Learning Shared Semantic Information from Multimodal Bio-signals for Brain-Muscle Modulation Analysis

Tian-Yu Xiang
Xiao-Hu Zhou*
Xiao-Liang Xie
State Key Laboratory of Multimodal Artificial Intelligence Systems, CASIA
University of Chinese Academy of Sciences, Beijing, China
xiangtianyu2021@ia.ac.cn
xiaohu.zhou@ia.ac.cn

Shi-Qi Liu
Hong-Jun Yang
Zhen-Qiu Feng
Mei-Jiang Gui
State Key Laboratory of Multimodal Artificial Intelligence Systems, CASIA
University of Chinese Academy of Sciences, Beijing, China

Hao Li
De-Xing Huang
Zeng-Guang Hou*
State Key Laboratory of Multimodal Artificial Intelligence Systems, CASIA
University of Chinese Academy of Sciences, Beijing, China
zengguang.hou@ia.ac.cn

ABSTRACT

This paper presents a novel learning-based algorithm to investigate the high-level shared semantic information between electroencephalography (EEG) and electromyography (EMG) signals, for understanding brain-muscle modulation during movement execution. The proposed algorithm incorporates a spatial encoder that condenses spatial information obtained from EEG/EMG signals into unified temporal tokens using a learnable correlation matrix. These tokens are then encoded and decoded via a siamese temporal encoder and classification head to extract joint semantic information presented in cross-modal signals. Additionally, an analysis pipeline is designed to examine brain-muscle modulation based on the proposed algorithm. Experimental results from a self-collected multimodal bio-signals dataset validate the efficacy of the proposed algorithm in extracting and analyzing high-level latent semantic information shared in EEG and EMG signals, outperforming the state-of-the-art model by 5.35% in accuracy, 4.69% in precision, and 8.65% in recall. Notably, the designed analysis pipeline can also reveal low-level relationships, such as those related to time and space, between multimodal bio-signals. This research provides neuroscientists with a valuable tool for obtaining enhanced insights into brain-muscle modulation.

CCS CONCEPTS

• **Computing methodologies** → **Cognitive science**; • **Human-centered computing** → *Human computer interaction (HCI).*

KEYWORDS

spatial-temporal representation learning, multimodal bio-signal analysis, brain-muscle modulation

*Xiao-Hu Zhou and Zeng-Guang Hou are the corresponding authors. xiaohu.zhou@ia.ac.cn, zengguang.hou@ia.ac.cn.

MM '23, October 29-November 3, 2023, Ottawa, ON, Canada

ACM ISBN 979-8-4007-0108-5/23/10.
https://doi.org/10.1145/3581783.3613794

ACM Reference Format:
Tian-Yu Xiang, Xiao-Hu Zhou*, Xiao-Liang Xie, Shi-Qi Liu, Hong-Jun Yang, Zhen-Qiu Feng, Mei-Jiang Gui, Hao Li, De-Xing Huang, and Zeng-Guang Hou*. 2023. Learning Shared Semantic Information from Multimodal Bio-signals for Brain-Muscle Modulation Analysis. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23), October 29-November 3, 2023, Ottawa, ON, Canada.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3581783.3613794

1 INTRODUCTION

Physical movements in individuals rely on a complex interplay between the brain and muscular systems. In this process, neurons generate electrical signals in the brain, which can be recorded using electroencephalography (EEG) [20]. Motor nerves then relay these commands to the muscles, resulting in contractions that can be characterized using electromyography (EMG) [8]. Understanding this interaction is critical for rehabilitation engineering [17] and human-machine interaction [31]. Despite the challenges in investigating these interactions, examining multimodal bio-signals, including EEG and EMG, offers valuable insights into the underlying mechanisms.

Motivated by the potential to enhance the understanding of brain-muscle modulation, numerous studies have explored this relationship using multimodal electronic bio-signals. Investigations have covered aspects like frequency properties [19, 32], temporal properties [29], and spatial properties [33, 34]. However, these low-level properties often rely on pre-defined statistical indices designed for specific properties, which may not fully capture the intricate modulation between brain and muscle systems [19, 29, 32–34]. Consequently, their ability to reveal the complete complexity of the relationship is limited. To address this issue, it is crucial to investigate high-level properties, such as semantic relationships, to enable more targeted analysis. This paper proposes a novel learning-based algorithm to explore these high-level associations.

To analyze high-level semantic information in bio-signals, current methods mainly focus on extracting spatial-temporal representations [5, 14, 21, 26, 28, 35, 40]. This can be achieved using statistical algorithms like filter banks and common space filters [2, 5, 14], or learning-based models such as convolutional and graph neural networks [14, 21, 26–28, 35, 38–40]. These methods face challenges when attempting to simultaneously extract and analyze signals from different sources, as they are primarily designed for single signal source analysis. However, to understand brain-muscle modulation, it's important to jointly analyze bio-signals from the brain and muscle system.

EEG and EMG signals exhibit significant shared semantic information during the execution of specific movements [4, 6, 18]. Previous studies have emphasized the shared semantic information by designing separate decoders for EEG and EMG corresponding to the same movement and demonstrating improved accuracy when combining their performance. However, using separate decoders prevents the extraction of joint EEG/EMG representations, which hinders a more in-depth analysis of brain-muscle modulation. Although EEG and EMG signals share similar semantic information during movement execution, the corresponding components mainly reside in the temporal dimension. The complex spatial properties of bio-signals, arising from electrode distribution, make it difficult to learn a shared representation space across these signals. This limitation underscores the need for a more robust approach to jointly learn shared semantic information from multimodal bio-signals, which remains a challenging task.

To address the limitations, this paper introduces a novel algorithm for learning shared semantic information between electroencephalography (EEG) and electromyography (EMG), to investigate brain-muscle modulation during movement execution. Multimodal bio-signals are first transformed into unified temporal tokens using a spatial encoder, which leverages a learnable correlation matrix to capture the spatial information. These tokens are then further encoded by a siamese temporal encoder to extract shared semantic information. As the primary objective of the proposed method is to learn shared semantic information, a motor classification problem is employed as the proxy-downstream task. Experimental results on a self-collected multimodal bio-signals dataset demonstrate the effectiveness of the proposed approach.

In summary, this paper presents an algorithm for extracting shared semantic information by learning unified temporal-spatial representations of multimodal bio-signals. Additionally, an analysis pipeline for brain-muscle modulation is designed, highlighting the method's ability to explore low-level relationships between multimodal bio-signals. The contributions of this paper can be summarized as follows:

- This paper proposes a novel algorithm for analyzing shared semantic information in multimodal bio-signals. To the best of our knowledge, this is the first study that demonstrates the joint learning of spatially encoded EEG and EMG signals using a siamese structure.
- Evaluation on a self-collected dataset demonstrates superior performance compared to state-of-the-art models, with an improvement of 5.35% in accuracy, 4.69% in precision, and 8.65% in recall for the EEG signal analysis.
- This work offers a learning-based tool for the neuroscience community, enabling the analysis of both high-level and low-level brain-muscle modulation during movement execution. This contribution has the potential to enhance the understanding of brain-muscle modulation and support advancements in various applications.

2 RELATED WORKS

This section discusses the related work in the areas of brain-muscle modulation analysis and bio-signal decoding.

2.1 Brain-muscle Modulation Analysis using Bio-signals

Analyzing brain-muscle modulation has garnered significant interest due to potential applications in emerging fields such as rehabilitation engineering [17] and human-machine interaction [31]. The methodology of existing bio-signals-based analysis often involves an index for an interested property and subsequently investigating relevant components based on the pre-defined index [19, 29, 32–34]. However, these methods face challenges in capturing the complex relationship between brain and muscle systems due to their focus on low-level relationships from limited perspectives. In contrast, the proposed analysis pipeline in this work investigates both high-level and low-level relationships, enabling a more comprehensive understanding of brain-muscle modulation.

2.2 Bio-signal Decoding

Bio-signal decoding algorithms can be broadly described as processes for extracting spatial-temporal representations. Among various methods, those based on statistical models generally exhibit relatively lower performance [2, 5, 14], as they fail to extract joint spatial-temporal representations. For instance, while the common space filter [14] is an effective tool for extracting spatial information, it contains minimal temporal features. Learning-based methods can jointly learn spatial-temporal representations based on various network structures, such as convolutional neural networks (CNN) [26, 40], mixed CNN [10, 21, 38, 39], graph neural networks [27, 35], and transformers [28, 36]. However, a primary limitation of existing learning-based methods is their inability to effectively handle and extract shared semantic information from multimodal bio-signals. The proposed method, as the first of its kind, learns representations for multimodal bio-signals in a joint space using a signal model, allowing for the extraction of shared semantic information to further advance the field.

3 METHODS

3.1 General Structure

The proposed model integrates two electronic bio-signals: EEG ($\mathbf{X}_{\text{EEG}}$) and EMG ($\mathbf{X}_{\text{EMG}}$), as depicted in Figure 1. The design of the structure is to capture the synchronous activation of certain brain regions or muscles during specific operations and the cessation of activation post-operation. This phenomenon is called Event-Related Synchronization/Desynchronization (ERS/ERD) in EEG studies [25].

It begins with a spatial encoder, which captures the spatial information of electrode placement while preserving the temporal structure of the raw signals. This encoder creates unified temporal representations with consistent dimensions for both signals:

$$\mathbf{T}_{\text{EEG}} = \alpha(\mathbf{X}_{\text{EEG}}) \tag{1}$$

$$\mathbf{T}_{\text{EMG}} = \beta(\mathbf{X}_{\text{EMG}}) \tag{2}$$

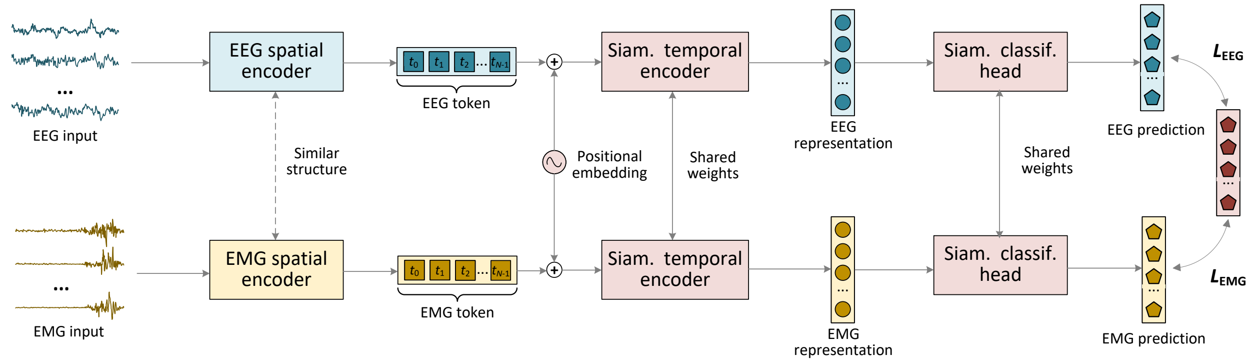

Figure 1: Overall structure of the proposed algorithm.

where α and β are spatial encoders for EEG/EMG signals, and $\mathbf{T}_{\mathrm{EEG/EMG}}$ refers to the encoded EEG/EMG tokens. Although α and β do not share parameters, they possess similar structures to encode raw bio-signals in the same dimension. Furthermore, a temporal encoder is incorporated to extract additional temporal representation from the EEG/EMG tokens. In order to capture the temporal order of the tokens more effectively, a positional embedding is added to each EEG/EMG token.

$$\mathbf{R}_{\mathrm{EEG/EMG}} = g(\mathbf{T}_{\mathrm{EEG/EMG}} + \mathbf{P}) \tag{3}$$

where $\mathbf{P}$ denotes the positional embedding, $\mathbf{R}_{\mathrm{EEG/EMG}}$ are deep representations for EEG/EMG signals, and g represents a siamese temporal encoder, which shares for $\mathbf{T}_{\mathrm{EEG}}$ and $\mathbf{T}_{\mathrm{EMG}}$. The siamese temporal encoder is chosen for its ability to extract shared features and temporal dependencies from both types of signals in a joint representation space effectively. Notably, the temporal encoder and positional embedding are shared for both EEG and EMG signal streams. To supervise the proposed framework, a proxy-downstream task is utilized. For instance, in this study, a classification task is chosen as an example, and a siamese classification head is applied to classify the EEG/EMG representations.

$$\hat{\mathbf{y}}_{\mathrm{EEG/EMG}} = h_c(\mathbf{R}_{\mathrm{EEG/EMG}}) \tag{4}$$

$$L = L_{\mathrm{EEG}}(\mathbf{y}, \hat{\mathbf{y}}_{\mathrm{EEG}}) + L_{\mathrm{EMG}}(\mathbf{y}, \hat{\mathbf{y}}_{\mathrm{EMG}}) \tag{5}$$

where h_c is a siamese classification head which is shared for EEG and EMG stream, $\hat{\mathbf{y}}_{\mathrm{EEG/EMG}}$ is the prediction from EEG/EMG signals, $\mathbf{y}$ is the ground truth, and L is the cost function. By minimizing L, the unified semantic information in EEG/EMG can be extracted.

3.2 Spatial Encoder

The spatial encoder focuses on the frequency and relationship among electrodes while preserving the temporal structure. This module consists of two steps: multi-scale feature extraction and spatial feature fusion, as illustrated in Figure 2.

3.2.1 Multi-scale feature extraction. This step processes raw input bio-signals using multiple 1D CNNs that act as dynamic filters. It aims to capture high/low-frequency patterns within the signals. The extraction of high/low-frequency features is accomplished by varying the kernel length of the dynamic filters, with k different scales of filters employed in the module. These dynamic filters with varied kernel lengths can be treated as filter banks, which are commonly used for extracting high/low-frequency patterns present in bio-signal feature extraction. Given that the maximum kernel length is L, the multi-scale kernel lengths are set to L, $L/2$, $L/4$, ..., $L/2^{(k-1)}$. The filtered signals are subsequently concatenated for further processing, creating a more comprehensive representation of the input signals.

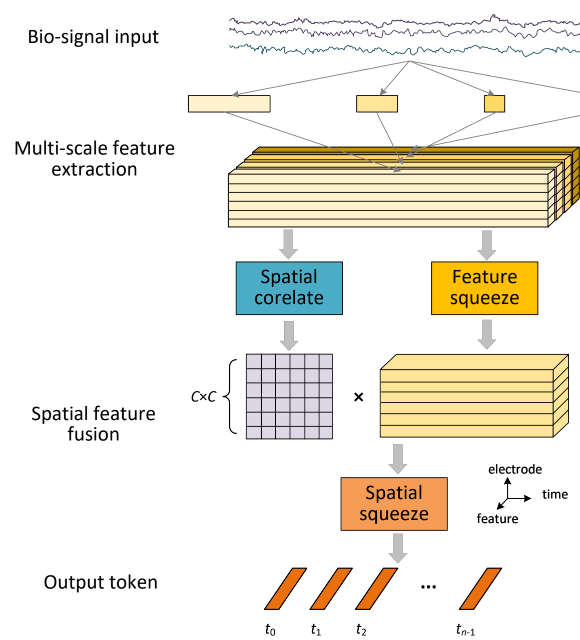

Figure 2: Structure of the spatial encoder for bio-signals.

3.2.2 Spatial feature fusion. This step leverages the inherent relationships among the electrodes in order to extract meaningful spatial features from the bio-signals. As functional connection and 3D spatial structure inherent in the bio-signals are lost in the structured data, a learnable correlation matrix is introduced to describe the relationships among different electrodes based on the multi-scale features. This matrix enables an efficient feature fusion process from the electrodes based on spatial correlations.

The correlation matrix is learned from the multi-scale feature. Firstly, a 1×1 CNN is employed to summarize the multi-scale feature along the feature axis. Next, a max pooling operation is performed to extract notable features along the time axis. In order to fully investigate relationships among electrodes, two-layer fully connected neural networks are utilized to obtain the correlation matrix. To model the correlation between electrodes, the following operation is conducted:

$$\mathbf{M}_{\text{sym}} = \frac{1}{2}\left(\mathbf{M} + \mathbf{M}^{T}\right) \tag{6}$$

where $\mathbf{M}$ denotes the correlation matrix generated by the neural networks and $\mathbf{M}_{\text{sym}}$ is the corresponding symmetric correlation matrix, ($\mathbf{M}, \mathbf{M}_{\text{sym}} \in \mathbb{R}^{C\times C}$), with C being the number of bio-signals' electrodes. The normalized correlation matrix of $\mathbf{M}_{\text{sym}}$ is computed for enhancing the relationships among electrodes:

$$\bar{\mathbf{M}} = \mathbf{I} + \mathbf{D}^{-1/2}\mathbf{M}_{\text{sym}}\mathbf{D}^{-1/2} \tag{7}$$

where $\mathbf{I}$ is a identity matrix, $\mathbf{D}$ represents the degree matrix of $\mathbf{M}_{\text{sym}}$, $\bar{\mathbf{M}}$ is the normalized correlation matrix.

The feature squeeze operation is performed by a 1×1 CNN, which models the relationships among the multi-scale features. The spatial feature fusion is carried out by the matrix multiplication operation between $\bar{\mathbf{M}} + \mathbf{I}$ and the squeezed feature.

$$\mathbf{F}_{f} = (\bar{\mathbf{M}} + \mathbf{I}) \times \mathbf{F}_{s} \tag{8}$$

where $\mathbf{F}_s$ is the squeezed feature and $\mathbf{F}_f$ denotes fused feature. As $\mathbf{I}$ is an identity matrix, this operation can be regarded as either a residual connection mentioned in ResNet [15], or as a self-connection where the electrode with the highest weight is itself.

The electrode squeezing operation is accomplished by depthwise convolution across the channels. This step simulates the traditional electrophysiological signal feature extraction algorithm Common Spatial Patterns (CSP) [14], initially utilized in EEGNet [21] and widely adopted thereafter.

3.3 Siamese Temporal Encoder

Given the underlying physical processes of generating EEG and EMG signals, it is evident that a time delay exists between the semantic context encoded in these two signals. While EEG originates from the brain and is closely related to the processes of perception and control, EMG reflects the resulting muscular movements following such command.

To fully explore both the long-term global dependencies and short-range local dependencies inherent in the bio-signals, a siamese temporal encoder based on the long short-term memory networks (LSTM) [16] and transformer [30] is formulated. The LSTM and transformer structures are parallelized to encode EEG/EMG tokens with positional embedding, which adds information about the position of the tokens within the sequence, allowing the model to capture temporal relationships. While the LSTM is capable of capturing short-range local dependencies, the transformer-based structure aims to summarize long-term global dependencies.

The encoded tokens are concatenated as EEG/EMG representations. Notably, the EEG and EMG tokens share the same temporal encoder to obtain deep representations. This siamese encoder is implemented using a joint architecture, which enables the model to learn a unified representation in the same space that effectively captures the inherent relationships between the two sources of signals.

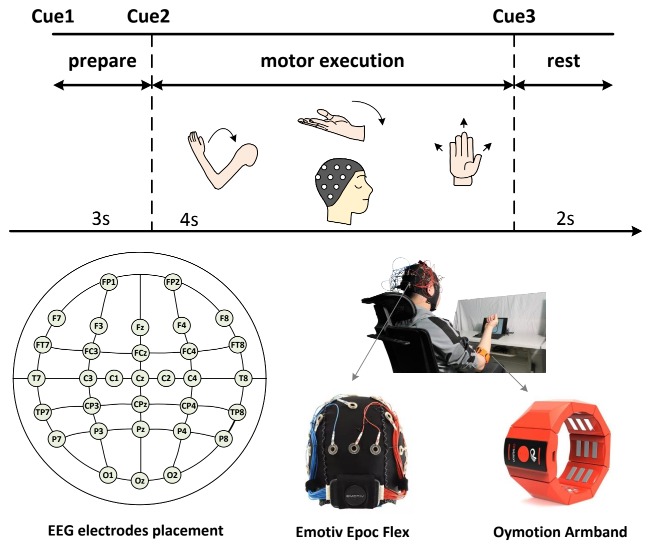

Figure 3: Electrodes placement and data acquisition pipeline used in the self-collected multimodal dataset.

3.4 Proxy-downstream Task

The design of the proxy-downstream task depends on the research objectives. Here, the primary aim is to investigate the shared high-level semantic information between EEG and EMG signals during motor execution which reflects movement types. Therefore, a motor classification task is selected as the proxy-downstream task.

To build a model for the designated proxy-downstream task, a two-layer fully connected neural network is formulated as the classification head. This classification head is placed after the siamese temporal encoder to classify the EEG/EMG representations. It is important to note that the EEG and EMG representations share the same classification head in this study, emphasizing the joint learning of both types of signals. This shared classification head enables the model to learn a common representation space for both EEG and EMG signals, which is useful for capturing the inherent relationships between the two modalities.

Since the proxy-downstream task is a classification problem, the smoothed cross-entropy loss [23] is applied as L_{EEG} and L_{EMG} to mitigate the over-fitting that often occurs in bio-signal processing. The smoothed cross-entropy loss is a variation of the standard cross-entropy loss, which adds a smoothing term to the loss function to regularize the model and prevent overfitting, particularly in situations where the available data is limited or noisy.

Table 1: Ablation study on the self-collected multimodal dataset under 'Only EEG' and 'EEG-EMG' training paradigms.

Structure		'Only EEG'						'EEG-EMG'					
SFF	STE	Accuracy		Precision		Recall		Accuracy		Precision		Recall	
×	×	0.423	9.22% ↑	0.420	10.71% ↑	0.416	12.98% ↑	0.446	6.05% ↑	0.444	5.63% ↑	0.437	12.13% ↑
✓	×	0.428	8.41% ↑	0.427	8.90% ↑	0.427	10.07% ↑	0.447	5.82% ↑	0.443	5.87% ↑	0.437	12.13% ↑
×	✓	0.441	5.22% ↑	0.440	5.68% ↑	0.452	3.98% ↑	0.458	3.27% ↑	0.456	2.85% ↑	0.452	8.41% ↑
✓	✓	**0.464**		**0.465**		**0.470**		**0.473**		**0.469**		**0.490**	

4 EXPERIMENTS AND RESULTS

4.1 Dataset Description

To investigate the brain-muscular modulation in participants during performing specific movements, we have acquired a multimodal bio-signal dataset, as illustrated in Figure 3. The dataset comprises 4, 000 EEG and EMG recordings from 10 participants, each performing designated motor tasks. EEG data were collected using a 32-channel Emotiv Epoc Flex EEG head cap with a sampling frequency of 128 Hz. Simultaneously, EMG data were acquired through an 8-channel Oymotion Armband at a frequency of 500 Hz. The experimental paradigm encompassed three distinct motor tasks executed by the right hand/arm: (1) twisting the wrist, (2) opening the hand, and (3) flexing the elbow. Each task was performed about 100 times by the participants. Additionally, a resting state was recorded, serving as an additional condition for analysis purposes.

4.2 Implementation Details

4.2.1 Experimental Setup. A 5 seconds time window of EEG and EMG data is selected, sliced between 1 second before Cue 2 up to Cue 3, as illustrated in Figure 3. The proposed approach is evaluated through subject-dependent experiments. The entire dataset is randomly split into three sets: training set (70% of trials), testing set (20% of trials), and validation set (10% of trials) for each training phase. The model is trained for 200 epochs and the final model tested on the testing set is the one that exhibits the best performance on the validation set. To ensure that all trials are evaluated, a five-fold cross-validation on the whole dataset is conducted.

4.2.2 Preprocessing. The EEG signals are pre-processed by applying a band-pass filter with a range from 8 Hz to 30 Hz, focusing on the most movement-relative frequency bands (alpha, beta), and a notch filter at 50 Hz to eliminate powerline interference. Furthermore, an independent component analysis (ICA) is performed to remove artifacts before analysis. The EMG signals are resampled at a rate of 512 Hz. Prior to inputting the signals into the model, all EEG and EMG signals in both datasets are normalized by setting the mean to 0 and the standard deviation to 1.

4.2.3 Model Configuration. The model is implemented using the TensorFlow framework and optimized by the Adamw optimizer [22]. The learning rate is set to $1e{-4}$, while the batch size is set to 64. To prevent overfitting, a dropout rate of 0.3 is applied. The kernel length for the 1D CNN in the spatial encoder is set to 1 second to correspond with the respective signals, which are 128 and 512 for EEG and EMG on the self-collected multimodal dataset. The siamese temporal encoder consists of a 3-layer transformer and LSTM. The hidden size for the forward layer in the transformer is set to 256, the hidden units for the LSTM is 64, while the hidden size for the general classifier is set to 32.

Table 2: Comparison with other representative methods on the self-collected multimodal dataset for EEG decoding

Method	Accuracy	Precision	Recall
WT [5]	0.311	0.311	0.306
CSP [14]	0.333	0.334	0.332
FBCSP [2]	0.353	0.353	0.352
DeepNet [26]	0.428	0.424	0.450
EEGNet [21]	0.449	0.448	0.451
GraphEEG [35]	0.366	0.364	0.365
Conformer [28]	0.412	0.411	0.404
Ours (only EEG)	**0.464**	**0.465**	**0.470**
Ours (EEG-EMG)	**0.473**	**0.469**	**0.490**

4.3 Experimental Results

As outlined in the Method section, the proxy-downstream task is selected as the motor classification for extracting high-level semantic information. The classification results are reported in Table 2. Given that the accuracy of the classification results for EMG is approximately 100%, this section focuses on the EEG results. Two training paradigms of the proposed model are discussed:

- Only EEG: Both the training and inference rely solely on EEG data (i.e., the EMG temporal encoder is not trained).
- EEG-EMG: The training process utilizes both EEG and EMG signals, but the inference process employs only EEG data.

The proposed model are compared to both representative traditional methods [2, 5, 14] and learning-based methods [21, 26, 28, 35]. These methods are chosen due to their varied techniques for modeling the temporal and spatial representations in the EEG signals: wavelet transform [5], common spatial filter [14], filter bank common spatial filter [2], CNN [26], mixed-CNN [21], GNN [35], and transformer [28]. The performance of the proposed model exceeds the best model among the compared models by 3.34%/5.35% in accuracy, 3.79%/4.69% in precision, and 4.21%/8.65% in recall under the 'Only EEG'/'EEG-EMG' training paradigms.

It is worth noting the significant improvement from 'Only EEG' to 'EEG-EMG' (1.94% in accuracy, 0.86% in precision, and 4.3% in recall). The results indicate that the inclusion of EMG signals enhances the model's ability to decode EEG signals. These findings

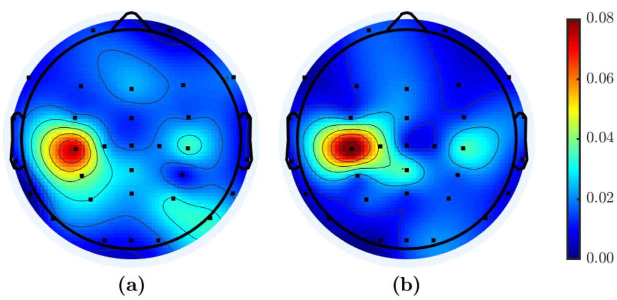

Figure 4: Visualization of the importance score I for different EEG electrodes under the (a) 'EEG-EMG' and (b) 'Only EEG' training paradigm.

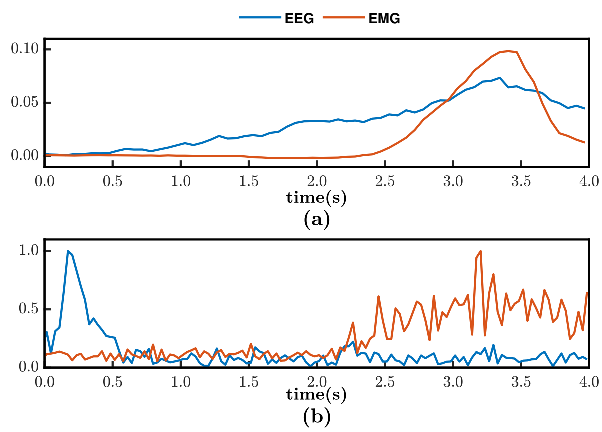

Figure 5: (a) Importance score I for EEG and EMG signals through time; (b) normalized envelope for average EEG and EMG signals. The timeline of the figures begins after Cue 2 in Figure 3.

provide evidence that the proposed model can effectively encode EEG and EMG signals into a joint representation space and extract their shared semantic information.

4.4 Ablation Study

The ablation study investigates the impact of two primary modules in our proposed method: Spatial Feature Fusion (SFF) and Siamese Temporal Encoder (STE). Under the 'Only EEG'/'EEG-EMG' training paradigms, the proposed method exhibits improvements of 9.22%/6.05% in accuracy, 10.71%/5.63% in precision, and 12.98%/12.13% in recall, respectively, compared to the baseline (without STE and SFF). These substantial improvements underscore the effectiveness of the ablated components in learning the semantic information from bio-signals in our proposed model.

5 ANALYSIS OF LOW-LEVEL RELATIONSHIPS

This section explores the application of the proposed algorithm in examining the low-level relationships among multimodal signals. The analysis pipeline is first outlined and validated on some well-established low-level relationships. Comparisons of the analysis results are presented for certain low-level relationships under the 'Only EEG' and 'EEG-EMG' paradigms. These comparisons show the distinctions between representations space for joint EEG-EMG and pure EEG, illustrating that the 'EEG-EMG' paradigm enables a more comprehensive understanding of motor execution.

5.1 Analysis Pipeline

As with previous studies on EEG/EMG signals, the low-level relationships between time, space, and frequency are the main focus of the investigation. Here a pipeline is outlined for analyzing these low-level relationships based on the proposed algorithm. A function N, which aims to remove a certain property of the raw signals, is introduced in this pipeline. The impact of the removed content on the model's performance reflects the importance of the content in learning shared semantic information, allowing for studying the low-level brain-muscle modulation.

$$\mathbf{X}'_{\mathrm{EEG/EMG}}(t, s, f) = N(\mathbf{X}_{\mathrm{EEG/EMG}}, t, s, f) \tag{9}$$

here, $\mathbf{X}'_{\mathrm{EEG/EMG}}(t, s, f)$ represents the modified data, where t represents a range of time, s represents a set of electrodes, and f represents a certain range of frequencies. For time and space, when interested in a specific time range or electrodes, the function N replaces the interested signals with Gaussian noises that conform to the same distribution as the original signals. For frequency, a notch filter is implemented within function N to remove the interested frequency.

The modified data is then input into the trained model described earlier in the study, and the predicted result is denoted by $\mathbf{y}'_{\mathrm{EEG/EMG}}$:

$$\mathbf{y}'_{\mathrm{EEG}} = h_c\left\{g\left[\alpha\left(\mathbf{X}'_{\mathrm{EEG}}(t, s, f)\right) + \mathbf{P}\right]\right\} \tag{10}$$

$$\mathbf{y}'_{\mathrm{EMG}} = h_c\left\{g\left[\beta\left(\mathbf{X}'_{\mathrm{EMG}}(t, s, f)\right) + \mathbf{P}\right]\right\} \tag{11}$$

As the modified data suppresses the interested time, space, or frequency in the EEG/EMG data, if one component is highly correlated with another source of data, the encoded data in the EEG-EMG representation space will change significantly and result in changes in $\mathbf{y}'_{\mathrm{EEG/EMG}}$. To evaluate the importance of certain time, space, or frequency components, an evaluation function for the classification results is defined. The expectation of the change in the evaluation function between the original prediction and the prediction by the modified data is treated as the importance score:

$$I_{\mathrm{EEG/EMG}}(t, s, f) = \mathbb{E}\left[e\left(\hat{\mathbf{y}}_{\mathrm{EEG/EMG}}, \mathbf{y}\right) - e\left(\hat{\mathbf{y}}'_{\mathrm{EEG/EMG}}, \mathbf{y}\right)\right] \tag{12}$$

where e is the evaluation function, which is chosen as accuracy as the downstream task is a classification problem, and I represents the importance score for a certain time (t), space (s), or frequency (f). This analysis pipeline allows for understanding the low-level brain-muscle modulation based on the high-level representations.

5.2 Spatial Relationship

The main findings in the topographical map of the importance score (Figure 4) indicate that the electrodes with high importance are located around the central part of the brain and the parietal

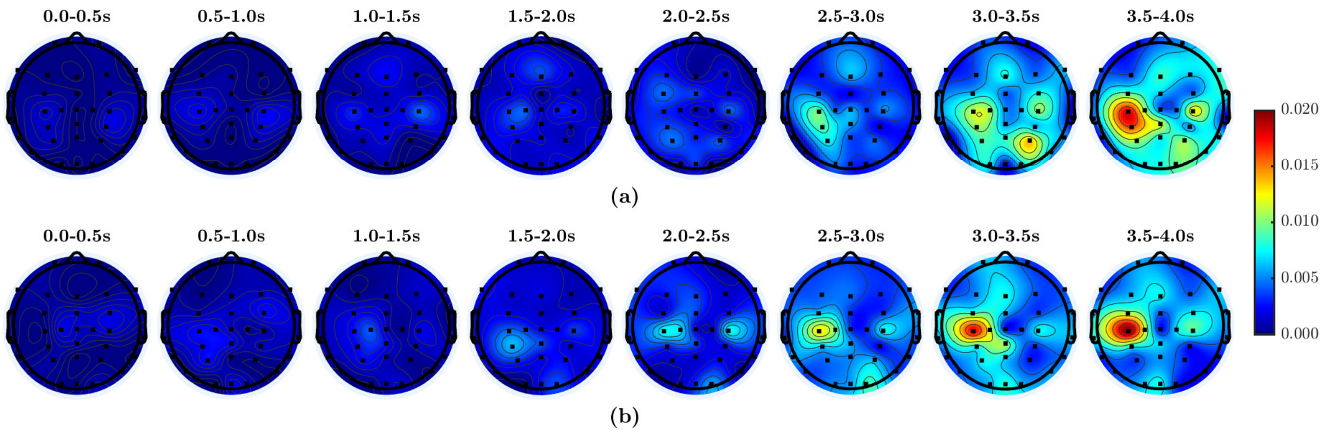

Figure 6: A comprehensive analysis of the importance score (*I*) through temporal-spatial under the (a) 'EEG-EMG' and (b) 'Only EEG' training paradigms. The timeline begins after Cue 2 in Figure 3.

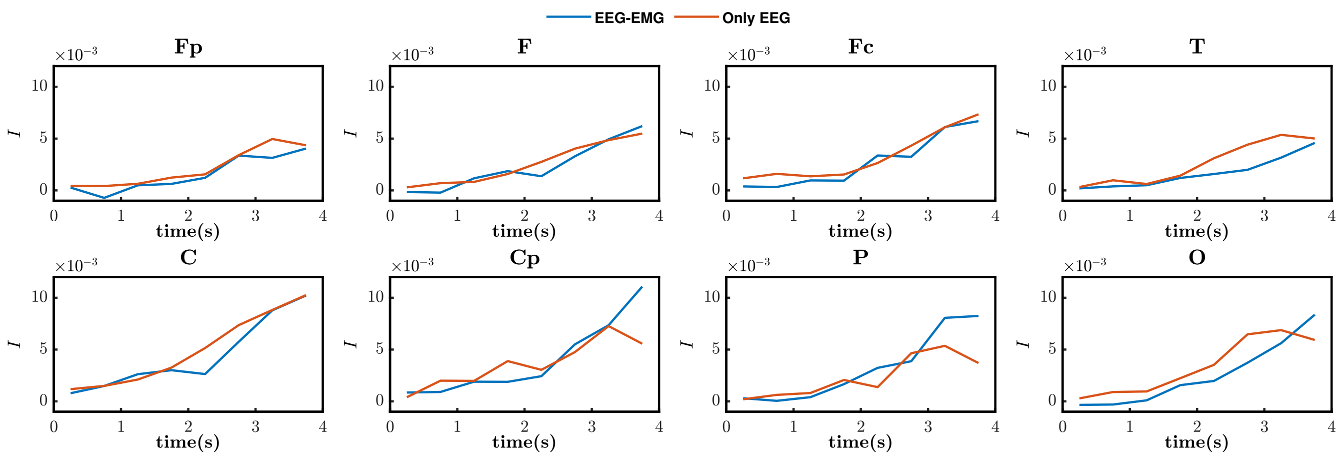

Figure 7: A analysis of the average importance score (*I*) for different brain areas through temporal-spatial under the (a) 'EEG-EMG' and (b) 'Only EEG' training paradigms. Fp: pre-frontal; F: frontal; Fc: frontal-central; T: temporal; C: central; CP: central-parietal; P: parietal; O: occipital. The timeline begins after Cue 2 in Figure 3.

lobe. These areas have been proven to be crucial for motor organization [11, 12]. Additionally, the contralateral activation phenomenon can be observed, with the left hemisphere of the brain exhibiting a relatively higher importance score than the right. This can be attributed to right-hand movements leading to more pronounced characteristics in the left brain hemisphere, which is consistent with existing neuroscience research [3] and practical applications [9].

When comparing the spatial relationship results under different training paradigms, the electrodes with high importance scores are more widely distributed under the 'EEG-EMG' paradigm in the parietal lobe. This observation further supports the notion that joint training with EEG and EMG signals can effectively extract latent semantic information about movements. The parietal lobe is closely associated with motor function, and electrodes in this region are often manually selected for motor decoding [10].

5.3 Temporal Relationship

The temporal relationship analysis focuses on comparing EEG and EMG signals under the 'EEG-EMG' training paradigm, as shown in Figure 5. A notable observation is that the importance score for the EEG signals increases much earlier than that of the EMG signals (EEG at approximately 0.5 seconds after Cue 2 and EMG at 2.4 seconds). This significant time delay between EEG and EMG signals can primarily be attributed to the visual cue carrying movement instructions for participants. A more precise metric for evaluating the time delay between brain and muscle activity is the interval between the maximum importance scores in EEG and EMG signals, as it reflects the most critical time related to motor execution in both signal types. The time interval is approximately 100−200 ms, which is consistent with previous studies [7, 19].

Comparing the importance score over time with the envelope of the average EEG and EMG signals reveals another interesting

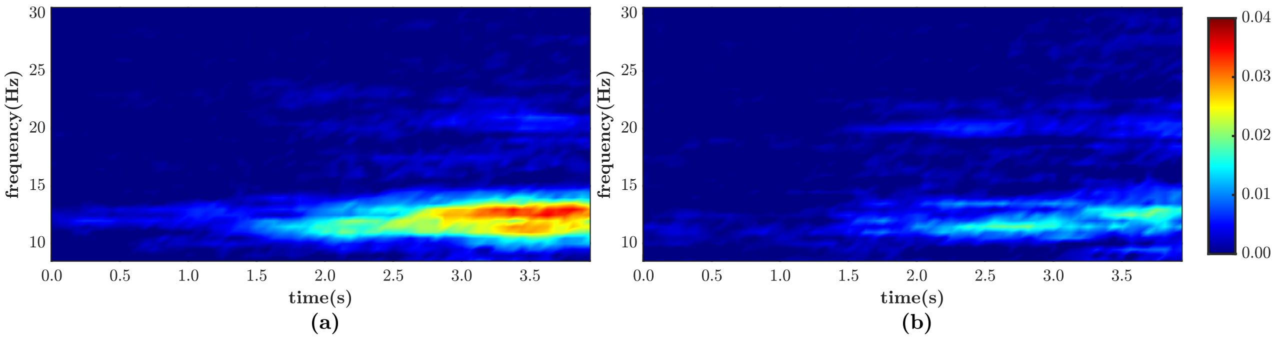

Figure 8: Time-frquency analysis using importance score I under the (a) 'EEG-EMG' and (b) 'Only EEG' training paradigm. The timeline begins after Cue 2 in Figure 3.

phenomenon: both the importance scores of EEG and EMG signals exhibit a more correlated trend with the EMG envelope. Although the EEG envelope shows a significant pattern stimulated by the visual cue at the beginning, it contributes little, as reflected by the importance score I. This observation not only supports the assertion that the proposed model can extract shared semantic information in the EEG and EMG signals related to movement but also demonstrates the model's ability to identify meaningful temporal relationships from the data.

5.4 Spatial-Temporal Relationship

As shown in Figure 6, the topographical maps' changes over time provide valuable insights into the variations in spatial information. For example, at 2 s after the Cue 2, the importance score in the central and parietal lobes increases, highlighting their role in movement organization [11, 12]. The frontal lobe, encompassing the premotor cortex and the primary motor cortex, also exhibits a high importance score during movement execution [24].

Furthermore, upon examining the average importance score for different areas of electrodes (Figure 7), it is observed that under the 'Only EEG' paradigm, the pre-frontal lobe (Fp), temporal lobe (T), and occipital lobe (O) exhibit slightly higher importance scores than under the 'EEG-EMG' paradigm. However, these areas are more connected with cognitive functions rather than direct movement execution [1, 13]. Conversely, under the 'EEG-EMG' training paradigm, the more significant areas are located around the central part and parietal lobes. This observation indicates that the proposed model can effectively focus on movement-related semantic information by jointly training with EEG and EMG signals, which may lead to a more accurate analysis of brain-muscle modulation.

5.5 Time-Frequency Relationship

As depicted in Figure 8, the frequency band with the highest importance score is within the range of 10 to 15 Hz, covering the sensorimotor rhythm (a sub-band of μ and β band, ranges between 12 to 15 Hz) that is phenomenologically associated with movements [37].

When comparing different training paradigms, it comes that under the 'EEG-EMG' paradigm, the proposed model can better concentrate on the movement-related frequency band in the EEG signals. This improved focus can be attributed to the joint training with EEG and EMG signals, which allows the model to learn and exploit the shared semantic information in both signal types. This observation not only further supports the effectiveness of the proposed model in capturing relevant information but also highlights its potential for better understanding the time-frequency relationship of brain activities during movement execution.

6 CONCLUSION

This paper addresses the challenge of understanding the shared semantic information between EEG and EMG signals during movement execution. To this end, a novel learning-based algorithm is designed, which contains a spatial encoder to generate EEG/EMG tokens and a siamese temporal encoder and classification head to learn the analogous semantic information inherent in these signals. Experiments conducted on a self-collected multimodal dataset demonstrated the effectiveness of the proposed model in capturing the shared semantic information between EEG and EMG signals. The proposed analysis pipeline has been validated for investigating brain-muscle modulation from multiple perspectives, offering valuable insights for neuroscience. The approach presented in this paper holds promise for training general models applicable to various bio-signals, with potential benefits for a wide range of applications, such as brain-computer interfaces, neurorehabilitation, and prosthetics. In the future, we plan to further develop the model to investigate brain-muscle modulation during more natural movements, expanding its applicability and enhancing its impact on the field.

7 ACKNOWLEDGMENTS

This work was sponsored in part by the National Natural Science Foundation of China under Grant 62003343, Grant 62222316, Grant U1913601, Grant 62073325, Grant 61720106012, Grant U20A20224 and Grant U1913210; in part by the CIE-Tencent Robotics X Rhino-Bird Focused Research Program; in part by the Beijing Natural Science Foundation under Grant M22008; in part by the Youth Innovation Promotion Association of Chinese Academy of Sciences (CAS) under Grant 2020140; in part by the Strategic Priority Research Program of CAS under Grant XDB32040000; and in part by Beijing Science and Technology Program under Grant Z211100007921021.

REFERENCES

[1] Cettina Allone, Viviana Lo Buono, Francesco Corallo, Laura Rosa Pisani, Patrizia Pollicino, Placido Bramanti, and Silvia Marino. 2017. Neuroimaging and cognitive functions in temporal lobe epilepsy: A review of the literature. *Journal of the Neurological Sciences* 381 (2017), 7–15.

[2] Kai Keng Ang, Zheng Yang Chin, Haihong Zhang, and Cuntai Guan. 2008. Filter bank common spatial pattern (FBCSP) in brain-computer interface. In *In Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. IEEE, 2390–2397.

[3] Javier M Antelis, Luis Montesano, Ander Ramos-Murguialday, Niels Birbaumer, and Javier Minguez. 2016. Decoding upper limb movement attempt from EEG measurements of the contralesional motor cortex in chronic stroke patients. *IEEE Transactions on Biomedical Engineering* 64, 1 (2016), 99–111.

[4] Jordan J Bird, Jhonatan Kobylarz, Diego R Faria, Anikó Ekárt, and Eduardo P Ribeiro. 2020. Cross-domain MLP and CNN transfer learning for biological signal processing: EEG and EMG. *IEEE Access* 8 (2020), 54789–54801.

[5] Duo Chen, Suiren Wan, and Forrest Sheng Bao. 2016. Epileptic focus localization using discrete wavelet transform based on interictal intracranial EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25, 5 (2016), 413–425.

[6] Anirban Chowdhury, Haider Raza, Yogesh Kumar Meena, Ashish Dutta, and Girijesh Prasad. 2019. An EEG-EMG correlation-based brain-computer interface for hand orthosis supported neuro-rehabilitation. *Journal of Neuroscience Methods* 312 (2019), 1–11.

[7] Brian L Day and Peter Brown. 2001. Evidence for subcortical involvement in the visual control of human reaching. *Brain* 124, 9 (2001), 1832–1840.

[8] Carlo De Luca. 2006. Electromyography. *Encyclopedia of Medical Devices and Instrumentation* (2006).

[9] Ranjan Debnath and Elizabeth A Franz. 2016. Perception of hand movement by mirror reflection evokes brain activation in the motor cortex contralateral to a non-moving hand. *Cortex* 81 (2016), 118–125.

[10] Zhijie Fang, Weiqun Wang, Shixin Ren, Jiaxing Wang, Weiguo Shi, Xu Liang, Chen-Chen Fan, and Zengguang Hou. 2021. Learning regional attention convolutional neural network for motion intention recognition based on EEG data. In *Proceedings of the International Conference on International Joint Conferences on Artificial Intelligence*. 1570–1576.

[11] Leonardo Fogassi, Pier Francesco Ferrari, Benno Gesierich, Stefano Rozzi, Fabian Chersi, and Giacomo Rizzolatti. 2005. Parietal lobe: from action organization to intention understanding. *Science* 308, 5722 (2005), 662–667.

[12] Leonardo Fogassi and Giuseppe Luppino. 2005. Motor functions of the parietal lobe. *Current Opinion in Neurobiology* 15, 6 (2005), 626–631.

[13] Naomi P Friedman and Trevor W Robbins. 2022. The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology* 47, 1 (2022), 72–89.

[14] Moritz Grosse-Wentrup and Martin Buss. 2008. Multiclass common spatial patterns and information theoretic feature extraction. *IEEE Transactions on Biomedical Engineering* 55, 8 (2008), 1991–2000.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 770–778.

[16] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.

[17] Mads Jochumsen, Muhammad S Navid, Usman Rashid, Heidi Haavik, and Imran K Niazi. 2019. EMG-versus EEG-triggered electrical stimulation for inducing corticospinal plasticity. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27, 9 (2019), 1901–1908.

[18] Salman Mohd Khan, Abid Ali Khan, and Omar Farooq. 2019. Selection of features and classifiers for EMG-EEG-based upper limb assistive devices—A review. *IEEE Reviews in Biomedical Engineering* 13 (2019), 248–260.

[19] Byeongnam Kim, Laehyun Kim, Yun-Hee Kim, and Sun K Yoo. 2017. Cross-association analysis of EEG and EMG signals according to movement intention state. *Cognitive Systems Research* 44 (2017), 1–9.

[20] Timo Kirschstein and Rüdiger Köhling. 2009. What is the source of the EEG? *Clinical EEG and neuroscience* 40, 3 (2009), 146–149.

[21] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. 2018. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering* 15, 5 (2018), 056013.

[22] Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. *Proceedings of the International Conference on Learning Representations.*

[23] Rafael Müller, Simon Kornblith, and Geoffrey E Hinton. 2019. When does label smoothing help?. In *Proceedings of the Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.), Vol. 32.

[24] Deepak N Pandya and Clifford L Barnes. 2019. Architecture and connections of the frontal lobe. In *The frontal lobes revisited*. Psychology Press, 41–72.

[25] Gert Pfurtscheller and FH Lopes Da Silva. 1999. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical neurophysiology* 110, 11 (1999), 1842–1857.

[26] Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. 2017. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping* 38, 11 (2017), 5391–5420.

[27] Tengfei Song, Wenming Zheng, Suyuan Liu, Yuan Zong, Zhen Cui, and Yang Li. 2021. Graph-embedded convolutional neural network for image-based EEG emotion recognition. *IEEE Transactions on Emerging Topics in Computing* 10, 3 (2021), 1399–1413.

[28] Yonghao Song, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. 2022. EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* (2022), 710–719.

[29] Nyi Nyi Tun, Fumiya Sanuki, and Keiji Iramina. 2021. EEG-EMG Correlation Analysis with Linear and Nonlinear Coupling Methods Across Four Motor Tasks. In *Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 783–786.

[30] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Proceedings of the Advances in Neural Information Processing Systems* 30.

[31] Zhongpeng Wang, Beibei He, Yijie Zhou, Long Chen, Bin Gu, Shuang Liu, Minpeng Xu, Feng He, and Dong Ming. 2022. Incorporating EEG and EMG patterns to evaluate BCI-based long-term motor training. *IEEE Transactions on Human-Machine Systems* 52, 4 (2022), 648–657.

[32] Xugang Xi, Jinsuo Ding, Junhong Wang, Yun-Bo Zhao, Ting Wang, Wanzeng Kong, and Jingqi Li. 2022. Analysis of Functional Corticomuscular Coupling Based on Multiscale Transfer Spectral Entropy. *IEEE Journal of Biomedical and Health Informatics* 26, 10 (2022), 5085–5096.

[33] Xugang Xi, Ziyang Sun, Xian Hua, Changmin Yuan, Yun-Bo Zhao, Seyed M Miran, Zhizeng Luo, and Zhong Lü. 2021. Construction and analysis of cortical–muscular functional network based on EEG-EMG coherence using wavelet coherence. *Neurocomputing* 438 (2021), 248–258.

[34] Xugang Xi, Xiangxiang Wu, Yun-Bo Zhao, Junhong Wang, Wanzeng Kong, and Zhizeng Luo. 2021. Cortico-muscular functional network: an exploration of cortico-muscular coupling in hand movements. *Journal of Neural Engineering* 18, 4 (2021), 046084.

[35] Tian-Yu Xiang, Xiao-Hu Zhou, Xiao-Liang Xie, Shi-Qi Liu, Hong-Jun Yang, Zhen-Qiu Feng, Jian-Long Hao, and Zeng-Guang Hou. 2022. A Novel Two-Stream Model for Human Motor Characteristics Learning. In *Proceedings of the 2022 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 339–344.

[36] Jin Xie, Jie Zhang, Jiayao Sun, Zheng Ma, Liuni Qin, Guanglin Li, Huihui Zhou, and Yang Zhan. 2022. A Transformer-based approach combining deep learning network and spatial-temporal information for raw EEG classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 30 (2022), 2126–2136.

[37] Han Yuan and Bin He. 2014. Brain–computer interfaces using sensorimotor rhythms: current state and future perspectives. *IEEE Transactions on Biomedical Engineering* 61, 5 (2014), 1425–1435.

[38] Dalin Zhang, Lina Yao, Kaixuan Chen, Sen Wang, Xiaojun Chang, and Yunhao Liu. 2019. Making sense of spatio-temporal preserving representations for EEG-based human intention recognition. *IEEE Transactions on Cybernetics* 50, 7 (2019), 3033–3044.

[39] Dalin Zhang, Lina Yao, Xiang Zhang, Sen Wang, Weitong Chen, Robert Boots, and Boualem Benatallah. 2018. Cascade and parallel convolutional recurrent neural networks on EEG-based intention recognition for brain computer interface. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32. 1703–1710.

[40] Yongxiang Zou and Long Cheng. 2021. A transfer learning model for gesture recognition based on the deep features extracted by CNN. *IEEE Transactions on Artificial Intelligence* 2, 5 (2021), 447–458.