Inter-participant Transfer Learning with Attention based Domain Adversarial Training for P300 Detection

Inter-participant Transfer Learning with Attention based Domain Adversarial Training for P300 Detection

Shurui Li[a], Ian Daly[b], Cuntai Guan[c], Andrzej Cichocki[d,e,f], Jing Jin[a,g,*]

[a] Center of Intelligent Computing, School of Mathematics, East China University of Science and Technology, Shanghai, 200237, China
[b] Brain-Computer Interfacing and Neural Engineering Laboratory, School of Computer Science and Electronic Engineering, University of Essex, Colchester, CO4 3SQ, United Kingdom
[c] School of Computer Science and Engineering, Nanyang Technological University, 639798, Singapore
[d] Systems Research Institute, Polish Academy of Science, Warsaw, 01-447, Poland
[e] RIKEN Advanced Intelligence Project, Tokyo, 103-0027, Japan
[f] Tokyo University of Agriculture and Technology, Tokyo, 184-8588, Japan
[g] Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai, 200237, China

* Corresponding author

Abstract

A Brain-computer interface (BCI) system establishes a novel communication channel between the human brain and a computer. Most event related potential-based BCI applications make use of decoding models, which requires training. This training process is often time-consuming and inconvenient for new users. In recent years, deep learning models, especially participant-independent models, have garnered significant attention in the domain of ERP classification. However, individual differences in EEG signals hamper model generalization, as the ERP component and other aspects of the EEG signal vary across participants, even when they are exposed to the same stimuli. This paper proposes a novel One-source domain transfer learning method based Attention Domain Adversarial Neural Network (OADANN) to mitigate data distribution discrepancies for cross-participant classification tasks. We train and validate our proposed model on both a publicly available OpenBMI dataset and a Self-collected dataset, employing a leave one participant out cross validation scheme. Experimental results demonstrate that the proposed OADANN method achieves the highest and most robust classification performance and exhibits significant improvements when compared to baseline methods (CNN, EEGNet, ShallowNet, DeepCovNet) and domain generalization methods (ERM, Mixup, and Groupdro). These findings underscore the efficacy of our proposed method.

Keywords: Brain-computer interface, P300 detection, cross participant task, domain generalization

1. Introduction

Brain-computer interface (BCI) systems establish a direct interactive pathway between the human brain and external devices through decoding users' neural activity into control commands [1]. BCI systems have been suggested to have considerable potential to improve aspects of the quality of life for people with conditions such as amyotrophic lateral sclerosis (ALS). As a non-invasive and safe modality, electroencephalography (EEG) has been widely used to monitor brain activity within BCI systems [2]. Moreover, it provides an excellent temporal resolution of less than a millisecond.

Event-related potentials (ERPs) are time-locked components of the EEG [3] that represent neural responses to specific stimuli or events. The low signal-to-noise ratio of the ERP means they are typically collected over multiple repetitions of stimulus presentation. Examples of ERPs that are widely used in BCI systems include the N200, P300, and N400 potentials [4]. Of these ERPs, the P300 is the most widely used in BCI systems. This P300 ERP is a positive deflection in the amplitude of the EEG over the parietal and occipital regions of the cortex that occurs approximately 300 milliseconds after the onset of an uncommon stimulus [5, 6]. The P300 ERP is most commonly elicited by the use of the oddball experimental paradigm, in which users are asked to focus on an infrequent stimulus and ignore other common stimuli presented in a sequence. The well-known P300 speller is based on this paradigm and was designed by Farewell and Donchin in 1988 [7]. In this speller paradigm, the BCI user is presented with a selection of options arranged on-screen within a 6×6 matrix containing 36 characters. Each row and column of this matrix is sequentially highlighted in a pseudo-randomised order for a given number of times. Due to inevitable external noise, the collected ERPs have a low signal-to-noise ratio. Consequently, it is crucial to design an effective algorithm to recognize these ERP signal components. In the aspect of feature extraction, structure constrained semi-nonnegative matrix factorization (semi-NMF) was used to extract the key patterns of EEG data in time domain. It has been reported that a human behavior data representation method based on structure constraint and semi-NMF achieved excellent sequential segmentation [8].

In recent years, deep learning methods have shown promising results and are able to automatically extract complex features from raw data, learning hierarchical representations of the input at different levels [9]. In the field of intelligent medical care, an automatic fetal ultrasound standard plane recognition model based on deep learning was extended to the Industrial Internet of Things platform to achieve efficient data analysis [10]. Convolutional neural networks (CNN) have also been utilized in other fields such as computer vision and speech recognition in recent years [11, 12]. To mine features from multiple spatiotemporal frequencies, a multiscale feature fusion octave convolution neural network was proposed for EEG classification [13]. To date, most studies have explored simple model architectures based on CNN and recurrent neural networks (RNN) [14]. For

example, Cecotti et al. [15] developed a 4-layer CNN for use in BCI to decode P300 ERPs for the first time. Liu et al. [16] combined the idea of a one-dimensional convolution with the traditional Caps Net model to construct a 1D-CapsNet model, which achieved superior detection performances compared to traditional machine learning. Borra et al. [17] investigated a Bayesian-optimized interpretable CNN to analyze P300 spectral and spatial features, which demonstrates that a CNN can be designed to be both accurate and interpretable for P300 decoding. Tortora et al. [18] trained a Long-Short Term Memory (LSTM) deep neural network to deal with time-dependent information within brain signals during locomotion.

Despite much research and impressive progress, there still remain some major challenges for BCI systems. For instance, different participants have different neural responses to the same stimulus, and even for the same participant, the distribution of data varies over time resulting in differences in ERPs over sessions and days [19]. In a word, EEG signals have been discovered to display large inter-individual variation. Thus, most P300 BCI system requires a long time of training and offline calibration data, which is time-consuming and inconvenient in real applications. To address the above problem, current work develops cross-participant transfer learning method to detect P300 signals without any training data from the target subject to diminish the influence of participant variability on decoding performance [20]. In a classic study, EEGNet was constructed by using Depthwise and Separable Convolutions, which could produce interpretable features and achieve better decoding performances than other CNN models when applied to cross-participant classification [21]. Subsequently, Inception modules were efficiently integrated into an EEG-Inception method to facilitate the extraction of feature maps at different temporal scales, which reduced the amount of calibration data needed to obtain a good decoding accuracy with new participants [22]. Bhatt et al. created a graph-based dual-attention convolutional recurrent model to enhance the detection of ERP signal, particularly for visual object recognition in cross-participant classification task [23]. To improve BCI performance by using the uncertainty information, a Bayesian convolutional neural network (BCNN) can efficiently estimate prediction uncertainty, which provides more reliable classification results [24]. Under uncertain conditions, the situational assessment scheme based on uncertainty risk awareness proposed by Gao et al. has improved the cognitive ability of intelligent vehicles in the environment [25]. Interacting multiple models for short-term and long-term trajectory prediction was developed to achieve high effectiveness [26].

Domain adaptation (DA) and domain generalization (DG) are two popular branches of transfer learning. Among the DA approaches, adversarial learning-based methods, such as Generative Adversarial Networks (GAN), and Adversarial Discriminative Domain Adapation (ADDA) have shown great potential and achieved significant improvements in EEG decoding performance. For example, Panwar et al. [27] investigated a conditioned Wasserstein GAN with gradient penalty to generate EEG data in a rapid serial visual

presentation (RSVP) task and achieved improved intra-participant cross-session performances over EEGNet. Li et al. [28] proposed a bi-hemisphere domain adversarial neural network (BiDANN) model in which domain discriminators work adversarially with a classifier to learn discriminative emotional features and alleviate the domain differences between source and target domains. Considering adversarial security, alignment based adversarial training integrated data alignment and adversarial training, which can simultaneously reduce their distribution discrepancies and robustifies the classification boundary [29, 30]. However, DA methods require collecting EEG data relating to tasks from new participants in advance and retraining the model. These are time-consuming and resource-intensive processes. In contrast, DG methods aim at robust performances when dealing with unknown domains without the need for extra information [31], which can address the above problems presented by DA methods and is preferable for practical applications. Since DG methods conduct vigilance estimation over multiple new participants with only well-trained models, it is critical to enhance the generalization ability of the DG models. There are several approaches available to improve model generalization, such as data augmentation, adversarial training, and meta-training. However, less attention has been paid to the use of DG for cross-participant EEG based ERP detection.

In this paper, we propose a domain adversarial neural network designed to address the domain generalization problem. The entire pretraining process is conceptualized as a binary classification task, where all data from source participants is amalgamated and treated as a unified source domain for training classifier. Specifically, we leverage a deep neural network to generate effective and non-handcrafted deep representations, employing adversarial learning to achieve cross-participant P300 classification. The main contributions of this paper include:

- Development of a one-source domain transfer learning method, termed one-source domain adversarial neural networks (ODANN), which utilizes domain adversarial neural networks to assimilate common features from source domains, enhancing the separability between target and non-target data.
- Introduction of an attention mechanism with a convolutional neural network (CNN) to recalibrate the weights of different channels based on the deep convolution structure in the feature extractor, effectively addressing channel interactions.
- Validation of the effectiveness of our proposed model through comparative experiments conducted on both a publically available OpenBMI dataset and a Self-collected dataset. The results confirm that our proposed methodology exhibits superior recognition accuracy.

The remainder of this work is arranged as follows. Section 2 details related method and proposed framework. Section 3 introduces the datasets, experimental setting, and baseline methods we use to compare with our method. Section 4 presents the experimental results and visualization. Section 5 includes a detailed discussion and Section 6 concludes

our work.

2. Materials and Methods

2.1 Domain Adversarial Neural Network

The DANN model is the first work to attempt to match the data distribution using an adversarial training strategy [32] and was proposed to deal with the domain adaptation problem. The DANN model consists of three components, a feature extractor G_f, a label classifier G_y, and a domain discriminator G_d [33]. Both classifiers share the feature extractor that extracts domain-invariant feature representations from the source domain and target domain. The label classifier is used for classifying the source domain, and the domain classifier distinguishes whether the signal belongs to the source domain or the target domain. This approach introduces a gradient reversal layer (GRL) [34] between the domain classifier and the feature extractor. The loss function is divided into two parts, label prediction loss, and domain prediction loss. Therefore, the DANN model attempts to minimize label prediction loss and maximize the domain prediction loss. The definition of each loss is as follows. Given the source domain data $X_s = ({x_i}^s, {y_i}^s)_{i=1}^{n_s}$ and the target domain data $X_t = ({x_i}^t, {y_i}^t)_{i=1}^{n_t}$, n_s and n_t are the number of samples in the labeled source domain and the unlabeled target domain, respectively. The label loss operation L_y can be defined by:

$$L_y(\theta_f, \theta_y) = \frac{1}{n_s}\sum_{i=1}^{n_s} {L_y}^i(\theta_f, \theta_y) \tag{1}$$

where θ_f and θ_y represent the parameters of the feature extractor and label classifier. The negative log-probability of the correct label can be used to represent the loss:

$${L_y}^i(\theta_f, \theta_y) = \log\frac{1}{G_y(G_f({x_i}^s), {y_i}^s)} \tag{2}$$

The domain discriminator loss operation L_d is

$$L_d = \frac{1}{n_s}\sum_{i=1}^{n_s} {L_d}^i(\theta_f, \theta_d) + \frac{1}{n_t}\sum_{j=1}^{n_t} {L_d}^j(\theta_f, \theta_d) \tag{3}$$

where θ_d denotes the parameters of the domain discriminator. According to Ganin et.al. [32], the loss of the domain discriminator can be defined as:

$${L_d}^i(\theta_f, \theta_d) = d_i\log\frac{1}{G_d(G_f(x_i))} + (1-d_i)\log\frac{1}{1-G_d(G_f(x_i))} \tag{4}$$

where d_i is a binary domain label for sample x_i. The domain label can be defined as 0 if x_i belongs to source domain, and the domain label can be defined as 1 if x_i belongs to target domain. The overall objective function is:

$$L(\theta_f, \theta_y, \theta_d) = L_y - \lambda L_d \tag{5}$$

The hyper-parameter λ is used to balance the trade-off of these two terms. Moreover,

the optimization for all parameters can be organized as follows:

$$\hat{\theta}_f, \hat{\theta}_y = arg\min_{\theta_f,\theta_y} L(\theta_f, \theta_y, \hat{\theta}_d) \tag{6}$$

$$\hat{\theta}_d = arg\max_{\theta_d} L(\hat{\theta}_f, \hat{\theta}_y, \theta_d) \tag{7}$$

After optimization, the feature extractor should find a mapping to the feature space where task-related information is retained and most of the domain-variant features are excluded.

2.2 Source Domain Transfer Based on ADANN Method

Due to large inter-participant variability, it is challengable to enhance the classification performance of participant-independent model. To address the above problem, we focus on domain generalization and put forward a novel adversarial structure, attention based DANN method (ADANN), to improve the generalization ability of the model. We design two different strategies for dealing with inter-participant variability. With this, the final model can help to extract domain-invariant class features for EEG classification tasks. The architecture of our ADANN model is illustrated in Figure 1. The overall framework can be divided into four branches: data preprocessing, feature extraction, label classification, and domain discrimination.

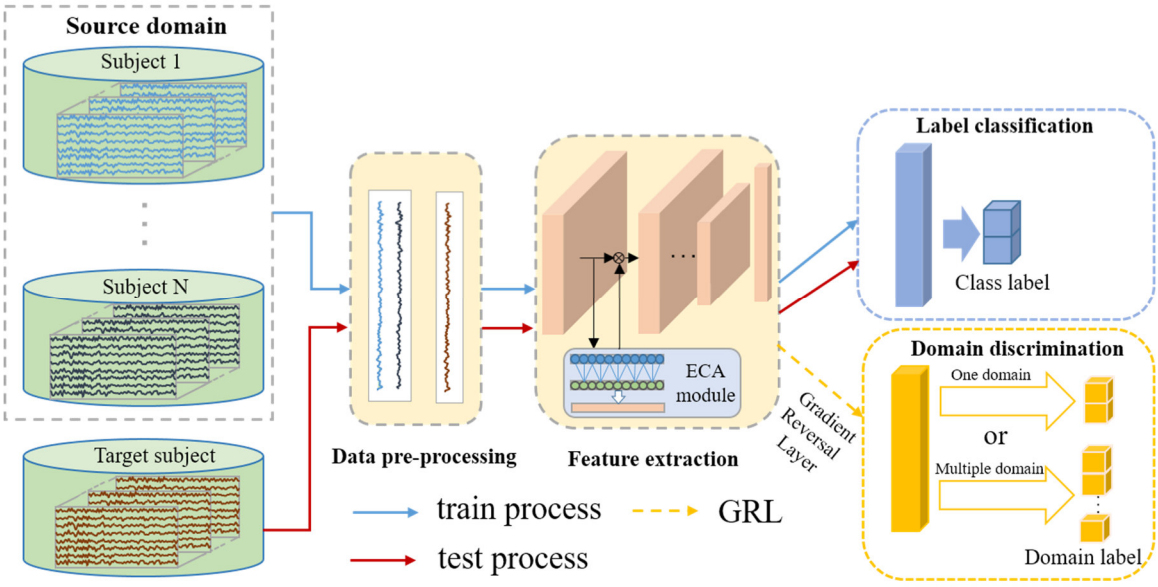

Figure 1. The overall structure of the proposed ADANN method based on two strategies

Given that this work transfers knowledge from the source domain, raw EEG data from multiple participants is first preprocessed and concatenated as the input data in the first step. Data preprocessing part includes data segmentation, filtering and downsampling. Then, the input data is fed into a feature extractor to obtain the domain-invariant feature space. To be specific, we introduce an attention mechanism based deep convolution

structure as a feature extractor $G_f(\theta_f)$ due to its powerful learning capability. A shallow yet efficient classifier $G_y(\theta_y)$ is used for EEG classification. However, the training process can easily cause overfitting to the source distribution. In this domain generalization strategy, we generalize the domain discriminator $G_d(\theta_d)$ as different source domain classifiers reduce the domain shift. At last, two kinds of losses, label classifier loss and domain discriminator loss are exerted during the training process. As shown in Eq.(3), the loss of the domain discriminator in the domain adaptation problem is related to the source and target data. In order to transfer it into a domain generalization problem we have generalized and re-designed the domain discriminator as follows:

$$L'_d = \frac{1}{N}\sum_{i=1}^{N} L'_d{}^{i}(\theta_f, \theta'_d) \tag{8}$$

$$L'_d{}^{i}(\theta_f, \theta'_d) = \log\frac{1}{G_d(G_f(x_i))_{d_i}} \tag{9}$$

where $d_i \in \mathbb{R}^k$, k is the number of source domains and N is the sample of labeled source domains. Accordingly, the overall loss function of the ADANN method can be represented as:

$$L(\theta_f, \theta_y, \theta'_d) = \frac{1}{N}\sum_{i=1}^{N} L'_y{}^{i}(\theta_f, \theta_y) - \lambda\frac{1}{N}\sum_{i=1}^{N} L'_d{}^{i}(\theta_f, \theta'_d) \tag{10}$$

where L'_y is the cross entropy function, which is adopted to minimize the difference between the predicted label and the corresponding ground truth label.

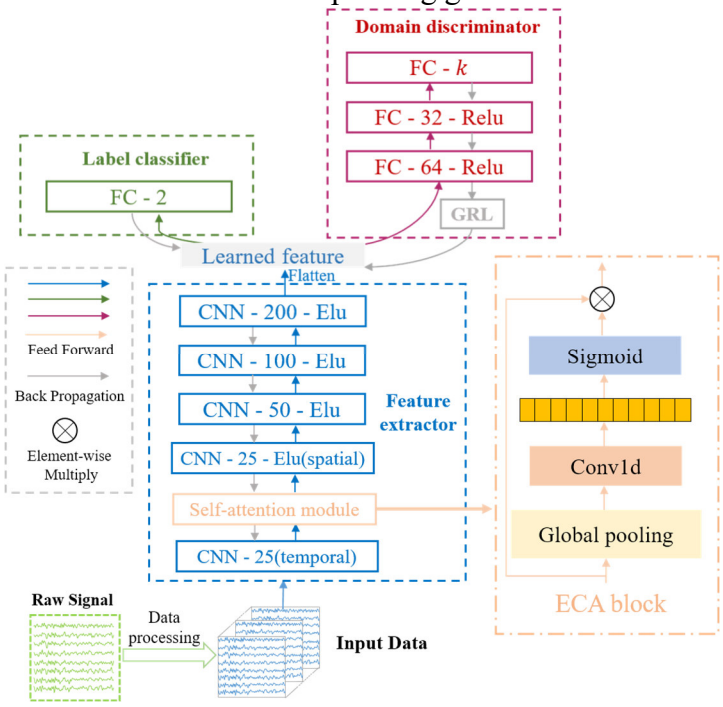

Figure 2. The learning process of our ADANN method.

Depending on the use of the domain discriminator G_d, two strategies can be used to apply our proposed method. Specifically, we can either apply our method using a one-

source domain transfer learning based ADNN method (OADANN) or using a multiple source domain transfer learning based ADNN method (MADANN). These two approaches are shown in Figure 1. In the OADANN method, data from all the source participants is regarded as belonging to one source domain, thus the source domain labels can be considered as the same one. Alternatively, in the MADANN method, data from all the source participants may be classified into multiple domains, with the total number of domains determined by the total number of participants. Therefore, the source domain labels can be considered multi-class labels. The detailed configuration of each block is presented in Figure 2.

In our work, we employ an efficient channel attention (ECA) [35] based deep convolutional neural network as a feature extractor. The deep convolutional neural network we use in this work has five blocks. The first two of these blocks executes a temporal convolution and a spatial convolution, the following performs standard convolutional-max-pooling. During feature extraction, the ECA block is followed by temporal convolution, which can avoid dimensionality reduction and capture cross-channel interactionss [35]. Figure 2 also illustrates the overview of the ECA block. First, global average pooling (GAP) is applied without dimensionality reduction [36]. Then we determine that the kernel size is 3 and perform 1D convolution followed by a Sigmoid function to learn channel attention. Furthermore, a simple label classifier with one fully connected layer is applied to perform ERP classification. A domain discriminator with three fully connected layers is then employed to distinguish the domain. Note that k is set to 1 in the OADANN method while k is set to the number of participants of source domain in the MADANN method. In order to make the proposed algorithm clear, we describe the pseudo-code of the OADANN method in Algorithm 1.

Algorithm 1 The OADANN method

Input: Training process: The source domain data $X_s = (x_i{}^s, y_i{}^s)$, and the corresponding domain label $y'_i{}^s$; // Test process: The target participant data $x_i{}^t$

Output: The corresponding predicted label $\hat{y}_i{}^t$.

1: Initilize the parameters $\theta_f, \theta_y, \theta_d$.

2: Preprocessing raw data and obtain processed data $x'_i{}^s$.

2: **for** $epoch$ in range ($maxepoch$) **do**

3: **foreach** minibatch **do**

4: Train the attention based domain adversarial neural network $G = \{G_f, G_y, G_d\}$ by processed source domain data $X_s = \{(x'_1{}^s, y_1{}^s), \dots, (x'_N{}^s, y_N{}^s)$, $X_d = \{(x'_1{}^s, y'_1{}^s), \dots, (x'_N{}^s, y'_N{}^s)$ where $x'_N{}^s$, $y_N{}^s$ and $y'_N{}^s$ are the data, class label and domain label of m-th participant, respectively; G is the model trained on all the source domain data. $f_s = G_f(x'_i{}^s)$; $y_s = G_y(f_s, y_i{}^s)$; $d_s = G_d(f_s, y'_i{}^s)$

5: maximize the loss function L'_d of the domain discriminator.
6: minimize the loss function L_y of the feature classifier.
7: end
8: end
9: Predict the label $\hat{y}_i{}^t$ of processed target participant data $x'_i{}^t$, where $\hat{y}_i{}^t = G_y\left(G_f\left(x'_i{}^t\right), y_i{}^s\right)$

3. Experiments

3.1 Datasets

We compare our proposed method to other state of the art methods on two ERP datasets: 1) a public OpenBMI dataset; 2) a Self-collected dataset.

OpenBMI dataset containing EEG and EMG data was collected via an ERP-BCI paradigm from 54 healthy participants (S01-S54). The dataset contains EEG signals recorded via 62 channels (the hollow circles and Cz in Figure 4) and the reference and ground electrodes are positioned on the nasion (Nz) and at position AFz. There are two sessions recorded on different days, and each session contains both offline and online phases. In this study, we only use one offline ERP phase for further analysis. Before the ERP experiment began, participants were instructed to sit comfortably in a chair with armrests positioned at a distance of approximately 60 (±5) cm in front of a 21-inch LCD monitor. The approximate horizontal and vertical visual angles are 37.7 and 28.1 degrees respectively. During the process of the experiment, participants were instructed to relax and minimize their eye and muscle movements.

The interface layout of the paradigm is shown in Figure 3. A grid of letters and numbers containing 6 rows and 6 columns and including 36 symbols (A-Z, 1-9, and the underscore character '_') was displayed on the screen. To evoke stronger ERP responses, random-set presentation and face stimuli are used in this paradigm. The stimulus-time interval is set to 80ms, and the inter-stimulus interval (ISI) to 135ms. A single iteration of stimulus presentation in all rows and columns is considered a sequence. Therefore, one sequence consists of 12 stimulus flashes. In the offline phase of our experiment, each target character is presented over five rounds, that is, there are 60 flashes in total (12 stimulus flashes by 5 repetitions ('rounds') per stimulus). In addition, a given sentence composed of 33 characters, "NEURAL_NETWORKS_AND_DEEP_LEARNING" is spelled by the participants by fixing on the target character on the screen. More details about the dataset can be found on the following website: http://deepbci.korea.ac.kr/wp-content/uploads/2020/11/Big-Data-of-ERP-Speller.pdf.

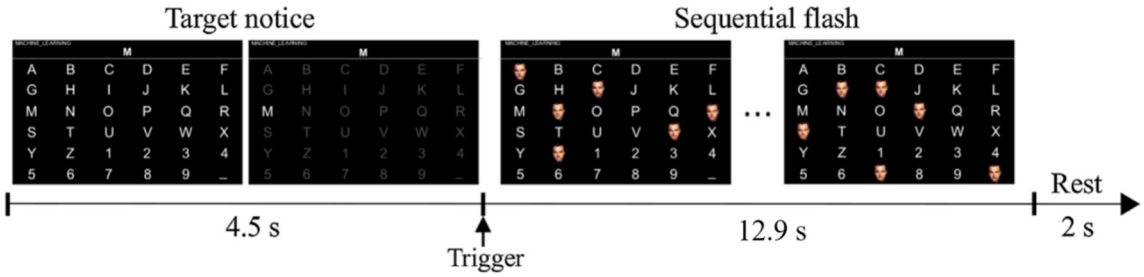

Figure 3. The flow chart of the ERP paradigm used in the OpenBMI dataset.

In our Self-collected dataset, EEG signals were collected from 15 participants (P01-P15) using 59 electrodes (all orange circles in Figure 4). For each participant, the experiment only includes a single offline block, which includes 36 targets (A-Z, 1-9, and the underscore character '_'). The graphical interface used in the paradigm is the same as shown in Figure 3. Each target is arranged to be presented over 5 repetitions (rounds) and each round consists of a sequence of 12 stimuli flashes. The stimulus presentation pattern is based on binomial coefficients [2]. The stimulus onset asynchrony (SOA) was set to 150 ms, and the stimulus interval was set to 75 ms throughout all stages of the experiment.

3.2 Data Preprocessing

In the preprocessing stage, the selected EEG signals are first band-pass filtered from 0.5Hz to 40Hz via a fourth-order Butterworth filter. Then, the temporal features from -200ms to 800ms from the stimulus presentation onset time from each channel are extracted. We down-sampled the EEG signals to 100 Hz and then baseline-corrected the signals by subtracting the mean amplitudes from the -200 ms to 0 ms pre-stimulus interval.

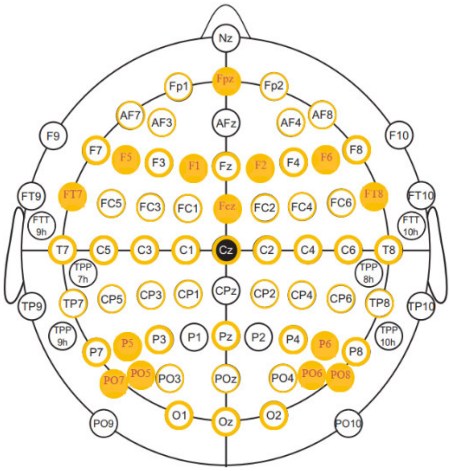

Figure 4. The channel configuration of the International 10-20 system for EEG electrode placement.

In our Self-collected dataset, a band pass filter is applied to filter the EEG between 0.5 and 35 Hz to reduce high frequency noise. The filtering algorithm we applied is a third-order Butterworth filter. In order to decrease the dimensionality of the data and complexity of the classification model, the sampling rate is downsampled to 250 Hz. For both datasets, EEG data from each trial is extracted using the same time window [0, 800ms]

after stimulus presentation.

3.3 Experimental Evaluation

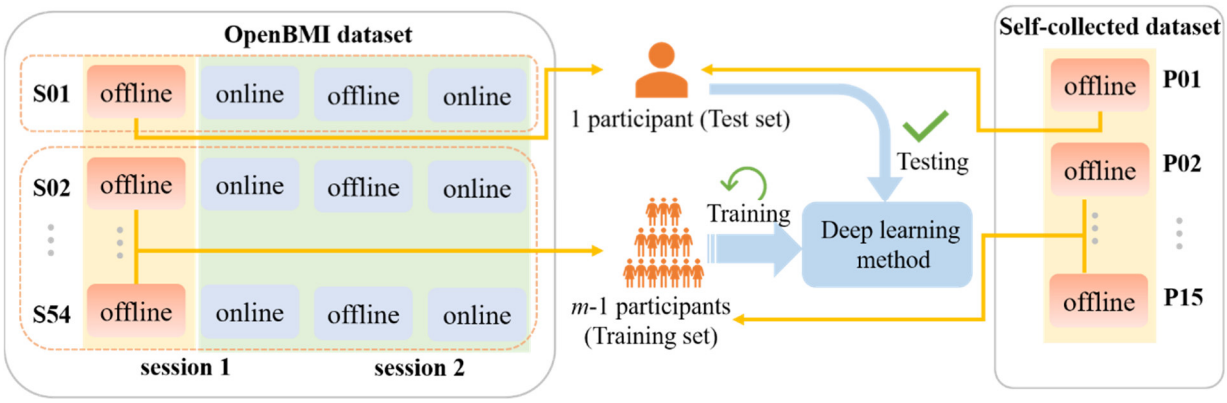

Figure 5. Illustration for a LOPO-CV scheme for training the classification models

To demonstrate the validity and generalizability of our proposed method in a P300 classification model, we execute our experiment in a participant-independent manner. Figure 5 presents an example of how we selected the training samples and testing samples for each of our two datasets. For the OpenBMI dataset, the offline data recorded in session 1 (orange box) is used in this study. Leave one participant out cross validation (LOPO-CV) scheme adopted in this study is one of cross-participant transfer learning tasks. For instance, suppose that S1 is the test participant, the offline data of the remaining participants should be used to train the classifier. Note that S52 includes error labels, which has no corresponding character during the experiment. Therefore, we only retain 53 participants as test participants. In accordance with widely used standardized metrics for assessing BCI performances, the character recognition accuracy and information transfer rate (ITR) are evaluated in a public dataset [37]. Character recognition accuracy can be computed as follows:

$$Acc = \frac{N_C}{N_{TOTAL}} \tag{11}$$

where N_C denotes the number of correctly predicted characters and N_{TOTAL} is the total number of characters. ITR is defined as:

$$ITR = (log_2 N + Acc log_2 Acc + (1 - Acc) log_2 \frac{1 - Acc}{N - 1}) \frac{60}{T} \tag{12}$$

where N represents the total number of classes, Acc denotes the classification accuracy, and T represents the time taken by the participant to perform each trial.

3.4 Experimental Setting

There are two transfer evaluation methods in our experiment. One participant is selected as the target and the rest as source domain in both methods, that is LOPO-CV scheme. Source data may either be regarded as one domain (S→O, OADANN) or source data may be divided into multiple domains according to the number of participants (S→M,

MADANN):
(1) (S→O): S→O is developed to evaluate the performance of the proposed OADANN method. In this regard, each participant is taken as the target participant, and the rest of the participants are taken as the source participants (one domain). If t represents the number of participants in the dataset than there are data from $t-1$ participants in one source domain.
(2) (S→M): S→M is designed to evaluate the efficacy of our proposed OADANN method in the case of multiple domains. We select one participant as the target and the rest as source participants (multiple domains). Let m represent the number of participants in the dataset, there are data from $t-1$ participants in $t-1$ source domains. In contrast, our proposed OADANN method analyzes the relationship between one source domain and the target domain, which is simplified as a binary classification problem for domain discriminator.

The differences between methods are that the source participants is regarded as one domain in the domain discriminator of OADANN and as multiple domains in the domain discriminator of MADANN, therefore there are different classification problems in two strategies.

3.5 Baseline Methods

In order to demonstrate the effectiveness of our proposed method, four traditional deep learning models are adopted as baseline methods to classify the P300 signal: CNN, EEGNet, ShallowNet, and DeepConvNet. CNN has been employed in computer vision and speech recognition and achieved great performance in many situations. Inspired by this, CNN has been adopted for use in the detection of P300 waves in the time domain. Subsequently, EEGNet was introduced for within and cross-participant classification tasks. EEGNet has been demonstrated to be robust enough to learn interpretable features over various BCI tasks. ShallowNet and Deep ConvNet have also been proposed to recognize imagined or executed tasks from raw EEG signals. Each of these methods are described below.

1) CNN (BASIC-CNN) This is the first model based on a convolutional neural network that has been deployed to detect P300 ERPs. This network consists of five layers, and each layer is composed of one or more maps. The first hidden layer is a channel combination layer, while the second hidden layer down-samples and transforms the signal in the time domain, the third hidden layer contains one map of 100 neurons, which is fully connected to the second layer. Finally, the output layer has only one map of 2 neurons denoting two classes (target and non-target), which is fully connected to the third layer. A detailed description of this CNN is provided by Cecotti and Graser [15].

2) EEGNet This is a compact CNN for classification and interpretation of EEG within BCI systems. There are three main blocks in this network. First, a temporal convolution is used to capture frequency information within the EEG. Second, a depthwise convolution is adopted to learn frequency-specific spatial filters. Then separable

convolution is carried out to learn temporal information for each feature map individually, followed by pointwise convolution. In the classification stage of the model the feature is flattened and sent into a fully connected layer with a softmax function for classification. This method has been evaluated across different BCI paradigms and the results demonstrate the effectiveness and generalizability of this model. More details about EEGNet are reported by Lawhern et al. [21].

3) ShallowNet It is inspired by the Filter Bank Common Spatial Patterns (FBCSP) pipeline, which is designed to decode band power features from the EEG. Specifically, the first two layers of this model consist of temporal and spatial convolution layers followed by a mean pooling layer. Since there are several pooling regions in one trial, shallow ConvNet can capture the temporal information of band power changes in one trial, which is helpful for classification. Detailed descriptions are reported by Schirrmeister et al. [38].

4) DeepConvNet It is inspired by the successful architecture DeepConvNet, first applied in the fielf of computer vision. It is implemented by four convolution-max-pooling blocks, which include a specific first block to address the input feature, followed by three standard convolution-max-pooling blocks and a dense softmax classification layer. The first convolution block is composed of two layers. The first layer performs a temporal convolution, while the second layer performs a spatial convolution. There is no activation function in the first block and we use exponential linear units (ELUs) as activation functions in the rest of the blocks. More details about Deep ConvNet are reported by Schirrmeister et al. [38].

3.6 Model Training

The whole workflow is implemented in the Pytorch 1.9.0 library and the whole experiment is run on an Intel(R) Xeon(R) platform with NVIDIA GeForce RTX 2080Ti GPU. In the training process, the loss function was optimized via an adaptive moment estimation (ADAM) optimizer. The learning rate of η was set to 0.0005, the weighted decay was set to 0.001, and a 25% dropout rate was used in the training process [39]. We set a batch size of 64 samples for participant-independent classification. Finally, the number of training iterations was set to 100.

4. Results

Our proposed method is a competitive model for effective feature extraction with an attention module. In this section, we compare the ERP identification performances achieved by our proposed model and four deep learning methods. We also construct the domain adversarial neural network framework. We then carry out comparison experiments with several DG algorithms. Moreover, we also perform an ablation study to compare the

classification performance among different methods under the same experimental settings to highlight the contributions of each block of our proposed method on P300 classification performance. Finally, we visualize the feature distributions associated with two sample participants using the t-SNE embedding method.

4.1 Deep Learning Methods

Tables I and II present the average classification performances in the LOPO-CV scheme applied to the two datasets. The experiments illustrate that our proposed OADANN method obtains the highest classification accuracy across the five rounds of stimuli flashes within our experiments. In addition, as the number of rounds of stimuli flashes increases, the classification performances also increases. In this regard, our proposed OADANN method outperforms BASIC-CNN, EEGNet, ShallowNet, DeepCovNet, and MADANN methods by an average of 5.72%, 3.6%, 4.63%, 0.97%, and 1.14% after five rounds of stimuli flashes in the OpenBMI dataset. We also compare the ITR achieved by our OADANN method with the other methods. This reveals an improvement of up to 2.66 bits/min for BASIC-CNN, 1.39 bits/min for EEGNet, 2.72 bits/min for ShallowCovNet, and 0.37 bits/min for DeepCovNet after two rounds, respectively. For Table II it may be seen that the accuracy of our method is 2.78% and 3.15% higher than that of EEGNet and DeepCovNet after 5 rounds of stimuli flashes.

Table I Average classification performances (accuracy±standard deviation (%) and ITR±standard deviation (bits/min)) in the LOPO-CV scheme with OpenBMI dataset.

		Methods					
Rounds		BASIC-CNN	EEGNet	ShallowNet	DeepCovNet	MADANN	OADANN
1	ACC	56.43±20.17***	61.41±21.14***	58.83±22.13***	65.41±19.57	61.29±20.58	**65.64±18.25**
	ITR	17.57±8.86***	20.10±9.77**	18.94±10.10***	**21.94±9.74**	19.95±9.73	21.91±9.01
2	ACC	80.33±21.47***	83.08±20.96*	79.93±22.51***	85.88±17.04	84.39±18.07	**86.62±16.58**
	ITR	22.66±8.42***	23.93±8.33*	22.60±8.87***	24.95±7.28	24.33±7.79	**25.32±7.29**
3	ACC	88.56±19.13**	90.91±18.86*	87.82±19.48***	93.42±11.84	91.54±13.21	**94.05±10.17**
	ITR	20.90±6.06***	21.76±5.85*	20.70±6.36***	22.54±4.29	21.82±4.85	**22.71±3.79**
4	ACC	90.74±18.18**	92.51±17.69*	91.37±17.86**	95.48±10.27	94.74±9.98	**96.17±8.24**
	ITR	18.02±4.94***	18.63±4.55*	18.25±4.86**	19.33±3.08	19.05±3.18	**19.51±2.64**
5	ACC	91.82±18.05**	93.94±16.50*	92.91±16.11**	96.57±8.41**	96.40±8.00**	**97.54±6.80**
	ITR	15.69±4.17**	16.26±3.72*	15.92±3.79**	16.77±2.26**	16.69±2.19**	**17.04±1.88**

ACC defines classification accuracy, ITR defines information transfer rate. The outcomes of the significance tests are reported in terms of p values between each of the methods and OADANN: * indicates ($p < 0.05$), ** indicates ($p < 0.01$), *** indicates ($p < 0.001$). Bold highlighting denotes the best numerical values.

Table II Average classification performances (accuracy±standard deviation) (%) and ITR±standard deviation (bits/min)) in the LOPO-CV scheme with Self-collected dataset.

		Methods					
Rounds		BASIC-CNN	EEGNet	ShallowNet	DeepCovNet	MADANN	OADANN
1	ACC	47.04±9.12	49.44±12.43	49.63±15.03	49.07±13.06	**50.18±11.48**	48.89±10.17
	ITR	92.62±26.36	**104.31±27.84**	97.77±30.31	94.46±33.49	98.47±29.35	97.39±34.21
2	ACC	64.81±11.49	**69.81±11.15**	66.85±13.13	65.19±14.81	67.22±12.52	66.48±14.03
	ITR	63.06±17.95	**71.02±18.96**	66.57±20.64	64.32±22.80	67.05±19.98	66.31±23.29
3	ACC	72.59±11.73	**77.59±10.20**	76.67±10.22	77.04±12.87	75.74±11.18	77.04±11.03
	ITR	57.39±14.85	**63.82±14.03**	62.56±14.01	63.62±17.28	61.51±15.39	63.26±15.52
4	ACC	78.70±9.19*	83.33±8.33	83.70±8.65	81.30±11.13	81.67±8.96*	**84.26±9.37**

5	ITR	52.43±9.91*	57.66±9.65	58.19±10.34	55.68±12.53	55.80±10.35*	**58.99±11.42**
	ACC	84.26±9.19*	86.48±7.99*	86.67±7.29	86.11±10.39	86.85±9.71*	**89.26±7.78**
	ITR	49.28±8.97*	51.42±8.08*	51.64±7.91	51.37±10.19	52.11±10.01	**54.37±8.08**

ACC denotes classification accuracy, and ITR denotes information transfer rate. The p value indicates the corresponding results between our proposed method and OADANN: * indicating ($p < 0.05$). Bold denotes the best numerical values.

We observe that the MADANN method achieves higher accuracies and ITRs than the BASIC-CNN, EEGNet, and ShallowConvNet methods in both datasets, while achieving slightly lower accuracies and ITRs than the DeepCovNet method when applied to the OpenBMI dataset. The MADANN method generalizes the domain discriminator as an t-1 class domain classifier, resulting in poor generalization. The better recognition performance of the OADANN method is likely due to the fact that the participant discriminator within the OADANN method can effectively reduce the participants' identity information and address the participant-independent P300 recognition problem. To further describe the the error results of each participant in detail, we also present the character classification accuracies achieved by the six algorithms across participants in the OpenBMI dataset, when participants are ordered based on the results achieved via the BASIC-CNN method from smallest to largest. From Figure 6, we can easily find that most participants can obtain zero error when using the OADANN method on the OpenBMI dataset. In essence, the variability among participants (inter-participant variability) can be expressed as differences in the amplitude and latency of ERP signal. Therefore, a high level of inter-participant variability can lead to differences in classification performances and data distribution, which brings great challenges in cross-participants classification tasks. We can take the OpenBMI dataset as an example. As shown in Figure 6, classification performances vary between participants when using the same classification model. However, the proposed method shows stable classification performances compared with baseline methods, that is, higher average classification accuracy and fewer outliers are presented in the OADANN method. It should be noted that the standard deviations of the accuracies and ITRs are high for all models due to the high levels of inter-participant variability. Nevertheless, our proposed method achieves the lowest standard deviation of the methods, demonstrating its' greater robustness to individual differences.

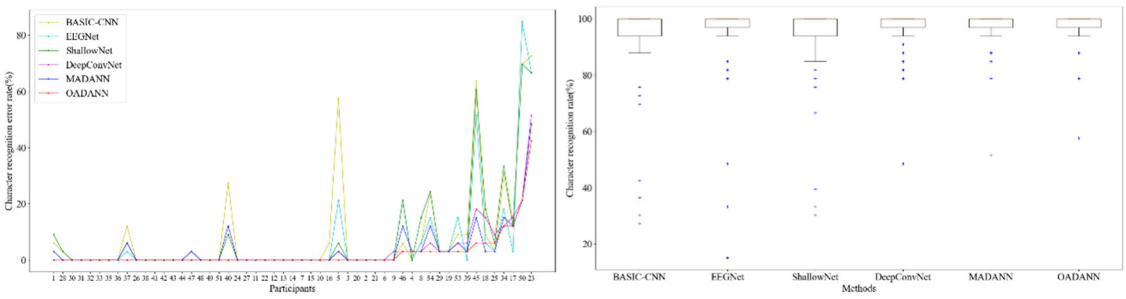

Figure 6. The character classification error rates obtained when combining ERPs from five

rounds of repeated stimuli flashes across 53 participants (left); The box plots of character classification accuracies for the six methods and blue colored plus signs denote outliers (right).

To evaluate the statistical significance of the performance differences between methods we perform a paired *t*-test (significant level of 0.05) with Bonferroni correction for multiple comparisons applied to avoid type I errors. The results in Table I indicate that our proposed OADANN method obtains statistically significant improvements when compared with the other deep learning algorithms in terms of both classification accuracy ($p < 0.05$) and ITR ($p < 0.05$) after five rounds of stimuli flashes. Concretely, our OADANN method has significantly higher performances than BASIC-CNN, EEGNet, and ShallowNet across all the rounds of stimuli flashes, which demonstrates the effectiveness of our proposed method. When considering our Self-collected dataset, significant performance differences ($p < 0.05$) are seen between both our method and BASIC-CNN and EEGNet after 5 rounds of stimuli flashes. Noted that the term 'inter-participant' refers to interactions, relationships occurring between participants within the experiment, typically used with 'variability', 'correlation', and 'transfer learning', such as inter-participant variability. The term 'participant-independent' typically refers to elements of the experiment that are unaffected by the individual characteristics, actions, or behaviors of the participants involved, typically used with 'model','task' and 'strategy', such as participant-independent model, participant-independent strategy. Both terms are used in the cross-participant transfer learning task, that is, the training and test data come from different participants.

4.2 Domain Generalization Algorithms

To evaluate the effectiveness of our OADANN method, we first introduce some data manipulation approaches. All methods are then evaluated via the same training strategy. Data generation based DG methods have been widely utilized to boost the generalization ability of models by generating supplementary data. In this regard, inter-domain Mixup (Mixup) [40] is included for comparison, which can augment the dataset by generating virtual feature-target vectors from real feature-target vectors. Specifically, it performs linear interpolations between any two samples and their labels with a weight sampled from a Beta distribution. In addition, a general learning strategy, referred to as group distributionally robust optimization (Groupdro) [41], aims to optimize feature extraction to be robust by minimizing the worst-case loss over the groups in the training data and then generalizing to the target domain. In addition to those methods, attention based DeepCovNet is also used as a baseline algorithm, referred to as empirical risk minimization (ERM).

Table III summarizes the results of these DG approaches when used in our LOPO-

CV scheme. It can be seen that the classification performance of our OADANN method still outperforms other DG methods, and when ERM, Groupdro, and Mixup are used, the average accuracy after five rounds dropped by 0.68%, 0.80%, and 1.6% on the OpenBMI dataset, respectively. At the same time, our proposed method achieves the highest accuracy with a difference of up to 12.12% compared to the ERM model in terms of classification accuracy for all participants. Statistically speaking, our OADANN method obtains significant improvements ($p < 0.05$) in classification accuracy and ITR when including data from all five rounds of repeated stimuli flashes. When considering our self-collected dataset, a significant difference ($p < 0.001$) can be seen between our proposed method and the DG methods in terms of ITR. On the whole, our OADANN method obtains higher performances than the ERM, Mixup, and Groupdro methods in terms of ACC and ITR with a low standard deviation of results over participants for both of the datasets, which indicates that our proposed DG models are more stable than the other models we compare them to.

Table III Average classification performances (accuracy±standard deviation (%) and ITR±standard deviation (bits/min)) obtained by the DG algorithms as well as our proposed algorithm in the LOPO-CV scheme after 5 rounds of repeated stimuli flashes.

Methods	OpenBMI		Self-collected	
	ACC	ITR	ACC	ITR
ERM	96.86±7.95*	16.85±2.16*	85.92±8.94**	51.02±9.09***
Groupdro	96.74±8.55*	16.83±2.28*	87.04±8.64	52.15±8.98***
Mixup	95.94±8.73*	16.57±2.41*	85.56±9.97	50.85±10.34***
OADANN	**97.54±6.80**	**17.04±1.88**	**89.26±7.78**	**54.37±8.08**

ACC denotes classification accuracy, ITR denotes information transfer rate. The p value indicates the corresponding result between each of the methods and our OADANN method: * indicating ($p < 0.05$), ** indicating ($p < 0.01$), *** indicating ($p < 0.001$). Bold denotes the best numerical values.

4.3 Ablation Study

We conduct ablation experiments to evaluate the effectiveness of our proposed model. We iteratively remove the domain adversarial attention modules from our OADANN method, and then apply the remaining modules to attempt to detect the ERP. Our proposed method, aggregating the attention mechanism, DeepCovNet, and DANN strategy, contributes to the best performances with an average accuracy of 97.54% and an average ITR of 25.32 bits/min after 2 rounds on the OpenBMI dataset. The three models we used for comparison in the ablation study are as follows:

- w/o ADANN: DeepCovnet used for the feature extraction block and the classification block.
- w/o DANN: Attention based DeepCovnet is used as the feature extraction block and the classification block.
- w/o ECA: DeepCovnet is fused into a DANN strategy.

Table IV indicates that our OADANN model is superior to the three kinds of ablation study. When one of the modules is ablated, the classification performance decreases slightly. Concretely, the introduction of domain adversarial mechanisms makes it possible for our OADANN model to capture more information, which demonstrates the effectiveness of our proposed model. According to our experimental results, the DANN mechanism can significantly enhance the performance of our model. Moreover, the performance of our model increases with the help of the attention module, which proves the feasibility of using our OADANN method for extracting participant-invariant P300 features. Considering that ECA module can capture cross-channel interaction without dimensionality reduction, it is observed from Table IV that OADANN method achieves higher performances in terms of character recognition accuracy and ITR compared with w/o ECA in both datasets. Besides, OADANN method can obtain statistically significant improvements when compared with w/o ADANN. It can be easily concluded that domain adverisal learning framework with attention mechanism can bring great classification performances.

Table IV Classification performances (accuracy±standard deviation (%) and ITR±standard deviation (bits/min)) in the LOPO-CV scheme after 5 rounds of repeated stimuli flashes.

Methods	OpenBMI		Self-collected	
	ACC	ITR	ACC	ITR
w/o ADANN	96.57±8.41**	16.77±2.26**	86.85±10.42	52.08±10.07***
w/o DANN	96.86±7.95*	16.85±2.16*	85.92±8.94**	51.02±9.09***
w/o ECA	97.31±7.39	16.98±1.99	85.56±9.97*	50.79±10.16***
OADANN	**97.54±6.80**	**17.04±1.88**	**89.26±7.78**	**54.37±8.08**

ACC denotes classification accuracy, ITR denotes information transfer rate. The result of significance testing is reported in terms of p values between each of the methods and our OADANN method: * indicating ($p < 0.05$), ** indicating ($p < 0.01$), *** indicating ($p < 0.001$). Bold highlighting denotes the best numerical values.

4.4 Visualization

In order to compare the capability of each of the methods to extract highly discriminative features from EEG signals, the t-distributed stochastic neighbor embedding (t-SNE) method [42] is adopted to project high-dimension data into a two-dimensional scatter plot. The corresponding evaluation criterion is that the more separable the classes, the better the related features perform. We visualize the feature embeddings for both our proposed method and the other deep learning methods in Figures 7 and 8. The red color denotes the target samples and the blue color denotes the nontarget samples, the visualization of feature distributions for example participants S45 and S54 are shown in Figures 7 and 8 respectively. These two participants were picked due to the interesting performance differences across the different methods. It can be seen that, compared with other deep learning methods, the separability between P300 and non-P300 samples becomes much easier when using our proposed OADANN method. In addition, compared with the MADANN method, the OADANN method obtains better separability in most cases. The better separability of the OADANN method compared to the MADANN

method can be mainly attributed to the discriminator used in the OADANN method, which can reduce the influences of inter-participant variability and hence improve the discriminative ability of the features.

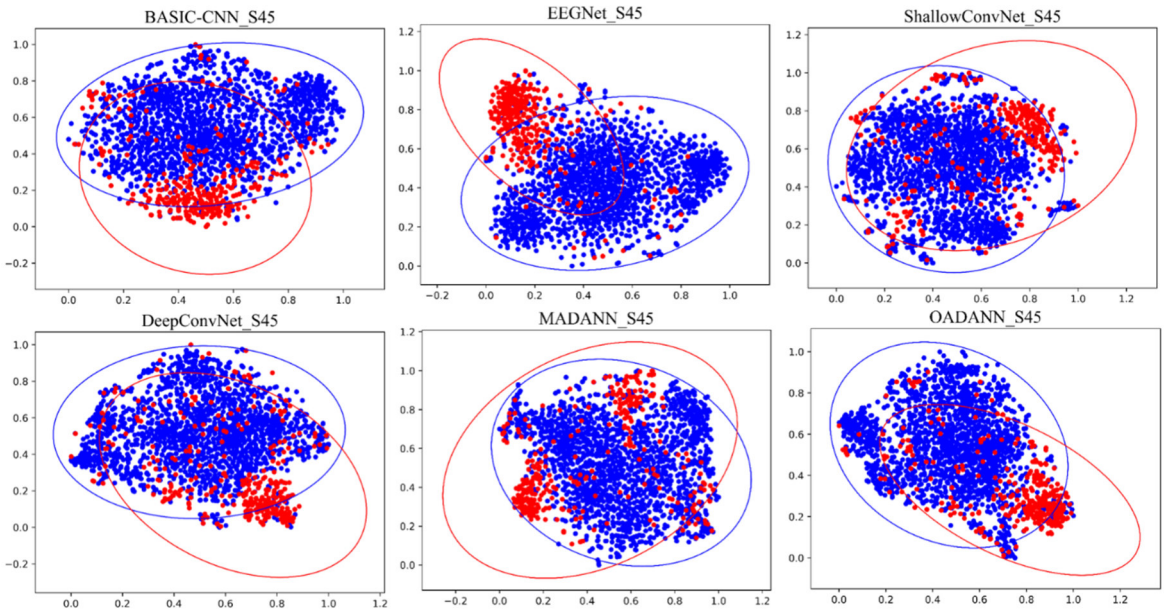

Figure 7. The t-SNE visualization of feature distributions between the target (red) and non-target samples (blue) for participant S45.

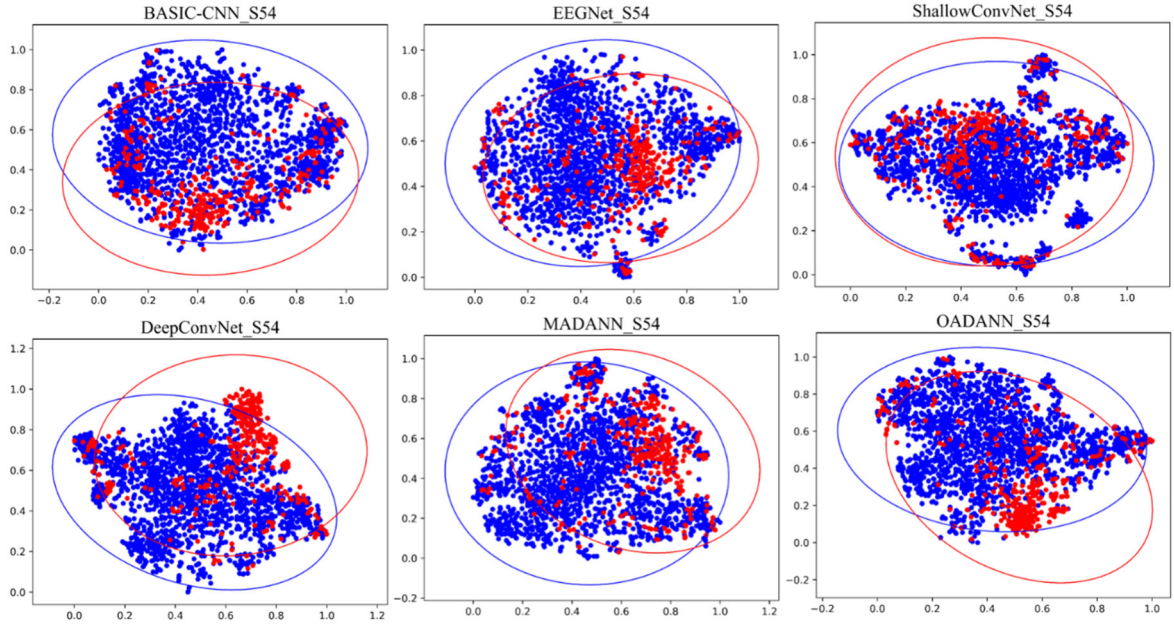

Figure 8. The t-SNE visualization of feature distributions between the target (red) and non-target samples (blue) for participant S54.

5. Discussion

5.1 Effect of the Number of Training Participants

The Self-collected dataset only includes 15 participants, this results in a lower

performances than when compared with the OpenBMI dataset. Therefore, we investigate the influences of training data on the performance of the OADANN method with the OpenBMI dataset. Three training cases have been analyzed with different numbers of participants in the training set ranging from 20 participants' training data, 30 participants' training data, and 40 participants' training data, respectively. Figure 9 shows the classification recognition accuracy and ITR obtained by our OADANN method with these different number of participants included in the training set. More specifically, Table V shows the average classification performances with different numbers of participants in the training set. We can see that the performance of our OADANN method declines slightly as the size of the training set declines. The average classification accuracy is 95.21% when using 40 participants in the training set, which increases by 2.23% when using 53 participants in the training set. Meanwhile, when increasing the number of rounds of stimuli flashes used in the training set, classification performance among the four cases increases accordingly.

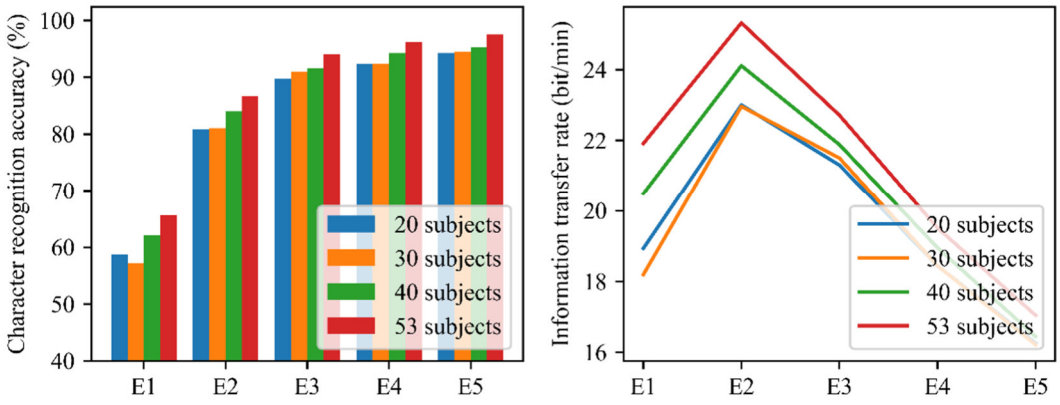

Figure 9. The classification performances obtained by our proposed algorithm with different numbers of participants in the training set as the number of rounds of repeated stimuli flashes is increased from 1 (E1) to 5 (E5).

Table V. Average classification performances (accuracy±standard deviation (%) and ITR±standard deviation (bits/min)) obtained by our proposed algorithm with different numbers of participants in the LOPO-CV scheme with increasing training set and numbers of rounds of repeated stimuli flashes (from 1 round to all 5 rounds).

Number of participants		Rounds				
		1	2	3	4	5
20 participants	ACC	58.72±22.64	80.90±21.83	89.82±16.12	92.40±15.04	94.28±14.23
	ITR	18.93±10.19	23.00±8.86	21.30±5.67	18.44±4.30	16.25±3.37
30 participants	ACC	57.18±22.71	81.07±20.47	90.94±15.71	92.45±14.82	94.45±11.99
	ITR	18.19±10.04	22.96±8.57	21.50±5.58	18.44±4.25	16.21±3.06
40 participants	ACC	62.09±21.83	83.99±18.04	91.60±14.15	94.28±11.56	95.31±10.66
	ITR	20.49±10.22	24.11±7.69	21.88±4.92	18.95±3.51	16.44±2.76
53 participants	ACC	65.64±18.25	86.62±16.58	94.05±10.17	96.17±8.24	97.54±6.80
	ITR	21.91±9.01	25.32±7.29	22.71±3.79	19.51±2.64	17.04±1.88

5.2 Analysis of the Comparison Results

Basic-CNN provides a new way of analyzing brain activities due to the receptive field of the CNN models. It can be seen that CNN obtains the worst classification performances among baseline methods. EEGNet with a compact structure required fewer training weights/parameters. To be concrete, the parameter size of DeepCovNet is two orders of magnitude larger than EEGNet, which has been demonstrated in our previous work. ShallowNet architecture was designed specifically for oscillatory signal classification (by extracting features related to log band-power). In addition, ShallowNet and DeepCovNet were proposed by Schirrmeister et al. in the same work. When using ShallowNet, only 6 of 53 participants and 4 of 15 participants presented higher performances compared with DeepCovNet in two datasets.

We first apply the attention based DeepCovNet mechanism as a feature extractor to learn common features from the source domains via a domain adversarial strategy. and put forward an attention mechanism based deep adversarial neural network framework in which all source participants are regarded as one domain, referred to as OADANN. In addition to this, we use two loss functions to improve the separability between target and nontarget data and decrease the separability of the domains. In our framework, the feature extractor can collectively leverage both the spatial and temporal information from the EEG. To implement a generalized evaluation of our proposed method, a leave one participant out cross validation training strategy is applied to two datasets. From our results we can see that our OADANN method achieves the best classification performance, with an average accuracy of 97.54% on the OpenBMI dataset, and 89.26% on our Self-collected dataset after five rounds of repeated stimuli flashes. It is worth noting that a classification accuracy of 100% is obtained by 71.69% (38 out of 53) participants on the OpenBMI dataset, which is higher than obtained by traditional deep learning models with cross-participant classification tasks. In addition, significant improvements have been gained when compared with BASIC-CNN, EEGNet, ShallowNet, and DeepCovNet models after using data from five rounds of repeated stimuli flashes. It is crucial to point out that one reason for the higher ITR obtained with our Self-collected dataset is due to the shorter duration between targets in our experiment design. More specifically, 4.5 s were given to the user for identifying, locating, and gazing at the next target character when collecting the OpenBMI dataset, while we set this duration to 0.85s in our Self-collected dataset. Baseline methods used in the current work are all based on CNN architecture, which only includes a feature extractor and label classifier and does not consider the domain distribution shift problem. OADANN method incorporated with domain discriminator can learn the domain invariant features through mitigating the feature discriminative ability. Therefore, our proposed method outperforms the other models without the need for any calibration data from target/test participants.

Our experiments on transfer learning between participants from the perspective of DG tasks allow us to make some observations. First, we find that a deep convolutional neural network obtains the best generalization performance of all deep learning models, so we choose it as the backbone of our proposed method. Second, when comparing three

DG algorithms, ERM outperforms GroupDRO and Mixup, and our proposed method achieves superior detection performances compared to the other three DG algorithms. We only test the performance of our OADANN method to detect ERPs elicited by the traditional P300 speller paradigm, whose main components are P300 and N400 potentials. Future work will focus on other stimulation paradigms to evaluate the performance of our proposed method when applied to different cognitive events.

6. Conclusion

Most BCI studies focus on intra-participant classification tasks to learn participant-specific features to reach robust performances [43]. However, this process can be time-consuming and inconvenient. Moreover, non-stationarity of EEG data over days and weeks can reduce the generalizability of models trained in this way. As a consequence, researchers have developed transfer learning methods to address the inter-participant variability problem in order to improve the BCI system's reliability. In our experiments, model training is carried out in a participant-independent scheme. We aim to construct a participant-independent model for online spelling for new users. The proposed model will be trained in advance and then utilized directly in the online system. Therefore, the training process will not affect the practical use. In the practical life, OADANN model as a generic model can directly be used to execute online speller tasks without the need for any calibration data from target/test participants. To reduce the inter-participant variability, we have introduced the concept of domain generalization, where models can be trained without any information from target participants. During the training process, the existing data from other participants is considered to be the source domain, and the new user is designated as the target participant. We proposed the use of attention-based deep adversarial training DG models and also applied some conventional DG methods for comparison. Table I shows that the classification results achieved with our baseline deep learning methods are relatively poor due to the effect of individual differences on model training. Hence, we consider the use of an adversarial mechanism to improve model generalization. Furthermore, two kinds of methods are compared: i) deep learning methods; ii) domain generalization methods. Finally, we performed an ablation study to explore the contribution of each module.

In summary, we propose a one-source domain transfer based attention domain adversarial neural network (OADANN) to analyze the EEG signals for participant-independent ERP classification. OADANN cascades the deep convolutional module, attention module, and domain adversarial module to simultaneously learn the spatial and temporal representations of EEG signals. We have compared a series of existing state-of-the-art methods and domain generalization methods on two datasets. The comparison results demonstrated that our OADANN method can extract the underlying invariant features of EEG signals for cross-participant transfer learning tasks so that the new participants can use ERP based BCI systems directly without the need for calibration data. In the future, we will focus on developing a more effective model on a variety of ERP paradigms to enhance the generalization of BCI systems.

Acknowledge

This work was supported by Young Scientists Fund of the National Natural Science Foundation of China under Grant 62306111, the China Postdoctoral Science Foundation under Grant 2023M741177, and Postdoctoral Fellowship Program of CPSF under Grant GZB20230216, in part by the Grant National Natural Science Foundation of China under Grant 62176090 and STI 2030-major projects 2022ZD0208900; in part by Shanghai Municipal Science and Technology Major Project under Grant 2021SHZDZX, in part by the Program of Introducing Talents of Discipline to Universities through the 111 Project under Grant B17017; This research is also supported by Project of Jiangsu Province Science and Technology Plan Special Fund in 2022 (Key research and development plan industry foresight and key core technologies) under Grant BE2022064-1.

References

[1] A. D. Degenhart, W. E. Bishop, E. R. Oby, E. C. Tyler-Kabara, S. M. Chase, A. P. Batista, and B. M. Yu, "Stabilization of a brain–computer interface via the alignment of low-dimensional spaces of neural activity," *Nature biomedical engineering,* vol. 4, no. 7, pp. 672-685, 2020.

[2] S. Li, J. Jin, I. Daly, C. Zuo, X. Wang, and A. Cichocki, "Comparison of the ERP-based BCI performance among chromatic (RGB) semitransparent face patterns," *Frontiers in Neuroscience*, pp. 54, 2020.

[3] J. Li, F. Wang, H. Huang, F. Qi, and J. Pan, "A novel semi-supervised meta learning method for subject-transfer brain–computer interface," *Neural Networks,* vol. 163, pp. 195-204, 2023.

[4] Á. Fernández-Rodríguez, M. T. Medina-Juliá, F. Velasco-Álvarez, and R. Ron-Angevin, "Effects of spatial stimulus overlap in a visual P300-based brain-computer interface," *Neuroscience,* vol. 431, pp. 134-142, 2020.

[5] S. Li, J. Jin, I. Daly, C. Liu, and A. Cichocki, "Feature selection method based on Menger curvature and LDA theory for a P300 brain–computer interface," *Journal of Neural Engineering,* vol. 18, no. 6, pp. 066050, 2022.

[6] S. Li, J. Jin, I. Daly, X. Wang, H.-K. Lam, and A. Cichocki, "Enhancing P300 based character recognition performance using a combination of ensemble classifiers and a fuzzy fusion method," *Journal of Neuroscience Methods,* vol. 362, pp. 109300, 2021.

[7] L. A. Farwell, and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and clinical Neurophysiology,* vol. 70, no. 6, pp. 510-523, 1988.

[8] H. Gao, C. Lv, T. Zhang, H. Zhao, L. Jiang, J. Zhou, Y. Liu, Y. Huang, and C. Han, "A structure constraint matrix factorization framework for human behavior segmentation," *IEEE Transactions on Cybernetics,* vol. 52, no. 12, pp. 12978-12988, 2021.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature,* vol. 521, no. 7553, pp. 436-444, 2015.

[10] B. Pu, K. Li, S. Li, and N. Zhu, "Automatic fetal ultrasound standard plane recognition based on deep learning and IIoT," *IEEE Transactions on Industrial Informatics,* vol. 17, no. 11, pp. 7771-7780, 2021.
[11] J. Mao, S. Qiu, W. Wei, and H. He, "Cross-modal guiding and reweighting network for multi-modal RSVP-based target detection," *Neural Networks,* vol. 161, pp. 65-82, 2023.
[12] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE transactions on neural networks and learning systems,* vol. 29, no. 11, pp. 5619-5629, 2018.
[13] J. Jin, R. Xu, I. Daly, X. Zhao, X. Wang, and A. Cichocki, "MOCNN: A Multiscale Deep Convolutional Neural Network for ERP-Based Brain-Computer Interfaces," *IEEE Transactions on Cybernetics*, 2024.
[14] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: a review," *Journal of neural engineering,* vol. 16, no. 3, pp. 031001, 2019.
[15] H. Cecotti, and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE transactions on pattern analysis and machine intelligence,* vol. 33, no. 3, pp. 433-445, 2010.
[16] X. Liu, Q. Xie, J. Lv, H. Huang, and W. Wang, "P300 event-related potential detection using one-dimensional convolutional capsule networks," *Expert Systems with Applications,* vol. 174, pp. 114701, 2021.
[17] D. Borra, E. Magosso, M. Castelo-Branco, and M. Simões, "A Bayesian-optimized design for an interpretable convolutional neural network to decode and analyze the P300 response in autism," *Journal of Neural Engineering,* vol. 19, no. 4, pp. 046010, 2022.
[18] S. Tortora, S. Ghidoni, C. Chisari, S. Micera, and F. Artoni, "Deep learning-based BCI for gait decoding from EEG with LSTM recurrent neural network," *Journal of neural engineering,* vol. 17, no. 4, pp. 046011, 2020.
[19] J. Jin, S. Li, I. Daly, Y. Miao, C. Liu, X. Wang, and A. Cichocki, "The study of generic model set for reducing calibration time in P300-based brain–computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 28, no. 1, pp. 3-12, 2019.
[20] D. Wu, X. Jiang, and R. Peng, "Transfer learning for motor imagery based brain–computer interfaces: A tutorial," *Neural Networks,* vol. 153, pp. 235-253, 2022.
[21] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of neural engineering,* vol. 15, no. 5, pp. 056013, 2018.
[22] E. Santamaria-Vazquez, V. Martinez-Cagigal, F. Vaquerizo-Villar, and R. Hornero, "EEG-inception: A novel deep convolutional neural network for assistive ERP-based brain-computer interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 28, no. 12, pp. 2773-2782, 2020.
[23] M. W. Bhatt, and S. Sharma, "Next Generation Imaging in Consumer Technology for ERP Detection based EEG Cross-Subject Visual Object Recognition," *IEEE Transactions on Consumer Electronics*, 2024.

[24] R. Ma, H. Zhang, J. Zhang, X. Zhong, Z. L. Yu, Y. Li, T. Yu, and Z. Gu, “Bayesian uncertainty modeling for P300-based brain-computer interface,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2023.

[25] H. Gao, J. Zhu, T. Zhang, G. Xie, Z. Kan, Z. Hao, and K. Liu, “Situational assessment for intelligent vehicles based on stochastic model and Gaussian distributions in typical traffic scenarios,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* vol. 52, no. 3, pp. 1426-1436, 2020.

[26] H. Gao, Y. Qin, C. Hu, Y. Liu, and K. Li, “An interacting multiple model for trajectory prediction of intelligent vehicles in typical road traffic scenario,” *IEEE transactions on neural networks and learning systems,* vol. 34, no. 9, pp. 6468-6479, 2021.

[27] S. Panwar, P. Rad, T.-P. Jung, and Y. Huang, “Modeling EEG data distribution with a Wasserstein generative adversarial network to predict RSVP events,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 28, no. 8, pp. 1720-1730, 2020.

[28] Y. Li, W. Zheng, Y. Zong, Z. Cui, T. Zhang, and X. Zhou, “A bi-hemisphere domain adversarial neural network model for EEG emotion recognition,” *IEEE Transactions on Affective Computing,* vol. 12, no. 2, pp. 494-504, 2018.

[29] J. Jin, Z. Wang, R. Xu, C. Liu, X. Wang, and A. Cichocki, “Robust similarity measurement based on a novel time filter for SSVEPs detection,” *IEEE Transactions on Neural Networks and Learning Systems,* vol. 34, no. 8, pp. 4096-4105, 2021.

[30] X. Chen, Z. Wang, and D. Wu, “Alignment-Based Adversarial Training (ABAT) for Improving the Robustness and Accuracy of EEG-Based BCIs,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2024.

[31] B.-Q. Ma, H. Li, Y. Luo, and B.-L. Lu, "Depersonalized cross-subject vigilance estimation with adversarial domain generalization." pp. 1-8.

[32] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, “Domain-adversarial training of neural networks,” *The journal of machine learning research,* vol. 17, no. 1, pp. 2096-2030, 2016.

[33] E. Jeon, W. Ko, J. S. Yoon, and H.-I. Suk, “Mutual information-driven subject-invariant and class-relevant deep representation learning in BCI,” *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[34] Z. Jia, X. Cai, and Z. Jiao, “Multi-modal physiological signals based squeeze-and-excitation network with domain adversarial learning for sleep staging,” *IEEE Sensors Journal,* vol. 22, no. 4, pp. 3464-3471, 2022.

[35] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks." pp. 11534-11542.

[36] Y. Ding, N. Robinson, S. Zhang, Q. Zeng, and C. Guan, “Tsception: Capturing temporal dynamics and spatial asymmetry from EEG for emotion recognition,” *IEEE Transactions on Affective Computing*, 2022.

[37] Y. Zhang, E. Yin, F. Li, Y. Zhang, D. Guo, D. Yao, and P. Xu, “Hierarchical feature fusion framework for frequency recognition in SSVEP-based BCIs,” *Neural Networks,* vol. 119, pp. 1-9, 2019.

[38] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K.

Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, “Deep learning with convolutional neural networks for EEG decoding and visualization,” *Human brain mapping,* vol. 38, no. 11, pp. 5391-5420, 2017.

[39] D. Borra, S. Fantozzi, and E. Magosso, "Convolutional neural network for a P300 brain-computer interface to improve social attention in autistic spectrum disorder." pp. 1837-1843.

[40] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, “mixup: Beyond empirical risk minimization,” *arXiv preprint arXiv:1710.09412*, 2017.

[41] S. Sagawa, P. W. Koh, T. B. Hashimoto, and P. Liang, “Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization,” *arXiv preprint arXiv:1911.08731*, 2019.

[42] S. Lall, D. Sinha, A. Ghosh, D. Sengupta, and S. Bandyopadhyay, “Stable feature selection using copula based mutual information,” *Pattern Recognition,* vol. 112, pp. 107697, 2021.

[43] D. Zhao, F. Tang, B. Si, and X. Feng, “Learning joint space–time–frequency features for EEG decoding on small labeled data,” *Neural Networks,* vol. 114, pp. 67-77, 2019.