Decoding multiclass motor imagery EEG from the same upper limb by combining Riemannian geometry features and partial least squares regression

Yaqi Chu1,2,*, Xingang Zhao1,6, Yijun Zou1,3, Weiliang Xu4, Guoli Song1,2, Jianda Han5 and Yiwen Zhao1,2

1 State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, People’s Republic of China
2 Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang, People’s Republic of China
3 University of Chinese Academy of Sciences (UCAS), Beijing, People’s Republic of China
4 Department of Mechanical Engineering, University of Auckland, Auckland, New Zealand
5 College of Artificial Intelligence, Nankai University, Tianjin, People’s Republic of China

*E-mail: zhaoxingang@sia.cn

Keywords: motor imagery EEG, same upper limb, Riemannian geometry features, partial least squares regression, brain-machine interface

Abstract

Objective. Due to low spatial resolution and poor signal-to-noise ratio of electroencephalogram (EEG), high accuracy classifications still suffer from lots of obstacles in the context of motor imagery (MI)-based brain-machine interface (BMI) systems. Particularly, it is extremely challenging to decode multiclass MI EEG from the same upper limb. This research proposes a novel feature learning approach to address the classification problem of 6-class MI tasks, including imaginary elbow flexion/extension, wrist supination/pronation, and hand close/open within the unilateral upper limb. Approach. Instead of the traditional common spatial pattern (CSP) or filter-bank CSP (FBCSP) manner, the Riemannian geometry (RG) framework involving Riemannian distance and Riemannian mean was directly adopted to extract tangent space (TS) features from spatial covariance matrices of the MI EEG trials. Subsequently, to reduce the dimensionality of the TS features, the algorithm of partial least squares regression was applied to obtain more separable and compact feature representations. Main results. The performance of the learned RG feature representations was validated by a linear discriminative analysis and support vector machine classifier, with an average accuracy of 80.50% and 79.70% on EEG dataset collected from 12 participants, respectively. Significance. These results demonstrate that compared with CSP and FBCSP features, the proposed approach can significantly increase the decoding accuracy for multiclass MI tasks from the same upper limb. This approach is promising and could potentially be applied in the context of MI-based BMI control of a robotic arm or a neural prosthesis for motor disabled patients with highly impaired upper limb.

1. Introduction

Brain-machine interface (BMI) technology based on electroencephalogram (EEG) is a promising tool that enables users to directly control or interact with external environment or devices by using their brain activity [1], especially for the people who have severe motor disorders related with neuromuscular injuries, such as spinal cord injury, amyotrophic lateral sclerosis, brain stroke and cerebral palsy [2].

Although people with paralysis exhibit different EEG or magnetoencephalography patterns between motor imagery (MI) and motor attempt, research evidence suggests that tetraplegic people can generate similar cortical activations during imagined movements as able-bodied people [3, 4]. Accordingly, applying MI-based BMI system as a rehabilitation intervention is efficient and suitable for tetraplegic people to restore neuromuscular connections and motor function by using endogenous EEG [3–5]. With the ability to provide an alternative interaction, some prominent MI-based BMI systems are developed and applied.
in assistive or rehabilitation devices control, such as neural prosthesis [6], functional electrical stimulation (FES) [7], exoskeleton [8], and robotic arm [9]. However, most of these BMI systems are in the stage of laboratory testing, which cannot be widely used in practical applications.

For current MI-based BMI systems, decoding different MI tasks form the same limb is extremely challenging mainly due to the limitations of low spatial resolution and poor signal-to-noise ratio (SNR) of the EEG signals [10–13]. Particularly, most of existing MI-based BMI systems can only decode spatially well separated imaginary movements from different limbs with the traditional spatial filter method, which results in a limited number of control commands. Generally, left/right hand, and foot movements are the most commonly adopted MI tasks during the BMI systems [6–11, 14]. Moreover, these BMI systems only enable users to control low degree-of-freedom (DOF) devices (generally 1 DOF), such as orthosis or FES for hand closing/opening [7], neuroprosthesis for hand or elbow movements [6]. To obtain more control commands, decoding additional information about the same limb is necessary for the MI-based BMI systems, including wrist, elbow and hand imaginary movements. However, for the fact that different MI tasks from the same limb activate and share nearby regions with high spatial proximity on the cortical motor area [12, 13], it is hard to decode these MI tasks with high accuracy by the existing spatial filter method.

At present, there is relatively few research that have been carried out to tackle this challenge. Navarro et al extracted time, frequency and independent component analysis (ICA) based features to decode 4-class imaginary wrist movements, namely wrist flexion/extension and pronation/supination, with a low average accuracy of 34% [15]. Furtherly, Vuckovic and Sepulveda adopted an Elman neural network with time-frequency features extracted by Gabor transform method to classify six pairs of wrist movements (e.g. flexion against extension, flexion against pronation or flexion against supination, etc.), with 71.3% average accuracy ranging from 60% to 80% for each pair [16]. Recently, Eldman et al combined time-frequency features with extended EEG source imaging method to predict the same 4-class imaginary wrist movements, achieved with remarkable average accuracy of 81.4% [17]. Additionally, a few studies for decoding non-wrist MI tasks from the same limb were also investigated. Liao et al used power spectral features with a support vector machine (SVM) classifier to decode ten pairs of finger movements within the same hand (e.g. index versus thumb, index versus middle, etc.), with acceptable accuracy of 77.1% by using high density 128-channel EEG [18]. Yong et al conducted a preliminary study for classifying rest against 2-class MI tasks within the same limb (hand grasp and elbow movements), using three state-of-the-art feature extraction methods, i.e. common spatial pattern (CSP), filter-bank CSP (FBCSP), and band power [19]. The average accuracy for 3-class classification was 58.4%, while 2-class MI accuracy (grasp against elbow) was 63.9%. In a subsequent study, Tavakolan et al proposed the use of time-domain features to classify these 3-class tasks, including autoregressive model coefficients, waveform length, and root mean square. The average accuracy of 74.2% for 12 healthy participants was obtained by using a multiclass SVM with optimal parameters [20]. Despite the success of those studies, decoding accuracy was relatively low for 3-class [19, 20] or 4-class [15, 17] MI tasks from the same limb. Furthermore, the classification performance highly depended on EEG signal processing, especially advanced feature learning methods [21, 22].

Recently, as a novel learning tool for feature representation, Riemannian geometry (RG) approach has been rapidly applied in signal decoding and image processing [23–27]. Due to its ability for direct manipulation of covariance matrices from the EEG signal, the RG was successfully adopted into EEG-based BMI fields [28] and demonstrated its superiority in many applications, such as sleep/respiratory state classification [29], EEG pattern decoding [30, 31] and regression problems [26]. Particularly, Barachant et al first introduced the RG for MI EEG classification and proposed two approaches for feature learning [27, 32]. The first approach, named Minimum Distance to Riemannian Mean (MDRM), is to apply RG to directly identify EEG covariance matrices in the manifold space of symmetric and positive definite (SPD) matrices [33, 34]. The second approach is to map the covariance matrices into Riemannian tangent space (TS) and perform TS feature extraction and classification. The RG approach was tested on dataset Ila from the BCI competition IV, which comprised of left/right hand, foot and tongue MI tasks. The classification results showed that it outperformed the CSP algorithm for feature extraction. Nevertheless, there is no research for decoding MI tasks from the same upper limb by using the RG at present.

In this work, we proposed a novel feature learning approach composed of RG feature extraction and partial least squares (PLS) regression for feature reduction to address the classification problem of 6-class MI tasks from the same upper limbs, including imaginary elbow flexion/extension, wrist supination/pronation, and hand close/open. Firstly, the Riemannian TS features were extracted for the covariance matrices of each MI EEG using the RG framework. Due to the fact that the TS features become less efficient with the increase of covariance matrices dimension [27, 31], we adopted a PLS algorithm to
perform supervised feature reduction for the original TS features by maximizing the classification accuracy of the classifier. The performance of the proposed approach was respectively validated by a linear discriminative analysis (LDA) and SVM classifier on MI EEG dataset collected from 12 participants. Additionally, the comparative study was also conducted between the proposed approach and other well-known feature extraction methods such as CSP and FBCSP, which were also followed by the LDA and SVM classifier for MI EEG classification.

The major innovations and contributions of this paper are four aspects: (1) An advanced method combined RG framework with PLS was proposed to learn new feature representations for decoding multiclass MI EEG signals from the same upper limb. To our knowledge, this is the first study that the RG framework has been adopted in decoding multiple imaginary movements related to the same limb. (2) Focused on the difference analysis between the TS features extracted by RG framework and spatial features from CSP or FBCSP method, we conducted a range of comparative studies. The results showed that the TS features were more separable than the CSP or FBCSP spatial features for discriminating different imaginary movements within the same limb, especially for those movements with adjacent cortical activation areas, such as supination movement and pronation movement of the same wrist. (3) To effectively alleviate the overfitting and bias problem caused by the TS features with high-dimensional space [31], a supervised dimensionality reduction algorithm (PLS) was used to efficiently identify more separable and compact feature representations from the original TS features. Thereby, higher decoding accuracy can be obtained for multiclass MI EEG signals. (4) Compared with the traditional MI-based BMI systems, which usually only decode left/right hand and foot MI tasks, the proposed method provides an alternative to EEG decoding with higher accuracy for the multiclass MI-based BMI systems. These multiclass MI-based BMI systems can increase the number of control commands and provide additional control for multi-DOF devices.

The remainder of this paper is organized as the following parts: section 2 briefly introduces the decoding pipeline for MI tasks from the same upper limb. Section 3 describes the experimental paradigm and EEG dataset for multiclass MI tasks from the same upper limb. Section 4 introduces the basic concepts of the RG framework and TS feature extraction algorithm for MI-based BMI classification problem. Section 5 gives the PLS method for feature dimension reduction and the multiclass classifiers. Section 6 presents the evaluations and discussions for the comparative results. Finally, section 7 provides conclusions and outlines promising study directions for multiclass MI EEG classification.

2. Proposed decoding pipeline

The main objective of this study is to utilize the strength of the RG framework to decode multiclass MI tasks from the same upper limb. The block diagram of the proposed decoding pipeline is illustrated in figure 1, which comprises of three steps: EEG signal preprocessing, RG feature extraction and feature reduction, and MI classification. The first step was the preprocessing stage, which can enhance the SNR and temporal-spatial resolution of the EEG signal by combining band filtering, ICA and sliding window segmentation. The second step involved RG feature extraction and feature reduction. The preprocessed EEG signals were transformed as a series of covariance matrices, which were SPD matrices and contained spatial information relative to diverse MI tasks. By manipulating these EEG covariance matrices, the Riemannian distance and Riemannian mean were estimated by the RG framework. Then, vector-based features were obtained by utilizing TS mapping of these covariance matrices. To overcome high dimensionality problem in the native TS features, a supervised feature reduction method (PLS) was used to obtain better separable feature representation. The effectiveness of the PLS was also compared with an un-supervised feature reduction method (principal component analysis, PCA). Finally, a vector-based classifier was used to identify MI tasks from these RG features, such as SVM and LDA.

3. Multiclass motor imagery EEG experiments

3.1. Participants recruitment

Twelve able-bodied volunteers with no history of neuromuscular disorders or mental diseases were recruited to participate the study, including nine males and three females (age range: 25 ± 3 years, named P01-P12 respectively). All the participants except P08 were right-handed. Certainly, all the volunteers have never experienced such MI EEG study. Prior to the study, the detailed experimental procedures were clearly explained to the participants and written consent form was signed for each volunteer. Moreover, to be familiar with the selected different MI tasks from the same upper limb, the participants also conducted extensive exercise for these MI tasks. The whole experimental procedure was carried out in accordance with the declaration of Helsinki and ethically permission was provided by the local ethics committee of the University of Chinese Academy of Sciences, China.

3.2. Experimental paradigm design

Inspired by [35], we respectively selected three pairs of motions from the right upper limb, which comprised of flexion/extension for elbow joint, supin-
Figure 1. The block diagram of the proposed feature learning approach for decoding multiclass MI EEG based on Riemannian geometry.

Figure 2. Motor imaginary tasks and experimental paradigm setup. (A) Experimental situation (left side): participant sat in a chair with a neutral posture and performed MI tasks according to cues presented on a screen. Six different imaginary motions from the right upper limb (right side): elbow joint flexion/extension, wrist joint supination/pronation, right hand close/open. (B) The experimental time sequences in one session, including a total of 36 trials repeated in random order. Each MI task was repeated six times. In every trial, one of six motion cues was randomly chosen and presented in the MI period.

Each experimental session consisted of 36 trials (6 trials for each MI task) and lasted approximately 30 min. Figure 2(B) illustrates the time sequence of the experimental paradigm in one session. Each trial with the duration of 8 s was divided in three phases by prompt information. The first phase was a relax/prepare period, which corresponded to a 2 s black screen. In this period, the participant should maintain neutral intention without any MI. After that, at the center of the screen, a green cross with the duration of 1 s was emerged to remind the participant to prepare for imagining movements. Immediately, the third phase was followed, which was motor imaginary period. In this period, one of six different motion cues was randomly chosen and displayed on the screen for 5 s, during which the volunteers executed sustained MI task with their right upper limb, such as imaginary sustained kinesthetic elbow flexion from the neutral posture at a repetition rate of 2 s/movement cycle, followed by elbow extension as quickly as possible. Each MI task was repeated six times and randomly presented in one session. Meanwhile, in order to reduce the exogenous noises, the volunteers should try not to blink eyes and swallow or move their body during EEG recording. At the end of each trial, the volunteers relaxed to prepare the next MI trial. In each session, the 36 trials of MI tasks were presented in random order to reduce the subject’s adaptability. For each participant in the whole MI experiment, a total of 360 EEG trials (36 trials/session × 10 sessions) with length 5 s were recorded, corresponding to 60 trials per MI class.
3.3. EEG data recording
During this study, non-invasive EEG data were recorded by a customized 64-channel grid cap with Ag/AgCl active scalp electrodes (Greentek Ltd., China), which were arranged according to the international 10-10 recording system as Fp1, Fpz, Fp2, AF7, AF3, AFz, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T9, T7, C5, C3, C1, Cz, C2, C4, C6, T8, T10, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO6, O1, Oz, O2, and Iz. The electrode was wet using conductive gel respectively and the cap was fixed by a chinstrap to avoid electrode deviation. The right and left mastoid electrodes were regarded as ground and reference channels. The impedance of 64 electrodes was kept below 30 kΩ. The EEG signals were sampled at 1 kHz and digitized using the OmnimPlex Neural Recording Data Acquisition System (Plexon Inc., USA), including front-end amplification, low-pass filtering and 24-bit A/D conversion. Furthermore, a 60 Hz notch filter was applied to reduce the power line interference. The recorded EEG data were transmitted to a computer across TCP/IP network using PlexNet (Plexon Inc., USA) and stored as 64 × 5000 × 360 (a format of Channel × Time × Trial). To obtain separable MI pattern, a 4 s period from 0.5 s to 4.5 s after a cue appeared was split out for each trial. Then, the size of MI EEG datasets for each participant was 64 × 4000 × 360.

3.4. EEG data preprocessing
Commonly, the recorded scalp EEG signals contain many noises and artifacts, such as eye blinks or movements, head motion, and muscle activity caused by face, neck and shoulder motions [36]. In order to improve the SNR of the MI EEG signal, the preprocessing pipeline was used, specifically consisting of (a) band-pass filtering, (b) ICA, and (c) sliding window segmentation.

3.4.1. Band-pass filtering.
During the MI period, the event-related synchronization and desynchronization (ERS/ERD) phenomenon mainly emerged in the frequency band 8–30 Hz, which encompassed both the mu (8–13 Hz) and beta (14–28 Hz) rhythms [37]. Generally, a band-pass filter was used to highlight interested frequency component while attenuate some artifacts. In our study, a zero-phase Butterworth filter (fifth-order, gain 2) was designed with the cutoff frequency of 8 and 35 Hz. This band-pass filter can eliminate a large part of muscle activity, such as EMG (higher than 35 Hz) resulting from swallow or face movements. Additionally, it can reduce some ocular artifacts, which mainly affect the low frequency components of the MI EEG signal, such as EOG from eye blinks and movements (lower than 8 Hz). On the other hand, the baseline shift of the EEG signal caused by head or limb motions can also be eliminated using the designed Butterworth band-pass filter.

3.4.2. Independent component analysis (ICA).
A kind of Blind Source Separation (BSS) algorithm (ICA) was used to furtherly remove the EMG and EOG artifacts between 8 and 30 Hz. Moreover, the ICA method was also applied to separate and remove the ECG artifacts caused by heart-beat. In this study, FastICA toolbox as an EEGLAB plug-in [38] was executed for the band-pass filtered MI EEG. The MI EEG trials were firstly decomposed into 32 independent components (ICs) by applying high-order statistics. In total, 4, 2, and 4 ICs corresponding to EMG, EOG, and ECG artifacts were respectively selected and rejected by visual inspection. Finally, the remaining ICs were mixed and projected back onto the original channel space to construct EMG/EOG/ECG-free MI EEG trials. Notably, after the ICA procedure, the recorded MI trials which still contained noises would be excluded. By checking the EEG signal processed by the ICA, no MI trial was discarded in this study.

3.4.3. Sliding window segmentation.
To improve the temporal resolution of the EEG trial and increase the dataset size, a sliding window was usually utilized to divide the MI EEG trial into overlapping or non-overlapping segments [11, 39]. Additionally, for the CSP or CSP-like method to obtain stable spatial filters, the length of sliding window should not be selected too small or too big [39]. In this paper, for a trial of MI EEG signal with 4 s, a 1-second sliding window with 80% overlap was adopted to divide into 16 segments with the length of 1 s [39]. Then, for each participant, the MI EEG dataset was extended to 64 × 1000 × 5760, where 5760 was the trial number (960 trials per MI class). The subsequent TS features were extracted on those MI EEG dataset by using RG method in this paper. Additionally, the CSP and FBCSP methods were also used to extract spatial features on the same MI EEG dataset, respectively.

4. Riemannian geometry features
In MI-based BMI literature, the classical CSP and filter-bank CSP (FBCSP) were widely used for feature extraction, which aim to extract spatial features by maximizing the variance of MI trials for one class and minimizing that for the other class simultaneously [40, 41]. The spatial features extracted by CSP or FBCSP depend on the selection of the frequency bands. Generally, a wide band 8–30 Hz is adopted for CSP or FBCSP, where the ERD/ERS for MI EEG is prominent. To enhance the separability of MI features, the CSP and FBCSP need to learn optimal spatial filter from the covariance matrices of EEG trials [40, 41]. As mentioned in [27, 32], the RG provides
alternative way to extract features by directly manipulating those covariance matrices without the need for spatial filter. In this section, the basic concepts and tools of RG applied in MI-based BMI are briefly introduced. The algorithms for MDRM classifier and TS feature extraction are also presented.

4.1. Riemannian geometry (RG) framework

Let a matrix $X_i \in \mathbb{R}^{N \times T_i}$ represents a short-time segment of the EEG signal, which corresponds to the $i$th trial of MI tasks. Here, $N$ denotes the number of channels and $T_i$ is the number of sampled time points. Note that the MI EEG trial $X_i$ was time-centered with a zero mean [27, 28]. The spatial covariance matrix for each MI EEG trial can be estimated and denoted by the unbiased sample covariance matrix (SCM) $C_i \in \mathbb{R}^{N \times N}$, which belongs to a SPD matrix. The SCM can be computed as:

$$C_i = \frac{1}{T_i-1}X_iX_i^T,$$

where the superscript $T$ represents the transposed operator. Then, the space of these SPD matrices forms a differentiable Riemannian manifold $\mathcal{M}$ with the dimensionality of $N(N+1)/2$ (illuminated in figure 3). In addition, the derivatives of each point $C$ on this manifold form a TS $T_C \mathcal{M}$, which is constituted by a set of tangent vector $S_i$ [27, 31].

Instead of the Euclidean metric, the Riemannian metric was adopted in this space of the SPD matrices, particularly Riemannian distance and Riemannian mean. For the two spatial covariance matrices $C \in \mathbb{R}^{N \times N}$ and $C_i \in \mathbb{R}^{N \times N}$ on the manifold, the Riemannian distance $\delta_R(C, C_i)$ is given by

$$\delta_R(C, C_i) = \left\| \log(C^{-1}C_i) \right\|_F = \sqrt{\sum_{n=1}^{N} \log^2 \lambda_n},$$

where $\log(\bullet)$ is the logarithmic operator, $\left\| \bullet \right\|_F$ is the Frobenius norm operator, and $\{\lambda_n\}_{n=1}^N$ are the real eigenvalues of $C^{-1}C_i$. This distance is the length of the shortest curve joining two covariance matrices on the manifold, also namely the geodesic distance. Because the TS $T_C \mathcal{M}$ at $C$ is Euclidean and locally homomorphic to the manifold $\mathcal{M}$ [27, 31, 32], the Riemannian distance can be equivalently approximated by a Euclidean distance between tangent vectors in $T_C \mathcal{M}$, such as

$$\delta_R(C, C_i) = \left\| S_i \right\|_F = \left\| \log_C(C_i) \right\|_F = \left\| \text{upper}(C^{-1} \log_C(C_i)C^{-\frac{1}{2}}) \right\|_2,$$

where $S_i$ is the mapped tangent vector of $C_i \in \mathbb{R}^{N \times N}$ in the TS $T_C \mathcal{M}$ at $C$ by using logarithmic mapping operator $\log_C(\bullet)$. The $\text{upper}(\bullet)$ operator is used to reserve the upper triangular elements of a symmetric matrix and synchronize vectorization by adding weight $\sqrt{2}$ for the off-diagonal elements and weight 1 for the diagonal elements [28, 42]. The logarithmic mapping operator is defined as

$$S_i = \log_C(C_i) = C^\frac{1}{2} \log(C^{-\frac{1}{2}}C_iC^{-\frac{1}{2}})C^\frac{1}{2}.$$
where \( \exp(\bullet) \) denotes the exponential operator.

For a set of \( I \geq 1 \) covariance matrices, the Riemannian mean or geometric mean is obtained by solving the minimization problem of sum of squared Riemannian distances:

\[
\Theta(C_1, C_2, \ldots, C_I) = \arg\min_C \sum_{i=1}^{I} \delta^2_S(C, C_i). \tag{6}
\]

An iterative optimization algorithm can be applied to estimate the Riemannian mean (see Algorithm 1) [26, 32].

### 4.2. Minimum Distance to Riemannian Mean (MDRM)

Inspired from the KNN classifier in Euclidean space, a simple and efficient supervised classifier can be derived by the RG framework, namely MDRM [27, 28, 31]. For a set of MI EEG trials with known class, the intra-class Riemannian mean \( \Theta^{(s)} \) for each MI task can be obtained, where \( \{ k \}_{k=1}^{K} \) represents the MI class. For each new MI EEG trial, the Riemannian distance between this unknown SCM and each intra-class Riemannian mean is computed, respectively. Then, the minimum distance criterion is used to determine which class the new trial belongs to. Formally, more details of the MDRM classifier are summarized in Algorithm 2.

### 4.3. Riemannian tangent space (TS) features

Although the MDRM classifier can give equivalent performance when compared to the reference method (CSP with LDA classifier) in traditional MI-based BMI classification [27], the SPD matrices (raw covariance matrices for EEG trials) cannot be directly fed to the vector-based classifier, such as LDA or SVM in BMI fields. Based on the concept of the TS in the RG framework, the SPD matrices can be mapped as a vector-based feature in the TS. For a set of covariance matrices for all EEG trials \( C_i, i = 1, \ldots, I \), the overall Riemannian mean \( \Theta \) for all classes is first computed. Then, each covariance matrix \( C_i \) is projected as tangent vector \( S_i \) on the local TS around the point \( \Theta \) by the following formula:

\[
S_i = \text{upper}(\Theta^{-\frac{1}{2}} \log(S_i)(\Theta^{-\frac{1}{2}})). \tag{7}
\]

It should be noted that the \( \text{upper}(\bullet) \) operator is applied to keep the Riemannian distance between \( \Theta \) and \( C_i \) equal to the Euclidean norm. These tangent vectors form the set of Riemannian TS features with the dimensionality of \( N(N + 1)/2 \). The specific procedure for TS feature extraction is presented in Algorithm 3. In this study, the dimension of extracted Riemannian TS features was 2080 for each MI trial with \( N = 64 \) channels.

### 5. Feature dimension reduction and classification

As the TS is a high-dimensional space, the dimensionality of the extracted TS features might exceed the number of the MI EEG trials, which can result in the problem of 'curse of dimensionality' (2080 features \( \times 960 \) samples \( \times 6 \) classes) in BMI classification. Hence, it is usually needed to conduct a feature reduction procedure for these TS features before training a classifier, in order to avoid overfitting and improving the generalization ability of the classifier. Accordingly, more separable feature representations can be obtained to decode multiclass MI EEG signals.

In this study, a supervised feature reduction method (PLS regression algorithm) was adopted. PLS is a kind of multivariate statistical technique based on latent variable (LV), which explores the covariation between predictor variables and target variables by finding a new set of LVs with maximal correlation [43]. In this study, the extracted TS features...
$S = [S_1, S_2, \ldots, S_I]^T \in \mathbb{R}^{I \times P}$ represented the predictor variables, where $I$ was the number of the MI EEG trials and $P = N(N + 1)/2$ was the number of the TS features ($N=64$ was the number of channels). Correspondingly, the class labels for MI EEG trials represented the target variables, denoted by $Y = [Y_1, Y_2, \ldots, Y_I]^T \in \mathbb{R}^{I \times K}$, where $K = 6$ was the number of the MI classes. Based on the idea of maximizing the LVs between TS features and MI class labels, the mathematical prototype of PLS can be formulated by

$$
\text{argmax} \langle Sw_1, Yw_2 \rangle
$$

$$
s.t. \ |w_1| = 1, |w_2| = 1,
$$
where $w_i (i = 1, 2)$ are the weight vectors for extracting the LVs of $S$ and $Y$, respectively. A typical solution of PLS can be performed by using singular value decomposition (SVD) algorithm [43, 44]:

$$
Y^T S = W_2 \Lambda W_1^T,
$$
where $\Lambda$ is a diagonal matrix with the singular values of $Y^T S$. Generally, according to the first $m \ (m < P$, $P$ was the number of the original TS features before PLS reduction) singular values of the $\Lambda$, thempairs of right singular vectors ($w_{1i}, i = 1, \ldots, m$) and left singular vectors ($w_{2i}, i = 1, \ldots, m$) were selected to construct the transform matrix $W_1 \in \mathbb{R}^{P \times m}$ and $W_2 \in \mathbb{R}^{I \times m}$, respectively [44]. Then, a set of LVs for TS features and target labels can be reconstructed by

$$
L_S = SW_i,
$$
$$
L_Y = YW_2,
$$
where $L_S \in \mathbb{R}^{I \times m}$ represents a set of the LVs for the predictor variables, that is the reduced feature representations for the original TS features $S$. The column vectors of $L_S$ were also called the PLS components. $L_Y$ represents a set of the LVs for the target labels. Finally, the reconstructed $L_S$ was used as new predictors for classification in the next classifier training step. Additionally, since the new predictors $L_S$ need to be combined with a vector-based classifier to obtain the target class labels, the column dimension of the new predictors $L_S$, that is the number $m$ of PLS components, can be optimized by maximizing the classification accuracy using 5-fold cross-validation method in the training phase [44]. Notably, before the implementation of PLS, the predictor variables (original TS features) should be normalized by z-scores that subtracting mean and dividing by the standard deviation. Furthermore, to validate the effectiveness of PLS in feature reduction, an un-supervised method (PCA) was also conducted on those original TS features $S$ by comparing the classification accuracy and the feature dimension after reduction. For each participant, the feature dimension after PLS or PCA reduction was obtained by selecting an optimal number of PLS components or principal components, respectively.

To compare the performance of different decoders with the proposed Riemannian features, two classifiers widely used in MI-based BMI system were employed, namely LDA and SVM [21]. Since LDA or SVM is an intrinsic binary classifier, a pairwise strategy is necessary to decode the multiple classes, such as one-versus-one or one-versus-rest [21, 45, 46]. To implement 6-class MI classification using LDA or SVM, we adopted a one-versus-rest strategy yielding six binary classifiers. More specifically, every binary classifier were trained considering the samples from a given MI class as positive labels and all other samples from the rest MI classes as negative labels. The six binary classifiers were combined to construct a multi-class classifier to predict the test samples by using a maximum confidence strategy, where the output of the classifier with the largest confidence was taken as the final class for the test samples [45, 46]. In this study, to evaluate the performance of the classifiers, a repeated 5-fold cross-validation procedure was employed. Thus, the whole dataset was randomized and divided into 5-folds of equal size, where 4 folds were used for training the classifier (training set) and the remaining fold was used for testing the classifier (test set). Each fold had the same proportion of samples for each MI class as in the original dataset as a whole. This procedure was repeated five times to guarantee that each fold was used once as the test set. In addition, to alleviate the imbalanced class problem of each binary classifier in a one-versus-rest strategy, class weighting method was also applied in the cost function of LDA or SVM to boost predictive performance of the minority class [45, 46]. Criteria such as decoding accuracy, true positive rate (TPR) or sensitivity, false positive rate (FPR) were used to validate the performance of the classifier, which were defined by

$$
\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \times 100 \%,
$$

Algorithm 3. The Riemannian TS feature extraction.

| Input: | a set of $I$ covariance matrices $C_i$ for all MI trials. |
| Output: | a set of $I$ TS feature vectors $S_i$. |
| 1: Compute $\Theta = \Theta(C_1, C_2, \ldots, C_I)$ by (6), the Riemannian mean for all MI classes |
| 2: for $i = 1$ to $I$ do |
| 3: $S_i = \text{upper}(\Theta^{-1} \text{Log}_{\Theta}(C_i) \Theta^{-1})$ by (7). |
| 4: end for |
| 6: Return $S_i$ |
where $TP$ is the number of true positive MI trials, $FP$ is the number of false positive MI trials, $TN$ is the number of true negative MI trials, $FN$ is the number of false negative MI trials. Accuracy is a rate of the number of correctly predicted MI trials relative to the total number of targeted MI trials. The TPR is the percentage of positive MI trials that are predicted as positive. The FPR is the percentage of false positives predicted as positive from negative MI class.

To study the effect of the proposed method on decoding accuracy, one-way Analysis of Variance (ANOVA) was carried out. When a significant difference ($p < 0.05$) was observed, multiple pairwise comparisons were applied to perform post-hoc t-test analysis by using Bonferroni correction method. The significance level of all post-hoc comparisons was below 0.05 after correction. For a 6-class EEG classification problem, the probabilistic chance level is given by $100/6 = 16.67\%$, which is valid only for an infinite number of MI trials. However, for the finite MI trials, the theoretical chance level in statistical terms is crucial for evaluating the decoding performance, which can be obtained by using binomial cumulative distribution (analytical method) or permutation tests (empirical method) [47]. The significant chance level of the analytical method is 17.48\%, with the Matlab (Mathworks Inc. MA, USA) function `binoinv` by

\[ \text{chance level} = \text{binoinv}(1 - \alpha, n, 1/c) \times 100/n, \]

where $\alpha = 0.05$ is the confidence level, $n = 5760$ is the total number of samples, $c = 6$ is the given number of classes. Additionally, by using permutation tests (a bootstrapping method) which performs 1000 random permutations of the labels (classes) in the raw samples at the start of the analysis and computes the decoding accuracy for each permutation, the obtained significant chance level is 18.10\% with the 95 percentile of empirical distribution. Eventually, for the 6-class EEG classification problem, the significant chance level for decoding accuracy was determined as 18.10\% in this study.

6. Results evaluations and discussions

6.1. Topographical and ERD/ERS analysis

The findings of previous researches have revealed that differences in the EEG neural correlates of diverse movements from the same hand/arm [17, 19, 20, 35, 48, 49], or different grasp types [30]. Particularly, the pioneering work of Ofner et al have proved that single imagined or attempted movements from the same upper limb (e.g. elbow flexion/extension, wrist supination/pronation, and hand close/open) could be decoded based on the low-frequency time-domain EEG signals, which can be captured by discriminative information of motor-related cortical potentials (MRCPs) [35, 48]. Here, from the perspective of the frequency-domain and spatial-domain of the EEG signals, we conducted topographical and ERD/ERS pattern analysis to further investigate the separability of those imagined movements of the same upper limb.

It is well known that the MI period is usually accompanied by the spectral power changes of the EEG in the mu (8–13 Hz) and beta (14–28 Hz) rhythm over the sensorimotor cortex, especially decrease (ERD) in the contra-lateral region and increase (ERS) in the ipsi-lateral region [37]. To illustrate the topographical distribution of the EEG signal on the scalp, the power values for the mu and beta frequency bands at each electrode are calculated for all participants. The spatial distribution maps of the EEG power for different MI tasks from the same limb are presented in figure 4(A). Generally, the analysis of the EEG activity shows obvious ERD/ERS in the motor-related frequency bands. As shown in figure 4(A), we can see that the spatial distribution of the EEG power is difference during six different MI tasks from the same limb. From the view of different joint parts of the same limb, for elbow flexion/extension imagined movements and hand close/open imagined movements, larger decrease of the EEG power (ERD) in the mu and beta rhythm is observed in the contra-lateral region (especially on the sensorimotor cortical area) when comparing wrist supination/pronation imagined movements. And for hand close/open imagined movements, the increase of the EEG power (ERS) in the ipsi-lateral region is not obvious compared with elbow flexion/extension imagined movements and wrist supination/pronation imagined movements. This indicates that the elbow and hand MI tasks of the same limb show a more bilateral and stronger activation than wrist MI tasks. Notably, the EEG power change of wrist supination/pronation imagined movements are characterized by relative low spatial resolution in the sensorimotor cortex, which is in line with [17]. Moreover, we have found as in [17, 20, 35, 48] that although the sensorimotor cortex is the most important cortical area related to hand/arm or elbow movements, the posterior parietal and frontal or occipital cortical areas are also involved in these motor control. Furthermore, the spatial differences of the EEG power for opposite imagined movements belonging to the same joint are also observable. For example, the EEG power decrease for the elbow extension MI task is more pronounced than that for the elbow flexion MI task, especially for the mu rhythm in the contra-lateral motor cortex area. The prominent decrease of the EEG power for hand open MI task occupies larger contra-lateral regions than that for hand close MI.
Figure 4. Spatial distribution of the EEG power and ERD/ERS analysis for different MI tasks from the same limb. (A) The topographical distribution of the EEG power in the mu rhythm (8–13 Hz) and beta rhythm (14–28 Hz). Black dots denote the 64 channels. (B) ERD/ERS time course for the mu rhythm (8–13 Hz) and beta rhythm (14–28 Hz) at the C3 channel. The vertical dashed line shows the time when the visual cue was presented to the participant.

6.2. Multiclass EEG decoding results comparison

To evaluate the classification performance of these 6-class MI tasks from the same upper limb, we firstly compared six decoding methods, respectively the proposed (TS-PLS) + LDA method (TS features with PLS for reduction and LDA for classification), the (TS-PCA) + LDA method (TS features with PCA for reduction and LDA for classification), the TS + LDA (TS features without any reduction followed by LDA for classification), the MDRM method, and two standard baseline methods adopted in traditional MI-based BMI classification (CSP + LDA and FBCSP + LDA) [19, 40, 41]. For the CSP + LDA method, the number of extracted spatial features was selected to 12 using the CSP algorithm with 5-fold cross-validation. For the FBCSP + LDA method, the number of extracted spatial features was set to 8 using the FBCSP algorithm, which was an extension of the CSP. The FBCSP algorithm was performed by three progressive stages: (1) A filter bank with type II Chebyshev bandpass filters was used to filter the EEG into six frequency bands, corresponding to 8–12 Hz, 12–16 Hz, 16–20 Hz, 20–24 Hz, 24–28 Hz, and 28–32 Hz respectively. (2) For each EEG frequency band, four spatial filters were extracted using the CSP algorithm. A total of 24 spatial filters can be obtained. (3) Then, a mutual information-based
Figure 5. The decoding accuracies for each participant by applying CSP + LDA, FBCSP + LDA, MDRM, TS + LDA, (TS-PCA) + LDA, and the proposed method (TS-PLS) + LDA. The probabilistic chance level is 16.67% (horizontal black solid line) and the significant chance level using a bootstrapping method is 18.10% (horizontal red dashed line) for the 6-class EEG classification.

Figure 6. The average, minimum and maximum decoding performance for each method across 12 participants. (A) The average decoding accuracies with the 95% confidence interval. The error bar represents the lower and upper bounds to the average accuracies. The asterisk notation denotes the significant difference ($p < 0.05$) between the methods. (B) The minimum and maximum decoding accuracies.

feature selection with eight components was applied, resulting in eight spatial features for the FBCSP [19, 41].

Figure 5 presents the decoding accuracies of six MI tasks from the same upper limb for each participant by using six methods, respectively. From this figure, we can observe that the proposed approach (TS-PLS) + LDA performed the best decoding accuracy for each participant, with maximum accuracy of 86.34% (P11) and minimum accuracy of 75.30% (P08). Moreover, for all participants except participant P11, the accuracies of the MDRM, TS + LDA, (TS-PCA) + LDA, and (TS-PLS) + LDA methods were higher than those of the standard baseline methods (CSP + LDA and FBCSP + LDA). And for the case of P11, the use of PCA and PLS feature reduction for TS features can dramatically increase the decoding accuracy by 11.25% and 16.08%, respectively. In general analysis, we also conducted 500 bootstrap iterations on the MI dataset for each method. The
Table 1. Statistical decoding performances (%) for six different MI tasks by FBCSP + LDA, TS + LDA and (TS-PLS) + LDA.

| MI Class | FBCSP + LDA | TS + LDA | (TS-PLS) + LDA |
|----------|-------------|----------|----------------|
| Flexion  | TPR 69.25  | FPR 5.57 | TPR 72.40  | FPR 4.84 | TPR 82.09  | FPR 3.59 |
| Extension| TPR 72.50  | FPR 6.11 | TPR 74.00  | FPR 5.56 | TPR 83.70  | FPR 3.91 |
| Supination| TPR 56.70 | FPR 8.27 | TPR 66.15  | FPR 6.77 | TPR 78.70  | FPR 4.55 |
| Pronation| TPR 60.05  | FPR 8.81 | TPR 67.00  | FPR 7.19 | TPR 78.40  | FPR 4.45 |
| Close    | TPR 68.55  | FPR 6.35 | TPR 70.85  | FPR 5.63 | TPR 80.60  | FPR 3.56 |
| Open     | TPR 70.20  | FPR 5.44 | TPR 71.50  | FPR 5.54 | TPR 79.45  | FPR 3.35 |
| Average  | TPR 66.21  | FPR 6.76 | TPR 70.31  | FPR 5.94 | TPR 80.49  | FPR 3.90 |

average (mean ± standard deviation) with the 95% confidence interval, minimum and maximum decoding performance for each method in the group-level of twelve participants are shown in figure 6. Since the performance of each method in term of decoding accuracy satisfied the normal distribution and the homoscedasticity, a parametric test (one-way ANOVA) was performed to evaluate the effectiveness of each method. Furthermore, a Bonferroni correction for multiple pairwise comparisons was also applied to perform the post-hoc t-test analysis. If the p-value was below the significant level of 0.05, the performance comparison was regarded as statistically significant difference and marked by asterisk notation in figure 6(A). CSP + LDA and FBSCSP + LDA methods obtained relatively low average decoding accuracies of 62.55 ± 2.45% (minimum accuracy of 59.04% and maximum accuracy of 67.15%) and 66.20 ± 3.08% (minimum accuracy of 61.30% and maximum accuracy of 70.45%) over all participants, respectively. Meanwhile, the FBCSP + LDA method significantly outperformed the CSP + LDA method across participants with p = 0.0066 < 0.01. Furthermore, compared with the CSP + LDA and FBCSP + LDA methods, the RG methods (MDRM and TS + LDA) can achieve higher decoding performance with average accuracies of 69.45 ± 2.50% (minimum accuracy of 64.88% and maximum accuracy of 73.70%) and 70.31 ± 1.70% (minimum accuracy of 67.50% and maximum accuracy of 73.03%) respectively. The performance improvement was statistically significant for CSP + LDA versus MDRM (p = 1.25 × 10−08 < 0.0001), CSP + LDA versus TS + LDA (p = 3.50 × 10−09 < 0.0001), FBCSP + LDA versus MDRM (p = 0.0060 < 0.01), and FBCSP + LDA versus TS + LDA (p = 0.0012 < 0.01). This indicated that the RG features were superior than CSP or FBCSP features in decoding multiple MI tasks from the same upper limb. Notably, the average decoding performance of the MDRM method was very close to that of the TS + LDA method and there was no statistically significant difference between these two methods (p = 0.65 > 0.05). Furthermore, the decoding accuracy can be remarkably improved by using the feature reduction for the TS features, with average accuracies of 75.80 ± 3.55% (minimum accuracy of 70.20% and maximum accuracy of 82.00%) for (TS-PCA) + LDA and 80.50 ± 3.25% (minimum accuracy of 75.85% and maximum accuracy of 85.90%) for (TS-PLS) + LDA respectively. The performance difference of the TS + LDA method between with and without feature reduction was statistically significant, with TS + LDA versus (TS-PCA) + LAD (p = 0.0002 < 0.001) and TS + LDA versus (TS-PLS) + LDA (p = 1.22 × 10−08 < 0.0001) respectively. In addition, there was a statistically significant difference between (TS-PCA) + LDA and (TS-PLS) + LDA (p = 0.0045 < 0.01). It suggested that the TS feature representations after PLS reduction were superior for decoding MI tasks from the same upper limb than that after PCA reduction.

To show the group-level decoding results for each MI task, the confusion matrices of FBCSP + LDA, TS + LDA and (TS-PLS) + LDA methods are respectively presented in figure 7. Table 1 provides the statistical decoding performance (TPR and FPR) for each of six MI tasks by FBCSP + LDA, TS + LDA and (TS-PLS) + LDA, separately. The FBCSP + LDA method achieved relative higher FPR of 6.76% and lower rates of correct classification, respectively 69.25% and 72.50% for elbow flexion/extension, 56.70% and 60.05% for wrist supination/pronation, 68.55% and 70.20% hand close/open MI tasks (see figure 7(A)). Especially, the misclassification rates for each MI task were obviously different by using the FBCSP + LDA. Such as in contrast to elbow flexion/extension and hand close/open, wrist supination was most frequently misclassified as wrist pronation (26%), and wrist pronation was frequently misclassified as wrist supination (23.35%). This result can be attributed to the fact that the separability of the spatial features extracted by the FBCSP was low due to the wrist supination/pronation MI tasks from the same upper limb activated and occupied the nearby area at the sensorimotor cortex. Compared with the FBCSP + LDA method, the rate of correct classification for the TS + LDA can be increased by 9.45%/6.95% for wrist supination/pronation MI tasks, 3.15%/1.50% for elbow flexion/extension MI tasks, and 2.30%/1.30% for hand close/open MI tasks, respectively (see figure 7(B)). Correspondingly, the overall FPR of the TS + LDA was reduced to 5.94% across six MI tasks (table 1). It implied that the TS features extracted by the RG framework had more separable distribution than spatial features extracted by FBCSP. Furthermore, the proposed approach (TS-PLS) + LDA performed lowest FPR of 3.90% and superior rates of correct classification, respectively 82.09% and 83.70% for elbow flexion/extension MI tasks, 78.70% and 78.40% for wrist supination/pronation MI tasks, 80.60% and 79.45% for and hand close/open MI tasks, respectively (see figure 7(C)). These results showed that the use of PLS
Figure 7. The confusion matrices for 6-class MI tasks with average rates (%) of correct classification and misclassification across all participants. (A) FBCSP + LDA. (B) TS + LDA. (C) (TS-PLS) + LDA.

Figure 8. The number of features after reduction using PCA and PLS methods with LDA classifier for each participant.

6.3. Feature reduction performance comparison
It has been shown in section 6.2 that the decoding accuracy of (TS-PLS) + LDA method was superior than that of (TS-PCA) + LDA method. To further compare the ability of feature reduction for the TS features between PCA and PLS methods, the number of TS features after reduction for each participant was obtained and presented in figure 8. Moreover, the average number of features after reduction over all participants was also presented in the last group of figure 8.

For the original TS features, the dimensionality was extremely large with $P = N(N + 1)/2$ ($N = 64$ was the number of channels in our EEG data).
By applying PCA method, the maximum and minimum dimensionality of TS features after reduction was 110 for participant 1 (P01) and 90 for participant 8 (P08), respectively. Whereas, the maximum and minimum dimensionality of TS features after PLS reduction was 46 for participant 1 (P01) and 35 for participant 3 (P03), respectively. It indicated that the original TS features contained plenty of redundant and irrelevant information (non-generalizable or insignificant features) which may greatly hamper the overall performance of the classifiers for decoding MI tasks. Totally, the average number of TS features after PLS reduction (40.65 ± 3.40) was significant smaller than that after PCA reduction (100.10 ± 6.24), with \( p = 0.0023 < 0.01 \) (paired t-test). This result revealed that PLS method obviously outperformed PCA method for TS features reduction, where more separable and generalizable feature representations with lower dimensions can be obtained by PLS. Consequently, the decoding performance of (TS-PLS) + LDA method was higher than that of (TS-PCA) + LDA method. It can be attributed to the fact that PLS method were implemented with the corresponding targeted class labels of MI tasks, while PCA method ignored the MI class information.

6.4. Classifier decoding performance comparison

To further investigate if the different classifiers could affect the performance of the PLS feature reduction and determine which classifier types and feature extraction methods provide maximal decoding accuracy, the SVM and LDA classifier were compared with the same FBCSP, TS-PCA, and TS-PLS feature representations, respectively. Similar to the LDA classifier, six one-versus-rest binary SVM classifiers with Radial Basis Function (RBF) kernels were trained by 5-fold cross-validation and combined to construct a multiclass SVM classifier [45, 46]. The optimal regularization constraint \( \omega \) and kernel width \( \gamma \) of the RBF in each SVM classifier were tuned by using a grid-search strategy [51]. In this paper, the search ranges were set to \( \omega = [10^{-2}, 10^{2}] \), \( \gamma = [10^{-3}, 10^{3}] \) respectively.

Figure 9 presents the decoding performance of FBCSP, TS-PCA, and TS-PLS methods respectively combined with LDA and SVM classifiers for each participant. Additionally, 500 bootstrap iterations on the MI dataset for each method were also applied. Accordingly, the average (mean ± standard deviation) with the 95% confidence interval, minimum and maximum decoding performance for each method in the group-level of 12 participants are shown in figure 10. One-way ANOVA analysis and Bonferroni correction for multiple pairwise comparisons were sequentially performed to evaluate the significant difference of the performance between LDA and SVM classifiers. Instead of using LDA classifier, the FBCSP + SVM, (TS-PCA) + SVM, and (TS-PLS) + SVM methods achieved an average decoding accuracy of 68.82 ± 2.25%, 75.90 ± 3.45%, and 79.70 ± 3.35% with the 95% confidence interval, respectively. For the same FBCSP features, the SVM classifier performed slightly higher decoding accuracy with an increase of 2.62% and significantly outperformed the LDA classifier with \( p = 0.033 < 0.05 \). The minimum and maximum accuracy were 61.30% and 70.45% for FBCSP + LDA, 64.35% and 72.08% for FBCSP + SVM, respectively. On the contrary, for the features extracted by TS-PCA or TS-PLS method, the SVM classifier cannot increase or even decrease the decoding accuracy such as the case of participant 1, 3, 7, 9, and 11 between (TS-PCA) + LDA and (TS-PCA) + SVM. The minimum and maximum accuracy for (TS-PCA) + SVM (69.50% and 81.15%) were separately lower than those for (TS-PCA) + LDA (70.20% and 82.00%). Similarly, the minimum and maximum accuracy for (TS-PLS) + SVM (75.20% and 84.38%) were also separately lower than those for (TS-PLS) + LDA (75.85% and 85.90%). Furthermore, the performance difference between LDA and SVM for TS-PCA (\( p = 0.80 > 0.05 \)) or TS-PLS \( (p = 0.42 > 0.05) \) method was non-significant. These results suggested that with the same spatial features extracted by FBCSP, the SVM classifier was more suitable for decoding MI tasks from the same limb compared with the LDA classifier. Whereas, for the TS-PCA or TS-PLS features, the SVM classifier provided no obvious superior performance than the LDA classifier. This also indirectly indicates that the TS-PCA or TS-PLS features have more generalization than the spatial features.

In addition, we also compared the performance of feature reduction between (TS-PLS) + LDA and (TS-PLS) + SVM methods. The number of features after reduction using (TS-PLS) + LDA and (TS-PLS) + SVM methods for each participant were severally illustrated in figure 11. Due to the number of features after PLS reduction was determined based on the maximum classification accuracy of LDA or SVM using 5-fold cross-validation, the number of reduced features varied slightly for LDA and SVM classifier of some participants. However, we can see that the number of features after reduction were approximately equivalent for (TS-PLS) + LDA and (TS-PLS) + SVM methods, such as the feature dimension after reduction was both 46 for participant 1, 45 for participant 9, and 43 for participant 11, respectively. From the group-level analysis, the average number of features after reduction using (TS-PLS) + SVM method over all participants was 40.65 ± 2.75, and there was no statistically significant difference between (TS-PLS) + LDA and (TS-PLS) + SVM methods in feature reduction (\( p = 0.82 > 0.05 \)). This indicated that the performance of the PLS method for TS features reduction was not affected by the designed classifier.
Figure 9. The decoding accuracies for each participant by applying FBCSP + LDA, FBCSP + SVM, (TS-PCA) + LDA, (TS-PCA) + SVM, (TS-PLS) + LDA, and (TS-PLS) + SVM. The probabilistic chance level is 16.67% (horizontal black solid line) and the significant chance level using a bootstrapping method is 18.10% (horizontal red dashed line) for the 6-class EEG classification.

Figure 10. The average, minimum and maximum decoding performance for LDA and SVM classifier across 12 participants. (A) The average decoding accuracies with the 95% confidence interval. The error bar represents the lower and upper bounds to the average accuracies. The asterisk notation denotes the significant difference (\(p < 0.05\)) between the methods. (B) The minimum and maximum decoding accuracies.

7. Conclusions and future study

Our work expands the pioneering study of Ofner et al who have demonstrated the successful decoding of single executed and imagined movements of the same upper limb based on the time-domain EEG signals (MRCPs) [35, 48]. Although there is a need for a more elaborate temporal-spatial analysis to explain the neuroscience principles behind the EEG signals as it is still unclear how the brain areas and activations are involved in different imagined movements of the same limb, especially for the opposite imagined movements from the same joint, we can draw that the elbow, wrist, and hand MI tasks from the same limb contain discriminative information in the frequency-domain and spatial-domain of the EEG signals from the perspective of the ERD/ERS analysis. Here, to improve the decoding accuracy, we proposed a novel decoding scheme for extracting new EEG feature representations and classifying six MI tasks from the same upper limb, including imaginary elbow flexion/extension, wrist supination/pronation, and hand
close/open, which are similar in [35]. The major contribution of this study was that a novel feature learning approach composed of RG framework and PLS algorithm was proposed to address the classification problem of 6-class MI tasks within the same limb. Firstly, the RG framework involving Riemannian distance and Riemannian mean was applied to extract the original TS features from the covariance matrices of MI EEG trials. To solve the problem of ‘curse of dimensionality’, the PLS algorithm was used to reduce the dimensionality of the TS features and obtain more separable feature representations.

To evaluated the performance of the proposed feature learning approach, a multi-class LDA and multi-class SVM classifier based on one-versus-rest strategy were separately adopted to perform classification for MI EEG dataset collected from 12 participants. In addition, CSP, FBCSP, MDRM classifier, original TS, and TS-PCA methods were also validated on the same dataset, respectively. The results showed that the decoding of multiclass MI tasks from the same upper limb can be achieved. Moreover, the proposed approach performed significantly higher decoding accuracy than that obtained by applying CSP, FBCSP, MDRM classifier, original TS, and TS-PCA methods, with an average accuracy of 80.50% for (TS-PLS) + LDA and 79.70% for (TS-PLS) + SVM respectively. It indicated that compared with spatial features extracted by the CSP or FBCSP method, the TS features extracted by the RG framework were more separable for discriminating different imaginary movements within the same limb. Particularly, the TS features can improve the decoding performance of those movements with proximal cortical activation areas, such as supination movement and pronation movement of the same wrist. Furthermore, the comparison results of feature dimension showed that the PLS method provided a superior performance than the PCA method in the context of TS feature dimensionality reduction.

Generally, rhythm brain activation modulation, that is the sensorimotor rhythms (SMR) can be observed during movement-related tasks [7]. However, the brain activation patterns between tetraplegic people and able-bodied people are different during imagined or attempted movements. More specifically, people with paralysis present a decrease in amplitude and spatial extent of the desynchronization of the SMR (i.e. ERD) during attempted movements compared with able-bodied people [3]. Whereas, people with paralysis generate similar cortical activation patterns during imagined movements as in able-bodied control [3]. Consequently, developing MI-based BMI system as a rehabilitation intervention is efficient and suitable for people with paralysis to restore neuromuscular connections and motor function [3–5]. Compared to the classical MI-based BMI systems which provide only limited number of control commands, decoding 6-class MI tasks from the same upper limb with higher accuracy using the proposed approach can provide more control commands for multi-DOF devices (i.e. 6-DOF exoskeleton robot). Therefore, the proposed approach is promising and could potentially be applied in the context of MI-based BMI control. Future studies need to confirm if different MI tasks from the same limb can also be decoded for people with paralysis by the proposed approach and if decoding performance is sufficient to control a robotic arm or a neural prosthesis for motor disabled patients with highly impaired upper limb.

On the other hand, there are some limitations that affect the application of our study to MI-based BMI systems for people with paralysis. For instance, in a daily-life scene, tetraplegic people can perform a drinking task by the assistance of 6-DOF exoskeleton robotic arm with MI-based BMI control, which includes reaching for and grasping a cup from a table, lifting to the mouth, taking a drink, returning the cup to the table and releasing the cup. The MI con-
trols of the drinking task consist of a sequence of elbow extension, hand open, hand close, elbow flexion, wrist supination, wrist pronation, elbow extension and hand open. However, lifting to the mouth and returning the cup to the table usually involve simultaneous hand close and elbow movements meanwhile maintaining the wrist movement, which are not considered in the here presented work. Hence, the simultaneous imaginary movements and relaxed state should be evaluated in the further study. Furthermore, considering simultaneous movements of neighboring body parts (i.e. hand close and elbow flexion, hand close and wrist supination) have overlapping spatial representations on the cortex [52], researches are needed to identify more separable EEG features for discriminating simultaneous movements with the RG framework.

In the future study, we will focus on conducting real-time online MI-based BMI control experiments by using the proposed approach. For instance, more MI tasks from different joints within the same limb will be investigated. And furtherly, the number of control commands for BMI-driven robotic devices could be increased. Furthermore, before the RG feature extraction, CSP or FBCSP methods could be used in preprocessing procedure to increase the spatial separability of EEG covariance matrices for diverse MI tasks. To reduce computational complexity of the RG feature extraction, channel selection method could be applied to exclude the redundant channels and select the optimal channels related to different MI tasks within the same limb. More importantly, further experiments need to be performed to verify the efficacy and feasibility of the proposed approach for the online MI-based BMI system, such as testing the stability of the decoder and subjects’ learning capacity. As mentioned in the previous researches [28, 53, 54], due to the two properties of affine invariance and inverse invariance of the Riemannian metric (Riemannian mean and Riemannian distance), the Riemannian approaches were robust to outliers in the BMI decoders [28, 53] and provided a significant improvement in the BMI transfer learning problem [54]. Hence, to validate the stability and robustness of the proposed decoding approach in this paper, long-term real-time classification for multiclass MI tasks will be conducted in the condition of outdoor setting with noise interference. Furthermore, as experimentally studied in [54], the proposed approach will be separately tested on the dataset of cross-session within the same subject and the dataset of cross-subject to validate the subject-to-subject transfer learning.

Acknowledgments

This work was mainly funded by the National Natural Science Foundation of China (Grants Nos. 61573340, 61773369, and U1813214), and partly supported by the Frontier Science research project of the Chinese Academy of Sciences (Grant No. QYZDY-SSW-JSC005). The authors would thank the twelve volunteers for participating the MI experiments. The authors have confirmed that any identifiable participants in this study have given their consent for publication.

ORCID iD

Yaqi Chu https://orcid.org/0000-0002-8975-3847

References

[1] Chaudhary U, Birbaumer N and Ramos-Murguialday A 2016 Brain–computer interfaces for communication and rehabilitation Nat. Rev. Neuro. 12 513–25
[2] Lazarou I, Nikolopoulos S, Petrantoniakis P C, Kompatsiari I and Tsolaki M 2018 EEG-based brain–computer interfaces for communication and rehabilitation of people with motor impairment: a novel approach of the 21st century Frontiers Hum. Neurosci. 12 14
[3] Foldes S T, Weber D J and Gollinger J L 2017 Altered modulation of sensorimotor rhythms with chronic paralysis J. Neurophysiol. 118 2412–20
[4] Mateo S, Di Rienzo E, Bergeron V, Guillet A, Collet C and Rode G 2015 Motor imagery reinforces brain compensation of reach-to-grasp movement after cervical spinal cord injury Frontiers Behav. Neurosci. 9 234
[5] Pichiorri F et al 2015 Brain–computer interface boosts motor imagery practice during stroke recovery Ann. Neurol. 77 851–65
[6] Rohm M, Schneider M, Müller C, Kreilinger A, Kaiser V, Müller-Putz G R and Rupp R 2013 Hybrid brain–computer interfaces and hybrid neuroprostheses for restoration of upper limb functions in individuals with high-level spinal cord injury Artif. Intell. Med. 59 133–42
[7] Pfurtscheller G, Müller G R, Pfurtscheller J, Gerner H J and Rupp R 2003 ‘Thought’—control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia Neurosci. Lett. 351 33–36
[8] López-Larráez E, Trincado-Alonso E, Rajasekaran V, Pérez-Nombela S, Del-Amo A J, Aranda J, Mínguez J, Gil-Aguado A and Montesano L 2016 Control of an ambulatory exoskeleton with a brain–machine interface for spinal cord injury gait rehabilitation Frontiers Neurosci. 10 359
[9] Meng J, Zhang S, Bekyo A, Olsoe J, Baxter B and He B 2016 Noninvasive electroencephalogram based control of a robotic arm for reach and grasp tasks Sci. Rep. 6 38565
[10] He B, Baxter B, Edelman B I, Cline C C and Ye W W 2015 Noninvasive brain–computer interfaces based on sensorimotor rhythms Proc. IEEE 103 907–25
[11] Chu Y, Zhao X, Zou Y, Xu W, Han J and Zhao Y 2018 A decoding scheme for incomplete motor imagery EEG with deep belief network Frontiers Neurosci. 12 680
[12] Plow E B, Arora P, Pliner M A, Binnewes M T and Carey J R 2010 Within-limb somatotopy in primary motor cortex-revealed using MfRI MRI 46 310–21
[13] Shiman E, López-Larráez E, Sarosas-Sanz A, Irastorza-Landa N, Spüler M, Birbaumer N and Ramos-Murguialday A 2017 Classification of different reaching movements from the same limb using EEG J. Neural Eng. 14 046018
[14] Yuan H and He B 2014 Brain–computer interfaces using sensorimotor rhythms: current state and future perspectives IEEE Trans. Biomed. Eng. 61 1425–35
[15] Navarro I, Hubais B and Sepulveda F 2005 A comparison of time, frequency and ICA based features and five classifiers
for wrist movement classification in EEG signals 2005 27th Annual Int. Conf. of the IEEE Eng. Med. Biol. Soc. pp 2118–21

[16] Vuckovic A and Sepulveda F 2012 A two-stage four-class BCI based on imaginary movements of the left and the right wrist Med. Eng. Phys. 34 964–71

[17] Edelman B J, Baxter B and He B 2016 EEG source imaging enhances the decoding of complex right-hand motor imagery tasks IEEE Trans. Biomed. Eng. 63 6–14

[18] Liu K, Xiao R, Gonzalez J and Ding L 2014 Decoding individuals finger movements from one hand using human EEG signals PLoS One 9 e85192

[19] Yong X and Menon C 2013 EEG classification of different imaginary movements within the same limb PLoS One 10 e0121896

[20] Tavakolan M, Frehlick Z, Yong X and Menon C 2017 Classifying three imaginary states of the same upper extremity using time-domain features PLoS One 12 e0174161

[21] Lotte F, Bougrain L, Cichocki A, Clerc M, Congedo M, Rakotomamonjy A and Yger F 2018 A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update J. Neural Eng. 15 031005

[22] Xu M, Xiao X, Wang Y, Qi H, Jung T P and Ming D 2018 A brain–computer interface based on miniature-event-related potentials induced by very small lateral visual stimuli IEEE Trans. Biomed. Eng. 65 1166–75

[23] Barachant A, Bonnet S, Congedo M and Jutten C 2010 Riemannian geometry applied to BCI classification Int. Conf. on Latent Variable Analysis and Signal Separation pp 629–36

[24] Barachant A and Bonnet S 2011 Channel selection procedure using Riemannian distance for BCI applications 2011 5th Int. IEEE/EMBS Conf. on Neural Engineering pp 348–51

[25] Devanne M, Wannous H, Berrettì S, Pala P, Daoudi M and Del Bimbo A 2015 3-d human action recognition by shape analysis of motion trajectories on riemannian manifold IEEE Trans. Cybern. 45 1340–52

[26] Wu D, Lance B J, Lawhern V J, Gordon S, Jung T P and Lin C T 2017 EEG-based user reaction time estimation using Riemannian geometry features IEEE Trans. Neural Syst. Rehabil. Eng. 25 2157–68

[27] Barachant A, Bonnet S, Congedo M and Jutten C 2012 Multiclass brain–computer interface classification by Riemannian geometry IEEE Trans. Biomed. Eng. 59 920–8

[28] Yger F, Berar M and Lotte F 2017 Riemannian approaches in brain-computer interfaces: a review IEEE Trans. Neural Syst. Rehabil. Eng. 25 1753–62

[29] Navarro-Sune X, Hudson A L, De Vico Fallani F, Martinerie J, Wibon A, Pouget P, Raux M, Similowski T and Chavez M 2017 Riemannian geometry applied to detection of respiratory states from EEG signals: the basis for a brain-ventilator interface IEEE Trans. Biomed. Eng. 64 1138–48

[30] Kalunga E K, Chevallier S, Barthélémy Q, Djouani K, Monacelli E and Hamam Y 2016 Online SSVEP-based BCI using Riemannian geometry Neurocomputing 191 55–68

[31] Nguyen C H and Artemiadis P 2018 EEG feature descriptors and discriminant analysis under Riemannian Manifold perspective Neurocomputing 275 1871–83

[32] Barachant A, Bonnet S, Congedo M and Jutten C 2013 Classification of covariance matrices using a Riemannian-based kernel for BCI applications Neurocomputing 112 172–8

[33] Pennec X, Fillard P and Ayache N 2006 A Riemannian framework for tensor computing Int. J. Comput. Vision 66 41–66

[34] Barbaresco F 2008 Innovative tools for radar signal processing based on Cartan's geometry of SPD matrices & information geometry 2008 IEEE Radar Conf. pp 1–6

[35] Ofner P, Schwarz A, Pereira J and Müller-Putz G R 2017 Upper limb movements can be decoded from the time-domain of low-frequency EEG PLoS One 12 e0182578

[36] Urgüzen A and García-Zapirain B 2015 EEG artifact removal—state-of-the-art and guidelines J. Neural Eng. 12 031001

[37] Pfurtscheller G, Brunner C, Schlögl A and Da Silva F L 2006 Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks Neuroimage 31 153–9

[38] Delorme A and Makeig S 2004 EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics J. Neurosci. Methods 134 9–21

[39] Asensio-Cubero J, Gan I Q and Palaniappan R 2013 Extracting optimal temporal-spatial features using local discriminant bases and common spatial patterns for brain computer interfacing Biom. Signal Process. Control 8 772–8

[40] Blankertz B, Tomioka R, Lemm S, Kawanabe M and Müller K 2008 Optimizing spatial filters for robust EEG Single-Trial analysis IEEE Signal Process Mag. 25 41–56

[41] Ang K G, Chin Z Y, Wang C, Guan C and Zhang H 2012 Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b Frontiers Neurosci. 6 39

[42] Tuzel O, Porikli F and Meer P 2008 Pedestrian detection via classification on Riemannian manifolds IEEE Trans. Pattern Anal. Mach. Intell. 30 1713–27

[43] Mwangi B, Tian T S and Soares J C 2014 A review of feature reduction techniques in neuroimaging Neuroinformatics 12 229–44

[44] Abdi H and Williams L J 2013 Partial least squares methods: partial least squares correlation and partial least square regression Methods Mol. Biol. 930 549–79

[45] Galar M, Fernández A, Barrenechea E, Bustince H and Herrera F 2011 An overview of ensemble methods for binary classifiers in multi-class problems: experimental study on one-vs-one and one-vs-all schemes Pattern Recognit. 44 1761–76

[46] Hsu C W and Lin C J 2002 A comparison of methods for multiclass support vector machines IEEE Trans. Neural Networks 13 415–25

[47] Combrisson E and Jerbi K 2015 Exceeding chance level by one-vs-one and one-vs-all schemes J. Neural Eng. 15 031005

[48] Zanini P, Congedo M, Barachant A, Bonnet S, Congedo M and Jutten C 2010 Riemannian-based kernel for BCI applications IEEE Trans. Biomed. Eng. 57 306–14

[49] Herrera F 2011 An overview of ensemble methods for binary classifiers in multi-class problems: experimental study on one-vs-one and one-vs-all schemes Pattern Recognit. 44 1761–76

[50] Fuggetta G, Tavakolan M, Takeoka A and Barachant A 2017 Support vector machines to detect Upper limb movements can be decoded from the time-domain of low-frequency EEG PLoS One 12 e0153842–52

[51] Schwarz A, Ofner P, Pereira J, Shurling A and Müller-Putz G R 2017 Decoding natural reach-and-grasp actions from human EEG J. Neural Eng. 15 016005

[52] Quinn M R, Cavrini F, Sernini L, Rillo F, Bianchi L, Sarsi S and Saggio G 2017 Support vector machines to detect physiological patterns for EEG and EMG-based human–computer interaction: a review J. Neural Eng. 14 011001

[53] Foldes S T and Taylor D M 2011 Offline comparison of spatial filters for two-dimensional movement control with noninvasive field potentials J. Neural Eng. 8 046022

[54] Congedo M, Barachant A and Bhatia R 2017 Riemannian geometry for EEG-based brain–computer interfaces; a primer and a review Brain Comput. Interfaces 4 153–74

[55] Zanini P, Congedo M, Jutten C, Said S and Berthoumieu Y 2017 Transfer learning: a Riemannian geometry framework for EEG-based brain-computer interfaces IEEE Trans. Biomed. Eng. 65 1107–16