EEG Signal Separation for Multi-Class Motor Imagery using Common Spatial Patterns Based on Joint Approximate Diagonalization

S. R. Liyanage, J. -X. Xu, C. T. Guan, K. K. Ang and T. H. Lee

***Abstract*— The design of multiclass BCI is a very challenging task because of the need to extract complex spatial and temporal patterns from noisy multidimensional time series generated from EEG measurements. This paper proposes a Multiclass Common Spatial Pattern (MCSP) based on Joint Approximate Diagonalization (JAD) for multiclass BCIs. The proposed method based on fast Frobenius diagonalization (FFDIAG) is compared with another method based on Jacobi angles on the BCI competition IV dataset 2a. The classification accuracies obtained from 10×10-fold cross-validations on the training dataset are compared using K-Nearest Neighbor, Classification Trees and Support Vector Machine classifiers. The proposed MCSP based on FFDIAG yields an averaged accuracy of 53.6% compared to 32.8% given by the method based on Jacobi angles and 27.8% of the one versus rest CSP methods.**

I. Introduction

Brain–Computer Interfaces (BCIs) translate brain signals into a control signal without using muscles or peripheral nerves. They provide a direct communication channel between brain and computer or external devices. A review of current BCI systems is found in [1]. A typical BCI system is composed of a set of sensors and signal processing components (displays and sensory stimulators) that translate a person's brain activity directly into useful control or communication signals.

Typical noninvasive electroencephalogram (EEG) based brain-computer communication devices are composed of three subsystems, namely, EEG acquisition, EEG signal processing and the output subsystems. The acquired EEG signals can be regarded as complex time series signals that have multiple factors intricately intertwined. Therefore, signal processing and classification methods are essential tools in the development of improved BCI technology. One of the main problems in this context is the low signal to noise ratio (SNR) of the recorded EEG data. This has motivated research on spatial filters that are designed to extract those components of the EEG/MEG data that provide most information on the intention of the BCI user. One algorithm that is very frequently used for this purpose is the common spatial patterns (CSP) algorithm. CSP is a technique to analyze multichannel data based on recordings from two classes (conditions). CSP was first proposed in the context of EEG/MEG analysis in [2], and introduced to the BCI community in [3].

The CSP algorithm is capable of computing spatial filters that maximize the ratio of the variance of the data conditioned on one class to the variance of the data conditioned on the other class, when the EEG/MEG data of two different classes are provided. The CSP algorithm computes optimal features for binary classification [15]. A fundamental limitation of CSP is that it can only handle two classes. There is no canonical method for computing the relevant CSP patterns for multiclass classification [16]. This is because simultaneous diagonalization, upon which CSP is based, can be carried out only for two matrices.

Several approaches have been proposed to extend the CSP algorithm to multiclass paradigm [15],[16]. One approach of extending CSP to multiclass paradigms is by performing two-class CSP on different combinations of classes (e.g., by computing CSPs for all combinations of classes or by computing CSP for one class versus all the other classes).

An extension of CSP for multiclass case has been proposed in [18] where the M-class problem is decomposed into a set of M binary problems. Spatial patterns for each class against all others are calculated in this approach. Classification is then performed on the variances of the projections of the EEG signals on all these CSP patterns [18]. However, the performance of one versus rest CSP in general is still limited [17,18].

Another approach for extending CSP for multiple classes is to approximate the joint diagonalization. JAD makes use of approximate optimization methods to diagonalize more than two matrices simultaneously. Given EEG data from M different classes, the goal of CSP by JAD is to find a transformation $W \in R^{N\times N}$ that diagonalizes the covariance matrices.

CSP by joint approximate diagonalization had been shown to be equivalent to independent component analysis (ICA) in [17] and a method to choose those independent components (ICs) that approximately maximize mutual information of ICs and class labels has also been proposed in [17]. A linear Least Squares algorithm for joint diagonalization had been attempted in [16].

The two implementations of multiclass CSP discussed in this paper are based on two approximate joint diagonalization

Manuscript received February 7, 2010. Sidath R. Liyanage is with the National University of Singapore Graduate School for Integrative Sciences and Engineering (e-mail: sidath@nus.edu.sg). Jianxin Xu is with the Department of Electrical & Computer Engineering, National University of Singapore (email: elexujx@nus.edu.sg). C. Guan and K. K. Ang are with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore 119613, Singapore (e-mail: ctguan@i2r.a-star.edu.sg; kkang@i2r.a-star.edu.sg). T.H.Lee is with the National University of Singapore Graduate School for Integrative Sciences and Engineering and the Department of Electrical & Computer Engineering, National University of Singapore (e-mail: eleleeth@nus.edu.sg).

978-1-4244-8126-2/10/$26.00 ©2010 IEEE

methods: fast Frobenius algorithm and Jacobi angles for joint diagonalization.

The paper is organized as follows: Section II provides brief descriptions of the CSP algorithm and the JAD methods. In Section III the methodologies synthesizing MCSP and classifiers are described. The Data and experimental paradigm are presented in Section IV, and followed by comparative results in Section V. In Section VI the conclusions are drawn up with a brief discussion on the results.

II. Common Spatial Patterns

The ultimate success of a learning machine relies typically on the proper preprocessing of the data. In practice, we can discard non-informative dimensions of the data and thus select the features of interest for classification [3].

The CSP algorithm had proven to be highly successful in computing spatial filters for EEG data. The CSP algorithm had first been presented by Koles [2] as a method to extract the abnormal components from EEG, using a set of patterns that are common to both the normal and the abnormal recordings and have a maximally different proportion of the combined variances. Later CSP was used to extract features for classification from EEG signals [3]. The first and last few CSP components (the spatial filters that maximize the difference in variance) are used to classify the trials with high accuracy.

Let the random variable $\vec{x} \in R^N$ represent the EEG data, recorded through N electrodes, from which the intention of the BCI user $c \in C = \{c_1, \dots, c_M\}$ is to be inferred. Denote the class probability by $P(c_i), i = 1, \dots, M$ and assume that the EEG data conditioned on any class follows a Gaussian distribution with zero mean, i.e.

$P(\vec{x}|C_i) = N\big(0, R_{X|Ci}\big), i = 1, \dots, M.$

Then a linear transformation $w \in R_{N\times L}$ can be found where $L \ll N$, such that for finite training data using the reduced dimension $\hat{x} = W^T x$. This reduced dimension would lead to an increased classification accuracy in comparison to using $\vec{x}$.

A. Two-Class CSP

Consider a two-class paradigm, i.e., $C = \{c_1, c_2\}$. The CSP algorithm then solves the optimization problem

$$\vec{w}^* = \underset{\vec{w} \in R^N}{argmax} \left\{\frac{\vec{w}^T R_{x/c1}\vec{w}}{\vec{w}^T R_{x/c2}\vec{w}}\right\}, \quad (1)$$

where $R_{x/c1}$ and $R_{x/c1}$ are covariance matrices of $\vec{x}$ with given c_1 and c_2. Solutions to (1), which is in the form of Rayleigh quotient,can be found by solving the generalized eigenvalue problem,

$$R_{x/c1}\vec{w} = \lambda R_{x/c2}\vec{w}. \quad (2)$$

The eigenvectors of (2) thus correspond to the desired spatial filters. The corresponding eigenvalue determines the value of the cost function

$$\lambda^* = \left\{\frac{\vec{w}^{*T} R_{x/c1}\vec{w}^*}{\vec{w}^{*T} R_{x/c2}\vec{w}^*}\right\}.$$

The eigenvalues are the measure of the quality of the obtained spatial filters, i.e., the eigenvalue associated with a spatial filter expresses the ratio of the variance between conditions of the component of the EEG data extracted by the spatial filter. Preprocessing is then done by combining the L eigenvectors of (2) with the smallest/largest eigenvalues to form $W \in R^{N\times L}$ and computing $\hat{x} = W^T x$, where x denotes the original EEG signal and $\hat{x}$ is the signal with reduced dimensions.

B. Fast Frobenius Algorithm for Joint Diagonalization

The fast algorithm for joint diagonalization (FFDIAG) is based on the Frobenius norm formulation. Frobenius norm formulation had been used in various joint diagonalization approaches [7-9].Define,

$$F^k = VC^kV^T, \quad (3)$$

which denotes the result of applying transformation V to matrix C^k. Joint diagonalization can be defined as the following optimization problem:

$$\underset{V \in R^{NXN}}{min} \sum_{k=1}^{K} M_D(F^k),$$

where the diagonality measure M_D is the Frobenius norm of the off-diagonal elements in F^k:

$$M_D\left(F^k\right) = off(F^k) = \textstyle\sum_{i\neq j}(F_{ij}^k)^2. \quad (4)$$

The FFDIAG proposed by [5] is an iterative scheme to approximate the solution of the following optimization problem:

$$\underset{V \epsilon R^{NXN}}{min} \textstyle\sum_{k=1}^{K}\sum_{i\neq j}((VC^kV^T)_{ij})^2. \quad (5)$$

The invertibility of the matrix V is used as a constraint preventing convergence of the cost function to the trivial solution of V=0. Invertibility can be enforced by carrying out the update of V in multiplicative form as

$V_{(n+1)} \leftarrow \big(I + W_{(n)}\big)V_{(n)},$

where I denotes the identity matrix. The update matrix $W_{(n)}$ is constrained to have zeros on the main diagonal, and n is the iteration number. In order to maintain invertibility of V it is sufficient to enforce invertibility of $I+W_{(n)}$.

According to the *Levi-Desplanques Theorem*, if an $n\times n$ matrix A is strictly diagonally-dominant, then it is invertible [5]. An $n\times n$ matrix A is said to be strictly diagonally dominant if,

$|a_{ii}| > \sum_{j\neq i}|a_{ii}|$, for all i=1,…,n.

The Levi-Desplanques theorem can be used to control invertibility of $I + W(n)$. The diagonal entries in $I + W(n)$ are all equal to 1. Therefore, it suffices to ensure that

$$\underset{i}{max} \sum_{j\neq i}|W_{ij}| = ||W_{(n)}||_\infty < 1.$$

This can be done by dividing $W(n)$ by its infinity norm whenever the latter exceeds some fixed $\theta < 1$. An even stricter condition can be imposed by using a Frobenius norm in the same way as

$$W_{(n)} \leftarrow \frac{\theta}{||W_{(n)}||_F} W_{(n)}.$$

To determine the optimal updates $W(n)$ at each iteration, first-order optimality constraints for the objective (5) are used. A special approximation of the objective function enables efficient computation of $W(n)$.

Let $D_{(n)}^k$ and $E_{(n)}^k$ denote the diagonal and off-diagonal parts of $C_{(n)}^k$, respectively. In order to simplify the optimization problem we assume that the norms of $W(n)$ and $E_{(n)}^k$ are small, i.e. quadratic terms in the expression for the new set of matrices can be ignored.

$$C_{(n+1)}^k = \left(I + W_{(n)}\right)\left(D_{(n)}^k + E_{(n)}^k\right)\left(I + W_{(n)}\right)^T,$$
$$C_{(n+1)}^k \approx D_{(n)}^k + W_{(n)}D_{(n)}^k + D_{(n)}^k W_{(n)}^T + E_{(n)}^k.$$

With these simplifications, and ignoring already diagonal terms D_k, the diagonality measure (4) can be computed using expressions linear in W,

$$F^k \approx \tilde{F}^k = WD^k + D^k W^T + E^k . \quad (6)$$

The linearity of terms in (6) allows to explicitly compute the optimal update matrix $W(n)$ minimizing the approximated diagonality criterion,

$$\min_W \sum_{k=1}^{K} \sum_{i \neq j} ((WD^k + D^k W^T + E^k)_{ij})^2.$$

The FFDIAG algorithm is able to approximate the joint diagonal matrix owing to the sparseness introduced by (6). If the $N(N-1)$ off-diagonal entries of the update matrix W are arranged as a vector

$$w = (W_{12}, W_{21}, \dots W_{ij}, W_{ji}, ..)^T.$$

where the order of elements in w reflects the pairwise relationship of the elements in W. If the $KN(N-1)$ off-diagonal entries of the matrices E_k are also arranged as,

$$e = \left(E_{12}^1, E_{21}^1, .. E_{ij}^1, E_{ji}^1, .., E_{ij}^k, E_{ji}^k ..\right).$$

A large but very sparse, $KN(N-1) \times N(N-1)$ matrix J is built in the following form

$$J = \begin{pmatrix} J_1 \\ \vdots \\ J_k \end{pmatrix} \; with\, J_k = \begin{pmatrix} D_{12}^k & & \\ & \ddots & \\ & & D_{ij}^k \end{pmatrix},$$

where each J_k is block-diagonal, containing $N(N-1)/2$ matrices of dimension 2×2.

$$D_{ij}^k = \begin{pmatrix} D_j^k & D_i^k \\ D_j^k & D_i^k \end{pmatrix}, i, j = 1, \dots, N, i \neq j,$$

where D_i^k is a short-hand notation for the ii-th entry of a diagonal matrix D^k. The approximate cost function can be re-written as the linear least-squares problem

$$L(w) = \sum_k \sum_{i \neq j} (\tilde{F}_{ij}^k)^2 = (jw + e)^T (jw + e).$$

The solution to this problem put forward by [10] is

$$w = -(J^T J)^{-1} J^T e. \quad (7)$$

Using the sparseness of J and e to enable the direct computation of the elements of w in (7), the matrix product $J^T J$ can be written as a block-diagonal matrix

$$J^T J = \begin{pmatrix} \sum_k {(D_{12}^k)}^T D_{12}^k & & \\ & \ddots & \\ & & \sum_k {(D_{ij}^k)}^T D_{ij}^k \end{pmatrix},$$

whose blocks are 2x2 matrices. Thus the system (7) actually consists of decoupled equations,

$$\begin{pmatrix} W_{ij} \\ W_{ji} \end{pmatrix} = -\begin{pmatrix} z_{jj} & z_{ij} \\ z_{ij} & z_{ii} \end{pmatrix}^{-1} \begin{pmatrix} y_{ij} \\ y_{ji} \end{pmatrix}, \; i, j = 1, \dots, N, i \neq j$$

where $z_{ij} = \sum_k D_i^k D_j^k$ and

$y_{ij} = \sum_k D_j^k \frac{E_{ij}^k + E_{ji}^k}{2} = \sum_k D_j^k E_{ij}^k$.

The matrix inverse can be computed in closed form, leading to the following expressions for the update of the entries of W

$$W_{ij} = \frac{z_{ij} y_{ji} - z_{ii} y_{ij}}{z_{jj} z_{ii} - z_{ij}^2}$$

$$W_{ji} = \frac{z_{ij} y_{ij} - z_{jj} y_{ji}}{z_{jj} z_{ii} - z_{ij}^2}.$$

Therefore only the off-diagonal elements ($i \neq j$) need to be computed and the diagonal terms of W are set to zero. This makes this algorithm faster than other JAD methods [5].

C. Jacobi Angles for Simultaneous Diagonalization

Another approach for joint approximate diagonalization (JAD) is known as Jacobi angles for joint diagonalization. This method is based on the Jacobi technique which is a joint diagonality criterion optimized iteratively under plane rotations [7].

Consider a set, $C = \{C_k | k = 1, K\}$ of K, N×N matrices. The off-diagonal elements of C can be defined as

$$off(C) = \sum_{1 \leq i \neq j \leq N} |c_{ij}|^2 \quad (8)$$

where a_{ij} denotes the (i,j)-th entry of matrix C. Simultaneous diagonalization can be obtained by minimizing the composite objective $\sum_{k=1,K} off(UC_k U^H)$, by a unitary matrix U where the superscript H denotes the Hermitian transpose. The extended Jacobi technique for simultaneous diagonalization constructs U as a product of plane rotations globally applied to all the matrices in C. A plane rotation in the (i,j)-plane is a unitary matrix $R = R(i,j,c,s)$ defined as

$$R = I + (c-1)e_i e_i^T - s e_i e_j^T + s e_j e_i^T + (c-1)e_j e_j^T$$

where $c, s \in C$ and $|c|^2 + |s|^2 = 1$.

It is desired for each choice of $i \neq j$, finding complex angles c and s that minimizes the following objective function:

$$O(c,s) = \sum_{k=1,K} off\left(R(i,j,c,s) C_k R^H(i,j,c,s)\right).$$

For a given pair (i,j) of indices, a 3×3 real symmetric matrix G is defined as

$$G = Real\left(\sum_{k=1,K} h^H(C_k) h(C_k)\right).$$

For any set A of N×N matrices the following theorem allows the Jacobi angles to be computed [7]. Under constraint $|c|^2 + |s|^2 = 1$, the objective function $O(c,s)$ is minimized at,

$$c = \sqrt{\frac{x+r}{2r}},\ s = \frac{y-iz}{\sqrt{2r(x+r)}} \text{ and } r = \sqrt{x^2+y^2+z^2}, \quad (9)$$

where $[x,y,z]^T$ is any eigenvector associated with the largest eigenvalue of G [7]. Proof of this theorem can be found in [7].

Thus, the minimization of $O(c,s)$ under the constraint $|c|^2+|s|^2=1$ is equivalent to maximization of real 3×3 quadratic form under unit norm constraint. The solution is given by unit norm eigenvector of G associated with the maximum eigenvalue. More theoretical analysis of this method can be found in [20].

When C_k is a set of real symmetric matrices, the rotation parameters c and s become real: the last component of each vector $h(C_k)$ then is zero and G can be reduced to a 2×2 matrix by deleting the last row and last column.

III. Synthesized Methods

We investigate the use of the FFDIAG algorithm and Jacobi angles method for approximate diagonalization to develop multiclass common spatial patterns.

The first algorithm is implemented by utilizing the FFDIAG method to jointly diagonalize M number of covariance matrices. The covariance matrices C^k are given in (3). The Frobenius norm is calculated according to (4) and the minimization problem shown in (5) is iteratively deduced as explained in the section II. The resulting eigenvectors are employed to spatially filter the covariance matrices.

The second method based on Jacobi angles also takes the multiple covariance matrices as inputs. This corresponds to the matrix A in equation (8). The real part of the resulting diagonalized matrix is used to spatially filter the covariance matrices.

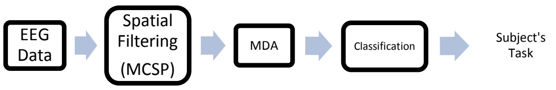

Fig. 1: Architecture of Multiclass Common Spatial Pattern

Multiple discriminant analysis (MDA) is carried out in order to select the most discriminating features from the filtered covariance data. Thirteen features were selected in order to distinguish the four classes. These selected features were used to train the classification algorithms and the 10×10 cross-validation accuracies are calculated. Fig. 1 depicts the architecture of the proposed multiclass CSP method.

The performances of the implemented spatial filters are compared with one another and one versus rest multiclass CSP using three multiclass classifiers. K-Nearest Neighbour, Classification and Regression Trees, and Support Vector Machine classifiers are implemented and the performances are compared.

A. k-NN algorithm

The k-nearest neighbor (k-NN) [14] is a classifier that assigns the class label of a new data based on the class with the most occurrences in a set of k nearest training data points usually computed using a distance measure such as the Euclidean distance. The k-nearest neighbor implementation in the Matlab Bioinformatics toolbox with k=5 is used in this paper.

B. CART algorithm

Decision tree is a classifier which uses symbolic tree like representations of finite sets of if-then-else questions that are natural, intuitive and interpretable. They are multistage decision systems in which classes are sequentially rejected until we reach a finally accepted class. The feature space is split into unique regions, corresponding to the classes, in a sequential manner. Upon the arrival of a feature vector, the searching of the region to which the feature vector will be assigned is achieved via a sequence of decisions along a path of nodes of an appropriately constructed tree. Such schemes offer advantages when a large number of classes are involved [11].

The Classification and Regression Tree (CART) [12] implementation in the Matlab Statistics toolbox is used in this work.

C. SVM algorithm

The Support Vector Machine (SVM) [13] is a linear discriminant that maximizes the separation between two classes based on the assumption that it improves the classifier's generalization capability. In this implementation a one versus rest multiclass SVM was applied to classify the four classes of data. A Gaussian kernel with penalty parameter (Bound on the lagrangian multipliers) of 45 was found to give the highest cross-validation accuracies.

The goal of pattern classification is to find a rule that assigns an object to one of several possible classes. According to the "No Free Lunch" theorem, there is no general superiority of any approach over the others in pattern classification. If one approach seems to outperform another in a particular situation, it is a consequence of its fitness to the particular pattern recognition problem [11]. Therefore, the performances of the classification algorithms were also analyzed in combination with the two aforementioned multiclass CSP algorithms for multiclass motor imagery-based BCI.

IV. Data and Experimental Procedure

The data set 2a of the fourth BCI Competition IV (2008) [4] is considered in this study. This data set is composed of EEG data collected from 9 subjects that have been recorded during two sessions on different days for each subject. The synchronous BCI data had been collected for four different motor imagery tasks. The imagination of movements of the left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4) had been considered as the four motor imagery tasks. Each session had been made up of 6 runs separated by short breaks. One run had included 48 trials (12 for each of the four possible classes), amounting to a total of 288 trials per session.

The subjects had been seated on an armchair in front of a computer screen and at the beginning of a trial (t = 0 s), a fixation cross had appeared on the black screen. Short acoustic warning tones had also been presented at the start of the trial. After two seconds (t = 2 s), a cue had been presented. This cue could have been in the form of an arrow pointing either to the left, right, down or up (corresponding to one of the four classes left hand, right hand, foot or tongue). The cue had appeared and stayed on the screen for 1.25 seconds and this was expected to induce the subjects to perform the desired motor imagery task. The subjects had been instructed to carry out the motor imagery tasks until the fixation cross disappeared from the screen at t = 6 sec. without any feedback on their performance. A short break had been given before the next trial and this procedure had been repeated for each of the 6 runs in a session. The timing scheme of this paradigm is depicted in Fig.2.

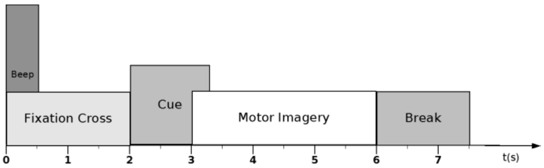

Fig. 2: Timing scheme of the BCI paradigm

EEG signals are recorded from 22 scalp positions, mainly covering the primary motor cortices bilaterally. The signals are sampled at 250 Hz and subject to a bandpass filter between 0.5 Hz and 100 Hz. The sensitivity of the amplifier is set to 100 µV. An additional 50 Hz notch filter had been utilized to suppress line noise.

EEG Signals are first band-pass filtered in the range of 8–30 Hz and then samples were partitioned into ten parts before building the multiclass CSP spatial filter. Each part was used as test set only once in the following way. The spatial filters were calculated on the basis of the 90% portion (nine parts) and were then multiplied to these data. Next, the selected features were subjected to MDA.

For the classification of the four classes of EEG data three classifiers that support multiclass classification were considered. First a k-Nearest Neighbor Classifier was attempted. Classification Trees (CART) and Support Vector Machines (SVM) were also applied. The time slice between 0.5 and 2.5 sec. was used to train the classifiers and to calculate the spatial filters in MCSP. The performance of the proposed framework was assessed using cross-validation. 10 by 10 Cross-validation was carried out on the training data set.

Each classifier was provided with the best 13 features selected from MDA. The classifier weights were calculated and these classifiers and the spatial filter were then applied to the remaining 10% of the data. The whole procedure was repeated ten times, i.e., a 10×10-fold cross-validation procedure [11] was performed and classification accuracies were determined.

V. Results

Cross-validation results obtained for the proposed methods of multiclass CSP based on FFDIAG and Jacobi Angles with k-NN classifier are depicted in table 1. One over rest application of the binary CSP is also presented in order to compare the performances.

Table 2 shows the cross validation results obtained for the same multiclass CSP method where the classification is carried out by Classification and Regression Trees (CART) algorithm. Results obtained for the classification by Support Vector Machines (SVM) is presented in table 3.

TABLE 1

CLASSIFICATION ACCURACIES OF CROSS VALIDATION (K-NN CLASSIFIER)

Subject	1	2	3	4	5	6	7	8	9	Avg.
CSP –FFDIAG	49.3	40.3	49.4	49.3	48.6	49.3	48.2	50.1	49.2	48.2
CSP-Jacobi	29.1	27.4	28.9	29.2	28.7	29.4	27.9	32.1	29.1	29.1
CSP (OVR)	26.3	25.1	26.2	25.1	26.9	24.3	26.1	27.0	25.8	25.9

TABLE 2

CLASSIFICATION ACCURACIES OF CROSS VALIDATION (CART CLASSIFIER)

Subject	1	2	3	4	5	6	7	8	9	Avg.
CSP –FFDIAG	43.8	35.6	44.2	43.4	43.1	43.5	41.9	44.7	43.9	42.7
CSP-Jacobi	25.4	24.7	25.2	26.5	25.1	26.7	24.8	29.6	25.3	25.9
CSP (OVR)	26.1	24.8	25.9	24.5	26.3	24.1	25.4	26.8	25.3	25.5

TABLE 3

CLASSIFICATION ACCURACIES OF CROSS VALIDATION (SVM CLASSIFIER)

Subject	1	2	3	4	5	6	7	8	9	Avg.
CSP –FFDIAG	63.2	58.8	64.2	42.1	39.4	42.6	56.3	69.3	45.9	53.6
CSP-Jacobi	33.4	30.9	31.2	33.7	32.4	33.1	31.8	35.3	33.5	32.8
CSP (OVR)	26.9	23.3	28.9	27.6	27.8	28.1	28.9	29.5	29.8	27.8

The highest average classification accuracy of 53.6% is recorded by the JAD method based on FFDIAG when the classification is carried out by multi-class SVM. The same JAD method records average accuracies of 48.2 and 42.7 under k-NN and CART classification methods respectively. The classification accuracies of 10×10-fold cross-validation indicate that the JAD method based on FFDIAG clearly outperforms others. The Jacobi angles based method slightly outperforms the one versus rest binary CSP.

VI. Conclusions

In this paper two machine learning approaches are adopted for multiclass Common Spatial Patterns for processing EEG measurements in multiclass motor imagery-based BCI based on JAD methods. MCSP extends the binary CSP technique to a truly multiclass paradigm and proves to be better than one versus rest application of the binary CSP.

The proposed JAD methods are compared on the BCI Competition IV for dataset 2a. Experimental results show that the proposed MCSP based on FFDIAG yields superior classification accuracy compared to the alternative MCSP methods. Furthermore, the FFDIAG method is much faster than the Jacobi angles method.

In the analysis carried out on the three classification algorithms it is identified that the SVM algorithm consistently gives a higher accuracy than the other two classification methods. Though k-NN also performs quite well in this data set, the nonlinearity of the implemented SVM classifier might have given it the edge over the other two linear classifiers.

Future work in this area would include the extension of the FFDIAG based multiclass CSP to the Filter Bank CSP (FBCSP) [19] method which was the winning method for the BCI Competition IV for dataset 2a.

ACKNOWLEDGMENTS

The authors would like to thank J.-F. Cardoso for providing the source code for the implementation of the Jacobi angles algorithm.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," Clin. Neurophysiol., vol. 113, no. 6, pp. 767–791, 2002.

[2] Z. J. Koles, "The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG," Electroencephalogr. Clin. Neurophysiol., vol. 79, pp. 440–447, 1991.

[3] H. Ramoser, J. Mueller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," IEEE Trans. Rehabil. Eng., vol. 8, no. 4, pp. 441–446, Dec. 2000.

[4] B. Blankertz, "BCI Competition IV", Fraunhofer FIRST.IDA,http://ida.first.fraunhofer.de/projects/bci/competition_iv/.

[5] A. Ziehe, P. Laskov, G. Nolte, K-R. Muller ,"A Fast Algorithm for Joint Diagonalization with Non-orthogonal Transformations and its Application to Blind Source Separation", Journal of Machine Learning Research 5, pp.777–800. 2004.

[6] J.-F. Cardoso and A. Souloumiac. "Blind beamforming for non Gaussian signals", IEE Proceedings, Part F, 140(6):362–370, 1993.

[7] J.-F. Cardoso and A. Souloumiac. Jacobi angles for simultaneous diagonalization. SIAM journal on matrix analysis and applications, 17(1):161–164, January 1996.

[8] M. Joho and K. Rahbar, "Joint diagonalization of correlation matrices by using Newton methods with application to blind signal separation", In Proceedings of IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), pp. 403–407, 2002.

[9] G. Hori, "Joint diagonalization and matrix differential equations", In Proceedings of International Symposium on Nonlinear Theory and its Applications (NOLTA'99), pp. 675–678, 1999.

[10] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical Recipies in C", Cambridge University Press., Cambridge, 1992.

[11] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern Classification", 2nd ed. New York:John Wiley, 2001.

[12] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, "Classification and regression trees", Belmont, CA:Wadsworth International Group, 1984.

[13] V. N. Vapnik. "Statistical learning theory", New York:Wiley, 1998.

[14] T. Cover and P. Hart, "Nearest neighbor pattern classification," IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21-27, 1967.

[15] J. Mueller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," Clinical Neurophysiology, vol. 110, no.5, pp. 787-798, 1999.

[16] G. Dornhege, B. Blankertz, G. Curio, and K. R. Muller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," IEEE Transactions on Biomedical Engineering, vol. 51, pp. 993-1002, 2004.

[17] M. Grosse-Wentrup and M. Buss, "Multiclass Common Spatial Patterns and Information Theoretic Feature Extraction," IEEE Transactions on Biomedical Engineering, vol. 55, no. 8, pp. 1991–2000, Aug 2008.

[18] K. Fukunaga, "Introduction to statistical pattern recognition", (2nd ed.). San Diego, CA, USA: Academic Press Professional, Inc., 1990.

[19] K. K. Ang, Z. Y. Chin, and C. Guan, "Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface," in Proceedings of the International Joint Conference on Neural Networks, (IJCNN 2008), pp. 2391-2398,2008.

[20] A.Bunse-Gerstner, R. Byers, and V. Mehrmann, "Numerical methods for simultaneous diagonalization", SIAM journal on matrix analysis and applications, vol. 14, pp.927-949,1993.