Enhancing Subject-Independent P300 Classification in RSVP-Based BCIs with Deep Learning

Muhammad Ahsan Awais
Insight Research Ireland Centre for Data Analytics
School of Computing, Dublin City University
Dublin, Ireland
muhammad.awais2@mail.dcu.ie

Tomás Ward
Insight Research Ireland Centre for Data Analytics
School of Computing, Dublin City University
Dublin, Ireland
tomas.ward@dcu.ie

Graham Healy
School of Computing
Dublin City University
Dublin, Ireland
graham.healy@dcu.ie

***Abstract*—Brain-computer interfaces offer transformative potential across a variety of fields, such as assistive technologies and neurorehabilitation. Traditional machine learning methods for P300 classification are typically subject-specific, which can lead to reduced generalizability. This study explores the subject-independent classification of P300 responses elicited through RSVP across 20 subjects. Three models — Bayesian Ridge, CNN-based, and EEGNet — were evaluated for their performance. The results revealed that EEGNet outperformed both Bayesian Ridge (ROC-AUC: 0.732) and CNN-based approaches (ROC-AUC: 0.763), attaining an average ROC-AUC score of 0.767. Additionally, the impact of varying the amount of training data was examined, demonstrating that larger training datasets significantly improved classification performance. Furthermore, fine-tuning EEGNet on individual test subjects significantly enhanced its performance, increasing the average ROC-AUC to 0.813. A paired t-test confirmed the statistical significance of the improvement, highlighting EEGNet's robust potential for generalizable P300 classification.**

***Index Terms*—Brain-computer interfaces, subject-independent, electroencephalography (EEG), RSVP**

I. INTRODUCTION

Brain-computer interfaces (BCIs) represent a revolutionary frontier in the interaction between humans and technology. By enabling direct communication between the brain and external devices, BCIs hold the potential to transform diverse applications, including assistive technologies, neurorehabilitation, gaming, and communication for individuals with severe motor impairments. Among various BCI paradigms, the Event-Related Potential (ERP)-based BCI, in particular using the Rapid Serial Visual Presentation (RSVP) paradigm, has gained significant attention. RSVP-based BCI systems leverage the P300 ERP, a prominent positive deflection in the EEG signal approximately 300 milliseconds after a target stimulus, to classify responses to visual stimuli into target and non-target categories [1]. The efficiency and reliability of these systems are highly dependent on the accurate classification of P300 signals.

P300-based brain-computer interfaces offer significant opportunities in both non-clinical applications [2], such as gaming, adaptive learning systems, and user authentication, and clinical applications [3] [4], particularly in assistive communication for individuals with severe motor disabilities. For instance, P300 spellers allow individuals with locked-in syndrome, such as those affected by amyotrophic lateral sclerosis (ALS), to communicate by selecting letters or words through brain activity alone [5]. This non-invasive technology offers a critical bridge to independence and enhances the quality of life for patients who have lost motor function.

Furthermore, RSVP-based event-related potentials hold clinical relevance for the early diagnosis of neurological disorders and cognitive rehabilitation. Altered P300 responses are biomarkers for neurological disorders such as Alzheimer's and schizophrenia, making them valuable for early diagnosis and monitoring cognitive states. Subject-independent models streamline deployment in clinical environments, reducing the need for personalization and enhancing usability for rehabilitation, neurofeedback, and health assessments across diverse patient populations.

In addition to assistive communication, P300 BCIs also play a significant role in cognitive and neurological diagnostics. The P300 waveform reflects cognitive processes such as attention and decision-making. By analyzing P300 responses, clinicians can gain insights into medical conditions such as Parkinson's disease [6], ADHD [7], Alzheimer's disease [8], and traumatic brain injuries. These BCI applications extend to monitoring cognitive load and fatigue, offering insights for therapy planning and rehabilitation [9].

Moreover, P300 BCIs are currently being explored as tools for neurofeedback and cognitive training, aiding in recovery for patients with neurological impairments by improving their capacity to regulate brain activity. These versatile applications underscore the significance of continued research into P300

This work was funded by Science Foundation Ireland under Grant Number SFI/12/RC/2289_P2 and CHIST-ERA under Grant Number (CHISTERA IV 2020 - PCI2021-122058-2A).

BCIs and their translation to practical clinical settings [10].

Despite recent advancements in machine learning and signal processing techniques, training models for robust P300 classification continues to be a substantial challenge. One crucial obstacle lies in the variability of EEG signals across individuals. Factors such as differences in brain anatomy, cognitive state, electrode placement, and external noise significantly influence the recorded signals [11]. Consequently, machine learning models trained on subject-specific data often fail to generalize effectively to new users, limiting the scalability and usability of BCIs in real-world scenarios [12].

Traditionally, machine learning models for P300 classification are trained on subject-specific data. While this approach delivers high accuracy within the training set, it often results in models that are overly specialized to the specific characteristics of the training subject's EEG responses. Such models often exhibit poor performance when applied to data from new subjects due to inter-subject variability. This limitation poses a significant barrier to developing generalized BCIs capable of serving diverse populations without requiring extensive retraining for each user.

Most existing BCIs rely on subject-specific training protocols, requiring significant time and effort to calibrate for each new user. These protocols limit scalability and hinder the widespread adoption of BCI technology. To address this, subject-independent classification methods have emerged as a promising solution. By training models on data from multiple users, these methods aim to learn invariant features of P300 signals, enabling BCIs to generalize across individuals without extensive retraining. This approach not only minimizes the need for lengthy calibration sessions but also makes BCIs more practical and accessible for real-world applications.

Subject-independent classification also aligns with the goal of building inclusive and adaptive technologies by enabling generalization across different users. This study investigates the effectiveness of deep learning and other performant machine learning algorithms for subject-independent P300 classification in RSVP paradigms, highlighting their potential to improve classification accuracy and adaptability across individuals.

II. Methodology

A. Dataset

The dataset used in this study comprises data from 20 subjects as part of the "AMBER: Advancing Multimodal BCIs for Enhanced Robustness—A Dataset for Naturalistic Settings" [13]. Data collection was conducted at Dublin City University, Ireland, with ethical approval obtained from the Dublin City University Research Ethics Committee (DCUREC/2021/175). The study involved 20 healthy participants aged 20–35 years, and EEG signals were recorded using a 32-channel ANT-Neuro eego sports mobile EEG system. Electrodes were positioned according to the 10–20 international system, with CPz serving as the reference channel. Data was sampled at 1000 Hz.

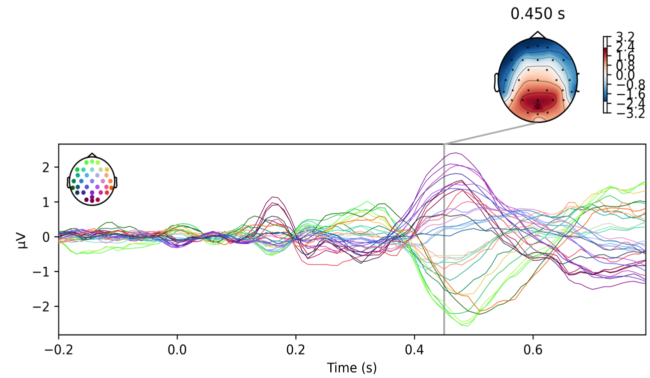

Fig. 1. Butterfly plot illustrating the difference in ERP averages between target epochs and the average standard epochs across all subjects.

The first 10 subjects performed RSVP tasks presented on a monitor display, while another 10 subjects completed the same tasks using a head-mounted display.

This dataset was originally designed to evaluate BCIs under both clean and noisy conditions by incorporating RSVP-BCI tasks in naturalistic settings, where a protocol involving instructed movements was used in order to systematically contaminate particular blocks of trials. For the purpose of this study, however, only the clean RSVP P300 trials from blocks where the subject was stationary (ideal condition) were used.

During the RSVP task, subjects engaged in a 90-second image search task, viewing 360 images per block, comprising 324 standard (non-target) trials and 36 target trials, across four sessions. The butterfly plot (Figure 1) illustrates the difference between the grand averages of target and standard epochs across all subjects. A prominent P300 component can be seen at posterior scalp sites approximately between 300ms and 600ms post-stimulus, indicating a significant neural response to target stimuli. This difference highlights the discriminative information present in the data, which is crucial for subject-independent classification.

B. Data Preparation and Classification

Data preparation involved several preprocessing steps to ensure the quality of the input for classification. EEG epochs were extracted between 0 ms and 800 ms post-stimulus, followed by downsampling to 50 SPS (samples per second) to reduce computational load. Common average referencing was applied, and the data was scaled using min-max normalization to standardize the input range.

Single-trial subject-independent classification, employing a leave-one-subject-out (LOSO) approach, was performed using three models: (1) Bayesian Ridge Regression [14] [15], a commonly used machine learning approach with P300-ERP data, and two CNN-based architectures, (2) CNN-1 [16] and (3) EEGNet [17]. As there was a class imbalance between target and non-target trials, model performance was evaluated using the Receiver Operating Characteristic—Area Under Curve (ROC-AUC) metric. This metric provides a robust assessment of the model's discriminative ability that is insensitive to class imbalances.

Experiments were conducted using an NVIDIA GeForce RTX 3090 Ti GPU using the Keras (version 3.8.0) and TensorFlow (version 2.18.0) frameworks for CNN-based model development. Further details about the models used are provided below.

1) Bayesian Ridge Regression: Bayesian Ridge Regression is a widely used approach in BCIs, combines Bayesian principles with ridge regression to handle uncertainty, reduce overfitting, and address challenges like multicollinearity [18]. By incorporating prior information into parameter estimation, it enhances model stability and generalizability [19].

2) CNN-1: In 2010, Cecotti et al. introduced CNN-1, a 4-layer convolutional neural network designed for P300 detection. The architecture begins with a spatial convolution layer that learns combinations of input channels, followed by a temporal 1D convolution layer that filters and subsamples the feature maps from the first layer. Both convolutional layers use a Scaled Hyperbolic Tangent activation function. The flattened output from the second layer is passed to a fully connected layer, and the final classification layer consists of two Sigmoid neurons. While the original architecture was largely retained in this study, hyperparameter tuning was performed to optimize model performance.

3) EEGNet: Lawhern et al. in 2018 proposed a compact and versatile convolutional neural network model specifically designed for BCI applications. This CNN model is effective for various tasks, including motor imagery classification [20], event-related potentials [21], and steady-state visual evoked potential analysis [22]. EEGNet employs separable and depthwise convolutions to create an EEG-specific model, incorporating established feature extraction principles for BCI. It is highly efficient, with significantly fewer trainable parameters compared to other deep learning architectures commonly used in BCI classification.

EEGNet consists of three main layers: a temporal convolution layer that captures time-domain features, a depthwise convolution layer that extracts spatial features by operating on each EEG channel independently, and a separable convolution layer that further refines feature extraction while reducing computational complexity. The model includes batch normalization, dropout, and activation functions to enhance stability and prevent overfitting.

While the core architecture of EEGNet was preserved, hyperparameter tuning was performed to optimize performance with learning rate, batch size, and dropout rate parameters.

C. Impact of Training Data Size

Furthermore, to examine the effect of training data size on model performance, we conducted additional experiments by reducing the number of training subjects from 19 to 15, 10, and 5 while keeping one subject for testing. The training subjects were selected randomly (excluding the test subject), and this process was repeated five times for each test subject to obtain a reliable average performance.

TABLE I
SUBJECT-INDEPENDENT SINGLE-TRIAL P300 CLASSIFICATION ROC-AUC SCORES FOR 20 SUBJECTS

Test Subject	**Bayesian Ridge Regression**	**CNN-1**	**EEGNet**	**EEGNet-FT**
1	0.622	0.638	0.632	0.654
2	0.646	0.764	0.773	0.913
3	0.672	0.668	0.679	0.803
4	0.744	0.801	0.785	0.835
5	0.844	0.857	0.845	0.873
6	0.773	0.804	0.794	0.835
7	0.776	0.725	0.778	0.813
8	0.602	0.656	0.671	0.718
9	0.799	0.851	0.845	0.895
10	0.866	0.836	0.856	0.922
11	0.763	0.846	0.848	0.846
12	0.833	0.840	0.834	0.912
13	0.671	0.720	0.709	0.732
14	0.806	0.811	0.819	0.847
15	0.777	0.814	0.795	0.801
16	0.512	0.528	0.571	0.569
17	0.668	0.726	0.740	0.751
18	0.716	0.744	0.735	0.787
19	0.786	0.837	0.860	0.855
20	0.762	0.803	0.761	0.898
Average	**0.732**	**0.763**	**0.767**	**0.813**

This analysis is significant as it provides insights into how the amount of available training data impacts subject-independent EEG classification. By systematically varying the training size and averaging across multiple trials, we ensure that the observed trends are not due to a specific subset of training data. Understanding these effects is crucial for designing EEG-based models in real-world scenarios where the amount of training data may be limited.

D. Fine-tuning

Fine-tuning was applied to the best-performing model to further enhance performance. In the subject-independent models (using the LOSO approach), 30% of each test subject's data was used for fine-tuning, while the remaining 70% was reserved for final evaluation. This approach allowed the model to adapt to subject-specific variations, making it more tailored to individual characteristics while maintaining the subject-independent framework.

III. RESULTS AND DISCUSSION

The performance of three different models was evaluated in a leave-one-subject-out classification approach for P300 RSVP detection across 20 subjects. For example, when evaluating Subject 1, the model was trained on data from the remaining 19 subjects (Subjects 2 to 20) and tested on Subject 1. The results shown in Table I indicate that deep learning models outperformed Bayesian Ridge Regression, with EEGNet achieving the highest average ROC-AUC score of 0.767, followed by CNN-1 (0.763) and Bayesian Ridge (0.732).

A closer look at individual subject results shows that EEGNet performed better than the other models across most test subjects. Notably, EEGNet achieved its highest ROC-AUC for subject 19 (0.860), demonstrating its ability to generalize well to unseen data.

To assess the statistical significance of the performance differences between the three classifiers, paired t-tests were performed on their ROC-AUC scores while keeping the significance level of 0.05 for all comparisons.

The results showed that CNN-1 (M = 0.763, SD = 0.088) performed significantly better than Bayesian Ridge Regression (M = 0.732, SD = 0.091), $t(19) = -3.73$, $p = 0.0014$, $d = 0.353$. Similarly, EEGNet (M = 0.767, SD = 0.080) demonstrated a significantly higher performance compared to Bayesian Ridge Regression, $t(19) = -4.26$, $p = 0.0004$, $d = 0.40$. While EEGNet also outperformed CNN-1, the difference was not statistically significant, $t(19) = -0.62$, $p = 0.544$, $d = 0.03$.

These results confirm that the EEGNet model achieves the highest performance among the three models. It significantly outperformed Bayesian Ridge and showed comparable results to CNN-1, reinforcing its robustness and suitability for subject-independent P300 classification tasks.

To examine the performance of fine-tuning LOSO subject-independent models (i.e., making them not subject-dependent), we adapted EEGNet by using 30% of the test subject's data for model fine-tuning while reserving the remaining 70% for final testing. The findings support further investigation into this approach, as Table I (last column) shows an improvement in the average ROC AUC score from 0.766 (EEGNet) to 0.813 (EEGNet Fine-Tuned).

To further verify the improvement in ROC-AUC scores, a paired t-test was conducted to compare EEGNet and Fine-Tuned EEGNet. The results showed that Fine-Tuned EEGNet (M = 0.813, SD = 0.092) performed significantly better than EEGNet (M = 0.767, SD = 0.080), $t(19) = -4.74$, $p = 0.00014$, $d = 0.54$.

These enhancements suggest that incorporating a small portion of subject-specific data during training allows the model to adapt to individual EEG variations, leading to more accurate classification.

The impact of training size on EEGNet's performance in LOSO subject-independent classification was evaluated using different numbers of training subjects (Figure 2). With 19 training subjects, the model achieved an average ROC-AUC score of 0.767. As the number of training subjects decreased to 15, 10, and 5, the average scores dropped to 0.757, 0.748, and 0.722, respectively. This decline in performance underscores the importance of a larger training dataset for better classification accuracy.

Fine-tuning was also evaluated using configurations of a reduced number of training subjects. Post fine-tuning, the average ROC-AUC scores improved to 0.813, 0.803, 0.799, and 0.795 for 19, 15, 10, and 5 training subjects, respectively. Although fine-tuning consistently boosted results, the performance decline with reduced training size was still evident, though less noticeable.

The box plots (Figure 2) demonstrate these trends, showing the variability in the results and the consistent improvements achieved with fine-tuning. These findings emphasize the critical role of larger training datasets in achieving robust subject-independent classification, while also demonstrating the potential of fine-tuning as an effective method for mitigating performance losses when training data is limited.

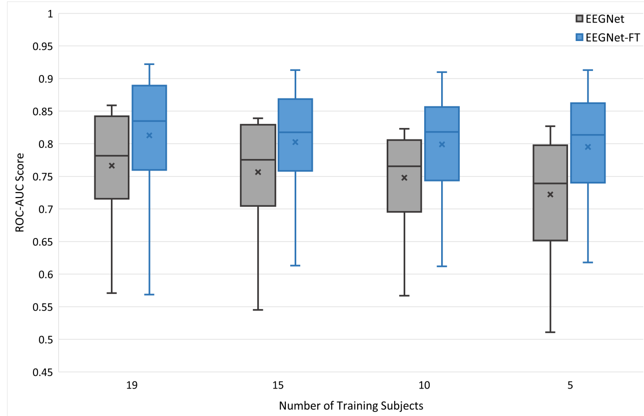

Fig. 2. Effect of training data size and fine-tuning on P300 classification performance, where each boxplot is generated using the (averaged-per-subject) ROC-AUC for 20 subjects for LOSO (EEGNet) / LOSO-but-later-fine-tuned (EEGNet-FT) configurations.

IV. Conclusion

This study demonstrates the effectiveness of subject-independent classification for P300 RSVP using a Leave-One-Subject-Out approach. Although Bayesian Ridge Regression has a proven history of performing well for RSVP-based P300 prediction, EEGNet outperformed it in the subject-independent classification setting. The results underscore EEGNet's ability to generalize across different subjects, making it a promising model for BCI applications that require high accuracy and reliability.

The further improvement in performance through fine-tuning indicates that EEGNet is highly adaptable, even in settings where data variability between subjects is significant.

In addition, different configurations with varying numbers of training subjects highlighted the importance of training data size, as larger datasets consistently led to improved classification performance, emphasizing their role in achieving robust subject-independent classification.

This work paves the way for more advanced, out-of-the-lab, real-time BCI applications, particularly in assistive technologies and neurorehabilitation, where robust and generalizable models are crucial. Future work will explore the trade-offs associated with the amount of subject-specific data required, assess the robustness of these approaches in the presence of noise, and investigate alternative architectures, such as transformers, that could benefit from larger datasets.

References

[1] M.A. Awais, T. Ward, P. Redmond and G. Healy, "From lab to life: assessing the impact of real-world interactions on the operation of rapid serial visual presentation-based brain-computer interfaces", Journal of Neural Engineering, vol. 21, no. 4, p. 046011, 2024.

[2] K. Värbu, N. Muhammad and Y. Muhammad, "Past, present, and future of EEG-based BCI applications", Sensors, vol. 22, no. 9, p. 3331, 2022.

[3] P. John, "Clinical application of the P300 event-related brain potential", Physical Medicine and Rehabilitation Clinics, vol. 15, no. 01, pp. 133-161, 2004.

[4] S. K. Mudgal, S. K. Sharma, J. Chaturvedi and A. Sharma, ”Brain computer interface advancement in neurosciences: Applications and issues”, Interdisciplinary Neurosurgery, vol. 20, p. 100694, 2020.

[5] R. Bettencourt, M. Castelo-Branco, E. Gonçalves, U.J. Nunes, and G. Pires, ”Comparing Several P300-Based Visuo-Auditory Brain-Computer Interfaces for a Completely Locked-in ALS Patient: A Longitudinal Case Study”, Applied Sciences, vol. 14, no. 18, p. 3464, 2024.

[6] N. Ferrazoli, C. Donadon, A. Rezende, P.H. Skarzynski and M.D. Sanfins, ”The application of P300-long-latency auditory-evoked potential in Parkinson disease”, International archives of otorhinolaryngology, vol. 26, no. 01, pp. 158-166, 2022.

[7] M. Tao, J. Sun, S. Liu, Y. Zhu, Y. Ren, Z. Liu, X. Wang, W. Yang, G. Li, X. Wang and W. Zheng, ”An event-related potential study of P300 in preschool children with attention deficit hyperactivity disorder”, Frontiers in Pediatrics, vol. 12, p. 1461921, 2024.

[8] M. Mohamed, N. Mohamed and J.G. Kim, ”P300 Latency with Memory Performance: A Promising Biomarker for Preclinical Stages of Alzheimer’s Disease”, Biosensors, vol. 14, no. 12, p. 616, 2024.

[9] V.V. Fateeva, A.B. Kushnir, A.V. Grechko, and L.A. Mayorova, 2024. ”Rehabilitation of Patients with Post-Stroke Cognitive Impairments Using a P300-Based Brain–Computer Interface: Results of a Randomized Controlled Trial”, Neuroscience and Behavioral Physiology, pp. 1-6, 2024.

[10] A. Haider and R. Fazel-Rezai, ”Application of P300 event-related potential in brain-computer interface”, in Event-Related Potentials and Evoked Potentials, vol. 29, no. 01, pp. 19-36, 2017.

[11] L. M. Zhao, X. Yan and B. L. Lu, ”Plug-and-play domain adaptation for cross-subject EEG-based emotion recognition”, in Proceedings of the AAAI conference on artificial intelligence, vol. 35, no. 1, pp. 863-870, 2021.

[12] B. Abibullaev, K. Kunanbayev and A. Zollanvari, ”Subject-independent classification of P300 event-related potentials using a small number of training subjects”, IEEE Transactions on Human-Machine Systems, vol. 52, no. 5, pp. 843-854, 2022.

[13] M.A. Awais, P. Redmond, T. Ward, and G. Healy, ”AMBER: advancing multimodal brain-computer interfaces for enhanced robustness—A dataset for naturalistic settings”, Frontiers in Neuroergonomics, vol. 4, p. 1216440, 2023.

[14] D.J. MacKay, ”Bayesian interpolation. Neural computation”, vol. 4, no. 3, pp. 415-447, 1992.

[15] M.E. Tipping, ”Sparse Bayesian learning and the relevance vector machine”, Journal of machine learning research, vol. 1, pp. 211-244, 2001.

[16] H. Cecotti, and A. Graser, ”Convolutional neural networks for P300 detection with application to brain-computer interfaces”, IEEE transactions on pattern analysis and machine intelligence, vol. 33, no. 3, pp.433-445, 2010.

[17] 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”, Journal of neural engineering, vol. 15, no. 5, p. 056013, 2018.

[18] H. Cecotti and A.J. Ries, ”Best practice for single-trial detection of event-related potentials: Application to brain-computer interfaces”, International Journal of Psychophysiology, vol. 111, pp. 156-169, 2017.

[19] J. Han, S.Y. Kim, J. Lee and W.H. Lee, ”Brain age prediction: A comparison between machine learning models using brain morphometric data”, Sensors, vol. 22, no. 20, p. 8077, 2022.

[20] H. Deng, M. Li, J. Li, M. Guo and G. Xu, ”A robust multi-branch multi-attention-mechanism EEGNet for motor imagery BCI decoding”, Journal of Neuroscience Methods, vol. 405, p. 110108, 2024.

[21] H. Zhang, Z. Wang, Y. Yu, H. Yin, C. Chen, and H. Wang, ”An improved EEGNet for single-trial EEG classification in rapid serial visual presentation task”, Brain Science Advances, vol. 8, no. 2, pp. 111-126, 2022.

[22] Y. Zhu, Y. Li, J. Lu and P. Li, ”EEGNet with ensemble learning to improve the cross-session classification of SSVEP based BCI from ear-EEG”, IEEE Access, 9, pp. 15295-15303, 2021.