ETST: EEG Transformer for Person Identification

Yang Du
Nanfang Hospital
Yongling Xu
Naolu Technology
Xiaoan Wang
Naolu Technology
Li Liu
Nanfang Hospital
Pengcheng Ma (pc.ma@foxmail.com)
Nanfang Hospital

Article

Keywords:

Posted Date: April 18th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1545508/v1

ETST: EEG Transformer for Person Identification

Yang Du[1,+], Yongling Xu[2,+], Xiaoan Wang[2,+,*], Li Liu[1,*], and Pengcheng Ma[1,*]

[1]Nanfang Hospital, Southern Medical University, Guangzhou, 510515, China
[2]Research Institute, Naolu Technology Co., Ltd. , Beijing, 100124, China
[*]wangxiaoan@naolubrain.com, liuli@i.smu.edu.cn, pc.ma@foxmail.com
[+]these authors contributed equally to this work

ABSTRACT

An increasing number of studies have been devoted to the use of electroencephalogram(EEG) for identity recognition due to the properties of EEG signals that are not easily stolen. Most of the existing studies on EEG person identification have only studied brain signals in a single state, requiring specific and repetitive sensory stimuli. However, the reality of human states is diverse and rapidly changing, which limits the use of their methods in realistic conditions. This demonstrates the excellent ability of the attention mechanism to model temporal signals. In this paper, we propose a transformer-based approach that extracts features in the temporal and spatial domains using a self-attention mechanism for the EEG person identification task. We conduct an extensive study to evaluate the generalization ability of the proposed method among different states. Our method is compared with the most advanced EEG biometrics techniques and the results show that our method reaches state-of-the-art results. Notably, we do not need to extract any features manually.

Introduction

In today's globalized world of information, the security of personal information has become particularly important[1], leading to the need for identification technologies. Today's identification technologies are widely used in everyday life, often using fingerprints, iris, or face recognition[2–4] and achieving high recognition accuracy rates. However, the problem with these biometrics is that they can be easily stolen and users often inadvertently reveal their identity information. The security of these technologies is not effectively guaranteed. In contrast to the above-mentioned conventional biometrics, emerging biometrics, called cognitive biometrics[5], are gradually being studied.

Unlike conventional biometrics, which rely on physiological or behavioral characteristics, cognitive biometrics rely on how people "think" and are a type of biometrics that measure human brain activity[6]. There are various methods used to measure human brain activity, and these methods are based on different principles reflecting brain activity. Functional magnetic resonance imaging (fMRI), which can use magnetic resonance imaging to measure changes in hemodynamics caused by neuronal activity, is measured by the concentration of oxyhemoglobin and deoxyhemoglobin. Positron emission tomography (PET) measures neuronal metabolism by injecting a radioactive substance into the subject's body. Near-infrared spectroscopy (NIRS) measures the concentration of oxyhemoglobin and deoxyhemoglobin by the intensity of reflection of infrared light from the cerebral cortex to reflect brain activity. Magnetoencephalography (MEG) collects the magnetic field generated by brain currents while electroencephalography (EEG) collects the electric fields generated by brain currents. They are the result of the flow of ionic currents in brain neurons in response to a specific task or a specific mental state.

We chose EEG for the identification task. Compared to other techniques, EEG is acquired by portable and relatively inexpensive devices[7,8]. The amplitude of EEG signal in normal humans ranges from 10-200uV, the frequency from 0.5-40Hz, and the temporal resolution is high, in the millisecond order[6], and it has a low spatial resolution due to the limitation of the size of the acquisition device and the interplay of the electric fields between regions of the brain. Individual variability is the basis of person identification,[9,10] have demonstrated that EEG signals have strong individual variability, especially in alpha waves where individual variability is more significant[11]. Permanence is also very important for identification, and the identity biometrics used require test-retest reliability[12,13]. The EEG signal is also highly secure, which is especially important for person identification, as it requires specialized acquisition equipment and amplifiers to collect, and cannot be inadvertently leaked by subjects or accessed remotely. At the same time, EEG-based identification is safer in crime scenarios. Users will not be forced to perform identification by criminals, as nervousness can lead to authentication failure. In addition, EEG signals are an internal trait in humans and can only be generated when the brain is active, therefore, EEG signals naturally carry the function of liveness detection[14]. Last but not least, EEG signals are universal, and everyone produces EEG signals unless some pathology causes structural damage to the brain that prevents the production of EEG signals.

In summary, EEG person identification shows great promise for application. However, most of the current researches only studied the recognition in a single state, which are still unable to guarantee the accuracy and robustness of recognition.

Therefore, we applied the attention mechanism to construct a network for identification tasks and made great progress. The main contributions of this paper are described below:

- We propose a transformer encoder-based neural network model ETST, EEG Temporal-Spatial Transformer, which can extract the information of EEG signals about individual differences in time and space domains well and ensure the accuracy of identification even in the case of cross-state.
- The ability of the transformer to extract features in the time and space domains is investigated. We also explored the effect of different position encoding on the EEG transformer.
- A data augmentation method is used to compensate for the deterioration in the effect of the transformer under the lack of data.

Related Works

The current **EEG-based biometrics systems** are broadly divided into two approaches, one is to extract distinguishable features first and then use traditional machine learning methods for classification, and the other is to use an end-to-end deep learning approach, which accomplishes both feature extraction and classification. Kong et al. assume that task-related EEG can be decomposed into two parts, including background EEG (BEEG) and residue EEG (REEG). BEEG contains a person's distinctive features and REEG is composed of task-evoked EEG and noises. They used the LRMD-based identification algorithm to decompose the EEG signal and then used the MCC algorithm to complete the classification[15]. Wang et al. argued that the functional connectivity of the brain reflects individual specificity. They computed the connectivity of the EEG signal by calculating metrics of EEG signals as feature vectors and then used a discriminant model based on Mahalanobis distance to identify people[16]. Moctezuma et al. used EMD to decompose EEG signals into a set of intrinsic mode functions (IMFs), then select the closest two IMFs and decompose them into 4 features, in this way each channel will return 8 features. Eventually, they use SVM with RBF as a classifier[17]. Also using SVM as a classifier, Alyasseri et al. used FPAβ-hc, which is a hybrid optimization technique based on binary flower pollination algorithm (FPA) and β-hill climbing to extract features[18]. Yıldırım et al. used a 1D CNN stacked with many layers to extract deep-level features of EEG signals about individual specificity. Wilaiprasitporn et al. tried to combine CNN and RNN, where CNN is used to extract spatial features and RNN is used to extract temporal features[19]. Özdenizci et al. tried an adversarial inference approach within a deep convolutional network structure, which is able to learn session-invariant and person discriminative features[20].

Currently, **Transformer** has shown good results in both NLP and CV fields[21–23]. Transformer is able to model long-range dependencies and has a faster computation speed compared with RNN or LSTM because of its parallel computing characteristic. Therefore, Transformer has taken the lead in the NLP field and has attracted the interest of researchers. However, the ability of Transformer to process EEG signals has yet to be investigated by scholars. Arjun et al. directly migrate Vit, which performs well on images, to EEG signals. The EEG signal in 1D is cut into different patches in the time dimension and used as input to the Vit model[24]. Lee et al. combined EEGNet and transformer, using an EEGNet-based convolutional neural network to obtain the temporal-spectral-spatial features[25]. Tao et al. proposed a gated Transformer, which combines the self-attentive mechanism and the gating mechanism in GRU to obtain the information of EEG signals on time series[26]. Song et al. used a CSP-based method to extract the spatial features of the EEG signals and then used a self-attentive algorithm to decode them. This method achieves the effect of state-of-the-art[27]. These approaches show that the self-attentive mechanism can improve the performance of brain-computer interface(BCI) systems.

Therefore, we designed our model based on the self-attention mechanism.

Methodology

In this paper, we propose an EEG person identification model based on the attention mechanism[21], and the overall framework diagram is shown in Fig. 1. Unlike other models, our approach does not require additional extraction of artificial features of EEG signals, and only raw EEG data was used for the identification task. We model the network in both time and space domains, taking into account the continuity of the sampled EEG signal in time and the functional connectivity among different channels in space. The model consists of two main parts, containing a temporal transformer encoder(TTE) and a spatial transformer encoder(STE). Firstly, we preprocess the EEG signal and then feed it as input to ETST. In the TTE part, we use the attention mechanism in time domain to calculate the correlation among different sampling points in samples, which is used to extract the time-domain features of the EEG. Since there is individual specificity in the coupling relationship of channels between individuals, we design the STE part to calculate the spatial domain attention for channels to capture the coupling relationship among different channel signals, which enables the model to identify different individuals based on the specific coupling relationship and the identification is more stable. Finally, a simple fully connected layer is used to aggregate global information and perform classification. In the following, we will explain the preprocessing method and ETST model in detail.

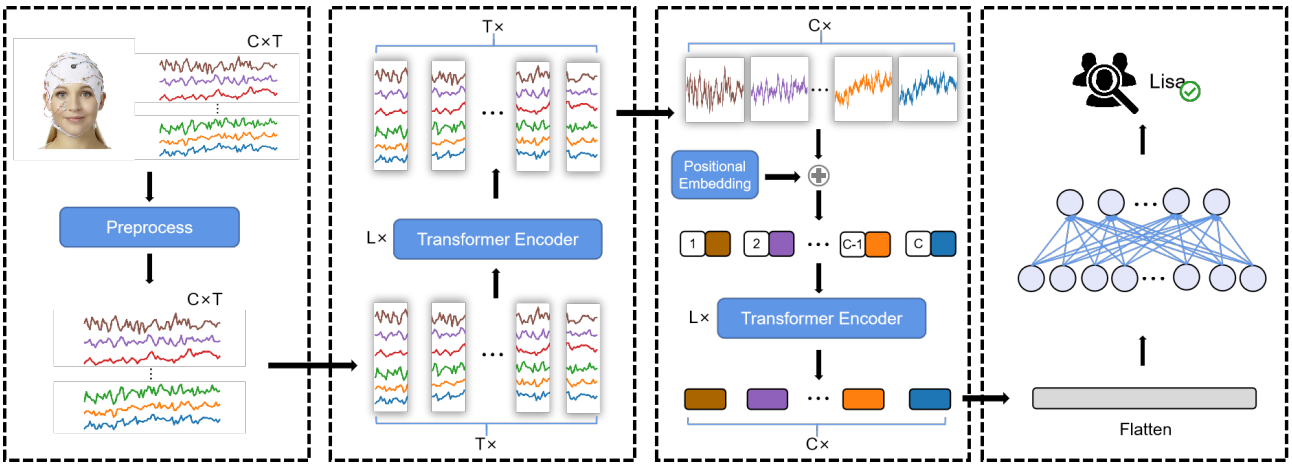

Figure 1. The architecture of the ETST model.

Preprocessing. The raw EEG signal is filtered using a [0.5 42] Hz bandpass filter for removing low and high-frequency noise, and we remove the ocular and muscular artifacts using independent component analysis (ICA). The size of each sample is T × C, where T is the number of sampling points and C is the number of EEG channels. For each sample, the following z-score standardization over time for each channel will be employed:

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

where t , c denotes the sampling point and channel of the sample, $\bar{x}_c$ denotes the mean of the sample on channel c, and σ_c denotes the standard deviation of the sample on channel c. After standardization, the mean of the data on each channel of the sample is 0 and the standard deviation is 1.

Temporal Transformer Encoder. The correlation among sampling points in each sample represents the time-domain information of EEG signals, which can represent the interrelationship of EEG signals at different times. Inspired by the attention mechanism[21], we use multiple transformer blocks to encode the time-domain information of the EEG, which can consider the long-range dependence of the EEG time-domain. For a given input $X = [x^1, x^2, \cdots, x^T] \in \mathbb{R}^{T \times C}$, we compute self-attention in the transformer block to estimate the similarity between sampled points in different samples and weight the sum to obtain the new representation. Self-attention is computed as follows:

$$Attention(Q,K,V) = Softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{2}$$

where Q, K, and V are all matrices obtained by linear projections of the input and d_k is a scalar factor. To enhance the capability of self-attention, we compute multiple self-attention on the input to obtain better representation, i.e., we adopt the multi-head attention mechanism. Each transformer encoder contains two parts: multi-head attention (MHA) and multi-layer perceptron (MLP). Each part employs residual connection[28] and layer normalization (LN)[29] to improve the speed of training and robustness of the model. Fig. 2 illustrates the above calculation process. The TTE part can be expressed by :

$$h_l^t = LN(MHA(z_{l-1}^t) + z_{l-1}^t) \quad l = 1, 2, \cdots, L \tag{3}$$

$$z_l^t = LN(MLP(h_l^t) + h_l^t) \quad l = 1, 2, \cdots, L \tag{4}$$

Spatial Transformer Encoder. The different channels in the EEG signal represent electrodes with different locations on the scalp, and the functional connectivity between different brain regions can be calculated by considering the dependencies among different channels. Similar to TTE, in STE we also used the attention mechanism to model the spatial information among

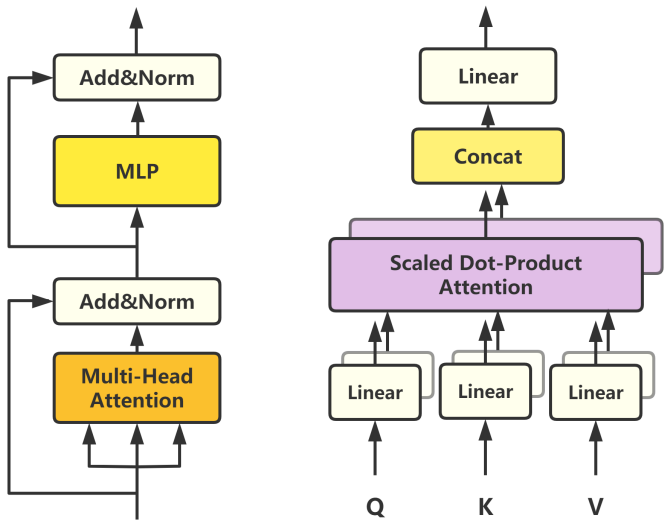

Figure 2. (left) The architecture of a transformer encoder. (right) Multi-Head Attention.

different channels. In order to preserve the spatial location information, we added the position encoding of spacial domain to the input and then fed the result to STE:

$$z_0^s = tran(z_L^t) + E_{pos} \tag{5}$$

where $tran()$ representation is a transpose operation and the $E_{pos} \in \mathbb{R}^{C\times T}$ representation is a position encoding. In this paper, we use the position encoding in the form of a trigonometric function at a fixed position, z_0^s representing the representation with the addition of spatial position information. In the STE, we use a similar structure to that in the TTE to learn the spatial information on the different channels of the EEG. The process equation is expressed as:

$$h_l^s = LN(MHA(z_{l-1}^s) + z_{l-1}^s) \quad l = 1,2,\cdots,L \tag{6}$$

$$z_l^s = LN(MLP(h_l^s) + h_l^s) \quad l = 1,2,\cdots,L \tag{7}$$

Classification Layer. ETST learns the time-domain information of the EEG data on the different sampling points in TTE. In the subsequent STE, ETST learns the spatial information among channels. Thus, the output of the transformer encoder layer yields a better representation containing both time-domain and space-domain features. To fuse the global information learned by the representation and use it for classification, a simple fully-connected layer with only one layer is used to obtain the final classification output which is optimized using the cross-entropy loss function.

$$L = -\frac{1}{N}\sum_{n=1}^{N}\sum_{c=1}^{C} y_n^c log(\hat{y}_n^c) \tag{8}$$

where N denotes the number of batch sizes and C denotes the number of categories. y_n^c is the true one hot label, $\hat{y}_n^c$ is the predicted probability of the corresponding category.

Ethical approval. This paper does not contain any studies with human or animals participants performed by any of the authors.

Experiments

Dataset. We validated our method on an EEG dataset provided by PhysioNet[30]. This dataset was recorded using the BCI2000 system with over 1500 segments of one- and two-minute EEG data containing 109 subjects. the sampling frequency of the

system was 160 Hz. These EEG data were recorded with 64 electrodes, which conformed to the 10-10system. Subjects were asked to do motor/imagery tasks while the EEG signal was recorded by the system, and each subject completed 14 experimental runs including 2 one-minute baseline runs and 12 two-minute task runs. The specific experimental procedure is described below. In the baseline runs, the EEG signals were recorded while the subjects kept their eyes open (EO) and eyes closed (EC), respectively. In the task runs, subjects were asked to complete four motor/imagery tasks, where subjects were asked to actually complete the corresponding physical action (PHY) or imagine completing the corresponding action (IMA) when the target appeared on the computer, and rest when the target disappeared. Task 1 was to open and clench the corresponding fist when a target is on the left or right side of the computer screen, task 2 is to imagine opening and clenching the corresponding fist when a target is on the left or right side of the computer screen, task 3 is to open and clench both fists when a target appears on the top or bottom of the computer, task 4 is to imagine opening and clenching both fists when a target appears at the top or bottom of the computer. The four tasks form a loop, and a total of 3 loops are executed, for a total of 12 task runs. A 1-second window with 50% overlap of each channel was used to generate samples. Therefore, the shape of a sample is 160×64.

Experiment Design. To make EEG person identification technology realistic and feasible, the stability and robustness of the system must be able to be guaranteed. This also means that the model needs to be able to consistently and accurately identify subjects by their EEG signals, even if the subjects are in different states, such as happy or calm, or even thinking about something. We conducted several experiments to verify the effectiveness and practicability of TST on EEG biometrics. The EEG signal in the Physionet Dataset contains four states, EO, EC, PHY, and IMA. We designed various experiments based on these four different states to test the performance of ETST in diverse scenarios. For example, we train different models in datasets with different preprocessing for a more comprehensive comparison. The experiments we conducted are described below.

1.We compare our model with state-of-the-art EEG identification methods and also with traditional neural network methods such as Convolution Neuron Network(CNN), MLP, and traditional machine learning methods such as Support Vector Machine(SVM), Random Forest(RF), and Linear Discriminant Analysis(LDA). In the comparison experiments with other methods, we set up three sub-experiments, corresponding to three different application scenarios. The first one is training and testing in a single human state, and we conducted training and testing in four states, EC, EO, IMA, and PHY, which corresponds to the case of EEG person identification in a fixed state. The second one is to train in one state and test in another state, we will train under EC and EO data and test under IMA and PHY. This type of task is the most challenging, and it tests whether the model obtained by training under one EEG paradigm can be generalized to other EEG paradigms. The third one is a mixture of EC, EO, IMA, and PHY datasets for training and testing. This experiment verifies that the trained model can guarantee reliable accuracy even under multiple human states.

2.We performed ablation experiments to explore the effect of each part of the model on the results. Position encoding is an important component of the model. The EEG signal contains position information in both the time and space domains. Transformer ensures that the model retains the location information by adding location encoding to the input species. We investigate the effect of adding time-domain location encoding and space-domain location encoding on person identification separately. In addition to comparing the location encoding, we also conducted ablation experiments on the encoder part of ETSF and investigated the performance of ETSF when removing TTE and STE respectively, to explore the role of different parts.

3.In other EEG identification methods, the segmentation length of samples is not uniform. For example, the segmentation length used by Wang et al. is 1s[31], while the segmentation length used by Thiago Schons et al. is 12s[32], and there may be a large gap between the sample segmentation lengths of different methods. Since the Transformer can model long-distance representation, Namuk Park et al. suggested that weak inductive bias would have some impact on the performance of transformer[33]. Therefore, we divided the dataset with different split lengths in our experiments for exploring the performance of ETST with different sample split lengths.

In addition to different segmentation lengths, the sample overlap rate also directly affects the size of the resulting sample size and the degree of information overlap among different samples. The loss function of Transformer is smoother than that of CNN[33], which may be more difficult to converge with smaller sample sizes, resulting in worse performance. Therefore, we designed experiments with different sample overlap lengths and obtained training datasets with different sample sizes to explore the effect of sample size on our model.

Experiment Detail. All experiments in this paper were performed on a GeForce TITAN XP GPU, Python 3.8 and Pytorch 1.10.2 configuration. The number of TTE layers, the number of heads of TTE layers, the number of STE layers, and the number of heads of STE layers in the model were set to 2, 8, 2, and 8, respectively. we used the AdamW[34] optimizer with learning rate, weight decay, and batch size of 4e-5, 1e-6, and 256, respectively, to optimize the network.

Method	EO	EC	PHY	IMA
FuzzEn+SVM	84.14±0.83	83.73±0.71	77.93±0.59	80.84±0.18
PSD+RF	91.82±0.80	89.72±0.28	96.05±0.21	95.79±0.81
PLV+MLP	97.34±0.54	96.14±0.56	98.44±0.36	98.50±0.31
Raw+CNN	96.89±0.77	67.43±47.36	97.96±1.55	97.42±0.83
PLV+CNN	99.36±0.03	99.02±0.04	99.41±0.02	99.84±0.00
COR+GCNN	99.75±0.11	99.48±0.26	99.94±0.02	99.98±0.02
PLV+GCNN	99.97±0.03	99.88±0.03	**99.99±0.02**	**100.00±0.00**
Ours	**100.00±0.00**	**99.96±0.06**	99.97±0.01	**100.00±0.00**

Table 1. Results of models training and testing within each human state.Results are accuracy in testing stage(average ± standard deviation)%

Method	PHY	IMA
FuzzEn+SVM	16.16±0.01	15.61±0.00
PSD+RF	23.15±0.74	22.13±1.04
PLV+MLP	56.74±1.32	57.12±0.87
Raw+CNN	49.26±3.85	52.51±2.26
PLV+CNN	80.52±4.39	82.73±3.35
COR+GCNN	86.99±2.37	87.18±2.86
PLV+GCNN	85.40±1.62	87.03±2.53
Ours	**97.29±0.03**	**97.45±0.13**

Table 2. Results of models training on resting states and testing on diverse states.Results are accuracy in testing stage(average ± standard deviation)%

Method	Results
FuzzEn+SVM	73.45±0.10
PSD+RF	89.40±0.12
PLV+MLP	97.71±2.25
Raw+CNN	99.85±0.06
PLV+CNN	99.79±0.03
COR+GCNN	99.15±0.56
PLV+GCNN	**99.98±0.02**
Ours	99.90±0.03

Table 3. Results of models training on diverse states and testing on diverse states.Results are accuracy in testing stage(average ± standard deviation)%

Results and Discussion

Evaluation and Comparison with Baseline. Currently, EEG-based person identification algorithms are broadly classified into two categories. One is the traditional machine learning algorithms, which generally require manual feature extraction including power spectral density (PSD), auto-regressive coefficient (AR), and fuzzy entropy (FuzzEn). Another category is deep learning algorithms, such as CNN-based or RNN-based neural network models. Since the concept of graph fits well with the functional connectivity in neuroscience, where graph features are used to represent the relationships among brain regions, graph convolutional neural networks are beginning to be used in the field of EEG. The results of graph convolution in EEG biometrics have also achieved state-of-the-art result. The most effective graph convolutional neural networks use the graph feature Phase Locking Value(PLV) as input. We compared our method with SVM, RF, MLP, and the state-of-the-art approach[31].

EEG is the result of the corresponding ionic currents generated by brain neurons in response to a specific task or a specific mental state. EEG signals in different states are characterized differently, for example, δ waves are associated with increased attention, alpha waves are enhanced in a resting state, while beta waves become stronger when we imagine or actually move body parts in motion. We investigated the performance of ETST in the same single state. We trained and tested ETST on a single-state dataset to evaluate the mentioned performance.The results are shown in Table 1.The experimental results show that our proposed method outperforms all methods when the data are in the same state , except for one result which is slightly lower

Models	PHY	IMA
Non PE	95.84±0.11	96.07±0.03
with Temporal PE	79.98±13.03	80.89±12.76
with Spatial PE	**97.45±0.13**	**97.29±0.03**
with Temporal+Spatial PE	90.56±1.94	91.20±1.92

Table 4. Results of the ETST model with different position encoding

Models	PHY	IMA
with TTE	72.98±0.39	75.19±0.09
with STE	68.98±0.34	70.22±0.47
with TTE+STE	**95.84±0.11**	**96.07±0.03**

Table 5. Ablation study on the ETST model(without position encoding)

than that of GCNN, only 0.2% lower.

The EEG in different states is variable. But for EEG biometrics to be practically applied in real life, the algorithm needs to be robust to states and able to recognize the identity of the user even in a different state. We evaluate the generalization ability of our proposed method in different states by training and testing ETST on different datasets. EO and EC data are used as training sets and tested on PHY and IMA data, respectively. Table 2 shows the experimental results of this experiment, which is the training set and test set are across different states. The results show that ETST has a significant improvement compared to other methods in the condition of different states. Compared with GCN, the improvement are 10.3% in PHY and 10.27% in IMA. When the states in the training and test sets were different, all methods suffered different degrees of performance degradation, with GCN decreasing by about 13%, SVM by about 40%, and the accuracy of the remaining methods dropping to less than 30%. This indicates that the other models are limited to extracting features from the same states and have weak generalization ability for different states. In contrast, the ETST model only decreases by about 3%, which indicates that the ETST is able to extract features that are valid across diverse states.

To enhance the model's ability to generalize to EEG signals in various states, in addition to the strong generalization ability of the model itself, another approach is to include multiple states right in the training set and make the model learn to extract distinguishable features that apply to diverse states. ETST also achieves close to the best results when all states were contained in both the training and test sets, including EO, EC, PHY, and IMA, and the specific experimental results are shown in Table 3. Compared to experiment on diverse states, the results of this experiment show less decrease in accuracy, and only SVM has a considerable decrease, down to 73%. It shows that different algorithms can guarantee certain results in case the training and test sets contain the same state data. However, this method of enhancing the effect is not applicable to the use of realistic scenarios for lots of states. It is impossible to contain data of all states in the train set. Therefore, the key to solving the EEG-based person identification problem is to improve the generalization ability of the model between different states. And our proposed ETST possesses a strong generalization ability among different states.

Ablation Experiment. In Transformer, self-attention calculates attention weights for all inputs simultaneously and sums the weights to obtain the output. In this process, self-attention considers the global information and discards the location information of the input data. For EEG data, the signal contains location information in both the time and space domains, representing different temporal sampling points and various brain regions, respectively. To investigate the effect of location information in EEG on person identification, we tried to retain the location information present in the EEG data by adding location encoding to the input of TTE and STE layers, respectively. We compare the effect of adding positional encoding to ETST in the time and space domains under the cross-state dataset, and the results are shown in Table 4. It shows that adding only the spatial positional encoding is the best result, which achieves the best performance of our model (97% in IMA, 97% in PHY). Adding both temporal and spatial positional encoding yields the next best result (96% in IMA, 95% in PHY). We find that the model performance can be improved by adding the location encoding information in space, and conversely, adding the location encoding information in time would make the model perform worse. In addition, by observing the training process of the model, we discovered that adding the location information in the time domain also affects the training efficiency to a certain extent, making the model more likely to converge to worse minima, which leads to worse results. We believe that the inclusion of absolute position encoding in the time domain breaks the translational invariance in the time domain, thus making it more difficult for the model to extract time-domain features. The absolute spatial position encoding retains the position information of different channels. Unlike the same sampling point which may appear in different positions in different samples, the channel positions in samples are fixed, and the inclusion of absolute position encoding in the space domain could instead improve the

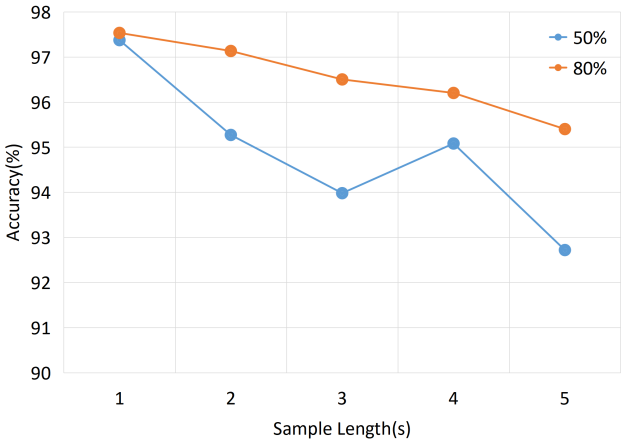

Figure 3. Results of the ETST model in different segment length and overlap

model's ability of spatial feature extraction .

The ETST model contains two parts, the TTE layer, and the STE layer, for extracting time-domain features and space-domain features, respectively. To illustrate the importance of the two distinct features on the experimental results, we conducted ablation experiments under cross-state for the model to reflect the necessity of each part of our model. As can be seen in Table 5, we compared the results under the TTE, STE, and TTE+STE models. The results indicate that using only the TTE layer or only the STE layer both make the results significantly lower. Moreover, the results show that the TTE layer has a little higher classification accuracy than STE (75.19% in IMA and 72.98% in PHY vs. 70.22% in IMA and 68.98% in PHY). Further, it can be shown that for person identification time domain information is more important than space domain information. In order to simultaneously EEG temporal and spatial information, our model consists of TTE and SPE layers, which can considerably improve the performance of the model and thus achieve the state-of-the-art effect.

Effect of Sample Lengths and Sample Size. The sample segmentation lengths used in different methods vary. As a result, some methods may only work with shorter sample segmentation lengths, while others do the opposite. Training the same method with samples of different split lengths may yield widely varying results. To illustrate the generalizability in sample length of our method, we compare the classification accuracy of the model under different segmentation length samples. Namuk Park et al.[33] mentioned that for Transformer, the size of the dataset directly affects the final training results due to its smoother loss function, i.e., transformer performs worse with fewer samples.

We attempt to increase the number of samples by increasing the overlap rate of the sliding window. Data augmentation of the samples is performed using an overlap rate of 80% and the effect of ETST is compared for different size training sets. Fig. 3 shows the effect of the model after training with different training sets. From the figure, it can be seen that the 1s length sample sequence achieves the best results with the same overlap rate, and the longer the time, the lower the classification accuracy. When the overlap ratio is constant, increasing the sample length leads to a reduction of training samples, and the sample size of 5s segmentation length is only about one-fifth of that of 1s, which leads to a decrease of ETST model accuracy. When the overlap ratio was increased to 80%, the sample size of the dataset increased a lot, about two times, and the model effect of the dataset picked up, with the accuracy of 5s rising to 95.44%, slightly below the accuracy of 1s, which dropped about 2%. This demonstrates that the reason for the deterioration in the effectiveness of the transformer-based model in the experiments we designed is the insufficient sample size of the data. As a whole, training samples by either length division can achieve state-of-the-art results for our model under cross-state as long as they have sufficient training data.

Conclusion

In this paper, we propose ETST, a deep learning model based on the attention mechanism. We used a multi-headed attention mechanism to extract the temporal and spatial features of EEG signals. The temporal transformer encoder in the model is able to extract long-range distinguishable representations, and the spatial transformer encoder is able to acquire spatial dependencies among channels, which characterize the functional connectivity among different brain regions.In this way, through several rounds of attention weighting, the model is able to focus its attention on the features that are most relevant to the classification results. The experimental results indicate that our method achieves state-of-the-art results on person identification, which also validates the feasibility of EEG on biometrics. The model is also robust to different states. The results of the ablation

experiments show that the temporal features have a slightly more significant effect on the outcome of the EEG biometrics. The experimental results also show that absolute position encoding in space enhances the model, indicating that specific channels have an effect on person identification, not just the correlation among channels.The use of longer EEG data causes a slight decrease in the effectiveness of the attention mechanism. The experiments demonstrate that the application of Transformer in EEG requires sufficient data to ensure its effectiveness. Therefore, it is necessary to investigate the data argument method for EEG data in future studies. In addition, the choice of hyper-parameters for our model is not yet optimal due to the limitation of time, which indicates that the model performance is not yet the best.

The stability and permanence problems are two key issues in implementing EEG biometrics into practical use, and there is a need to ensure that the model can re-identify users correctly in different states and at different times. This requires the model to be able to extract time-invariant and state-invariant distinguishable features. In future work, we will try more ways to extract the features of EEG signals, such as filtering the alpha band features of EEG signals, which have a strong inter-individual difference in the resting state, or selecting the channels with strong correlation to person identification and removing the effect of redundant channels. At the same time, experiments on person identification based on subjects' EEG on different days are yet to be conducted.

Data availability

The dataset used for this study is publicly available and accessible online at PhysioNet Database [https://physionet.org/content/eegmmidb/1.0.0/][30].

References

1. Soomro, Z. A., Shah, M. H. & Ahmed, J. Information security management needs more holistic approach: A literature review. *Int. J. Inf. Manag.* **36**, 215–225 (2016).
2. Cappelli, R., Ferrara, M. & Maltoni, D. Minutia cylinder-code: A new representation and matching technique for fingerprint recognition. *IEEE transactions on pattern analysis machine intelligence* **32**, 2128–2141 (2010).
3. Masek, L. *et al. Recognition of human iris patterns for biometric identification*. Ph.D. thesis, Citeseer (2003).
4. Guillaumin, M., Verbeek, J. & Schmid, C. Is that you? metric learning approaches for face identification. In *2009 IEEE 12th international conference on computer vision*, 498–505 (IEEE, 2009).
5. Revett, K., Deravi, F. & Sirlantzis, K. Biosignals for user authentication-towards cognitive biometrics? In *2010 International Conference on Emerging Security Technologies*, 71–76 (IEEE, 2010).
6. Campisi, P. & La Rocca, D. Brain waves for automatic biometric-based user recognition. *IEEE transactions on information forensics security* **9**, 782–800 (2014).
7. Tan, D. & Nijholt, A. Brain-computer interfaces and human-computer interaction. In *Brain-Computer Interfaces*, 3–19 (Springer, 2010).
8. Min, B.-K., Marzelli, M. J. & Yoo, S.-S. Neuroimaging-based approaches in the brain–computer interface. *Trends biotechnology* **28**, 552–560 (2010).
9. Berkhout, J. & Walter, D. O. Temporal stability and individual differences in the human eeg: An analysis of variance of spectral values. *IEEE Transactions on Biomed. Eng.* 165–168 (1968).
10. Vogel, F. The genetic basis of the normal human electroencephalogram (eeg). *Humangenetik* **10**, 91–114 (1970).
11. Van Dis, H., Corner, M., Dapper, R., Hanewald, G. & Kok, H. Individual differences in the human electroencephalogram during quiet wakefulness. *Electroencephalogr. clinical neurophysiology* **47**, 87–94 (1979).
12. Henry, C. E. Electroencephalographic individual differences and their constancy: I. during sleep. *J. Exp. Psychol.* **29**, 117 (1941).
13. Henry, C. E. Electroencephalographic individual differences and their constancy: Ii. during waking. *J. Exp. Psychol.* **29**, 236 (1941).
14. Ruiz-Blondet, M. V., Jin, Z. & Laszlo, S. Cerebre: A novel method for very high accuracy event-related potential biometric identification. *IEEE Transactions on Inf. Forensics Secur.* **11**, 1618–1629 (2016).
15. Kong, X., Kong, W., Fan, Q., Zhao, Q. & Cichocki, A. Task-independent eeg identification via low-rank matrix decomposition. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 412–419 (IEEE, 2018).

16. Wang, M., Hu, J. & Abbass, H. A. Brainprint: Eeg biometric identification based on analyzing brain connectivity graphs. *Pattern Recognit.* **105**, 107381 (2020).

17. Moctezuma, L. A. & Molinas, M. Multi-objective optimization for eeg channel selection and accurate intruder detection in an eeg-based subject identification system. *Sci. Reports* **10**, 1–12 (2020).

18. Alyasseri, Z. A. A., Khader, A. T., Al-Betar, M. A. & Alomari, O. A. Person identification using eeg channel selection with hybrid flower pollination algorithm. *Pattern Recognit.* **105**, 107393 (2020).

19. Yıldırım, Ö., Baloglu, U. B. & Acharya, U. R. A deep convolutional neural network model for automated identification of abnormal eeg signals. *Neural Comput. Appl.* **32**, 15857–15868 (2020).

20. Özdenizci, O., Wang, Y., Koike-Akino, T. & Erdoğmuş, D. Adversarial deep learning in eeg biometrics. *IEEE signal processing letters* **26**, 710–714 (2019).

21. Vaswani, A. *et al.* Attention is all you need. *Adv. neural information processing systems* **30** (2017).

22. Dosovitskiy, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).

23. Liu, Z. *et al.* Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10012–10022 (2021).

24. Arjun, A., Rajpoot, A. S. & Panicker, M. R. Introducing attention mechanism for eeg signals: Emotion recognition with vision transformers. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 5723–5726 (IEEE, 2021).

25. Lee, Y.-E. & Lee, S.-H. Eeg-transformer: Self-attention from transformer architecture for decoding eeg of imagined speech. In *2022 10th International Winter Conference on Brain-Computer Interface (BCI)*, 1–4 (IEEE, 2022).

26. Tao, Y. *et al.* Gated transformer for decoding human brain eeg signals. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 125–130 (IEEE, 2021).

27. Song, Y., Jia, X., Yang, L. & Xie, L. Transformer-based spatial-temporal feature learning for eeg decoding. *arXiv preprint arXiv:2106.11170* (2021).

28. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778 (2016).

29. Ba, J. L., Kiros, J. R. & Hinton, G. E. Layer normalization. *arXiv preprint arXiv:1607.06450* (2016).

30. Goldberger, A. L. *et al.* Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation* **101**, e215–e220 (2000).

31. Wang, M., El-Fiqi, H., Hu, J. & Abbass, H. A. Convolutional neural networks using dynamic functional connectivity for eeg-based person identification in diverse human states. *IEEE Transactions on Inf. Forensics Secur.* **14**, 3259–3272 (2019).

32. Schons, T., Moreira, G. J., Silva, P. H., Coelho, V. N. & Luz, E. J. Convolutional network for eeg-based biometric. In *Iberoamerican Congress on Pattern Recognition*, 601–608 (Springer, 2017).

33. Park, N. & Kim, S. How do vision transformers work? *arXiv preprint arXiv:2202.06709* (2022).

34. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017).

Author contributions

Yang Du and Yongling Xu proposed the method, performed the experiments, and wrote the manuscript. Xiaoan Wang, Li Liu, and Pengcheng Ma gave guidance on the experiment and reviewed the manuscript.

Competing interests

The authors declare no competing interests.