Dean&Francis

ISSN 2959-6157

Research on Decoding of Neural Network Algorithm for Upper Limb Rehabilitation Robot Based on Motor Imagery

Zhenkun Tian

Miami College, Henan University, Kaifeng, Henan, 475000, China
tiantian199808@gmail.com

Abstract:

Brain-computer interfaces (BCIs) enable direct communication with computers by encoding and decoding brain electrical signals to generate control signals and interact directly with external devices, thereby assisting patients with damaged motor nerves. Research on brain electrical signals based on motor imagery has always received widespread attention. To effectively extract and classify complicated brain electrical signal features, a system architecture that can withstand a low ratio between signal and noise, instability, and physiological artifacts must be constructed. Advancements in motor imagery electroencephalography have been made possible by the rapid development of deep learning over the past few years. This paper aims to introduce and analyze three decoding models that use convolutional neural network architectures (CNN) and three combinations of CNN and Transformer architectures. Through similar comparisons and cross-comparisons, by analyzing the accuracy of each model on the current mainstream motor imagery-related datasets BCI IV 2A and 2B, the advantages and disadvantages of each model are explored. Through comparison, it can be seen that models based on the CNN architecture still occupy a dominant position due to their fewer parameters and better adaptability to various situations. However, the confusion decoding model CTNet performs very well on the BCI IV 2A and 2B, indicating that the decoding model with the Transformer architecture has higher performance development potential and is the main direction of future development.

Keywords: Brain-Computer Interfaces (BCIs); Motor Imagery; Convolutional Neural Network; Transformer

ISSN 2959-6157

1. Introduction

With the continuous acceleration of the aging process, the number of patients with neurological diseases such as stroke and cerebral palsy is increasing. Among them, 70% of stroke patients have varying degrees of upper limb dysfunction [1]. The number of patients with upper limb dysfunction caused by other neurological diseases or congenital factors is also relatively high. To restore the upper limb function of patients, rehabilitation training is indispensable, and upper limb rehabilitation has become a hot research field [2, 3]. Traditional therapies rely on professional therapists. With the growth of rehabilitation needs, there is a significant shortage of professional rehabilitation personnel. The gradual increase of disabled people and the elderly population, and the problem of many patients and few therapists, has become increasingly prominent. Therefore, using rehabilitation robots to provide auxiliary training for patients, accelerating the rehabilitation process, reducing medical burdens, and optimizing rehabilitation efficacy is more urgent.

Motor Imagery (MI) is the act of imagining specific movements in the brain without actually performing them physically [4]. Electroencephalography (EEG) is a method that doesn't require invasive procedures for recording brain electrical activity, and can capture two-dimensional data of brain electricity on the scalp surface [5]. The establishment of a communication and control channel between the human brain and computers or other electronic devices is called BCI or Brain Computer Interface. EEG is one of the main signal sources of BCI [6]. Motor Imagery Electroencephalogram (MI-EEG) is a type of electroencephalogram that does not require external stimulation and can be self-regulated. It can be detected through electrode channels and is a multi-dimensional long-term sequence point. The low signal-to-noise ratio (SNR), instability, and physiological artifact interference of MI-EEG affect its decoding. Therefore, the current challenge for brain-computer interface technology is to accurately recognize human intentions. Progress in MI-EEG has been promoted by the rapid development of deep learning in recent years. The purpose of this article is to study the research on decoding and control of the current mainstream neural network algorithms.

2. Overview of Relevant Theories

2.1 Decoding Flowchart

Fig. 1 shows the flowchart for decoding the upper limb robot using motion imagination. The decoding part mainly consists of three processing stages. The initial stage of data collection involves recording neural data. Signal processing is the second step, during which recorded data is cleaned and preprocessed. The third stage involves extracting features and classifying neural data, which provides meaningful information and allows for decision-making. The commonly used methods in this step include support vector machines, Bayesian decision, decision trees, random forests, neural networks, and deep learning etc.

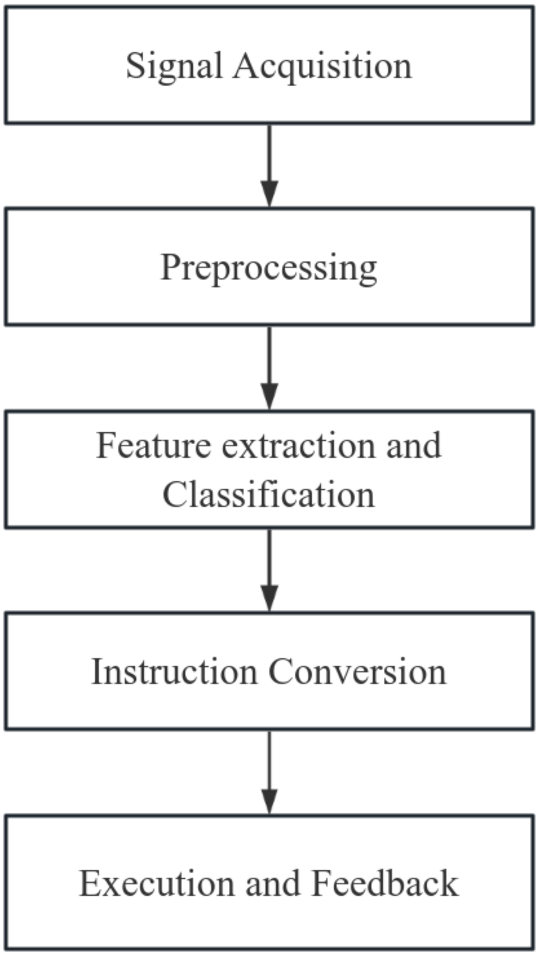

Fig. 1. Decoding Flowchart of the Upper Limb Rehabilitation Robot (Photo/Picture credit: Original).

2.2 Dataset and Paradigm

Currently, brain-computer interfaces are classified into two types based on the characteristics of the electroencephalogram (EEG): event-related type and oscillatory type [7]. The ERP brain-computer interface is designed to identify high-amplitude, low-frequency EEG responses to external stimuli that have a time limit. ERP's time course can be efficiently modeled due to their high stability across different individuals and clear waveforms [8]. Oscillatory brain-computer interface external control is achieved through the signal power of specific EEG frequency bands that are usually asynchronous [9]. Event-related spectral perturbation (ERSP) analysis can be used to represent the oscillation signal that is time-limited to the external stimulus [10]. Occasional brain-computer interface training is more difficult, primarily due to its lower SNR and good differences between individuals [11]. The motor imagery EEG signal is the SMR potential, which is an oscillatory potential. The data of this type used in the several models introduced in this paper comes from the datasets named BCI IV 2A and 2B.

3. Case Analysis

3.1 Application Cases of Convolutional Neural Networks in Brain Electrocode Analysis

3.1.1 Shallow convNet and deep convNet

A CNN can be used to extract local features from images. Such networks usually have a fixed hierarchical structure. Convolutional layers, pooling layers, activation functions, and fully connected layers are all part of a typical CNN. Schirrmeister et al. designed two CNN models, Deep ConvNet and Shallow ConvNet [12]. These two models contain temporal convolutional layers and spatial convolutional layers, and their main purpose is to decode and classify the initial EEG signals on motor imagery. Deep ConvNet features four convolution-max pooling modules, with two dedicated for processing the input electroencephalogram, three standard ones, and an integral softmax classification layer. Shallow ConvNet's first two layers perform temporal convolution and spatial filtering. In contrast to Deep ConvNet, Shallow ConvNet's temporal convolution has a more substantial convolution kernel.

3.1.2 EEGnet

Lawhern et al. designed the EEGnet model [13]. This is a widely used compact convolutional neural network structure. The network first performs frequency filtering through temporal convolution, and then uses deep convolution, which is connected with each feature map respectively, to conduct spatial filtering at specific frequencies. Separable convolution combines deep convolution and pointwise convolution. Each feature map has its deep convolution that generates summaries on the time dimension, while the pointwise convolution learns how to fuse these feature maps optimally.

3.2 Case of the Brain Electrocode Method Combining CNN and Transformer

3.2.1 CTNet for EEG decoding

The Transformer model, with its global perception ability, performs exceptionally well in the fields of natural language and image processing [14]. In recent years, it has also been introduced into the EEG decoding field, achieving certain results by leveraging time dependence. However, this model neglects local feature learning (which is crucial for EEG decoding), and requires additional reliance on spatial filtering and other compensation methods [15, 16]. At the same time, its working principle lacks detailed analysis and visualization. Therefore, Transformer is still in the exploration stage in the EEG field and cannot yet be used as an end-to-end original EEG classification backbone model. Using a convolutional module like EEGNet, CTNet was proposed by Zhao et al to extract local and spatial features from the EEG time series with specific precision [17]. After that, it merges the encoder module called Transformer and utilizes the Mechanism for multi-headed attention to recognize the global dependencies of advanced EEG features. Ultimately, it categorizes EEG signals using a straightforward classifier module that comprises fully interconnected layers.

3.2.2 Conformer for EEG decoding

Song and his associates proposed a convolutional transformer named EEG Conformer that can encapsulate features that are local and global in a single system that categorizes EEG [18]. The convolutional module, the self-attention module, and the classifier are three components that make up the overall framework. Local temporal and spatial features are captured through time and spatial convolution in the convolutional module, and time feature parts are segmented using an average pooling layer. This decreases the model's complexity and eliminates unnecessary information. Next, the self-attention module learns about global temporal dependencies through the self-attention layer after recognizing every point in the time dimension as a token. To obtain the decoding result, a fully connected layer is employed in the final stage.

3.2.3 RCA-Conformer for EEG decoding

Li et al. proposed RCA-Conformer, which combines CNN and Transformer. It extracts local features through multi-scale temporal convolution (MSTCN) and residual channel attention (RCA), and enhances global features by using the multi-head attention mechanism [19, 20]. A convolutional unit, a transformer encoder module, and a classification device are the three components of the model. The convolutional module combines MSTCN, spatial

ISSN 2959-6157

convolution, and RCA modules, respectively extracting multi-scale temporal features, spatial features, and enhancing spatial information selectivity. The Transformer encoder module receives the rearranged convolutional feature maps. The Transformer encoder module captures global context information through a multi-head attention mechanism (MHA) and further processes the features using the feed-forward network (Feed-Forward Network, FFN).

4. Discussion

4.1 Discussion and Analysis of CNN's Decoding Model

The design differences between the two CNN models are shown in Table 1:

Table 1. Design of Two CNN Models

Name	ConvNet	EEGNet
Convolution type	Standard two-dimensional convolution	Depthwise separable convolution
Feature extraction logic	Multi-layer stacked convolution (For example, Deep ConvNet has 5 layers)	Block design: Time filtering → Spatial filtering → Feature fusion
Parameter quantity	Higher	Extremely low

Furthermore, in terms of performance, compared with EEGNet, Deep ConvNet is more dependent on a large amount of data [11]. Moreover, Deep ConvNet requires the adoption of training data augmentation strategies to achieve good classification performance on the SMR dataset. However, EEGNet performs well on all test datasets and does not require data augmentation, which makes this model more user-friendly in practical applications, so there are more improved versions of EEGNet.

The performance of Shallow ConvNet on event-related potential brain-computer interface datasets (ERP) is often inferior to that on oscillatory brain-computer interface datasets (SMR), while Deep ConvNet shows the opposite situation. The performance of EEGNet in SMR classification is comparable to that of Shallow ConvNet, and in ERP classification, it is comparable to that of Deep ConvNet (and performs better in group-level MRCP, ERN, and SMR classification), indicating that EEGNet is more stable and can learn multiple features in various BCI tasks.

4.2 Discussion and Analysis of the Decoding Model Combined with CNN and Transformer

The design differences of the three types of combined architecture models are shown in Table 2 as follows:

Table 2. Design of Models Combining Three Types of CNN and Transformer

Name	CTNet	Conformer	RCA-Conformer
Type	EEGNet+Transformer	ConvNet+Transformer	MSTCN+RCA+Transformer
Record set	BCI IV 2A and 2B	variety	BCI IV 2A and 2B
Parameter quantity	Higher	Higher	Higher

Conformer's design involves combining Shallow ConvNet with the self-attention mechanism. Local features are captured by the self-attention mechanism while the convolutional module extracts global dependencies on them. CTNet combines EEGNet and uses CNN to extract local features and learns global features through the Transformer encoder. RCA-Conformer combines CNN and Transformer, introducing multi-scale temporal convolution and residual channel attention (RCA) to extract local features.

In terms of performance, Conformer can be applied to multiple types of datasets. CTNet and RCA-Conformer are applied in the BCI IV 2A and 2B. Conformer has a wide range of application scenarios, while CTNet and RCA-Conformer have relatively limited application scenarios. In experiments targeting the BCI IV 2A and 2B, the accuracy of CTNet in subject-specific evaluation and cross-subject-specific evaluation is the highest among the three. RCA-Conformer outperforms Conformer in subject-specific evaluation, but Conformer has also been tested in the SEED dataset, achieving an accuracy rate of 95.3%, demonstrating excellent performance.

4.3 Cross-Comparison Analysis

As can be seen from Table 3, in the experiments conducted on specific subjects in the BCI IV 2A and 2B, CTNet achieved extremely high average accuracy rates. The same type of Conformer and RCA-Conformer also demonstrated excellent performance. Compared with the CNN architecture, the architecture combining CNN and Transformer was at the leading level in the experiments on specific subjects.

Table 3. Subject-specifc classifcation accuracy (in percentage %)

Name	BCI IV-2a	BCI IV-2b
Shallow ConvNet	75.69	85.13
Deep ConvNet	77.78	85.21
EEGnet	77.39	87.71
Conformer	77.66	85.87
CTNet	82.52	88.49
RCA-Conformer	80.29	85.74

Table 4 indicates that CTNet had an average classification accuracy of 58.64% in the cross-subject experiments of BCI IV-2a, which was just slightly less than Deep ConvNet. CTNet's average classification accuracy on BCI IV-2b was 76.27%. For the Conformer, which is also a combination of CNN and Transformer architecture, its accuracy in cross-subject experiments was not as high as that of the CNN architecture model.

Table 4. Cross-subject classification accuracy (in percentage %)

Name	BCI IV-2a	BCI IV-2b
Shallow ConvNet	56.75	74.28
Deep ConvNet	60.15	75.18
EEGnet	56.85	75.13
Conformer	53.41	73.52
CTNet	58.64	76.27

5. Conclusion

This article reviews the three mainstream CNN architecture models and the three models combining CNN and Transformer architectures in recent years. Through comparative analysis, it can be seen that the CNN architecture models still hold the dominant position due to their efficiency in extracting local features and their simple structure. The models incorporating the Transformer architecture have their core advantage in that the self-attention mechanism can dynamically capture the long-range dependencies between electrodes and enhance the global representation of complex brain electrical patterns. However, these hybrid models still need to overcome two major challenges: first, their performance highly depends on the optimization of hyperparameters, and the number of Transformer layers and channel dimensions need to be precisely configured for different datasets; second, the cross-subject generalization performance has not yet fully led the way.

In conclusion, the decoding of brain electrical signals still faces many challenges. On one hand, due to the special nature of brain electrical signals, it brings two problems. Firstly, even in the laboratory environment, brain electrical signals are extremely prone to noise interference and have a low SNR. Additionally, as biological signals, the data of brain electrical signals often varies from person to person, and the data distribution has significant differences. On the other hand, unlike traditional image recognition and speech recognition, the acquisition of brain electrical data is more difficult, resulting in limited training data and the need to achieve high-precision classification and recognition from limited data. Therefore, future brain-computer interface decoding technologies need to break through in three major directions: multimodal fusion, adaptive architecture, and lightweight hardware, and also need to promote large-scale applications in medical rehabilitation

ISSN 2959-6157

and education fields.

References

[1] Longhi M, Merlo A, Prati P, Giacobbi M, Mazzoli D. Instrumental indices for upper limb function assessment in stroke patients: a validation study. Journal of NeuroEngineering and Rehabilitation, 2016, 13(1): 52.
[2] Karamians R, Proffitt R, Kline D, Gauthier L V. Effectiveness of virtual reality-and gaming-based interventions for upper extremity rehabilitation poststroke: a meta-analysis. Archives of Physical Medicine and Rehabilitation, 2020, 101(5): 885-896.
[3] Afsar S I, Mirzayev I, Yemisci O U, Saracgil S N C. Virtual reality in upper extremity rehabilitation of stroke patients: a randomized controlled trial. Journal of Stroke and Cerebrovascular Diseases, 2018, 27(12): 3473-3478.
[4] Gao D, Yang W, Li P, Liu S, Liu T, Wang M, Zhang Y. A multiscale feature fusion network based on attention mechanism for motor imagery EEG decoding. Applied Soft Computing, 2024, 151: 111129.
[5] Ling L, Li W, Feng B. A spatiotemporal feature decoding method for motor imagery EEG based on multi-model fusion. Journal of Nanjing University (Natural Sciences), 2024, 60(1): 65-75.
[6] Wang X, Liesaputra V, Liu Z, Wang Y, Huang Z. An in-depth survey on deep learning-based motor imagery electroencephalogram (EEG) classification. Artificial Intelligence in Medicine, 2024, 147: 102738.
[7] Lotte F. Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain-computer interfaces. Proceedings of the IEEE, 2015, 103(6): 871-890.
[8] Fazel-Rezai R, Allison B Z, Guger C, Sellers E W, Kleih S C, Kübler A. P300 brain computer interface: current challenges and emerging trends. Frontiers in Neuroengineering, 2012, 5: 14.
[9] Pfurtscheller G, Neuper C. Motor imagery and direct brain-computer communication. Proceedings of the IEEE, 2001, 89(7): 1123-1134.
[10] Makeig S. Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. Electroencephalography and Clinical Neurophysiology, 1993, 86(4): 283-293.
[11] Pfurtscheller G, Da Silva F L. Event-related EEG/MEG synchronization and desynchronization: basic principles. Clinical Neurophysiology, 1999, 110(11): 1842-1857.
[12] Lawhern V J, Solon A J, Waytowich N R, Gordon S M, Hung C P, Lance B J. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. Journal of Neural Engineering, 2018, 15(5): 056013.
[13] Schirrmeister R T, Springenberg J T, Fiederer L D J, Glasstetter M, Eggensperger K, Tangermann M, Hutter F, Burgard W, Ball T. Deep learning with convolutional neural networks for EEG decoding and visualization. Human Brain Mapping, 2017, 38(11): 5391-5420.
[14] Vaswani A. Attention is all you need. Advances in Neural Information Processing Systems, 2017, 30: 1.
[15] Bagchi S, Bathula D R. EEG-ConvTransformer for single-trial EEG-based visual stimulus classification. Pattern Recognition, 2022, 129: 108757.
[16] Zheng Y, Zhao X, Yao L. Copula-based transformer in EEG to assess visual discomfort induced by stereoscopic 3D. Biomedical Signal Processing and Control, 2022, 77: 103803.
[17] Zhao W, Jiang X, Zhang B, Xiao S, Weng S. CTNet: a convolutional transformer network for EEG-based motor imagery classification. Scientific Reports, 2024, 14(1): 20237.
[18] Song Y, Zheng Q, Liu B, Gao X. EEG conformer: convolutional transformer for EEG decoding and visualization. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023, 31: 710-719.
[19] Li X, Aierken H. EEG decoding research based on CNN and Transformer. Microelectronics and Computer, 2024, 1-12.
[20] Ma X, Chen W, Pei Z, Zhang Y, Chen J. Attention-based convolutional neural network with multi-modal temporal information fusion for motor imagery EEG decoding. Computers in Biology and Medicine, 2024, 175: 108504.