Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface

Kai Keng Ang, Zheng Yang Chin, Haihong Zhang, and Cuntai Guan

Abstract—**In motor imagery-based Brain Computer Interfaces (BCI), discriminative patterns can be extracted from the electroencephalogram (EEG) using the Common Spatial Pattern (CSP) algorithm. However, the performance of this spatial filter depends on the operational frequency band of the EEG. Thus, setting a broad frequency range, or manually selecting a subject-specific frequency range, are commonly used with the CSP algorithm. To address this problem, this paper proposes a novel Filter Bank Common Spatial Pattern (FBCSP) to perform autonomous selection of key temporal-spatial discriminative EEG characteristics. After the EEG measurements have been bandpass-filtered into multiple frequency bands, CSP features are extracted from each of these bands. A feature selection algorithm is then used to automatically select discriminative pairs of frequency bands and corresponding CSP features. A classification algorithm is subsequently used to classify the CSP features. A study is conducted to assess the performance of a selection of feature selection and classification algorithms for use with the FBCSP. Extensive experimental results are presented on a publicly available dataset as well as data collected from healthy subjects and unilaterally paralyzed stroke patients. The results show that FBCSP, using a particular combination feature selection and classification algorithm, yields relatively higher cross-validation accuracies compared to prevailing approaches.**

I. INTRODUCTION

A Brain-Computer Interface (BCI) is a system that empowers a user to command external devices, such as a computer, wheelchair control [1] or prostheses [2], using scalp-recorded electroencephalogram (EEG) measurements. Studies have shown that the EEG measurements recorded during mental imagination of movements can be translated into commands. Thus, motor imagery-based BCI provides a promising control and communication control to people suffering from motor disabilities [3].

The Common Spatial Pattern (CSP) algorithm [4],[5] is effective in constructing optimal spatial filters that discriminates two classes of EEG measurements in motor-imagery-based BCI [5]. However, the performance of this spatial filter is dependent on its operational frequency band. Classification performed on the CSP features generally yields poor accuracies when the EEG measurements are either unfiltered or have been filtered with an inappropriately selected frequency range [6]. Hence, setting a broad frequency range or manually selecting a subject-specific frequency range are commonly used with the CSP algorithm [7].

To address the problem of manually selecting the operational subject-specific frequency band for the CSP algorithm, several approaches have been proposed. One approach is the Common Spatio-Spectral Pattern (CSSP), which optimizes a simple filter that employs a one time-delayed sample with the CSP algorithm [8]. Another approach is the Common Sparse Spectral Spatial Pattern (CSSSP). It improves the CSSP algorithm by performing simultaneous optimization of an arbitrary (Finite Impulse Response) FIR filter within the CSP algorithm [7]. However, due to the inherent nature of the optimization problem, the solution of filter coefficients is also strongly dependent on the choice of initial parameters [6].

Recently, an alternative approach called the Sub-band Common Spatial Pattern (SBCSP) was proposed and has been shown to yield superior classification accuracy compared against CSSP and CSSSP on a publicly available dataset [6]. Instead of optimizing a single arbitrary FIR filter within the CSP algorithm, SBCSP uses a Gabor filter bank that decomposes the EEG measurements into multiple sub-bands. Spatial filters that use the CSP algorithm are then employed on each of these sub-bands. After obtaining sub-band scores, recursive band elimination or a classification algorithm is employed to fuse the sub-band score. Another classification algorithm is then employed to classify the fused sub-band score. Although SBCSP can use different sub-band score fusion techniques and classification algorithms, only the results from the use of the Support Vector Machine (SVM) to fuse the sub-band score as well as to perform classification are presented in [6]. Hence, a comparative study of using different sub-band score fusion techniques and classification algorithms are not available.

In this paper, a novel machine learning approach called the Filter Bank Common Spatial Pattern (FBCSP) is proposed for processing EEG measurements in motor imagery-based BCI. FBCSP comprises four stages: frequency filtering, spatial filtering, feature selection and classification. In the first stage, the EEG measurements are bandpass-filtered into multiple frequency bands. In the second stage, CSP features are extracted from each of these bands. In the third stage, a feature selection algorithm is used to automatically select discriminative pairs of frequency bands and corresponding CSP features. In the fourth stage, a classification algorithm is used to classify the CSP features. This paper also presents extensive

This work was supported by the Science and Engineering Research Council of A*STAR (Agency for Science, Technology and Research), and The Enterprise Challenge, Prime Minister's Office, Singapore.

The authors are with Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), 21 Heng Mui Keng Terrace, Singapore 119613. (email: {kkang, zychin, hhzhang, ctguan}@i2r.a-star.edu.sg).

978-1-4244-1821-3/08/$25.00©2008 IEEE

experimental results using a selection of feature selection and classification algorithms for use in FBCSP. The purpose of this study is to recommend suitable feature selection and classification algorithms for motor imagery-based BCI.

Section II describes the FBCSP approach and compares it with the SBCSP approach. A brief description of the CSP algorithm is then given in Section III. The feature selection and classification algorithms studied in this paper are briefly described in Section IV and Section V respectively. Extensive experimental results of using FBCSP on a publicly available dataset and experimental data collected from healthy subjects as well as unilaterally paralyzed stroke patients are given in Section VI. The results are also compared against the SBCSP and the CSP with manually configured bandpass filter parameters. Section VII concludes this paper.

II. FILTER BANK COMMON SPATIAL PATTERN

The proposed *Filter Bank Common Spatial Pattern* (FBCSP) machine learning approach is illustrated in Fig. 1. It comprises four progressive stages of EEG measurements processing: multiple bandpass filters using zero-phase Chebyshev Type II filters, spatial filtering using the CSP algorithm, feature selection of the CSP features, and classification of the selected CSP features.

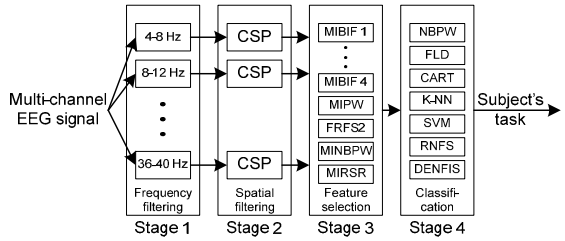

Fig. 1. Architecture of the proposed Filter Bank Common Spatial Pattern (FBCSP) machine learning approach

The first stage employs a filter bank that bandpass filters the EEG measurements into multiple bands. The second stage performs spatial filtering on each of these bands using the CSP algorithm. Thus, each pair of bandpass and spatial filter yields CSP features that are specific to the frequency range of the bandpass filter. The third stage employs a feature selection algorithm to select the discriminative CSP features from the filter bank. The fourth stage employs a classification algorithm to model and classify the selected CSP features.

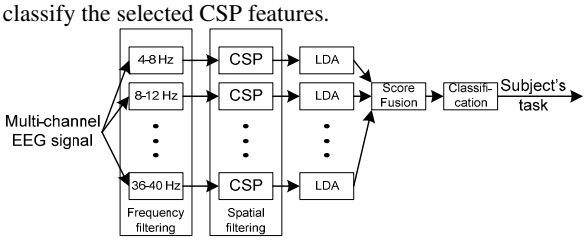

Fig. 2. Architecture of Sub-Band Common Spatial Pattern (SBCSP) [6]

Fig. 2 shows the architecture of SBCSP [6]. Comparing Fig. 1 against Fig. 2, the first, second, and fourth stages of FBCSP are similar to SBCSP [6]. However, SBCSP employs a Gabor fourier-based filterbank whereas FBCSP employs a zero-phase Chebyshev Type II Infinite Impulse Response (IIR) filterbank. The use of zero-phase filtering in the FBCSP overcomes the non-linear phase shift caused by the IIR filter. Furthermore, SBCSP computes a sub-band score for each spatial filter, followed by recursive band elimination or a classification algorithm to fuse the sub-band score. SBCSP then employs another classification algorithm to model and classify the fused sub-band score. In contrast, the third stage of FBCSP in this paper employs a feature selection algorithm to select discriminating CSP features; the fourth stage employs a classification algorithm to model and classify the selected CSP features. Hence, FBCSP is more generalized because any feature selection and classification algorithms from the machine learning and computational intelligence literature can be employed. In addition, SBCSP deploys all the spatial filters whereas FBCSP deploys only those effective spatial filters whose pairs of CSP features are selected. Hence, FBCSP employs only a small subset of effective spatial filters instead. This reduces the computational complexity as compared against using the entire set of spatial filters.

The next section provides a brief description of the CSP algorithm, followed by sections IV and V that provide a brief description of the feature selection and classification algorithms.

III. COMMON SPATIAL PATTERN ALGORITHM

The neurophysiological background of motor-imagery based BCIs is that motor activity, both actual and imagined [9],[10], causes an attenuation or increase of localized neural rhythmic activity called Event-Related Desynchronization or Event-Related Synchronization (ERS) respectively [11]. The *Common Spatial Pattern* (CSP) algorithm is highly successful in calculating spatial filters for detecting ERD and ERS [4]. The objective of spatial filtering employing the CSP algorithm [12] in BCI is to compute features whose variances are optimal for discriminating two classes of EEG measurements [12],[13].

The method employed by the CSP algorithm is based on the simultaneous diagonalization of two covariance matrices [12],[14]. In summary, the spatially filtered signal $\mathbf{Z}$ of a single trial EEG $\mathbf{E}$ is given as

$$\mathbf{Z} = \mathbf{W}\mathbf{E} . \quad (1)$$

where $\mathbf{E}$ is an $N \times T$ matrix representing the raw EEG measurement data of a single trial; N is the number of channels; T is the number of measurement samples per channel. $\mathbf{W}$ is the CSP projection matrix. The rows of $\mathbf{W}$ are the stationary spatial filters and the columns of $\mathbf{W}^{-1}$ are the common spatial patterns.

The spatial filtered signal $\mathbf{Z}$ given in (1) maximizes the differences in the variance of the two classes of EEG measurements. However, the variances of only a small number m of the spatial filtered signal are generally used as features for classification [12]. The m first and last rows of $\mathbf{Z}$ i.e. $\mathbf{Z}_p$, $p \in \{1..2m\}$ form the feature vector X_p given in (2) as inputs to a classifier.

$$X_p = \log\left(\mathrm{var}\left(\mathbf{Z}_p\right) \Big/ \sum_{i=1}^{2m} \mathrm{var}\left(\mathbf{Z}_p\right) \right) \quad (2)$$

From the concept of cortical homunculus [15], different areas of the cerebral cortex control movements of different parts of the body [8]. In addition, the distinct areas of the cerebal cortex control movements on the contralateral side of the body [16]. For example, the region that controls the feet is at the center of the vertex, the region that controls the left hand is on the right hemisphere, and the region that controls the right hand is on the left hemisphere [8]. Hence, the spatial patterns of a motor action are verifiable with the specific region that controls the motor action [4].

IV. FEATURE SELECTION ALGORITHM

Feature selection in pattern classification is defined as: given a set of *d* features, select a subset of size *k* that leads to the smallest classification errors. There are mainly two feature selection approaches in the literature [17]: the *wrapper* approach where features are selected using the classifier; and the *filter* approach where the features are selected independent from the classifier. Although the wrapper approach may yield better performance, increased computational effort is often required [17].

In Mutual Information (MI)-based feature selection, the problem is defined as, given an initial set *F* with *d* features, find the subset $S \subset F$ with *k* features that maximizes Mutual Information $I(S;\Omega)$ [18]. The MI between the two random variables is [19]

$$I(\mathbf{X};\mathbf{Y}) = H(\mathbf{Y}) - H(\mathbf{Y} \mid \mathbf{X}). \quad (3)$$

where the entropy is of a *d*-dimensional random variable $\mathbf{X}$ $\mathbf{X} = \{X_1, X_2, \ldots X_d\}$ is

$$H(\mathbf{X}) = -\sum_{x \in \mathbf{X}} p(x) \log_2 p(x); \quad (4)$$

the conditional entropy of random variables $\mathbf{X}$ and $\mathbf{Y}$ is

$$H(\mathbf{Y} \mid \mathbf{X}) = -\sum_{x \in \mathbf{X}} \sum_{y \in \mathbf{Y}} p(x,y) \log_2 p(y \mid x); \quad (5)$$

and $p(\cdot)$ are probability functions.

In pattern classification problems, the input features are usually continuous variables and the class has discrete values. Thus the MI between input features X and the class Ω is

$$I(\mathbf{X};\Omega) = H(\Omega) - H(\Omega \mid \mathbf{X}), \quad (6)$$

where $\omega \in \Omega = \{1,\ldots,N_\omega\}$; and the conditional entropy is

$$H(\Omega \mid \mathbf{X}) = -\int_{\mathbf{X}} \sum_{\omega=1}^{N_\omega} p(\omega \mid x) \log_2 p(\omega \mid x) dx, \quad (7)$$

where N_ω is the number of classes.

In the wrapper approach, the conditional entropy in (7) is simply

$$H(\Omega \mid \mathbf{X}) = -\sum_{\omega=1}^{N_\omega} p(\omega \mid \mathbf{X}) \log_2 p(\omega \mid \mathbf{X}), \quad (8)$$

where $p(\omega|\mathrm{X})$ is easy to estimate from the number of data sample classified as class ω using the classifier over the total number of data samples. In the filter approach, this conditional entropy is relatively harder to compute since it is not easy to estimate $p(\omega|\mathrm{X})$. The method of using Parzen Window to estimate $p(\omega|\mathrm{X})$ is briefly reviewed in section V.

The following feature selection algorithms are used in this paper:

A. MIBIF algorithm

The *Mutual Information based Best Individual Feature* (MIBIF) algorithm that is based on the filter approach is described as follows:

- Step 1: Initialization
  Initialize set of *d* features $F = \{f_1, f_2, \ldots f_d\}$, set of selected features $S = \varnothing$.
- Step 2: Compute the MI of features with the output class
  Compute $I(f_i;\Omega)\ \forall i = 1..d, f_i \in F$.
- Step 3: Select the best *k* features
  Repeat
  Select the feature f_i that maximizes $I(f_i;\Omega)$ using
  $$F = F \setminus \{f_i\}, S = \{f_i\} \mid I(f_i;\Omega) = \max_{j=1..d, f_j \in F} I(f_j;\Omega). \quad (9)$$
  Until $|S| = k$

The MIBIF feature selection algorithm requires a user-defined parameter *k* to select the *k* best features.

B. MINBPW algorithm

The *Mutual Information-based Naïve Bayesian Parzen Window* (MINBPW) [20] algorithm that is based on the wrapper approach for the *Naïve Baysian Parzen Window* (NBPW) classifier [20], is described as follows:

- Step 1: Initialization
  Initialize set of *d* features $F = \{f_1, f_2, \ldots f_d\}$, set of selected features $S = \varnothing$.
- Step 2: Compute the MI of features with the output class
  Compute $I(f_i;\Omega)\ \forall i = 1..d, f_i \in F$.
- Step 3: Select the first feature
  Select the feature f_i that maximizes $I(f_i;\Omega)$ using
  $$F = F \setminus \{f_i\}, S = \{f_i\} \mid I(f_i;\Omega) = \max_{j=1..d, f_j \in F} I(f_j;\Omega). \quad (10)$$
  where the MI is computed using (6) and the conditional entropy is estimated using (8).
- Step 4: Greedy selection
  Repeat
  a) Compute $I(f_i \cup S;\Omega)\ \forall i = 1..d, f_i \in F$, the joint MI between the feature *i* and selected features with the output class where the conditional entropy is estimated using (8).
  b) Select next feature using
  $$F = F \setminus \{f_i\}, S = \{f_i\} \mid I(f_i \cup S;\Omega) = \max_{j=1..d, f_j \in F} I(f_j \cup S;\Omega). \quad (11)$$
  Until $(|S| = k) \vee (I(f_i \cup S;\Omega) - I(S;\Omega) < \delta)$

The MINBPW algorithm also requires a user-defined parameter *k* or a small δ to stop the selection of features. The

parameter δ is set to 0.1 in this paper.

C. MIFS algorithm

The Sequential Forward Selection method or greedy selection scheme is adopted by Battiti in the *Mutual Information based feature selection* (MIFS) algorithm [18]. The MIFS algorithm is based on the filter approach [18]. The MIFS algorithm requires a user defined parameter k to set the maximum number of features to select, or a small δ to stop the selection of features once there are no further increase in MI beyond δ. The user-defined value of 0.1 is used for δ in this paper

D. FRFS2 algorithm

In addition to the prevailing application of MI in feature selection, Rough set theory (RST) [21] is also potentially feasible in feature selection and significantly reduce the pattern dimensionality [22]. However, RST is only capable of handling discretized attribute values. The Fuzzy-Rough set-based Feature Selection (FRFS) is a new approach that extends data reduction in RST theory for crisp and real-value attributes [23]. The new Fuzzy-Rough set-based Feature Selection (FRFS2) algorithm [23],[24] that is based on the filter approach is used in this paper.

The FRFS2 algorithm does not require any user-defined parameters.

E. MIRSR algorithm

The *Mutual Information-based Rough Set Reduction* (MIRSR) that is based on the wrapper approach for the *Rough set-based Neuro-Fuzzy System* (RNFS), employs the MI to select attributes with high relevance and the concept of knowledge reduction in rough set theory to select attributes with low redundancy [20].

V. CLASSIFICATION ALGORITHMS

A classifier is one that estimates the class label $\omega \in \Omega$ from the trained model given a data sample $\mathbf{X}=\{X_1,X_2,\ldots,X_d\}$ with d features where Y is the true class label. The trained model is constructed from training data that comprises n samples $\{\mathbf{X}_1,\mathbf{X}_2,\ldots\mathbf{X}_n\}$ with respective class labels $\{Y_1,Y_2,\ldots Y_n\}$. The following classifiers are used in this paper:

A. NBPW algorithm

The *Naïve Bayesian Parzen Window* (NBPW) classifier estimates $p(\mathbf{X}|\omega)$ and $P(\omega)$ from training data samples and predicts the class ω with the highest posterior probability $p(\omega|\mathbf{X})$ using Bayes rule

$$p(\omega \mid \mathbf{X}) = \frac{p(\mathbf{X} \mid \omega) P(\omega)}{p(\mathbf{X})}, \tag{12}$$

where $p(\omega|\mathbf{X})$ is the conditional probability of class ω given the data sample $\mathbf{X}$; $p(\mathbf{X}|\omega)$ is the conditional probability of $\mathbf{X}$ given class ω; $P(\omega)$ is the prior probability of class ω; and $p(\mathbf{X})$ is

$$p(\mathbf{X}) = \sum_{\omega \in \Omega} p(\mathbf{X} \mid \omega) P(\omega). \tag{13}$$

The computation of $p(\omega|\mathbf{X})$ is rendered feasible by a naïve assumption that all the features $X_1,X_2,\ldots,X_d$ are conditionally independent given class ω in

$$p(\mathbf{X} \mid \omega) = \prod_{i=1}^{d} p(X_i \mid \omega). \tag{14}$$

The NBPW classifier employs Parzen Window to estimate the conditional probability $p(X_i|\omega)$ in

$$\hat{p}(X_i \mid \omega) = \frac{1}{n_\omega} \sum_{j \in I_\omega} \phi(X_i - X_{i,j}, h), \tag{15}$$

where $\omega=1,\ldots,N_\omega$; n_ω is the number of data samples belonging to class ω; I_ω is the set of indices of the data samples belonging to class ω; and ϕ is a smoothing kernel function with a smoothing parameter h. The NBPW classifier employs the univariate Gaussian kernel given by

$$\phi(x,h) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x^2}{2h^2}\right)}, \tag{16}$$

and *normal optimal smoothing strategy* [25] given by

$$h^{opt} = \left(\frac{4}{3n}\right)^{1/5} \sigma, \tag{17}$$

where σ denotes the standard deviation of the distribution.

The classification rule of the NBPW classifier is given by

$$\omega = \arg\max_{\omega \in \Omega} p(\omega \mid \mathbf{X}). \tag{18}$$

B. FLD algorithm

The Linear Discriminant classification rule is given by

$$\omega \begin{cases} \in k & \mathbf{W}'\mathbf{X} \geq b \\ \notin k & \mathbf{W}'\mathbf{X} < b \end{cases} \tag{19}$$

where class $k \in \Omega$ is discriminated against the rest; $\mathbf{W}$ is an adjustable weight vector or projection vector for class k; $'$ denotes transpose operator; and b is a bias. This classification rule maps multi-dimensional data to one-dimensional using a linear function.

The *Fisher Linear Discriminant* (FLD) [26] is a linear discriminant that maximizes the ratio of between-class scatter to within-class scatter given by

$$J(\mathbf{W}) = \frac{\mathbf{W}'\mathbf{S_B}\mathbf{W}}{\mathbf{W}'\mathbf{S_W}\mathbf{W}} \tag{20}$$

where $\mathbf{S_B}$ is the between class scatter matrix; and $\mathbf{S_W}$ is the within class scatter matrix.

C. SVM algorithm

The Support Vector Machine (SVM) [27] is a linear discriminant that maximizes the separation between two classes based on the assumption that it improves the classifier's generalization capability. This is achieved by minimizing the cost function

$$J(\mathbf{W}) = \frac{1}{2}\|\mathbf{W}\|^2, \tag{21}$$

subject to the constraint

$$Y_i(\mathbf{W}' \cdot \mathbf{X}_i - b) \geq 1 \; \forall i = 1..n, \tag{22}$$

where $\mathbf{X}_1,\mathbf{X}_2,\ldots\mathbf{X}_n$ are the training data, and b is a bias.

The SVM employs the linear discriminant classification rule given in equation (19). The SVM implementation in the Matlab Bioinformatics toolbox is used in this paper.

D. 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 [26]. The *Classification And Regression Tree* (CART) [28] implementation in the Matlab Statistics toolbox is used in this paper.

E. k-NN algorithm

The k-nearest neighbor (k-NN) [29] 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=3 is used in this paper.

F. RNFS and DENFIS algorithms

Neuro-Fuzzy Systems are hybrid intelligent systems that synergize the human-like reasoning style of fuzzy systems with the learning and connectionist structure of neural networks. Neuro-Fuzzy Systems can be classified into linguistic or precise model [30]. The linguistic Neuro-Fuzzy System used in this paper is the *Rough set-based Neuro-Fuzzy System* (RNFS) [20]. RNFS is a novel hybrid intelligent system that synergizes the concept of knowledge reduction in rough set theory with neuro-fuzzy systems [20]. The precise NFS used in this paper is the *Dynamic Evolving Neural-Fuzzy Inference System* (DENFIS) [31].

VI. EXPERIMENTAL RESULTS

The "No Free Lunch" theorem states that there is no general superiority of any approach over the others in pattern classification; and if one approach seems to outperform another in a particular situation, it is a consequence of its fit to the particular pattern recognition problem [26]. Therefore, this section assesses the performance of combining the various feature selection algorithms described in section IV and classification algorithms described in section V for use with the FBCSP in motor imagery-based BCI. Filter-based feature selection algorithms, namely, MIBIF, MIFS and FRFS2, work with any classification algorithms. However, wrapper-based feature selection algorithms, namely, MINBPW and MIRSR work only with the RNFS and NBPW classification algorithms respectively. Hence, MINBPW and MIRSR are not combined with other classification algorithms. The proposed FBCSP approach is applied to the publicly available BCI competition III dataset IVa [32],[33] and data collected from healthy subjects and unilaterally paralyzed stroke patients. The FBCSP employs a filter bank that covers 4-40Hz, which comprises 9 bandpass filters that covers 4Hz each; and the CSP algorithm with m=2.

A. Publicly available BCI Competition III dataset IVa

The BCI Competition III dataset IVa [33] is collected from 5 subjects (labeled 'aa', 'al', 'av', 'aw', 'ay') who performed right hand and right foot imagination [32]. The data for each subject comprises 280 trials of EEG measurements from 118 electrodes. The data is extracted from selected electrodes, starting from 0.5s to 2.5s after the visual cue. The time segment and electrode selections are consistent with the experiment performed in [6].

Fig. 3 presents the experimental results of unbiased 10×10-fold cross-validations performed using FBCSP with various feature selection and classification algorithms described in section IV and V respectively. The results in Fig. 3 shows that among the classification algorithms, NBPW, FLD and SVM yield superior test accuracies compared against the CART, k-NN, RNFS and DENFIS. The results also show that among the feature selection algorithms used with NBPW, the MIBIF that selects 4 pairs of CSP features yields superior test accuracy. Specifically, the FBCSP with MIBIF4 and NBPW yields a test accuracy of 90.3±0.7%; whereas FLD yields 89.9±0.9%, and SVM yields 90.0±0.8%.

Fig. 4 shows the results of unbiased 10×10-fold cross-validations performed using the FBCSP with the NBPW classification algorithm and the MINBPW wrapper-based feature selection algorithm (labeled as FBCSP_w), as well as the MIBIF filter-based feature selection algorithm that selects 4 pairs of CSP features (labeled as FBCSP_f). The FBCSP_w and FBCSP_f have been shown to yield superior test accuracies for wrapper and filter-based approaches respectively compared to the rest in Fig. 3. Hence, only FBCSP_w and FBCSP_f are presented in this experiment. The SBCSP has been shown to yield superior results on this dataset compared against existing approaches such as CSSP and CSSSP in [6]. Therefore, this experiment compares only with CSP and SBCSP. In the experiment, both FBCSP and SBCSP employed 9 bandpass filters and used the same NBPW classifier. This similar configuration is used to avoid ad-hoc tuning of the classifiers in order to make a fair comparison between FBCSP and SBCSP. CSP is manually configured with a broad bandpass filter of 8-30Hz [12]. The results in Fig. 4 show that the FBCSP_f yields superior averaged test accuracy of 90.3±0.7%, whereas FBCSP_w yields 89.2±1.1%, SBCSP yields 86.3±1.1%, and CSP yields 86.6±0.7%. Hence, the results show that both the FBCSP_f and FBCSP_w yield statistically superior results than SBCSP and CSP.

Fig. 5 shows the most significant CSP using the frequency range that is autonomously selected in FBCSP_f. The results show that the centre of the vertex and the left hemisphere discriminates the right foot action and the right hand action respectively for subjects 'aa', 'al', 'aw', and 'ay'. These results verify the neurophysiological plausibility of the CSP projection matrix computed for these subjects. However, the result for subject 'av' does not show such patterns. This is a plausible reason for the relatively inferior test accuracy obtained for subject 'av'.

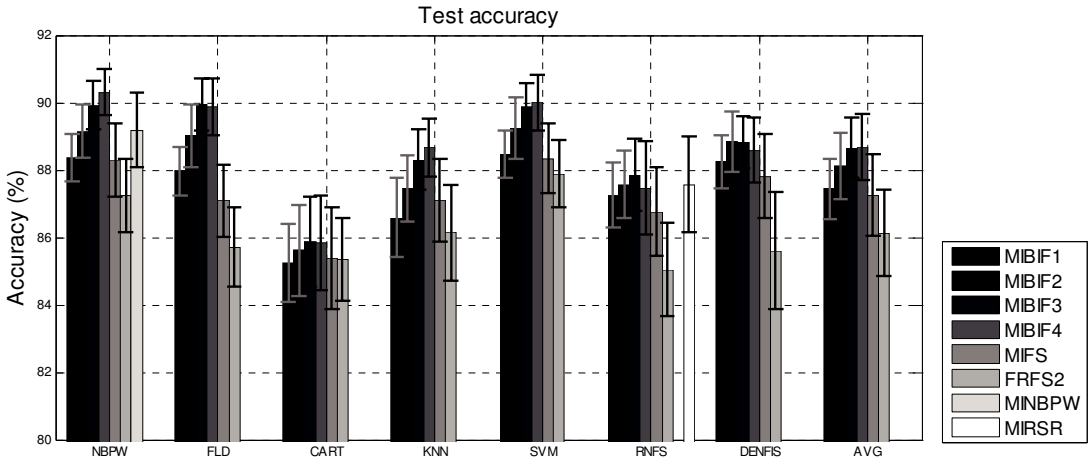

Fig. 3. Experimental results on the test accuracies of 10×10-fold cross-validations performed using FBCSP with various feature selection and classification algorithms on publicly available BCI Competition III dataset IVa

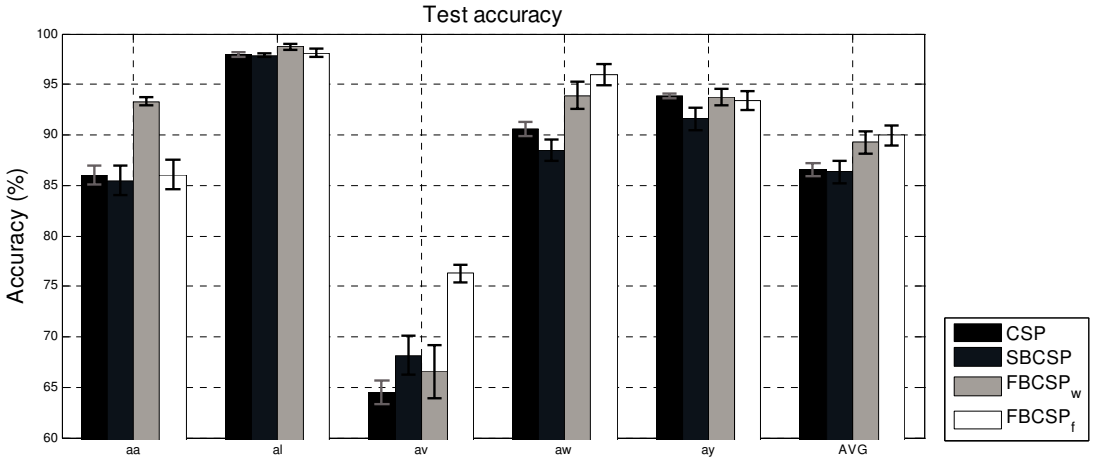

Fig. 4. Experimental results on the test accuracies of 10×10-fold cross-validations performed using CSP, SBCSP, FBCSP$_w$ and FBCSP$_f$ on BCI Competition dataset IVa

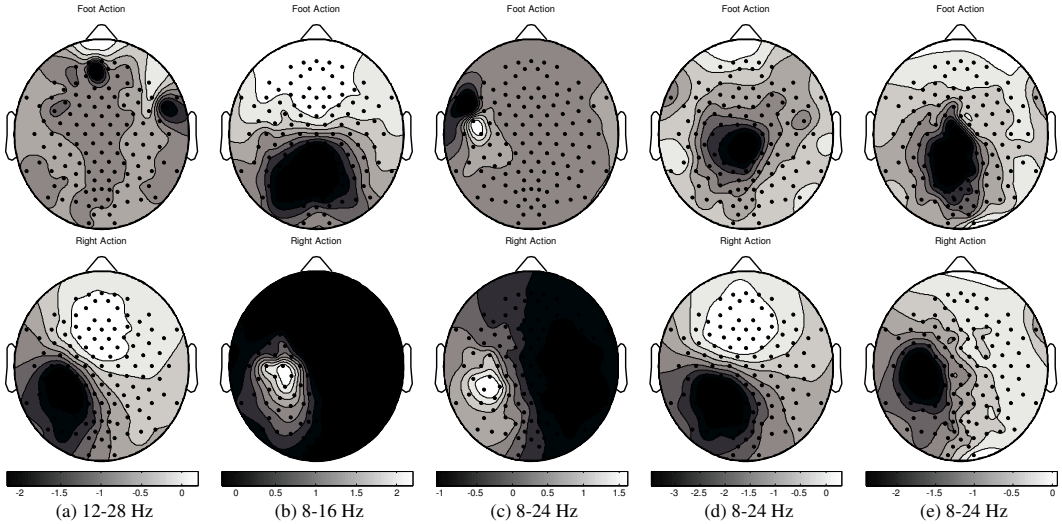

Fig. 5. Spatial Patterns of using FBCSP$_f$ on BCI Competition dataset IVa for subjects (a) 'aa', (b) 'al', (c) 'av', (d) 'aw', and (e) 'ay' respectively

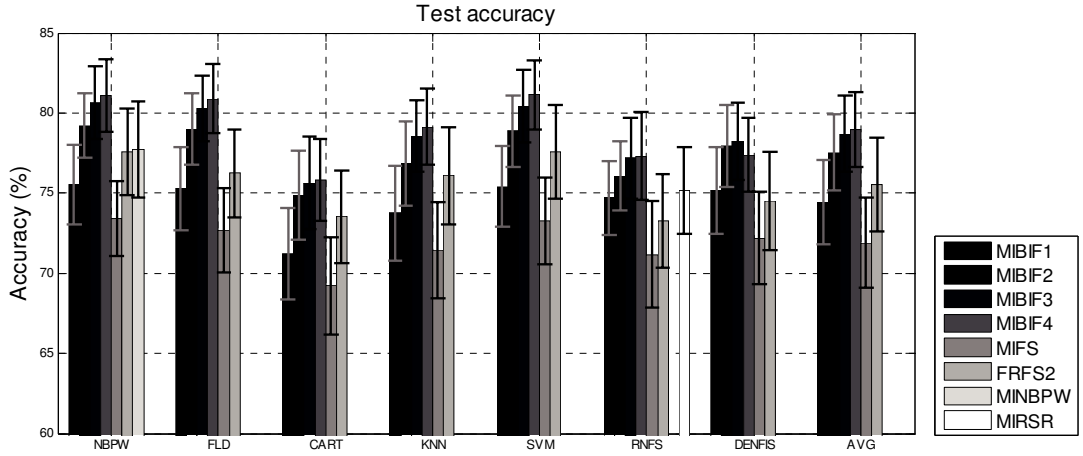

Fig. 6. Experimental results on the (a) training and (b) test accuracies of 10×10-fold cross-validations performed using FBCSP with various feature selection and classification algorithms on data collected from healthy subjects and unilaterally paralyzed stroke patients

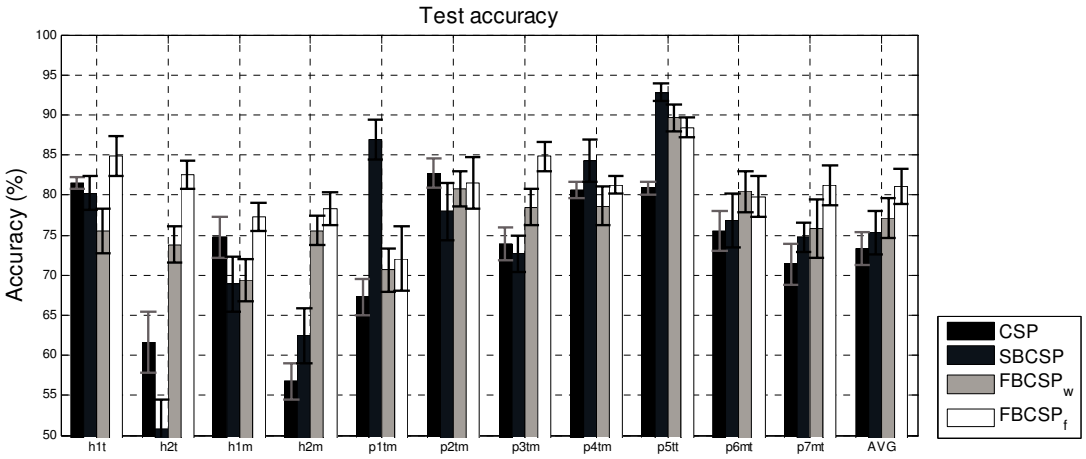

Fig. 7. Experimental results on the test accuracies of 10×10-fold cross-validations performed using CSP, SBCSP, FBCSP$_w$ and FBCSP$_f$ on data collected from healthy subjects and unilaterally paralyzed stroke patients

B. Finger tapping and motor-imagery data collected from healthy subjects and unilaterally paralyzed stroke patients

This dataset is collected using Neuroscan NuAmps from 2 healthy subjects (labeled 'h1~', 'h2~') and 7 unilaterally paralyzed stroke patients (labeled 'p1~', 'p2~'). The data is collected with approval from the Ethics Approval Board. The healthy subjects performed left and right hand finger tapping (labeled '~tt) and motor imagery (labeled '~mm'). The unilaterally paralyzed stroke patients performed motor imagination of their disabled arm and finger tapping using their unaffected arm. The data for each subject comprises 160 trials of EEG measurements from 21 electrodes (F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, C3, Cz, C4, TP7, CP3, CPz, CP4, TP8, P3, Pz, P4) starting from 0.5s to 2.5s after the visual cue.

Fig. 6 presents the experimental results of unbiased 10×10-fold cross-validations performed using FBCSP with various feature selection and classification algorithms described in section IV and V respectively. The results in Fig. 6 again shows that NBPW, FLD and SVM yield superior test accuracies than CART, k-NN, RNFS and DENFIS. The results also show that the FBCSP with MIBIF4 and NBPW as well as SVM yield a superior test accuracy of 81.1±2.2%, whereas FLD yields 80.9±2.1%.

Fig. 7 shows the results of unbiased 10×10-fold cross-validations performed using the FBCSP with NBPW and the wrapper-based MINBPW (labeled as FBCSP$_w$), as well as the filter-based MIBIF4 (labeled as FBCSP$_f$). The results are again compared against CSP and SBCSP. In the experiment, both FBCSP and SBCSP employed 9 bandpass filter bands and used the same NBPW classifier. However, CSP is manually configured with subject-specific operational frequency ranges. The results in Fig. 7 shows that the FBCSP$_f$ yields superior averaged test accuracy of 81.1±2.2%, whereas FBCSP$_w$ yields 77.7±3.0%, SBCSP yields 75.3±2.7%, and CSP yields 73.3±2.0%. Hence, the results show that both the FBCSP$_f$ and FBCSP$_w$ yield statistically superior results than SBCSP and CSP.

VII. CONCLUSIONS

This paper proposed a novel machine learning approach called Filter Bank Common Spatial Pattern (FBCSP) for processing EEG measurements in motor imagery-based BCI. FBCSP addresses the problem of selecting an appropriate operational frequency band for extracting discriminating CSP features. FBCSP employs a feature selection algorithm to select discriminative CSP features from a bank of multiple bandpass filters and spatial filters, and a classification algorithm to classify the selected features. FBCSP is capable of learning subject-specific patterns from the high-dimensional EEG measurements without operator intervention as well as yielding relatively high classification accuracies. Since any feature selection and classification algorithms from the machine learning and computational intelligence literature can be employed, FBCSP is more generalized than existing approaches such as the Sub-band Common Spatial Pattern (SBCSP).

Experimental results showed that the proposed FBCSP yields superior classification accuracy compared against SBCSP and CSP with manually selected operational frequency bands. Based on the results, the Mutual Information Best Individual Feature (MIBIF) selection algorithm that selects 4 pairs of CSP features; and the Naïve Bayes Parzen Window (NBPW), Fisher Linear Discriminant (FLD) or Support Vector Machines (SVM) classification algorithms are recommended for use with FBCSP in motor imagery-based BCIs..

ACKNOWLEDGMENT

The authors would like to thank Beng Ti Ang from National Neuroscience Institute, Karen Chua and Christopher Kuah from Tan Tock Seng Hospital, Wang Chuanchu and Phua Kok Soon from Institute for Infocomm Research A*STAR for their support in the data collection from the stroke patients.

REFERENCES

[1] B. Rebsamen, E. Burdet, C. Guan, H. Zhang, C. L. Teo, Q. Zeng, C. Laugier, and M. H. Ang Jr., "Controlling a Wheelchair Indoors Using Thought," *IEEE Intelligent Systems*, vol. 22, no. 2, pp. 18-24, 2007.

[2] N. Birbaumer, "Brain-computer-interface research: Coming of age," *Clin. Neurophysiol.*, vol. 117, no. 3, pp. 479-483, 2006.

[3] 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.

[4] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Muller, and G. Curio, "The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539-550, 2007.

[5] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123-1134, 2001.

[6] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band Common Spatial Pattern (SBCSP) for Brain-Computer Interface," *3rd International IEEE/EMBS Conference on Neural Engineering, 2007. CNE '07,* pp. 204-207, 2007.

[7] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Muller, "Combined Optimization of Spatial and Temporal Filters for Improving Brain-Computer Interfacing," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 11, pp. 2274-2281, 2006.

[8] S. Lemm, B. Blankertz, G. Curio, and K.-R. Muller, "Spatio-Spectral Filters for Improving the Classification of Single Trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 9, pp. 1541-1548, 2005.

[9] G. Pfurtscheller and A. Aranibar, "Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement," *Electroencephalography and Clinical Neurophysiology*, vol. 46, no. 2, pp. 138-146, 1979.

[10] A. Schnitzler, S. Salenius, R. Salmelin, V. Jousmaki, and R. Hari, "Involvement of Primary Motor Cortex in Motor Imagery: A Neuromagnetic Study," *NeuroImage*, vol. 6, no. 3, pp. 201-208, 1997.

[11] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842-1857, 1999.

[12] H. Ramoser, J. Muller-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, 2000.

[13] J. Muller-Gerking , G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clin. Neurophysiol.*, vol. 110, no. 5, pp. 787-798, 1999.

[14] K. Fukunaga. *Introduction to Statistical Pattern Recognition*, 2nd ed. New York:Academic Press, 1990.

[15] W. Penfield and T. Rasmussen. *The cerebral cortex of man: a clinical study of localization of function*. New York :Macmillan, 1950.

[16] E. R. Kandel, J. H. Schwartz, and T. M. Jessell. *Essentials of neural science and behavior*. Norwalk, CT:Appleton & Lange, 1995.

[17] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 273-324, 1997.

[18] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Trans. Neural Networks*, vol. 5, no. 4, pp. 537-550, 1994.

[19] T. M. Cover and J. A. Thomas. *Elements of Information Theory*, 2nd ed. New York:Wiley, 2006.

[20] K. K. Ang and C. Quek, "Rough Set-based Neuro-Fuzzy System," *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN '06),* pp. 742-749, 2006.

[21] Z. Pawlak. *Rough Sets: Theoretical Aspects of Reasoning about Data*. Dordrecht, Boston:Kluwer Academic Publishers, 1991.

[22] R. W. Swiniarski and A. Skowron, "Rough set methods in feature selection and recognition," *Patt. Rec. Lett.*, vol. 24, no. 6, pp. 833-849, 2003.

[23] R. Jensen and Q. Shen, "Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches," *IEEE Trans. Know. Data Eng.*, vol. 16, no. 12, pp. 1457-1471, 2004.

[24] R. Jensen and Q. Shen, "New Approaches to Fuzzy-Rough Feature Selection," *IEEE Trans. Fuzzy Systems*, pp. in press 2007.

[25] A. W. Bowman and A. Azzalini. *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*. New York:Oxford University Press, 1997.

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

[27] V. N. Vapnik. *Statistical learning theory*. New York:Wiley, 1998.

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

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

[30] J. Casillas, O. Cordón, F. Herrera, and L. Magdalena. *Interpretability Issues in Fuzzy Modeling* . Berlin:Springer-Verlag, 2003.

[31] N. K. Kasabov and Q. Song, "DENFIS: dynamic evolving neural-fuzzy inference system and its application for time-series prediction," *IEEE Trans. Fuzzy Systems*, vol. 10, no. 2, pp. 144-154, 2002.

[32] 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 Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993-1002, 2004.

[33] B. Blankertz, "BCI Competition III*",* Fraunhofer FIRST.IDA, http://ida.first.fraunhofer.de/projects/bci/competition_iii, 2005.