral sclerosis (ALS). Electroencephalography (EEG)-based BCIs, such as the P300 speller, are widely adopted since they are noninvasive, safe, and affordable. This work examines the trade-off between training data quality and quantity in P300 speller performance. Using a curated EEG dataset, we train SWLDA and EEGNet models under varying data conditions to compare the impact of high-quality, low-volume datasets versus larger, noisier datasets. Results show that high-quality, domain-matched datasets consistently outperform larger mismatched datasets, underscoring the primacy of data quality in P300 speller performance.

1 Introduction

A P300 speller presents users with a grid of characters that flash in rows and columns. When the target symbol flashes, a P300 event-related potential emerges in the EEG, appearing as a positive peak around 300 ms post-stimulus. Classifiers detect this response and rank characters to infer the user's intended choice. Figure 1 illustrates this effect, showing average EEG responses to target and non-target stimuli. The P300 peak appears clearly in the target condition, demonstrating the neural basis for BCI operation.

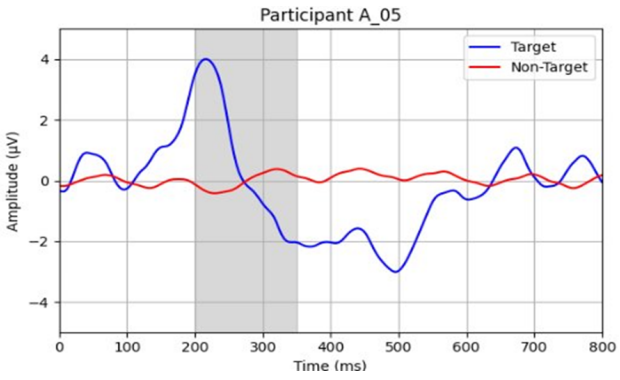


Figure 1: Grand-average EEG waveforms showing the P300 positive peak in response to target stimuli compared to non-targets.

2 Methods

Experiments were conducted using the bigP3BCI dataset [3], which contains EEG recordings from both ALS patients and non-ALS participants performing the same P300 speller task. Our methodology was organized in two parts: defining the experimental conditions and establishing a consistent data processing pipeline.

Training Conditions

Three experimental conditions were defined to represent different trade-offs between quality and quantity: *user-specific*, *within-domain*, and *cross-domain*. User-specific models were trained and tested on the same individual, within-domain models used data from other ALS patients, and cross-domain models were trained on non-ALS participants before testing on ALS patients.

Figure 2 summarizes these conditions. The left panel shows subject assignments and inter-stimulus interval adjustments, while the right panel illustrates the imbalance in training data size across the three setups.

Condition	Inter-Stimulus Interval	Training Data	Testing Data
User-Specific	250 ms	Participant L##	Participant L##
Within-Domain	250 ms	Study B	
Cross-Domain, Small	125 ms	Studies A and C	
	250 ms*		
Cross-Domain, Large	125 ms	Studies A, C, H, K, and M	
	250 ms*		

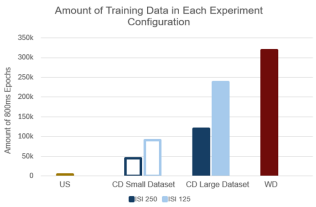


Figure 2: Overview of training conditions. Left: subject partitions and inter-stimulus intervals. Right: training data volume in each condition.

Processing Pipeline

Building on these conditions, the second part of the methodology was a consistent pipeline for preprocessing and model training. Raw EEG signals were notch and bandpass filtered, segmented into stimulus-locked epochs, and resampled for uniformity. Two models were evaluated: SWLDA and EEGNet. Models were trained under all three experimental conditions and tested for character-level speller accuracy.

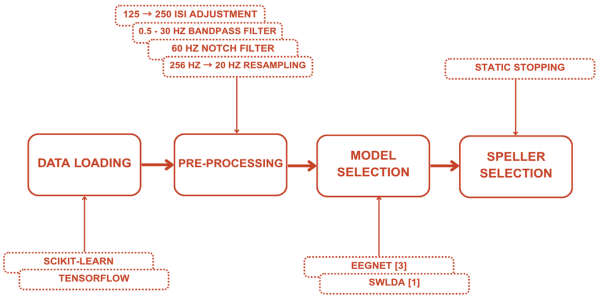


Figure 3: Pipeline for P300 speller experiments, including preprocessing, classifier training, and speller output evaluation.

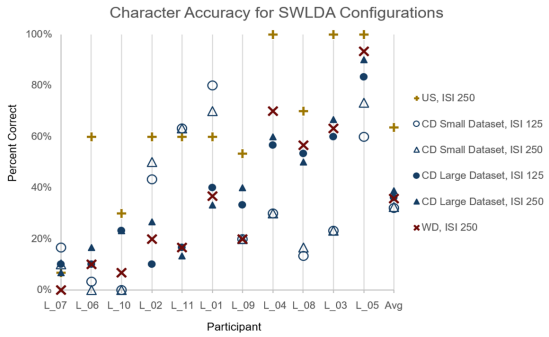


Figure 5: Character accuracy across participants using SWLDA with user-specific, within-domain, and cross-domain training data.

3 Results

Results were obtained separately for EEGNet and SWLDA to compare how classifier choice interacts with training conditions.

As shown in Figure 4, EEGNet achieved the highest accuracy in the user-specific condition, followed by within-domain, while cross-domain training produced the lowest performance. Larger cross-domain datasets improved accuracy but never reached the levels of domain-matched training.

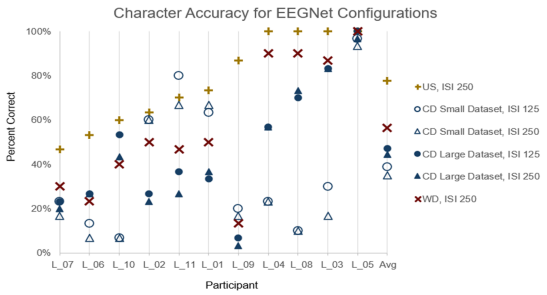


Figure 4: Character accuracy across participants using EEGNet with user-specific, within-domain, and cross-domain training data.

Figure 5 shows the corresponding results for SWLDA. Although absolute performance was lower than EEGNet, the same trend was observed: user-specific training performed best, within-domain provided intermediate performance, and cross-domain was lowest. Increasing the amount of cross-domain data led to modest gains but did not close the gap with domain-matched conditions.

The consistency between Figures 4 and 5 indicates that training data quality, rather than classifier type, is the dominant factor in determining P300 speller performance.

4 Discussion

The results demonstrate that the central factor in P300 BCI performance is not the algorithm but the quality and relevance of the training data. User-specific models consistently outperformed all others, highlighting the importance of personalization. Within-domain training, using ALS participants to train for other ALS users, provided close performance and suggests that pooling data from similar patient populations can be effective. Cross-domain training, even with large volumes of healthy-subject data, was consistently the weakest. This reflects the strong domain mismatch between EEG signals of ALS and non-ALS populations, which cannot be resolved by sheer quantity of data.

EEGNet outperformed SWLDA overall, showing the benefits of modern deep learning architectures for EEG analysis, but both models exhibited the same dependency on training data domain. This suggests that improvements in algorithms alone cannot compensate for poor data alignment. These findings carry strong implications for practical deployment of P300 spellers, where resources should prioritize targeted, high-quality data collection from the intended user population.

5 Conclusions

This report shows that small, domain-matched datasets can outperform larger mismatched datasets for P300 BCI classification. User-specific models remain the gold standard, while within-domain models offer a strong alternative when individual calibration data are limited. Cross-domain transfer from non-ALS to ALS populations remains problematic despite larger training sets. The findings emphasize that future P300 BCI research and development should focus on collecting representative data from target populations rather than building ever-larger generic datasets.

6 Future Work

This work represents an initial report draft, and further experiments will expand on these findings. In particular, we will examine how the amount of additional training data relates to performance improvements, investigate optimized data selection strategies using autoencoder-based embeddings with k-nearest neighbor filtering for within-domain and cross-domain transfer, and systematically assess the combined impact of these approaches.

Acknowledgments

We express our sincere gratitude to our advisors Aniketh Shenoy Kota, Dr. Boyla Mainsah, and Dr. Leslie Collins for their invaluable guidance throughout this research project. Their expertise in BCI research and machine learning provided essential direction for our investigations. We acknowledge KAUST Academy for providing this research opportunity and Duke University Applied Machine Learning Laboratory for access to the bigP3BCI dataset and computational resources that made this work possible.

References

[1] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, no. 6, pp. 510–523, 1988.

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

[3] B. Mainsah, C. Fleeting, T. Balmat, E. Sellers, and L. Collins, "bigP3BCI: An open, diverse and machine-learning-ready P300-based brain–computer interface dataset," *PhysioNet*, 2025.

[4] R. Lin et al., "Assessing the impact of population data domain differences on transfer learning in P300-based brain-computer interfaces," *Proc. AAAI Conf. Artif. Intell.*, vol. 39, no. 28, pp. 29415–29417, 2025.