

Cross-Subject DD: A Cross-Subject Brain-Computer Interface Algorithm

Xiaoyuan Li^{1,2}, Xinru Xue^{1,2}, Bohan Zhang³, Ye Sun^{1,2}, Shoushuo Xi^{1,2}, and Gang Liu^{*1,2}

¹School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou, 450001, China

²Henan Key Laboratory of Brain Science and Brain Computer Interface Technology, Zhengzhou, 450001, China

³School of Computer Science and Engineering, University of New South Wales, Sydney, NSW 2052, Australia

July 9, 2025

^{*}Correspondence: gangliu_@zzu.edu.cn

Abstract

Brain-computer interface (BCI) based on motor imagery (MI) enables direct control of external devices by decoding the electroencephalogram (EEG) generated in the brain during imagined movements. However, due to inter-individual variability in brain activity, existing BCI models exhibit poor adaptability across subjects, thereby limiting their generalizability and widespread application. To address this issue, this paper proposes a cross-subject BCI algorithm named Cross-Subject DD (CSDD), which constructs a universal BCI model by extracting common features across subjects. The specific methods include: 1) training personalized models for each subject; 2) transforming personalized models into relation spectrums; 3) identifying common features through statistical analysis; and 4) constructing a cross-subject universal model based on common features. The experiments utilized the BCIC IV 2a dataset, involving nine subjects. Eight of these subjects were selected for training and extracting the common features, and the cross-subject decoding performance of the model was validated on the remaining subject. The results demonstrate that, compared with existing similar methods, our approach achieves a 3.28% improvement in performance. This paper introduces for the first time a novel method for extracting pure common features and constructing a universal cross-subject BCI model, thereby facilitating broader applications of BCI technology.

1 Introduction

Brain-computer interface (BCI) enables the direct translation of brain signals into control commands for computers or external devices, offering vast potential across diverse fields such as neural rehabilitation, neuroscience, augmented reality, and more[1][2][3].

Central to the functionality of BCI technology is the real-time decoding of neural activity, often relying on techniques like electroencephalography (EEG), functional magnetic resonance imaging (fMRI), and other brain imaging modalities[4]. Among these technologies, EEG-based BCI has emerged as a leading area of research and application, primarily due to its

non-invasive nature, low cost, and strong real-time capabilities[5]. Owing to the enhancement of computational capabilities and continuous advancements in machine learning technologies, BCI systems have witnessed a significant improvement in performance, leading to breakthroughs in various practical applications.

However, despite these advancements, the widespread deployment of BCI technology in real-world scenarios continues to face numerous challenges. One of the foremost obstacles is the issue of poor cross-subject generalization[6]. Current BCI models are typically trained for individual users and often struggle to generalize to new users. This is primarily due to inherent differences in brain structures, neural activity patterns, and electrophysiological signals across individuals[7]. As a result, extensive user-specific adaptation and retraining are required, necessitating large datasets and considerable computational resources. These limitations significantly impede the commercialization and broader adoption of BCI technology.

To address this challenge, research on cross-subject BCI has gained increasing attention. An ideal cross-subject BCI system should be capable of operating seamlessly across multiple users without the need for extensive individual calibration. However, developing an effective cross-subject model remains a significant challenge due to the substantial inter-subject variability in brain electrophysiological activity. Existing methods often struggle to extract stable and distinctive common features across individuals, leading to poor adaptability and limited performance in cross-subject[8].

Several approaches have been proposed to improve cross-subject BCI performance, including deep learning (DL), transfer learning (TL), domain adaption (DA), data augmentation, among others. Among these, deep learning has been particularly prominent, due to its powerful feature extraction capabilities[9][10][11]. Models like EEGNet, introduced by Lawhern et al., leverage deep learning to automatically learn hierarchical representations from raw, noisy, and non-stationary EEG signals, thus enhancing generalization across subjects while maintaining interpretability even with limited data[9]. In

addition to deep learning, transfer learning has been widely explored as a way to reduce the data collection burden for new users. This method adapts pre-trained models from source subjects to new individuals, facilitating more efficient adaptation[12][13]. Zhang et al. proposed HDNN-TL, an end-to-end hybrid deep neural network combining convolutional neural networks (CNN) and long short-term memory (LSTM) with transfer learning[14]. The network was fine-tuned using rectified linear units (RELU) in the fully connected layer to accommodate new tasks. The testing showed that the framework achieved satisfactory results for new subjects, reducing both time and training data requirements in the motor imagery task. Domain adaptation is a key technique in transfer learning, with the primary goal of reducing the data distribution gap between the source domain and the target domain. This enables the model to maintain good generalization performance, even when the target domain has little or no labeled samples. Fu et al. developed the SCDAN model[15], utilizing domain adaptation techniques to learn shared neural activation patterns, thereby improving the transferability of motor imagery EEG decoding. To enhance model robustness, data augmentation techniques have been used to generate synthetic EEG samples that mimic inter-subject variability[16][17].

Despite these advancements, several challenges remain. While deep learning models are effective at identifying common features across subjects using large-scale annotated data, they are still susceptible to overfitting due to the variability inherent in individual subjects[18]. Transfer learning and domain adaptation can enhance model performance by incorporating subject-specific features, but they introduce additional computational complexity and data requirements[19][20]. Furthermore, although data augmentation techniques can simulate inter-subject variability, the synthetic samples may lack physiological realism, which limits their overall effectiveness[21].

A major limitation of most existing approaches lies in focusing on subject-specific adaptation, which fine-tunes models for individual users rather than systematically extracting stable and distinguishable neural representations across subjects. We argue that the most direct approach to cross-subject adaptation

is to extra common features among individuals and construct a universal model to enable broader applicability. However, to date, there is no well-established method for systematically extracting inter-subject commonalities in BCI applications, which significantly limits the generalization ability of current models.

To address these issues, this paper introduces a novel approach that explicitly distinguishes between subject-specific and common features. We propose a cross-subject BCI algorithm based on spectral analysis, referred to as Cross-Subject DD (CSDD). Unlike conventional methods, our core innovation lies in extracting common components across subjects to construct a widely applicable BCI model. Our framework consists of four key steps. (see Fig. 1)

- (1) Training personalized BCI models for each subject.
- (2) Transforming personalized models into relation spectrums.
- (3) Applying statistical analysis to identify common features present across multiple subjects.
- (4) Constructing a generalized BCI model based on the extracted common features.

Through this approach, our model enhances adaptability and decoding accuracy for new users without requiring extensive subject-specific training data.

The primary contributions of this paper are as follows.

- (1) This paper proposes for the first time a method for extracting both common and subject-specific features in human intentions, along with visualization and selection techniques.
- (2) This paper proposes for the first time a model based on common features across subjects during human intention generation, demonstrating that common feature modeling enhances model generalization across subjects.
- (3) Filter: This paper proposes a novel method for decomposing a system into relation spectrums and then “extracting” the necessary components to reconstruct a new system (system filter). This method is similar to the Fourier filter in signal analysis, which decomposes a complex signal into its frequency spectrum, and from which the desired components are extracted to reconstruct the signal.

The remainder of this paper is organized as follows.

Section 2 provides a detailed description of the proposed CSDD model. Section 3 introduces the datasets and experimental setups. Section 4 presents the experimental results. Section 5 discusses the implications of our findings. Section 6 concludes the paper.

2 CSDD

CSDD is a model designed to extract common features and build models based on these features. The model aims to solve the cross-subject generalization problem by learning common features across subjects while filtering out subject-specific features. The framework of the CSDD model consists of four key components, which are explained in detail in the following sections: subject-specific transfer learning based on pre-training and fine-tuning (SSTL-PF), transformation of personalized models based on relation spectrums (TPM-RS), extraction of common features based on statistical analysis (ECF-SA), building cross-subject decoding model based on common features (BCSDM-CF). The overall flowchart of the CSDD model is illustrated in Fig. 1.

2.1 Subject-specific transfer learning based on pre-training and fine-tuning (SSTL-PF)

In recent years, transfer learning has become a widely used approach to address the problem of generalizing brain-computer interface models across different subjects. One of the key techniques within transfer learning is pretraining on a large, diverse dataset followed by fine-tuning on a subject-specific dataset. This approach has been shown to be effective in extracting common features across subjects while adapting to the individual characteristics of each subject. At this stage, subject-specific transfer learning (SSTL-PF) is employed, which includes two sub-components: universal feature extraction (UFE) and dendrite net filter (DD Filter).

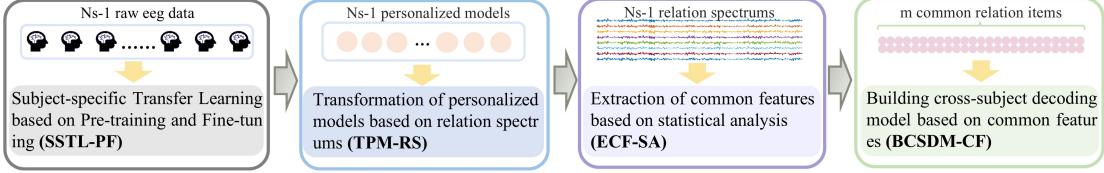


Figure 1: Flow chart of the proposed cross-subject decoding model(CSDD).

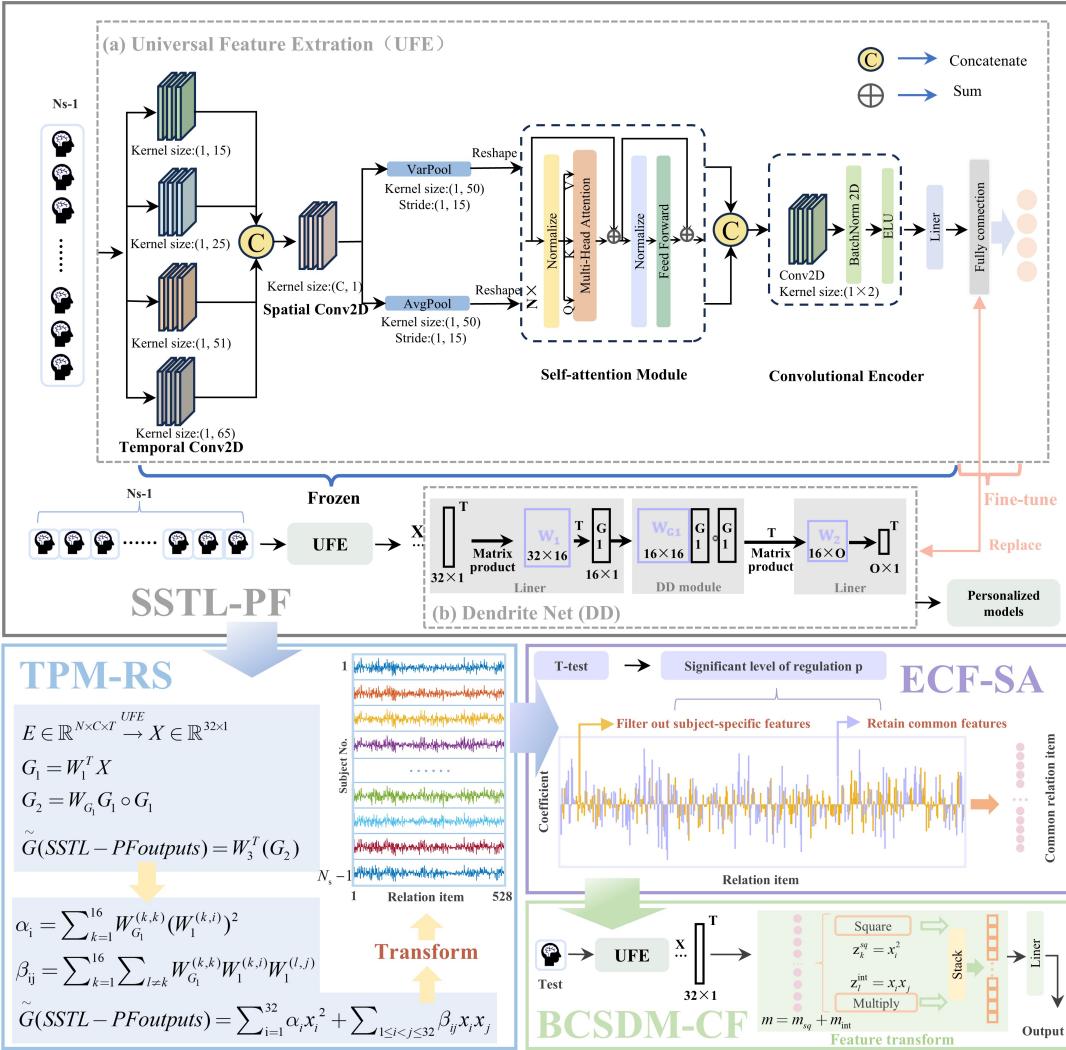


Figure 2: Overall model architecture of the proposed CSDD for cross-subject decoding. The proposed CSDD model consists of four major steps. 1. SSTL-PF. 2. TRM-RS. 3. ECF-SA. 4. BCSDM-CF. N_s : the number of subjects. m : the number of common relation items. O : the number of outputs. Q : query. K : key. V : value.

2.1.1 Universal feature extraction (UFE) as shown in Fig. 2(a)

The Universal Feature Extraction (UFE) process is based on an end-to-end convolutional neural network architecture, adapted from the model proposed by Ma et al.[22] , with modifications to support subsequent fine-tuning. During the pre-training phase, EEG data from $N_s - 1$ subjects are used to train the universal model.

Raw EEG signals from $N_s - 1$ subjects are represented as $E \in \mathbb{R}^{N \times C \times T}$, where N denotes the number of trials, C represents the number of electrodes, and T indicates the number of sample points. These signals are then processed through a series of temporal convolutional layers with kernel sizes of (1, 15), (1, 25), (1, 51), and (1, 65) to extract features at multiple temporal scales. The resulting feature maps are then passed through spatial convolutional layers with a kernel size of (C, 1), where C represents the number of electrode channels, to capture spatial relationships across the different EEG channels. Following the spatial convolution, the feature maps undergo dimensionality reduction and feature selection via varpool and avgpool layers, which help retain important information while reducing computational complexity. The pooled features are then reshaped and input into a shared self-attention module, which utilizes a multi-head attention mechanism and feed-forward networks to weight and integrate features, enhancing their representational capacity. Next, the features are processed by a convolutional encoder using a (1, 2) kernel, followed by batch normalization and activation (ELU) to prepare the model for the final classification task. The processed features are then passed through a fully connected (FC) layer to generate the final output.

To facilitate the subsequent fine-tuning with DD Filter, a linear dimensionality reduction layer is added before the FC layer, reducing the feature space to 32 dimensions and performing normalization. Let $X \in \mathbb{R}^{32 \times 1}$ represent the resulting 32-dimensional feature vector, which serves as the input for the subsequent processing stages.

2.1.2 DD Filter as shown in Fig. 2(b)

Dendrite Net, or DD, is a fundamental machine learning algorithm and a white-box model used for classification, regression, and system identification[23][24]. It is characterized by its white-box nature, controllable accuracy, and low computational complexity, which enhance its generalization performance. It resembles the Taylor series expansion and can converge to a function (or system) that closely approximates the target. Unlike traditional neural networks, it does not require non-linear functions but instead utilizes multiple DD modules to construct logical expressions among input features.

In this study, the author team has made significant innovations to the previously proposed DD model, introducing for the first time a filtering framework based on the DD algorithm, namely the DD Filter. The innovation of the DD Filter lies in its ability to transform the trained DD model into a relation spectrum. By utilizing statistical analysis, the DD Filter effectively identifies and retains common relation items across subjects while filtering out subject-specific features, which are treated as noise. This section focuses on the model representation of the DD Filter, with additional details on its implementation provided in Sections 2.2 and 2.3.

Within the CSDD framework, the DD Filter plays a crucial role in converting extracted EEG features into meaningful representations, facilitating subsequent filtering processes. During the fine-tuning stage, all layers of the pretrained model, except the fully connected layer, are frozen. The fully connected layer is replaced with the DD module for feature learning, and the model is subsequently trained using data from the current subject. This results in $N_s - 1$ fine-tuned models, also known as personalized models. It receives a 32-dimensional feature vector from the preceding layer, obtained through a universal feature extraction process.

The DD Filter in this paper consists of three key components: the dimensionality reduction module, the DD module, and the linear module. To enhance computational efficiency while preserving essential information, a dimensionality reduction module projects this feature vector onto a 16-dimensional space. The

formula of the dimensionality reduction module is as follows.

$$G_1 = W_1^T X \quad (1)$$

where $G_1 \in \mathbb{R}^{16 \times 1}$ expresses the outputs of the dimensionality reduction module. $W_1 \in \mathbb{R}^{32 \times 16}$ is the weight matrix. $X \in \mathbb{R}^{32 \times 1}$ expresses the inputs of the dimensionality reduction module and the outputs of the UFE module.

DD modules aim at no-linear mapping. The degree of interaction items is adjusted by the number of DD modules. Each DD module is equivalent to a first-order approximation, and stacking multiple DD modules results in a higher-order approximation. Therefore, by adjusting the number of DD modules, the mapping capability of DD can be controlled, preventing overfitting or underfitting. The DD module in this paper is expressed as follows.

$$G_2 = W_{G_1} G_1 \circ G_1 \quad (2)$$

where $G_1 \in \mathbb{R}^{16 \times 1}$ and $G_2 \in \mathbb{R}^{16 \times 1}$ express the inputs and outputs of the DD module, respectively. $W_{G_1} \in \mathbb{R}^{16 \times 16}$ is the weight matrix. \circ expresses Hadamard product.

The linear module, whose dimensions are determined by the number of output categories. The formula is expressed as follows.

$$G^{\text{SSTL-PFoutputs}} = W_2^T G_2 \quad (3)$$

where $G_2 \in \mathbb{R}^{16 \times 1}$ express the outputs of the DD module. $W_2 \in \mathbb{R}^{16 \times O}$ is the weight matrix. $G^{\text{SSTL-PFoutputs}} \in \mathbb{R}^{O \times 1}$ is the outputs of SSTL-PF. O is the number of categories determines.

2.2 Transformation of personalized models based on relation spectrums (TPM-RS)

The relation spectrum is a polynomial expression based on the dendrite network[23][25][26]. Unlike traditional spectral analysis (such as Fourier transform), it emphasizes the relation items between the input signals and the output decoding. By analyzing the polynomial and its coefficients, the relation spectrum provides a deeper insight into neural decoding.

In this study, the fine-tuned personalized models are transformed into relation spectrums through symbolic simplification. This transformation allows us to analyze feature interactions and identify cross-subject stable components for building generalized BCI decoding module.

2.2.1 Symbolic simplification of the DD model

To generate individualized relation spectrum for multi-subject analysis, the linear classification module is modified by replacing the class-specific weight matrix $W_2 \in \mathbb{R}^{16 \times O}$ with a subject-specific vector $W_3 \in \mathbb{R}^{16 \times 1}$, where all elements are set to 1. This operation consolidates the nonlinear interaction items across hidden units, enabling explicit extraction of per-subject relation spectrum. The revised model is formalized as:

$$\begin{cases} G_1 = W_1^T X \\ G_2 = W_{G_1} G_1 \circ G_1 \\ \tilde{G}^{\text{SSTL-PFoutputs}} = W_3^T (W_{G_1} G_1 \circ G_1) \end{cases} \quad (4)$$

where $\tilde{G}^{\text{SSTL-PFoutputs}} \in \mathbb{R}^{1 \times 1}$ represents the revised outputs of SSTL-PF, which are used for transforming relation spectrums. And the expanded output can also be expressed as follows. The relation spectrum is derived by symbolically expanding equation.

$$\begin{cases} \alpha_i = \sum_{k=1}^{16} W_{G_1}^{(k,k)} (W_1^{(k,i)})^2 \\ \beta_{ij} = \sum_{k=1}^{16} \sum_{l \neq k} W_{G_1}^{(k,l)} W_1^{(k,i)} W_1^{(l,j)} \\ \tilde{G}^{\text{SSTL-PFoutputs}} = \sum_{i=1}^{32} \alpha_i x_i^2 + \sum_{1 \leq i < j \leq 32} \beta_{ij} x_i x_j \end{cases} \quad (5)$$

This polynomial is automated using symbolic tools (e.g., MATLAB Symbolic Toolbox).

2.2.2 Controlling the number of common features

The completeness of the relation spectrum is mathematically guaranteed through its construction, which explicitly preserves all quadratic interactions in the feature space. For a d-dimensional input vector $X \in$

$\mathbb{R}^{d \times 1}$, the relation spectrum comprises d squared items (x_i^2) and $\frac{d(d-1)}{2}$ pairwise interaction items ($x_i x_j$), resulting in a deterministic theoretical term count.

$$N_t = d + \frac{d(d-1)}{2} = 528 \quad (d = 32). \quad (6)$$

This constraint is enforced during symbolic simplification, where the dendrite network's hierarchical operations are systematically expanded into polynomial form. Redundant items are eliminated through algebraic consolidation, while coefficients of identical items are merged. Consequently, the final spectrum contains precisely N_t non-redundant items, ensuring exhaustive representation of nonlinear feature interactions without information loss.

2.3 Extraction of common features based on statistical analysis (ECF-SA)

Statistical analysis is employed to identify significant patterns and relationships within data. In this paper, it is employed to extract common features from $N_s - 1$ relation spectrums corresponding to $N_s - 1$ subjects, allowing for the identification of common components across subjects.

For each quadratic term $\alpha_i x_i^2$ or $\beta_{ij} x_i x_j$ interaction term in the spectrum, coefficient values $\{\alpha_i^{(1)}, \alpha_i^{(2)}, \dots, \alpha_i^{(N_s-1)}\}$ or $\{\beta_{ij}^{(1)}, \beta_{ij}^{(2)}, \dots, \beta_{ij}^{(N_s-1)}\}$ are collected across subjects. A one-sample t -test is applied to determine whether the mean coefficient significantly deviates from zero. The null and alternative hypotheses are defined as follows.

$$H_0 : \mu_{R_i} = 0, \quad H_1 : \mu_{R_i} \neq 0 \quad (7)$$

The t-statistic is calculated as:

$$t = \frac{\bar{R}_i}{s_i / \sqrt{N_s - 1}} \quad (8)$$

where \bar{R}_i and s_i are the sample mean and standard deviation of the coefficient values, $N_s - 1$ is the number of subjects. Relation items with p-values below a predefined threshold $\alpha_{\text{threshold}}$ are retained as statistically significant common features.

As summarized in [Table 1](#), common features are characterized by consistent patterns across subjects, whereas subject-specific features exhibit more random and individual-specific variation. In this context, the features correspond to the relation items.

[Table 1](#): Definitions of common features and subject-specific features

| Feature type | Definition |
|---------------------------|---|
| common features | Features that are consistent across subjects, where the relationships are similar (mostly positive or negative). The accumulated average significantly differs from zero, and the random component is non-zero. |
| subject-specific features | Features that are specific to each subject, with relationships that are random (can be positive or negative). The accumulated average approaches zero, and the random component is zero. |

2.4 Building cross-subject decoding model based on common features (BCSDM-CF)

Using the extracted common features across subjects, a generalized BCI model is then constructed.

2.4.1 Feature transformation layer

The common features are derived by transforming each subject's signal space. To map all subjects' information into a unified feature space, a feature transformation layer is designed. This layer transforms the input brain signal $X \in \mathbb{R}^{d \times 1}$ into a new feature vector $Z \in \mathbb{R}^{m \times 1}$, where $m \leq N_t$ and $N_t = d + \frac{d(d-1)}{2}$ is the theoretical maximum relation items. The transformation involves two types of operations: squared items and interaction items. For each significant squared item x_i^2 , it can be described as follows.

$$z_k^{\text{sq}} = x_i^2 \quad (9)$$

where i belongs to the set of indices S_{sq} corresponding to statistically validated squared features. Similarly, for each significant interaction term $x_i x_j$, it is represented as follows.

$$z_l^{\text{int}} = x_i x_j \quad (10)$$

where the pair (i, j) is included in the set S_{int} of significant interactions. The transformed feature vector Z is constructed by concatenating all retained components.

$$Z = [z_1^{\text{sq}}, \dots, z_{m_{\text{sq}}}^{\text{sq}}, z_1^{\text{int}}, \dots, z_{m_{\text{int}}}^{\text{int}}]^T \in \mathbb{R}^{m \times 1} \quad (11)$$

where $m = m_{\text{sq}} + m_{\text{int}}$, ensuring that only cross-subject common features are preserved.

2.4.2 Linear classifier

The linear classifier then maps the m -dimensional feature vector Z to class probabilities through a linear transformation followed by softmax activation. The linear transformation is defined as follows.

$$h = \mathbf{W}^\top Z + b \quad (12)$$

where $\mathbf{W} \in \mathbb{R}^{m \times O}$ is the weight matrix, $\mathbf{b} \in \mathbb{R}^{O \times 1}$ is the bias vector, and O is the number of categories. The output probabilities are computed via the softmax function. The model is trained by minimizing the cross-entropy loss.

3 Experiments

3.1 Experimental Details

3.1.1 Dataset

To assess the feasibility and generalizability of CSDD, it is evaluated using a publicly available dataset. The BCIC IV 2a dataset contains EEG signals from 22 electrodes, recorded from nine healthy subjects over two sessions conducted on two different days. Each session consists of 288 four-second motor imagery trials per subject, involving the imagined movements of the left hand, right hand, feet, and tongue. This study uses data from 0 to 4 seconds. Before release, the signals

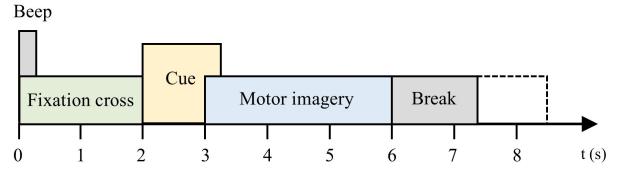


Figure 3: The experimental paradigm for BCIC IV 2a

were sampled at 250 Hz and bandpass filtered between 0.5 Hz and 100 Hz. The original dataset uses 288 trials from the first session (“T”) for training and 288 trials from the second session (“E”) for testing. However, in the cross-subject scenario, the original dataset needs to be re-split by subject using a leave-one-subject-out approach for modeling based on common features. Consequently, nine evaluation datasets (A01-A09) are obtained, each containing 576 trials (288 trials \times 2 sessions) for testing one subject, with the remaining 4608 trials (288 trials \times 2 sessions \times 8 subjects) used for training. The BCIC IV 2a dataset is accessible at <http://bnci-horizon-2020.eu/database/data-sets>. The experimental paradigm of the BCIC IV 2a dataset is shown in Fig. 3.

3.1.2 Performance metric

In the experiment, classification accuracy and Cohen’s kappa coefficient are employed to evaluate the decoding performance of different networks, thereby comprehensively assessing the performance of the model in cross-subject tasks. The method for calculating accuracy is described as follows.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

where TP and TN represent the number of correct positive and correct negative samples predicted by the model, respectively, while FP and FN denote the number of false positive and false negative samples predicted by the model, respectively.

The method for calculating Cohen’s kappa coeffi-

cient is described as follows.

$$k = \frac{P_o - P_e}{1 - P_e} \quad (14)$$

where P_o denotes the accuracy of the model, and P_e represents the probability or accuracy of a random guess.

We evaluate the statistical significance between the proposed model and other baseline models using paired t -tests and p -values. A p -value below 0.01 (**) denotes a statistically significant difference, $p < 0.05$ (*) indicates a significant difference, and $p > 0.05$ suggests no significant difference.

3.1.3 Evaluation baselines

We compare the proposed network with six baseline models.

ShallowConvNet [27] consists of two convolutional layers, which are used for temporal and spatial filtering, respectively.

EEGNet [9] is a lightweight neural network that first learns through temporal filtering and then performs global spatial filtering using deep convolutions. After extracting initial temporal and spatial features, it utilizes separable convolutions to perform deep temporal feature extraction, demonstrating a reduced parameter count.

CRAM [11] is a graph-based convolutional recurrent attention model, which first introduces a graph structure to represent the EEG node location information. Subsequently, the convolutional recurrent attention model learns EEG features from both spatial and temporal dimensions.

CCNN [10] is a multilayer CNN approach that integrates CNNs with different characteristics and structures, utilizing convolutional features from different layers to capture spatial and temporal features from raw EEG data.

A novel EEG 3D representation is introduced in the multi-branch 3D CNN [28], which preserves both spatial and temporal information. Based on this 3D representation, a multi-branch 3D CNN is used to extract MI-related features.

CTNet [29] is a model for EEG-based MI classification that combines convolutional and Transformer

architectures. It first extracts local and spatial features using a convolutional module, then captures global dependencies with a Transformer encoder, and finally classifies the signals using fully connected layers.

3.2 Experimental Setups

All experiments in this paper are conducted on a single GPU, specifically the Nvidia RTX 3080, using PyTorch, while the extraction of common relation items is performed using MATLAB. The cross-subject setting for the entire experiment is illustrated in Fig. 4.

For the upcoming experiments, this paper adopts the Leave-One-Subject-Out (LOSO) validation strategy. This method is widely regarded as the most rigorous approach for cross-subject classification, as it completely excludes certain subjects during training and evaluates the model’s performance on unseen subjects[9][30][31][32]. It provides an objective measure of generalization ability. In contrast, other commonly used strategies, such as mix-up-all-subjects and random selection, either merge data from all subjects or partition the data randomly, which may artificially inflate classification performance and fail to reflect real-world adaptability. Given that an ideal BCI system should decode EEG signals from entirely unseen subjects, leave-n-subjects-out remains the most appropriate choice for evaluating cross-subject generalization[33][34][35].

To avoid ambiguity, the Table 2 specifies the notational conventions used for different subject configurations.

Table 2: Notation used for subject configurations.

| Notation | Definition |
|------------|---|
| subject i | Refers to subject A0i ($i = 1, 2, \dots, 9$) |
| Subject -i | Indicates that data from subject A0i is used for testing, while data from the remaining eight subjects is used for training (e.g., leave-one-subject-out setting) ($i = 1, 2, \dots, 9$) (see Fig. 4) |

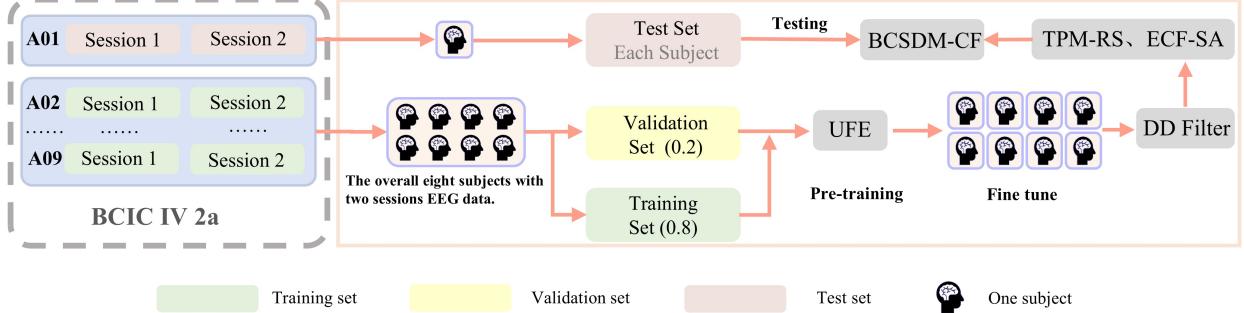


Figure 4: Cross-subject setting.

3.2.1 Personalized model training and acquisition

To eliminate inter-subject variability and extract features directly relevant to the motor imagery task, we designed a transfer learning strategy consisting of two stages: pretraining and fine-tuning.

In the pretraining stage, the LOSO strategy was employed, where EEG data from eight subjects were divided into a training set and a validation set in an 8:2 ratio. A five-fold cross-validation was conducted for iterative model optimization and parameter tuning.

In the fine-tuning stage, the pretrained model was adapted to individual subjects using the training data from eight subjects. This module was then independently optimized for each subject, leading to the generation of eight personalized models. Finally, the fine-tuned model was applied to the test set of the left-out subject, and the classification accuracy of the motor imagery task was evaluated to assess the decoding performance of the personalized model. Hyperparameter optimization was performed using grid search.

During the pretraining stage, the learning rate was set to 0.0002, the batch size was 64, and the model was trained for 500 epochs. In the fine-tuning stage, the learning rate remained at 0.0002, but the batch size was increased to 256 to accelerate the optimization process, with training conducted for 200 epochs.

3.2.2 Extraction of common relation items

Using the best parameters obtained during the training phase, the relation spectrums of the eight personalized models were expanded, resulting in eight corresponding relation spectrums. These relation spectrums are expressed in polynomial form, where the relation items are identical across subjects, but the coefficients vary for each subject. To identify common relation items, a one-sample t -test was conducted on each of the 528 relation items to assess the difference between their mean values and zero. By setting different significance levels, we selected the statistically significant relation items as common features.

To investigate the impact of the number of common relation items on cross-subject decoding performance, we analyzed the common relation items with predefined significance thresholds set to $\alpha_{\text{threshold}}$ values of 0.5, 0.1, 0.01, and 0.001.

3.2.3 Building cross-subject models with common features

To validate the cross-subject performance of the proposed CSDD model, we employ LOSO for evaluation.

Specifically, in each experiment, eight subjects are selected as the training set, and the data from the remaining subject are used as the test set, ensuring that the model is evaluated on unseen individuals. Through nine iterations of experiment, the cross-subject decoding results based on the CSDD

model are calculated for each subject, and the average performance of the model in unseen individuals is statistically analyzed. In this stage, the learning rate was set to 0.0001, the batch size was set to 32, and the number of epochs was set to 200. The decoding results of different subjects were obtained by selecting the common features that yielded the best decoding performance for each subject.

The statistical significance of the differences in decoding performance achieved by different methods was assessed by the paired sample t -test for the BCIC IV 2a dataset.

4 Results

4.1 Relation spectrum of the personalized model

Fig. 5 (a) depicts the relation spectrum of the personalized models for subject 1 to subject 8, with each curve representing the relation spectrum of an individual subject. The x-axis corresponds to 528 relation items, and the y-axis represents the coefficients associated with these relation items. Upon observation, significant differences can be observed between the relation spectrums of different subjects, reflecting the influence of individual differences on EEG signal decoding. This shows that the model training process successfully captured the features of each subject.

Furthermore, **Fig. 5** (b) presents the correlation matrix of the eight personalized models. In the matrix, the intensity of the color blocks indicates the similarity of relation spectrums between subjects, with darker blocks representing higher correlation and lighter blocks representing lower correlation. It can be observed that, while there is some similarity between subjects, individual differences remain prominent. In particular, the relation spectrums of certain subjects are highly similar, highlighting the potential for cross-subject feature sharing. This result further corroborates the effectiveness of the CSDD model in extracting common features and underscores the importance of distinguishing between common and subject-specific features for enhancing cross-subject decoding performance.

4.2 Common Features

4.2.1 Visualization of common features

To validate the effectiveness of the proposed cross-subject decoding method, we performed a key visualization analysis, focusing on the extraction of common features. Feature visualization not only provides an intuitive view of the data distribution but also offers valuable support for subsequent classification tasks.

Fig. 6 displays the extracted common features across different subjects, based on data from subject 1 to subject 8 (Subject-9), along with the relation spectrum analysis results. This visualization allows us to observe the common relationships between different subjects intuitively. These features represent stable components shared by multiple subjects, providing visual evidence to confirm whether common features with broad applicability have been successfully extracted. The figure clearly highlights the stability of the extracted common features, suggesting that they are effective and representative in cross-subject decoding tasks.

4.2.2 The number of common features corresponding to different predefined thresholds $\alpha_{\text{threshold}}$

As shown in **Fig. 7**, the number of common relation items decreases with stricter predefined thresholds $\alpha_{\text{threshold}}$. When $p < 0.1$, the number of features ranges from 361 to 406 across subjects. As the threshold is reduced to 0.05, 0.01, and 0.001, fewer features are retained, with the number of features dropping to as low as 125 for Subject-3 when $p < 0.001$. This indicates that stricter thresholds lead to fewer but more statistically significant features.

4.3 Decoding Performance Evaluation

4.3.1 Decoding performance across different models

Table 3 comprehensively presents the performance of the proposed cross-subject decoding strategy compared with other baseline models on the BCIC IV 2a dataset. It includes the classification accuracy for

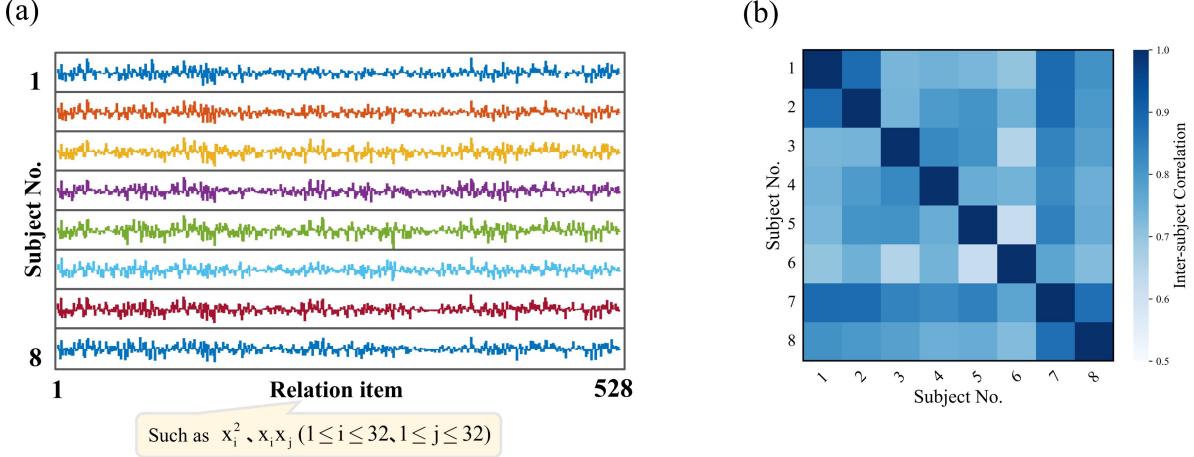


Figure 5: Transform SSTL-PF models. (Test data for subject 9 and train data for the remaining eight subjects). (a) The relation spectrum of subject 1-subject 8. (b) Correlation matrix between eight personalized models.

Table 3: Cross-subject classification results for BCIC IV 2a dataset. Bold denotes the best numerical values. * denotes a significant difference ($p < 0.05$), while ** denotes a highly significant difference ($p < 0.01$).

| Methods | Subject-1 | Subject-2 | Subject-3 | Subject-4 | Subject-5 | Subject-6 | Subject-7 | Subject-8 | Subject-9 | Accuracy | Standard | Kappa | <i>p</i> -value |
|---------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|----------|-------|-----------------|
| EEGNET | 53.76 | 39.54 | 54.88 | 43.02 | 51.8 | 48.96 | 60.70 | 61.38 | 47.82 | 51.32 | 7.34 | - | ** |
| ShallowConvNet | 68.92 | 43.75 | 75.17 | 55.73 | 43.75 | 45.31 | 74.65 | 80.03 | 66.84 | 61.57 | 14.63 | - | - |
| Multi-branch 3D CNN | 49.51 | 40.74 | 64.5 | 44.56 | 54.29 | 40.46 | 58.87 | 59.75 | 56.83 | 52.17 | 8.76 | 0.453 | ** |
| CCNN | 62.07 | 42.44 | 63.12 | 52.09 | 49.96 | 37.16 | 62.54 | 59.32 | 69.43 | 55.35 | 10.66 | - | ** |
| CRAM | 61.02 | 42.35 | 73.11 | 50.43 | 50.74 | 51.48 | 67.26 | 69.72 | 66.85 | 59.22 | 10.74 | - | ** |
| CTNet | 69.27 | 43.92 | 79.34 | 55.38 | 43.92 | 36.11 | 65.10 | 70.66 | 64.06 | 58.64 | 14.61 | 0.449 | * |
| CSDD | 69.10 | 49.31 | 80.38 | 55.03 | 63.02 | 54.17 | 76.04 | 70.83 | 65.80 | 64.85 | 10.47 | 0.531 | - |

each subject and the average accuracy comparison. The proposed CSDD model demonstrates excellent performance in cross-subject decoding tasks, achieving the highest average classification accuracy and significantly surpassing other models.

Firstly, in terms of average classification accuracy, the CSDD model outperforms EEGNet by 13.53%, with a statistically significant difference ($p < 0.01$). This indicates that the CSDD model exhibits stronger robustness and generalization ability when extracting common features and decoding across different subjects. Additionally, compared to CTNet, the CSDD model achieved a 6.21% higher average classification accuracy and a higher kappa value, while also demonstrating a lower standard deviation ($p < 0.05$).

When compared with other baseline models, such as the Multi-branch 3D CNN, CCNN, and CRAM, the CSDD model achieved classification accuracy improvements of 12.68%, 9.23%, and 5.63%, respectively, with these differences also being statistically significant ($p < 0.01$). These results further validate the superiority of the CSDD model in cross-subject tasks, particularly in its effective utilization of common features during feature extraction and model training to reduce the impact of individual differences on decoding performance. Although the CSDD model showed a 3.28% improvement in accuracy compared with ShallowConvNet, this difference was not statistically significant ($p > 0.05$). This suggests that these models may perform similarly in some cases, which may be at-

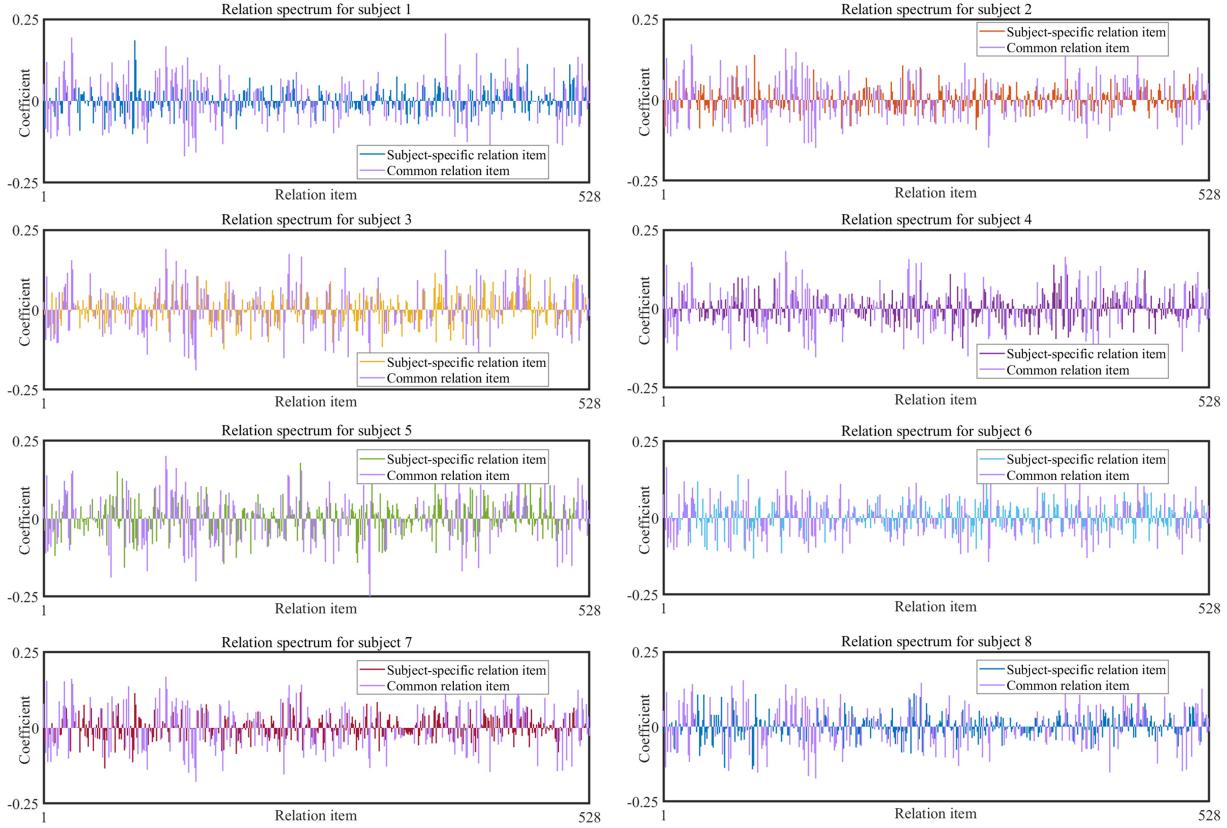


Figure 6: Visualization of common and subject-specific features. (e.g. Subject-9)

tributed to the limitations of these baseline models in handling inter-subject differences. This further highlights the advantages of the CSDD model in terms of adaptability and generalization.

4.3.2 Effect of common feature extraction on cross-subject decoding performance

This section investigates the impact of varying pre-defined significance thresholds for common features extraction on the cross-subject decoding performance of the CSDD model, using the BCIC IV 2a dataset. The common features were extracted at different pre-defined thresholds $\alpha_{\text{threshold}}$ (0.1, 0.05, 0.01, and 0.001), and the performance of the model was evaluated.

The classification accuracy results based on these

thresholds are summarized in Fig. 8. When $p < 0.1$, the average accuracy was 62.96%, with subject-specific accuracies ranging from 49.31% (Subject-2) to 65.62% (Subject-1). Performance improved as the p -value threshold was reduced: when $p < 0.05$, accuracy increased to 63.58%, and when $p < 0.01$, it rose to 64.93%, with Subject-4, Subject-6, and Subject-8 showing substantial improvement. However, when $p < 0.001$, accuracy reached 63.79%, showing a marginal improvement over when $p < 0.01$, indicating that further reduction in features may not provide significant performance gains.

Paired sample t -test was conducted to assess the statistical significance of the performance differences across different predefined thresholds. Results showed

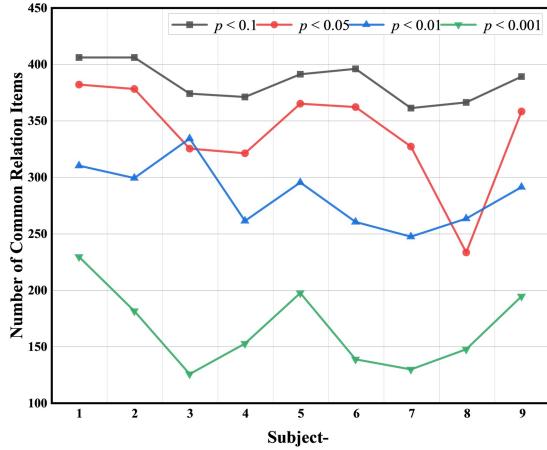


Figure 7: The change in the number of common relation items in the CSDD on the BCIC IV 2a dataset is observed as the predefined threshold $\alpha_{\text{threshold}}$ is gradually increased from 0.001 to 0.1.

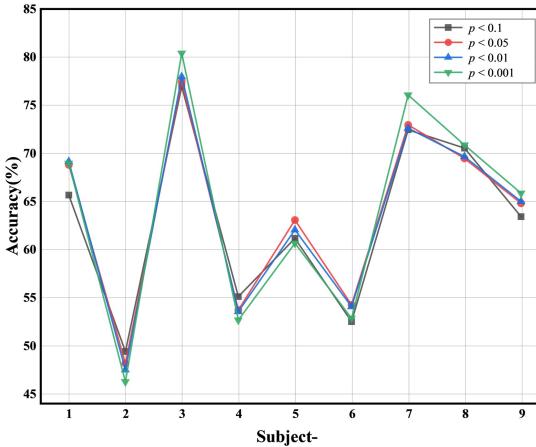


Figure 8: Cross-subject decoding results based on the BCIC IV 2a dataset, corresponding to the number of common relation items under different predefined thresholds $\alpha_{\text{threshold}}$.

significant improvements when $p < 0.05$ and $p < 0.01$, indicating that stricter feature selection enhances the model's generalization ability. However, the difference between when $p < 0.01$ and when $p < 0.001$ was not

statistically significant, suggesting diminishing returns at very strict thresholds.

5 Discussion

In this study, we proposed the CSDD model, which leverages common feature extraction to improve cross-subject decoding performance in BCI tasks. The CSDD model demonstrated superior decoding results across subjects, significantly outperforming baseline models, particularly in terms of classification accuracy and generalization ability. The results show that the CSDD model effectively handles inter-subject variability by extracting shared features, leading to improved decoding performance across subjects.

This paper involves three research directions. The first is the study of cross-subject universal BCI. The second is the method for distinguishing between common and subject-specific features. The third is the system filter.

5.1 Cross-subject universal BCI decoding model

In cross-subject BCI research, traditional BCI models are mainly based on personalized training strategies, where models are individually trained for specific brain activity patterns of each subject[36][37]. Although this approach performs well in individual tasks, it has significant limitations. Personalized training typically requires large amounts of labeled data and extensive training time, which becomes computationally expensive and difficult to scale, especially when dealing with a large number of subjects. Additionally, individual differences introduce biases in each subject's model training, resulting in poor generalization across subjects. In recent years, with advancements in technology, methods such as transfer learning and domain adaptation have been proposed to reduce the need for personalized training and improve cross-subject adaptability[38][39][40]. However, these methods do not yet fully address the problem of poor adaptability, primarily because they rely more on calibrating individual differences and lack an effective mechanism to extract common features across subjects.

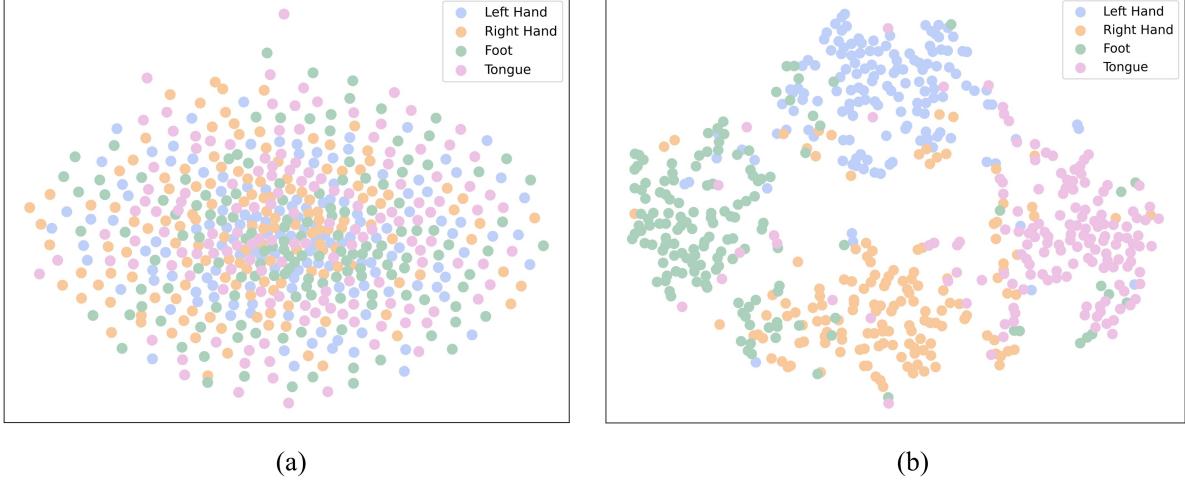


Figure 9: The visualization for Subject-3. (a) T-SNE visualization of raw data. (b) T-SNE visualization of the feature-extracted data. Different colors represent different categories.

Unlike traditional personalized BCI models, the cross-subject universal BCI model (CSDD) proposed in this paper introduces the concept of extracting common features across subjects. The innovation of this model lies in its reliance on common features shared across subjects rather than subject-specific features, constructing a universal model. This approach not only reduces the dependence on personalized data but also significantly improves the adaptability and decoding accuracy for new subjects, thus achieving better generalization across subjects.

Fig. 9 uses t-SNE (t-distributed Stochastic Neighbor Embedding) [41] to reduce the dimensionality of the data from subject 3 and displays the distribution of extracted common features across different subjects. The experimental results show that the t-SNE method successfully separates the four motor intentions (left hand, right hand, foot, and tongue), with clear separation between feature distributions of different classes. Notably, the cross-subject accuracy for Subject-3 was 80.38%, emphasizing the effective feature clustering for different motor intentions in this subject. The features extracted for subject 3 are well-clustered in the feature space, indicating that the proposed method can effectively extract common features across sub-

jects, thereby enhancing the generalization ability of the model and achieving accurate decoding. In contrast, raw EEG signals without processing show a significant overlap in the feature space, making classification more difficult. This phenomenon further validates the robustness of the proposed method, demonstrating its ability to extract easily classifiable features from complex and noisy EEG signals, significantly improving cross-subject decoding accuracy.

5.2 Distinguish between common and subject-specific features

Traditional BCI decoding methods typically rely on optimizing the temporal and spatial dimensions or adjusting the decoding functions to enhance the learning of common features across subjects[42][43][44]. Deep learning methods, on the other hand, approximate these common features through large-scale data training, thereby improving decoding accuracy[45][46][47]. Although these methods improve model generalization by increasing common features across individuals and reducing subject-specific features, they typically rely on a single strategy and lack clear methods for distinguishing and optimizing common and subject-specific

features. In contrast, the DD algorithm, as a white-box model, offers lower computational complexity and strong generalization capability, capable of expressing both the independent and interactive effects of input variables on output through the relation spectrum.

Therefore, this paper proposes DD Filter, which can simplify the formula of the trained personalized model into a polynomial form, using random probability statistical tests to select statistically significant common relation items and filter out subject-specific relation items.

The significance level is controlled through one sample t -test, which regulates the number of common relation items. As shown in Fig. 7, as the predefined threshold $\alpha_{\text{threshold}}$ decreases from 0.1 to 0.001, the number of extracted common relation items progressively decreases. This indicates that more stringent significance levels select fewer features. In Fig. 8, decoding accuracy fluctuates with variations in the p -value. Although a slight improvement is observed when $p < 0.1$, the accuracy change is minimal and does not lead to a significant performance enhancement under stricter thresholds. This suggests that under excessively strict significance levels (e.g., $p < 0.001$), although the number of features is reduced, their contribution to cross-subject decoding is limited and may even detract from the performance of the model. This phenomenon further underscores the necessity of finding an appropriate balance between feature quantity and decoding performance during model optimization. Proper feature selection and avoidance of overfitting are essential for effectively enhancing cross-subject decoding performance.

5.3 System component filtering method with a filter-like structure

Traditional filtering methods, especially in signal processing, typically rely on Fourier transforms to convert time-domain signals into the frequency domain. This allows for the removal of noise while retaining useful frequency components. However, despite its success in signal level applications, this concept has not yet been effectively applied to filtering out irrelevant components at the system level.

The main contribution of this paper is the extension

of the filtering concept to relation system processing. Specifically, a relation filtering mechanism is introduced to identify and remove irrelevant or redundant relation components within a system.

In the proposed framework, the personalized model for each subject is initially transformed into a relation spectrum, similar to the transformation of time-domain signals into frequency spectra in traditional signal processing. This transformation allows subject-specific EEG decoding features to be represented within a unified relation spectrum, making it easier to compare and analyze patterns across subjects. By applying statistical analysis, shared components across individuals are identified, while individual differences are treated as “noise” and filtered out.

Unlike conventional time-frequency feature filtering methods, the proposed approach focuses on filtering at the relational level of the system rather than merely extracting time frequency features. By capturing shared relational structures across subjects and eliminating redundant information caused by individual variability, the proposed filtering method aims to optimize the cross-subject representation of EEG data. The goal is to remove unnecessary subject-specific information and retain only those common features that generalize well across subjects, thus enabling the construction of a more universal and adaptive BCI model.

6 Conclusion

In summary, this study proposes a novel cross-subject BCI decoding model based on the extraction of common features. By training personalized models for each subject and subsequently extracting common relation items across subjects through the system filter proposed in this study, a generalized cross-subject BCI model has been successfully constructed, achieving complete cross-subject decoding. Experimental results demonstrate that the proposed method effectively improves decoding accuracy for new subjects, offering better generalization compared to traditional methods. Under cross-subject conditions, the method achieved an average accuracy of 64.85% on the BCIC IV 2a dataset.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (62303423), the STL 2030-Major Project (2022ZD0208500), Postdoctoral Science Foundation of China (2024T170844, 2023M733245), the Henan Province key research and development and promotion of special projects (242102311239), Shaanxi Provincial Key Research and Development Program (2023GXHLH-012).

Data availability

This study utilizes publicly available datasets, with the corresponding data links provided in the manuscript.

References

- [1] Dennis J McFarland and Jonathan R Wolpaw. Brain-computer interfaces for communication and control. *Communications of the ACM*, 54(5):60–66, 2011.
- [2] Bram van de Laar, Danny Plass-Oude Bos, Boris Reuderink, Mannes Poel, and Anton Nijholt. How much control is enough? influence of unreliable input on user experience. *IEEE transactions on cybernetics*, 43(6):1584–1592, 2013.
- [3] Jinyi Long, Yuanqing Li, Hongtao Wang, Tianyou Yu, Jiahui Pan, and Feng Li. A hybrid brain computer interface to control the direction and speed of a simulated or real wheelchair. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(5):720–729, 2012.
- [4] Aya Rezeika, Mihaly Benda, Piotr Stawicki, Felix Gembler, Abdul Saboor, and Ivan Volosyak. Brain-computer interface spellers: A review. *Brain sciences*, 8(4):57, 2018.
- [5] Gabriel Alves Mendes Vasiljevic and Leonardo Cunha De Miranda. Brain-computer interface games based on consumer-grade eeg devices: A systematic literature review. *International Journal of Human-Computer Interaction*, 36(2):105–142, 2020.
- [6] Kaishuo Zhang, Neethu Robinson, Seong-Whan Lee, and Cuntai Guan. Adaptive transfer learning for eeg motor imagery classification with deep convolutional neural network. *Neural Networks*, 136:1–10, 2021.
- [7] Jordy Thielen, Pieter Marsman, Jason Farquhar, and Peter Desain. From full calibration to zero training for a code-modulated visual evoked potentials for brain-computer interface. *Journal of Neural Engineering*, 18(5):056007, 2021.
- [8] Phairot Autthasan, Rattanaphon Chaisaen, Thapanun Sudhawiyangkul, Phurin Rangpong, Suktipol Kiatthaveephong, Nat Dilokthanakul, Gun Bhakdisongkhram, Huy Phan, Cuntai Guan, and Theerawit Wilaiprasitporn. Min2net: End-to-end multi-task learning for subject-independent motor imagery eeg classification. *IEEE Transactions on Biomedical Engineering*, 69(6):2105–2118, 2021.
- [9] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain-computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.
- [10] Syed Umar Amin, Mansour Alsulaiman, Ghulam Muhammad, Mohamed Amine Mekhtiche, and M Shamim Hossain. Deep learning for eeg motor imagery classification based on multi-layer cnns feature fusion. *Future Generation computer systems*, 101:542–554, 2019.
- [11] Dalin Zhang, Lina Yao, Kaixuan Chen, and Jessica Monaghan. A convolutional recurrent attention model for subject-independent eeg signal analysis. *IEEE signal processing letters*, 26(5):715–719, 2019.
- [12] Vinay Jayaram, Morteza Alamgir, Yasemin Altun, Bernhard Scholkopf, and Moritz Grosse-Wentrup. Transfer learning in brain-computer in-

- terfaces. *IEEE Computational Intelligence Magazine*, 11(1):20–31, 2016.
- [13] Rito Clifford Maswanganyi, Chungling Tu, Pius Adewale Owolawi, and Shengzhi Du. Multi-class transfer learning and domain selection for cross-subject eeg classification. *Applied Sciences*, 13(8):5205, 2023.
- [14] Ruilong Zhang, Qun Zong, Liqian Dou, Xinyi Zhao, Yifan Tang, and Zhiyu Li. Hybrid deep neural network using transfer learning for eeg motor imagery decoding. *Biomedical Signal Processing and Control*, 63:102144, 2021.
- [15] Boxun Fu, Fu Li, Youshuo Ji, Yang Li, Xuemei Xie, and Guangming Shi. Scdn: Learning common feature representation of brain activation for intersubject motor imagery eeg decoding. *IEEE Transactions on Instrumentation and Measurement*, 72:1–15, 2023.
- [16] Guanghai Dai, Jun Zhou, Jiahui Huang, and Ning Wang. Hs-cnn: a cnn with hybrid convolution scale for eeg motor imagery classification. *Journal of neural engineering*, 17(1):016025, 2020.
- [17] Fang Wang, Sheng-hua Zhong, Jianfeng Peng, Jianmin Jiang, and Yan Liu. Data augmentation for eeg-based emotion recognition with deep convolutional neural networks. In *MultiMedia Modeling: 24th International Conference, MMM 2018, Bangkok, Thailand, February 5–7, 2018, Proceedings, Part II 24*, pages 82–93. Springer, 2018.
- [18] Eunjin Jeon, Wonjun Ko, Jee Seok Yoon, and Heung-Il Suk. Mutual information-driven subject-invariant and class-relevant deep representation learning in bci. *IEEE Transactions on Neural Networks and Learning Systems*, 34(2):739–749, 2021.
- [19] Yunyuan Gao, Mengting Li, Yun Peng, Feng Fang, and Yingchun Zhang. Double stage transfer learning for brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:1128–1136, 2023.
- [20] Peiyin Chen, He Wang, Xinlin Sun, Haoyu Li, Celso Grebogi, and Zhongke Gao. Transfer learning with optimal transportation and frequency mixup for eeg-based motor imagery recognition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:2866–2875, 2022.
- [21] Oleksandr Zlatov and Benjamin Blankertz. Towards physiology-informed data augmentation for eeg-based bcis. *arXiv preprint arXiv:2203.14392*, 2022.
- [22] Xinzhi Ma, Weihai Chen, Zhongcai Pei, Yue Zhang, and Jianer Chen. Attention-based convolutional neural network with multi-modal temporal information fusion for motor imagery eeg decoding. *Computers in Biology and Medicine*, 175:108504, 2024.
- [23] Gang Liu and Jing Wang. Dendrite net: A white-box module for classification, regression, and system identification. *IEEE Transactions on Cybernetics*, 52(12):13774–13787, 2021.
- [24] Gang Liu. It may be time to improve the neuron of artificial neural network. *TechRxiv. Preprint*, 2020.
- [25] Gang Liu and Jing Wang. Egg: An analytic brain-computer interface algorithm. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:643–655, 2022.
- [26] Gang Liu and Jing Wang. A relation spectrum inheriting taylor series: Muscle synergy and coupling for hand. *Frontiers of Information Technology & Electronic Engineering*, 23(1):145–157, 2022.
- [27] Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human Brain Mapping*, 38:5391–5420, 2017.
- [28] Xinqiao Zhao, Hongmiao Zhang, Guilin Zhu, Fengxiang You, Shaolong Kuang, and Lining Sun.

- A multi-branch 3d convolutional neural network for eeg-based motor imagery classification. *IEEE transactions on neural systems and rehabilitation engineering*, 27(10):2164–2177, 2019.
- [29] Wei Zhao, Xiaolu Jiang, Baocan Zhang, Shixiao Xiao, and Sujun Weng. Ctnet: a convolutional transformer network for eeg-based motor imagery classification. *Scientific Reports*, 14(1):20237, 2024.
- [30] Elias Chaibub Neto, Abhishek Pratap, Thanneer M Perumal, Meghasyam Tummalacherla, Phil Snyder, Brian M Bot, Andrew D Trister, Stephen H Friend, Lara Mangravite, and Larsson Omberg. Detecting the impact of subject characteristics on machine learning-based diagnostic applications. *NPJ digital medicine*, 2(1):99, 2019.
- [31] Dalin Zhang, Kaixuan Chen, Debao Jian, and Lina Yao. Motor imagery classification via temporal attention cues of graph embedded eeg signals. *IEEE journal of biomedical and health informatics*, 24(9):2570–2579, 2020.
- [32] Hauke Dose, Jakob S Møller, Helle K Iversen, and Sadasivan Puthusserypady. An end-to-end deep learning approach to mi-eeg signal classification for bcis. *Expert Systems with Applications*, 114: 532–542, 2018.
- [33] Xiaoxi Wei, Pablo Ortega, and A Aldo Faisal. Inter-subject deep transfer learning for motor imagery eeg decoding. In *2021 10th international IEEE/EMBS conference on neural engineering (NER)*, pages 21–24. IEEE, 2021.
- [34] Alberto Zancanaro, Giulia Cisotto, João Ruivo Paulo, Gabriel Pires, and Urbano J Nunes. Cnn-based approaches for cross-subject classification in motor imagery: From the state-of-the-art to dynamicinet. In *2021 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)*, pages 1–7. IEEE, 2021.
- [35] Xianheng Wang, Veronica Liesaputra, Zhaobin Liu, Yi Wang, and Zhiyi Huang. An in-depth survey on deep learning-based motor imagery electroencephalogram (eeg) classification. *Artificial intelligence in medicine*, 147:102738, 2024.
- [36] Hanrui Wu, Zhengyan Ma, Zhenpeng Guo, Yanxin Wu, Jia Zhang, Guoxu Zhou, and Jinyi Long. Online privacy-preserving eeg classification by source-free transfer learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2024.
- [37] Demetres Kostas and Frank Rudzicz. Thinker invariance: enabling deep neural networks for bci across more people. *Journal of Neural Engineering*, 17(5):056008, 2020.
- [38] Chun-Shu Wei, Yuan-Pin Lin, Yu-Te Wang, Chin-Teng Lin, and Tzyy-Ping Jung. A subject-transfer framework for obviating inter-and intra-subject variability in eeg-based drowsiness detection. *NeuroImage*, 174:407–419, 2018.
- [39] Lichao Xu, Minpeng Xu, Yufeng Ke, Xingwei An, Shuang Liu, and Dong Ming. Cross-dataset variability problem in eeg decoding with deep learning. *Frontiers in human neuroscience*, 14: 103, 2020.
- [40] Xingfu Wang, Wenjie Yang, Wenxia Qi, Yu Wang, Xiaojun Ma, and Wei Wang. Starnet: A spatio-temporal and riemannian network for high-performance motor imagery decoding. *Neural Networks*, 178:106471, 2024.
- [41] Olawunmi George, Roger Smith, Praveen Madiraju, Nasim Yahyasoltani, and Sheikh Iqbal Ahamed. Data augmentation strategies for eeg-based motor imagery decoding. *Heliyon*, 8(8), 2022.
- [42] Junjian Chen, Zhuliang Yu, Zhenghui Gu, and Yuanqing Li. Deep temporal-spatial feature learning for motor imagery-based brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(11):2356–2366, 2020.

- [43] Jing Jin, Tingnan Qu, Ren Xu, Xingyu Wang, and Andrzej Cichocki. Motor imagery eeg classification based on riemannian sparse optimization and dempster-shafer fusion of multi-time-frequency patterns. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31: 58–67, 2022.
- [44] Yangang Li, Yu Qi, Yiwen Wang, Yueming Wang, Kedi Xu, and Gang Pan. Robust neural decoding by kernel regression with siamese representation learning. *Journal of Neural Engineering*, 18(5): 056062, 2021.
- [45] Chang Wang, Yang Wu, Chen Wang, Yu Zhu, Chong Wang, Yanxiang Niu, Zhenpeng Shao, Xudong Gao, Zongya Zhao, and Yi Yu. Mi-eeg classification using shannon complex wavelet and convolutional neural networks. *Applied Soft Computing*, 130:109685, 2022.
- [46] Hongli Li, Hongyu Chen, Ziyu Jia, Ronghua Zhang, and Feichao Yin. A parallel multi-scale time-frequency block convolutional neural network based on channel attention module for motor imagery classification. *Biomedical Signal Processing and Control*, 79:104066, 2023.
- [47] Kahoko Takahashi, Zhe Sun, Jordi Solé-Casals, Andrzej Cichocki, Anh Huy Phan, Qibin Zhao, Hui-Hai Zhao, Shangkun Deng, and Ruggero Micheletto. Data augmentation for convolutional lstm based brain computer interface system. *Applied Soft Computing*, 122:108811, 2022.