An interpretable contrastive learning transformer for EEG-based person identification

Xinghan Shao, C. Chang, John Q. Gan, and Haixian Wang, *Senior Member, IEEE*

***Abstract*—Research on electroencephalogram (EEG)-based person identification is increasing because EEG signals must be collected from the living body, making them difficult to steal or alter. However, EEG signals are greatly influenced by subjects' states, and most studies on EEG-based person identification have overlooked this influence. In this study, we proposed an interpretable contrastive learning transformer to tackle the impact of state changes on EEG-based person identification. Contrastive learning transformers construct pairs of EEG signal feature samples to capture state-independent and identity-distinct features. Specifically, the power spectral density (PSD) of EEG signals from the same user in different paradigms is used as positive samples, while the PSD from other users is used as negative samples. Pairs of samples are encoded to obtain corresponding features and then projected into a contrastive space through a multi-layer perceptron. Then, the NT-Xent loss function minimizes the distance between positive samples within the same batch and maximizes the distance between negative samples. Finally, to eliminate bias between positive sample pairs from different paradigms, we introduced the cross-paradigm alignment loss for the first time to capture individual consistency. We evaluated our model on two datasets. Dataset 1 contains EEG signals from 109 individuals, recorded across multiple paradigms designed to elicit different states. Dataset 2 consists of EEG signals from 71 individuals, collected across two sessions, with each session including two paradigms. We evaluated the accuracy of both single-paradigm and cross-paradigm recognition. Our proposed model outperforms state-of-the-art models for EEG-based person identification. We also conducted experiments on electrode attention visualization to capture the brain regions that the model focuses on, and the results demonstrate that, unlike in a single-paradigm, models trained in cross-paradigm focus on fewer electrodes and more concentrated regions.**

***Index Terms*—EEG, biometrics, person identification, contrastive learning, transformer.**

I. Introduction

In the era of the information explosion, safeguarding an individual's identity information has emerged as a critical concern, necessitating the development of innovative, precise, and secure identity recognition technologies. While several biometric recognition technologies, such as fingerprint [1], face [2], and iris [3] recognition, have achieved remarkable accuracy, the progression of AI technology has raised concerns about the vulnerability of these biometric features to theft and replication. Electroencephalogram (EEG)-based person identification (PI), compared to traditional biometric characteristics, has garnered heightened attention from researchers [4][5]. This method captures brain electrical activity recorded through ion currents resulting from significant neural activations, with amplitudes ranging from 10-200 μV and frequencies between 0.5-40 Hz [6]. EEG serves as a biometric feature due to the distinctiveness and universality of EEG signals. EEG needs to be collected from the living body, making it hard to pilfer or imitate [7]. The inherent nature of EEG signals prevents users from subjectively or inadvertently disclosing their identifiers [8]. EEG collection is straightforward and can be accomplished using portable, cost-effective devices [9][10].

EEG-based PI typically consists of two components: paradigms and algorithms. Paradigms are utilized to elicit specific EEG patterns in experimental settings. The predominant paradigms encompass resting state, stimulus potentials, and cognitive tasks. The resting state does not necessitate any user response or task completion, including eyes-open (EO) resting-state and eyes-closed (EC) resting-state [11][12]. Stimulus potentials are produced by using external triggers to evoke brain responses that lead to neural activity, such as steady-state visually evoked potentials (SSVEP) [13] and event-related potentials (ERP) [14]. Cognitive tasks involve active user engagement, with a classic example being motor imagery, which entails users imagining motor intentions [15].

There are currently two mainstream approaches in EEG-based PI algorithms: the first involves combining feature extraction methods with classifiers, while the second utilizes deep learning techniques. Common feature extraction methods include autoregressive (AR) [16], power spectral density (PSD) [17], wavelet transform (WT) [18], fuzzy entropy (FE) [19], and phase locking value (PLV) [20]. EEG signals are inherently non-stationary, and brain functionality is often associated with specific frequency bands and brain regions. It has been proved that the frequency domain features of EEG, such as PSD and PLV, are better than the time domain features in EEG-based PI [20]. Classifier options encompass K-nearest neighbor (KNN) [21], support vector machine (SVM) [22][23], linear discriminant analysis (LDA) [24], and random forest [25]. With the advancement of deep learning, numerous neural network models are gaining popularity in the field of brain-wave recognition, including convolutional neural networks (CNN) [26][27], long short-term memory networks (LSTM) [5], graph convolutional neural networks (GCNN) [28], and transformer [29], among others.

EEG-based PI is currently in the early stages of development and has not yet been put into practical application. In labora-

This work was supported by the National Natural Science Foundation of China under Grants 92270113 and 62176054. *(Corresponding author: Haixian Wang.)*

Xinghan Shao and Haixian Wang are with the Key Laboratory of Child Development and Learning Science of Ministry of Education, School of Biological Science & Medical Engineering, Southeast University, Nanjing 210096, Jiangsu, PR China (e-mail: shaoxinghan@seu.edu.cn; hxwang@seu.edu.cn).

C. Chang is with the College of Electronic Engineering, National University of Defense Technology, Hefei 230031, Anhui, PR China.

John Q. Gan is with the School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, UK (e-mail: jqgan@essex.ac.uk).

tory settings, EEG-based PI often utilizes various paradigms to stimulate specific EEG signals. This method is effective in standardizing the state-related components of EEG signals and highlighting differences in identity-related information. However, in real-world environments, users may not adhere perfectly to the prescribed paradigms, resulting in the failure to record the expected EEG signals. Moreover, user states may vary at different times because EEG signals are prone to rapid changes influenced by various factors such as tasks [30], emotions [31], and dietary habits [32]. When paradigms or user states change, a considerable amount of new state data is typically required to train the model, wasting computational resources. From an algorithmic perspective, most methods focus solely on a user's single state, lacking generalizability. While limited literature [26][33] explores state-independent EEG-based PI, these studies mainly propose models without delving into state-independent identity features from an interpretability perspective. Thus, designing an interpretable novel cross-state recognition model is essential for applying brainwave recognition.

In response to the above issue, this study presents a contrastive learning model based on PSD features to mitigate the impact of different states on brainwave recognition. The model incorporates electrode attention, spectral attention, and a transformer module, customizes sample pairs, and extracts frequency and spatial information from EEG signals. Our model was evaluated on two datasets, with experimental results demonstrating superior performance in brainwave recognition and extraction of state-independent frequency and spatial features. Furthermore, by visually analyzing electrode attention, we explore brain regions that are minimally affected by states and possess sufficient identity recognition qualities. This study aids in feature selection for subsequent brainwave recognition research. In summary, our contributions are as follows:

- We proposed a self-supervised framework contrastive learning transformer (CLT), which can effectively extract individual differences in EEG signals in the frequency and spatial domains. Even for cross-paradigm EEG-based PI, fine-tuning can be performed using a small dataset to ensure recognition accuracy. Experimental results have demonstrated that the performance of CLT exceeds that of state-of-the-art models for EEG-based PI.
- We also explored the roles of electrode encoder and spectral encoder in EEG-based PI.
- We proposed the cross-paradigm alignment loss for the first time to capture individual consistency and improve model performance.
- We visualized electrode attention and explored the channels and frequency regions differentiating individuals in single-paradigm and cross-paradigm conditions.

II. Related work

A. EEG-Based Person Recognition

EEG-based PI systems are mainly divided into two branches. One is the traditional machine learning method, which extracts features first and then classifies them. This method has the advantage of being interpretable. Another is the deep learning approach, which has the advantage of high classification accuracy. Yildirim et al. developed a multi-layered stacked 1D CNN model to extract individual-specific features from EEG signals [34]. Kong et al. proposed that EEG comprises background EEG inherent in each person's brain and residue EEG caused by tasks and noise, suggesting the presence of identity-related features in the former. They decomposed the EEG using a low-rank matrix decomposition method, and the maximum correlation criterion algorithm was utilized for classification purposes [35]. Wang et al. suggested that the functional connectivity of the brain can reflect identity uniqueness, employing PLV and GCNN for recognition [20]. Du et al. introduced a transformer-based method utilizing self-attention mechanism to extract spatial features from EEG, achieving state-of-the-art accuracy levels [29]. Cai et al. presented an affective temporal-spatial transformer (AITST) designed to capture temporal and spatial characteristics of EEG signals through interconnected temporal and spatial attention modules, achieving an outstanding accuracy rate of 99.21±0.03% on a single state of the DEAP dataset [36].

B. Contrastive Learning

Contrastive learning is a self-supervised method that determines whether data pairs are similar. It achieves advanced performance in computer vision [37], natural language processing [38], and biometrics [39][40]. According to the way of constructing the contrast set, contrastive learning is mainly divided into two types: one is global-local comparison, such as comparing an instance (a specific data point) to the broader context to which it belongs; Another is sample-sample comparison, such as comparing the converted original image with the original image [41]. Contrastive learning has good generalization ability, flexibility and adaptability, and can improve feature representation. Mohsevand et al. enhanced the similarity between different views of samples in the same original data by using time masking, linear scaling, and Gaussian noise to enhance the samples. It has achieved excellent results in sleep stage classification, clinical abnormal detection, emotion recognition, and other aspects [42]. Shen et al. proposed a contrastive learning of subject-invariant approach for cross-subject emotion recognition that minimizes inter-individual differences by maximizing the similarity of the cross-individual EEG representation under the same emotional stimulus, thus achieving the most advanced cross-individual emotion recognition performance on THU-EP dataset and publicly available SEED dataset [43]. Wang et al. employed multiple self-supervised contrastive tasks to enable the model to extract semantically rich, subject-independent features, thereby helping it extract meaningful and robust EEG data representations from both tinnitus patients and healthy controls [44]. Li et al. combined self-supervised contrastive learning with supervised classification learning to create a joint learning model, which demonstrated exceptional accuracy on the SEED dataset, highlighting its effectiveness in emotion recognition and its potential applicability to other EEG-based classification tasks [45]. Cheng et al. addressed inter-subject differences by calculating subject-based contrastive loss, ensuring that the learned representations effectively capture

individual characteristics. They also introduced adversarial training to enhance the model's subject invariance, reducing the impact of subject differences and enabling the model to better learn subject-general representations [46]. Song et al. proposed a self-supervised framework that verifies the feasibility of learning image representations from EEG signals by using image and electroencephalogram (EEG) encoders to extract paired features of image stimuli and EEG responses and applying contrastive learning to ensure the similarity of the two modes. The framework achieved significant results above the chance level across 200 zero-sample tasks [47]. For EEG-based PI, changes in user status may lead to model performance degradation. In this study, we proposed a method to compare different states, which can capture the identity characteristics of the same user in different states.

III. Methods

This paper proposes a self-supervised framework, the CLT, for EEG-based PI. The overall framework is shown in Fig. 1. The CLT framework is mainly divided into the enrollment stage and the identification stage, corresponding to model training and testing, respectively. During enrollment stage, pairs of EEG signals from the same individual in different states are fed into the framework. EEG signals undergo preprocessing and feature extraction to obtain the corresponding PSD features. The proposed encoder then encodes the PSD features to obtain the corresponding embeddings. Contrastive learning is used to optimize the similarity between matched feature pairs and reduce it for unmatched feature pairs, which increases the similarity between different states of the same subject and decreases the similarity between different subjects. The multilayer perceptron (MLP) comprises a linear transformation layer, a ReLU activation layer, and another linear transformation layer. The MLP projects embeddings into the contrastive space, which is trained using the normalized temperature-scaled cross entropy loss (NT-Xent) function [37]. Before the identification stage, the MLP was discarded, and the fully connected layer (FCL) was fine-tuned for optimal performance. The FCL, a linear transformation layer, transforms the intermediate representation of the model into the final classification result.

A. Dataset

The dataset provided by PhysioNet [48] was collected using the BCI2000 system and includes EEG data from six paradigms involving 109 subjects. The six paradigms consist of two baseline runs and four task runs. The paradigms for the two baseline runs are to record EEG signals when the subjects' eyes are open and closed, respectively. The paradigm of task 1 is to open and close the corresponding fist when the target is located on the left or right side of the computer screen. The paradigm of task 2 is to imagine opening and closing the corresponding fist when the target is on the left or right side of the computer screen. The paradigm of task 3 is to open and close both fists when the target appears at the top or bottom of the computer screen. The paradigm of task 4 is to imagine opening and closing both fists when the target appears at the top or bottom of the computer screen. EO and EC have 1 session each lasting 1 minute, while the other four paradigms have 3 sessions each lasting 2 minutes, making a total of 14 sessions. Raw EEG data were recorded using 64 channels according to the 10-10 system, with an initial sampling frequency of 160 Hz, which was later downsampled to 125 Hz. In this study, we simplified these six paradigms into four, including EO, EC, actual completion of corresponding physical actions (PHY) including task 1 and task 3, and imagined completion of corresponding actions (IMA) including task 2 and task 4. After splitting the training and testing sets according to the temporal sequence, we used a 1-second window with 50% overlap to generate the samples. For each subject, each session of the EO and EC paradigms has 118 samples, while each session of the other four paradigms has 236 samples. The size of each sample is 64×125, where 64 represents the number of channels and 125 represents the number of sampling points.

The second dataset provided by Xu et al. [49] was collected using the EEG system (Brain Products GmbH, Steingrabenstr, Germany, 61 electrodes) and contains EEG data from 71 subjects in 2 experimental sessions, with each session including two paradigms: EO and EC. These two sessions were collected after normal sleep (NS) and under sleep deprivation (SD) conditions, with the original EEG signal lasting for 300 seconds and a sampling frequency of 500Hz. We downsampled the data to 125Hz and segmented it into 1-second time windows with a 0% overlap. This approach aims to simulate real-world scenarios in order to prevent data leakage. Each paradigm consists of 300 samples for each subject. The size of each sample is 61×125, where 61 represents the number of channels and 125 represents the number of sampling points.

B. Preprocessing and Feature Extraction

Before extracting power spectral density (PSD) features from the original EEG signal, we preprocessed the raw EEG data. This involved applying a bandpass filter to limit the raw EEG signal to the frequency range of 0.5 Hz to 40 Hz, removing artifacts using independent component analysis (ICA), and performing z-score normalization:

$$\hat{x}_{s,c} = \frac{x_{s,c} - \bar{x}_c}{\sigma_c}, \tag{1}$$

where s, c denotes the sampling point and the channel, $\bar{x}$ and σ_c represents the average value and standard deviation of the sample on channel c.

The original EEG signal was decomposed into five frequency bands using bandpass filters for the feature extraction: delta (0.5-4Hz), theta (4-8Hz), alpha (8-13Hz), beta (13-30Hz), and gamma (30-40Hz). Therefore, the dimensions of the EEG signals per second for the two datasets become 64×125×5 and 61×125×5, respectively. Next, the Welch method was used to compute the power spectral density (PSD) of EEG signals across five distinct frequency bands. The shapes of the PSD features per second are 64×63×5 and 61×63×5, respectively, where the three dimensions represent channels, features, and frequency bands.

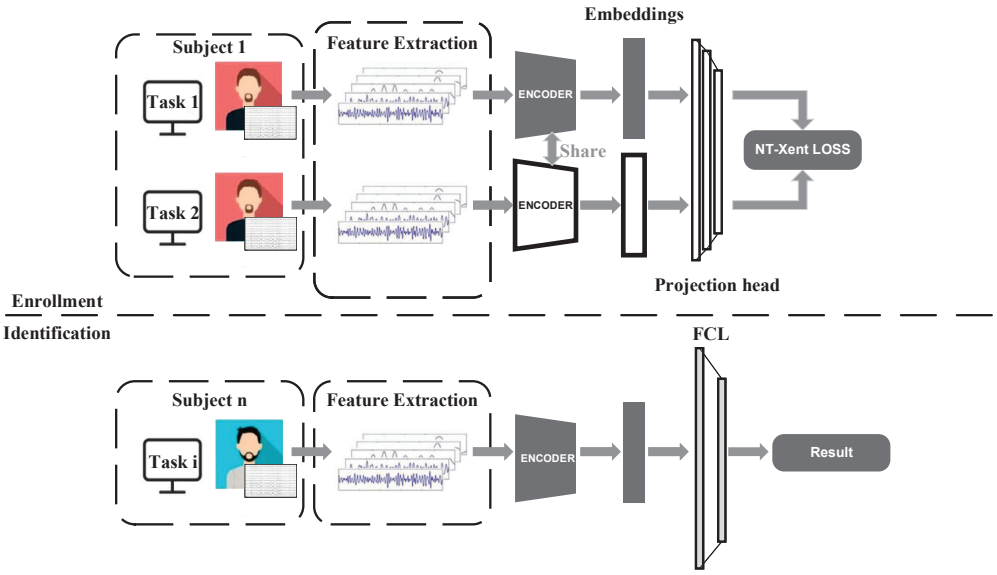


Fig. 1. Overview of the architecture of the CLT.

C. The Proposed Encoder

Figure 2 shows the architecture of the proposed encoder. It consists of an electrode attention encoder, a spectral attention encoder, patching and position embedding, and a transformer encoder, which help preserve the frequency and spatial features, thus reflecting the intrinsic patterns of brain activity.

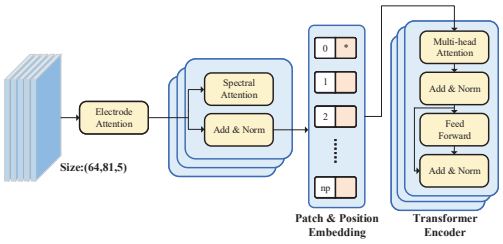


Fig. 2. Architecture of the proposed encoder.

1) The Electrode Attention Encoder: Inspired by the research of graph attention [50], we regard the PSD features $x_p \in R^{C\times P}$ from each frequency band as C nodes $c_i \in R^{1\times P}$, $i = 1, \ldots\ldots, C$, and the connections between electrodes as edges $\mathcal{C}_i$, where P is the number of PSD samples. We utilize the information from all nodes to update one of the nodes c_i:

$$c_i' = \alpha_{i,i} W c_i + \sum_{j\in\mathcal{C}_i} \alpha_{i,j} W c_j, \tag{2}$$

where c_i' represents the updated features of electrode i, $\alpha_{i,j}$ denotes the attention weight between electrodes i and j, and W is the coefficients of the linear transformation. $\alpha_{i,j}$ is calculated as follows:

$$\alpha_{i,j} = \frac{\exp\left(a^T \,\text{LeakyReLU}\left(W\left[c_i \| c_j\right]\right)\right)}{\sum_{k\in\mathcal{C}_i\cup\{i\}} \exp\left(a^T \,\text{LeakyReLU}\left(W\left[c_i \| c_k\right]\right)\right)}, \tag{3}$$

where the attention parameter a represents the weight of the feedforward layer. LeakyReLU is a variant of the rectified linear unit (ReLU) activation function that allows a non-zero gradient of 0.2 when the input is negative. In this encoder, we utilize residual connections to facilitate stable training.

2) The Spectral Attention encoder: As each PSD segment represents a specific spectral response, we use two convolutional layers and a batch normalization layer to focus on spectral information, as shown in Fig 3. To enhance the robustness of the encoder and save computational costs, the first convolutional layer reduces the dimensionality of electrodes, and the second convolutional layer restores the dimension of electrodes. The size of feature map x_p changed from C×63×5 to (C/r)×63×5 and then back to C×63×5, where C is the number of channels and r is the reduction ratio. The convolution kernels in two convolutional layers are 6×6 matrices. To avoid overfitting, we add a dropout layer at the end of the module with a dropout rate set to 0.1.

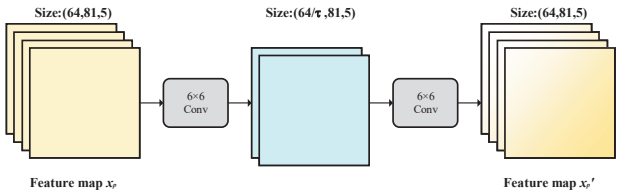


Fig. 3. The structure of the spectral attention encoder.

3) Patching and Position Embedding: To feed the feature sequence into the transformer, we divide the PSD features $x_p \in R^{C\times P}$ processed by electrode attention encoder and spectral attention encoder into a fixed-size map of patches $x_m \in R^{N\times(P_w\times P_h\times F)}$, where N, F, P_w, P_h represent the number of patches, the number of the frequency bands, width and height of each patch respectively. Then, we map

the features to the embedding vector through a linear mapping denoted by E and introduce ordered one-dimensional positional embeddings into the embedding vector. Finally, the sequence of embedded vectors is obtained as $z_0 = \left[z_{cls}, x_1^p E, x_2^p E, \ldots, x_{n_p}^p E\right] + E_{pos}$, where n_p denotes the length of the feature block.

4) The Transformer Encoder: We employ a multi-head transformer module to encode the PSD features across various frequency bands to capture the influence and dependency relationships between patches at different positions. Each encoding module comprises alternating multi-head self-attention (MSA) and MLP layers. Each module's beginning and end are equipped with residual connections and normalization layers. For a given input, self-attention within the transformer is computed to estimate the characteristics of each frequency band. Then, we weight and obtain a new representation. The calculation of self-attention is as follows:

$$\text{Attention } (Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V, \tag{4}$$

where Q, K, and V are all matrices obtained by linear input projections, and d_k is a scalar factor. The process equation is expressed as:

$$\begin{aligned} z'_p &= \text{LN}\left(\text{MSA}\left(z_{p-1}\right) + z_{p-1}\right) \\ z_p &= \text{LN}\left(\text{MLP}\left(z'_p\right) + z'_p\right), \end{aligned} \tag{5}$$

where z_{p-1} is the input to the pth encoding module and LN represents layernorm [51].

D. Contrastive Learning

The framework is constructed with contrastive learning, as shown in Algorithm 1. First, construct the dataset $\{\tilde{\boldsymbol{x}}_{2k}\}_{k=1}^{2N}$ using the signal sets $\{\boldsymbol{a}_k\}_{k=1}^{N}$ and $\{\boldsymbol{b}_k\}_{k=1}^{N}$ from two paradigms. Then, $\{\tilde{\boldsymbol{x}}_{2k}\}_{k=1}^{2N}$ is processed through the proposed encoder t to obtain the corresponding representations. Next, these representations $\{\tilde{\boldsymbol{z}}_{2k}\}_{k=1}^{2N}$ are projected into the contrastive space using the projection head m. In the contrastive space, the pairwise similarity is calculated as shown in (6):

$$\mathrm{s}_{i,j} = \boldsymbol{z}_i^{\top} \boldsymbol{z}_j / \left(\|\boldsymbol{z}_i\| \, \|\boldsymbol{z}_j\|\right) \tag{6}$$

Next, the NT-Xent loss is used to minimize the distance between positive samples, as shown in (7):

$$\begin{aligned} \ell(i,j) &= -\log \frac{\exp\left(s_{i,j}/\tau\right)}{\sum_{k=1}^{2N} \mathbf{1}_{[k \neq i]} \exp\left(s_{i,k}/\tau\right)} \\ \mathcal{L} &= \frac{1}{2N} \sum_{k=1}^{N} [\ell(2k-1, 2k) + \ell(2k, 2k-1)], \end{aligned} \tag{7}$$

where $\mathbf{1}_{[k \neq i]}$ means that the value is 1 only when $k \neq i$, and the value is 0 when $k = i$. By updating the networks t and m, the NT-Xent loss is minimized. After training, the projection head $m(\cdot)$ is discarded, and the proposed encoder $t(\cdot)$ is retained for identification tasks [37].

E. Cross-Paradigm Alignment Loss

The function $\ell(i, j)$ discussed in the previous section reduces the distance between positive samples while increasing the distance between negative samples. However, it overlooks the fact that differences still exist between positive sample pairs obtained from different paradigms. To address this issue, we introduce the cross-paradigm alignment loss for the first time, aiming to minimize the differences in representations of the same subject across different paradigms, thereby enabling the model to better capture individual consistency. Specifically, the cross-paradigm alignment loss maximizes the cosine similarity $s_{2k,2k-1}$ using the Kullback-Leibler divergence, where $2k$ and $2k-1$ represent positive samples from the same user across different paradigms. Maximizing s is equivalent to minimizing the cosine distance $d = 1 - s$, thus, for each $\ell(2k, 2k-1)$, we incorporate a penalty term:

$$D_{KL}\left(\{p_{2k}\} \,\|\, \{p_{2k-1}\}\right) = \sum_{e=1}^{E} p_{2k}^e \log \frac{p_{2k}^e}{p_{2k-1}^e}, \tag{8}$$

$$p_{2k}^e = \frac{\exp(z_{2k}^e)}{\sum_{j=1}^{E} \exp(z_{2k}^j)}, \tag{9}$$

where E represents the length of each representation in the projection space. In practice, their symmetric version $\ell_{\text{CPA}} = D_{KL}\left(\{p_{2k}\} \,\|\, \{p_{2k-1}\}\right) + D_{KL}\left(\{p_{2k-1}\} \,\|\, \{p_{2k}\}\right)$ can be used, which is the Jeffreys divergence. For each batch:

$$\mathcal{L}_{\text{CPA}} = \frac{1}{2N} \sum_{k=1}^{N} \ell_{\text{CPA}} \tag{10}$$

Here,we propose to minimize the following objective function,where λ is positive hyperparameters:

$$\mathcal{L}_{\text{CPA}} = \mathcal{L} + \lambda \mathcal{L}_{\text{CPA}}, \tag{11}$$

when there is no paradigm difference between positive sample pairs, λ = 0; when a paradigm difference exists between positive sample pairs, λ = 0.1.

F. Experiment Setup

In this study, all experiments were conducted using the NVIDIA GeForce RTX 3090 GPU. For the two datasets, the ratio of the training set to the testing set was 80% to 20%. The numbers of the spectral encoder and transformer were both set to 8 [36]. In the transformer encoder, the number of heads was set to 16. The reduction ratio was set to 4 in the spectral encoder, and the temperature parameter τ was set to 0.5 in NT-Xent. We employed 8-fold cross-validation in the training process, considering the size of the dataset. To optimize the network, we used the Adam optimizer with a learning rate set to 1e-4 and a batch size of S, where S represents the number of participants in the current dataset. For statistical testing, we employed ANOVA to assess the significance of the model comparisons.

Algorithm 1 Contrastive Learning Transformer Framework

INPUT : structure of the proposed encoder t, the projection head m, constant τ, batch size N
for sampled minibatch $\{\boldsymbol{a}_k\}_{k=1}^{N}$ $from\ task\ \boldsymbol{a}$ and $\{\boldsymbol{b}_k\}_{k=1}^{N}$ $from\ task\ \boldsymbol{b}$ **do**
  for all $k \in \{1, \ldots, N\}$ **do**
    # the first mapping
    $\tilde{\boldsymbol{x}}_{2k-1} = \boldsymbol{a}_k$
    $\boldsymbol{h}_{2k-1} = t\left(\tilde{\boldsymbol{x}}_{2k-1}\right)$ # representation
    $\boldsymbol{z}_{2k-1} = m\left(\boldsymbol{h}_{2k-1}\right)$ # projection
    # the second mapping
    $\tilde{\boldsymbol{x}}_{2k} = \boldsymbol{b}_k$
    $\boldsymbol{h}_{2k} = t\left(\tilde{\boldsymbol{x}}_{2k}\right)$ # representation
    $\boldsymbol{z}_{2k} = m\left(\boldsymbol{h}_{2k}\right)$ # projection
  for all $i \in \{1, \ldots, 2N\}$ and $j \in \{1, \ldots, 2N\}$ **do**
    $s_{i,j} = \boldsymbol{z}_i^{\top} \boldsymbol{z}_j / \left(\|\boldsymbol{z}_i\| \, \|\boldsymbol{z}_j\|\right)$ # pairwise similarity
  end for
  define $\ell(i,j)$ **as** $\ell(i,j) = -\log \frac{\exp(s_{i,j}/\tau)}{\sum_{k=1}^{2N} \mathbf{1}_{[k \neq i]} \exp(s_{i,k}/\tau)}$
  $\mathcal{L} = \frac{1}{2N} \sum_{k=1}^{N} [\ell(2k-1, 2k) + \ell(2k, 2k-1)]$
  update networks t and m to minimize $\mathcal{L}$
end for
return encoder network $t(\cdot)$, and throw away $m(\cdot)$

IV. Experiments and Results

To validate the proposed CLT, we employed several advanced methods as baseline models and conducted comparisons on two datasets. These comparisons included both training and testing in same paradigm and cross-paradigms. We also captured the attention weights and visualized them across channels and spectral, in same paradigm and cross-paradigms. Finally, we conducted ablation studies on the proposed model to confirm the functionality of each module.

A. Baseline Models

- Fuzzy entropy and SVM: The fuzzy entropy derived from the feature extraction module serves as input for the SVM [29].
- CNN: A classic CNN model with input as raw EEG time-series [29].
- GCNN: A graph convolutional neural network model with an input of phase-locking value (PLV) [20].
- AITST: The EEG temporal–spatial transformers for EEG-based personal identification, which includes spatial–temporal attention mechanism [36].

The aforementioned methods will be evaluated using the same dataset as in this study.

B. Training and Testing in Same Single-Paradigm

Following comprehensive evaluations of the proposed model, we have obtained results for training and testing on same single state. In this experimental scenario, we selected EEG signals of the same paradigm as sample pairs for contrastive learning with epochs to 500 and fine-tuning for 20 epochs. The four paradigms tested separately in this experiment include EO, EC, PHY, and IMA from Dataset 1, as well as EO and EC under NS and SD conditions from Dataset 2 (denoted as NS-EO, NS-EC, SD-EO, and SD-EC, respectively). Table I shows the results. Among the four different paradigms, GCNN, AIRST, and our proposed algorithm have good performance. The average accuracy of these three algorithms exceeds 98%.

Since our proposed algorithm shares commonalities with the AITST structure, both containing a transformer encoder, we compared the confusion matrices of our proposed method and AITST on Dataset 1, as shown in Fig. 4, which displays the confusion matrices used to observe the correct and incorrect classifications in each category of the study. The dark diagonal in the confusion matrix represents a large number of correctly classified samples for each category, whereas the light color off-diagonal signifies a small number of incorrectly classified samples. The high number of correct classifications by CLT and AITST demonstrates their effectiveness when trained and tested within a single paradigm.

C. Training and Testing in Cross-Paradigms

EEG signals exhibit pronounced fluctuations corresponding to the user's state. For example, beta waves augment when the user is concentrated, and alpha waves augment during relaxation or when the eyes are closed. Although the results in of training and testing in the same single paradigm are favorable, it is crucial to ensure that the model retains adequate accuracy across different paradigms to facilitate its use in daily settings. In this part of the experiment, we assessed the robustness of the proposed model by utilizing diverse datasets in the training and testing phases. Specifically, during the training phase, we used EC and EO datasets for comparison to capture identity features that are independent of the state. Specifically, for Dataset 1, during the testing phase, the PHY and IMA datasets were utilized for evaluation. In Dataset 2, we used the NS-EC and NS-EO datasets for comparison and the SD-EC and SD-EO datasets for evaluation. It is noteworthy that, due to 37 users in Dataset 2 not undergoing the EC experiment, the

TABLE I
RESULTS OF MODELS TRAINED AND TESTED WITHIN EACH PARADIGM. RESULTS ARE TESTING ACCURACY (AVERAGE ± STANDARD DEVIATION)%.

	Dataset 1				Dataset 2			
Method	EO	EC	PHY	IMA	NS-EO	NS-EC	SD-EO	SD-EC
Fuzzy entropy [29]	82.07±0.71	81.12±0.52	79.43±0.45	79.97±0.25	85.22±0.17	84.31±0.79	84.90±0.98	85.73±0.91
CNN [29]	93.54±0.87	94.78±1.44	95.28±1.65	95.86±0.61	96.54±0.70	95.72±1.29	97.52±0.96	97.58±0.52
GCNN [20]	**98.92±0.12**	97.88±0.09	98.67±0.07	**99.06±0.12**	**99.12±0.09**	98.84±0.06	98.89±0.15	99.31±0.04
AITST	97.60±0.17	97.73±0.08	**99.18±0.18**	97.28±0.21	98.64±0.19	98.86±0.12	**99.27±0.09**	98.65±0.17
CLT	98.68±0.09	**99.32± 0.03**	99.10±0.06	98.79±0.04	98.75±0.33	**98.99±0.18**	99.12±0.04	**99.82±0.02**

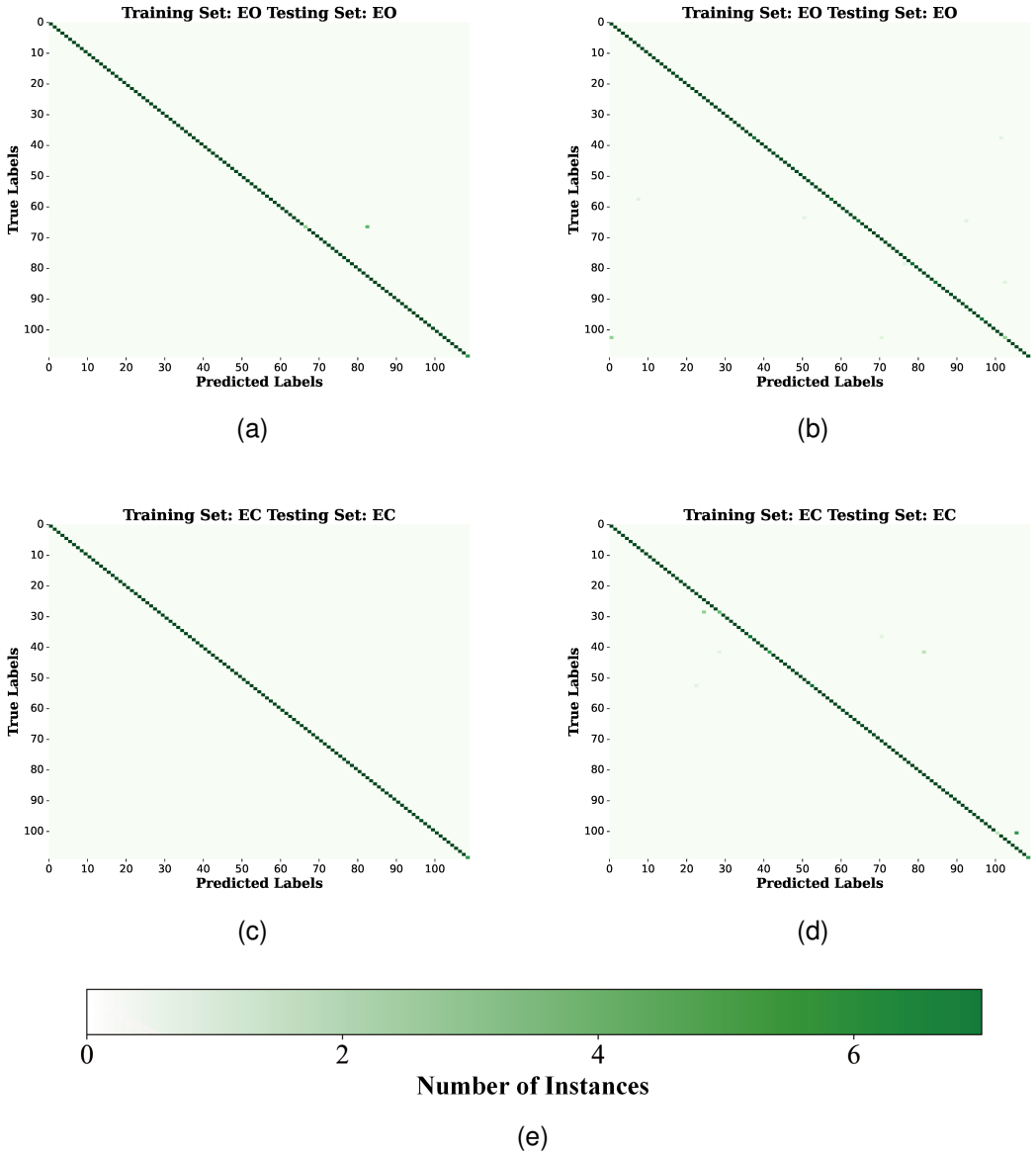


Fig. 4. Confusion matrices for the identification results of 109 subjects. (a) Results when the model is CLT and both the training and test sets are EO; (b) Results when the model is AITST and both the training and test sets are EO; (c) Results when the model is CLT and both the training and test sets are EC; (d) Results when the model is AITST and both the training and test sets are EC; (e) Color bar.

TABLE II
RESULTS OF MODELS TRAINED AND TESTED IN CROSS-PARADIGMS. RESULTS ARE TESTING ACCURACY (AVERAGE ± STANDARD DEVIATION)%.

	Dataset 1		Dataset 2	
Training set	EO&EC	EO&EC	NS-EO&NS-EC	NS-EO&NS-EC
Testing set	PHY	IMA	SD-EO	SD-EC
Fuzzy entropy [29]	18.64±2.21	15.45±3.75	30.40±1.63	35.37±0.82
CNN [29]	47.93±2.02	50.84±0.47	54.51±0.52	57.60±1.01
GCNN [20]	83.13±1.67	84.75±0.73	86.43±0.57	87.24±0.78
AITST	80.15±1.37	80.88±0.40	84.09±1.35	83.25±1.13
CLT (without $\mathcal{L}_{\text{CPA}}$)	91.11±0.62	91.24±0.43	93.53±0.73	94.73±0.53
CLT	**92.05±0.56**	**92.14±0.28**	**95.86±0.66**	**96.91±0.71**

total number of participants in the second scenario is 34. Table II displays the results of this experiment, revealing notable enhancements in the proposed model when compared with the baseline models. The results indicate that, for Dataset 1, the accuracy of CLT exceeds 92%, whereas for Dataset 2, the accuracy of CLT exceeds 95%. Compared to GCNN, our proposed CLT demonstrated performance improvements of 8.92%, 7.93%, 9.43%, and 9.67% on the four evaluation test sets, respectively. When the paradigms in the training and testing sets differ, all methods exhibited varying levels of performance degradation. Specifically, the performance of AITST decreased by approximately 12%, that of CNN by about 42%, and the accuracy of fuzzy entropy dropped below 40%. This indicates that the model proposed in this paper possesses the capability to extract identity features across diverse states.

Additionally, on Dataset 1, CLT with $\mathcal{L}_{\text{CPA}}$ improved by 0.94% and 0.90% on the two evaluation datasets, respectively (p <0.05), compared to the version without $\mathcal{L}_{\text{CPA}}$. On Dataset 2, this improvement reached 2.53% and 2.23% on the two evaluation datasets, respectively (p <0.05), demonstrating that $\mathcal{L}_{\text{CPA}}$ contributes positively to improving accuracy.

Figure 5 shows the confusion matrices in cross-paradigm scenarios of Dataset 1. Figures 5a and 5c indicate that the number of correct classifications by the CLT model is significantly higher than the number of incorrect classifications. Figures 5b and 5d show that although the AITST model has an overall higher correct classification rate than the incorrect classification rate, the number of incorrect classifications increases significantly. The numbers of noise points in Fig. 5a and Fig. 5c are significantly fewer than those in Fig. 5b and Fig. 5d, confirming the effectiveness of the CLT model in cross-state scenarios.

D. Electrode Attention Visualization

In this section, we visualize the importance of the electrodes in EEG-based PI by extracting the attention parameters a in the electrode encoder, as shown in equation (2). For electrode i, we consider $A_i = \sum_j a_{j,i}$ as the attention of electrode i. According to equation (2), $a_{j,i}$ represents the influence weight of electrode i on electrode j. Figures 6 and 7 present the visualization results of electrode attention across different frequency bands for Datasets 1 and 2 during contrastive training using the same paradigm samples. The color red indicates strong attention of the model to that region, while blue represents the opposite. The electrode attention of various frequency bands of IMA is significantly focused on FC5, while other paradigms have multiple electrodes receiving attention. These focused electrodes are task-specific. For example, attention increases in the occipital area and the frontal lobe during eye closure, and in motor imagery, attention near the motor cortex increases. After sleep deprivation, during eye closure, attention increases in the central region (near the Cz channel). This indicates that in a single-paradigm, the identity features captured by the model come from task-induced EEG, and this attentional method is not sufficiently stable when the paradigm changes. It is important to note that the EEG acquisition devices used for Dataset 1 and Dataset 2 are different. Additionally, Dataset 2 is missing five electrode channels (AFz, FCz, Iz, P9, P10) compared to Dataset 1, and includes two additional electrode channels (TP9 and TP10), which may result in slight differences in the electrode attention captured by the model. Figures 8 and 9 display the visualization results of electrode attention across different frequency bands for Datasets 1 and 2 during contrastive training using different paradigm samples, the number of electrodes receiving attention decreases significantly, and the attention range becomes noticeably focused. When comparing the EC and EO states, electrode positions such as FC1, PO3, CP2, AF8, and T8 are highlighted at different frequency bands, while in the comparison between PHY and IMA, electrode positions like FC5, PO7, and Pz are emphasized. When comparing the NS-EC and NS-EO states, electrode positions such as Pz, P2, P4, FC1, T8, and AF8 are highlighted across different frequency bands. In contrast, during the comparison between SD-EC and SD-EO, electrode positions such as CP3, AF8, CP1, and AF4 are emphasized. Compared to Fig. 6, some electrodes that receive attention in Fig. 8 are the same, such as FC5 and FC1. Some electrodes that receive attention in Fig. 8 do not receive special attention in Fig. 6. However, their weights increase in Fig. 8. This phenomenon is also observed when comparing Fig. 7 and Fig. 9. This indicates that during the comparison process, the model has learned spatial features that better represent identity under different paradigms.

E. Performance Comparison in Different Frequency Bands

To explore the contribution of different frequency bands to the performance of CLT in single-paradigm scenarios, we conducted an experiment to investigate the performance of the CLT model across five frequency bands. The results in

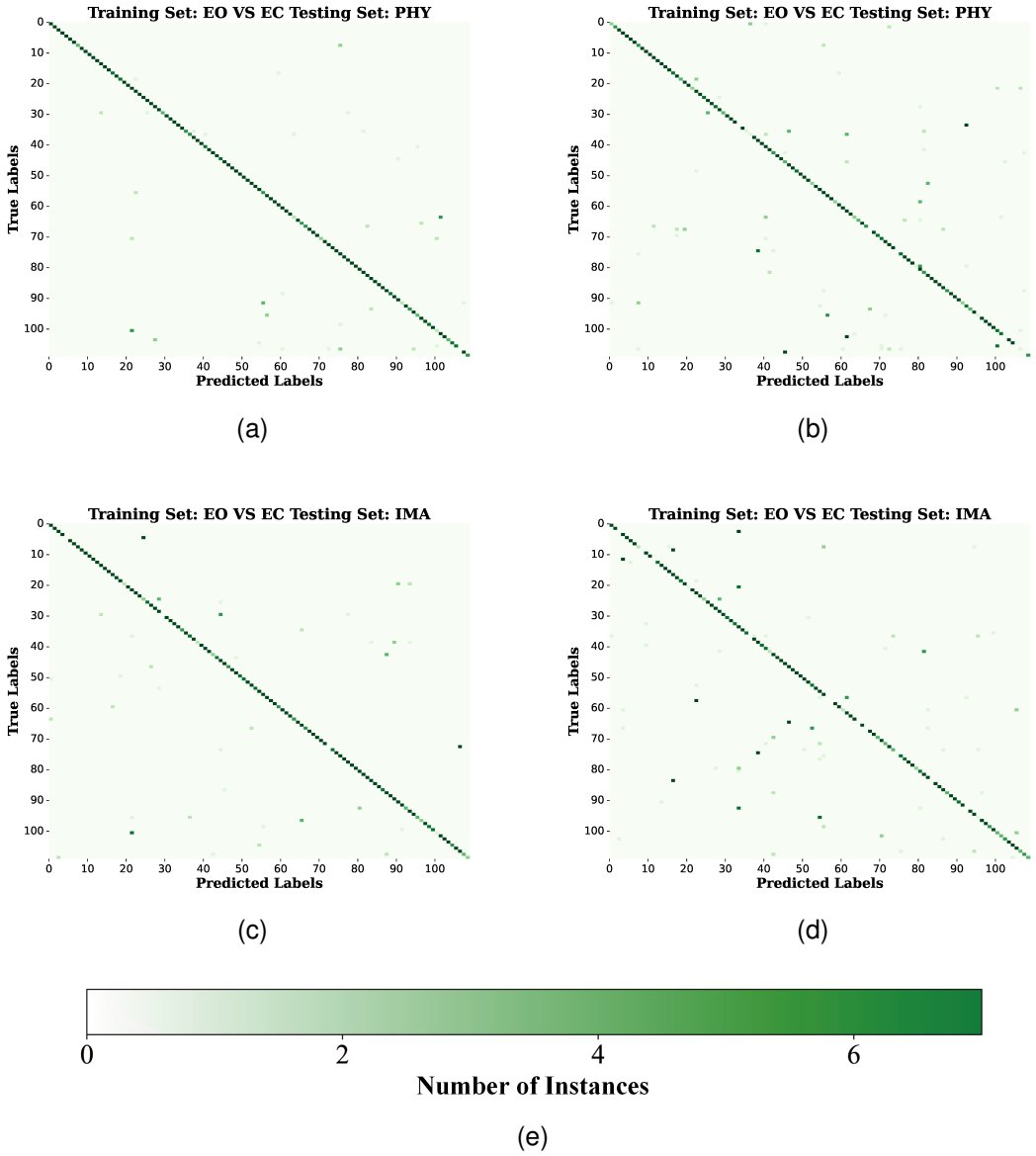


Fig. 5. Confusion matrices for the cross-paradigm identification results of 109 subjects. (a) Results when the model is CLT and the training set uses EO, EC comparison, and the test set uses PHY; (b) Results when the model uses AITST and the training set is EO and EC, and the test set uses PHY; (c) Results when the model uses CLT and the training set is EO, EC comparison, and the test set uses IMA; (d) Results when the model is AITST and the training set uses EO and EC, and the test set uses IMA; (e) Color bar.

Fig. 10 indicate that for the four paradigms, EEG data in the alpha band contributes more to recognition performance. Additionally, in the EC paradigm, the gamma band also plays an important role, while in the IMA paradigm, both the beta and gamma bands are equally important.

F. Ablation Experiment

The proposed CLT consists of a contrastive framework, electrode attention encoder, spectral attention encoder and transformer. Electrode attention focuses on the importance of different electrodes; spectral attention focuses on the importance of different frequency bands and spectral features, and transformer focuses on global context information. We further conducted ablation experiments on the various modules in the model to evaluate the contributions of each module when training and testing in diverse paradigms, as shown in Table III. We compared the model without the contrastive framework, without the electrode attention encoder, without the spectral encoder, and the whole model. The case of ”without the contrastive framework” is divided into two situations: training on a single paradigm and joint training. The results show that each module has a positive impact on the performance of the model. When any module is missing, the accuracy of the model will be reduced, among which the contrastive framework and the spectral module have the most significant impact. Without the contrastive framework (single paradigm training), the cross-paradigm recognition accuracy is reduced by 38.92% and 37.87% respectively in PHY and IMA test sets,

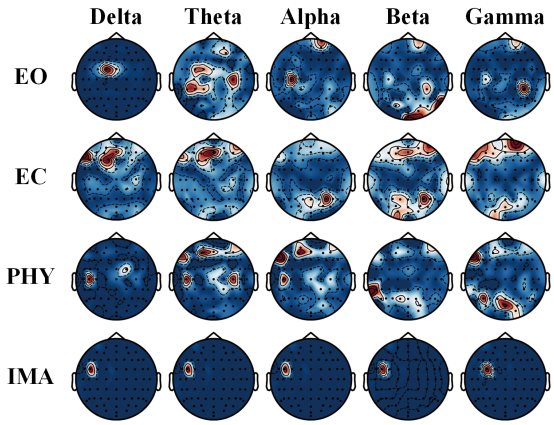


Fig. 6. Electrode attention visualization in Dataset 1 for sample comparison under the same paradigm.

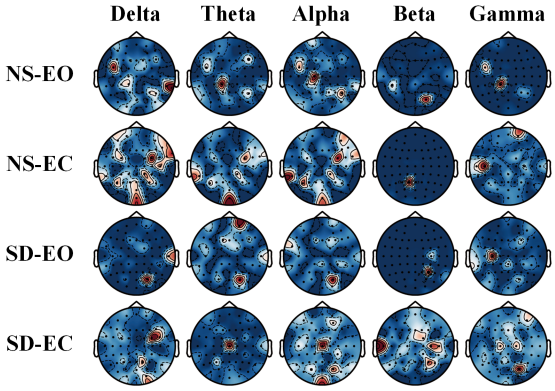


Fig. 7. Electrode attention visualization in Dataset 2 for sample comparison under the same paradigm.

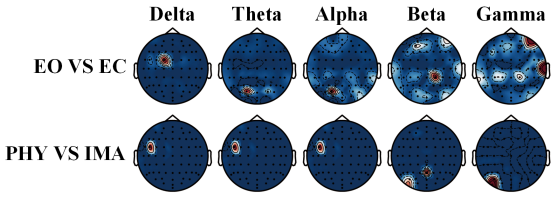


Fig. 8. Electrode attention visualization in Dataset 1 for sample comparison under the different paradigms.

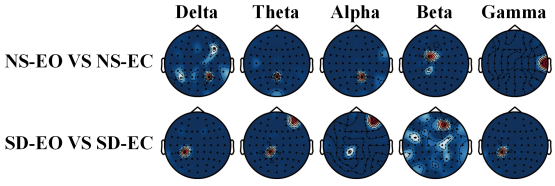


Fig. 9. Electrode attention visualization in Dataset 2 for sample comparison under the different paradigms.

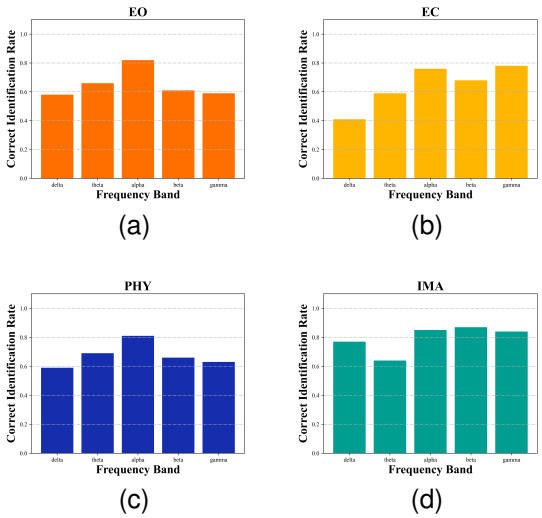


Fig. 10. Results of the CLT model in different frequency bands in individual paradigms: (a) EO, (b) EC, (c) PHY, (d) IMA.

in the case of joint training, the recognition accuracy is reduced by 11.68% and 11.38%, respectively. The spectral attention module has the second largest impact due to the PSD feature used as input. When the spectral attention module is missing, the accuracy decreases by 22.27% and 21.63% respectively. When the electrode attention encoder is absent, the accuracy decreases by 6.69% and 5.73%, respectively. These results indicate that each module is necessary for the CLT.

G. Effect of Sample Lengh

In this section, we conducted experiments using samples of different time lengths in cross-paradigm contexts, and the results obtained are the mean accuracy when the test sets are PHY and IMA. To reduce the overlap of longer samples, we set the sample size for all durations to 120. To save computational resources and time, we adjusted the number of training epochs to 100. Figure 11 shows the recognition rate of the CLT model for different sample lengths after 100 epochs of training. The results indicate that the recognition rate increases as the sample length increases. When the sample length was extended from 1s to 3s, the accuracy was increased by 6.74%.

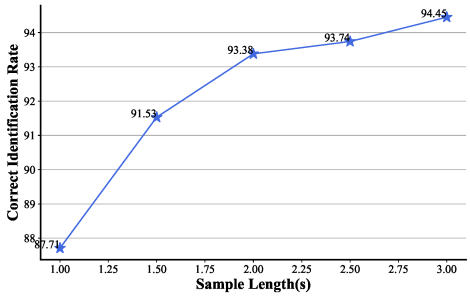


Fig. 11. Results of the CLT model in different segment lengths

TABLE III
ABLATION STUDY ON THE CLT MODEL WHEN TRAINED AND TESTED IN CROSS-PARADIGMS.

Models	PHY	IMA
Without contrastive framework (single paradigm training)	53.13	54.27
Without contrastive framework (joint training)	80.37	80.76
Without electrode attention encoder	85.36	86.41
without spectral attention encoder	69.78	70.51
CLT	92.05	92.14

V. CONCLUSION

In this study, a CLT was proposed for EEG-based PI, which achieves significant performance in cross-paradigm PI. The CLT was evaluated using two datasets: one comprised 109 subjects, while the other included 71 subjects. The model extracts both frequency domain and spatial features, utilizing contrastive learning and cross-paradigm alignment loss to maximize the similarity of features from the same subject across different paradigms. The electrode attention decoder acquires important electrodes, while the frequency attention decoder focuses on acquiring essential spectra. Additionally, the transformer is employed to extract discernible representations. Experimental results demonstrate that our proposed model achieves superior accuracy in single-paradigm and cross-paradigm EEG-based PI. Electrode attention visualization reveals differences in electrode focus between same-paradigm and cross-paradigm contrastive learning, with fewer electrodes receiving attention and more concentrated brain areas during cross-paradigm contrastive learning. This suggests that identity classification features are not entirely dependent on a user's state change. Ablation experiments confirm that our contrastive framework and the electrode and spectral attention encoders are indispensable components of the proposed CLT. Validation in real-world non-overlapping cases also demonstrates the great potential of CLT in practical applications. Furthermore, sufficient negative samples are necessary for ensuring performance when applying contrastive learning to EEG data, thus necessitating further exploration into methods for EEG data augmentation in future studies.

ACKNOWLEDGEMENT

We wish to thank the anonymous reviewers for their insightful comments.

REFERENCES

[1] X. Zhu, T. Qiu, W. Qu, X. Zhou, M. Atiquzzaman, and D. O. Wu, "BLS-location: A wireless fingerprint localization algorithm based on broad learning," *IEEE Transactions on Mobile Computing*, vol. 22, no. 1, pp. 115–128, 2021.

[2] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215–244, 2021.

[3] C.-W. Lien and S. Vhaduri, "Challenges and opportunities of biometric user authentication in the age of iot: A survey," *ACM Computing Surveys*, vol. 56, no. 1, pp. 1–37, 2023.

[4] Y. Miao, W. Jiang, N. Su, J. Shan, T. Jiang, and N. Zuo, "MLDA: Multi-loss domain adaptor for cross-session and cross-emotion EEG-based individual identification," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 12, pp. 5767–5778, 2023.

[5] K. Gorur, E. Olmez, Z. Ozer, and O. Cetin, "EEG-driven biometric authentication for investigation of fourier synchrosqueezed transform-ICA robust framework," *Arabian Journal for Science and Engineering*, vol. 48, no. 8, pp. 10 901–10 923, 2023.

[6] P. Campisi and D. La Rocca, "Brain waves for automatic biometric-based user recognition," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 782–800, 2014.

[7] H.-L. Chan, P.-C. Kuo, C.-Y. Cheng, and Y.-S. Chen, "Challenges and future perspectives on electroencephalogram-based biometrics in person recognition," *Frontiers in Neuroinformatics*, vol. 12, p. 66, 2018.

[8] M. V. Ruiz-Blondet, Z. Jin, and S. Laszlo, "CEREBRE: A novel method for very high accuracy event-related potential biometric identification," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 7, pp. 1618–1629, 2016.

[9] D. Tan and A. Nijholt, *Brain-computer interfaces and human-computer interaction*. Springer, 2010.

[10] B.-K. Min, M. J. Marzelli, and S.-S. Yoo, "Neuroimaging-based approaches in the brain-computer interface," *Trends in Biotechnology*, vol. 28, no. 11, pp. 552–560, 2010.

[11] C. Xiang, X. Fan, D. Bai, K. Lv, and X. Lei, "A resting-state EEG dataset for sleep deprivation," *Scientific Data*, vol. 11, no. 1, p. 427, 2024.

[12] S. Lopez, H. Hampel, P. A. Chiesa, C. Del Percio, G. Noce, R. Lizio, S. J. Teipel, M. Dyrba, G. González-Escamilla, H. Bakardjian *et al.*, "The association between posterior resting-state EEG alpha rhythms and functional MRI connectivity in older adults with subjective memory complaint," *Neurobiology of Aging*, vol. 137, pp. 62–77, 2024.

[13] Y. Liu, W. Dai, Y. Liu, D. Hu, B. Yang, and Z. Zhou, "An SSVEP-based BCI with 112 targets using frequency spatial multiplexing," *Journal of Neural Engineering*, vol. 21, no. 3, p. 036004, 2024.

[14] G. Zhang and S. J. Luck, "Variations in ERP data quality across paradigms, participants, and scoring procedures," *Psychophysiology*, vol. 60, no. 7, p. e14264, 2023.

[15] K. Lakshminarayanan, R. Shah, S. R. Daulat, V. Mood-

ley, Y. Yao, and D. Madathil, "The effect of combining action observation in virtual reality with kinesthetic motor imagery on cortical activity," *Frontiers in Neuroscience*, vol. 17, p. 1201865, 2023.

[16] S. Zhao, K. Liu, and X. Deng, "EEG identification based on brain functional network and autoregressive model," in *IEEE 9th Data Driven Control and Learning Systems Conference (DDCLS)*, 2020, pp. 474–479.

[17] T. Waili, G. M. Johar, K. A. Sidek, N. S. H. M. Nor, H. Yaacob, and M. Othman, "EEG based biometric identification using correlation and MLPNN models," *International Journal of Online and Biomedical Engineering (IJOE)*, no. 10, 2019.

[18] L. A. Moctezuma and M. Molinas, "Towards a minimal EEG channel array for a biometric system using resting-state and a genetic algorithm for channel selection," *Scientific Reports*, vol. 10, no. 1, p. 14917, 2020.

[19] Z. Mu, J. Hu, and J. Min, "EEG-based person authentication using a fuzzy entropy-related approach with two electrodes," *Entropy*, vol. 18, no. 12, p. 432, 2016.

[20] M. Wang, H. El-Fiqi, J. Hu, and H. A. Abbass, "Convolutional neural networks using dynamic functional connectivity for EEG-based person identification in diverse human states," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 12, pp. 3259–3272, 2019.

[21] A. B. Tatar, "Biometric identification system using EEG signals," *Neural Computing and Applications*, vol. 35, no. 1, pp. 1009–1023, 2023.

[22] S. Bak and J. Jeong, "User biometric identification methodology via EEG-based motor imagery signals," *IEEE Access*, vol. 11, pp. 41 303–41 314, 2023.

[23] J. Chen, Z. Mao, W. Yao, and Y. Huang, "EEG-based biometric identification with convolutional neural network," *Multimedia Tools and Applications*, vol. 79, pp. 10 655–10 675, 2020.

[24] S. A. Valizadeh, R. Riener, S. Elmer, and L. Jäncke, "Decrypting the electrophysiological individuality of the human brain: Identification of individuals based on resting-state EEG activity," *NeuroImage*, vol. 197, pp. 470–481, 2019.

[25] Y. Zhang, H. Shen, M. Li, and D. Hu, "Brain biometrics of steady state visual evoked potential functional networks," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 15, no. 4, pp. 1694–1701, 2022.

[26] B. Bandana Das, S. Kumar Ram, K. Sathya Babu, R. K. Mohapatra, and S. P. Mohanty, "Person identification using autoencoder-CNN approach with multitask-based EEG biometric," *Multimedia Tools and Applications*, pp. 1–21, 2024.

[27] H. Vadher, P. Patel, A. Nair, T. Vyas, S. Desai, L. Gohil, S. Tanwar, D. Garg, and A. Singh, "EEG-based biometric authentication system using convolutional neural network for military applications," *Security and Privacy*, vol. 7, no. 2, p. e345, 2024.

[28] T. Behrouzi and D. Hatzinakos, "Understanding power of graph convolutional neural network on discriminating human EEG signal," in *2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2021, pp. 1–7.

[29] Y. Du, Y. Xu, X. Wang, L. Liu, and P. Ma, "EEG temporal-spatial transformer for person identification," *Scientific Reports*, vol. 12, no. 1, p. 14378, 2022.

[30] L. Wang, L. Feng, T. Tang, D. Yang, and Y. Wei, "Brainprint recognition based on the stable SSVEP space-frequency energy distribution," in *45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2023, pp. 1–4.

[31] P. Arnau-González, M. Arevalillo-Herráez, S. Katsigiannis, and N. Ramzan, "On the influence of affect in EEG-based subject identification," *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 391–401, 2018.

[32] F. Su, L. Xia, A. Cai, and J. Ma, "Evaluation of recording factors in EEG-based personal identification: A vital step in real implementations," in *2010 IEEE International Conference on Systems, Man and Cybernetics*, 2010, pp. 3861–3866.

[33] H. Liu, X. Jin, D. Liu, W. Kong, J. Tang, and Y. Peng, "Affective EEG-based cross-session person identification using hierarchical graph embedding," *Cognitive Neurodynamics*, pp. 1–12, 2024.

[34] Ö. Yıldırım, U. B. Baloglu, and U. R. Acharya, "A deep convolutional neural network model for automated identification of abnormal EEG signals," *Neural Computing and Applications*, vol. 32, no. 20, pp. 15 857–15 868, 2020.

[35] X. Kong, W. Kong, Q. Fan, Q. Zhao, and A. Cichocki, "Task-independent EEG identification via low-rank matrix decomposition," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2018, pp. 412–419.

[36] H. Cai, J. Jin, H. Wang, L. Li, Y. Huang, and J. Pan, "AITST—affective EEG-based person identification via interrelated temporal-spatial transformer," *Pattern Recognition Letters*, vol. 174, pp. 32–38, 2023.

[37] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International Conference on Machine Learning*. Proceedings of Machine Learning Research, 2020, pp. 1597–1607.

[38] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[39] P. Li, J. Wang, Y. Qiao, H. Chen, Y. Yu, X. Yao, P. Gao, G. Xie, and S. Song, "An effective self-supervised framework for learning expressive molecular global representations to drug discovery," *Briefings in Bioinformatics*, vol. 22, no. 6, p. bbab109, 2021.

[40] X. Liu, Y. Luo, P. Li, S. Song, and J. Peng, "Deep geometric representations for modeling effects of mutations on protein-protein binding affinity," *PLoS Computational Biology*, vol. 17, no. 8, p. e1009284, 2021.

[41] X. Liu, F. Zhang, Z. Hou, L. Mian, Z. Wang, J. Zhang, and J. Tang, "Self-supervised learning: Generative or contrastive," *IEEE Transactions on Knowledge and Data*

Engineering, vol. 35, no. 1, pp. 857–876, 2021.

[42] M. N. Mohsenvand, M. R. Izadi, and P. Maes, "Contrastive representation learning for electroencephalogram classification," in *Machine Learning for Health*. Proceedings of Machine Learning Research, 2020, pp. 238–253.

[43] X. Shen, X. Liu, X. Hu, D. Zhang, and S. Song, "Contrastive learning of subject-invariant EEG representations for cross-subject emotion recognition," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 2496–2511, 2022.

[44] C.-D. Wang, X.-R. Zhu, X. Zhou, J. Li, L. Lan, D. Huang, Y. Zheng, and Y. Cai, "Cross-subject tinnitus diagnosis based on multi-band EEG contrastive representation learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 7, pp. 3187–3197, 2023.

[45] X. Li, J. Song, Z. Zhao, C. Wang, D. Song, and B. Hu, "A supervised information enhanced multi-granularity contrastive learning framework for EEG based emotion recognition," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 2325–2329.

[46] J. Y. Cheng, H. Goh, K. Dogrusoz, O. Tuzel, and E. Azemi, "Subject-aware contrastive learning for biosignals," *arXiv preprint arXiv:2007.04871*, 2020.

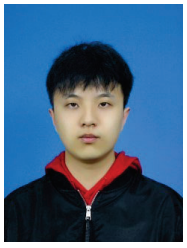
[47] Y. Song, B. Liu, X. Li, N. Shi, Y. Wang, and X. Gao, "Decoding natural images from EEG for object recognition," *arXiv preprint arXiv:2308.13234*, 2023.

[48] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[49] C. Xiang, X. Fan, D. Bai, K. Lv, and X. Lei, "A resting-state EEG dataset for sleep deprivation," *Scientific Data*, vol. 11, no. 1, 2024.

[50] S. Brody, U. Alon, and E. Yahav, "How attentive are graph attention networks?" *arXiv preprint arXiv:2105.14491*, 2021.

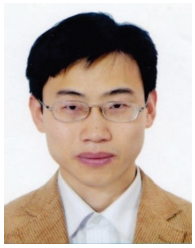
[51] Q. Wang, B. Li, T. Xiao, J. Zhu, C. Li, D. F. Wong, and L. S. Chao, "Learning deep transformer models for machine translation," *arXiv preprint arXiv:1906.01787*, 2019.



Xinghan Shao received the B.S. degree in machine engineering from Beijing Institute of Technology in 2017 and M.S. degree in mechanical and electronic engineering from Shandong University in 2020. He is currently pursuing the Ph.D. degree in biomedical engineering from the School of Biological Science & Medical Engineering, Southeast University, Nanjing, Jiangsu, China. His research interests focus on brain-computer interfaces and machine learning.

C. Chang received his Ph.D. degree in computer science from EEI of NUDT in 2017. His research interests focus on network security and machine learning.

John Q. Gan received the B.Sc. degree in electronic engineering from Northwestern Poly-technic University, China, in 1982, and the M.Eng. degree in automatic control and the Ph.D. degree in biomedical electronics from Southeast University, China, in 1985 and 1991, respectively. He is currently a professor in artificial intelligence with the University of Essex, U.K. He has coauthored a book and published over 200 research articles. His research interests include machine learning, artificial intelligence, signal and image processing, data and text mining, pattern recognition, brain-computer interfaces, and intelligent systems.



Haixian Wang received the B.S. and M.S. degrees in statistics and the Ph.D. degree in computer science from Anhui University, Hefei, Anhui, China, in 1999, 2002, and 2005, respectively. Currently, he is a full professor with the Key Laboratory of Child Development and Learning Science of Ministry of Education, School of Biological Science & Medical Engineering, Southeast University, Nanjing, Jiangsu, China. His research interests focus on biomedical signal processing, brain-computer interfaces, and machine learning.