Motor Imagery EEG Classification Based on Riemannian Sparse Optimization and Dempster-Shafer Fusion of Multi-Time-Frequency Patterns

Jing Jin, *Senior Member, IEEE*, Tingnan Qu, Ren Xu, Xingyu Wang, and Andrzej Cichocki, *Life Fellow, IEEE*

***Abstract*— Motor imagery-based brain-computer interfaces (MI-BCIs) features are generally extracted from a wide fixed frequency band and time window of EEG signal. The performance suffers from individual differences in corresponding time to MI tasks. In order to solve the problem, in this study, we propose a novel method named Riemannian sparse optimization and Dempster-Shafer fusion of multi-time-frequency patterns (RSODSF) to enhance the decoding efficiency. First, we effectively combine the Riemannian geometry of the spatial covariance matrix with sparse optimization to extract more robust and distinct features. Second, the Dempster-Shafer theory is introduced and used to fuse each time window after sparse optimization of Riemannian features. Besides, the probabilistic values of the support vector machine (SVM) are obtained and transformed to effectively fuse multiple classifiers to leverage potential soft information of multiple trained SVM. The open-access BCI Competition IV dataset IIa and Competition III dataset IIIa are employed to evaluate the performance of the proposed RSODSF. It achieves higher average accuracy (89.7% and 96.8%) than state-of-the-art methods. The improvement over the common spatial patterns (SFBCSP) are respectively 9.9% and 12.4% ($p < 0.01$, paired t-test). These results show that our proposed RSODSF method is a promising candidate for the performance improvement of MI-BCI.**

***Index Terms*— Brain-computer interface, motor imagery, Riemannian geometry, sparse optimization, Dempster-Shafer theory.**

I. Introduction

BRAIN-COMPUTER Interface (BCI) bridges people with physical disabilities to the outside sensual world by converting brain signals such as electrophysiological (EEG) to control instructions [1], [2], [3]. Motor imagery (MI) is the mental imagination of body movement without actual muscle movement and its corresponding rhythmic activities of the brain could be observed and applied as the input signals of BCI systems [4]. The regular rhythmic power changes in the sensorimotor area within mu (8-12 Hz) and beta (13–30 Hz) frequency bands are called event-related desynchronization/ synchronization (ERD/ERS), which can be detected to discriminate different kinds of MI tasks [5]. MI-based BCIs are flexible in their applications for consideration of MI not demanding external stimuli [6]. However, the problems of existing MI-BCIs decoding suffer from the instability and complexity of EEG signals [7]. It is pivotal and remains challenging to improve EEG decoding ability, extract discriminative information from low signal-to-noise EEG, and realize reliable classification of different MI tasks [8], [9], [10].

Various methodologies have been proposed to extract discriminant patterns from high-dimensional EEG signals [11], [12], [13]. The spatial patterns associated with ERD generated from different MI tasks could be identified [14]. The most frequently used method is the common spatial patterns (CSP) proposed in [15]. Whereas, the effectiveness of CSP is

Manuscript received 5 June 2022; revised 6 September 2022 and 4 October 2022; accepted 23 October 2022. Date of publication 26 October 2022; date of current version 30 January 2023. This work was supported in part by the Science and Technology Innovation 2030 Major Projects under Grant 2022ZD0208900; in part by the Grant National Natural Science Foundation of China under Grant 62176090; in part by the Shanghai Municipal Science and Technology Major Project under Grant 2021SHZDZX; in part by the Program of Introducing Talents of Discipline to Universities through the 111 Project under Grant B17017; in part by the Shuguang Project supported by the Shanghai Municipal Education Commission and the Shanghai Education Development Foundation under Grant 19SG25; in part by the Ministry of Education and Science of the Russian Federation under Grant 14.756.31.0001; in part by the Polish National Science Center under Grant UMO-2016/20/W/NZ4/00354; and in part by the National Government Guided Special Funds for Local Science and Technology Development, Shenzhen, China, under Grant 2021Szvup043. *(Corresponding author: Jing Jin.)*

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

Jing Jin is with the Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, Shanghai 200237, China, and also with the Shenzhen Research Institute, East China University of Science and Technology, Shenzhen 518063, China (e-mail: jinjingat@gmail.com).

Tingnan Qu and Xingyu Wang are with the Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, Shanghai 200237, China (e-mail: qutingnan838338691@163.com; xywang@ecust.edu.cn).

Ren Xu is with g.tec medical engineering GmbH, 4521 Schiedlberg, Austria (e-mail: xu@gtec.at).

Andrzej Cichocki is with the Skolkovo Institute of Science and Technology (SKOLTECH), 143026 Moscow, Russia, also with the Systems Research Institute, Polish Academy of Sciences, 01-447 Warsaw, Poland, and also with the Department of Informatics, Nicolaus Copernicus University, 87-100 Torun, Poland (e-mail: a.cichocki@skoltech.ru).

Digital Object Identifier 10.1109/TNSRE.2022.3217573

very sensitive to the selection of frequency bands and easily affected by irrelevant noise [16]. Thus, a fixed time window for feature extraction may lead to low classification accuracy [17]. There are numerous extensions of traditional CSP to optimize subject-specific operational time windows or frequency bands, such as sub-band CSP (SBCSP) [18], filter bank CSP (FBCSP) [19] and sliding window discriminative CSP (SWDCSP) [20]. Advanced algorithms by dividing time window and frequency bands effectively improve the classification accuracy, however, these methods still bring much information redundancy and multiply the decoding time cost [21]. To solve the problem of ineffectiveness, several feature selection algorithms emerge to automatically select the corresponding extracted features, such as sparse CSP (SCSP) [22] and sparse FBCSP (SFBCSP) [23] which select features through sparse learning.

With regard to further effectively utilizing discriminative information from multiple time windows and frequency bands, several heuristic methods have been proposed to automatically select one or a set of time windows and frequency bands [24], [25], [26]. Though these methods contrive to improve the final decoding performance, the feature selection procedure should not be an open/shut case [27]. In [24], the frequency bands are selected through the affinity propagation of multi-filter bank CSP features, which contain the cluster centers. In [25], time windows are selected based on correlation analysis and performance evaluation. In [26], the frequency band and time interval are coded with the particles and the optimal settings of them can be simultaneously detected by the evolution of particles for individual subjects. However, valuable messages in abandoned bands would be lost. Because the optimal time-frequency bands change rapidly and even vary between trials [28]. However, simply concatenating extracted features of all time-frequency bands would cause much redundancy in information and increase the difficulty of classification then multiplying the calculation time of algorithms [29]. Moreover, the Riemannian method has been wildly applied in MI-BCIs for its robustness [30]. A. Barachant et al. successfully utilized the symmetric positive-definite (SPD) spatial covariance matrix [31], to introduce the robustness of Riemannian distance and Riemannian mean into the BCI system and build more stable decoding models. They also combined the Riemannian method with state-of-the-art algorithms by mapping the data onto the Riemannian tangent space (RTS), where the matrices could be vectorized to fit traditional Euclidean models [32].

Aiming at solving the preceding problems, we propose a Riemannian sparse optimization and Dempster-Shafer fusion approach (RSODSF) for MI-BCIs classification to effectively utilize the underlying information of multiple time-frequency patterns. Raw EEG signals are segmented and filtered into multiple time windows and overlapping sub-bands respectively. Then discriminative features of each sub-band from Riemannian tangent space are extracted and then optimized through sparse learning to enhance the stationarity and robustness of the algorithm. Moreover, Dempster-Shafer theory (DST) is introduced to implement the classification fusion of multi-time windows. After we train a support vector machine (SVM) classifier with the linear kernel for the classification, we conduct a series of experiments to evaluate our proposed algorithms with different public datasets of BCI Competition. The major contributions of this work are:1) This work combines the Riemannian geometry of the spatial covariance matrix with sparse optimization to extract more robust and distinct features. 2) The Dempster-Shafer theory is introduced and used to effectively fuse the probabilistic outputs of each time window and reduce uncertainty of the final prediction.

The remainder of this paper is organized as follows. Section II elaborates on the materials and proposed method. Section III and IV describe the public datasets we use for experiments and the experimental setup, show the result and discussion, respectively, and conclude in Section V.

II. Methodology And Materials

A. Riemannian Feature Extraction

In MI-BCIs, Riemannian geometry is applied to the covariance matrix of EEG signal, which is one of the most frequently used statistical properties [33] and has the property of being symmetric positive-definite (SPD) and located in the Riemannian manifold [34].

Let the $X = [x(t), \ldots, x(t+L-1)] \in R^{N_C \times L}$ be the bandpass filtered EEG signal, N_c, L respectively demonstrates the number of channels and sample points. The N_c x N_c spatial covariance matrix C and Riemannian tangent space features can be calculated using the following steps.

1) Bandpass filter the raw EEG data at a specific time-frequency band.

2) Calculate the N_c x N_c spatial covariance matrix C of each trial using Eq. (1):

$$C = \frac{1}{L-1} X(t) X^T(t) \tag{1}$$

3) Use the iterative algorithm to calculate the Riemannian mean between N covariance matrices, which has a minimum sum of the squared distances to all covariance matrices [35]. N demonstrates the number of EEG trials.

$$\delta_R(C_1, C_2) = \|\log(C_1^{-1} C_2)\|_F = \left[\sum_{i=1}^{N} \log^2 \lambda_i\right]^{\frac{1}{2}} \tag{2}$$

$$\bar{C} = \arg\min_{C} \sum_{i=1}^{N} \delta_R^2(C, C_i) \tag{3}$$

where C_1, C_2 are two different covariance matrices respectively, λ_i denotes the i-th eigenvalue of $C_1^{-1} C_2$, $||\ ||_F$ denotes the Frobenius norm [36], and log() is the log-matrix operator.

4) Use the logarithmic mapping to project the data points from Riemannian manifold to its corresponding tangent space, as shown in **Fig. 1**. and expressed in Eq. (4). The projected data points could be mapped back to the manifold by the exponential mapping, i.e., the exponential function of the matrix [35].

$$\begin{aligned} S_i &= Log_{\bar{C}}(C_i) = \bar{C}^{\frac{1}{2}} Log(\bar{C}^{-\frac{1}{2}} C_i \bar{C}^{-\frac{1}{2}}) \bar{C}^{\frac{1}{2}} \\ C_i &= Exp_{\bar{C}}(S_i) = \bar{C}^{\frac{1}{2}} Exp(\bar{C}^{-\frac{1}{2}} S_i \bar{C}^{-\frac{1}{2}}) \bar{C}^{\frac{1}{2}} \end{aligned} \tag{4}$$

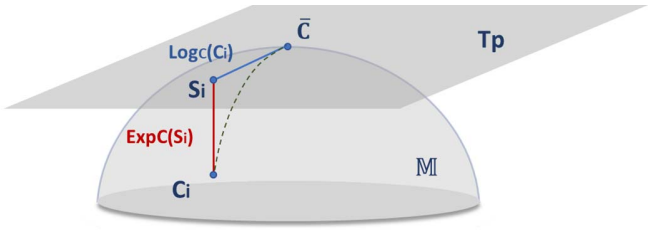

Fig. 1. Riemannian manifold $\mathbb{M}$ and corresponding local tangent space T_P at reference covariance point $\bar{C}$. (SPD matrix C_i ⇔ symmetric matrix S_i).

5) The projected data, i.e., the symmetric matrices can be vectorized into Euclidian objects and further used for classification, expressed in Eq. (5).

$$f = vect(C) = [C_{1,1}; \sqrt{2}\, C_{1,2}; \ldots ; C_{N_C,N_C}] \in \mathrm{R}^{(N_C+1)N_C/2} \quad (5)$$

where f denotes $N \times ((N_c+1) \times N_c/2)$ vectorized features from N covariance matrices C. Considering the symmetry property of the projected matrices, the non-diagonal elements of C need to multiply $\sqrt{2}$ to maintain the standard when vectorizing the projected matrices of each trial [37].

B. Probabilistic Support Vector Machine

Support Vector Machine (SVM) is one of the most popular machine learning algorithms and has been broadly used in binary classification situations [38], which separates differential data samples by constructing a set of optimal high-dimensional hyperplanes. These hyperplanes are capable of maximizing the margins between two features samples, and are least affected by local disturbance of samples [39]. Unlike advanced deep learning methods [40], SVM does not require large scales of data and is suitable for small and medium-sized nonlinear data samples [41]. Moreover, it has a potent generalization ability for unknown data and flexibility for decision making [38]. For consideration of expedience, we select the SVM with a linear kernel as our fundamental classification tool, as expressed in Eq. (6).

$$\min_{\omega,b,\xi} \frac{1}{2}\|\omega\|^2 + C \cdot \sum_{i=1}^{n} \xi_i$$
$$s.t.\ y_i(\omega^T \phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, i=(1,\ldots,n) \quad (6)$$

where the weight vector ω and an offset b of the hyperplane could be obtained through the minimum optimization, $\phi(x_i)$ is the function that maps x_i into a higher-dimensional space, C is the penalty parameter of the error term, and ξ denotes the slack variable [38].

To ensemble the SVM outputs trained through multiple time windows, we further introduce an SVM–Platt model to transform the outputs into posterior probabilities [42]. The transformation steps are as follows:

1) Calculate the distances d between the sample points and decision boundary.

$$d(x_i) = x_i^T \omega + b \quad (7)$$

2) Note that if the distance is a positive value, the sample belongs to a positive class [43]. The transformed posterior probability ρ_j is actually the probability that a sample belongs to a specific class and could be computed by fitting a two-parameter sigmoid function of the distance value d_j which matches the posterior that is empirically observed [44].

$$\rho_j = \frac{1}{1 + \exp(Ad_j + B)} \quad (8)$$

where $\exp(\bullet)$ is the exponential function, A and B are the parameters representing the slope and the intercept [42].

3) Transform the labels into the probability target t_j, in order to fit the afore sigmoid function. In our study, labels 1, and 2 are respectively demonstrating left-hand and right-hand MI tasks. As expressed in Eq. (9), the right-hand MI task, i.e., label = 2, is the positive target for example.

$$t_j = \begin{cases} 0 & \text{label=1} \\ 1 & \text{label=2} \end{cases} \quad (9)$$

4) Find the optimal parameters A and B through a cross-entropy function, as expressed in Eq. (10). The two-parameter minimization is fitting by using a maximum likelihood estimation from the training set, which finds the two optimal parameters by minimizing the negative log likelihood of the training data [42].

$$A, B = \arg\min_{A,B} -\sum t_j \log(\rho_j) + (1 - t_j)\log(1 - \rho_j) \quad (10)$$

Since we obtain the transformed probabilistic outputs of several trained SVM models, we contrive to find a way to maximumly use underlying information within and between classifiers and effectively fuse multiple posterior probabilities.

C. Dempster-Shafer Theory

Dempster-Shafer Theory (DST) is an evidence theory presented by Dempster [45] and further developed by Shafer [46], which is an extension of the Bayesian inference method by introducing the uncertainty of our knowledge into the probabilistic space. DST could combine all the available evidence come from different sources about a specific event [45] and applied in BCI context [47]. In our context, it is a binary classification event utilizing sources of multiple time windows.

In short, given a finite and exhaustive set of possible conclusions to a question $\Omega = \Theta_1, \Theta_2, \ldots, \Theta_n$. 2^Ω includes all possible subsets of Ω and could be presented as $\emptyset, \Theta_1, \Theta_2, \ldots, \Theta_n, \Theta_1, \Theta_2, \{\Theta_1, \Theta_3, \ldots$. DST would assign a probability for each subset, called mass function and satisfies the following conditions:

$$m(\emptyset) = 0 \sum_{EA \in 2^\Omega} m(EA) = 10 \leq m(EA) \leq 1, EA \in 2^\Omega \quad (11)$$

where $\emptyset$ is an empty set, the focal element E_A is a hypothesis to the event and may include one or more conclusions. DST also defines belief and plausibility functions to model the

uncertainty, expressed in Eq. (12).

$$Bel(EA) = \sum_{EB \subseteq EA} m(EB)$$
$$Pls(EA) = \sum_{EB \cap EA \neq \emptyset} m(EB) \tag{12}$$

The smaller the distance between the plausibility and the belief, the lower the uncertainty. In order to further mitigate the uncertainty, we can fuse two sets of evidence of focal element E_A through joint mass function, expressed in Eq. (13). E_B and E_C respectively demonstrate the hypotheses to the event from two sets of evidence.

$$m_{1,2}(EA) = \begin{cases} \dfrac{\sum_{EB \cap EC = EA} m_1(EB) \times m_2(EC)}{1 - \sum_{EB \cap EC = \emptyset} m_1(EB) \times m_2(EC)}, & EA \neq \emptyset \\ 0, & EA = \emptyset \end{cases} \tag{13}$$

D. Riemannian Sparse Optimization and Dempster-Shafer Fusion of Multi-Time-Frequency Patterns (RSODSF)

Since the responding time and optimal operational frequency bands vary between subjects. We contrive to fully take benefit of scarce EEG resources and to operate the decoding better. Given N trials of EEG signals $X \in R^{N \times N_c \times L}$, and X is segmented into a set of T overlapping time windows and K overlapping sub-frequency bands. After the preceding feature extraction procedure, the multiple time-frequency Riemannian patterns are obtained. Each time window of which could be donated as $F_j = [f_1, f_2, \ldots, f_i]^T \in R^{K \times (N_C+1)N_C/2}$, where $f_i \in R^{(N_C+1)N_C/2}$ stands for the extracted Riemannian patterns from i-th frequency band.

Motivated by SFBCSP [23], a multiband optimization method is proposed and developed to select discriminative features through sparse regression [48], we tend to incorporate the sparse learning with our previous extracted Riemannian tangent space features. $l1$-norm is widely used due to its better optimization solution characteristics than $l0$-norm and sparser weight parameters than $l2$-norm [49]. The sparse regression model used in our study would be defined as:

$$\mathrm{w} = \arg\min_{\mathrm{w}} \frac{1}{2} ||\mathrm{y} - \mathrm{Fw}||_2^2 + \lambda ||\mathrm{w}||_1 \tag{14}$$

where $y \in R^N$ is the label vector of EEG trials, $F \in R^{N \times (N_C+1)N_C/2}$ is the extracted Riemannian features and $w \in R^{(N_C+1)N_C/2}$ is the weight vector, $l1$-norm is introduced to penalize the inessential feature weights close to zero. λ denotes the regularization parameter to control the sparsity of w, which would be determined through experiments.

Now that the Riemannian features of multiple frequency bands from each time window are extracted. We tend to further effectively fuse the information of T time windows in the classification stage.

In our context, two elements are corresponding to each MI task class label and would generate the frame of discernment $\Omega = \Theta_1, \Theta_2$. First, we contrive to get the mass function from every single time window. We respectively set each MI label i.e., labels=1 or 2, as the probability target to obtain the posterior probabilities P_1, P_2 of each classifier. For the sake of expedience, we simply use the identity function to transform the probability values to mass functions.

$$\begin{cases} m_i = P_{\Theta_i} = P_i, \Theta_i \in \{\Theta_1, \Theta_2\} \\ m_j = P_{\Theta_j}, P_{\Theta_j} = P_{\bar{\Theta}_i} \end{cases} \tag{15}$$

where m_i and m_j are respectively the mass functions of Θ_i and Θ_j, Θ_j is the complement of Θ_i, and P_i is the probabilistic outputs of SVM when label i is the target probability.

Noting that when choosing one label as the target probability, the probability output of another label would be unknown and usually set as zero, the complement of the target would be the uncertainty. However, when the value of zero is involved, the fusion result will always be zero regardless of the basic mass function allocation value of the other evidence.

In order to address the above-mentioned problems, we first contrive to fuse the mass functions of a time window by respectively setting each one of the labels as our target probability. Then we could obtain three non-zero mass functions of one model through Eq. (14), which respectively stand for three focal elements in the frame of discernment m_{i1}, m_{i2} and m_{iu}. In our BCI context, we tend to classify two MI tasks. We preset the frame of discernment as {LH, RH, {LH, RH}} for example and the third element is the uncertainty needed to be mitigated. As shown in TABLE I, the D-S fusion procedure of one PSVM model is achieved, which actually degenerates into Bayesian probability. P1 and P2 are the probabilistic outputs of the PSVM when the positive targets are respectively left-hand and right-hand MI tasks, respectively. The values behind the elements are the basic mass functions allocated in the last section. The mass functions m1, m2 are transformed from the P1, P2 through Eq. (15).

TABLE I
THE D-S FUSION PROCEDURE OF ONE PSVM MODEL TRAINED BY EACH TIME WINDOW IN OUR BCI CONTEXT (LH/RH: LEFT/RIGHT-HAND MI; M1/M2: MASS FUNCTIONS)

		m2		
		LH (0)	RH (P2)	{LH, RH} (1-P2)
m1	LH (P1)	LH (0)	∅ (P1*P2)	LH (P1*(1-P2))
	RH (0)	∅ (0)	RH (0)	RH (0)
	{LH, RH} (1-P1)	LH (0)	RH (P1*(1-P2))	{LH, RH} ((1-P1)*(1-P2))

After we obtain mass functions m_{i1}, m_{i2} and m_{iu} of the i-th time window. We manage to further combine T segmented time windows using Eq. (13). As shown in TABLE II, the D-S fusion procedure of mass functions from PSVM models trained by two different time windows is achieved. P is the probabilistic output of the PSVM. Once the final three joint mass functions are calculated, we are now capable of implementing the decision-making by comparing two mass functions corresponding to the label. The outputs of the framework would be assigned with the larger mass function and be more credible with smaller uncertainty. The combination of the mass functions of all time windows is achieved through

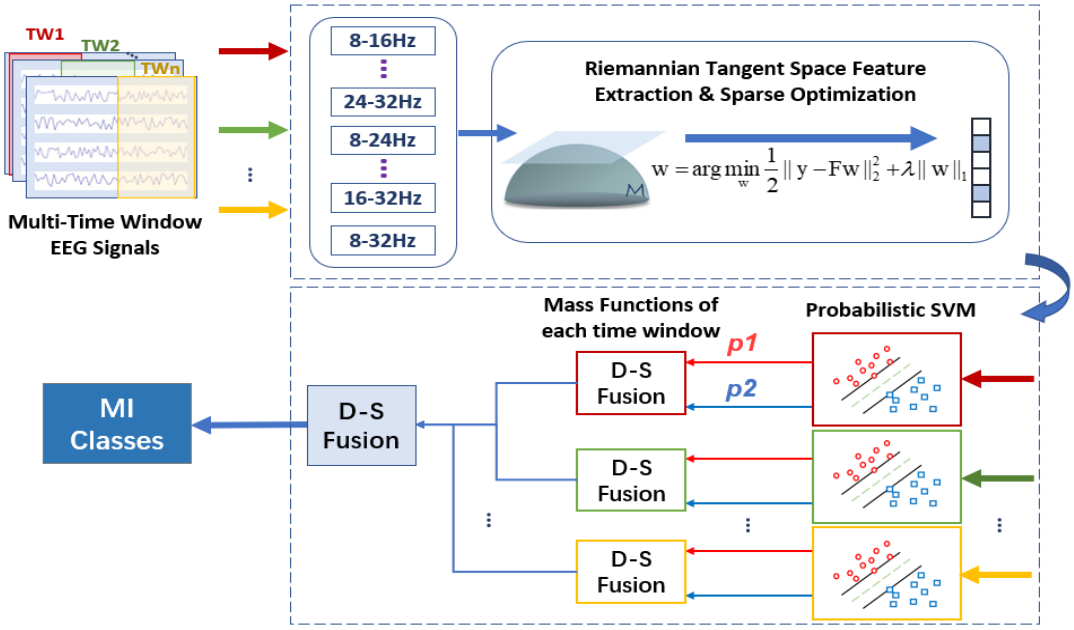

Fig. 2. The whole block diagram of the proposed RSODSF methods. (TW: Time Window; MI: Motor Imagery; D-S Fusion: Dempster-Shafer Fusion).

TABLE II
THE D-S FUSION PROCEDURE OF MASS FUNCTIONS FROM PSVM MODELS TRAINED BY TWO DIFFERENT TIME WINDOWS IN OUR BCI CONTEXT (LH/RH: LEFT/RIGHT-HAND MI; TW: TIME WINDOWS)

		TW2		
		LH (m21)	RH (m22)	{LH, RH} (m2u)
TW1	LH (m11)	LH (m11*m21)	∅ (m11*m22)	LH (m11*m2u)
	RH (m12)	∅ (m12*m21)	RH (m12*m22)	RH (m12*m2u)
	{LH, RH} (m1u)	LH (m1u*m21)	RH (m1u*m22)	{LH, RH} (m1u*m2u)

two together. To be specific, after one time window combines with another, the results of combination are further utilized to combine with a third one. The whole structure of the proposed method is shown in **Fig. 2**.

III. EXPERIMENTS AND RESULTS

A. Data Descriptions

1) Dataset 1: We choose Dataset IIa from BCI Competition IV [50] as the first dataset to evaluate the effectiveness of the proposed methods. It contains 22-channel EEG data recorded from 9 subjects performing 4 kinds of MI tasks, i.e., Left-Hand, Right-Hand, Feet, and Tongue mental tasks. Each participant completed two sessions, one for training and the other for evaluation [50]. We combine two sessions and choose Left and Right-Hand tasks to evaluate the binary classification performance, which contains 144 trials of each MI task. The training set and test set would be further divided for an 8-fold cross validation. The dataset is provided by the Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology. This work involves human subjects or animals in its research. We confirm all human/animal subject research procedures and protocols are exempt from review board approval. More details can be found on the following website: http://www.bbci.de/competition/iv/

2) Dataset 2: We choose Dataset IIIa from BCI Competition III [51] as the second dataset to evaluate the effectiveness. It contains 3 healthy subjects performing 4 kinds of MI tasks and recorded 60-channel EEG signals with a sampling rate of 250 Hz. We downsample them to 100Hz. In our experiments, a total of 180, 120, and 120 trials for subject k3, k6, and l1 were respectively included for two chosen different MI tasks. The training set and test set would be further divided for a 6-fold cross validation. This work involves human subjects or animals in its research. We confirm all human/animal subject research procedures and protocols are exempt from review board approval. Detailed information about this dataset can be found on: http://www.bbci.de/competition/iii/

B. Experimental Setup

In this study, EEG data of 0-4 s after the visual MI cue appeared has been extracted for signal processing. We firstly used 5th order Butterworth bandpass filters to capture EEG signals from 8-32Hz wide frequency bands. The length of the time window should not be too short to include the execution time of MI and also not be too long which would add irrelevant information. However, the response time fluctuates between different subjects. Moreover, time windows and filter banks of different lengths are adopted. The divided time windows are respectively, [0.5, 4], [0.5, 3.5], [1, 4], [0.5, 3], [1, 3.5] and [1.5, 4] after the visual MI cue appeared for DS1, and [0, 3.5], [0, 3], [0.5, 3.5], [0, 2.5], [0.5, 3] and [1, 3.5] for DS2 denoted as TW1-TW6. The 15 overlapping sliding filer banks are segmented with lengths of 8, 16, and 24 Hz. Moreover, in order to avoid the over-fitting of the trained

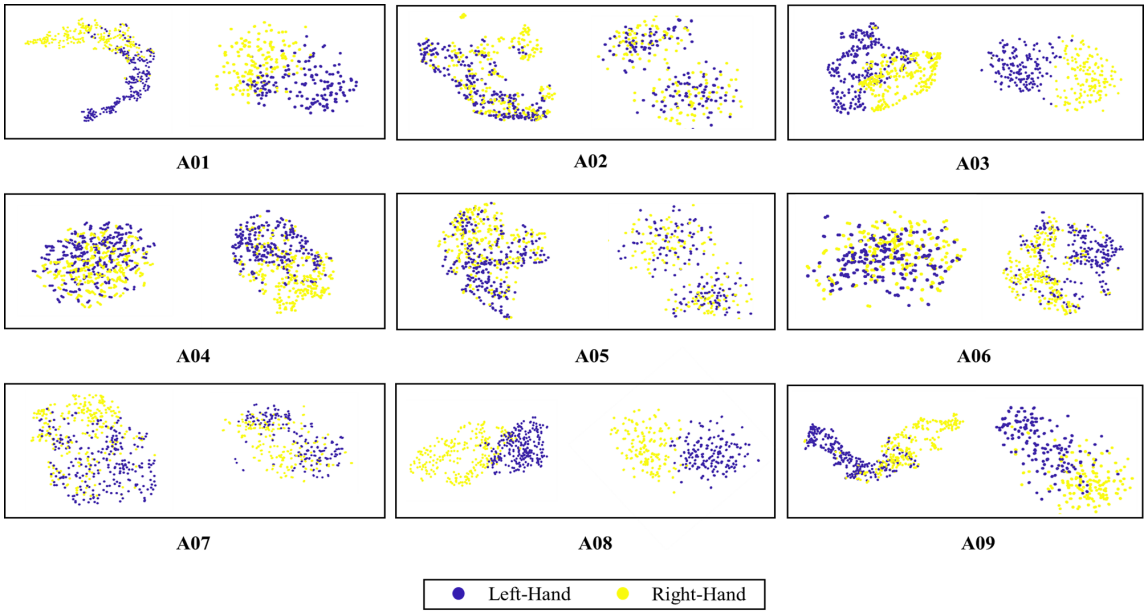

Fig. 3. Visualization of feature distribution extracted from subjects A01-A09. (left and right data clusters in each sub plot are respectively extracted by SFBCSP and the proposed method).

classification SVM models, the final classification accuracy is calculated within an 8-fold/6-fold cross-validation framework for DS1/DS2. We manually search a set of hyperparameters [10^−4, 10^−3, …, 10^1] and choose λ as 10^-2 through experiments to weigh the classification accuracies and computational cost. To validate the effectiveness of the proposed RSODSF, we implement extensive comparisons with several state-of-the-art methods.

(1) SFBCSP-SVM [23]: A fixed time window EEG of 0.5–4 s after the cue, filter banks are set the same as the proposed method and extracted by SFBCSP.

(2) SW-LCR [52]: It uses the longest consecutive repetition of the sequence of prediction of all the sliding windows.

(3) RoCSP-SRIT2NFIS [53]: A robust CSP with a self-regulated interval type-2 neuro-fuzzy inference system is used.

(4) AR-CSP+SRSG-FasArt [54]: An artifact rejected CSP and self-regulated adaptive resonance theory-based neuro-fuzzy classifier as "self-regulated supervised Gaussian fuzzy adaptive system Art (SRSG-FasArt)" is introduced.

(5) RSO-SVM: Sparse Optimization of Riemannian Tangent Space features of multi-frequency bands.

The entire evaluation is under the environment of MATLAB R2018a on a laptop with 2.60GHz CPU (i7-9750H, 8GB RAM). The classification tool is the SVM with linear kernel and default classifier parameters. The Covariance Toolbox is used which mainly focuses on Riemannian features extraction of SPD matrices and contains a series of MATLAB functions dedicated to covariance matrices estimation [55].

C. Comparison of Feature Distribution

To better interpret the experimental results, we utilize t-distributed stochastic neighbor embedding (t-SNE) [56] to visualize the extracted features. Fig. 3 presents the feature distribution extracted from subjects A01-A09 using SFBCSP and our proposed method. Compared with SFBCSP, the features extracted using our method have stronger discriminability and seem easier to be separated for the majority of subjects. As shown in Fig. 3, subjects A03, A04 and A06 have the most obvious improvement, especially A04 and A06 have an improvement of 14.3% and 11.9%, respectively. The feature distribution of A02 and A05 seems to have no improvement, yet the accuracies still improve a lot.

D. Classification Performance of Multiple Time-Frequency-Patterns Versus Canonical Time-Frequency Band

In MI-BCIs, canonical time-frequency bands are most commonly used to process the raw EEG signals, which is 0.5-4s after the visual cue and filtered within 8-32Hz. As shown in **Fig. 4**, the accuracies vary between time-frequency bands and the optimal operational time-frequency bands are not always the canonical ones, i.e., TW1-FB15 in **Fig. 4**.The multiple sub-bands would provide much more affluent band information that allows us to achieve more credible classification.

More specifically, as shown in Fig. 5 (a), each classifier corresponding to each time window has a relative accuracy. We can conclude from Fig. 5 (a) that the ensembled accuracies are always higher than results obtained from a single fixed time window, which also proves the effectiveness of our methods.

E. Classification Performance With Dempster-Shafer Fusion Versus Majority Voting Strategy

Numerous data fusion methods have been applied to combine outputs of multiple trained classifiers. The majority voting strategy (MVS) [57] is the most commonly adopted fusion method which is followed by the "Max Wins" rule and determines the decisions with the maximum votes. Whereas,

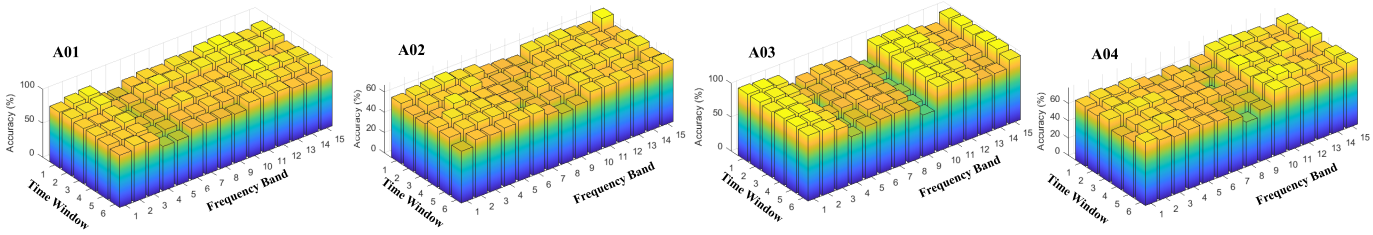

Fig. 4. Comparison of EEG Decoding accuracies of subject A01-A04 between divided multi-time-frequency Bands.

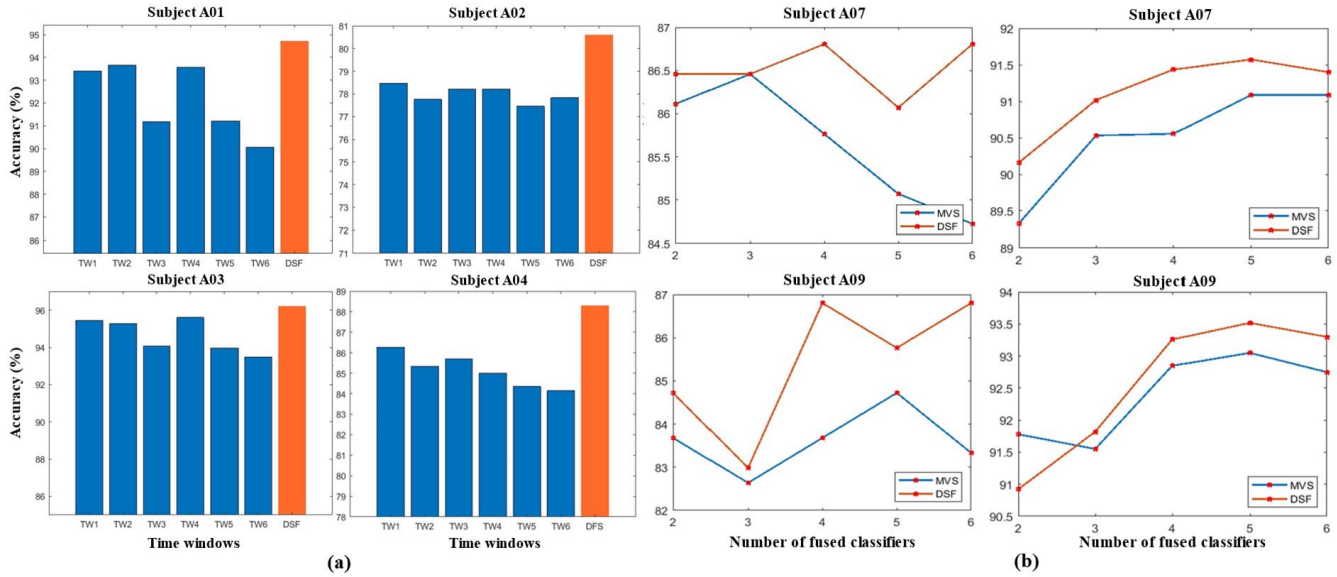

Fig. 5. (a) Classification accuracies of each time window and fused by proposed DSF from subjects A01-A04 (the blue and orange bars respectively denote the accuracies of each time window and fused by DSF). (b) Classification accuracies of MVS & DSF of CSP and RSO features from subject A07 and A09.

the unintelligible soft information between different classifiers would be lost. The proposed D-S fusion effectively resolves the ambiguity in determining the final class caused by the majority voting of binary classifier results. As shown in **Fig. 5 (b)**, the accuracy of subject A07 obviously decreases with applying MVS, as the number of classifiers increases while its accuracy increases with applying the proposed RSODSF. This might originate from the ambiguity caused by MVS. As the number of fused classifiers grows, DFS always outperforms MVS.

F. Classification Accuracies Comparison

The experimental results of binary classification are summarized in TABLE III. The mean accuracy improvement of the proposed RSODSF method, compared with the above SFBCSP/SW-LCR/RoCSP-SRIT2NFIS/AR-CSP+SRSG-Fas-Art/RSO-SVM are respectively 9.9%/12.4%, 9.7%/11.6%, 6.4%/4.2%, 4.2%/2.4% and 8.8%/11.3% for DS1/DS2, which demonstrates that the proposed RSODSF achieves the highest mean accuracy among them and outperforms other state-of-the-art methods in processing 9 of 12 subjects. All the p-values obtained by paired t-test between the RSODSF and any of others are less than 0.01 ($p<0.01$).

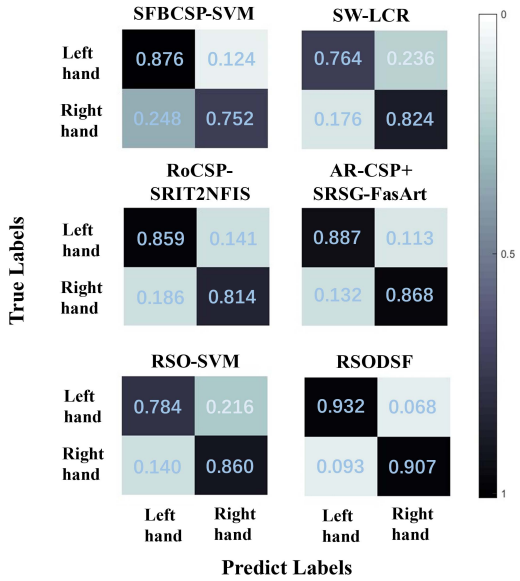

Fig. 6. Confusion matrices of the proposed rsodsf and the compared State-of-the-art methods.

As shown in TABLE III, the improvement of RSODSF is significantly obvious when applied in A02 and A05. The

TABLE III
THE CLASSIFICATION ACCURACIES (%), AVERAGE COMPUTATIONAL TIME(S) AND SIGNIFICANCE COMPARISONS (WHERE THE BEST RESULTS ARE MARKED IN BOLDFACE AND AVERAGE COMPUTATIONAL TIME IS DENOTED AS TIME)

Subjects		Methods					
		SFBCSP-SVM	SW-LCR	RoCSP-SRIT2NFIS	AR-CSP+SRSG-FasArt	RSO-SVM	RSODSF
Dataset 1	A01	88.2	86.8	93.1	95.1	90.3	**94.7**
	A02	56.9	64.6	68.8	68.8	61.8	**80.6**
	A03	96.9	95.8	97.2	**98.6**	97.2	96.2
	A04	76.4	67.4	75.0	81.3	74.3	**88.3**
	A05	60.4	68.1	66.0	72.2	64.6	**83.1**
	A06	72.9	67.4	72.2	75.0	74.0	**85.1**
	A07	86.1	80.6	86.1	86.1	85.1	**91.4**
	A08	**97.2**	97.2	97.2	97.9	**97.2**	95.0
	A09	83.3	92.4	93.8	94.4	83.3	**93.3**
	Mean±Std	79.8±14.5	80.0±13.5	83.3±12.8	85.5±11.6	80.9±13.1	**89.7±5.7**
Dataset 2	K3	97.6	95.6	**100.0**	**100.0**	94.0	99.6
	K6	60.7	66.7	78.8	84.0	71.0	**96.7**
	L1	94.8	93.3	98.9	**99.1**	91.6	**99.1**
	Mean±Std	84.4±20.5	85.2±16.1	92.6±11.9	94.4±9.0	85.5±12.6	**98.5±1.6**
Experimental results	Time	14.91	26.83	36.43	26.72	10.43	12.90
	p-value	<0.01	<0.01	<0.01	<0.01	<0.01	/

classification results of these two subjects are the worst ones among 9 subjects when processed by other methods. However, the improvement is not that much notable in those subjects who originally performed well with basic algorithms. Moreover, the accuracies of subjects A03 and A08 even slightly decline, which is probably on account of the well-performed EEG data of them and benefits from the simplicity of conventional algorithms. Since the decline in accuracies is not significant, the average performance of the proposed RSODSF still outperforms other state-of-the-art methods. The average confusion matrices of the proposed RSODSF and compared state-of-the-art methods (see from Fig.6) further evaluate the effectiveness of the proposed method. We also compare the computational time of the proposed method with other state-of-the-art methods. The average model training time and testing time is calculated and shown in TABLE III. The proposed RSODSF also outperforms in the matter of computational time. Compared with SFBCSP, the RSO method has improved the mean accuracies in both DS1 and DS2, especially for subject K6 in DS2 with an improvement of 10.3%. The use of Riemannian geometry has enhanced the robustness of the classification model. The consequent use of the SPD matrix as a feature may have led to the curse of dimensionality and increase the training time. However, the sparse optimization has been utilized to select the most discriminative features which reduces the influence of the large dimensionality of SPD matrix.

IV. DISCUSSION

Previous studies have demonstrated that Riemannian geometry could be effectively applied in BCI fields. Our proposed RSODSF method integrates advanced process algorithms and achieves the best performance among the abovementioned state-of-the-art method. However, the feature distribution of two different classes from subjects A02 and A05 in **Fig. 3** still seems mixing and confusing, but the classification accuracies are significantly increased with an improvement of 22.7% and 17.7% respectively. Moreover, we can see that the feature distribution of these two subjects is spatially divided into two clusters, which probably results from the intrinsic geometry structure of the EEG data. Moreover, the accuracy improvement in the feature layer (1.3% and 1.1%) is much less than in the classification layer (8.8% and 11.3%), the key step leading to the improvement in accuracy is at the classification layer. This has further explained the obvious accuracy improvement when the feature distribution has no enhancement.

Besides, see A01 and A03 in **Fig. 5 (a)**, we can observe the accuracies with TW2 of A01 and TW4 of A03 are higher than TW1, i.e., the most frequently used fixed time window in MI classification. These results further denote that TW1 is not always the optimal operational time interval. Due to the individual difference, the response time to the visual cue of subjects varies between subjects, and even between trials. Other achieved accuracies of time windows are obviously lower than TW1, however, the fusion of combined time windows always achieves the best accuracies which validates the necessity of utilizing multiple time windows as our research direction.

It is also worth noting that when evidence conflicts severely or completely, the D-S fusion of mass functions would cause ambiguity which would affect the classification accuracy. Nevertheless, this problem would be naturally resolved in the BCI context because of the design of paradigms. The subjects would only perform the cued MI tasks, and the possibility of occurring undesirable results would become negligible.

Although the RSODFS method has been evaluated and proved to outperform state-of-the-art methods, there still exists

leaving blank space in which could be further ameliorated. First, the Riemannian sparse optimization is realized by mapping the covariances points to the tangent space and vectorizing them, which may destruct the underlying structure of original data to some extent. This could be addressed through Riemannian dictionary learning [58], which treats atoms as covariance matrices to keep the intrinsic distribution of EEG data. In addition, the improvement of decoding performance benefits from the robustness of the introduction of Riemannian geometry. However, the succeeding surging of feature dimensionality also increases the computational costs and results in the inability of the sparsity hypermeter to be large, which should be considered in future studies. Moreover, when fusing the probabilistic outputs of the abovementioned one-versus-one SVM classifiers in a multi-class classification scenario, two-versus-two strategies could be together considered to improve the credibility of the final prediction.

V. Conclusion

In this study, we propose an integrated framework combining the robust Riemannian tangent space features extracted from multiple time-frequency bands with sparse optimization. Then the Dempster Shafer theory based fusion method is utilized to effectively fuse the transformed probabilistic values, which are obtained from trained linear kernel SVM of each divided time window to maximally exploit underlying soft information within and between multiple classifiers. Experimental results on the public dataset show that the proposed RSODSF achieves significantly higher classification accuracy compared with other state-of-the-art methods. In addition, our method is capable of extracting more discriminative features and solves the issue of ambiguity in determining the final class. In conclusion, the RSODSF is a promising candidate for improving the performance of MI-BCIs.

References

[1] J. R. 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, Jun. 2000.

[2] J. Jin et al., "The study of generic model set for reducing calibration time in P300-based brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 3–12, Nov. 2019.

[3] Z. Huang, J. Guo, W. Zheng, Y. Wu, Z. Lin, and H. Zheng, "A calibration-free approach to implementing P300-based brain–computer interface," *Cogn. Comput.*, vol. 14, no. 2, pp. 1–13, Mar. 2022.

[4] J. Jin, Z. Chen, R. Xu, Y. Miao, X. Wang, and T.-P. Jung, "Developing a novel tactile P300 brain-computer interface with a cheeks-stim paradigm," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 9, pp. 2585–2593, Sep. 2020.

[5] J. Decety, "The neurophysiological basis of motor imagery," *Behavioural Brain Res.*, vol. 77, nos. 1–2, pp. 45–52, May 1996.

[6] F. Pichiorri et al., "Brain–computer interface boosts motor imagery practice during stroke recovery," *Ann. Neurol.*, vol. 77, no. 5, pp. 851–865, 2015.

[7] D. Huang, K. Qian, D. Fei, W. Jia, X. Chen, and O. Bai, "Electroencephalography (EEG)-based brain–computer interface (BCI): A 2-D virtual wheelchair control based on event-related desynchronization/synchronization and state control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 3, pp. 379–388, May 2012.

[8] H. Altaheri et al., "Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: A review," *Neural Comput. Appl.*, pp. 1–42, Aug. 2021, doi: 10.1007/s00521-021-06352-5.

[9] C. Park, C. C. Took, and D. P. Mandic, "Augmented complex common spatial patterns for classification of noncircular EEG from motor imagery tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 1, pp. 1–10, Jan. 2013.

[10] I. Daly, R. Scherer, M. Billinger, and G. Müller-Putz, "FORCe: Fully online and automated artifact removal for brain–computer interfacing," *IEEE Trans. Neural. Syst. Rehabil. Eng.*, vol. 23, no. 5, pp. 725–736, Sep. 2014.

[11] J. Jin, C. Liu, I. M. Y. Daly, S. Li, X. Wang, and A. Cichocki, "Bispectrum-based channel selection for motor imagery based brain–computer interfacing," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 10, pp. 2153–2163, Oct. 2020.

[12] J. Su, Z. Yang, W. Yan, and W. Sun, "Electroencephalogram classification in motor-imagery brain–computer interface applications based on double-constraint nonnegative matrix factorization," *Physiol. Meas.*, vol. 41, no. 7, Aug. 2020, Art. no. 075007.

[13] J. Jin et al., "Optimization of model training based on iterative minimum covariance determinant in motor-imagery BCI," *Int. J. Neural Syst.*, vol. 31, no. 7, 2021, Art. no. 2150030.

[14] 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.

[15] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selelction in motor imagery based brain–computer interface," in *Proc. IEEE Eng. Med. Biol. 27th Annu. Conf.*, Jan. 2005, pp. 5392–5395.

[16] 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. 2020.

[17] J. Jin et al., "A novel classification framework using the graph representations of electroencephalogram for motor imagery based brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 20–29, 2021.

[18] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band common spatial pattern (SBCSP) for brain–computer interface," in *Proc. 3rd Int. IEEE/EMBS Conf. Neural Eng.*, May 2007, pp. 204–207.

[19] 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.

[20] G. Sun, J. Hu, and G. Wu, "A novel frequency band selection method for common spatial pattern in motor imagery based brain computer interface," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2010, pp. 1–6.

[21] M. Liu, D. Zhang, E. Adeli, and D. Shen, "Inherent structure-based multiview learning with multitemplate feature representation for Alzheimer's disease diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1473–1482, Jul. 2015.

[22] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing the channel selection and classification accuracy in EEG-based BCI," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1865–1873, Jun. 2011.

[23] Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Optimizing spatial patterns with sparse filter bands for motor-imagery based brain–computer interface," *J. Neurosci. Methods*, vol. 255, pp. 85–91, Nov. 2015.

[24] Y. Zhang et al., "Improving EEG decoding via clustering-based multitask feature learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 1–11, Feb. 2021.

[25] J. Feng et al., "Towards correlation-based time window selection method for motor imagery BCIs," *Neural Netw.*, vol. 102, pp. 87–95, Jun. 2018.

[26] P. Xu, T. Liu, R. Zhang, Y. Zhang, and D. Yao, "Using particle swarm to select frequency band and time interval for feature extraction of EEG based BCI," *Biomed. Signal Process. Control*, vol. 10, pp. 289–295, Mar. 2014.

[27] C. Zhu, B. Qin, F. Xiao, Z. Cao, and H. M. Pandey, "A fuzzy preference-based Dempster–Shafer evidence theory for decision fusion," *Inf. Sci.*, vol. 570, pp. 306–322, Sep. 2021.

[28] M. Arvaneh, C. Guan, K. Keng Ang, and C. Quek, "Optimizing spatial filters by minimizing within-class dissimilarities in electroencephalogram-based brain–computer interface," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 4, pp. 610–619, Apr. 2013.

[29] Y. Miao et al., "Learning common time-frequency-spatial patterns for motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 699–707, 2021.

[30] H. Fang, J. Jin, I. Daly, and X. Wang, "Feature extraction method based on filter banks and Riemannian tangent space in motor-imagery BCI," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 6, pp. 2504–2514, Jun. 2022.

[31] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Riemannian geometry applied to BCI classification," in *Proc. Int. Conf. Latent Variable Anal. Signal Separat.* Berlin, Germany: Springer, 2010, pp. 629–636.
[32] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications," *Neurocomputing*, vol. 112, pp. 172–178, Jul. 2013.
[33] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. London, U.K.: Chapman & Hall, 2018.
[34] H. L. Van Trees and K. L. Bell, *Covariance, Subspace, and Intrinsic CramrRao Bounds*. Hoboken, NJ, USA: Wiley, 2007.
[35] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *Int. J. Comput. Vis.*, vol. 66, no. 1, pp. 41–66, 2006.
[36] M. Hersche, T. Rellstab, P. D. Schiavone, L. Cavigelli, L. Benini, and A. Rahimi, "Fast and accurate multiclass inference for MI-BCIs using large multiscale temporal and spectral features," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2018, pp. 1690–1694.
[37] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on Riemannian manifolds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1713–1727, Oct. 2008.
[38] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2560–2574, 2007.
[39] S. Selim, M. M. Tantawi, H. A. Shedeed, and A. Badr, "A CSP\AM-BA-SVM approach for motor imagery BCI system," *IEEE Access*, vol. 6, pp. 49192–49208, 2018.
[40] M.-P. Hosseini, D. Pompili, K. Elisevich, and H. Soltanian-Zadeh, "Optimized deep learning for EEG big data and seizure prediction BCI via Internet of Things," *IEEE Trans. Big Data*, vol. 3, no. 4, pp. 392–404, Dec. 2017.
[41] A. Janani and M. Sasikala, "Classification of fNIRS signals for decoding right- and left-arm movement execution using SVM for BCI applications," in *Computational Signal Processing and Analysis*. Singapore: Springer, 2018, pp. 315–323.
[42] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Adv. Large Margin Classifiers*, vol. 10, no. 3, pp. 61–74, 1999.
[43] S. E. N. Fernandes, D. R. Pereira, C. C. O. Ramos, A. N. Souza, D. S. Gastaldello, and J. P. Papa, "A probabilistic optimum-path forest classifier for non-technical losses detection," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3226–3235, May 2018.
[44] H.-T. Lin, C.-J. Lin, and R. C. Weng, "A note on Platt's probabilistic outputs for support vector machines," *Mach. Learn.*, vol. 68, no. 3, pp. 267–276, 2007.
[45] A. P. Dempster, "The Dempster–Shafer calculus for statisticians," *Int. J. Approx. Reasoning*, vol. 48, no. 2, pp. 365–377, Jun. 2008.
[46] G. Shafer, "Dempster–Shafer theory," *Encyclopedia Artif. Intell.*, vol. 1, pp. 330–331, Jan. 1992.
[47] 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.
[48] C. Zuo, Y. Miao, X. Wang, L. Wu, and J. Jin, "Temporal frequency joint sparse optimization and fuzzy fusion for motor imagery-based brain–computer interfaces," *J. Neurosci. Methods*, vol. 340, Jul. 2020, Art. no. 108725.
[49] N. Kwak, "Principal component analysis based on L1-norm maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1672–1680, Sep. 2008.
[50] M. Tangermann et al., "Review of the BCI competition IV," *Frontiers Neurosci.*, vol. 6, p. 55, Jul. 2012.
[51] B. Blankertz et al., "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 153–159, Jun. 2006.
[52] 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 Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.
[53] A. K. Das, S. Sundaram, and N. Sundararajan, "A self-regulated interval type-2 neuro-fuzzy inference system for handling nonstationarities in EEG signals for BCI," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 6, pp. 1565–1577, Dec. 2016.
[54] A. Jafarifarmand, M. A. Badamchizadeh, S. Khanmohammadi, M. A. Nazari, and B. M. Tazehkand, "A new self-regulated neuro-fuzzy framework for classification of EEG signals in motor imagery BCI," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 3, pp. 1485–1497, Jun. 2017.
[55] T. Short, L. Alzapiedi, R. Brüschweiler, and D. Snyder, "A covariance NMR toolbox for MATLAB and OCTAVE," *J. Magn. Reson.*, vol. 209, no. 1, pp. 75–78, Mar. 2011.
[56] L. Van Der Maaten, "Accelerating t-SNE using tree-based algorithms," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3221–3245, Jan. 2014.
[57] A. C. Ramos, R. G. Hernandez, M. Vellasco, and P. Vellasco, "Ensemble of classifiers applied to motor imagery task classification for BCI applications," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 2995–3002.
[58] A. Cherian and S. Sra, "Riemannian dictionary learning and sparse coding for positive definite matrices," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 2859–2871, Dec. 2016.