Subject-Independent Wearable P300 Brain–Computer Interface Based on Convolutional Neural Network and Metric Learning

Li Hu, Wei Gao, Zilin Lu, Chun Shan, Haiwei Ma, Wenyu Zhang, and Yuanqing Li, *Fellow, IEEE*

***Abstract*— The calibration procedure for a wearable P300 brain-computer interface (BCI) greatly impact the user experience of the system. Each user needs to spend additional time establishing a decoder adapted to their own brainwaves. Therefore, achieving subject independent is an urgent issue for wearable P300 BCI needs to be addressed. A dataset of electroencephalogram (EEG) signals was constructed from 100 individuals by conducting a P300 speller task with a wearable EEG amplifier. A framework is proposed that initially improves cross- subject consistency of EEG features through a common feature extractor. Subsequently, a simple and compact convolutional neural network (CNN) architecture is employed to learn an embedding sub-space, where the mapped EEG features are maximally separated, while pursuing the minimum distance within the same class and the maximum distance between different classes. Finally, the model's generalization capability was further optimized through fine-tuning. Results: The proposed method significantly boosts the average accuracy of wearable P300 BCI to 73.23±7.62% without calibration and 78.75±6.37% with fine-tuning. The results demonstrate the feasibility and excellent performance of our dataset and framework. A calibration-free wearable P300 BCI system is feasible, suggesting significant potential for practical applications of the wearable P300 BCI system.**

***Index Terms*— Subject-independent, brain–computer interface (BCI), P300, convolutional neural network (CNN), wearable.**

Manuscript received 3 April 2024; revised 16 July 2024 and 23 August 2024; accepted 7 September 2024. Date of publication 10 September 2024; date of current version 25 September 2024. This work was supported in part by the Chinese Ministry of Science and Technology Project through STI 2030-Major Projects under Grant 2022ZD020 8900, in part by the National Natural Science Foundation of China under Grant 62273108, and in part by the National Key Research and Development Program under Grant 2022YFB3604502. *(Corresponding author: Yuanqing Li.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Guangdong Work Injury Rehabilitation Hospital under Approval No. AF/SC-07/2023.01.

Li Hu is with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510641, China, also with South China Brain-Computer Interface Technology Company Ltd., Guangzhou 510220, China, and also with the School of Electronic and Information, Guangdong Polytechnic Normal University, Guangzhou 510640, China (e-mail: 20225875@qq.com).

Wei Gao is with the School of Artificial Intelligence, South China Normal University, Guangzhou 510631, China, and also with the Research Center for Brain-Computer Interface, Pazhou Laboratory, Guangzhou 510330, China.

Zilin Lu and Yuanqing Li are with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510641, China, and also with Guangzhou Key Laboratory of Brain Computer Interface and Applications (no. 15180006), Guangzhou 510640, China (e-mail: auyqli@scut.edu.cn).

Chun Shan is with the School of Electronic and Information, Guangdong Polytechnic Normal University, Guangzhou 510640, China.

Haiwei Ma and Wenyu Zhang are with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510641, China, and also with South China Brain-Computer Interface Technology Company Ltd., China.

Digital Object Identifier 10.1109/TNSRE.2024.3457502

I. Introduction

BRAIN–COMPUTER interfaces (BCIs) provide direct communication channels between the brain and external devices [1]. These systems can translate brain activity into information for clinical applications or commands to control electronic devices, showing great potential in the fields of communication, control, and rehabilitation [2]. Among them, the EEG-based BCI system is the most widely used and exhibit the highest practical potential. P300, steady-state visual evoked potential (SSVEP), and motor imagery (MI) are three main paradigms for EEG-based BCI systems [3], [4]. Additionally, to achieve better detection performance and multi-degree control, hybrid BCIs developed by combining the aforementioned EEG signals have also become very popular [5], [6].

Over the past decade, although significant progress has been made in BCI systems, there remain several barriers preventing the transfer of BCI systems from the laboratory to real-world scenarios. Firstly, traditional BCI systems typically rely on desktop EEG recording devices, which are bulky and expensive, significantly limiting the use of BCI systems outside the laboratory. Therefore, BCI systems have not yet gained widespread acceptance among users. In recent years, wearable devices have gradually integrated into people's daily lives, due to improvements in their facilitation and comfort [7]. Meanwhile, the P300 potential recorded by EEG has exhibited excellent performance in areas such as speller, wheelchair control, and consciousness detection [9], [10], [11].

Achieving wearable P300 BCI systems can effectively execute BCI tasks and enhance people's acceptance of BCI systems simultaneously.

Furthermore, brain waves are constantly changing over time in response to an individual's psychological and physiological state [14]. This inter/intra-subject variability in the EEG makes it difficult for the classifier to identify the EEG features of different individuals or the same individual at different time periods [15]. Therefore, individuals using a wearable P300 BCI system require 10-15 minutes to establish their personalized brainwave decoder tailored to their unique profile. To address this issue, some studies attempt to develop subject-independent models through data augmentation. This process involves generating additional training samples or enhancing the quality of training data, thereby improving generalization capability of a classifier. For example, [16] used temporal, spatial, and rotational distortions to generate augmented data, which can improve the performance of the P300 BCI system by 1% to 6%. Reference [17] attempted to generate artificial EEG training data with deep convolutional generative adversarial networks (DCGAN) and wasserstein GAN with gradient penalty (WGAN-GP). The experimental results indicate that data augmentation is effective for subject-specific and subject-independent P300 BCI systems. Another idea for achieving zero calibration is through learning strategies. Some systems focus on building classical classifiers through training data, such as xDAWN [18], ridge regression [19], then fine-tune classifier parameters using test subject data for adapting to distribution shift of EEG signals. In contrast, other systems integrate multiple classifiers, then enhance the generalization capability of a classifier by weighting [21] or voting [22]. In addition, the elimination of the calibration process can be achieved by learning invariant brain patterns. For instance, in [14], discriminative spectral-spatial EEG features are employed as a general brain pattern for motor imagery. Similarly, [23] attempted to learn invariant representations in P300 EEG signals through CNN model and large datasets, achieving favorable results in online simulation tests. However, previous research on wearable calibration-free BCI systems has been constrained by limitations in datasets, lacking sufficient data for model training and testing. Additionally, the performance of existing BCI systems still demands further enhancement.

Recently, metric learning has emerged as a widely adopted machine learning method, demonstrating promising results in computer vision and natural language processing [26]. Metric learning assesses sample similarity through distance and refines this distance through learning tasks, potentially revealing latent invariant representations of the samples [27], [28]. In [21], an attempt was made to acquire a concise and distinctive latent representation from EEG signals through the utilization of an end-to-end multi-task autoencoder, this method involved utilizing metric learning to identify the optimal distances between invariant patterns while ensuring separability between different categories. At present, research on metric learning for BCI systems is still relatively limited. Further studies are necessary to explore the full potential of metric learning in enhancing BCI systems.

In this work, we built a large-scale wearable P300 EEG database and report a subject-independent wearable P300 BCI system framework. The main contributions of this article are as follows:

1) We constructed a large-scale wearable P300 EEG database. To the best of our knowledge, the dataset is the first and largest wearable P300 EEG database reported in the existing literature.

2) We propose a simple and compact CNN model integrated with metric learning to learn latent discriminative invariant brain patterns from EEG dataset for achieving wearable P300 BCI system without calibration.

3) We attempt to enhance model generalization capability and classification performance by improving sample distribution consistency and model fine-tuning.

The remainder of this article is organized as follows: Section II explains the materials and methods, including the participants, the equipment, the graphical user interface (GUI), the paradigm, and some backgrounds; Section III describes the detailed experimental results; Section IV evaluates and discusses the proposed method; Section V concludes the article and provides prospects for future research.

II. Materials and Methods

A. Participants

100 healthy subjects (58 females and 42 males, aged 18 to 40 years (mean±SD = 24.84±5.91)) were employed for this pilot study and all of them provided written informed consent. According to their reports, all the participants had normal or corrected-to-normal vision and None of the participants had a relevant history of mental or neurological disorders or other cerebral diseases. This work was approved by the Guangdong Work Injury Rehabilitation Hospital (approval number: AF/SC-07/2023.01).

B. Equipment

The EEG signals were recorded by an elaborately designed wearable EEG amplifier (For more details, please refer to [24]). The wearable EEG amplifier was sampled at 250Hz from three channels (Oz, P3, P4) following the 10-20 system. All electrode impedances were maintained at less than 5kΩ. During the experiment, the participants sat approximately 60 cm in front of a 15.6-inch LCD screen with a resolution of 2560 × 1600 pixels, 141 pixels per inch (PPI) and a refresh rate of 60 Hz.

C. GUI and Paradigm

1) GUI: In this study, we collect EEG signals though a P300 speller with a GUI shown in Fig. 1. The GUI consists of a text box and sixteen buttons. The text box is located at the top of the screen, used to present tips, messages, and results. The buttons are evenly arranged in 4 rows and 4 columns at the bottom of the screen, with each button corresponding to a character.

2) Paradigm: The paradigm employed for P300 speller is follow the single display (SD) paradigm [25]. In the SD paradigm, only one button highlighted each time, and a complete flicker sequence, during which all buttons are randomly

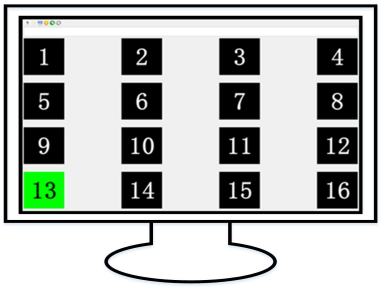

Fig. 1. GUI of P300 speller. The top text box shows cue character (target) and tips. The sixteen buttons provide standard visual stimuli and target visual stimuli (black turn to green).

highlighted only once, is referred to as a round. In our study, we illuminated each button for 10 ms, succeeded by an interstimulus interval (ISI) of 100 ms without any highlighted characters. Therefore, the total duration of a round amounted to 1.76 (110 ms×16). In the experiment, sixteen characters are randomly selected as targets exactly once. To acquire reliable P300 EEG signals, participants are directed to concentrate on target character for 10 rounds and silently count the number of flashes. This entire process involving one character is referred to as a trial. The EEG signals were recorded in the European Data Format (EDF).

D. Dataset Description

A participant's EEG data consists of 2560 epochs (16 trials × 10 rounds × 16 button). The duration of EEG signal for each epoch is 1 second. The EEG time segments (0–600ms) were isolated from EEG epochs. Therefore, there were 150 (250Hz × 600ms) sample points in each time segment. The baseline of EEG signal was adjusted by subtracting the mean amplitude between −200 and 0ms. Subsequently, the EEG signals were filtered by a 6th order Butterworth filter with a bandwidth ranging from 0.1 to 25 Hz. The collected EEG data is divided into two groups for analysis: some subjects' data are used for training and validation, while a portion of the subjects' data is used for subject-independent testing. Note that five subjects were instructed to repeat this P300 speller experiment 10 times over two weeks for cross-time analysis. In the remainder of this paper, the subsets of dataset are denote as shown in Tabel I.

E. Proposed Framework

In this section, we introduce the framework proposed to address subject-independent. The framework consists of four key components: feature extraction (Fig. 2.a), CNN model (Fig. 2.b), classifier (Fig. 2.c), metric learning (Fig. 2.d).

1) Feature Extraction: The EEG time fragments in the dataset are represented as $S^n_{k,c,r,j,l}$, where n, k, c, r, j,and l correspond to the nth subject($n \in [1, 100]$), the kth trial ($k \in [1, 16]$), the cth channel($c \in [1, 3]$), the rth round($r \in [1, 10]$), the jth character ($j \in [1, 16]$), and the number of samples l ($l = 150$), respectively. Subsequently, the EEG time fragments $S^n_{k,c,r,j,l}$ are fed to common feature extractor and down sampled by a factor of 5 ($l = 30$). Next, the EEG feature vectors $S^n_{k,j,r,l}$ ($l = 90$) were calculated by concatenating the time fragments of all selected channels and the arrangement order of the channels is Oz, P3, P4. Then, the EEG feature vectors $S^n_{k,j,l}$ ($l = 90$) were calculated by averaging the value across all rounds. Therefore, each subject can obtain 256 (16 $trials$ × 16 $characters$ = 256) EEG feature vectors S^n_l ($l = 90$) (Fig. 2.a), consisting of 16 target feature vectors and 240 non-target feature vectors. Finally, the EEG feature vectors were normalized as follows:

$$\tilde{S}_l = \frac{S_l - \bar{S}_l}{\sigma_l} \tag{1}$$

where S_l and $\tilde{S}_l$ refer to the EEG feature vectors before and after normalization respectively. $\bar{S}_l$ and σ_S are represented the mean value and the standard deviation of S_l as follows, respectively.

$$\bar{S}_c = \frac{1}{L}\sum_{l=1}^{L} S_l \tag{2}$$

$$\sigma_S = \sqrt{\frac{1}{L-1}\sum_{l=1}^{L}\left|S_l - \bar{S}_l\right|^2} \tag{3}$$

The P300 signal is widely acknowledged as the most stable and reproducible event-related potential (ERP). While the P300 signals of different subjects may vary in shape or latency, obtaining P300 waveforms through time-locked averaging of the EEG signals for each participant can improve the consistency of data distribution to some extent.

2) CNN Model: The CNN model is utilized to project the input feature vector S into an embedding space, where the input feature vector S is encoded into a latent invariant pattern z, denote as $z = h(S)$. This CNN model consists of three blocks, each containing a Conv1D layer for convolutional, a Batch- Norm1d for normalizing the input signal, a AveragePooling1D layer for pooling, a rectified linear unit (ReLU) for activation and a dropout for reducing the dependency between neurons and increase the generalization ability of the network. The convolutional layer serves as a spatial filter, effectively learning discriminative features from input feature vector, initially represented as a 1 × 90 matrix. The Conv1D operation convolutes the input feature vector along the time dimension with 8 convolutional kernels sized 3 × 3 and a stride of 1. This results in 8 feature maps sized 1 × 45. These feature maps can be regarded as new time-series signals, integrating spatial information from all channels simultaneously. Following the convolutional layer, The AveragePooling1D operation processes the feature maps, retaining the overall features of the data while reducing the number of parameters. This pooling operation contributes to dimensionality reduction, making the subsequent processing more efficient. The rectified linear unit (ReLU) is employed as an activation function between the adjacent blocks. ReLU introduces non-linearity to the model, enabling it to learn complex brain patterns and relationships within the EEG data. This architecture, consisting of convolution, pooling, and activation operations, is repeated for three times, enhancing the model's capacity to extract hierarchical and abstract features from the input feature vector. Ultimately, the feature maps undergo a flattening operation, transforming into latent invariant brain patterns.

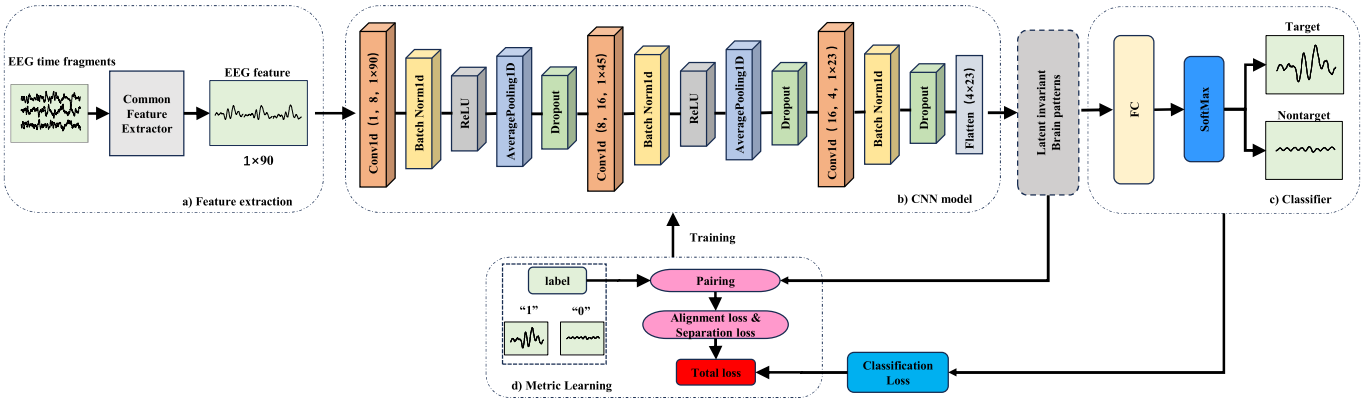

Fig. 2. Overall illustration of our framework. (a) exhibits feature extraction; the common feature extractor compresses the input EEG time fragments and produces the EEG feature. (b) display the CNN model architecture; The convolutional layer is denoted as Conv1D (Cin, Cout, h × w), where Cin and Cout are the number of input and output channels, respectively. The height and width of the kernel are denoted as h and w respectively. (c) illustrate the classifier; the latent invariant brain patterns were fed into a FC layer following a SoftMax function for classification. (d) shows metric learning that learns the latent invariant brain patterns by the classification loss, separation loss, and alignment loss.

3) Classifier: In this framework, the SoftMax function serves as a classifier to categorize the latent invariant brain patterns within the input feature vector. The latent invariant brain pattern, denoted as z, is fed to the FC layer. Then, the SoftMax function is applied to the output of the FC layer to obtain the probability for each class. This process is mathematically expressed as follows:

$$\hat{y} = SoftMax(Wz + b) \tag{4}$$

where $\hat{y}$, W,and b are the probability for each class, the weight matrix, and the bias vector, respectively. The classification loss is computed using the cross-entropy loss function:

$$\mathcal{L}_{ce}\left(y, \hat{y}\right) = -\sum_{i=1}^{m} y_i log \hat{y}_i \tag{5}$$

Here, m is the number of classes, y is the true label. The purpose of this classification loss is to ensure maximum separability between classes.

4) Metric Learning: To ensure the CNN model effectively maps EEG features into a latent invariant subspace, we introduced the metric learning (ML) module. With the ML module, the CNN model can preserve the separability of EEG feature while enhancing intra-class compactness and inter-class separability simultaneously [27] (Fig. 3). To achieve this, the ML module utilizes a triplet loss to train the CNN model, ensuring that the CNN model can learn the optimal distances among different classes of the latent invariant brain pattern [21].

Suppose that input EEG signal for each subject forms a domain D. $\Delta = \{D_{S1}, \cdots, D_{Sn}\}$ represent n distinct source domains. During the training, we consider every distinct unordered pair of source domains, denoted as (D_{Su}, D_{Sv}). A set of $\{S^p_{Su}, S^q_{Sv}, c\}$ are samples paired by labels from (D_{Su}, D_{Sv}). Where S^p_{Su} and S^q_{Sv} are EEG feature vectors from domain D_{Su} and D_{Sv}, respectively, and c indicates whether the labels corresponding to two EEG feature vectors are the same. while $\left\{z^p_{Su}, z^q_{Sv}, c\right\}$ is latent invariant brain patterns corresponding to the EEG signals. Depending on the value of c, alignment loss ($c = 1$) and separation loss ($c = 0$) will

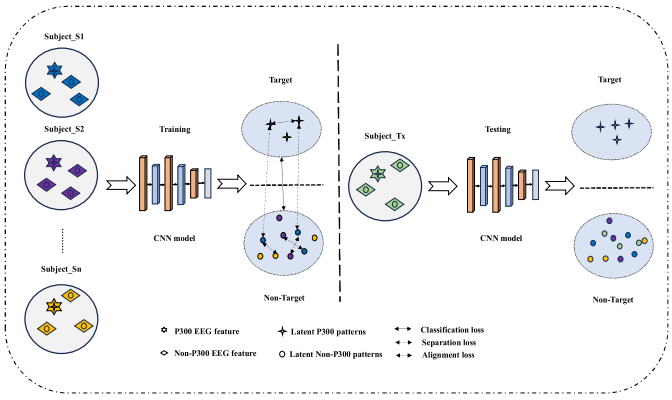

Fig. 3. Visualization of metric learning: The classification loss ensures maximum separability between classes, the alignment loss enhances intra-class compactness, and the separation loss ensures inter-class separability.

be calculated using the formulas as follows, respectively:

$$\mathcal{L}_{al} = \frac{1}{2}\left\|z^p_{Su} - z^q_{Sv}\right\|^2 \tag{6}$$

$$\mathcal{L}_{se} = \frac{1}{2}\max(0, m - \left\|z^p_{Su} - z^q_{Sv}\right\|)^2 \tag{7}$$

where $\|\cdot\|$ represent Euclidean distance, and $m > 0$ is represents a margin that the distance between samples of different classes should exceed this margin value. Finally, the total loss is the sum of the losses from (5), (6) and (7). Note that the $L_{al} + L_{se}$ corresponds to the famous contrastive loss defined in [28]. The total loss is expressed as follows:

$$\mathcal{L}_{total} = \alpha\mathcal{L}_{ce} + (1 - \alpha)\left(\mathcal{L}_{al} + \mathcal{L}_{se}\right) \tag{8}$$

where α is a coefficient used to balance the classification loss, separation loss and alignment loss.

F. Fine-Tuning

Fine-tuning facilitates the transfer of knowledge from a pre-trained model to a specific task, allowing the use of a small dataset to quickly improve the model's classification accuracy and generalization ability. In our framework, all information

originates from the source domain, and data from the test domain is unseen to us. While we have integrated metric learning to enable the model capture potential brain invariant patterns, fine-tuning the pre-trained model with a small set of test data remains essential. Fine-tuning enables our model to quickly adapt to the testing domain, achieving higher accuracy while greatly shorten the time needed for calibration, thereby improving the user experience of the BCI system.

G. Model Training

The proposed approach was implemented using the Pytorch frameworks with CUDA 11.7 and cuDNN v8.5. An NVIDIA GeForce RTX 2060 Ti GPU was employed for hardware acceleration. The well-known Xavier method is used to initialize the convolutional kernels and weights of the CNN. In each training epoch, the loss function is optimized using the Aam optimizer with a learning rate of 0.001. Considering the imbalance in the training dataset, characterized by a class sample ratio of 1:15 (target to non-target), the classification loss $\mathcal{L}_{ce}$ was weighted by a factor of 15 for the target class.

III. Experiments and Results

A. Result of Subject-Independent and Cross-Time Offline Test

The P300 signals are evoked by visual stimuli, known as oddball paradigm, and the iterative presentation of these stimuli progressively enhances the P300 signal waveform. Thereby improving the classification accuracy of P300 signal. To assess the performance of our model, we have devised an offline test. In offline tests, the EEG features of the test are the average of signals for each participant over 10 rounds and the character corresponding to the maximum classification probability is the target and the procedure can be denoted as follows:

$$D_k^n = \underset{j\in\{1,\ldots,16\}}{\arg\max}\ P(y=1|\bar{\bar{S}}_{k,j}^n) \tag{9}$$

where D_k^n represents the character output of the classifier for the kth test of the nth participant and $\bar{\bar{S}}$ represents the average of EEG feature vectors cross 10 rounds after normalization. In the subsequent sections of this paper, we denote our framework as CNN-ML with α set to 0.25.

For the subject-independent test, we randomly selected 30 participants to form the test set (Te-cs-30), while the brainwave data from the remaining 70 participants constituted the training and validation dataset (Tv-cs-70). Ten-fold cross-validation is employed to optimize parameters for model training. subject-independent testing was carried out using offline analysis. The classification accuracy with respect to subject are shown in Table II.

Furthermore, brain signals not only constantly change due to individual differences, but the brain signals of the same subject can also vary significantly at different time periods. To validate the model's adaptability to intra-subject EEG signals, we conducted cross-time tests on the model. We used the same training and validation set as in subject-independent test (Tv-cs-70). The five subjects EEG signal mentioned in section II-D are employed as test set (Te-ct-5). It is noteworthy that the EEG signals of each subject were collected from 10 different time periods. The cross-time experiment results are shown in Fig. 4.

B. Result of Hyperparameters Test

1) Training Epoch: The setting of the training epochs is crucial for the performance of the model. Too many training epochs may lead to overfitting and waste computational resources, while too few training epochs may not provide the model with sufficient time to learn, resulting in non-convergence and poor performance. In the test, the training epoch is set to 100 times. After each set of 10 training epochs, the model is tested once using the test set (Te-cs-30). The classification accuracy and training loss with respect to training epochs are illustrated in Fig. 5.

2) Dataset Size: In general, the more samples in a dataset, the greater the number of features available for the classification model to learn, resulting in higher classification accuracy [12]. To further explore the influence of sample size on our CNN-ML model for the wearable P300 BCI system, we attempted to train the model using different quantities of samples and conducted tests to evaluate the model's classification accuracy. Specifically, we use randomly selected 10 participants to form the test set (Te-cs-10) and the remaining 90 participants randomly contributing to the creation of training sets denoted as Tv-cs-10 to Tv-cs-90, respectively. The classification accuracy with respect to various sizes of training datasets are presented in Fig. 6.

3) Feature Maps and Kernel Size: Tables III and IV illustrate the performance variations of the proposed method according to the number of feature maps and the kernel size. The classification accuracies are 77.36±9.13%, 73.45±7.85%, and 70.79± 8.54% according to the kernel size, respectively. The classification accuracies are 69.34±5.91%, 76.28±7.45%, and 71.82± 9.41% according to the number of feature maps, respectively.

C. Comparison With Other Models

In this experiment, we introduce the classic support vector machine (SVM), CNN, MsCNN, and EEG-Net as the baseline models, The SVM classifier is implemented using the Scikit- Learn package (version 0.21.3) with the radial basis function (RBF) as the kernel function. In [31] and [32], SVM has been exhibited the capability to accurately classify P300 signals. The CNN, MsCNN, and EEG-Net models have demonstrated excellent performance in the processing of EEG signals [12], [34], [41]. The CNN model employs the same network parameters as our proposed method, and all methods underwent evaluation using identical training and testing sets. Specifically, training set and test set are both the same as in Section III-A, denoted as Te-cs-30 and Tv-cs-70, respectively. The classification accuracy with respect to various methods are presented in Table II.

D. Result of Online Test

During this experiment, we attempted to evaluate the performance of the subject-independent wearable BCI system in an

TABLE I
PARTITIONS OF THE DATASET

Task type	Training and validation set		Testing set	
	Symbol	Number of subjects	Symbol	Number of subjects
Subject independent (Cross-subject)	Tv-cs-*n*	*n*	Te-cs-n	*n*
Cross-time	Tv-ct-*n*	*n*	Te-ct-*n*	*n*

Note: In the cross-time task, each subject has 10 sets of data.

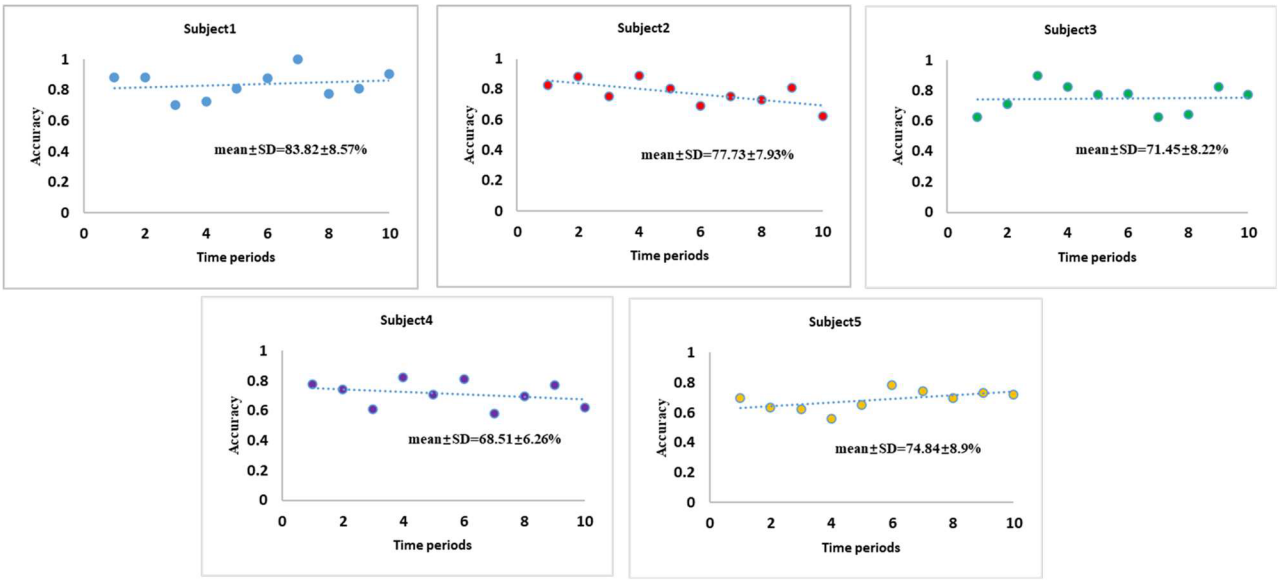

Fig. 4. Scatter plot of individual classification performance in the cross-time test. The horizontal axis is the different time periods for recording the EEG signal of subject, and the vertical axis is the classification accuracy.

TABLE II
RESULTS OF THE SUBJECT-INDEPENDENT TEST WITH DIFFERENT METHODS AND MODELS

Subject (Age)	Models			Methods	
	SVM	MsCNN	EEG-Net	CNN	CNN-ML
S1(22)	61.43%	69.36%	68.43%	61.03%	68.53%
S2(23)	68.32%	75.25%	70.33%	68.56%	72.65%
S3(21)	55.87%	71.01%	63.72%	58.81%	71.83%
S4(19)	67.88%	74.37%	81.53%	70.76%	67.14%
S5(19)	74.46%	75.77%	73.28%	72.98%	87.51%
S6(22)	75.83%	80.23%	83.51%	67.35%	84.98%
S7(24)	62.29%	73.79%	72.38%	63.84%	78.15%
S8(22)	63.35%	63.58%	71.91%	73.69%	79.37%
S9(25)	72.43%	78.86%	72.95%	85.51%	69.84%
S10(19)	76.54%	57.53%	65.79%	72.17%	78.74%
Mean ± SD	67.74 ± 6.57%	71.98 ± 6.61%	72.38 ± 5.89%	69.47 ± 7.2%	**75.87 ± 6.62%**

online free spelling task. All subjects were instructed to spell 16 predesignated characters (i.e., each of the 16 characters on the screen was selected once in a pseudorandom manner). At the beginning of the experiment, a pseudorandom sequence of 16 characters was displayed in the text box. The initial color of the characters in the sequence was black. If a character

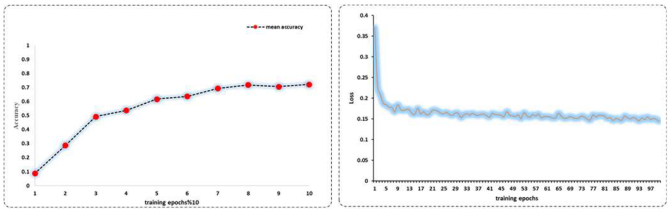

Fig. 5. The result of training epochs test. The left subplot shows mean classification accuracy of 30 subject with respect to the training epochs. The right subplot depicts the training loss with respect to the training epochs.

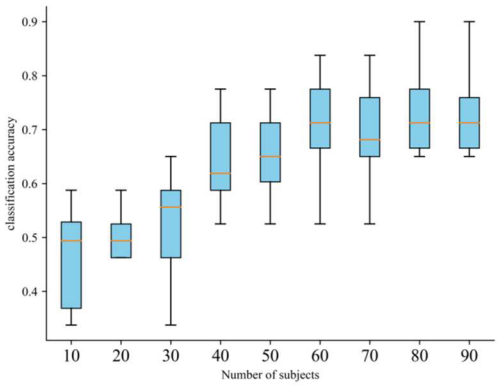

Fig. 6. Classification accuracy with respect to the number of subjects.

TABLE III
CLASSIFICATION ACCURACY ACCORDING TO THE KERNEL SIZE

kernel size	Mean ± SD	Median	Range (Min-Max)
1×3	77.36 ± 9.13%	75%	(62.75%-93.75%)
1×5	73.45 ± 7.85%	75%	(50%-87.5%)
1×7	70.79 ± 8.54%	75%	(50%-81.25%)

was spelled correctly, its color changed to green. If spelled incorrectly, its color changed to red. Note that none of the subjects participated in the data collection experiment.

The entire experiment consists of two stages. During the first stage, participants conducted free spelling tests directly using a pretrained model. In the second stage, the pretrained model underwent fine-tuning for 5 trials before participants conducted another free spelling test. As described in Section II, each round took 1.76 seconds. Therefore, the fine-tuning process in the second stage only took $1.76 \times 10 \times 5 = 88$ seconds, significantly reducing the time required to calibrate the wearable P300 BCI system.

To evaluate the online performance of subject-independent wearable BCI system, we calculated the accuracy, mean response time (RT) for single target selection, and information transfer rate (ITR) (Table V). ITR represents the total amount of information received by the BCI system per unit time and is widely used to assess BCI system performance. The ITR (bits per minute) is calculated as follows [24]:

$$ITR = 60\left(\log_2 M + P\log_2 P + (1-P)\log_2\left(\frac{1-P}{M-1}\right)\right)/T \tag{10}$$

TABLE IV
CLASSIFICATION ACCURACY ACCORDING TO THE NUMBER OF FEATURE MAPS

feature maps	Mean ± SD	Median	Range (Min-Max)
4,8,16	69.34 ± 5.91%	75%	(50%-87.5%)
8,16,4	76.28 ± 7.45%	75%	(50%-87.5%)
10,15.20	71.82 ± 9.41%	75%	(56.25%-87.5%)

where M is the number of characters, P is the accuracy of the BCI system, and T is the mean response time required to select a single character. Please note that the ITR calculation in this experiment does not include preparation time and fine-tuning time.

IV. DISCUSSION

The motivation of this study is to eliminate the cumbersome calibration process of wearable P300 BCI system. To achieve this, we developed a subject-independent CNN model based on metric learning. In this section, we discuss the model and experimental results. Firstly, we explore the subject-independent and cross-time test. Secondly, we evaluate the impact of Hyperparameters. Thirdly, we compare the proposed method with other models. Four, we analyze the online test results and the model fine-tuning. Finally, we introduce the limitations of the proposed method and presented ideas for future work.

A. Analysis of Subject-Independent and Cross-Time Offline Test

The P300 signal refers to the potential fluctuation generated by the brain in response to transient visual stimuli. This distinctive signal exhibits a latency of 300ms and demonstrates excellent reproducibility. The P300 signals of most individuals possess noteworthy recognizability, only showcasing variations in waveform shape, amplitude, and latency. Previous studies have demonstrated that both CNN [34] and SVM [32] models exhibit good classification capabilities for specific subjects' P300 signals. In our subject-independent testing, the classification accuracy of CNN and SVM models is 69.47±7.2% and 67.74±6.57%, respectively. However, the CNN-ML model achieves a higher classification accuracy of 75.87±6.62%. Deep learning, through increasing the sample size or adjusting the learning strategy, can better capture invariant brain patterns. In contrast, traditional machine learning struggles to adapt to subject-independent tasks.

The results of the cross-time test for wearable P300 EEG signals are presented in Fig. 4. The average classification accuracy of the cross-time test is slightly lower than that of the cross-subject test (75.27% compared to 75.87%). Surprisingly, in cross-time testing, the classification accuracy of S1 is significantly higher than the other four individuals. The ERP signals of S1 and S5 are represented in Fig. 7 and Fig. 9 respectively. In Fig. 8, we calculate the ERP signal correlation matrices for S1 across 10 different time periods. The correlation heatmap indicates that the ERP signals show strong

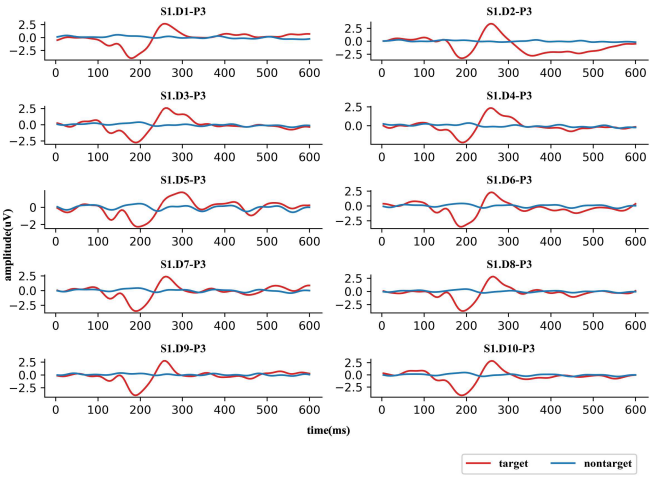

Fig. 7. The ERP waveforms of subject1 from 10 time periods. The red lines represent the time-locked average of all trials for target P300 signals, while the blue lines correspond to the time-locked average of all trials for nontarget P300 signals.

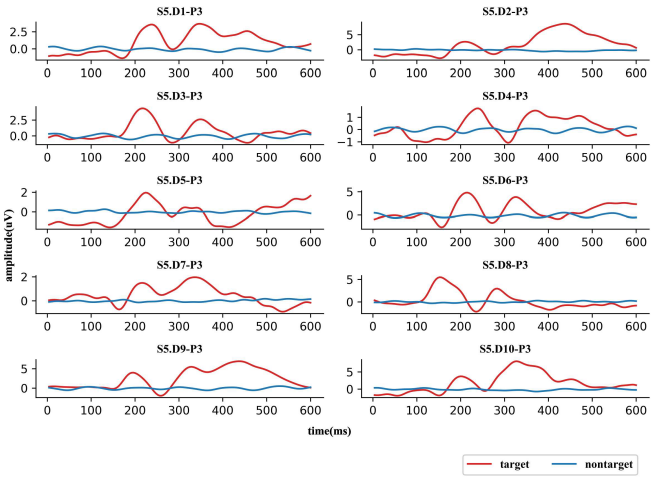

Fig. 9. The ERP waveforms of subject5 from 10 time periods. The red lines are obtained by taking the time-locked average of all trial of target P300 signals, and the blue lines are obtained by taking the time-locked average of all trials for nontarget P300 signal.

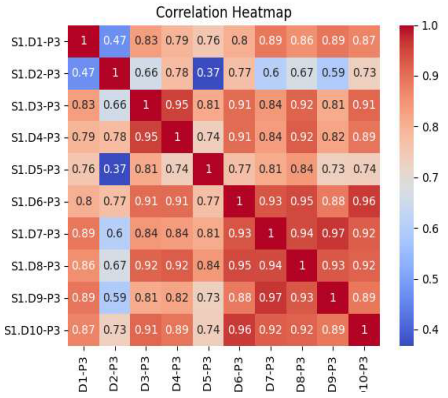

Fig. 8. Correlation heatmap of ERP signals for Subject 1 across 10 different time periods.

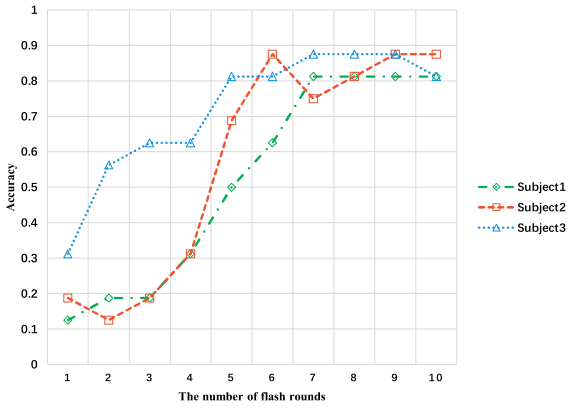

Fig. 10. Effect of the number of flash round on the classification accuracy.

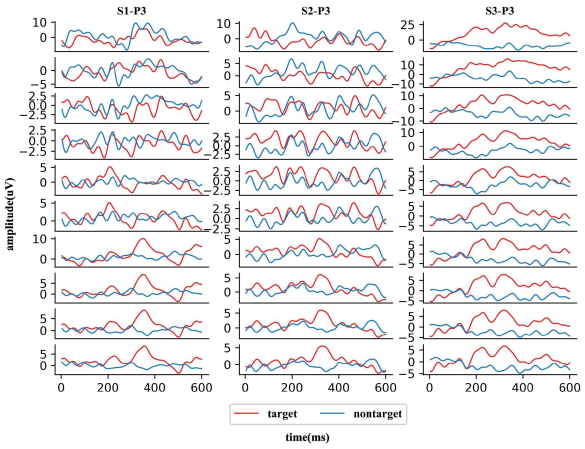

Fig. 11. Target (orange line) and nontarget (blue line) P300 signal waveforms corresponding to three subjects. the target and non-target P300 signal waveforms are obtained from channel P3. The ten EEG signals for each subject in the subplot correspond to stimulus repetition numbers 1 to 10, respectively.

correlations (correlation coefficient $>$ 0.7) across all time periods except D2, suggesting that the ERP signals of S1 have a good consistency in the shape, latency, and amplitude of the waveforms across time periods. On the contrary, in Fig. 9 the ERP waveforms of S5 from D2, D4, D5, D9, and D10 periods exhibit distinct brain patterns. Furthermore, the latency of the ERP signals from D2 and D9 is significantly longer than that of other subjects. Additionally, the waveform amplitude of D4 is noticeably smaller than that of other subjects. This indicates that our model is reliable and capturing subtle changes in the P300 signal. However, further research is needed to understand why the EEG signals of S1 show minimal changes across time.

Furthermore, Fig. 10 illustrates the classification accuracy of three subjects in respect to the number of flashing rounds from the cross-subject experiment. In the initial five flashing rounds, S3 demonstrates notably higher classification accuracy compared to S1 and S2. From the sixth flashing round onward, the classification accuracy changes for all three subjects are not substantial, and the values are quite similar. Fig. 11 depicts the ERP waveforms of three subjects, and we observed that during the initial five flashing rounds, S3 exhibited relatively noticeable waveform distinctiveness. With repetitions of seven, six, and four times, each of the three subjects exhibited distinct P300 EEG waveforms. This suggests the effectiveness of the proposed framework. The CNN-ML model can adeptly capture

the enhancement of P300 signals induced by repeated visual stimuli, thereby improving the real-time performance of the P300 BCI system.

B. Influence of Hyperparameters

Fig. 5 indicates that the mean classification accuracy with respect to training epochs and the training loss curve. As the model's training loss decreases, we observe a corresponding upward trend in the mean classification accuracy. This suggests that our model effectively learns latent brain invariant patterns and enabling accurate classification of P300 EEG signals cross subjects.

Due to the time-consuming and challenging nature of acquiring EEG signal databases, previous research on BCI systems using deep learning could only rely on relatively small subject-dependent datasets. Small datasets are susceptible to model overfitting and exhibit poor generalization ability. In our study, we constructed a large-scale P300 EEG database and attempted to validate how much EEG data is required to develop an acceptable subject-independent P300 deep learning model [23]. Fig. 5 illustrates the test results of models trained on datasets with varying sample sizes. When the training dataset contains 10 samples, the model's average classification accuracy is 46.35%. When the training dataset sample size increases to 60, the model's classification accuracy significantly improves to 70.63%. while the sample size increases from 60 to 90, the model's classification accuracy only increases by 1.87%. We believe that increasing the number of samples can effectively enhance the model's classification accuracy. However, when the sample size reaches a certain point, the improvement in classification accuracy becomes exceedingly limited. Although more samples may potentially provide additional information, achieving this may require more complex models, posing challenges to system hardware and real-time performance. Therefore, it is essential to innovate in capturing the invariant representation of P300 EEG signals while increasing the sample size to obtain a better subject-independent P300 BCI model.

The impact of the number of feature maps and kernel size of the CNN model on the classification accuracy of P300 signals is presented in Table III and Table IV. The number of feature maps needs to be matched with the sample size. Too many feature maps require more computational resources, which is not friendly for wearable devices. Smaller convolutional kernels lead to higher classification accuracy, but require more training time.

C. Analysis of the Comparison Performance of Methods

The classification accuracies of P300 EEG signals based on several models are listed in Fig. 12. All models utilized the same training and testing datasets, and we observe that the average classification accuracy of deep learning models is slightly higher than that of SVM. This also confirms the earlier statement that deep learning methods have greater potential than traditional machine learning methods in calibration free tasks. Furthermore, we observed that the performance of the CNN-ML method is significantly better than other baseline models. This can be explained as the CNN model, through integrating the ML module, effectively learns to classify EEG features and further clusters these features to extract invariant brain patterns. This enables the model to perform better on unseen subject data tests.

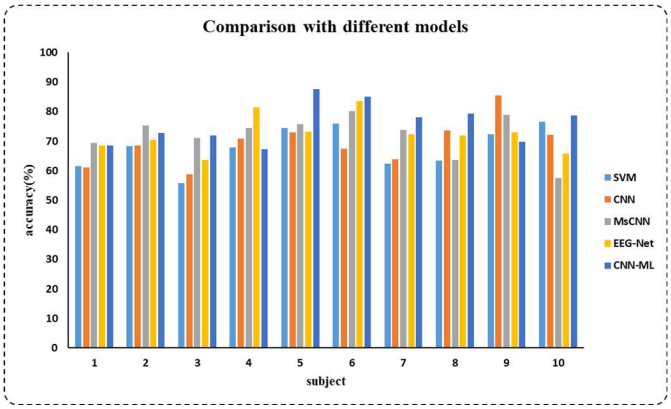

Fig. 12. The classification accuracy with respect to subject under different models.

TABLE V
ONLINE TESTING PERFORMANCE OF SUBJECT-INDEPENDENT WEARABLE BCI SYSTEM

Subject(age)	Accuracy (%)		Mean RT (S)	ITR (bits/min)
S1(23)	normal	68.75	6.7	16.86
	FT	75	7.1	18.69
S2(23)	normal	87.5	5.7	25.08
	FT	81.25	5.4	28.57
S3(30)	normal	62.5	6.3	15.05
	FT	68.75	5.9	19.15
S4(24)	normal	81.25	7.4	20.85
	FT	81.25	6.9	22.36
S5(20)	normal	75	5.8	22.88
	FT	81.25	7.2	21.43
S6(26)	normal	75	6.4	20.74
	FT	75	6.3	21.07
S7(25)	normal	68.75	6.1	18.52
	FT	75	6.9	19.23
S8(28)	normal	62.5	5.3	17.89
	FT	81.25	5.8	26.6
S9(32)	normal	68.75	6.2	18.22
	FT	75	5.4	24.58
S10(23)	normal	81.25	6.9	22.36
	FT	93.75	7.4	27.72
Mean ± SD	normal	73.23 ± 7.62	6.28 ± 0.58	19.85 ± 2.91
	FT	78.75 ± 6.37	6.43 ± 0.72	22.94 ± 3.51

RT: response time for selecting a single target, FT: with fine-tuning

D. Analysis of Online Test

Table V show the detailed online classification accuracy, response time, and information transfer rate of all subject in the experiment. The proposed model achieved an acceptable result. The response times ranged from 5.3 to 7.4 s, and an average accuracy of 73.23%±7.62% was achieved, leading to an average ITR of 19.85±2.91 bits/min. The minimal and maximal ITRs for all subjects were 15.05 bits/min and 22.88 bits/min. Table IV also illustrates the impact of five fine-tuning iterations on classification accuracy, response

time, and information transfer rate. We observed that after 1.5 minutes of calibration, the model's mean accuracy classification and average ITR are increased by 5.52% and 3.09 bits/min respectively. Specifically, the response times is between 5.4 and 7.2 seconds, with an average accuracy of 78.75%±6.37% and an average ITR of 22.94±3.51 bits/min. The minimal and maximal ITRs for all subjects were 18.69 bits/min and 28.57 bits/min, respectively. We believe that fine-tuning methods are equally effective for EEG signals. Adjusting the model's output layer with a small amount of test data can rapidly adapt the pre-trained model to an unknown domain. This approach not only improves classification accuracy but also reduces the user's calibration waiting time. It is a feasible method to implement a wearable P300 BCI system without calibration [29].

E. Limitations and Future Study

Although CNN-ML model shows acceptable performance, the present study has three main limitations. First, this model is designed specifically for processing P300 signals and cannot extract potential invariant brain patterns from raw EEG signals. Second, the standard deviation of classification accuracy is relatively high, indicating that further improvement is needed to enhance the robustness of the model [36]. Third, the loss functions in metric learning still require further optimization [37]. Further research on wearable P300 BCI system is necessary to develop innovative and stable invariant brain patterns. Moreover, the exploration of new loss functions holds great appeal in this context [38].

V. Conclusion

In this study, we constructed a large-scale wearable P300 EEG dataset and proposed a framework CNN-ML for wearable subject-independent P300 BCI system. Initially, feature extraction is applied to EEG signals to enhance their distribution consistency. Subsequently, the EEG features are projected into an embedding space through a CNN model. Metric learning optimizes the parameters of the CNN model based on the distance of brain patterns. iteratively refining the embedding space to approach the representation of invariant brain patterns. Ultimately, the CNN model learns features that can generalize to unseen domains and improves the model's classification accuracy through fine-tuning, thereby shortening or eliminating the calibration process in the wearable P300 BCI system. The results demonstrated that our approach significantly improved the classification accuracy of the subject-independent wearable P300 BCI system, achieving an accuracy of 73.23±7.62% without calibration and 78.75±6.37% with fine-tuning. The promising experimental results from our study indicates that this framework can be utilized to develop wearable BCI system without calibration and apply them to real-world scenarios.

References

[1] D. J. McFarland and J. R. Wolpaw, "Brain–computer interfaces for communication and control," *Commun. ACM*, vol. 54, no. 5, pp. 60–66, May 2011.

[2] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, "Brain–computer interfaces for communication and rehabilitation," *Nature Rev. Neurol.*, vol. 12, no. 9, pp. 513–525, Sep. 2016.

[3] R. Fazel-Rezai, B. Z. Allison, C. Guger, E. W. Sellers, S. C. Kleih, and A. Kübler, "P300 brain computer interface: Current challenges and emerging trends," *Frontiers Neuroeng.*, vol. 5, Jul. 2012, Art. no. 14.

[4] W. Zhang, C. Tan, F. Sun, H. Wu, and B. Zhang, "A review of EEG-based brain–computer interface systems design," *Brain Sci. Adv.*, vol. 4, no. 2, pp. 156–167, Dec. 2018.

[5] Y. Yu et al., "Self-paced operation of a wheelchair based on a hybrid brain–computer interface combining motor imagery and P300 potential," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 12, pp. 2516–2526, Dec. 2017.

[6] H. Wang, Y. Li, J. Long, T. Yu, and Z. Gu, "An asynchronous wheelchair control by hybrid EEG-EOG brain–computer interface," *Cogn. Neurodynamics*, vol. 8, no. 5, pp. 399–409, Oct. 2014.

[7] M. Xu, F. He, T.-P. Jung, X. Gu, and D. Ming, "Current challenges for the practical application of electroencephalography-based brain–computer interfaces," *Engineering*, vol. 7, no. 12, pp. 1710–1712, Dec. 2021.

[8] S. Saha et al., "Progress in brain computer interface: Challenges and opportunities," *Frontiers Syst. Neurosci.*, vol. 15, Feb. 2021, Art. no. 75.

[9] J. A. Ramirez-Quintana, L. Madrid-Herrera, M. I. Chacon-Murguia, and L. F. Corral-Martinez, "Brain–computer interface system based on P300 processing with convolutional neural network, novel speller, and low number of electrodes," *Cognit. Comput.*, vol. 13, no. 1, pp. 108–124, Jan. 2021.

[10] R. Zhang et al., "A BCI-based environmental control system for patients with severe spinal cord injuries," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1959–1971, Aug. 2017.

[11] J. Pan et al., "A hybrid brain–computer interface combining P300 potentials and emotion patterns for detecting awareness in patients with disorders of consciousness," *IEEE Trans. Cogn. Develop. Syst.*, vol. 15, no. 3, pp. 1386–1395, Sep. 2023.

[12] S. Kundu and S. Ari, "MsCNN: A deep learning framework for P300-based brain–computer interface speller," *IEEE Trans. Med. Robot. Bionics*, vol. 2, no. 1, pp. 86–93, Feb. 2020.

[13] H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain–computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 433–445, Mar. 2011.

[14] O.-Y. Kwon, M.-H. Lee, C. Guan, and S.-W. Lee, "Subject-independent brain–computer interfaces based on deep convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 10, pp. 3839–3852, Oct. 2020.

[15] K. Vo, T. Pham, D. N. Nguyen, H. H. Kha, and E. Dutkiewicz, "Subject-independent ERP-based brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 4, pp. 719–728, Apr. 2018.

[16] M. M. Krell, A. Seeland, and S. Kim, "Data augmentation for brain–computer interfaces: Analysis on event-related potentials data," 2018, *arXiv:1801.02730*.

[17] K. Kunanbayev, B. Abibullaev, and A. Zollanvari, "Data augmentation for P300-based brain–computer interfaces using generative adversarial networks," in *Proc. 9th Int. Winter Conf. Brain-Computer Interface (BCI)*, Feb. 2021, pp. 1–7.

[18] H. Woehrle, M. M. Krell, S. Straube, S. K. Kim, E. A. Kirchner, and F. Kirchner, "An adaptive spatial filter for user-independent single trial detection of event-related potentials," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 7, pp. 1696–1705, Jul. 2015.

[19] Z. Huang, J. Guo, W. Zheng, Y. Wu, Z. Lin, and H. Zheng, "A calibration-free approach to implementing P300-based brain–computer interface," *Cogn. Comput.*, vol. 14, no. 2, pp. 887–899, Jan. 2022.

[20] A. Mussabayeva, P. K. Jamwal, and M. T. Akhtar, "Ensemble voting-based multichannel EEG classification in a subject-independent P300 speller," *Appl. Sci.*, vol. 11, no. 23, Nov. 2021, Art. no. 11252.

[21] P. Autthasan et al., "MIN2Net: End-to-end multi-task learning for subject-independent motor imagery EEG classification," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 6, pp. 2105–2118, Jun. 2022.

[22] L. Wang et al., "Domain-invariant feature exploration for domain generalization," *Trans. Mach. Learn. Res.*, pp. 1–20, Jul. 2022.

[23] W. Gao et al., "Eliminating or shortening the calibration for a P300 brain–computer interface based on a convolutional neural network and big electroencephalography data: An online study," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1754–1763, 2023.

[24] L. Hu, J. Zhu, S. Chen, Y. Zhou, Z. Song, and Y. Li, "A wearable asynchronous brain–computer interface based on EEG-EOG signals with fewer channels," *IEEE Trans. Biomed. Eng.*, vol. 71, no. 2, pp. 504–513, Feb. 2024.

[25] C. Guan, M. Thulasidas, and J. Wu, "High performance P300 speller for brain–computer interface," in *Proc. IEEE Int. Workshop Biomed. Circuits Syst.*, Singapore, Dec. 2004, pp. 3–5.

[26] M. Ghifary, D. Balduzzi, W. B. Kleijn, and M. Zhang, "Scatter component analysis: A unified framework for domain adaptation and domain generalization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1414–1430, Jul. 2017.

[27] S. Motiian, M. Piccirilli, D. A. Adjeroh, and G. Doretto, "Unified deep supervised domain adaptation and generalization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5716–5726.

[28] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2006, pp. 1735–1742.

[29] P. Izmailov et al., "Averaging weights leads to wider optima and better generalization," in *Proc. Conf. Uncertainty Artif. Intell.*, 2018.

[30] A. Ramé, M. Kirchmeyer, T. Rahier, A. Rakotomamonjy, P. Gallinari, and M. Cord, "Diverse weight averaging for out-of-distribution generalization," 2022, *arXiv:2205.09739*.

[31] Y. Li, C. Guan, H. Li, and Z. Chin, "A self-training semi-supervised SVM algorithm and its application in an EEG-based brain computer interface speller system," *Pattern Recognit. Lett.*, vol. 29, no. 9, pp. 1285–1294, Jul. 2008.

[32] Y. Li and C. Guan, "A semi-supervised SVM learning algorithm for joint feature extraction and classification in brain computer interfaces," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 113, Aug. 2006, pp. 2570–2573.

[33] M. Liu, W. Wu, Z. Gu, Z. Yu, F. Qi, and Y. Li, "Deep learning based on batch normalization for P300 signal detection," *Neurocomputing*, vol. 275, pp. 288–297, Jan. 2018.

[34] H. Shan, Y. Liu, and T. Stefanov, "A simple convolutional neural network for accurate P300 detection and character spelling in brain computer interface," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 1604–1610.

[35] F. Fahimi, Z. Zhang, W. B. Goh, T.-S. Lee, K. K. Ang, and C. Guan, "Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI," *J. Neural Eng.*, vol. 16, no. 2, Apr. 2019, Art. no. 026007.

[36] P.-J. Kindermans, M. Tangermann, K.-R. Müller, and B. Schrauwen, "Integrating dynamic stopping, transfer learning and language models in an adaptive zero-training ERP speller," *J. Neural Eng.*, vol. 11, no. 3, Jun. 2014, Art. no. 035005.

[37] K. Muandet, D. Balduzzi, and B. Schölkopf, "Domain generalization via invariant feature representation," in *Proc. Int. Conf. Mach. Learn.*, 2013.

[38] J. Huang et al., "Cross-subject transfer method based on domain generalization for facilitating calibration of SSVEP-based BCIs," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 3307–3319, 2023.

[39] I. A. Fouad, F. E.-Z.-M. Labib, M. S. Mabrouk, A. A. Sharawy, and A. Y. Sayed, "Improving the performance of P300 BCI system using different methods," *Netw. Model. Anal. Health Informat. Bioinf.*, vol. 9, no. 1, pp. 1–13, Sep. 2020.

[40] Siddharth, A. N. Patel, T.-P. Jung, and T. J. Sejnowski, "A wearable multi-modal bio-sensing system towards real-world applications," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 4, pp. 1137–1147, Apr. 2019.

[41] 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," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.