

A comparison study on EEG signal processing techniques using motor imagery EEG data

Vangelis P. Oikonomou*, Kostas Georgiadis*, George Liaros*, Spiros Nikolopoulos* and Ioannis Kompatsiaris*

**Information Technologies Institute,*

Centre for Research and Technology Hellas, CERTH-ITI,

6th km Charilaou-Thermi Road, 57001 Thermi-Thessaloniki, Greece

Email: viknmu, kostas.georgiadis, geoliaros, nikolopo, ikom@iti.gr

Abstract—Brain-computer interfaces (BCIs) have been gaining momentum in making human-computer interaction more natural, especially for people with neuro-muscular disabilities. Among the existing solutions the systems relying on electroencephalograms (EEG) occupy the most prominent place due to their non-invasiveness. In this work, we provide a review of various existing techniques for the identification of motor imagery (MI) tasks. More specifically, we perform a comparison between Common Spatial Patterns (CSP) related features and features based on Power Spectral Density (PSD) techniques. Furthermore, for the identification of MI tasks, two well-known classifiers are used, the Linear Discriminant Analysis (LDA) and the Support Vector Machines (SVM). Our results confirm that PSD features demonstrate the most consistent robustness and effectiveness in extracting patterns for accurately discriminating between left and right MI tasks.

Keywords-Brain Computer Interface, Common Spatial Patterns, Power Spectral Density, Linear Discriminant Analysis, Support Vector Machine

I. INTRODUCTION

Brain Computer Interface (BCI) term is referred to as an artificial communication channel between human brain and the external environment by using machines (mostly computers) [1]. It is a channel out of the ordinary communication channel of human brain, the nerves. By using this channel, the human is able to send messages and commands to its external environment. A BCI system measures brain activity and translates it into control signals. These control signals can be used to construct new augmentative technologies. People with motor disabilities need augmentative technologies corresponding to natural ways of communications. Those who are totally paralyzed, or locked-in, cannot use conventional augmentative technologies, since some measure of muscle control is required. The immediate goal of a BCI system is to provide these users with basic communication capabilities, so that they can express their wishes to caregivers or even operate word processing programs or neuroprostheses. It is obvious that a BCI system could improve their quality of life, while at the same time reduce the cost of intensive care. Besides medical applications, a

BCI system provides us with the opportunity to facilitate the communication of human with machines/computers by constructing interfaces which are based on more natural ways of communication beyond mouse and keyboards. The brain responses can be measured by adopting various acquisition modalities such as Functional Magnetic Resonance Imaging (fMRI), functional Near-Infrared Spectroscopy (fNIRS) and electroencephalography (EEG). From the above acquisition modalities, the EEG signal is the most frequently used because of its noninvasiveness, its high time resolution, ease of acquisition, and cost effectiveness compared to other brain activity monitoring modalities [2].

BCI systems can be divided into two main categories. In the first category, belong systems where the user focuses attention on a set of stimuli, which produce an autonomic response that can be detected by the system, for example Steady State Visual Evoked Potential (SSVEP) BCI systems. In the second category, the user performs a mental task, such as imagined movement or sub-vocal counting, to create changes in brain signals that can be detected by a BCI, for example Motor Imagery (MI) BCI. The MI mental strategy detects brain activity patterns that are generated during the imagination of movement. This technique is based on the Sensorimotor Rhythms (SMR), brain waves that appear on EEG recordings from areas of the brain that are associated with planning, control, and execution of voluntary movements [3].

The input of a BCI system is the electrophysiological brain activity, while the output is a device command. The brain activity is recorded through the use of an EEG system. After that, the analysis of EEG signals is performed in order to extract the intended commands of the user. The goal of a BCI system is to translate brain activity into a command for a computer. To achieve this goal machine learning algorithms are used. More specifically, a BCI system depends on whether the following two requirements can be satisfied: (1) the extracted EEG features are able to differentiate the task-oriented brain states; and (2) the methods for classifying such features are efficient [4].

One major challenge of this field is thus to extract reliable

information from noisy data in real time in the form of relevant features. These can then be passed on to classification techniques for identifying the user's mental state frequencies (or parts) of EEG, depending on the application. With respect to EEG-BCI systems, feature extraction approaches are dominated by methods estimating the distribution of energy in various domains, such as the time domain, the frequency domain, the time - frequency (t-f) domain, the wavelet domain and the spatial domain. In order to extract time domain features, a digital filter is applied to EEG segments. The goal of filtering is to select the EEG rhythms components depending on the frequency band of interest and then calculate the energy of the filtered EEG segments [5] (band power features). Also in time domain, AutoRegressive (AR) and Adaptive AutoRegressive (AAR) parameters [5], [6] can be extracted and used as features by fitting in EEG segments an AR model (AR features) in conjunction with a Kalman Filter (AAR features). In [7] the EEG segments are represented in the frequency domain by estimating the Power Spectrum Density (PSD). The PSD provides us with an energy distribution across the frequencies bins (PSD features). In order to examine the time varying characteristics inside the EEG segment we need to transform the signal in the t-f domain or in the wavelet domain. Hence, t-f features and wavelet - based features have been also used in MI BCI [7], [8]. Finally, features that exploit the energy distribution of EEG into the spatial domain are used in [5], [9]–[11] (CSP features).

In MI BCI, the extracted features are fed into a classifier to identify the user's mental state. Many classifiers have been used to predict the user's mental state. A comparison between non linear and linear classifiers is provided in [12]. According to the results, the nonlinear classifiers present slightly better performance than the linear ones [12]. In [7], various classifiers have been used for the identification of motor tasks. More specifically, a comparison between Linear Discriminant Analysis (LDA) and various extensions of Support Vector Machines (SVM) is provided. The main outcome of [7] is that the SVM with a gaussian kernel is the most appropriate classifier for the examined problem. In addition, at the same work a genetic algorithm had to be used to fine-tune SVM. This fact increases considerably the overall tuning of a MI BCI system. To avoid any interference from the SVM training procedure in [13] the LDA classifier was used.

A comparison of different feature extraction techniques in the context of BCI was presented in [7], together with their use in classification of motor imagery. Also, in [13] a comparative study of different PSD feature extraction approaches was presented. While both studies [7], [13] use in some cases the same method for feature extraction, there are differences between them which are concentrated at how the EEG segment is processed. Furthermore in [13] the log-transform was applied to the extracted PSD

features. Our work aims at complementing, and reproducing independently, the observations made in those two studies. In the next sections, we detail how we extract the features and present the corresponding techniques we use to extract the relevant information. Then, we present the data set under study and the results of our experiments. Finally, some conclusions and recommendations, based on our results, are provided.

II. METHODS

The next subsections describe in details the feature extraction techniques and the classifiers and how they are used in the present study. For feature extraction, we considered using the Welch's method for spectrum estimation and the CSP-related features, while, for the classification stage, we used the LDA and SVM classifiers.

A. Feature Extraction Procedure

1) *Welch Spectrum*: A well-known non parametric method for PSD estimation is the Welch's method [14]. Let $x[n], n = 0, \dots, N - 1$ are the samples from an EEG segment. To estimate the Welch's spectrum of this segment three basic steps are applied:

- 1) First, divide the original N length sequence into K sections (possibly overlapped) of equal lengths M .

$$x_i[m] = x[m + iD], \quad \begin{matrix} i = 0, \dots, K - 1 \\ m = 0, \dots, M - 1 \end{matrix} \quad (1)$$

- 2) Apply a window to each section and then calculate the periodogram on the windowed sections (modified periodograms).

$$P_i(f) = \frac{1}{NU} \left| \sum_{n=0}^{N-1} w[m] \cdot x_i[m] e^{-j2\pi f m} \right|^2, \quad \begin{matrix} i = 0, \dots, L - 1 \end{matrix} \quad (2)$$

where $U = \frac{1}{M} \sum_{m=0}^{M-1} w[m]^2$ is a normalization constant with respect to the window function.

- 3) Average the modified periodograms from the K sections in order to obtain an estimator of the spectral density.

$$P^W(f) = \frac{1}{K} \sum_{i=0}^{K-1} P_i(f) \quad (3)$$

In our analysis, we use eight sections of equal length with 50% overlap for the first step and the Hamming window in the second step.

B. Common Spatial Patterns

Spatial filtering is the process of filtering the signals by using information from the spatial domain. More specifically, in the spatial filtering, "new" channels are created as a combination of the original ones. In EEG analysis, well-known spatial filters are the bipolar and Laplacian

which are local spatial filters. A bipolar filter is defined as the difference between two neighboring channels, while a Laplacian filter is defined as four times the value of a central channel minus the values of the four neighboring channels around. The above spatial filters are defined *a priori*, i.e., the filter coefficients are known and fixed. There are different ways to define spatial filters. In particular, the filter coefficients (or weights) can be fixed in advance or they can be data driven, i.e., the weights are obtained during a learning procedure.

The CSP algorithm is an algorithm that provides us with a set of spatial filters. These filters are obtained after performing a learning procedure, during which the variance of the spatially filtered signals is maximized for one class (e.g., one mental imagery task) and minimized for the other class. At the beginning the CSP algorithm was applied on multichannel data from two classes/conditions [9]. However, extensions of the algorithm to handle multi-class problems have been proposed [15], [16].

The CSP algorithm performs a decomposition of the signal through the matrix \mathbf{W} , which contains the spatial filters. More specifically, this algorithm transforms the EEG signal from the original into a new domain which is occupied by the "new" channels

$$\mathbf{x}^{(CSP)} = \mathbf{W}\mathbf{x}, \quad (4)$$

where $\mathbf{x} \in \mathbb{R}^{C \times 1}$ is the EEG signal at time point t , $\mathbf{x}^{(CSP)} \in \mathbb{R}^{C \times 1}$ is the decomposed "new" EEG signal and $\mathbf{W} \in \mathbb{R}^{C \times C}$ is the matrix with the spatial filters $\mathbf{w}_i, i = 1, \dots, C$, and C is the number of channels. The spatial filters are obtained by maximizing (or extremizing) the following function [10]:

$$\begin{aligned} J(\mathbf{w}) &= \frac{\mathbf{w}^T \mathbf{X}_1^T \mathbf{X}_1 \mathbf{w}}{\mathbf{w}^T \mathbf{X}_2^T \mathbf{X}_2 \mathbf{w}} \\ &= \frac{\mathbf{w}^T \mathbf{C}_1 \mathbf{w}}{\mathbf{w}^T \mathbf{C}_2 \mathbf{w}} \end{aligned} \quad (5)$$

where T denotes transpose, \mathbf{X}_i is the data matrix containing the trials from i -th class, \mathbf{C}_i is the covariance matrix of i -th class. The above maximization problem is equal to maximize $\mathbf{w}^T \mathbf{C}_1 \mathbf{w}$ subject to the constraints $\mathbf{w}^T \mathbf{C}_2 \mathbf{w} = 1$. The last problem is equivalent to the generalized eigenvalue problem $\mathbf{C}_1 \mathbf{w} = \lambda \mathbf{C}_2 \mathbf{w}$. So, the spatial filters \mathbf{w}_i are the generalized eigenvectors of the above problem. It is worth to note here that in most cases after the application of CSP algorithm for spatial filtering an additional step is performed in order to extract CSP-related features [10]. Once the spatial filters \mathbf{w}_i are obtained, CSP feature extraction consists in filtering the EEG signals using the \mathbf{w}_i and then computing the resulting signals variance. As reported in [10] it is common to select three pairs of CSP spatial filters, corresponding to the three largest and smallest eigenvalues, hence resulting in a trial being described by six CSP features. However, the above heuristic choice of the number of spatial filters depends

heavily from the nature of the data (i.e., number of channels), as well as from the data analysis perspective (i.e., using filterbanks or not)

In motor BCI experiments special interest attracts the brain activity in the motor cortex, which is reflected into the electrical activity that is acquired from two channels, C3 and C4. This information could be used in order to construct the feature vectors. More specifically, this information could be combined by concatenating the features from each channel. For this reason the features that are extracted from each channel, for both aforementioned methods, are concatenated into one feature vector \mathbf{f} .

C. MI Classification

1) *SVM classifier*: A popular classification algorithm is the SVM [17], which aims to find the optimal hyper-plane that separates the positive class from the negative class by maximizing the margin between the two classes. This hyperplane, in its basic linear form, is represented by its normal vector \mathbf{w} and a bias parameter b . These two terms are the parameters that are learnt during the training phase. Assuming that the data are linearly separable, there exist multiple hyper-planes that solve the classification problem. SVMs choose the one that maximizes the margin, assuming that this will generalize better to new unseen data. This hyper-plane is found by solving the following minimization problem:

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i$$

$$\text{s.t.}: y_i(\mathbf{w}^T \mathbf{f}_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \dots, N, \quad (6)$$

where ξ_i are slack variables relaxing the constraints of perfect separation of the two classes and C is a regularization parameter controlling the trade-off between the simplicity of the model and its ability to better separate the two classes.

In order to classify a new unseen test example \mathbf{f}_j , its distance from the hyper-plane is calculated by the following equation.

$$f(\mathbf{f}_j) = \mathbf{w}^T \mathbf{f}_j + b. \quad (7)$$

In many cases the use of SVM is followed by the kernel trick (a transformation of features), which allows for the hyper-plane to take various forms. This is accomplished by projecting the input data into a higher dimensional space using a Kernel function $K(f_i, f_j)$ to measure the distance between the training instances \mathbf{f}_i and \mathbf{f}_j .

2) *LDA*: Linear Discriminant Analysis (LDA) [17] works in a similar way with SVMs, by attempting to find the separating line between the two classes. However, LDA does consider the margin between the classes. More specifically, based on the assumption that the covariance matrices of the two classes are equal and have full rank ($\Sigma_0 = \Sigma_1 = \Sigma$),

the optimization problem degenerates to an analytic form for the optimal \mathbf{w} and b as a function of the covariance matrix (Σ) and the mean (μ_0, μ_1):

$$\mathbf{w} = \Sigma^{-1}(\mu_1 - \mu_0) \quad (8)$$

$$b = \frac{1}{2}(T - \mu_0^T \Sigma_0^{-1} \mu_0 + \mu_1^T \Sigma_1^{-1} \mu_1) \quad (9)$$

where T is a threshold separating the two classes.

III. RESULTS

1) *Graz dataset B*: This data set consists of EEG data from 9 subjects. The subjects were right-handed, had normal or corrected-to-normal vision and were paid for participating in the experiments. All volunteers were sitting in an arm-chair, watching at the screen monitor placed approximately 1m away from the eye level. For each subject 5 sessions are provided, whereby the first two sessions contain training data without feedback (screening), and the last three sessions were recorded with feedback. Three bipolar recordings (C3, Cz, and C4) were recorded with a sampling frequency of 250 Hz. They were bandpass-filtered between 0.5 Hz and 100 Hz, and a notch filter at 50 Hz was enabled. The placement of the three bipolar recordings (large or small distances, more anterior or posterior) was slightly different for each subject. The electrode position Fz served as EEG ground. Further information on this dataset can be acquired in [18].

A. Data analysis protocol

The time segment of 0.5 - 2.5s after the onset of the visual cue was used to train the algorithms. It is the same procedure as that used in [11]. To evaluate the algorithms a sliding window of 2 secs, from the beginning of the corresponding trial until the end of it, is used. A continuous classification output for each sample in the form of class labels (1, 2) is provided by each algorithm. A confusion matrix was built from all trials for each time point. From these confusion matrices, the time course of the accuracy as well as the kappa coefficient can be obtained. For each algorithm we provide the largest (or peak) kappa value and accuracy.

To construct the spectral features we applied Welch's method in each EEG trial for each channel (C3 and C4). Then the spectral features from each channel are concatenated to create the final spectral feature vector. For the extraction of CSP features we take into account the number of channels and in this case after the application of CSP filter and the calculation of log-variances we take two CSP - related features. Finally, the regularization parameter C of linear SVM was set to 1.

The Graz dataset B is composed by 5 sessions, 2 sessions have been created without the feedback procedure and the last 3 sessions with the feedback. To evaluate the various algorithms the following division between the sessions has been adopted [11]: the first 3 sessions have been used for the training (2 sessions without feedback and the 1st session of

Table I
CLASSIFICATION SCHEMES UNDER STUDY.

Feature Extraction	Classifier	Classification Scheme
CSP	LDA	CSP-LDA
CSP	SVM	CSP-SVM
Welch	SVM	PW-SVM
Welch	LDA	PW-LDA

using feedback), while the rest 2 sessions (with feedback) are used to evaluate the various algorithms. Also, the EEG data from channels C3 and C4 are used for further processing. Before applying the algorithms a band pass filter from 8 to 40 Hz have been applied on the data.

By combining the classifiers and the features extraction methods, we obtain the various classification schemes that are described in Table I. In addition to our presentation we have included the results that are reported in [11]. In [11] a filter bank in combination with the CSP filtering approach was used to extract the features. Then a filter selection algorithm based on mutual information was applied in order to select relevant features, and finally for the classification the Naive Bayesian Parzen Window (NBPW) classifier was used [11].

In Tables II and III, we show the results by using our algorithms as well as results reported in [11]. In all cases an identical procedure for data analysis is used (i.e., filtering, trial segmentation), the difference is on the machine learning algorithm. By observing the results on Tables II and III we can see that there is not an algorithm that performs best on all subjects. Furthermore, the algorithm reported in [11] provides the best mean kappa - value (0.59), while the Welch approach in conjunction with linear SVMs gives us the second best value (0.58). By applying a paired t-test (using `ttest` matlab function) we observe no significant statistical difference between the two methods ($p=0.5313$). Similar observations can be extracted by looking the classification results with respect to the measure of accuracy (see Table III).

By looking the results on Tables II and III, besides reporting the best performance, we can make additional observations. We can see that, when we use the CSP features, the performance between LDA and SVM is similar. However, when spectral features are used the behaviour of these two classifiers is very different. We can see that SVM outperforms LDA considerably. Furthermore, the performance of LDA, when spectral features are used, is the worst among all presented classification schemes. This fact can be explained by the high dimensionality of spectral features. Finally, it is worth to note that in our analysis no tuning of any type have been performed with respect to the regularization parameter C of SVM. This fact leaves space for further improvement in the performance of PW-SVM classification scheme.

Finally, it has to be noted that, for the particular dataset, the bipolar channel data at C3 and C4 differ between

participants with respect to the spatial location of the reference electrode, which has been selected based on the LDA classification accuracy [19]. In other words, the EEG channels that are available for the current analysis have been already optimized for LDA. It is expected that the accuracy of the SVM approach could be further increased by using other bipolar montages or channel data that has been referenced towards a common reference electrode.

IV. CONCLUSION

In this work, a comparison on the discriminative ability between the CSP related features and the spectral features was performed. Our results indicate that the spectral features of the signal are superior from CSP features if the appropriate classifier is used. The use of PSD estimation methods provides us with a feature set of high dimensionality. The classifier needs to take into account the above property in order to achieve its highest performance. This fact is evident when we compare the performance between the LDA and the SVM in the case where the input feature set is the PSD features.

ACKNOWLEDGMENT

This work is part of project MAMEM that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 644780.

REFERENCES

- [1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767 – 791, 2002.
- [2] P. Sajda, K. r. Muller, and K. V. Shenoy, "Brain-computer interfaces [from the guest editors]," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 16–17, 2008.
- [3] B. Graimann, B. Allison, and G. Pfurtscheller, *BrainComputer Interfaces: A Gentle Introduction*. Springer, 2010, ch. 1.
- [4] S.-M. Zhou, J. Q. Gan, and F. Sepulveda, "Classifying mental tasks based on features of higher-order statistics from eeg signals in brain - computer interface," *Information Sciences*, vol. 178, no. 6, pp. 1629 – 1640, 2008.
- [5] G. Pfurtscheller, C. Neuper, C. Guger, W. Harkam, H. Ramoser, A. Schlogl, B. Obermaier, and M. Pregenzer, "Current trends in graz brain-computer interface (bci) research," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 216–219, Jun 2000.
- [6] A. Schloegl, K. Lugger, and G. Pfurtscheller, "Using adaptive autoregressive parameters for a brain-computer-interface experiment," in *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE*, vol. 4, Oct 1997, pp. 1533–1535 vol.4.
- [7] P. Herman, G. Prasad, T. M. McGinnity, and D. Coyle, "Comparative analysis of spectral approaches to feature extraction for eeg-based motor imagery classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, no. 4, pp. 317–326, Aug 2008.
- [8] Q. Xu, H. Zhou, Y. Wang, and J. Huang, "Fuzzy support vector machine for classification of eeg signals using wavelet-based features," *Medical Engineering and Physics*, vol. 31, no. 7, pp. 858 – 865, 2009.
- [9] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Miller, "Optimizing spatial filters for robust eeg single-trial analysis," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2008.
- [10] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve bci designs: Unified theory and new algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, Feb 2011.
- [11] K. K. Ang, Z. 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*, vol. 6, no. 39, 2012.
- [12] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for eeg signal classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 141–144, June 2003.
- [13] N. Brodu, F. Lotte, and A. Lcuyer, "Comparative study of band-power extraction techniques for motor imagery classification," in *2011 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*, April 2011, pp. 1–6.
- [14] J. Proakis and D. Manolakis, *Digital Signal Processing (3rd Ed.): Principles, Algorithms, and Applications*. Prentice-Hall, Inc., 1996.
- [15] Z. Y. Chin, K. K. Ang, C. Wang, C. Guan, and H. Zhang, "Multi-class filter bank common spatial pattern for four-class motor imagery bci," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Sept 2009, pp. 571–574.
- [16] 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.
- [17] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [18] R. Leeb, C. Brunnera, G. R. Muller-Putz, A. Schloegl, and G. Pfurtscheller, "Bci competition 2008 - graz data set b," <https://lampx.tugraz.at/bci/database/002-2014/description.pdf>, 2008.
- [19] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, "Brain - computer communication: Motivation, aim, and impact of exploring a virtual apartment," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 15, no. 4, pp. 473–482, Dec 2007.

Table II
 CLASSIFICATION RESULTS WITH RESPECT TO KAPPA VALUE ON UNSEEN EVALUATION DATA ON DATASET GRAZ B

Sub no	CSP-LDA [11]	FBCSP-MIRSR [11]	CSP-SVM	PW-SVM	PW-LDA
1	0.3190	0.4000	0.2437	0.4625	0.3125
2	0.2290	0.2070	0.2929	0.2357	0.2071
3	0.1250	0.2190	0.1625	0.1687	0.1563
4	0.9250	0.9500	0.8938	0.8750	0.6500
5	0.5250	0.8560	0.4688	0.7500	0.6000
6	0.5000	0.6130	0.6313	0.6563	0.5313
7	0.5440	0.5500	0.4312	0.5375	0.3375
8	0.8560	0.8500	0.7937	0.8438	0.6375
9	0.6560	0.7440	0.7000	0.7500	0.3313
mean	0.5198	0.5987	0.5131	0.5866	0.4181

Table III
 CLASSIFICATION RESULTS WITH RESPECT TO ACCURACY (%) ON UNSEEN EVALUATION DATA ON DATASET GRAZ B

Sub no	CSP-LDA [11]	FBCSP-MIRSR [11]	CSP-SVM	PW-SVM	PW-LDA
1	65.95	70.00	62.18	73.12	65.62
2	61.45	60.35	64.64	61.78	60.35
3	56.25	60.95	58.12	58.43	57.81
4	96.25	97.50	94.69	93.75	82.50
5	76.25	92.80	73.44	87.5	80.00
6	75.00	80.65	81.56	82.81	76.56
7	77.20	77.50	71.56	76.87	66.87
8	92.80	92.50	89.68	92.19	81.87
9	82.80	87.20	85.00	87.50	66.56
mean	75.99	79.93	75.65	79.33	70.90