EEG-Infinity: A Mathematical Modeling-Inspired Architecture for Addressing Cross-Device Challenges in Motor Imagery

Chengxuan Qin, Rui Yang, *Senior Member, IEEE*, Longsheng Zhu, Zhige Chen, *Student Member, IEEE*, Mengjie Huang, *Member, IEEE*, Fuad E. Alsaadi, and Zidong Wang, *Fellow, IEEE*

***Abstract*— The distribution of electroencephalogram (EEG) data generally varies across datasets due to the huge difference between the physical structure of brain-computer interface devices, known as cross-device variability. Such variability poses great challenges in EEG decoding and hinders the standardized utilization of EEG datasets. In this study, we explore a new issue concerning the cross-device variability problem, pointing to the gap in the existing studies facing cross-device variability. To tackle this challenge, our paper is the first to model the cross-device variability problem through a "sequentially comprehensive formula" and a "spatial comprehensive formula". Inspired by this modeling, a novel deep domain adaptation network named EEG-Infinity is proposed, incorporating replaceable EEG feature extraction backbones with a novel structure named "alignment head". To show the effectiveness of the proposed EEG-Infinity, systematic experiments are conducted across four different EEG-based motor imagery datasets under 48 cases. The experimental results highlight the superior performance of the proposed EEG-Infinity over commonly used approaches with an average classification accuracy improvement of 1.51% across 34 cases, laying a foundation for research in large-scale EEG models. The code can be assessed at https://github.com/Baizhige/cd-infinity**

***Index Terms*— Electroencephalogram, cross-device variability, brain-computer-interface, mathematical modeling, transfer learning.**

Received 16 September 2024; revised 27 January 2025, 27 May 2025, and 21 October 2025; accepted 12 November 2025. Date of publication 19 November 2025; date of current version 2 December 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 72401233; in part by Jiangsu Provincial Scientific Research Center of Applied Mathematics under Grant BK20233002; in part by Jiangsu Provincial Qinglan Project; in part by the Natural Science Foundation of Jiangsu Higher Education Institutions of China under Grant 23KJB520038; in part by the Research Enhancement Fund of Xi'an Jiaotong-Liverpool University (XJTLU) under Grant REF-23-01-008; and in part by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, Saudi Arabia, under Grant GPIP194-135-2024. *(Corresponding authors: Rui Yang; Mengjie Huang.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the University Ethics Committee of Xi'an Jiaotong-Liverpool University.

Chengxuan Qin and Longsheng Zhu are with the School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China, and also with the School of Electrical Engineering, Electronics and Computer Science, University of Liverpool, L69 3BX Liverpool, U.K. (e-mail: C.Qin8@liverpool.ac.uk; Longsheng.Zhu22@alumni.xjtlu.edu.cn).

Rui Yang is with the School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China (e-mail: R.Yang@xjtlu.edu.cn).

Zhige Chen is with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, SAR, China (e-mail: zhige.chen@connect.polyu.hk).

Mengjie Huang is with the Design School, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China (e-mail: Mengjie.Huang@xjtlu.edu.cn).

Fuad E. Alsaadi is with the Communication Systems and Networks Research Group, Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia (e-mail: falsaadi@kau.edu.sa).

Zidong Wang is with the Department of Computer Science, Brunel University London, Uxbridge, UB8 3PH Middlesex, U.K. (e-mail: Zidong.Wang@brunel.ac.uk).

Digital Object Identifier 10.1109/TNSRE.2025.3635018

I. Introduction

THE brain-computer interface (BCI) is a system designed to establish a direct pathway between human neural activity and the external environment [1], serving various applications such as disease diagnosis [2], control of external devices [3], and sleep stage detection [4], [5]. Electroencephalogram (EEG) data is commonly utilized in BCI, involving the strategic placement of electrodes on the scalp to detect voltage fluctuations resulting from neuronal activity [1], [6]. EEG has emerged as a significant tool in the development of BCI systems due to its safety and high temporal resolution.

Enhancing the reliability of EEG signal decoding [7], [8], [9], [10] is critical for developing user-friendly BCI systems, where the EEG decoding represents the extraction of meaningful neural patterns associated with specific tasks in EEG signals. The key to distinctly observing neural patterns is introducing a carefully designed experimental setup—paradigm—aiming to activate specific brain activity patterns in the EEG acquisition. Motor imagery (MI) represents a specific EEG paradigm in which a subject imagines performing a physical movement without actual physical execution, thereby reflecting the brain activation during real movement. To date, MI decoding requires further research

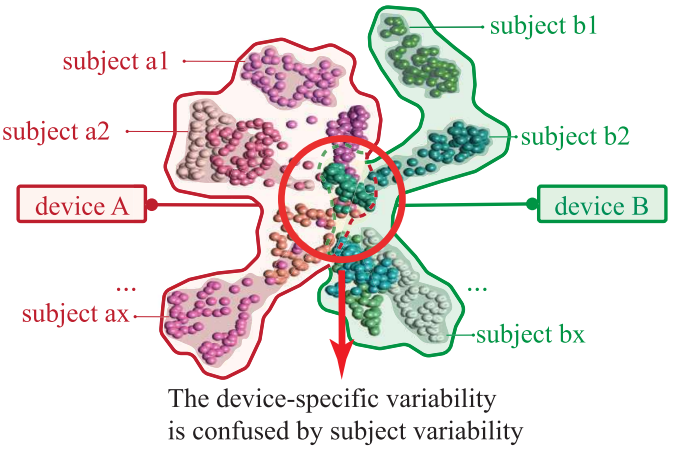

Fig. 1. Distribution Visualization in cross-device Issue with multiple subjects. The data utilized for visualization [22] are obtained from the openly available dataset Meng2019 [18]. The "device A" and "device B" refer to the Neuroscan SynAmps RT and the Biosemi Active Two EEG systems, respectively. The data from each device involves multiple subjects colored with different gradient colors. Each scatter point represents data from a single channel within the data for a single subject.

to promote the development of user-friendly BCI systems. A challenge in MI decoding arises from the absence of readily discernible features in the data, leading most decoding methods to rely on statistical approaches [11], [12], [13], [14]. In such a context, the quality of MI data is crucial for effective MI decoding. Unfortunately, the inherent variability of EEG data presents a significant obstacle, limiting the compatibility of statistical approaches across different contexts [1], [15]. The variability comes from various sources, such as individual differences in brain anatomy and function [15], [16], the dynamic nature of cognitive states [16], and variations in BCI systems [17], [18]. Addressing the issue of data variability is vital for advancing MI decoding techniques and significantly improving the efficacy and user-friendliness of BCI systems.

Variability in MI data is evident across several dimensions, such as cross-subject, cross-session, and cross-device differences [19]. The cross-device issue typically involves processing extensive data characterized by multiple sources of variability, complicating the observation of cross-device differences, including not only differences between devices but also cross-subject [16] or cross-session variability [20]. Fig. 1 demonstrates how the cross-subject variability influences the observation of cross-device variability. In such a context, approaches addressing cross-device compatibility in MI research remain scarce, demonstrating a significant gap in the field. Inspired by the above discussion, the cross-device issue is complex and challenging to observe.

While substantial research efforts have aimed to mitigate cross-subject [15] and cross-session variability [20], investigations into cross-device variability are still limited. Only one study has addressed cross-device data variability with the utilization of solely three common channels [21]. Overall, the current limitations of existing MI cross-device studies can be summarized as the following two main aspects: (1) the issue of cross-device variability in the MI paradigm has not been effectively resolved, with most studies focusing on the MI paradigm within a dataset from a single device; (2) the variability in data from cross-device interactions has yet to be systematically characterized through mathematical modeling techniques. No previous study has focused on explaining how the difference across BCI devices influences EEG signals.

According to the systematic literature review on the structural components function, the cross-device variability can be summarized into two factors:

- **EEG caps factor:** The EEG cap is a device designed to measure electrical signals on the surface of the brain, consisting of an array of electrodes, leads, and a montage that determines the positions and numbers of electrodes [23]. Serving as the initial point of contact for brain electrical signal data entering the device, the EEG cap influences the spatiotemporal structure of the EEG signals. Electrodes with various materials, such as dry, wet, and gel electrodes [24], [25], possessing varying impedances [26], make acquired EEG signals different across devices. Additionally, electrodes specified for reference voltage and grounding [27] in amplification circuits play a role in determining the acquired EEG signals. The montage controls the placement and quantity of electrodes, but even though the quantity of electrodes is the same, variations in placement standards and individual device differences have the potential to cause shifts in the spatial information of EEG data [28]. In summary, the EEG cap significantly influences the spatiotemporal structure of EEG signals.
- **Integrated circuits factor:** The integrated circuit serves to preprocess EEG signals and consists of various components including but not limited to amplifiers, filters, and analog-to-digital converters (ADCs) [23]. Once the electrical signals from the brain are collected by the EEG cap, they are transmitted to the integrated circuit for processes such as amplification, filtering, denoising, and analog-to-digital conversion [25], [29], [30], [31]. The configuration of amplifiers can vary from single-stage to multi-stage amplification, each with distinct characteristics [25], [29]. Filter implementation [32], [33] is not standardized, and factors such as the capacitance values of high-pass filters may differ. Different ADCs exhibit varying levels of precision and quantization errors. Moreover, the thermal and Flicker noise of electronic components can vary [29], [34], [35]. In summary, the design of integrated circuits can significantly influence the final acquired EEG signal, impacting aspects such as amplification ratio, amplitude and phase across different frequency bands, and introduced noise.

Because of the aforementioned two factors, the cross-device variability persists in practical MI experiments, hindering the utilization of EEG datasets across devices. To mitigate the cross-device variability and improve the performance of the existing cross-device decoding model, we propose a novel network named EEG-Infinity, incorporating the alignment head with different feature extractions. The main contributions of this paper are summarized as follows:

- Systematically demonstrating, for the first time, the impact of cross-device variability on EEG signals, providing novel theoretical guidance for interfacing with multiple BCI systems;
- Introducing a novel alignment head that aligns data across BCI systems, offering a precise optimization pathway for addressing cross-device discrepancies;
- Proposing an end-to-end deep learning architecture named EEG-Infinity for MI classification tasks, with robust potential for classifying diverse datasets.

The remaining sections of this paper are organized as follows: Section II comprehensively details the cross-device issues and gives formulas to demonstrate how cross-device settings influence EEG signals. Section III describes the proposed method of EEG-Infinity with an alignment head. Section IV details the experimental evaluations, including comparison and ablation studies, controlled analyses on sample size, ERD/ERS patterns, channel configuration, single-channel settings, electrode-contact robustness, paradigm generalization, and complexity analysis. Section V provides an overall discussion of the findings and limitations. Ultimately, Section VI provides the conclusion of the paper and outlines prospects for future research directions.

II. Mathematical Modeling for Cross-Device Variability

This section explores the impact of cross-device variability on EEG signals in three parts. The first part introduces the typical setup for MI-based BCI devices and provides an overview of the EEG data acquisition process. The second part focuses on analyzing single-channel EEG signal processing and illustrates how variations in electronic components affect the signals. The third part delves into the complexities of handling multiple data channels simultaneously, demonstrating the multi-channel EEG signal processing. At present, only the aggregate impact of cross-device variability can be characterized and mitigated, while its underlying components remain entangled.

A. Typical Setup for MI-Based BCI Devices

The MI-based BCI device acquires signals through two core modules: the EEG cap and the signal-processing integrated circuit, as depicted in Fig. 2. These core modules consist of the following sub-modules [25], [29], [34], [36]: the EEG cap consists of an array of electrodes, leads, and a montage that controls the placement of the electrodes, and the signal-processing integrated circuit includes a series of amplifiers, filters, ADCs, and signal processors.

The above description outlines a typical BCI device setup, and the sequence of "brain → electrode → lead → amplifier-filter-ADC → signal processor" forms the foundational hypothesis in this study, as depicted in Fig. 3. It is important to acknowledge that variations in the combination and sequence of these modules may be adopted, reflecting diverse design principles inherent in EEG acquisition devices [29]. For instance, the study [29] discusses a multi-stage amplification circuit structure as a strategy to mitigate noise amplification [34], [37]. Despite diverse design principles, the proposed modeling methods in this study remain applicable to such alternative structures.

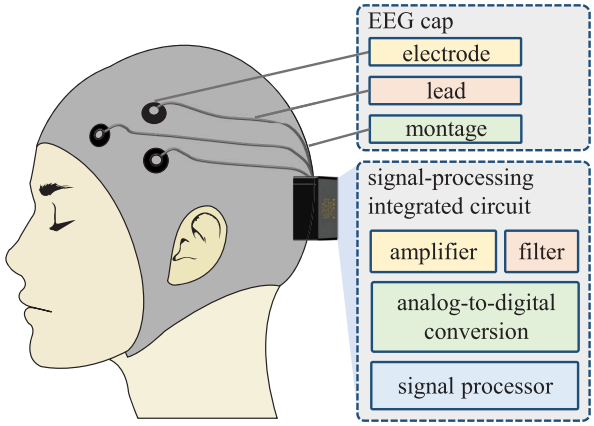

Fig. 2. Illustration for typical MI-based BCI devices setup.

B. Analysis of Single-Channel EEG Signal Acquisition

This part describes the acquisition of single-channel EEG signals. The original voltage signal from the brain is denoted as $U_{bra}(t)$, and the acquired signal voltage is denoted as $U_{out}(t)$. The relationship between $U_{bra}(t)$ and $U_{out}(t)$ is comprehensively detailed in subsequent stages. The first stage is the voltage dividing stage illustrated in Fig. 3A, including electrode contact impedance Z_e, lead impedance Z_l, and amplifier input impedance Z_{in}. The $U_{bra}(t)$ passes through the electrode and lead impedance before reaching the amplifier, forming a voltage divider. Consequently, the input voltage of amplifier $U_{in}(t)$ can be represented by $U_{bra}(t)$ using (1):

$$U_{in}(t) = U_{bra}(t) \times \frac{Z_{in}}{Z_{in} + Z_e + Z_l} + N_{in}(t) \tag{1}$$

where $N_{in}(t)$ represents the input noise of the voltage divider [29], [34].

The second stage is the amplifier stage, as depicted in Fig. 3B, where the input voltage U_{in} is amplified to produce U_{amp}. The amplifier applies varying amplification factors, $A(f)$, to signals of different frequencies, making the amplification factor a frequency-dependent variable. For clarity, the mathematical expression for the amplifier stage is presented in the frequency domain, as described in (2):

$$U_{amp}(f) = A(f) \times U_{in}(f) + N_{amp}(f) \tag{2}$$

where $U_{in}(f)$ and $N_{amp}(f)$ represent the input voltage and the noise of the amplifier respectively [29], [34], and $A(f)$ denotes the frequency-dependent amplification factor in a practical and non-ideal amplifier application.

The third stage is the filtering stage illustrated in Fig. 3C, where the amplified voltage $U_{amp}(f)$ passes through the filter to remove unnecessary components, resulting in the filtered signal $U_{fil}(f)$, as described in (3):

$$U_{fil}(f) = H(f, \phi(f)) \times U_{amp}(f) + N_{fil}(f) \tag{3}$$

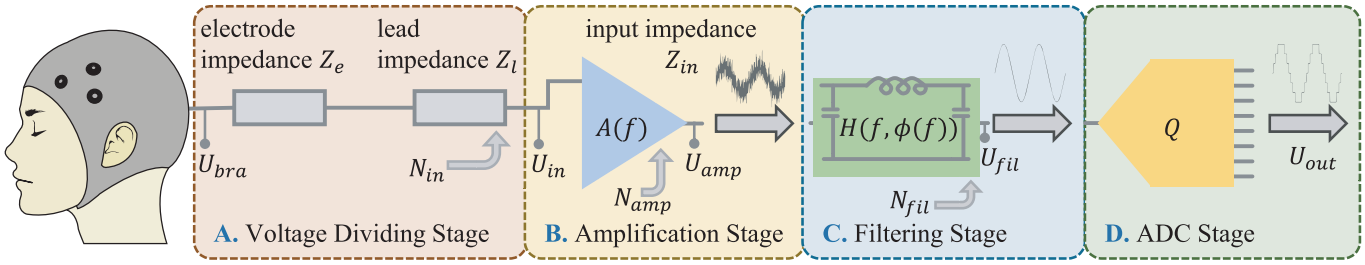

Fig. 3. Illustration for mathematical modeling for EEG acquisition.

where $H(f,\phi(f))$ and $N_{fil}(f)$ represent the frequency and phase response and the noise of the non-ideal filter, respectively.

The fourth stage involves the analog-to-digital conversion depicted in Fig. 3D. In this stage, the filtered signal is digitized as the actual output of signal-processing integrated circuit $U_{out}(t)$, as described in (4):

$$U_{out}(t) = Q\left(U_{fil}(t)\right) \tag{4}$$

where $Q(\cdot)$ denotes the quantization operation.

By integrating the four stages mentioned above, the relationship between the output of signal-processing integrated circuit $U_{out}(t)$ and the EEG signal from brain $U_{bra}(t)$ can be represented as (5):

$$U_{out}(t) = Q(\mathcal{F}^{-1}(H(f,\phi(f)) \times (A(f) \times \mathcal{F}(U_{bra}(t) \times \frac{Z_{in}}{Z_{in}+Z_e+Z_l} + N_{in}(t)) + N_{amp}(f)) + N_{fil}(f))) \tag{5}$$

The EEG signal acquisition process, as described in (5), involves multiple transformations encompassing various categories of analogous items. For example, $N_{in}(f)$, $N_{amp}(f)$, and $N_{fil}(f)$ all serve as noise components, while $H(f,\phi(f))$, $A(f)$, and $\frac{Z_{in}}{Z_{in}+Z_e+Z_l}$ exhibit amplification and filtering effects. To capture the aggregate impact of cross-device variability without disentangling its underlying components, functionally similar processes are aggregated into abstract terms. This facilitates a concise representation of the signal acquisition process and enables the derivation of a sequentially interpretable formulation, as presented in (6):

$$\begin{aligned} U_{out}(t) &= I_{<Q,\alpha,W,N>}(U_{bra}) \\ &= Q\left(\mathcal{F}^{-1}\{\alpha \times W(f,\phi(f)) \times \mathcal{F}\{U_{bra}(t)\} + N(f)\}\right) \end{aligned} \tag{6}$$

The sequentially comprehensive formula reveals the transformations involved in signal acquisition under non-ideal conditions. To further explain the formula, Fig. 4 shows the frequency-amplitude comparison between the Neuroscan SynAmps RT and Biosemi Active Two EEG systems, where the differences can be described by parameters $Q(\cdot)$, α, $W(f,\phi(f))$, and $N(f)$ in (6). For example, the Neuroscan system exhibits consistently higher magnitude from the delta through alpha range, which may be attributed to the scaling factor α. Variations in Shape across these bands could result from the filter $W(f,\phi(f))$. And spikes in the Beta band (such as spikes at around 8 Hz and 27.6 Hz) may be a result of noise N in different devices. Despite identical task protocols and bilaterally symmetric References, these differences persist,

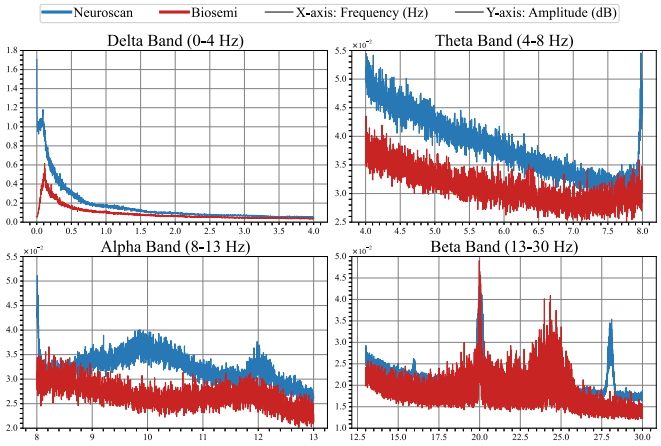

Fig. 4. Frequency-magnitude comparison between Neuroscan SynAmps RT and Biosemi active two EEG systems. The data utilized for visualization are obtained from the publicly available dataset Meng and He [18]. For each recording, we included all EEG channels and files, applied an FFT to each channel, took the magnitude, and interpolated the spectra onto a common frequency grid.

indicating systematic characteristics of the acquisition chains rather than task-induced effects.

Considering the general context, let's assume that there are two devices, s and t, from the source and target domain datasets, respectively. The transformations in these different devices can be simplified as a model I^s or I^t, involving four principal aspects: analog-to-digital conversion $Q(\cdot)$, scaling factor α, filtering $W(f,\phi(f))$, and noise $N(f)$. By consolidating various analogous items into these four aspects, the sequentially comprehensive formula provides a clear framework for understanding and addressing the variations in EEG data acquisition across different devices.

C. Analysis of Multi-Channel EEG Signal Processing

This part describes the processing of multi-channel EEG signals. Let the acquired multi-channel EEG signal be X_{out}:

$$\begin{aligned} X_{out} &= \begin{bmatrix} U^1_{out} \\ \vdots \\ U^c_{out} \end{bmatrix} = \begin{bmatrix} I_{<Q,\alpha,W,N>}(U^1_{bra}) \\ \vdots \\ I_{<Q,\alpha,W,N>}(U^c_{bra}) \end{bmatrix} \\ &= I_{<Q,\alpha,W,N>}(X_{bra}) \end{aligned} \tag{7}$$

where c is the number of channels, and U^i_{out} and U^i_{bra} are the acquired and brain EEG signal in the i^{th} channel respectively. The spatial distribution of multi-channel EEG signals X_{out} is

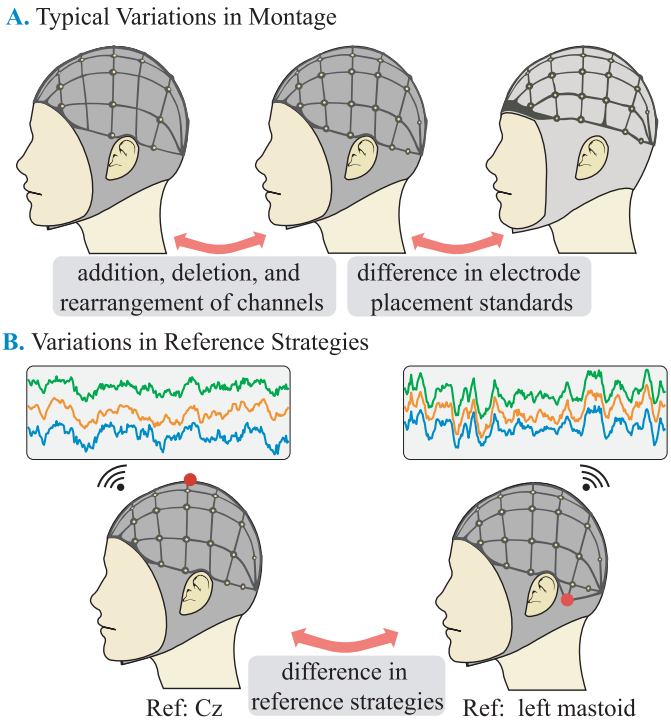

Fig. 5. Illustration for variations in multi-channel EEG signal.

significantly influenced by montage and reference electrode configuration, as explained below:

- The placement of electrodes is controlled by the montage [38]. In MI-based BCI devices, typical variations in montage include the addition, deletion, and rearrangement of channels and the difference in electrode placement standards [38], [39], as illustrated in Fig. 5A. The electrical signal recorded by each electrode is affected by its relative position to specific cortical areas of the brain; thereby, the placement of electrodes would straightforwardly result in spatial transformations to MI data.
- The reference electrode serves as a reference voltage for each channel. In MI-based BCI devices, common variations include the selection of reference electrodes [18], [40], [42], and the application of reference techniques [43], [44], as illustrated in Fig. 5B. Considering the influence on spatial distribution, reference electrodes, and techniques can be treated as a linear transformation applied to multi-channel EEG signals [43], [44].

A spatial comprehensive formula (8) is proposed to systematically express the influence of montage and reference electrode configurations on the EEG signal across different devices. This formula incorporates linear transformation to approximate the impact on spatial distribution:

$$\begin{aligned} X^t_{out} &= \begin{bmatrix} U^{<t,1>}_{out} \\ \vdots \\ U^{<t,c>}_{out} \end{bmatrix} \\ &= T_{<s,t>} \begin{bmatrix} U^{<s,1>}_{out} \\ \vdots \\ U^{<s,c>}_{out} \end{bmatrix} + \epsilon_{<s,t>} \\ &= T_{<s,t>} X^s_{out} + \epsilon_{<s,t>} \end{aligned} \tag{8}$$

where X^s_{out} and X^t_{out} denote the multi-channel EEG signals obtained from devices s and t respectively, $U^{<s,i>}_{out}$ and $U^{<t,i>}_{out}$ represent the EEG signals acquired from the i^{th} channel of devices s and t respectively, and $T_{<s,t>}$ and $\epsilon_{<s,t>}$ denote the linear transformation and approximation error from device s to device t respectively.

III. Proposed EEG-Infinity for Cross-Device Variability

This section presents the methodology for addressing cross-device variability—EEG-Infinity. The proposed EEG-Infinity employs transfer learning in both data and feature spaces, following a systematic learning schedule as illustrated in Fig. 6. The subsequent contents discuss the strategies for data alignment, feature alignment, and optimization process of EEG-Infinity.

A. Data Alignment by Alignment Head

A novel structure called alignment head is introduced in the data space to synchronize EEG data across spatial and time-frequency domains. This is illustrated in the first yellow-highlighted part in section A of Fig. 6, where a spatial filtering block (SFB) is designed to align the spatial distribution in EEG data. The SFB incorporates both prior spatial transformations (T^s_p and T^t_p) and fine-tuning spatial transformations (T^s_f and T^t_f). While these transformations employ similar matrix operations, they serve different roles within the model. Initially, the prior spatial transformations pre-align the data using predetermined weights derived from prior knowledge, including electrode information and interpolation algorithms. In contrast, the fine-tuning spatial transformations start with identity weights and are adjusted during training to correct any errors from the initial alignment. To effectively train the fine-tuning transformation matrix, a loss is computed using the transformed data from both the source and target domains, as shown in (9):

$$\begin{aligned} \mathcal{L}_{cov} &= \mathbb{E}\left[\|T^s_f T^s_p X^s_{out} - T^t_f T^t_p X^t_{out}\|\right] \\ &= \mathbb{E}\left[\|T^s_f T^s_p X^s_{out} - T^t_f T^t_p \left(T_{<s,t>} X^s_{out} + \epsilon_{<s,t>}\right)\|\right] \\ &= \mathbb{E}\left[\|T^s_f T^s_p X^s_{out} - T^t_f T^t_p T_{<s,t>} X^s_{out} + \epsilon'_{<s,t>}\|\right] \end{aligned} \tag{9}$$

where T^s_p and T^s_f refer to the prior and fine-tuning transformation matrices for the source domain, T^t_p and T^t_f refer to those for the target domain, $\|\cdot\|$ represents the measurement of distance, and $\mathbb{E}[\cdot]$ represents the expectation. Regarding $\mathcal{L}_{cov}$, the measurement of distance is computed via the cosine distance of row vectors in the average covariance matrix of mini-batches from the source and target domains. When minimizing $\mathcal{L}_{cov}$, the goal is to minimize the difference between transformations $T^s_f T^s_p$ and $T^t_f T^t_p T_{<s,t>}$.

As shown in the second yellow-highlighted part in section A of Fig. 6, a frequency filtering block (FFB) is designed to filter noise introduced by $I^s_{<Q,\alpha,W,N>}$ and $I^t_{<Q,\alpha,W,N>}$. The FFB utilizes a convolutional kernel, as the convolution theorem suggests that one-dimensional convolution can be conceptualized as a finite impulse response filter with adjustable weights [45]. Due to the difficulties in obtaining $I^s_{<Q,\alpha,W,N>}$ and $I^t_{<Q,\alpha,W,N>}$, data

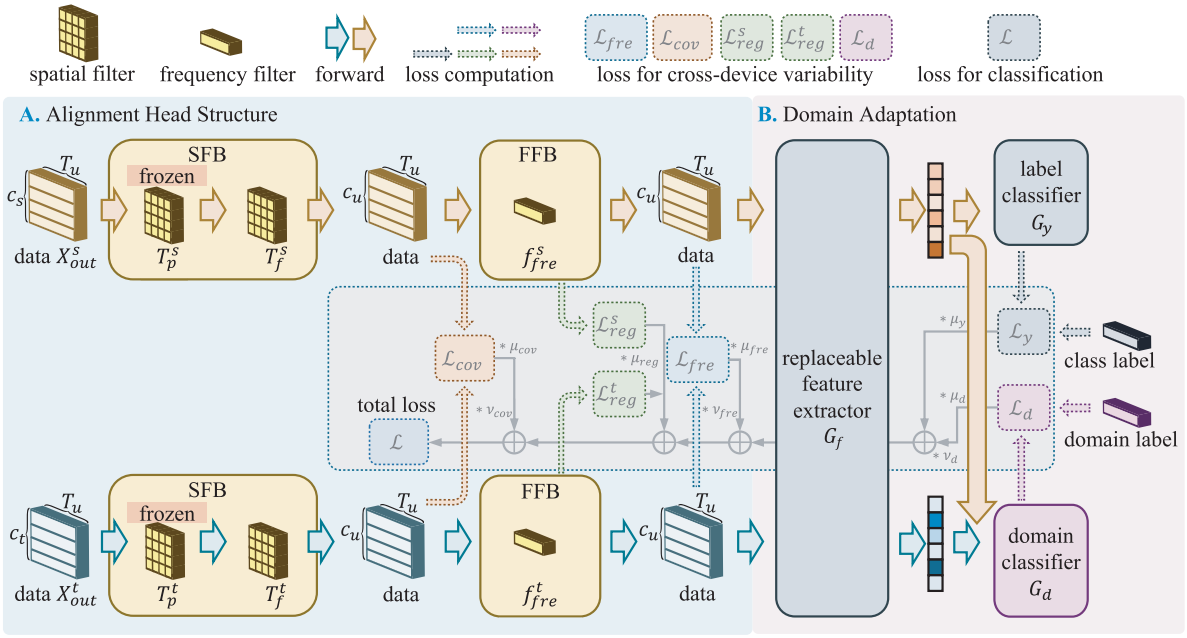

Fig. 6. Overview of EEG-infinity architecture.

from both source and target domains are utilized to compute a loss and find a solution, as shown in (10):

$$\begin{aligned}\mathcal{L}_{fre} &= \mathbb{E}\left[\| f^s_{fre}\left(T^s_f T^s_p X^s_{out}\right) - f^t_{fre}\left(T^t_f T^t_p X^t_{out}\right)\|\right] \\ &= \mathbb{E}\left[\| f^s_{fre}\left(T^s_f T^s_p I^s\left(X^s_{bra}\right)\right) - f^t_{fre}\left(T^t_f T^t_p I^t\left(X^t_{bra}\right)\right)\|\right] \end{aligned} \tag{10}$$

Regarding $\mathcal{L}_{fre}$, the distance measurement is computed by taking the average L1 distance between mini-batches from the source and target domains. During the minimization process of $\mathcal{L}_{fre}$, it is possible to reduce the difference between the filtering effects I^s and I^t. Furthermore, to prevent the introduction of a direct current bias or the attenuation of the EEG signal, the use of regularization losses $\mathcal{L}^s_{reg}$ and $\mathcal{L}^t_{reg}$ is necessary, as shown in (11) and (12):

$$\mathcal{L}^s_{reg} = \left(\sum_i^z \left[w^s_i\right] - 1\right)^2 \tag{11}$$

$$\mathcal{L}^t_{reg} = \left(\sum_i^z \left[w^t_i\right] - 1\right)^2 \tag{12}$$

where w^s_i and w^t_i denote the parameters of the convolutional kernel for the source and target domains, respectively.

B. Feature Alignment by Wasserstein Guided Domain Adaptation

To address the nonlinear effects inherent in cross-device variability captured by $N(f)$ and $\epsilon_{<s,t>}$, a Wasserstein-guided domain adaptation approach [46], [47], [48], [49] is utilized to align features within unified representations, as illustrated in section B of Fig. 6. This strategy specifically targets the removal of noise components N and $\epsilon_{<s,t>}$ that cannot be eliminated through spatial transformation and filtering techniques. The architecture for Wasserstein-guided domain adaptation comprises three key components [46]: a feature extractor for obtaining domain-representative features from EEG data, a label classifier for predicting task-related labels using the extracted features, and a domain classifier for determining the domain of the extracted features and measuring the feature distance between source and target domains.

For concurrent training of the feature extractor and the domain classifier, a gradient reversal layer is employed to reverse the gradients during training. In this study, any commonly used network in EEG research can serve as the feature extractor [11], [50], [51]. Both the label and domain classifiers are implemented using multi-layer perceptions, with the classification loss denoted by $\mathcal{L}_y$ in (13):

$$\begin{aligned}\mathcal{L}_y = -\mathbb{E}\big[&y \log\left(G_y\left(G_f\left(f^t_{fre}\left(T^t_f T^t_p X^t_{out}\right)\right)\right)\right) \\ &+ (1-y)\log\left(1 - G_y\left(G_f\left(f^t_{fre}\left(T^t_f T^t_p X^t_{out}\right)\right)\right)\right)\big]\end{aligned} \tag{13}$$

where y denotes the true label, G_y denotes the label classifier, and G_f denotes the feature extractor.

The domain classifier is trained to differentiate between samples from source and target domains, with the loss of domain classifier serving as an approximation of the Wasserstein distance [47], [48], [52] between the two distributions, as shown in (14):

$$\begin{aligned}\mathcal{L}_d = &\mathbb{E}\left[G_d\left(G_f\left(f^s_{fre}\left(T^s_f T^s_p X^s_{out}\right)\right)\right)\right] \\ &- \mathbb{E}\left[G_d\left(G_f\left(f^t_{fre}\left(T^t_f T^t_p X^t_{out}\right)\right)\right)\right]\end{aligned} \tag{14}$$

where G_d denotes the domain classifier.

Remark 1: Ensuring adherence to the 1-Lipschitz continuity condition is crucial for accurate approximation throughout the training process of the domain classifier. In this study, weight clipping during the training phase of the domain classifier fulfills such condition [47].

C. Optimization Process

The optimization process of EEG-Infinity involves a variety of loss functions. During the training process, the loss

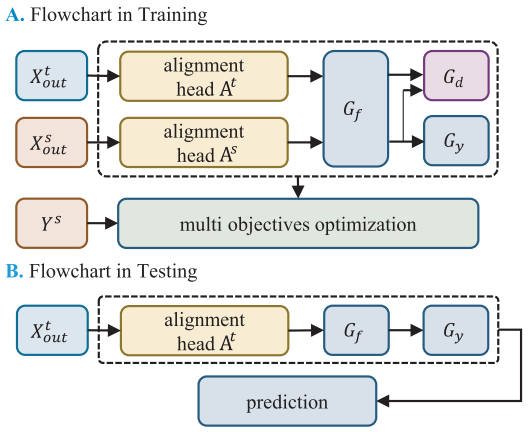

Fig. 7. Flowchart of proposed approach.

functions can be categorized into three types based on their objectives:

- classification loss $\mathcal{L}_y$;
- alignment losses, including $\mathcal{L}_{cov}$, $\mathcal{L}_{fre}$, and $\mathcal{L}_d$;
- regularization losses, $\mathcal{L}^s_{reg}$ and $\mathcal{L}^t_{reg}$.

In previous studies, solutions have been proposed to handle the challenge of dealing with multiple loss functions [53], [54], [55], [56]. Inspired by these preliminary studies, a specific multi-loss function training [57], [58] framework is designed for EEG-Infinity, incorporating both scaling coefficients and adjustment weights:

- Scaling coefficients: Each task-specific loss is initially scaled by a corresponding coefficient to ensure that gradient magnitudes are of comparable scale. The coefficients are estimated by computing the average gradient magnitude for each task during a testing epoch.
- Adjustment weights: In addition to initial scaling, each loss is multiplied by an adjustment weight. During training, the adjustment weights for classification and regularization losses remain fixed at 1. In contrast, alignment losses are assigned varying adjustment weights to form a composite loss. The adjustment weights allow the contributions of $\mathcal{L}_{cov}$, $\mathcal{L}_{fre}$, and $\mathcal{L}_d$ to be emphasized in a specific sequence throughout the training process.

In summary, the loss function of EEG-Infinity is constructed as illustrated in (15):

$$\mathcal{L} = \begin{bmatrix} \mu_y \\ \mu_{reg} \\ \mu_{cov} \\ \mu_{fre} \\ \mu_d \end{bmatrix}^\top \begin{bmatrix} \mathcal{L}_y \\ \mathcal{L}_{reg} \\ v_{cov}\mathcal{L}_{cov} \\ v_{fre}\mathcal{L}_{fre} \\ v_d\mathcal{L}_d \end{bmatrix} \tag{15}$$

where $\mathcal{L}_{reg} = \mathcal{L}^s_{reg} + \mathcal{L}^t_{reg}$. The scaling coefficients for each loss are represented by μ_y, μ_{reg}, μ_{cov}, μ_{fre}, and μ_d, while the adjustment weights for each loss are represented by v_{cov}, v_{fre}, and v_d. The alignment heads are designed to align the EEG data from different devices rather than for classification purposes; therefore, the gradients of alignment heads derived from $\mathcal{L}_y$ are set to zero.

Fig. 7 outlines the flowchart of the proposed approach. In the training phase, as shown in Fig. 7A, the EEG data X^s_{out} and X^t_{out} are aligned by alignment heads, then processed by feature extractor G_f. The features feed into G_d for domain discrimination and G_y for label prediction. The weights of the network are optimized via multi-objective optimization. In the testing phase, as shown in Fig. 7B, inputs X^t_{out} go through the alignment head and G_f, then G_y for prediction, skipping domain classifier.

Particularly, the total partial derivative with respect to the alignment head for the source and target domains, denoted as A^s/A^t, can be expressed as follows:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \theta_{A^s}} &= \begin{bmatrix} \mu_{\text{reg}} \\ \mu_{\text{cov}} \\ \mu_{\text{fre}} \\ \mu_d \end{bmatrix}^\top \begin{bmatrix} \frac{\partial \mathcal{L}_{\text{reg}}}{\partial \theta_{A^s}} \\ v_{\text{cov}} \frac{\partial \mathcal{L}_{\text{cov}}}{\partial \theta_{A^s}} \\ v_{\text{fre}} \frac{\partial \mathcal{L}_{\text{fre}}}{\partial \theta_{A^s}} \\ v_d \frac{\partial \mathcal{L}_d}{\partial \theta_{A^s}} \end{bmatrix} \\ \frac{\partial \mathcal{L}}{\partial \theta_{A^t}} &= \begin{bmatrix} \mu_{\text{reg}} \\ \mu_{\text{cov}} \\ \mu_{\text{fre}} \\ \mu_d \end{bmatrix}^\top \begin{bmatrix} \frac{\partial \mathcal{L}_{\text{reg}}}{\partial \theta_{A^t}} \\ v_{\text{cov}} \frac{\partial \mathcal{L}_{\text{cov}}}{\partial \theta_{A^t}} \\ v_{\text{fre}} \frac{\partial \mathcal{L}_{\text{fre}}}{\partial \theta_{A^t}} \\ v_d \frac{\partial \mathcal{L}_d}{\partial \theta_{A^t}} \end{bmatrix} \end{aligned} \tag{16}$$

where θ_{A^s} and θ_{A^t} denote the parameter of alignment heads of source and target domains, respectively.

When using stochastic gradient descent (SGD) as an optimizer, the parameter updates for A^s and A^t during one mini-batch can be concisely expressed as:

$$\begin{aligned} \theta^{(t+1)}_{A^s} &= \theta^{(t)}_{A^s} - \eta \begin{bmatrix} \mu_{\text{reg}} \\ \mu_{\text{cov}} \\ \mu_{\text{fre}} \\ \mu_d \end{bmatrix}^\top \begin{bmatrix} \frac{\partial \mathcal{L}_{\text{reg}}}{\partial \theta_{A^s}} \\ v_{\text{cov}} \frac{\partial \mathcal{L}_{\text{cov}}}{\partial \theta_{A^s}} \\ v_{\text{fre}} \frac{\partial \mathcal{L}_{\text{fre}}}{\partial \theta_{A^s}} \\ v_d \frac{\partial \mathcal{L}_d}{\partial \theta_{A^s}} \end{bmatrix} \\ \theta^{(t+1)}_{A^t} &= \theta^{(t)}_{A^t} - \eta \begin{bmatrix} \mu_{\text{reg}} \\ \mu_{\text{cov}} \\ \mu_{\text{fre}} \\ \mu_d \end{bmatrix}^\top \begin{bmatrix} \frac{\partial \mathcal{L}_{\text{reg}}}{\partial \theta_{A^t}} \\ v_{\text{cov}} \frac{\partial \mathcal{L}_{\text{cov}}}{\partial \theta_{A^t}} \\ v_{\text{fre}} \frac{\partial \mathcal{L}_{\text{fre}}}{\partial \theta_{A^t}} \\ v_d \frac{\partial \mathcal{L}_d}{\partial \theta_{A^t}} \end{bmatrix} \end{aligned} \tag{17}$$

where $\theta^{(t)}_{A^s}$ and $\theta^{(t)}_{A^t}$ denote the parameters of the alignment modules A^s and A^t before the t-th iteration respectively and η is the learning rate. The pseudocode for training and testing the EEG-Infinity model is shown in Algorithm 1.

IV. Experiment and Result Analysis

A. Dataset Description and Experimental Setting

To fully evaluate the effectiveness of the proposed approach, four MI datasets with similar tasks but different devices are used in this study, including BCI Competition IV 2a (B), Meng2019 experiment 3 (M3), Meng2019 experiments 1/2 (M12) (with differences between M3 and M12 in the devices and subjects involved), and PhysioNet motor imagery (P). Considering cross-device issues, only data related to similar tasks from each dataset are extracted, specifically trials involving tasks related to left and right hands. All data are downsampled to 128 Hz for consistency, and the length of the segmentation window is set to 3 seconds, as a compromise that is suitable for the majority of the datasets. A summarized detail is available in Table I. All datasets are divided using 5-fold cross-validation, with the training, validation, and testing

Algorithm 1 Training and Testing EEG-Infinity Model

Input: X^s_{out}, Y^s, X^t_{out}, T^s_p, T^t_p, model $\mathcal{M}$, max epochs E, optimizer $\mathcal{O}$
Output: Trained model $\mathcal{M}^*$
Initialize: Initialize $\mathcal{M}$ with T^s_p, T^t_p, and optimizer $\mathcal{O}$
for epoch $\leftarrow 1$ to E **do**
  for each minibatch $(X^s_{batch}, Y^s_{batch})$ and (X^t_{batch}) in X^s_{out}, Y^s and X^t_{out} **do**
    Forward Pass (Source Domain):
    Compute model output $\hat{Y}^s = \mathcal{M}(X^s_{batch}, T^s_p)$
    Compute Losses (Source Domain):
    Compute $\mathcal{L}_{cov}$, $\mathcal{L}_{fre}$, $\mathcal{L}^s_{reg}$, $\mathcal{L}_y$, $\mathcal{L}_d$ and $\mathcal{L}$ through (9)-(15)
    Backward Pass (Source Domain):
    Backpropagate gradients and delete $\mathcal{L}$
    Retain the gradients of model $\mathcal{M}$
    Set gradients of alignment head with respect to L_y to zero
    Forward Pass (Target Domain):
    Compute model output $\hat{Y}^t = \mathcal{M}(X^t_{batch}, T^t_p)$
    Compute Losses (Target Domain):
    Compute $\mathcal{L}_{cov}$, $\mathcal{L}_{fre}$, $\mathcal{L}^s_{reg}$, $\mathcal{L}_d$ and $\mathcal{L}$ through (9)-12, 14-15)
    Backward Pass (Target Domain):
    Backpropagate gradients and delete $\mathcal{L}$
    Update Model Parameters:
    update $\mathcal{M}$ by $\mathcal{O}$
    set gradients to zeros
  end for
  Validation:
  Evaluate $\mathcal{M}$ on validation set, save best parameters
end for
Testing:
Load best parameters into $\mathcal{M}^*$, evaluate on test set

datasets set at a ratio of 6:2:2. In the experiments, the performance of the method across all datasets is evaluated using the weighted accuracy metric Acc_w, incorporating weights based on the sample size of each dataset:

$$Acc_w = \frac{\sum_{i=1}^{m} w_i Acc_i}{\sum_{i=1}^{m} w_i} \tag{18}$$

where w_i represents the number of trials in the target domain for the i-th transfer case, and Acc_i represents the accuracy of the model on the i-th cases.

In both the comparison and ablation studies, the batch size and max epoch are set to 32 and 100, respectively. Additionally, the stochastic gradient descent optimizer with momentum is employed to train the model, with the learning rate μ specified by (19):

$$\mu = \frac{\mu_0}{1 + (\alpha p)^\beta} \tag{19}$$

where μ_0, β, and α are constant, and p denotes the progress of the training process, which ranges from 0 to 1.

TABLE I
DESCRIPTION OF DATASETS USED IN EXPERIMENTS

Dataset	#Subject / Channel	Device Manufacturer	#Trials	Hz	MI Class
B [40]	9 / 22	g.tec (Ag/AgCl) [41]	1296	250	right / left hand
M3 [18]	8 / 64	Neuroscan SynAmps RT	4923	1024	right / left cursor
M12 [18]	30 / 62	Biosemi Active Two	10045	1000	right / left cursor
P [42]	109 / 64	not provided (BCI2000)	4748	160	right / left fist

B. Comparison Study

1) Comparison Study Setting: A comparison study is designed to evaluate the performance of the proposed approach against four typical transfer learning approaches used in the field of EEG:

- Deep Domain Confusion (DDC) [59] aims to minimize discrepancies between domains by aligning the features of the source and target domains, utilizing a domain confusion loss.
- Deep Coral [60] addresses domain adaptation by minimizing the disparities in the second-order statistical characteristics between the source and target domain features.
- Domain-Adversarial Neural Network (DANN) [46] adopts an adversarial strategy for domain adaptation inspired by generative adversarial networks, incorporating a gradient reversal layer designed to smooth the training process.
- Wass-DANN [49], [52], an extension of DANN, integrates the Wasserstein distance into the adversarial training framework, providing a robust and theoretical sound training process.
- DANet [61] addresses domain adaptation by minimizing the discrepancy between the processed EEG data (obtained through a neural network) and the original EEG signals using the mean squared error metric.
- PSAT [62], which stands for prototype-supervised adversarial transfer learning, employs a domain discriminator to reduce discrepancies and a prototype mapper to constrain non-stationarity.

Prior to applying traditional transfer learning methods, additional preprocessing steps must be taken due to discrepancies in data structures. This involves aligning the channels from the source domain to match those of the target domain, and utilizing linear interpolation to address any missing channels. This process is referred to as the "Baseline" configuration.

2) Comparison Study Result Analysis: The comparison study evaluated the performance of eight transfer learning configurations across 48 cases, incorporating 12 transfer cases and four backbones. The performance results for each case are detailed in Table II, with the best performance highlighted in bold and the second best underlined. The weighted accuracy is provided in the last column, taking into account the sample

TABLE II
EXPERIMENTAL RESULTS FOR COMPARISON STUDY (P-VALUE < 0.05, EEG-INFINITY VS. OTHERS, USING WILCOXON SIGNED-RANK TEST)

Methods	Backbones	B ↓ M3	B ↓ M12	B ↓ P	M3 ↓ B	M3 ↓ M12	M3 ↓ P	M12 ↓ B	M12 ↓ M3	M12 ↓ P	P ↓ B	P ↓ M3	P ↓ M12	Weighted Accuracy
Baseline	EEGNet	65.56	72.17	**70.28**	56.28	77.11	74.51	57.71	70.13	72.22	63.97	74.58	76.03	71.52
	ShallowConvNet	63.95	66.72	65.15	55.63	71.10	71.34	**65.79**	68.30	71.07	61.18	75.73	75.48	
	DeepConvNet	63.44	72.61	65.46	56.09	74.53	74.39	59.64	64.97	73.03	61.23	76.68	76.12	
	InceptionEEG	66.81	73.37	67.87	57.52	76.89	73.64	56.97	71.24	75.14	66.64	75.41	76.52	
DDC	EEGNet	66.48	70.98	67.09	56.75	76.54	73.94	61.84	72.16	76.32	63.51	73.92	75.70	71.80
	ShallowConvNet	63.85	67.57	64.95	58.41	73.17	74.02	62.89	68.28	71.32	67.64	75.45	75.91	
	DeepConvNet	63.91	**72.72**	62.04	57.94	73.73	75.88	59.83	67.83	75.71	62.85	75.97	77.21	
	InceptionEEG	67.49	71.33	68.60	59.49	77.10	74.07	61.03	72.81	75.75	65.48	74.86	75.60	
Deep Coral	EEGNet	65.08	71.59	65.52	56.16	75.74	71.43	57.52	70.82	70.12	61.93	75.83	74.33	68.38
	ShallowConvNet	57.52	72.44	66.87	53.35	71.85	71.45	53.49	69.91	72.42	54.46	66.91	70.50	
	DeepConvNet	54.20	50.71	51.74	55.47	69.96	70.93	57.21	70.05	72.83	58.83	74.78	73.84	
	InceptionEEG	57.48	62.89	59.61	57.60	76.62	72.48	56.62	71.18	71.30	63.01	75.65	74.72	
DANN	EEGNet	**66.87**	73.01	68.10	56.75	77.10	74.47	62.50	75.33	75.65	65.75	**75.85**	76.38	71.76
	ShallowConvNet	61.17	67.19	64.00	54.47	71.66	72.83	55.43	67.05	72.46	61.30	74.76	75.68	
	DeepConvNet	64.41	69.30	64.30	57.44	72.70	74.70	57.75	71.02	75.46	59.92	76.66	76.76	
	InceptionEEG	66.73	72.87	68.31	59.45	77.26	75.31	60.65	72.87	76.47	63.89	75.25	76.08	
DANN-Wass	EEGNet	64.81	**73.81**	66.60	57.75	76.63	73.46	60.38	72.49	77.48	62.89	74.46	75.58	71.76
	ShallowConvNet	63.20	69.11	65.56	54.35	71.84	71.60	56.05	63.60	69.49	59.79	75.17	75.98	
	DeepConvNet	63.42	71.95	65.42	57.98	73.83	75.04	60.06	71.46	76.36	61.92	**76.86**	76.71	
	InceptionEEG	65.34	73.79	66.73	59.88	77.76	74.74	58.13	72.04	74.97	65.32	75.33	76.08	
DANet	EEGNet	65.66	72.01	67.64	58.10	76.42	74.23	59.80	72.12	75.35	66.64	73.98	76.02	71.65
	ShallowConvNet	63.36	68.97	66.11	57.06	71.55	71.62	57.59	64.87	70.31	60.34	75.77	75.86	
	DeepConvNet	64.43	69.51	65.38	57.60	74.06	75.71	59.14	70.79	73.79	61.77	76.25	75.95	
	InceptionEEG	67.45	73.42	68.69	60.11	76.62	74.42	62.00	71.32	74.93	63.47	74.98	77.04	
PSAT	EEGNet	66.04	73.45	67.95	58.45	76.83	75.77	**63.86**	73.27	75.96	65.67	73.68	76.33	72.87
	ShallowConvNet	**67.49**	**74.03**	**68.58**	53.73	71.92	73.08	54.50	68.78	72.61	60.03	**76.35**	76.47	
	DeepConvNet	**67.69**	72.28	**67.32**	59.99	74.59	75.95	58.10	69.99	76.56	61.57	76.29	77.12	
	InceptionEEG	**69.36**	**75.49**	69.19	60.50	76.95	75.23	61.81	73.61	76.03	64.90	**75.99**	76.36	
EEG-Infinity	EEGNet	64.01	72.90	69.97	**64.78**	**79.31**	**79.72**	63.04	**75.65**	**78.45**	**70.12**	74.66	**78.87**	**74.34**
	ShallowConvNet	60.10	68.61	66.79	**64.55**	**79.08**	**78.92**	61.77	**73.63**	**79.02**	**74.48**	74.36	**78.86**	
	DeepConvNet	62.42	71.81	63.71	**64.82**	**78.74**	**77.23**	**66.75**	**74.32**	**77.87**	**69.15**	74.04	**78.07**	
	InceptionEEG	66.32	74.91	**72.27**	**63.08**	**79.35**	**78.72**	**63.16**	**76.98**	**79.21**	**74.95**	75.06	**78.50**	

size of the target domain as the weights. The experimental results are analyzed in terms of performance and applicability:

- From the aspect of performance, the EEG-Infinity model demonstrated superior performance in most cross-device cases. Specifically, the EEG-Infinity achieved the best performance in 31 out of 48 cases and ranked second-best in 3 cases. The above findings highlight the overall superiority of EEG-Infinity in handling a wide range of cross-device variations. The weighted accuracy results in Table II further support the leading performance of the proposed method in these cases. Particularly, comparing the experimental results of Wass-DANN and EEG-Infinity reveals that the proposed alignment head structure significantly improves the stability and efficiency of handling cross-device variability. This is particularly important as cross-device cases involve other variabilities whose distribution differences can interfere with traditional transfer learning methods. The proposed alignment head structure, meticulously constructed based on mathematical modeling of cross-device issues, provides the model with a precise optimization direction for minimizing cross-device variability, ultimately leading to superior performance in these challenging cross-device issues.
- From the aspect of applicability, none of the methods demonstrate good performance when using the "B" dataset as the source domain, except for PSAT. In fact, many cases show negative impacts, as illustrated in the first three columns of Table II. However, when the "B" dataset served as the target domain, transfer learning still had an effect compared with "Baseline", as shown in columns 4, 7, and 10. This suggests the asymmetric difficulty of handling cross-device variability, since the source domain data does not completely cover the knowledge required to learn specific patterns in the target domain. Specifically, the mathematical modeling method highlights two potential limitations in "B" dataset transfer learning: first, the spatial transformations and filtering actions of the source domain data occur within a relatively compressed space; second, the noise in the source domain

TABLE III
EXPERIMENTAL RESULTS FOR ABLATION STUDY

Methods	Backbones	B ↓ M3	B ↓ M12	B ↓ P	M3 ↓ B	M3 ↓ M12	M3 ↓ P	M12 ↓ B	M12 ↓ M3	M12 ↓ P	P ↓ B	P ↓ M3	P ↓ M12	Weighted Accuracy
EEG-Infinity (A1)	EEGNet	63.58	72.22	69.06	64.71	78.72	78.75	62.69	76.09	**79.19**	**71.43**	74.18	78.82	74.16
	ShallowConvNet	60.74	68.80	67.21	69.07	79.00	78.33	60.30	73.76	77.57	74.44	74.10	78.80	
	DeepConvNet	**64.05**	68.52	64.16	67.18	78.54	77.82	66.13	74.64	77.46	**70.39**	**74.38**	77.97	
	InceptionEEG	67.15	74.45	72.19	64.32	79.61	**79.46**	63.66	76.68	77.46	74.18	73.94	**78.78**	
EEG-Infinity (A2)	EEGNet	**64.65**	72.54	63.37	60.53	78.68	79.57	62.97	74.16	77.30	57.67	73.82	78.59	72.50
	ShallowConvNet	60.68	63.71	63.65	**69.07**	77.12	76.31	61.07	71.26	75.37	66.79	71.60	76.00	
	DeepConvNet	54.96	70.53	59.93	64.09	77.84	**78.62**	64.12	73.41	78.85	61.96	73.63	76.77	
	InceptionEEG	64.57	72.96	64.79	63.16	79.03	78.71	**64.36**	75.06	76.93	63.70	73.23	77.96	
EEG-Infinity (A3)	EEGNet	63.87	71.97	65.36	60.99	**80.72**	78.89	64.05	75.11	77.95	61.61	73.27	77.76	73.52
	ShallowConvNet	61.75	66.98	64.62	66.75	77.66	78.01	66.13	73.13	78.05	64.86	**76.07**	77.90	
	DeepConvNet	60.54	69.91	62.23	**67.49**	**79.32**	77.29	66.10	73.15	78.45	62.96	74.18	**78.59**	
	InceptionEEG	64.75	73.61	68.01	60.53	80.89	79.38	61.26	76.47	77.57	60.99	73.66	78.37	
EEG-Infinity (A4)	EEGNet	64.49	71.68	69.54	63.82	80.63	78.66	**65.02**	**77.78**	77.50	67.80	71.80	76.95	74.32
	ShallowConvNet	**63.62**	**70.30**	**67.76**	67.45	**79.18**	78.54	**67.38**	**77.90**	**80.41**	73.17	74.66	77.75	
	DeepConvNet	61.53	67.53	**64.78**	65.56	77.59	78.09	**68.96**	**75.57**	**79.06**	65.90	73.25	77.65	
	InceptionEEG	**67.65**	**74.91**	71.75	**67.68**	**81.96**	78.77	63.31	**78.31**	78.73	73.64	71.54	77.28	
EEG-Infinity (A5)	EEGNet	64.01	**72.90**	**69.97**	**64.78**	79.31	**79.72**	63.04	75.65	78.45	70.12	**74.66**	**78.87**	**74.34**
	ShallowConvNet	60.10	68.61	66.79	64.55	79.08	**78.92**	61.77	73.63	79.02	**74.48**	74.36	**78.86**	
	DeepConvNet	62.42	**71.81**	63.71	64.82	78.74	77.23	66.75	74.32	77.87	69.15	74.04	78.07	
	InceptionEEG	66.32	74.91	**72.27**	63.08	79.35	78.72	63.16	76.98	**79.21**	**74.95**	**75.06**	78.50	

data is difficult to eliminate. Both factors prevent the model from effectively learning the patterns of the target domain.

- From the aspect of transfer learning performance, the transfer case from "P" dataset to "M3" dataset shows that the proposed method did not outperform other methods, indicating that the transformations applied to the data by devices associated with "P" dataset are difficult to map to the common space described by the "M3" dataset. The experimental results reveal an unexpected negative transfer effect associated with the Deep Coral method. Two potential explanations can account for the phenomenon: 1) Based on the mathematical modeling presented in this study, the EEG signals obtained from the two devices exhibit subtle pattern differences that are difficult to discern; 2) The similarity of the covariance matrices (on which the Deep Coral depends) may not be adequate to address the subtle differences across EEG devices. In this context, leveraging Deep Coral to mitigate such discrepancies may require extensive hyperparameter tuning or even be infeasible.

The observed improvement of weighted accuracy, while moderate in magnitude, holds practical value in application scenarios where reliability and precision are critical. In medical BCI systems and neural-controlled assistive devices, even slight gains in decoding performance can lead to a measurable reduction in misclassification events, thereby improving operational stability and user trust. For example, in motor imagery-based control tasks, enhanced accuracy contributes to more consistent device responses and reduces the risk of unintended actions, ultimately mitigating user fatigue and enhancing usability. These considerations underscore the real-world relevance of the proposed method, particularly in safety-sensitive or user-centered BCI applications.

C. Ablation Study

1) Ablation Study Setting: To verify the efficacy and rationality of the proposed configuration, an ablation study is conducted by comparing five different configurations mentioned below, with all other settings (optimizer, dataset, and backbone network) remaining consistent throughout the comparison study:

- using the traditional classification loss function for domain classifier, referred to as EEG-infinity (A1);
- utilizing a multi-objective loss function based on Frank-Wolfe-based optimizer [53], referred to as EEG-infinity (A2);
- the proposed configuration, excluding the SFB, referred to as EEG-infinity (A3);
- the proposed configuration, excluding the FFB, referred to as EEG-infinity (A4);
- the proposed configuration, referred to as EEG-infinity (A5).

2) Ablation Study Result Analysis: The ablation study evaluated the performance of five different EEG-Infinity configurations across 48 cases that were identical to those in the comparison study. The results of all the aforementioned ablation approaches are presented in Table III, with the best performance highlighted in bold and the second best underlined. From Table III, the proposed EEG-Infinity configuration (A5) achieves the best performance in 14 out of 48 cases and the second-best performance in 11 cases, demonstrating its superiority over the other four configurations. Specifically, Table III reveals the following findings:

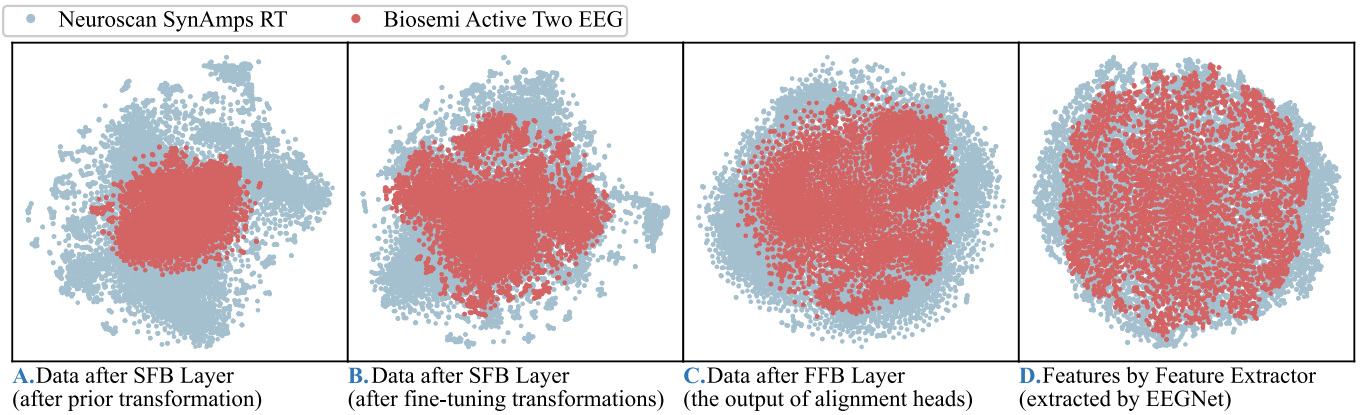

Fig. 8. Visualization of data and feature in EEG-infinity.

- **Adversarial Loss**: Compared with the configuration (A1), the proposed configuration (A5) leads to a 0.18% increase in classification accuracy. Although the difference between the configurations (A1) and (A5) may appear insignificant, the adversarial loss function based on Wasserstein distance (configuration (A5)) has been theoretically demonstrated to outperform the original loss function configuration (configuration (A1)). Thus, the observed phenomenon is consistent with theoretical expectations and highlights the importance of improving loss functions for model performance optimization, especially in complex cases of transfer learning environments.
- **Multi-objective Learning**: Compared with the configuration (A2), the proposed configuration (A5) further increases the classification accuracy by 1.84%. Such results demonstrate that the designed multi-objective learning method significantly improves the stability of the proposed EEG-Infinity, highlighting the effectiveness of the designed loss function over the multi-objective loss function based on a Frank-Wolfe-based optimizer. Experimental results indicate that the multi-objective optimization discussed in this paper involves a complex trade-off scenario. Theoretically, the dynamically scaled weights designed in configuration (A5) balance the weights of multiple objectives, minimizing conflicts and enhancing training stability. Consequently, this configuration improves average performance across multiple cases.
- **Validation of the Effectiveness of SFB and FFB**: Compared with the configuration (A3), which excludes the SFB, configuration (A5) resulted in a significant increase in classification accuracy (0.82%). The improvement indicates that the SFB consistently contributes positively to the performance of the model in most cases, demonstrating its universal applicability. Compared with configuration (A4), which excludes the FFB, configuration (A5) resulted in a slight increase in classification accuracy (0.02%). Although the improvement may seem negligible in terms of weighted accuracy, the performance difference varies across different cases, highlighting the complexity of transfer learning in this context. Specifically, when the source domain is set to "P", the FFB demonstrates a significant positive impact. However, when "M12" is used as the source domain, the FFB has a negative effect, suggesting that the frequency domain of the "M12" dataset differs substantially from that of other datasets, potentially hindering the effectiveness of the FFB.

The scatter plots in Fig. 8, visualized by t-SNE [63], demonstrate a continuous alignment effect in the proposed architecture. They are generated from the Meng2019 dataset, specifically using data from M12 and M3. M12 was recorded with the Neuroscan SynAmps RT system, while M3 used the Biosemi Active Two system, providing a suitable setting for evaluating cross-device generalization. Firstly, the inclusion of an interpolation matrix and a common average reference for prior spatial transformation initially aligns the data from the two devices, resulting in a similar center of distribution after dimension reduction, as shown in Fig. 8A. Then, further fine-tuning spatial transformations in the SFB layer leads to similar variance between two clusters, as shown in Fig. 8B. The progressive transfer effect is also evident in Fig. 8C. Finally, the data distributions processed by the feature extractor display insignificant differences, as shown in Fig. 8D, implying that the data from two different devices have been successfully mapped to the same space. Overall, the experimental result demonstrates the continuous alignment progression in the data processing workflow, demonstrating the effectiveness of the proposed approach.

Similarly, Fig. 9, visualized with t-SNE using features from the initial layers of the feature extractor, illustrates the alignment effect during training. The visualizations are generated from the B (source) and M3 (target) datasets. Three main observations can be drawn: 1) Figs. 9A-B compare task labels at epoch 0 and epoch 37. At epoch 37, the distribution is more separable between task labels than before, with dark dots clustered near the lower-right corner (similar to Fig. D), indicating improved class discrimination after training; 2) Figs. 9C-D compare domain labels at epoch 0 and epoch 37. At epoch 37, the distributions of the two domains (red and blue) become less separable than before, suggesting that features from different domains converge after training; 3) Fig. 9 E presents the Wasserstein distance trend across five-fold cross-validation. The curve shows the mean normalized

TABLE IV
EXPERIMENTAL RESULTS UNDER SAME SAMPLE SIZE CONTROL (SAMPLE SIZE FOR TRAINING DATASETS ARE SET TO 700)

Methods	Backbones	B ↓ M3	B ↓ M12	B ↓ P	M3 ↓ B	M3 ↓ M12	M3 ↓ P	M12 ↓ B	M12 ↓ M3	M12 ↓ P	P ↓ B	P ↓ M3	P ↓ M12	Weighted Accuracy
Baseline	EEGNet	**63.48**	**70.20**	63.38	56.63	73.60	73.86	59.41	71.74	73.71	56.90	**75.21**	74.98	69.00
	ShallowConvNet	**59.96**	62.83	57.84	50.41	66.82	63.53	54.31	66.75	65.21	52.60	**71.94**	73.02	
	DeepConvNet	55.04	**69.79**	57.48	55.16	71.93	74.09	57.75	**75.73**	75.63	57.36	**76.52**	75.24	
	InceptionEEG	60.36	71.18	58.78	56.24	73.28	71.87	55.63	72.91	74.55	57.78	**75.75**	75.27	
EEG-Infinity	EEGNet	62.90	69.22	**65.17**	**60.68**	**78.55**	**79.23**	**61.85**	**73.33**	**78.37**	**66.33**	72.87	**77.67**	**72.11**
	ShallowConvNet	56.29	**66.25**	**64.34**	**61.19**	**75.43**	**73.94**	**58.14**	**68.66**	**75.73**	**64.44**	70.82	**75.82**	
	DeepConvNet	**60.56**	67.33	**61.03**	**63.09**	**77.48**	**78.00**	**61.47**	73.59	**76.71**	**66.83**	72.91	**78.14**	
	InceptionEEG	**63.65**	**72.15**	**69.15**	**62.69**	**78.04**	**78.66**	**62.19**	**73.98**	**77.63**	**67.76**	72.51	**78.50**	

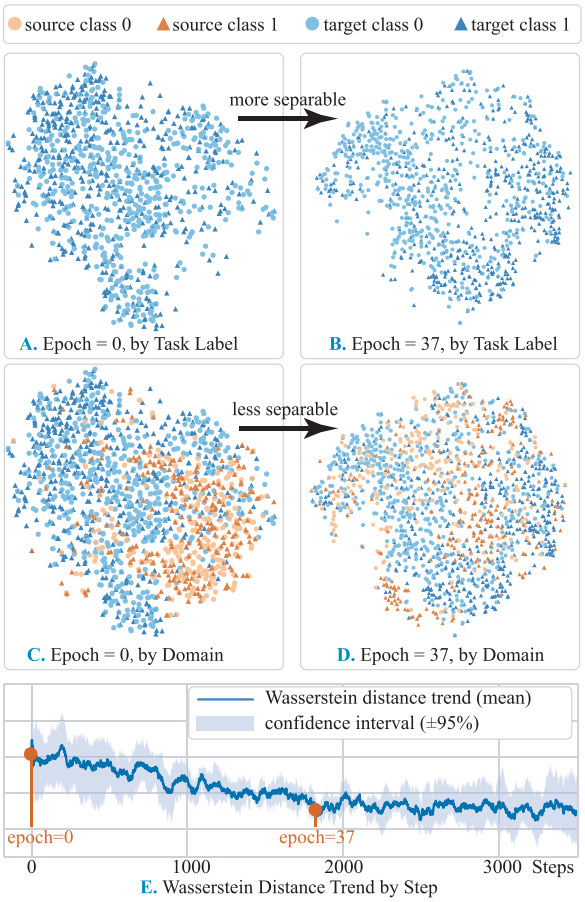

Fig. 9. Visualization of progressive alignment during the training process of EEG-infinity.

Wasserstein distance, which decreases stepwise, indicating progressive domain alignment.

D. Analysis of Experimental Design With Controlled Sample Size

To further evaluate the performance of the EEG-Infinity, we conducted an experiment controlling for sample size. The objective was to compare the differences between EEG-Infinity and the "Baseline" configuration using an identical number of experimental samples across each dataset. Specifically, in this experiment, both the source and target domain training sets were set to the same sample size. Considering the sample size limitations of each dataset, the training set size is set to 700 samples, while maintaining consistency in other experimental settings with the comparison study. The experimental results are presented in Table IV. As observed in Table IV, transfer tasks involving the "B" dataset exhibit higher transfer difficulty than other cases, possibly due to discrepancies in channel configurations within the "B" dataset. Compared to the other three datasets, the "B" dataset demonstrates significant variations in channel configurations, leading to larger differences in the marginal distributions between the source and target tasks. Theoretically, substantial differences in marginal distributions typically inhibit the effectiveness of transfer learning, thereby rendering transfer tasks involving the "B" dataset challenging. Besides, two key observations can be made from Table IV:

- **Changes in Weighted Accuracy**: Due to the reduction in sample size, the weighted accuracy of both the "Baseline" and EEG-Infinity decreased by 2.52% and 2.23%, respectively. The declining trends for both methods are similar, with a slight increase in performance disparity. The result indicates that the set sample size does not constrain the performance of the EEG-Infinity method.
- **Changes in Each Dataset**: In the transfer case from "P" to "M3", a decline in performance was observed, which can be attributed primarily to two reasons according to the theory of negative transfer: 1) The "P" or "M3" datasets exhibit internal uneven data distributions (e.g., due to cross-subject variability), leading to a significant increase in distribution differences between the two datasets, eventually increasing the difficulty of the task; 2) The "P" to "M3" transfer case is inherently complex, causing the model to struggle to effectively transfer useful information with small sample size, resulting in the occurrence of negative transfer phenomena.

E. Analysis of ERD and ERS Across Different Datasets

To further investigate the difference in transfer learning performance across datasets, event-related desynchronization (ERD) and event-related synchronization (ERS) features across various datasets are visualized. ERD/ERS refers to the dynamic changes in the synchrony or desynchrony of EEG signals in response to specific event stimuli [65]. In this study, a time-frequency decomposition method is used

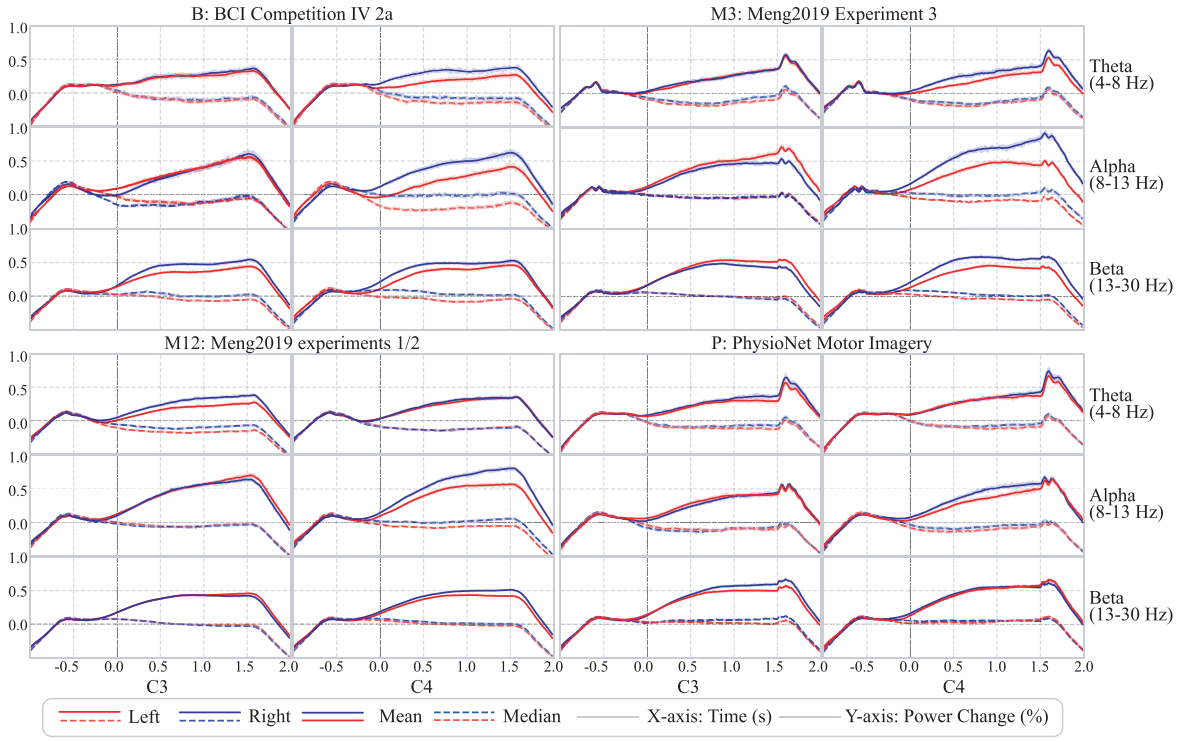

Fig. 10. Visualization of mean and median ERD/ERS across datasets.

to quantify ERD/ERS [65], [66]. The resulting quantified ERD/ERS across datasets are illustrated in Fig. 10. In each subplot of Fig. 10, the y-axis represents the percentage change in spectral time-domain power relative to the baseline interval (-1 to 0 seconds), while the x-axis represents the time range from -1 second to 2 seconds. To accurately capture the neural characteristics of motor imagery, the C3 and C4 electrode channels are used to analyze ERD/ERS features within the θ, α, and β frequency bands. The red and blue curves denote ERD/ERS features of all samples under left-hand and right-hand motor imagery conditions, respectively, while the shaded regions indicate the 95% confidence intervals. Specifically, the observations from Fig. 10 are as follows:

- **Commonalities in ERD/ERS Features across Datasets**: All four datasets demonstrate similar ERS characteristics during motor imagery tasks. Following the onset of an event, the spectral energy at both C3 and C4 electrodes shows an increasing trend in synchrony. The similarity provides a data foundation for implementing transfer learning. In all datasets, the ERS features generally correspond to the neural patterns associated with motor imagery, with ipsilateral electrodes exhibiting stronger ERS responses than contralateral electrodes [18]. Additionally, the median values across all datasets are generally lower than the means, indicating a right-skewed distribution. The right skew suggests that only a subset of samples exhibits pronounced ERS effects.
- **Impact of Cross-Device Factors on ERD/ERS Signals**: The "M3" and "M12" datasets maintain complete consistency in the motor imagery paradigms, with the only differences being the subjects and the devices used. Despite averaging across all samples, inconsistencies stemming from subject and device variability persist. These inconsistencies indicate that both individual differences among subjects and variations in device parameters influence the observed ERD/ERS signals, introducing uncertainties in cross-device transfer learning scenarios.

F. Analysis of Channel Configuration Influence on Model Adaptability

In this section, to further analyze how channel configuration has an influence on transfer learning performance, we quantify both the transfer learning performance and the layout differences between source and target domains. The transfer learning performance R_t is calculated by (20).

$$R_t = \frac{Acc_t - Acc_r}{Acc_s - Acc_r} \tag{20}$$

where Acc_t stands for the test accuracy when the dataset serves as the target domain in unsupervised learning; Acc_s the test accuracy when the dataset serves as the source domain; and Acc_r is the accuracy of a random prediction.

Before quantifying the differences in channel configuration, we define the electrode positions in the source and target domains as $\{p_i^s\}_{i=1,\ldots,N}$ and $\{p_i^t\}_{i=1,\ldots,M}$, respectively. The layout difference D_l is quantized by (21).

$$D_l = \frac{1}{2}\left(\frac{1}{N}\sum_{i=1}^{N}\min_j \|p_i^s - p_j^t\| + \frac{1}{M}\sum_{j=1}^{M}\min_i \|p_j^s - p_i^t\|\right) \tag{21}$$

Next, the difference in the number of channels D_c is quantized by (22).

$$D_c = \frac{|N - M|}{\max(M, N)} \tag{22}$$

TABLE V
CHANNEL CONFIGURATION AND TRANSFER LEARNING PERFORMANCE ANALYSIS

Tasks	M3 ↓ B	M12 ↓ B	B ↓ M3	B ↓ P	P ↓ B	B ↓ M12	M12 ↓ M3	M3 ↓ M12	P ↓ M12	M12 ↓ P	P ↓ M3	M3 ↓ P
EEG-Infinity $R_t(\%)$	51.44	46.75	43.41	54.67	79.18	67.32	81.49	88.22	86.26	86.61	78.92	86.59
			*57.13(**+8.03**)				*85.64(**+9.78**)				*82.75(**+1.99**)	
Others $R_t(\%)$	36.24	39.21	46.98	50.94	57.49	63.75	69.56	75.40	78.87	79.64	83.09	78.43
			*49.10				*75.86				*80.76	
$D_l(10^{-2})$	1.90	1.89	1.90	1.86	1.86	1.89	0.20	0.20	0.21	0.21	0.11	0.11
			*1.88				*0.21				*0.11	
$D_c(10^{-2})$	65.63	64.52	65.63	65.63	65.63	64.52	3.13	3.13	3.13	3.13	0.00	0.00
			*65.26				*3.13				*0.00	

Note: * indicates mean value; (·) indicates absolute improvement achieved by EEG-Infinity compared with others.

Given the aforementioned setting, we can generate Table V and observe the following about channel configurations:

- **Differences Across Transfer Cases**: Based on the quantification metrics, channel differences are categorized into three levels: 1) combinations involving dataset B; 2) combinations of datasets M12 and M13, as well as P and M12; and 3) the combination of P and M3. As the differences decrease, the transfer learning performance R_t consistently improves;
- **Comparison of EEG-Infinity with Other Transfer Configurations**: EEG-Infinity consistently outperforms other algorithms across three levels of channel differences. The greatest performance improvement is observed in the combinations of M12 and M3, and P and M12, with an absolute improvement of R_t up to 9.78%.

G. Evaluation on Single-Channel Configuration

To validate the effectiveness of the proposed method under minimal electrode settings, a single-channel evaluation is conducted using only the C3 electrode, which is commonly associated with motor imagery of hand movements. All experimental configurations, including model architectures, training procedures, and hyperparameters, are kept identical to the main comparison study to ensure consistency. Two representative cross-device scenarios, M3 →M12 and M3 →P, are adopted to demonstrate performance under realistic domain shifts.

As shown in Table VI, the proposed method achieves the highest weighted accuracy (67.30%) across all settings. These results suggest that the proposed method effectively retains discriminative and transferable representations, even when using only a single electrode. This demonstrates the robustness and generalization capabilities of the approach in low-density EEG configurations, highlighting its potential for use in resource-constrained or portable BCI applications.

TABLE VI
THE RESULTS IN SINGLE-CHANNEL CONFIGURATION

Methods	Backbones	M3 ↓ M12	M3 ↓ P	Weighted Accuracy
Baseline	EEGNet	68.12	61.74	65.45
	ShallowConvNet	67.66	54.51	
	DeepConvNet	68.39	61.61	
	InceptionEEG	68.82	60.33	
DDC	EEGNet	69.47	60.07	66.19
	ShallowConvNet	68.11	58.86	
	DeepConvNet	68.48	61.92	
	InceptionEEG	69.17	61.84	
Deep Coral	EEGNet	69.88	60.78	61.18
	ShallowConvNet	53.78	54.29	
	DeepConvNet	60.31	55.47	
	InceptionEEG	68.54	57.69	
DANN	EEGNet	69.38	60.61	66.31
	ShallowConvNet	66.58	**60.59**	
	DeepConvNet	70.12	60.70	
	InceptionEEG	69.54	61.46	
DANN-Wass	EEGNet	69.38	60.55	66.43
	ShallowConvNet	67.95	59.08	
	DeepConvNet	69.14	61.29	
	InceptionEEG	69.92	62.32	
DANet	EEGNet	70.00	59.69	66.46
	ShallowConvNet	68.06	58.83	
	DeepConvNet	69.71	61.82	
	InceptionEEG	69.05	62.37	
PSAT	EEGNet	69.00	61.17	*66.74*
	ShallowConvNet	**69.44**	60.05	
	DeepConvNet	69.30	61.27	
	InceptionEEG	70.19	61.35	
EEG-Infinity	EEGNet	**70.77**	**63.75**	**67.30**
	ShallowConvNet	66.49	58.67	
	DeepConvNet	**70.43**	**62.50**	
	InceptionEEG	**71.35**	**63.48**	

H. Evaluation on Electrode Contact Robustness

To examine the robustness of the proposed algorithm under realistic electrode contact issues (including loosening, positional shifts, and shorting), we conducted a controlled simulation study designed to simulate common forms of contact degradation. Three degradation modes were implemented: 1) Noise Injection, where zero-mean Gaussian noise (scaled to the signal amplitude) was added to a subset of chan-

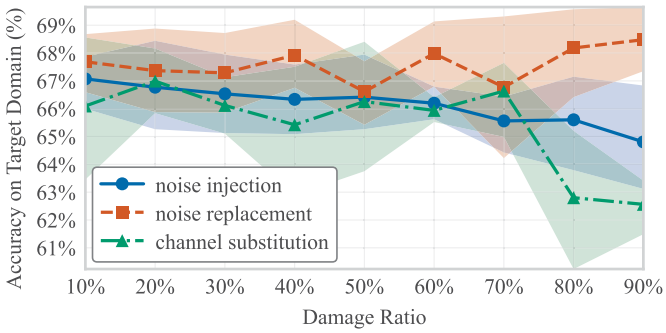

Fig. 11. Simulated electrode contact degradation in EEG-Infinity.

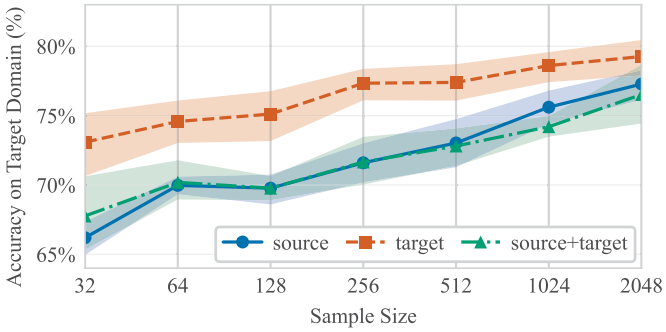

Fig. 12. Accuracy on the Target Domain versus sample size.

nels to simulate mild loosening or small positional drift; 2) Noise Replacement, where selected channels were completely replaced by Gaussian noise of matched scale to simulate full detachment or failed interpolation; and 3) Channel Substitution, where the signal of one channel was replaced by that of another to model shorting effects. For each degradation mode, the ratio of affected channels (damage ratio) was systematically varied to assess the tolerance of the model to increasing levels of disruption.

The perturbations were applied exclusively to the training data, while evaluation was consistently performed on the target-domain test set, ensuring that performance estimates reflect realistic deployment conditions rather than overfitting to damaged inputs. The resulting trends are summarized in Fig. 11, which plots target-domain accuracy against damage ratio for each degradation type. The shaded region indicates the 95% confidence interval across five cross-validation folds.

As shown in Fig. 11, the algorithm demonstrates distinct degradation patterns across the three modes. In both the noise injection and channel substitution settings, accuracy decreases as damage severity increases, but with different trajectories: noise injection leads to a gradual decline consistent with damage ratio reduction, whereas channel substitution remains relatively stable until higher damage ratios (around 80%), where accuracy drops sharply (from 66.5% to 62.6%) due to loss of spatial diversity. In contrast, the noise replacement condition shows minimal or even slightly improved performance, echoing prior findings [50] that controlled perturbations can act as implicit regularization, mitigating over-reliance on individual channels.

TABLE VII
THE RESULTS IN SLEEP STAGE DETECTION

Methods	Backbones	HMC ↓ CAP	CAP ↓ HMC	Weighted Accuracy
Baseline	EEGNet	50.77	50.54	50.74
	ShallowConvNet	53.41	50.91	
	DeepConvNet	50.22	50.12	
	InceptionEEG	49.85	50.77	
EEG-Infinity	EEGNet	**82.74**	**74.53**	**77.62**
	ShallowConvNet	**86.26**	**74.81**	
	DeepConvNet	**88.69**	**71.53**	
	InceptionEEG	**84.61**	**72.86**	

Overall, under mild degradation (10)-30% of channels affected), EEG-Infinity maintains consistent target-domain accuracy within narrow confidence bounds, demonstrating practical resilience to the kinds of electrode contact variability that commonly arise in real-world EEG recording scenarios. This evaluation confirms that the proposed method retains stable performance even in the presence of moderate spatial disturbances.

I. Evaluation on Sleep Stage Detection Paradigm

To further assess the generalization capability of the proposed EEG-Infinity framework across a broader range of paradigms, an extra experiment is conducted in the sleep stage detection task—a representative classification problem that differs in both task structure and signal characteristics from the motor imagery evaluated previously. In this experiment, two public datasets are selected—HMC Sleep Stage [67] and CAP Sleep [68]. To ensure experimental consistency, the data preprocessing pipeline (including segmentation, normalization, and train-test splits) follows the same protocol in the comparison study.

As summarized in Table VII, all backbone models exhibit significant performance gains when equipped with EEG-Infinity. EEG-Infinity improve weighted accuracy from 50.74% to 77.62%. This consistent performance boost across all backbones and datasets suggests that EEG-Infinity effectively captures robust task-invariant representations across datasets. These findings validate the applicability of EEG-Infinity to other tasks and suggest its broader utility in real-world EEG systems where task diversity and sensor heterogeneity are prevalent.

J. Evaluation on Sample Size Effects

To further investigate the impact of training sample size on the transfer learning performance achieved by EEG-Infinity, we conducted an additional experiment. Specifically, we varied the number of training samples under the representative transfer case (M12 → M3) to analyze how the model accuracy changes from data-scarce to data-rich conditions. The results are presented in Fig. 12, which plots target-domain accuracy versus the number of training samples. To examine this effect closely, we considered three different scenarios in limiting the sample size: 1) in the source domain only, marked by

TABLE VIII
COMPLEXITY SUMMARY OF KEY COMPONENTS

Structure	Time Complexity		Params (k)
	Big O	MACs (M)	
SFB	$\mathcal{O}(l(n_e n_c + (n_c)^2))$	$0.371 \sim 3.15$	$0.484 \sim 4.1$
FFB	$\mathcal{O}(l n_c n_f N_f)$	$0.144 \sim 1.67$	$0.017 \sim 0.068$
G_f	Varies	$11 \sim 118$	$2 \sim 119$
G_d/G_y	$\mathcal{O}(d_0 d_1 + d_1 d_2 + d_2 d_3)$	$0.018 \sim 0.111$	$18 \sim 111$

"source"; 2) in the target domain only, marked by "target"; and 3) both domains simultaneously, marked by "source+target". All curves in Fig. 12 mean across five cross-validation folds, with 95% confidence intervals. In Fig. 12, we can see that:

- Target-domain accuracy increases steadily as the number of training samples grows, approximately following a logarithmic trend (across all three curves). Across these settings, sample sizes in both domains influence transfer learning performance. In particular, the effect of the source-domain sample size is markedly stronger than the target domain, which indicates that expanding the source-domain dataset substantially enhances the ability of the model to generalize to unseen target domains.
- Within the tested range of sample sizes (32-2048), no clear saturation effect is observed. Due to dataset constraints, the 4096-sample configuration could not be evaluated. Regarding overfitting, although large models can be prone to this issue under empirical risk minimization, our training protocol employed 5-fold cross-validation with validation-based model selection. The model achieving the highest validation accuracy was subsequently tested once on the held-out target set, to provide a reliable performance estimate.

Overall, this evaluation demonstrates that the proposed algorithm benefits consistently from large sample sizes, particularly within the source domain, while maintaining stable generalization behavior without observable overfitting across the explored data scales.

K. Complexity Analysis

This subsection provides a theoretical and empirical complexity analysis of the key computational components in the proposed model. Table VIII summarizes time and space complexities in terms of asymptotic behavior and typical runtime statistics.

In Table VIII, the results are organized into four columns:

- *Structure* lists the modules analyzed: SFB, FFB, G_f, and G_d/G_y;
- *Big O* gives the asymptotic complexity with respect to symbolic input size;
- *MACs (M)* refers to the number of multiply-accumulate operations, measured in millions;
- *Params (k)* indicates the number of parameters in thousands.

The symbols used in Table VIII are defined as follows: l denotes the sequence length; n_e the channel count; n_c the channel count for projection; n_f the filter count; N_f the order of the filters; d_i the ith layer width of G_d/G_y.

Note: The reported MACs and parameter counts represent typical ranges under experimental configuration in this study, as these metrics can vary significantly depending on specific model hyperparameters and input dimensions. The Big O complexity of module G_f is not reported, as it depends on the specific backbone architecture and cannot be meaningfully defined in a general form.

The theoretical time complexities listed in the Big O column are derived as follows:

- For SFB, the complexity $\mathcal{O}(l(n_e n_c + (n_c)^2))$ arises from two consecutive matrix multiplications over a sequence of length l;
- For FFB, the expression $\mathcal{O}(l n_c n_f N_f)$ captures the cost of applying n_f filters of order N_f to each of the n_c channels;
- For the classifier networks G_d/G_y, the complexity follows a standard multi-layer perception structure, represented as $\mathcal{O}(d_0 d_1 + d_1 d_2 + d_2 d_3)$.

The analysis demonstrates that the computational and spatial costs of the alignment heads (SFB and FFB) are comparable to those of typical modules, underscoring their efficiency and suitability for the overall model design.

V. DISCUSSION

This paper models cross-device variability in EEG as transformations introduced by different recording systems and implements a practical solution—EEG-Infinity, that combines an alignment head with a deep domain adaptation framework. Across multiple transfer cases on MI datasets, EEG-Infinity demonstrates progressive alignment at both structural and temporal levels, achieving consistently strong performance in the target domain. Analyses of channel configuration (such as layout difference D_l and channel-count difference D_c), sample-size scaling, and simulated electrode-contact degradation jointly suggest that incorporating device topology into the alignment head is beneficial under realistic recording conditions.

The alignment head encodes approximate electrode positions and channel relationships, helping preserve spatial patterns that are informative for classification. Empirically, as the layout and channel-count differences between devices decrease, transfer learning performance R_t increases; the improvement is most evident in moderate-discrepancy settings. In addition, accuracy increases steadily with the number of training samples within the tested range and remains stable under mild contact perturbations, indicating that the approach is robust in common usage scenarios. Our results indicate that transfer performance is primarily driven by the source-domain distribution—data quality and sample size—while only a small amount of target-domain data is required for adaptation. However, large discrepancies in electrode number or placement expose an upper bound tied to channel configuration: 1) High-density EEG: A promising path is to adopt more expressive backbones (such as transformer-style models that treat multi-channels data as flexible input tokens. Such designs naturally ingest variable channel sets and might model long-range spatial dependencies; 2) For portable devices with fewer or varying electrodes, unsupervised auxiliary datasets from the

same or similar hardware can serve as intermediate alignment references for pre-adaptation. Meanwhile, channel masking and random electrode dropout during training encourage invariance to missing or permuted channels, enabling label-efficient adaptation in the realistic condition. In this study, the limitations include: 1) Sensitivity to extreme spatial mismatch: Without reliable electrode metadata or when geometry gaps are very large, spatial priors in the current alignment head may be insufficient; 2) Task scope: Empirical validation focuses on motor imagery due to limited public cross-device resources for other paradigms.

VI. Conclusion

The study systematically describes the cross-device issues of EEG through mathematical description and introduces an innovative alignment head structure with the application of the EEG-Infinity cross-device network for MI data. The proposed approach has been experimentally validated in twelve transfer cases, providing a new research perspective for exploring EEG cross-device issues and significantly enhancing the capability to handle compatibility between different devices through alignment head structure, thereby expanding the research scope of EEG data analysis. Although current research primarily focuses on MI datasets, this study demonstrates the potential for expansion to multimodal biometric data. In the field of BCI control, this finding can enhance the user's sense of control and reduce cognitive fatigue. Overall, this study serves as a preliminary attempt to address compatibility issues between EEG devices and lay the foundation for large-scale processing and analysis of future EEG data. Looking ahead, future work will focus on: 1) integrating token formulations with explicit spatial priors into the training objective to better address variable and sparse montages; 2) constructing a multi-device dataset for self-supervised pre-alignment and systematically evaluating channel masking/electrode dropout; and 3) expanding the evaluation to other paradigms and modalities, reporting performance-compute trade-offs, and electrode-count-accuracy curves to further quantify adaptability.

References

[1] D. Wu, Y. Xu, and B.-L. Lu, "Transfer learning for EEG-based brain–computer interfaces: A review of progress made since 2016," *IEEE Trans. Cognit. Develop. Syst.*, vol. 14, no. 1, pp. 4–19, Mar. 2022.

[2] M. Sharifshazileh, K. Burelo, J. Sarnthein, and G. Indiveri, "An electronic neuromorphic system for real-time detection of high frequency oscillations (HFO) in intracranial EEG," *Nature Commun.*, vol. 12, no. 1, p. 3095, May 2021.

[3] C. I. Penaloza and S. Nishio, "BMI control of a third arm for multitasking," *Sci. Robot.*, vol. 3, no. 20, p. 1228, Jul. 2018.

[4] E. Eldele et al., "An attention-based deep learning approach for sleep stage classification with single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 809–818, 2021.

[5] E. Eldele, M. Ragab, Z. Chen, M. Wu, C.-K. Kwoh, and X. Li, "Self-supervised learning for label-efficient sleep stage classification: A comprehensive evaluation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1333–1342, 2023.

[6] R. Wang and J. Liang, "The effect of multiscale parameters on the spiking properties of the morphological neuron with excitatory autapse," *Syst. Sci. Control Eng.*, vol. 12, no. 1, Dec. 2024, Art. no. 2313865.

[7] I. Carrara and T. Papadopoulo, "Classification of BCI-EEG based on the augmented covariance matrix," *IEEE Trans. Biomed. Eng.*, vol. 71, no. 9, pp. 1–17, Sep. 2024.

[8] S. Park, J. Ha, and L. Kim, "Improving performance of motor imagery-based brain–computer interface in poorly performing subjects using a hybrid-imagery method utilizing combined motor and somatosensory activity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1064–1074, 2023.

[9] X. Tang, C. Yang, X. Sun, M. Zou, and H. Wang, "Motor imagery EEG decoding based on multi-scale hybrid networks and feature enhancement," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1208–1218, 2023.

[10] Q. She, T. Chen, F. Fang, Y. Gao, and Y. Zhang, "Discriminative adversarial network based on spatial–temporal–graph fusion for motor imagery recognition," *IEEE Trans. Computat. Social Syst.*, vol. 12, no. 3, pp. 972–983, Jun. 2025.

[11] 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.

[12] L. Zhao and B. Li, "Adaptive fixed-time control for multiple switched coupled neural networks," *Int. J. Netw. Dyn. Intell.*, Sep. 2024, Art. no. 100018.

[13] F. Deng, Y. Ming, and B. Lyu, "CCE-Net: Causal convolution embedding network for streaming automatic speech recognition," *Int. J. Netw. Dyn. Intell.*, vol. 3, no. 3, Sep. 2024, Art. no. 100019.

[14] N. N. Ahmed, T. K. Bhat, and O. S. Powar, "Stacked ensemble machine learning approach for electroencephalography based major depressive disorder classification using temporal statistics," *Syst. Sci. Control Eng.*, vol. 12, no. 1, Dec. 2024, Art. no. 2427028.

[15] Z. Chen, R. Yang, M. Huang, Z. Wang, and X. Liu, "Electrode domain adaptation network: Minimizing the difference across electrodes in single-source to single-target motor imagery classification," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 8, no. 2, pp. 1–15, Apr. 2024.

[16] Y. Chen, R. Yang, M. Huang, Z. Wang, and X. Liu, "Single-source to single-target cross-subject motor imagery classification based on multisubdomain adaptation network," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1992–2002, 2022.

[17] J. Han, X. Wei, and A. A. Faisal, "EEG decoding for datasets with heterogenous electrode configurations using transfer learning graph neural networks," *J. Neural Eng.*, vol. 20, no. 6, Dec. 2023, Art. no. 066027.

[18] J. Meng and B. He, "Exploring training effect in 42 human subjects using a non-invasive sensorimotor rhythm based online BCI," *Frontiers Human Neurosci.*, vol. 13, Apr. 2019, Art. no. 128.

[19] Y. Gao, M. Li, Y. Peng, F. Fang, and Y. Zhang, "Double stage transfer learning for brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1128–1136, 2023.

[20] W. Liu, C. Guo, and C. Gao, "A cross-session motor imagery classification method based on Riemannian geometry and deep domain adaptation," *Expert Syst. Appl.*, vol. 237, Mar. 2024, Art. no. 121612.

[21] L. Xu, M. Xu, Y. Ke, X. An, S. Liu, and D. Ming, "Cross-dataset variability problem in EEG decoding with deep learning," *Frontiers Human Neurosci.*, vol. 14, Apr. 2020, Art. no. 103.

[22] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," 2020, *arXiv:1802.03426*.

[23] G. Niso, E. Romero, J. T. Moreau, A. Araujo, and L. R. Krol, "Wireless EEG: A survey of systems and studies," *NeuroImage*, vol. 269, Apr. 2023, Art. no. 119774.

[24] Y. Zhao et al., "Ultra-conformal skin electrodes with synergistically enhanced conductivity for long-time and low-motion artifact epidermal electrophysiology," *Nature Commun.*, vol. 12, no. 1, p. 4880, Aug. 2021.

[25] 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. 2023.

[26] G. Qian, Y. Suo, Q. Cai, Y. Lian, and Y. Zhao, "A 382nVrms 100GΩ@50Hz active electrode for dry-electrode EEG recording," *IEEE Trans. Biomed. Circuits Syst.*, vol. 19, no. 2, pp. 332–343, Apr. 2024.

[27] X. Lei and K. Liao, "Understanding the influences of EEG reference: A large-scale brain network perspective," *Frontiers Neurosci.*, vol. 11, Apr. 2017, Art. no. 205.

[28] M. Sazgar and M. G. Young, "Overview of EEG, electrode placement, and montages," in *Absolute Epilepsy and EEG Rotation Review: Essentials for Trainees*. Cham, Switzerland: Springer, 2019, pp. 117–125.

[29] M. Tohidi, J. Kargaard Madsen, and F. Moradi, "Low-power high-input-impedance EEG signal acquisition SoC with fully integrated IA and signal-specific ADC for wearable applications," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 6, pp. 1437–1450, Dec. 2019.

[30] D. Dai, J. Li, Y. Song, and F. Yang, "Event-based recursive filtering for nonlinear bias-corrupted systems with amplify-and-forward relays," *Syst. Sci. Control Eng.*, vol. 12, no. 1, Dec. 2024, Art. no. 2332419.

[31] F. Jin, L. Ma, C. Zhao, and Q. Liu, "State estimation in networked control systems with a real-time transport protocol," *Syst. Sci. Control Eng.*, vol. 12, no. 1, Dec. 2024, Art. no. 2347885.

[32] S. Liu, L. Wang, Y. Zhang, Y.-A. Wang, and H. Dong, "Recursive filtering of networked systems with communication protocol scheduling: A survey," *Int. J. Syst. Sci.*, vol. 56, no. 11, pp. 2499–2516, Aug. 2025.

[33] W. Wang, X. Kan, D. Ding, H. Liu, and X. Gao, "Distributed correntropy Kalman filtering over sensor networks with FlexRay-based protocols," *Int. J. Syst. Sci.*, vol. 56, no. 6, pp. 1347–1359, Apr. 2025.

[34] A. Bagheri, M. T. Salam, J. L. Perez Velazquez, and R. Genov, "Low-frequency noise and offset rejection in DC-coupled neural amplifiers: A review and digitally-assisted design tutorial," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 1, pp. 161–176, Feb. 2017.

[35] N. K. Yadav, A. Dhawan, M. Tiwari, and S. K. Jha, "A state-of-the-art survey on noise removal in a non-stationary signal using adaptive finite impulse response filtering: Challenges, techniques, and applications," *Int. J. Syst. Sci.*, vol. 56, no. 4, pp. 885–918, Mar. 2025.

[36] Y. Zhang et al., "A batteryless 19 μW MICS/ISM-band energy harvesting body sensor node SoC for ExG applications," *IEEE J. Solid-State Circuits*, vol. 48, no. 1, pp. 199–213, Jan. 2013.

[37] R. Caballero-Águila, J. Hu, and J. Linares-Pérez, "Filtering and smoothing estimation algorithms from uncertain nonlinear observations with time-correlated additive noise and random deception attacks," *Int. J. Syst. Sci.*, vol. 55, no. 10, pp. 2023–2035, Jul. 2024.

[38] R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution EEG and ERP measurements," *Clin. Neurophysiol.*, vol. 112, no. 4, pp. 713–719, Apr. 2001.

[39] G. Klem, H. Lüders, H. H. Jasper, and C. Elger, "The ten-twenty electrode system of the international federation. The international federation of clinical neurophysiology," *Electroencephalogr. Clin. neurophysiology. Suppl.*, vol. 52, pp. 3–6, 1999.

[40] M. Tangermann et al., "Review of the BCI competition IV," *Frontiers Neurosci.*, vol. 6, 2012, Art. no. 55.

[41] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of EOG artifacts in EEG recordings," *Clin. Neurophysiol.*, vol. 118, no. 1, pp. 98–104, Jan. 2007.

[42] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain–computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, Jun. 2004.

[43] D. Yao, "A method to standardize a reference of scalp EEG recordings to a point at infinity," *Physiological Meas.*, vol. 22, no. 4, pp. 693–711, Nov. 2001.

[44] J. Dien, "Issues in the application of the average reference: Review, critiques, and recommendations," *Behav. Res. Methods, Instrum., Comput.*, vol. 30, no. 1, pp. 34–43, Mar. 1998.

[45] X. Huang, S. Jing, Z. Wang, Y. Xu, and Y. Zheng, "Closed-form FIR filter design based on convolution window spectrum interpolation," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1173–1186, Mar. 2016.

[46] Y. Ganin et al., "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, pp. 59:1–59:35, Dec. 2016.

[47] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial network," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2017, pp. 1–12.

[48] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223.

[49] J. Shen et al., "Wasserstein distance guided representation learning for domain adaptation," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, 2018, pp. 1–12.

[50] S. Pérez-Velasco, E. Santamaría-Vázquez, V. Martínez-Cagigal, D. Marcos-Martínez, and R. Hornero, "EEGSym: Overcoming inter-subject variability in motor imagery based BCIs with deep learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1766–1775, 2022.

[51] R. T. Schirrmeister et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017.

[52] Q. She, T. Chen, F. Fang, J. Zhang, Y. Gao, and Y. Zhang, "Improved domain adaptation network based on Wasserstein distance for motor imagery EEG classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1137–1148, 2023.

[53] O. Sener and V. Koltun, "Multi-task learning as multi-objective optimization," in *Advances in Neural Information Processing Systems*, vol. 31. Red Hook, NY, USA: Curran Associates, 2018.

[54] S. Liu, E. Johns, and A. J. Davison, "End-to-end multi-task learning with attention," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Mar. 2019, pp. 1871–1880.

[55] N. Zeng, P. Wu, Z. Wang, H. Li, W. Liu, and X. Liu, "A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–14, 2022.

[56] H. Cheng, Z. Wang, Z. Wei, L. Ma, and X. Liu, "On adaptive learning framework for deep weighted sparse autoencoder: A multiobjective evolutionary algorithm," *IEEE Trans. Cybern.*, vol. 52, no. 5, pp. 3221–3231, May 2022.

[57] L. Tian, "Multi-dimensional adaptive learning rate gradient descent optimization algorithm for network training in magneto-optical defect detection," *Int. J. Netw. Dyn. Intell.*, Sep. 2024, Art. no. 100016.

[58] Q. Yang, J. Zhou, and Z. Wei, "Time perspective-enhanced suicidal ideation detection using multi-task learning," *Int. J. Netw. Dyn. Intell.*, Jun. 2024, Art. no. 100011.

[59] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," *CoRR*, vol. abs/1412.3474, pp. 1–9, Dec. 2014.

[60] B. Sun and K. Saenko, "Deep CORAL: Correlation alignment for deep domain adaptation," in *Proc. Eur. Conf. Comput. Vis. Workshops (ECCVW)*, vol. 9915, 2016, pp. 443–450.

[61] S.-Y. Chen, C.-M. Chang, K.-J. Chiang, and C.-S. Wei, "SSVEP-DAN: Cross-domain data alignment for SSVEP-based brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 2027–2037, 2024.

[62] J. Zhang, J. Fang, S. Liu, D. Liu, H. Wu, and J. Long, "Toward cross-brain–computer interface: A prototype-supervised adversarial transfer learning approach with multiple sources," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–13, 2024.

[63] L. v. d. Maaten and G. E. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605, 2008.

[64] Y. Ma, W. Zhao, M. Meng, Q. Zhang, Q. She, and J. Zhang, "Cross-subject emotion recognition based on domain similarity of EEG signal transfer learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 936–943, 2023.

[65] B. Graimann, J. E. Huggins, S. P. Levine, and G. Pfurtscheller, "Visualization of significant ERD/ERS patterns in multichannel EEG and ECoG data," *Clin. Neurophysiol.*, vol. 113, no. 1, pp. 43–47, Jan. 2002.

[66] A. Gramfort",", "MEG and EEG data analysis with MNE-Python," *Frontiers Neurosci.*, vol. 7, p. 267, Dec. 2013.

[67] D. Alvarez-Estevez and R. M. Rijsman, "Inter-database validation of a deep learning approach for automatic sleep scoring," *PLoS ONE*, vol. 16, no. 8, pp. 1–27, Aug. 2021.

[68] M. G. Terzano et al., "Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep," *Sleep Med.*, vol. 3, no. 2, pp. 187–199, Mar. 2002.