Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Multi-Task EEG Model Based on Wavelet Convolution for Virtual Reality Motion Sickness Detection

Chengcheng Hua
C-IMER, CICAEET, Nanjing University of Information Science & Technology
Qiuyu Wang
C-IMER, CICAEET, Nanjing University of Information Science & Technology
Min Zhu
C-IMER, CICAEET, Nanjing University of Information Science & Technology
Jingqi Fu
C-IMER, CICAEET, Nanjing University of Information Science & Technology
Ying Yan
C-IMER, CICAEET, Nanjing University of Information Science & Technology
Dapeng Chen
C-IMER, CICAEET, Nanjing University of Information Science & Technology
Jia Liu
C-IMER, CICAEET, Nanjing University of Information Science & Technology
Rongrong Fu
frr1102@aliyun.com

Yanshan University

Research Article

Keywords: Virtual Reality Motion Sickness (VRMS), Electroencephalography (EEG), wavelet convolution, attention mechanism, multi-task learning

Posted Date: June 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9897424/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Multi-Task EEG Model Based on Wavelet Convolution for Virtual Reality Motion Sickness Detection

Chengcheng Hua[1], Qiuyu Wang[1], Min Zhu[1], Jingqi Fu[1], Ying Yan[1], Dapeng Chen[1], Jia Liu[1], Rongrong Fu[2*]

[1]School of Automation, C-IMER, CICAEET, Nanjing University of Information Science & Technology, Nanjing 210044, China

[2]Measurement Technology and Instrumentation Key Lab of Hebei Province, Department of Electrical Engineering, Yanshan University, Qinhuangdao 066000, China

Corresponding author: Rongrong Fu(e-mail: frr1102@aliyun.com)

Abstract: Virtual Reality Motion Sickness (VRMS) has become the primary bottleneck hindering the widespread adoption of VR technology, and the precise detection of the VRMS is a prerequisite for studying the causes and treatments of it. This paper proposes an Electroencephalography (EEG) detection model named as WTConvNet, that integrates wavelet convolution, time convolution and lead convolution to extract the multiple scales and domains of features from EEG. And a multi-task framework is introduced to learn the shared features across the different tasks, thereby enhancing the robustness of feature extraction. The experiment utilized a self-built dataset, which includes the EEGs collected from the 22 participants. In this dataset, a VR simulated flight task is introduced to induce the VRMS of the participants, and an adaptive training stage is used to change the VRMS severity. This paper validates the performance of the proposed model on the multi-task combinations of five different recognition tasks on this dataset. The results show that the highest accuracies are achieved in dual-task learning for the VRMS severity classifications in the task state before and after the training stage (*Task1*) and in the two resting states before and after the VRMS-induced task before the training stage (*Task3*), at 99.57% and 97.65%, respectively, representing improvements of 0.56% and 2.17% over the single-task learning. The research results confirm that the model combines efficiency and reliability in VRMS detection, providing a potential solution for aiding to optimize user experience and accelerating the adoption of VR technology.

Keywords: Virtual Reality Motion Sickness (VRMS), Electroencephalography (EEG), wavelet convolution, attention mechanism, multi-task learning

1. Introduction

1.1. Background

Virtual reality (VR) technology is widely used in entertainment, healthcare, education, etc. However, it often causes Virtual Reality Motion Sickness (VRMS) when locomotions happen to the users in VR scenes, which reduces users' experience [1-3] and severely affects the promotion and popularization of VR technology. Therefore, it is an urgent issue needed to be addressed. Electroencephalography (EEG) is a high-time-resolution, low-cost, non-invasive neuroimaging technique that provides objective physiological indicators of the brain states unaffected by subjective consciousness. By analyzing EEG, it may be possible to identify neural pattern changes associated with the VRMS onset, enabling the early detection. Deep learning-based EEG analysis methods can automatically capture the neural activity patterns associated with the VRMS, offering a potential effective means for its detection. Additionally, these studies can provide technical

support for researches aimed at alleviating the VRMS, optimizing VR system design, and thereby driving the further development and widespread application of VR technology.

1.2. Related Work

Subjective questionnaires and objective physiological signals are currently the primary methods for detecting VRMS [4]. However, subjective questionnaires are susceptible to individual cognitive biases, leading to lower reliability. More importantly, subjective questionnaires can not reflect the changes of VRMS level in real time [5]. In contrast, objective physiological signals offer significant advantages, such as real-time, objective, non-interfering, etc. Among various physiological signals, EEG have gained significant attention due to its rich physiological and pathological information. This information holds great value in neurology, clinical testing, and brain-computer interface researches, offering a potential approach for the VRMS detection. Yang et al. [6] utilized a random forest algorithm to detect the VRMS using EEG, achieving a maximum accuracy of 84%. Hua et al. [4] used the absolute values of the differences in entropy between the interhemispheric electrodes and the differences in cross-frequency couplings of EEG as the features to detect the VRMS, achieving a classification accuracy of 99.5% using SVM. Demirel et al. [7] proposed a multitaper spectrum estimation method based on EEG, combining head motion signals with users' passively monitored EEG to achieve real-time continuous detection of the VRMS, with an accuracy of 76.83%. This indicates that numerous features in EEG are associated with VRMS. However, the effective features should be selected manually, which cost many time. And those manual features may fail in the different experimental conditions such as: different VR scenes, populations, motion types, etc. because the neural pattern may change with those conditions, which prompts researchers to explore more efficient methods.

Compared with the two-stage methods combining machine learning and manual feature, deep learning methods have gained widespread application in the EEG field due to their superior performance and automatic feature extraction capabilities. In addition to the classic models such as EEGNet [8], DeepConvNet [9] and ShallowConvNet [9], new models are continuously emerging. Bang et al. [10] utilized EEG collected from a real driving experiment, employing the normalized sample covariance matrices combined with Convolutional Neural Networks (CNNs) to establish a motion sickness prediction model, achieving an average accuracy of 89.05%. Wang et al. [11] proposed a new algorithm called SBLEST based on sparse Bayesian learning, which optimizes the spatio-temporal filters and classifiers and combines them with low-rank weight matrices, significantly improving the accuracy of EEG decoding across multiple EEG datasets. Li et al. [12] use self attention in the frequency band to learn the frequency information of each frequency band, and inter band mapping to learn the complementary frequency information, and use the parallel convolutional neural network (PCNN) layer to mine the spatial information of EEG signals to achieve efficient emotion recognition. Deep learning methods have achieved good results in EEG decoding in multiple fields, and hold potential for detecting VRMS.

EEG exhibit complex characteristics of non-stationarity and multi-frequency bands. As an important tool in signal processing, the Discrete Wavelet Transform (DWT) can decompose EEG into different frequency components, in which some components imply the features related to the VRMS. Traditionally, the DWTs have been primarily used for EEG denoising [13, 14] and frequency feature extraction [15-17]. Now, a new method combining DWT with CNN is proposed, which is called Wavelet Convolution Unit(WCU) . By incorporating wavelet analysis into a standard convolutional neural network, the WCU achieves a non-local receptive field and cross-scale feature fusion, thereby overcoming key limitations of traditional CNNs in medical image analysis. Chen [18] et al. proposed a novel pedestrian trajectory prediction model named WTGCN,

which effectively captures the spatio-temporal interaction features of pedestrian trajectories by combining DWTs and graph convolutional networks, achieving excellent prediction performance across multiple datasets. These successful cases demonstrates the advantages of wavelet convolution in multi-scale feature extraction. EEG also implies the multi-scale characteristics, and wavelet transforms have long been applied in EEG feature extraction. Therefore, wavelet convolution holds promise for EEG decoding. However, the existing wavelet convolution methods are designed for 2D data, whereas EEG is a one-dimensional time signal. To adapt wavelet convolution for EEG, both the DWT and the CNN must be modified to one-dimensional versions. Based on this, this study introduces the 1D wavelet convolution to extract the time-frequency features with both physical meaning and discriminative power from the EEG for the subsequent classification.

Currently, VRMS detection tasks are primarily categorized into task-based and resting-state categories. However, the existing researches typically focuses solely on a single one of them, failing to effectively achieve cross-task feature transfer and fusion. To overcome this limitation, multi-task learning can be introduced into the VRMS detection. Multi-task learning is increasingly attracting attention in the EEG analysis. Jia et al. [19] proposed a dual-task model with two tasks: emotion recognition and cognitive recognition, which demonstrated the superior performance compared to the single-task learning models on both the ECED and DEAP datasets. He et al. [20] proposed an end-to-end multi-modal multi-task model (M2NN) that integrates EEG and fNIRS signals, incorporating spatio-temporal feature extraction, multi-modal feature fusion, and multi-task learning modules. Compared to multitask unimodal EEG model, multitask unimodal HbR models, and multimodal single-task models, M2NN achieved improved classification accuracy. This provides a new idea for the VRMS detection. We attempted to decode the VRMS-related EEG under the different conditions: the task states, resting states, high/low VRMS, and with/without VRMS, and then using a multi-task learning framework to adopt the advantages of it in the multi-conditional VRMS detection. By learning the multiple related tasks simultaneously, a single model can share and leverage the common information and knowledge across the different tasks, thereby improving overall performance and generalization abilitys.

This study aims to propose a novel multi-task EEG model based on wavelet convolution for the efficient detection of the VRMS. The model first employs DWT and time convolution modules to perform multi-scale composite filtering on EEG. It then uses a time convolution module and a lead convolution module with attention mechanisms of the two dimmensions to further extract features from the filtered EEG. Finally, a multi-task approach is adopted to leverage the data correlations between the different tasks, thereby enhancing the model's robustness, generalisation ability and accuracy. The model's effectiveness is validated on an existed VRMS-related EEG dataset. The proposed method provide a theoretical basis and technical support for optimising VR technology and enhancing user experience in the future. The main contributions of this study are as follows:

1. WTConvNet based on wavelet convolution is proposed. Combining a time convolutional attention module and a lead convolutional attention module, the proposed model can effectively extract multi-scale spatio-temporal features from EEG and achieve efficient detection of the VRMS.

2. The study identifies five different VRMS detection tasks according to the different conditions such as: the task states, resting states, high/low VRMS, and with/without VRMS to make sure the EEG model can learn the VRMS-realated EEG features and avoid the interfare from the irrelated factors.

3. By introducing a multi-task framework to decode the VRMS-related EEG of the different tasks, the shared features across tasks are revealed, enhancing the model's classification accuracy and robustness, thereby providing an efficient and reliable solution for detecting the VRMS.

2. Materials and Methods

2.1. Dataset

2.1.1. Participants and Experimental Protocol

This study involved an existing VRMS-related EEG dataset [21]. A total of 22 participants were recruited for the experiment, including 16 males and 6 females, aged between 19 and 28 years old. The purpose of the experiment was to induce the VRMS through a specific VR task and collect the related EEG. The EEG was collected by Neuroscan Greal at a sampling rate of 1024 Hz. The EEG cap had 30 electrodes (FP1, FP2, F11, F7, F3, FZ, F4, F8, F12, FT11, FC3, FCZ, FC4, FT12, T7, C3, CZ, C4, T8, CP3, CPZ, CP4, P7, P3, PZ, P4, P8, O1, OZ, and O2) as well as two reference electrodes on the mastoids behind the ears. The experimental procedure is shown in Fig. 1. There were two phases: pre-training and post-training. In the two phases, the participants wore the EEG cap and VR headset to perform a VRMS exposure task in a standing position. The participants needed to manoeuvre an aircraft through seven targets located at different positions with the task duration set at five minutes. To ensure all the participants received the same stimulus exposure time, the task would start over if it was completed early, or it would end directly if it was not completed within five minutes. During the task, the task-state EEG (task) was collected, and before and after the task, two five-min resting-state EEG (rest1 and rest2) were collected respectively. Before the rest1 and after the rest2, the participants fill in the Virtual Reality Sickness Questionnaire (VRSQ)[22] to assess the VRMS severity. Between the two phases, a seven-day training was conducted, every participants need to be trained on the same task for 30 minutes every day.

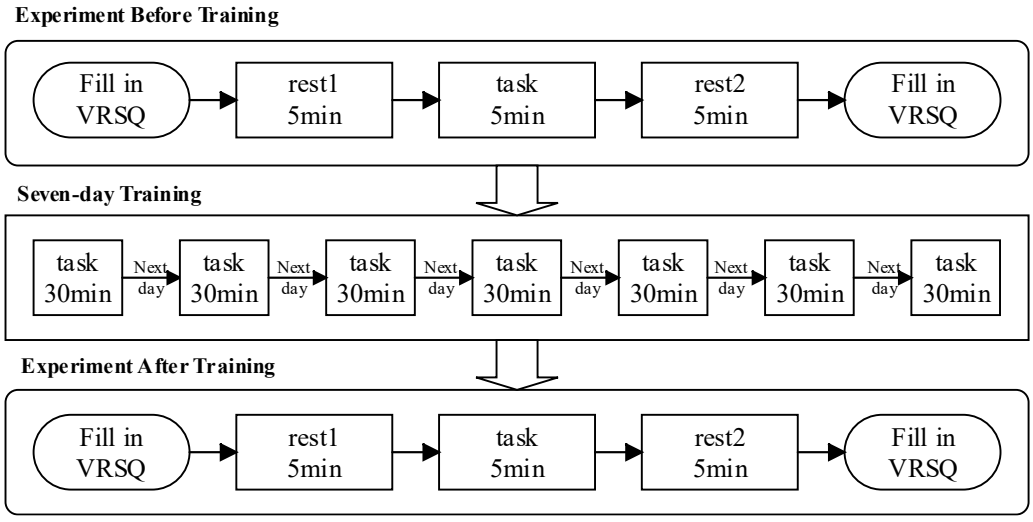


Fig. 1 Experimental procedure The diagram illustrates the experimental workflow, including pre-training tests, 7-day training, and post-training tests. Pre- and post-training follow identical procedures (VRSQ questionnaire → 5-min rest → 5-min task → 5-min rest → VRSQ questionnaire). The training phase involves daily 30-minute tasks. Arrows indicate procedural order; "VRSQ" denotes Valence Rating Scale Questionnaire, "rest" indicates rest periods, and "task" refers to the experimental task.

2.1.2. Pre-processing

Due to the low signal-to-noise ratio of EEG, it is necessary to preprocess the data to improve signal quality. The specific steps are as follows: (1) A 50Hz notch filter is used to eliminate power frequency interference; (2) A bandpass FIR filter (0.5–45Hz) is used for effectively suppressing baseline drift (<0.5Hz) and high-frequency noise (>45Hz); (3) Use ICA combined with the Electrooculogram (EOG) reference to remove the ocular artifacts; (4) Downsample the EEG to the

sampling rate of 256 Hz to reduce the following computatioal cost; (5) The continuous EEG is segmented into 1-second-long non-overlapping samples, and the signal of each lead is normalized using the z-score according to Equation (1).

$$\tilde{x}_i = \frac{x_i - \mu_i}{\sigma_i} (i = 1, 2, \cdots 30) \quad (1)$$

where μ_i and σ_i represent the mean and standard deviation of the EEG x_i in the *i*-th lead.

2.1.3. VRMS Detection Tasks

Due to the different experimental conditions, the VRMS severity and EEG features varies across the different conditions. Therefore, this study identifies five distinct tasks of VRMS detection under the different experimental stages and conditions. Through these analyses, the commonalities and differences of the EEG features among those tasks can be used to enhance the performance and generalization ability of the VRMS detection models by the multi-task learning stratage. The five specific tasks are as follows:

Task1: Based on the fact that training can reduce participants' sensitivity to the VRMS, the VRMS severities in the task before and after the training are different, and the pre-training task-state EEGs can be considered as high VRMS while the post-training ones can be considered as low VRMS.

Task2: Based on the same fact as *Task1*, the VRMS severities in the rest2 before and after the training are also different. Therefore, the pre-training rest2 can be considered as high VRMS while the post-training rest2 can be considered as low VRMS.

Task3: Compared to task-state, the participants' VRMS levels in the rest2 are reduced but not completely eliminated[23]. Therefore, the participants were in a normal state before the task (rest1) and in a VRMS state after the task (rest2). This task classifies the pre-training rest1 vs. rest2 to detect whether there is the VRMS or not.

Task4: Based on the same reasons as *Task3*, this task classifies the post-training rest1 vs. rest2 to detect whether there is the VRMS or not. Given that rest2 is in a lower VRMS level post the training, the difference between rest2 and rest1 should be smaller, making this task more challenging.

Task5: Merge the rest1 and rest2 from *Task3* and *Task4* and classify them as either VRMS or not. This task aims to verify whether there are universally applicable VRMS-related features before and after the training).

2.2. WTConvNet

This study introduces a 1D wavelet convolution into the EEG analysis, proposing a VRMS detection model that integrates wavelet convolution and attention mechanism named as WTConvNet (Fig. 2). The model primarily consists of a wavelet convolution module, a time convolutional attention module, and a lead convolutional attention module. The preprocessed data $x \in R^{C \times T}$ (where *C* represents the number of EEG leads and *T* represents the number of time points) is expanded to $x \in R^{1 \times C \times T}$ as the input.

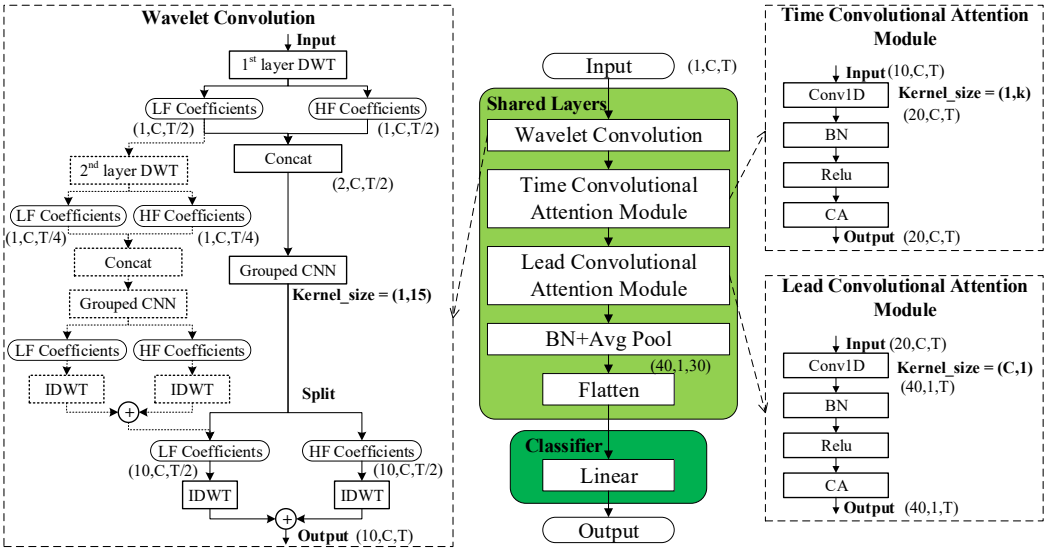


Fig. 2 Architecture of WTConvNet The left module (Wavelet Convolution) decomposes input signals via Discrete Wavelet Transform (DWT), processes low-frequency (LF) and high-frequency (HF) coefficients separately, and reconstructs features through Inverse DWT (IDWT). The central green-highlighted Shared Layers comprise wavelet convolution, Time Convolutional Attention Module, Lead Convolutional Attention Module, batch normalization with average pooling (BN+Avg Pool), and a flatten layer (Flatten), connected to a Classifier (Linear layer and Output). The right section details the structures of the time and lead attention modules. Tensor dimensions and data flow (dashed arrows) are annotated.

2.2.1. 1D Wavelet Convolution

First, the input signal $x[n]$ is decomposed into the low-frequency coefficients L_1 and high-frequency coefficients H_1 through the first level of DWT. And the low-frequency coefficients and high-frequency coefficients after the *j*-th level of DWT can be calculated as the following equations:

$$\begin{aligned} L_j[n] &= \sum_{\mathrm{k}} L_{j-1}[k] \cdot h[2n-k] \\ H_j[n] &= \sum_{\mathrm{k}} L_{j-1}[k] \cdot g[2n-k] \end{aligned} \tag{2}$$

where L_{j-1} is the low-frequency coefficients of the (*j*-1)-th layer, $h[n]$ is the low-pass filter, and $g[n]$ is the high-pass filter. The low-frequency and high-frequency coefficients are then concatenated along the lead dimension, and then a group CNN is applied as follows:

$$Y_w = \mathrm{Conv}_{\mathrm{group}}(\mathrm{Concat}(L_j, H_j), \theta) \tag{3}$$

where θ represents the convolution kernel parameters. Further, Y_w is splited into the low-frequency and high-frequency parts $\overline{L_j}$ and $\overline{H_j}$, which are reconstructed via inverse DWT and then added as the reconstructed signal $Y_{\mathrm{rec}} \in R^{10\times C\times T}$:

$$Y_{\mathrm{rec}} = \sum_{k} (\overline{L_j}[k] \cdot p[\tfrac{n}{2} - k] + \overline{H_j}[k] \cdot q[\tfrac{n}{2} - k]) \tag{4}$$

where p and q are the reconstructive low-pass and high-pass filters, respectively.

2.2.2. Time Convolutional Attention Module

In the time convolutional attention module, a 1D-CNN layer(kernel_size=(1, 15), channel_num=20), a Batch Normalisation (BN) and a ReLU activation function are first used to extract features $R_t \in R^{20\times C\times T}$ along the time dimension from Y_{rec}. R_t is then input into a

Coordinate Attention (CA) module [24], shown in Fig. 3, with the output of this module $Y_{temp} \in R^{20\times C\times T}$ calculated as follows:

$$\begin{aligned} R_t &= \mathrm{Relu}(\mathrm{BN}(\mathrm{Conv1D}(Y_{rec}))) \\ Y_{temp} &= R_t \cdot M_x \cdot M_y \end{aligned} \tag{5}$$

where M_x and M_y are the attention weights of two dimensions (lead and time) calculated as follows:

$$\begin{aligned} X_{avg} &= \mathrm{AdaptiveAvgPool2d}\left((1, None)\right) \\ Y_{avg} &= \mathrm{AdaptiveAvgPool2d}\left((None, 1)\right) \\ M &= \mathrm{Conv2D}(\mathrm{Concat}(Y_{avg}, \mathrm{permute}(X_{avg})), \mathrm{dim}=2) \in R^{ch/r\times 1\times (C+T)} \\ M_{nl} &= \mathrm{h_swish}(\mathrm{BN}(M)) \in R^{ch/r\times 1\times (C+T)} \\ M_1, M_2 &= \mathrm{split}\left(M_{nl},\ [h,\ w],\ dim=2\right) \\ M_x &= \sigma\left(\mathrm{Conv2D}(\mathrm{permute}(M_1))\right) \in R^{ch\times C\times 1} \\ M_y &= \sigma\left(\mathrm{Conv2D}(M_2)\right) \in R^{ch\times 1\times T} \end{aligned} \tag{6}$$

where X_{avg} and Y_{avg} represent the outputs of the average pooling along the time and lead dimensions, σ represents the Sigmoid function. The attention mechanism enables the model to automatically learn to assign the attention weights at different positions, focusing on the information critical to the current task and enhancing the model's ability to capture the important features, thereby improving its performance across the various tasks.

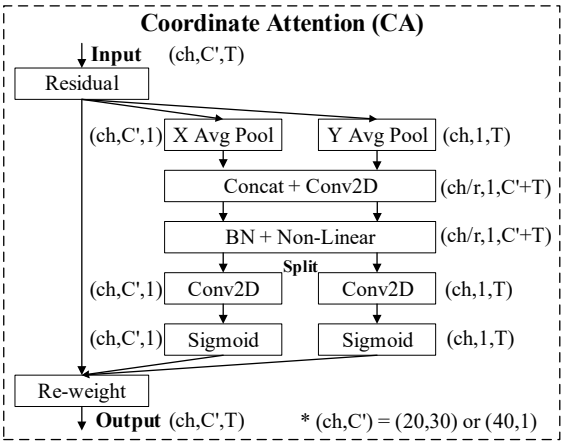


Fig. 3 Architecture of Coordinate Attention (CA) The diagram illustrates the architecture of the Coordinate Attention (CA) module.

2.2.3. Lead Convolutional Attention Module

This module is designed to integrate features in the lead dimension. Most of the structures of this module is like the time convolutional attention module. The only difference in this module is that the 1D-CNN layer is used along the lead dimension instead of the time dimension (kernel_size=(C, 1), channel_num=40). The output of this module is denoted as $F_{lead} \in R^{40\times 1\times T}$, which then passes through a Batch Normalization layer and a global average pooling layer.

2.2.4. Classifier and Loss Function

A fully-connected layer is employed as the classifier in the model, ultimately producing the binary classification results. During training, the model uses a cross-entropy loss function to quantify the difference between the predicted results and the ground truths:

$$\mathrm{L} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right] \tag{7}$$

where N denotes the number of samples, y_i and $\hat{y}_i$ represent the true label and predicted probability of the i-th sample, respectively.

2.3. Multi-Task Learning Framework

Multi-task learning is a method that utilises shared information from diverse tasks [25] to improve the generalisation ability of models by optimising the loss functions of those tasks simultaneously. The multi-task learning is typically achieved through two ways: hard parameter sharing or soft parameter sharing [26]. As showen in Fig 4, the core of hard parameter sharing is to allow multiple tasks to share the parameters of the extractor while maintaining the unique classifier for each task. This sharing mechanism can significantly reduce the risk of overfitting because the model needs to find a representation that can adapt to all tasks rather than simply fitting any individual one of them. Then the task-specific classifier is responsible for extracting task-specific information from the general representation.

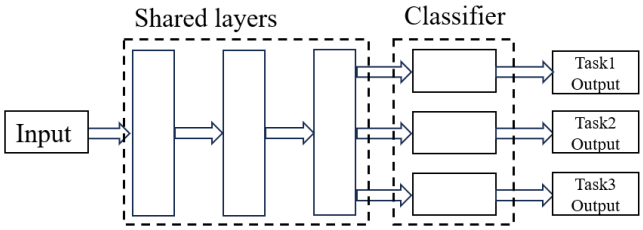


Fig. 4 The structure of multi-task learning in the hard parameters sharing way The diagram illustrates the typical architecture of hard parameter sharing in Multi-Task Learning (MTL). The input passes through shared layers to extract common features, then branches into three task-specific outputs (Task1 Output, Task2 Output, Task3 Output). The arrows indicate data flow.

For the soft parameter sharing, each task has its own model and parameters. Regularisation techniques[27] are used to constrain the distance of the parameters between the models of the different tasks, encouraging the parameters to converge[28], thereby encouraging the parameter similarity between the tasks. The purpose of this approach is to allow the models to learn from each other when tackling the different tasks while maintaining a degree of independence.

This study constructed single-task, dual-task, and triple-task learning models based on WTConvNet. Different task combinations of the five tasks mentioned in Section 2.1.3 were used to validate the models. Given the similarities or intrinsic correlations among these tasks or the corresponding EEG features, and the complexity of the model, the use of soft parameter sharing would result in excessive computational costs. Therefore, the hard parameter sharing was employed. The multi-task learning way not only reduces the number of parameters significantly by using the one shared feature extractor, but also enhances the model's adaptability and generalisation ability across the different tasks.

2.4. Training and Validation

The PyTorch was used to build, train and validate the models in this paper. During the model training, early stopping was adopted as a regularisation strategy to avoid overfitting and ensure the model's generalisation ability, andthe patience was set to 10 epochs. Additionally, the maximum iterations of the training was set to 100 to ensure the model had sufficient time to learn the complex patterns in the data. The Adam optimiser was adopted in this study. To validate the models' reliability and stability and obtain a more accurate estimate of their performance, a 10-fold cross-validation was employed.

Furthermore, this paper validates the effectiveness of the proposed model by comparing the different DWT parameters, different attention modules, and performing the ablation experiments

under the single-task. Furthermore, to validate the model's performance, this paper introduces the existing typical models such as EEGNet[8], DeepConvNet[9], ShallowConvNet[9], and state-of-art EEG models such as EEGConformer[29], FMCNN[30], and MSConvNet[31] for comparison under the multi-task.

3. Results

3.1. Model Optimization

In the DWT, there are various wavelet basis functions available for selection. El-Gindy et al. [32] compared the performance of different wavelet basis functions such as: Haar, Daubechies, Symlets, and Coiflets, in epilepsy prediction based on EEG. The results showed that the Daubechies wavelet performed best in the prediction performance. Therefore, the Daubechies wavelet was used for the wavelet convolution in this study.

Daubechies wavelets of different orders exhibit varying degrees of smoothness. Higher order of Daubechies wavelets are smoother, but this comes at the cost of increased support length and computational complexity. Specifically, low-order Daubechies wavelets demonstrate the unique advantages in capturing signal distortion, while high-order Daubechies wavelets are more suitable for processing smooth signals. Additionally, increasing the number of decomposition layers can decompose the signals into more components in the low frequency, which may also have a certain impact on the results. In this paper, the *Task1* was taken as the example to validate the effects of the different setings in the WTConvNet, using the different orders of the Daubechies wavelets and the different decomposition layer numbers (Table 1). The results showed that the single-layer db3 basis function, achieved the best accuracy of 98.10±0.43%. In contrast, the performance of two-layer db3 and three-layer db3 did not surpass that of a single-layer db3, indicating that increasing the number of decomposition layers does not significantly improve the model's performance.

Table 1 Accuracies of the WTConvNet for the different wavelet functions and decomposition levels (mean±std. %)

Levels	Wavelet function	Accuracy	Levels	Wavelet function	Accuracy
1	db1	97.70±0.57	1	db6	97.62±0.48
	db2	97.78±0.51		db8	97.78±0.37
	db3	**98.10±0.43**		db10	97.73±0.55
	db4	97.59±0.45	2	db3	97.76±0.53
	db5	97.76±0.65	3	db3	97.94±0.46

This study further compares the different attention mechanisms in the time convolutional attention module and lead convolutional attention module in the WTConvNet based on the single-layer db3 in the wavelet convolution: Squeeze-and-Excitation (SE) [33], Efficient Channel Attention (ECA) [34], Convolutional Block Attention Module (CBAM) [35], and CA, with the aim of further optimising the model performance. According to the results of subsequent ablation experiments (Table 2), SE, ECA, and CA all improve the model's classification accuracy to some extent. Among them, the introduction of the CA has the most significant improvement on the proposed model, achieving a accuracy of 99.04±0.33%. Additionally, this is a model without an attention mechanism.

Table 2 Accuracies of the WTConvNet with the different types of attention mechanisms (mean±std. %)

Attentions	SE	ECA	CBAM	CA
Accuracy	98.52±0.44	98.45±0.59	95.68±0.69	**99.04±0.33**

3.2. Ablation Experiment

In the ablation experiments, this section conducted an independent evaluation of the core modules of the WTConvNet. The ablation experiments were set up as follows: (1) Removing the wavelet convolution; (2) Replacing the wavelet convolution with a standard CNN; (3) Removing the CAs from the time convolutional attention module and the lead convolution attentional module; (4) Removing the time convolutional attention module; (5) Replacing the time convolutional attention module with a standard CNN; (6) Removing the lead convolutional atttention module; (7) Replacing the lead convolutional attention module with a standard CNN.

Table 3 Accuracies of the various models in the ablation experiments (mean±std. %)

Models	WTConvNet	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Accuracy	99.04 ±0.33	94.13 ±0.77	96.53 ±0.42	95.75 ±0.82	96.78 ±0.39	98.76 ±0.43	97.75 ±0.84	98.89 ±0.62

To verify the effectiveness of the core components in WTConvNet, an ablation study is conducted, and the results are presented in Table 3. The complete WTConvNet achieves optimal performance, attaining a classification accuracy of 99.04% with the lowest standard deviation. This outcome convincingly demonstrates the superiority of the overall architecture and the stability of its performance. Removing the wavelet convolution module leads to the most significant decline in model performance (approximately 4.91%), and the stability has significantly deteriorated. accompanied by a notable deterioration in stability. When this module is replaced with a standard CNN (2), the performance also decreases—though to a lesser extent compared to complete removal—suggesting that the wavelet convolution module plays a central role in capturing key time-frequency features related to the VRMS. The contribution of the attention mechanism is equally crucial. Removing the CAs from both the time convolutional attention module and the lead convolutional attention module (3) results in a noticeable yet smaller performance decline (approximately 3.29%) compared to (1). Notably, using a single CA (5 or 7) has a relatively limited impact on the overall classification performance. Besides, removing either the time convolutional attention module (4) or the lead convolutional atttention module also lead to a degree of performance degradation, demonstrating the importance of both modules in enhancing feature representation.

3.3. Single-Task Results

The performances of the different models in five single-task are compared with the proposed model, and the results are shown in Table 4. The WTConvNet achieves higher accuracies than other models in all five tasks. In addition, the accuracies of *Task1* was higher than those of other tasks in all the models.

Table 4 Accuracy of the single-task models across the various tasks (mean±std. %)

Models	*Task1*	*Task2*	*Task3*	*Task4*	*Task5*
EEGNet	77.21±2.54	74.22±1.97	63.24±2.77	70.38±2.05	65.15±1.32
DeepConv	87.81±1.66	86.87±1.26	74.70±1.59	80.87±1.95	73.81±0.93
ShallowConv	97.32±0.54	97.00±0.48	91.69±2.06	92.39±0.79	87.39±1.01
EEGConformer	88.75±10.10	87.18±9.07	83.90±10.33	85.89±9.44	76.74±11.67
FMCNN	91.33±0.71	89.83±1.67	84.71±0.76	85.62±0.96	83.99±1.48
MSConvNet	95.79±0.70	95.33±1.59	93.11±0.94	94.18±1.75	85.95±2.52
WTConvNet (ours)	**99.04±0.33**	**98.44±0.46**	**95.48±0.69**	**95.75±0.99**	**94.86±0.65**

3.4. Dual-Task Results

In dual-task learning, the different models exhibit distinct characteristics across the various task combinations (Table 5). The combinations of *Task1+Task3* and *Task3+Task4* demonstrate significant improvements on the performances compared to the corresponding single-task learning across all the models. Among these, the combination of *Task3+Task4* yields the greatest improvements compared to the corresponding single-task learning. At the same time, compared to

the single task of *Task5*, whose dataset actually combines the data from *Task3* and *Task4*, this combination also shows a significant improvements. Among them, the WTConvNet achieves significant improvements over the corresponding single tasks in the most task combinations. The combinations of *Task1+Task3* achieve 99.57% and 97.65%, respectively, representing improvements of 0.56% and 2.17% over the corresponding single tasks. Moreover, some models demonstrated both improved accuracy and reduced standard deviation across the specific task combinations. Moreover, while some task combinations yielded higher accuracy, they were accompanied by increased standard deviation, suggesting heightened overfitting.

Table 5 Accuracies of the different models on the different dual-task combinations (mean±std. %)

Models	*Task1+Task2*		*Task1+Task3*		*Task1+Task4*		*Task1+Task5*	
	Task1	*Task2*	*Task1*	*Task3*	*Task1*	*Task4*	*Task1*	*Task5*
EEGNet	84.69 ±1.40	83.97 ±4.54	84.16 ±1.64	72.01 ±3.93	85.22 ±1.92	71.55 ±2.24	87.27 ±2.27	63.73 ±2.90
DeepConv	66.28 ±11.48	54.49 ±8.69	90.11 ±3.14	76.96 ±8.87	93.49 ±2.23	72.42 ±4.56	92.99 ±4.12	69.07 ±5.38
ShallowConv	98.25 ±0.43	94.55 ±1.97	98.41 ±0.37	92.75 ±1.79	97.09 ±0.52	91.43 ±1.84	98.35 ±0.39	89.04 ±1.05
MSConvNet	96.31 ±0.63	95.12 ±0.84	95.84 ±0.79	95.53 ±0.96	92.98 ±2.01	94.19 ±0.52	96.57 ±0.60	86.64 ±0.83
EEGConformer	85.01 ±9.97	86.24 ±12.05	90.23 ±11.48	84.97 ±9.90	88.51 ±13.64	84.62 ±9.53	88.08 ±12.58	76.39 ±13.61
FMCNN	92.19 ±0.66	88.62 ±0.96	92.19 ±0.44	85.58 ±1.13	89.86 ±1.85	83.71 ±1.54	91.05 ±1.14	80.88 ±0.40
WTConvNet (ours)	98.19 ±3.55	97.99 ±1.87	99.57 ±0.27	97.65 ±1.61	97.76 ±1.86	97.68 ±0.82	99.23 ±0.28	96.83 ±0.75

Continued Table 6 Accuracies of the different models on the different dual-task combinations (mean±std. %)

Models	*Task2+Task3*		*Task2+Task4*		*Task2+Task5*		*Task3+Task4*	
	Task2	*Task3*	*Task2*	*Task4*	*Task2*	*Task5*	*Task3*	*Task4*
EEGNet	87.61 ±2.09	77.03 ±2.13	86.58 ±2.12	80.49 ±2.31	89.14 ±1.72	68.33 ±1.58	77.29 ±2.49	80.50 ±1.75
DeepConv	88.86 ±7.21	77.65 ±7.99	86.78 ±8.47	87.47 ±4.82	83.67 ±13.07	74.96 ±11.25	90.19 ±1.32	88.48 ±3.20
ShallowConv	97.28 ±0.90	94.06 ±1.38	96.08 ±1.34	94.50 ±0.93	98.14 ±0.39	85.12 ±1.52	95.13 ±0.59	93.96 ±0.77
MSConvNet	92.37 ±1.03	92.74 ±1.24	97.66 ±1.40	94.59 ±1.56	96.9 4±0.69	85.51 ±1.09	94.39 ±1.47	95.61 ±1.09
EEGConformer	88.64 ±13.78	88.77 ±11.36	88.77 ±10.43	87.09 ±12.55	90.40 ±10.82	75.83 ±13.67	84.99 ±14.09	85.99 ±13.80
FMCNN	90.61 ±1.32	87.58 ±1.04	89.96 ±1.51	85.43 ±1.81	89.75 ±2.30	82.75 ±1.87	84.85 ±1.83	86.07 ±1.77
WTConvNet (ours)	99.28 ±0.78	99.07 ±0.24	99.08 ±0.72	98.60 ±0.82	99.27 ±0.37	94.76 ±2.18	98.24 ±0.55	98.47 ±0.52

3.5. Triple-Task Results

The results of the triple-task learning are shown in Table 7. Compared with the results of the single-task learning in Table 3, In the combination of*Task1+Task2+Task4*, EEGNet, WTConvNet and FMCNN could achieve the superior performances across all the three tasks compared to the single-task approaches. Similarly, when substituting with the combination of *Task2+Task4+Task5*, EEGNet, ShallowConvNet and WTConvNet also demonstrated overall performance improvements. Beyond these configurations, no other task combinations yielded simultaneous benefits across all three tasks. At the same time, compared with the dual-task learning, their performances decline in the most task combinations. Moreover, similar to the dual-task approach, only certain models demonstrated both improved accuracy and reduced standard deviation across the specific task combinations

Table 6 Accuracy of the different models on the different triple-task combinations (mean±std. %)

Models	*Task1+Task2+Task3*			*Task1+Task2+Task4*			*Task1+Task2+Task5*		
	Task1	*Task2*	*Task3*	*Task1*	*Task2*	*Task4*	*Task1*	*Task2*	*Task5*
EEGNet	84.04 ±1.54	84.34 ±2.57	70.54 ±3.32	84.53 ±1.49	79.66 ±5.25	69.39 ±3.44	84.96 ±2.15	82.79 ±3.39	61.06 ±1.45
DeepConv	60.23 ±3.15	51.06 ±2.13	50.29 ±0.69	70.16 ±14.13	56.88 ±8.78	51.80 ±3.11	67.47 ±11.09	52.76 ±2.63	51.23 ±2.99
ShallowConv	98.29 ±0.42	94.67 ±2.24	93.23 ±1.38	96.44 ±0.75	97.32 ±0.71	90.64 ±1.80	97.55 ±0.60	96.79 ±0.52	88.54 ±1.09
MSConvNet	97.04 ±1.07	94.98 ±1.11	94.18 ±1.55	89.53 ±0.96	96.43 ±1.50	91.68 ±1.38	95.02 ±0.92	96.46 ±1.04	86.47 ±1.43
EEGConformer	85.56 ±12.04	86.75 ±13.31	85.26 ±10.04	84.29 ±13.39	87.45 ±13.74	85.59 ±10.66	82.23 ±13.07	87.01 ±9.95	76.81 ±11.40
FMCNN	94.14 ±1.60	90.90 ±1.19	86.46 ±1.47	91.32 ±1.75	89.75 ±1.12	85.14 ±1.66	93.40 ±1.37	89.67 ±1.02	83.53 ±1.81
WTConvNet (ours)	99.17 ±0.51	99.04 ±0.32	98.50 ±0.37	96.72 ±2.50	98.07 ±0.95	96.74 ±1.20	98.60 ±1.46	99.10 ±0.40	96.36 ±1.04

Continued Table 7 Accuracy of the different models on the different triple-task combinations (mean±std. %)

Models	*Task1+Task3+Task4*			*Task2+Task3+Task4*		
	Task1	*Task3*	*Task4*	*Task2*	*Task3*	*Task4*
EEGNet	84.72±1.45	70.97±1.88	71.30±3.09	86.47±1.34	68.83±2.20	78.12±1.98
DeepConv	62.35±2.73	51.41±4.77	51.12±1.80	52.06±1.08	52.20±1.33	52.13±1.22
ShallowConv	97.56±0.64	93.24±0.65	92.37±1.00	97.79±0.62	93.87±1.04	93.87±0.72
MSConvNet	94.18±1.30	94.16±0.98	94.05±1.14	95.82±1.20	93.35±1.33	92.74±1.21
EEGConformer	85.58±12.27	83.81±9.76	83.17±13.79	86.83±13.37	83.69±11.98	82.96±13.51
FMCNN	90.63±1.68	86.49±0.65	83.33±1.86	91.46±1.03	84.61±1.52	84.64±1.87
WTConvNet (ours)	98.25±1.17	96.61±2.47	97.05±0.60	99.08±1.02	96.00±5.01	97.55±0.78

4. Discussion

Wavelet convolution has demonstrated powerful capabilities in processing EEG.As showen in Table 3 , removing wavelet convolution resulted in a sharp drop in accuracy of 4.91%, significantly higher than that of the removel of other modules, confirming that multi-scale time-frequency decomposition is the foundation of the proposed model. In the wavelet convolution (Table 1), the single-layer db3 configuration achieves the highest performance (98.10%), outperforming both two-

layer and three-layer db3 (97.76% and 97.94%), this superiority can be attributed to the flat amplitude-frequency response of the single-layer db3's low-frequency coefficients within the 0-56 Hz range, which fully covers the δ (0.5–4 Hz) and θ (4–8 Hz) frequency bands that are highly correlated with VRMS [36]. Moreover, a consistent accuracy reduction of approximately 0.5% is observed when using other wavelet basis functions compared to db3. Specifically, compared to db3, the transition band between db1 and db2 exhibit an excessively wide transition band, leading to insufficient attenuation. In contrast, the transition bands of db4 through db10 narrow further inward, resulting in noticeable attenuation near the upper edge of the θ band (around 7–8 Hz), which negatively impacts model performance. Besides, according to the results shown in Table 1 and 2, the CA performs collaborative weights along the lead and time dimensions, improving accuracy from 98.10% to 99.04%. The CA mechanism focuses on performing collaborative weighting across lead and temporal dimensions. It enables the model to automatically learn and assign importance weights to feature dimensions related to leads and time, thereby enhancing the model's emphasis on discriminative features that are more conducive to distinguishing tasks or categories, while mitigating the influence of irrelevant and redundant features.In summary, the wavelet convolution provides a compact frequency-domain representation aligned with the physiological mechanisms of VRMS, while CA further maps the representation precisely to the most discriminative lead-time combinations, which explains the significant performance of the WTConvNet.

For the *Task1* , all the models have better performances, indicating that the task-state EEG patterns before and after the seven-day training are more different compared with the those in other tasks. This difference can be attributed to the training effectively improving the participants' tolerance to the VRMS. For the *Task2*, the models also demonstrated high accuraiesy but little worse than those in the *Task1*, indicating that the training reduced the intensity of the persistent impact of VRMS on the participants' recovery process, and at the same time, the VRMS natural declines after the task. Notably, despite integrating the post-task resting-state EEGs before and after training, *Task5* exhibited significantly lower classification accuracy than *Task3* and *Task4*. This suggests that the adaptive training significantly altered the neural feature associated with VRMS

After introducing the multi-task learning, the model shares the information at the feature level and makes independent decisions at the classification level, effectively addressing the issue of significant differences in the distribution of the task-state, resting-state, pre- and post-training EEGs in the feature space. Experimental results show that, the multi-task learning enhances the model's performance and generalization ability for some certain but not all the task combinations. The combinations of *Task1*+*Task3* and *Task3*+*Task4* demonstrate significant improvements over single tasks across all the models, indicating that *Task1* and *Task3*, as well as *Task3* and *Task4*, exhibit the higher correlation and more discriminated shared features. In the *Task1*+*Task5* combination, the WTConvNet achieved accuracy improvements of 0.22% and 1.97% for *Task1* and *Task5*, respectively, compared to single-task models, while the EEGNet and DeepConvNet experienced the performance declines. This suggests that the proposed model offer unique advantages in identifying the shared features across tasks. When the number of tasks in the multi-task learning increases to three, only the WTConvNet shows improvements in each task compared to the single-tasks in the *Task2*+*Task3*+*Task4* combination, with the increases of 0.07%, 0.52%, and 1.80%, respectively. This suggests that as the number of tasks increases, the difficulty of extracting the shared features rises non-linearly, and making the reasonable selection of task combinations more critical than expanding the number of tasks.

This study has certain limitations. First, the relatively small sample size may affect statistical power. Although the model's robustness was validated through cross-validation strategies, the

insufficient sample diversity may still constrain its generalizationg ability across populations. In particular, the regulatory effects of biological variables such as age and gender were not fully elucidated. Second, at the experimental design level, the study did not systematically examine the differential effects of different virtual reality content types (e.g., linear motion versus rotational scenes) on neural features. This limits the model's adaptability in complex application scenarios. Additionally, operational differences among the participants during tasks and the resulting variations in the stimulus may induce the EEG changes that could affect the recognition outcomes. Future research should expand the sample sizes to include participants with diverse ages, genders, and health statuses, fully accounting for the potential influence of biological variables to enhance the model's universality and generalization across populations. Second, subsequent studies should design more targeted and diverse experimental paradigms to deeply analyze the differential shaping mechanisms of EEG features across different virtual scenarios. Additionally, exploring lightweight model architectures could reduce computational resource consumption while maintaining high performance, thereby enhancing feasibility and efficiency in practical applications.

5. Conclusion

The WTConvNet developed in this study integrates the wavelet convolution, two-dimensional convolutional attention modules with the multi-task learning, achieving a high-precision detection of VRMS on a self-built EEG dataset. In a single-task learning, the proposed model outperforms the existing baselines with an accuracy of 99.04±0.33%. In a dual-task learning for the classifying VRMS in task and resting states before and after training, the accuracies improved to 99.57% and 97.65%, respectively, validating the wavelet convolution's precise capturing of multi-scale spatio-temporal features and the multi-task framework's enhanced role in cross-task feature sharing. This work not only provides a new EEG decoding paradigm that combines interpretability and robustness, but also lays the methodological foundation for the real-time VRMS monitoring and intervention in VR, potentially accelerating the safe adoption of VR technology in applications such as education and healthcare.

6. Declaration

Funding

This work is supported in part by the National Natural Science Foundation of China under Grant 62206130, Grant 62073282, Grant 62473200; in part by the Natural Science Foundation of Jiangsu Province under Grant BK20200821; in part by the Startup Foundation for Introducing Talent of NUIST under Grant 2020r075.

Competing interests

The authors have no competing interests to declare that are relevant to the content of this article.

Ethics approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. This experiment is approved by the Ethics Committee of Qinhuangdao First Hospital (2022A024). Clinical trial number: not applicable

Clinical trial number: not applicable

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Data availability

The datasets generated and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

Authors' contribution

Chengcheng Hua: Funding acquisition, Conceptualization, Supervision.

Qiuyu Wang: Writing – original draft, Software.

Min Zhu: Software, Data curation.

Jingqi Fu: Visualization.

Yin Yan: Supervision.

Dapeng Chen: Funding acquisition, Supervision.

Jia Liu: Supervision.

Rongrong Fu: Funding acquisition, Supervision.

Reference

[1] E. Chang, H. Kim, and B. Yoo, "Virtual Reality Sickness: A Review of Causes and Measurements," (in English), *Int. J. Hum.-Comput. Interact.,* Review vol. 36, no. 17, pp. 1658–1682, Oct 2020, doi: 10.1080/10447318.2020.1778351.

[2] G. Gersak, H. Lu, and J. Guna, "Effect of VR technology matureness on VR sickness," (in English), *Multimed. Tools Appl.,* vol. 79, no. 21-22, pp. 14491–14507, Jun 2020, doi: 10.1007/s11042-018-6969-2.

[3] C. Yildirim, "Don't make me sick: investigating the incidence of cybersickness in commercial virtual reality headsets," (in English), *Virtual Real.,* vol. 24, no. 2, pp. 231–239, Jun 2020, doi: 10.1007/s10055-019-00401-0.

[4] C. C. Hua *et al.*, "Detection of virtual reality motion sickness based on EEG using asymmetry of entropy and cross-frequency coupling," *Physiology & Behavior,* vol. 284, Oct 2024, Art no. 114626, doi: 10.1016/j.physbeh.2024.114626.

[5] Z. Yang and H. Ren, "Feature Extraction and Simulation of EEG Signals during Exercise-Induced Fatigue," *Ieee Access,* vol. 7, pp. 46389–46398, 2019 2019, doi: 10.1109/access.2019.2909035.

[6] Y. F. Yang *et al.*, "Effects of social interaction on virtual reality cybersickness," (in English), *Displays,* vol. 80, p. 12, Dec 2023, Art no. 102512, doi: 10.1016/j.displa.2023.102512.

[7] B. U. Demirel, A. H. Dogan, J. Rossie, M. Möbus, and C. Holz, "Beyond Subjectivity: Continuous Cybersickness Detection Using EEG-based Multitaper Spectrum Estimation," (in English), *IEEE Trans. Vis. Comput. Graph.,* vol. 31, no. 5, pp. 2525–2534, May 2025, doi: 10.1109/tvcg.2025.3549132.

[8] V. Lawhern, A. Solon, N. Waytowich, S. Gordon, C. Hung, and B. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces," *Journal of Neural Engineering,* vol. 15, no. 5, Oct 2018, Art no. 056013, doi: 10.1088/1741-2552/aace8c.

[9] R. T. Schirrmeister *et al.*, "Deep Learning With Convolutional Neural Networks for EEG Decoding and Visualization," *Human Brain Mapping,* vol. 38, no. 11, pp. 5391–5420, Nov 2017, doi: 10.1002/hbm.23730.

[10] J. S. Bang, D. O. Won, T. E. Kam, and S. W. Lee, "Motion Sickness Prediction Based on Dry EEG in Real Driving Environment," (in English), *Ieee Transactions on Intelligent Transportation Systems,* vol. 24, no. 5, pp. 5442–5455, May 2023, doi: 10.1109/tits.2023.3240407.

[11] W. Wang *et al.*, "Sparse Bayesian Learning for End-to-End EEG Decoding," *Ieee Transactions on Pattern Analysis and Machine Intelligence,* vol. 45, no. 12, pp. 15632–15649, Dec 2023, doi: 10.1109/tpami.2023.3299568.

[12] D. Li, L. Xie, B. Chai, Z. Wang, and H. Yang, "Spatial-frequency Convolutional Self-Attention Network for EEG Emotion Recognition," *Applied Soft Computing,* vol. 122, 2022, doi: 10.1016/j.asoc.2022.108740.

[13] Z. A. A. Alyasseri *et al.*, "Multi-objective flower pollination algorithm: a new technique for EEG signal denoising," *Neural Computing & Applications,* vol. 35, no. 11, pp. 7943–7962, Apr 2023, doi: 10.1007/s00521-021-06757-2.

[14] N. Kaongoen and S. Jo, "Adapting Artifact Subspace Reconstruction Method for SingleChannel EEG using Signal Decomposition Techniques," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference,* ; Research Support, Non-U.S. Gov't vol. 2023, pp. 1–4, 2023–07 2023, doi: 10.1109/embc40787.2023.10340077.

[15] M. Zhang *et al.*, "Wavelet Entropy Analysis for Ictal Electroencephalogram Signals of Child Absence Epilepsy," *Journal of biomedical engineering,* vol. 35, no. 4, pp. 530–538, 2018, doi: 10.7507/1001-5515.201701002.

[16] M. Mohseni, V. Shalchyan, M. Jochumsen, and I. Niazi, "Upper limb complex movements decoding from pre-movement EEG signals using wavelet common spatial patterns," (in English), *Comput. Meth. Programs Biomed.,* vol. 183, p. 10, Jan 2020, Art no. 105076, doi: 10.1016/j.cmpb.2019.105076.

[17] S. Khare and V. Bajaj, "A self-learned decomposition and classification model for schizophrenia diagnosis," (in English), *Comput. Meth. Programs Biomed.,* vol. 211, p. 10, Nov 2021, Art no. 106450, doi: 10.1016/j.cmpb.2021.106450.

[18] W. Chen, H. Sang, J. Wang, and Z. Zhao, "WTGCN: wavelet transform graph convolution network for pedestrian trajectory prediction," *International Journal of Machine Learning and Cybernetics,* vol. 15, no. 12, pp. 5531–5548, Dec 2024, doi: 10.1007/s13042-024-02258-5.

[19] Z. Jia, Y. Ouyang, X. Kong, Y. Guo, Z. Li, and H. Zeng, "A Novel Dual-Task Model for EEG-Based Emotion and Cognition Recognition," *Ieee Transactions on Instrumentation and Measurement,* vol. 74, 2025 2025, Art no. 2503314, doi: 10.1109/tim.2024.3522413.

[20] Q. He, L. Feng, G. Jiang, and P. Xie, "Multimodal Multitask Neural Network for Motor Imagery Classification With EEG and fNIRS Signals," *IEEE Sens. J.,* vol. 22, no. 21, pp. 20695–20706, Nov 1 2022, doi: 10.1109/jsen.2022.3205956.

[21] C. Hua *et al.*, "Dual-pathway EEG model with channel attention for virtual reality motion sickness detection," *J Neurosci Methods,* vol. 418, p. 110425, Jun 2025, doi: 10.1016/j.jneumeth.2025.110425.

[22] H. K. Kim, J. Park, Y. Choi, and M. Choe, "Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment," *Appl Ergon,* vol. 69, pp. 66–73, May 2018, doi: 10.1016/j.apergo.2017.12.016.

[23] N. O. Conner *et al.*, "Virtual reality induced symptoms and effects: concerns, causes, assessment & mitigation," in *Virtual Worlds*, 2022, vol. 1, no. 2: MDPI, pp. 130–146.

[24] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 13713–13722.

[25] S. Vandenhende, S. Georgoulis, W. Van Gansbeke, M. Proesmans, D. Dai, and L. Van Gool, "Multi-task learning for dense prediction tasks: A survey," *IEEE transactions on pattern analysis and machine intelligence,* vol. 44, no. 7, pp. 3614–3633, 2021.

[26] S. Ruder, "An overview of multi-task learning in deep neural networks," *arXiv preprint arXiv:1706.05098,* 2017.

[27] L. Duong, T. Cohn, S. Bird, and P. Cook, "Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser," in *Proceedings of the 53rd annual meeting of the Association for Computational Linguistics and the 7th international joint conference on natural language processing (volume 2: short papers)*, 2015, pp. 845–850.

[28] Y. Yang and T. M. Hospedales, "Trace norm regularised deep multi-task learning," *arXiv preprint arXiv:1606.04038,* 2016.

[29] Y. Song, Q. Zheng, B. Liu, and X. Gao, "EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization," *Ieee Transactions on Neural Systems and Rehabilitation Engineering,* vol. 31, pp. 710–719, 2023 2023, doi: 10.1109/tnsre.2022.3230250.

[30] W. Wang *et al.*, "EEG-FMCNN: A fusion multi-branch 1D convolutional neural network for EEG-based motor imagery classification," *Medical & Biological Engineering & Computing,* vol. 62, no. 1, pp. 107–120, Jan 2024, doi: 10.1007/s11517-023-02931-x.

[31] G. Zhang, B. Yang, X. Dong, J. Ma, P. Zan, and X. Xia, "A multiscale convolutional neural network based on time-frequency features for decoding rat exercise fatigue LFP," *Biomedical Signal Processing and Control,* vol. 95, Sep 2024, Art no. 106333, doi: 10.1016/j.bspc.2024.106333.

[32] S. A.-E. El-Gindy *et al.*, "Efficient communication and EEG signal classification in wavelet domain for epilepsy patients," *Journal of Ambient Intelligence and Humanized Computing,* vol. 12, no. 10, pp. 9193–9208, Oct 2021, doi: 10.1007/s12652-020-02624-5.

[33] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[34] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11534–11542.

[35] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[36] E. Chang, M. Billinghurst, and B. Yoo, "Brain activity during cybersickness: a scoping review," *Virtual Real.,* vol. 27, no. 3, pp. 2073–2097, 2023.