

Feature Extraction Method of EEG Signals Evaluating Spatial Cognition of Community Elderly With Permutation Conditional Mutual Information Common Space Model

Dong Wen, *Member, IEEE*, Bingbing Liang, Jingjing Li, Lingyu Wu, Xianglong Wan, *Member, IEEE*, Xianling Dong^{ib}, Xifa Lan, Haiqing Song, and Yanhong Zhou^{ib}

Abstract—In order to improve the traditional common space pattern (CSP) algorithm pattern in EEG feature extraction, this study proposes a feature extraction method of EEG signals based on permutation conditional mutual information common space pattern (PCMICSP), which used the sum of the permutation condition mutual information matrices of each lead to replacing the mixed spatial covariance matrix in the traditional CSP algorithm, and its eigenvectors and eigenvalues are used to construct a new spatial filter. Then the spatial features in the different time domains and frequency domains are combined to construct the two-dimensional pixel map. Finally, a convolutional neural network (CNN) is used for binary classification. The EEG signals of 7 community elderly before and after spatial cognitive training in virtual reality (VR) scenes were used as the test data set. The average classification accuracy of the PCMICSP algorithm for pre-test and post-test EEG signals is 98%, which was higher than that of CSP based on CMI (conditional mutual information), CSP based on MI (mutual

information), and traditional CSP in the combination of four frequency bands. Compared with the traditional CSP method, PCMICSP can be used as a more effective method to extract the spatial features of EEG signals. Therefore, this paper provides a new approach to solving the strict linear hypothesis of CSP and can be used as a valuable biomarker for the spatial cognitive evaluation of the elderly in the community.

Index Terms—Permutation conditional mutual information common space pattern, EEG signals, community elderly, spatial cognitive evaluation, virtual reality.

I. INTRODUCTION

SPATIAL cognition is a relatively independent spatial element, which uses visual memory encoding to understand the higher cognitive function of the spatial environment. Spatial cognitive impairment can lead to many diseases, such as Alzheimer's disease [1]. The incidence rate of Alzheimer's disease is increasing along with the aging of the population, which can lead to memory loss, impaired attention, impaired problem-solving ability, and varying degrees of degeneration in spatial memory, visual-spatial structure, orientation, and other abilities in older adults [2]. At present, there is no feasible specific drug for such problems. Considering that community hospitals can provide auxiliary training services with portable devices, it is an urgent problem to explore and promote the diagnosis and treatment methods of spatial cognitive impairment of the elderly based on the spatial cognitive training and evaluation of the elderly in the community. Also, the development of spatial cognition training and evaluation methods is of great significance to prevent the decline of spatial cognition in healthy subjects. Relevant research results show that there is a close correlation between spatial cognition and EEG signals [3], [4]. EEG signals are often used to evaluate the correlation between spatial cognitive function and brain response.

In recent years, spectral analysis [5], [6], [7], [8], blind source separation (BSS) [9], [10], brain region synchronization analysis [11], [12], [13], [14], [15], event-related potential (ERP) [16], [17], [18], and brain function network analysis [15], [19], [20] have been mainly used to explore the changes

Manuscript received 26 May 2022; revised 27 October 2022; accepted 29 November 2022. Date of publication 4 May 2023; date of current version 25 May 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62276022, Grant 62206014, Grant 61876165, and Grant 61503326; and in part by the National Key Research and Development Program of China under Grant 2021YFF1200603. (Corresponding author: Yanhong Zhou.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of First Hospital of Qinhuangdao in Hebei Province, China, under Application No. 2018B006, in 2018.

Dong Wen, Lingyu Wu, and Xianglong Wan are with the School of Intelligence Science and Technology, University of Science and Technology Beijing, Haidian, Beijing 100083, China.

Bingbing Liang and Jingjing Li are with the Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province, School of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China.

Xianling Dong is with the Hebei Key Laboratory of Nerve Injury and Repair, Chengde Medical University, Chengde 067050, China.

Xifa Lan is with the Department of Neurology, First Hospital of Qinhuangdao, Qinhuangdao 066000, China.

Haiqing Song is with the Department of Neurology, Xuanwu Hospital, Capital Medical University, Beijing 100053, China.

Yanhong Zhou is with the School of Mathematics and Information Science and Technology, Hebei Normal University of Science and Technology, Qinhuangdao, Hebei 066004, China (e-mail: yhzhou168@163.com).

Digital Object Identifier 10.1109/TNSRE.2023.3273119

of large brain related to spatial cognition. However, these methods mainly analyze spatial cognitive EEG signals from the correlation perspective of the time domain and frequency domain, without considering the spatial characteristics. As we all know, spatial features play an important role in the field of two classification analyses of EEG signals, which can describe the spatial distribution of multi-lead EEG signals in each category [21]. CSP is the most widely used spatial filtering technique, which can effectively extract the features of different states of the brain [22], [23]. The classical CSP algorithm focuses on the linear relationship between EEG signals of different channels [24], while EEG signals are nonlinear and non-stationary [25]. Therefore, a few researchers have introduced the nonlinear relationship of EEG signal into the CSP algorithm to extend the nonlinear spatial filter. Sun and Zhang [26] proposed a kernel CSP (KCSP) method based on kernel optimization feature extraction. However, the KCSP method is difficult to generalize because it requires the same sample size for the two kinds of data. Nasihatkon et al. [27] proposed to use linear and KCSP to construct spatial filters. However, this method requires a large amount of kernel matrix eigenvalue decomposition and high computational complexity.

Mutual information (MI) uses the concept of entropy to quantify the nonlinear synchronization relationship between EEG signals [28]. There is high demand for enough data length, associated with good statistical significance [29]. Moreover, the coupling strength of mutual information calculation is unsigned, which is not enough to indicate the effect driving between brain regions [30]. Conditional mutual information (CMI) quantifies the relationship between the two variables and eliminates the influence of the third variable, which can better reflect the information drive between brain regions [30], but the noise of EEG signals will have a great impact on the results [31]. Permutation condition mutual information (PCMI) [32] can measure the linear or nonlinear coupling strength of two-time series, and it has strong robustness to signal noise. Relevant studies show that the analysis performance of PCMI on EEG signals is better than CMI [32]. Yuan [33] used PCMI to extract the features of EEG signals before and after spatial cognitive training. Through the effective evaluation of training results, it has been proved that PCMI is effective in analyzing spatial cognitive EEG signals.

This study proposed a permutation conditional mutual information common space pattern (PCMICSP) algorithm. The mixed space covariance matrix of two data series, calculated by the traditional CSP method, is replaced by the mixed spatial permutation conditional mutual information matrix. So that the obtained spatial filter contains both linear and nonlinear features. The algorithm is verified by comparing the EEG signals of the elderly before and after spatial cognitive training. First, the PCMICSP method is used to calculate the spatial features of multiple frequency bands, and then the spatial features in the different time- and frequency domains are combined. Finally, we use CNN to compare the performance of the proposed PCMICSP with MI CSP (MICSP), CMI CSP (CMICSP), and traditional CSP algorithms, respectively.

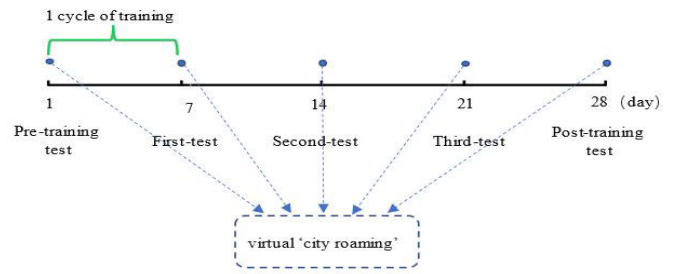


Fig. 1. Experimental training time point distribution.

II. MATERIALS AND METHODS

A. Subject Information

Supported by the Ethics Committee of Qinhuangdao First People's Hospital, 7 healthy subjects with an average age of 67 ± 6.81 years were recruited from nearby communities to participate in virtual reality spatial cognition training and testing tasks for 28 days [34]. The subjects signed the informed consent, had normal visual acuity or corrected visual acuity, had no history of major mental illness and had never participated in such spatial memory game training. The data collection was approved by the Ethics Committee of First Hospital of Qinhuangdao in Hebei Province, China (The approval number is 2018B006 in 2018).

In this study, two popular and widely used scales, Minimal State Examination (MMSE) [35] and Montreal Cognitive Assessment (MoCA) [36], were used to determine whether the subjects had cognitive dysfunction. The mean MMSE of the subjects was 28.29 ± 1.25 and the mean MoCA was 24.7 ± 2.56 . The MMSE results showed that the cognitive function of the subjects was normal. Three spatial cognition scales, namely Guilford Zimmerman spatial orientation test (GZSOT) [37], Perspective Taking /Spatial Orientation Test (PTSOT) [38], and Corsi Block-Tapping Task (CBTT) [39], were used to detect the users' spatial positioning ability.

B. Experimental Design Scheme

In this paper, the 'virtual community' spatial cognitive training game and the virtual 'city roaming' test task, designed in literature [40], were used to conduct the spatial cognitive training task for 28 days in a relatively quiet laboratory. The experiment was carried out in four cycles, each cycle of training for seven days. The training process (as shown in Fig. 1) is divided into three stages.

(1) Pre-training test. Before the first training cycle, all subjects wore electrode caps and head-mounted VR glasses to conduct a pre-training spatial cognitive ability assessment test using a spatial cognitive scale. During the test, the EEG signals of the subjects were mainly collected.

(2) Training and testing. The subjects used 'virtual community' training during each cycle. To ensure the participation enthusiasm of the elderly, the experimental group used the three-target points version of the virtual community in the first week and the six-target points version in the last three weeks. The training session recorded the time spent by the subjects to complete each task.

(3) Post-training test. Three spatial cognition scales (GZSOT, PTSOT, CBTT) were used to evaluate the training



Fig. 2. VCW test route.

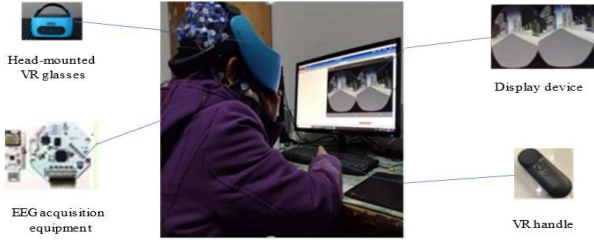


Fig. 3. Experimental setup and scenario.

effect of all subjects, and the VCW test route was used to complete the post-training task. The evaluation results of the spatial cognition scale, test game behavior data, and EEG signals were archived to evaluate the effectiveness of training in the later stage. The VCW test route is shown in Fig. 2. The test task is divided into two modes: learning mode and testing mode. In the learning stage, users are required to try their best to remember reference buildings and routes along the virtual city in accordance with certainly visible route guidance. This process is the coding learning stage. In the test stage, the original route guide is hidden, and then the user is asked to repeat the original route based on spatial memory and positioning ability. To keep the difficulty of each test consistent, each test path is treated symmetrically.

Fig. 3 indicates the experimental setup and scenario for the spatial cognition test, which consists of an OpenBCI EEG acquisition device, HTC Vive Focus headset VR glasses, supporting control handles, and user perspective. OpenBCI device is used to collect users' EEG signals, and bluetooth communication is realized between OpenBCI USB dongle and EEG acquisition client. HTC Vive Focus head-mounted VR glasses provide users with virtual scenes required by test tasks, and the matching control handle is used for the control required in tasks. In addition, TCP/IP protocol is used to interact with behavioral data acquisition clients to complete data transfer. The subject view can be shared by using the screen assistant of 360 mobile assistant software in USB debugging mode of headset VR glasses, so that the user's perspective in headset VR glasses can be shared in time, which is convenient for experimental staff to guide their work.

C. Recording and Preprocessing of EEG Signals

1) *Recording*: The sampling rate is 125Hz and the electrode impedance is less than 10k Ω . Two 8-channel Cyton amplifiers are used to realize the synchronous acquisition of 16-channel EEG signals, and the data is transmitted to the computer via

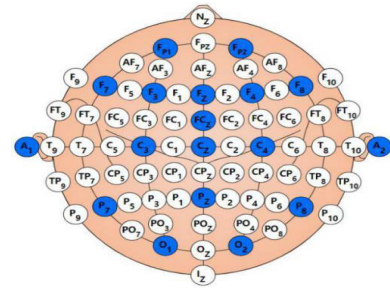


Fig. 4. Sampling electrode position distribution.

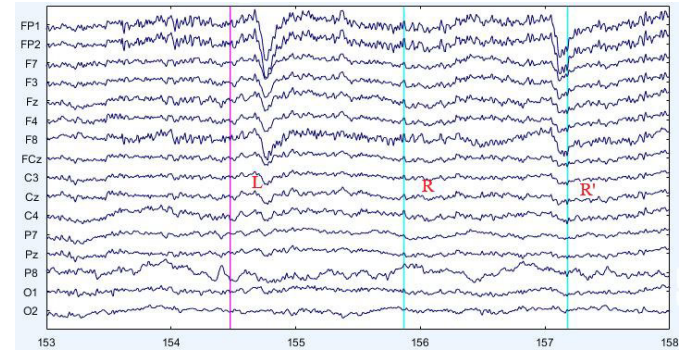


Fig. 5. Synchronous acquisition results of EEG signals and key events.

Bluetooth. Based on the literature regarding the evaluation of spatial coding and retrieval [15], [41], [42], the electrodes in this experiment were determined as shown in blue in Fig. 4. In this experiment, the key response code of the subject in the VR task is used as an event marker to synchronize the EEG signal, align the subject's key response with the EEG signal on the time axis, and facilitate the offline analysis of the subject's EEG signal at the time of action response. Fig. 5 shows the EEG signals of the subjects when performing the test task, where L, R, and R' represent the left, right, and right key event markers respectively, and the abscissa represents the time dimension of EEG signals, and the ordinate represents the 16 sampling channels. The data of each channel is the potential amplitude of EEG signals.

2) *Preprocessing*: In this study, EEGLAB [43] toolbox was used to preprocess EEG signals, mainly including the following steps:

(1) Channel location. To ensure that independent component analysis (ICA) can estimate the source location of independent components of data, the relative coordinate information of each channel needs to be imported.

(2) Band pass filtering. A band-pass filter of 1-49hz is used to filter the attenuation of signals outside this frequency range (50Hz electrical frequency interference, etc.).

(3) Remove artifacts. Firstly, the artifacts caused by the small relative motion between the scalp and the electrode are removed by visual inspection; Secondly, ICA is used to remove the artifacts of EEG and EMG.

(4) Data segmentation. The key events in the test game were taken as event markers, and the EEG signals were divided according to one second before and after the event markers.

(5) Baseline correction. The data before 0ms is used as the baseline. The average value of each point data before 0ms can

be subtracted from the data after 0ms to eliminate part of the spontaneous EEG noise.

(6) Extract the frequency band. we divided EEG signals into seven frequency bands: Delta (1-4 Hz), Theta (4-8 Hz), Alpha1 (8-10.5 Hz), Alpha2 (10.5-13 Hz), Beta1 (13-20 Hz), Beta2 (20-30 Hz) and Gamma (30-40 Hz).

D. Relevant Feature Extraction Methods

1) *CSP Method*: The CSP algorithm diagonalizes the covariance matrix of the two types of samples and uses the method of principal component analysis to find the part with the largest difference between the two types of samples to construct the optimal spatial filter. After the two types of sample data are processed by the spatial filter, the energy difference between the two types of samples in spatial components is the largest [44]. But the correlation can only reflect the linear relationship between the characteristics of EEG signals and cannot measure the nonlinear relationship between the two-time series.

2) *MI, CMI, PCMI Methods*: MI based on information theory reflects the information that a random variable carries with another random variable and can measure the degree of interdependence of two linear or nonlinear time series. CMI can quantify the coupling relationship between the two channels and eliminate the influence of the third channel, to better reflect the coupling information drive between brain regions [46]. In addition, the coupling direction index of brain regions can be analyzed [30]. MI and CMI can be used to calculate the coupling degree and interdependence of task state brain regions [47]. PCMI is another effective method to analyze the coupling strength in the information theory method. This method combines the permutation mode method and CMI to analyze the linear and nonlinear coupling strength of different brain regions of EEG signals. Its effectiveness in EEG feature extraction has been verified by many people [48], [49]. In principle, given any two-channel time series and observation data at a specific time, this method determines the probability distributions of arrangement mode, joint arrangement mode, and conditional arrangement mode in multi-dimensional space. Permutation entropy and PCMI value of EEG signal in these three modes can be obtained [32]. In addition, PCMI is more robust to signal noise than the CMI method [32], [33].

E. Permutation Conditional Mutual Information Common Space Pattern Method

1) *Method Description*: The original EEG signals are segmented according to the pre-test and post-test categories. The two types of EEG signals are expressed as $E_1 = \{e_{11}, e_{11}, \dots, e_{1p-1}, e_{1p}\}$ and $E_2 = \{e_{21}, e_{21}, \dots, e_{2q-1}, e_{2q}\}$ respectively, and the dimensions are $N \times T \times q$. Where N represents the number of source signal channels, $N = 16$. T represents the number of sampling points on each channel, p and q represent the number of two types of signal segments respectively.

(1) Calculate the mixed spatial permutation condition mutual information of two kinds of EEG signals

Take any two time-series X and Y , expressed as $X = (X_1 X_2, \dots, X_n)^T$ and $Y = (Y_1, Y_2, \dots, Y_n)^T$ respectively,

where n is the number of observation points of EEG signal. They are embedded into m -dimensional space to obtain new vectors, as shown in formulas (1) and (2).

$$X_i = (x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}), i = 1, \dots, n - (m-1)\tau \quad (1)$$

$$Y_j = (y_j, y_{j+\tau}, \dots, y_{j+(m-1)\tau}), j = 1, \dots, n - (m-1)\tau \quad (2)$$

where m represents the embedding dimension and τ represents the delay time. The elements $X_i Y_j$ are sorted in ascending order. If there are equivalent elements in X_i or Y_j , the sorting order is determined according to the size of subscript i or j . The vectors expressed in formulas (1) and (2) correspond to a sorting mode. There are $m!$ kinds of sorting patterns for vectors in m -dimensional space. For two EEG signals X and Y from different channels, analyze the vector $X_k, k = 1, \dots, n$ and $Y_k, k = 1, \dots, n$ sorting modes $\pi_i, i = 1, \dots, m!$ and $\pi_j, j = 1, \dots, m!$, there are $m! * m!$ joint patterns for vectors in m -dimensional space. For two EEG signals X and Y from different channels, analyze the vector $X_k, k = 1, \dots, n$ and $Y_k, k = 1, \dots, n$ sorting modes $\pi_i, i = 1, \dots, m!$ and $\pi_j, j = 1, \dots, m!$, there are $m! * m!$ joint sorting modes. Take the vectors with the same sorting pattern as a class, and the occurrence probability of each joint sorting pattern can be obtained according to the occurrence times of each pattern C_{ij} :

$$P(X = \pi_i, Y = \pi_j) = C_{ij}/n - (m-1)\tau \quad (3)$$

$P(x_i, y_j)$ is defined as the joint probability distribution of X and Y ordering patterns of EEG signals. Similarly, the conditional probability distribution $P(x_i | y_j)$ of the sorting mode X under the assumption that the EEG signal Y exists can be calculated:

$$P(x_i | y_j) = P(X = \pi_i | Y = \pi_j) = C_{ij}/n - (m-1)\tau \quad (4)$$

According to formulas (3) and (4), combined with the Shannon entropy method, the conditional ranking entropy X under the premise of the existence of Y can be calculated, as shown in the following formula:

$$PE(X|Y) = - \sum_{i=1}^{m!} \sum_{j=1}^{m!} P(x_i, y_j) \log(P(x_i | y_j)) \quad (5)$$

Now, Y_δ is set to represent the observation value of Y the δ time in the future. According to the above formula (5), the ranking condition entropy $PE(Y_\delta | Y) Y_\delta$ under the premise of assuming the existence of Y can be obtained; Then, the joint ranking modes of vectors X_i and Y_j at the δ th time in the future are analyzed. At this time, $m! * m! * m!$ joint ranking modes can be obtained at most, and the joint ranking entropy of the above vectors can be obtained.

$$PE(X, Y_\delta, Y) = - \sum_{i=1}^{m!} \sum_{j=1}^{m!} \sum_{k=1}^{m!} P(x_i, y_j, y_k) \log(P(x_i, y_j, y_k)) \quad (6)$$

Assuming the existence of Y , the conditional joint ranking entropy of X and Y_δ is:

$$PE(X, Y_\delta | Y) = PE(X, Y_\delta, Y) / PE(Y) \quad (7)$$

Finally, the $PCMI_{X \rightarrow Y}^\delta$ used to calculate the permutation condition mutual information of EEG signals X and Y is obtained, as shown in formula (8):

$$\begin{aligned} PCMI_{X \rightarrow Y}^\delta &= PCMI(X; Y_\delta | Y) \\ &= PE(X | Y) + PE(Y_\delta | Y) - PE(X, Y_\delta | Y) \end{aligned} \quad (8)$$

Similarly, $PCMI_{Y \rightarrow X}^\delta$ can be obtained. At subsequent time points, the coupling strength of the X and Y is calculated accordingly. The transmission between such signals is defined as:

$$PCMI_{X \rightarrow Y} = \frac{1}{N} \sum_{\delta=1}^N PCMI_{X \rightarrow Y}^\delta \quad (9)$$

Similarly, $PCMI_{Y \rightarrow X}$ can be obtained, which represents the coupling strength from X to Y and Y to X , respectively. In (9), N refers to the maximum step size. This paper N takes the optimal empirical value of 15. The final calculated permutation condition mutual information matrix is shown in formula (10):

$$C_{ij} = \begin{bmatrix} PCMI_{1 \rightarrow 1} & \cdots & PCMI_{1 \rightarrow N} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ PCMI_{N \rightarrow 1} & \cdots & PCMI_{N \rightarrow N} \end{bmatrix} \quad (10)$$

$= 1, 2; j = 1, 2, \dots, p, \text{ or } q$

where i is the classification category, and the values are 1 and 2, indicating two classifications; j is the sample size of each category; N is the number of channels; $PCMI$ means calculating the coupling strength of the two channels. Then calculate the matrix expectation of the two types of raw data after segmentation:

$$C_1 = \frac{C_{11} + C_{12} + \dots + C_{1p}}{p}, C_2 = \frac{C_{21} + C_{22} + \dots + C_{2p}}{q} \quad (11)$$

Formula (11) shows that the permutation condition mutual information matrix after obtaining their average normalization. C_1 is the expectation of the permutation condition mutual information matrix of the first type of samples and C_2 is the expectation of the permutation condition mutual information matrix of the second type of samples. C_c represents the mixed space matrix of two types of data.

$$C_c = C_1 + C_2 \quad (12)$$

(2) The whitened eigenvalue matrix P is obtained by principal component analysis. Eigenvalue decomposition of C_c :

$$C_c = U_c \lambda U_c^T \quad (13)$$

In formula (13): U_c is the eigenvector matrix of the matrix C_c , λ is the diagonal matrix composed of corresponding eigenvalues, arrange the eigenvalues in descending order, and obtain the eigenvalue matrix after whitening:

$$P = \frac{1}{\sqrt{\lambda}} \cdot U_c^T \quad (14)$$

At the same time, the diagonalization of C_1 and C_2 is obtained:

$$S_1 = PC_1P^T, S_2 = PC_2P^T \quad (15)$$

S_1 and S_2 respectively represent two types of sample orthogonal whitening transformation matrices and have common eigenvectors. Therefore, S_1 and S_2 can be

decomposed into two diagonal matrices λ_1, λ_2 and an eigenvector matrix B respectively:

$$S_1 = B\lambda_1B^T, S_2 = B\lambda_2B^T \quad (16)$$

The two diagonal matrices λ_1, λ_2 meet the requirements of formula (17) and I is the identity matrix.

$$\lambda_1 + \lambda_2 = I_{N \times N} \quad (17)$$

(1) Calculate the projection matrix

For the eigenvector matrix B , when one type S_1 can take the maximum eigenvalue, the other type S_2 can take the minimum eigenvalue. Therefore, the matrix B can be used for binary classification calculation. The projection matrix can be obtained by the following formula:

$$W_{N \times N} = (B^T P)^T \quad (18)$$

(2) Eigenvalues are calculated by projection

The original EEG signal $E_{N \times T}$ is projected through the projection matrix W of formula (18) to obtain the characteristic matrix Z , whose dimension is $N \times T$:

$$Z = W_{N \times N} \times E_{N \times T} \quad (19)$$

Normalize the calculated characteristic matrix, as shown in formula (20), where f_i is the normalized characteristic of the i th sample. According to the previous research conclusions, the spatial feature information is mainly concentrated in the head and tail of the feature matrix [50]. Therefore, we selected the first n rows and the last n rows of data as the feature matrix calculated by the PCMICSP algorithm, where $2n \leq N$. Z_i represents the characteristic matrix of the i th sample, and Z_{in} represents the j -th eigenvalue of the i th sample.

$$f_i = \log \left(\frac{\text{var}(Z_i)}{\sum_{j=1}^{2n} \text{var}(Z_{ij})} \right) \quad (20)$$

F_1 represents the normalized spatial features of the first type of samples obtained by the PCMICSP algorithm, and p represents the number of the first type of samples.

$$F_1 = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_p \end{bmatrix}_{p \times 2n} \quad (21)$$

Similarly, the normalized spatial feature F_2 of the second type of samples obtained by the PCMICSP algorithm can be obtained.

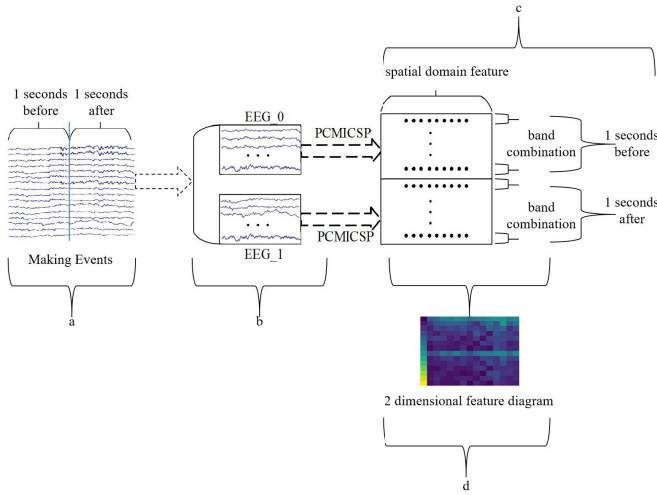


Fig. 6. Flow chart of feature combination. (a. original data; b. segmented into 1s before/after the raw data according to event markers; c. different frequency band combinations; d. visualization of the characteristic matrix of the previous step).

2) *Feature Combination Method*: CSP can effectively extract spatial features from EEG signals that are easy to distinguish and classify. However, Alotaiby et al. [51] proved that the performance of the CSP spatial filter has an important relationship with the working frequency band of EEG, so this paper will analyze the spatial characteristics of PCMICSP in multiple frequency bands. In this paper, the feature combination method [52] designed by Zhang et al. is used to combine the spatial features extracted by PCMICSP with the time and frequency features of the source signal. The time domain, spatial domain and frequency domain features of each category can be extracted from multi-channel EEG signals. The spatial domain feature is the normalized spatial domain feature obtained in subsection I); The time-domain feature is the time feature before and after the source EEG signal is divided into 1 second; The frequency-domain feature is the combined feature of multiple frequency bands of the source signal. Fig. 6 is a feature combination flow chart of the method applied to the data in this paper.

In this experiment, the preprocessed data are the EEG signals one second before and one second after the subjects press the key in the test game. Its length is, 16×250 , 16 is the number of EEG leads, and 250 is the sampling point of the 2s event sequence. Firstly, the data is divided into EEG_0 and EEG_1 according to event markers, respectively representing the data one second before and one second after the button pressing. Then, the PCMICSP algorithm proposed in this paper is used to obtain the normalized eigenmatrix of the two segments of data, whose dimension is,

1×16 and respectively represent the weights of 16 channels' characteristics in the spatial space. Finally, the 2-dimensional eigenmatrix is constructed according to the frequency band combination. In this paper, the combination of five frequency bands [45] is selected. Making the final two-dimensional feature matrix form a square matrix is made in the last step of the PCMICSP algorithm, that is, the first five rows and the last five rows of data are selected as the feature matrix for PCMICSP feature extraction. EEG of the previous

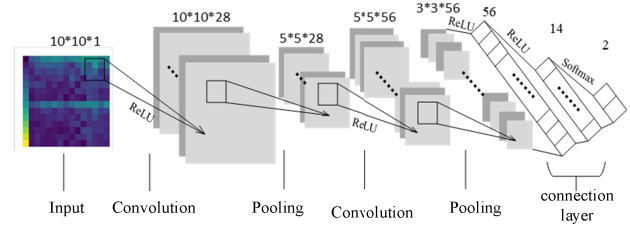


Fig. 7. CNN model structure.

second and the next second according to the above EEG_0 and EEG_1 will get 5×10 characteristic matrix respectively, 5 represents 5 frequency bands, and 10 represents the selected spatial characteristic matrix. The two feature matrices are placed up and down and combined into a two-dimensional feature matrix as the input value of the subsequent convolution neural network classifier. The matrix contains the time domain, spatial domain, and frequency domain characteristics of EEG signals of each key event before and after spatial cognitive training.

F. Classification and Statistical Methods

1) *Classification Model*: CNN model is used to classify EEG signals before and after spatial cognitive training. CNN is widely used in computer vision and natural language processing [53]. This paper uses Keras deep learning library to build the CNN model, in which the optimization function is Adam, the learning rate is 10^{-4} , the batch size is set to 64, and 200 epochs per iteration. The two-dimensional feature matrix is used as the input of shallow CNN to construct the classification model of EEG signals before and after spatial cognitive training. Fig. 7 shows the structure of the CNN model.

The two-dimensional feature matrix obtained by the above feature combination method is used as the input layer of the CNN model, and the dimension is $10 \times 10 \times 1$. The second layer is the convolution layer. For the two-dimensional characteristic matrix of the input layer, 28 convolution cores with a size of 3×3 and a moving step of 1 are used for convolution operation, and the activation function is the corrected linear unit (RELU). The third and fifth layers select the maximum pooling layer with the size of 2×2 and the moving step size of 2. The pooling layer realizes the data simplified sampling processing of the output matrix data of the previous layer. The fourth layer is the convolution layer, which uses 56 convolution cores with the size of 3×3 and the moving step size of 1 for convolution operation, and the RELU function for nonlinear transformation. The sixth and seventh layers are fully connected layers, which contain 56 and 14 neurons respectively. The nonlinear transformation is carried out by using the RELU activation function. The last layer is the output layer, which contains two neurons. The second classification is realized by using the Softmax activation function. To increase the robustness and generalization of the model, a Dropout layer with a probability coefficient of 0.5 was added to the fully connected layers of the sixth and seventh layers.

2) *Statistical Method*: In this study, 5-fold cross-validation was used to evaluate the comprehensive performance of the

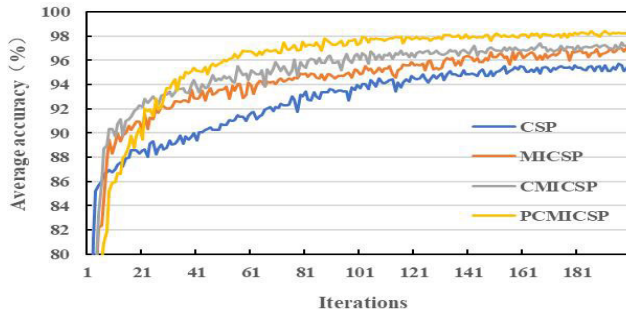


Fig. 8. Average accuracy curve of Delta-Theta-Alpha1-Beta1-Beta2 frequency band combination.

model. Four evaluation indicators were compared: precision, F1 value, recall rate, and AUC value. Since the accuracy curve and the loss curve can be used to evaluate the fitting ability of the CNN model, this study gives the average accuracy curve and the average loss curve obtained through 5-fold cross-validations. The precision curve refers to the accuracy change curve of the CNN model in the training iteration process, and the loss curve refers to the loss change curve of the model in the verification data set.

III. RESULTS

In this study, combined with the feature combination method, the traditional CSP, mutual information CSP, conditional mutual information CSP, and the proposed sequential condition mutual information PCMICSP (CSP, MICSP, CMICSP, and PCMICSP) were used to convert the EEG signals before and after spatial cognitive training into two-dimensional images. Among them, the MICSP algorithm is a CSP feature extraction algorithm that fuses mutual information. The main difference between MICSP and PCMICSP is whether the mutual information matrix is used to calculate the mixed space matrix of the two kinds of data. The same applies to the CMICSP algorithm. Then the image data transformed by the four methods are input into the shallow CNN model for classification. The superiority of the PCMICSP feature extraction algorithm in spatial cognitive training effect evaluation was verified by CNN classification results. There are many combinations of PCMICSP features in the different time domains and frequency domains. This paper only selects the classification results of some frequency bands [33], including Delta-Theta-Alpha1-Beta1-Beta2, Theta-Alpha1-Beta1-Beta2-Gamma, Delta-Alpha2-Beta1-Beta2-Gamma, Delta-Theta-Alpha2-Beta1-Gamma combination.

A. Delta-Theta-Alpha1-Beta1-Beta2 Band Combination

Fig. 8 and Fig. 9 show the variation curves of average accuracy and loss in the CNN model of CSP, MICSP, CMICSP, and PCMICSP feature sets obtained under the Delta-Theta-Alpha1-Beta1-Beta2 frequency band combination method. The figure shows that when the training iterations of CNN model are 110-120, 120-130, 150-160, and 160-170, the average accuracy of PCMICSP, CMICSP, CSP, and MICSP feature set is stable at 98%-98.5%, 96%-97%, 94.5%-95.5%, and 96%-97%, respectively. The corresponding

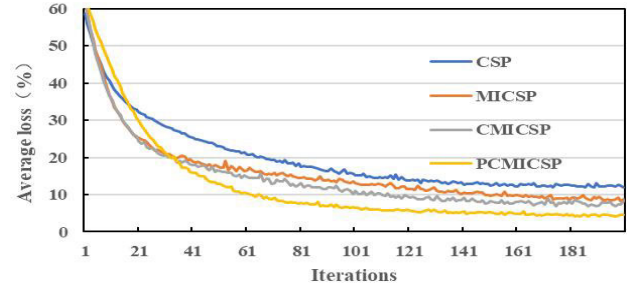


Fig. 9. Average loss curve of Delta-Theta-Alpha1-Beta1-Beta2 band combination.

TABLE I

AVERAGE EVALUATION INDEX OF CNN 5-FOLD CROSS-VALIDATION

Different feature extraction methods	Precision	Recall	F1-score	AUC
CSP	95.45%	95.45%	95.45%	0.9884
MICSP	96.30%	96.29%	96.29%	0.9911
CMICSP	96.35%	96.34%	96.35%	0.9931
PCMICSP	96.84%	96.84%	96.84%	0.9943

TABLE II

AVERAGE CLASSIFICATION ACCURACIES OF THE INDIVIDUAL PARTICIPANTS (%)

Methods	S1	S2	S3	S4	S5	S6	S7
CSP	94.2	94.5	94.8	95.5	95.5	94.9	95.2
MICSP	96.7	96.3	97.0	96.9	96.3	96.7	96.5
CMICSP	96.5	96.4	96.5	96.6	96.8	97.0	96.4
PCMICSP	98.0	98.5	98.3	98.6	98.5	98.4	98.6

average loss rate steady at 4%-5%, 7%-8%, 12%-13%, and 8%-9%, respectively.

Table I shows the average evaluation values of the above four feature sets in the CNN model under the combination of the Delta-Theta-Alpha1-Beta1-Beta2 frequency band. As can be seen from Table I, each evaluation index of the PCMICSP feature set classified by CNN classification is superior to that of CMICSP, MICSP, and CSP.

Table II shows the classification accuracy of the CNN model under the combination of the Delta-Theta-Alpha1-Beta1-Beta2 frequency band. As can be seen from Table II, each accuracy index of the PCMICSP feature set classified by CNN classification is superior to that of CMICSP, MICSP, and CSP.

B. Theta-Alpha1-Beta1-Beta2-Gamma Band Combination

Fig. 10 and Fig. 11 show the average accuracy and loss curves of CSP, MICSP, CMICSP, and PCMICSP feature sets obtained by the CNN model in the combination of Theta-Alpha1-Beta1-Beta2-Gamma bands, respectively. The figure shows that when the training iterations of CNN reach 160-170, the average accuracy rate and average loss rate of MICSP and CMICSP are stable at 96%-96.5% and 10%-11%, respectively. In addition, when the training iterations of the CNN model are 170 ~ 180 and 190 ~ 200, the average accuracy of the

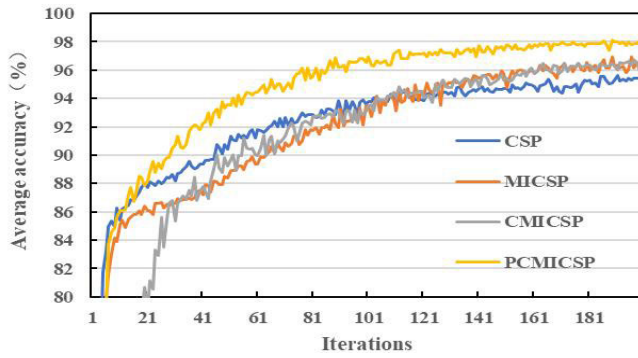


Fig. 10. Average accuracy curve of Theta-Alpha1-Beta1-Beta2-Gamma band combination.

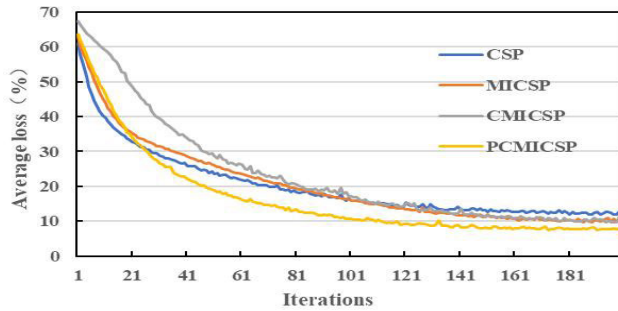


Fig. 11. Average loss curve of Theta-Alpha1-Beta1-Beta2-Gamma band combination.

TABLE III

AVERAGE EVALUATION INDEX OF CNN 5-FOLD CROSS-VALIDATION

Different feature extraction methods	Precision	Recall	F1-score	AUC
CSP	95.96%	95.95%	95.96%	0.9885
MICSP	96.25%	96.25%	96.25%	0.9896
CMICSP	96.38%	96.37%	96.37%	0.9912
PCMICSP	96.61%	96.84%	96.61%	0.9934

TABLE IV

AVERAGE CLASSIFICATION ACCURACIES OF THE INDIVIDUAL PARTICIPANTS (%)

Methods	S1	S2	S3	S4	S5	S6	S7
CSP	94.5	95.5	94.6	94.6	94.7	94.9	95.5
MICSP	96.4	96.4	96.5	96.3	96.5	96.4	96.5
CMICSP	96.5	96.0	96.3	96.5	96.4	96.5	96.5
PCMICSP	98.5	98.5	98.3	98.2	98.5	98.5	98.5

PCMICSP and CSP feature set is stable at 98%-98.5% and 94.5%-95.5%, respectively. And the corresponding average loss rate was steady at 8%-9% and 12%-13%, respectively.

Table III shows the average evaluation value of the above four feature sets in the CNN model under the Theta-Alpha1-Beta1-Beta2-Gamma band combination. It can be seen from Table III that each evaluation index of the PCMICSP feature set classified by CNN classification is superior to that of CMICSP, MICSP, and CSP.

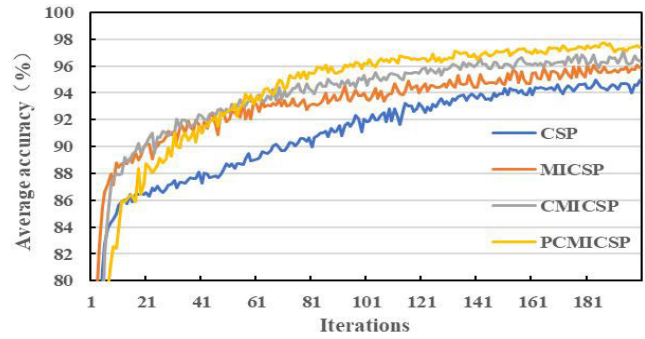


Fig. 12. Average accuracy curve of Delta-Alpha2-Beta1-Beta2-Gamma band combination.

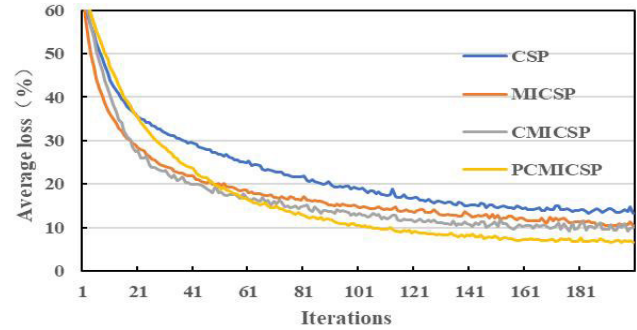


Fig. 13. Average loss curve of Delta-Alpha2-Beta1-Beta2-Gamma band combination.

Table IV shows the classification accuracy of the CNN model under the combination of the Theta-Alpha1-Beta1-Beta2-Gamma frequency band. As can be seen from Table IV, each accuracy index of the PCMICSP feature set classified by CNN classification is superior to that of CMICSP, MICSP, and CSP.

C. Delta-Alpha2-Beta1-Beta2-Gamma Band Combination

Fig. 12 and Fig. 13 respectively display the variation curves of average accuracy and loss in the CNN model of CSP, MICSP, CMICSP, and PCMICSP feature sets under the Delta-Alpha2-Beta1-Beta2-Gamma band combination. The figure shows that when the training iterations of CNN reach 181-190, the average accuracy rate of CSP and PCMICSP are stable at 94%~95% and 97%~97.5%, respectively. And the corresponding average loss rate is steady at 11%~12% and 7%~8%, respectively. Also, when the training iterations of the CNN model are 180 ~ 190 and 190 ~ 200, the average accuracy of MICSP and CMICSP feature set is stable at 96.5%-97%, and the corresponding average loss rate steady at 10%-11% and 9%-10%, respectively.

Table V shows the average evaluation values of the above four feature sets in the CNN model under the combination of the Delta-Alpha2-Beta1-Beta2-Gamma frequency band. It can be seen from Table V that each evaluation index of feature sets classified by CNN is better than that of MICSP, CMICSP, and CSP.

Table VI shows the classification accuracy of the CNN model under the combination of the Delta-Alpha2-Beta1-Beta2-Gamma frequency band. As can be seen from Table VI, each accuracy index of the PCMICSP feature set classified by

TABLE V

AVERAGE EVALUATION INDEX OF CNN 5-FOLD CROSS-VALIDATION

Different feature extraction methods	Precision	Recall	F1-score	AUC
CSP	94.94%	94.95%	94.94%	0.9882
MICSP	95.83%	95.83%	95.83%	0.9910
CMICSP	96.18%	96.18%	96.18%	0.9918
PCMICSP	96.57%	96.56%	96.57%	0.9933

TABLE VI

AVERAGE CLASSIFICATION ACCURACIES OF THE INDIVIDUAL PARTICIPANTS (%)

Methods	S1	S2	S3	S4	S5	S6	S7
CSP	94.5	94.2	95.5	95.2	95.0	95.2	94.5
MICSP	97.0	96.6	96.1	96.5	96.3	96.4	97.0
CMICSP	96.5	97.3	97.2	96.3	96.1	96.5	96.2
PCMICSP	97.2	97.0	97.5	96.4	96.9	96.7	96.2

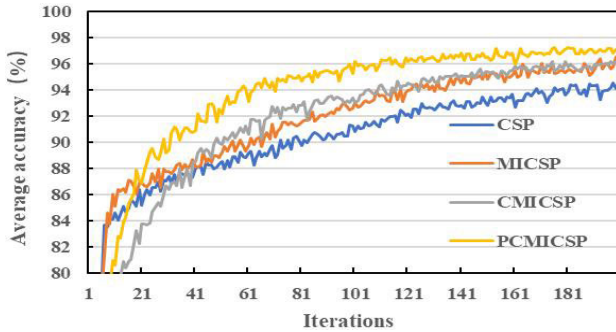


Fig. 14. Average accuracy curve of Delta-Theta-Alpha2-Beta1-Gamma band combination.

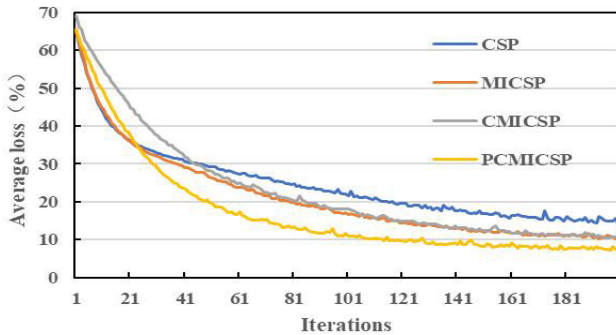


Fig. 15. Average accuracy curve of Delta-Theta-Alpha2-Beta1-Gamma band combination.

CNN classification is superior to that of CMICSP, MICSP, and CSP.

D. Delta-Theta-Alpha2-Beta1-Gamma Band Combination

Fig. 14 and Fig. 15 respectively show the average accuracy and loss curves of CSP, MICSP, CMICSP, and PCMICSP feature sets obtained by using the above feature combination method in the Delta-Theta-Alpha2-Beta1-Gamma band combination in the CNN model. The figure shows that when the training iterations of the CNN model reach 150 ~ 160, the

TABLE VII

AVERAGE EVALUATION INDEX OF CNN 5-FOLD CROSS-VALIDATION

Different feature extraction methods	Precision	Recall	F1-score	AUC
CSP	94.43%	94.45%	94.44%	0.9852
MICSP	95.24%	95.26%	95.25%	0.9893
CMICSP	95.53%	95.54%	95.55%	0.9899
PCMICSP	95.92%	95.93%	95.92%	0.9926

TABLE VIII

AVERAGE CLASSIFICATION ACCURACIES OF THE INDIVIDUAL PARTICIPANTS (%)

Methods	S1	S2	S3	S4	S5	S6	S7
CSP	94.3	94.5	94.1	94.6	93.8	93.8	94.5
MICSP	96.0	95.3	96.1	95.5	95.3	95.4	96.0
CMICSP	96.5	95.3	95.2	95.3	95.1	95.5	96.2
PCMICSP	96.5	96.4	96.5	95.7	96.0	96.3	96.2

TABLE IX

EXPERIMENTAL GROUP BEFORE AND AFTER TRAINING SCALE STATISTICS

Scale	Before training	After training	<i>P</i> value
CBTT	37.14±16.48	50±6.83	0.048
GZSOT	4.86±5.16	9.89±6.20	0.034
PTSOT	35.75±9.81	19.32±7.72	0.0009

average accuracy of the MICSP and CMICSP feature set is stable at 95%-96%, and the corresponding average loss rate steady at 11%-12% and 10%-11%, respectively. In addition, the CSP feature set is stable when the CNN model is iterated to 181 ~ 190 times, the average accuracy is stable at 94% ~ 94.5%, and the average loss is stable at 17% ~ 18%. The PCMICSP feature set is stable when the CNN model is iterated to 160 ~ 170 times, the average accuracy is stable at 96% ~ 96.5%, and the average loss is stable at 10% ~ 11%.

Table VII shows the average evaluation value of the above four feature sets in the CNN model under the Delta-Theta-Alpha2-Beta1-Gamma band combination. It can be seen from Table VII that each evaluation index of feature sets classified by CNN is better than that of MICSP, CMICSP, and CSP.

Table VIII shows the classification accuracy of the CNN model under the combination of the Delta-Theta-Alpha2-Beta1-Gamma frequency band. As can be seen from Table VIII, each accuracy index of the PCMICSP feature set classified by CNN classification is superior to that of CMICSP, MICSP, and CSP.

Table IX shows that the *P* values of the three spatial cognition scales of the 7 individuals before and after training were all less than 0.05 before and after training, and the statistical results showed significant differences. The average CBTT score increased from 37.14 points before training to 50 points after training, showing a wide range of improvement. The average score of GZSOT increased from 4.86 points before

training to 9.89 points after training, with a wide variation range. The mean error Angle obtained by PTSOT decreased from 35.75 points before training to 19.32 points after training, with a wide range of variation.

IV. DISCUSSION

A. Feasibility and Effectiveness of PCMICSP

In this section, the CSP feature extraction method and convolutional neural network classification method were combined to classify and evaluate the EEG signals of the spatial cognitive training group before and after the training. The results of convolutional neural network classification show that the classification accuracy and other performance indexes of MICSP, CMICSP, and PCMICSP are better than that of traditional CSP. The common space pattern feature extraction algorithm combining the MI theory [54] method can effectively improve classification accuracy. MI theory method can measure the degree of linear and nonlinear correlation of EEG signals, which is very important for non-stationary and nonlinear EEG signals. By comparing the classification accuracy and other performance indexes of MICSP, CMICSP, and PCMICSP, it is found that the results of PCMICSP proposed in this paper are the best ones under the combination of multiple frequency bands, while the results of MICSP and CMICSP are not significantly different. This is consistent with the conclusion that the permutation condition mutual information algorithm is superior to the conditional mutual information method described in Li and Ouyang literature [32]. PCMI can more effectively detect the interaction delay between the event sequences of two brain regions. In addition, PCMI has strong robustness to EEG, so the EEG noise will not seriously damage the inherent timing pattern, especially when there is strong coupling strength between channels [48], [49], [55].

B. The Classification Performance of PCMICSP

The classification results of the PCMICSP feature set under different frequency band combinations were analyzed separately. It was found that the CNN model of the Delta-Theta-Alpha1-Beta1-Beta2 frequency band combination had the best classification effect in terms of average loss, average accuracy, and various evaluation indexes. The Theta-Alpha1-Beta1-Beta2-Gamma band combination is next, but not far apart. The variation of the Theta frequency band is consistent with the research results of Guilford and Zimmerman [37]: the Theta frequency band oscillation of humans is correlated with the coding and retrieval of spatial information. Koenig et al. found an increase in EEG synchronous measurements in the Delta band in their study of spatial cognitive network function [56]. The Theta and Alpha1 frequency bands are consistent with the results obtained in literature [40]: the brain network attributes of Theta and Alpha1 frequency bands change most significantly before and after training. Yuan [33] showed a good EEG classification effect in Beta1 and Beta2 bands before and after spatial cognitive training, which was consistent with the frequency band combination obtained in this study. It can be concluded that the PCMICSP EEG feature extraction method based on the combination of the

Delta-Theta-Alpha1-Beta1-Beta2 frequency band is the most effective method to evaluate the effect of spatial cognitive training.

However, the sample size of this paper is relatively small, and the deep learning model with more hidden layers cannot be used. Therefore, the rationality of the PCMICSP algorithm can be further verified by adding more EEG signals in the future.

V. CONCLUSION

In this paper, the common spatial patterns feature extraction algorithm based on the mutual information of sorting conditions is proposed. The covariance matrix in the original CSP algorithm is replaced by the mutual information matrix of sorting conditions so that CSP can construct spatial filters according to the linear and nonlinear correlation degree of EEG signals simultaneously. The EEG signals before and after spatial cognitive training were used as the data set, and the feature combination method and convolutional neural network classification method were used to verify the performance of the proposed feature extraction algorithm. Experiments show that PCMICSP has the highest classification accuracy under the combination of multiple frequency bands. The algorithm combines the advantages of PCMI and solves the strict linear hypothesis of traditional CSP. The experimental results show that the PCMICSP algorithm can be used for the classification and evaluation of spatial cognitive task state EEG.

ACKNOWLEDGMENT

The authors have no potential conflicts.

REFERENCES

- [1] B. Dubois et al., "Clinical diagnosis of Alzheimer's disease: Recommendations of the International Working Group," *Lancet Neurol.*, vol. 20, no. 6, pp. 484–496, Jun. 2021.
- [2] A. Ata, B. Yesilkaya, O. K. Cura, and A. Akan, "Control of serious games designed for Alzheimer's and dementia patients by EEG signals," in *Proc. Med. Technol. Congr. (TIPTEKNO)*, 2019, pp. 149–152.
- [3] J. Bures and P. Lánský, "From spreading depression to spatial cognition," *Physiological Res.*, vol. 53, pp. S177–S185, 2004.
- [4] T. Chiu, K. Gramann, L. Ko, J. Duann, T. Jung, and C. Lin, "Alpha modulation in parietal and retrosplenial cortex correlates with navigation performance," *Psychophysiology*, vol. 49, no. 1, pp. 43–55, Jan. 2012.
- [5] S. Chouinard, M.-È. Brière, C. Rainville, and R. Godbout, "Correlation between evening and morning waking EEG and spatial orientation," *Brain Cognition*, vol. 53, no. 2, pp. 162–165, Nov. 2003.
- [6] N. Nishiyama, H. Mizuhara, F. Miwakeichi, and Y. Yamaguchi, "Theta episodes observed in human scalp EEG during virtual navigation-Spatial distribution and task dependence," in *Proc. 9th Int. Conf. Neural Inf. Process.*, 2002, pp. 428–432.
- [7] C.-T. Lin, T.-C. Chiu, and K. Gramann, "EEG correlates of spatial orientation in the human retrosplenial complex," *NeuroImage*, vol. 120, pp. 123–132, Oct. 2015.
- [8] J. Park et al., "Role of low- and high-frequency oscillations in the human hippocampus for encoding environmental novelty during a spatial navigation task," *Hippocampus*, vol. 24, no. 11, pp. 1341–1352, Nov. 2014.
- [9] D. J. White, M. Congedo, J. Ciorciari, and R. B. Silberstein, "Brain oscillatory activity during spatial navigation: Theta and gamma activity link medial temporal and parietal regions," *J. Cognit. Neurosci.*, vol. 24, no. 3, pp. 686–697, Mar. 2012.
- [10] G. Sharma, A. A. Salam, S. Chandra, V. Singh, and A. Mittal, "Influence of spatial learning perspectives on navigation through virtual reality environment," in *Proc. Int. Conf. Brain Inform.*, in Lecture Notes in Artificial Intelligence, 2016, pp. 346–354.

- [11] J. Snider, O. J. Ahmed, E. Halgren, H. Poizner, and S. S. Cash, "Human intracranial recordings during spatial exploration of a 3D virtual environment," in *Proc. 6th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, Nov. 2013, pp. 464–467.
- [12] S. E. Kober, J. Kurzmann, and C. Neuper, "Cortical correlate of spatial presence in 2D and 3D interactive virtual reality: An EEG study," *Int. J. Psychophysiology*, vol. 83, no. 3, pp. 365–374, Mar. 2012.
- [13] J. Ramos-Loyo and L. M. Sanchez-Loyo, "Gender differences in EEG coherent activity before and after training navigation skills in virtual environments," *Hum. Physiol.*, vol. 37, no. 6, pp. 68–75, 2011.
- [14] D. P. Estefan et al., "Coordinated representational reinstatement in the human hippocampus and lateral temporal cortex during episodic memory retrieval," *Nature Commun.*, vol. 10, no. 1, p. 2255, May 2019.
- [15] G. Sharma, K. Gramann, S. Chandra, V. Singh, and A. P. Mittal, "Brain connectivity during encoding and retrieval of spatial information: Individual differences in navigation skills," *Brain Informat.*, vol. 4, no. 3, pp. 207–217, Sep. 2017.
- [16] X. Li, Y. Yan, and W. Wei, "Identifying patients with poststroke mild cognitive impairment by pattern recognition of working memory load-related ERP," *Comput. Math. Methods Med.*, vol. 2013, pp. 1–10, Oct. 2013.
- [17] V. Pergher, B. Wittevrongel, J. Tournoy, B. Schoenmakers, and M. M. Van Hulle, "N-back training and transfer effects revealed by behavioral responses and EEG," *Brain Behav.*, vol. 8, no. 11, Nov. 2018, Art. no. e01136.
- [18] T. E. Baker and C. B. Holroyd, "The topographical n170: Electrophysiological evidence of a neural mechanism for human spatial navigation," *Biol. Psychol.*, vol. 94, no. 1, pp. 90–105, Sep. 2013.
- [19] F. Protopapa, D. Mylonas, K. Spiliotis, C. Siettos, N. Smyrnis, and I. Evdokimidis, "Dynamic analysis of EEG signals during spatial working memory used for either perception discrimination or planning of action," in *Proc. IEEE Eng. Med. Biol. Soc. Conf.*, Sep. 2011, pp. 5896–5899.
- [20] J. Toppi et al., "Different topological properties of EEG-derived networks describe working memory phases as revealed by graph theoretical analysis," *Frontiers Human Neurosci.*, vol. 11, p. 637, Jan. 2018.
- [21] C. Y. Xu, "Comparative study on classification method of multitask brain-computer interfaces based on CSP and ICA," M.S. thesis, Nanchang Univ., 2012.
- [22] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clin. Neurophysiol.*, vol. 110, no. 5, pp. 787–798, May 1999.
- [23] X. Liao, D. Yao, D. Wu, and C. Li, "Combining spatial filters for the classification of single-trial EEG in a finger movement task," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 821–831, May 2007.
- [24] J. Wang, L. Zhang, and B. Hu, "Research on the methods for multi-class kernel CSP-based feature extraction," *J. Biomed. Eng.*, vol. 29, no. 2, pp. 217–222, 2012.
- [25] S. Y. Zhao, "Synchronous intensity and direction analysis of regional EEG signals," M.S. thesis, Yanshan Univ., 2017.
- [26] S. Sun and C. Zhang, "An optimal kernel feature extractor and its application to EEG signal classification," *Neurocomputing*, vol. 69, nos. 13–15, pp. 1743–1748, Aug. 2006.
- [27] B. Nasihatkon, R. Boostani, and M. Z. Jahromi, "An efficient hybrid linear and kernel CSP approach for EEG feature extraction," *Neurocomputing*, vol. 73, nos. 1–3, pp. 432–437, Dec. 2009.
- [28] Q. Tong, J. Pang, and S. Xu, "Information entropy analysis of EEG," *Acta Biophysica Sinica*, vol. 4, pp. 645–650, 1993.
- [29] H. Li, D. Wen, Z. Wei, and Y. Zhou, "Advances in the extraction and classification of EEG dynamic features in patients with mild cognitive impairment," *Chin. J. Biomed. Eng.*, vol. 38, no. 3, pp. 348–354, 2019.
- [30] R. A. A. Ince, B. L. Giordano, C. Kayser, G. A. Rousselet, J. Gross, and P. G. Schyns, "A statistical framework for neuroimaging data analysis based on mutual information estimated via a Gaussian copula," *Hum. Brain Mapping*, vol. 38, no. 3, pp. 1541–1573, Mar. 2017.
- [31] Y. Li, X. L. Li, G. X. Quyang, and X. P. Guan, "Neural information flow based on Bayesian estimation," *Chin. Sci. Bull.*, vol. 52, no. 7, pp. 848–854, 2007.
- [32] X. Li and G. Ouyang, "Estimating coupling direction between neuronal populations with permutation conditional mutual information," *NeuroImage*, vol. 52, no. 2, pp. 497–507, Aug. 2010.
- [33] J. P. Yuan, "Research on spatial cognitive training and evaluation system integrating brain-computer interface and virtual reality," M.S. thesis, Yanshan Univ., 2019.
- [34] B. B. Liang, "Research on spatial cognitive training and EEG analysis of community elderly based on BCI-VR," M.S. thesis, Yanshan Univ., 2020.
- [35] T. N. Tombaugh, I. McDowell, B. Kristjansson, and A. M. Hubley, "Mini-mental state examination (MMSE) and the modified MMSE (3MS): A psychometric comparison and normative data," *Psychol. Assessment*, vol. 8, no. 1, pp. 48–59, Mar. 1996.
- [36] L. R. B. Ramirez, R. Saracco-Alvarez, R. Escamilla-Orozco, and A. F. Orellana, "Validity of the Montreal Cognitive Assessment scale (MoCA) to detect cognitive impairment in schizophrenia," *Salud Mental*, vol. 37, no. 6, pp. 517–522, Nov-Dec. 2014.
- [37] J. P. Guilford and W. S. Zimmerman, "The Guilford-Zimmerman aptitude survey," *J. Appl. Psychol.*, vol. 32, no. 1, pp. 24–34, 1948.
- [38] M. Hegarty, "A dissociation between mental rotation and perspective-taking spatial abilities," *Intelligence*, vol. 32, no. 2, pp. 175–191, Apr. 2004.
- [39] R. P. C. Kessels, M. J. E. van Zandvoort, A. Postma, L. J. Kappelle, and E. H. F. de Haan, "The corsi block-tapping task: Standardization and normative data," *Appl. Neuropsychology*, vol. 7, no. 4, pp. 252–258, Dec. 2000.
- [40] D. Wen et al., "The EEG signal analysis for spatial cognitive ability evaluation based on multivariate permutation conditional mutual information-multi-spectral image," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 10, pp. 2113–2122, Oct. 2020.
- [41] W. Bischof and P. Boulanger, "Spatial navigation in virtual reality environments: An EEG analysis," *Cyberpsychology Behav., Impact Internet, Multimedia Virtual Reality Behav. Soc.*, vol. 6, pp. 95–487, Jan. 2003.
- [42] S. E. Kober and C. Neuper, "Sex differences in human EEG theta oscillations during spatial navigation in virtual reality," *Int. J. Psychophysiology*, vol. 79, no. 3, pp. 347–355, Mar. 2011.
- [43] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.
- [44] N. Ye, Y. G. Sun, and X. Wang, "Recognition based on common spatial patterns and ANN for brain-computer interface signal," *J. Northeastern Univ.*, vol. 31, no. 1, pp. 12–15, 2010.
- [45] J. Xu and G. Zuo, "Motor imagery electroencephalogram feature selection algorithm based on mutual information and principal component analysis," *J. Biomed. Eng.*, vol. 33, no. 2, pp. 201–207, 2016.
- [46] S. Y. Peng, D. Zhou, J. Q. Zhang, Y. Wang, and J. F. Gao, "Research on mutual information-based brain network and lie detection," *Acta Electronica Sinica*, vol. 47, no. 7, pp. 1551–1556, 2019.
- [47] M. Vejmelka and M. Palus, "Inferring the directionality of coupling with conditional mutual information," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 77, no. 2, Art. no. 026214, Feb. 2008.
- [48] J. Li, X. Z. Liu, and G. X. Ouyang, "Permutation conditional mutual information and its application to epileptic EEG," in *Proc. Int. Conf. Comput. Sci. Appl.*, 2013, pp. 733–736.
- [49] D. Wen, Z. Bian, Q. Li, L. Wang, C. Lu, and X. Li, "Resting-state EEG coupling analysis of amnesic mild cognitive impairment with type 2 diabetes mellitus by using permutation conditional mutual information," *Clin. Neurophysiol.*, vol. 127, no. 1, pp. 335–348, Jan. 2016.
- [50] J. Chen, R. Zheng, X. Jia, and P. Zhang, "EEG emotion classification based on common spatial patterns and wavelet packet decomposition," *Comput. Eng. Appl.*, vol. 55, no. 1, pp. 149–153, 2019.
- [51] T. N. Alotaiby, S. A. Alshebeili, F. M. Alotaibi, and S. R. Alrshoud, "Epileptic seizure prediction using CSP and LDA for scalp EEG signals," *Comput. Intell. Neurosci.*, vol. 2017, pp. 1–11, Oct. 2017.
- [52] Y. Zhang, Y. Guo, P. Yang, W. Chen, and B. Lo, "Epilepsy seizure prediction on EEG using common spatial pattern and convolutional neural network," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 465–474, Feb. 2020.
- [53] F. T. Li, "Convolutional neural network and its applications," M.S. thesis, Dalian Univ. Technol., 2020.
- [54] K. M. Hassan, M. R. Islam, T. T. Nguyen, and M. K. I. Molla, "Epileptic seizure detection in EEG using mutual information-based best individual feature selection," *Exp. Syst. Appl.*, vol. 193, May 2022, Art. no. 116414.
- [55] Z. Li and X. Li, "Causality of spike trains based on entropy," in *Signal Processing in Neuroscience*, X. Li, Ed. Singapore: Springer, 2016, pp. 39–55.
- [56] T. Koenig et al., "Decreased EEG synchronization in Alzheimer's disease and mild cognitive impairment," *Neurobiol. Aging*, vol. 26, no. 2, pp. 165–171, Feb. 2005.