2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Wavelet-Based Cross-Frequency and Cross-Region Interaction Convolutional Neural Network for Working Memory Load Level Detection

Congming Tan[1], Yahong Ma[3] and Yin Tian[1,2[✉]]

[1] School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[2] School of Life and Health Information Science and Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[3] Xi'an Key Laboratory of High Precision Industrial Intelligent Vision Measurement Technology, School of Electronic Information, Xijing University, Xi'an 710123, China
✉ Corresponding author: tianyin@cqupt.edu.cn

Abstract. The detection of Working Memory Load (WML) plays a crucial role in neurofeedback processes and the treatment of disorders such as ADHD. However, the performance of existing detection methods remains unsatisfactory. Neuropsychology research indicates that high-level cognitive processes are driven by both inter-regional collaborations across different brain functional areas and cross-frequency couplings. To comprehensively capture brain activities spanning both frequency domains and intra/inter-regional interactions, we propose a novel cognitively-inspired neural network – the Wavelet-based Cross-Frequency and Cross-Region Interaction Convolutional Neural Network (CFCRNet) – for WML decoding. Specifically, CFCRNet first employs predefined wavelet kernels to perform 1D convolution for time-frequency feature extraction, followed by multi-branch learning to model cross-frequency feature coupling with varying scales, and finally integrates intra- and inter-regional information interactions through spatial attention mechanisms. This architecture systematically fuses neurophysiologically meaningful cross-frequency coupling mechanisms with functional integration principles across brain regions, constructing a network model capable of simultaneously resolving dynamic characteristics of neural signals across different frequency bands and complex interactive relationships within/between functional areas. Experimental validation on our collected working memory dataset and public benchmarks demonstrates that incorporating neuroscientific priors into neural network design enhances classification performance. Collectively, our findings establish an advanced framework for accurate WML detection that can be extended to explore detection tasks associated with other cognitive behaviors and neurological disorders.

Keywords: working memory, cross-frequency, cross-region interaction.

1 Introduction

Cognitive working memory load (WML) refers to the physiological and psychological demands generated during the execution of single or multiple tasks, serving as a quantifiable indicator of the proportion of the human brain's information processing capacity engaged in given cognitive activities [1],[2]. With the rapid advancement of contemporary technologies and escalating complexity of occupational requirements, WML has inevitably intensified. However, given the finite nature of human cognitive resources, prolonged exposure to high-intensity WML may compromise both mental and physical health, subsequently diminishing task performance. Current research environments for cognitive workload recognition encompass diverse domains including but not limited to educational settings [3] (e.g., student online learning [4]), transportation systems (e.g., vehicle operation [5], aviation [6]), and more recently, computer-assisted diagnostics for medical conditions such as cancer [7], depression [8], schizophrenia [9], and autism spectrum disorders [10]. Consequently, precise detection of neural activity under varying WML levels not only facilitates the development of adaptive strategies to enhance operational efficiency, but also provides an empirical foundation for parameter optimization in neuromodulation interventions and objective evaluation of therapeutic outcomes.

To date, the measurement of cognitive WML can be classified into subjective and objective approaches [11]. Subjective measures rely on participants' self-perception and self-evaluation through standardized questionnaires [12],[13]. While these approaches are easily implementable, their reliability predominantly hinges on participants' honesty, thereby failing to provide real-time, objective, and accurate assessments. In contrast, objective measures primarily utilize task performance metrics and physiological signal recordings, which minimize task interference and address the aforementioned limitations. Commonly employed physiological modalities include electroencephalography (EEG), functional magnetic resonance imaging (fMRI), and functional near-infrared spectroscopy (fNIRS) [14]. Notably, EEG has emerged as one of the most widely adopted physiological measurement techniques for estimating cognitive WML due to its portability, cost-effectiveness, and capacity for real-time monitoring of cortical electrical activity [3]. Its applications extend to diverse cognitive domains such as emotion recognition [15],[16], mental arithmetic tasks [17], and n-back paradigms [18]. Building on these methodological advantages, this study focuses on EEG-based assessment of cognitive WML.

EEG signals pose significant challenges for robust feature extraction due to their weak amplitude, high noise susceptibility, and non-stationary nature. Traditional analytical approaches predominantly rely on statistical validation of feature differences, including spectral power variations in specific frequency bands [19] (e.g., theta-band power augmentation and alpha-band suppression under increased workload [20],[21], along with δ-, β-, and γ-band correlations [22],[23],[24]), time-domain event-related potential characteristics (e.g., P300 amplitude attenuation with load escalation [20],[25]), and complexity metrics (LZC, SEn, and their fusion variants [26]). These methods typically employ machine learning classifiers like SVM and LDA for pattern recognition, yet suffer from limited generalization due to EEG's temporal variability,

while manual feature engineering introduces computational complexity and scalability constraints in large-scale data processing. The advent of deep learning has enabled end-to-end convolutional neural network (CNN) frameworks to autonomously extract multi-domain features [27],[28]. Specifically, convolutional filter banks synchronously capture spatio-temporal-spectral patterns (e.g., EEGNet's triple-convolution architecture for efficient representation in P300/ERN paradigms [29]), pooling layers achieve feature dimensionality reduction and semantic aggregation, and fully-connected layers integrate high-order representations. Notable implementations include frequency-transformed multiband 2D-CNN mappings [30], 3D convolutional attention-based decoders [31], and deep architectures for driving load classification [5]. Compared to conventional two-stage paradigms (handcrafted features + classifiers), deep learning mitigates modeling difficulties arising from EEG's variability and dynamic fluctuations through hierarchical feature abstraction, demonstrating superior representational capacity and generalizability in cognitive workload recognition applications.

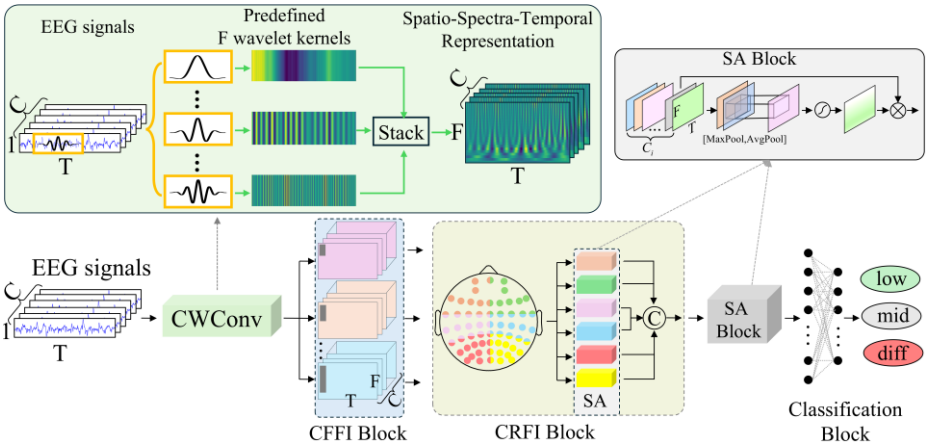


Fig. 1.. Overview of the proposed CFCRNet architecture

Current research on EEG-based feature learning for WML predominantly focuses on spatio-temporal information. In temporal dimension modeling, researchers have achieved effective extraction of temporal dynamics through one-dimensional convolutional neural networks (1D-CNNs) [29],[32] and their multi-scale extended architectures [33]. However, existing paradigms exhibit representational limitations in spatial dimension modeling: while 1D convolutional operations along electrode sequences capture global spatial correlations, they inadequately resolve local topological properties of brain regions. Conversely, image-like 2D mapping approaches [34] emphasize localized spatial patterns but potentially neglect cross-regional functional connectivity. The integration of neuropsychological mechanisms into deep learning architectures presents novel opportunities for enhancing EEG-based mental state decoding. Neurophysiological studies reveal that the human brain, as a prototypical multi-scale complex system, demonstrates hierarchical spatial organization spanning neuronal ensembles, local circuits, and functional zones, coupled with dynamic cross-frequency neural

oscillatory coupling. Notably, during advanced cognitive task execution, information processing involves both region-specific activation of local functional areas and coordinated interactions across distributed brain networks. Consequently, the design of neurophysiologically plausible decoding networks that effectively simulate these multiscale brain activities becomes imperative for advancing WML assessment.

To address these challenges, we propose a wavelet-based cross-frequency and cross-region interaction convolutional neural network (CFCRNet) for WML decoding, as illustrated in Fig.1. Specifically, based on the characteristics of EEG signals, we predefined wavelet convolution kernels for 1D convolution to capture the temporal-frequency characteristics of each electrode during the working memory task. Subsequently, multi-branch cross-scale convolution was employed to achieve dynamic cross-frequency coupling. Drawing on neuroscientific prior knowledge, the brain was divided into distinct functional regions, where nodes within each region were fully connected to reflect intra-regional brain activity. A spatial attention module was then applied to aggregate all nodes within each functional region into a high-level representation, which subsequently interacts with the high-level representations of other brain regions to capture distinct WML patterns. Finally, the aggregated representations from all branches were fused and passed through a fully connected layer for WML detection. By integrating neuroscientific prior knowledge of cross-frequency coupling mechanisms and functional network topology into the deep learning framework, the proposed model effectively enhances the ability to capture differences across various WML levels.

In summary, the main contributions of this study are as follows:

1. We designed an end-to-end neural network model, CFCRNet, for WML detection based on prior knowledge. This model effectively integrates the concepts of cross-frequency and cross-region interactions.

2. We developed a multi-branch cross-scale module to capture cross-frequency information interactions. Additionally, leveraging predefined functional brain regions, we proposed a cross-region feature interaction module based on a spatial attention mechanism to learn high-level representations both within and between brain regions.

3. We validated the proposed model on a laboratory-collected working memory dataset based on the delayed match-to-sample paradigm and a publicly available N-back working memory dataset. The results demonstrated the effectiveness of the proposed model.

2 Datasets and preprocessing

Working memory tasks involve the temporary storage and manipulation of information, with common experimental paradigms including the n-back task and the delayed match-to-sample task (DMT). In this study, we validated the effectiveness of the proposed model using a publicly available dataset based on the n-back paradigm and a self-collected dataset based on the DMT paradigm.

2.1 Delayed Match-to-Sample Task Dataset

A total of 20 healthy participants were recruited for this study (mean age: 21 years; all males; right-handed). All participants had no history of cognitive impairment, psychiatric, or neurological disorders, and had normal or corrected-to-normal vision. This study was approved by the Ethics Committee of the authors' university, and all participants provided written informed consent prior to the experiment. Participants received monetary compensation for their time and effort upon completing the experiment. All experimental procedures were conducted in accordance with the ethical guidelines set forth by the National Ministry of Health, Labor and Welfare and the Declaration of Helsinki (BMJ 1991; 302:1194).

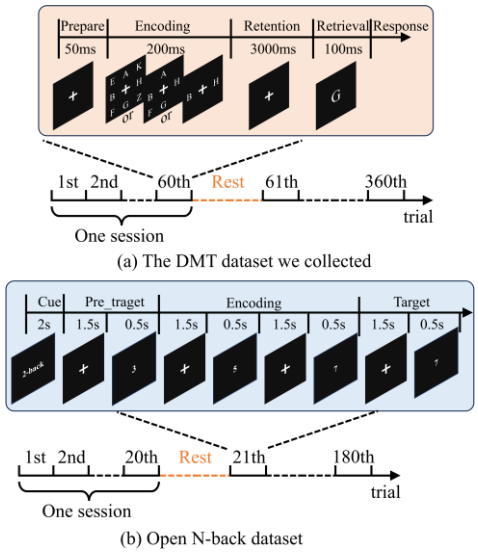


Fig. 2. The working memory experimental paradigms used in this study: (a) Delayed matching to sample paradigm; (b) N-back paradigm.

Throughout the experiment, EEG data were recorded using a 64-channel NeuroScan system (Quik-Cap, band-pass filter range: 0.05–100 Hz, sampling rate: 1000 Hz, impedance < 5 kΩ). E-prime software was employed to present experimental stimuli and record participants' behavioral responses. The experimental task was a DMT consisting of three difficulty levels: low (2 memory items), middle (4 memory items), and high (8 memory items), with a total of six sessions. Each session included 60 trials, with memory load items from the three difficulty levels presented with equal probability. During the experiment, participants were instructed to fixate on a central fixation cross, remain relaxed, minimize blinking, eye movements, and head motion, and respond to the stimuli via button presses. The experimental paradigm is illustrated in Fig. 2(a).

At the beginning of each trial, a central fixation cross flashed for 50 ms, indicating the start of the trial. Subsequently, a memory sequence was presented for 200 ms, during which participants were instructed to memorize the sequence as accurately as

possible. The memory sequence consisted of randomly selected uppercase letters from A to Z and included three difficulty levels presented with equal probability. Following a 3000 ms interval (retention period), a test sequence (probe item) was displayed for 100 ms (retrieval period). Participants were required to determine whether the probe item had appeared in the previously presented memory sequence. If the probe item was absent from the memory sequence, participants were instructed to press the "F" key; if the probe item was present, they were to press the "J" key. Each participant completed a total of 360 trials, with 120 trials assigned to each difficulty level.

2.2 Public N-back Dataset

This dataset includes EEG recordings from 26 participants (9 males; mean age: 26.1 ± 3.5 years; all right-handed). EEG data were collected using 30 electrodes, including Fp1, Fp2, AFF5, AFF6, AFz, F1, F2, FC1, FC2, FC5, FC6, Cz, C3, C4, T7, T8, CP1, CP2, CP5, CP6, Pz, P3, P4, P7, P8, POz, O1, O2, TP9 (reference electrode), and TP10 (ground electrode), with a sampling rate of 1000 Hz. The experimental paradigm is illustrated in Fig. 2(b), and additional experimental details can be found in [35]. All participants were informed about the experimental procedure and provided written informed consent prior to the experiment. This study adhered to the ethical principles outlined in the Declaration of Helsinki and was approved by the Ethics Committee of the Institute of Psychology and Human Factors at Technische Universität Berlin (SH_01_20150330).

2.3 Data Preprocessing

The raw EEG signals from both datasets underwent the following preprocessing steps: 1. Unrelated electrodes were removed (DMT dataset: "CB1," "CB2," "M1," and "M2"), and the data were downsampled to 256 Hz. 2. A band-pass filter of 0.5–45 Hz was applied, and a notch filter was used to eliminate power line noise. 3. Signal segments with noticeable artifacts, particularly those caused by large muscle activity, were visually inspected and discarded. Damaged electrode data were interpolated using spherical spline interpolation. 4. The data were re-referenced using the average reference method. Independent component analysis (ICA) was then employed to identify and remove artifacts such as electromyographic (EMG) and electrooculographic (EOG) noise. 5. Following data cleaning, the EEG signals were segmented to retain only the retention period data (DMT dataset: 3 s; N-back dataset: 2 s), with each epoch including a 0.1 s baseline period preceding the stimulus. 6. Each epoch was baseline-corrected using the baseline period data to reduce baseline drift effects.

3 Methodology

This section first introduces the predefined 1D wavelet kernel function designed to transform the raw EEG signals into spatial-spectral-temporal (SST) representations. Subsequently, the modules for learning cross-frequency and cross-region interactions

are described. Finally, the extracted features are passed through a fully connected layer for load level detection. The overall framework of the proposed model is illustrated in Fig. 1.

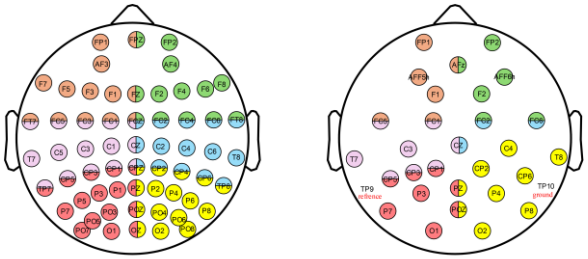
3.1 Spatial-Spectral-Temporal Representation of EEG Signals

The CFCRNet network proposed in this study employs a wavelet-based time convolutional neural network [36] to learn the SST representation for the EEG signals from C electrodes after preprocessing. Specifically, as shown in Fig. 1, the continuous wavelet convolution layer (CWConv) is responsible for extracting SST features from the EEG signals. This convolutional layer uses F wavelet time convolution kernels $\psi_s(t)$ to specifically analyze the corresponding frequency components f in the signal. The mathematical definition is as follows:

$$\psi_s(t) = \frac{1}{\sqrt{s}}\psi(\frac{t-u}{s}), s = \frac{\omega r}{2\pi f} > 0, \quad (1)$$

Here, the real-valued Morlet wavelet $\psi(\cdot)$ [37] is used as the basis function, where u represents the time-domain translation parameter, s is the scale parameter (inversely related to the frequency component f), ω is the central frequency constant of the mother wavelet (used to adjust the balance between time resolution and frequency resolution), and r is the sampling rate of the EEG signal. By convolving the EEG signal from the i-th electrode $X_i \in R^{1\times T}$ with F convolution kernels of different scale parameters, it is transformed into a two-dimensional spectral-temporal representation, denoted as $H_i \in R^{F\times T}$. Specifically, the scale parameters regulate feature extraction through the following mechanism: when s decreases, the wavelet function is compressed in the time domain, and its convolution with the signal effectively captures high-frequency components; conversely, when s increases, the wavelet function is stretched, facilitating the extraction of low-frequency features. Finally, by performing CWConv on the signals from C electrodes, the network constructs an SST feature space $H \in R^{C\times F\times T}$ with discriminative power for classification.

It is important to note that the inherent frequency smoothing properties of wavelet convolution may lead to autocorrelation effects in the time-frequency representation. To reduce the complexity of network training, this study uses predefined wavelet convolution kernels. By initializing the scale parameters of the CWConv layer, the frequency range is set to cover the common cognitive frequency range (1-45 Hz), with an expansion interval of 2 Hz, covering a total of 23 frequencies. This design, by incorporating frequency domain prior knowledge, ensures the completeness of the convolution kernels in the cognitively relevant frequency bands while effectively guiding the network to focus on physiologically meaningful frequency bands.



(a) (b)

Fig. 3. Predefined brain regions in this study. (a) DMT dataset, (b) Nback dataset.

3.2 Cross-Frequency Feature Interaction Module

Previous studies in cognitive neuroscience have shown that high-level cognitive tasks (such as working memory and emotion) involve cross-frequency coupling at different frequencies [38]. Based on this, the present study proposes a cross-frequency feature interaction module (CFFI), as shown in Fig. 1, specifically designed to extract cross-scale frequency interaction information from the SST representation data extracted by CWConv. This module utilizes a multi-branch structure, with each branch having a unique frequency receptive field, allowing it to capture frequency interactions across different ranges and achieve cross-frequency information coupling. In this study, multiple branches with receptive fields ranging from narrow to wide frequencies are constructed to cover local frequency interactions and global frequency feature coupling in the SST data. To preserve the spatial structure, the cross-frequency feature interaction is implemented through separable convolutions in each branch, meaning that cross-frequency interactions are independently performed at each electrode.

$$Y_{i,f} = \sum_{j=1}^{k} W_{i,j} \cdot X_{i,f-j+1} \tag{2}$$

where X and Y represent the input and output features, respectively, $W_{i,j}$ is the j-th convolution kernel parameter of the i-th electrode, with a size of k, and f is the frequency step. In this study, four branches are used, with convolution kernel sizes set as (1,1), (5,1), (9,1), and (13,1), where the first dimension corresponds to the direction of the frequency convolution.

3.3 Cross-Region Feature Interaction Module

In the cross-frequency feature interaction module, each branch performs cross-region feature interaction (CRFI). In this study, based on the anatomical structure of the cerebral cortex, the brain is divided into six regions (frontal lobe, central region, occipital-parietal region), as shown in Fig. 3. Each brain region is responsible for different cognitive functions, yet they cooperate to accomplish cognitive tasks. Therefore, we hypothesize that the features extracted from the electrodes in each brain region are similar and contain redundancy. To address this, we apply the proposed spatial attention mechanism to extract high-level representations of each brain region for dimensionality

reduction, effectively learning a virtual electrode to represent the information from all electrodes within a given brain region. We obtain six sets of high-level representations to describe the entire brain's information, and these sets are then cascaded to achieve information interaction and fusion. We assume that the input features for the CRFI module are denoted as $X \in R^{C\times F\times T}$, and we first partition them into six brain regions, as shown in Fig. 3.

$$X_s \in R^{C_s\times F\times T}, (s=1,2,\cdots,6; C_1+C_2+\cdots+C_6=C) \tag{3}$$

Then, a spatial attention mechanism (SA) is applied to each brain region, and the specific process is as follows: First, global max pooling and global average pooling are performed on X_s along the spatial dimension, yielding two feature maps of size $X_{s,mean}, X_{s,max} \in R^{1\times F\times T}$. Next, the results of the global max pooling and global average pooling are concatenated along the channel dimension, resulting in a feature map of size $X_{s,1} \in R^{2\times F\times T}$. A 1×1 convolution operation is applied to the concatenated result to reduce the dimensionality, obtaining a feature map of size $X_{s,2} \in R^{1\times F\times T}$. Finally, the Sigmoid activation function is applied to generate the spatial attention weight matrix $M(X_s)$, with the following calculation formula:

$$\begin{aligned} M(X_s) &= \sigma(f^{1\times1}[AvgPool(X_s); MaxPool(X_s)])) \\ &= \sigma(f^{1\times1}[X_{s,mean}; X_{s,max}]) \end{aligned} \tag{4}$$

where AvgPool and MaxPool represent the max and average pooling operations performed along the spatial electrode dimension, $f^{1\times1}(\cdot)$ is the convolution operation, and $\sigma(\cdot)$ is the Sigmoid function.

Finally, the spatial attention weight matrix $M(X_s)$ is applied to the input features X_s, and the entire region's information is represented by performing an averaging operation.

$$Y = mean(X_s \cdot M(X_s)), Y \in R^{1\times F\times T} \tag{5}$$

Similarly, the same cross-region feature interaction operation is performed in each branch.

3.4 Classification block

The features obtained from each branch are fused and passed to the classification block to determine the sample's category. Specifically, the features from each branch are first merged to form a combined feature map $Y_{all} \in R^{4\times F\times T}$. Then, the proposed SA mechanism is applied to extract global information from each fused scale. After that, all feature maps are flattened into a one-dimensional vector and fed into a fully connected layer. The output of the fully connected layer is then passed through a softmax function to convert it into category probabilities. Finally, the category label with the highest probability is selected as the network's final classification result.

3.5 Experimental Setup

The model training in this study adopts the following optimization strategy: the Adam optimizer (with an initial learning rate of 1×10^{-3} is used to minimize the cross-entropy loss function. Additionally, ℓ_2 regularization is applied to constrain the model

complexity, and a cosine annealing learning rate scheduler is integrated to implement dynamic learning rate decay. The batch size is set to 32. For the continuous wavelet convolution layer (CWConv), the core parameter configuration is as follows: the wavelet kernel time domain length s = 150 samples, the translation parameter u=0, and the central frequency constant ω=5.15.

In terms of the validation scheme, a within-subject ten-fold cross-validation strategy is used for the DMT dataset. Previous related studies [39] have suggested that EEG experimental decoding based on block design might be influenced by block effects. Therefore, in this study, a "leave-one-session-out" cross-validation is adopted for the open N-back dataset to demonstrate the effectiveness of our model.

4 Results and Discussion

This section primarily validates the effectiveness of the proposed model through comparisons with other models and ablation experiments. Standard evaluation metrics are used to assess the model's performance in detecting WML levels, including AUC, Macro_F1 (MF1), Macro_Precision (MP), Macro_Sensitivity (MS), and Accuracy (ACC, %).

Table 1: Comparison of the proposed model with other models on the DMT dataset.

Model	ACC	AUC	MF1	MP	MS
ShallowNet	71.16±5.56	56.70±8.33	70.81±5.63	72.19±5.56	71.22±5.61
DeepNet	71.55±5.77	57.26±8.61	71.26±5.77	72.55±5.85	71.54±5.75
EEGNet	67.26±6.58	50.58±9.84	66.76±6.52	68.31±6.54	67.31±6.60
LMDA-Net	69.51±6.39	54.23±9.54	69.21±6.56	70.56±6.46	69.56±6.41
JMNet	70.84±6.44	56.19±9.64	70.33±6.65	72.07±6.46	70.85±6.48
LGGNet	64.44±5.49	46.56±8.19	63.64±5.67	65.48±5.81	64.45±5.53
EEGNex	73.15±5.10	59.68±7.62	72.76±5.20	74.25±5.10	73.18±5.13
CFCRNet	78.08±5.54	66.99±8.31	77.77±5.65	78.94±5.44	78.05±5.56

4.1 Comparison with Other Models

In this study, the proposed algorithm is compared with several commonly used deep learning models for EEG data classification. The commonly used EEG data classification models include ShallowNet and DeepNet [32], EEGNet [29], LMDA-Net [40], JMNet [41], LGGNet [42], and EEGNex [43]. These models have been applied to various brain-computer interface paradigms such as P300, movement-related cortical potentials, sensorimotor rhythms, and other tasks. Below is a brief description of each model:

ShallowNet and DeepNet (2017): These models consist of a series of convolutional and pooling layers designed to decode task-related information from EEG signals. ShallowNet is a simpler model with fewer layers, while DeepNet employs deeper architectures for more complex feature extraction.

EEGNet (2018): EEGNet is a compact CNN for EEG analysis. It utilizes deep convolutions and depthwise separable convolutions to construct an EEG feature extraction model.

LMDA-Net (2023): LMDA-Net is a lightweight decoding network based on a multi-dimensional attention mechanism. It includes a baseline network for EEG feature extraction and classification, a channel attention module that expands the spatial information of the input EEG across the depth dimension using tensor products, and a deep attention module to promote high-dimensional feature interaction.

JMNet (2024): JMNet introduces a multi-branch attention CNN framework for decoding EEG signals. This model automatically generates task-specific EEG signal representations within a particular subject's sub-band and effectively extracts local and global time-frequency features through its multi-branch structure.

LGGNet (2024): LGGNet is a new neural-inspired graph neural network (GNN). It consists of a series of multi-scale one-dimensional convolution kernels in the time convolution module and a kernel-level attention fusion mechanism. This network is designed to capture the temporal dynamic features of EEG signals. These features are then input into the designed local and global graph filtering layers. The model uses a set of locally and globally meaningful graphs to model the complex interactions within the brain.

EEGNex (2024): EEGNex uses EEGNet as the backbone network and introduces three modules to enhance the extraction of spatial representation information from EEG input signals. It also replaces depthwise separable convolutions with dilated convolutions to increase the receptive field of the model.

These models have been widely used in brain-computer interface applications and are selected as baseline models for comparison with the proposed CFCRNet.

4.2 Results on the DMT dataset.

The comparison results between the proposed model and other baseline algorithms are shown in Table 1 and Fig. 4. Specifically, from Table 1, it can be seen that the CFCRNet model outperforms all other models in all evaluation metrics. For example, CFCRNet achieves higher accuracy than the most advanced baseline model, EEGNex, by 4.93% (78.08% vs. 73.15%). This highlights the effectiveness of cross-frequency and cross-region interactions in working memory tasks. This aligns with previous studies [38], which suggest that completing higher-order cognitive tasks (such as working memory) involves cross-frequency coupling and information interactions between different brain regions. The network model designed in this study is more consistent with such cognitive processes.

Furthermore, to better analyze the model's classification performance at different load levels, we compared the classification accuracy at various load levels using confusion matrices, as shown in Fig. 4. The results indicate that most models show higher accuracy for classifying low-load levels, while significant confusion occurs between middle and high load levels, especially at the middle load level, where classification accuracy is the lowest. Often, models confuse middle and high load levels. However, our model alleviates this issue to some extent. Still, at the medium load level, there is a

17.31% confusion with high load levels. This could be due to the fact that the cognitive resources required for medium and high load levels are very similar, making it challenging to distinguish them using EEG data.

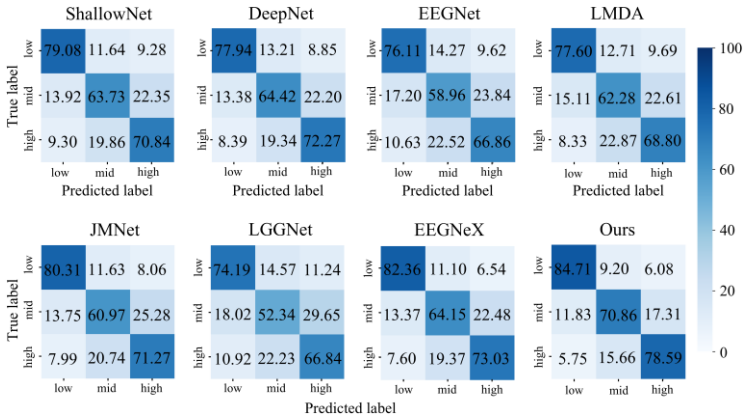


Fig. 4. Confusion matrix results of CFCRNet compared with other algorithms on the DMT dataset.

4.3 Results on the N-back dataset

To validate the generalization capability of the model, we conducted comparative experiments on the public N-back dataset. Previous research has shown that EEG experiment decoding based on block designs may be affected by block effects. Therefore, we adopted leave-one-block cross-validation to avoid this issue. The comparison results are shown in Table 2. Similar to the results on the DMT dataset, the proposed algorithm still achieves the best performance across all evaluation metrics. Specifically, in terms of accuracy, our model outperforms the best comparative algorithm by 2.88% (70.35% vs. 67.47%), and it achieves a 3.11% improvement on the MF1 metric. These results further suggest that the CFCRNet proposed in this study is more in line with the cognitive processes involved in working memory.

Table 2: Comparison of the proposed model with other models on the N-back dataset.

Model	ACC	AUC	MF1	MP	MS
ShallowNet	63.79±6.03	45.59±8.98	62.51±6.10	63.97±6.24	63.54±5.84
DeepNet	60.65±5.59	40.85±8.25	58.85±5.66	61.50±5.33	60.40±5.30
EEGNet	64.12±5.54	46.09±8.24	63.07±5.54	64.57±5.36	63.86±5.37
LMDA-Net	64.58±5.86	46.73±8.70	63.33±5.85	65.21±5.56	64.25±5.58
JMNet	63.06±5.22	44.38±7.83	61.18±5.47	64.52±4.59	62.67±5.10
LGGNet	61.78±5.56	42.54±8.29	60.00±6.11	62.89±5.21	61.44±5.45
EEGNex	67.47±4.84	51.08±7.23	66.18±4.95	68.08±4.77	67.13±4.72
CFCRNet	70.35±4.81	55.44±7.19	69.29±5.00	71.36±4.64	70.11±4.65

As before, the comparison results of the confusion matrix are shown in Fig. 5. In almost all algorithms, the distinction between the middle-load and high-load levels (corresponding to 2-back and 3-back in the experiment) is difficult, while the low-load level is easily separable. This could be due to the fact that the cognitive resources required for the two load levels are very close. In our daily experience, it is also noticeable that there is no significant difference in terms of both accuracy and response time between 2-back and 3-back tasks, which is why they are difficult to distinguish. This might also explain why the performance difference between the proposed algorithm and other algorithms on the N-back dataset is smaller than that on the DMT dataset.

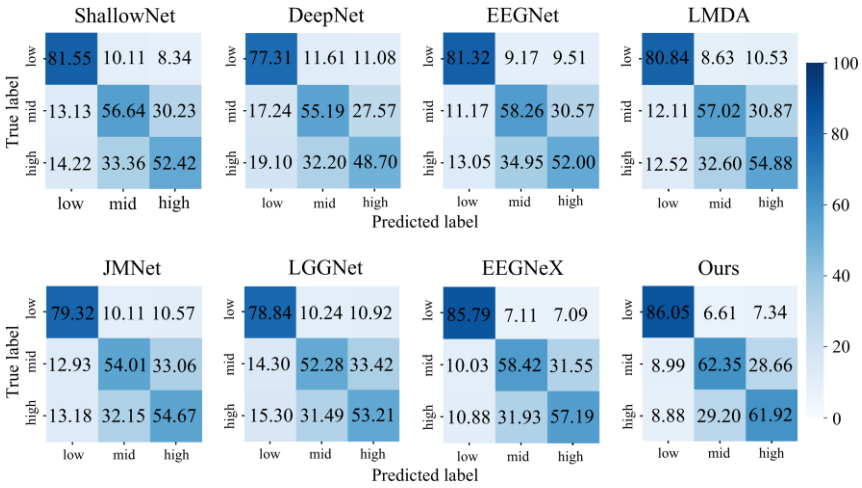


Fig. 5. Confusion matrix results of CFCRNet compared with other algorithms on the N-back dataset.

4.4 Ablation Study

To verify the indispensability of the individual modules proposed in this study, an ablation study was conducted on the DMT dataset. The ablation experimental design is as follows:

Base: This is the baseline model of this study, where cross-frequency feature interaction is replaced by a 1×1 convolution layer, while maintaining the multi-branch structure. The entire brain region is treated as a single region, with the rest of the structure unchanged.

Base+CFFI: This model adds the cross-frequency feature interaction module to the Base model.

Base + CRFI: This model adds the cross-brain-region feature fusion module to the Base model.

The experimental comparison results are shown in Fig. 6. We observed that both the CFFI and CRFI modules lead to significant improvements over the Base model, especially when both modules are integrated into the network. The performance of CRCFNet, with both modules, was enhanced by 5.49% (78.02% vs. 72.53%). This

indicates that the network achieves optimal performance only when both cross-frequency information interaction and cross-brain-region information interaction are simultaneously applied. This further supports previous research, which suggests that high-level cognitive tasks (such as working memory) involve cross-frequency coupling and information interaction between different brain regions.

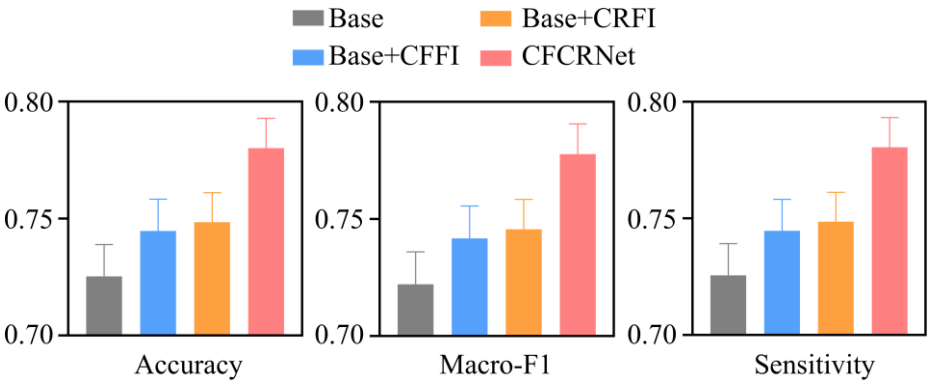


Fig. 6. Ablation experiment results on the DMT dataset

5 Conclusion

This study designs a cross-frequency and cross-region interaction convolutional neural network for WML decoding, named CFCRNet, based on the cognitive process of working memory. Specifically, CFCRNet first extracts SST features using a custom wavelet kernel via 1D convolution. Then, it employs a multi-branch structure to learn cross-frequency features coupling over different frequency ranges. Finally, the network uses a spatial attention mechanism to learn cross-brain region information interaction, achieving local and global feature fusion. CFCRNet integrates cross-frequency coupling mechanisms with cross-brain region functional integration principles, constructing a network model capable of analyzing dynamic characteristics of brain signals in different frequency bands and complex interactions both within and between functional regions. We validated the model through experiments on a working memory dataset collected in our lab, a public dataset, and ablation studies. The results demonstrate that incorporating neuroscientific prior knowledge into neural network design improves classification performance. In conclusion, our findings provide an advanced framework for accurately detecting working memory load, and the approach can be extended to exploring detection tasks related to other cognitive behaviors and diseases.

Acknowledgments. This work was supported by the Chongqing Natural Science Fund Innovation and Development Joint Fund (Municipal Education Commission) project under Grant CSTB2024NSCQ-LZX0121; and the funding of Institute for Advanced Sciences of Chongqing University of Posts and Telecommunications under Grant E011A2022327.

Disclosure of Interests. authors have no competing interests.

References

1. Durantin, G., Gagnon, J. F., Tremblay, S., Dehais, F. Using near infrared spectroscopy and heart rate variability to detect mental overload. Behavioural brain research 259, 16-23 (2014).
2. Arvidsson, H., Larsson, G., Larsolle, A., Neely, G., Hansson, P. A. Easily applicable methods for measuring the mental load on tractor operators. Journal of agricultural safety and health 26(1), 5-14 (2020).
3. Antonenko, P., Paas, F., Grabner, R., Van Gog, T. Using electroencephalography to measure cognitive load. Educational psychology review 22, 425-438 (2010).
4. Mazher, M., Abd Aziz, A., Malik, A. S., & Amin, H. U. An EEG-based cognitive load assessment in multimedia learning using feature extraction and partial directed coherence. IEEE Access 5, 14819-14829 (2017).
5. Almogbel, M. A., Dang, A. H., Kameyama, W. Cognitive workload detection from raw EEG-signals of vehicle driver using deep learning. In: 21st International Conference on Advanced Communication Technology, pp. 1-6. IEEE, Phoenix Park, Korea (2019).
6. Dehais, F., Duprès, A., Blum, S., Drougard, N., Scannella, S., Roy, R. N., Lotte, F. Monitoring pilot's mental workload using ERPs and spectral power with a six-dry-electrode EEG system in real flight conditions. Sensors 19(6), 1324 (2019).
7. Mathan, S., Smart, A., Ververs, T., Feuerstein, M. Towards an index of cognitive efficacy EEG-based estimation of cognitive load among individuals experiencing cancer-related cognitive decline. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, pp. 6595-6598. IEEE, Buenos Aires, Argentina (2010).
8. Wu, W., Zhang, Y., Jiang, J., Lucas, M. V., Fonzo, G. A., Rolle, C. E., Etkin, A. An electroencephalographic signature predicts antidepressant response in major depression. Nature biotechnology 38(4), 439-447 (2020).
9. Zhang, L., Wade, J., Bian, D., Fan, J., Swanson, A., Weitlauf, A., Sarkar, N. Cognitive load measurement in a virtual reality-based driving system for autism intervention. IEEE transactions on affective computing 8(2), 176-189 (2017).
10. Wang, S., Gwizdka, J., Chaovalitwongse, W. A. Using wireless EEG signals to assess memory workload in the $ n $-back task. IEEE Transactions on Human-Machine Systems 46(3), 424-435 (2015).
11. Heard, J., Harriott, C. E., Adams, J. A. A survey of workload assessment algorithms. IEEE Transactions on Human-Machine Systems 48(5), 434-451 (2018).
12. Hart, S. G., Staveland, L. E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Advances in psychology, North-Holland (1988).
13. Reid, G. B., Nygren, T. E. The subjective workload assessment technique: A scaling procedure for measuring mental workload. In Advances in psychology. North-Holland (1988).
14. Tao, D., Tan, H., Wang, H., Zhang, X., Qu, X., Zhang, T. A systematic review of physiological measures of mental workload. International journal of environmental research and public health 16(15), 2716 (2019).
15. Hu, L., Tan, C., Xu, J., Qiao, R., Hu, Y., Tian, Y. Decoding emotion with phase–amplitude fusion features of EEG functional connectivity network. Neural Networks 172, 106148 (2024).
16. Xu, J., Qian, W., Hu, L., Liao, G., Tian, Y. EEG decoding for musical emotion with functional connectivity features. Biomedical Signal Processing and Control 89, 105744 (2024).
17. Chakladar, D. D., Roy, P. P., Iwamura, M. EEG-based cognitive state classification and analysis of brain dynamics using deep ensemble model and graphical brain network. IEEE Transactions on Cognitive and Developmental Systems 14(4), 1507-1519 (2021).

18. Chakladar, D. D., Samanta, D., Roy, P. P. Multimodal deep sparse subspace clustering for multiple stimuli-based cognitive task. In: 26th International Conference on Pattern Recognition, pp. 1098-1104. IEEE, Haikou, China (2022).
19. Chai, R., Naik, G. R., Nguyen, T. N., Ling, S. H., Tran, Y., Craig, A., Nguyen, H. T. Driver fatigue classification with independent component by entropy rate bound minimization analysis in an EEG-based system. IEEE journal of biomedical and health informatics 21(3), 715-724 (2016).
20. Brouwer, A. M., Hogervorst, M. A., Van Erp, J. B., Heffelaar, T., Zimmerman, P. H., Oostenveld, R. Estimating workload using EEG spectral power and ERPs in the n-back task. Journal of neural engineering 9(4), 045008 (2012).
21. Mühl, C., Jeunet, C., Lotte, F. EEG-based workload estimation across affective contexts. Frontiers in neuroscience, 8, 114 (2014).
22. Zarjam, P., Epps, J., Chen, F. Characterizing working memory load using EEG delta activity. In: 19th European signal processing conference, pp. 1554-1558. IEEE, Barcelona (2011).
23. Dijksterhuis, C., De Waard, D., Brookhuis, K. A., Mulder, B. L., de Jong, R. Classifying visuomotor workload in a driving simulator using subject specific spatial brain patterns. Frontiers in neuroscience 7, 149 (2013).
24. Zarjam, P., Epps, J., Chen, F., Lovell, N. H. Estimating cognitive workload using wavelet entropy-based features during an arithmetic task. Computers in biology and medicine 43(12), 2186-2195 (2013).
25. Kok, A. On the utility of P3 amplitude as a measure of processing capacity. Psychophysiology 38(3), 557-577 (2001).
26. Tian, Y., Zhang, H., Jiang, Y., Li, P., Li, Y. A fusion feature for enhancing the performance of classification in working memory load with single-trial detection. IEEE Transactions on Neural Systems and Rehabilitation Engineering 27(10), 1985-1993 (2019).
27. Jiao, Z., Gao, X., Wang, Y., Li, J., Xu, H. Deep convolutional neural networks for mental load classification based on EEG data. Pattern Recognition 76, 582-595 (2018).
28. Zhang, P., Wang, X., Zhang, W., Chen, J. Learning spatial–spectral–temporal EEG features with recurrent 3D convolutional neural networks for cross-task mental workload assessment. IEEE Transactions on neural systems and rehabilitation engineering 27(1), 31-42 (2018).
29. 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 15(5), 056013 (2018).
30. Bashivan, P., Rish, I., Yeasin, M., Codella, N. Learning representations from EEG with deep recurrent-convolutional neural networks. arXiv preprint arXiv:1511.06448 (2015).
31. Zhang, Y., Cai, H., Nie, L., Xu, P., Zhao, S., Guan, C. An end-to-end 3D convolutional neural network for decoding attentive mental state. Neural Networks 144, 129-137 (2021).
32. Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., Ball, T. Deep learning with convolutional neural networks for EEG decoding and visualization. Human brain mapping 38(11), 5391-5420 (2017).
33. Ding, Y., Robinson, N., Zeng, Q., Chen, D., Wai, A. A. P., Lee, T. S., Guan, C. Tsception: a deep learning framework for emotion detection using EEG. In: 2020 international joint conference on neural networks, pp. 1-7. IEEE, Glasgow, UK (2020).
34. Jung, T. P., Sejnowski, T. J. Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing. IEEE Transactions on Affective Computing 13(1), 96-107 (2022).
35. Shin, J., Von Lühmann, A., Kim, D. W., Mehnert, J., Hwang, H. J., Müller, K. R. Simultaneous acquisition of EEG and NIRS during cognitive tasks for an open access dataset. Scientific data 5(1), 1-16 (2018).

36. Li, T., Zhao, Z., Sun, C., Cheng, L., Chen, X., Yan, R., Gao, R. X. WaveletKernelNet: An interpretable deep neural network for industrial intelligent diagnosis. IEEE Transactions on Systems, Man, and Cybernetics: Systems 52(4), 2302-2312 (2021).
37. Liu, W., Tang, B. A hybrid time-frequency method based on improved Morlet wavelet and auto terms window. Expert Systems with Applications 38(6), 7575-7581 (2011).
38. Daume, J., Gruber, T., Engel, A. K., Friese, U. Phase-amplitude coupling and long-range phase synchronization reveal frontotemporal interactions during visual working memory. Journal of Neuroscience 37(2), 313-322 (2017).
39. Li, R., Johansen, J. S., Ahmed, H., Ilyevsky, T. V., Wilbur, R. B., Bharadwaj, H. M., Siskind, J. M. The perils and pitfalls of block design for EEG classification experiments. IEEE Transactions on Pattern Analysis and Machine Intelligence 43(1), 316-333 (2020).
40. Miao, Z., Zhao, M., Zhang, X., Ming, D. LMDA-Net: A lightweight multi-dimensional attention network for general EEG-based brain-computer interfaces and interpretability. NeuroImage 276, 120209 (2023).
41. Kim, J. M., Heo, K. S., Shin, D. H., Nam, H., Won, D. O., Jeong, J. H., Kam, T. E. A learnable continuous wavelet-based multi-branch attentive convolutional neural network for spatio–spectral–temporal EEG signal decoding. Expert Systems with Applications 251, 123975 (2024).
42. Ding, Y., Robinson, N., Tong, C., Zeng, Q., Guan, C. Lggnet: Learning from local-global-graph representations for brain–computer interface. IEEE Transactions on Neural Networks and Learning Systems 35(7), 9773-9786 (2024).
43. Chen, X., Teng, X., Chen, H., Pan, Y., Geyer, P. Toward reliable signals decoding for electroencephalogram: A benchmark study to EEGNeX. Biomedical Signal Processing and Control, 87, 105475 (2024).