Channel Reflection: Knowledge-Driven Data Augmentation for EEG-Based Brain-Computer Interfaces

Ziwei Wang[a,b,], Siyang Li[a,b,], Jingwei Luo[c], Jiajing Liu[d], Dongrui Wu[a,b,]

[a]*Key Laboratory of the Ministry of Education for Image Processing and Intelligent Control, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China*
[b]*Shenzhen Huazhong University of Science and Technology Research Institute, Shenzhen 518063, China*
[c]*China Electronic System Technology Co., Ltd., Beijing 100089, China*
[d]*School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China*

Abstract

A brain-computer interface (BCI) enables direct communication between the human brain and external devices. Electroencephalography (EEG) based BCIs are currently the most popular for able-bodied users. To increase user-friendliness, usually a small amount of user-specific EEG data are used for calibration, which may not be enough to develop a pure data-driven decoding model. To cope with this typical calibration data shortage challenge in EEG-based BCIs, this paper proposes a parameter-free channel reflection (CR) data augmentation approach that incorporates prior knowledge on the channel distributions of different BCI paradigms in data augmentation. Experiments on eight public EEG datasets across four different BCI paradigms (motor imagery, steady-state visual evoked potential, P300, and seizure classifications) using different decoding algorithms demonstrated that: 1) CR is effective, i.e., it can noticeably improve the classification accuracy; 2) CR is robust, i.e., it consistently outperforms existing data augmentation approaches in the literature; and, 3) CR is flexible, i.e., it can be combined with other data augmentation approaches to further increase the performance. We suggest that data augmentation approaches like CR should be an essential step in EEG-based BCIs. Our code is available online.

Keywords: Brain-computer interface, electroencephalogram, informed machine learning, integration of data and knowledge, data augmentation

1. Introduction

A brain-computer interface (BCI) enables direct communication between the human brain and an external device (Rosenfeld and Wong, 2017). BCIs serve diverse purposes, encompassing research, mapping, augmentation, assistance, and restoration of human cognitive and/or sensory-motor functions (Krucoff et al., 2016).

According to the proximity of the electrodes to the brain cortex, BCIs can be categorized into three types: non-invasive, partially invasive, and invasive (Wu et al., 2020). The latter two require the surgical placement of sensors, and hence are predominantly employed in clinical applications. Non-invasive BCIs are the preferred choice for able-bodied individuals. They could use different input signals, e.g., electroencephalography (EEG), magnetoencephalography, functional magnetic resonance imaging, and functional near-infrared spectroscopy. Among them, EEG stands as the most popular choice due to its convenience and cost-effectiveness. Classical paradigms of EEG-based non-invasive BCIs include motor imagery (MI) (Pfurtscheller and Neuper,
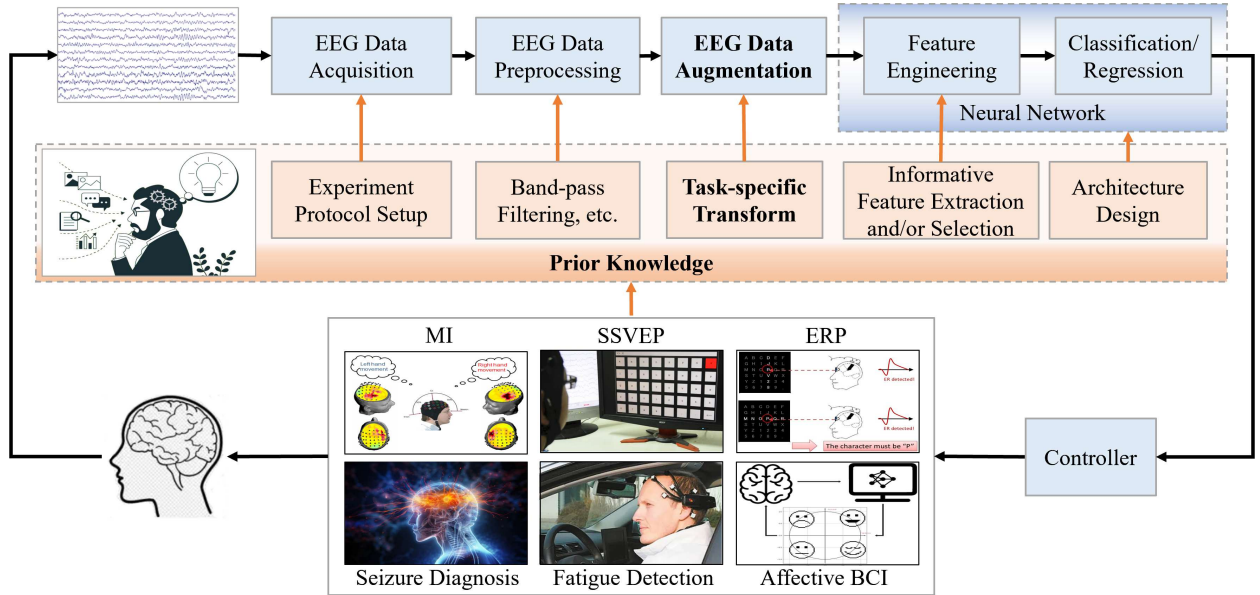
Figure 1: Knowledge-driven machine learning pipeline for EEG-based BCIs, which includes EEG data acquisition, data preprocessing, data augmentation (optional), feature engineering, and classification/regression. The latter two can be integrated into a single end-to-end neural network. Task-related prior knowledge can be integrated into all blocks of the pipeline.

2001), steady-state visual evoked potential (SSVEP) (Friman et al., 2007), and event-related potential (ERP) (Hoffmann et al., 2008). Additionally, BCIs have also been used for epileptic seizure detection (Acharya et al., 2013), driver-drowsiness estimation (Lin et al., 2005), emotion analysis (Wu et al., 2023), and so on.

The development of BCIs heavily relies on comprehending how the human brain functions. The 'homunculus' model from the 20th century confirmed the presumed relationship between each part of the human body and a corresponding region in the primary motor/somatosensory areas of the neocortex (Jasper and Penfield, 1949), which becomes the basis for MI-based BCIs (Pfurtscheller and Neuper, 2001). Through differentiating the patterns in brain signals of imaged movements of different body parts, MI-based BCIs can assist, augment, or even repair human sensory-motor functions in various applications (Krucoff et al., 2016). The mechanism relies on the neuroscience discovery that processing of motor commands or somatosensory stimuli causes an attenuation of the rhythmic activity termed event-related desynchronization, while an increase is termed as event-related synchronization (Blankertz et al., 2008).

Similarly, other paradigms also follow specific neuroscience discoveries. SSVEPs are electrical brain responses that occur in synchrony with repetitive visual stimuli, such as flashing lights or flickering images, typically elicited in the visual cortex of the brain, specifically in the occipital region located at the back of the head (Friman et al., 2007). ERPs are triggered by specific events or stimuli, offering valuable insights into cognitive processes, such as attention, memory, and perception. The P300 ERP is distinguished by a prominent positive deflection within the EEG signal, typically ~300 milliseconds after the presentation of a sensory stimulus (Hoffmann et al., 2008). The P300 ERP is prominently associated with neural activity in the parietal cortex, particularly in the parietal midline (centroparietal) region (Abiri et al., 2019).

EEG-based seizure classification emerges as a pivotal focus within the field of neurology, with positive implications for patients affected by epilepsy (Acharya et al., 2013). Typically, seizures can originate in various brain

2

regions, including the temporal, frontal, parietal and occipital lobes. Seizure activity can spread from the onset zone to other brain parts. The path and extent of propagation can vary, leading to different types of seizures and clinical manifestations.

Accurate brain signal decoding is critical for successful BCI applications. Although EEG-based BCIs have various advantages as discussed, there are still many challenges for their wide-spread real-world applications (Lance et al., 2012; Makeig et al., 2012), including non-stationarity of EEG signals, individual differences, and inter-environment differences. The last refers to the differences in EEG headsets and experiment protocols across datasets. Such differences limit the current analysis of EEG within a given paradigm and dataset, limiting the number of training samples.

Data augmentation is the most commonly used strategy to cope with the small data problem (Shorten and Khoshgoftaar, 2019; Li et al., 2022). For EEG signal analysis, approaches from both signal processing and deep learning have been attempted. Nevertheless, such approaches are usually based on the characteristics of time series, EEG signals, or neural networks in general. Without adequately utilizing the paradigm-specific knowledge, such approaches usually have unstable performance and are not generalizable across paradigms and datasets. A simple, effective and generalizable data augmentation approach remains to be found. Notably, the relationships among paradigms are complex yet vital. Delving into the connections among channels, especially in different brain regions, is essential for comprehending the brain and contributing to constructing machine learning models.

The role of prior knowledge has received much attention, particularly for interpretable and explainable models (Lisboa et al., 2023). von Rueden et al. (2021) pointed out that "*Despite its great success, machine learning can have its limits when dealing with insufficient training data. A potential solution is the additional integration of prior knowledge into the training process*." In EEG-based BCIs, solely data-driven machine learning approaches are becoming the current trend for signal decoding. Nevertheless, effective integration of prior knowledge can profoundly influence model performance at various stages, as illustrated in Figure 1.

This paper proposes a straightforward yet effective knowledge-driven channel reflection (CR) data augmentation approach to generate high-quality task-specific augmented data, enriching the training dataset without introducing additional hyperparameters. Inspired by the left/right hand MI paradigm design, CR constructs new samples by reflecting left and right brain electrodes/channels of EEG signals and simultaneously exchanging the labels. For other BCI paradigms, CR assumes task-invariability when reflecting the hemispheres and sticks with the original labels. Extensive experiments on eight public EEG datasets demonstrated that CR could effectively improve performance on four BCI paradigms, namely MI, SSVEP, P300, and seizure classification. Our Python code is available on GitHub[1].

The remainder of this paper is organized as follows: Section 2 introduces related works. Section 3 proposes CR. Section 4 presents experimental results to show the effectiveness of CR. Finally, Section 5 draws conclusions and points out future research directions.

2. Related Works

This section introduces related works on integrating prior knowledge into machine learning in EEG-based BCIs, and discusses current data augmentation approaches for EEG classification.

2.1. Integration of Prior Knowledge in EEG-based BCIs

Previous works have demonstrated the power of prior knowledge in various stages in brain signal decoding:

1. Data Preprocessing. Filtering and denoising procedures (Pedroni et al., 2019), vital for data quality enhancement, depend extensively on the specific characteristics and requirements of the dataset. Prior knowledge aids in the development of robust data preprocessing methods, ensuring that the data fed into the model is clean, reliable, and aligned with the objectives of the task. As an example, the passband of filters differs across BCI paradigms.

2. Feature Engineering. Features can be extracted in the time domain (Boonyakitanont et al., 2020), frequency domain, time-frequency domain, etc

---

[1]https://github.com/sylyoung/DeepTransferEEG

3

(Jenke et al., 2014). However, without integration of task specific knowledge, general features for time series would not work well for a particular BCI paradigm. For example, in EEG-based BCIs, selecting features or spatial patterns relies on prior knowledge of neurophysiological processes and specific cognitive tasks. For MI-based BCIs, Common Spatial Pattern (CSP) (Ramoser et al., 2000) is the most widely used supervised spatial filter. It aims to find a set of spatial filters to maximize the ratio of variance between two different classes, i.e., left hand and right hand imaginations (Wu et al., 2022). CSP is motivated by the neuroscience discovery that motor activities, both actual and imagined, modulate the μ-rhythm (Blankertz et al., 2008), and can trigger noticeable pattern changes in EEG from different brain regions.

3. Classification. Representative neural networks for EEG signal classification, e.g., EEGNet (Lawhern et al., 2018), EEGWaveNet (Thuwajit et al., 2022), and SSVEPNet (Pan et al., 2022), are motivated by our understanding of the corresponding BCI paradigm. For example, EEGNet approximates CSP in designing its convolution layers.

2.2. EEG Data Augmentation

Existing EEG data augmentation approaches mainly consider time, frequency, and/or spatial domain transformations.

For time domain transformation, Wang et al. (2018) added random Gaussian white noise to the original signals; Mohsenvand et al. (2020) set a random portion of the EEG signal to zero; and, Rommel et al. (2022) randomly flipped the signals or reversed the axis of time of all channels.

For frequency domain transformation, Schwabedal et al. (2018) randomized the phases of Fourier transforms of all channels; Mohsenvand et al. (2020) and Cheng et al. (2020) randomly filtered a narrow frequency band of all channels; and, Rommel et al. (2022) randomly shifted all channels' power spectral density by a small value.

For spatial domain transformation, Saeed et al. (2021) set the values of some randomly picked channels to zero,

or performed random permutation, and Krell and Kim (2017) interpolated channels on randomly rotated positions.

Deep learning techniques, e.g., generative adversarial networks (Luo and Lu, 2018), have also been used for EEG data augmentation (Lashgari et al., 2020).

However, existing EEG data augmentation approaches integrated little task-related knowledge into the data transformation process, which is the challenge to be solved by this paper. The data augmentation approach most similar to ours was Deiss et al. (2018), which exchanged the left and right hemisphere channels. Section 4.2 gives detailed comparisons.

3. Channel Reflection for EEG Data Augmentation

Assume the training data include m labeled EEG trials $\{(X^i, y^i)\}_{i=1}^m$, where $X^i \in \mathbb{R}^{C \times T}$ is the i-th EEG trial (C is the number of EEG channels, and T the number of time samples), and $y^i \in \{0, 1\}$ the corresponding label. Let $X_c^i \in \mathbb{R}^T$ be the c-th channel of X^i, where $c \in \mathbb{N}_C = \{1, 2, ..., C\}$. Then, X^i can also be denoted as $X^i = [X_1^i, X_2^i, ..., X_C^i]$.

For simplicity, we assume exact symmetric placement of electrodes, i.e., K channels with indices $L = \{L_1, ..., L_K\} \subset \mathbb{N}_C$ are placed on the left hemisphere, and K channels with indices $R = \{R_1, ..., R_K\} \subset \mathbb{N}_C$ on the right hemisphere, where L_k and R_k ($k = 1, ..., K$) are symmetrical electrodes on the left and right hemisphere, respectively.

CR constructs new training trials by exchanging the symmetrical left and right hemisphere channels. More specifically, the left and right hemisphere channels are exchanged, while the middle line channels stay fixed, i.e., the c-th channel of the transformed trial $\tilde{X}^i$ becomes

$$\tilde{X}_c^i = \begin{cases} X_{R_k}^i, & c = L_k \\ X_{L_k}^i, & c = R_k \\ X_c^i, & c \notin L \cup R \end{cases} . \qquad (1)$$

The label of $\tilde{X}^i$ is modified in MI classification, but stay identical to the label of X^i for other BCI paradigms, i.e.,

$$\tilde{y}^i = \begin{cases} 1 - y^i, & \text{left/right hand MI classification} \\ y^i, & \text{other BCI paradigms} \end{cases} \qquad (2)$$
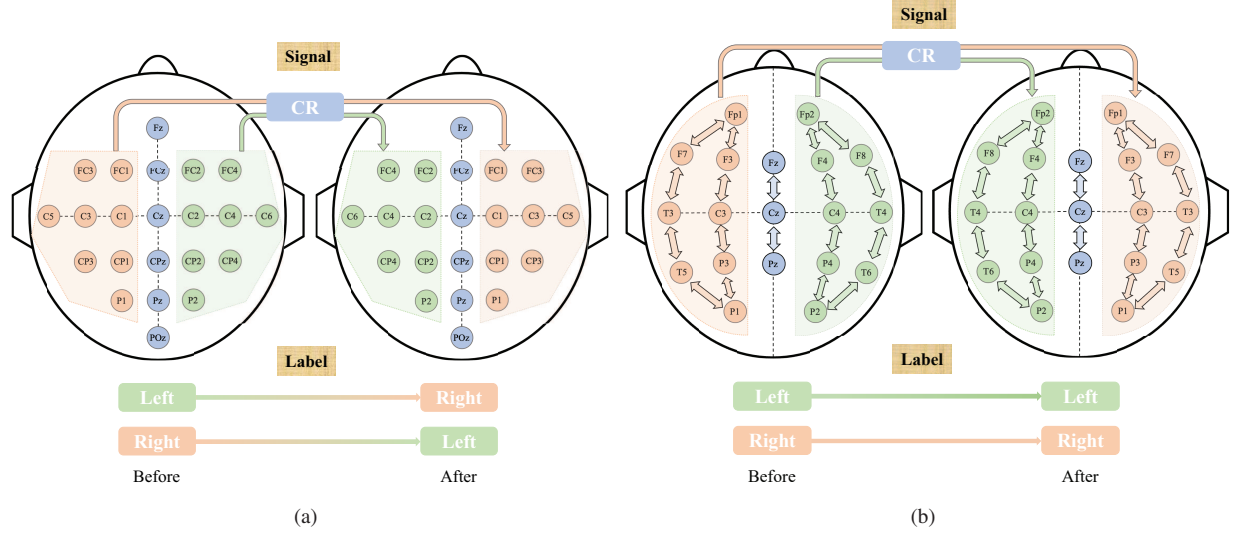
Figure 2: CR data augmentation for (a) unipolar channels, using MI-I dataset (Tangermann et al., 2012) as an example; and, (b) bipolar channels, using Seizure-I dataset (Stevenson et al., 2019) as an example.

Figure 2 illustrates the details of CR augmentation in MI, SSVEP and P300 paradigms with unipolar channels, and seizure classification with bipolar channels. A unipolar channel [Figure 2(a)] measures the potential difference between an electrode and a common reference, whereas a bipolar channel [Figure 2(b)] outputs the potential difference between two adjacent channels (Yao et al., 2019). CR augmentation flips all unipolar or bipolar channels from left to right, and right to left.

Figure 3 shows the flowchart of using CR in a closed-loop EEG-based BCI system, by extending the cross-subject transfer learning pipeline proposed in Wu et al. (2022). After CR, the training data become the combination of the original EEG trials and the augmented EEG trials.

4. Experiments and Results

Extensive experiments were performed to validate the superior performance of CR. This section presents the experiment settings, results and analyses.

4.1. Datasets

Eight datasets from four different BCI paradigms were used to verify the effectiveness of CR:
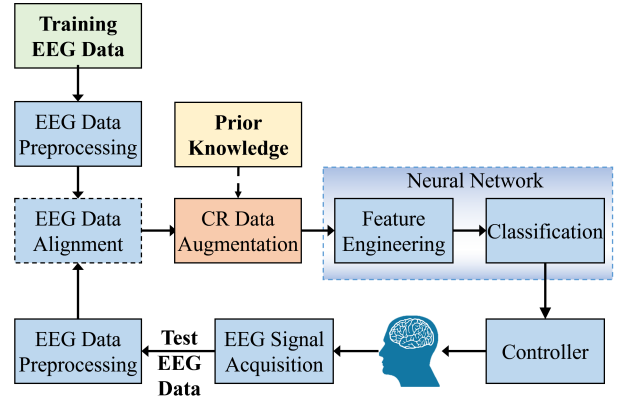

Figure 3: Closed-loop EEG-based BCI system, including the proposed CR data augmentation block.

Table 1: Summary of the eight EEG datasets from MI, SSVEP, P300, and seizure classification paradigms.

Dataset	Number of Subjects	Number of EEG Channels	Sampling Rate (Hz)	Trial Length (seconds)	Number of Total Trials	Task Types
MI-I	9	22	250	4	1,296	left/right hand
MI-II	9	3	250	4.5	1,080	left/right hand
MI-III	7	59	250	3	1,400	left/right hand, or feet/left hand
SSVEP	10	8	256	1	1,800	12 different class
P300-I	8	8	256	1	29,400	target/non-target
P300-II	10	16	256	0.8	17,280	target/non-target
Seizure-I	39	18	256	4	52,534	seizure/normal
Seizure-II	27	18	500 or 1000	4	21,237	seizure/normal

1. MI: Three datasets, namely MI-I (Tangermann et al., 2012) and MI-II (Leeb et al., 2007) from MOABB (Jayaram and Barachant, 2018), and MI-III from BCI Competition IV-1 (Blankertz et al., 2007), were used.

2. SSVEP: The dataset in Nakanishi et al. (2015) was used, and is simply referred to as the SSVEP dataset.

3. P300: Two EEG-based P300 datasets, P300-I (Riccio et al., 2013) and P300-II (Aricò et al., 2014), were used.

4. Seizure classification: Seizure-I (Stevenson et al., 2019) and Seizure-II (Wang et al., 2023) datasets were used.

Their characteristics are summarized in Table 1. Table 2 also shows the channel locations for each dataset.

The following data preprocessing procedures were used:

1. MI-I and MI-II: The standard preprocessing steps in MOABB, including notch filtering, band-pass filtering, etc., were followed to ensure the reproducibility. For MI-I, only two classes (left/right hand MIs) were used. For both datasets, only EEG trials from the first session of each subject were used for training and test.

2. MI-III: The EEG data were first [8, 30] Hz band-pass filtered, and then downsampled to 250 Hz to match the other two datasets. Note that Subjects S0 and S5 in MI-III conducted feet/left hand tasks instead of left/right hand tasks; so, their training data were not CR augmented. Again, only EEG trials from the first session of each subject were used for training and test.

3. SSVEP: We followed the preprocessing steps in Pan et al. (2022). All signals were first downsampled to 256 Hz and [6, 80] Hz band-pass filtered via fourth-order forward-backward Butterworth band-pass filter, and then split into 1-second length trials.

4. P300-I and P300-II: We followed the standard preprocessing steps in MOABB, including band-pass, high-pass and low-pass filtering, etc. Only the first session of P300-I was used, whereas all three sessions of P300-II were used, due to its smaller size.

5. Seizure-I and Seizure-II: We followed the preprocessing steps in (Wang et al., 2023). Each bipolar EEG channel was preprocessed by a 50 Hz notch filter and a [0.5,50] Hz band-pass filter. EEG signals of Seizure-II were further downsampled to 500 Hz. The signals were then segmented into 4-second long non-overlapping trials.

4.2. Algorithms

CR was compared with five popular data augmentation baselines. Four of them were described in Freer and Yang (2020), which were also used in Zhang et al. (2022). The remaining one was proposed in Deiss et al. (2018).

1. No augmentation (Baseline), which does not use any data augmentation.

2. Noise adding (Noise), which adds uniform noise to an EEG trial.

Table 2: Channel locations of the eight datasets.

Dataset	Left Hemisphere	Right Hemisphere
MI-I	[FC3, FC1, C5, C3, C1, CP3, CP1, P1]	[FC4, FC2, C6, C4, C2, CP4, CP2, P2]
MI-II	[C3]	[C4]
MI-III	[AF3, F5, F3, F1, FC5, FC3, FC1, CFC7, CFC5, CFC3, CFC1, T7, C5, C3, C1, CCP7, CCP5, CCP3, CCP1, CP5, CP3, CP1, P5, P3, P1, PO1, O1]	[AF4, F6, F4, F2, FC6, FC4, FC2, CFC8, CFC6, CFC4, CFC2, T8, C6, C4, C2, CCP8, CCP6, CCP4, CCP2, CP6, CP4, CP2, P6, P4, P2, PO2, O2]
SSVEP	[P7, P3, O1]	[P8, P4, O2]
P300-I	[P3, PO7]	[P4, PO8]
P300-II	[F3, C3, CP3, P3, PO7]	[F4, C4, CP4, P4, PO8]
Seizure-I	[Fp1-F3, F3-C3, C3-P3, P3-O1, Fp1-F7, F7-T3, T3-T5, T5-O1]	[Fp2-F4, F4-C4, C4-P4, P4-O2, Fp2-F8, F8-T4, T4-T6, T6-O2]
Seizure-II	[Fp1-F3, F3-C3, C3-P3, P3-O1, Fp1-F7, F7-T3, T3-T5, T5-O1]	[Fp2-F4, F4-C4, C4-P4, P4-O2, Fp2-F8, F8-T4, T4-T6, T6-O2]

3. Data flipping (Flip), which flips the amplitude of an EEG trial.
4. Data multiplication (Scale), which scales the amplitude of an EEG trial by a coefficient close to 1.
5. Frequency shift (Freq), which uses Hilbert transform (Freeman, 2007) to shift the frequency of an EEG trial.
6. Channel symmetry (Symm), which reflects left and right hemisphere channels.

Symm seems similar to our proposed CR, but there are two differences:

1. Symm does not alter the labels, which is crucial for prior knowledge guided data augmentations. For paradigms that are relying detecting event-related desynchronization/synchronization in the left/right hemisphere, the operations of Symm and CR are different.
2. Symm does not explicitly require the left and right hemisphere channels to be strictly symmetric, whereas CR does.

Table 3 shows the details of the seven approaches. $C_{\text{noise}} = 2$, $C_{\text{Scale}} = 0.05$ and $C_{\text{freq}} = 0.2$ were used, following Freer and Yang (2020). Note that CR does not require any hyperparameters.

4.3. Experiment Settings

EEG signals usually exhibit large inter-subject variations (individual differences), and are non-stationary. As a result, collecting labeled calibration data from the target user is generally required for satisfactory classification accuracy (Wu et al., 2022). Three experiment settings with different amounts of training data were used, as illustrated in Figure 4:
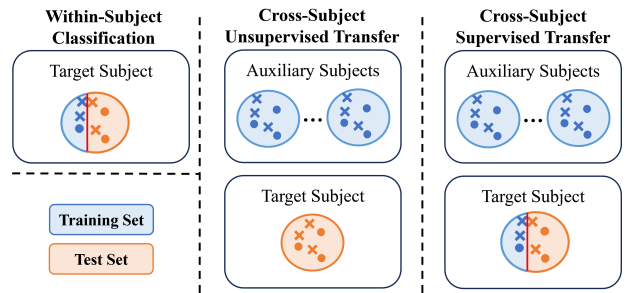

Figure 4: Illustration of the three experiment scenarios.

1. Within-subject classification, where the training set contains only a few labeled data from the target subject, but no data from the source subjects were used.

Table 3: Comparison of different data augmentation strategies.

Strategy	Formulation	Label	Hyper-parameters
Noise	$\tilde{X}_c = X_c + rand * std(X_c)/C_{\text{noise}}$	Fixed	$C_{\text{noise}} = 2$
Flip	$\tilde{X}_c = \max(X_c) - X_c$	Fixed	–
Scale	$\tilde{X}_c = X_c * (1 \pm C_{\text{scale}})$	Fixed	$C_{\text{scale}} = 0.05$
Freq	$\tilde{X}_c = F_{\text{shift}}(X_c, \pm C_{\text{freq}})$	Fixed	$C_{\text{freq}} = 0.2$
Symm	$\tilde{X}_c = \begin{cases} X_{c'}^i, & c \in L, c' \in R \\ X_{c'}^i, & c \in R, c' \in L \\ X_c^i, & c \notin \{L \cup R\} \end{cases}$	Fixed	–
CR (ours)	$\tilde{X}_c = \begin{cases} X_{R_k}^i, & c = L_k \\ X_{L_k}^i, & c = R_k \\ X_c^i, & c \notin \{L \cup R\} \end{cases}$	Exchanged (based on task)	–

2. Cross-subject unsupervised transfer, where the training set contains a decent amount of labeled data from the source subjects, but no labeled data from the target subject.

3. Cross-subject supervised transfer, where the training set contains a large amount of labeled data from the source subjects, and a small amount of labeled data from the target subject.

In all three scenarios, only the training set was used to tune the algorithms, and the test set was inaccessible during the training phase. The labels of the test set were used only during the test phase to compute the performance measures. For each dataset, each subject was treated as the target subject once, and all remaining subjects as the source subjects. For the target subject, different amounts of labeled trials in a continuous block [as in (Li et al., 2021)] were used in training, with the remaining trials in testing.

In all three scenarios, data augmentation was applied to all labeled trials in the training set to generate extra training data. The amount of training data was doubled for Noise, Flip, Symm and CR, and tripled for Scale and Freq.

The raw classification accuracy was used as the performance measure for MI and SSVEP classification, which usually do not have significant class-imbalance. However, significant class-imbalance exists in the other two paradigms, i.e., the non-target class overwhelmed the tar-

get class in P300, and the normal class overwhelmed the seizure class in seizure classification; thus, the raw classification accuracy may be misleading. So, balanced classification accuracy (BCA), defined as the average of recall obtained on each class, was used as the performance measure for P300 and seizure classification.

All experiments using neural networks were repeated five times to accommodate randomness.

4.4. Main Results

4.4.1. MI Classification

For the MI paradigm, the classic CSP (Blankertz et al., 2008) filtering with Linear Discriminant Analysis (LDA) classifier was used in within-subject classification. CSP used ten spatial filters. In cross-subject unsupervised and supervised transfers, EEGNet (Lawhern et al., 2018), a compact convolutional neural network (CNN) with four convolutional layers, was used. In both within-subject classification and cross-subject supervised transfer, the number of labeled trials per class from the target subject increased from 5 to 45 with step 5. The batch size was 32 for Baseline, doubled for Noise, Flip and CR, and tripled for Scale and Freq. All models were trained with 100 epochs using Adam optimizer with learning rate 10^{-3}, identical to those in Li et al. (2024). The final update steps of gradient descent for each evaluation instance stayed the same.

Euclidean Alignment (EA) (He and Wu, 2020), an effective unsupervised EEG data alignment approach for

8

MI (Wu et al., 2022), was applied before all approaches to align the EEG trials for each source subject. For the target subject, the reference matrix of EA was calculated on the labeled trials, then incrementally updated as each new test trial arrived, as in (Li et al., 2024).

Tables 4-6 show the results:

1. Comparing the Baseline performance in the three settings, we can find that when the labeled target data were small, cross-subject unsupervised transfer generally outperformed or performed comparably with within-subject transfer, and cross-subject supervised transfer always outperformed the former two, suggesting the benefits of using source data in transfer learning.

2. The five existing data augmentation approaches did not noticeably improve over Baseline; particularly, Symm significantly degraded the performance, because switching the left/right channels without switching the label contradicts the neuroscience principle of MI.

3. In terms of average performance, our proposed CR always outperformed all other approaches in all three classification scenarios, indicating the effectiveness and robustness of incorporating prior knowledge into data augmentation.

4.4.2. SSVEP Classification

Three different settings were also considered for SSVEP.

SSVEPNet (Pan et al., 2022) trained with 500 epochs and Adam optimizer was used as the classifier. The number of labeled target trials per class increased from 1 to 9 with step 1. For baseline, batch size 30 and learning rate 10^{-2} were used in within-subject classification, and batch size 64 and learning rate 10^{-3} in cross-subject transfers. The batch size was doubled for Noise, Flip and CR, and tripled for Scale and Freq.

Table 7 shows the results:

1. Unlike the case in MI classification, here cross-subject unsupervised transfer always had worse performance than within-subject classification. This may be because SSVEP had 12 classes, much more than the two classes in MI, so it was more sensitive to the data discrepancies between the source and target subjects.

2. Although most of the four existing data augmentation approaches were ineffective, our proposed CR always outperformed Baseline, again indicating the effectiveness and robustness of incorporating prior knowledge into data augmentation.

4.4.3. P300 Classification

Due to page limit, only cross-subject unsupervised transfer was consider in P300. EEGNet with batch size 256 was used. All other hyperparameters were identical to those in MI.

Tables 8 and 9 show the results. All data augmentation approaches were effective, but our proposed CR achieved the best or second best average BCAs, comparable to Noise. However, Noise has one hyper-parameter, whereas CR is completely parameter-free.

4.4.4. Seizure Classification

EEGNet and EEGWaveNet (Thuwajit et al., 2022) were used in cross-subject unsupervised transfer in seizure classification. Same as Wang et al. (2023), both networks were trained for 100 epochs. Batch size 256 was used for Baseline. It was doubled for Noise, Flip and CR, and tripled for Scale and Freq.

Tables 10 and 11 show the results. Again, our proposed CR achieved the best average BCAs. Particularly, it was the only effective data augmentation approach for EEG-Net.

4.5. Visualizations

Figure 5 uses t-SNE (Van der Maaten and Hinton, 2008) to visualize the feature distributions of the left and right hand imaginations from some subject in the three MI datasets. Observe that the augmented samples may occur in regions where no original samples were present, which may not be possible without utilizing prior knowledge.

Figure 6 shows t-SNE visualization of feature distributions of all subjects on the three MI datasets. Integrating the CR augmented samples added additional information beyond the original samples.

Figures 5 and 6 show that, for each subject, the CR augmented samples generally had consistent distributions with the original samples. There are a few exceptions:

1. Figure 6(a) shows that, on MI-I, the CR augmented samples from S1 and S4 were far away from their

Table 4: Classification accuracies (%) on MI-I using different data augmentation approaches. The best average performance in each panel is marked in bold, and the second best by an underline.

Scenario	Approach	S0	S1	S2	S3	S4	S5	S6	S7	S8	Avg.
	Baseline	70.44	59.33	81.33	64.11	48.78	66.89	60.78	89.11	71.11	67.99
	Noise	69.89	58.56	81.11	63.78	49.22	67.44	62.56	89.89	70.22	68.07
Within	Flip	71.78	57.89	80.89	64.67	52.33	66.44	61.78	88.78	71.44	68.44
-subject	Scale	70.00	58.56	82.22	62.33	49.78	66.44	61.33	89.33	70.67	67.85
classification	Freq	70.00	59.44	81.22	63.44	50.78	67.56	62.00	90.33	71.22	68.44
	Symm	52.00	56.11	52.00	58.11	52.11	51.22	57.22	68.78	63.78	56.81
	CR	75.67	56.00	88.22	69.78	46.78	71.22	71.00	87.11	73.67	**71.05**
	Baseline	82.22	60.69	93.19	68.06	56.11	72.64	65.00	85.42	80.14	$73.72_{\pm1.14}$
Cross	Noise	79.17	62.08	95.69	67.08	60.42	70.28	65.14	84.17	81.81	$73.98_{\pm1.06}$
-subject	Flip	75.69	59.58	94.72	65.83	57.36	75.83	62.22	83.61	82.78	$73.07_{\pm0.84}$
unsupervised	Scale	80.56	61.11	92.78	65.56	56.94	71.25	63.75	85.42	81.53	$73.21_{\pm0.76}$
Transfer	Freq	83.47	62.22	93.61	68.33	58.75	72.22	62.92	85.83	82.08	$74.38_{\pm0.89}$
	Symm	52.78	52.64	54.86	54.17	51.53	52.22	52.64	55.00	53.75	$53.29_{\pm0.45}$
	CR	84.03	63.06	91.39	71.67	60.83	75.00	70.00	85.00	80.97	$\mathbf{75.77}_{\pm1.58}$
	Baseline	84.96	60.93	94.53	72.51	63.60	74.11	68.01	86.94	89.39	$77.22_{\pm0.81}$
Cross	Noise	83.85	59.25	93.59	72.76	63.34	72.50	68.43	86.20	88.52	$76.49_{\pm1.37}$
-subject	Flip	85.40	60.01	94.05	71.27	59.12	73.40	65.06	86.15	90.89	$76.15_{\pm1.00}$
supervised	Scale	83.04	59.54	92.40	71.49	67.81	73.73	69.06	84.77	89.23	$76.78_{\pm1.46}$
transfer	Freq	85.12	63.18	94.26	70.88	68.68	74.31	69.80	87.29	90.65	$78.24_{\pm1.30}$
	Symm	57.01	52.43	69.54	52.27	51.10	53.00	52.61	73.86	59.39	$57.91_{\pm1.20}$
	CR	91.35	64.30	95.18	75.76	70.60	76.06	78.51	87.31	85.12	$\mathbf{80.46}_{\pm1.02}$

Table 5: Classification accuracies (%) on MI-II using different data augmentation approaches. The best average performance in each panel is marked in bold, and the second best by an underline.

Scenario	Approach	S0	S1	S2	S3	S4	S5	S6	S7	S8	Avg.
	Baseline	82.56	59.78	48.56	77.56	58.78	70.11	61.78	55.11	66.78	64.56
	Noise	82.89	60.11	49.44	77.78	59.00	70.11	61.56	55.78	66.56	64.80
Within	Flip	81.11	56.22	43.00	64.11	49.78	49.67	54.78	55.44	69.11	58.14
-subject	Scale	82.44	59.00	49.22	77.89	58.78	70.33	62.00	55.78	66.78	64.69
classification	Freq	82.56	59.78	48.56	77.56	58.78	70.11	61.78	55.11	66.78	64.56
	Symm	63.22	56.33	40.78	46.56	49.44	51.11	52.00	52.44	48.33	51.14
	CR	83.44	56.00	56.00	78.89	57.89	71.89	59.67	50.78	73.33	**65.32**
	Baseline	81.33	60.83	63.00	72.67	63.83	71.67	64.00	65.67	66.33	$67.70_{\pm0.91}$
Cross	Noise	81.67	62.00	64.17	72.00	66.33	70.83	64.33	64.67	67.50	$68.17_{\pm0.77}$
-subject	Flip	84.00	57.83	62.50	74.33	63.17	70.50	66.00	67.83	67.67	$68.20_{\pm0.53}$
unsupervised	Scale	79.00	58.83	66.33	69.83	65.33	70.17	62.67	68.17	64.83	$67.24_{\pm1.32}$
transfer	Freq	82.67	57.83	64.00	71.83	65.17	67.50	66.67	63.50	68.83	$67.56_{\pm0.55}$
	Symm	49.00	48.00	49.83	49.33	48.67	46.50	50.33	49.00	47.50	$48.69_{\pm2.03}$
	CR	84.83	59.33	66.83	73.50	64.17	73.00	65.33	64.00	71.00	$\mathbf{69.11}_{\pm0.84}$
	Baseline	88.73	57.11	56.44	76.41	66.27	75.89	68.25	59.04	71.76	$68.88_{\pm1.16}$
Cross	Noise	87.80	56.89	57.20	76.04	68.47	74.58	67.03	58.49	70.34	$68.54_{\pm0.68}$
-subject	Flip	88.99	57.71	54.03	80.51	65.85	75.26	73.01	57.35	73.24	$69.55_{\pm1.01}$
supervised	Scale	87.46	56.55	54.93	73.32	67.11	71.80	67.33	58.22	70.74	$67.49_{\pm0.87}$
transfer	Freq	87.91	57.23	56.39	80.05	65.77	72.78	68.53	57.43	73.36	$68.83_{\pm0.94}$
	Symm	53.62	48.81	52.52	57.20	48.18	46.50	53.16	52.11	46.90	$51.00_{\pm2.31}$
	CR	90.13	62.35	59.22	76.56	64.80	76.95	70.89	57.62	71.65	$\mathbf{70.02}_{\pm1.13}$

Table 6: Classification accuracies (%) on MI-III using different data augmentation approaches. The best average performance in each panel is marked in bold, and the second best by an underline.

Scenario	Approach	S0	S1	S2	S3	S4	S5	S6	Avg.
Within -subject classification	Baseline	66.89	57.33	62.89	81.67	84.22	74.89	80.78	72.67
	Noise	68.00	55.56	61.11	82.00	84.56	74.67	78.67	72.08
	Flip	69.22	54.67	61.44	83.78	87.89	73.78	77.11	72.56
	Scale	67.22	55.56	62.89	83.44	84.56	75.22	81.22	$\underline{72.87}$
	Freq	67.56	55.00	62.67	83.00	84.56	72.22	79.11	72.02
	Symm	66.89	54.78	65.11	74.67	57.78	74.89	57.67	64.54
	CR	66.89	59.22	61.33	81.78	85.00	74.89	89.22	**74.05**
Cross -subject unsupervised transfer	Baseline	66.00	69.30	65.70	58.30	92.20	74.80	71.90	$71.17_{\pm0.87}$
	Noise	69.60	69.70	65.80	59.40	92.80	75.20	70.60	$\underline{71.87}_{\pm0.55}$
	Flip	67.50	66.70	66.10	54.50	91.50	75.00	67.70	$69.86_{\pm1.49}$
	Scale	68.80	69.40	66.10	60.50	91.20	73.30	72.60	$71.70_{\pm0.73}$
	Freq	68.10	68.30	66.00	58.90	91.30	73.70	70.40	$70.96_{\pm0.82}$
	Symm	53.50	49.00	49.80	50.60	47.40	53.80	42.50	$49.51_{\pm0.80}$
	CR	66.40	75.60	64.50	68.10	92.90	72.10	85.20	$\mathbf{74.97}_{\pm0.96}$
Cross -subject supervised transfer	Baseline	75.16	74.34	69.98	75.57	93.48	81.51	80.13	$\underline{78.60}_{\pm1.13}$
	Noise	76.26	72.70	71.37	74.86	92.85	78.85	79.69	$78.08_{\pm1.00}$
	Flip	73.55	73.41	70.55	73.99	94.16	77.69	77.11	$77.21_{\pm1.66}$
	Scale	76.16	71.11	68.61	70.50	91.03	77.56	75.86	$75.83_{\pm1.04}$
	Freq	75.54	72.42	69.62	71.92	90.86	78.02	77.36	$76.53_{\pm0.86}$
	Symm	56.88	50.33	49.33	59.29	75.28	60.66	64.31	$59.44_{\pm1.78}$
	CR	76.86	77.65	67.31	82.11	94.32	79.14	85.91	$\mathbf{80.47}_{\pm0.82}$

Table 7: Classification accuracies (%) on SSVEP using different data augmentation approaches. The best average performance in each panel is marked in bold, and the second best by an underline.

Scenario	Approach	S0	S1	S2	S3	S4	S5	S6	S7	S8	S9	Avg.
Within -subject classification	Baseline	79.88	60.75	80.46	93.90	92.66	94.06	92.91	97.50	92.32	74.42	$85.89_{\pm0.15}$
	Noise	65.95	47.29	69.85	90.34	88.6	91.38	92.16	97.71	88.37	68.1	$79.97_{\pm0.50}$
	Flip	66.48	47.87	69.5	90.48	88.77	91.68	91.98	97.86	89.06	68.26	$80.19_{\pm0.39}$
	Scale	81.57	62.44	83.6	93.22	94.28	94.92	93.06	97.32	93.28	75.36	$\mathbf{86.91}_{\pm0.07}$
	Freq	71.55	54.49	78.38	91.51	91.66	91.86	92.59	92.56	89.87	63.91	$81.84_{\pm0.30}$
	CR	81.56	64.07	79.95	94.57	94.25	93.83	92.86	97.22	92.19	76.26	$\underline{86.68}_{\pm0.28}$
Cross -subject unsupervised transfer	Baseline	52.52	43.77	61.87	95.81	95.03	93.88	94.39	97.90	95.27	86.97	$\underline{81.74}_{\pm0.37}$
	Noise	42.78	46.52	47.37	97.21	93.32	94.16	92.21	93.75	91.24	78.20	$77.68_{\pm0.20}$
	Flip	32.22	34.03	42.64	89.58	89.45	89.72	85.14	95.97	87.91	74.17	$72.08_{\pm0.38}$
	Scale	46.67	45.37	50.74	97.59	94.63	94.26	91.30	95.19	89.07	80.56	$78.54_{\pm0.46}$
	Freq	45.18	38.34	49.08	95.00	93.52	93.33	92.78	94.26	81.30	81.30	$76.41_{\pm0.23}$
	CR	56.89	48.11	67.33	95.33	96.11	91.67	94.78	98.67	93.66	88.67	$\mathbf{83.12}_{\pm0.17}$
Cross -subject supervised transfer	Baseline	67.88	54.11	79.03	97.30	97.88	95.35	93.28	99.09	95.57	81.25	$86.07_{\pm0.30}$
	Noise	69.04	52.28	78.83	97.26	97.80	95.88	92.49	98.32	95.36	81.40	$85.87_{\pm0.23}$
	Flip	63.66	43.83	65.64	94.52	84.68	91.74	92.64	96.16	94.85	77.05	$80.48_{\pm0.27}$
	Scale	72.34	56.52	83.85	97.07	97.20	96.50	94.11	98.05	95.87	85.69	$87.72_{\pm0.11}$
	Freq	73.05	54.69	86.52	97.54	97.95	94.29	96.25	97.53	95.34	84.95	$\underline{87.81}_{\pm0.22}$
	CR	74.54	58.33	84.64	97.33	97.93	93.99	95.74	98.97	94.59	87.99	$\mathbf{88.40}_{\pm0.28}$

Table 8: BCAs (%) of different data augmentation approaches in cross-subject unsupervised transfer on P300-I. The best average performance is marked in bold, and the second best by an underline.

Approach	S0	S1	S2	S3	S4	S5	S6	S7	Avg.
Baseline	71.65	69.35	75.11	70.31	73.46	69.94	74.18	72.78	$72.10_{\pm 0.17}$
Noise	71.17	70.27	75.59	70.33	74.00	70.29	74.56	74.29	$\underline{72.56}_{\pm 0.23}$
Flip	71.43	69.61	75.92	70.19	73.99	70.67	74.14	73.42	$72.42_{\pm 0.21}$
Scale	71.09	70.36	75.45	70.34	73.89	70.08	74.13	74.88	$72.53_{\pm 0.21}$
Freq	71.60	70.44	74.42	69.86	74.73	69.57	74.76	73.92	$72.41_{\pm 0.23}$
CR	71.50	70.35	75.43	70.21	74.36	70.66	74.82	74.33	$\mathbf{72.71}_{\pm 0.13}$

Table 9: BCAs (%) of different data augmentation approaches in cross-subject unsupervised transfer on P300-II. The best average performance is marked in bold, and the second best by an underline.

Approach	S0	S1	S2	S3	S4	S5	S6	S7	S8	S9	Avg.
Baseline	76.04	85.51	79.26	85.52	87.15	79.06	76.92	74.28	87.95	87.71	$81.94_{\pm 0.16}$
Noise	78.36	85.10	79.66	85.65	88.08	79.15	80.18	76.29	88.35	89.50	$\mathbf{83.03}_{\pm 0.08}$
Flip	76.70	85.26	79.11	84.54	88.78	79.86	78.65	75.54	88.73	87.54	$82.47_{\pm 0.20}$
Scale	78.52	85.13	79.76	85.32	88.09	78.63	79.38	75.25	88.26	89.14	$82.75_{\pm 0.28}$
Freq	78.45	86.01	79.63	84.91	88.17	80.05	79.50	75.70	88.73	88.69	$82.98_{\pm 0.12}$
CR	77.91	85.39	79.69	85.92	88.32	79.51	79.81	75.52	88.63	89.53	$\underline{83.02}_{\pm 0.11}$

original samples. This is because these two subjects had lower data quality, as indicated by the low classification accuracies in Table 4 and previous research (Zanini et al., 2018). Figure 5(a) also shows that their two classes had large overlaps. So, the CR approach may perform less well when the original data quality is low.

2. Figure 5(c) shows that, on MI-III, the distributions of CR augmented samples and the original samples from S0 and S5 were much separated than others. This is because S0 and S5 performed right hand versus feet MI tasks, whereas other subjects performed left hand versus right hand MI tasks.

4.6. Necessity of Symmetry in CR

Random shuffle (RS) was performed to verify the necessity of symmetry in CR. RS randomly exchanges the left and right hemisphere channels, without considering the channel symmetry. For example, in Figure 2(a), FC3 may be switched with P2 in RS, whereas FC3 must be switched with FC4 in CR.

The results are shown in Table 12. CR always achieved much better performance, indicating the necessity of maintaining strict symmetry of the switched channels.

4.7. Effect of Transfer Learning

Wu et al. (2022) demonstrated the benefits of utilizing source subjects' data to facilitate the calibration for the target subject. Figure 7 shows the performance of CR as the number of labeled target samples increased:

1. Generally, as the number of labeled target samples increased, the performance in both scenarios increased, regardless of whether data augmentation was used or not, which is intuitive.

2. Cross-subject supervised transfer almost always outperformed within-subject classification (the solid curves are almost always higher than the dashed curves of the same color), especially when the number of labeled target samples was small, indicating again the benefits of utilizing source subjects' data for target subject model learning.

3. As the number of labeled target samples increased, the performance improvement of transfer learning

Table 10: BCAs (%) of different data augmentation approaches on Seizure-I in cross-subject unsupervised transfer. The best average performance is marked in bold, and the second best by an underline.

ID	EEGNet						EEGWaveNet					
	Baseline	Noise	Flip	Scale	Freq	CR	Baseline	Noise	Flip	Scale	Freq	CR
S1	50.98	62.07	62.00	56.72	61.38	55.73	59.67	54.21	57.47	58.58	62.53	65.55
S4	81.01	60.37	60.55	60.60	57.02	66.03	70.02	67.59	72.85	67.77	70.09	69.52
S5	69.26	79.00	78.27	78.92	78.78	77.38	70.39	65.28	73.08	70.45	64.58	74.47
S7	57.38	55.82	63.15	60.86	61.16	66.30	70.78	65.68	62.72	63.18	61.99	66.96
S9	74.30	55.94	57.12	52.09	55.53	54.94	64.39	65.56	57.35	65.51	67.64	67.62
S11	59.65	63.95	70.85	64.32	71.29	70.01	62.03	61.57	61.40	60.50	70.68	58.74
S13	60.37	54.75	57.18	59.33	55.51	67.13	65.92	60.46	56.67	64.03	57.67	71.20
S14	55.02	56.56	56.36	55.30	49.10	49.31	47.17	54.25	56.85	56.42	55.81	52.98
S15	51.21	55.60	53.34	54.32	53.30	50.27	51.04	51.64	53.90	50.50	49.80	52.42
S16	51.54	51.37	53.85	54.31	54.57	59.61	56.48	62.91	57.78	61.39	62.02	56.81
S17	59.55	57.47	63.54	57.65	60.09	52.10	49.98	53.81	50.89	55.09	55.06	53.58
S19	62.16	51.65	52.90	48.49	50.40	56.60	52.44	55.08	56.27	57.05	58.03	61.59
S20	50.88	59.80	59.90	55.64	61.65	56.78	60.63	58.75	56.76	60.18	53.87	58.26
S21	59.23	56.36	87.48	86.19	88.74	89.00	58.97	74.93	68.00	66.09	66.41	61.67
S22	40.38	52.79	52.84	51.82	50.38	54.73	53.14	49.92	52.72	55.29	52.02	57.04
S25	64.29	71.02	65.89	69.14	66.20	74.60	62.64	59.46	55.93	65.44	62.36	61.77
S31	78.75	55.22	57.21	52.87	53.26	59.20	76.75	60.18	59.14	82.31	67.31	87.11
S34	75.29	64.86	69.42	57.60	56.62	70.90	76.68	86.57	81.75	93.03	90.07	91.82
S36	69.61	78.63	84.57	63.11	75.09	74.00	77.25	73.63	62.48	80.67	79.84	71.97
S38	66.76	61.98	58.02	57.39	51.48	54.34	59.70	58.13	58.28	55.38	56.69	60.80
S39	74.43	66.78	65.85	64.49	65.09	58.13	69.54	64.02	67.54	71.56	64.47	64.41
S40	51.15	49.44	50.62	50.21	50.27	58.84	62.07	56.56	59.85	53.39	64.14	61.99
S41	55.26	63.23	61.85	53.02	57.53	64.96	70.64	61.74	66.09	67.81	67.42	65.22
S44	68.79	45.39	64.38	42.51	66.24	64.85	64.57	79.30	52.59	69.43	71.01	61.58
S47	83.12	63.45	59.04	53.27	62.49	63.73	70.34	64.71	66.71	82.84	71.09	84.48
S50	69.92	65.61	69.58	65.37	68.64	73.96	73.85	78.82	70.84	73.62	64.58	68.48
S51	54.03	44.26	47.87	46.81	48.37	46.62	55.35	50.41	39.81	61.28	51.48	56.30
S52	47.28	60.41	57.31	67.96	56.17	63.85	54.14	50.46	51.41	59.99	56.79	63.88
S62	90.19	51.30	53.17	50.05	52.32	46.00	68.78	83.73	76.66	88.12	93.98	88.09
S63	50.10	52.71	51.36	54.65	54.19	51.49	48.33	51.69	48.40	54.33	50.95	53.49
S66	63.95	71.12	56.07	70.19	65.85	75.69	64.20	66.64	68.66	63.69	60.71	67.71
S67	59.71	67.83	70.78	71.32	65.35	76.54	74.97	72.60	63.95	65.32	58.11	74.49
S69	50.05	49.98	50.64	68.48	49.93	50.08	83.02	84.68	86.52	89.66	87.90	90.91
S73	50.03	60.97	59.50	51.08	50.67	67.94	51.28	60.33	67.20	57.25	56.51	64.82
S75	52.47	51.51	50.35	52.87	52.59	56.34	56.01	52.24	58.61	59.15	56.64	59.04
S76	53.95	51.79	51.24	55.00	50.65	53.82	53.37	54.67	53.51	52.77	52.98	59.41
S77	50.58	58.29	56.61	58.58	54.84	45.98	48.97	52.80	54.11	49.66	49.74	49.43
S78	67.16	67.51	69.64	66.34	63.74	80.51	74.13	66.41	63.15	68.56	74.64	76.40
S79	57.67	49.57	46.56	49.96	46.32	60.95	62.12	60.79	57.60	62.25	54.40	63.57
Avg.	<u>61.22</u>	58.88	60.43	58.69	58.79	**62.03**	62.86	62.88	61.17	<u>65.12</u>	63.38	**66.04**
	±0.54	±0.29	±0.24	±0.43	±0.51	±0.40	±0.68	±0.50	±0.47	±0.71	±0.53	±0.16

13

Table 11: BCAs (%) of different data augmentation approaches on Seizure-II in cross-subject unsupervised transfer setting. The best average performance is marked in bold, and the second best by an underline.

ID	EEGNet						EEGWaveNet					
	Baseline	Noise	Flip	Scale	Freq	CR	Baseline	Noise	Flip	Scale	Freq	CR
S1	89.41	86.37	84.80	86.28	84.22	89.31	83.18	83.27	87.93	91.00	83.49	85.22
S2	84.52	81.78	71.15	82.97	78.71	86.04	76.60	76.28	72.85	88.74	66.53	75.00
S3	99.17	96.50	93.87	98.83	98.33	97.50	92.36	98.67	95.33	94.12	96.50	98.25
S4	92.47	93.55	94.62	93.01	93.01	94.09	94.62	95.16	89.25	92.47	97.31	93.55
S5	85.42	80.16	73.05	81.16	77.67	78.35	70.20	75.68	81.57	75.14	73.96	88.54
S6	70.67	70.02	76.76	78.13	67.65	78.24	83.38	86.79	74.41	86.81	76.67	84.14
S7	93.75	79.58	89.58	93.75	93.75	79.17	85.83	91.67	87.50	67.92	78.33	85.83
S8	84.62	78.21	76.23	85.90	76.92	88.46	92.79	90.22	90.22	87.66	94.07	83.23
S9	91.03	88.78	94.87	94.55	93.27	93.59	95.19	95.83	91.76	95.83	87.50	91.03
S10	82.52	84.53	62.27	67.91	79.43	81.25	57.74	70.70	61.87	66.39	62.48	63.93
S11	86.60	91.18	92.75	94.93	94.55	96.50	90.33	85.18	76.03	86.71	84.23	85.39
S12	54.11	53.56	52.00	54.26	50.60	60.50	55.06	52.99	52.96	55.95	60.94	66.31
S13	85.89	88.74	88.33	94.30	87.22	83.26	66.67	75.00	78.33	62.78	74.44	81.11
S14	89.50	89.75	89.88	84.53	85.59	96.55	78.96	71.79	94.67	71.55	93.63	85.13
S15	92.19	91.67	92.83	87.76	88.31	67.81	91.31	70.49	72.04	70.56	70.80	55.75
S16	82.83	89.39	87.37	89.90	85.86	85.86	78.28	77.27	81.31	81.82	78.28	82.32
S17	94.39	93.26	89.54	92.50	90.67	92.05	86.56	81.00	84.22	84.16	80.94	89.77
S18	95.42	82.31	97.14	90.67	94.22	97.86	97.22	99.08	92.86	98.78	97.14	84.14
S19	82.67	78.47	71.34	80.63	76.32	91.35	64.01	59.87	76.44	66.05	79.73	81.18
S20	69.11	62.12	60.04	59.73	60.12	71.02	88.91	86.74	80.84	84.56	81.35	76.77
S21	72.06	91.36	75.07	80.37	84.17	86.14	64.49	57.36	73.30	80.96	65.24	80.03
S22	89.82	87.50	81.94	88.89	86.11	85.83	87.96	88.15	91.11	84.44	84.35	88.06
S23	57.60	58.82	56.87	59.74	55.66	59.69	49.92	59.33	56.39	59.04	50.27	48.05
S24	83.38	78.50	81.85	81.57	81.30	84.91	76.62	78.62	81.13	74.37	80.37	82.39
S25	92.48	95.51	94.01	95.51	89.14	88.39	93.74	88.61	89.97	94.16	90.29	97.93
S26	46.91	45.13	51.79	49.40	55.29	62.00	56.01	64.30	65.26	71.91	56.17	56.97
S27	63.45	60.31	65.43	59.88	63.62	59.00	49.46	51.72	54.22	59.68	51.66	55.66
Avg.	81.92	80.63	79.46	81.74	80.43	**82.77**	78.05	78.21	79.03	79.02	77.65	**79.47**
	±0.76	±0.81	±0.78	±0.31	±0.93	±0.40	±0.34	±0.97	±0.70	±1.56	±1.31	±0.57

Table 12: Performance comparison of RS and CR in cross-subject unsupervised transfer.

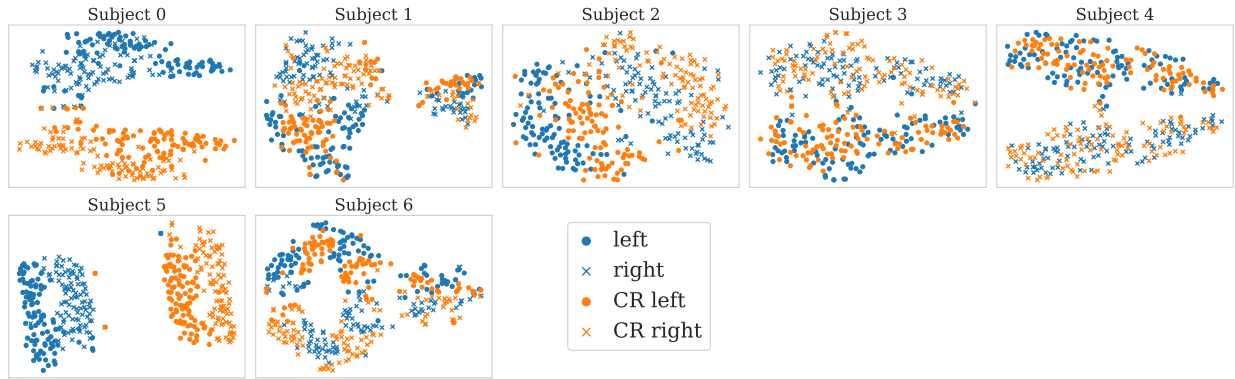| Approach	MI-I	MI-II	MI-III	SSVEP	ERP-I	ERP-II	Seizure-I	Seizure-II
RS	$60.54_{\pm0.45}$	$48.69_{\pm2.03}$	$67.44_{\pm0.80}$	$76.25_{\pm0.32}$	$71.76_{\pm0.18}$	$79.50_{\pm0.21}$	$63.39_{\pm0.23}$	$75.25_{\pm1.09}$
CR	$75.77_{\pm1.58}$	$69.11_{\pm0.84}$	$74.97_{\pm0.96}$	$83.12_{\pm0.17}$	$72.71_{\pm0.13}$	$83.02_{\pm0.11}$	$66.04_{\pm0.16}$	$79.47_{\pm0.57}$
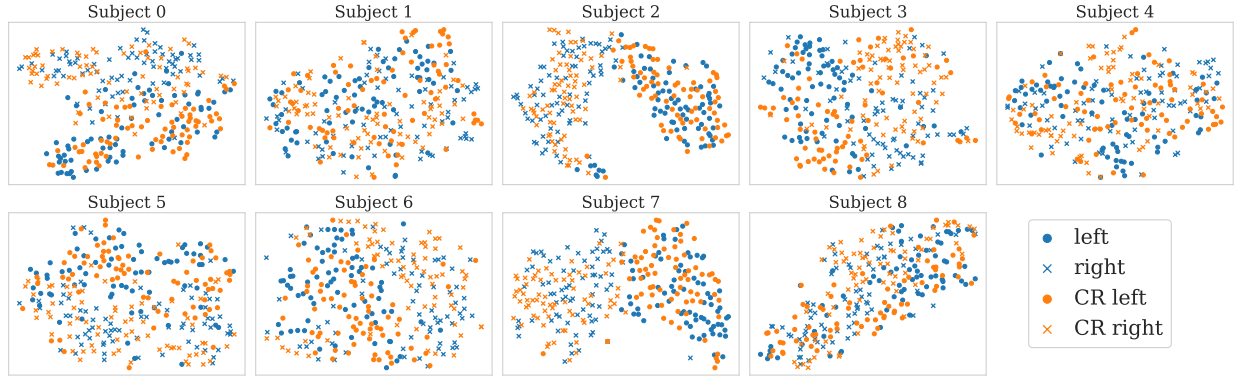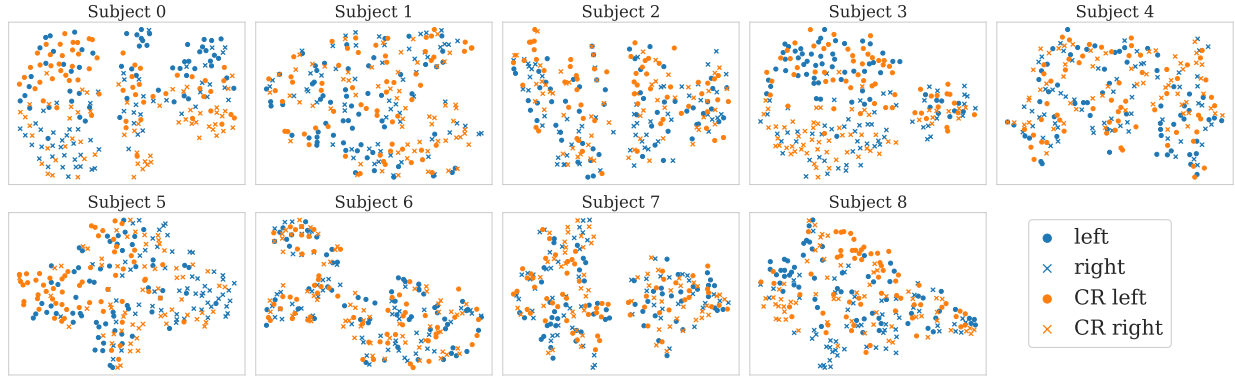
Figure 5: *t*-SNE visualization of CSP features from the original EEG trials and the CR augmented EEG trials. (a) Subjects S0-S8 in MI-I; (b) Subjects S0-S8 in MI-II; and, (c) Subjects S0-S6 in MI-III. Different colors represent different classes. The dots represent the original trials, and the crosses represent the CR augmented trials. Note that for S0 and S5 in MI-III, the classification task was right hand versus feet.
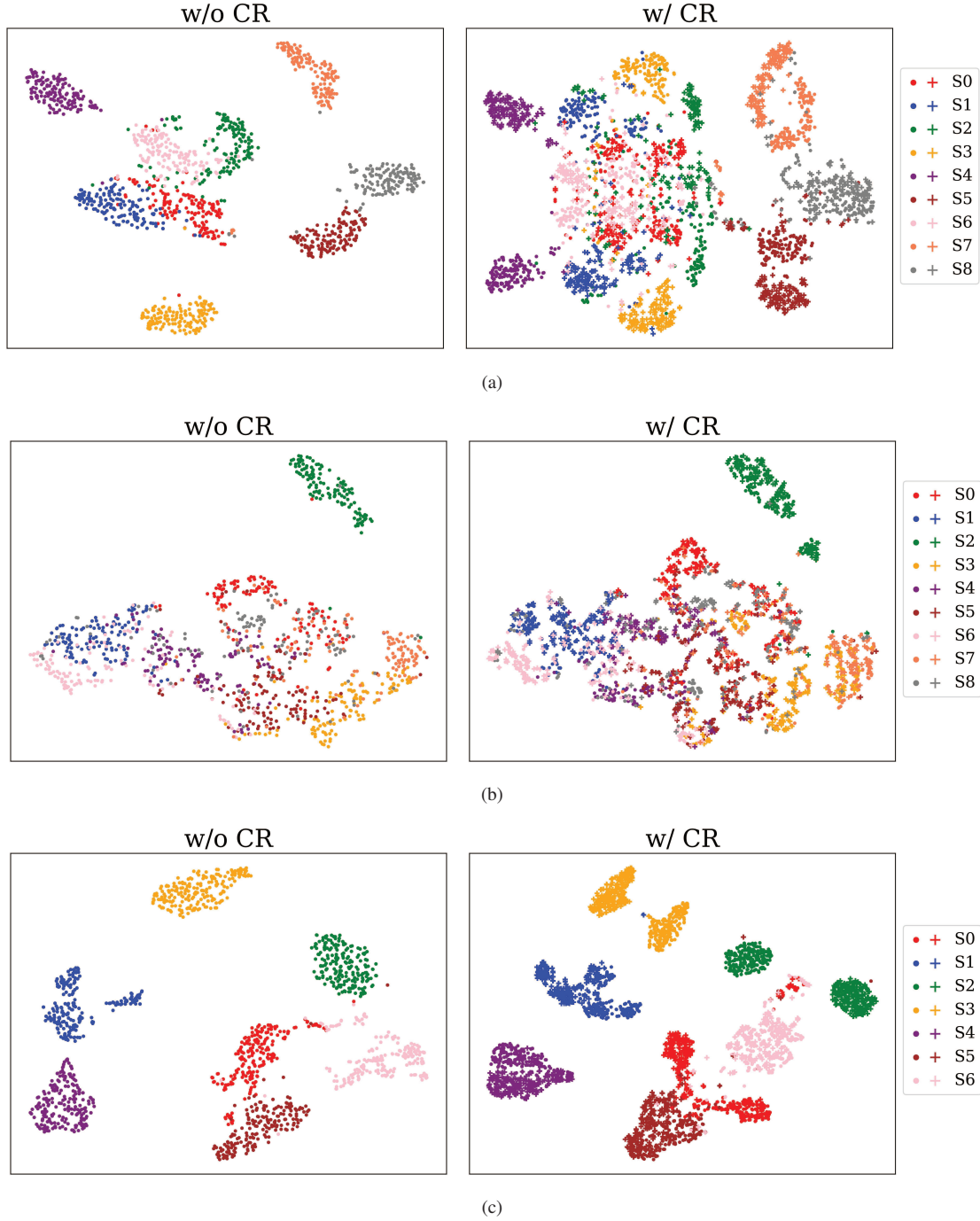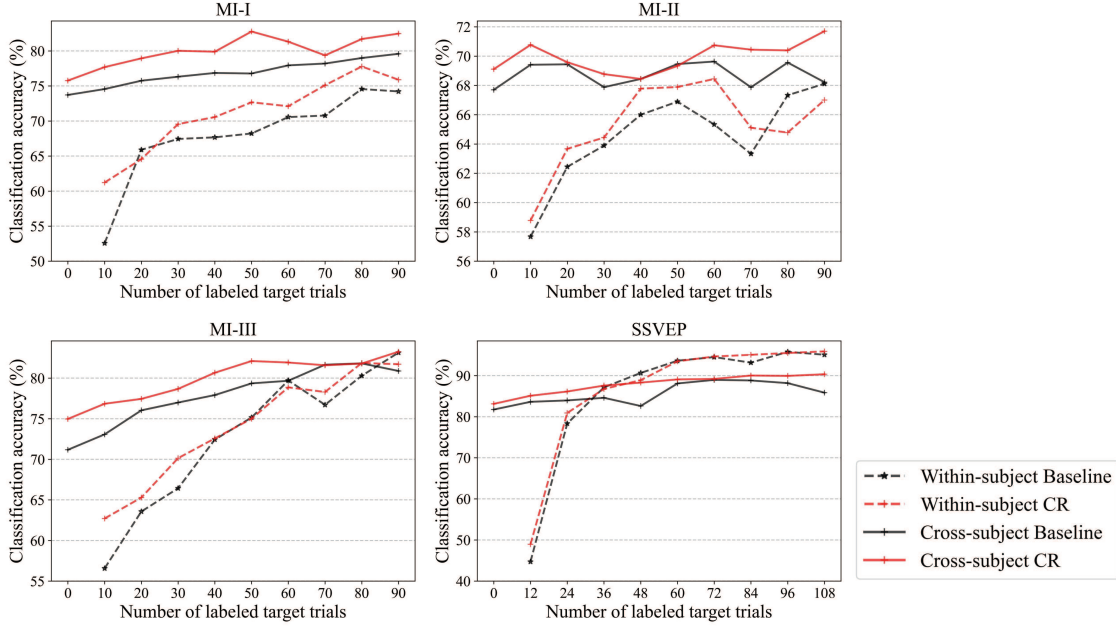
Figure 6: *t*-SNE visualization of CSP features from the original EEG trials and the CR augmented ones. (a) MI-I; (b) MI-II; and, (c) MI-III. Different colors represent different subjects. The dots represent the original trials, and the crosses represent the CR augmented trials.

Figure 7: Performance of CR as the number of labeled target samples increased on MI-I, MI-II, MI-III and SSVEP datasets.

diminished. It indicates that if we have access to an adequate amount of labeled target data, then auxiliary data from other subjects are less beneficial. For example, on MI-I, when there were 10 labeled target samples, the performance improvement of cross-subject CR over within-subject CR was 77.69%-61.22%=16.47%; however, when there were 90 labeled target samples, the performance improvement of cross-subject CR over within-subject CR reduced to 82.47%-75.89%=6.58%.

4. CR almost always outperformed Baseline (the red curves are almost always higher than the corresponding black curves of the same line style), indicating again the effectiveness and robustness of incorporating prior knowledge in data augmentation.

4.8. Hyperparameter Analysis

CR is hyperparameter-free; however, other data augmentation approaches like Noise, Freq and Scale all have hyperparameters, which may affect their performance. Figure 8 shows how their performance changed with the hyperparameters. Although better performance of the other three data augmentation approaches may be obtained by adjusting their hyperparameters, our proposed CR almost always outperformed them.

4.9. Combination of Different Data Augmentations

It is also interesting to study if our proposed CR can be combined with other data augmentation approaches for further performance improvement.

Figure 8 shows that Freq has the second-best performance in MI-I when $C_{freq} = 0.4$, and in MI-II and MI-III when $C_{freq} = 0.5$. We combined these best configurations of Freq with CR, expanding the training data size by a factor of 6. The results are shown in Table 13. CR+Freq always outperformed Freq, and achieved comparable or better performance than CR, suggesting the effectiveness, robustness and flexibility of CR.

5. Conclusions

To cope with the typical calibration data shortage challenge in EEG-based BCIs, this paper has proposed a
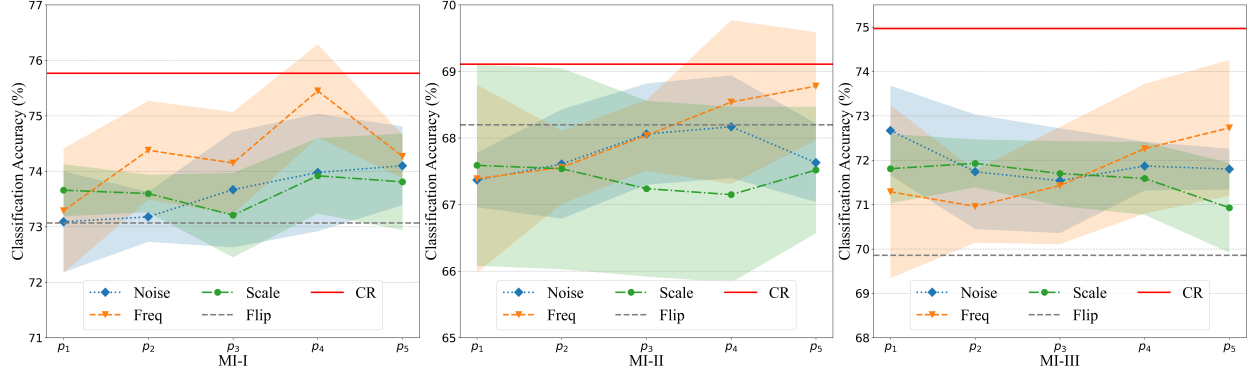
Figure 8: Parameter sensitivity analysis for Noise, Freq, Scale and Flip on MI-I, MI-II and MI-III in cross-subject unsupervised transfer. $C_{\text{noise}} \in \{0.25, 0.5, 1, 2, 4\}$, $C_{\text{freq}} \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$, and $C_{\text{scale}} \in \{0.005, 0.01, 0.05, 0.1, 0.2\}$. CR and Flip do not have hyperparameters.

Table 13: The performance of combining CR with Freq on MI datasets in cross-subject unsupervised transfer setting. The best average performance is marked in bold.

Approach	MI-I	MI-II	MI-III
Freq	$75.45_{\pm 0.89}$	$68.78_{\pm 0.55}$	$72.73_{\pm 0.82}$
CR	$\mathbf{75.77}_{\pm 1.58}$	$69.11_{\pm 0.84}$	$74.97_{\pm 0.96}$
CR+Freq	$75.45_{\pm 0.62}$	$\mathbf{69.70}_{\pm 0.73}$	$\mathbf{77.27}_{\pm 0.84}$

parameter-free CR data augmentation approach that incorporates prior knowledge on the channel distributions of different BCI paradigms in data augmentation. Experiments on eight public EEG datasets across four different BCI paradigms (MI, SSVEP, P300, and Seizure classifications) using different decoding algorithms demonstrated that: 1) CR is effective, i.e., it can noticeably improve the classification accuracy; 2) CR is robust, i.e., it consistently outperforms other data augmentation approaches in the literature; and, 3) CR is flexible, i.e., it can be combined with other data augmentation approaches to further increase the performance. We suggest that data augmentation approaches like CR should be an essential step in EEG signal classification.

CR also has some limitations, which will be investigated in our future research:

1. Although CR is effective for augmenting left/right hand MI trials, it cannot be directly applied to other classes like feet or tongue MIs.
2. This paper assumes strict symmetry between left and right hemisphere EEG electrodes; however, in practice the electrodes may not always be perfectly symmetric.

References

Abiri, R., Borhani, S., Sellers, E.W., Jiang, Y., Zhao, X., 2019. A comprehensive review of EEG-based brain-computer interface paradigms. Journal of Neural Engineering 16, 011001.

Acharya, U.R., Sree, S.V., Swapna, G., Martis, R.J., Suri, J.S., 2013. Automated EEG analysis of epilepsy: a review. Knowledge-Based Systems 45, 147–165.

Aricò, P., Aloise, F., Schettini, F., Salinari, S., Mattia, D., Cincotti, F., 2014. Influence of P300 latency jitter on event related potential-based brain–computer interface performance. Journal of Neural Engineering 11, 035008.

Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.R., Curio, G., 2007. The non-invasive Berlin brain–computer interface: fast acquisition of effective performance in untrained subjects. NeuroImage 37, 539–550.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., Muller, K.r., 2008. Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Processing Magazine 25, 41–56.

Boonyakitanont, P., Lek-Uthai, A., Chomtho, K., Songsiri, J., 2020. A review of feature extraction and performance evaluation in epileptic seizure detection using EEG. Biomedical Signal Processing and Control 57, 101702.

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

Deiss, O., Biswal, S., Jin, J., Sun, H., Westover, M.B., Sun, J., 2018. HAMLET: interpretable human and machine co-learning technique. arXiv preprint arXiv:1803.09702 .

Freeman, W.J., 2007. Hilbert transform for brain waves. Scholarpedia 2, 1338.

Freer, D., Yang, G.Z., 2020. Data augmentation for self-paced motor imagery classification with C-LSTM. Journal of Neural Engineering 17, 016041.

Friman, O., Volosyak, I., Graser, A., 2007. Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces. IEEE Trans. on Biomedical Engineering 54, 742–750.

He, H., Wu, D., 2020. Transfer learning for brain-computer interfaces: A Euclidean space data alignment approach. IEEE Trans. on Biomedical Engineering 67, 399–410.

Hoffmann, U., Vesin, J.M., Ebrahimi, T., Diserens, K., 2008. An efficient P300-based brain–computer interface for disabled subjects. Journal of Neuroscience Methods 167, 115–125.

Jasper, H., Penfield, W., 1949. Electrocorticograms in man: effect of voluntary movement upon the electrical activity of the precentral gyrus. Archiv für Psychiatrie und Nervenkrankheiten 183, 163–174.

Jayaram, V., Barachant, A., 2018. MOABB: trustworthy algorithm benchmarking for BCIs. Journal of Neural Engineering 15, 066011.

Jenke, R., Peer, A., Buss, M., 2014. Feature extraction and selection for emotion recognition from EEG. IEEE Trans. on Affective computing 5, 327–339.

Krell, M.M., Kim, S.K., 2017. Rotational data augmentation for electroencephalographic data, in: IEEE Engineering in Medicine and Biology Society, Jeju Island, Korea. pp. 471–474.

Krucoff, M.O., Rahimpour, S., Slutzky, M.W., Edgerton, V.R., Turner, D.A., 2016. Enhancing nervous system recovery through neurobiologics, neural interface training, and neurorehabilitation. Frontiers in Neuroscience 10, 584.

Lance, B.J., Kerick, S.E., Ries, A.J., Oie, K.S., McDowell, K., 2012. Brain-computer interface technologies in the coming decades. Proc. of the IEEE 100, 1585–1599.

Lashgari, E., Liang, D., Maoz, U., 2020. Data augmentation for deep-learning-based electroencephalography. Journal of Neuroscience Methods 346, 108885.

Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J., 2018. EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces. Journal of Neural Engineering 15, 056013.

Leeb, R., Lee, F., Keinrath, C., Scherer, R., Bischof, H., Pfurtscheller, G., 2007. Brain–computer communication: Motivation, aim, and impact of exploring a virtual apartment. IEEE Trans. on Neural Systems and Rehabilitation Engineering 15, 473–482.

Li, B., Hou, Y., Che, W., 2022. Data augmentation approaches in natural language processing: A survey. AI Open 3, 71–90.

Li, R., Johansen, J.S., Ahmed, H., Ilyevsky, T.V., Wilbur, R.B., Bharadwaj, H.M., Siskind, J.M., 2021. The perils and pitfalls of block design for EEG classification experiments. IEEE Trans. on Pattern Analysis and Machine Intelligence 43, 316–333.

Li, S., Wang, Z., Luo, H., Ding, L., Wu, D., 2024. T-TIME: Test-time information maximization ensemble for plug-and-play BCIs. IEEE Trans. on Biomedical Engineering 71, 423–432.

Lin, C.T., Wu, R.C., Liang, S.F., Chao, W.H., Chen, Y.J., Jung, T.P., 2005. EEG-based drowsiness estimation for

safety driving using independent component analysis. IEEE Trans. on Circuits and Systems I: Regular Papers 52, 2726–2738.

Lisboa, P., Saralajew, S., Vellido, A., Fernández-Domenech, R., Villmann, T., 2023. The coming of age of interpretable and explainable machine learning models. Neurocomputing 535, 25–39.

Luo, Y., Lu, B.L., 2018. EEG data augmentation for emotion recognition using a conditional Wasserstein GAN, in: 40th Annual Int'l Conf. IEEE Engineering in Medicine and Biology Society, Honolulu, HI. pp. 2535–2538.

Van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. Journal of Machine Learning Research 9, 2579–2605.

Makeig, S., Kothe, C., Mullen, T., Bigdely-Shamlo, N., Zhang, Z., Kreutz-Delgado, K., 2012. Evolving signal processing for brain-computer interfaces. Proc. of the IEEE 100, 1567–1584.

Mohsenvand, M.N., Izadi, M.R., Maes, P., 2020. Contrastive representation learning for electroencephalogram classification, in: Proc. Advances in Neural Information Processing Systems Machine Learning for Health Workshops, Vancouver, Canada. pp. 238–253.

Nakanishi, M., Wang, Y., Wang, Y.T., Jung, T.P., 2015. A comparison study of canonical correlation analysis based methods for detecting steady-state visual evoked potentials. PloS one 10, 1–18.

Pan, Y., Chen, J., Zhang, Y., Zhang, Y., 2022. An efficient CNN-LSTM network with spectral normalization and label smoothing technologies for SSVEP frequency recognition. Journal of Neural Engineering 19, 056014.

Pedroni, A., Bahreini, A., Langer, N., 2019. Automagic: standardized preprocessing of big EEG data. NeuroImage 200, 460–473.

Pfurtscheller, G., Neuper, C., 2001. Motor imagery and direct brain-computer communication. Proc. of the IEEE 89, 1123–1134.

Ramoser, H., Muller-Gerking, J., Pfurtscheller, G., 2000. Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans. on Rehabilitation Engineering 8, 441–446.

Riccio, A., Simione, L., Schettini, F., Pizzimenti, A., Inghilleri, M., Belardinelli, M.O., Mattia, D., Cincotti, F., 2013. Attention and P300-based BCI performance in people with amyotrophic lateral sclerosis. Frontiers in Human Neuroscience 7, 732.

Rommel, C., Moreau, T., Paillard, J., Gramfort, A., 2022. CADDA: class-wise automatic differentiable data augmentation for EEG signals, in: Proc. Int'l Conf. Learning Representations, Virtual. pp. 1–24.

Rosenfeld, J.V., Wong, Y.T., 2017. Neurobionics and the brain-computer interface: current applications and future horizons. Medical Journal of Australia 206, 363–368.

von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., Kirsch, B., Walczak, M., Pfrommer, J., Pick, A., Ramamurthy, R., Garcke, J., Bauckhage, C., Schuecker, J., 2021. Informed machine learning - a taxonomy and survey of integrating prior knowledge into learning systems. IEEE Trans. on Knowledge and Data Engineering 35, 614–633.

Saeed, A., Grangier, D., Pietquin, O., Zeghidour, N., 2021. Learning from heterogeneous EEG signals with differentiable channel reordering, in: Proc. IEEE Int'l Conf. on Acoustics, Speech and Signal Processing, Toronto, Canada. pp. 1255–1259.

Schwabedal, J.T., Snyder, J.C., Cakmak, A., Nemati, S., Clifford, G.D., 2018. Addressing class imbalance in classification problems of noisy signals by using fourier transform surrogates. arXiv preprint arXiv:1806.08675 .

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. Journal of Big Data 6, 1–48.

Stevenson, N.J., Tapani, K., Lauronen, L., Vanhatalo, S., 2019. A dataset of neonatal EEG recordings with seizure annotations. Scientific Data 6, 1–8.

Tangermann, M., Müller, K.R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., Leeb, R., Mehring, C., Miller, K.J., Mueller-Putz, G., et al., 2012. Review of the BCI Competition IV. Frontiers in Neuroscience 6, 55.

Thuwajit, P., Rangpong, P., Sawangjai, P., Autthasan, P., Chaisaen, R., Banluesombatkul, N., Boonchit, P., Tatsaringkansakul, N., Sudhawiyangkul, T., Wilaiprasitporn, T., 2022. EEGWaveNet: multiscale CNN-based spatiotemporal feature extraction for EEG seizure detection. IEEE Trans. on Industrial Informatics 18, 5547–5557.

Wang, F., Zhong, S.h., Peng, J., Jiang, J., Liu, Y., 2018. Data augmentation for EEG-based emotion recognition with deep convolutional neural networks, in: Proc. 24th Int'l Conf. Multimedia Modeling, Bangkok, Thailand. pp. 82–93.

Wang, Z., Zhang, W., Li, S., Chen, X., Wu, D., 2023. Unsupervised domain adaptation for cross-patient seizure classification. Journal of Neural Engineering, early access, 1–20.

Wu, D., Jiang, X., Peng, R., 2022. Transfer learning for motor imagery based brain-computer interfaces: A tutorial. Neural Networks 153, 235–253.

Wu, D., Lu, B.L., Hu, B., Zeng, Z., 2023. Affective brain–computer interfaces (aBCIs): a tutorial. Proc. of the IEEE , 1–19.

Wu, D., Xu, Y., Lu, B.L., 2020. Transfer learning for EEG-based brain–computer interfaces: A review of progress made since 2016. IEEE Trans. on Cognitive and Developmental Systems 14, 4–19.

Yao, D., Qin, Y., Hu, S., Dong, L., Bringas Vega, M.L., Valdés Sosa, P.A., 2019. Which reference should we use for EEG and ERP practice? Brain topography 32, 530–549.

Zanini, P., Congedo, M., Jutten, C., Said, S., Berthoumieu, Y., 2018. Transfer learning: A Riemannian geometry framework with applications to brain–computer interfaces. IEEE Trans. on Biomedical Engineering 65, 1107–1116.

Zhang, W., Wang, Z., Wu, D., 2022. Multi-source decentralized transfer for privacy-preserving BCIs. IEEE Trans. on Neural Systems and Rehabilitation Engineering 30, 2710–2720.