Evaluation of the Electroencephalogram Conformer for the P300 Signal

Bruno Aristimunha, Grégoire Cattan, Anton Andreev

Evaluation of the Electroencephalogram Conformer for the P300 Signal

Bruno Aristimunha
Université Paris-Saclay, France
Federal University of ABC, Brazil
b.aristimunha@gmail.com

Gregoire H. Cattan
IBM, Poland
Hother, Poland
gregoire.cattan@ibm.com

Anton Andreev
GIPSA-lab, France
andreev.anton@gipsa-lab.grenoble-inp.fr

Abstract

The electroencephalogram (EEG) conformer is a transformer model that uses self-attention mechanisms for EEG feature extraction. It has been successfully applied to various datasets, demonstrating excellent performance in motor imagery and sentiment analysis. In this article, we explore whether the EEG conformer can be adapted for P300 data analysis.

Disclaimer

This is a work in progress. Although we have made every effort to present all the details of our implementation, this paper is a synthesis of different pieces of work that were completed at different points in time, by the three authors of this paper, and in different environments. Training this transformer model is also resource intensive, and we had a limited number of opportunities to run it. Reproduction of the results can also be affected by differences between versions of the software libraries we have used, including changes to underlying algorithms, bugs, and bug fixes.

1 Introduction

Attention-based transformer models have led to significant advances in natural language and image processing, thanks to their ability to capture global dependencies [26]. Transformers have also gained traction in electroencephalogram (EEG) decoding, achieving strong performance by effectively leveraging long-term temporal relationships [27, 23, 18]. A similar attention-based approach was explored in hypergraph case-based reasoning [19], a binary classification method that can operate in spaces without metrics. It learns metrics through a Bayesian approach and can intuitively be interpreted as an attention mechanism applied to self-organizing maps. However, this method did not yield conclusive results with EEG data [5]. In this paper, we explore the EEG conformer [24], a hybrid architecture that combines the strengths of convolutional neural networks and transformers, hence their name, which is an abbreviation of "convolution" and "transformer". The EEG conformer has shown outstanding results in motor imagery [17, 4] and sentiment analysis [29, 9] datasets, as shown in [24]. Here, we focus on the P300 signal, an event-related potential (ERP) elicited by the brain in response to a stimulus. Our goal is to evaluate whether the EEG conformer can effectively classify epochs of EEG data according to whether they contain P300 signal. Specifically, we aim to assess if the EEG conformer can serve as a valid transfer learning method across subjects using the P300 paradigm. The remainder of the paper is structured as follows: Section 2 describes our methodology, Section 3 presents our results and discussion, and Section 4 concludes.

Working paperPreprint. Under review.

2 Method

2.1 Datasets

We used eight data sets, which are available for public use and contain epochs of EEG signals with and without P300 (Table 1).

Number	Name	Number of subjects	Reference
1	BNCI2014_008	8	[20]
2	BNCI2014_009	10	[2]
3	BNCI2015_003	10	[11]
4	bi2013a	24	[25]
5	bi2014a	64	[15]
6	bi2014b	38	[16]
7	bi2015a	43	[13]
8	bi2015b	44	[14]

Table 1: Datasets.

2.2 Compound datasets

Two compound datasets were artificially created for this study. A compound dataset, a concept we introduced in MOABB, is formed by combining complete or partial data from one or more datasets into a new, artificial dataset. MOABB [3, 12] is a framework that standardizes the preprocessing and evaluation of EEG datasets. In a compound dataset, an artificial subject can be composed of real subjects, sessions, or even runs sourced from different datasets. The key condition is that all data must originate from experiments conducted under the same paradigm. For instance, data from a P300 experiment must not be combined with data from a motor imagery experiment. The first compound dataset we created contains the full contents of three BNCI datasets. The second includes the complete data from four Brain Invaders (BI) experiments (Table 2). BNCI is a European funded project, aiming to support and foster communication and collaboration between stakeholders in the field of Brain-Computer interfaces (BCIs). BI datasets experiments are a series of P300 experiments conducted at the GIPSA-lab (France) starting from 2011 [8, 1].

Number	Name	Number of subjects	Original datasets
9	D1	28	BNCI2014_008, BNCI2014_009, BNCI2015_003
10	D2	175	bi2013a, bi2014a, BI2015a, BI2015b

Table 2: Compound datasets.

2.3 Model: EEG Conformer

The EEG conformer [24] uses convolution to learn local temporal and spatial features and then uses self-attention to learn global temporal features. The overall framework is depicted in Fig. 1. The architecture has three components: a convolution module, a self-attention module, and a fully connected classifier. The convolution module takes raw two-dimensional EEG trials as input, applying temporal and spatial convolutional layers along the time and the number of electrodes dimensions. An average pooling layer follows to reduce noise interference and enhance generalization. Next, the spatiotemporal representation generated by the convolution module is passed into the self-attention module, which further extracts long-term temporal features by measuring global correlations between

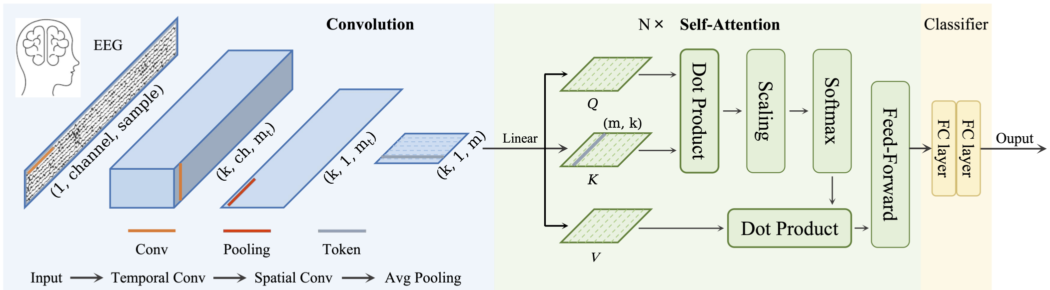

Figure 1: Framework of the EEG conformer, including a convolution module, a self-attention module, and a classifier module (reproduced from [24]).

different time positions in the feature maps. Finally, a compact classifier consisting of several fully connected layers is used to output the decoding results [24].

2.4 EEG Decoding and Evaluation

We extracted epochs of 1 second after the onset of stimulation for all datasets except for D1, for which the epoch length was set to 0.8 seconds. In D1, 1-second epochs may or may not include the last sample, leading to slight variations in the number of samples per epoch (e.g., 127 or 128 samples in a 1-second epoch). By using 0.8-second epochs, we ensured a stable number of samples across all epochs. An infinite impulse response bandpass filter between 1 and 24 Hz was applied to the data. Electrodes were arranged according to the 10–20 international system, but only a subset of the channels was selected for analysis. For the BNCI datasets, the selected channels were Cz, O1, O2, P3, P4, P7, P8, Pz, T7, and T8. For the BI datasets, the selected channels were Cz, Oz, P3, P4, PO7, PO8, and Pz. The first and second subsets correspond to the intersection of all channels within the BNCI and BI datasets, respectively. The preprocessed data were then passed to the EEG conformer model. In the convolutional layer, the number of temporal filters was set to 40, with each filter having a length of 25. The length and stride of the temporal pooling filter were set to 75 and 15, respectively. A dropout rate of 0.5 was applied. In the self-attention module, we used six self-attention layers with ten attention heads each and a dropout rate of 0.5. The dimension of the fully connected layer was set to "auto", meaning it was automatically calculated based on the sampling frequency and the number of input channels. Finally, we set the patience parameter to 50 and the learning rate to 1×10^{-4}, the weight decay to 1×10^{-5}, and we used a validation split of 20%.

Next, we applied the following two pipelines. The first pipeline was used to study the performance of the classifier with different batch sizes (32 or 64), resampling frequencies (128, 256 or 512 Hz), and training epochs (120, 200, 300, 500 and early stop). Its performance was evaluated using a cross-subject validation with one subject left out, and the area under the curve (AUC) was used as the evaluation metric. In the second pipeline, data were resampled to 128 Hz. The batch size of the transformer was fixed at 32, and the number of training epochs was set to 300. Its performance was evaluated using 5-fold cross-subject validation, with AUC as the metric. The cross-validation setup included 5 distinct splits, in which each fold utilized approximately 80% of the subjects for training and the remaining 20% for validation. The performance of the EEG conformer was compared with state-of-art methods based on Riemannian geometry as described in [7]. We used either ERP covariances (also referred to as "super-trials" in the literature) [6] or xDAWN covariances. xDAWN refers to the spatial filter of the same name [21], which is applied before computing the covariance matrices. We then directly classified the covariance matrices using the minimum-distance-to-means (MDM) classifier, or by projecting the covariance matrices to the tangent space before applying linear discriminant analysis (LDA) or a support vector machine (SVM). We used MOABB, Braindecode [22], and MNE [10] to ensure standardization in the implementation and evaluation of these pipelines.

3 Results and Discussion

Table 3 presents the results we obtained for BNCI2014_008, BNCI2014_009, bi2013a, and bi2015a using a cross-subject validation with one subject left out. The AUC increases from 0.75 in Test 2 to

Test	Dataset	First n subjects	Batch size	Resampling frequency (Hz)	Training epochs	xDAWN covariances + MDM	EEG conformer
1	BNCI2014_008	8/8	32	256	300	0.73	0.64
2	BNCI2014_009	6/10	64	128	200	0.85	0.75
3	BNCI2014_009	6/10	64	256	200	0.85	0.80
4	BNCI2014_009	10/10	64	128	300	0.85	0.69
5	bi2013a	10/24	64	512	500	0.75	0.60
6	bi2013a	18/24	32	512	300	0.76	0.66
7	bi2013a	24/24	32	512	early stop	0.76	0.53
8	bi2015a	43/43	32	128	120	-	0.54

Table 3: Cross-subject validation with one subject out.

0.8 in Test 3, while the resampling frequency increased from 128Hz to 256Hz. This indicates that a higher resampling frequency does improve the performance.

On the other hand, the AUC decreases from 0.75 in Test 2 to 0.69 in Test 4, while the number of subjects increases from 6 to 10 and the number of training epochs increases from 200 to 300. This suggests that either the number of epochs was insufficient when including more subjects or that the model struggled to generalize with the additional subjects.

A similar observation can be made by comparing the “First n subjects” column with the last two AUC score columns. Datasets with more subjects do not exhibit the best performance, which could be due to either an insufficient number of epochs or the model’s limited ability to generalize when more data are included. The EEG conformer demonstrated the ability to generalize with all datasets (AUC > 0.6), except perhaps for Tests 7 and 8.

Dataset	EEG conformer	ERP covariance + MDM	xDAWN covariances + MDM	xDAWN covariances + TS + SVM	xDAWN covariances + TS + LDA
BNCI2015-003	51.47	52.63	49.93	58.29	59.70
BNCI2014-008	53.53	59.34	63.15	74.56	74.49
BNCI2014-009	54.96	70.82	77.62	86.90	86.96
bi2013a	55.59	65.28	68.84	76.65	78.01
bi2014a	50.79	56.38	56.79	64.57	63.12
bi2014b	58.74	63.75	68.08	81.72	74.35
bi2015a	58.56	69.18	69.52	77.86	75.90
bi2015b	56.43	60.87	64.15	72.88	74.39
D1	50.63	57.89	53.16	77.08	73.69
D2	49.07	62.80	64.34	80.95	72.06

Table 4: Group 5-fold cross-subject validation.

Table 4 presents the results we obtained for BNCI2014_003, BNCI2014_008, BNCI2014_009, bi2013a, bi2014a, bi104b, bi2015a, bi2015b, D1, and D2 and D2 using 5-fold cross-subject validation. The columns in the table represent the score for the five pipelines we evaluated. As a reminder, D1 is a compound dataset resulting from the merging of all BNCI datasets. Similarly, D2 comprises all the BI datasets.

The AUC differs from Table 3, even on the same datasets and pipelines. This is expected as the evaluation method is different between these two results. The results from these two tables are not intended to be compared. However, similar conclusions to those drawn from Table 3, can be derived from Table 4 when considering the AUC of the EEG conformer (first column). In fact, the AUC of D1 is lower than the minimum AUC observed in the BNCI datasets, and the AUC of D2 is lower than the minimum AUC for the BI dataset. This indicates that the EEG conformer struggles to generalize effectively when the number of subjects is high. The overall low results in 4 can be explained by the fact that performing cross-subject classification usually yields lower results than a within-session classification. Additionally, EEG data has been fed in raw format without the use of normalization - a common step when using CNNs and transformers - which can also explain the lower score results. We also recommend the use of specific transfer learning techniques for BCI such as [28].

3.1 Limitations and Future Work

Our work has two key limitations. First, the number of samples is low. In fact, in the original study, the sampling frequency ranged from 200 to 250 Hz, and the signal epoch length varied from 1 to 7 seconds. Such an epoch length is common while using motor imagery BCI. In our case, some datasets had a lower sampling frequency of 128 Hz, and we limited the epoch length to 1 second due to our use of the P300 ERPs. Indeed, an ERP generally occurs 0–1 s after stimulation. In the cases of D1 and D2 (table 4), we narrowed down the channels to those that were common across all embedded datasets. Additionally, the number of samples could be increased by interpolating time samples and adding missing channels. Such interpolation could be implemented with MNE. The performance tends to diminish as the number of subjects increases, and the overall performance is low as we use cross-subject evaluations, without transfer learning technics. In addition, in the original study, the authors employed data augmentation techniques, specifically segmentation and reconstruction in the time domain, to generate new data. This data augmentation did not increase the number of samples but rather involved cutting and randomly mixing the time samples to create new epochs. In our study, we did not apply such data augmentation techniques.

Second, regarding the model hyperparameters, namely the depth and number of heads in the attention module and the length of the pooling layer, in the original study, it was demonstrated that the attention heads and the pooling layer had minimal impact on the classification accuracy. However, a linear relationship between accuracy and the depth of the attention layer was observed. In our case, we maintained the depth at its default value of 6, despite the potential for it to vary between 0 and 15. Our suggestion for further research would be to set this parameter to 15.

4 Conclusion

In this work, we integrated the EEG conformer into Braindecode, introduced the concept of compound datasets in MOABB, and tested the EEG conformer on eight datasets for a total of 241 subjects, as well as two compound datasets. Results of the cross-subject evaluation, show that the EEG conformer can be successfully applied using Braindecode in conjunction with MOABB. The results show that the method does not outperform state-of-the-art methods like MDM. At least on one dataset, the algorithm was able to generalize well. We have also explored how several parameters affect the performance. However, further work is required to explore a larger parameter space.

References

[1] A. Andreev, A. Barachant, F. Lotte, and M. Congedo. *Recreational Applications of OpenViBE: Brain Invaders and Use-the-Force*, volume chap. 14. John Wiley ; Sons, Hoboken, New Jersey, USA, 2016.

[2] P. Aricò, F. Aloise, F. Schettini, S. Salinari, D. Mattia, and F. Cincotti. Influence of P300 latency jitter on event related potential-based brain–computer interface performance. *Journal of Neural Engineering*, 11(3):035008, May 2014. Publisher: IOP Publishing.

[3] B. Aristimunha, I. Carrara, P. Guetschel, S. Sedlar, P. Rodrigues, J. Sosulski, D. Narayanan, E. Bjareholt, B. Quentin, R. T. Schirrmeister, E. Kalunga, L. Darmet, G. Cattan, A. Abdul Hussain, R. Gatti, V. Goncharenko, J. Thielen, T. Moreau, Y. Roy, V. Jayaram, A. Barachant, and S. Chevallier. Mother of all BCI Benchmarks, Oct. 2023.

[4] C. Brunner, R. Leeb, and G. R. Muller-Putz. BCI Competition 2008–Graz data set A, 2008.

[5] G. H. Cattan and A. Quemy. Case-Based and Quantum Classification for ERP-Based Brain–Computer Interfaces. *Brain Sciences*, 13(2):303, Feb. 2023. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

[6] M. Congedo. *EEG Source Analysis*. Habilitation à diriger des recherches, Université de Grenoble, Oct. 2013.

[7] M. Congedo, A. Barachant, and A. Andreev. A New Generation of Brain-Computer Interface Based on Riemannian Geometry. *arXiv:1310.8115 [cs, math]*, Oct. 2013.

[8] M. Congedo, M. Goyat, N. Tarrin, G. Ionescu, L. Varnet, B. Rivet, R. Phlypo, N. Jrad, M. Acquadro, and C. Jutten. "Brain Invaders": a prototype of an open-source P300- based video game working with the OpenViBE platform. In *5th International Brain-Computer Interface Conference 2011 (BCI 2011)*, pages 280–283, Styria, Austria, Sept. 2011.

[9] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu. Differential entropy feature for EEG-based emotion classification. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 81–84, Nov. 2013. ISSN: 1948-3554.

[10] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. Hämäläinen. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7, 2013.

[11] C. Guger, S. Daban, E. Sellers, C. Holzner, G. Krausz, R. Carabalona, F. Gramatica, and G. Edlinger. How many people are able to control a P300-based brain-computer interface (BCI)? *Neuroscience Letters*, 462(1):94–98, Oct. 2009.

[12] V. Jayaram and A. Barachant. MOABB: trustworthy algorithm benchmarking for BCIs. *Journal of Neural Engineering*, 15(6):066011, Sept. 2018.

[13] L. Korczowski, M. Cederhout, A. Andreev, G. CATTAN, P. L. C. Rodrigues, V. Gautheret, and M. Congedo. Brain Invaders calibration-less P300-based BCI with modulation of flash duration Dataset (bi2015a). Research Report, GIPSA-lab, 2019.

[14] L. Korczowski, M. Cederhout, A. Andreev, G. CATTAN, P. L. C. Rodrigues, V. Gautheret, and M. Congedo. Brain Invaders Cooperative versus Competitive: Multi-User P300- based Brain-Computer Interface Dataset (bi2015b). Research Report, GIPSA-lab, 2019.

[15] L. Korczowski, E. Ostaschenko, A. Andreev, G. CATTAN, P. L. C. Rodrigues, V. Gautheret, and M. Congedo. Brain Invaders calibration-less P300-based BCI using dry EEG electrodes Dataset (bi2014a). Research Report, GIPSA-lab, 2019.

[16] L. Korczowski, E. Ostaschenko, A. Andreev, G. CATTAN, P. L. C. Rodrigues, V. Gautheret, and M. Congedo. Brain Invaders Solo versus Collaboration: Multi-User P300-based Brain-Computer Interface Dataset (bi2014b). Research Report, GIPSA-lab, 2019.

[17] R. Leeb, C. Brunner, G. R. Muller-Putz, and A. Schlogl. BCI Competition 2008 – Graz data set B, 2008.

[18] Q. Li, T. Zhang, Y. Song, and M. Sun. Transformer-based spatial-temporal feature learning for P300. In *2022 16th ICME International Conference on Complex Medical Engineering (CME)*, pages 310–313, Nov. 2022.

[19] A. Quemy. Binary Classification in Unstructured Space With Hypergraph Case-Based Reasoning. *arXiv:1806.06232 [cs, stat]*, Mar. 2019. arXiv: 1806.06232.

[20] A. Riccio, L. Simione, F. Schettini, A. Pizzimenti, M. Inghilleri, M. Olivetti Belardinelli, D. Mattia, and F. Cincotti. Attention and P300-based BCI performance in people with amyotrophic lateral sclerosis. *Frontiers in Human Neuroscience*, 7, Nov. 2013. Publisher: Frontiers.

[21] B. Rivet and A. Souloumiac. Optimal linear spatial filters for event-related potentials based on a spatio-temporal model: Asymptotical performance analysis. *Signal Processing*, 93(2):387–398, 2013.

[22] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, 38(11):5391–5420, Nov. 2017. arXiv:1703.05051 [cs].

[23] Y. Song, X. Jia, L. Yang, and L. Xie. *Transformer-based Spatial-Temporal Feature Learning for EEG Decoding*. June 2021.

[24] Y. Song, Q. Zheng, B. Liu, and X. Gao. EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:710–719, 2023. Conference Name: IEEE Transactions on Neural Systems and Rehabilitation Engineering.

[25] E. Vaineau, A. Barachant, A. Andreev, and M. Congedo. *Brain Invaders 2013a*. Zenodo, Nov. 2018.

[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, u. Kaiser, and I. Polosukhin. Attention is All you Need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[27] J. Xie, J. Zhang, J. Sun, Z. Ma, L. Qin, G. Li, H. Zhou, and Y. Zhan. A Transformer-Based Approach Combining Deep Learning Network and Spatial-Temporal Information for Raw EEG Classification. *IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society*, 30:2126–2136, 2022.

[28] P. Zanini, M. Congedo, C. Jutten, S. Said, and Y. Berthoumieu. Transfer Learning: A Riemannian Geometry Framework With Applications to Brain-Computer Interfaces. *IEEE transactions on bio-medical engineering*, 65(5):1107–1116, May 2018.

[29] W.-L. Zheng and B.-L. Lu. Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Transactions on Autonomous Mental Development*, 7:1–1, Sept. 2015.