Fine-Grained Spatial-Frequency-Time Framework for Motor Imagery Brain-Computer Interface

Guoyang Liu, *IEEE, Member*, Rui Zhang, Lan Tian, Weidong Zhou

***Abstract*—The Motor Imagery Brain-Computer Interfaces (MI-BCIs) have shown considerable promise for applications in neural rehabilitation. However, improving the practicality and interpretability of MI-BCIs remains a critical challenge. Unlike previous methods that focus generally on either spatial, frequency, or temporal domains with coarse-grained segmentation schemes, this study proposes a novel fine-grained spatial-frequency-time (FGSFT) framework, aiming to enhance the efficiency and reliability of MI-BCIs. Multi-channel MI EEG recordings are firstly processed through multiscale time-frequency segmentation and spatial segmentation schemes, yielding fine-grained spatial-frequency-time segments (SFTSs). The key SFTSs are then selected with a tailored wrapper-based feature selection approach. Discriminative MI EEG features are extracted using a divergence-based common spatial pattern algorithm with intra-class regularization and classified using an efficient linear support vector machine (SVM). The proposed framework was evaluated on the BCI IV IIa and SDU-MI datasets, demonstrating state-of-the-art performance in terms of information transfer rate (ITR) Meanwhile, the proposed spatial segmentation strategy can significantly improve the performance of MI-BCIs when using a larger number of electrodes. Additionally, the fine-grained Motor Imagery Time-Frequency Reaction Map (MI-TFRM) and time-frequency topographical map can be obtained with the proposed framework enabling visualization of the subject-specific dynamic neural process during motor imagery tasks, facilitating the devising of personalized MI-BCIs. The FGSFT framework significantly advances the accuracy, ITR, and interoperability of MI-BCIs, paving the way for future neuroscientific research and clinical applications in neural rehabilitation and assistive technologies.**

***Index Terms*—Motor imagery, brain-computer interface, EEG, common spatial pattern.**

I. Introduction

In the field of neural engineering, Motor Imagery Brain-Computer Interfaces (MI-BCIs) offer a transformative approach to bridging human cognition with external systems by decoding electroencephalogram (EEG) signals [1], [2]. By facilitating non-muscular interaction, MI-BCIs offer individuals with severe motor impairments renewed opportunities for communication and control [3]. This transformative paradigm not only improves the quality of their lives but also lays a foundation for innovative rehabilitation strategies [4]. Despite significant progress, improving the accuracy, robustness, and interpretability of MI-BCI systems remains challenging.

Deep learning approaches, including convolutional and recurrent neural networks, have emerged as powerful tools due to their end-to-end feature learning capabilities and their capacity to model complex temporal dependencies [5]. Nevertheless, these models often require large, high-quality datasets for optimized performance, a condition rarely met in MI-BCI practice due to the inherently noisy EEG signals and the limited EEG data obtained from controlled MI tasks [6]. Furthermore, while attention-based and hybrid deep learning architectures have been applied to identify discriminative features and key temporal or spatial locations [7], [8], the underlying decision-making processes remain unclear [9], making it difficult to map their learned representations onto the neurophysiological underpinnings of motor imagery. This limitation will be further compounded by focusing only on either spatial or temporal aspects, or segmenting frequency bands at a coarse-grained level, possibly resulting in missing the subtle but critical neural oscillatory patterns. Such interpretability deficits not only hinder the neuroscientific understanding of dynamic brain mechanisms but also pose practical barriers to clinical adoption and personalization. In scenarios where subject-specific adaptations are important, the inability to elucidate how these models make their decisions impedes the development of individualized neural decoding strategies and limits the MI-BCI system's credibility and efficacy.

Although some previous studies have taken initial steps toward incorporating time-frequency domain segmentation, these strategies often remain at the coarse-grained stage and fail to integrate with spatial segmentation methods [10], [11]. While several studies have explored multi-scale time and frequency segmentation strategies [12], [13], Few have presented a holistic approach that unifies finely granulated segmentation across spatial, frequency, and temporal domains. Without such integrated fine-grained strategies, current methods struggle to identify and preserve the fine-grained and critical spatial-frequency-time features that are pivotal for achieving both high classification accuracy and interpretability. The absence of a comprehensive framework that leverages multi-scale time-frequency segmentation with a spatially segmented architecture underscores a key unmet need in MI-BCI research. This gap motivates us to develop approaches capable of representing the complex, subject-specific neural dynamics that govern motor imagery tasks in a more interpretable and performance-optimized manner.

The support in part by the National Natural Science Foundation of China (No. 62401342, No. 62271291), the Key Program of Natural Science Foundation of Shandong Province (No. ZR2020LZH009), the Shenzhen Science and Technology Program (No. GJHZ20220913142607013) and the Natural Science Foundation of Shandong Province (No. ZR2024QF092, No. ZR2021ZD40, No. ZR2021MF065) was gratefully acknowledged. *(Corresponding author: Weidong Zhou; Lan Tian.)*

Guoyang Liu, Rui Zhang, Lan Tian, and Weidong Zhou are with the School of Integrated Circuits, Shandong University, Jinan, China (e-mail: gyliu@sdu.edu.cn; 202320399@mail.sdu.edu.cn; tianlan65@sdu.edu.cn; wdzhou@sdu.edu.cn).

Manuscript received xxx xx, 20xx; revised xxx xx, 20xx.

To address these limitations, we propose a novel fine-grained spatial-frequency-time (FGSFT) framework. Our approach introduces a multiscale time-frequency domain segmentation combined with a comprehensive spatial segmentation strategy to generate discriminative Spatial-Frequency-Time Segments (SFTSs) from multi-channel EEG data. By creating multiple subsets of electrodes corresponding to various spatial domains using both manual and automatic channel selection strategies, our method allows for a more comprehensive capture of the complex spatial patterns inherent in EEG data. A wrapper-based feature selection algorithm is then employed to identify the most informative SFTSs, enabling superior classification accuracy and improved computational efficiency. Notably, this work not only focuses on improving the accuracy of subject-specific brain-computer interfaces but also aims to investigate the subject-specific neural mechanisms implicated in motor imagery tasks through fine-grained time-frequency topographical maps generated by the proposed framework. These visualizations improve interpretability and provide insights into individualized neural activities, enabling the design of more personalized and effective MI-BCIs. The main contributions of this work are:

- We propose a novel fine-grained spatial-frequency-time (FGSFT) framework that significantly enhances the extraction and classification of MI-BCI signals. The FGSFT framework integrates a multiscale time-frequency segmentation with a spatial segmentation strategy, enabling the accurate extraction of discriminative SFTSs from multi-channel EEG data.
- We introduce a wrapper-based feature selection algorithm to identify the most informative STFSs, improving both the accuracy and efficiency of MI-BCI classification.
- We achieve state-of-the-art performance on two benchmark MI-BCI datasets in terms of information transfer rate (ITR), demonstrating the reliability and efficiency of the proposed approach.
- The fine-grained Motor Imagery Time-Frequency Reaction Map (MI-TFRM) and time-frequency topographical map can be obtained with the proposed framework to allow for visualization of the subject-specific dynamic neural process and the individual variability in neural responses during motor imagery tasks, thereby significantly improving the model interpretability and facilitating the design of personalized MI-BCIs.

The remainder of this paper is organized as follows: In Section II, we comprehensively review the existing literature on MI-BCI methods. Section III details the experimental materials and the proposed methods. Section IV presents the evaluation results on two MI-BCI datasets. Section V provides the visualization and discussion of the obtained findings. Finally, Section VI draws the conclusion.

II. Related Works

Early efforts in MI-BCI research focused mainly on optimizing spatial filters to extract discriminative features from EEG signals. The Common Spatial Pattern (CSP) algorithm, which relies on the covariance between spatially filtered EEG signals, has been widely used for feature extraction [14]. Despite its popularity, the CSP method is prone to overfitting and sensitivity to noise, leading to the development of Regularized CSP (RCSP) [15]. RCSP incorporates quadratic penalties based on prior knowledge, yielding improved robustness and generalization. Subsequently, advanced variants have explored L_p-norms to further refine spatial filters. For example, Park et al. [16] introduced CSP-L_p to identify optimal p values for improved MI performance, while Cai et al. [17] developed CSP-$L_{p/q}$ to enhance generalization across multiple BCI competition datasets. Complementary approaches have redefined CSP as a divergence maximization problem. Samek et al. [18], [19] introduced a divergence-based CSP framework that utilizes gradient descent on an orthogonal manifold, enabling sparse spatial filters that have shown effectiveness in multi-class MI paradigms [20]–[22]. Parallel to these endeavors, Riemannian geometry-based methods have gained traction by formulating MI-BCI feature extraction as a problem within a geometric manifold [23], [24]. Although these Riemannian geometry-based methods often demonstrate robust performance, their high computational complexity and extensive feature dimensionality limit their feasibility in real-world applications [25].

With the advent of powerful computing resources, deep learning has emerged as another dominant approach for MI-EEG feature extraction and classification [26]. Convolutional neural networks (CNNs) have been extensively explored, ranging from compact architectures like EEGNet [27] to deeper networks that extract complex spatial-temporal patterns [28]–[30]. Recurrent neural networks (RNNs), with their inherent capability to model temporal dependencies, have also been employed [31], [32]. Moreover, attention-based architectures have been proposed to highlight the most discriminative features [8], [33]. Recently, transformer-based models such as Vision Transformer (ViT) have gained considerable attraction in MI-BCI fields due to their global dependency capturing capability, further enhancing the accuracy of MI-BCI [7], [34], [35]. Nevertheless, these deep learning architectures also depend heavily on substantial training data and often exhibit limited interpretability. Meanwhile, the attention patterns generated by attention modules cannot straightforwardly translate into human-understandable explanations of the underlying neural mechanisms.

To address the limited exploration of frequency dynamics, numerous approaches have adopted various frequency-domain strategies. The sub-band CSP (SBCSP) [10] pioneered decomposing EEG signals into multiple frequency bands, while filter-bank CSP (FBCSP) [11] further integrated feature selection algorithms to identify key frequency bands. Afterward, it was found that applying the time domain segmentation strategy effectively enhanced the performance of CSP-based MI-BCI algorithms [36]. Subsequently, more sophisticated time-frequency segmentation methods have been introduced. For instance, Miao et al. [12] and Wang et al. [13] subdivided EEG data into multiple time-frequency segments to capture a richer set of discriminative features. Recently, Pei et al. [37] greatly extended the FBCSP with a novel tensor-based frequency features combination method, which combined tensor-to-vector projection, fast Fourier transform, CSP, and feature fusion,

resulting in robust improvements in MI-BCI accuracy. The work visualized the personalized useful frequency bands and confirmed the importance of frequency information in MI-BCI systems. Chen et al. [38] developed an optimization strategy based on dynamic time windows, improving the efficiency of the SSVEP-BCI system. Additionally, Riemannian geometry-based methods have also incorporated similar segmentation strategies [39], [40], and multiscale CNN architectures have been proposed to exploit multi-band temporal patterns [41]. Despite these advancements, they mainly rely on coarse-grained segmentation schemes and fail to fully exploit fine-grained multi-dimensional EEG feature extraction.

In addition to time-frequency segmentation, spatial optimization has also attracted extensive attention. Several studies proposed data augmentation approaches based on channel-level and time-level perturbation [8], [42], enhancing the deep learning-based MI-BCI algorithms. On the other hand, CSP-based time-frequency optimization using multiple intervals and bands has been demonstrated in [43]–[46], while others integrated channel selection with RCSP [47], or combined channel selection with deep learning models [48], [49]. Jiang et al. [50] proposed a spatio-temporal filtering strategy for CSP algorithm, which can update the spatial and temporal filters automatically. More recent work [51] presented a method to extract the multi-view time-frequency decomposed spatial feature matrix using CSP-based and Riemannian-based algorithms, achieving improved accuracy of MI-BCI. These methods mark important steps toward capturing localized spatial features and improving classification performance. However, they generally applied channel weighting strategy (e.g., CSP-based methods) or combined it with single-scale spatial segmentation strategy (e.g., channel selection-based methods), and the multi-scale fine-grained spatial features of EEG have not been fully explored.

These gaps in the literature motivate us to investigate the fine-grained spatial-frequency-time (FGSFT) framework. By simultaneously and systematically segmenting the EEG data in spatial, frequency, and temporal domains and coupling these strategies with effective feature selection methods, our proposed approach significantly improves both performance and interpretability of MI-BCIs.

III. Materials and Methods

In this work, a fine-grained spatial-frequency-time framework is proposed for motor imagery brain-computer interface. Fig. 1 illustrates the overall architecture of the proposed framework. In the training phase, the pre-processed EEG data are first segmented into fine-grained SFTSs, and then the discriminative EEG features are extracted by divCSP. After that, an effective wrapper-based feature selection approach is applied, and the information of the selected SFTs is saved. Finally, the optimized ensembled SVM model is trained. In the testing phase, the testing EEG data are segmented into optimized SFTSs according to the selected SFTs information obtained from the training phase, and they are then sent into the optimized model to obtain the predicted label. Moreover, the thorough time-frequency-spatial analyses using the interpretable feature selection approach are discussed in the Discussion section.

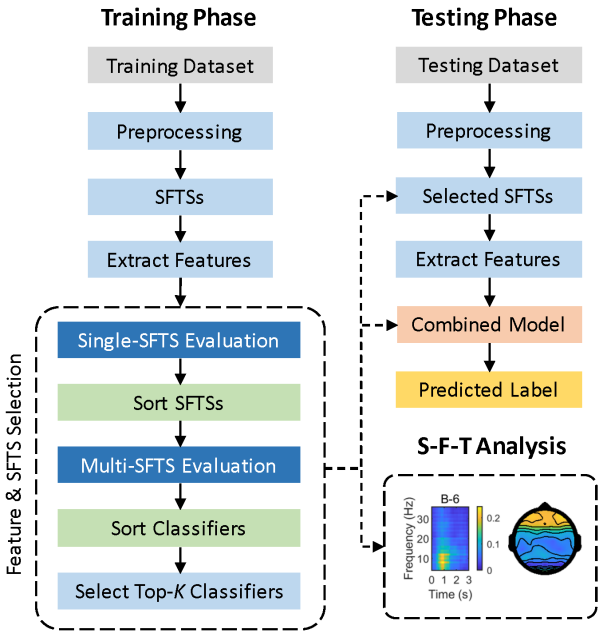

Fig. 1. The overall architecture of the proposed BCI framework.

A. EEG datasets

Two motor imagery EEG datasets were adopted in this study to evaluate the proposed algorithms comprehensively, where the first one was the BCI IV IIa dataset [52], [53], a widely used and publicly available MI-BCI dataset, and another was a BCI dataset collected by ourselves. The BCI IV IIa dataset comprised 25-channel EEG data from nine subjects, where 22 of which were scalp EEG channels (as shown in Fig. 2(a)), and the remaining three were monopolar electrooculography (EOG) channels. The EEG data was sampled at a frequency of 250 Hz and pre-filtered by a bandpass filter in the range of 0.5 Hz to 100 Hz, including a 50 Hz notch filter to avoid line noise. This study focuses on the EEG channels, omitting the EOG channels to concentrate on the brain activity relevant to motor imagery tasks. The dataset was collected with a typical motor imagery experimental paradigm, and the details of which are illustrated in Fig. 3(a). This paradigm consisted of four types of motor imagery tasks: left hand, right hand, tongue, and feet, and these four tasks are equally and randomly distributed in the collection procedure of 288 trials per subject during the training session, with the same number of trials conducted in the test session. Note that these sessions were conducted on separate days. Each session consisted of six blocks, with each block containing 48 trials. Subjects were given short rest periods between blocks to prevent fatigue. To simplify the task, we only consider the left/right-hand imagery task in this work. Each trial in this dataset is segmented from 2 seconds to 5 seconds post-stimulus onset.

The SDU-MI (Shandong University - Motor Imagery) dataset was collected for a more comprehensive evaluation

of the proposed framework. This dataset contained EEG recordings from ten participants (7 males and 3 females), all aged 18 to 25 years and affiliated with Shandong University. These recordings were captured using the NeuSen-W64 EEG data acquisition system, equipped with 59 electrodes aligned according to the extended 10-20 international system (see Fig. 2(b)). The EEG signals were sampled at 1000Hz, with a reference electrode located between Pz and Cz, and a ground electrode between Fpz and Fz. Three separate binary motor imagery tasks were performed: imagining left versus right-hand movements (with a fixed elbow position to concentrate on hand motion), left versus right elbow movements (with a fixed hand position to concentrate on elbow motion), and combined left elbow and hand versus right elbow and hand movements (e.g., simulating actions like reaching out and grasping a cup). These tasks were chosen to reflect distinct motor control areas, enriching the applicability of datasets in developing brain-computer interfaces. Participants completed these tasks in three separate sessions conducted on the same day, with a rest period of 10-30 minutes between sessions. The order of the tasks was randomized and determined by the participants' preferences. Experimentally, each motor imagery task followed a consistent paradigm, similar to the paradigm in BCI IV IIa dataset. The paradigm for each binary task comprised five blocks, each containing 40 trials where the two classes of motor imagery were equally and randomly presented. As shown in Fig. 3(b), a trial began with a preparatory phase (0-2 seconds), where cues in the form of images depicting the hand, elbow, or hand plus elbow were displayed, followed by a directive phase (2-4 seconds), where cues indicating the direction of imagined movement (left or right) were shown. Participants were instructed to conduct their motor imagery following these cues. The trial concluded with a rest period (6-9 seconds), signified by a blank screen, allowing for mental recovery before proceeding to the next trial. This study was approved by the Ethics Committee of Qilu Hospital of Shandong University.

In the preprocessing phase, the raw EEG data from both datasets were uniformly resampled to a frequency of 250Hz. To simulate real-world conditions and validate the robustness of our model, we refrained from excluding any trials or channels, preserving the authenticity and complexity of EEG datasets. To evaluate the efficacy of the proposed algorithm, we adopted a Leave-One-Block-Out (LOBO) cross-validation strategy in both two MI databases for within-session evaluation. This validation approach ensures that each block of data is used once as the test set, with the remaining blocks forming the training set. Specifically, a 6-fold LOBO cross-validation was conducted on the BCI IV IIa dataset. Concurrently, the SDU-MI dataset, with five blocks per session, was subjected to a 5-fold LOBO cross-validation. Considering that the BCI IV IIa database has two sessions, and the second session was determined as the test set in the BCI competition, we also report the accuracy and ITR on its test set for cross-session evaluation. It is important to note that both the classifier training and feature ranking procedures were only performed on the training set, avoiding information leakage.

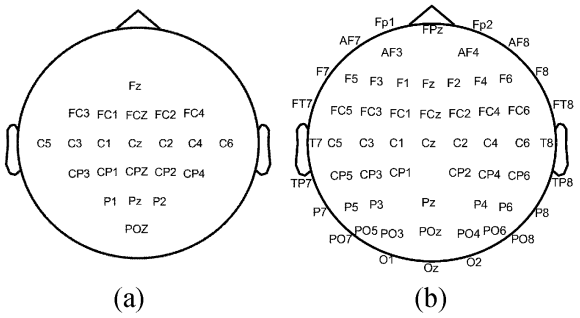

Fig. 2. The electrode placement of (a) BCI IV IIa dataset and (b) SDU-MI dataset

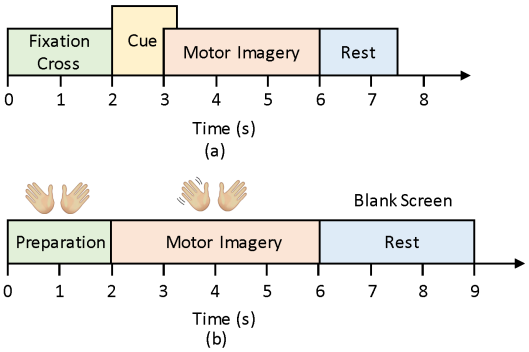

Fig. 3. The EEG collection paradigm of two databases. (a) Paradigm of BCI IV IIa dataset. (b) Paradigm of SDU-MI dataset.

B. S-F-T segmentation scheme

To comprehensively capture the temporal, spectral, and spatial characteristics of EEG signals, we introduce a novel framework for multi-scale, fine-grained segmentation across time, frequency, and space dimensions. In the time domain, the scale is defined by the length of the time window, and the time segmentation is executed by windowing. We have developed a multi-scale temporal windowing approach with window lengths of 0.2, 0.4, 0.7, 1.0, 1.25, and 1.5 seconds, each overlapping by 50%. For instance, Figure 4(a) illustrates the multi-scale temporal windows corresponding to EEG data with a duration of 1.5 seconds. All temporal segments are obtained using rectangular windows, ensuring that for each scale, the last window terminates at the end of the EEG data. Furthermore, this multi-scale temporal windowing strategy, featuring six distinct window lengths, also applies to EEG data exceeding 1.5 seconds.

In the frequency domain, the scale corresponds to the bandwidth of the frequency, and the frequency segmentation is implemented by band-pass filtering. We proposed a multi-scale frequency band set with bandwidths of 4, 8, 16, and 32 Hz, each with 50% overlap. Figure 4(b) depicts these bands, where each line represents the passband of a bandpass filter, with the left and right endpoints indicating the start and stop frequencies, respectively. Notably, the final scale with a 4 Hz bandwidth encompasses the range of 30-34 Hz. For different frequency bands, we utilize a 5th-order Butterworth bandpass filter to process the EEG signals, thereby acquiring the segmented signals in the frequency domain. To

preserve the time-domain information and eliminate the phase distortion, zero-phase filtering was employed. Specifically, the input signal was initially filtered in the forward direction, and the output after this forward filtering step was reversed and processed again through the same filter in the backward direction, resulting in a zero-phase filtered signal.

In the spatial domain (channel domain) for EEG analysis, the selection of an optimal set of electrodes is an NP-hard problem. This is because the number of possible combinations of electrodes grows exponentially with the number of electrodes, thus making exhaustive search impractical for larger sets. Accordingly, this study introduces a hybrid approach that combines manual and data-driven channel selection to conduct the spatial segmentation of EEG signals. The manual selection of electrodes is grounded in the well-established ERD/ERS phenomenon during motor imagery tasks. Specifically, motor imagery of left and right-hand movements induces contralateral desynchronization (ERD) over the sensorimotor cortex, while ipsilateral synchronization (ERS) can also occur [3]. This contralateral ERD and ipsilateral ERS symmetry is a key neurophysiological foundation leveraged in the manual channel selection process. Given this principle, our manual selection of channels was based on creating symmetrical (left-right) electrode subsets to capture and exploit this ERD/ERS symmetry effectively. Fig. 5(a) illustrates ten different electrode configurations for the 22-channel EEG data in the BCI IV IIa dataset. The first eight configurations are systematically designed to explore the spatial dynamics of motor-related ERD/ERS phenomena. The last two configurations are generated using a correlation-based channel selection algorithm proposed in [54], comprising 8 and 12 channels, respectively. Furthermore, Fig. 5(b) displays 20 electrode subsets used for the 59-channel EEG data from the SDU-MI database. The first 18 electrode sets are manually designed, while the final two sets are also generated by the same channel selection algorithm mentioned in [54], containing 9 and 18 channels, respectively. The manually designed subsets systematically target symmetrical ERD/ERS patterns. For instance, some sets focus on the frontal lobes, as shown in set 4 of Fig. 5(b), and the occipital lobes, as in set 10 of Fig. 5(b). They explore how these regions contribute to motor imagery tasks, even though primary ERD/ERS phenomena occur in the central sensorimotor areas. Furthermore, electrode subsets such as set 12 apply decimation strategies to reduce redundancy while maintaining hemispheric symmetry. By designing electrode sets that correspond to different spatial patterns and complementing them with automatic channel selection algorithms, we can effectively enrich the spatial features extracted for various EEG-based tasks.

C. Motor imagery EEG feature extraction

In this study, discriminative motor imagery EEG features are extracted by incorporating a divergence-based CSP algorithm with within-class regularization. This approach is designed for a binary classification scenario with EEG data comprising D channels. The aim of our methodology is the optimization of an orthogonal rotation matrix $R \in \mathbb{R}^{D\times D}$, which is calibrated

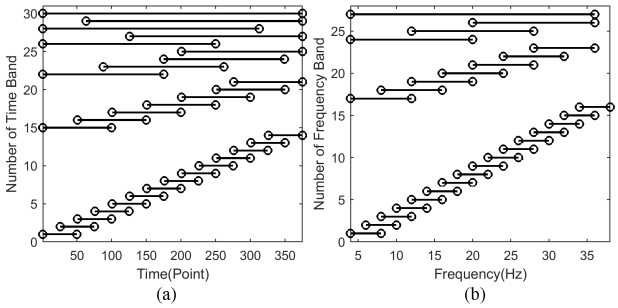

Fig. 4. The fine-grained segmentation schemes in time and frequency domains. (a) The fine-grained temporal segmentation scheme. (b) The fine-grained frequency segmentation scheme.

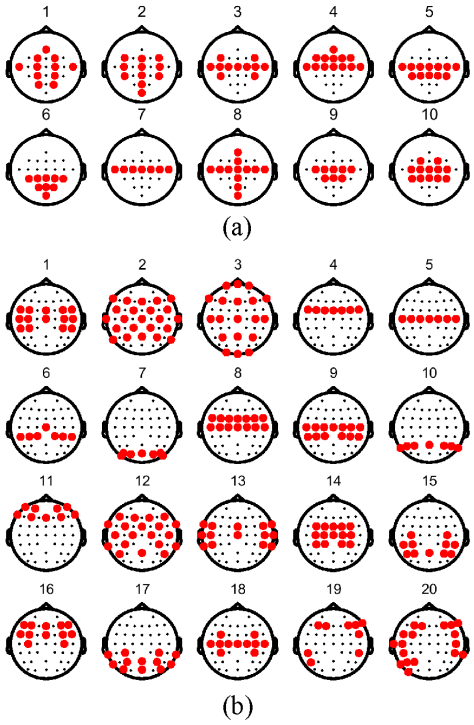

Fig. 5. The electrode selection strategy for two datasets. (a) Ten electrode groups sets for BCI IV IIa dataset. (b) Twenty electrode groups sets for SDU-MI dataset.

to simultaneously maximize the inter-class Kullback-Leibler (KL) divergence of covariance matrices and minimize the intra-class KL divergence.

The objective function $J(R)$ is as follows:

$$J(R) = (1-\lambda)\cdot D_K\left(R_d^T\Sigma_1 P_d \,\middle\|\, R_d^T\Sigma_2 R_d\right) - \lambda\cdot\frac{1}{2N}\sum_{c=1}^{C}\sum_{i=1}^{N} D_K\left(R_d^T\Sigma_c^i R_d \,\middle\|\, R_d^T\Sigma_c R_d\right), \quad (1)$$

where D_K represents the KL divergence, I_d is the identity matrix truncated to d dimensions, $R_d = I_d R$ is the truncated R matrix. Σ_c, Σ_c^i denote the average and trial-specific covariance matrices of class c, respectively, both pre-whitened by the matrix $P = (\Sigma_1+\Sigma_2)^{-\frac{1}{2}}$. Unlike the original CSP method which can be solved by a generalized eigenvalue decomposition and thus yields a closed-form solution, maximizing the divergence

Algorithm 1 The algorithm of divCSP

Require: X_1 (EEG data for class 1), X_2 (EEG data for class 2), dimensionality d of the extracted subspace
Ensure: Optimal spatial filter matrix W_{sf}
Calculate covariance matrices Σ_1 for X_1 and Σ_2 for X_2
Compute whitening matrix $P = (\Sigma_1 + \Sigma_2)^{-\frac{1}{2}}$
Initialize random orthogonal rotation matrix R_0
Apply whitening and rotation $\Sigma_1^* = R_0 P \Sigma_1 P^T R_0^T$, $\Sigma_2^* = R_0 P \Sigma_2 P^T R_0^T$
repeat
    Compute the gradient of $J(R)$ at the identity on the manifold
    Compute optimal step size U using line search
    Update the rotation matrix $R_{k+1} = U R_k$
    Rotate the matrices $\Sigma_1^* = U \Sigma_1^* U^T$, $\Sigma_2^* = U \Sigma_2^* U^T$
until maximum iterations are reached
Compute spatial filter matrix $V^T = I_d R_{k+1}^T P$
Compute eigenvectors E of $V^T \Sigma_1 V$
Apply PCA: $W_{sf} = VE$

between distributions and integrating additional regularization terms does not yield a direct closed-form solution. Consequently, gradient-based methods, such as manifold-constrained gradient descent, become indispensable. The comprehensive analysis of the gradient computation for the objective function and the detailed derivation process is thoroughly elaborated in [18]. To optimize $J(R)$, a subspace method is employed (see Algorithm 1), aiming to derive d spatial filters within a d-dimensional subspace, thereby refining the rotation matrix R. This is achieved through iterative updates, guided by the gradient of $J(R)$, within a line search strategy on the manifold of the orthogonal matrices [55]. The process iterates until convergence.

After optimization, the spatial filters W_{sf} are utilized to extract features, which are defined as the normalized logarithm of the variance of the spatially filtered EEG signals. Formally, the feature vector for the u-th Spatial-Frequency-Time (SFT) segment is calculated as:

$$F^{(u)} = -\ln\left(\frac{\mathrm{Var}(W_{sf}^{(u)} X^{(u)})}{\frac{1}{d}\sum_{i=1}^{d}\mathrm{Var}(W_{sf}^{(i)} X^{(u)})}\right), \tag{2}$$

where $\mathrm{Var}(\cdot)$ signifies the variance operation over the spatially filtered data. For our experimental setup, we empirically set the dimensionality $d = 4$ and the regularization coefficient $\lambda = 0.1$ to balance the computational efficiency and feature discriminability effectively.

D. Feature selection approach

In this study, we proposed a wrapper-based feature selection algorithm for motor imagery EEG feature optimization. The core objective of the algorithm is to construct an optimized ensemble model with multiple sub-classifiers in spatial, temporal, and frequency domains. The detailed procedure of the algorithm is described in Algorithm 2. Initially, the algorithm segments the EEG data across the spatial, frequency, and temporal domains, creating an array of SFTSs. Then, the discriminative MI EEG features of each segment are extracted using the divCSP algorithm for motor imagery EEG data. The extracted features are then classified using a linear SVM with a LOBO cross-validation scheme, providing a robust estimation of the classification performance for each segment. Subsequently, the algorithm ranks these segments based on their classification accuracies and iteratively merges the top-performing segments to enhance the feature space with a step size of D. This process involves concatenating the high-dimensional features from selected SFTSs and re-evaluating the performance of the resemble model using linear SVM classifiers. The final ensemble model comprises the top-K best-performing classifiers, each corresponding to a distinct set of SFTSs, ensuring a comprehensive representation of the EEG data. Considering the balance between accuracy and efficiency, the hyperparameters D and K are set to 5 and 3, respectively. Note that the proposed wrapper-based feature selection algorithm only outputs the information of the selected SFTSs and the trained model weights in training stage, and no training data were used in the testing phase.

E. Classification method

SVM as a classic machine learning classifier has been extensively applied in various motor imagery classification tasks [20]. It excels by optimizing the separation margin between distinct classes, effectively identifying the optimal decision boundary or hyperplane within the feature space. In this study, the efficient linear SVM is leveraged to classify linearly separable EEG features. The linear kernel is defined as:

$$K(x_i, x_j) = x_i^T x_j, \tag{3}$$

where K represents the kernel function, x_i and x_j are the i-th and j-th feature vectors. All hyperparameters employed match the default configuration of the 'fitclinear' function in Matlab.

In the testing phase of our proposed model, the classification procedure is executed on testing sets to evaluate the generalizability and robustness of the trained ensemble model. Each classifier in the optimized ensemble model, denoted as the top-K classifiers, is utilized to predict the class labels of the test data independently. Specifically, for each classifier, the corresponding critical SFTS information is identified, and the SFTSs are extracted accordingly from the test data. Then, the divCSP features are derived from the extracted SFTSs. Subsequently, these features are fed into the respective linear SVM classifier, yielding probability scores for specific classes. The final prediction is determined by averaging these probabilities across all classifiers, thereby integrating their collective intelligence to enhance the prediction accuracy. For a detailed description of the testing phase, refer to Algorithm 3.

IV. Results

This section illustrates the performance of the proposed MI-BCI approach on both BCI IV IIa dataset and SDU-MI dataset. Two metrics are used for model performance evaluation, namely, classification accuracy and information transfer

Algorithm 2 The Proposed SFTSs Selection Algorithm

Require:
EEG training dataset with N_s segments, each with predefined length and channels
Corresponding labels of the EEG dataset
N_c predefined channel groups
N_f predefined frequency band groups
N_t predefined time window groups
Ensure:
Optimized ensemble of K classifiers
Initialize $N = N_c \times N_f \times N_t$ SFTSs
Create struct array Seg of size N each with specific SFTS information
Initialize accuracy array Acc of size N
for $i = 1$ to N **do**
  Extract SFTS based on $Seg[i]$
  - Select channel group
  - Apply band-pass filter for frequency band
  - Truncate signal using time window
  Extract subset EEG dataset for this SFTS
  Extract d-dimensional divCSP features
  Classify features using Linear SVM with LOBO cross-validation
  Store the average accuracy in $Acc[i]$
end for
Sort Acc in descending order to obtain Acc^*, reorder Seg to get Seg^*
Initialize merged accuracy array $AccM$
Set counter $count = 0$
for $j = D$ to N step D **do**
  $count \leftarrow count + 1$
  Initialize feature set $FeaSet$
  for $k = 1$ to j **do**
    Extract SFTS corresponding to $Seg^*[k]$
    Extract divCSP features and add them to $FeaSet$
  end for
  Concatenate features in $FeaSet$ to form $FeaMerged$
  Classify $FeaMerged$ using Linear SVM with LOBO cross-validation
  Store the average accuracy in $AccMerged[count]$
  Save SFTS info and trained model
end for
Sort $AccMerged$ in descending order to get $AccMerged^*$
Select top-K accuracies with corresponding SFTSs info and models as final model

Algorithm 3 Testing Phase for the FGSFT-based MI-BCI

Require:
Test EEG dataset
Top-K optimized sub-classifiers with corresponding SFTS information
Ensure:
Predicted class for the test EEG dataset
Initialize an empty list $ClassProb$
for each classifier C_i in the top-K ensemble **do**
  Identify the SFTSs corresponding to C_i
  Extract and concatenate divCSP features from the test EEG data for the selected SFTSs
  Use C_i to classify the features and obtain the class probabilities
  Append the obtained class probabilities to $ClassProb$
end for
Compute the average of probabilities in $ClassProb$ to get $ProbAvg$
Determine the predicted class based on the highest value in $ProbAvg$
return Predicted class

rate (ITR). Detailly, the classification accuracy quantifies the proportion of correct predictions made by the model, while the ITR reflects both accuracy and speed, measuring the efficiency of a BCI system in bits per minute [56].

A. Results on BCI IV IIa dataset

Since this study focuses on the development of efficient BCI algorithms, we mainly utilized the initial 1.5 seconds of motor imagery EEG data for training and testing, reporting the mean and overall accuracies achieved. To comprehensively demonstrate the efficacy of the proposed method, we conducted seven ablation experiments, and the results are shown in Table I. These experiments individually segmented EEG data in the spatial domain, frequency domain, and time domain (denoted as experiments 'S', 'F', 'T' in Table I), as well as in spatial-frequency, frequency-time, and spatial-time domains (represented as experiments 'SF', 'FT', 'ST' in Table I). The last experiment evaluates the performance of our proposed SFT method (indicated by experiment 'SFT' in Table I). For each experimental configuration, the mean accuracies of LOBO cross-validation on both validation and test datasets are reported. We can see from Table I that, on the test dataset, the experiment of the proposed SFT method achieves the highest mean accuracy of 83.68% on the test set, which is significantly superior to that of the 'ST' experiment at 77.66% ($p = 0.0329$, pairwise t-test) and the 'SF' experiment at 76.45% ($p < 0.001$, pairwise t-test). Owing to the limited number of EEG channels in the BCI IV IIa database (22 channels), the impact of spatial segmentation is not particularly significant. Although the accuracy of the 'SFT' experiment (83.68%) exceeds that of the 'FT' (82.90%), the difference is not statistically significant ($p = 0.55$, pairwise t-test). Due to the same reason, the overall accuracy of the 'STF' method (85.55%) is slightly lower than that of the 'FT' method (85.59%) (not statistically significant, $p = 0.96$, pairwise t-test). Additionally, the results of the 'S', 'F', 'T' experiments suggest that for this dataset, the temporal segmentation of EEG data has the most significant effect on classification accuracy, followed by frequency domain segmentation, and finally spatial domain segmentation.

The length of used EEG data affects the ITR performance. A longer EEG data length can include more useful EEG features but will also reduce the ITR performance, resulting in a higher delay for the designed BCI system. Fig. 6 illustrates the classification accuracy under different lengths of used EEG data. It is obvious that the longer EEG data length

TABLE I
THE AVERAGE ACCURACIES (%) OF DIFFERENT SEGMENTATION SCHEMES ON VALIDATION AND TEST SETS OF BCI IV IIA.

Experiments	Evaluation Dataset	1	2	3	4	5	6	7	8	9	Mean	Overall
S	Validation set	72.22	61.81	87.5	67.36	58.33	68.75	65.28	84.72	88.19	72.69	71.25
	Test set	68.75	62.85	78.59	64.81	60.07	71.18	58.68	90.63	72.69	69.8	
F	Validation set	87.5	66.67	83.33	70.14	79.86	77.08	85.42	89.58	89.58	81.02	77.95
	Test set	82.99	64.47	77.66	76.04	76.04	67.36	67.82	86.46	75	74.87	
T	Validation set	78.47	59.72	94.44	77.78	71.53	72.92	84.72	95.14	99.31	81.56	79.23
	Test set	70.37	61.92	90.86	83.1	62.04	70.95	71.99	95.6	85.19	76.89	
SF	Validation set	82.64	70.83	85.42	67.36	78.47	69.44	81.94	90.28	89.58	79.55	78.00
	Test set	82.87	66.2	77.66	76.27	78.24	72.45	76.97	84.72	72.69	76.45	
FT	Validation set	88.89	74.31	93.06	81.94	89.58	79.86	93.75	95.14	97.92	88.27	85.59
	Test set	88.08	67.25	88.08	79.63	87.15	74.54	84.72	95.14	81.48	82.9	
ST	Validation set	77.08	64.58	94.44	80.56	71.53	74.31	81.94	91.67	97.92	81.56	79.61
	Test set	76.39	64	89.12	83.33	63.19	71.88	72.8	96.99	81.25	77.66	
SFT	Validation set	87.5	74.31	90.97	84.72	88.19	79.17	91.67	95.14	95.14	87.42	85.55
	Test set	87.96	67.01	88.08	88.77	81.83	76.97	85.88	95.95	80.67	83.68	

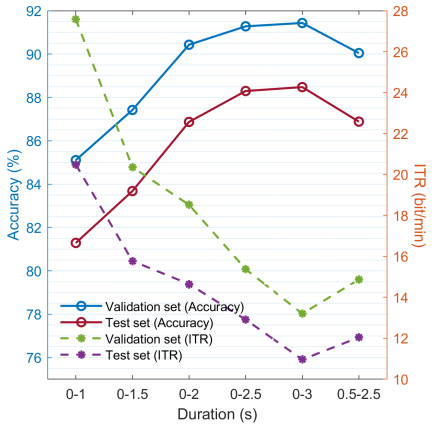

Fig. 6. The average accuracies and ITRs of different EEG lengths on validation and test sets of BCI IV IIa dataset.

corresponds to the higher mean accuracy of both validation and testing accuracy. Employing EEG data spanning 0-3 seconds yields mean testing and validation accuracies of 88.48% and 91.44%, respectively. We also selected EEG data spanning 0.5-2.5 seconds commonly used in previous studies, yielding a validation accuracy of 90.05% and a testing accuracy of 86.88%. However, when we consider the ITR performance, a shorter length of used EEG data is better. Utilizing EEG data from 0-1 seconds achieved an ITR of 27.59 bits/min and 20.49 bits/min on the validation and testing sets respectively, while data spanning 0-1.5 seconds resulted in ITRs of 20.36 bits/min and 15.77 bits/min correspondingly. A significant decrement in ITR is observed with EEG data durations extending to 0-3 seconds, culminating in ITRs of 13.20 bits/min and 10.96 bits/min for the validation and testing sets respectively.

B. Results on SDU-MI dataset

For the SDU-MI dataset, six ablation experiments are also conducted to evaluate the effectiveness of the segmentation strategies across various domains, and the results are demonstrated in Table II. All three motor imagery tasks, including left-hand and elbow (LHE) vs. right-hand and elbow (RHE), left-hand (LH) vs. right-hand (RH), left-elbow (LE) vs. right-elbow (RE), are considered. Consistent with the results on BCI IV IIa dataset, our SFT approach obtains the best classification performance for all three motor imagery tasks with a mean accuracy of 78.25%, 80.80%, and 76.95%, respectively. In the last column of Table II, the overall mean accuracy over all three motor imagery tasks is computed. Statistical analysis suggests that the overall accuracy of our 'SFT' method (78.67%) is significantly higher than that of the 'ST' method (77.00%, $p = 0.018$, pairwise t-test), the 'FT' method (70.88%, $p < 0.001$, pairwise t-test), and the 'SF' method (72.30%, $p < 0.001$, pairwise t-test). Aligning with findings from the BCI IV IIa dataset, the single-domain segmentation results on this dataset (i.e., results from 'S', 'F', and 'T') also highlight that the time segmentation strategy contributes most significantly to the enhancement of classification accuracy for all three motor imagery tasks. However, Differing from the results on the BCI IV IIa dataset where spatial segmentation showed no significant contribution to model accuracy improvement, the results from the SDU-MI dataset indicate that spatial segmentation substantially aids in accuracy enhancement, outperforming frequency segmentation. The overall accuracy with the 'S' method is significantly higher than that of the 'F' method (68.07% vs. 64.12%, $p = 0.0238$, pairwise t-test). Furthermore, the joint-domain segmentation results (i.e., results from 'SF', 'FT', and 'ST') suggested that the 'ST' method is significantly higher than the 'FT' method (77.00% vs. 70.88%, $p = 9.23 \times 10^{-6}$, pairwise t-test). These results demonstrate the effectiveness of spatial domain segmentation in enriching the feature set for enhanced model performance, particularly in EEG data with a large number of electrodes.

Fig. 7 depicts the mean accuracy and ITR metrics of models with different lengths of used EEG data on SDU-MI dataset. In this experiment, the LHE vs. RHE motor imagery task is considered, and 50% of the training set is used for model evaluation to speed up the experiment. It can be seen that the longest-used EEG data corresponds to the highest mean accuracy (76.05%) but the lowest mean ITR (6.17 bit/min), while the shortest-used EEG data corresponds to the lowest mean accuracy (71.75%) but the highest mean ITR (12.82 bit/min). Balanced performance is observed with EEG data segments spanning 0 to 1.5 seconds, which provides a balanced compromise between mean classification accuracy (74.90%) and ITR (11.13%).

TABLE II
THE AVERAGE ACCURACIES (%) OF DIFFERENT SEGMENTATION SCHEMES ON SDU-MI DATASET.

Experiments	Paradigm	1	2	3	4	5	6	7	8	9	10	Mean	Overall
S	LHE vs. RHE	94.5	77	64.5	73.5	63	84	64	63.5	58.5	53	69.55	68.07
	LH vs. RH	89.5	83	71.5	71	63.5	65	64	52	66	54.5	68	
	LE vs. RE	84	79.5	68	77.5	72.5	53.5	50.5	68.5	62.5	50	66.65	
F	LHE vs. RHE	94	87.5	57	57	62.5	86.5	49.5	56	59	54.5	66.35	64.12
	LH vs. RH	93.5	90	55	51.5	60.5	59.5	47	53.5	73	51.5	63.5	
	LE vs. RE	81	93	67.5	52.5	69	50.5	47.5	58	54	52	62.5	
T	LHE vs. RHE	93	89.5	67	82	69.5	85.5	49	61.5	67	68.5	73.25	71.75
	LH vs. RH	95	91	76	68	70	59.5	55	60.5	83	66.5	72.45	
	LE vs. RE	80	93.5	70.5	69.5	74	52.5	59.5	71.5	62	62.5	69.55	
SF	LHE vs. RHE	95	89	71.5	76	63	80	64.5	62.5	68	57	72.65	72.3
	LH vs. RH	95	92.5	69.5	68.5	72.5	62	63.5	57	84.5	66.5	73.15	
	LE vs. RE	86	92	70.5	77	77.5	55.5	62	69.5	64.5	56.5	71.1	
FT	LHE vs. RHE	96	90	79.5	76.5	66	80	55	53	64	59.5	71.95	70.88
	LH vs. RH	97.5	94.5	82	60.5	73.5	53	53	58	90	54.5	71.65	
	LE vs. RE	80.5	95.5	75	66.5	78	56	53	70	59	57	69.05	
ST	LHE vs. RHE	95	93.5	77.5	82	75	84	60	68.5	69.5	70.5	77.55	77
	LH vs. RH	95.5	92.5	79.5	74	76	60.5	66	68.5	83	72.5	76.8	
	LE vs. RE	87	94	76.5	82	83.5	53	70.5	81.5	65	73.5	76.65	
SFT	LHE vs. RHE	95	92	83	85.5	72	82	65.5	60.5	73.5	73.5	78.25	78.67
	LH vs. RH	96	93	87.5	78	83	61.5	72.5	69.5	90.5	76.5	80.8	
	LE vs. RE	87.5	95	81.5	78.5	83	53	71.5	81	66.5	72	76.95	

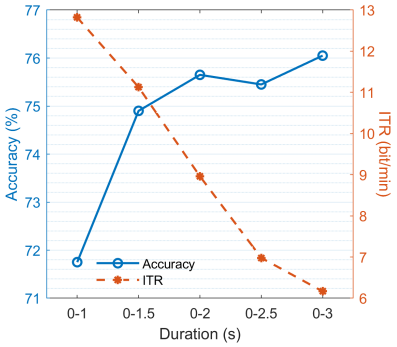

Fig. 7. The average accuracies and ITRs of different EEG lengths on SDU-MI database.

V. DISCUSSIONS

In this work, a fine-grained spatial-frequency-time segmentation framework is proposed for efficient motor imagery BCI. The proposed FGSFT method first decomposes the input EEG signals into multiple SFTSs and then the divCSP with intra-class regularization algorithm is applied to those SFTSs to obtain robust EEG features during motor imagery. A dedicated wrapper-based SFTS selection algorithm is designed to select the most significant features in time, frequency, and spatial domains. Results show that the proposed framework can effectively improve the overall performance of MI-BCI.

Although some previous works also proposed various segmentation strategies in time domain [12], [20], their selected window length was not narrow enough to capture fine-grained and multi-scale EEG features, challenging to achieve optimal accuracy. In this study, we propose to segment the raw EEG signals with a fine-grained and multi-scale strategy in time domains, where the shortest window length is defined as 200ms. Fig. 8 illustrates the performance of the proposed method under various temporal segmentation strategies. Both multi-scale and single-scale segmentation strategies were considered. The multi-scale segmentation strategy considered the shortest duration scale of a segment and the time segmentation strategies of all scales above this scale. For example, under the multi-scale segmentation strategy, the minimal duration of a 0.2s segment includes all segmentation strategies in time domain at the scales of 0.2 s, 0.4 s, 0.7 s, 1 s, 1.25 s, and 1.5 s. In contrast, single-scale segmentation only considers all time segmentation strategies at the shortest duration scale of a segment. It is shown in Fig. 8 that the single-scale of 0.2 s achieves the best performance in cross-session evaluation (The model is evaluated on the test set), and the more fine-grained scale of 0.1 s is too short to obtain useful EEG features. In within-session evaluation (The model is evaluated on the validation set), the single-scale of 0.7 s has the best performance. It can also be observed that the multi-scale segmentation strategy can enhance the model accuracy, especially in within-session evaluation. Although the multi-scale time segmentation strategy with a scale of 0.1 s achieves a slightly higher accuracy on both the validation set and the test set, its computational complexity is substantial and the performance improvement is not significant. Therefore, we chose the multi-scale time segmentation strategy of 0.2 s in this study. On the other hand, our previous study suggested that the multi-scale frequency segmentation method significantly improved the accuracy of MI-BCI [47]. Accordingly, we adopted the same multi-scale segmentation strategy in the frequency domain in this work. Besides, a novel segmentation strategy implemented by manual and automatic channel selection methods is proposed, effectively enriching EEG features. The results described in Table I and Table II suggest that the segmentation strategy in time domain plays a key role, and the spatial segmentation strategy is essential when the number of EEG electrodes is high. These results manifest the significance of our fine-grained and multi-scale segmentation strategies to MI-BCI research, providing a foundational framework for

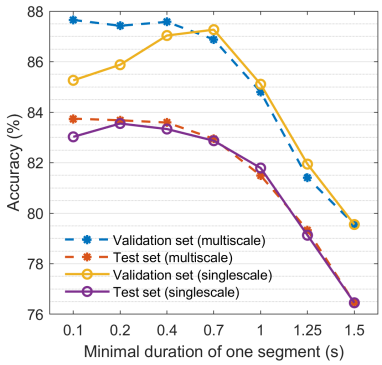

Fig. 8. The performance of the proposed method on the validation set and test set of BCI IV IIa database with different fine-grained segmentation strategies in the time domain. The dashed lines indicate the performance obtained by multi-scale temporal segmentation strategies, while the solid lines indicate the performance achieved by single-scale temporal segmentation strategies.

subsequent studies to enhance the accuracy and functionality of future MI-BCI systems.

Table III illustrates the comprehensive performance comparison between the proposed FGSFT method and other recently proposed and classic MI-BCI algorithms on BCI IV IIa dataset. It can be observed that the proposed FGSFT methods are superior to other MI-BCI methods in both accuracy and ITR metrics. Particularly, the average accuracy of FGSFT method with a selected window of 0-3 seconds (88.48%) is significantly higher than other comparison methods ($p < 0.05$, pairwise t-test), while the average ITR of FGSFT method with a selected window of 0-1.5 seconds (15.64 bit/min) is significantly higher than other compared methods ($p < 0.05$, pairwise t-test). Among these works, some also adopted segmentation strategies in time, frequency, or spatial domain. Zhang et al. [43] and Zheng et al. [44] applied the CSP-based time-frequency optimization with 5 time intervals and 17 frequency bands. Malan et al. [45] presented a CSP-based time-frequency optimization strategy with 6 time intervals and 3 frequency bands. Concurrently, Gaur et al. [46] proposed a sliding window CSP method where 9 overlapped time intervals were used for temporal optimization. Besides, Tiwari et al. [47] integrated channel selection algorithms with the RCSP method, while Ghorbanzadeh et al. [48] and Wang et al. [49] combined the channel selection algorithms with deep learning models, achieving spatial optimization of EEG. However, these methods cannot obtain the fine-grained spatial-frequency-time features as attained in this work, limiting their interpretability as well as classification accuracy and ITR. These comparative results demonstrated that the proposed FGSFT framework sets a new benchmark in binary motor imagery tasks, achieving state-of-the-art performance. Its ability to accurately segment and utilize the EEG signal components significantly enhances the overall effectiveness of MI-BCI systems, paving the way for more accurate and efficient BCI applications.

The visualization of the fine-grained MI-TFRM reflects the reaction intensity of a subject in time and frequency domains [20]. In the fine-grained MI-TFRM, each pixel is weighted by the mean LOBO cross-validation accuracy value of all its corresponding SFTSs. The fine-grained MI-TFRMs for subjects in the two datasets are depicted in Fig. 9. In our previous work, the MI-TFRMs were proposed, but the time-frequency resolution was relatively low. The comparison between the traditional MI-TFRMs presented in our previous work [20] and the fine-grained MI-TFRMs for subjects in the BCI IV IIa dataset is shown in Fig. 10. Clearly, with the fine-grained segmentation strategy applied in time and frequency domains, the resolution of the fine-grained MI-TFRMs is significantly enhanced. With these high-resolution fine-grained MI-TFRMs, more evident individual differences can be observed and analyzed. For example, the A-3 subject mainly activates the alpha band (around 10 Hz) during the motor imagery task, while A-5 exhibits spectral activations in the gamma band (around 30 Hz). Because a fine-grained spatial segmentation is conducted in this study, a more informative time-frequency topographical map can be visualized for each subject. From Fig. 11 (a), it can be found that the cerebral cortex is significantly activated in the alpha band during 1.5-2.0 seconds, and the occipital area contributes the most. The time-frequency topographical map shown in Fig. 11 (b) reveals that in 0.5-1.0 seconds, the theta band is activated, with a whole-brain activation pattern. From 1.5 to 3.0 seconds, the high-beta band and gamma band in the pre-frontal lobe contain the most discriminative EEG features of motor imagery tasks. These informative time-frequency topographic maps reveal fine-grained individual differences in the neural mechanisms underlying motor imagery tasks. It is suggested that the brain regions activated during motor imagery tasks vary across individuals in terms of timing and EEG frequency bands, indicating the necessity of applying segmentation in spatial domain. Neural activity associated with motor imagery is not confined to the central brain regions but may also manifest in the prefrontal and occipital lobes. This phenomenon might be linked to the participants' dominant hand preferences and the involvement of complex limb motor imagery [57], [58]. Therefore, the time-frequency topographical map can be used to effectively analyze the dynamic pattern of cognitive processes during motor imagery tasks, offering profound insights into the neural correlation undergoing motor imagery.

To better illustrate the feature selection procedure of the proposed method, Fig. 12 visualizes the details of the feature selection procedure in SDU-MI dataset. It can be seen that although a large amount of SFTSs are generated by the spatial-frequency-time framework, the proposed feature selection method effectively selects a small portion (3%-20%) of the SFTSs with discriminative EEG features, significantly reducing the feature dimensionality, and thus making the motor imagery BCI system more efficient.

VI. Conclusions

In this work, a novel MI-BCI algorithm based on a fine-grained spatial-frequency-time framework has been proposed. The algorithm was evaluated on the publicly available BCI IV IIa dataset and SDU-MI dataset collected by ourselves with multiple binary motor imagery paradigms, obtaining a state-of-the-art average ITR. An EEG spatial segmentation

TABLE III
PERFORMANCE COMPARISONS ON BCI IV IIA DATASET

No.	Author	Year	Method	Selected Window (s)	S1	S2	S3	S4	S5	S6	S7	S8	S9	Average Accuracy (%)	Average ITR (bit/min)
1	Lotte et al. [59]	2010	SRCSP	0.5-2.5	88.89	63.19	96.53	70.14	63.19	63.89	78.47	95.83	92.36	78.2	8.44
2	Arvaneh et al. [60]	2011	SSCSP	0.5-2.5	92.4	68	99.3	78.5	69.4	66	81.9	96.5	91	82.55	9.98
3	Alimardani et al. [61]	2017	GSW+MNN	-	85.11	76.27	83.72	81.85	74.87	78.13	78.6	74.87	82.32	79.53	-
4	Gaur et al. [62]	2018	SS-MEMDBF	0.5-2.5	91.49	60.56	94.16	76.72	58.52	68.52	78.57	97.01	93.85	79.93	8.94
5	Zhang et al. [43]	2018	TSGSP	≥ 2	87	64.7	93.8	74.3	90.4	63.9	91.4	95.8	81.3	82.5	-
6	Singh et al. [63]	2019	RMDRM	0.5-2.5	91.61	63.28	97.2	72.91	64.08	69.71	81.25	96.52	92.3	80.98	9.26
7	Yu et al. [64]	2020	p-LTCS	0.5-2.5	92.36	65.28	97.22	71.53	78.47	71.53	85.42	95.14	92.36	83.26	10
8	Malan et al. [45]	2021	DTCWT-CSP	-0.5-4	85.6	66.7	97.2	77.1	82.6	69.4	79.1	95.4	86.1	82.13	5.64
9	Gaur et al. [46]	2021	SW-LCR	0-3	86.81	64.58	95.83	67.36	68.06	67.36	80.56	97.22	92.36	80.02	7.2
10	Zheng et al. [44]	2022	PSPD	0-3	91.43	66.07	94.64	75.36	78.57	66.79	97.14	95.36	87.86	83.69	8.64
11	Tiwari et al. [47]	2022	RCSPA+SVM	-	83.79	74.18	73.92	94.01	69.32	84.71	89.36	79.11	82.18	81.71	-
12	Ghorbanzadeh et al. [48]	2023	DGAFF	-	91.44	78.77	93.97	78	78.37	73.78	88.51	85.86	87.68	84.04	-
13	Srimadumathi et al. [65]	2024	CMW+CNN	0-4	84.15	67.95	97.5	80.36	84.29	67.86	81.5	80.26	96.00	82.20	6.48
14	Wang et al. [49]	2024	MI-BMInet	0-3	86.98	72.65	94.95	76.66	93.84	81.11	91.17	98.27	93.26	87.65	10.17
-	Liu et al.	-	FGSFT	0-1.5	87.96	67.01	88.08	88.77	81.83	76.97	85.88	95.95	80.67	83.68	15.64
-	Liu et al.	-	FGSFT	0.5-2.5	92.94	67.25	96.64	81.48	87.04	75.58	92.48	98.61	89.93	86.88	11.99
-	Liu et al.	-	FGSFT	0-3	93.75	66.55	97.22	89.81	89	76.97	95.02	98.26	89.7	88.48	10.92

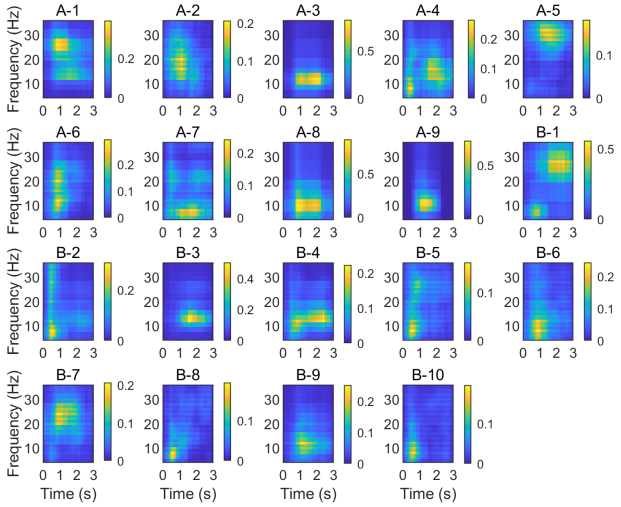

Fig. 9. The fine-grained weighted MI-TFRMs for two databases.

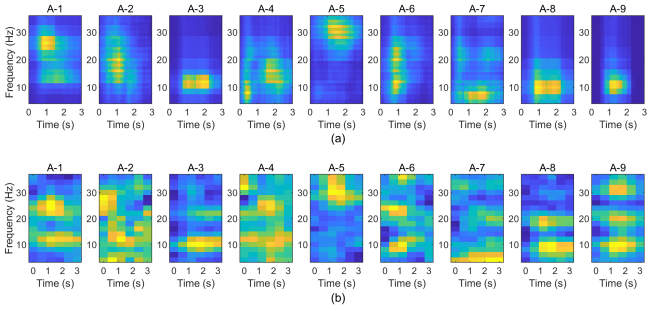

Fig. 10. The resolution comparison between fine-grained MI-TFRMs and traditional MI-TFRMs in BCI IV IIa dataset. (a) The fine-grained MI-TFRMs generated with the proposed FGSFT method. (b) The traditional MI-TFRMs generated using multiscale time-frequency method [20].

technique combined with manual and automatic channel selection processes was proposed, while fine-grained and multi-scale EEG segmentation strategies in time and frequency domains were presented. Meanwhile, the divCSP with intra-class regularization was employed to extract robust EEG features. The iterative refinement and selection of top-performing SFTSs were adopted to construct a robust ensemble of SVM

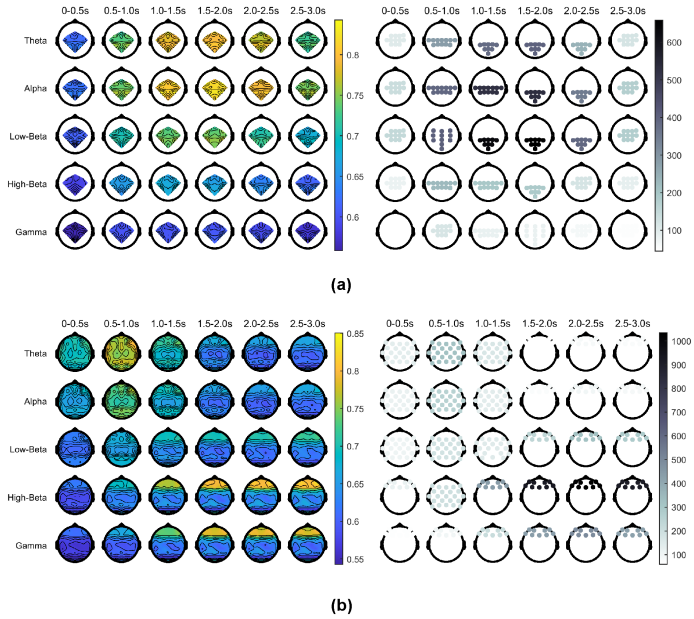

Fig. 11. The time-frequency topographical maps. (a) The time-frequency topographical map (left) and the electrode group significance map (right) for the 8-th subject in BCI IV IIa dataset. (b) The time-frequency topographical map (left) and the electrode group significance map (right) for the first subject in SDU-MI dataset.

classifiers to ensure the effectiveness of our model. Additionally, the high-resolution fine-grained MI-TFRMs and time-frequency topographical maps were visualized according to the proposed feature ranking algorithms, significantly enhancing the interpretability of the model by depicting the dynamic cognitive processes during motor imagery tasks. This work not only offers a more accurate and efficient BCI framework but also can contribute to elucidating subject-specific neural mechanisms related to motor imagery. Through the proposed fine-grained spatial-frequency-time feature selection approach, the individual differences in the brain activities during the motor imagery task can be well visualized, facilitating the design of personalized MI-BCI systems. Future work will focus on enhancing cross-subject generalization by transfer learning fused with time-frequency topographical maps and extending the framework to multi-class paradigms and real-

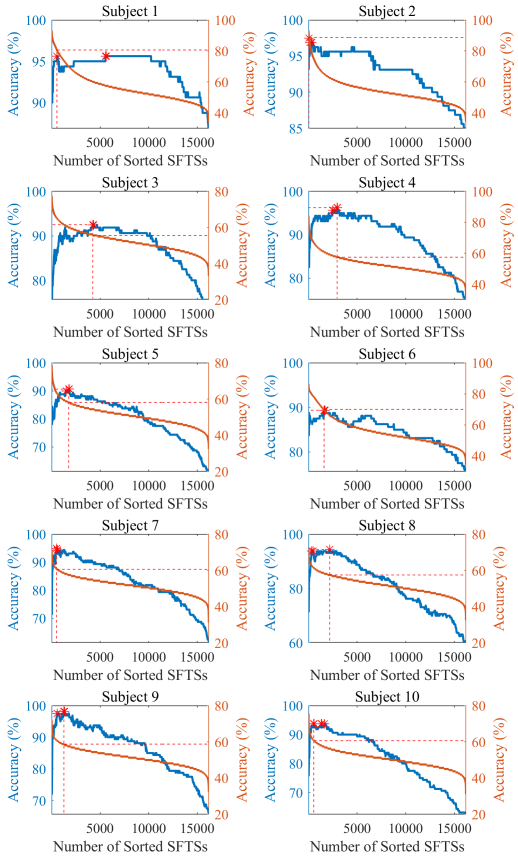

Fig. 12. The feature selection procedure for all subjects in SDU-MI dataset. The red line indicates the accuracy of sorted single SFTSs, while the blue line indicates the accuracy of merged SFTSs. The stars point out the selected top-3 accuracy, and the dashed line indicates the peak accuracy of merged SFTSs.

world applications, paving the way for more effective and adaptable BCI systems.

REFERENCES

[1] C. Guger, H. Ramoser, and G. Pfurtscheller, "Real-time eeg analysis with subject-specific spatial patterns for a brain-computer interface (bci)," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 447–456, 2000.

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

[3] G. Pfurtscheller, "Eeg event-related desynchronization (erd) and synchronization (ers)," *Electroencephalography and Clinical Neurophysiology*, vol. 103, p. 26, 1997.

[4] M. A. Khan, R. Das, H. K. Iversen, and S. Puthusserypady, "Review on motor imagery based bci systems for upper limb post-stroke neurorehabilitation: From designing to application," *Computers in Biology and Medicine*, vol. 123, p. 103843, 2020.

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

[6] Z. Khademi, F. Ebrahimi, and H. M. Kordy, "A review of critical challenges in mi-bci: From conventional to deep learning methods," *Journal of Neuroscience Methods*, vol. 383, p. 109736, 2023.

[7] W. Zhao, X. Jiang, B. Zhang, S. Xiao, and S. Weng, "Ctnet: a convolutional transformer network for eeg-based motor imagery classification," *Scientific Reports*, vol. 14, no. 1, p. 20237, 2024.

[8] R. Zhang, G. Liu, Y. Wen, and W. Zhou, "Self-attention-based convolutional neural network and time-frequency common spatial pattern for enhanced motor imagery classification," *JOURNAL OF NEUROSCIENCE METHODS*, vol. 398, OCT 1 2023.

[9] B. Xua and G. Yang, "Interpretability research of deep learning: A literature survey," *Information Fusion*, p. 102721, 2024.

[10] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band common spatial pattern (sbcsp) for brain-computer interface," in *2007 3rd International IEEE/EMBS Conference on Neural Engineering*, 2007, pp. 204–207.

[11] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (fbcsp) in brain-computer interface," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008, pp. 2390–2397.

[12] Y. Y. Miao, J. Jin, I. Daly, C. L. Zuo, X. Y. Wang, A. Cichocki, and T. P. Jung, "Learning common time-frequency-spatial patterns for motor imagery classification," *Ieee Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 699–707, 2021.

[13] J. Wang, Z. R. Feng, X. D. Ren, N. Lu, J. Luo, and L. Sun, "Feature subset and time segment selection for the classification of eeg data based motor imagery," *Biomedical Signal Processing and Control*, vol. 61, 2020.

[14] Z. J. Koles, M. S. Lazar, and S. Z. Zhou, "Spatial patterns underlying population differences in the background eeg," *Brain Topography*, vol. 2, no. 4, p. 275 – 284, 1990.

[15] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve bci designs: Unified theory and new algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2011.

[16] J. Park and W. Chung, "Common spatial patterns based on generalized norms," in *2013 International Winter Workshop on Brain-Computer Interface (BCI)*. IEEE, 2013, Conference Proceedings, pp. 39–42.

[17] Q. Cai, W. Gong, Y. Deng, and H. Wang, "Single-trial eeg classification via common spatial patterns with mixed lp-and lq-norms," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–13, 2021.

[18] W. Samek, M. Kawanabe, and K.-R. Müller, "Divergence-based framework for common spatial patterns algorithms," *IEEE Reviews in Biomedical Engineering*, vol. 7, pp. 50–72, 2013.

[19] W. Samek, D. Blythe, K.-R. Müller, and M. Kawanabe, "Robust spatial filtering with beta divergence," in *Advances in Neural Information Processing Systems*, C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, Eds., vol. 26. Curran Associates, Inc., 2013.

[20] G. Liu, L. Tian, and W. Zhou, "Multiscale time-frequency method for multiclass motor imagery brain computer interface," *Computers in Biology and Medicine*, vol. 143, p. 105299, 2022.

[21] G. Liu, J. H. Hsiao, W. Zhou, and L. Tian, "Martmi-bci: A matlab-based real-time motor imagery brain-computer interface platform," *SoftwareX*, vol. 22, p. 101371, 2023.

[22] G. Liu, L. Tian, and W. Zhou, "Multiscale time-frequency method for multiclass motor imagery brain computer interface," *Computers in Biology and Medicine*, vol. 143, p. 105299, 2022.

[23] F. Yger, M. Berar, and F. Lotte, "Riemannian approaches in brain-computer interfaces: a review," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, pp. 1753–1762, 2016.

[24] T. Qu, J. Jin, R. Xu, X. Wang, and A. Cichocki, "Riemannian distance based channel selection and feature extraction combining discriminative time-frequency bands and riemannian tangent space for mi-bcis," *Journal of Neural Engineering*, vol. 19, no. 5, p. 056025, sep 2022. [Online]. Available: https://dx.doi.org/10.1088/1741-2552/ac9338

[25] F. Yger, F. Lotte, and M. Sugiyama, "Averaging covariance matrices for eeg signal classification based on the csp: an empirical study," in *2015 23rd European signal processing conference (EUSIPCO)*. IEEE, 2015, Conference Proceedings, pp. 2721–2725.

[26] H. Altaheri, G. Muhammad, M. Alsulaiman, S. U. Amin, G. A. Altuwaijri, W. Abdul, M. A. Bencherif, and M. Faisal, "Deep learning techniques for classification of electroencephalogram (eeg) motor imagery (mi) signals: a review," *Neural Computing and Applications*, vol. 35, no. 20, pp. 14 681–14 722, 2023.

[27] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces," *arXiv e-prints*, p. arXiv:1611.08024, Nov. 2016.

[28] F. Li, F. He, F. Wang, D. Zhang, Y. Xia, and X. Li, "A novel simplified convolutional neural network classification algorithm of motor imagery eeg signals based on deep learning," *Applied Sciences*, vol. 10, no. 5, 2020.

[29] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Mekhtiche, and M. S. Hossain, "Deep learning for eeg motor imagery classification based on multi-layer cnns feature fusion," *FUTURE GENERATION COMPUTER SYSTEMS-THE INTERNATIONAL JOURNAL OF ESCIENCE*, vol. 101, pp. 542–554, DEC 2019.

[30] M.-A. Li, J.-F. Han, and L.-J. Duan, "A novel mi-eeg imaging with the location information of electrodes," *IEEE Access*, vol. 8, pp. 3197–3211, 2020.

[31] P. Wang, A. Jiang, X. Liu, J. Shang, and L. Zhang, "Lstm-based eeg classification in motor imagery tasks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 11, pp. 2086–2095, 2018.

[32] S. Kumar, A. Sharma, and T. Tsunoda, "Brain wave classification using long short-term memory network based optical predictor," *Scientific Reports*, vol. 9, no. 1, p. 9153, 2019.

[33] H. Altaheri, G. Muhammad, and M. Alsulaiman, "Physics-informed attention temporal convolutional network for eeg-based motor imagery classification," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 2249–2258, 2023.

[34] X. Shi, B. Li, W. Wang, Y. Qin, H. Wang, and X. Wang, "Eeg-vttcnet: A loss joint training model based on the vision transformer and the temporal convolution network for eeg-based motor imagery classification," *Neuroscience*, vol. 556, pp. 42–51, 2024.

[35] K. Liu, T. Yang, Z. Yu, W. Yi, H. Yu, G. Wang, and W. Wu, "Msvtnet: Multi-scale vision transformer neural network for eeg-based motor imagery decoding," *IEEE Journal of Biomedical and Health Informatics*, 2024.

[36] J. Jiang, C. H. Wang, J. H. Wu, W. Qin, M. P. Xu, and E. W. Yin, "Temporal combination pattern optimization based on feature selection method for motor imagery bcis," *Frontiers in Human Neuroscience*, vol. 14, 2020.

[37] Y. Pei, Z. G. Luo, H. Y. Zhao, D. K. Xu, W. G. Li, Y. Yan, H. J. Yan, L. Xie, M. P. Xu, and E. Yin, "A tensor-based frequency features combination method for brain-computer interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 465–475, 2022.

[38] L. L. Chen, P. F. Chen, S. K. Zhao, Z. G. Luo, W. Chen, Y. Pei, H. Y. Zhao, J. Jiang, M. P. Xu, Y. Yan, and E. W. Yin, "Adaptive asynchronous control system of robotic arm based on augmented reality-assisted brain-computer interface," *Journal of Neural Engineering*, vol. 18, no. 6, 2021.

[39] J. Jin, T. Qu, R. Xu, X. Wang, and A. 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*, vol. 31, pp. 58–67, 2022.

[40] X. Ding, L. Yang, and C. Li, "Study of mi-bci classification method based on the riemannian transform of personalized eeg spatiotemporal features," *Mathematical Biosciences and Engineering*, vol. 20, no. 7, pp. 12 454–12 471, 2023.

[41] A. M. Roy, "An efficient multi-scale cnn model with intrinsic feature integration for motor imagery eeg subject classification in brain-machine interfaces," *Biomedical Signal Processing and Control*, vol. 74, p. 103496, 2022.

[42] Y. Pei, Z. G. Luo, Y. Yan, H. J. Yan, J. Jiang, W. G. Li, L. Xie, and E. R. Yin, "Data augmentation: Using channel-level recombination to improve classification performance for motor imagery eeg," *Frontiers in Human Neuroscience*, vol. 15, 2021.

[43] Y. Zhang, C. S. Nam, G. X. Zhou, J. Jin, X. Y. Wang, and A. Cichocki, "Temporally constrained sparse group spatial patterns for motor imagery bci," *Ieee Transactions on Cybernetics*, vol. 49, no. 9, pp. 3322–3332, 2019.

[44] L. S. Zheng, Y. Ma, P. C. Lian, Y. Xiao, Z. K. Yi, Q. Z. Song, W. Feng, and X. Y. Wu, "A power spectrum pattern difference-based time-frequency sub-band selection method for mi-eeg classification," *Ieee Sensors Journal*, vol. 22, no. 12, pp. 11 928–11 939, 2022.

[45] N. S. Malan and S. Sharma, "Time window and frequency band optimization using regularized neighbourhood component analysis for multi-view motor imagery eeg classification," *Biomedical Signal Processing and Control*, vol. 67, 2021.

[46] P. Gaur, H. Gupta, A. Chowdhury, K. McCreadie, R. B. Pachori, and H. Wang, "A sliding window common spatial pattern for enhancing motor imagery classification in eeg-bci," *Ieee Transactions on Instrumentation and Measurement*, vol. 70, 2021.

[47] A. Tiwari and A. Chaturvedi, "Automatic eeg channel selection for multiclass brain-computer interface classification using multiobjective improved firefly algorithm," *Multimedia Tools and Applications*, vol. 82, no. 4, pp. 5405–5433, 2023.

[48] G. Ghorbanzadeh, Z. Nabizadeh, N. Karimi, P. Khadivi, A. Emami, and S. Samavi, "Dgaff: Deep genetic algorithm fitness formation for eeg bio-signal channel selection," *Biomedical Signal Processing and Control*, vol. 79, 2023.

[49] X. Y. Wang, M. Hersche, M. Magno, and L. Benini, "Mi-bminet: An efficient convolutional neural network for motor imagery brainmachine interfaces with eeg channel selection," *IEEE Sensors Journal*, vol. 24, no. 6, pp. 8835–8847, 2024.

[50] A. Jiang, J. Shang, X. Liu, Y. Tang, H. K. Kwan, and Y. Zhu, "Efficient csp algorithm with spatio-temporal filtering for motor imagery classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 4, pp. 1006–1016, 2020.

[51] T.-j. Luo, "Selective multi–view time–frequency decomposed spatial feature matrix for motor imagery eeg classification," *Expert Systems with Applications*, vol. 247, p. 123239, 2024.

[52] K. K. Ang, Z. Y. Chin, C. C. Wang, C. T. Guan, and H. H. Zhang, "Filter bank common spatial pattern algorithm on bci competition iv datasets 2a and 2b," *Frontiers in Neuroscience*, vol. 6, 2012.

[53] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "Bci competition 2008–graz data set a," *Institute for knowledge discovery (laboratory of brain-computer interfaces), Graz University of Technology*, vol. 16, pp. 1–6, 2008.

[54] J. Jin, Y. Y. Miao, I. Daly, C. L. Zuo, D. W. Hu, and A. Cichocki, "Correlation-based channel selection and regularized feature optimization for mi-based bci," *Neural Networks*, vol. 118, pp. 262–270, 2019.

[55] P. von Bünau, "Stationary subspace analysis: towards understanding non-stationary data," Thesis, 2012.

[56] D. J. McFarland, W. A. Sarnacki, and W. Jr, "Brain-computer interface (bci) operation: optimizing information transfer rates," *Biological Psychology*, vol. 63, no. 3, pp. 237–251, 2003.

[57] W. Yi, S. Qiu, H. Qi, L. Zhang, B. Wan, and D. Ming, "Eeg feature comparison and classification of simple and compound limb motor imagery," *Journal of neuroengineering and rehabilitation*, vol. 10, pp. 1–12, 2013.

[58] D. Zapała, P. Iwanowicz, P. Francuz, and P. Augustynowicz, "Handedness effects on motor imagery during kinesthetic and visual-motor conditions," *Scientific reports*, vol. 11, no. 1, p. 13112, 2021.

[59] F. Lotte and C. Guan, "Spatially regularized common spatial patterns for eeg classification," in *2010 20th International Conference on Pattern Recognition*. IEEE, 2010, Conference Proceedings, pp. 3712–3715.

[60] M. Arvaneh, C. T. Guan, K. K. Ang, and H. C. Quek, "Spatially sparsed common spatial pattern to improve bci performance," *2011 Ieee International Conference on Acoustics, Speech, and Signal Processing*, pp. 2412–2415, 2011.

[61] F. Alimardani, R. Boostani, and B. Blankertz, "Weighted spatial based geometric scheme as an efficient algorithm for analyzing single-trial eegs to improve cue-based bci classification," *Neural Networks*, vol. 92, pp. 69–76, 2017.

[62] P. Gaur, R. B. Pachori, H. Wang, and G. Prasad, "A multi-class eeg-based bci classification using multivariate empirical mode decomposition based filtering and riemannian geometry," *Expert Systems with Applications*, vol. 95, pp. 201–211, 2018.

[63] A. Singh, S. Lal, and H. W. Guesgen, "Small sample motor imagery classification using regularized riemannian features," *Ieee Access*, vol. 7, pp. 46 858–46 869, 2019.

[64] Z. H. Yu, T. Ma, N. Fang, H. X. Wang, Z. L. Li, and H. Fan, "Local temporal common spatial patterns modulated with phase locking value," *Biomedical Signal Processing and Control*, vol. 59, 2020.

[65] V. Srimadumathi and M. R. Reddy, "Classification of motor imagery eeg signals using high resolution time-frequency representations and convolutional neural network," *Biomedical Physics and Engineering Express*, vol. 10, no. 3, 2024.