Received: 22 April 2023 | Revised: 15 June 2023 | Accepted: 26 June 2023 | Published online: 26 June 2023

Journal of Data Science and Intelligent Systems
2024, Vol. 2(1) 1–13
DOI: 10.47852/bonviewJDSIS3202983

RESEARCH ARTICLE

BON VIEW PUBLISHING

3D-STCNN: Spatiotemporal Convolutional Neural Network Based on EEG 3D Features for Detecting Driving Fatigue

Bo Peng[1,2,†], Dongrui Gao[1,3,†], Manqing Wang[1,3] and Yongqing Zhang[1,*]

[1]*School of Computer Science, Chengdu University of Information Technology, China*

[2]*School of Software, Sichuan Vocational College of Information Technology, China*

[3]*School of Life Sciences and Technology, University of Electronic Science and Technology, China*

Abstract: Fatigue driving has become one of the main causes of traffic accidents, and driving fatigue detection based on electroencephalogram (EEG) can effectively evaluate the driver's mental state and avoid the occurrence of traffic accidents. This article evaluates a feature extraction method for extracting multiple features of EEG signals and establishes a spatiotemporal convolutional neural network (STCNN) to detect driver fatigue. Firstly, we constructed a three-dimensional feature of the EEG signal, which includes the frequency domain, time domain, and spatial features of the EEG signal. Then, we use STCNN for fatigue state classification. STCNN is composed of an attention time network based on attention mechanism and an attention convolutional neural network based on attention mechanism. In addition, we conducted fatigue driving experiments and collected EEG signals from 14 subjects in both awake and fatigued states, ultimately collecting EEG data under three different driving task loads. We conducted extensive experiments on this basis and compared the effectiveness of STCNN and six competitive methods. The results show that the classification accuracy of STCNN is 87.55%, which can effectively detect the fatigue status of drivers.

Keywords: fatigue testing, electroencephalogram, feature extraction, deep learning

1. Introduction

With the development of transportation and the continuous acceleration of urbanization, driving has become an indispensable part of modern people's lives. According to the World Health Organization, the number of deaths caused by fatigue driving exceeds 10,000 annually, posing a huge threat to road safety (Racioppi et al., 2004). People often experience fatigue driving due to factors such as prolonged driving, lack of sufficient sleep, or high-intensity physical labor. Fatigue leads to a decrease in the driver's reaction speed, attention, decision-making ability, vision, and predictive ability. Fatigue can lead to drivers being unable to respond to unexpected situations in a timely manner, losing control of vehicle safety, and increasing the risk of traffic accidents. In 2011, a study by Coetzer and Hancke (2011) showed that if drivers were warned in advance, 90% of traffic accidents could be avoided. Therefore, developing a reliable driving fatigue detection system has become an important research direction. By analyzing the driver's mental state through a fatigue detection system and timely reminding, the driver can effectively reduce the probability of traffic accidents.

So far, several fatigue detection methods related to human factors have been developed, including head state and facial expressions, such as continuously collecting driver nods, yawns, and blink frequencies (Sikander & Anwar, 2019). However, these non-physiological sources are easily influenced by personal differences and environmental factors, and the accuracy and reliability of detection are to some extent unreliable. Physiological signals are directly collected from the human body, such as Electroencephalography, Electromyography, ECG and EOG. These signals contain all the information about the state of the human body and have attracted wide attention. After long-term research and practice, EEG signals have become the most important physiological signal source in fatigue detection. EEG signals are directly collected from the human scalp and have high temporal resolution, high sensitivity, and other characteristics. When the mental state of the human body changes, EEG will undergo changes under fatigue (Risqiwati et al., 2020), such as a decrease in the power of alpha waves and an increase in the power of theta waves. Therefore, by analyzing EEG signals, it is possible to identify whether the driver is in a fatigue state. The collection of EEG signals requires professional EEG equipment. Various forms of EEG equipment and electrode technology are rapidly developing (Chi et al., 2012), capable of adapting to EEG signal acquisition tasks in different scenarios, with good application prospects.

Extracting valuable features from EEG signals plays an important role in improving the accuracy of fatigue detection. Machine learning algorithms and deep learning algorithms are often used to extract the features of EEG signals. Sharma et al.

*Corresponding author: Yongqing Zhang, School of Computer Science, Chengdu University of Information Technology, China. Email: zhangyq@cuit.edu.cn
†These authors contributed equally to this work.

(2021) extracted features from different frequency bands using wavelet transform. Zheng et al. (2022) combined integrated empirical mode decomposition with power spectral density (PSD) to explore new EEG features for driving fatigue detection. Abu Farha et al. (2022) proposed a new wavelet independent component analysis for processing EEG signals to reduce the impact of artifacts. This method outperforms existing independent component analysis (ICA) methods in feature extraction. Zhang et al. (2021b) proposed a new structure for convolutional autoencoder and convolutional neural network, which is used to extract useful features from data and has good performance. Zhang et al. (2021a) proposed a neural network based on convolution and short-term memory (CNN+LSTM), which can quickly learn the information before and after the time series. Du et al. (2021) proposed a deep learning framework for Takagi-Sugeno-Kang (TSK) type convolutional recursive fuzzy network, which can extract spatiotemporal features from EEG signals. Zhang et al., (2023) proposed a hybrid neural network for extracting features from single-channel EEG. In 2023, Peng et al. (2023) evaluated an entropy feature extraction strategy and proposed a new T-A-MFFNET model that can extract multidimensional features from multi-channel EEG signals, achieving good classification results.

Machine learning algorithms have excellent performance in fatigue state classification tasks. In 2016, Bashivan et al. (2015) used an Linear Discriminant Analysis (LDA) classifier to classify EEG signals, particularly in tasks such as finger movement and visual stimulation. In 2017, Miao et al. (2017) used a weighted naive Bayes classifier in the EEG classification task, which achieved good results on two EEG datasets. Djamal et al. (2017) used fast Fourier transform for feature extraction to classify EEG signals. Compared to machine learning algorithms, deep learning algorithms can automatically learn features in EEG signals without any prior selection (Schirrmeister et al., 2017). Deep learning technology has achieved excellent results in classification such as image processing, speech recognition, and object detection (Mandavifar & Ghorbani, 2019; Marino et al., 2018), and many researchers have begun to use deep learning methods to analyze EEG data. In particular, the convolutional neural network (CNN) transforms EEG signals into different forms by changing their dimensions. This not only allows for valuable information to be learned from EEG signals but also does not cause any significant information loss. These methods have analyzed EEG signals from different aspects and achieved certain results, but the extracted features of EEG signals are relatively single, making it difficult to achieve better performance.

In summary, there are still many shortcomings in the EEG signal-based driving fatigue detection method, mainly including the following issues:

(1) Extracting the characteristics of the EEG signal is too single: EEG signals have a low signal-to-noise ratio, are complex and non-linear, and are prone to interference from the human body and external environment during the acquisition process. If only a single feature is extracted, it will significantly affect the final classification effect.

(2) Low classification accuracy: Fatigue detection methods based on EEG signals are fundamental in every aspect, from signal acquisition, pre-processing, and feature extraction to classification. If some areas are not handled properly, the final classification accuracy will not be ideal.

(3) Model performance needs to be more stable: Many factors affect the performance of a model, such as each person's fatigue level. Therefore, improving the stability of model performance under different levels of fatigue is particularly important.

(4) Different frequency band combinations affect classification accuracy: The most common types of EEG signals are delta wave, theta wave, alpha wave, beta wave, and gamma wave. These frequency bands have different effects on classification accuracy. Therefore, exploring the impact of different frequency band combinations on classification performance is essential.

To address the above issues, we constructed 3D features of EEG signals and then utilized the proposed spatiotemporal convolutional neural network (STCNN) for fatigue state classification. Specifically, (1) this article evaluates a 3D feature construction method that includes frequency domain and spatial features, aiming to analyze EEG signals from multiple perspectives and improve the accuracy of EEG signal classification. (2) This article proposes a STCNN for fatigue state classification. STCNN consists of attention time network (ATNet) and attention-based convolutional neural network (ACNN), ATNet consists of attention module and bidirectional long- and short-term memory (BiLSTM), and ACNN consists of attention module and ConvNeXt. (3) This article inputs 1D features into ATNet, which can focus on spatiotemporal features in EEG signals and discard some useless information. By inputting 3D features into ACNN, this network can further analyze valuable information in EEG signals, making classification performance more stable. This article conducted extensive experiments using this method on three datasets, with the aim of verifying whether the method can still maintain stability under different levels of fatigue. Finally, the influence of different frequency band combinations on fatigue state classification was verified.

The main contributions of this study are as follows: (1) The proposed 3D feature construction method includes multiple features from EEG data, being able to learn more useful information for classification; (2) a new STCNN is proposed to classify fatigue states; (3) a large number of experiments were carried out on three self-collected sleep-deprived driving datasets. Compared with other excellent in-depth learning methods, our method performs well in the secondary classification task.

The remaining work of this article is as follows: In the second section, a brief review of relevant work is provided. In the third section, we provided a detailed introduction to our data collection process. In the fourth section, we introduced the technical method we proposed. In the fifth section, the experiment and result analysis were introduced. In the sixth section, we summarized the work of this article.

2. Related Work

EEG signals have the characteristics of high dimensionality and high sampling rate. After dimensionality reduction and compression, they can extract features related to the target task. These features can better represent and characterize the information of EEG signals, helping to classify, recognize, and analyze important information in EEG signals. Meanwhile, different features may have different importance for different tasks and applications, so it is necessary to select appropriate features based on specific application scenarios.

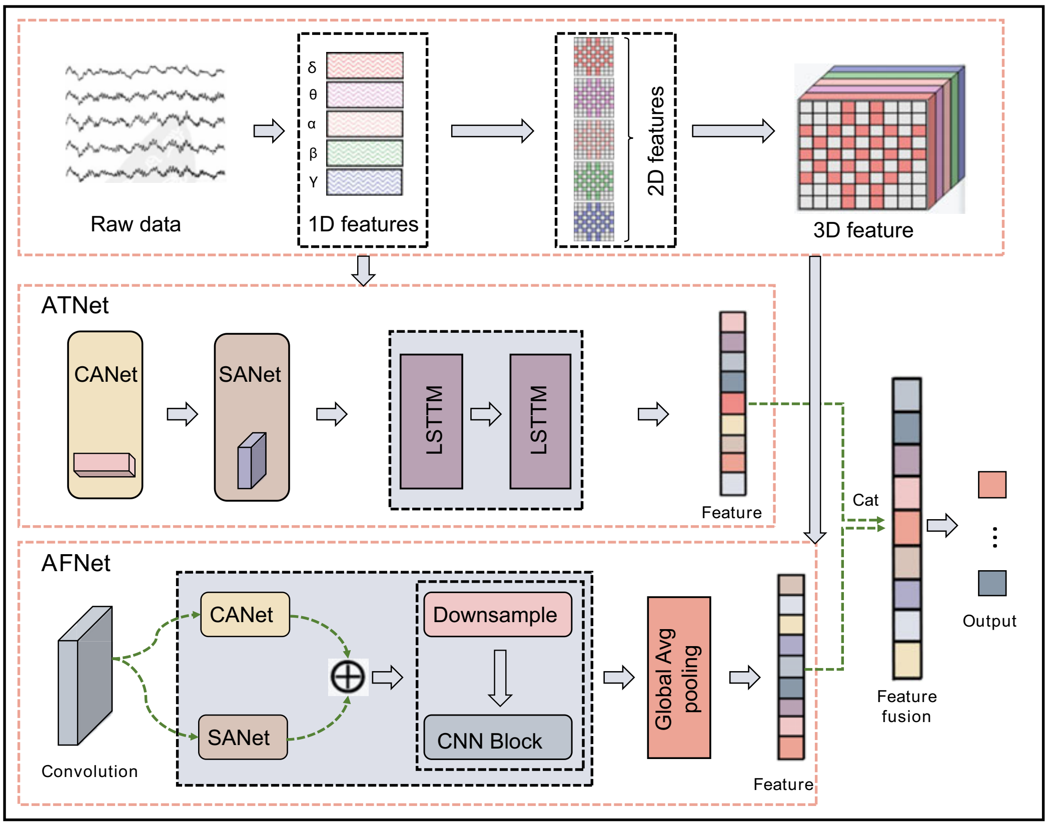
Zhang et al. (2019) proposed a cyclic 3D convolutional neural network (R3DCNNs) for learning EEG features of different tasks without prior knowledge. This network can simultaneously learn EEG features from spatial, spectral, and temporal dimensions. Yang et al. (2018) proposed a three-dimensional representation of EEG bands to combine signal features from different frequency bands while preserving spatial information between channels. Zhao et al. (2020) proposed a three-dimensional CNN model for automatically extracting spatiotemporal features from EEG signals, which performed well in binary and quaternary classification tasks based on emotion recognition datasets. Lin et al. (2020) proposed that 3D-CNN and 3D-LSTM can remove noise from sEMG signals with a classification accuracy of 90.52%. Shalsh (2021) proposed a new CNN deep network model for identifying fatigue or normal fatigue, which utilizes only one EEG channel signal to estimate the driver's fatigue state and has good flexibility. Sheykhivand et al. (2022) proposed an automatic system for two-stage classification of driving fatigue based on EEG signals. The compressed EEG data are fed into the proposed deep CNN for automatic feature extraction and classification, which can effectively detect driver fatigue.

In summary, the following aspects can be summarized: (1) It is feasible to construct 3D features of EEG signals, which include various features of EEG signals. From the multidimensional features, the state displayed by EEG signals can be more comprehensively analyzed. (2) The use of deep learning algorithms can automatically and comprehensively analyze the features of EEG data, playing a crucial role in identifying fatigue states. (3) Using CNN and attention mechanism module to classify fatigue states can achieve better performance.

3. Technical Method

Figure 1 shows the entire workflow. In order to analyze EEG signals more accurately, we evaluated a 3D feature extraction method. First, the original signal is divided into five frequency bands, and then, the PSD of each frequency band is extracted to obtain 1D features. Then, based on the distribution position of the electrodes, the PSD features are mapped to a two-dimensional plane to get two-dimensional features. Finally, five 2D planes are stacked to obtain 3D elements, mainly temporal and spatial features. This paper inputs the extracted features into a STCNN. STCNN consists of an attention-based time domain network (ATNet) and ACNN. This paper inputs 1D features into ATNet, composed of attention mechanisms (channel attention and spatial attention) and BiLSTM. It is mainly used to extract one-dimensional time domain signal features and focus on some valuable information. Input 3D features into ACNN, which comprises attention mechanisms (channel attention and spatial attention) and CNN. The main body of CNN is the ConvNeXt

Figure 1
Experimental process of the entire work



pure CNN, as proposed by Liu et al. (2022). This article is mainly used to analyze the spatiotemporal characteristics of EEG. Finally, this paper fuses ATNet and ACNN output characteristics and performs fatigue state prediction based on the obtained features.

3.1. 3D features

We have constructed three-dimensional features aimed at integrating both temporal and spatial information of EEG signals. The construction process is shown in Figure 2. Since the dataset in this article calculates a tag every 10 s, to make each feature input correspond to the label, this article divides the original EEG data into several non-overlapping 10-second EEG signal segments, matching the tag with the corresponding EEG signal segment. The figure shows the details of extracting signals from an EEG segment to construct 3D features.

First, the T second signal segment (in this article, T is 10 s) is divided into T time windows, each with a length of 1 second. Then, using a Butterworth filter, the EEG signal segments within each time window are decomposed into five frequency bands (Delta, Theta, Alpha, Beta, and Gamma). Next, the PSD of different frequency bands is calculated separately, and PSD can describe more detailed frequency distribution information to obtain five 1D features. Finally, the dataset used in this article contains 31 channels. According to the 2D graph of the 31 channels, the PSD is arranged and superimposed to obtain 3D features. In summary, each raw EEG signal can be represented as a 3D feature: $X_n \in R^{h*w*d}$, $n=1,2,\ldots,N$. N is the total number of samples, h and w represent the height and width of the two-dimensional graph of electrode positions, respectively, and d represents the number of frequency bands. In this article, h is 9, w is 9, and d is 5.

Assume that the original signals of the three datasets in this article are represented as $I_n \in R^{c*r}$, where c represents the number of electrode channels and r represents the sampling rate. In this article, c is 31, and r is 1000. First, each EEG fragment is divided into T 1-second time windows. Then, this article uses a Butterworth filter to filter out five frequency bands: Delta (0.5–4 Hz), Theta (4–8 Hz), Alpha (8–12 Hz), Beta (12–30 Hz), and Gamma (30–50 Hz). Then, calculate the PSD of different frequency bands. PSD is a measurement that describes the energy distribution of a signal in the frequency domain. It is used to analyze the frequency domain characteristics of signals and can convert time domain signals into frequency domain signals. The higher the PSD, the greater the energy in the signal at that frequency. In EEG signal processing, PSD is often used to analyze brain wave activity in different frequency bands (Demiralp et al., 2001). If the time domain signal is $x(t)$, its Fourier transform is:

$$X(f) = \int_{-\infty}^{\infty} X(t) e^{-j2\pi ft} dt \tag{1}$$

where f is the frequency and $X(f)$ is the complex amplitude of the signal at frequency f. According to the properties of the Fourier transform, it can be obtained that the PSD function of $x(t)$ is:

$$P(f) = \lim_{T \to \infty} \frac{1}{T} \left| \int_{-T/2}^{T/2} X(t) e^{-j2\pi ft} dt \right|^2 \tag{2}$$

where $P(f)$ represents the power density of the signal at frequency f. T represents the length of the time window. When $T \to \infty$ is used, this formula represents the PSD estimation of the entire signal.

The original EEG signal fragment is converted into a PSD fragment and used as an input to ATNet, which can be expressed as $S_n \in R^{c*d*T}$, where c is the number of channels, d is the number of frequency bands, and T is the length of the signal segment. In this article, c is 31, d is 5, and T is 10.

Figure 2
Construction process of 3D features

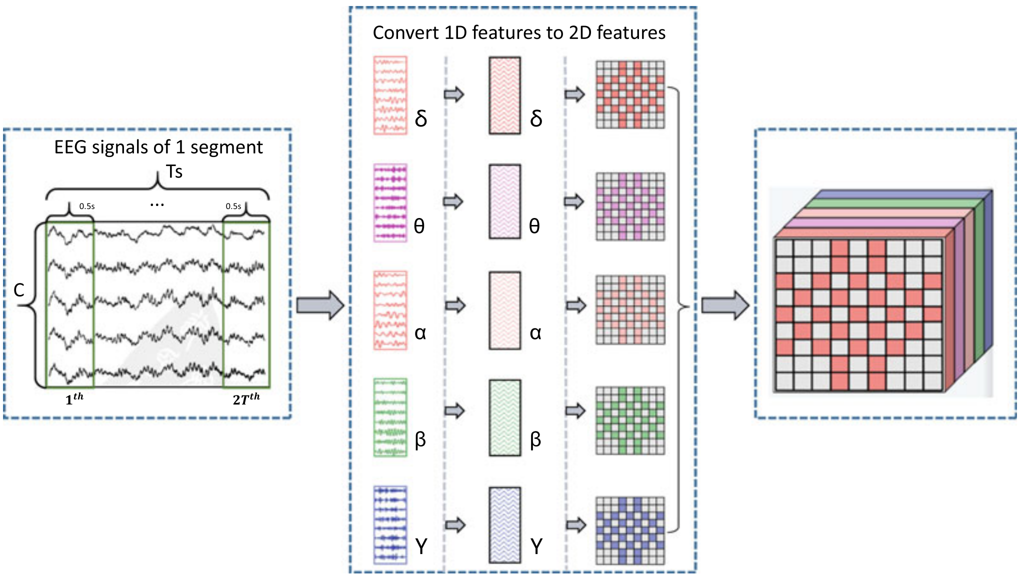
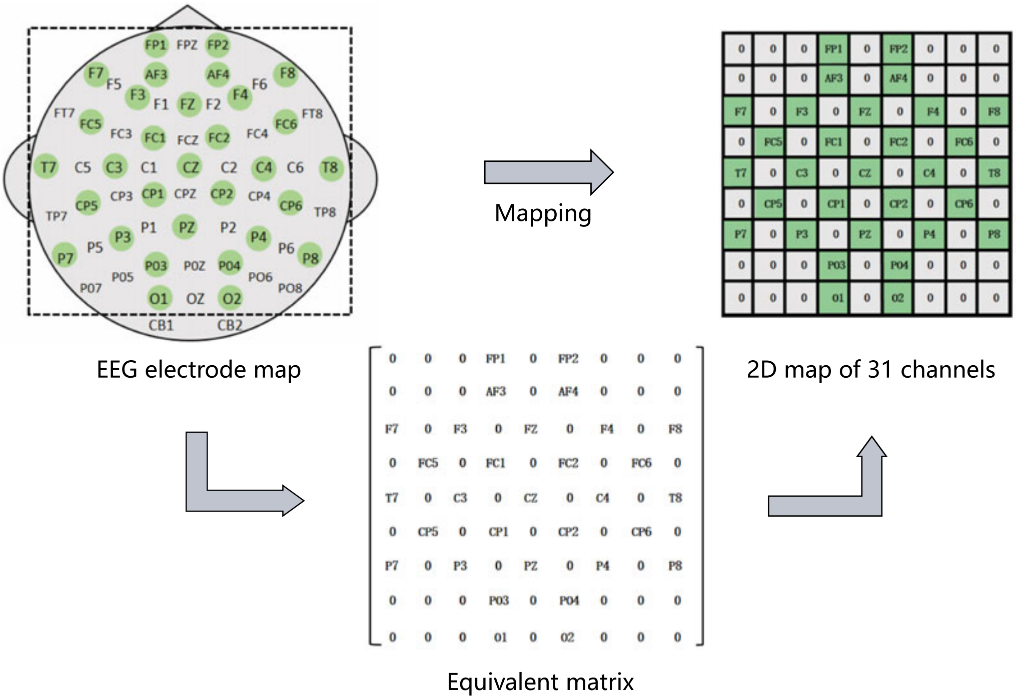


Figure 3
Building a 2D plane



The International 10–20 System is a standard method for describing electrode locations and subcortical regions, as shown in Figure 3. The electrodes marked green are the electrodes used in EEG experiments. Generally speaking, EEG electrodes are spatially distributed around the head, with multiple electrodes adjacent. Different electrodes may interact to generate specific information or record information on particular areas. One-dimensional PSD features are transformed into two-dimensional planes based on the electrode distribution map to preserve spatial information between adjacent channels. The width and height of the 2D plane are 9, and unused channels are represented by 0.

Using the above method, five 2D planes for each time window within the EEG segment can be obtained (i.e., each frequency segment corresponds to a 2D PSD feature map). Yang et al. (2018) proposed a 3D design method. Specifically, it refers to using the representation method of color images to construct the three-dimensional input of features using the sigma representation. Referring to this method, this article stacks five 2D planar elements S_n into 3D EEG features X_n and uses them as input to ACNN.

Table 1 lists the corresponding operations. Therefore, X_n can be expressed as $X_n \in R^{h*w*d}$, where h and w are the height and width of the 2D graph, respectively, and d represents the number of frequency bands. In this article, h is 9, w is 9, and d is 5.

After the above discussion, 3D features mainly include the frequency domain features of the five frequency bands of the EEG signal and the spatial domain features between the electrodes. For the original EEG signal, the PSD of five frequency bands represents the frequency domain characteristics. The 2D mapping plane of 31 channels means the spatial information between the electrodes.

3.2. STCNN

In the work of this article, STCNN consists of ATNet and ACNN. ATNet comprises an attention mechanism and a BiLSTM. The attention mechanism module consists of channel and spatial attention modules. The ACNN consists of a ConvNeXt network and an attention module. These sections will be described in detail next.

3.2.1. Attention module

Attention networks have been widely used in deep learning in recent years (Woo et al., 2018). Especially in the fields of classification, anomaly detection, feature extraction, etc., it has good performance. When using attention networks to process EEG signals, they can learn the importance of EEG signals at specific points in time, thereby achieving precise modeling and analysis of EEG signals.

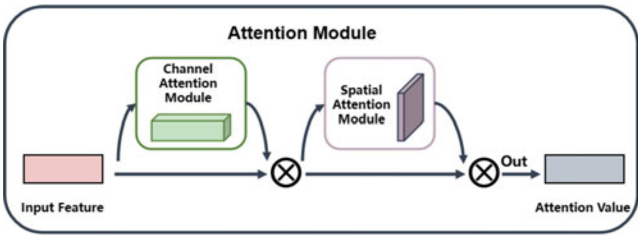
Table 1
Corresponding terms

	Territory	
	Computer vision	EEG
Term description	Color image	EEG cube
	Color channel (R G B)	Frequency band (δ θ α β γ)
	Color intensity	PSD characteristic value

It is worth noting that EEG signals are collected from the scalp, and the electrodes present a spatial distribution pattern on the head, dependency between channels. Attention networks can learn more valuable information and discard other useless information, thereby improving the performance of the model. Therefore, this article uses channel and spatial attention networks to learn inter-channel dependency and spatial features from EEG signals.

Attention Module: This article uses the attention module in both ATNet and ACNN. In hybrid attention mechanisms, attention mechanisms can be used in series or parallel. This article's attention modules are built in series, as shown in Figure 4. This module first processes the input features through the channel attention module, then processes them through the spatial attention module, and finally obtains the output features.

Figure 4
Attention module



Channel Attention: The module first performs average and maximum pooling operations on the input characteristics. This allows compression of the spatial dimensions of features, aggregation of spatial information, and learning of correlation features between channels. Next, the pooled features are subjected to upsampling and downsampling processing. The purpose of downsampling is to reduce computing resources while upsampling is to maintain a constant number of channels.

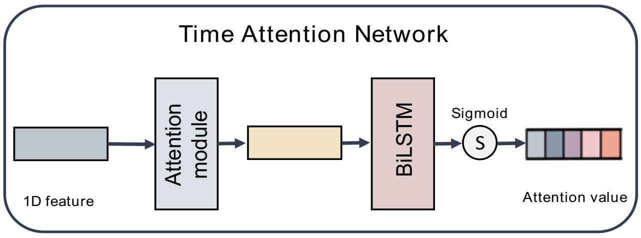
Spatial Attention: The module first performs maximum pooling and average pooling on input features to obtain two 1-D global receptive fields. Then, concatenate the two features to get a 2-D spatial attention graph, which can bring more valuable information. Using filters reduces the dimension of features and the amount of computation.

3.2.2. Time domain network based on attention mechanism (ATNet)

BiLSTM was initially proposed by Schuster and Paliwal (1997). BiLSTM is an extension of LSTM, which can better capture the context in a sequence when processing sequence data. EEG signals have continuity and timing in time series, and BiLSTM can better capture the temporal dependencies between signals, thereby more accurately predicting and classifying them. For example, the EEG signals of each channel can be input as a time series into BiLSTM to organize the EEG signals in different states. Therefore, this paper proposes a time domain network based on an attention mechanism, which aims to learn more valuable information from 1D feature input and improve classification performance. The structure is shown in Figure 5.

Time Network: We define the input 1-D signal $X = (x_1, x_2, \ldots, x_N) \in R^{N*C*w*h}$, where N represents the batch size, C represents the number of channels, and w and h represent the width and height of the feature. Firstly, through the attention module, the spatial attention mechanism can weigh EEG signals in the time domain, allowing greater attention to spatial regions that play an important role in time series. The channel attention mechanism can weigh EEG signals in the frequency domain, allowing greater attention to frequency bands that play an important role in the spectrum. The output of the attention module is then sent to BiLSTM. Through a two-layer BiLSTM network, each layer contains 128 computing units, which can better focus on the time dependence between signals. The process of entering a sequence through ATNet can be described as follows:

Figure 5
ATNet network structure diagram



$$Q_t = \sigma(T(\sigma(S(C(X))))) \in R^{N*C*w*h} \tag{3}$$

where C represents channel attention, S represents spatial attention, σ represents the sigmoid activation function, and T represents the BiLSTM network.

3.2.3. CNN based on attention mechanism (ACNN)

We input 3D features into ACNN. The ACNN consists of an attention module and a CNN. CNN is based on the improved ConvNeXt pure CNN to fuse and classify features. The structure of CNN is shown in Figure 6, mainly composed of a convolutional layer, Layer Norm, ConvNeXt Block, Downsample, pooled layer, and fully connected layer.

Figure 6
ACNN network structure diagram

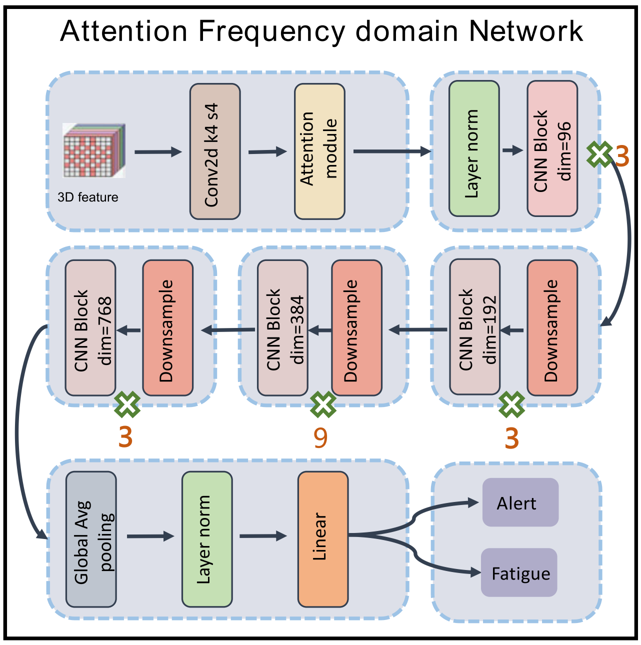
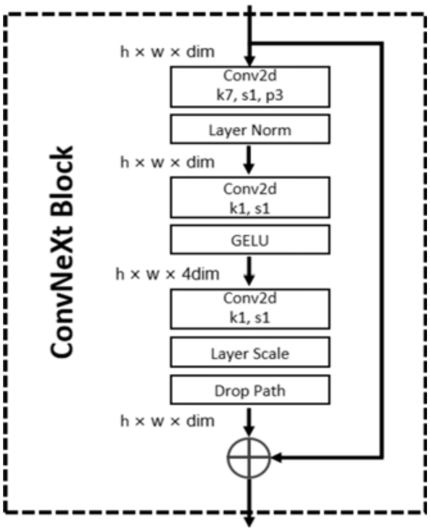


Figure 7
Structural diagram of ConvNeXt block

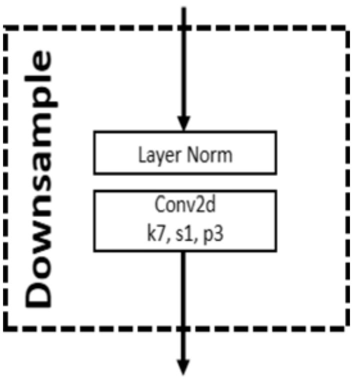


Convolution layer: The feature input first passes through a convolution layer with a convolution kernel size of 4 × 4. The primary purpose of this layer is to receive the feature input and obtain the feature vector.

Layer Norm: This layer is used after the fully connected layer and is normalized on a single sample, which can stabilize the feature distribution, improve the robustness and generalization ability of the model, and accelerate the convergence speed of the model.

ConvNeXt Block: The architecture of this module is shown in Figure 7. The selection of activation functions and normalization layers is the focus of this module. Firstly, this module only uses an activation function, GELU, instead of ReLU. This function can be seen as a smoothing variant of ReLU, which can enhance the model's performance. The normalization layer is only used once after the first convolution, and instead of using the conventional BatchNorm layer, the Layer Norm layer is selected. This is done to improve convergence and reduce overfitting. In addition, the Layer Scale operation is a special normalization operation that can adjust features based on the scale information of the sample, thereby improving the expression and generalization capabilities of the network. The final function of DropPath is to regularize cross-layer connections, reduce network complexity, and reduce the risk of overfitting.

Figure 8
Downsample structure diagram



Downsample: The architecture of this module is shown in Figure 8. The lower sampling layer comprises a Layer Norm and a convolution layer with a convolution kernel size of 2 steps and a distance of 2. The primary purpose is to downsample input data and to reduce the spatial dimension of the data, thereby reducing the amount of computation and parameters of the network while increasing the network's receptive field and improving the network's generalization ability.

Pooling layer: The pooling layer uses global average pooling. The feature map is dimensionally reduced to convert spatial information in the height and width directions into features in the channel direction.

Full connection layer: Finally, the characteristics output from ATNet and ACNN are combined and input to the full connection layer for fatigue state classification.

4. Data Preparation

4.1. Experimental subjects

In this experiment, 14 healthy subjects participated in the complete investigation, all aged between 23 and 26 years, including 13 males and one female. The investigation requires that all subjects, be physically and mentally healthy, have no mental illness, have normal vision, obtain a driving license, and have some driving experience. To complete the experiment, the subjects also need to pay attention to the following points throughout the investigation:

1. All subjects need to ensure normal work and rest one week before the start of the experiment. Do not exercise vigorously or work excessively, ensure adequate sleep, and do not stay up late.
2. Do not drink caffeinated or alcoholic beverages before the start of the experiment.
3. The subjects who participated in the experiment in the morning should go to bed on time the night before and not stay up late to ensure sufficient sleep. The subjects who participated in the investigation in the afternoon or evening had a normal rest at noon, and those who had a habit of napping had a normal lunch break.
4. You must wash and blow dry your hair 1 h before the experiment begins. Before charging the electrode cap, the hair is dry and should not be too hot or wet.
5. Before the experiment begins, subjects must undergo a pre-test to ensure they understand and are familiar with the experimental process. They feel well due to the driving environment or instrumentation.
6. The entire experiment takes about 150–180 min. Before the investigation, eating regularly and using the bathroom correctly were necessary. The whole process is not allowed to play with mobile phones or do other things.
7. During the experiment, corresponding tasks and tests must be completed, and the subjects should complete them as required. If there are any questions, they should communicate with relevant personnel promptly.
8. Each subject must complete three driving tasks: low, medium, and high.
9. The experimental process is tedious and time-consuming, requiring patience, not impatience, and being able to complete the experiment usually.

4.2. Experimental protocol

In order to ensure the safety of the subjects and ensure that the collected data is closer to reality. In this experiment, a simulated driving platform was built, and a simulated driver was purchased to make the subjects feel like they were driving in real life. The entire simulated driving platform is shown in Figure 9:

Each participant needs to complete three different driving tasks: low load, medium load, and high load, with each experiment completing one task. Low-load task driving scenario: Driving on spacious roads with a speed limit of below 80 km/h (to avoid collisions), participants can follow the runway and enjoy the scenery. Medium-load task driving scenario: Driving on a highway loop in a city involves tasks such as entering a tunnel and avoiding vehicles traveling back and forth, requiring participants to maintain attention at all times. High-load task driving scenario: Driving on rugged mountain roads (one-way mountain climbing), maintaining attention and speed control at all times to avoid collisions. Three different driving scenarios were simulated to simulate different driving environments in real situations, and our goal was to collect EEG signals from each participant at different levels of fatigue. The entire experiment takes about 175 min, and the specific details are shown in Figure 10:

Figure 9
Simulated driving platform

1.0–15, 75–90, and 150–165 min: Subjects are required to fill out the fatigue self-test scale and the Dundee stress scale respectively in the first, middle, and last stages of the driving task and complete behavioral testing tasks such as Pyramid Vision Transformer (PVT) and decision deviation, with the purpose of assisting in judging the fatigue status of the subjects. In addition, we placed a camera in front of the subjects to collect video face data.

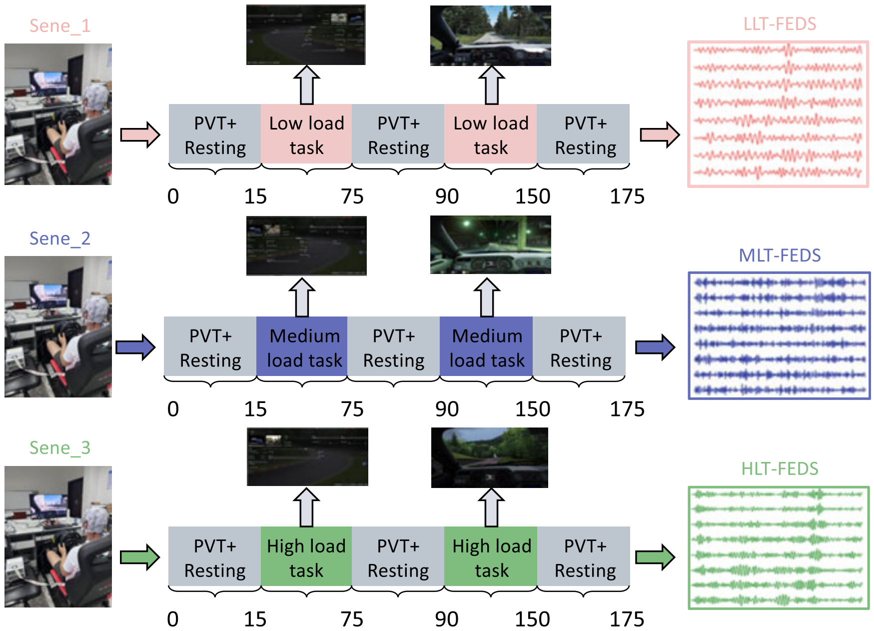
2.15–75 and 90–150 min: In these two stages, subjects only need to drive the vehicle to complete corresponding driving tasks.

4.3. Data acquisition and pre-processing

The EEG signal of this experiment was obtained using internationally recognized Back Propagation (BP) acquisition equipment. In addition to one reference electrode, data from 31 electrode channels can be obtained at a sampling frequency of 1000 Hz. In the experiment of three driving tasks, each subject must complete a two h driving tasks, collecting 7200s of EEG data and totaling 7,200,000 sampling points.

This experiment extracts eye movement features from collected face images, and PERCLOS indicators are calculated as data labels (Du et al., 2022). The specific formula is as follows:

Figure 10
EEG signal acquisition experimental process



$$PERCLOS = \frac{blink + CLOS}{interval} \quad (4)$$

where *interval* means the total time, including blinking, closing, and other states. The calculation formula is as follows:

$$interval = blink + CLOS + fixation + saccade \quad (5)$$

where *fixation* and *saccade* mean the gaze and saccade states of the eye.

The three datasets are labeled every 10 s, totaling (7200/10) 720 tags. The tag value is between 0 and 1, with a smaller value indicating that the subject is more awake and has a lower level of fatigue. In comparison, a more significant value indicates a higher level of fatigue. Labels are divided into two categories based on PERCLOS values, with 0–0.35 indicating an awake state and 0.35–1 indicating a fatigued state.

We used a validated method to pre-process the three collected datasets. Specifically, using a bandpass filter to filter out raw data from 1 to 60 Hz, ICA (Jung et al., 2000) is used to remove noise and artifacts from the data. The pre-processing process is completed based on the EEGLAB toolkit (Delorme & Makeig, 2004). Due to the excessive data corruption of one subject, the data of 13 subjects were used for subsequent research. The three datasets are fatigue EEG dataset for low-load tasks (LLT-FEDS), fatigue EEG dataset for medium-load tasks (MLT-FEDS), and fatigue EEG data assigned for high-load tasks (HLT-FEDS).

5. Experimental Results and Analysis

This article has conducted many experiments on three datasets: LLT-FEDS, MLT-FEDS, and HLT-FEDS. First, we performed different verifications on the time window length for dividing EEG segments and selected the method with the best effect. Then, this article uses the HLT-FEDS dataset to do a model ablation experiment, proving that each step of this method benefits the fatigue detection effect. Then, this article compares the proposed model and some popular models in recent years on three datasets to verify the performance of the proposed method. Finally, this paper demonstrates the impact of different frequency band combinations on fatigue detection classification.

5.1. Model performance evaluation index

In order to demonstrate the performance of the proposed model through experimental results, we used the five most commonly used evaluation indicators in the field of deep learning. They are accuracy, precision, recall, special effects, and F1 score. Firstly, regarding the concepts of TP, TN, FP, and FN, TP represents positive sample determination as positive, TN represents negative sample determination as negative, FP represents negative sample determination as positive, and FN represents positive sample determination as negative.

Among them, accuracy refers to the probability of predicting positive and negative categories in the total number of samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Precision refers to the proportion of correct predictions being positive to all predictions being positive.

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

Recall rate refers to the proportion of correctly predicted positive results to all actual positive results.

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

Specificity refers to the probability of accurately predicting among all negative categories.

$$Specificity = \frac{TN}{FP + TN} \quad (9)$$

F1score is the harmonic average of accuracy and recall.

$$F1score = \frac{2*Precision*Recall}{Precision + Recall} \quad (10)$$

5.2. Method introduction time window comparative experiment

In constructing 3D features, this article divides each EEG segment into time windows of the same length for subsequent operations. Many experiments have been conducted on time window partitions of 0.5, 1, 2, and 3 s to verify and select the most suitable time window partition length. As shown in Table 2, the impact of different partition methods on various performance indicators is shown.

The results in Table 2 show that when the time window is set to 1 second, the proposed model achieves the best performance, with only one indicator of accuracy lower than the time window setting of 0.5 s and the other four indicators achieving the best results. The model's performance gradually deteriorates as the time window length increases. The setting of the time window length significantly impacts the final classification result. If it is too short, it can easily split the features of the signal and cannot effectively learn the time dependence in the EEG signal. If it is too long, it can easily lead to inaccurate extracted features. Therefore, in the later experiment of this article, the time window length is set to 1 second.

In addition, after many experiments, other parameters most beneficial to fatigue detection performance have also been selected. The training iteration is set to 200 rounds, the learning rate is 0.001, and the data per batch is 128. The Mean Squared Error (MSE) loss function is assigned to calculate the loss value, and the optimizer chooses the Adam optimizer. In the proposed model, a variety of activation function, such as Relu and Gelu, and convolution kernels of different sizes, such as 1 * 1, 3 * 3, and 7 * 7, are also used.

Table 2
Experimental results of different time window partitioning

Time window	Specificity	Recall	Precision	F1-Score	Accuracy
0.5 s	0.8975	0.8398	**0.7812**	0.8088	0.8733
1s	**0.9011**	**0.8436**	0.7792	**0.8178**	**0.8755**
2s	0.8702	0.8139	0.7607	0.7811	0.8593
3s	0.8594	0.7981	0.7403	0.7681	0.8389

5.3. Model ablation experiment

This article conducts ablation experiments on the proposed model based on the parameter settings in Section 5.2, in order to verify that each part of the model has a positive impact on fatigue detection. This article first verifies the effectiveness of constructing 3D features and then proves the impact of ATNet and ACNN on classification detection. The experimental results are shown in Table 3. The specific operation is as follows:

First, the original signal without pre-processing and feature extraction is input into STCNN and iterated 200 times. The final classification accuracy of the model is 72.46%, and several other indicators are also relatively low. The main reason is that the unprocessed signal contains much noise and is contaminated by various artifacts. Although STCNN can learn some spatiotemporal features and achieve individual classification results, it is only possible for STCNN to perform well by directly inputting processed EEG signals into the model.

Next, this paper divides the original signal into a time window of 1 second, performs Butterworth bandpass filtering on the signal to obtain five common frequency bands, and extracts PSD features of each of the five frequency bands to get one-dimensional frequency domain features. After repeated training and verification, the classification accuracy rate has risen to about 80.32%. The other indicators also have varying degrees of improvement due to the pre-processing of raw data and PSD feature extraction, indicating that extracting one-dimensional features improves the model's classification performance.

If only 1D features are input into the ATNet model, only 78.52% classification accuracy can be achieved. This indicates that the classification effect is affected when removing the ACNN module. The ACNN model is mainly used to fuse multiple features and further learn more valuable information, which is beneficial to improving classification accuracy.

Then, this article inputs the constructed 1D and 3D features into the ACNN network, respectively. If 1D features are input, an accuracy rate of 81.08% can be achieved. If you input 3D features, all indicators have significantly improved, with an accuracy rate of 84.57%. The experimental results show that 3D features include frequency and spatial domain features, which provide a more comprehensive analysis of EEG signals than 1D feature inputs, and ultimately achieve good results.

Finally, this article inputs 1D and 3D features into the STCNN network. Specifically, the proposed method achieves the best classification performance by inputting 1D features into the ATNet module and 3D parts into the ACNN module.

The experimental results show that the pre-processing operation of EEG data cannot be ignored, and removing pollution from the signal can effectively improve the quality of the signal. Simply analyzing the characteristics of EEG signals can achieve specific results, but the accuracy rate is low. Therefore, a more comprehensive analysis of various aspects of EEG signals is needed. The 1D features and 3D features constructed in this article can be input into ATNet and ACNN, respectively, which can effectively extract the spatiotemporal features of EEG signals and improve the accuracy of fatigue state classification.

5.4. Comparative experiment

To verify the performance of the proposed model, this article compares it with six highly competitive models. These models are EEGConv (Zeng et al., 2018), EEGConvR (Zeng et al., 2018), ESTCNN (Gao et al., 2019), R3DCNNs (Zhang et al., 2019), 3D-CNN (Lin et al., 2020), and 3D-LSTM (Lin et al., 2020). This article reproduces these models by introducing models, parameters, and some details in the original literature. Then, use the same method to conduct experiments on three datasets: LLT-FEDS, MLT-FEDS, and HLT-FEDS. The results are shown in Tables 4, 5, and 6.

By synthesizing the results of the three tables, it can be seen that among the seven models, except for the ESTCNN, all the other models have an accuracy rate of over 80%. The accuracy of the three models, 3D-CNN, 3D-LSTM, and STCNN, has reached over 85%. According to Table, the method proposed in this paper

Table 3
Ablation experiment of 3D-STCNN model

Methods	Specificity	Recall	Precision	F1-Score	Accuracy
STCNN+Raw	0.6031	0.6360	0.6740	0.6825	0.7246
STCNN+1D	0.6802	0.7258	0.7522	0.7438	0.8032
ATNet+1D	0.7006	0.7417	0.7555	0.7749	0.7852
ACNN+1D	0.7011	0.7343	0.7686	0.8207	0.8108
ACNN+3D	0.7565	0.7755	0.7970	0.8682	0.8457
STCNN+1D+3D	**0.7792**	**0.8178**	**0.8436**	**0.9011**	**0.8755**

Table 4
Experimental results based on the LLT-FEDS dataset

Methods	Specificity	Recall	Precision	F1-Score	Accuracy
ESTCNN (Gao et al., 2019)	0.6079	0.6524	0.7003	0.7965	0.7708
EEGConvR (Zeng et al., 2018)	0.6643	0.6966	0.7443	0.8101	0.8136
EEGConv (Zeng et al., 2018)	0.6886	0.7264	0.7823	0.8321	0.8238
R3DCNNs (Zhang et al., 2019)	0.7358	0.7484	0.7811	0.8555	0.8333
D-CNN (Lin et al., 2020)	0.7685	0.7977	0.8404	0.8792	0.8658
D-LSTM (Lin et al., 2020)	0.7678	0.7913	0.8195	0.8633	0.8587
STCNN	**0.7746**	**0.8035**	**0.8422**	**0.8894**	**0.8715**

Table 5
Experimental results based on MLT-FEDS dataset

Methods	Specificity	Recall	Precision	F1-Score	Accuracy
ESTCNN (Gao et al., 2019)	0.6125	0.6556	0.7021	0.7977	0.7802
EEGConvR (Zeng et al., 2018)	0.6687	0.7003	0.7498	0.8132	0.8228
EEGConv (Zeng et al., 2018)	0.6908	0.7277	0.7843	0.8335	0.8267
R3DCNNs (Zhang et al., 2019)	0.7403	0.7506	0.7844	0.8578	0.8361
D-CNN (Lin et al., 2020)	**0.7794**	0.8008	0.8431	0.8847	0.8718
D-LSTM (Lin et al., 2020)	0.7685	0.7934	0.8201	0.8712	0.8668
STCNN	0.7763	**0.8065**	**0.8456**	**0.8987**	**0.8737**

Table 6
Experimental results based on HLT-FEDS dataset

Methods	Specificity	Recall	Precision	F1-Score	Accuracy
ESTCNN (Gao et al., 2019)	0.6187	0.6645	0.7079	0.8004	0.7820
EEGConvR (Zeng et al., 2018)	0.6722	0.7063	0.7554	0.8115	0.8237
EEGConv (Zeng et al., 2018)	0.6832	0.7343	0.7806	0.8320	0.8327
R3DCNNs (Zhang et al., 2019)	0.7465	0.7579	0.7863	0.8558	0.8397
D-CNN (Lin et al., 2020)	**0.7817**	0.8081	**0.8456**	0.8863	0.8732
D-LSTM (Lin et al., 2020)	0.7792	0.8005	0.8289	0.8745	0.8693
STCNN	0.7792	**0.8178**	0.8436	**0.9011**	**0.8755**

achieves the best performance in EEG datasets based on high-load driving tasks, with an accuracy rate of up to 87.55%. Although the recall rate and accuracy are lower than those of 3D-CNN, the other three indicators are all higher. In addition, the proposed method performs better in the other two datasets. In addition, the three datasets are based on EEG datasets collected under different task loads, which means that participants participating in experiments with different tasks have different fatigue levels during the actual experimental process. When subjects complete relatively easy and low-load driving functions, they feel less tired and sleepy. When completing a high-load task, the subject will feel very tired, and due to the high concentration of mental energy for a long time, the degree of fatigue will be heavier. This is also similar to real life, where different people have significantly different fatigue levels due to personal reasons, work stress, or duration of driving time.

Therefore, the algorithm's stability in different situations is also worth studying. The experimental results in the three tables show that all models perform less well on the dataset of low-load tasks than on the dataset of high-load assignments. The accuracy of the STCNN model is 87.15%, 87.37%, and 87.55%, respectively, with a fluctuation range of 0.4 percentage points. The fluctuation ranges of the six models of ESTCNN, EEGConvR, EEGConv, R3DCNNs, 3D-CNN, and 3D-LSTM are 1.12, 1.01, 0.89, 0.64, 0.74, and 1.06 percentage points, respectively. According to these data, compared to mild fatigue, the more severe the fatigue, the higher the classification accuracy of any model. This indicates that the degree of fatigue does affect the final test results. The fluctuation range of classification accuracy shows that the fluctuation range of the model proposed in this article is only 0.4 percentage points, which is the smallest among all models. Therefore, if we can comprehensively analyze the EEG signal from data pre-processing and feature extraction to classification and propose a model with better performance, the stability of the final category can be guaranteed in either case.

5.5. Frequency band experiment

In order to evaluate the impact of frequency bands (such as Delta) on fatigue state recognition, this article uses the same method to construct features separately or in combination for 5 frequency bands and then inputs them into the proposed model for experimentation. The performance of this experiment in the HLT-FEDS dataset is shown in Table 7.

Table 7
Experimental results in different frequency bands

Bands	Specificity	Recall	Precision	F1-Score	Accuracy
δ	0.7221	0.7374	0.7412	0.7956	0.7689
θ	0.7584	0.7797	0.7816	0.8329	0.8047
α	0.7635	0.7803	0.7898	0.8479	0.8131
β	0.7884	**0.8167**	**0.8155**	0.8642	0.8309
γ	**0.8004**	0.8089	0.8072	**0.8711**	**0.8406**
$\beta+\gamma$	0.7663	0.8067	0.8336	0.8857	0.8655
$\alpha+\beta+\gamma$	0.7722	0.8103	0.8387	0.8873	0.8678
$\theta+\alpha+\beta+\gamma$	0.7709	**0.8194**	**0.8482**	0.8975	0.8742
All	**0.7792**	0.8178	0.8436	**0.9011**	**0.8755**

From the results in the table, it can be seen that when constructing features separately for 5 frequency bands, the higher the frequency of the frequency band, the higher the final classification accuracy, among them, Beta (β) And Gamma (γ) The performance is similar, but significantly higher than the classification performance of the other three frequency bands. It is not difficult to find that the features constructed in a single frequency band have overall low performance and the highest accuracy is only about 84%, which has a significant impact on the recognition of fatigue states.

In addition, this article conducts combination experiments on different frequency bands, and there are many combination methods, which are not limited to this. This article mainly selects Beta (β) And Gamma (γ) Two frequency bands are combined with other frequency bands. From the results, it can be seen that the overall performance of the multi-band feature construction method is significantly better than that of the single-band feature construction method. The combination of five frequency bands showed the best performance on four indicators, and the combination of four frequency bands had one indicator that was the best. Therefore, it is not difficult to find that the combination of frequency bands has a significant impact on fatigue state recognition, and multi-frequency bands have better performance.

This is an experiment that analyzes the importance of frequency bands from the perspective of neural networks. In experiments with a single frequency band, Beta (β) And Gamma (γ), the frequency band has the most significant impact on the fatigue classification results. The overall performance of the multi-band combination method is better and more stable, and it is hoped that this study can provide some new ideas for future research on fatigue detection.

6. Conclusions

This paper evaluates a method for constructing 3D features and proposes a STCNN for classifying fatigue states. 3D features include frequency domain features and spatial features. STCNN consists of ATNet and ACNN, which can further focus on valuable information in EEG signals, thereby improving classification accuracy. This paper uses the proposed method to conduct much experimental verification on three datasets: LLT-FEDS, MLT-FEDS, and HLT-FEDS. The experimental results show that the feature extraction method can extract the spatiotemporal features of EEG signals, and the proposed STCNN can improve classification performance. Moreover, it can maintain good stability in different fatigue classification tasks, and this method is superior to some existing fatigue detection methods based on deep learning. Finally, this paper uses the proposed approach to verify the impact of different frequency band combinations on classification performance. The method proposed in this paper mainly analyzes the spatiotemporal characteristics of EEG signals, which improves the classification accuracy of fatigue detection to a certain extent, providing some ideas for future research.

However, current research has certain limitations. Firstly, due to the expensive and cumbersome wearing of traditional EEG devices, EEG data collection could be better; Secondly, the complexity of the algorithm proposed in this paper is relatively high, and processing data requires a certain amount of time. Therefore, the real-time performance could be better. Finally, due to inconsistent methods of processing data, it is difficult to compare different studies. Consequently, it is necessary to specify future unified data processing and feature extraction standards.

Funding Support

This work is supported by the National Natural Science Foundation of China under Grant No. 62272067; the Sichuan Science and Technology Program under Grant Nos. 2023NSFSC0499, 2023YFG0018; the LOST 2030 Brain Project No. 2022ZD0208500; the Scientific Research Foundation of Chengdu University of Information Technology under Grant Nos.KYQN202208, KYQN202206; and the 2011 Collaborative Innovation Center for Image and Geospatial Information of Sichuan Province.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

Abu Farha, N., Al-Shargie, F., Tariq, U., & Al-Nashash, H. (2022). Improved cognitive vigilance assessment after artifact reduction with wavelet independent component analysis. *Sensors*, *22*(8), 3051.

Bashivan, P., Rish, I., Yeasin, M., & Codella, N. (2015). Learning representations from EEG with deep recurrent-convolutional neural networks. *arXiv Preprint*:1511.06448.

Chi, Y. M., Wang, Y. T., Wang, Y., Maier, C., Jung, T. P., & Cauwenberghs, G. (2012). Dry and noncontact EEG sensors for mobile brain–computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *20*(2), 228–235.

Coetzer, R. C., & Hancke, G. P. (2011). Eye detection for a real-time vehicle driver fatigue monitoring system. In *2011 IEEE Intelligent Vehicles Symposium, (IV)*, 66–71.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Demiralp, T., Ademoglu, A., Comerchero, M., & Polich, J. (2001). Wavelet analysis of p3a and p3b. *Brain Topography*, *13*(4), 251–267.

Djamal, E. C., Abdullah, M. Y., & Renaldi, F. (2017). Brain computer interface game controlling using fast fourier transform and learning vector quantization. *Journal of Telecommunication, Electronic and Computer Engineering*, *9*(2–5), 71–74.

Du, G., Wang, Z., Li, C., & Liu, P. X. (2021). A TSK-type convolutional recurrent fuzzy network for predicting driving fatigue. *IEEE Transactions on Fuzzy Systems*, *29*(8), 2100–2111.

Du, G., Zhang, L., Su, K., Wang, X., Teng, S., & Liu, P. X. (2022). A multimodal fusion fatigue driving detection method based on heart rate and perclos. *IEEE Transactions on Intelligent Transportation Systems*, *23*(11), 21810–21820.

Gao, Z., Wang, X., Yang, Y., Mu, C., Cai, Q., Dang, W., & Zuo, S. (2019). EEG-based spatio–temporal convolutional neural network for driver fatigue evaluation. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(9), 2755–2763.

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., Mckeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(2), 163–178.

Lin, M. W., Ruan, S. J., & Tu, Y. W. (2020). A 3DCNN-LSTM hybrid framework for sEMG-based noises recognition in exercise. *IEEE Access*, *8*, 162982–162988.

Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A ConvNet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11976–11986.

Mandavifar, S., & Ghorbani, A. A. (2019). Application of deep learning to cybersecurity: A survey. *Neurocomputing*, *347*, 149–176.

Marino, M., Liu, Q., Koudelka, V., Porcaro, C., Hlinka, J., Wenderoth, N., & Mantini, D. (2018). Adaptive optimal basis set for BCG artifact removal in simultaneous EEG-fMRI. *Scientific Reports*, *8*(1), 8902.

Miao, M., Zeng, H., Wang, A., Zhao, C., & Liu, F. (2017). Discriminative spatial-frequency-temporal feature extraction and classification of motor imagery EEG: An sparse regression and weighted nave Bayesian classifier-based approach. *Journal of Neuroscience Methods*, *278,* 13–24.

Peng, B., Zhang, Y., Wang, M., Chen, J., & Gao, D. (2023). TA-MFFNet: Multi-feature fusion network for EEG analysis and driving fatigue detection based on time domain network and attention network. *Computational Biology and Chemistry, 104*.

Racioppi, F., Eriksson, L., Tingvall, C., & Villaveces, A. (2004). *Preventing road traffic injury: A public health perspective for Europe*. Denmark: World Health Organization Regional Office for Europe.

Risqiwati, D., Wibawa, A. D., Pane, E. S., Islamiyah, W. R., Tyas, A. E., & Purnomo, M. H. (2020). Feature selection for EEG-based fatigue analysis using Pearson correlation. In *2020 International Seminar on Intelligent Technology and Its Applications,* 164–169.

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., . . . , & Ball, T. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, *38*(11), 5391–5420. https://doi.org/10.1002/hbm.23730.

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, *45*(11), 2673–2681.

Shalsh, W. M. (2021). A deep learning CNN model for driver fatigue detection using single Eeg channel. *Journal of Theoretical and Applied Information Technology*, *99*(2), 462–477.

Sharma, S., Khare, S. K., Bajaj, V., & Ansari, I. A. (2021). Improving the separability of drowsiness and alert EEG signals using analytic form of wavelet transform. *Applied Acoustics*, *181*, 108164.

Sheykhivand, S., Rezaii, T. Y., Meshgini, S., Makoui, S., & Farzamnia, A. (2022). Developing a deep neural network for driver fatigue detection using EEG signals based on compressed sensing. *Sustainability*, *14*(5), 2941.

Sikander, G., & Anwar, S. (2019). Driver fatigue detection systems: A review. *IEEE Transactions on Intelligent Transportation Systems*, *20*(6), 2339–2352.

Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision*, 3–19.

Yang, Y., Wu, Q., Fu, Y., & Chen, X. (2018). Continuous convolutional neural network with 3D input for EEG-based emotion recognition. In *Neural Information Processing: 25th International Conference,*, 433–443.

Zeng, H., Yang, C., Dai, G., Qin, F., Zhang, J., & Kong, W. (2018). EEG classification of driver mental states by deep learning. *Cognitive Neurodynamics*, *12*(6), 597–606.

Zhang, P., Wang, X., Zhang, W., & Chen, J. (2018b). 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.

Zhang, S., Zhang, Z., Chen, Z., Lin, S., & Xie, Z. (2021a). A novel method of mental fatigue detection based on CNN and LSTM. *International Journal of Computational Science and Engineering*, *24*(3), 290–300.

Zhang, Y., Cao, W., Feng, L., Wang, M., Geng, T., Zhou, J., & Gao, D. (2023). SHNN: A single-channel EEG sleep staging model based on semi-supervised learning. *Expert Systems with Applications*, *213*, 119288.

Zhang, Y., Qiao, S., Zeng, Y., Gao, D., Han, N., & Zhou, J. (2021b). CAE-CNN: Predicting transcription factor binding site with convolutional autoencoder and convolutional neural network. *Expert Systems with Applications*, *183*, 115404.

Zhao, Y., Yang, J., Lin, J., Yu, D., & Cao, X. (2020). A 3D convolutional neural network for emotion recognition based on EEG signals. In *2020 International Joint Conference on Neural Networks*, 1–6.

Zheng, Y., Ma, Y., Cammon, J., Zhang, S., Zhang, J., & Zhang, Y. (2022). A new feature selection approach for driving fatigue EEG detection with a modified machine learning algorithm. *Computers in Biology and Medicine*, *147*.

How to Cite: Peng, B., Gao, D., Wang, M., & Zhang, Y. (2024). 3D-STCNN: Spatiotemporal Convolutional Neural Network Based on EEG 3D Features for Detecting Driving Fatigue. *Journal of Data Science and Intelligent Systems* 2(1), 137–149, https://doi.org/10.47852/bonviewJDSIS3202983