2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

Residual Attention Module on EEGNet for Brain-Computer Interface

Davi Esteves dos Santos
Federal University of Juiz de Fora
Juiz de Fora, Brazil
0009-0003-3572-6071

Gabriel Henrique de Souza
Federal University of Juiz de Fora
Juiz de Fora, Brazil
0000-0002-5931-654X

Heder Bernardino
Federal University of Juiz de Fora
Juiz de Fora, Brazil
0000-0003-2012-7802

Alex Borges Vieira
Federal University of Juiz de Fora
Juiz de Fora, Brazil
0000-0003-0821-126X

Luciana Paixão Motta
Federal University of Juiz de Fora
Juiz de Fora, Brazil
0000-0001-5865-4830

***Abstract*—Brain-computer interfaces (BCI) allow for the brain to communicate with electronic devices. Concerning the BCI paradigms, motor imagery uses brain signals to decode an imagined movement. However, this is a hard task given the low signal-to-noise ratio. Usually, the main steps in BCI models are pre-processing, feature extraction, and classification. In recent years, Convolutional Neural Networks (CNNs) have been gaining relevance in several areas of science due to their feature extraction, translation invariance, and parameter sharing capabilities. Another, more recent way of feature extraction is using attention mechanisms, which are layers of neural networks based on human attention and have the ability to highlight important features. A variation of the attention mechanism is the Convolutional Block Attention Module, which combines the CNN structure with the attention mechanism. In this work, we propose a new model that joins the core architecture of EEGNet, a compact CNN widely used in the literature, with the Convolutional Block Attention Module and residual connections. The residual connections were introduced to lower data degradation throughout the model. The results highlight the residual connection's importance for the performance of the model. The proposed model obtained a kappa result 5.2% better than the EEGNet with a p-value less than 0.01 on BCI Competition IV dataset 2a, which is a well-known dataset for Motor Imagery. Furthermore, the proposal was better than EEGNet for most subjects and had the best-worst case.**

***Index Terms*—Brain-Computer Interface, Motor Imagery, Convolutional Neural-Network, Attention Mechanism, Electroencephalogram**

I. Introduction

Brain-Computer Interfaces (BCIs) allow for direct communication between the brain and electronic systems [1]. It Improves a wide range of fields, including rehabilitation, healthcare, neuroscience, and entertainment [2]. Many brain functions can be used in BCIs, such as visual information [3] and motor imagery (MI) [4]. Each brain function has a scope of usually used applications, for example, MI is used in post-stroke motor rehabilitation [5] and prosthesis control [6].

The standard approach to BCI can be divided into five main steps [7]: (i) signal acquisition: Brain signals acquisition; (ii) Preprocessing: signal denoising and filtering procedures; (iii) Extraction and selection: procedures to capture essential properties from the signals; (iv) Classification: classifies the signals using the features; and (v) Task Execution: the classification label is used by the application. Each of these steps requires distinct approaches depending on the specific paradigm under investigation.

Several established approaches can be used in each step presented above, such as filter bank [8], common spatial pattern (CSP) [9], linear discriminant analysis (LDA) [10], Support Vector Machine (SVM) [11], and Independent Component Analysis (ICA) [12]. Filter Bank Common Spatial Patterns (FBCSP) [8] is an example of a BCI model that uses a filter bank as temporal filtering, CSP as spatial filtering, LogPower as Feature Extraction, and Naive-Bayes Parzen-Window (NBPW) as a classifier. Thus, expertise in the field and careful consideration of the signal's properties are important to minimize losses of information.

Convolutional Neural Networks (CNNs) have emerged as a popular choice for MI tasks due to their pattern recognition [13]. Notably, the EEGNet model [14] stands out as an example of CNN architecture that effectively integrates the FBCSP steps within its convolutional structure. EEGNet also has a large number of works with Field Programmable Gate Arrays (FPGA) due to its compact size [15], [16].

By leveraging neural network techniques, such as CNNs, researchers can enhance the performance of BCI in some applications. The attention mechanism [17] has gained popularity due to its powerful capabilities, especially in the transformer structure [18]. The attention mechanism allows for the network to focus on relevant features or regions of interest within the input data, enhancing its ability to capture important information [18]. A very successful variant of the attention mechanism is the Convolutional Block Attention Module (CBAM) proposed by Woo et al. [19]. Woo, also proposed the use of residual connection [20] jointly with CBAM, enabling the models to increase in depth and complexity without degradation.

The authors thank the support provided by CAPES, CNPq, FAPEMIG, FAPESP, UFJF, and OpenBCI.

In this paper, we propose a novel EEGNet-based model approach that uses CBAM and Residual connections, aiming to increase the quality of the solutions without compromising EEGNet's small size. The residual connections in our model control the deterioration of the signal. Moreover, the proposal extracts new features from the signal by using the attention mechanism. The results indicate that the combination of attention and EEGNet improves the previous models when the residual connection is included. In addition, the proposed model showed to have better results and with less variation between different subjects.

The remaining of this paper is as follows: Section II presents a summary of literature for EEGNet, Attention mechanism, and their applications in BCI; Section III describes the dataset used in this work; Section IV describes the EEGNET and CBAM models; Section V presents the proposed method EEGRCBAM; Section VI describes the experiment's procedures and discusses the results obtained; and Section VII - concludes this study and presents some future works.

II. Related Works

This section describes published articles related to EEGNet and attention mechanisms in BCI applications.

Attention mechanisms in neural networks have their roots in natural language models and machine translation. The concept of attention was first proposed by Bahdanau et al. in 2015 [17]. In 2017, Vaswani et al. [18] extended the application of attention mechanisms to the transformer architecture. Transformers have achieved remarkable success in various domains, including natural language processing [18], computer vision [21]–[23], and speech recognition [24]. The attention mechanism in transformers enabled the models to capture long-range dependencies more effectively and improved their ability to process sequential data.

In 2018, Woo et al. [19] introduced the convolutional block attention module (CBAM), leveraging the strengths of CNNs and attention mechanisms. CBAM enabled improved localized feature extraction, enhancing the network's ability to capture fine-grained details and spatial relationships.

Building upon these advancements, attention mechanisms started to find applications in BCIs. Zhang et al. [25] were among the first to apply attention mechanisms in the BCI domain, demonstrating its efficacy. Similarly, Li et al. [4] explored the application of attention mechanisms in Motor Imagery (MI) tasks, highlighting its potential to enhance performance in BCI paradigms. After that, it was introduced temporal and channel-wise attention-based neural networks for MI analysis [26]–[28]. Zhang et al. [29] proposed a filter bank approach together with attention mechanisms to improve feature extraction. Lastly, Wen et al. [30] put forth an attention-based 3d densely connected cross-stage-partial network architecture. The use of attention mechanisms in BCI already points to improvements in models, both in accuracy and robustness between the subjects. However, more experiments still need to be performed given the number of possibilities that attention enables.

Therefore, several models have emerged as popular choices with the growing use of neural networks in BCI research. Among these models, DeepConvNet, and ShallowConvNet proposed by Schirrmeister et al. [31] and EEGNet proposed by et al. [14] have gained significant attention. Notably, EEGNet stands out as a compact model that manages to maintain meaningful accuracy levels. Following that, Yu et al. [32] analyzed the impact of a CBAM layer with EEGNet.

III. Dataset

The BCI competition IV 2A dataset [33] was selected due to its wide use in motor imagery (MI) researches. The experiments used a setup of 22 electrodes, with a sampling rate of 250 Hz. The bandpass filter was set between 0.5-100 Hz. During the MI tasks, each trial lasted for 6 seconds followed by a short break time. During the initial 2 seconds, a fixation cross is presented on the screen, followed by a 1.25-second cue. The cue overlapped with the subsequent 4-second interval where the subjects performed the motor imagery. Four classes of MI tasks were considered: right hand, left hand, feet, and tongue. A total of 9 subjects participated in 2 separate sessions, with each session consisting of 288 trials equally distributed among the 4 classes.

IV. Methods

In this section, we present the foundation core models for our EEGRCBAM proposal, EEGNet, and CBAM.

A. EEGNet

EEGNet [14] is a CNN Inspired by the FBCSP algorithm, highlighted by its concise structure of 2 main blocks. The structure can be seen in further detail in Figure 1. In block 1, the main steps are: (i) Reshape; (ii) Temporal 2D convolution, with a kernel size of (1, 64), the length is half the sampling rate (128hz), with F_1 temporal filters; (iii) Batch normalization; (iv) DepthWiseConv2D, with a kernel size of number of channels and $D * F_1$ filters; (v) Batch normalization; (vi) ELU activation function; (vii) AveragePool2D with size 4; and (viii) Dropout.

In block 2: (i) SeparableConv2d with a kernel size of 16 and F_2 filters; (ii) Batch normalization; (iii) ELU activation function; (iv) AveragePool2D with size (1, 8); (v) Dropout; (iv) Flatten; and (v) Softmax.

B. Convolutional Block Attention Module

A variant of the attention mechanism, Convolutional Block Attention Module (CBAM) [19], can be integrated into CNN modules, combining the global feature extraction of the attention mechanism with CNN's inherent pattern recognition capacity [34].

CBAM consists of 2 main modules, the first one is the channel attention module, which applies the following steps: (i) Parallel AvgPool and MaxPool; (ii) Both pooling are passed through a shared Multilayer Perceptron (MLP); (iii) Element-wise summation of the MLP outputs; (iv) Sigmoid activation function; and (v) Initial input undergoes element-wise multiplication with step's iv output.

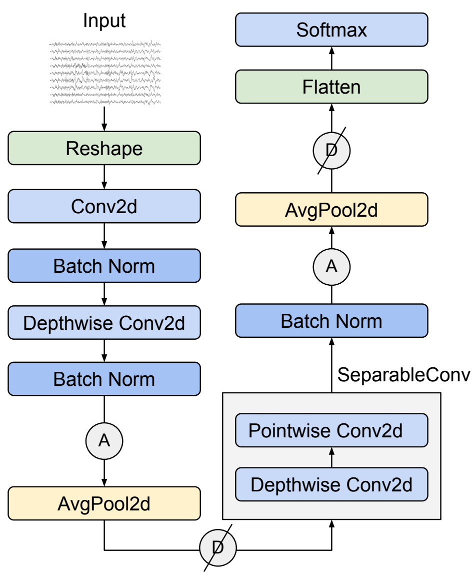

Fig. 1. EEGNet structure. The letter A stands for Elu activation function. D stands for dropout. D is crossed by a line as it's not used in the test process. Adapted from [14].

And the spatial attention module that has the following steps: (i) Parallel AvgPool and MaxPool; (ii) 2D convolution that goes through both pools; (iii) Sigmoid activation function; and (iv) The channel-attention output undergoes element-wise multiplication with step Sigmoid layer output.

CBAM and its modules can be seen in Figure 2. It's worth noting that, the inputs are represented as 2D in BCI, despite CBAM's usual 3D inputs otherwise. This representation choice is a consequence of the depthwise convolution in EEGNet, which makes the EEG channels have always depth 1 and, as such, can be considered a 2D input.

V. Proposed Approach

The proposed EEGNet-Residual-CBAM (EEGRCBAM) is a variant of the original EEGNet, built upon its core themes of compactness and efficacy. Inspired by the architecture used in EEGNet-CBAM (EEGCBAM) [32]. The main differences from EEGRCBAM's to EEGCBAM are the addition of a residual connection and a concatenation layer. The inclusion of CBAM accentuates features captured by the EEGNet-Block, which is the EEGNet architecture without its final flatten and softmax layers.

The residual connections [20] alleviate degradation as the architecture goes deeper while facilitating the flow of gradients, enabling efficient training. Finally, by concatenating the refined features with EEGNet-block's output, it improves the softmax classification layer, by not relying solely on the attention layer. These processes can be summarized into the following steps: (i) EEGNet-block; (ii) Layer Normalization [35]; (iii) Residual CBAM; (iv) Layer Normalization; (v) Concatenation of the normalization layers outputs; (vi) Flatten; and (vi) Softmax. The Normalization layers are important to parameter regularization and are not dependent on mini-batches. Their inclusion ensures robust regularization and enhances the generalization capability of the network.

EEGRCBAM and EEGCBAM structures can be seen in further detail in Figure 3.

VI. Computational Experiments

The BCI Competition dataset presented in Section III was chosen to perform the experiments. Its pre-processing was: downsampling to 128Hz and a bandpass filter of 4-40Hz, following the experimental setup of EEGNet [14].

The evaluated models are EEGnet, EEGCBAM, and EEGCRBAM. EEGNet and EEGCBAM were chosen as they are networks that present good results for MI-BCI and have similar sizes to the proposed model.

The temporal kernel sizes of 32 and 64 were chosen as both were used in [14]. We defined the model with kernel size equal to 32, 64 as $(\cdot)_{32}$ and $(\cdot)_{64}$.

The implementation and training were performed using the PyTorch framework [36]. A batch size of 64 was used as this value is widely adopted in the literature [37] and 1000 iterations, with a learning rate of 9×10^{-4}. We tested and trained the model for each subject using both sessions and 5-fold stratified cross-validation. For training and test, the data was set from 0.5 to 2.5 seconds after the cue as used in [14]. The training procedure employed the cross-entropy loss function and the Adam optimizer. For statistical hypothesis testing, the Wilcoxon test was used. The codes and results are publicly available at https://github.com/Davi-Esteves-dos-Santos/EEGRCBAM.

A. Results and Discussion

Table I presents the kappa values obtained for the models tested here. The kappa score was chosen due to it being used as the performance metric in BCI Competition IV 2a.

Overall, the proposed EEGRCBAM$_{32}$ reached the best average results when compared to the other approaches, outperforming EEGNet$_{64}$ in average kappa by 5.22%. One can observe that EEGRCBAM$_{32}$ is the only one that obtained results statistically different to those found by (1) EEGNet$_{64}$ considering p-value ¡ 0.01, and (2) the other methods with p-value ¡ 0.001. Moreover, EEGRCBAM$_{32}$ maintained a higher or equal average kappa for most subjects. The EEGRCBAM$_{64}$ did not have a good result. A possible reason is the increases in the number of trainable parameters, thus, making it more susceptible to overfitting. However, more experiments need to be performed to validate this hypothesis.

EEGNet$_{64}$ and EEGCBAM$_{64}$ had similar results, as in [32]. Another noteworthy comparison is its relatively modest increase in size, with EEGNet$_{64}$ having 1716 parameters, EEGRCBAM$_{64}$ with 3286, and EEGRCBAM$_{32}$ 3030. EEGRCBAM$_{32}$ is 76.6% larger than EEGNet$_{64}$, but it still is concise compared to other popular approaches from the literature, up to 13 times smaller than ShallowConvNet, and 50 times smaller than DeepConvNet.

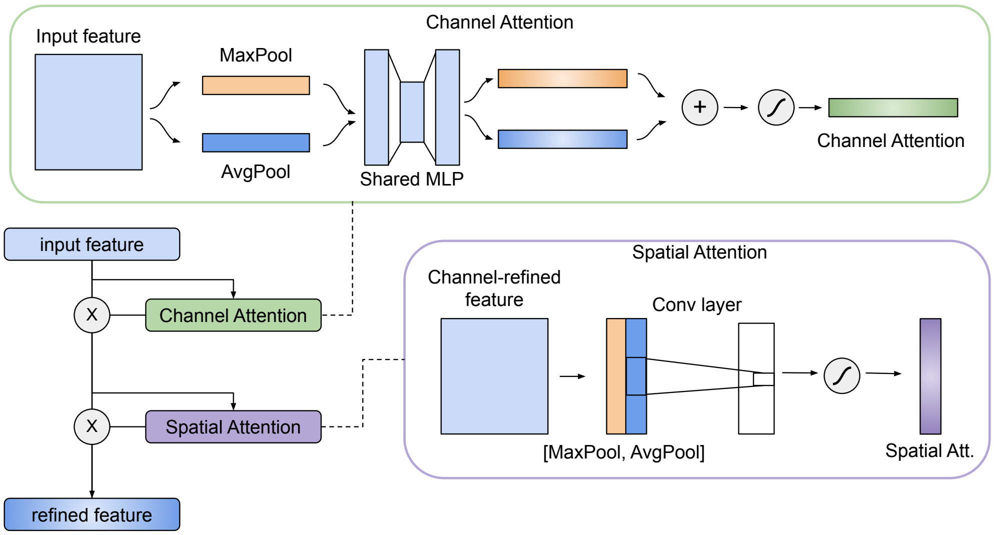

Fig. 2. Convolutional Block Attention Module. Adapted from [19].

TABLE I
KAPPA SCORE RESULTS FOR 5-FOLD STRATIFIED CROSS-VALIDATION. THE BEST VALUE FOR EACH SUBJECT IS BOLDFACE. P-VALUES FOR EACH METHOD AGAINST $EEGRCBAM_{32}$: (1) P-VALUE < 0.01 AND (2) P-VALUE < 0.001

Subject	Method					
	EEGNet_{64}	EEGNet_{32}	EEGCBAM_{64}	EEGRCBAM_{64}	EEGCBAM_{32}	EEGRCBAM_{32}
1	0.6759	**0.7037**	0.6759	0.6505	0.6898	**0.7037**
2	0.2940	0.2986	0.2940	0.2523	0.2917	**0.3426**
3	0.8009	0.7847	0.7963	0.8218	0.8287	**0.8356**
4	0.4329	0.4329	0.4282	0.4120	0.4005	**0.4699**
5	**0.4236**	0.3403	0.2708	0.2500	0.2870	0.3681
6	**0.3796**	0.3171	0.3542	0.3565	0.3657	0.3773
7	0.6042	0.6111	**0.7176**	0.7153	0.6829	0.6944
8	0.7384	0.7292	0.7500	0.7685	0.7593	**0.7778**
9	0.7894	0.8125	0.8241	0.8194	0.8125	**0.8380**
Average	0.5710 (1)	0.5589 (2)	0.5679 (2)	0.5607 (2)	0.5687 (2)	**0.6008**

For a better overall picture of the models, we used performance profiles [38]. We defined the set S of models s_i, and P as the set of problems p_j. $t_{p,s}$ is the inverse average kappa score results of each method and problem, as suggested by Souza et al. [7].

The performance ratio $r_{p,s}$ is defined as:

$$r_{p,s} = \frac{t_{p,s}}{\min\{t_{p,s} : s \in S\}}, \quad (1)$$

Defining the cardinality of a set by its absolute value, we have:

$$\rho_s(\tau) = \frac{1}{n_p}|\{p \in P : r_{p,s} \leq \tau\}|, \quad (2)$$

The three key insights shown in the performance profiles are: (i) the largest $\rho(1)$ is observed in the approach that obtained the best results; (ii) the most robust approach is that one that reaches $\rho(\tau) = 1$ with the smaller tau; and (iii) the best overall approach is to have the largest area under the performance profile curves.

Figure 4 shows that EEGRCBAM_{32} has the best performance for most subjects. Also, it is the most robust method as it is the first to reach $\rho(\tau) = 1$. Furthermore, it is the best approach as it has the largest area. EEGNet_{64} is the second most robust method, having very similar results to the other approaches when $\tau < 1.2$. The worst approach concerning robustness is EEGCBAM_{64}.

VII. CONCLUSION AND FUTURE WORKS

Brain-computer interfaces (BCI) have a wide range of applications and uses, such as Field Programmable Gate Arrays (FPGA), one of the tools to enable this is the Neural Network EEGNet, which is a well-known approach in the literature. One way to improve EEGNet is to use attention mechanisms like Convolutional Block Attention Module (CBAM) which highlights global features with the attention mechanism while keeping itself compact. However, residual connections were needed to keep some features throughout the model and mitigate degradation.

In this paper, we proposed the EEGNet-Residual-CBAM (EEGRCBAM) by introducing two new steps in EEGNet-CBAM (EEGCBAM): (i) a residual connection in

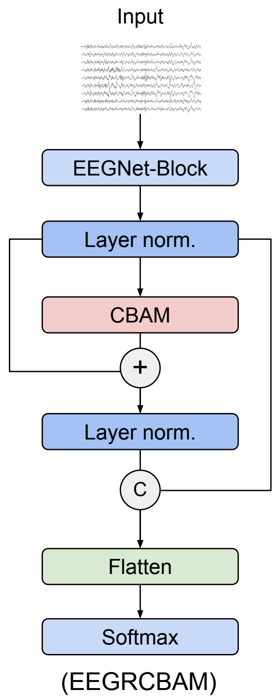

Fig. 3. Proposed methods. The symbol + stands for element-wise summation. C stands for concatenation.

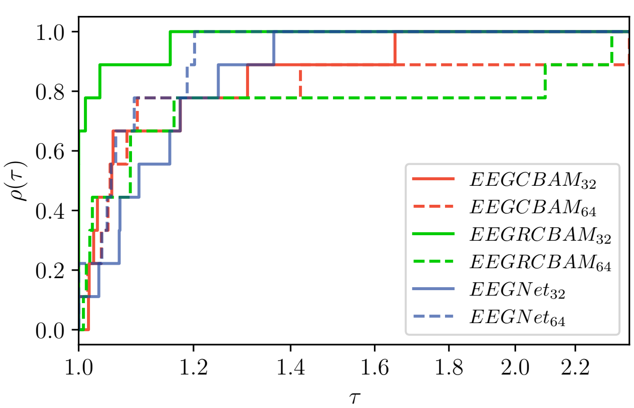

Fig. 4. Performance profiles using the subject average of the kappa results obtained.

CBAM and (ii) a concatenation of the normalized layers result. The proposed EEGRCBAM was evaluated for Motor Imagery (MI) through the BCICIV2a dataset.

EEGRCBAM performed better in all evaluated subjects compared to EEGCBAM and performed better in most of the evaluated subjects compared to EEGNet. In addition, EEGRCBAM presented the highest robustness between the subjects among the evaluated models, having the best-worst case. These results point to the importance of the residual layer with CBAM in BCI applications.

All evaluated models showed sensitivity to the size of the temporal kernel. EEGNet performed better with larger kernels while EEGCBAM and EEGRCBAM performed better with smaller kernels. More parameter sensitivity analysis can be performed to improve the model. Additionally, further testing with different datasets can be conducted to validate the model's generalization capabilities across various subjects.

EEGNet has a large number of works with FPGA, this is due to its compact architecture when compared to other CNNs in the literature. Such as EEGNet, EEGRCBAM has a low number of parameters and keeps competent accuracy further enhancing its suitability for practical implementation, even in FPGA, highlighting its potential for real-time applications and contexts with limited resources.

References

[1] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, and T. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, 2002.

[2] L. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, 2012.

[3] C. Verbaarschot, D. Tump, A. Lutu, M. Borhanazad, J. Thielen, P. van den Broek, J. Farquhar, J. Weikamp, J. Raaphorst, J. Groothuis, and P. Desain, "A visual brain-computer interface as communication aid for patients with amyotrophic lateral sclerosis," *Clinical Neurophysiology*, 2021.

[4] D. Li, J. Xu, J. Wang, X. Fang, and Y. Ji, "A multi-scale fusion convolutional neural network based on attention mechanism for the visualization analysis of eeg signals decoding," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2020.

[5] M. Cervera, S. Soekadar, J. Ushiba, J. Millán, M. Liu, N. Birbaumer, and G. Garipelli, "Brain-computer interfaces for post-stroke motor rehabilitation: a meta-analysis," *Annals of Clinical and Translational Neurology*, 2018.

[6] D. Murphy, O. Bai, A. Gorgey, J. Fox, W. Lovegreen, B. Burkhardt, R. Atri, J. Marquez, Q. Li, and D.-Y. Fei, "Electroencephalogram-based brain-computer interface and lower-limb prosthesis control: A case study," *Frontiers in Neurology*, 2017.

[7] G. H. De Souza, A. B. Vieira, H. S. Bernardino, and H. J. Barbosa, "Differential evolution based spatial filter optimization for brain-computer interface," in *Proceedings of the Genetic and Evolutionary Computation Conference*, 2019.

[8] K. Ang, Z. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (fbcsp) in brain-computer interface," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008.

[9] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust eeg single-trial analysis," *IEEE Signal Processing Magazine*, 2008.

[10] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, "Single-trial analysis and classification of erp components - a tutorial," *NeuroImage*, 2011.

[11] L. Bruzzone and M. Marconcini, "Domain adaptation problems: A dasvm classification technique and a circular validation strategy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010.

[12] I. Winkler, S. Haufe, and M. Tangermann, "Automatic classification of artifactual ica-components for artifact removal in eeg signals," *Behavioral and Brain Functions*, 2011.

[13] K. Hossain, M. Islam, S. Hossain, A. Nijholt, and M. Ahad, "Status of deep learning for eeg-based brain–computer interface applications," *Frontiers in Computational Neuroscience*, 2023.

[14] 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*, 2018.

[15] A. Tsukahara, Y. Anzai, K. Tanaka, A. Homma, and Y. Uchikawa, "A design and trial production of eegnet based eeg pattern recognition processor using fpga," in *2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC)*, 2020.

[16] L. Feng, H. Shan, Y. Zhang, and Z. Zhu, "An efficient model-compressed eegnet accelerator for generalized brain-computer interfaces with near sensor intelligence," *IEEE Transactions on Biomedical Circuits and Systems*, 2022.

[17] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *International Conference on Learning Representations*, 2015.
[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017.
[19] S. Woo, J. Park, J.-Y. Lee, and I. Kweon, "Cbam: Convolutional block attention module," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018.
[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
[21] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021.
[22] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020.
[23] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations*, 2021.
[24] A. Gulati, J. Qin, C.-C. Chiu, N. Parmar, Y. Zhang, J. Yu, W. Han, S. Wang, Z. Zhang, Y. Wu, and R. Pang, "Conformer: Convolution-augmented Transformer for Speech Recognition," in *Proc. Interspeech 2020*, 2020.
[25] D. Zhang, K. Chen, L. Yao, and S. Wang, "Ready for use: Subject-independent movement intention recognition via a convolutional attention model," in *CIKM 2018 - Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2018.
[26] X. Liu, R. Shi, Q. Hui, S. Xu, S. Wang, R. Na, Y. Sun, W. Ding, D. Zheng, and X. Chen, "Tcacnet: Temporal and channel attention convolutional network for motor imagery classification of eeg-based bci," *Information Processing and Management*, 2022.
[27] X. Liu, Q. Hui, S. Xu, S. Wang, R. Na, Y. Sun, X. Chen, and D. Zheng, "Tacnet: Task-aware electroencephalogram classification for brain-computer interface through a novel temporal attention convolutional network," in *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, 2021.
[28] X. Liu, Y. Shen, J. Liu, J. Yang, P. Xiong, and F. Lin, "Parallel spatial–temporal self-attention cnn-based motor imagery classification for bci," *Frontiers in Neuroscience*, 2020.
[29] Y. Zhang, S. Qiu, W. Wei, X. Ma, and H. He, "Dynamic weighted filter bank domain adaptation for motor imagery brain-computer interfaces," *IEEE Transactions on Cognitive and Developmental Systems*, 2022.
[30] Y. Wen, W. He, and Y. Zhang, "A new attention-based 3d densely connected cross-stage-partial network for motor imagery classification in bci," *Journal of Neural Engineering*, 2022.
[31] 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*, 2017.
[32] H. Yu, Y. Deng, F. Yan, Z. Guan, and F. Peng, "An improved schema of brain-computer interface based on motor imagery with eye movement," in *Journal of Physics: Conference Series*, 2022.
[33] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. Miller, G. Müller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the bci competition iv," *Frontiers in Neuroscience*, 2012.
[34] R. Sicre and F. Jurie, "Discriminative part model for visual recognition," *Computer Vision and Image Understanding*, 2015.
[35] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
[36] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," 2019.
[37] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, "On large-batch training for deep learning: Generalization gap and sharp minima," 2017.
[38] E. Dolan and J. Moré, "Benchmarking optimization software with performance profiles," *Mathematical Programming, Series B*, 2002.