A Tensor-Based Frequency Features Combination Method for Brain–Computer Interfaces

Yu Pei, Zhiguo Luo, Hongyu Zhao, Dengke Xu, Weiguo Li, Ye Yan, Huijiong Yan, Liang Xie, Minpeng Xu, *Member, IEEE*, and Erwei Yin

***Abstract*— With the development of the brain-computer interface (BCI) community, motor imagery-based BCI system using electroencephalogram (EEG) has attracted increasing attention because of its portability and low cost. Concerning the multi-channel EEG, the frequency component is one of the most critical features. However, insufficient extraction hinders the development and application of MI-BCIs. To deeply mine the frequency information, we proposed a method called tensor-based frequency feature combination (TFFC). It combined tensor-to-vector projection (TVP), fast fourier transform (FFT), common spatial pattern (CSP) and feature fusion to construct a new feature set. With two datasets, we used different classifiers to compare TFFC with the state-of-the-art feature extraction methods. The experimental results showed that our proposed TFFC could robustly improve the classification accuracy of about 5% ($p < 0.01$). Moreover, visualization analysis implied that the TFFC was a generalization of CSP and Filter Bank CSP (FBCSP). Also, a complementarity between weighted narrowband features (wNBFs) and broadband features (BBFs) was observed from the averaged fusion ratio. This article certificates the importance of frequency information in the MI-BCI system and provides a new direction for designing a feature set of MI-EEG.**

***Index Terms*— Brain–computer interface, electroencephalogram, motor imagery, common spatial pattern, tensor-to-vector projection, fast fourier transformation.**

I. Introduction

THE brain-computer interface (BCI) is a communication control system, which is directly established between the brain and external electronic devices, using signals generated during brain activity [1]. The BCIs have shown great potentials applying in various fields such as communication, synchronous control, asynchronous control [2], and rehabilitation [3]. Electroencephalogram (EEG) is one of the most common signals used for building a BCI system because of its cost-effectiveness, noninvasive implementation, and portability. Throughout numerous BCI studies, growing attention has been dedicated to analyzing EEG of motor imagery (MI) [4]–[9]. This interest is due to MI-BCI's ability to allow both healthy and disabled people to control electronic devices using self-regulate brain signals without an external stimulus. No external stimulus, which is a distinctive feature of the MI-BCI, brings it excellent application potentials.

The MI-BCI system's framework is based on the fact that the brain's activity in a specific area could be changed when the patients (or subjects) imagine moving a certain part of their bodies [10]. The spatial distribution and frequency-band's energy are two essential aspects to characterize MI-EEG [11]. Around the design of the spatial filter, a large amount of works have been produced to decode the MI-EEG signals [12]–[14].

Among these works, the most important one was the common spatial pattern (CSP) algorithm, which was the root node of other algorithms, and a large amount of works were derived from it. CSP-rank was a filtering channel selection algorithm [15], which was used to remove blackundant channels while improving BCI systems' performance. It sorted the importance of channels according to the coefficients of spatial filters, and the optimal channel set was determined by the highest cross-validation accuracy. For the same channel selection problem, Correlation based channel selection CSP (CCS-CSP) used a correlation-based method to select the channels that contained more correlated information [16]. To include a spatial priority in the learning process, Spatially regularized CSP (SRCSP) added a regularization term to spatially penalize non-smooth filters [17]. Because the

Manuscript received June 22, 2021; revised October 14, 2021; accepted November 1, 2021. Date of publication November 4, 2021; date of current version March 8, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62076250, Grant 61703407, and Grant 61901505; and in part by the National Innovation Platform Open Fund under Grant 2019YJ192. *(Yu Pei and Zhiguo Luo contributed equally to this work.) (Corresponding author: Erwei Yin.)*

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Tianjin University under Approval No. TJUE-2021-138.

Yu Pei is with the School of Software, Beihang University, Beijing 100191, China, and also with the Tianjin Artificial Intelligence Innovation Center (TAIIC), Tianjin 300450, China.

Zhiguo Luo, Ye Yan, Huijiong Yan, Liang Xie, and Erwei Yin are with the Defense Innovation Institute, Academy of Military Sciences (AMS), Beijing 100071, China, and also with the Tianjin Artificial Intelligence Innovation Center (TAIIC), Tianjin 300450, China (e-mail: yinerwei1985@gmail.com).

Hongyu Zhao is with the Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China.

Dengke Xu is with the Institute of Communication Signals, China Academy of Railway Sciences, Beijing 100081, China.

Weiguo Li is with the School of Software, Beihang University, Beijing 100191, China.

Minpeng Xu is with the Laboratory of Neural Engineering & Rehabilitation, Department of Biomedical Engineering, College of Precision Instruments and Optoelectronics Engineering, Tianjin University, Tianjin 300072, China, and also with the Tianjin International Joint Research Center for Neural Engineering, Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin 300072, China.

Digital Object Identifier 10.1109/TNSRE.2021.3125386

L_1-norm has lower sensitivity to the outlier, Difference and ratio of average L1-Norm CSP (DRL1-CSP) [13] defined a new L_1-norm-based feature ranking function to select the more efficient features. Also, DRL1-CSP further fused the L_1-norm-based feature ranking and Fisher scores ranking by Dempster–Shafer Theory [18]–[20] to built a more robust feature selection mechanism.

We found a commonality after sorting out these works: almost all of these methods used 8-30 Hz band-pass filters, which was a relatively broad frequency band. Therefore, in this article, the features extracted by these methods were called broadband features (BBFs).

Moreover, the Filter Bank CSP (FBCSP) architecture created another trend [21], [22]. This architecture used multiple narrowband bandpass filters rather than the classic single bandpass filter and successfully extracted narrowband features (NBFs) related to MI. Many works have been derived [23]–[25], based on the FBCSP framework. This architecture also have achieved good results on public datasets. These methods calculated features in each narrow frequency band and then selected k features with the most discriminative ability. Note that the selection operation may be an unsafe operation. Because of that, the features on the narrow frequency bands may lost the possibility of linearly weighting each other.

Through the analyses of previous works, we found two significant shortcomings in these works. First, in the CSP framework, many works only focused on 8-30 Hz broadband information but ignoblack narrowband information [12]–[14]. Second, even though some subsequent researches focused on narrowband information based on the filter bank strategy, some potentially useful information was lost due to their crude feature selection [22], [26]–[28]. Almost all the FBCSP-based methods would inevitably use feature selection methods, which could cause narrowband information to be selected or discarded. In other words, the fused features among these narrow bands were ignoblack. In addition, increasing numbers of studies have provided a clue that there may be complementarity between broadband and narrowband information [6], [29]. Combining them may enable the classifier to obtain better classification capabilities. If there is a feature extraction method that can fuse the two parts of information without losing the information fusion potential among the narrow frequency bands, the performance of the brain-computer interface system could be further improved.

To solve these shortcomings discussed above, we firstly constructed a spatial-spectrum tensor based on FFT and used a tensor analysis method, named Uncorrelated Multilinear Discriminant Analysis (UMLDA) [30], to extract more discriminative weighted NBFs (wNBFs) hidden in narrow frequency bands. Secondly, we fused these features with BBFs to obtain a more robust feature set named TFFC. Classifiers trained on TFFC have better classification performance. The novelty and contribution of this article can be summarized as follows:

1) A spatial-spectrum tensor was constructed by FFT.
2) The wNBFs were extracted by the tensor analysis method named UMLDA.
3) The BBFs and wNBFs were fused to obtain a more robust feature set named TFFC.
4) Implying the complementarity of BBFs and wNBFs in MI tasks.

The remainder of this article is organized as follows. Section II shows the methods used in our study. In section III, we briefly introduce our dataset. The baseline methods and the experimental results are discussed in section IV. In section V, the reasons for the success of TFFC and the shortcomings of TFFC are discussed. Finally, section VI concludes this article.

II. Method

This paper follows the notation conventions in multilinear algebra, pattern recognition, and adaptive learning literature. Vectors are denoted by lowercase boldface letters, e.g., $\mathbf{x}$; matrices by uppercase boldface, e.g., $\mathbf{X}$; and tensors by calligraphic letters, e.g., $\mathcal{X}$. Their elements are denoted with indices in parentheses. Indices are denoted by lowercase letters, spanning the range from 1 to the uppercase letter of the index, e.g., $p = 1, 2, \ldots, P$. In addressing part of a vector/matrix/tensor, ":" denotes the full range of the respective index and $n_1{:}n_2$ denotes indices ranging from n_1 to n_2. In this paper, only real-valued data are consideblack. The overall architecture of our method is shown in Fig. 1, and more details are as below:

A. Tensor Algebra

An Nth-order tensor is denoted as $\mathcal{A} \in \mathbb{R}^{I_1 \times I_1 \times \ldots \times I_N}$ [31]. It is addressed by N indices $i_n, n = 1, 2, \ldots, N$, and each i_n addresses the n-mode of $\mathcal{A}$. The n-mode product of a tensor $\mathcal{A}$ by a matrix $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n \mathbf{U}$, is a tensor with entries

$$(\mathcal{A} \times_n \mathbf{U})(i_1, \ldots, i_{n-1}, j_n, i_{n+1}, \ldots i_N) = \sum\nolimits_{i_n} \mathcal{A}(i_1, \ldots, i_N) \cdot \mathbf{U}(j_n, i_n). \quad (1)$$

The scalar product of two tensors, $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I_1 \times I_1 \times \ldots \times I_N}$ is defined as

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum\nolimits_{i_1} \cdots \sum\nolimits_{i_N} \mathcal{A}(i_1, \ldots, i_N) \cdot \mathcal{B}(i_1, \ldots, i_N) \quad (2)$$

A rank-one tensor $\mathcal{A}$ equals to the outer product of N vectors: $\mathcal{A} = \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \cdots \circ \mathbf{u}^{(N)}$, which means that $\mathcal{A}(i_1, \ldots, i_N) = \mathbf{u}^{(1)}(i_1) \cdot \mathbf{u}^{(2)}(i_2) \cdots \mathbf{u}^{(N)}(i_N)$ for all values of indices.

B. Spatial-Spectrum Tensor Generation

The original segmented EEG data is M-order tensor $\mathcal{X}_{eeg} \in \mathbb{R}^{I_T \times I_C}$, where the number of modes M is 2, I_C is the number of channels, and I_T is the number of sample points. A transformation from spatial-temporal formulation to spatial-spectrum formulation is desperately needed to conduct a spatial-spectrum tensor analysis. We use Fast Fourier Transform (FFT) to complete the purpose.

$$\mathcal{X}'_{eeg} = fft(\mathcal{X}_{eeg}) \quad (3)$$

where $fft(\bullet)$ calculates the one-sided FFT along with the first mode with 1024 points. Note that the frequency components in $\mathcal{X}'$ are under f_s according to Nyquist sampling theory, where $f_s = \frac{F_s}{2}$ and the F_s is the sampling frequency. The motor imagery-related frequency components are not high. Since the

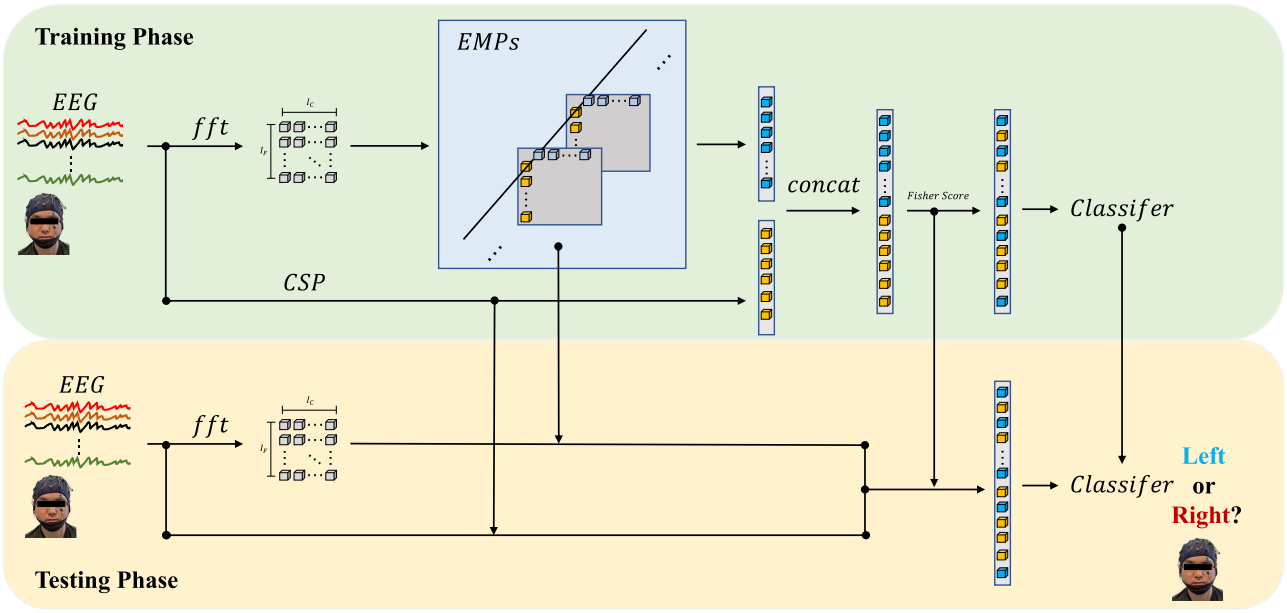

Fig. 1. Overall architecture. The training phase is divided into two branches. In the first branch, the segmented MI-EEG trial is processed by band-pass filtering and FFT to construct the spatial-spectrum tensor. The UMLDA algorithm projects the spatial-spectrum tensor into a feature vector, which is the wNBFs. In the second branch, the CSP algorithm extracts a feature vector of length 6, the BBFs, from the segmented and band-pass filteblack MI-EEG trial. The classifier is trained on the TFFC, which is a combination of wNBFs and BBFs. The downward arrow (↓) in the picture represents the parameters that need to be learned in the training set. In the test phase, only the feature index calculated from the training set is used.

TABLE I
THE NARROWBANDS USED IN THIS ARTICLE

Frequency Index	Frequency Band (Hz)
1	[1, 4]
2	[4, 8]
3	[8, 12]
4	[12, 16]
5	[16, 20]
6	[20, 24]
7	[24, 30]
8	[30, 35]
9	[35, 40]

frequency components related to motor imagery are mainly located between 8 Hz and 30 Hz, we need to extract the components of interest from $\mathcal{X}'$ and discarded the unrelated parts. Following the FBCSP algorithm's setting, we established nine narrow frequency bands to extract the interested components. The extraction process is defined below.

$$\mathcal{X}_F(i,j) = log((\frac{\sum_{k=a_i}^{b_i} \mathcal{X}'_{eeg}(k,j)}{b_i - a_i + 1})^2)$$
$$a_i = \frac{l_i \times 1023}{f_s} + 1, \quad b_i = \frac{h_i \times 1023}{f_s} + 1 \tag{4}$$

where l_i and h_i are the lower cut-off frequency and the upper cut-off frequency of the i-th frequency band. $\mathcal{X}_F(i,j)$ is the Differential Entropy (DE) of $\mathcal{X}_{eeg}$ in the i_{th} frequency band of the j_{th} channel [32], [33]. The $\mathcal{X}_F \in \mathbb{R}^{I_F \times I_C}$ is the spatial-spectrum tensor we wanted, where the I_F is nine. The frequency bands used in this article are shown in Table. I.

C. Common Spatial Pattern Extraction

The CSP algorithm is an effective spatial filtering method commonly used to extract features in MI-based BCI systems. The spatial filters are consideblack projection vectors and are calculated to maximize the variance of one class while minimizing the other class's variance. Consider two classes of EEG signals $\mathbf{X}_{i,1}, \mathbf{X}_{i,2} \in \mathbb{R}^{I_T \times I_C}$ from the experimental ith trial, where I_C is the number of channels, and I_T denotes the number of sampling points. The spatial covariance matrix of class c is computed as follows:

$$\sum_c = \frac{1}{n_c} \sum_{i=1}^{n_c} \mathbf{X}_{i,c}^T \mathbf{X}_{i,c} \tag{5}$$

where n_c represents the number of trials in class c. Then, the spatial filter that maximizes the variance of one class and minimizes the variance of the other can be calculated by

$$\mathbf{J}_c(w) = \frac{w^T \sum_{\mathbf{c}} w}{w^T \sum_{\bar{\mathbf{c}}} w} \quad s.t. \|w\|_2 = 1 \tag{6}$$

where w is the spatial filter. The optimization of the Rayleigh quotient can be converted to the generalized eigenvalue problem

$$\sum_c w = \lambda \sum_{\bar{c}} w \tag{7}$$

where λ and w are the generalized eigenvalue and eigenvector. The spatial filters $\mathbf{W}_{csp}$ are formed by eigenvectors corresponding to m maximum and minimum eigenvalues.

The projection signal $\mathbf{Z}$ of the signal trial is given by

$$\mathbf{Z} = \mathbf{W}_{csp} \mathbf{X} \tag{8}$$

Then, the pth feature of the single trial can be obtained as follows:

$$f_p = log(\mathbf{Z}_p \mathbf{Z}_p^T) \tag{9}$$

where $\mathbf{Z}_p$ is the pth row of $\mathbf{Z}$ ($p = 1, 2, \ldots, 2m$). And $\mathbf{f}_{csp}$ is the final feature vector, where $\mathbf{f}_{csp} = [f_1, f_2, \cdots, f_{2m}] \in \mathbb{R}^{2m}$.

D. Multilinear Spatial-Spectrum Feature Extraction

The tensor's projection has two types. Tensor-to-tensor projection (TTP) [34] and tensor-to-vector projection (TVP) [30]. We used the UMLDA algorithm, a TVP method, to project a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_1 \times \ldots \times I_N}$ to a vector $\mathbf{y} \in \mathbb{R}^P$ by a set of elementary multilinear projections (EMPs) learned from data, denoted as $\{\mathbf{u}_p^{(1)^T}, \mathbf{u}_p^{(2)^T}, \cdots, \mathbf{u}_p^{(N)^T}\}$, $p = 1, 2, \cdots, P$. P is the number of components we wanted and the pth component of $\mathbf{y}$ is obtained from the pth EMP as $\mathbf{y}(p) = \mathcal{X} \times_1 \mathbf{u}_p^{(1)^T} \times_2 \mathbf{u}_p^{(2)^T} \cdots \times_N \mathbf{u}_p^{(N)^T}$. The UMLDA's objective is to determine a set of P EMPs $\{\mathbf{u}_p^{(n)^T}, n = 1, \cdots, N\}_{p=1}^P$ that maximize the scatter ratio while producing features with zero correlation. In this study, N is 2 corresponding to space mode and spectrum mode, and P is set to be 50. We apply the UMLDA algorithm on the space-spectrum tensor $\mathcal{X}_F$ as below:

$$\mathbf{f}_{ss}(p) = \mathcal{X}_F \times_1 \mathbf{u}_p^{(1)^T} \times_2 \mathbf{u}_p^{(2)^T}, \quad p = 1, 2, \cdots, P. \quad (10)$$

where $\mathbf{f}_{ss} \in \mathbb{R}^P$ is extracted feature vector.

E. Features Fusion and Fisher Score Rank

We use vector concatenation to fuse features from CSP algorithm and UMLDA algorithm.

$$\mathbf{F} = concat(\mathbf{f}_{csp}, \mathbf{f}_{ss}) \quad (11)$$

where $\mathbf{F} \in \mathbb{R}^{2m+P}$ is the wNBFs.

As blackundant features cannot improve the training accuracy, we need a feature selection mechanism to find out the most discriminative features [35]. After feature extraction, we employed a Fisher score strategy for feature selection [36]. The Fisher score is defined as

$$S_F = \frac{|\mu_1 - \mu_2|^2}{\sigma_1 + \sigma_2} \quad (12)$$

where μ_i and σ_i denote the mean and standard deviation of class i over an individual feature. We sort these features in descending order by the value of S_F, and then take the top k features to train the SVM classifier. The optimal k is calculated by 10-fold cross-validation.

F. Classifier Tools

To obtain reliable experimental conclusions, we selected four mainstream classifiers for comparative experiments. They are the support vector machine (SVM), the k-nearest neighbor classifier (KNN), linear discriminator (LDA), and naive bayes classifier (NB). We used the libsvm toolbox [37] to implement the first one, and the parameters setting string was '-t 2'. For the other three classifiers, we used the functions, which are fitcknn(), fitcdiscr(), and fitcnb() [38] in Matlab2020a, to implement them and used the default parameters.

TABLE II
STATISTICS OF THE TWO DATASETS

	Number of Subjects	Days per Subject	Number of Channels	Sample Rate (Hz)	Trials per Subject
Dataset 1	7	3	59	1000	400
Dataset 2	7	1	59	100	200

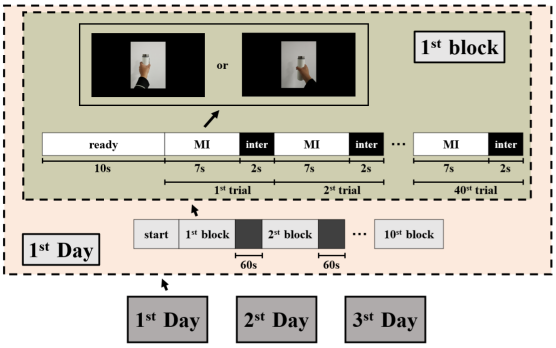

Fig. 2. Each subject needed to complete a motor imagery experiment three times on different days. Each experiment included 10 blocks. Each subject had 10-second preparation time before the start of each block. There was a 60-second interval between blocks.

III. Data Acquisition and Pre-Processing

A. Participants in Dataset 1

Seven subjects (aged 23-26: one female and six males) participated in the experiments with written consent. Among them, 6 subjects (S2-S7) were naive BCI users, and only one (S1) had a previous BCI experiment. The overall design of our experiment was shown in Fig.2. Each subject was asked to conduct experiments in three successive days. Each day of the experiment was divided into 10 blocks. And each block included 40 trials, where 20 trials were for left-hand motor imagery, and 20 trials were for right-hand motor imagery. We obtained 400 (40×10) trials for one subject in one day. $Sx.y$ represented the experiment of the xth subject in the yth day. This data set was collected using Neuroscan acquisition system with SynAmps RT 64-channel Amplifier.

At the beginning of the experiment, the subjects were seated comfortably in a chair with armrests 60 (± 5) cm from a 20-in LCD monitor (refresh rate: 60 Hz and resolution: 1600×1200). During the experiment, participants were asked to relax and minimized their eye's and muscle's movements.

B. Participants in Dataset 2

Dataset 2 was from BCI-Competition-IV-1 [39]. Dataset 2 was recorded from seven subjects (A, B, C, D, E, F, G), including four real human subjects (named A, B, F, G), and three artificially generated "participants" (named C, D, E). Two motor imagery classes were selected for each subject from the three classes: left hand, right hand, and foot. Here we only used the calibration data because of the complete marker information. There were 200 trials for each subject. We randomly split these trials into ten blocks for 10-fold cross-validation. The recording was made using BrainAmp MR plus amplifiers and a Ag/AgCl electrode cap. The details of the

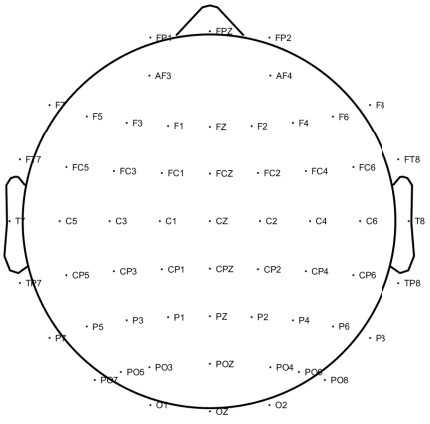

Fig. 3. Location of electrodes.

competition, including ethical approval and the raw data, can be downloaded from http://www.bbci.de/competition/iv/.

C. Pre-Processing

In each experiment, 59 EEG electrodes were used to record the EEG signals, with a sample rate of 1000 Hz for dataset 1 and 100 Hz for dataset 2. The EEG electrodes' location configuration is shown in Fig.3. The EEG electrodes' impedances were maintained below 20 $k\Omega$ during the entire experiment and checked before the start of each experiment. Each sample was segmented from [0, 4] s for dataset 1 and [0, 2.5] s for dataset 2 by marks. To blackuce the high computational burden from the sample rate of 1000 Hz for dataset 1, we downsampled the sample rate into 200 Hz. For the CSP algorithm, the recorded EEG signals were band-passed with a 10-order Butterworth filter between 8 Hz and 30 Hz. For the UMLDA algorithm, the recorded EEG signals were band-passed filteblack with a 10-order Butterworth filter between 0.1 Hz and 40 Hz.

IV. Results

A. Baselines and Comparison Criteria

In this work, the decoding accuracies of the different methods were evaluated in Matlab2020a. Computer hardware resources included 32 GB RAM and 2.21 GHz Intel Core I7 CPU. Because each experiment included 10 blocks in dataset 1, each experiment was divided into 10 folds according to the 10 blocks, 9 folds were used as the training set and 1 remaining fold was used as the testing set. We did not use the random division method for cross-validation, instead, we divided the block as the smallest unit to avoid randomness on the experimental results as much as possible. For dataset 2, we used 10-fold cross-validation to evaluate the performance of each algorithm. Specifically, each time we selected one fold as the testing set and all other folds as the training set and then performed the machine learning process. Repeated this process for ten times. The averaged decoding accuracy over all folds was used as the final metric.

The following methods were chosen for performance comparison.

Algorithm 1 Tensor-Based Frequency Features Combination (TFFC)

Input:
The training dataset $\{(\mathcal{X}_{eeg}^{(i)}, \mathbf{Y}^{(i)})\}_{i=1}^{n_{tr}}$;
The number of selected features, k;
The number of EMPs, P;
Output:
$\mathbf{F}'$, $\mathbf{W}_{csp}$ and $\{\mathbf{u}_p^{(n)^T}, n = 1, \cdots, N\}_{p=1}^{P}$.
1: Use (3) and (4) to obtain the $\mathcal{X}_F$.
2: Obtain $\{\mathbf{u}_p^{(n)^T}, n = 1, \cdots, N\}_{p=1}^{P}$. Use the same method as [30] to train EMPs on the training set.
3: Extract the spatial-spectrum bilinear weighted features $\mathbf{f}_{ss}$, using (10).
4: Use (8) and (9) to obtain $\mathbf{W}_{csp}$ and $\mathbf{f}_{csp}$.
5: Concatenate $\mathbf{f}_{csp}$ and $\mathbf{f}_{ss}$ to form $\mathbf{F}$ by (11).
6: Sort $\mathbf{F}$ in descending order with the Fisher-Score calculated by (12).
7: Select the best k features from $\mathbf{F}$ to construct the $\mathbf{F}'$.

1) CSP feature extraction. Because of the CSP algorithm's unshakable position in the field of MI-BCIs, we chose it as the baseline algorithm for this research. In the CSP algorithm process, each segmented EEG signal was first filteblack by a pass-band filter with [8, 30] Hz and projected by a spatial projection matrix. The column number of the spatial projection matrix was set to be 6 referring to [12].
2) FBCSP, an extended version of the CSP. In the FBCSP algorithm, each segmented EEG signal was passed through multiple narrowband filters, and the CSP algorithm extracted features in each frequency band. These features formed a feature set, and then k optimal features were selected as the final feature set through a feature selection method. In this article, we used the same feature selection method to selected 4 pairs of CSP features as [21] did. The source code can be downloaded from https://github.com/stupiddogger/FBCSP.
3) UMLDA, a framework for the recognition of multidimensional objects, known as tensor objects [30]. We used this method to extract wNBFs and shown the classifiers' decoding performance when only wNBFs were used.
4) Tensor-based Frequency Features Combination (TFFC) method, which was a feature extraction and fusion method that focused on wNBFs and BBFs, simultaneously. For wNBFs, this method constructed a spatial-spectrum tensor and used UMLDA, a tensor analysis method, to extract spatial-spectrum bilinear weighted features. For BBFs, this method integrated the features extracted by CSP. 10-fold cross-validation was used to determined the optimal parameters for each subject. The overall proposed algorithm was shown in Algorithm 1. We have opened part of the source code and template data to promote the development of BCI field, which could be downloaded from https://github.com/iuype/TFFC.

TABLE III
SIGNIFICANCE TEST ($\alpha = 0.01$) FOR TABLE IV. **BOLDFACE** HIGHLIGHTS SIGNIFICANT DIFFERENCES

α = 0.01		SVM			KNN			LDA			NB		
		CSP	UMLDA	FBCSP	CSP	UMLDA	FBCSP	CSP	UMLDA	FBCSP	CSP	UMLDA	FBCSP
TFFC	SVM	**4.87E-05**	**3.86E-05**	**3.86E-05**	**7.55E-06**	**1.21E-05**	**3.52E-07**	**1.10E-04**	**1.87E-05**	**1.43E-04**	**3.97E-05**	**2.52E-05**	**7.55E-06**
	KNN	**1.02E-03**	**1.51E-04**	**1.10E-03**	**5.62E-05**	**5.49E-05**	**1.13E-06**	**9.87E-04**	**9.86E-05**	1.60E-02	**3.25E-04**	**1.68E-04**	**1.20E-04**
	LDA	**3.25E-04**	**1.81E-04**	**1.73E-03**	**2.17E-05**	**7.70E-05**	**2.73E-06**	**7.65E-04**	**1.11E-04**	1.56E-02	**2.24E-04**	**1.84E-04**	**1.86E-04**
	NB	**1.10E-04**	**3.86E-05**	**1.16E-04**	**9.23E-06**	**1.21E-05**	**3.52E-07**	**1.84E-04**	**1.87E-05**	**3.44E-04**	**6.71E-05**	**2.52E-05**	**1.21E-05**

TABLE IV
MEAN (%) AND STANTARD DEVIATION (%) OF TFFC METHOD AND BASELINES METHODS. **BOLDFACE** HIGHLIGHTS THE MAXIMUM CLASSIFICATION ACCURACY AND MINIMUM STANTARD DEVIATION FOR INDIVIDUAL CLASSIFIERS

SUBJECT	SVM				KNN				LDA				NB			
	CSP	UMLDA	FBCSP	**TFFC**	CSP	UMLDA	FBCSP	**TFFC**	CSP	UMLDA	FBCSP	**TFFC**	CSP	UMLDA	FBCSP	**TFFC**
A	63.00	56.50	79.00	**88.50**	60.00	67.00	78.00	**87.50**	66.50	60.50	79.50	**84.50**	64.50	63.00	74.50	**85.00**
B	68.50	50.50	69.50	**70.00**	**69.50**	60.00	59.00	68.50	65.00	51.50	57.50	**68.50**	65.50	53.50	56.00	**65.50**
C	**84.50**	44.00	78.00	83.00	**81.00**	55.50	72.50	79.00	82.50	54.00	80.50	**82.50**	82.00	58.50	76.00	**83.50**
D	94.50	69.50	84.00	**95.00**	91.50	68.00	82.50	**93.00**	96.50	66.50	87.50	**96.50**	**96.00**	67.00	83.50	95.50
E	95.50	74.00	83.00	**96.50**	**97.00**	72.50	78.50	95.00	96.00	72.50	82.50	**96.00**	96.00	75.50	79.00	**96.00**
F	74.50	67.00	82.50	**94.50**	71.00	64.50	83.50	**91.50**	72.50	70.00	86.00	**92.00**	70.50	70.50	82.50	**92.50**
G	93.50	70.50	91.00	**94.00**	90.50	70.50	87.50	**94.50**	95.00	74.50	87.00	**94.00**	**92.00**	74.00	86.50	91.50
MEAN	82.00	61.71	81.00	**88.79**	80.07	65.43	77.36	**87.00**	82.00	64.21	80.07	**87.71**	80.93	66.00	76.86	**87.07**
STD	13.39	11.43	**6.59**	9.53	13.69	**5.97**	9.39	9.86	14.11	**9.06**	10.44	10.09	14.11	**8.13**	10.12	10.66
S1.1	59.75	63.75	58.00	**71.50**	63.25	63.00	54.50	**69.00**	**73.75**	63.75	71.50	67.00	**73.25**	64.00	67.75	72.50
S1.2	70.75	70.75	74.50	**77.50**	65.00	68.75	67.75	**72.25**	72.75	68.25	69.25	**73.00**	71.25	70.25	66.75	**75.25**
S1.3	60.00	63.00	65.75	**69.75**	59.25	63.75	60.50	**65.50**	67.00	64.50	**70.50**	65.50	67.25	64.00	**71.75**	69.25
S2.1	61.25	69.75	71.00	**72.50**	59.25	70.75	63.75	**69.25**	63.75	71.00	68.75	**70.50**	60.75	70.00	66.75	**72.25**
S2.2	50.50	60.75	**64.50**	64.25	53.75	60.25	60.25	**65.75**	57.25	62.25	61.00	**63.00**	55.75	61.25	63.25	**63.75**
S2.3	65.25	65.50	67.50	**71.00**	**66.00**	64.00	64.50	65.75	52.50	66.75	68.75	**69.00**	53.00	68.25	66.50	**70.50**
S3.1	54.25	58.75	**64.75**	62.00	54.00	59.00	59.00	**59.25**	54.25	59.25	**65.75**	60.25	53.50	61.50	62.50	**64.00**
S3.2	54.75	62.00	**62.25**	62.00	53.75	**62.75**	57.50	62.25	51.75	63.00	**64.25**	63.25	51.50	63.50	61.50	**63.75**
S3.3	57.00	62.00	51.25	**63.50**	52.75	61.25	53.25	**62.00**	56.25	61.25	55.50	**62.25**	54.00	**62.50**	55.50	62.25
S4.1	82.50	88.75	95.50	**96.75**	85.00	86.50	**97.50**	95.00	63.25	87.00	93.50	**96.00**	65.00	90.00	91.25	**94.50**
S4.2	80.00	91.25	**94.50**	94.25	77.25	89.75	**93.75**	**93.75**	68.00	89.00	91.25	**94.25**	65.75	90.50	89.50	**94.25**
S4.3	94.50	90.00	94.00	**97.50**	95.50	90.75	94.50	**98.25**	65.50	89.50	92.50	**96.50**	65.75	90.00	91.75	**97.00**
S5.1	**67.00**	63.00	52.50	63.00	**68.25**	61.75	53.75	61.50	58.25	**65.00**	55.75	64.00	59.25	65.00	50.25	**65.75**
S5.2	**63.25**	60.25	57.25	61.75	57.50	59.25	**62.75**	60.75	58.50	59.75	**64.00**	61.25	58.25	60.00	**63.75**	61.50
S5.3	**76.00**	63.25	61.50	74.25	70.75	64.25	55.25	**72.25**	79.25	63.75	**80.25**	71.75	76.25	65.25	**77.75**	70.25
S6.1	54.00	**59.50**	59.25	57.50	52.75	58.50	59.00	**63.75**	52.25	56.50	58.25	**58.75**	52.75	56.50	57.25	**58.00**
S6.2	62.25	61.00	**72.00**	64.50	56.50	61.00	59.75	**63.25**	52.25	59.00	**63.50**	59.50	52.25	61.25	64.00	**65.00**
S6.3	**66.75**	60.50	61.75	66.25	**61.50**	59.25	56.25	61.25	**67.00**	60.25	60.75	62.50	**65.00**	60.75	57.50	63.75
S7.1	57.25	62.50	57.00	**66.25**	52.25	**61.75**	54.00	60.75	51.75	62.00	**63.50**	63.25	54.00	62.25	57.50	**64.75**
S7.2	57.00	59.50	57.50	**62.25**	53.25	61.00	58.00	**61.75**	56.75	60.75	58.25	**61.00**	57.75	60.75	56.25	**63.75**
S7.3	52.25	57.00	60.25	**64.50**	55.00	57.25	59.25	**64.25**	62.00	58.75	**63.25**	61.00	58.25	58.50	59.00	**65.00**
MEAN	64.11	66.32	66.79	**70.61**	62.50	65.93	64.04	**68.93**	61.14	66.25	68.57	**68.74**	60.50	66.95	66.57	**70.33**
STD	11.24	**10.43**	13.07	11.75	11.57	**10.18**	13.59	11.78	**8.06**	9.91	11.51	11.91	**7.44**	10.29	11.82	11.22

To obtain a more objective comparison, the dataset's division was consistent among different algorithms.

B. Classification Performance Comparison

All experimental results were shown in Table. IV. The first column of Table. IV showed the information of the subject's name, and the second to fifth columns showed the performance of the four feature sets when SVM, KNN, LDA, and NW were used as classifiers.

The second column of Table. IV showed the individual day's accuracy for the four methods when the SVM was used as the classifier. We observed that our TFFC method outperformed the other three approaches. In dataset 1, CSP, UMLDA, FBCSP, and TFFC achieved the classification performance of 70.61%, 66.79%, 66.32% and 64.11% respectively. In dataset 2, they achieved classification performances of 88.79%, 81.00%, 61.71%, and 82.00%. It had better accuracy than FBCSP in 22 out of 28 subjects, UMLDA in 25 out of 28 subjects, and CSP in 23 out of 28 subjects. Similar results were obtained when using KNN, LDA, and NW as shown in Table. IV. These experimental results showed that classification accuracy is significantly enhanced by using the proposed TFFC method compared with the baseline methods.

The sub-column of Table. IV could be represented as O-C, where O is the feature set name and C is the classifier's name, e.g., TFFC-SVM.

TABLE V
THE COMPARISON WITH OTHER ARTICLES

SUBJECT	METHOD			
	JIN-1 [13]	JIN-2 [13]	JIANG [40]	**TFFC**
A	64.50	66.50	84.50	**88.50**
B	62.00	58.50	**82.00**	70.00
C	-	-	**87.00**	83.00
D	-	-	**97.50**	95.00
E	-	-	**98.50**	96.50
F	60.50	63.50	91.50	**94.50**
G	85.00	85.00	79.00	**94.00**
MEAN	68.00	68.38	88.57	**88.79**
STD	11.45	11.56	**7.54**	9.53
p-value	0.029	0.018	0.474	-

To verify the reliability of our experimental results, a paiblack t-test was used between the baseline algorithm and our proposed algorithm. Specifically, the matlab fuction $ttest$ was used to test the Table. IV. All improvements were statistically significant ($p < 0.01$). The paiblack t-test's results were shown in Table. III. These p-values are corrected by FDR [42]. Moreover, we compablack our method with Jin's [13] and Jiang's [40] on dataset 2. As shown in Table. V, our results are comparable with Jiang's and much better than Jin's. Note that Jiang's method consideblack a more time-domain feature than Jin's and ours.

Fig. 4. Feature selection on training dataset. The dotted line represents the experimental results of each subject on each day. The solid line is the average of all experimental results.

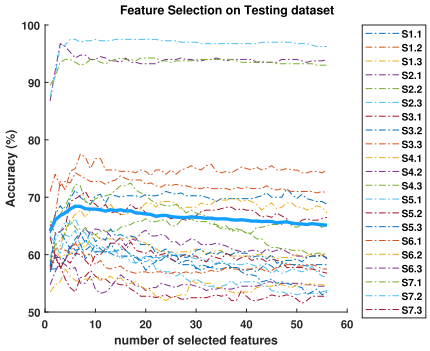

Fig. 5. Feature selection on testing dataset. The dotted line represents the experimental results of each subject on each day. The solid line is the average of all experimental results.

C. Visualization of Feature Selection

Fig.4 showed the influence of the number of selected features on the proposed method's classification accuracy on training set. Fig.5 showed the influence of the number of selected features on the proposed method's classification accuracy on testing set. The dotted line represented the experimental results of each subject on each day. The solid line was the average accuracy of all experimental results. Noticed that we selected the peak value on each dotted line as the final metric for each experiment.

D. Visualization of Feature Fusion

After repeated experiments, the optimal parameters of feature selection were determined for each subject. The optimal parameter of feature selection for subject A was 3, which contained two BBFs and one wNBF obtained by (8), (9) and (10). The selected two CSP projection vectors and EMP were visualized in Fig.6. In Fig.6, the pictures in the upper row were the selected CSP projection vectors, and the pictures in the bottom row were the selected EMP. We observed that the value of each element in the vector $\mathbf{u}_2^{(2)}$ was not equal, which supported our viewpoint that better features can be obtained by linearly weighting narrowband components.

Fig.7 shows the fusion percentage between wNBFs and BBFs. The fusion percentage was calculated by

$$Ratio = \frac{n_1}{n_1 + n_2} * 100\% \tag{13}$$

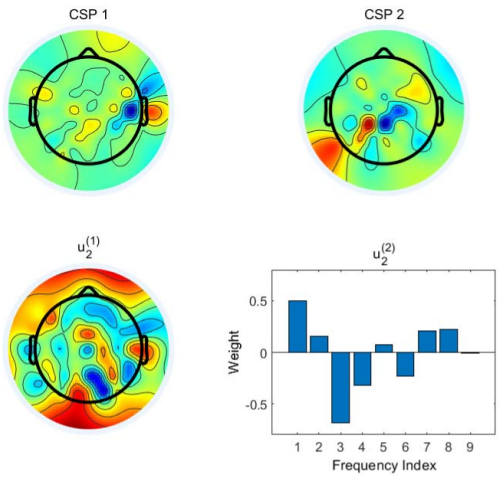

Fig. 6. The selected CSP projection vectors and EMP for subject A in dataset 2. The sub-pictures in the upper row are the selected CSP projection vectors. The sub-pictures in bottom row are the selected EMP.

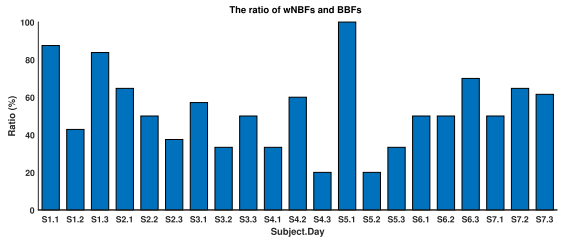

Fig. 7. The ratio between the number of selected wNBFs and the number of selected BBFs.

where n_1 was the number of selected wNBFs and n_2 was the number of selected BBFs. We observed that the average fusion percentage was around 50%, which indicated that both types of features were equally important.

E. Investigate the Performance Influenced by the Choice of P and the Number of Channels

In [30], the correct recognition rate (CRR) was maintained at a stable level when $P = [30, 60]$. Inspiblack by this, we heuristically set P to be 50. Although under this setting, our proposed method has achieved quite good results, it is still very meaningful to study the influence of different values of P on the classification performance of the TFFC. To strengthen persuasiveness of our study, we added additional experiments on dataset 1, where the $P = [0, 1, 2, 3, 4, 5, 10, 20, 30, 40, 50]$ and SVM as the classifier. The experimental results were shown in Fig. 8. When $P = 0$, the feature set degenerated into a CSP feature set. We observed that as P increased, the classification performance of SVM also steadily increased. When P was greater than 5, the classification performance of SVM reached a stable level.

In addition to the value of P, the number of channels was another important factor that affected the performance of TFFC. We choose three different settings of channels to explore the performance of the TFFC feature set under different channel settings. The first one only coveblack the motor cortices, which included C3, C4, and Cz. The second one included 20 channels, which were FC-5/3/1/2/4/6, C-5/3/1/z/2/4/6, and CP-5/3/1/z/2/4/6 [6]. The final

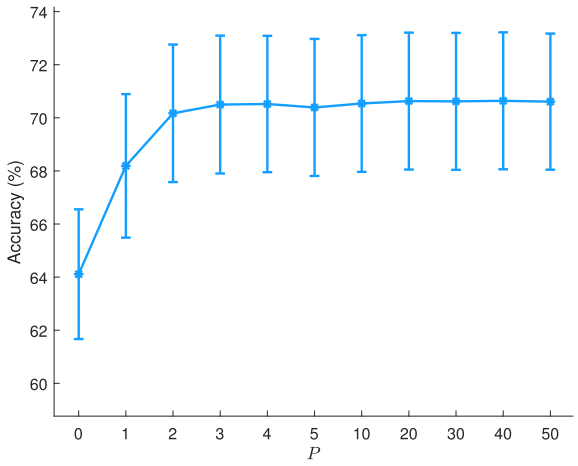

Fig. 8. The performance influenced by *P*. The error bar is standard error of mean (SEM).

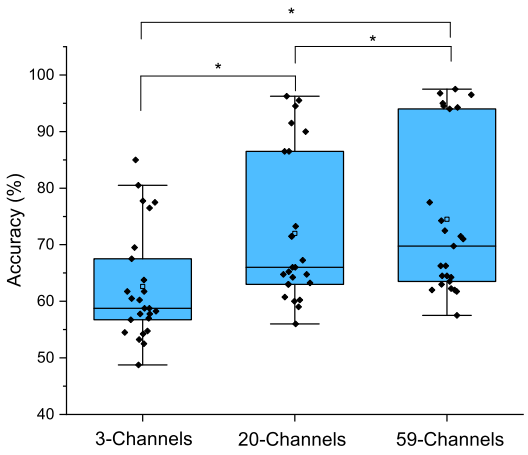

Fig. 9. The performance influenced by the setting of channels. The difference between each pair of the three settings of channels was indicated by paiblack t-tests ($*p < 0.001$).

one included 59 channels, which were plotted in Fig. 3. We conducted experiments with different settings of channels on dataset 1 and dataset 2, and the experimental results were shown in Fig. 9. We observed that the first channel setting achieved the worst performance, and the third channel setting achieved the best performance. And the difference among the three was statistically significant ($p < 0.001$).

V. Discussion

Concerning MI-BCI systems, the frequency is one of the essential feature dimensions apart from spatial distribution. Optimization techniques [22], [26]–[28] for frequency bands can often bring definite classification performance improvements. However, the state-of-the-art methods have some shortcomings, which have been discussed in the section I. Focusing on these problems, we constructed a new feature set, named TFFC, based on CSP, FFT, and UMDAL. To verify the TFFC method, we conducted many experiments on a public dataset and a private dataset. Experimental results and corresponding significance test were shown in Table.IV and Table.III.

The UMLDA is a method designed for tensor objects published in TNN in 2009 [30]. This article gave a very comprehensive description of the tensor calculation in the UMLDA method. In mathematics, a tensor is an algebraic object that describes a multilinear relationship between sets of algebraic objects related to a vector space. Objects that tensors may map between include vectors and scalars, and even other tensors. There are many types of tensors, including scalars and vectors, which are the simplest tensors. So, a matrix is a particular case of a tensor, it is a second-order tensor, but it cannot be said that it is not a tensor [43]. The operation of projecting a second-order tensor into a scalar is the tensor calculation. Moreover, in the original article of UMLDA and other studies, the experiment of a second-order tensor is also carried out [30], [44], [45]. Other studies that regard matrices as tensors and have conducted some research using tensor analysis methods [46]–[48]. Meanwhile, we consider that our proposed method incorporates the features of BBFs and wNBFs, so we named this method **T**ensor-Based **F**requency **F**eatures **C**ombination (**TFFC**). What's more, our method has a certain potential for expansion. When more factors are consideblack and the dimensionality of the tensor increases, our method still has the potential to extract the features. However, because the calculation of extracting high-order tensors' features is very time-consuming, how to speed up this calculation has become a significant problem in tensor calculations, and further research is needed.

A. Why TFFC Works

It is vital to explore the reasons for the success of TFFC and we believe that there are two main reasons for it.

First, we solved the problem discussed in section I. In the establishment of the feature set, we simultaneously extracted features of two aspects, broadband features (BBFs), and weighted narrowband features (wNBFs), and merged them together. For BBFs, we followed the classic CSP algorithm that needed to work in the frequency band of 8-30 Hz. We selected 6 features as part of TFFC. Inspiblack by the differential entropy matrix, mostly used in emotion classification [49], and the FBCSP algorithm, we calculated the wNBFs using UMLDA. In pattern recognition tasks, including MI-EEG signal recognition, features that were not related to each other were desirable since they contained minimum blackundancy and ensublack the independence of features [30]. The UMLDA was an algorithm designed specifically for tensor data for this purpose. The UMLDA method was used to extract wNBFs from the 9-narrowband differential entropy matrix. These wNBFs were extracted by supervised linear projections on spatial and frequency directions, and they were not related to each other theoretically [30]. The linear combination of features among the narrow frequency bands produced new features with more discriminative ability. Fig. 10 and Fig. 11 plotted the highest rank score EMPs for all subjects in dataset 1. We observed that most frequencies weighted vectors' elements were not equal, which showed that the features calculated by EMPs did contain linear combination information among narrow frequency bands. However, due to the differences of EEG across subjects, the critical features may not be hidden in narrowbands for all subjects. To build a robust feature set, we fused the wNBFs with BBFs and named it TFFC. Experiments showed that the TFFC

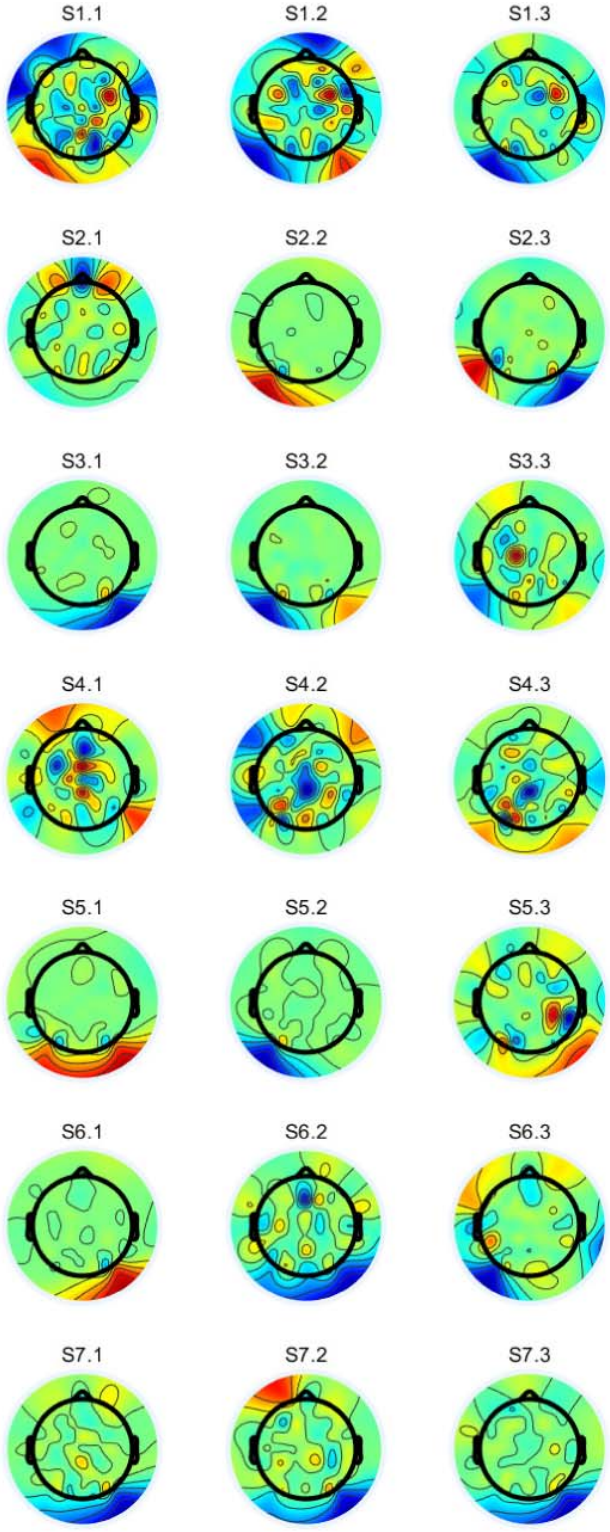

Fig. 10. Spatial weighted vectors learned by UMLDA for all subjects and all days in dataset 1.

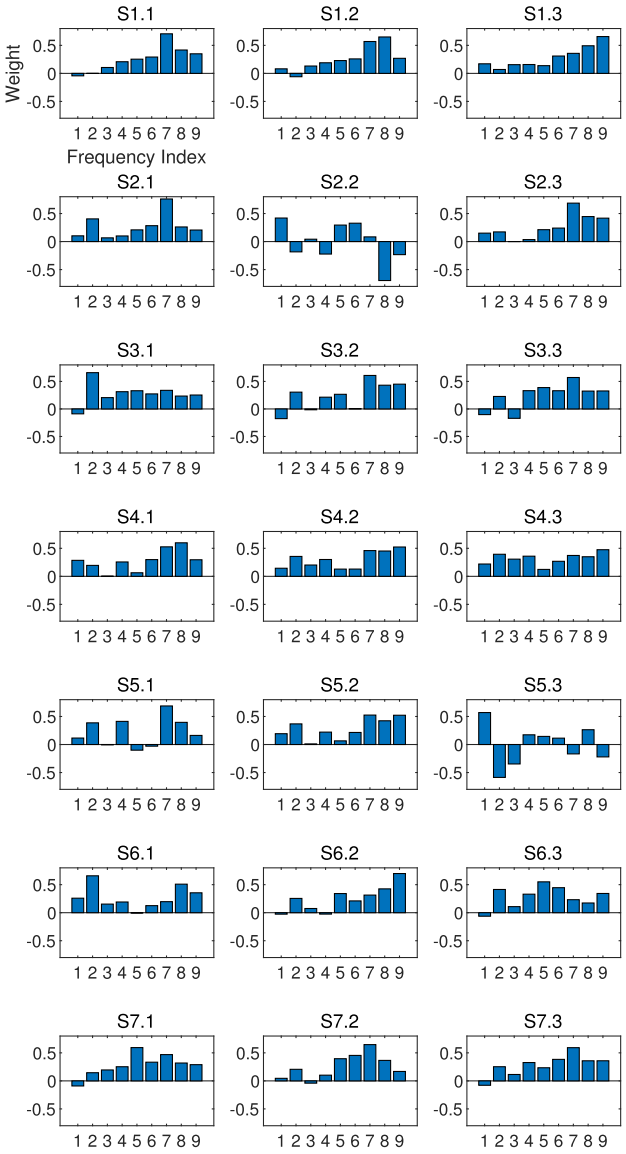

Fig. 11. Frequency weighted vectors learned by UMLDA for all subjects and all days in dataset 1.

feature set had better classification performance, not unexpectedly, which indicated that it had a more vital discriminative ability.

Second, CSP and FBCSP were corner cases of TFFC. As discussed in the introduction section, CSP extracted BBFs and FBCSP extracted NBFs. The core competence of TFFC was that it contained the above-described two kinds of features. When the elements in the learned frequency weighted vectors were equal or approximately equal, we believed that the information of the features is similar to the BBFs. For instance, sub-picture *S*4.3 in Fig. 11 was just the case discussed above, where the elements in the frequency weighted vectors were more consistent than those in other sub-pictures. This showed that the information in the feature calculated by this EMP was more similar to the information in the feature calculated by CSP. When some elements in the learned frequency weighted vectors were close to 0, we believed that the information of the features was similar to the NBFs. There were many instances in Fig. 11.

B. Performance With Different Classifier

The academia hoped to find a feature set with robustness in the machine learning project for MI-BCI systems. This could blackuce classifier selection for researchers and speed up the development cycle of MI-BCI systems. We chose

KNN as the representative of the classifier with a simple structure and SVM as the representative of the classifier with a more complicated structure. We observed that TFFC-SVM, TFFC-KNN, TFFC-LDA, and TFFC-NW, respectively, had a significantly higher than CSP/UMLDA/FBCSP-SVM's, CSP/UMLDA/FBCSP-KNN's, CSP/UMLDA/FBCSP-LDA's, CSP/UMLDA/FBCSP-NW's ($p < 0.05$), which meant that the proposed TFFC was robust and superior.

C. The Feature Interpretability and Biological Meaning

Like most MI feature extractors [15]–[17], our proposed TFFC method included spatial filters. However, the difference was that our method included frequency band filters. The frequency band filter was a vector that performed linear fusion among different frequency band components. The frequency band filters trained by the supervised method can effectively fuse the information hidden in the narrow frequency bands. In Fig. 6, we visualized the spatial filters and frequency band filters trained on the data of subject A. The spatial filters were divided into two parts, one part came from the CSP algorithm, and the other came from the UMLDA algorithm. In Fig. 6, CSP2 was mainly responsible for collecting features from the motor cortices, which was consideblack the main area for MI signals [11]. Although CSP2 looked more regular than UMDAL, the CSP2 spatial filter had no discriminant ability for the frequency band and could only be improved by choosing a specified bandpass filter based on expert knowledge. However, the most efficient frequency band had noticeable differences among subjects in the MI task, and the fixed frequency band of 8-30 Hz could not be a good choice [28]. Fortunately, the UMLDA method could select and even fuse the features among different frequency bands. Therefore, the features calculated by TFFC include two aspects, BBFs and wNBFs. In the lower right corner of Fig. 6, it can be observed that frequency band 1 and frequency band 3 were more concerned by the UMLDA algorithm. Since UMLDA was a supervised feature extraction method, band 1 and band 3 were more important for the classification task of MI in the case of subject A. In Fig. 10 and Fig. 11, we performed many visualizations of UMLDA's spatial filters and frequency band filters. We can observe that the spatial filters of UMLDA were more inclined to capture the features of the motor cortices on subject 4 of Fig. 10. For Subject 1 of Fig. 10, this tendency can also be seen. In Fig. 11, we observed that the frequency band filters of subjects 1 and 7 had a certain degree of stability across the three days, while the other subjects did not have this phenomenon. This indicated that there might be a small number of people in the population with cross-day stability in the frequency features of the MI task.

In the branch of calculating BBFs, we used a fixed setting of a bandpass filter of 8-30 Hz. Previous studies have proved that the 8-30 Hz frequency band features had certain cross-subject stability in MI tasks [50], [51]. In the branch of calculating wNBFs, TFFC used the UMLDA algorithm to adaptively let the projection vector on the frequency band learn each individual's weighting coefficient. This idea is relatively common in the machine learning community, that machine learning models can learn more efficient and robust performance by learning on coarse and fine features [52]. The coarse features give the classifier the basic ability to classify, while the fine features enable the classifier to learn the individual features of each subject, which further increases the classification ability.

D. Limitations

In the MI-BCI system, cross-subject and cross-time is a tough challenge, which requires algorithms to find stable and unchanging patterns [53]. Unfortunately, our method cannot do this either. In Fig. 7, Fig. 10 and Fig. 11, we can observe that fusion percentage, space and frequency weighted vectors are unstable across subjects and across days. In addition, a low decoding accuracy (around 75%) is also a shortcoming of our method. There is still much room for improvement here.

VI. Conclusion

The CSP algorithm only extracts one single broadband features, and the FBCSP extracts feature in multiple narrow frequency bands, which is improved compablack to CSP. However, each feature is either selected or discarded, which may lead to the neglect of discriminative fusion information between these features. To further explore the discriminative information related to MI in the frequency bands, we used FFT to construct a spatial-spectrum tensor and applied tensor analysis to mine the discriminative information. We successfully extracted wNBFs and used feature cascade and feature selection to fuse them with BBFs. The final feature set was named TFFC. The TFFC method significantly improved the classifiers' classification performance for MI-EEG. Also, Our experiment constructed a more robust feature set with low dependence to classifiers. Moreover, the complementarity between BBFs and wNBFs was verified. This article certificated the importance of frequency information in the MI-BCI system and provided a new direction for designing a feature set of MI-EEG.

References

[1] J. Wolpaw *et al.*, "Brain-computer interface technology: A review of the first international meeting," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 164–173, Feb. 2000.

[2] Y. Yu, Y. Liu, J. Jiang, E. Yin, Z. Zhou, and D. Hu, "An asynchronous control paradigm based on sequential motor imagery and its application in wheelchair navigation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 12, pp. 2367–2375, Dec. 2018.

[3] S. N. Abdulkader, A. Atia, and M.-S. M. Mostafa, "Brain computer interfacing: Applications and challenges," *Egyptian Inform. J.*, vol. 16, pp. 213–230, Jul. 2015.

[4] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[5] M.-H. Lee, S. Fazli, J. Mehnert, and S.-W. Lee, "Subject-dependent classification for robust idle state detection using multi-modal neuroimaging and data-fusion techniques in BCI," *Pattern Recognit.*, vol. 48, pp. 2725–2737, Aug. 2015.

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

[7] W. Zhang and D. Wu, "Manifold embedded knowledge transfer for brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 5, pp. 1117–1127, May 2020.

[8] L. Xu, M. Xu, Y. Ke, X. An, S. Liu, and D. Ming, "Cross-dataset variability problem in EEG decoding with deep learning," *Frontiers Hum. Neurosci.*, vol. 14, p. 103, Apr. 2020, doi: 10.3389/fnhum.2020.00103.

[9] K. Wang, M. Xu, Y. Wang, S. Zhang, L. Chen, and D. Ming, "Enhance decoding of pre-movement EEG patterns for brain–computer interfaces," *J. Neural Eng.*, vol. 17, no. 1, Jan. 2020, Art. no. 016033, doi: 10.1088/1741-2552/ab598f.

[10] D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw, "Mu and beta rhythm topographies during motor imagery and actual movements," *Brain Topogr.*, vol. 12, no. 3, pp. 177–186, Mar. 2000.

[11] M. Lotze and U. Halsband, "Motor imagery," *J. Physiol. Pairs*, vol. 99, pp. 386–395, Feb. 2006.

[12] Z. J. Koles, M. S. Lazar, and S. Z. Zhou, "Spatial patterns underlying population differences in the background EEG," *Brain Topogr.*, vol. 2, no. 4, pp. 275–284, 1990.

[13] J. Jin, R. Xiao, I. Daly, Y. Miao, X. Wang, and A. Cichocki, "Internal feature selection method of CSP based on L1-norm and Dempster–Shafer theory," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4814–4825, Nov. 2021.

[14] Q. Wang *et al.*, "A motor-imagery channel-selection method based on SVM-CCA-CS," *Meas. Sci. Technol.*, vol. 32, no. 3, Mar. 2021, Art. no. 035701.

[15] T. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, and I. Ahmad, "A review of channel selection algorithms for EEG signal processing," *EURASIP J. Adv. Signal Process.*, vol. 2015, no. 1, pp. 1–21, Dec. 2015.

[16] J. Jin, Y. Miao, I. Daly, C. Zuo, D. Hu, and A. Cichocki, "Correlation-based channel selection and regularized feature optimization for MI-based BCI," *Neural Netw.*, vol. 118, pp. 262–270, Oct. 2019.

[17] F. Lotte and C. Guan, "Spatially regularized common spatial patterns for EEG classification," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 3712–3715.

[18] B. R. Mathon, M. M. Ozbek, and G. F. Pinder, "Dempster–Shafer theory applied to uncertainty surrounding permeability," *Math. Geosci.*, vol. 42, no. 3, pp. 293–307, Apr. 2010.

[19] S. Razi, M. R. K. Mollaei, and J. Ghasemi, "A novel method for classification of BCI multi-class motor imagery task based on Dempster–Shafer theory," *Inf. Sci.*, vol. 484, pp. 14–26, May 2019.

[20] A. Yazdani, T. Ebrahimi, and U. Hoffmann, "Classification of EEG signals using Dempster–Shafer theory and a k-nearest neighbor classifier," in *Proc. 4th Int. IEEE/EMBS Conf. Neural Eng.*, Apr. 2009, pp. 327–330.

[21] K. Keng Ang, Z. Yang Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.)*, Jun. 2008, pp. 2390–2397.

[22] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers Neurosci.*, vol. 6, no. 1, p. 39, 2012.

[23] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) algorithm using online adaptive and semi-supervised learning," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2011, pp. 392–396.

[24] W. Abbas and N. A. Khan, "FBCSP-based multi-class motor imagery classification using BP and TDP features," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 215–218.

[25] B. Yang, J. Tang, C. Guan, and B. Li, "Motor imagery EEG recognition based on FBCSP and PCA," in *Proc. Int. Conf. Brain Inspired Cognit. Syst.*, Springer, 2018, pp. 195–205.

[26] K. P. Thomas, C. Guan, L. C. Tong, and V. A. Prasad, "An adaptive filter bank for motor imagery based brain computer interface," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2008, pp. 1104–1107.

[27] K. P. Thomas, C. Guan, L. C. Tong, and A. P. Vinod, "Discriminative FilterBank selection and EEG information fusion for brain computer interface," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2009, pp. 1469–1472.

[28] Y. Zhang, Y. Wang, J. Jin, and X. Wang, "Sparse Bayesian learning for obtaining sparsity of EEG frequency bands based feature vectors in motor imagery classification," *Int. J. Neural Syst.*, vol. 27, no. 2, Mar. 2017, Art. no. 1650032.

[29] E. Gysels, P. Renevey, and P. Celka, "SVM-based recursive feature elimination to compare phase synchronization computed from broadband and narrowband EEG signals in brain–computer interfaces," *Signal Process.*, vol. 85, no. 11, pp. 2178–2189, 2005.

[30] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 103–123, Jan. 2009.

[31] A. Cichocki, R. Zdunek, A. H. Phan, and S.-I. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation*. Hoboken, NJ, USA: Wiley, 2009.

[32] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," in *Proc. 6th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, Nov. 2013, pp. 81–84.

[33] L.-C. Shi, Y.-Y. Jiao, and B.-L. Lu, "Differential entropy feature for EEG-based vigilance estimation," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 6627–6630.

[34] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Multilinear principal component analysis of tensor objects for recognition," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2006, pp. 776–779.

[35] Y. Liu, Q. Zhao, and L. Zhang, "Uncorrelated multiway discriminant analysis for motor imagery EEG classification," *Int. J. Neural Syst.*, vol. 25, no. 4, Jun. 2015, Art. no. 1550013.

[36] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[37] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 27:1–27:27, 2011. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[38] *MATLAB Version 9.10.0.1613233 (R2021a)*, The Mathworks, Inc., Natick, MA, USA, 2021.

[39] M. Tangermann *et al.*, "Review of the BCI competition IV," *Frontiers Neurosci.*, vol. 6, no. 1, p. 55, 2012, doi: 10.3389/fnins.2012.00055.

[40] J. Jiang, C. Wang, J. Wu, W. Qin, M. Xu, and E. Yin, "Temporal combination pattern optimization based on feature selection method for motor imagery BCIs," *Frontiers Hum. Neurosci.*, vol. 14, p. 231, Jun. 2020.

[41] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern Recognit.*, vol. 40, no. 7, pp. 2038–2048, Jul. 2007.

[42] J. D. Storey, "A direct approach to false discovery rates," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 64, no. 3, pp. 479–498, 2002.

[43] J. C. Kolecki, *An Introduction to Tensors for Students of Physics and Engineering*. Washington, DC, USA: United States National Aeronautics and Space Administration, Glenn Research Center, 2002.

[44] C. Liu, K. He, J.-L. Zhou, and C.-B. Gao, "Discriminant orthogonal rank-one tensor projections for face recognition," in *Proc. Asian Conf. Intell. Inf. Database Syst.*, Springer, 2011, pp. 203–211.

[45] C. Liu, K. He, J.-L. Zhou, and F.-N. Lang, "Discriminant orthogonal rank-one tensor projections in extended graph embedding framework," *Int. Inf. Inst. (Tokyo). Inf.*, vol. 15, no. 8, p. 3423, 2012.

[46] W. Hu, J. Gao, J. Xing, C. Zhang, and S. Maybank, "Semi-supervised tensor-based graph embedding learning and its application to visual discriminant tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 1, pp. 172–188, Jan. 2017.

[47] J. Ye, R. Janardan, and Q. Li, "Two-dimensional linear discriminant analysis," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 17, 2004, pp. 1569–1576.

[48] X. Li, M. K. Ng, Y. Ye, E. K. Wang, and X. Xu, "Block linear discriminant analysis for visual tensor objects with frequency or time information," *J. Vis. Commun. Image Represent.*, vol. 49, pp. 38–46, Nov. 2017.

[49] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.

[50] Y. Pei *et al.*, "Data augmentation: Using channel-level recombination to improve classification performance for motor imagery EEG," *Frontiers Hum. Neurosci.*, vol. 15, p. 113, Mar. 2021.

[51] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, Jun. 2018, Art. no. 031005.

[52] J.-H. Kim and J.-P. Heo, "Learning coarse and fine features for precise temporal action localization," *IEEE Access*, vol. 7, pp. 149797–149809, 2019.

[53] P. L. C. Rodrigues, C. Jutten, and M. Congedo, "Riemannian procrustes analysis: Transfer learning for brain–computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 8, pp. 2390–2401, Aug. 2019.