Tangent Space Features based Transfer Learning Classification Model for Two-Class Motor Imagery Brain-Computer Interface

Pramod Gaur
Intelligent Systems Research Centre,
Ulster University, Derry/Londonderry, U.K.
E-mail: p.gaur@ulster.ac.uk

Karl McCreadie
Intelligent Systems Research Centre,
Ulster University, Derry/Londonderry, U.K.

Ram Bilas Pachori
Discipline of Electrical Engineering,
Indian Institute of Technology Indore, Indore, India.

Hui Wang
Computer Science Research Institute,
Ulster University, Jordanstown, U.K.

Girijesh Prasad
Intelligent Systems Research Centre,
Ulster University, Derry/Londonderry, U.K.

The performance of a brain-computer interface (BCI) will generally improve by increasing the volume of training data on which it is trained. However, a classifier's generalization ability is often negatively affected when highly non-stationary data are collected across both sessions and subjects. The aim of this work is to reduce the long calibration time in BCI systems by proposing a transfer learning model which can be used for evaluating unseen single trials for a subject without the need for training session data. A method is proposed which combines a generalization of the previously proposed subject-specific 'multivariate empirical mode decomposition' preprocessing technique by taking a fixed band of 8-30 Hz for all four motor imagery tasks, and a novel classification model which exploits the structure of tangent space features drawn from the Riemannian geometry framework, that is shared among the training data of multiple sessions and subjects. Results demonstrate comparable performance improvement across multiple subjects without subject-specific calibration, when compared with other state-of-the-art techniques.

Keywords: Motor imagery; brain-computer interface (BCI); tangent space, covariance matrix, multivariate empirical mode decomposition (MEMD), subject-specific multivariate empirical mode decomposition based filtering (SS-MEMDBF).

1. Introduction

The brain-computer interface (BCI) is gaining popularity due to its applicability to a number of diverse fields ranging from medicine to gaming. BCI not only allows an alternative control mechanism for healthy users but can also allow those with a disor-

der of consciousness the means for communication or control.[1–5]

Vidal[6] first introduced the term BCI in 1973 before a follow up study by the same author in 1977 demonstrated on-screen cursor control using the visual evoked potential phenomenon. It wasn't until 1988 that Farwell and Donchin utilised the P300 event-related potential to successfully spell a word. In the same year, Bozinovska et al.[7,8] were able to control a buzzer using their Contingent Negative Variation (CNV) potential, whilst Bozinovski et al.[9] were able to demonstrate robotic control using changes in the alpha frequency band for the very first time. More details about the early works carried out in BCI can be found in Bozinovski and Bozinovska.[10]

The field of Artificial Intelligence (AI) has seen a resurgence in recent years along with its subfields of machine learning and artificial neural networks[11–13] with an increasing body of literature appearing on deep networks.[14] Techniques such as these and others used for neural modelling[15,16] attempt to mimic cortical processes and have uses not only in neuroscience but also in engineering[17] and computer science. A BCI is complex device and often utilises techniques from many of these fields along with the collective expertise of physicians, psychologists, hardware, and software engineers to produce a system which is sufficiently accurate and fast as to be practical.

Although the field of BCI can generally be subdivided into those which focus on endogenous and those on exogenous methods,[18] this paper focuses on the former principle of imagined movement, i.e. motor imagery (MI) BCI.[19] A perennial problem in MI based BCI, is the length of time it typically takes to train a machine learning model to classify modulations in a user's brainwaves. Additionally, it is generally accepted that just because one classifier produces a high accuracy for one participant, this does not necessarily mean that it will do so in another, or even for that matter, on the same participant on a different day. In other words, there is rarely a one-size-fits-all solution in machine learning, or at least it has yet to be discovered.

Initially in MI BCI, subjects had to adapt to a classifier with invariant parameters, and although early studies demonstrated some level of success using this approach, it would often take an impractically long time to train the user. For instance, Bozinovska et al. (1990, 1992),[20,21] in their closed-loop paradigm using the contingent negative variation (CNV) phenomena,[20] were able to exploit the expectation process to effect a change.

These attempts, though largely successful, required a classifier to be retrained before each testing period adding further unnecessary delay. One method was to train on all previous sessions data, which has the side-effect of being drawn from differing probability distributions, leading to a significant estimation bias.[22] A modern approach to this problem is to find a pattern or shared structure in the differences in these data taken both from different sessions and often from different subjects.[23]

Bozinovska et al.[20,21] used a stimulus based closed-loop CNV paradigm which needed no initial training and is a good example of a zero-training BCI. Common spatial patterns (CSP), which frequently appear in the literature, utilise covariance matrices as a feature space, as do studies based on Riemannian geometry which however, appear much less often. In the past, researchers have applied several different transfer learning methods in conjunction with CSP. In a move towards the coveted "zero training BCI", Krauledat et al.[24] developed an online system which demonstrated that performance did not suffer even in the absence of a calibration session, due to the application of spatial filters generated from previous session data, which generalise better than those obtained from a single session.

Similarly, others[25] have also demonstrated their zero training subject-independent system by using a general classifier trained on a large dataset taken from forty-five subjects. By using subject-specific spatial and temporal filters, Fazli et al.[25] allowed new BCI users immediate control in real time with only a slight drop in performance. Others have performed regularisation of spatial filters[26,27] to achieve effective subject-to-subject transfer, whilst more recently, researchers have turned their attention to Riemannian geometry[28–30] as a tool to achieve the same ends. The non-stationarity inherent in electroencephalogram (EEG)/magnetoencephalography (MEG) signals is another significant problem in BCI and is often exhibited in inter-session transfers. Also known as a covariate shift, this can cause classifier performance to deteriorate over time. As both EEG and MEG are often recorded using multiple channels over multiple trials, this can lead to a high dimensionality further impacting on classification ac-

curacy. Sugiyama et al.[31] use covariate shift adaptation in conjunction with an importance weighted cross-validation method to address the inherent biasedness of traditional model selection techniques. Others such as Satti et al.,[32] have also used similar methods by applying a covariate shift minimization technique to variances in CSP filtered channels. These techniques often rely on the relabelling of offline data as does the similar process of employing boosting-based learning algorithms as seen in Dai et al.[33] EEGNet[34] was also proposed recently as a type of compact convolutional neural network for EEG-based brain-computer interfaces which works well, even when limited training samples are available.

Transfer learning is the process of applying the knowledge gained from one task, to another related activity. The first discussion of transfer learning in machine learning was by Bozinovski.[35]

BCI research is primarily concerned with the challenge of gleaning useful information from a recorded dataset through the identification of some invariant feature(s), often referred to as domain adaptation.[36] It consists of first finding a domain-independent structure (a domain-independent model, feature set, or set of rules) and then adapting it to the new domain of interest. An alternative approach, is to discover some structure in the rules for classification between different sessions or subjects, sometimes referred to as rule adaptation.[23]

This work describes a novel BCI classification technique which exploits the structure of tangent space features drawn from the Riemannian geometry framework which is shared among the training data of multiple sessions and subjects. This is done by proposing a classification model which can be used for evaluating unseen single EEG MI trials of a new subject without creating a classification model using its training session data.

In the training session, a novel tangent space based transfer learning (TS-TL) classification model has been created by exploiting the features obtained from the fixed band multivariate empirical mode decomposition (FB-MEMD) filtering (8-30 Hz) to enhance EEG signals using leave-one-subject-out cross-validation (LOSO-CV). The features share some common information because the feature set is generated whenever a subject is asked to perform the same MI task. The proposed method exploits the shared structure of the tangent space features among the training data of eight subjects and then uses the shared structure model to identify unseen single EEG MI trials from the remaining ninth subject, while applying the LOSO-CV procedure.

The overall aim of this work is to reduce the long calibration time in BCI systems by proposing a model which can be used for evaluating unseen EEG MI single trials for a subject with no training session data. Therefore, the following objectives are proposed:

(1) To study the subject-to-subject non-stationarity present in the EEG signals;
(2) To present the results in terms of classification accuracy when single trials are classified.

2. Dataset

The publicly available BCI competition IV dataset 2A[37] is used for this study. The dataset contains EEG signals from twenty-two EEG channels and three EOG channels with left mastoid used as a reference. Four MI tasks: right hand, left hand, foot, and tongue movements were performed in this dataset. The dataset contains data from nine healthy subjects with each subject having two sessions: one for training, and one for testing. Each session contains 288 trials of MI data with 72 trials for each MI task. The EEG signals contained in this dataset were originally sampled at 250 Hz and then band-passed between 0.5 Hz and 100 Hz. Additionally, a 50 Hz notch filter was applied to remove power-line noise. The selection of time-interval in a single trial period plays an important role in MI classification.

A 2 sec data window, between 0.5 sec and 2.5 sec (Figure 1), is taken after the onset of the visual cue in the training stage, as used by the competition winner.[38] Further details about the BCI competition IV dataset 2A can be obtained from.[37] All twenty-two channels have been considered in this study as shown in Figure 2.

Table 1 shows the number of correct trials (CT) and rejected trials (RT) from all subjects as marked by event 1023 in the evaluation session[37] which denotes trials containing artefacts.

Subject A02 has the least number of the rejected trials. The RT corresponding to each MI tasks are as follow: left hand (LH): 1 trial, right hand (RH): 1 trial, foot: 3 trials and tongue: 0 trials, respectively.
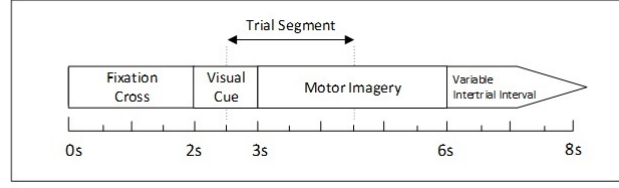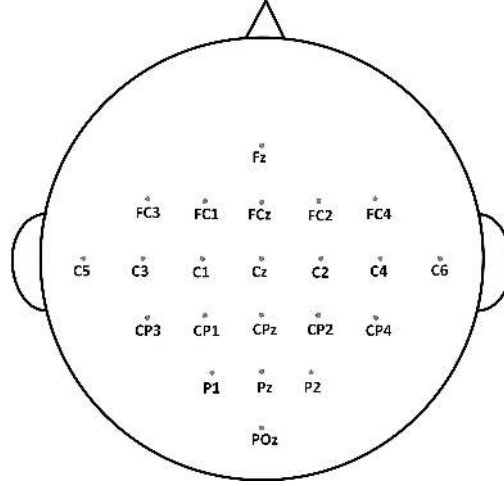
Figure 1.   Trial timing.

Figure 2.   Head plot showing twenty-two channel considered for this study from BCI competition IV dataset 2A.

This gives a total of 5 trials rejected for subject A02 in the evaluation session and similarly Table 1 lists rejected trials for all other subjects.

Table 1.   Subject wise breakdown of correct and rejected trials.

Subject	Number of trials						
	Total	CT	RT	LH	RH	Foot	Tongue
A01	288	281	7	1	2	3	1
A02	288	283	5	1	1	3	0
A03	288	273	15	5	2	4	4
A04	288	228	60	13	15	13	19
A05	288	276	12	2	7	0	3
A06	288	215	73	19	17	18	19
A07	288	277	11	1	3	1	6
A08	288	271	17	6	4	3	4
A09	288	264	24	7	7	3	7

3.   Methods

Motion intentions are classes in an MI BCI. Other BCI tasks may classify into "switch motor", "letter A", "disable buzzer etc. One of the tasks in a BCI is to classify motion intentions into one of two classes. However, the recorded EEG signals are highly subject-specific, sensitive to noise, and have inherent non-stationarities due to changes in the signal properties not only over time, but also within the session as well as across sessions. They may require a long training time, which limits the use of BCI in both patients and healthy individuals. Some BCI paradigms typically require the collection of numerous trials for machine learning techniques to be effective. Hence, this work presents a novel 'tangent space-based transfer learning' (TS-TL) method to improve classification performance. This classification technique exploits the shared structure in the tangent space features among the training data of multiple sessions and subjects, instead of simply combining the training data as in previous studies.[39] Whereas previous work by Gaur et al.,[40] described in the following section, demonstrates the effectiveness of a novel 'subject specific multivariate empirical mode decomposition' (SS-MEMD) based filtering method in the preprocessing stage (Figure 3), the current work, although utilising SS-MEMD, focuses instead on the application of the novel TS-TL

method at the classification stage. Please refer to Figure 4 for an overview of the proposed method.

The enhanced EEG signals from all twenty-two channels have been used to extract the sample covariance matrix as a feature set which contains $n(n+1)/2$ features where n denotes the number of channels. Here, in this study, the number of channels are twenty-two, so the features obtained from the enhanced EEG signals are 253.

3.1. *Multivariate Empirical Mode Decomposition*

EEG signals tend to have a poor signal-to-noise (SNR) ratio and suffer from interference from both external sources, such as power lines, and physiological sources such as electrooculography (EOG) and electromyography (EMG).[41] The EEG signals of interest corresponding to a particular movement (such as μ and β rhythms in the case of motor imagery) are often obscured by these noise and interference sources leading to erroneous results. Therefore, a preprocessing technique is required to filter this noise without weakening the original signal. In 1998, Huang et al. proposed empirical mode decomposition (EMD)[42] which decomposes the original signal into a finite group of band-limited basis functions which are known as intrinsic mode functions (IMFs). Based on this EMD approach, the original signal can be represented in terms of IMFs and a residual as follows:

$$Z(t) = \sum_{k=1}^{p} I_k(t) + Rs_p(t) \qquad (1)$$

where $Z(t)$ denotes the recorded EEG signal in the time domain, $I_k(t)$ represents the kth IMF, and $Rs_p(t)$ gives the residual. Thus, a summation of the selected IMFs can be used to reconstruct the signal of interest.[43] The remaining IMFs are then discarded which may contain artefacts and noise. However, univariate EMD suffers from what is known as a 'mode mixing' issue.[42] To overcome this issue, another research group proposed an extension to EMD which they call ensemble empirical mode decomposition (EEMD).[44] Unfortunately, this is also unsuitable for real-time implementations as it adds white noise to all channels making this method time-consuming. Additionally, a multichannel version of the EMD method has been proposed which utilizes cross-channel information called multivariate

empirical mode decomposition (MEMD).[45–48] The mean $A(t)$ is computed through multivariate envelope curves, given as:[49]

$$A(t) = \frac{1}{p} \sum_{j=1}^{p} e^{\eta_j}(t) \qquad (2)$$

where j gives the length of vectors. $e^{\eta_j}(t)$ denotes the multivariate envelope curves for the entire set of direction vectors. More details can be obtained from.[45] Further, the candidate IMF $Rd(t)$ is computed as $Rd(t) = Z(t) - A(t)$. If the candidate IMF satisfies the stoppage criterion, then it becomes the multichannel IMF. Otherwise, the input $Z(t)$ is set equal to the remainder $Rd(t)$ and the complete process will be repeated again until the remaining multivariate IMFs (MIMFs) have been extracted. All of the MIMFs must satisfy two criteria: 1) the number of maxima and minima should be zero or differ by one, 2) the mean value of the envelope defined by the local maxima and local minima should be zero at any point. For more details please refer to.[45] Recently, Gaur et al.[50] proposed a method for automatic selection of MIMFs with reduced number of channels (FC1, C1, CP1, FCz, Cz, FC3, C3, CP3, FC2, FC4, C4, CP4, C2, CP2, and CPz) to enhance the performance of MI based BCI.

To get a better insight of the MEMD decomposition of EEG signals, a trial of each MI task is considered to decompose the EEG signal of three channels FC3, CP1, and POz simultaneously of the provided twenty-two channels. In the final analysis, all of the twenty-two channels were considered. The feet, tongue, left hand, and right hand MI EEG signals were decomposed using this MEMD method. The decomposition mechanism of the left hand MI task is shown in Figure 5 decomposed using single trials from subject A01T.

In this work, an MEMD filtering of fixed band 8-30 Hz is used to enhance the EEG signals in the preprocessing step for all motor imagery tasks. This filtering process helps to filter the *mu* and *beta* bands by retaining the IMFs which contributes to them and by discarding the remaining IMFs. The discarded IMFs contain high and low frequency components > 30 Hz and < 8 Hz, artefacts, and noise. The selection criterion used to select the IMFs is to compute the mean frequency of all the obtained IMFs and retain the IMFs whose mean frequency falls between 8-30
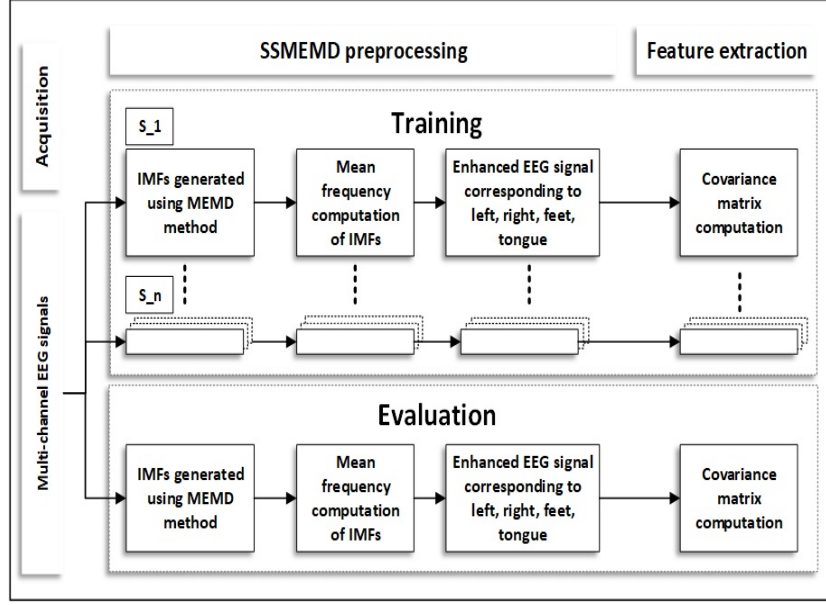
Figure 3.   Overview of our previous work showing Subject-Specific Multivariate Empirical Mode Decomposition (SS-MEMD) preprocessing stage before extracting features using covariance matrix computation.
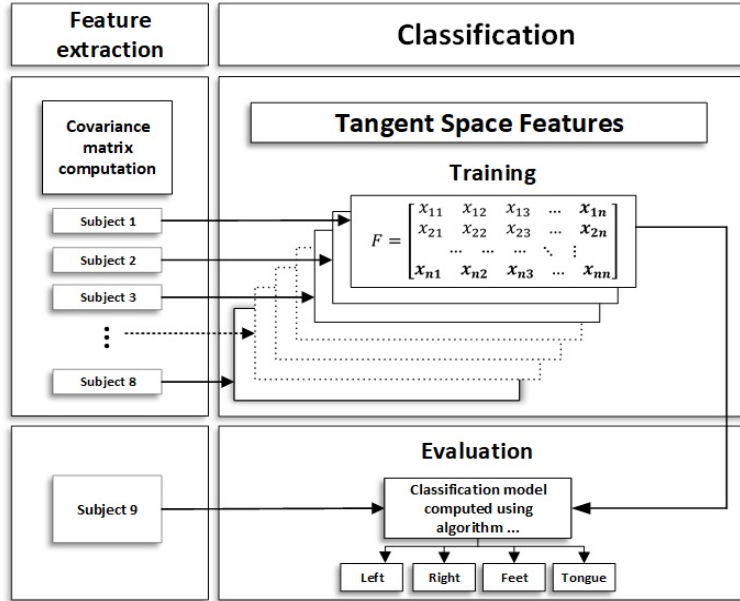

Figure 4.   Block diagram of the proposed methodology.

Hz. The mathematical expression of mean frequency is denoted as:

$$MeanF_{IMF} = \frac{\sum_{j=1}^{n} P_j f_j}{\sum_{j=1}^{n} P_j} \tag{3}$$

where n denotes the length of the frequency bin, and P_j gives the power spectrum at the frequency bin i, whereas f_j represents the frequency value at the frequency bin i. The obtained IMFs satisfying the above criteria were summed to obtain the enhanced EEG signal. Whereas in Gaur et al.[40] a subject specific mean frequency range is identified , the current work takes a fixed band of 8-30 Hz for all subjects and MI tasks, as is more common in the literature in
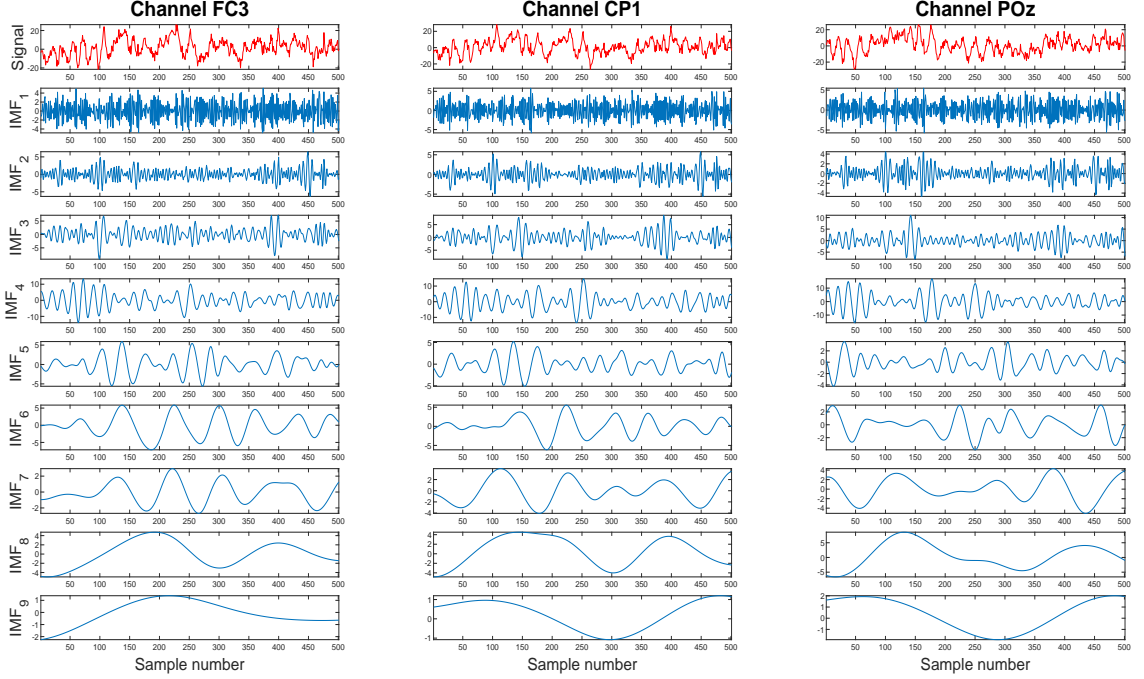
Figure 5. The EEG signals of channels FC3, CP1, and POz and their corresponding MIMFs of Subject A01T for left hand MI task.

an attempt to maintain consistency.[26,38]

As the MI task begins at 2 seconds, the covariance matrix feature is computed for the EEG signals between 2.5 and 4.5 seconds. In this study, the mean frequency measure is also calculated to identify the MIMFs corresponding to the right hand, left hand, both feet, and tongue MI tasks.

4. Proposed model: Tangent Space Based Transfer Learning (TS-TL) classification model

As mentioned previously, although not affecting all systems, one problem in BCI is the long calibration time needed where subjects are involved firstly in a training session, followed by an evaluation session. This work aims to address this issue by proposing a novel classification model which can be used for evaluating unseen single trials for a subject without any previously recorded training session data.

The training sessions from multiple subjects are indexed as $s = \{1, ..., S\}$ and have n_s^{tr} trials, $Z_s = \{(x_s^i, y_s^i)\}_{i=1}^{n_s^{tr}}$. The tangent space concept in the Riemannian geometry framework and the logarithmic

mapping gives the inverse mapping which is defined as,

$$\text{Log}_c(Q_i) = P_i = \\ ||\text{lower}\{Q^{1/2}\log(Q^{-1/2}Q_iQ^{-1/2})Q^{1/2}\}|| \quad (4)$$

where Q denotes the $r \times r$ symmetric positive definite matrices. P_i is defined as the derivative at $t = 0$ of the geodesic distance between exponential mapping $Q_i(= \text{Exp}_c(P_i))$ and Q. P_i denotes a tangent vector at Q in the TS. See Figure 6 for more information and refer to[40,51] for detailed explanation.

These features are derived from the tangent space and named as tangent space features. Only $n(n+1)/2$ are considered by taking the lower triangular matrix of the provided $n \times n$ features as given in eq. (5).

$$F = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{ij} & \dots & x_{nn} \end{bmatrix} \quad (5)$$
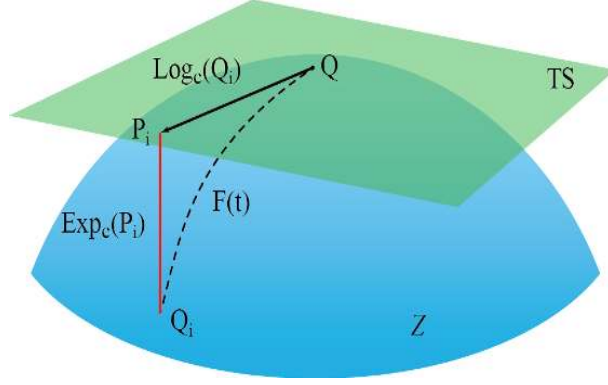
Figure 6.   The tangent space (TS) at point Q. P_i gives a tangent vector at Q.

where, x_{ij} denotes the tangent space feature extracted between channel i and channel j in the feature matrix F for each trial and $x_s^i \in \mathbb{R}^d$ denotes the features derived from the EEG signals of the training session for subject s during trial i, with d denoting the number of features selected from the feature matrix F. Since the sample covariance matrices (SCM) are symmetric, the lower triangular matrix is considered for this study giving a total of $d = n(n+1)/2$ features. The x_s^i consists of tangent space features computed at different scalp locations. Variable y_s^i denotes the label of either 'left' or 'right' hand imagination, in trial i for session s. This approach is applicable for solving two-class classification problems. Here, it is solved as a linear regression problem with $y_s^i \in \{+1, -1\}$ for all i and s. Based on this assumption, the model is linear with a noise term ν, the linear function model is denoted by:

$$y_s^i = w_s^T x_s^i + \nu \tag{6}$$

related to the subject/training session s. Here, a linear classifier is used for each subject s with features x_s^i and weights w_s. The parameters w_s shows the weights assigned to the individual features which are further used to evaluate the class label for unseen trials in the evaluation session of a new subject s. Given a new EEG signal x for subject/ evaluation session s, the class label is evaluated by,

$$\hat{y}_s^{i+1} = w_s^T x_s^i \tag{7}$$

where, $\hat{y}_s^{i+1}$ denotes the class assigned to unseen trials. Each subject/training session has a shared structure (Σ, μ) that represents the invariant properties for class prediction. To be specific, (Σ, μ) represents the covariance and mean vectors of features respectively. The divergence of the training session model

from the shared structure of each subject $||w_s - \mu||$ gives the session specific properties of the class prediction.[36] This shared structure is unknown, thus the main goal is to find the shared structure across all the subjects. This is achieved by combining the optimization problem as:

$$\min_{K,\Sigma,\mu} LF(K, \Sigma, \mu; Z, \theta) = \min_{K,\Sigma,\mu} \frac{1}{\theta} \sum_1^s ||X_s w_s - y_s||^2 + \sum_1^s \Psi(w_s; \Sigma, \mu) \tag{8}$$

where, LF denotes the least square loss function with $N(0, \sigma_2)$. The estimated label is expected to be normally distributed with mean zero and variance σ_2. In this work the σ_2 symbol is replaced with the θ symbol. The input matrix for each subject training data in matrix form is defined as $X_s = [x_s^{1T}, ..., x_s^{nT}]^T$. Here nT denotes the training session features for a particular subject, T denotes the transpose of the matrix, $K = [w_1, ..., w_s]^T$, and $Z = \{Z_s\}_{s=1}^S$. It is a common phenomenon that models trained without cross-validation may over-fit and lead to poor generalization of unseen evaluation data. So Ψ is used as a regularization term to handle overfitting.

Equation 8 has two parts. The first part is the summation of the losses from each session. All the sessions fit well is ensured by minimizing it. The second part of the equation 8 is solved using minimization with respect to K and (Σ, μ) in an iterative manner by alternatively holding K and (Σ, μ) constant. Now, the optimization over w_s for fixed Σ and μ is computed for all subjects/ training session data and

hence it is optimized independently. The new w_s is obtained in each iteration by computing the derivative with respect to w_s for all subjects s and then, equating it to 0. For each w_s, the closed form update is given by:

$$w_s = \left(\frac{1}{\theta}X_s^T X_s + \mu^{-1}\right)^{-1}\left(\frac{1}{\theta}X_s^T y_s + \mu^{-1}\Sigma\right) \quad (9)$$

The w_s is computed iteratively by solving the above optimization problem. w_s refers to the weights of the linear regression problem for a single MI task. The structure shared across multiple subjects' training data is computed using multi-task BCI. The algorithm steps are as follow:

(1) Input: Z,θ
(2) Set : $\left\{(\Sigma, \mu)\right\} = \{I, 0\}$
(3) **Repeat**
(4) Update $w_s = \frac{\left(\frac{1}{\theta}X_s^T y_s \mu + \Sigma\right)}{\left(\frac{1}{\theta}X_s X_s^T \mu + I\right)}$
(5) Update Σ using $\hat{\Sigma} = \frac{\sum_{s=1}^{S}\left(w_s - \mu\right)\left(w_s - \mu\right)^T}{Tr\left(\left(w_s - \mu\right)\left(w_s - \mu\right)^T\right)} + \varepsilon I$
(6) Update μ using $\hat{\mu} = \frac{1}{S}\sum_{s=1}^{S}w_s$
(7) **Until** converge
(8) **Output :** $\left\{(\Sigma, \mu)\right\}$

ε denotes a small number used for diagonal loading and I represents the identity matrix. The aforementioned algorithm is used for computing the shared structure and w_s across training sessions using leave-one-subject-out-cross-validation (LOSO-CV) for all subjects. More details about different variations of the optimization problem used to compute w_s can be obtained in the literature.[36,52]

5. Results and discussion

In the evaluation session, the tangent space feature is computed on all unknown test trials of the corresponding evaluation session A0SE on a trial-by-trial basis identified by LOSO-CV, where S represents the subject number and E represents the evaluation session number. These features are classified using the proposed method and assigned a particular class. Currently, the proposed method can be applied to motion intentions into two class problems. In this dataset, there are four MI tasks, it gives a total of six combinations of two-class classification problems which are as follows: left vs. right (LvR), left vs. foot (LvF), left vs. tongue (LvT), right vs. foot (RvF),

right vs. tongue (RvT), and foot vs. tongue (FvT). The classification accuracy is calculated for all nine subjects for each evaluation session.

Table 2. LOSO-CV classification accuracy (%) for the proposed classification method with one vs. one scheme applied on BCI competition IV dataset 2A.

Subject	Accuracy with proposed method (%)					
	LvR	LvF	LvT	RvF	RvT	FvT
A01	88.65	93.57	97.89	96.4	100	68.57
A02	61.27	62.86	65.03	79.29	67.83	70.21
A03	91.24	88.15	85.19	97.83	97.83	78.68
A04	74.14	57.63	83.93	77.59	82.73	62.5
A05	57.04	66.2	78.42	64.96	60.45	60.28
A06	69.44	70.09	71.7	67.89	58.33	55.14
A07	60	86.62	81.02	89.29	87.41	70.07
A08	94.03	80.74	91.79	83.94	90.44	85.4
A09	83.85	94.03	97.69	78.36	83.85	88.81
Average	75.52	77.77	83.63	81.73	80.99	71.07
Std	14.39	13.83	11.14	11.42	15.39	11.34

Table 2 reports the LOSO-CV results obtained in terms of classification accuracy using a one vs. one scheme when studied on BCI competition IV datatset 2A. The LvR MI task provided an average of 75.52% group accuracy across all nine subjects whereas the LvF MI task gives an average of 77.77% group classification accuracy for nine subjects. The LvT MI task demonstrates a maximum group classification accuracy of 83.63% when compared with all other MI tasks for all nine subjects. The RvF MI task gives an average of 81.73% group accuracy whereas the RvT MI task reported an average of 80.99% group classification accuracy for all nine subjects. The FvT MI task resulted in the lowest group classification accuracy of 71.07% when compared with all other MI tasks. It should be noted that although an increase in the number of participants would lead to a more statistically significant result, this restriction is due to the nature of the dataset used being BCI Competition IV Dataset IIa.

Table 3 compares the classification accuracy of the proposed method against competing comparable methods available in the literature. Std denotes standard deviation in Table 3.

Method 1 (M1) displays results from the evaluation session from previously published work by Gaur et al.[40] which uses a subject-specific MEMD based filtering method (SS-MEMDBF) for preprocessing and implements a Riemannian geometry framework

devised separately for each subject for classification. In M1, for each subject, a subject-specific filtering range is identified in the training session before a classification model is created and used to classify each unseen single trial of the evaluation session. The proposed method uses a generalised 8-30 Hz frequency band to cover the *mu* and *beta* bands, and the LOSO-CV mechanism to build the classification model before classifying unseen single trials from the evaluation session not used in the LOSO-CV procedure. The average classification accuracy of the proposed method (TS-TL) (75.52±14.39) is comparable to method M1 (79.93±14.99). Two of the nine subjects show a slight improvement of >1% when classified using the proposed method whilst five are within 3% of the results presented in M1.

Table 3. Classification accuracy (%) for a left versus right motor imagery task when compared with other published works on BCI competition IV dataset 2A.

Subject	Proposed method	M1	M2	M3	FBCSP
A01	88.65	91.49	88.89	90.28	84.4
A02	61.27	60.56	51.39	57.64	57.04
A03	91.24	94.16	96.53	95.14	97.08
A04	74.14	76.72	70.14	65.97	64.66
A05	57.04	58.52	54.86	61.11	69.63
A06	69.44	68.52	71.53	65.28	62.96
A07	60	78.57	81.25	61.11	77.14
A08	94.03	97.01	93.75	91.67	93.28
A09	83.85	93.85	93.75	86.11	92.31
Average	75.52	79.93	78.01	74.92	77.61
Std	14.39	14.99	17.01	15.43	14.82

Method 2 (M2)[26] implements CSP on band-pass filtered EEG between 8 and 30 Hz before calculating the log variance from three pairs of filters for extraction of features and uses linear discriminant analysis (LDA)[26] for a binary classification problem. One subject shows an improvement of almost 10% whilst three others show varying levels of improvement.

Method 3 (M3)[53] again implements CSP and uses detection of the covariate shift and adaptive learning. The proposed method outperforms M3 by >0.5%. Subject A04 has also improved by > 8% with an additional three subjects showing an improvement of >3%.

A filter bank common spatial pattern (FBCSP) technique implements filter bank CSP on EEG signals between 8 and 32 Hz taking a frequency band of 4 Hz. The frequency bands considered are 4-8, 8-12, 12-16, 16-20, 20-24, 24-28 and 28-32. Thereafter, the log variance from three pairs of filters for extraction of features is computed and uses linear discriminant analysis (LDA)[26] for a binary classification problem. Four subjects show an improvement of almost 4% whilst one subject shows a slight improvement of less than 1%.

Although the results of this current study are presented here in comparison with other state-of-the-art methods, it is an unfair comparison. Results obtained using the TS-TL method are based on a different approach whereby a generalised model is created using the LOSO-CV mechanism described above. In other state-of-the-art methods, a subject specific classification model is first created using training session data before a subject specific classification model is used to classify evaluation data for the same subject. The results herein demonstrate that it is possible to create a generalised model and then further classify the unseen trials of a new subject's evaluation data by using the proposed novel method.

Figs. 7, 8 and 9 display the difference between the classification accuracy computed using (1) SS-MEMDBF with the Riemannian geometry framework[40] (M1 in Table 3), (2) FB-MEMDBF (filtered using a fixed frequency band of 8-30 Hz) with tangent space based transfer learning (TS-TL) and (3) FBCSP method obtained in the evaluation session for all of the six possible binary MI tasks. The performance improvement is demonstrated with bar graphs for all nine subjects.

Fig. 7(a) shows the classification accuracy comparison using SS-MEMDBF with the Riemannian geometry framework (M1), FB-MEMDBF with TS-TL and other state-of-the-art methods (M2, M3 and FBCSP) obtained in the evaluation session for the left hand and right hand MI tasks. With the proposed tangent space based transfer learning method, comparable results were achieved when compared with these other state-of-art methods. Although, the results obtained with SS-MEMDBF[40] and the Riemannian geometry framework are impressive, a few of the subjects still perform badly. This may be due to the effect of non-stationarity, even after developing a subject specific classification model. Hence, there is a need to create a generalized model using transfer learning which may help to overcome this issue.
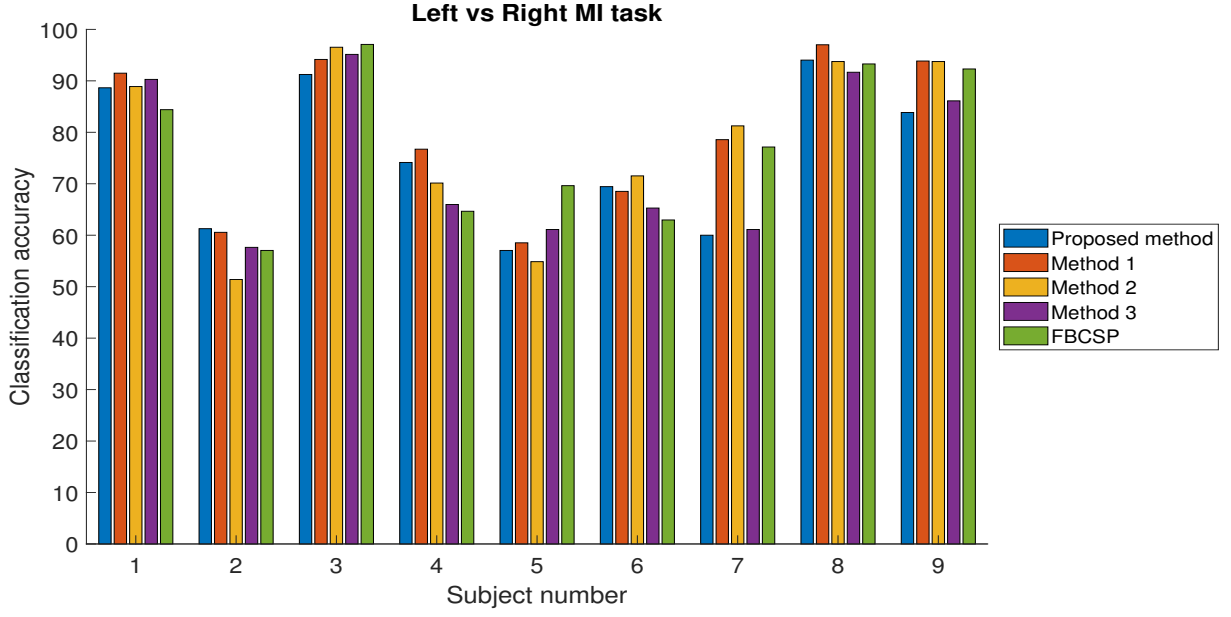
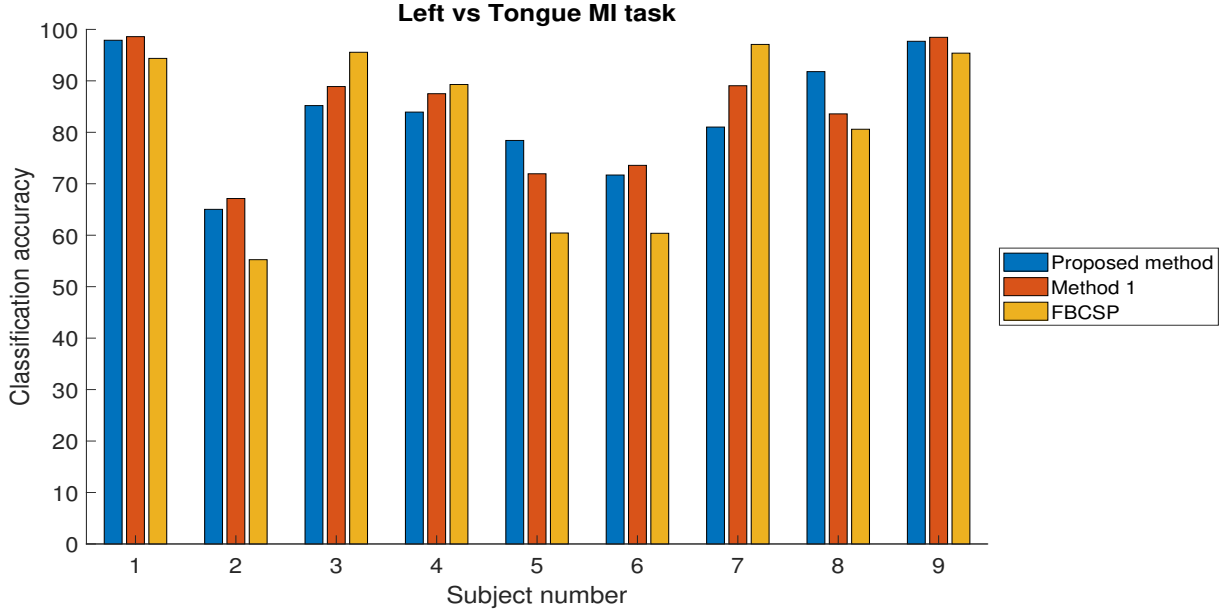(a)Comparison of classification accuracy for left hand and right hand MI task.
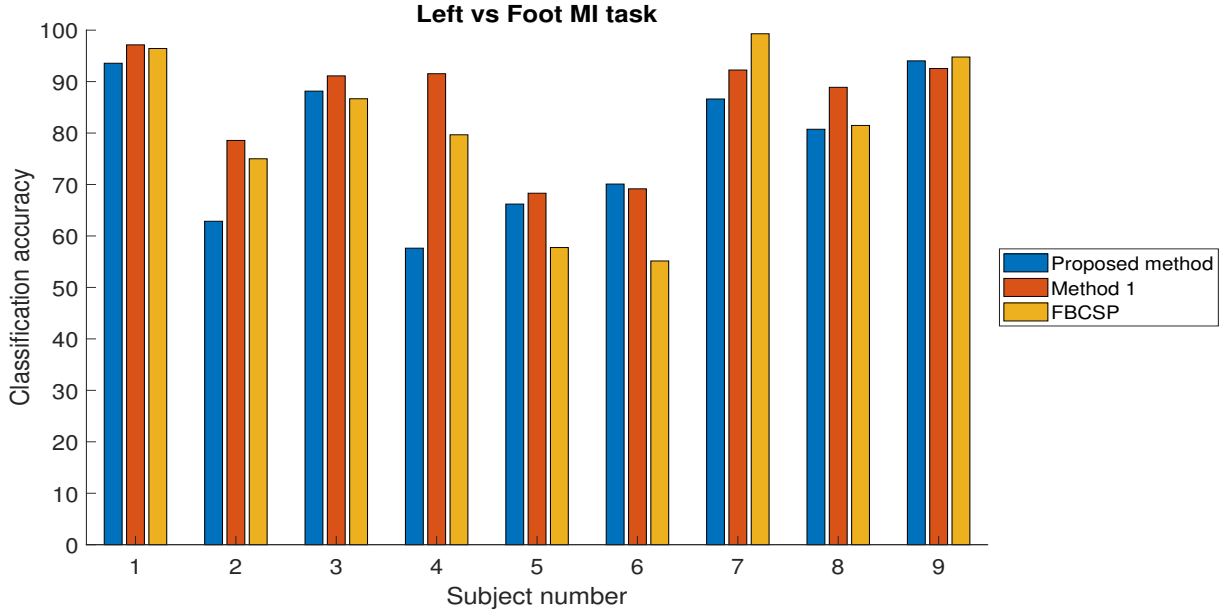
(b)Comparison of classification accuracy for left hand and foot MI task.

Figure 7. The bar graph displays the classification accuracy comparison using SS-MEMDBF with Riemannian geometry framework, FBCSP and FB-MEMDBF (8-30 Hz) with TS-TL and other state-of-the-art methods obtained in the evaluation session for (a) left hand and right hand MI tasks (b) left hand and foot MI tasks.

With the proposed method, a fixed band of 8-30 Hz is considered in the preprocessing step and the features were classified using a LOSO-CV mechanism demonstrating results comparable with other state-of-the-art methods. One of the nine subjects (A02) shows a slight improvement in classification accuracy when compared with all four of the competing methods.

The bar graph shown in Figure 7(b) displays the classification accuracy comparison using SS-MEMDBF with the Riemannian geometry framework, FBCSP method and FB-MEMDBF (8-30 Hz)

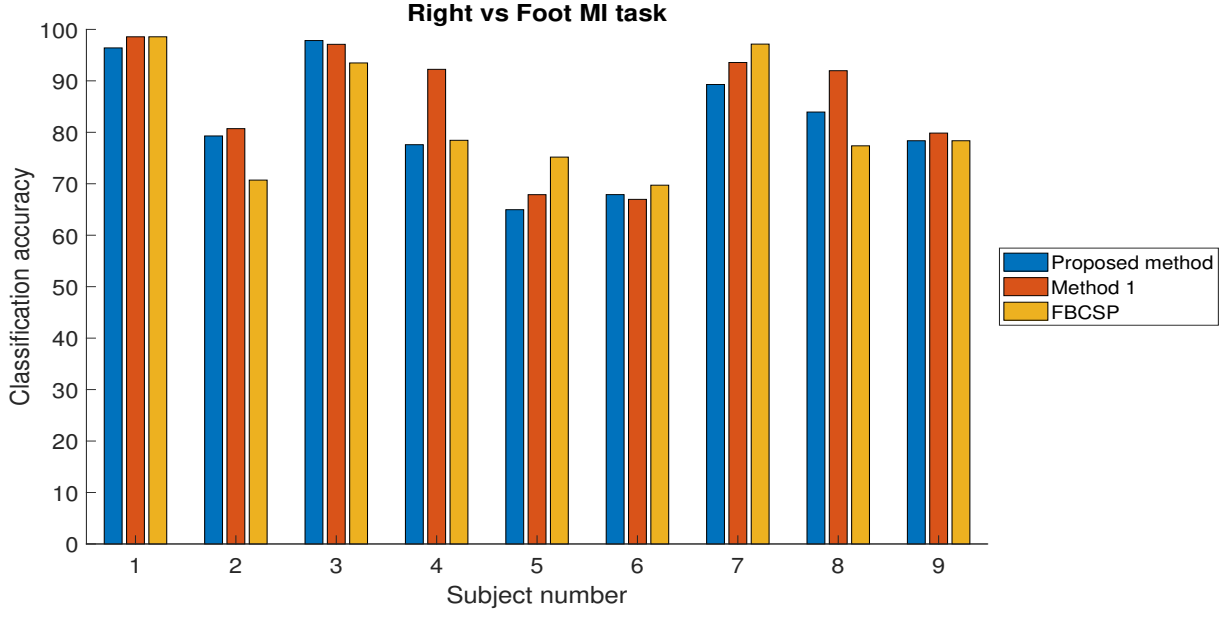(a)Comparison of classification accuracy for left hand and tongue MI task.
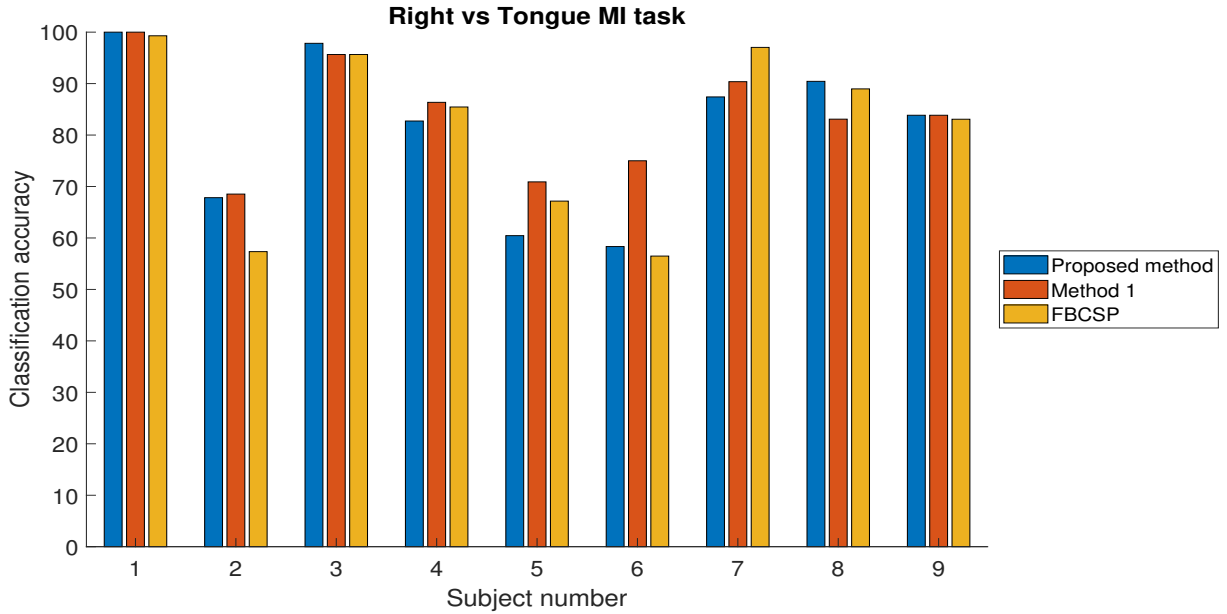
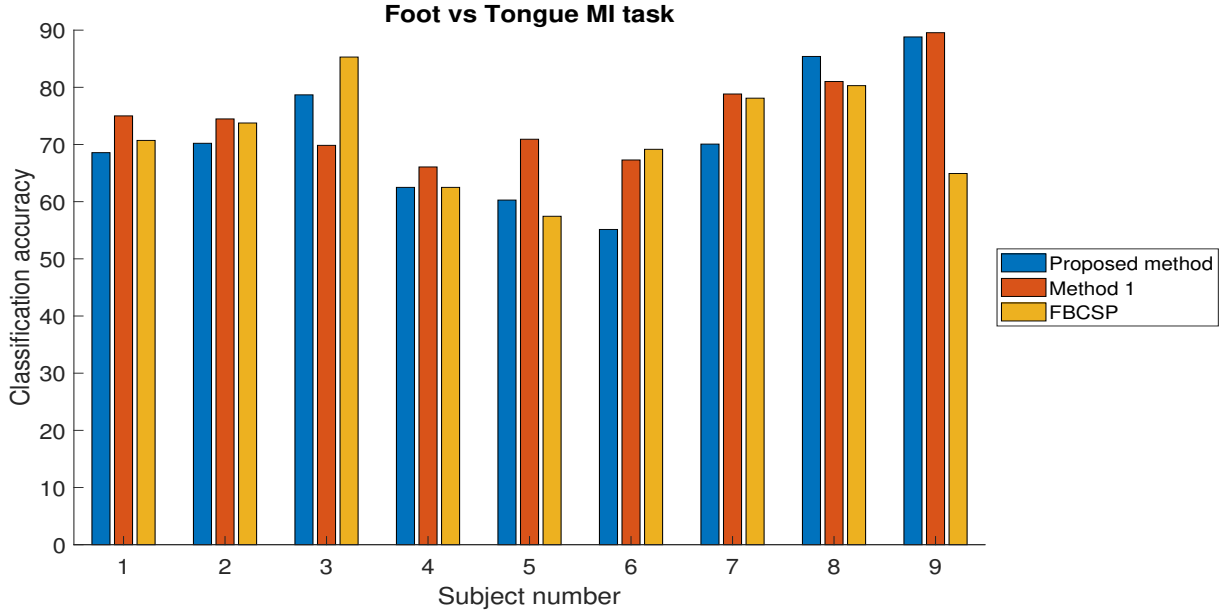(b)Comparison of classification accuracy for right hand and tongue MI task.

Figure 8.   (a) The bar graph displays the classification accuracy comparison using SS-MEMDBF with Riemannian geometry framework, FBCSP and FB-MEMDBF (8-30 Hz) with TS-TL obtained in the evaluation session for (a) left hand and tongue MI task (b) right hand and tongue MI task.

with TS-TL obtained in the evaluation session for left hand and foot MI tasks. There is an improvement in classification accuracy for two of the nine subjects using the FB-MEMDBF (8-30 Hz) with TS-TL method when compared to SS-MEMDBF with the Riemannian geometry framework. Subjects A06 and A09 show an improvement in the evaluation session. A total of three subjects of the nine subjects haves shown improvement in the classification accuracy when compared with FBCSP method. Subjects A05 and A06 show an improvement of >8% in the evaluation session.

(a)Comparison of classification accuracy for right hand and foot MI task.

(b)Comparison of classification accuracy for foot and tongue MI task.

Figure 9.   (a) The bar graph displays the classification accuracy comparison using SS-MEMDBF with Riemannian geometry framework, FBCSP and FB-MEMDBF (8-30 Hz) with TS-TL obtained in the evaluation session for (a) right hand and foot MI task (b) foot and tongue MI task.

Fig. 8(a) presents the classification accuracy when comparing SS-MEMDBF with the Riemannian geometry framework, FBCSP method and FB-MEMDBF (8-30 Hz) with TS-TL obtained in the evaluation session for the left hand and tongue MI tasks. Two of the nine subjects show improvement

in classification accuracy using the FB-MEMDBF (8-30 Hz) with TS-TL. The difference between average classification accuracy of the two methods is within 1%. Subjects A05 and A08 show a maximum improvement of $> 6\%$. A total of six of the nine subjects show improvement in the classification accuracy

when compared to FBCSP method. Subjects A03, A05, A06 and A08 give an improvement of >9% in the evaluation session.

Fig. 8(b) shows the classification accuracy comparison using SS-MEMDBF with the Riemannian geometry framework, FB-MEMDBF (8-30 Hz) with TS-TL and FBCSP method obtained in the evaluation session for the right hand and tongue MI tasks. Two of the nine subjects show improvement in classification accuracy using the proposed method. The difference in average classification accuracy of the two methods is within 3%. Subjects A03 and A08 show an improvement of > 2% and > 7% when compared to SS-MEMDBF with the Riemannian geometry framework. Six of the nine subjects have improved in classification accuracy when compared to FBCSP. Subject A02 has shown maximum improvement of >10% whilst other four subjects have shown an improvement within 2%.

Fig. 9(a) similarly shows the classification accuracy comparison using SS-MEMDBF with the Riemannian geometry framework, FB-MEMDBF (8-30 Hz) with TS-TL and FBCSP method obtained in the evaluation session but for the right hand and foot MI tasks. Two of the nine subjects show improvement in classification accuracy using the proposed method. The average classification accuracy with the proposed method is > 81%. The greatest improvement is seen in subjects A03 and A06 of < 1% when proposed method is compared to SS-MEMDBF with the Riemannian geometry framework. Additionally, the classification accuracy of four of the nine subjects have improved when the proposed method is compared with the FBCSP method. Subjects A02, A03 and A08 show an improvement of >4% in the evaluation session while subject A09 improved marginally.

Fig. 9(b) again compares SS-MEMDBF with the Riemannian geometry framework, FB-MEMDBF (8-30 Hz) with TS-TL and FBCSP obtained in the evaluation session but this time for the foot and tongue MI tasks. In this case also, the classification accuracy of two subjects improved using the proposed method with an average of 71.07%. Subject A03 showed significant improvement of > 8.8% with subject A08 improving by > 4% when proposed method is compared to SS-MEMDBF with the Riemannian geometry framework. While three of the nine subjects shows an improvement in the classification accuracy when proposed method is compared to

FBCSP method. Subject A09 has shown maximum improvement of >23%.

6. Conclusion

BCI classifier performance is often linked to the volume of training data. However, this data which is often recorded across sessions as well as across subjects, leads not only to a highly non-stationary training dataset, but can also be very time consuming to collect. A method is presented to enhance performance of a two-class MI based BCI using sample covariance as a feature set by combining a previously reported SS-MEMDBF preprocessing technique[40] with a novel tangent space based learning (TS-TL) classification method. Whereas FB-MEMDBF preprocessing helps to obtain enhanced feature separability and reduce the error rates due to intrinsic non-stationarities present in EEG signals, the addition of TS-TL handles non-stationarities more efficiently. This novel method allows for the classification of unseen trials from evaluation session data using LOSO-CV resulting in comparable improvement across subjects. The problem of lengthy recording sessions could thus be solved by classifiers trained using transfer learning techniques. Future work will focus on applying this method to real-time data and multi-class problems on EEG and MEG datasets whilst reducing the computational time complexity of the proposed method.

Bibliography

1. J. Li, J. Liang, Q. Zhao, J. Li, K. Hong and L. Zhang, Design of assistive wheelchair system directly steered by human thoughts, *International Journal of Neural Systems* **23**(03) (2013) p. 1350013.
2. J. Li, H. Ji, L. Cao, D. Zang, R. Gu, B. Xia and Q. Wu, Evaluation and application of a hybrid brain computer interface for real wheelchair parallel control with multi-degree of freedom, *International Journal of Neural Systems* **24**(04) (2014) p. 1450014.
3. J. Jin, E. W. Sellers, S. Zhou, Y. Zhang, X. Wang and A. Cichocki, A P300 brain–computer interface based on a modification of the mismatch negativity paradigm, *International Journal of Neural Systems* **25**(03) (2015) p. 1550011.
4. A. Burns, H. Adeli and J. A. Buford, Brain–computer interface after nervous system injury, *The Neuroscientist* **20**(6) (2014) 639–651.
5. A. Ortiz-Rosario and H. Adeli, Brain-computer interface technologies: from signal to action, *Reviews in the Neurosciences* **24**(5) (2013) 537–552.

6. J. J. Vidal, Toward direct brain-computer communication, *Annual review of Biophysics and Bioengineering* **2**(1) (1973) 157–180.

7. L. Bozinovska, M. Sestakov, G. Stojanovski and S. Bozinovski, Intensity variation of the CNV potential during the biofeedback training guided by a personal computer, *Neurologija* **37**(suppl 2) (1988) p. 76.

8. L. Bozinovska, S. Bozinovski, G. Stojanov and M. Sestakov, Introduction of biofeedback in the CNV paradigm, *Proceedings of Conference on ETAN, Novi Sad, Yugoslavia*, **12**1989, pp. 93–98.

9. S. Bozinovski, M. Sestakov and L. Bozinovska, Using EEG alpha rhythm to control a mobile robot, *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE1988, pp. 1515–1516.

10. S. Bozinovski and L. Bozinovska, Brain–Computer Interface in Europe: the thirtieth anniversary, *Automatika* **60**(1) (2019) 36–47.

11. S. Iliya and F. Neri, Towards artificial speech therapy: A neural system for impaired speech segmentation, *International journal of neural systems* **26**(06) (2016) p. 1650023.

12. M. A. Molina-Cabello, R. M. Luque-Baena, E. López-Rubio and K. Thurnhofer-Hemsi, Vehicle type detection by ensembles of convolutional neural networks operating on super resolved images, *Integrated Computer-Aided Engineering* (Preprint) (2018) 1–13.

13. P. Wang and X. Bai, Regional parallel structure based CNN for thermal infrared face identification, *Integrated Computer-Aided Engineering* **25**(3) (2018) 247–260.

14. M. Koziarski and B. Cyganek, Image recognition with deep neural networks in presence of noise-dealing with and taking advantage of distortions, *Integrated Computer-Aided Engineering* **24**(4) (2017) 337–349.

15. L. Pan, G. Păun, G. Zhang and F. Neri, Spiking neural P systems with communication on request, *International journal of neural systems* **27**(08) (2017) p. 1750042.

16. T. Wu, F.-D. Bîlbîe, A. Păun, L. Pan and F. Neri, Simplified and yet Turing universal spiking neural P systems with communication on request, *International journal of neural systems* **28**(08) (2018) p. 1850013.

17. S. Wu, G. Zhang, M. Zhu, T. Jiang and F. Neri, Geometry based three-dimensional image processing method for electronic cluster eye, *Integrated Computer-Aided Engineering* **25**(3) (2018) 213–228.

18. L. F. Nicolas-Alonso and J. Gomez-Gil, Brain computer interfaces, a review, *sensors* **12**(2) (2012) 1211–1279.

19. M.-C. Corsi, M. Chavez, D. Schwartz, L. Hugueville, A. N. Khambhati, D. S. Bassett and F. D. V. Fallani, Integrating EEG and MEG Signals to Improve Motor Imagery Classification in Brain–Computer Interface, *International journal of neural systems* (2018) p. 1850014.

20. L. Bozinovska, G. Stojanov, M. Sestakov and S. Bozinovski, CNV pattern recognition: Step toward a cognitive wave observation, *5. European Signal Processing Conference.*, **3**1990, pp. 1659–1662.

21. L. Bozinovska, S. Bozinovski and G. Stojanov, Electroexpectogram: Experimental Design and Agorithms, *Biomedical Engineering Days, Proceedings of the 1992 International*, IEEE, pp. 55–60.

22. M. Sugiyama, M. Yamada and M. C. du Plessis, Learning under nonstationarity: covariate shift and class-balance change, *Wiley Interdisciplinary Reviews: Computational Statistics* **5**(6) (2013) 465–477.

23. D. Wu, Online and offline domain adaptation for reducing BCI calibration effort, *IEEE Transactions on Human-Machine Systems* (2017).

24. M. Krauledat, M. Tangermann, B. Blankertz and K.-R. Müller, Towards zero training for brain-computer interfacing, *PloS one* **3**(8) (2008) p. e2967.

25. S. Fazli, F. Popescu, M. Danóczy, B. Blankertz, K.-R. Müller and C. Grozea, Subject-independent mental state classification in single trials, *Neural networks* **22**(9) (2009) 1305–1312.

26. F. Lotte and C. Guan, Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms, *IEEE Transactions on biomedical Engineering* **58**(2) (2011) 355–362.

27. D. Devlaminck, B. Wyns, M. Grosse-Wentrup, G. Otte and P. Santens, Multisubject learning for common spatial patterns in motor-imagery BCI, *Computational intelligence and neuroscience* **2011** (2011) p. 8.

28. A. Barachant, S. Bonnet, M. Congedo and C. Jutten, BCI Signal Classification using a Riemannian-based kernel, *20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2012)*, Michel Verleysen2012, pp. 97–102.

29. M. Congedo, A. Barachant and A. Andreev, A new generation of brain-computer interface based on riemannian geometry, *arXiv preprint arXiv:1310.8115* (2013).

30. M. Congedo and A. Barachant, A special form of SPD covariance matrix for interpretation and visualization of data manipulated with Riemannian geometry, *AIP Conference Proceedings*, **1641**(1), AIP2015, pp. 495–503.

31. M. Sugiyama, M. Krauledat and K.-R. MÃžller, Covariate shift adaptation by importance weighted cross validation, *Journal of Machine Learning Research* **8**(May) (2007) 985–1005.

32. A. Satti, C. Guan, D. Coyle and G. Prasad, A covariate shift minimisation method to alleviate non-stationarity effects for an adaptive brain-computer interface, *Pattern Recognition (ICPR), 2010 20th In-*

ternational Conference on*, IEEE2010, pp. 105–108.

33. W. Dai, Q. Yang, G.-R. Xue and Y. Yu, Boosting for transfer learning, *Proceedings of the 24th international conference on Machine learning*, ACM2007, pp. 193–200.

34. V. Lawhern, A. Solon, N. Waytowich, S. M. Gordon, C. Hung and B. J. Lance, EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces, *Journal of neural engineering* (2018).

35. S. Božinovski, A representation theorem for linear pattern classifier training, *IEEE transactions on systems, man, and cybernetics* (1) (1985) 159–161.

36. V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf and M. Grosse-Wentrup, Transfer learning in brain-computer interfaces, *IEEE Computational Intelligence Magazine* **11**(1) (2016) 20–31.

37. C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl and G. Pfurtscheller, BCI Competition 2008–Graz data set A, *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology* (2008) 136–142.

38. K. K. Ang, Z. Y. Chin, C. Wang, C. Guan and H. Zhang, Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b, *Frontiers in Neuroscience* **6** (2012).

39. P. Gaur, R. B. Pachori, H. Wang and G. Prasad, A multivariate empirical mode decomposition based filtering for subject independent BCI, *Signals and Systems Conference (ISSC), 2016 27th Irish*, IEEE2016, pp. 1–7.

40. P. Gaur, R. B. Pachori, H. Wang and G. Prasad, A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry, *Expert Systems with Applications* **95** (2018) 201–211.

41. G. Pfurtscheller, C. Neuper, D. Flotzinger and M. Pregenzer, EEG-based discrimination between imagination of right and left hand movement, *Electroencephalography and Clinical Neurophysiology* **103**(6) (1997) 642–651.

42. N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung and H. H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, **454**(1971), The Royal Society1998, pp. 903–995.

43. P. Gaur, R. B. Pachori, H. Wang and G. Prasad, An empirical mode decomposition based filtering method for classification of motor-imagery EEG sig-nals for enhancing brain-computer interface, *Neural Networks (IJCNN), 2015 International Joint Conference on*, IEEE2015, pp. 1–7.

44. Z. Wu and N. E. Huang, Ensemble empirical mode decomposition: a noise-assisted data analysis method, *Advances in adaptive data analysis* **1**(01) (2009) 1–41.

45. C. Park, D. Looney, A. Ahrabian, D. P. Mandic *et al.*, Classification of motor imagery BCI using multivariate empirical mode decomposition, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **21**(1) (2013) 10–22.

46. N. E. Huang, M.-L. C. Wu, S. R. Long, S. S. Shen, W. Qu, P. Gloersen and K. L. Fan, A confidence limit for the empirical mode decomposition and Hilbert spectral analysis, *Proceedings of the royal society of london a: Mathematical, physical and engineering sciences*, **459**(2037), The Royal Society2003, pp. 2317–2345.

47. G. Rilling, P. Flandrin, P. Goncalves *et al.*, On empirical mode decomposition and its algorithms, *IEEE-EURASIP workshop on nonlinear signal and image processing*, **3**, IEEER, Grado, Italy2003, pp. 8–11.

48. N. Rehman and D. P. Mandic, Multivariate empirical mode decomposition, *Proceedings of The Royal Society of London A: Mathematical, Physical and Engineering Sciences*, The Royal Society2009, p. rspa20090502.

49. N. ur Rehman, C. Park, N. E. Huang and D. P. Mandic, EMD via MEMD: multivariate noise-aided computation of standard EMD, *Advances in Adaptive Data Analysis* **5**(02) (2013) p. 1350007.

50. P. Gaur, R. B. Pachori, H. Wang and G. Prasad, An Automatic Subject Specific Intrinsic Mode Function Selection for Enhancing Two-Class EEG based Motor Imagery-Brain Computer Interface, *IEEE Sensors Journal* (2019).

51. A. Barachant, S. Bonnet, M. Congedo and C. Jutten, Multiclass brain–computer interface classification by Riemannian geometry, *IEEE Transactions on Biomedical Engineering* **59**(4) (2012) 920–928.

52. M. Alamgir, M. Grosse-Wentrup and Y. Altun, Multitask learning for brain-computer interfaces, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 17–24.

53. H. Raza, H. Cecotti, Y. Li and G. Prasad, Adaptive learning with covariate shift-detection for motor imagery-based brain–computer interface, *Soft Computing* (2015) 1–12.