Proceedings of the
9th Graz Brain-Computer Interface Conference 2024
10.3217/978-3-99161-014-4-010

DEEP LEARNING FOR MOTOR IMAGERY-BASED BCIS USING SEEG SIGNALS

Zhichun Fu[1], Xiaolong Wu[1], Xin Gao[1], Dingguo Zhang[1]

[1] Department of Electronic and Electrical Engineering, University of Bath, Bath, United Kingdom

E-mail: d.zhang@bath.ac.uk

ABSTRACT: Motor imagery (MI) is the most popular paradigm for brain-computer interfaces (BCIs) based on scalp electroencephalography (EEG), while this paradigm is missing for stereo-electroencephalography (sEEG)-based BCIs. Recently, the first public dataset of sEEG has become available for MI-based BCIs. However, the performance using traditional methods is still inferior. In this study, we employed some state-of-the-art methods based on deep learning to improve the classification accuracy of MI for sEEG-based BCIs. Six different deep learning models were developed, which include Shallow ConvNet, DeepNet, ResNet20, conformer, vision transformer (ViT) and ViT with pre-trained parameters. Among six deep learning models, we achieved an average accuracy of 0.83 in the hand open/closed binary classification task with the conformer model. Compared to the available work, our approach demonstrated a remarkable 16% increase in accuracy.

INTRODUCTION

Brain-computer interface (BCI) technology serves as a promising solution, enabling direct communication between the human brain and external devices or computer systems. In general, there are two categories of BCI technology, i.e., invasive and non-invasive. Non-invasive BCI relies on capturing brain signals from the scalp in a user-friendly way. Invasive BCI, on the other hand, involves direct implantation into the brain for signal acquisition, which can result in intracranial signals with less noise interference and a higher spatial-temporal resolution. Examples of invasive BCIs include signals such as stereo-electroencephalography (sEEG) and electrocorticography (ECoG) [1], [2].

Current sEEG-based BCIs primarily focus on motor-related decoding, such as various hand gestures, tongue movements, and foot movements [3], [4]. Combrisson et al demonstrated that motor execution, intention movement and rest status can be differentiated by decoding sEEG signals [5]. The authors discovered a relationship between phase, amplitude and PAC during the planning and execution phases of the goal-directed movement. Additionally, they were able to predict continuously changing grasp force through decoding sEEG signals [6].

However, there have been relatively few studies on sEEG decoding of imagined movements using sEEG. Murphy et al. employed a Support Vector Machine (SVM) to classify the imagined force and rest status in two different grasp configurations, achieving an average accuracy of over 0.6, which was higher than the chance level [7]. When analysing imagined single feature modulation, the alpha band showed a higher modulation level compared to other bands. Ottenhoff et al. demonstrated that non-motor areas contain sufficient information for motor decoding [8]. To avoid the effect of the motor area, they excluded all electrodes originating from the central sulcus and its adjacent area. They used a Riemannian decoder as the classifier, which achieved an average area under curve of 0.68 for imagined movements, with details extracted from the beta band. Individuals with movement disorders or speech impairments often rely on imagery movement as a means of communication. This work aims to enhance the accuracy of sEEG imagined movement decoding by using a deep learning model.

Considering the capability of deep learning models to extract sophisticated features without manual feature extraction, we propose using the same to decode imagined movements. Furthermore, after the advent of the Transformer model, it was demonstrated to be highly effective in sequence-to-sequence tasks due to its attention mechanism. Recent research has shown promising results for deep learning models based on Transformers in reconstructing trajectories of imagined movement [9]. Therefore, the purpose of this study is to evaluate whether a deep learning model can enhance BCI performance for each participant. By utilizing algorithms that have previously been successful in executed movement decoding and regression tasks, they can improve the classification and recognition of imaginary motions with some optimizations. By comparing six different deep learning models with different structures, we have identified a more suitable structure for recognizing imaginary movements which will be valuable for future studies. In summary, our main contributions can be outlined as follows:

1) We explore the application of deep learning methods on sEEG motor imagery datasets.
2) We demonstrate improvements in recognition results compared to previous studies.

The remainder of the article is organized as follows. In the Methods section, we introduce the deep learning models utilized in this study, along with details regarding the dataset and data preprocessing methods. The Result section presents the experimental findings of the models. Finally, we discuss the outcomes and summarize the

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-010

contributions of this article.

METHODS

To assess the performance of deep learning on MI-based BCIs using sEEG, this study implemented five different state-of-the-art (SOTA) models and an improved model with pre-trained weights. The results were then compared with those obtained from the Riemannian classifier in reference [8].

Shallow ConvNet

The ConvNet shown in Fig. 1 is a model that utilises temporal convolution and spatial filtering in its initial layers, similar to the bandpass phase in the filter bank common spatial pattern (FBCSP) [10]. The shallow ConvNet's use of a larger kernel size in temporal convolution allows for a broader range of transformations. Additionally, incorporating multiple pooling regions per trial enables the learning of the temporal structure of band power changes, thereby enhancing classification.

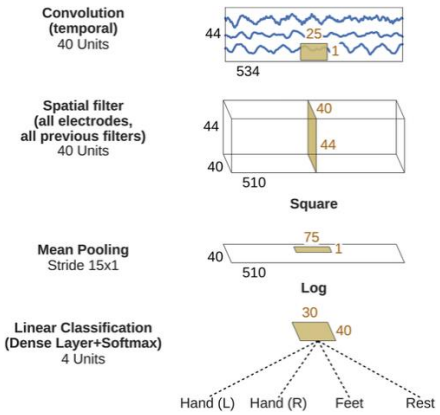

Figure 1: The model structure for shallow ConvNet [10].

DeepNet

The deepNet model utilized in this study featured a more intricate architecture with a substantial increase in the number of layers compared to the shallow model [10]. The architecture includes temporal convolution, spatial convolution, a fully connected layer and basic blocks which are used to extract the spatial features. Fig. 2 shows the model structure with one basic block. A dropout rate of 0.5 was employed to improve the model's robustness. The process of optimization involves experimenting with different quantities of basic blocks to determine the optimal configuration of the model. For this work, we utilized the deepNet model with two basic blocks.

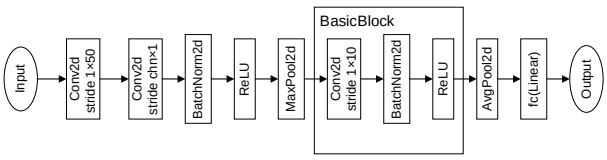

Figure 2: The model structure for DeepNet model with changing depth.

ResNet20

ResNet architectures have demonstrated success in various computer vision tasks due to their ability to mitigate the vanishing gradient problem and facilitate the training of exceptionally deep networks [11]. The ResNet model, with its residual connections, aims to leverage these advantages to enhance the performance of the imagining motion task. As the ResNet model has been previously used for emotion classification based on EEG image recognition, we incorporated a 20-layer Residual Network (ResNet) into our model, as shown in Fig. 3 [12].

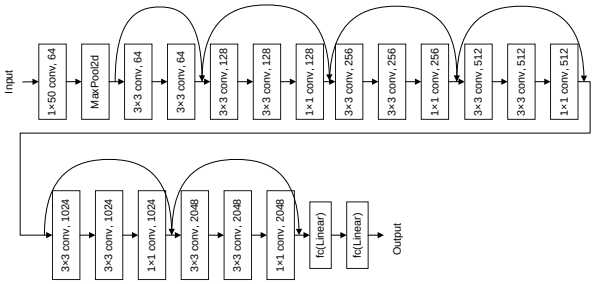

Figure 3: The model structure for Resnet20.

Conformer

The Conformer model comprises three modules: a convolution module, a self-attention module, and a classifier module, as shown in Fig. 4 [13]. The convolution module uses both spatial and temporal convolutions to capture local spatial and temporal features of EEG signals. This is followed by an average pooling layer to reduce feature dimension and mitigate noise interference. The self-attention module utilizes multi-head attention mechanisms to capture global temporal dependencies of EEG features, complementing the local features learned by the convolution module. The classifier module includes two fully connected layers to output the probability of different EEG categories, such as motor imagery or emotion recognition tasks.

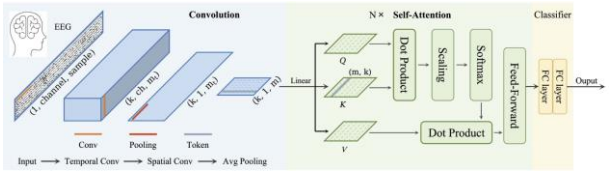

Figure 4: The model structure for conformer [13].

ViT (Vision Transformer)

ViT is a hybrid model that combines a two-step convolution block with a transformer block, depicted in Fig. 5 [14]. The two-step convolution block is composed of two convolutional layers, one for the temporal dimension and one for the channel (spatial) dimension. This block generates patch embeddings that capture the frequency and spatial information of the sEEG data. The transformer block utilises the ViT architecture, which divides the input into patches and processes them as a sequence using self-attention and multi-layer perceptron. Additionally, it also captures global dependencies and patterns in the patch sequence. The final representation of the input is the hidden state of a special token.

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-010

The transformer block was pre-trained on the ImageNet dataset, which contains millions of natural images. This pre-training allows the model to utilise prior knowledge learned from image data and transfer it to the BCI field.

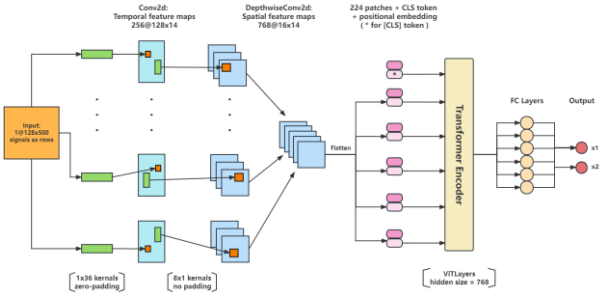

Figure 5: The model structure for ViT [14].

Dataset

The sEEG dataset was collected from eight subjects performing two motor imagined tasks [8]. Specifically, subjects were instructed to imagine opening and closing their left and right hands, with each action lasting for 3 seconds. Tab. 1 provides additional details about the dataset, including the distribution of contacts for each subject, as well as the number of electrodes in the left and right hemispheres and the presence of electrodes in the motor area. Each participant imagined 30 trials of opening and closed actions for each of their hand, with each action lasting 3 seconds. Consequently, each participant generated a total of 60 movements and 60 rest events throughout the entire experiment.

Table 1: Electrode details of subjects after removing noisy and abnormal signals. Number of contacts also includes the electrodes which are located in the motor areas.

Patient ID	Motor Left	Motor Right	Contacts Left	Contacts Right
1	4	6	37	90
2	0	0	103	24
3	9	0	66	0
4	0	0	54	0
5	6	0	117	0
6	0	0	63	63
7	3	5	67	60
8	0	3	40	75

Data processing

For each subject, any abnormal signals, including flat signals and signals with abnormal amplitudes, were removed. In brief, the logarithm of the root mean square (LRMS) of each channel's signal was calculated. Then, we normalized these LRMS values, and calculated the corresponding p-values based on a normal distribution assumption. Channel with p-values less than or equal to 0.05 were flagged as having abnormal amplitudes. For the channel where 50Hz frequency band power exceeded two times the interquartile range of the signal, it would be removed as well. The remaining signals underwent detrending, mean removal, and were subjected to a notch filter at 50Hz, 100Hz, 150Hz, and 200Hz to minimize interference from noise.

Subsequently, the data was downsampled to 500Hz, and each experiment was segmented from -0.5s to 3s. The entire dataset is segmented by selecting fixed window size and stride size and stored as DataLoader formatted data for subsequent input into deep models for training and evaluation. By utilizing a fixed stride size, the optimal window size was identified among 200, 400, 600, 800, 1000 and 1200. With the best window size, best stride size can be found among 20, 50, 100, 200, 300, 400, 500. Based on the performance of all models, 800 and 400 are selected as the window size and stride size respectively. In this work, a learning rate of 0.0001 was employed with a weight decay set to 0.0005. The optimizer used was Adadelta, and the cross-entropy loss was used for loss function calculation. To prevent data leakage from affecting model training and prediction, trials were classified before splitting. 60% of the trials were assigned to the training set, 20% to the validation set, and 20% to the test set. After determining these sets, each trial was further divided into one-second overlapping intervals to simulate data obtained during online experiments. The desktop computer that was used in the tests has the following configuration: 11th Gen Intel i9-11900 16 core CPU, 64 GB of RAM and a NVidia RTX 3080 GPU.

RESULT

The performance evaluation of six deep learning models was conducted to investigate their effectiveness in the task of imagining motion. Tab. 2 summarizes the performance metrics for each model, including the performance of the original Riemannian decoder. Our results indicate that shallow ConvNet and deep models have a relatively lower performance in imagining motion, with average test accuracy of 0.52 and 0.62, respectively, slightly above the chance level (0.5). Among the other deep learning models, ResNet, Conformer and ViT achieved performance levels of 0.75, 0.83 and 0.71, respectively, demonstrating superior effectiveness in the task. The ViT model with pre-trained parameters achieved an accuracy of 0.76, higher than the ViT model without pre-trained.

Given that the experiment involves movements of both hands, the binary classification only focuses on distinguishing hand movements, neglecting the distinction between left and right hands. Hence, Fig. 6 presents the classification performance of four gestures across six models, considering both the left- and right-hand movements. Additionally, Fig. 7 illustrates the classification performance of the Conformer model on the dataset from Subject 8 (S8).

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-010

Table 2: Comparison of classification accuracy results among 6 deep learning mode. The chance level for the classification is 0.5. ViT_p model refers to ViT model with pretrained data.

Subject	Shallow ConvNet	DeepNet	ResNet20	Conformer	ViT	ViT_p	Ref [8]
S1	0.45	0.55	0.7	1.0	0.8	0.8	0.82
S2	0.4	0.5	0.7	0.8	0.5	0.8	0.7
S3	0.7	0.7	0.8	0.7	0.7	0.7	0.64
S4	0.35	0.65	0.8	0.9	0.6	0.8	0.64
S5	0.4	0.6	0.8	0.8	0.8	0.7	0.58
S6	0.65	0.65	0.8	0.7	0.7	0.8	0.65
S7	0.7	0.6	0.7	0.9	0.7	0.8	0.61
S8	0.5	0.7	0.7	0.8	0.9	0.7	0.7
Avg	0.52	0.62	0.75	0.83	0.71	0.76	0.67

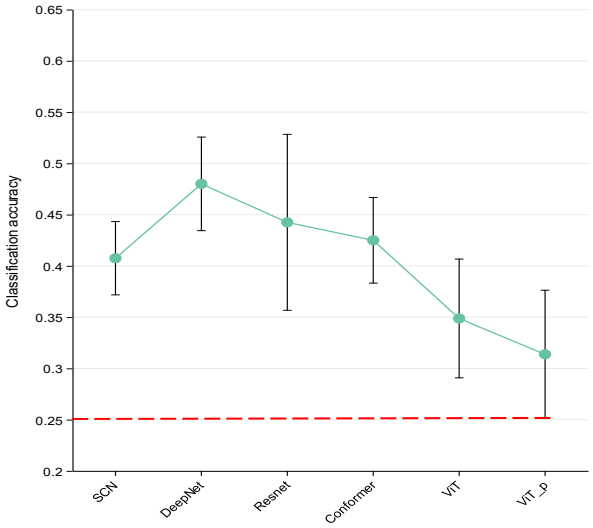

Figure 6: Comparison of 4-gestures classification accuracy results among 6 deep learning mode with red dot line represents for the chance level of 0.25.

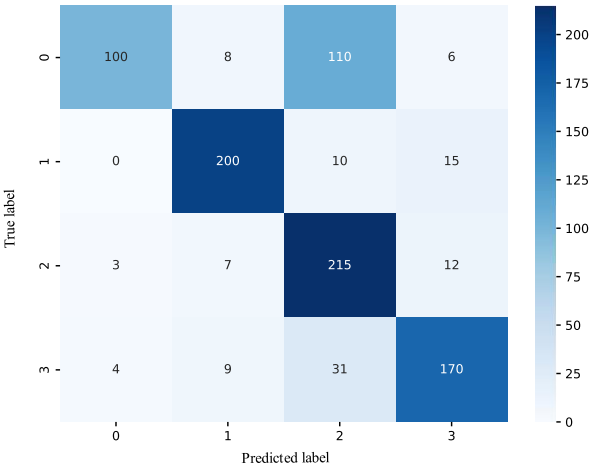

Figure 7: Confusion matrix for 4 gestures classification on S8 with conformer model. Label 0 and label 1 represents the close and open status for left hand respectively. Label 2 and label 3 represents the close and open status for right hand respectively.

DISCUSSION

For the 2-gesture classification, it suggests the need for sophisticated feature extraction capabilities, which transformer architectures appear to provide, especially when no electrode is located in the motor area. And the result from the ViT with pre-trained data suggests that leveraging pre-training significantly enhances the ViT model's performance in the imagining motion task. Therefore, these results have displayed potential advantages of deep learning models in the imagery motion task. By comparing our findings to previous results, it can be suggested that given limited sEEG dataset, not only the model expressiveness can be improved through data augmentation on dataset itself but also through pretraining on other datasets, such as ImageNet for an EEG regression task [15]. This approach proves effective in improving the classification performance of MI tasks. While transformer-based models may not perform as well as models relying solely on convolutional modules for four-class classification, this can possibly be explained by the electrode distribution. For example, some participants only have electrode implantation on single side of the brain and only some have a limited number of electrodes presented in the motor cortex. Due to contralateral control, the left-hand movements are dominated by the right hemisphere of the brain. As shown in Fig. 7, more than half of the left-hand closing gestures are incorrectly recognized as right-hand closing gestures.

While our study makes valuable contributions, it is essential to acknowledge certain limitations. Due to the limited availability of public sEEG motor imagery datasets, our research focused solely on evaluating the model's performance in classifying two types of gestures within a single dataset. Future investigations should aim for a more comprehensive exploration of task specificity, dataset characteristics, and the impact of model hyperparameters on the ultimate performance. This would allow the exploration of various deep learning architectures, particularly the advantages and limitations of transformer-based models in motor imagery tasks. However, with the varied performance of different

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-010

models, especially the enhanced accuracy with pre-trained ViT, we provide more opportunities for further explorations and optimisation.

CONCLUSION

In this work, we evaluate the performance of sEEG-based imagery motion classification by using multiple deep learning models. By comparing six different deep learning models, we used the conformer model to achieve an accuracy of 0.83 in the binary classification of imagined movements, which is 0.16 higher than the performance of the previous work. This work provides a reference for using deep learning models in BCI imagery movements with sEEG signals.

ACKNOWLEDGEMENTS

This work is supported by the EPSRC New Horizons Grant of UK (EP/X018342/1).

REFERENCES

[1] Z. Xie, O. Schwartz, and A. Prasad, ‘Decoding of finger trajectory from ECoG using deep learning’, J. Neural Eng., vol. 15, no. 3, p. 036009, Feb. 2018, doi: 10.1088/1741-2552/aa9dbe.

[2] P. Z. Soroush, C. Herff, S. K. Ries, J. J. Shih, T. Schultz, and D. J. Krusienski, ‘The nested hierarchy of overt, mouthed, and imagined speech activity evident in intracranial recordings’, NeuroImage, vol. 269, p. 119913, Apr. 2023, doi: 10.1016/j.neuroimage.2023.119913.

[3] G. Li et al., ‘Assessing differential representation of hand movements in multiple domains using stereo-electroencephalographic recordings’, NeuroImage, vol. 250, p. 118969, Apr. 2022, doi: 10.1016/j.neuroimage.2022.118969.

[4] M. A. Jensen et al., ‘A motor association area in the depths of the central sulcus’, Nat. Neurosci., vol. 26, no. 7, Art. no. 7, Jul. 2023, doi: 10.1038/s41593-023-01346-z.

[5] E. Combrisson et al., ‘From intentions to actions: Neural oscillations encode motor processes through phase, amplitude and phase-amplitude coupling’, NeuroImage, vol. 147, pp. 473–487, Feb. 2017, doi: 10.1016/j.neuroimage.2016.11.042.

[6] X. Wu et al., ‘Decoding continuous kinetic information of grasp from stereo-electroencephalographic (SEEG) recordings’, J. Neural Eng., vol. 19, no. 2, p. 026047, Apr. 2022, doi: 10.1088/1741-2552/ac65b1.

[7] B. A. Murphy, J. P. Miller, K. Gunalan, and A. B. Ajiboye, ‘Contributions of Subsurface Cortical Modulations to Discrimination of Executed and Imagined Grasp Forces through Stereoelectroencephalography’, PLOS ONE, vol. 11, no. 3, p. e0150359, Mar. 2016, doi: 10.1371/journal.pone.0150359.

[8] M. C. Ottenhoff et al., ‘Decoding executed and imagined grasping movements from distributed non-motor brain areas using a Riemannian decoder’, Front. Neurosci., vol. 17, 2023, Accessed: Nov. 25, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2023.1283491

[9] P. Wang, P. Gong, Y. Zhou, X. Wen, and D. Zhang, ‘Decoding the Continuous Motion Imagery Trajectories of Upper Limb Skeleton Points for EEG-Based Brain–Computer Interface’, IEEE Trans. Instrum. Meas., vol. 72, pp. 1–12, 2023, doi: 10.1109/TIM.2022.3224991.

[10] R. T. Schirrmeister et al., ‘Deep learning with convolutional neural networks for EEG decoding and visualization’, Hum. Brain Mapp., vol. 38, no. 11, pp. 5391–5420, 2017, doi: 10.1002/hbm.23730.

[11] K. He, X. Zhang, S. Ren, and J. Sun, ‘Deep Residual Learning for Image Recognition’. arXiv, Dec. 10, 2015. Accessed: Dec. 28, 2023. [Online]. Available: http://arxiv.org/abs/1512.03385

[12] T. Tian, L. Wang, M. Luo, Y. Sun, and X. Liu, ‘ResNet-50 based technique for EEG image characterization due to varying environmental stimuli’, Comput. Methods Programs Biomed., vol. 225, p. 107092, Oct. 2022, doi: 10.1016/j.cmpb.2022.107092.

[13] Y. Song, Q. Zheng, B. Liu, and X. Gao, ‘EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization’, IEEE Trans. Neural Syst. Rehabil. Eng., vol. 31, pp. 710–719, 2023, doi: 10.1109/TNSRE.2022.3230250.

[14] R. Yang and E. Modesitt, ‘ViT2EEG: Leveraging Hybrid Pretrained Vision Transformers for EEG Data’. arXiv, Aug. 01, 2023. Accessed: Dec. 08, 2023. [Online]. Available: http://arxiv.org/abs/2308.00454

[15] J. Chen, D. Wang, W. Yi, M. Xu, and X. Tan, ‘Filter bank sinc-convolutional network with channel self-attention for high performance motor imagery decoding’, J. Neural Eng., vol. 20, no. 2, p. 026001, Mar. 2023, doi: 10.1088/1741-2552/acbb2c.

Published by
Verlag der Technischen Universität Graz