Attention-Based CNN Capturing EEG Recording's Average Voltage and Local Change

Long Yi[1], Xiaodong Qu[2]

[1] Brandeis University, Waltham MA 02453, USA
longyi@brandeis.edu
[2] Swarthmore College, Swarthmore PA 19081, USA
xqu1@swarthmore.edu

Abstract. The attention mechanism is one of the most popular deep learning techniques in recent years and it is arguably able to produce human-interpretable results. In this research, we developed a classification model combining two self-attention modules and a convolutional neural network. This model achieved benchmark or superior performance on two electroencephalography (EEG) recording datasets. Moreover, we demonstrated that the self-attention modules were able to capture features, including average voltage of signal features and instant voltage change of the EEG signals, by visualizing the attention maps they produced.

Keywords: Self-Attention, CNN, EEG Classification, Brain-Machine Interface

1 Introduction

Brain-machine interface (BMI) aims to recover or enhance human abilities. BMI has benefited from the recent development of machine learning and deep learning [11]. For example, BMI enabled a paralyzed man to "type" 90 characters per minute in recent research [28]. BMI can also identify a variety of cognitive tasks in healthy subjects such as learning programming language [18,19,21]. Most research used the convolutional neural network (CNN), a type of deep neural network (DNN), as the model of choice [9,15]. Despite its general superior performance, CNN is notorious for its black-box nature. In other words, researchers are unable to understand the inner mechanism of the neural network or connect the latent features CNN learns to meaningful and human-understandable neural patterns [31]. More and more researchers are considering interpretability as equally, if not more, important than prediction performance in current BMI research, as it could enable researchers to: 1. verify the reliability of the neural networks, 2. discuss and improve the neural networks, 3. even gain insight into complex neurobiological patterns that human experts can not interpret [2,16,31].

Attention has been one of the most impactful machine learning techniques in recent years, and researchers have used it to achieve a number of breakthroughs since its invention, especially in the field of natural language processing. Self-attention is one of the attention mechanisms and it gains its name from the analogy to the human cognitive attention. As its name suggests, self-attention assigns weights (i.e., paying different levels of attention) to every element within a sequence or collection based on their degrees of importance. Therefore, human researchers can easily gain insights into self-attention models' inner operations. Moreover, Zhang et al. [31] stated that a general BMI framework requires the ability to both focus on and capture the most informative features of input data under different sizes and conditions, and the attention mechanism and CNN are respectively the ideal deep learning techniques for these two jobs. Based on this philosophy, we implemented a lightweight model combining two self-attention modules and a CNN classifier in this work. We experimented with the model on two electroencephalography (EEG) classification datasets, and it achieved comparable results to another type of state-of-the-art attention-based CNN model on both datasets. The two self-attention modules work on the temporal and spectral dimensions, respectively. We visualized the outputs from the two self-attention modules and demonstrated that they were able to capture global statistics and local changes of the EEG signals.

2 Related Work

The attention-like mechanism was first proposed in the 1960s [4], yet it was not until 2014 that this concept was successfully implemented as a machine learning module that reached meaningful results on real-world problems [1]. Since then, researchers have created different variations of the attention mechanism and incorporated them into traditionally powerful models. The self-attention mechanism was introduced in 2016 as a way to learn semantic correlations between words in a sentence [6]. Shortly after that, Vaswani et al. [26] created the Transformer model that is solely built on the attention mechanism.

Regarding the attention-based CNN (i.e., combination models of the attention mechanism and CNN) that this work focuses on, researchers have devised various configurations of them, and they excelled in performance on different tasks, including computer vision [29], natural language processing [30], and so on [5].

Within the scope of BMI research, researchers have recently used the attention modules or Transformer models to achieve high performance as well, on the grounds that there are correlations between time steps and signal sources of EEG recordings [13,25]. There were also a few studies that combined the attention mechanism with CNN and achieved state-of-the-art level prediction accuracy on EEG classification tasks [7,14,25]. However, to the best of the authors' knowledge, the incorporation of the attention mechanism into CNN in previous research can be categorized in two ways: 1. attention modules operate within the convolution layers and help capture latent features [7,14], 2. Transformer models receive inputs from CNN and generate predictions [25]. Different from these two approaches, we used the self-attention modules to learn important features and used the CNN to capture them. In other words, we used the self-attention modules to generate attention maps and fed them into the CNN as a classifier to make the classification predictions.

Regarding the work that wrestled with the interpretability problem in BMI, there were two teams of researchers who combined the interpretable Layer-wise Relevance Propagation modules with the powerful but black-box DNNs to enhance the model interpretability [2,24]. We incorporated the self-attention mechanism and CNN in a similar fashion, that the two parts of the model worked complementarily and were respectively designated to achieve high accuracy and enhance interpretability. Besides, there were other research groups who used the attention mechanism to visualize the correlation maps of subjects' brain activities on different motor imagery tasks (i.e., imagining in head but not performing a movement or task) [7,14,25]. Different from their work, we managed to interpret the model operation and distinguish distinct neurobiological features the model learned in our work.

Previous researches have used the attention-mechanism to capture features from the temporal, spectral, and spatial dimensions [13]. Based on the condition of the datasets we used, we employed two self-attention modules to capture features on the temporal and spectral dimensions [14,25]. We will call them Temporal Attention and Spectral Attention in this paper.

3 Experiment

Dataset We experimented with our proposed model on two EEG datasets that differed in various conditions to show the validity and generalization power of our model.

The first dataset consists of 14 subjects' EEG recordings. Each subject performed five cognitive tasks during the recording in random order — reading, copying, answering, writing, and typing. Subjects were recorded using the Muse headbands with four data acquisition nodes, and the signal transmission rate was 10Hz. Each recording node's signal was decomposed into the five classic bands of brainwaves (i.e., α, β, γ, δ, and θ), thus producing 20 channels. The absolute Band Powers (BP) features were computed and used as our model inputs. We used non-overlapping 500-millisecond-long EEG recordings to feed into the model. In other words, the model inputs were 5 by 20 matrices. For more detail about the dataset, please refer to the original work that collected and analyzed the dataset [20].

The second dataset we used was the publicly available BCI competition III dataset V [3]. During recording, subjects performed three different motor imagery tasks — left-hand movement, right-hand movement, and word generation. The dataset contained three subjects' recordings, but we only used the first subject's. We used the power spectral density (PSD) features, and each PSD sample was a vector of 96 values (8 channels $\times$ 12 frequency components). Similarly, we used non-overlapping 500-millisecond-long recordings to feed into the model. The model inputs were 8 by 96 matrices.

We will refer to the first dataset as RWT in this paper. We used 90% randomly-split data to train the model and the remaining 10% to test it. We will refer to the second dataset using the abbreviation of BCI III, and we used 75% of the data for training and 25% for testing. A summary and comparison of the two datasets can be found in Table 1. Please refer to the original papers for more details.

Model and Training Fig. 1 A depicts the model's architecture. The EEG inputs will first be fed in parallel into the two self-attention modules, which operate on the temporal and spectral dimensions, respectively. Within each self-attention module, a position encoding is first concatenated to the inputs, then the dot-product computation is performed, as: $Output = Softmax(Q \cdot K^T) \cdot V$ [17]. Both of the self-attention modules produce outputs of the same size as the input. Each of the original inputs is added to

Table 1: Summary of Datasets

	RWT	BCI III
Recording Signal	EEG	EEG
Category Of Task	Cognitive	Motor imagery
Number of Subjects	14	1
Recording Device	Muse Headband	Biosemi System
Artifact Removal	Muscle Movement	None
Signal Feature	BP	PSD
Data Size For Each Subject	10528×96	2022×20
Recording Length for each Input	500 ms	500ms

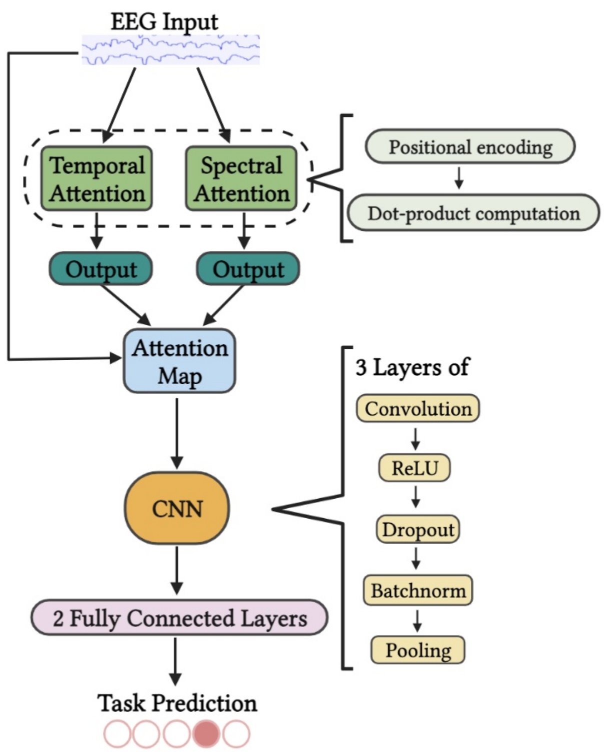


Fig. 1: Model architecture. Every EEG input is fed into the Temporal and Spectral Attention modules. The two outputs of the self-attention modules are added to the original input with weights and produce an attention map, which is then sent into the three-layers CNN. Two fully connected layers form the end of the CNN and make the task prediction (one of three classes for the BCI III dataset or one of five classes for the RWT dataset).

its corresponding two attention outputs with weights to produce an attention maps as a residual connection [12]. The attention maps are then fed into the CNN. The CNN has three convolution layers; each layer contains a convolution kernel sequentially followed by a ReLU activation function, a Dropout layer, a Batchnorm layer, and a pooling layer. Two fully connected layers are appended to the end of the convolution layers and form the classifier, where the final predictions are generated. We used the Adam optimizer and Cross-entropy loss for all training and testing. The model structure and training method were the same for both datasets, except for necessary adjustment on the convolution kernel size to accommodate the input sizes. We varied the learning rates between 5e-4 and 5e-5 based on the model sizes. More hyperparameters and training processing can be found in our publicly available code [1].

4 Result

Performance We used prediction accuracy as our performance evaluation metric, as both datasets were well-balanced over classes of tasks. On the BCI III dataset, the chance of randomly guessing the true class was 33%, and our proposed model reached 79% prediction accuracy, which was comparable to the benchmark result (79%) [3]. On the RWT dataset, the random guessing chance was 20%, and our model reached 46%. The difference between the two performances was presumably caused by the different baseline random guessing accuracy as well as the non-stationarity of EEG signals suffered in the cross-subject setting [9]. The original paper collected RWT did not report a benchmark performance on the cross-subject setting. Therefore, we propose the performance we obtained as a benchmark for future comparison. To verify the effectiveness of the self-attention modules, we did ablation experiments on both datasets, where we compared the performance of our full model (i.e., CNN + Temporal Attention + Spectral Attention) to those of CNN alone, Temporal Attention alone, Spectral Attention alone, Temporal Attention combining CNN, and lastly Spectral Attention combining CNN. Note that the self-attention modules in the original model generated outputs of the same size as the inputs. For the purpose of this experiment, we appended two fully connected layers to both of them to make them full classification models.

Table 2 summarizes the results of the ablation experiment. It shows that every component of the model helped boost the performance. Moreover, we can observe several general patterns on both datasets: 1. CNN alone performed better than either self-attention modules alone (except for the Spectral Attention on RWT), despite that the CNN had fewer parameters than either of the self-attention modules alone, proving the power of CNN; 2. the Spectral Attention performed better than the Temporal Attention, given the inputs from both datasets carried more information on the spectral dimension; 3. both self-attention modules and CNN gained in performance when being concatenated to each other. The combination of CNN with either self-attention module beat CNN alone or self-attention modules alone. The performance was further boosted in the full model.

Furthermore, we also compared the performance of our model to that of a three-layers CNN built on the state-of-the-art Convolutional Block Attention Module (CBAM)

[1] https://github.com/longyi1207/Attention_Based_CNN/blob/main/Attention_Based_CNN.py

Table 2: Ablation Experiment

	Accuracy on RWT (%)	Accuracy on BCI III (%)
Full Model	46	79
CNN	36	73
Temporal Attention	33	69
Spectral Attention	38	70
Temporal Attention + CNN	39	75
Spectral Attention + CNN	41	77

[29]. Our proposed model had slightly more parameters, but the two models achieved comparable accuracy. Refer to Table 3 for the concrete values.

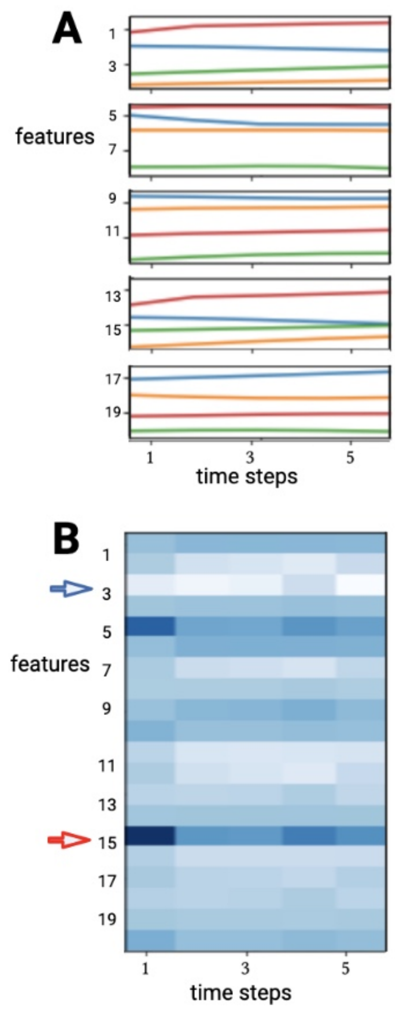
Table 3: Comparison with CBAM CNN

	Our Model	CBAM-based CNN
Performance on RWT (%)	46	46
Number of Parameters on RWT	5155	2532
Performance on BCI III (%)	79	78
Number of Parameters on BCI III	63011	53658

In sum, the two experiments above demonstrate the effectiveness and power of our model. We provided a lightweight implementation of this model structure, and it can be incorporated or upgraded with other techniques to achieve higher performance.

Interpretability In this section, we extracted the two self-attention modules from the original model, visualized their outputs on instance data, and offered insights into their model operation.

Fig. 2A is a visualization of an instance input from RWT whose size is 5 by 20. The twenty colored lines represent the 20 features, and the x-axis indicates the five time steps. For visual clarity, the concrete values of the 20 features are not provided, and the 20 lines are partitioned into five subplots and plotted in four colors corresponding to the five brainwaves and four data acquisition nodes. Panel B shows the attention map for this instance input. We illustrated it as a heatmap and we used a blue and white color palette with blue indicating a larger value. The attention map has the same size as the input. We located where the most and least attention were paid to by the self-attention modules, mapped them on the input, and attempted to interpret the model behavior. Visually inspecting Panel B, there is a horizontal blue band on feature 15 across the five time steps (indicated by the red arrow), indicating that it was assigned the largest weight, while feature 3 (indicated by the blue arrow) was assigned the least. In other words, the model determined that feature 15 provided the most reliable information for class prediction, while feature 3 provided the least. The ground truth label for this



C

Task	Average	95% CI
I	0.90	0.05
II	0.41	0.07
III	0.27	0.10
IV	0.11	0.05
V	**-0.19**	0.06

D

Feature	Average	95% CI
1	0.55	0.04
2	0.53	0.03
3	0.08	**0.47**
4	-0.07	0.04
5	0.05	0.08
6	0.22	0.05
7	-0.17	0.03
8	0.26	0.07
9	0.47	0.03
10	0.50	0.05
11	0.01	0.10
12	0.46	0.05
13	0.73	0.06
14	0.69	0.07
15	0.47	0.09
16	0.15	0.16
17	0.51	0.05
18	0.48	0.09
19	0.33	0.04
20	-0.14	0.14

Fig. 2: **A.** An input from RWT whose label is task V. **B.** The corresponding attention map. Features 3 and 15 are assigned the largest and least weights, indicated by the blue and red arrows. **C.** All tasks' feature 15. Task V has a negative value; it does not overlap the 95% confidence interval (CI) of the four other tasks. **D.** Task V's all features. Feature 3 has the largest CI; it overlaps the CI of most of the other features.

instance input was the fifth task. We compared value of feature 15 of the fifth task with those of the four other tasks, on the premise that the self-attention modules learned to classify tasks by focusing on features that differentiated the five tasks. We will denote the fifth task as Task V in Roman numerals later, likewise for the other four tasks.

Panel C shows the average values as well as 95% confidence intervals (CI) of the five tasks in the RWT training set. We state that the reliance on feature 15 was reasonable since it was negative only for Task V and was non-negative for every other task (circled in red); the CI of task V does not overlap those of the other tasks.

To justify the least weight given to feature 3, we computed the average voltage and CI for all 20 features of Task V. We can observe in Panel D that feature 3 has the largest confidence interval among all the 20 features and therefore the least indicative power (circled in blue), explaining the least reliance given by the model.

Moreover, we observed that despite most inputs were temporally steady, the Temporal Attention module was able to catch distinguishable local voltage changes. Fig.3 shows another instance data from RWT (Panel A) and its corresponding attention map (Panel B). Pinpointed by the two red arrows on the input and attention map, we can see that time steps 2 and 3 were assigned the heaviest weight by the self-attention modules, when the EEG signal fluctuated distinctly. Summarizing the two cases above, we state that the model behavior can be plausibly justified in hindsight. Specifically, the Spectral and Temporal Attention modules were able to learn the average voltage and distinct temporal variations, respectively.

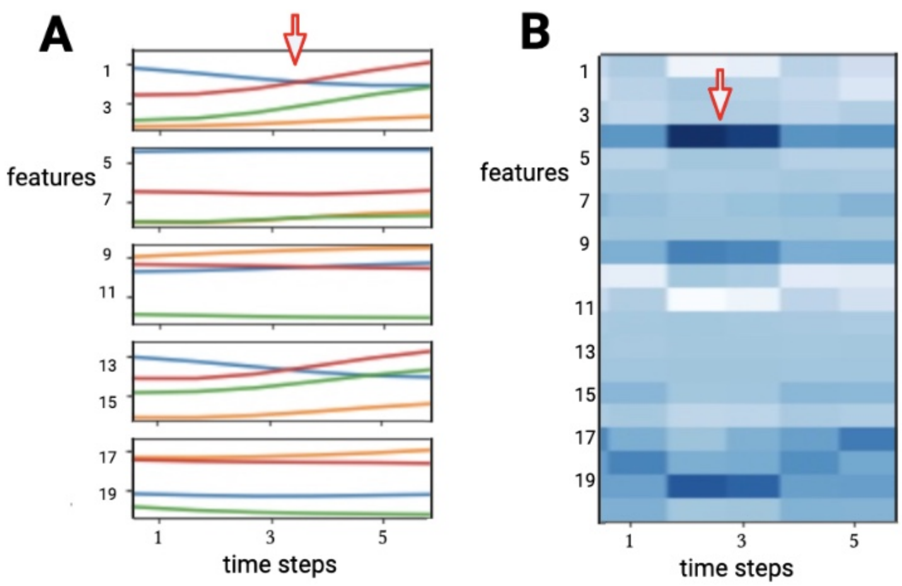


Fig. 3: **A.** An input from RWT. Voltage fluctuates distinctly at time steps 2 and 3. **B.** The corresponding attention map. Time steps 2 and 3 are assigned the largest weights.

5 Discussion

In this work, we implemented a lightweight attention-based CNN model and we experimented with the model on two datasets of different conditions. The model performed well on both datasets. Experiments on more datasets of other settings can be done in the future to test the robustness of this model. For example, both datasets we used are EEG recordings, which have low signal-to-noise ratios (SNR) and also limited dimensionality; CNN is generally powerful at handling data with these qualities. We would expect a larger advantage of the self-attention modules in our model on other types of recordings that have larger SNR and higher dimensionality, such as functional magnetic resonance imaging recordings. Incidentally, we aligned with previous study and used 500-milliseconds-recordings as model inputs[3], but long-range temporal correlations have been found in EEG recordings of MI tasks [27]. CNN is structurally enforced to extract local features and thus weak in capturing long-distance dependency, whereas the attention mechanism was invented to solve this problem. Therefore, we would also expect an advantage in performance if we lengthen the data inputs.

In addition to achieving satisfactory performance, we visualized our self-attention modules' outputs and gained insight into distinguishable neural characteristics underlying different cognitive tasks by interpreting the operation inside the two modules. However, we could only identify simple features such as average voltage of features and instant temporal oscillation and we were unable to recognize more complex neurobiological patterns. Moreover, we did not have prior knowledge of the neural patterns underlying the different MI and cognitive tasks in the datasets we used, namely what really differs in the EEG signals of different tasks, and thus we do not have an objective metric to evaluate how well the self-attention modules perform. One solution to this conundrum is to apply this model on types of EEG recordings that are well-studied and more recognizable by human experts (such as on epilepsy with its characteristic wave discharge [23]) and inspect model operation based on human knowledge.

Zhang et al. [31] proposed that CNN can work in complement with the attention mechanism and this model architecture would conceptually suffice a general BMI framework. However, there is currently no experimental framework that assesses the degree of compatibility of two model structures. Moreover, after researchers used attention models to achieve competitive performance on computer vision tasks, there were questions on the necessity of CNN or even arguments that CNN can be replaced by attention models based on experimental and theoretical evidence [8,10]. The nature of various model structures are yet to be determined, and deeper understandings into the model operation are required.

As the last word, BMI has benefited greatly from machine learning techniques, and researchers have developed advanced BMI applications in an engineering approach in recent years, such as the one described at the beginning of this paper. However, the future development of BMI is indispensable of and also currently hindered by the insufficiency of knowledge and understanding in neuroscience. In fact, the invention of CNN, the attention mechanism, and even the neural network are inspired by understanding of human organism or behavior. Fortunately, researchers can also and have been using computational models as a tool to gain more understanding in various neurobiological subjects, by interpreting the structure and parameters of models that fit well on recording data. For example, neuroscientists proved the existence of nonlinear computation in retinal bipolar cells in a recent study by using a linear-nonlinear model to prove that the previously assumed linear integration can not accurately define the intracellular recordings [22]. Equipped with stronger and stronger deep learning techniques and neural networks, we would like to encourage researchers to pursue interpretable as well as high-performance results in the future.

6 Conclusion

In this research, we developed a model combining two self-attention modules with a CNN, where the self-attention modules focus on the important features and the CNN captures them. This model achieves satisfactory prediction accuracy on two EEG datasets. We verified the effectiveness of our model structure by an ablation experiment; moreover, our model's performance is comparable with that of a CBAM-based CNN. Lastly, we visualized the self-attention modules' outputs as well as the attention maps and made attempts to interpret them. We demonstrated that the self-attention modules were able to learn meaningful features of the EEG signal, and this helped us to gain insight into the data, such as the relative importance of the features and time steps. This model can be future experimented on epilepsy EEG recordings and more datasets to assess its generalizing ability and practical values.

References

1. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate (2014), http://arxiv.org/abs/1409.0473, cite arxiv:1409.0473Comment: Accepted at ICLR 2015 as oral presentation
2. Bang, J.S., Lee, S.W.: Interpretable convolutional neural networks for subject-independent motor imagery classification (2021)
3. Blankertz, B., Müller, K.R., Krusienski, D.J., Schalk, G., Wolpaw, J.R., Schlögl, A., Pfurtscheller, G., Millán, J.R., Schröder, M., Birbaumer, N.: The bci competition. iii: Validating alternative approaches to actual bci problems. IEEE Trans Neural Syst Rehabil Eng **14**(2), 153–159 (Jun 2006). https://doi.org/10.1109/TNSRE.2006.875642
4. Chaudhari, S., Mithal, V., Polatkan, G., Ramanath, R.: An attentive survey of attention models (2021)
5. Chen, K., Wang, J., Chen, L.C., Gao, H., Xu, W., Nevatia, R.: Abc-cnn: An attention based convolutional neural network for visual question answering (11 2015)
6. Cheng, J., Dong, L., Lapata, M.: Long short-term memory-networks for machine reading (2016)
7. Cisotto, G., Zanga, A., Chlebus, J., Zoppis, I., Manzoni, S., Markowska-Kaczmar, U.: Comparison of attention-based deep learning models for eeg classification (2020)
8. Cordonnier, J.B., Loukas, A., Jaggi, M.: On the relationship between self-attention and convolutional layers (2020)
9. Craik, A., He, Y., Contreras-Vidal, J.L.: Deep learning for electroencephalogram (eeg) classification tasks: a review. Journal of neural engineering **16**(3), 031001 (2019)
10. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale (2021)

11. Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. MIT press (2016)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition (2015)
13. Lee, Y.E., Lee, S.H.: Eeg-transformer: Self-attention from transformer architecture for decoding eeg of imagined speech (2021)
14. Liu, X., Shen, Y., Liu, J., Yang, J., Xiong, P., Lin, F.: Parallel spatial–temporal self-attention cnn-based motor imagery classification for bci. Frontiers in Neuroscience **14** (2020). https://doi.org/10.3389/fnins.2020.587520, https://www.frontiersin.org/article/10.3389/fnins.2020.587520
15. Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., Yger, F.: A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update. Journal of neural engineering **15**(3), 031005 (2018)
16. Lotte, F., Lécuyer, A., Guan, C.: Towards a Fully Interpretable EEG-based BCI System (Jul 2010), https://hal.inria.fr/inria-00504658, working paper or preprint
17. Luong, M.T., Pham, H., Manning, C.D.: Effective approaches to attention-based neural machine translation (2015)
18. Qu, X., Hall, M., Sun, Y., Sekuler, R., Hickey, T.J.: A personalized reading coach using wearable eeg sensors-a pilot study of brainwave learning analytics. pp. 501–507 (2018)
19. Qu, X., Liu, P., Li, Z., Hickey, T.: Multi-class time continuity voting for eeg classification. In: International Conference on Brain Function Assessment in Learning. pp. 24–33. Springer (2020)
20. Qu, X., Mei, Q., Liu, P., Hickey, T.: Using eeg to distinguish between writing and typing for the same cognitive task. In: International Conference on Brain Function Assessment in Learning. pp. 66–74. Springer (2020)
21. Qu, X., Sun, Y., Sekuler, R., Hickey, T.: Eeg markers of stem learning pp. 1–9 (2018). https://doi.org/10.1109/FIE.2018.8659031
22. Schreyer, H.M., Gollisch, T.: Nonlinear spatial integration in retinal bipolar cells shapes the encoding of artificial and natural stimuli. Neuron **109**(10), 1692–1706 (May 2021). https://doi.org/10.1016/j.neuron.2021.03.015
23. Smith, S.J.M.: Eeg in the diagnosis, classification, and management of patients with epilepsy. Journal of Neurology, Neurosurgery & Psychiatry **76**(suppl 2), ii2–ii7 (2005). https://doi.org/10.1136/jnnp.2005.069245, https://jnnp.bmj.com/content/76/suppl_2/ii2
24. Sturm, I., Bach, S., Samek, W., Müller, K.R.: Interpretable deep neural networks for single-trial eeg classification (2016)
25. Sun, J., Xie, J., Zhou, H.: Eeg classification with transformer-based models. 2021 IEEE 3rd Global Conference on Life Sciences and Technologies (2021)
26. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L.u., Polosukhin, I.: Attention is all you need **30** (2017), https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
27. Wairagkar, M., Hayashi, Y., Nasuto, S.J.: Dynamics of long-range temporal correlations in broadband eeg during different motor execution and imagery tasks. Frontiers in Neuroscience **15** (2021). https://doi.org/10.3389/fnins.2021.660032, https://www.frontiersin.org/article/10.3389/fnins.2021.660032
28. Willett, F.R., Avansino, D.T., Hochberg, L.R., Henderson, J.M., Shenoy, K.V.: High-performance brain-to-text communication via imagined handwriting. Nature **593**, 249–254 (May 2021)
29. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: Convolutional block attention module (2018)
30. Yin, W., Schütze, H., Xiang, B., Zhou, B.: Abcnn: Attention-based convolutional neural network for modeling sentence pairs (2018)
31. Zhang, X., Yao, L., Wang, X., Monaghan, J., McAlpine, D., Zhang, Y.: A survey on deep learning-based non- invasive brain signals: Recent advances and new frontiers. Journal of Neural Engineering (March 2021)