Multiband tangent space mapping and feature selection for classification of EEG during motor imagery

IOP Publishing Journal of Neural Engineering

J. Neural Eng. **15** (2018) 046021 (14pp) https://doi.org/10.1088/1741-2552/aac313

Multiband tangent space mapping and feature selection for classification of EEG during motor imagery

Md Rabiul Islam[1], Toshihisa Tanaka[1,3,4] and Md Khademul Islam Molla[2]

[1] Department of Electronic and Information Engineering, Tokyo University of Agriculture and Technology, Koganei–shi, Tokyo 184–8588, Japan
[2] Department of Computer Science and Engineering, University of Rajshahi, Rajshahi-6205, Bangladesh
[3] RIKEN Brain Science Institute, Wako-shi, Saitama 351–0198, Japan

E-mail: tanakat@cc.tuat.ac.jp

Received 7 November 2017, revised 31 March 2018
Accepted for publication 8 May 2018
Published 6 June 2018

Abstract

Objective. When designing multiclass motor imagery-based brain–computer interface (MI-BCI), a so-called tangent space mapping (TSM) method utilizing the geometric structure of covariance matrices is an effective technique. This paper aims to introduce a method using TSM for finding accurate operational frequency bands related brain activities associated with MI tasks. *Approach.* A multichannel electroencephalogram (EEG) signal is decomposed into multiple subbands, and tangent features are then estimated on each subband. A mutual information analysis-based effective algorithm is implemented to select subbands containing features capable of improving motor imagery classification accuracy. Thus obtained features of selected subbands are combined to get feature space. A principal component analysis-based approach is employed to reduce the features dimension and then the classification is accomplished by a support vector machine (SVM). *Main results.* Offline analysis demonstrates the proposed multiband tangent space mapping with subband selection (MTSMS) approach outperforms state-of-the-art methods. It acheives the highest average classification accuracy for all datasets (BCI competition dataset 2a, IIIa, IIIb, and dataset JK-HH1). *Significance.* The increased classification accuracy of MI tasks with the proposed MTSMS approach can yield effective implementation of BCI. The mutual information-based subband selection method is implemented to tune operation frequency bands to represent actual motor imagery tasks.

Keywords: brain–computer interfaces (MI-BCIs), electroencephalogram (EEG), motor-imagery (MI), multiband tangent space mapping (MTSM), tangent space mapping (TMS), mutual information

(Some figures may appear in colour only in the online journal)

1. Introduction

Designing of brain–computer interface (BCI) applications based on electroencephalogram (EEG) is one of the most challenging tasks; they translates the mental imagination of movement to commands without depending on any muscle or peripheral nervous system activities [1–3]. These BCI applications are of great value in the field of neuroscience, neural engineering, and rehabilitation in which the use of prosthetics, robots, and other devices are fully controllable by mental intentions [1, 4–6]. However, it is well-known that motor imagery has been used to encourage neuroplasticity in patients' brains after a stroke [7]. Therefore, recent application, of BCI with appropriate feedback offers neurorehabilitation, which assists stroke patients to restore impaired motor functions [8, 9]. Patients can receive visual, auditory, or kinesthetic feedback to promote the brain's response to the MI tasks after stroke.

[4] Author to whom any correspondence should be addressed.

BCI researchers' fundamental goal is to determine the correct intention from brain activities durng MI tasks convert them into the corresponding control signals for BCI applications. To achieve this goal, appropriate feature extraction and machine learning (ML) strategies have been employed to classify MI tasks. Pfurtscheller *et al* first proposed an adaptive autoregressive (AAR) model with linear discriminant analysis (LDA) to classify binary responses corresponding right and left motor imagery [10]. To decode motor imagery, Ramoser *et al* proposed using optimal spatial filters with a common spatial pattern (CSP) estimated from a set of data to perform a weighting of the electrodes according to their importance for the classification task [11]. Blankertz *et al* also notably improved the performance of MI-based BCI through the CSP with different physiological phenomena such as event-related desynchronization (ERD) and lateralized readiness potential (LRP) [12].

Promising features used in MI-BCI are mu (μ) and central beta (β) rhythms evoked while imagining different movements. They have been observed in the sensorimotor area. For instance, the real and imaginary movements of hands and feet elicit changes in different brain regions' μ rhythms [13]. Pfurtscheller *et al* also demonstrated changes in EEG activity in low-frequency bands, including μ and β rhythms, caused by voluntary movements [14]. The studies in [15, 16] demonstrated that μ and β rhythms have distinct typographic responses to limb movements. Hence, μ and β rhythms should be individually considered when implementing MI-BCI. Such observed changes to μ and β rhythms are utilized by localizing the CSP method into narrow bands of EEG signals to extract reliable features. However, the effectiveness of CSP depends on the subject-specific frequency bands to find a reliable representation of brain signals, which is the most challenging task with BCI [17–19].

Several methods to utilize the narrow-band features have been developed, such as subband CSP (SBCSP) [20], filter-bank CSP (FBCSP) [17, 18], discriminative filter-band CSP (DFBCSP) [21], and sparse filter-band CSP (SFBCSP) [22], which lead to more suitable for specifying subject-specific frequency bands, thereby improving BCI performance due to autonomously selecting discriminative features.

In various fields, the Riemannian manifold of symmetric positive definite (SPD) matrices has attracted increasing attention due to a rich framework to manipulate the covariance structure of the data as the feature of interest. The concept of covariance matrices in a Riemannian manifold has been successfully applied to radar signal processing [23], diffusion tensor imaging [24], and computer vision [25]. The similar method has also been used to classify EEGs into different sleep states by using the *k*-nearest neighbors approach [26]. In motor-imagery based BCI, Barachant *et al* proposed the use of the Riemannian distance between SPD matrices in BCI applications to manipulate EEG spatial covariance matrices in their native space called the Riemannian manifold [27]. The method named tangent space linear discriminant analysis was used to implement multi-class BCI. All sample covariance matrices are mapped onto the linear tangent space at their geometric mean. Since the tangent space is linear, this mapping operation allows utilizing many pupular classification methods within the Euclidean framework. Kumar *et al* proposed a framework using the combination of CSP and TSM-based approach in the MI-based BCI [28]. The underlying idea of the method was to increase the performance as well as reduce the computational time. Although the method could successfully decode MI activities, it is only suitable to implement binary classification problems. The coefficients of the filter using CSP are estimated so that the variance of the signals extracted by the spatial filter differs between two opposite tasks, limiting multi-class MI-based BCI.

Recently, Uehara *et al* proposed trimmed based averaging of sample covariance matrices (SCMs) instead of the geometric mean to improve the performance of two-class MI-BCIs [29]. Trimmed based averaging can estimate robust centrality from training data by eliminating those exhibiting the largest distance from the average covariance. In addition, TSM at the geometric mean may not be appropriate due to noise effects from EEG signals. The same authors also extended their first idea into multi-class cases to observe the effectiveness of the trimmed averaging technique compared to conventional averaging [30].

By considering the problems of noise effects as well as subject-specific frequency bands, multiband approach with tangent space mapping (MTSM) was proposed in [31] to further improve MI-BCI's classification performance. The idea was to extract effective noise-robust features with respect to narrow band signals in a multiband tangent space framework. This motivated us to develop an autonomous method for frequency band selection in the framework of TSM. In this study, we address the problem of subband selection, which can extract effective subject-specific frequency bands based on the mutual information approach between features and class labels to further improve the performance of the MI-BCI. EEG signals are decomposed into subbands with a filter bank and some subbands are chosen based on maximum values of average mutual information scores from the feature space in a tangent space. The features dimension is reduced and a support vector machine (SVM) is used for classification.

2. Methods

A multiband approach using the TSM method with subband selection is introduced here. A block diagram of the proposed method is illustrated in figure 1. For both training and test sessions, the following steps are performed. First, a multi-channel EEG is decomposed into subbands. Second, the sample covariance matrix (SCM) of each subband is calculated. Third, individual SCMs at each subband are mapped onto the *tangent space* corresponding to a central point determined by a set of SCMs. Finally, only dominant subbands are selected. Then, a dimensionality reduction is applied to the feature vector consisting of all features of the selected subbands.

Subband selection and dimensionality reduction are performed by using a training dataset. For the subband selection, the mutual information between features and class labels is

used. For dimensionality reduction, the principal component analysis (PCA) with one-way ANOVA based approach is employed [27]. Finally, a SVM is used to perform the classification.

2.1. Classification methods for spatial covariance matrices

Pfurtscheller *et al* reported in [14] that the topological representation and changes in the band power of mental tasks exhibited in different motor cortex areas known as event-related desynchronisation (ERD) and event-related synchronisation (ERS), which has practical significance in implementing BCI. In the BCI system, EEG recording of brain activity is usually split into trials, which consist of mental tasks or class. To implement the calibration of MI-based BCI, some trails are used as training to map a model that can correctly determine class in unseen instances. In this context, CSP is a well-known approach for implementing MI-BCI, which provides spatial weights determined from training trails so the variances of the signal extracted by the spatial weights differ between two mental tasks or classes. In this subsection, we first review the tangent space mapping methods, and then describe the method proposed in this paper.

2.1.1. TSM. Recently, there has been increasing interest in using the spatial covariance matrices (SCMs) approach in the study of MI-BCI due to multi-class implementation [27, 30, 31]. SCMs are used to extract features from MI data since they contain spatial information embedded in the EEG signals. Therefore, SCMs containing spatial information from EEG signals are mapped in their native space, also known as tangent space mapping (TSM). Thus, any machine learning algorithm working in the Riemannian distance can be applied. Details of covariance matrices used for mapping the tangent space are illustrated herein.

Let us denote $X_i \in \mathbb{R}^{m\times T}$ as a data-matrix consisting of an m-channel EEG signal where T represents time samples corresponding to the ith trial of a MI mental task. The spacial covariance matrix of an ith trial can be estimated as a sample covariance matrix

$$P_i = \frac{1}{T}\left(X_i - \bar{X}\right)\left(X_i - \bar{X}\right)^T \in \mathbb{R}^{m\times m}, \tag{1}$$

where i represents an index of the trials ranging from 1 to the total number of trials denoted by I.

To apply distance-based classification algorithms in MI-based BCI, Barachant *et al* proposed to map the covariance matrices onto the tangent space of the Riemannian manifold where SCMs can be served as vectors [27]. Since a SCM is an instance of a symmetric, positive definite matrix, we can define the set of all $m \times m$ real symmetric matrices as $P(m) = \left\{P \in S(m) \mid u^T P u > 0, \forall u \in \mathbb{R}^m, u \neq 0\right\}$ in the space of real symmetric matrices $S(m)$, which forms a differentiable Riemannian manifold $\mathcal{M}$ of dimension $m(m+1)/2$ [27].

Let P_1 and P_2 be the two SPD matrices of MI trails. The natural metric of the manifold $\mathcal{M}$ of SPD matrices is defined

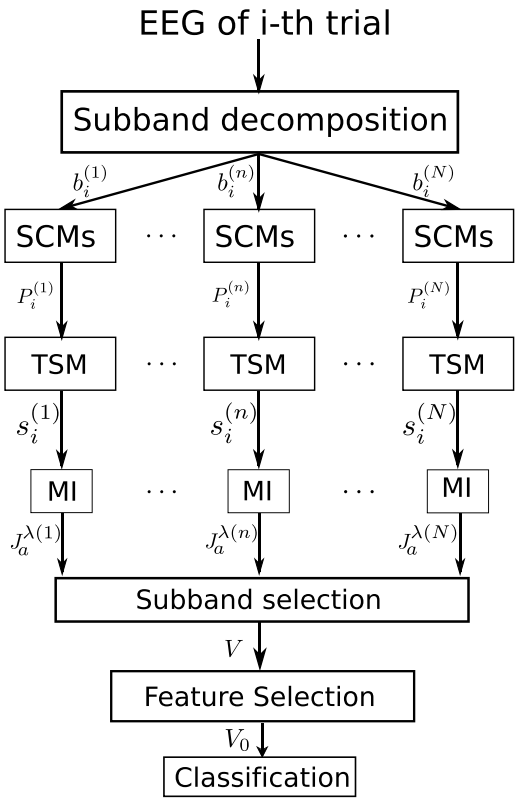

Figure 1. Proposed MI-BCI system with different components where the values $b_i^{(1)}, \ldots, b_i^{(n)}, \ldots, b_i^{(N)}$ represent the subbands proposed in [25] and are obtained by applying a Chebyshev Type II filter. N is the number of subbands.

by the local inner product induces the Riemannian distance between the two SPD matrices as:

$$\delta_e(P_1, P_2) = \left\|\log(P_1^{-1/2} P_2 P_1^{1/2})\right\|_F = \left[\sum_{i=1}^{m} \log^2 \lambda_i\right]^{1/2} \tag{2}$$

where $\|\cdot\|_F$ is the Frobenius norm. The λ_i and $i = 1, \ldots, m$ represent the real eigenvalues of $P_1^{-1}P_2$. To map from the Riemannian manifold $\mathcal{M}$ to the tangent space, the robust reference point is first estimated from the whole training data using averaging approaches [32], and each SPD matrix P_i is projected on the tangent space of the Riemannian manifold at point c and vectorized as:

$$s_i = \text{upper}(\log(Q_c^{-1/2} P_i Q_c^{-1/2})) \tag{3}$$

where Q_c represents the reference point determined by one of the averaging techniques described in section 2.4, and the upper$(\cdot)$ is an operator that keeps the upper triangular part of a symmetric matrix and vectorizes it by applying the weight $2^{1/2}$ only to the off-diagonal entries. A graphical representation of tangent space mapping from Riemannian manifold $\mathcal{M}$ is shown in figure 2. The tangent space features, denoted as V, for the whole set of I trials can be defined as:

$$V = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_I \end{bmatrix} \in \mathbb{R}^{I\times d} \tag{4}$$

where $d = m(m+1)/2$ is the dimensional of feature vector.

2.1.2. Multiband tangent space mapping (MTSM). To improve the performance of MI-based BCI, Islam *et al* proposed multiband tangent space mapping [31]. In the multiband TSM approach, there are three major procedures: (1) multiband analysis; (2) tangent space mapping of covariance matrices on Riemannian manifold $\mathcal{M}$ to extract features; and (3) classification of MI using linear SVM. First, multiband analysis performed a mixed EEG signal into multiple subband components, which is widely used for several practical applications [17, 18, 33–36]. Let us consider $S_1, \ldots, S_N$ are the subbands decomposed from EEG signals, where N is the total number of subbands. After applying a filter bank analysis of the m channel EEG signal, spatial covariance matrices are estimated for each subband component separately to obtain a set of sample covariance matrices $P_i^1, \ldots, P_i^n, \ldots, P_i^N$. According to [31], tangent space mapping of the nth subband can be defined as:

$$s_i^n = \text{upper}(\log(Q_c^n)^{-1/2} P_i^n (Q_c^n)^{-1/2}) \tag{5}$$

where $n = 1, \ldots, N$. The MTSM feature vector of the ith trial is then combined as follows:

$$v_i = [s_i^1, \ldots, s_i^n, \ldots, s_i^N] \in \mathbb{R}^{1 \times N \cdot d}. \tag{6}$$

The tangent space feature V for the whole set of trials I can be defined as:

$$V = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_I \end{bmatrix} \in \mathbb{R}^{I \times N \cdot d}. \tag{7}$$

2.1.3. Multiband tangent space mapping with subband selection (MTSMS). Recorded EEG signals are often corrupted by noise, which is significantly stronger than the original EEG signals. Hence only some subbands are useful for recognition purposes. Therefore, the selection of subbands would intuitively provide a more accurate classification with respect to the whole set of subbands. To select a subband with relevant features, mutual information was used in this paper to estimate the score of each subband and the subbands are rearranged according to their maximum values of mutual information scores. The mutual information-based feature selection approach was first proposed by Ang *et al* to extract the best features for MI-based BCI [17, 18]. The main purpose of the mutual information method is to measure whether the values of a feature are dependent on the associated class labels without assuming the distributions of the features. In this study, we applied mutual information to quantify the relevance scores of the features based on a feature scoring algorithm proposed in different studies [37, 38]. For s_i^n mapped by (5), mutual information $\mathcal{I}_{mi}^{(n)}$ of nth subband for the training feature F and true class labels C was calculated formally as:

$$\mathcal{I}_{mi}^{(n)}(F;C) = \sum_{f \in \mathcal{F}} \sum_{c \in C} p(f,c) \log \frac{p(f,c)}{p(f)\,p(c)} \tag{8}$$

where $p(f,c)$ is the joint probability function of F and C, and $p(f)$ and $p(c)$ are the marginal probability density function of F and C. For class lebel C, the mutual information score for a feature f_k is defined using following criteria proposed in [38]:

$$J_{mi}^{(n)}(f_k) = \mathcal{I}_{mi}^{(n)}(f_k;C) - \lambda \sum_{f_j \in F_L} \mathcal{I}_{mi}(f_k;f_j) \tag{9}$$

where F_L represents the set of currently selected features. The λ represents weighting factor to control the influence inter feature redundancy. We used the formula (9) with ($\lambda = 0.5$). The term $\mathcal{I}_{mi}^{(n)}(f_k;C)$ is to ensure feature relevance and the penalty $\sum_{f_j \in S} \mathcal{I}_{mi}(f_k;f_j)$ is used to enforce low correlations with features already selected in F_L. The average mutual information score $J_a^{(n)}$ of nth subband for training feature can be defined as:

$$J_a^{(n)} = \sum_{k=1}^{d} J_{mi}^{(n)}(f_k). \tag{10}$$

The set of average mutual information, $J_a^{(n)}$ for N subbands was rearranged in descending order such that $J_a^{\lambda(1)} \geqslant \ldots \geqslant J_a^{\lambda(n)} \geqslant \ldots \geqslant J_a^{\lambda(N)}$, where $\lambda(n)$ is the permutation function. The tangent space feature with selected subbands for ith trial can be rearranged as:

$$\tilde{v}_i = [\tilde{s}_i^{T(\lambda(1))}, \ldots, \tilde{s}_i^{T(\lambda(G))}] \tag{11}$$

where G is the maximum number of subbands to be selected. In MTSMS, the tangent space feature V for the whole set of trials I can be defined as:

$$V = \begin{bmatrix} \tilde{v}_1 \\ \tilde{v}_2 \\ \vdots \\ \tilde{v}_I \end{bmatrix} \in \mathbb{R}^{I \times G \cdot d}. \tag{12}$$

2.2. Feature selection

In machine learning and statistics, feature selection is a very important part of choosing the best set of features from all available features. The principal component analysis (PCA)-based feature selection approach was performed on high dimensional tangent space to select more discriminative tangent features. The vectors V can be orthogonalized and the number of dimensions is reduced as:

$$V_0 = U^T V, \tag{13}$$

where U is the matrix of the principal orthogonal vectors of V, which are drived from the training set. After applying the PCA on tangent space features, a one-way analysis of variance (ANOVA) is applied to automatically select the minimum number of variables (components) in the same way

Table 1. Different types of mean depend on the distance function $\delta_e(\cdot,\cdot)$.

Distance	Definition
Euclidean distance	$\delta_E(A,B) = \|A-B\|_F$
Riemannian geodesic distance	$\delta_R(A,B) = \left\|\log(A^{-1/2}BA^{-1/2})\right\|_F$
Log-Euclidean distance	$\delta_L(A,B) = \|\log(A) - \log(B)\|_F$
Harmonic distance	$\delta_H(A,B) = \left\|A^{-1} - B^{-1}\right\|_F$

as Barachant *et al* [27]. All components applied in the PCA are ranked according to their p-values and the minimum number of varables (componets) is set using a weighted false discovery rate (FDR) [27, 39] with the expected proportion of false rejections with respect to all rejections. In order to provide rank to the components, singular values are used as weight for the FDR procedure, which represent the best structure of the data. Finally, the variables (components) are selected in ascending order of p-values.

2.3. Robust averaging of sample covariance matrices

In MI-based BCI, spatial covariance matrices play important roles in feature extraction. Therefore, estimation of spatial covariance matrices potentially improve recognition performance. Many studies suggest different approaches for averaging sample covariance matrices (SCM) based on the chosen geometry to achieve more effective tangent space mapping. Details of averaging techniques are described in the following paragraphs.

2.3.1. Mean. The most common types of averaging approach for SCMs are arithmetic, geometric, and harmonic mean; these estimate the centrality c of manifold $\mathcal{M}$. The general form of mean can be defined for the set of SPD matrices $P_1,\ldots,P_I$ as:

$$Q_{mean} = \arg\min_{P\in P(m)} \sum_{i=1}^{I} \delta_e^2(P,P_i) \tag{14}$$

where $\delta_e(\cdot,\cdot)$ denotes a distance function in space $P(m)$. Depending on distance function $\delta_e(\cdot,\cdot)$, we can represent the above equation (14) as the different types of means (arithmetic mean $\delta_E(A,B)$, Riemannian geometric mean $\delta_R(A,B)$, log-Euclidean mean $\delta_L(A,B)$, and harmonic mean $\delta_H(A,B)$) shown in table 1.

2.3.2. Median. Another robust averaging technique for SCMs to estimate the centrality c of manifold $\mathcal{M}$. To deal with outliers, the squared distances in geometric mean may cause problems when estimating centrality [40]. Therefore, the square is removed from the formula and the Riemannian geometric median is defined through the following optimization problem:

$$Q_{med} = \arg\min_{P\in P(m)} \sum_{i=1}^{I} \delta_e(P,P_i). \tag{15}$$

In [41] Weiszfeld algorithm computed geometric median in the space $P(m)$ using an Euclidean metric. The solution of the minimization problem with kth iteration for the set of SPD matrices P_i can be computed as [41]:

$$Q_{med}^{(k+1)} = R_e(Q_{med}^{(k)}, v^{(k)}) \tag{16}$$

$$v^{(k)} = \frac{\sum_{i=1}^{I} \frac{R_e^{-1}(Q_{med}^{(k)}),P_i}{\delta_R(Q_{med}^{(k)},P_i)}}{\sum_{i=1}^{I} \frac{1}{\delta_R(Q_{med}^{(k)}),P_i}} \tag{17}$$

where $Q_{med}^{(k)}$ represents the value of the Euclidean geometric median for the kth iteration. R_e and R_e^{-1} are a retraction map and its inverse (lifting) operator, respectively, defined as [41]:

$$R_e(P_{ref},S) = \exp\left(\log\left(P_{ref}\right) + S\right) \tag{18}$$

$$R_e^{-1}(P_{ref},S) = \log\left(P\right) - \log\left(P_{ref}\right) \tag{19}$$

where $P_{ref}, P \in P(m)$ and $S \in S(m)$ are m-dimensional matrices.

2.3.3. Triming-based averaging of SCMs. Tangent space mapping is an effective method used for extracting discriminative patterns, which strongly depends on the selection of a reference covariance matrix. All sample covariance matrices (SCMs) are mapped onto the tangent space at their reference point. In previous work of MI-BCI, the reference point was estimated based on the geometric mean or median of SCMs, which may not be the best choice due to the outliers in EEG signals [29]. EEG signals yield very noisy measurements with low spatial resolution as well as voluntary components related to other brain activities. To address this problem, Uehara *et al* first proposed the trimmed based averaging of SCMs in MI-BCI by discarding the matrices, which are far apart from the geometric mean or median [29, 30]. The Trimmed Riemannian and trimmed log-Euclidean mean of SPD matrices $P_1,\ldots,P_I$ can be defined according to [29, 30] as:

$$Q_{\hat{g}} = Q_g[\zeta_e^g(P_1,\ldots,P_I)] \tag{20}$$

$$Q_{\hat{l}} = Q_l[\zeta_e^l(P_1,\ldots,P_I)] \tag{21}$$

where $Q_g[.]$ and $Q_l[.]$ denote a Riemannian geometric and a log-Euclidean mean operator, respectively. $\zeta_e^g(.)$, and $\zeta_e^l(.)$ are the trimming operators, which discard e (%) SPD matrices of the largest distance from the mean. In the same way, the trimmed Riemannian and trimmed log-Euclidean median can be defined for the set of SPD matrices $P_1,\ldots,P_I$ according to [29, 30]:

$$Q_{\hat{g}^*} = Q_{g^*}[\zeta_e^{g^*}(P_1,\ldots,P_I)] \tag{22}$$

$$Q_{\hat{l}^*} = Q_{l^*}[\zeta_e^{l^*}(P_1,\ldots,P_I)] \tag{23}$$

where $Q_g^*[.]$ and $Q_l^*[.]$ are the operators of a geometric and a log-Euclidean median. The trimming operators $\zeta_e^{g^*}(.)$ and $\zeta_e^{l^*}(.)$ discard the e (%) of covariance matrices with the largest distances from the median.

3. Experimental results

3.1. EEG data description

We used the datasets JK-HH 1, which was approved by the research ethics committee of the Tokyo University of Agriculture and Technology in Tokyo, Japan [42]. Aside from this dataset, the performance of our proposed methods are also evaluated using the well-known publicly available datasets from BCI competition III and IV (http://bbci.de/competition/). Details of these datasets are discribed in the following sections.

3.1.1. Dataset JK-HH 1 (four-class). This dataset contains four classes of EEG signals during left hand, right hand, foot, and idle states [42]. The EEG signals were recorded from five subjects, denoted as sa, sb, sc, sd, and se. All subjects were male and their average age was 23.2 years with a standard deviation of 1.6 years. An international 10–20 system [43] recorded 29 electrodes placed at F3, Fz, F4, FC5, FC3, FC6, FCz, FC2, FC4, FC6, T7, C5, C3, C1, Cz, C2, C4, C6, T8, CP5, CP3, CP1, CPz, CP2, CP4, CP6, P3, Pz, and P4. Signals observed in electrodes were amplified using a bio-amplifier (MEG- 6116 produced by Nihon Kohden) and sampled using an A/D converter (AIO-163202F-PE produced by Contec) with a sampling rate of 256 Hz. The converted signals were recorded under Matlab (Mathworks, Inc.) using the data acquisition toolbox.

As a preprocessing step, we applied a fourth-order Butterworth low-pass filter to the acquired signals, with a cutoff frequency of 50 Hz, and downsampled the signals to 128 Hz. The dataset for each subject consisted of signals of 100 trials per class. The signals in each trial are extracted from the period of 4 s after a visual cue.

3.1.2. BCI competition III dataset IIIa (four-class). The dataset consists of recordings from three subjects (k3b, k6b, and l1b). Each subject sat in a relaxing chair with armrests and was asked to perform imagery movements with four different tasks (left hand, right hand, foot, and tongue) according to a cue. The order of cues was random. Recordings were made with a 64-channel EEG amplifier from Neuroscan with the left mastoid for reference and the right mastoid as ground. EEG signals with 60-channels were recorded with a sampling rate of 250 Hz and filtered between 1–50 Hz with the Notch filter on. Each subject completed 60 trials per class.

3.1.3. BCI Competition III Dataset IIIb (two-class). The dataset IIIb was recorded from three healthy subjects, denoted as O3, S4, and X11, which provided by the Graz University of Technology in Austria. It consists of cued motor imagery with two different MI tasks, namely imagining movement of the left (class 1) and right (class 2) hands. The experiment contains three sessions of each subject with online feedback. Two subjects labelled S4 and X11 are used in our experiment due to errors in data of subject O3. The recordings were taken by a bipolar EEG amplifier from g.tec. with sampled at 125 Hz, and 540 trials were recorded from each subject.

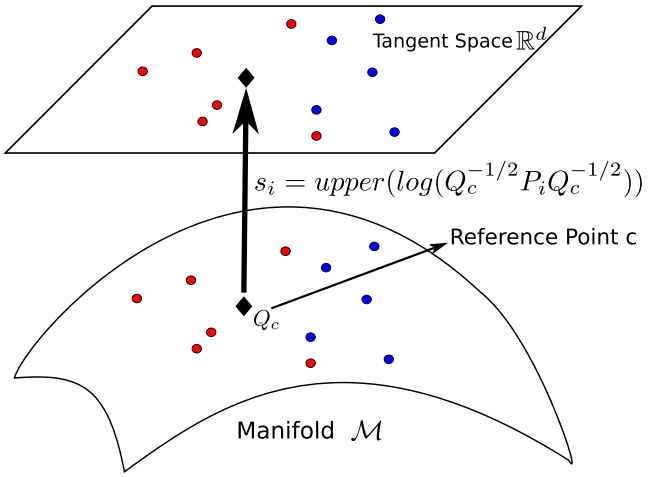

Figure 2. An example of the tangent space mapping from Riemannian manifold $\mathcal{M}$.

3.1.4. BCI competition IV dataset 2a (four-class). Dataset IIa consists of four-class MI EEG data, which was provided by the Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces) at the Graz University of Technology, Austria. The dataset was recorded from nine healthy subjects While they imagined movement of their left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4). The purpose of the BCI competition IV datasets was to address the new challanging problems to validate the signal processing and classification methods that are highly relevant for practical BCI systems. Therefore, the eye movement artifacts were included in dataset 2a to observe the performance of the methods used for designing BCI systems. Twenty-two Ag/AgCl electrodes (inter-electrode distances of 3.5 cm) were used to record EEG monopolarly with the left mastoid serving as the reference and the right mastoid as the ground. The EEG signals were sampled at 250 Hz and band-pass filtered between 0.5 and 100 Hz, with the 50 Hz notch filter on. Data includes 288 trials recorded from each subject.

3.2. Results

Features were extracted using different algorithms mentioned below and classification was performed by a SVM classifier [30]. Performance in terms of the classification accuracy (in %) of different feature extraction techniques with ten-fold cross-validation were compared. The feature extraction algorithms are briefly described as follows:

TSM [27]: The EEG signals are pre-filtered using a butterworth bandpass filter whose passband is 7–30 Hz. In MI-based BCI, several studies used the frequency band within 7 to 30 Hz for effective use of the CSP algorithm [44–46]. By considering this general setting, studies in [27, 29, 30] used TSM approach with 7–30 Hz frequency

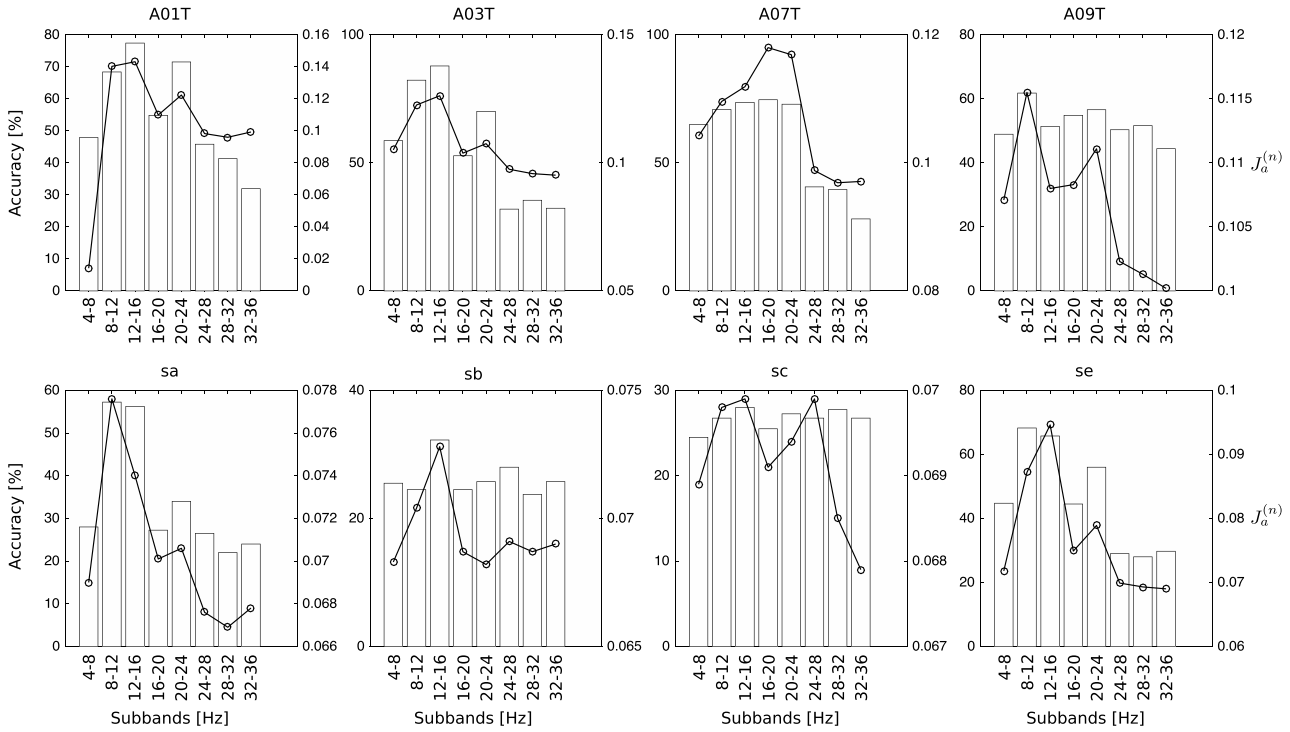

Figure 3. Average recognition accuracies of different subbands ($n = 1, \ldots, N$) with TSM and its average mutual information scores ($J_a^{(n)}$) for dataset IV-II and JK-HH 1. In both cases, a ten-fold cross-validation was performed.

band. The spatial information of multichannel EEG is embedded in the spatial covariance matrices. Therefore, all distance-based classification algorithms can be extended by applying this new Riemannian distance. Tangent space mapping sends the covariance matrices, belonging to a manifold, into Euclidean space where they can be treated as vectors. This mapping operation allows the use of state-of-the-art classifiers within the Riemannian framework.

MTSM [31]: The EEG signals of each subject are decomposed into N subbands by applying a filterbank in the frequency range from 4 to $4N + 4$ Hz, where N is the maximum number of subbands to be optimal and the bandwidth of each subband was 4 Hz. The N bandpass filters are zero-phase Chebyshev Type II. A spatial covariance matrix (SCM) is derived for each subband signal, leading to tangent space mapping. The mutual information-based potential feature vector-extraction is not used.

MTSMS(proposed): We use N subbands for each subject (where $N = 8$ and the bandwidth of each subband was 4 Hz) with covering frequency range from 4 to $4N + 4$ Hz using zero-phase Chebyshev Type II bandpass filters. The selection of operational bands based on the mutual information is used.

3.2.1. Subband selection based on average mutual information score. To justify the use of the average mutual information score with respect to each subband as a selection criterion, the classification accuracies of different subbands of BCI competition IV datasets IIa and JK-HH1 are calculated, as shown in figure 3. In this case, the arithmetic mean induced from equation (13) with $\delta E(A, B)$ (see table 1) was used. Average mutual information scores of each subband are also estimated as shown in figure 3 (black circles). It is observed that higher ranked (higher value of mutual information) subbands perform better than those of lower rank. Hence, using the mutual information based subband selection approach is justified. The subbands are sorted according to rank and the top G subbands are employed here for MI classification.

3.2.2. BCI results. To observe the performance of the methods, results were obtained from different datasets. In MI-based BCI, it is unknown whether the TSM with multiband approach can be done in an online operation due to its subject dependency of the performances. To fairly compare the performance with the proposed MTSMS, optimal subband with multiband TSM was used. In the proposed MTSMS method, we chose $G = 4$ and $G = 7$ in equation (12) for 4- and 2-class problems, respectively. The optimal subband with MTSM approach is estimated based on maximum average accuracy across subjects for each dataset. We used $N = 6$ and $N = 7$ in equation (7) for 4- and 2-class problems, respectively.

Performance is evaluated regarding the classification accuracy using the BCI competition IV dataset 2a as shown in table 2. It is a 4-class classification problem. We set the maximum number of subbands with ($G = 4$) for dataset 2a. All the mentioned algorithms (TSM, MTSM and MTSMS) are tested with different SCM averaging approaches. It is found that the highest accuracies for subjects A01T, A03T, A08T, and

Table 2. Classification accuracy (%) of TSM, MTSM, and MTSMS for BCI competition IV dataset 2a (4-class problem). Average approaches of SCMs indicate in this paper as: Arithmetic mean (A_{mean}), Riemannian geometric mean (RG_{mean}), log-Euclidean mean ($L\text{-}E_{mean}$), Harmonic mean (H_{mean}), Euclidean geometric median (EG_{med}), Riemannian geometric median (RG_{med}), log-Euclidean geometric median ($L\text{-}E_{med}$), trimmed Riemannian mean (TR_{mean}), trimmed log-Euclidean mean ($TL\text{-}E_{mean}$), Trimmed Riemannian median (TR_{med}), and trimmed log-Euclidean median ($TL\text{-}E_{med}$).

		Subjects					
Methods	Averaging functions	A01T	A03T	A07T	A08T	A09T	AVG
TSM	A_{mean}	75.4 ± 0.015	83.7 ± 0.031	79.9 ± 0.018	83.3 ± 0.047	57.6 ± 0.072	75.98 ± 10.80
	RG_{mean}	75.0 ± 0.033	80.2 ± 0.021	77.4 ± 0.051	85.1 ± 0.052	57.6 ± 0.042	75.06 ± 10.45
	$L\text{-}E_{mean}$	75.4 ± 0.027	83.0 ± 0.027	76.4 ± 0.033	84.8 ± 0.053	59.7 ± 0.033	75.86 ± 9.90
	H_{mean}	72.6 ± 0.027	83.0 ± 0.020	81.2 ± 0.026	84.8 ± 0.060	58.0 ± 0.089	75.92 ± 11.05
	EG_{med}	74.7 ± 0.010	83.7 ± 0.051	81.6 ± 0.036	86.1 ± 0.0535	57.0 ± 0.069	76.62 ± 11.76
	RG_{med}	74.7 ± 0.033	80.2 ± 0.020	77.8 ± 0.055	85.4 ± 0.057	56.6 ± 0.024	74.94 ± 10.97
	$L\text{-}E_{med}$	74.0 ± 0.048	83.3 ± 0.021	77.8 ± 0.031	85.4 ± 0.047	59.7 ± 0.027	76.04 ± 10.18
	TR_{mean}	76.7 ± 0.029	83.7 ± 0.026	81.2 ± 0.030	86.8 ± 0.060	62.8 ± 0.054	78.24 ± 9.38
	$TL\text{-}E_{mean}$	76.0 ± 0.030	83.7 ± 0.018	82.0 ± 0.032	86.1 ± 0.058	63.5 ± 0.058	78.26 ± 9.05
	TR_{med}	77.4 ± 0.028	83.0 ± 0.023	81.6 ± 0.032	85.8 ± 0.060	61.1 ± 0.055	77.64 ± 9.66
	$TR\text{-}E_{med}$	76.4 ± 0.027	84.4 ± 0.022	82.3 ± 0.028	86.4 ± 0.062	64.2 ± 0.049	78.74 ± 8.94
MTSM	A_{mean}	83.7 ± 0.012	88.5 ± 0.031	91.3 ± 0.045	89.6 ± 0.076	64.6 ± 0.037	83.54 ± 10.95
	RG_{mean}	85.8 ± 0.015	87.5 ± 0.036	92.4 ± 0.052	89.9 ± 0.094	63.5 ± 0.015	83.82 ± 11.62
	$L\text{-}E_{mean}$	86.1 ± 0.027	87.8 ± 0.036	90.9 ± 0.026	90.6 ± 0.106	63.5 ± 0.027	83.78 ± 11.51
	H_{mean}	86.6 ± 0.031	87.5 ± 0.033	92.7 ± 0.041	88.5 ± 0.070	62.5 ± 0.0104	83.56 ± 12.00
	EG_{med}	86.6 ± 0.026	87.8 ± 0.006	90.6 ± 0.020	89.2 ± 0.075	65.3 ± 0.018	83.90 ± 10.50
	RG_{med}	84.7 ± 0.021	88.2 ± 0.041	92.4 ± 0.047	89.6 ± 0.094	64.2 ± 0.012	83.82 ± 11.31
	$L\text{-}E_{med}$	87.5 ± 0.031	88.2 ± 0.031	92.4 ± 0.037	90.6 ± 0.095	64.2 ± 0.021	84.58 ± 11.55
	TR_{mean}	87.5 ± 0.020	88.9 ± 0.040	93.4 ± 0.039	91.0 ± 0.083	64.2 ± 0.028	85 ± 11.83
	$TL\text{-}E_{mean}$	86.8 ± 0.017	90.6 ± 0.035	93.0 ± 0.028	92.0 ± 0.087	67.0 ± 0.032	85.88 ± 10.81
	TR_{med}	87.2 ± 0.023	89.2 ± 0.037	**93.8 ± 0.034**	91.0 ± 0.083	67.0 ± 0.030	85.64 ± 10.69
	$TR\text{-}E_{med}$	87.5 ± 0.018	90.3 ± 0.028	93.4 ± 0.031	92.0 ± 0.081	67.0 ± 0.029	86.04 ± 10.86
MTSMS	A_{mean}	87.6 ± 0.026	90.3 ± 0.047	87.8 ± 0.043	90.3 ± 0.078	66.7 ± 0.030	84.54 ± 10.05
	RG_{mean}	84.4 ± 0.006	91.3 ± 0.053	87.2 ± 0.053	89.2 ± 0.052	67.0 ± 0.042	83.82 ± 9.74
	$L\text{-}E_{mean}$	86.8 ± 0.010	89.9 ± 0.058	88.2 ± 0.031	88.9 ± 0.063	68.8 ± 0.037	84.52 ± 8.85
	H_{mean}	84.7 ± 0.021	90.3 ± 0.042	89.9 ± 0.045	89.6 ± 0.068	68.1 ± 0.045	84.52 ± 9.14
	EG_{med}	86.8 ± 0.021	91.3 ± 0.042	88.9 ± 0.024	89.2 ± 0.086	64.6 ± 0.069	84.16 ± 11.05
	RG_{med}	84.0 ± 0.006	90.3 ± 0.039	87.8 ± 0.042	88.5 ± 0.082	65.6 ± 0.057	83.24 ± 0.12
	$L\text{-}E_{med}$	86.6 ± 0.006	89.6 ± 0.047	88.5 ± 0.039	88.2 ± 0.073	67.7 ± 0.033	84.12 ± 9.24
	TR_{mean}	86.5 ± 0.022	**91.7 ± 0.045**	89.2 ± 0.041	**93.0 ± 0.079**	68.1 ± 0.044	85.70 ± 10.14
	$TL\text{-}E_{mean}$	87.2 ± 0.014	91.0 ± 0.047	90.0 ± 0.039	90.0 ± 0.081	69.0 ± 0.047	85.44 ± 9.28
	TR_{med}	86.8 ± 0.021	91.0 ± 0.043	89.2 ± 0.037	**93.0 ± 0.080**	**69.1 ± 0.049**	85.82 ± 9.63
	$TR\text{-}E_{med}$	**87.9 ± 0.019**	91.0 ± 0.043	92.0 ± 0.038	**93.0 ± 0.083**	68.8 ± 0.046	**86.54 ± 10.08**

A09T are achieved by using the proposed MTSMS method, whereas the subject A07T exhibits the highest accuracy using the MTSM method proposed in [31].

Table 3 presents the classification accuracy of the 4-class problem for Dataset JK-HH1. The overall performance with this dataset is lower compared to others rather than the subjects sa and se. The TSM method performs better than MTSM for only one subject, sd. The performance of subjects sa and sb is close to MTSM. The extension with subband selection approach ($G = 4$) exhibits the highest classification accuracy for all subjects (sa, sb, sc, sd, and se). It is observed that the trimmed based averaging techniques with MTSMS perform better than simple mean and median.

The BCI competition III Dataset IIIa, which is very popular and publicly available, is used to test the performance of the proposed methods on the 4-class paradigm. The performance of three subjects taken from the dataset is illustrated in table 4 in which the comparative results of different methods are studied. We set the maximum number of subbands to be selected with ($G = 4$). The proposed method MTSMS exhibits the best accuracy for all subjects. Hence, the proposed methods outperform recent algorithms. It is also noted that most of the higher accuracies are achieved with the trimmed-based averaging methods.

Comparative performance is also evaluated for the 2-class problem in BCI competition III dataset IIIb as shown in table 4. We select oparational bands with ($G = 7$). The classification

Table 3. Classification accuracy (%) of TSM, MTSM, and MTSMS for dataset JK-HH1(4-class problem).

Methods	Averaging functions	Subjects					AVG
		sa	sb	sc	sd	se	
TSM	A_{mean}	60.3 ± 0.071	34.8 ± 0.043	29.0 ± 0.061	44.5 ± 0.045	71.3 ± 0.009	46.82 ± 18.85
	RG_{mean}	62.5 ± 0.077	33.0 ± 0.031	31.5 ± 0.033	36.5 ± 0.027	70.8 ± 0.056	46.86 ± 18.39
	$L\text{-}E_{mean}$	59.5 ± 0.069	33.5 ± 0.017	29.3 ± 0.057	40.8 ± 0.060	71.3 ± 0.044	46.88 ± 17.89
	H_{mean}	60.5 ± 0.073	33.3 ± 0.031	29.8 ± 0.069	37.8 ± 0.035	71.5 ± 0.028	46.58 ± 18.36
	EG_{med}	59.3 ± 0.071	31.5 ± 0.038	30.0 ± 0.050	38.3 ± 0.038	69.8 ± 0.025	45.78 ± 17.80
	RG_{med}	62.0 ± 0.079	33.0 ± 0.035	31.3 ± 0.033	37.0 ± 0.026	70.0 ± 0.058	46.66 ± 17.99
	$L\text{-}E_{med}$	61.8 ± 0.091	35.3 ± 0.025	29.0 ± 0.054	40.0 ± 0.041	72.5 ± 0.041	47.72 ± 18.55
	TR_{mean}	62.5 ± 0.069	35.5 ± 0.050	32.0 ± 0.034	44.3 ± 0.045	73.0 ± 0.043	49.46 ± 17.68
	$TL\text{-}E_{mean}$	60.8 ± 0.074	38.5 ± 0.042	30.3 ± 0.047	42.8 ± 0.050	73.5 ± 0.039	49.18 ± 17.59
	TR_{med}	63.0 ± 0.069	36.3 ± 0.035	31.8 ± 0.035	42.5 ± 0.043	72.8 ± 0.039	49.28 ± 17.75
	$TR\text{-}E_{med}$	63.5 ± 0.083	39.5 ± 0.048	30.8 ± 0.042	43.5 ± 0.040	73.5 ± 0.041	50.16 ± 17.71
MTSM	A_{mean}	61.0 ± 0.025	35.5 ± 0.033	29.8 ± 0.051	39.5 ± 0.069	72.8 ± 0.029	47.72 ± 17.89
	RG_{mean}	60.5 ± 0.0574	36.3 ± 0.038	30.3 ± 0.049	37.8 ± 0.067	74.5 ± 0.043	47.88 ± 18.97
	$L\text{-}E_{mean}$	59.8 ± 0.031	37.0 ± 0.035	31.0 ± 0.070	38.8 ± 0.050	72.0 ± 0.022	47.72 ± 17.15
	H_{mean}	61.3 ± 0.068	35.3 ± 0.040	31.5 ± 0.078	40.0 ± 0.077	71.5 ± 0.025	47.92 ± 17.78
	EG_{med}	59.5 ± 0.045	37.3 ± 0.029	29.8 ± 0.042	37.3 ± 0.065	72.3 ± 0.032	47.24 ± 17.75
	RG_{med}	60.8 ± 0.055	36.5 ± 0.045	30.0 ± 0.056	37.8 ± 0.065	73.8 ± 0.033	47.78 ± 18.74
	$L\text{-}E_{med}$	60.8 ± 0.028	38.0 ± 0.028	30.5 ± 0.071	37.8 ± 0.043	72.3 ± 0.034	47.88 ± 17.49
	TR_{mean}	63.5 ± 0.041	39.3 ± 0.036	31.5 ± 0.051	40.0 ± 0.061	74.5 ± 0.029	49.76 ± 17.81
	$TL\text{-}E_{mean}$	62.8 ± 0.049	39.0 ± 0.026	32.3 ± 0.051	39.3 ± 0.062	73.3 ± 0.025	49.34 ± 17.26
	TR_{med}	63.0 ± 0.043	39.5 ± 0.039	32.5 ± 0.054	40.8 ± 0.063	73.8 ± 0.029	49.92 ± 17.37
	$TR\text{-}E_{med}$	62.0 ± 0.054	39.5 ± 0.031	32.3 ± 0.045	40.0 ± 0.060	74.3 ± 0.024	49.62 ± 17.31
MTSMS	A_{mean}	61.8 ± 0.060	36.3 ± 0.062	32.5 ± 0.053	46.0 ± 0.063	72.8 ± 0.045	49.88 ± 17.09
	RG_{mean}	62.8 ± 0.041	34.5 ± 0.038	30.8 ± 0.041	35.3 ± 0.059	74.5 ± 0.029	47.58 ± 19.74
	$L\text{-}E_{mean}$	62.3 ± 0.028	35.3 ± 0.012	31.3 ± 0.038	40.3 ± 0.069	72.0 ± 0.035	48.24 ± 17.88
	H_{mean}	62.3 ± 0.066	35.3 ± 0.020	32.8 ± 0.031	40.0 ± 0.066	71.5 ± 0.026	48.38 ± 17.40
	EG_{med}	63.3 ± 0.057	37.3 ± 0.031	30.8 ± 0.022	39.3 ± 0.071	72.3 ± 0.039	48.60 ± 18.08
	RG_{med}	63.8 ± 0.045	36.3 ± 0.031	30.8 ± 0.054	39.5 ± 0.041	73.8 ± 0.025	48.84 ± 18.81
	$L\text{-}E_{med}$	62.3 ± 0.034	36.0 ± 0.023	30.3 ± 0.053	42.3 ± 0.047	72.3 ± 0.042	48.64 ± 17.78
	TR_{mean}	**65.8 ± 0.061**	39.0 ± 0.037	**35.5 ± 0.040**	**46.8 ± 0.047**	**74.5 ± 0.030**	**52.32 ± 17.11**
	$TL\text{-}E_{mean}$	65.0 ± 0.055	**40.3 ± 0.027**	33.5 ± 0.049	44.3 ± 0.051	73.3 ± 0.026	51.28 ± 17.01
	TR_{med}	64.3 ± 0.062	39.5 ± 0.032	33.3 ± 0.033	44.0 ± 0.048	73.8 ± 0.028	50.98 ± 17.20
	$TR\text{-}E_{med}$	64.0 ± 0.059	39.8 ± 0.027	35.3 ± 0.046	46.0 ± 0.051	74.3 ± 0.026	51.88 ± 16.62

accuracy of the proposed subband selection-based algorithm is the highest in subject S4b. The trimmed based averaging methods also perform better than simple mean and median. It is noted that the multiband TSM method [31] performs best for subject X11b.

3.2.3. Performance analysis of the datasets with statistical test. The performances of the mentioned methods regarding average classification for the individual datasets (IIa, IIIa, IIIb, and JK-HH1) are evaluated as illustrated in figure 4. This summarization is performed using the results of tables 2–4 concerning arithmetic mean. Average accuracy is computed for each subject in the respective dataset. In all cases, accuracy of the subband selection-based approach is higher than other methods. In dataset 2a, the performance of MTSM is very close to that of the proposed MTSMS. TSM has the next highest accuracy rate in dataset IIIa but the performance of TSM and MTSM are close to each other. The average accuracy of TSM is higher than MTSM, and both are close to MTSMS for JK-HH1. With dataset IIIb, the classification accuracy of MTSM is very close to MTSMS, whereas the performance of TSM is considerably lower than MTSM. All the methods perform worst with dataset JK-HH1. To test the statistical significance of the methods, the Wilcoxon signed rank test was performed. From the results of a paired Wilcoxon signed rank test, the proposed MTSMS were significantly improves the performance of MI-BCI across subjects of the datasets than the other methods (MTSMS versus TSM: $p < 0.001$; MTSMS versus MTSM: $p < 0.02$). However, the significance of performance using TSM, MTSM, and the proposed MTSMS approaches was examined with arithmetic mean as a reference method proposed in [27]. To observe the effect of the performances according to the various averaging functions with respect to methods (TSM, MTSM, and MTSMS), we examined the variation of average accuracy shown in figure 5. The average accuracy is estimated across subjects of the datasets.

Table 4. Classification accuracy (%) of TSM, MTSM, and MTSMS for BCI competition III: dataset IIIa (4-class problem) and dataset IIIb (2-class problem).

Methods	Averaging function	Dataset IIIa Subjects k3b	k6b	l1b	AVG	Dataset IIIb Subjects S4b	X11b	AVG
TSM	A_{mean}	91.1 ± 0.024	57.5 ± 0.081	73.3 ± 0.051	73.96 ± 16.80	50.9 ± 0.046	65.0 ± 0.054	57.95 ± 9.97
	RG_{mean}	91.1 ± 0.015	55.8 ± 0.095	65.0 ± 0.051	70.63 ± 18.31	50.9 ± 0.046	65.2 ± 0.055	58.05 ± 10.11
	$L\text{-}E_{mean}$	91.7 ± 0.043	55.8 ± 0.114	70.8 ± 0.090	72.76 ± 18.03	50.7 ± 0.046	65.0 ± 0.055	57.85 ± 10.11
	H_{mean}	91.1 ± 0.015	56.7 ± 0.093	70.0 ± 0.095	72.60 ± 17.34	50.5 ± 0.043	65.3 ± 0.052	57.90 ± 10.46
	EG_{med}	91.1 ± 0.031	55.9 ± 0.051	67.5 ± 0.061	71.50 ± 17.93	51.1 ± 0.044	65.2 ± 0.054	58.15 ± 9.97
	RG_{med}	91.1 ± 0.015	55.9 ± 0.095	66.6 ± 0.051	71.20 ± 18.04	50.9 ± 0.046	65.3 ± 0.054	58.10 ± 10.18
	$L\text{-}E_{med}$	91.7 ± 0.049	53.3 ± 0.047	71.6 ± 0.078	72.20 ± 19.20	50.7 ± 0.043	65.0 ± 0.055	57.85 ± 10.11
	TR_{mean}	93.3 ± 0.026	59.2 ± 0.117	71.6 ± 0.067	74.70 ± 17.26	50.9 ± 0.045	65.3 ± 0.055	58.10 ± 10.18
	$TL\text{-}E_{mean}$	92.8 ± 0.037	62.5 ± 0.089	72.5 ± 0.052	75.93 ± 15.43	50.9 ± 0.045	65.2 ± 0.056	58.05 ± 10.11
	TR_{med}	93.3 ± 0.029	61.7 ± 0.112	70.8 ± 0.066	75.26 ± 16.26	50.9 ± 0.045	65.3 ± 0.054	58.10 ± 10.18
	$TR\text{-}E_{med}$	92.8 ± 0.034	60.8 ± 0.092	73.3 ± 0.055	75.63 ± 16.12	50.9 ± 0.044	65.3 ± 0.053	58.10 ± 10.18
MTSM	A_{mean}	92.2 ± 0.063	58.3 ± 0.109	67.5 ± 0.057	72.66 ± 17.53	74.4 ± 0.069	79.3 ± 0.075	76.85 ± 3.46
	RG_{mean}	92.2 ± 0.034	58.3 ± 0.090	65.0 ± 0.050	71.83 ± 17.95	74.6 ± 0.069	79.1 ± 0.071	76.85 ± 3.18
	$L\text{-}E_{mean}$	90.6 ± 0.044	55.0 ± 0.076	64.2 ± 0.028	69.93 ± 18.47	74.6 ± 0.069	79.3 ± 0.074	76.95 ± 3.32
	H_{mean}	90.6 ± 0.041	55.0 ± 0.101	71.7 ± 0.038	72.43 ± 17.81	74.4 ± 0.066	79.6 ± 0.075	77.00 ± 1.55
	EG_{med}	91.7 ± 0.048	52.5 ± 0.114	65.8 ± 0.094	70.00 ± 19.93	74.4 ± 0.069	79.3 ± 0.075	76.85 ± 3.46
	RG_{med}	92.2 ± 0.038	57.5 ± 0.101	63.3 ± 0.028	71.00 ± 18.58	75.0 ± 0.069	79.3 ± 0.077	77.15 ± 3.04
	$L\text{-}E_{med}$	90.6 ± 0.044	56.7 ± 0.076	65.0 ± 0.025	70.76 ± 17.67	74.6 ± 0.069	79.4 ± 0.071	77.00 ± 3.39
	TR_{mean}	93.3 ± 0.041	63.3 ± 0.083	73.3 ± 0.049	76.63 ± 15.27	75.2 ± 0.069	**80.2 ± 0.073**	77.70 ± 3.53
	$TL\text{-}E_{mean}$	**95.6 ± 0.033**	59.2 ± 0.084	69.2 ± 0.051	74.66 ± 18.80	75.2 ± 0.068	80.0 ± 0.073	77.60 ± 3.39
	TR_{med}	94.4 ± 0.034	63.3 ± 0.088	73.3 ± 0.050	77.00 ± 15.87	75.2 ± 0.069	**80.2 ± 0.078**	77.70 ± 3.53
	$TR\text{-}E_{med}$	95.0 ± 0.035	57.5 ± 0.086	68.3 ± 0.051	73.60 ± 19.30	75.0 ± 0.069	**80.2 ± 0.078**	77.60 ± 3.67
MTSMS	A_{mean}	91.7 ± 0.065	67.5 ± 0.120	76.7 ± 0.052	78.63 ± 12.21	80.6 ± 0.074	77.2 ± 0.053	78.90 ± 2.40
	RG_{mean}	92.2 ± 0.040	68.3 ± 0.095	77.5 ± 0.061	79.33 ± 12.05	81.1 ± 0.071	76.9 ± 0.059	79.00 ± 2.96
	$L\text{-}E_{mean}$	90.0 ± 0.036	62.5 ± 0.082	81.7 ± 0.091	78.06 ± 14.10	80.9 ± 0.077	76.7 ± 0.059	78.80 ± 2.96
	H_{mean}	91.1 ± 0.039	62.5 ± 0.108	79.2 ± 0.097	77.60 ± 14.36	78.9 ± 0.073	77.2 ± 0.061	78.05 ± 1.20
	EG_{med}	93.3 ± 0.049	65.0 ± 0.119	80.0 ± 0.063	79.43 ± 14.15	80.2 ± 0.074	77.2 ± 0.061	78.70 ± 2.12
	RG_{med}	92.2 ± 0.039	66.7 ± 0.100	78.3 ± 0.063	79.06 ± 12.76	79.4 ± 0.076	77.2 ± 0.059	78.30 ± 1.55
	$L\text{-}E_{med}$	91.7 ± 0.048	65.0 ± 0.086	78.3 ± 0.080	78.33 ± 13.35	79.8 ± 0.071	77.0 ± 0.060	78.40 ± 1.97
	TR_{mean}	93.3 ± 0.036	**70.0 ± 0.087**	**84.2 ± 0.068**	82.50 ± 11.74	**81.5 ± 0.075**	78.0 ± 0.058	**79.75 ± 2.47**
	$TL\text{-}E_{mean}$	95.3 ± 0.040	66.7 ± 0.092	81.7 ± 0.061	81.23 ± 14.30	81.3 ± 0.076	78.0 ± 0.059	79.65 ± 2.33
	TR_{med}	94.4 ± 0.038	**70.0 ± 0.091**	83.3 ± 0.069	82.56 ± 12.21	80.6 ± 0.076	77.4 ± 0.055	79.00 ± 2.26
	$TR\text{-}E_{med}$	95.3 ± 0.040	69.2 ± 0.089	83.3 ± 0.058	**82.60 ± 13.06**	80.6 ± 0.074	77.4 ± 0.056	79.00 ± 2.26

The result shows that the proposed trimmed based averaging approaches outperform the simple mean- and median-based approaches.

3.2.4. Computational cost. The computational time at the testing phase is measured using MATLAB R2014b on MacBook Pro (with Intel Core i5 processor and 8 GB RAM). The average computational time at the testing phase required in single-trial analysis for the proposed MTSMS with a different number of subbands are simulated in figure 6. Moreover, the proposed method was the extension of TSM with multiband approach, therefore the computation time increases linearly with the number of subbands. Note that the case when $N = 1$ with MTSMS corresponds to the TSM excluding mutual information stage during processing time. BCI competition III dataset IIIa for subject k3b (recording was performed with a 64-channel EEG for 4-class BCI implementation) and dataset IIIb for subject s4b (EEG was recored with two channels for 2-class MI-BCI) were considered, since the computational time increases with increased number of electrodes. The average computational time for subject k3b with 60 channels data was 0.1612 s for first subband (first subband equivalents to TSM) shown in figure 6 (a). Since, the recognition time increases linearly with the number of subbands increased, the time complexity of our proposed methods is same as the TSM method. Similarly, the average computational time with TSM for subject s4b was 0.010 s shown in figure 6 (b). Therefore, the average computational time suggests that the proposed methods are possible to implement the MI-BCI for real time applications.

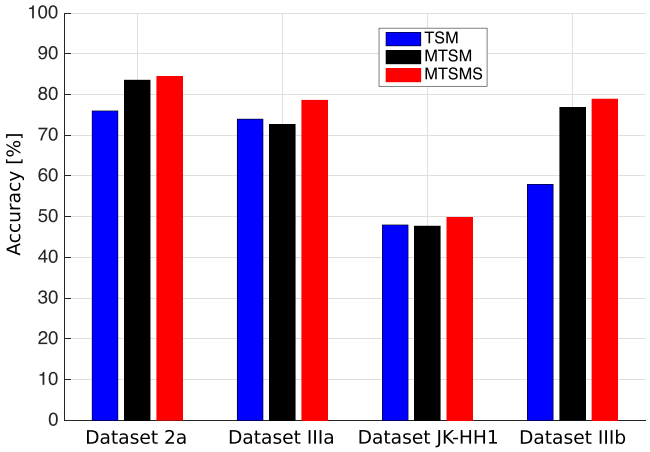

Figure 4. Average recognition accuracy obtained on four datasets: BCI III (dataset 2a, dataset IIIa, and dataset IIIb) and dataset JK-HH1 using three different methods, TSM, MTSM, and MTSMS (proposed method).

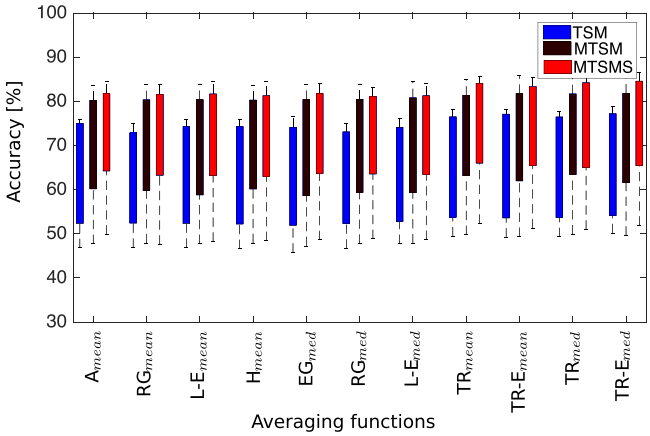

Figure 5. Box plots of average recognition accuracy of the different averaging approaches with respect to TSM, MTSM, and proposed MTSMS. The horizontal-axis represents the different averaging approaches and colors represent methods (TSM, MTSM, and proposed MTSMS).

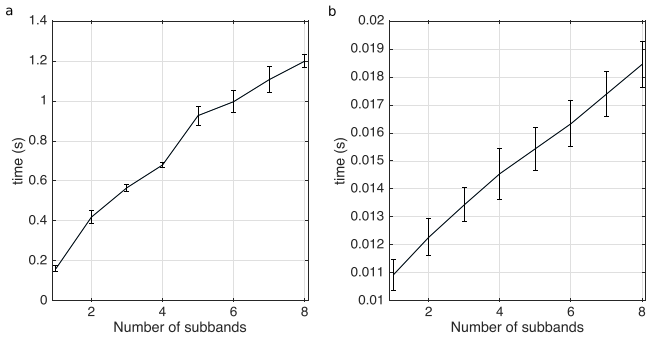

Figure 6. Average computational time (s) at the testing phase for single-trial detection. (a) BCI competition dataset IIIa with 60-channels for subject k3b and (b) BCI competition dataset IIIb with 2-channels data for subject s4b. In both case, the results for 200 runs were averaged.

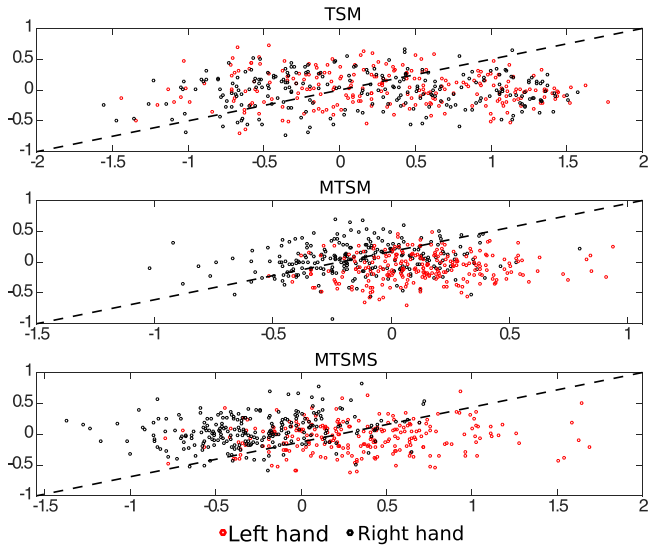

Figure 7. Scatter plot of features with TSM, MTSM, and MTSMS in case of arithmetic mean for BCI competition dataset IIIb (subject S4b): the horizontal-axis represrnts the value of features $(V_0)_1$ correspond to left hand movement and the vertical-axis represents the value of features $(V_0)_2$ correspond to right hand movement.

4. Discussion and conclusion

With statistical analysis, the performance of the proposed MTSMS approach exhibits its superiority for all four dataset used in the experiment. The proposed MTSMS method chooses the most dominant subbands for MI classification using a support vector classifier. Several studies have employed linear discriminant analysis (LDA) after feature selection for motor imagery classification [27, 47, 48]. It was proven in [30], TSM with SVM classifier outperformed the LDA classifier. However, in the case of non-invasive BCI, the major challenge is to extract reliable information as well as the inequality of the energy distribution of EEG signal over channels. To address these problems, several studies introduced multiband approaches to enhance the performance of MI-BCI considering subject specific frequency bands [17, 18, 20–22, 31].

To compare the effective features for each of the methods, the distribution of two features $(V_0)_1$ and $(V_0)_2$ with TSM, MTSM, and MTSMS are depicted in figure 7, where $(V_0)_1$ and $(V_0)_2$ indicate the first and second components of V_0 defined in (13), respectively. BCI competition III dataset IIIa for subject S4b is used for left and right hand movement representing by red and black open circles, respectively. It is observed from figure 7 that the extracted features of MTSMS method enhanced discriminability between two classes. This can be the underlying reason for the improvement in recognition accuracy. Moreover, selecting subbands by analyzing mutual information works well in increasing the discriminability. An interesting feature found in figure 3 is that the value of mutual information is relatively higher for alpha or beta bands and sometimes in both bands. Using these bands improves MI

classification accuracy and hence it confirms influential activities of alpha and beta in motor imagery. In this study, the real world data with noise were used to address the robustness of the methods that are highly relevant with practical BCI systems. Considering the accuracy, the statistical significance of the methods as well as the separability of classes (in figure 7), it is clearly revealed that the proposed method is more robust compared with TSM and MTSM methods.

With BCI datasets for 2-class problems, a comparison study with variants of CSP including CSP [11], FBCSP [17, 18], DFBCSP [21], SFBCSP [22], a sparse Bayesian learning of filter-banks (SBLFB) [49], and CSP-TSM approaches has been reported by Kumar *et al* [28]. In [28], TSM approach was implemented in the case of simple mean (arithmetic mean) to compare CSP, TSM, and CSP-TSM. The performance of TSM and CSP-TSM approaches outperformed the CSP method. Moreover, a recent study [30] has compared the performance of TSM and CSP-based MI-BCI. In the experiment, TSM and CSP methods were applied on a partial set of classes (considering 2-class problem for fair comparison) with BCI dataset 2a, IIIa, IIb and dataset JKHH-1. The classification performances of TSM with arithmetic mean were 5.16%, 3.13%, and 6.4% higher than CSP for BCI dataset 2a, IIIa, and JKHH-1 respectively expect only 1.95% lower than CSP for BCI dataset IIIb (2-class dataset). Compared with our proposed methods and TSM with arithmetic mean, the average performances of proposed system with trimmed based averaging approaches for 4-class problems are 10.56%, 8.64%, and 5.50% higher than TSM with simple mean (see table 2–4 with BCI datasets 2a, IIIa, and dataset JK-HH1). With dataset IIIb for 2-class problems, the performance of the proposed methods is approximately 22% higher than TSM with simple mean (see table 4). Therefore, the use of trimmed-based averaging methods can enhance the classification accuracy of MI-based BCI, which is evaluated in proposed approaches. The reason of using trimmed averaging is to deal with real world EEG signals by estimating outliers according to the distance from averages. It is more robust than the traditional arithmetic mean [23, 30].

Further research is required to reduce the computational cost without compromising performance toward various kinds of practical real-world BCI applications. A recent promising application of MI-BCI studies is a close loop neurofeedback [50–52]. In such a technology, the estimation of average computational time for single-trial detection at the testing phase is the major factor to implement MI-BCI. Compared to other frequency decomposition approaches such as wavelets and EMD-based methods [3, 53–55], which need huge computational cost for decomposition, the proposed approach needs less computational complexity and is suitable for practical BCI applications. Barachant *et al* reported that the computational time dramatically increases with the number of electrodes. However, the computational cost of a Riemannian distance could be reduced using metrics approximating the Riemannian metric [32]. Kumar *et al* also has observed that use of TSM with spatial filter signal using CSP can reduce the computational cost [28]. The reason of using spatial filters with CSP can reduce the size of sample covariance matrices, which facilitate the reduction of computation load, even though the number of electrodes was large. However, use of spatial filter using CSP can restrict multi-class BCI. Therefore, to design multi-class MI-BCI as more usability, further investigations are necessary to reduce sample covariance matrix (SCM) of EEG. The reduction of dimensions with graph Fourier transform (GFT) [56–58] would be more usability to improve the proposed method, implementing multi-class MI-BCI.

It is also important to detect a user's resting state in the neurorehabilitation applications [59, 60]. The CSP-based methods were used to detect a resting state [61, 62]. Further investigation is required to classify the observed brain activation as either movement or rest using our proposed methods. Therefore, there are several avenues for further research to implement online multi-class MI-BCI with reduced calibration time.

Acknowledgments

This work was supported in part by the Japan Society for Promotion of Science (JSPS) under KAKENHI, grant number 15H04002.

ORCID iDs

Md Rabiul Islam https://orcid.org/0000-0002-4514-8370
Toshihisa Tanaka https://orcid.org/0000-0002-5056-9508

References

[1] Vidal J J 1973 Toward direct brain-computer communication *Annu. Rev. Biophys. Bioeng.* **2** 157–80
[2] Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain-computer interfaces for communication and control *Clin. Neurophysiol.* **113** 767–91
[3] Wolpaw J R, Birbaumer N, Heetderks W J, McFarland D J, Peckham P H, Schalk G, Donchin E, Quatrano L A, Robinson C J and Vaughan T M 2000 Brain-computer interface technology: a review of the first international meeting *IEEE Trans. Rehabil. Eng.* **8** 164–73
[4] Posner J B, Saper C B, Schiff N and Plum F 2007 *Plum and Posner's Diagnosis of Stupor and Coma* 4th edn (New York: Oxford University Press)
[5] Cecotti H 2011 Spelling with non-invasive brain-computer interfaces-current and future trends *J. Physiol.-Paris* **105** 106–14
[6] Nuyujukian P, Fan J M, Kao J C, Ryu S I and Shenoy K V 2015 A high-performance keyboard neural prosthesis enabled by task optimization *IEEE Trans. Biomed. Eng.* **62** 21–9
[7] Mulder T 2007 Motor imagery and action observation: cognitive tools for rehabilitation *J. Neural Transm.* **114** 1265–78
[8] van Dokkum L, Ward T and Laffont I 2015 Brain computer interfaces for neurorehabilitation its current status as a rehabilitation strategy post-stroke *Ann. Phys. Rehabil. Med.* **58** 3–8

[9] Soekadar S R, Birbaumer N, Slutzky M W and Cohen L G 2015 Brain machine interfaces in neurorehabilitation of stroke *Neurobiol. Dis.* **83** 172–9

[10] Pfurtscheller G, Neuper C, Schlogl A and Lugger K 1998 Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters *IEEE Trans. Rehabil. Eng.* **6** 316–25

[11] Ramoser H, Muller-Gerking J and Pfurtscheller G 2000 Optimal spatial filtering of single trial EEG during imagined hand movement *IEEE Trans. Rehabil. Eng.* **8** 441–6

[12] Dornhege G, Blankertz B, Curio G and Muller K R 2004 Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms *IEEE Trans. Biomed. Eng.* **51** 993–1002

[13] Sanei S and Chambers J A 2007 *EEG Signal Processing* (New York: Wiley)

[14] Pfurtscheller G and da Silva F L 1999 Event-related EEG/MEG synchronization and desynchronization: basic principles *Clin. Neurophysiol.* **110** 1842–57

[15] McFarland D J, Miner L A, Vaughan T M and Wolpaw J R 2000 Mu and beta rhythm topographies during motor imagery and actual movements *Brain Topogr.* **12** 177–86

[16] Pfurtscheller G, Pregenzer M and Neuper C 1994 Visualization of sensorimotor areas involved in preparation for hand movement based on classification of μ and central β rhythms in single EEG trials in man *Neurosci. Lett.* **181** 43–6

[17] Ang K K, Chin Z Y, Zhang H and Guan C 2008 Filter bank common spatial pattern (FBCSP) in brain-computer interface *IEEE Int. Joint Conf. on Neural Networks (IEEE World Congress on Computational Intelligence)* pp 2390–7

[18] Ang K K, 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

[19] Brodu N, Lotte F and Lcuyer A 2011 Comparative study of band-power extraction techniques for motor imagery classification *IEEE Symp. on Computational Intelligence, Cognitive Algorithms, Mind and Brain (CCMB)* pp 1–6

[20] Novi Q, Guan C, Dat T H and Xue P 2007 Sub-band common spatial pattern (SBCSP) for brain-computer interface *3rd Int. IEEE/EMBS Conf. on Neural Engineering* pp 204–7

[21] Thomas K P, Guan C, Lau C T, Vinod A P and Ang K K 2009 A new discriminative common spatial pattern method for motor imagery brain-computer interfaces *IEEE Trans. Biomed. Eng.* **56** 2730–3

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

[23] Barbaresco F 2008 Innovative tools for radar signal processing based on cartan geometry of spd matrices and information geometry *IEEE Radar Conf.* pp 1–6

[24] Fletcher P T and Joshi S 2004 *Principal Geodesic Analysis on Symmetric Spaces: Statistics of Diffusion Tensors* (Berlin: Springer) pp 87–98

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

[26] Li Y, Wong K M and deBruin H 2009 EEG signal classification based on a riemannian distance measure *IEEE Toronto Int. Conf. Science and Technology for Humanity* pp 268–73

[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] Kumar S, Mamun K and Sharma A 2017 CSP-TSM: optimizing the performance of riemannian tangent space mapping using common spatial pattern for MI-BCI *Comput. Biol. Med.* **91** 231–42

[29] Uehara T, Tanaka T and Fiori S 2016 Robust averaging of covariance matrices by Riemannian geometry for motor-imagery brain–computer interfacing *Advances in Cognitive Neurodynamics (V):* Proc. of the 5th Int. Conf. on Cognitive Neurodynamics—2015 (Singapore: Springer) pp 347–53

[30] Uehara T, Sartori M, Tanaka T and Fiori S 2017 Robust averaging of covariances for EEG recordings classification in motor imagery brain-computer interfaces *Neural Comput.* **29** 1631–66

[31] Islam M R, Tanaka T, Akter M S and Molla M K I 2017 Classification of motor imagery BCI using multiband tangent space mapping *IEEE Int. Conf. on Digital Signal Processing* pp 1–5

[32] Arsigny V, Fillard P, Pennec X and Ayache N 2007 Geometric means in a novel vector space structure on symmetric positive definite matrices *SIAM J. Matrix Anal. Appl.* **29** 328–47

[33] Islam M R, Rashed-Al-Mahfuz M, Ahmad S and Molla M K I 2012 Multiband prediction model for financial time series with multivariate empirical mode decomposition *Discrete Dyn. Nat. Soc.* **2012** 21

[34] Chen X, Wang Y, Gao S, Jung T-P and Gao X 2015 Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain computer interface *J. Neural Eng.* **12** 046008

[35] Islam M R, Molla M K I, Nakanishi M and Tanaka T 2017 Unsupervised frequency-recognition method of SSVEPs using a filter bank implementation of binary subband CCA *J. Neural Eng.* **14** 026007

[36] Higashi H and Tanaka T 2013 Simultaneous design of FIR filter banks and spatial patterns for EEG signal classification *IEEE Trans. Biomed. Eng.* **60** 1100–10

[37] Peng H, Long F and Ding C 2005 Feature selection based on mutual information: criteria of max-dependency, max-relevance and min-redundancy *IEEE Trans. Pattern Anal. Mach. Intell.* **27** 1226–38

[38] Battiti R 1994 Using mutual information for selecting features in supervised neural net learning *IEEE Trans. Neural Netw.* **5** 537–50

[39] Benjamini Y and Hochberg Y 1997 Multiple hypotheses testing with weights *Scand. J. Stat.* **24** 407–18

[40] Yger F, Lotte F and Sugiyama M 2015 Averaging covariance matrices for EEG signal classification based on the CSP: an empirical study *23rd European Signal Processing Conf.* pp 2721–5

[41] Weiszfeld E 1937 Sur le point pour lequel la somme des distances de n points donnes est minimum *Tohoku Math. J.* **43** 355–86

[42] Tomida N, Tanaka T, Ono S, Yamagishi M and Higashi H 2015 Active data selection for motor imagery EEG classification *IEEE Trans. Biomed. Eng.* **62** 458–67

[43] Oostenveld R and Praamstra P 2001 The five percent electrode system for high-resolution EEG and ERP measurements *Clin. Neurophysiol.* **112** 713–9

[44] Blankertz B and Curio G 2007 The non-invasive berlin Brain-computer interface: fast acquisition of effective performance in untrained subjects *NeuroImage* **37** 539–50

[45] Blankertz B, Tomioka R, Lemm S, Kawanabe M and Müller K R 2008 Designing optimal spatial filters for single-trial EEG classification in a movement task *IEEE Signal Process. Mag.* **25** 41–56

[46] Müller-Gerking J, Pfurtscheller G and Flyvbjerg H 2008 Optimizing spatial filters for robust EEG single-trial analysis *Clin. Neurophysiol.* **110** 787–98

[47] Kumar S, Sharma R, Sharma A and Tsunoda T 2016 Decimation filter with common spatial pattern and fishers

discriminant analysis for motor imagery classification *Int. Joint Conf. on Neural Networks* pp 2090–5

[48] Kim H S, Chang M H, Lee H J and Park K S 2013 A comparison of classification performance among the various combinations of motor imagery tasks for brain-computer interface *6th Int. IEEE/EMBS Conf. on Neural Engineering* pp 435–8

[49] Zhang Y, Wang Y, Jin J and Wang X 2017 Sparse bayesian learning for obtaining sparsity of EEG frequency bands based feature vectors in motor imagery classification *Int. J. Neural Syst.* **27** 1650032

[50] Gomez-Rodriguez M, Peters J, Hill J, Schlkopf B, Gharabaghi A and Grosse-Wentrup M 2011 Closing the sensorimotor loop: haptic feedback facilitates decoding of motor imagery *J. Neural Eng.* **8** 036005

[51] Ramos-Murguialday A *et al* 2013 Brain-machine interface in chronic stroke rehabilitation: a controlled study *Ann. Neurol.* **74** 100–8

[52] Remsik A *et al* 2016 A review of the progression and future implications of brain-computer interface therapies for restoration of distal upper extremity motor function after stroke *Expert Rev. Med. Devices* **13** 445–54

[53] Park C, Looney D, ur Rehman N., Ahrabian A and Mandic D P 2013 Classification of motor imagery BCI using multivariate empirical mode decomposition *IEEE Trans. Neural Syst. Rehabil. Eng.* **21** 10–22

[54] Molla M K I, Tanaka T, Rutkowski T M and Tanaka K 2013 Phase synchronization analysis of EEG channels using bivariate empirical mode decomposition *2013 IEEE Int. Conf. on Acoustics, Speech and Signal Processing* pp 1182–6

[55] Diez P F, Mut V, Laciar E, Torres A and Avila E 2009 Application of the empirical mode decomposition to the extraction of features from EEG signals for mental task classification *Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society* pp 2579–82

[56] Tanaka T, Uehara T and Tanaka Y 2016 Dimensionality reduction of sample covariance matrices by graph Fourier transform for motor imagery brain-machine interface *IEEE Statistical Signal Processing Workshop* pp 1–5

[57] Higashi H, Tanaka T and Tanaka Y 2014 Smoothing of spatial filter by graph fourier transform for EEG signals *Asia-Pacific Signal and Information Processing Association Annual Summit and Conf.* pp 1–8

[58] Shuman D I, Narang S K, Frossard P, Ortega A and Vandergheynst P 2013 The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains *IEEE Signal Process. Mag.* **30** 83–98

[59] Ahmad R F, Malik A S, Amin H U, Kamel N and Reza F 2016 Classification of cognitive and resting states of the brain using EEG features *IEEE Int. Symp. on Medical Measurements and Applications* pp 1–5

[60] Cao Z H, Ko L W, Lai K L, Huang S B, Wang S J and Lin C T 2015 Classification of migraine stages based on resting-state EEG power *Int. Joint Conf. on Neural Networks* pp 1–5

[61] Hurtier J, Dokkum L V, Dalhoumi S, Coffey A, Perrey S, Jourdan C, Dray G, Ward T, Froger J and Laffont I 2016 A closed-loop bci system for rehabilitation of the hemiplegic upper-limb: a performance study of the systems ability to detect intention of movement *Ann. Phys. Rehabil. Med.* **59** e88

[62] Tavakolan M, Yong X, Zhang X and Menon C 2016 Classification scheme for arm motor imagery *J. Med. Biol. Eng.* **36** 12–21