

Article

Classification of EEG Signals Using a Multiple Kernel Learning Support Vector Machine

Xiaoou Li ^{1,2}, Xun Chen ^{3,*}, Yuning Yan ⁴, Wenshi Wei ⁴ and Z. Jane Wang ⁵

¹ Shanghai Medical Instrumentation College, Shanghai 200093, China

² School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China; E-Mail: xiaouli@gmail.com

³ Department of Biomedical Engineering, Hefei University of Technology, Hefei 230009, China

⁴ Department of Neurology, Huadong Hospital Affiliated to Fudan University, Shanghai 200040, China; E-Mails: quack008@126.com (Y.Y.); wenshiwei@medmail.com.cn (W.W.)

⁵ Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; E-Mail: zjanew@ece.ubc.ca

* Author to whom correspondence should be addressed; E-Mail: xunchen@ece.ubc.ca; Tel.: +86-21-6548-5551; Fax: +86-21-5562-0106.

Received: 19 February 2014; in revised form: 20 June 2014 / Accepted: 10 July 2014 /

Published: 17 July 2014

Abstract: In this study, a multiple kernel learning support vector machine algorithm is proposed for the identification of EEG signals including mental and cognitive tasks, which is a key component in EEG-based brain computer interface (BCI) systems. The presented BCI approach included three stages: (1) a pre-processing step was performed to improve the general signal quality of the EEG; (2) the features were chosen, including wavelet packet entropy and Granger causality, respectively; (3) a multiple kernel learning support vector machine (MKL-SVM) based on a gradient descent optimization algorithm was investigated to classify EEG signals, in which the kernel was defined as a linear combination of polynomial kernels and radial basis function kernels. Experimental results showed that the proposed method provided better classification performance compared with the SVM based on a single kernel. For mental tasks, the average accuracies for 2-class, 3-class, 4-class, and 5-class classifications were 99.20%, 81.25%, 76.76%, and 75.25% respectively. Comparing stroke patients with healthy controls using the proposed algorithm, we achieved the average classification accuracies of 89.24% and 80.33% for 0-back and 1-back tasks respectively. Our results indicate that the proposed approach is

promising for implementing human-computer interaction (HCI), especially for mental task classification and identifying suitable brain impairment candidates.

Keywords: brain computer interface; mental task; stroke patients; multiple kernel learning; polynomial kernel; radial basis function kernel

1. Introduction

A brain computer interface (BCI) system is generally composed of a set of sensors and signal processing units, which can establish an information communication channel between a subject's brain and an external device. The realization of electroencephalography (EEG)-based BCI systems mainly involves three processes: first, the brain activity is recorded by means of electrodes located on the scalp and then a pre-processing step is applied to remove artifacts in order to enhance the signal-to-noise ratio; second, a feature extraction step is performed to extract meaningful information from raw EEG signals; the last step is conducted to translate such specific features into effective control commands and drive the external device. Since the implementation of BCI does not depend on peripheral nerves and muscles, it is particularly useful for individuals who suffer from the motor disorder with the cognitive ability [1–6].

EEG-based BCI technologies can be divided into many categories. In this study, we concentrate on the mental task recognition [7] and the P300 evoked related potential (ERP) recognition [8]. Keirn and Aunon [9] proposed that it was possible to accurately distinguish between various mental tasks using only the EEG. Many classification algorithms have been developed to improve mental task-based BCI performance due to the difficulty in obtaining high classification accuracy [10]. Garrett *et al.* [11] applied one linear (*i.e.*, linear discriminant analysis (LDA)) and two nonlinear (neural network (NN) and support vector machine (SVM)) classifiers to the classification of spontaneous EEG during five mental tasks. Palaniappan [10] used the spectral power differences in four bands and a NN classifier to classify different mental tasks. Li *et al.* [12] studied the classification of mental task EEG signals using SVM. Gupta *et al.* [13] used the relevant features with the SVM and LDA techniques to classify mental tasks.

Some studies have been also motivated by the goal of using EEG with effective algorithms to identify cognitive impairment patients. Lehmann *et al.* explored the ability of a multitude of linear and non-linear classification algorithms (*i.e.*, LDA, NN, SVM) to discriminate between EEG signals of patients with varying degrees of cognitive impairment [14]. Dauwels *et al.* used LDA and quadratic discriminant analysis (QDA) to classify cognitive impairment [15]. Gallego-Jutglà *et al.* used theta band power and LDA to achieve the best accuracy for diagnosing cognitive impairment [16].

In those EEG-based classification algorithms, although LDA has a very low computational requirement, it is not suitable for solving nonlinear problems [17]. NN can make the classification more flexible. However, the NN classifier requires a large number of training data and is also sensitive to over-fitting, especially for noisy and nonstationary data such as the EEG. This is because NN based on empirical risk minimization cannot control the learning mode well and requires more empirical

parameters [18]. SVM based on structural risk minimization yields good performance in many applications, especially for solving problems with high dimension, nonlinearity and small dataset [19].

In SVM, a kernel function is an essential element, which maps samples in one feature space to another feature space. However, it is often unclear what the most suitable kernel for a task at hand is, and hence the user may wish to combine several possible kernels. One problem with simply adding kernels is that using equal weights is possibly not optimal. For instance, if one kernel is not correlated with the classification problem at all, then assigning it a great weight will degrade the performance. Multiple kernel learning SVM (MKL-SVM) is an efficient way of optimizing kernel weights [20,21]. Compared with a single kernel SVM, MKL-SVM can enhance the interpretability of the decision function and improve the performance [22–24].

For a BCI system, a major component is efficient feature extraction from multi-channel EEG signals. In recent years, as useful tools for analyzing the complexity and rhythm information of EEG signals, several popular approaches for the extraction of quantitative EEG features have been introduced, such as wavelet packet entropy (WPE), causality analysis, *etc.* Since EEG signals are non-stationary (both time-varying and space-varying), these methods are excellent candidates for the feature extraction process [25,26].

In this study, WPE was extracted from the mental task EEG. Granger causality flow was extracted from the cognitive task EEG. Then these features and the MKL-SVM classifier were used to perform the recognition of different tasks. The mental tasks were baseline task, visual counting task, mental letter composing task, mathematical multiplication task, and geometric figure rotation task, respectively. The recognition of 2-class, 3-class, 4-class, and 5-class cases of mental task EEG signals were performed. The cognitive tasks based on working memory that may elicit a P300 ERP component were 0-back and 1-back tasks. The recognition of stroke patients and healthy controls was also performed.

2. Classification Algorithm

The BCI performance mainly depends on the features and the classification algorithm adopted. The recognition efficiency of the classification algorithm is a key issue. Several popular algorithms have been developed to serve certain scenarios, such as LDA, NN and SVM [10–16]. These classical algorithms will be briefly reviewed. Comparatively, SVM is a powerful approach for pattern recognition especially for high dimensional, nonlinear problems. SVM has achieved better classification results in mental task and cognitive task classifications. Recent developments on SVM have shown the need to consider multiple kernels [23]. This provides flexibility and reflects the fact that typical learning problems often involve multiple, heterogeneous data sources. For appropriately designed kernels, the optimized combination weights can be used to understand which features are important for a specific recognition problem. In this study, we applied the MKL-SVM algorithm to the classifications of mental task and cognitive task EEG signals.

2.1. Brief Review of Popular Classification Algorithms

The LDA classifier is a dimension reduction method which finds an optimal linear transformation that maximizes the class separability. It performs well when the feature vector is multivariate normally

distributed in each class group and different groups have a common covariance. However, these assumptions are rare in practice.

The NN classifier is a data driven self-adaptive method and provides a direct estimation of the posterior probabilities. It can be viewed as a mapping function, $F: R^d \rightarrow R^M$, where d -dimensional input is submitted to the network and an M -vector network output is obtained to make the classification decision. Although the NN classifier can detect complex nonlinear relationships, its disadvantages include higher computational cost and the empirical nature of the model, *etc.*

The SVM algorithm is based on the statistical learning theory and the Vapnik-Chervonenkis dimension [27]. It maps the input patterns into a higher dimension feature space through some nonlinear mapping where a linear decision surface is then constructed [28]. In SVM, a kernel function implicitly maps samples to a feature space given by a feature mapping [29]. Since there are different kernel types, it is often unclear what the most suitable kernel for a task at hand is, and hence the user may wish to combine several possible kernels. It is thus desirable to use an optimized kernel function that can fit the data structure well [21]. To optimally combine multiple kernels in SVM, the MKL-SVM approach is introduced as follows.

2.2. Multiple Kernel Learning SVM

We employed an algorithm called SimpleMKL that Rakotomamonjy *et al.* [30] proposed to solve the MKL-SVM problem. The kernel of SimpleMKL is defined as a linear combination of multiple kernels. The algorithm is essentially based on a gradient descent on the SVM objective value through a weighted 2-norm regularization formulation with an additional constraint on the weights. It can be applied to multiclass classification beyond binary classification, meanwhile it has rapid convergence and favourable efficiency compared with other algorithms [31–33].

SVM is developed from the optimal hyperplane in the case of linear separating. In MKL-SVM, any dataset (x_i, y_i) , where $x_i \in \mathbb{R}^n$, $y_i \in \{-1, 1\}$, can be separated by an optimal separating hyperplane: $\sum_m f_m(x) + b = 0$ with the maximum margin between two classes. The constrained optimization problem for MKL-SVM is described as [30]:

$$J(d) = \begin{cases} \min_d J(d) \sum_{m=1}^M d_m = 1, d_m \geq 0, \forall m \\ \min_{\{f_m\}, b, \xi, d} \frac{1}{2} \sum_m \frac{1}{d_m} \|f_m\|_{\mathcal{H}_m}^2 + C \sum_i \xi_i \quad \forall i \\ \text{s. t. } y_i \sum_m f_m(x_i) + y_i b \geq 1 - \xi_i \\ \xi_i \geq 0 \quad \forall i \end{cases} \quad (1)$$

where ξ_i is a loose variable. $\xi_i = 0$ represents a linear separable case, $\xi_i > 0$ represents a linear nonseparable case with certain misclassification. The penalty factor C controls the degree of misclassification. M is the total number of kernels used. d_m is the weight coefficient of the kernel function. Each basic kernel K_m corresponds to one f_m . Each d_m controls the squared norm of f_m in the objective function. The smaller d_m is, the smoother f_m will be. Equation (1) can be transformed into the following dual form using Lagrange multipliers:

$$J(d) = \begin{cases} \max_{\alpha} -\frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \sum_m d_m K_m(x_i, x_j) + \sum_i \alpha_i \forall i \\ \text{s. t. } \sum_i \alpha_i y_i = 0 \\ C \geq \alpha_i \geq 0 \forall i \end{cases} \quad (2)$$

This is a typical dual problem. Only a small number of α_i are nonzero, and they correspond to some data points that are support vectors.

For the nonlinear case, the original problem can be solved by mapping the original data space into a high dimension feature space \mathcal{H} with a mapping $\Phi: \mathbb{R}^n \rightarrow \mathcal{H}$. It is unnecessary to exactly know the mapping $\Phi(x)$ if we use the kernel function $K(x_i, x_j) = (\Phi(x_i) \cdot \Phi(x_j))$, which is a symmetric function and satisfies the Mercer condition [21]. The classical kernel functions include polynomial function (Poly) $K_{Poly}(x, y) = ((x \cdot y) + 1)^d$, radial basis function (RBF) $K_{RBF}(x, y) = \exp(-\|x - y\|^2 / 2\delta^2)$ and sigmoidal function (Sig) $K_{Sig}(x, y) = \tanh(\kappa x \cdot y + \nu)$, etc.

d_m is updated with the descent direction. The updating scheme is $d \leftarrow d + \gamma D$, where γ is the step size. This procedure is repeated until the objective value stops decreasing. The descent direction D is defined as:

$$D_m = \begin{cases} 0 & d_m = 0, \frac{\partial J}{\partial d_m} - \frac{\partial J}{\partial d_\mu} > 0 \\ -\frac{\partial J}{\partial d_m} + \frac{\partial J}{\partial d_\mu} & d_m > 0, m \neq \mu \\ \sum_{v \neq \mu, d_v > 0} \left(\frac{\partial J}{\partial d_v} - \frac{\partial J}{\partial d_\mu} \right) & m = \mu \end{cases} \quad (3)$$

where $\frac{\partial J}{\partial d_m} = -\frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K_m(x_i, x_j)$. d_μ is the maximal nonzero component of d . d_v is a component corresponding to the maximal admissible step size.

The algorithm is terminated at the following duality gap [30]:

$$\max_m \sum_{i,j} \alpha_i \alpha_j y_i y_j K_m(x_i, x_j) - \sum_{i,j} \alpha_i \alpha_j y_i y_j \sum_m d_m K_m(x_i, x_j) \leq \varepsilon \quad (4)$$

The optimal classification function is:

$$\begin{cases} f(x) = \text{sgn}\left(\sum_{sv} \alpha_i y_i K(x, x_i) + b\right) \\ K(x, x_i) = \sum_{m=1}^M d_m K_m(x, x_i) \end{cases} \quad (5)$$

3. Methods

The presented approach consisted of three main parts: (1) the purpose of the pre-processing step was to improve the general signal quality of the EEG in order to get more accurate rhythm analysis and measurements; (2) the features including WPE and Granger causality were extracted from the EEG signals to compose a feature vector for further classification; (3) MKL-SVM was employed to perform

the different task classification, and the classification accuracies were used to evaluate the performance of the proposed algorithm.

3.1. Human Subjects

3.1.1. EEG Data of Mental Task

The EEG data used in this study were collected by Keirn and Aunon [9]. Seven subjects, 21 to 48 years old, participated in the experiment. An elastic electrode cap was used to record EEG signals. The electrodes were placed on the scalp at C3, C4, P3, P4, O1, and O2 based on the international 10–20 system. They were referenced to two electrically linked mastoids, A1 and A2. The data were sampled at 250 Hz. Signals were recorded for 10 s during each task. Each task was repeated for two sessions. Each session contained five trials. The data are available online at <http://www.cs.colostate.edu/~anderson>. Five different mental tasks were involved [10], namely:

Baseline Task (denoted by task B): The subjects were told to relax and try to think of nothing.

Visual Counting Task (denoted by task C): The subjects were told to imagine a blackboard and visualise numbers being written on the board sequentially, with the previous number being erased before the next number was written.

Mental Letter Composing Task (denoted by task L): The subjects were told to mentally compose a letter to a friend or a relative without vocalising.

Mathematical Multiplication Task (denoted by task M): The subjects were given nontrivial multiplication problems, e.g., 24 times 14 and were told to solve them without vocalising or making any other physical movements.

Geometric Figure Rotation Task (denoted by task R): The subjects were given 30 s to study a particular three-dimensional block object, after which the drawing was removed and the subjects were told to visualise the object being rotated about an axis.

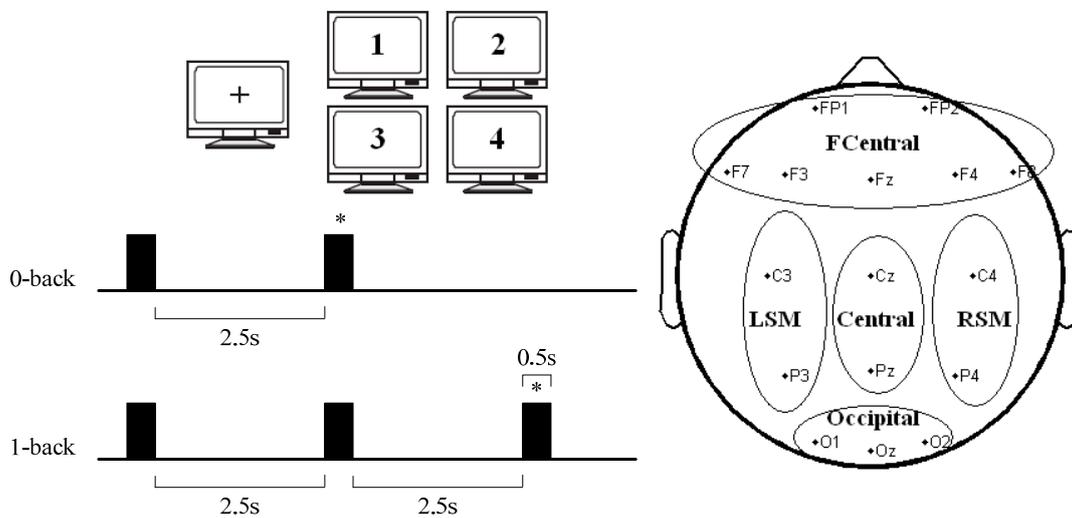
3.1.2. EEG Data of Cognitive Task

Consecutive patients aged 50 years or older with a first-ever acute ischemic stroke at Huadong Hospital Affiliated to Fudan University were recruited. All patients underwent cognitive testing, and those who met the criteria for mild cognitive impairment were included ($n = 13$). 13 age- and sex-matched healthy controls were enrolled in this cross-sectional study. All subjects were right handed and had normal vision. This study was approved by Huadong Hospital Affiliated to Fudan University Ethics Board, and all subjects gave written, informed consents before participation.

As shown in Figure 1, the working memory was assessed using a verbal N-back task [34]. A pseudorandom set of 4-digit numbers was displayed on a monitor, and the subjects were instructed to determine whether a specific digit—one appeared on the screen (0-back task); or the currently displayed number at any given time had been already displayed in the previous presentation (1-back task). Stimuli consisted in a 0.5 s. Inter stimulus interval (ISI) was 2.5 s in all conditions. Subjects had to distinguish between targets and non-targets by pressing a keyboard. Continuous ERP signals were

acquired using an EEG/ERP amplifier system (NATION[®] Inc., Shanghai, China). For all ERP recordings, 18 electrodes were placed according to the 10–20 international system. The chosen electrode positions were EOG1, EOG2, Fp1, Fp2, F3, F4, F7, F8, Fz, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2 (Figure 1). The data were sampled at 256 Hz. Signals were recorded for 120 s during each task. Each task was repeated for three sessions. Each session contained 40 trials with a 1:1 target/non-target relation. Namely, the total number of targets was 60, the same as that of non-targets.

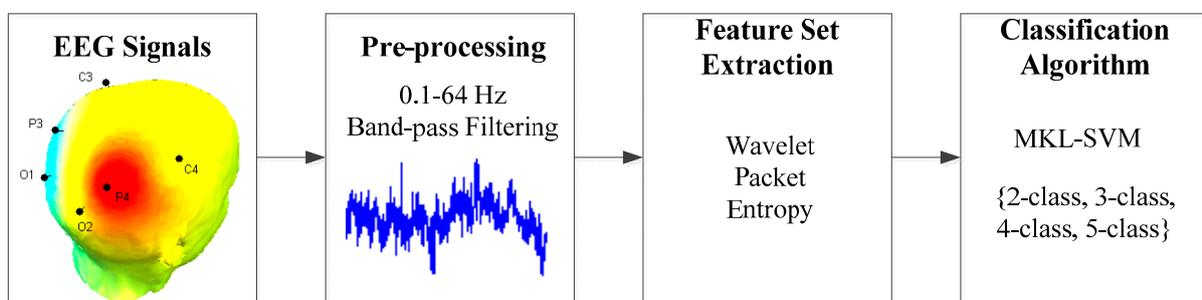
Figure 1. N-back task timeline and electrode positions. Here trial time sequences for 0-back and 1-back conditions. Black squares represent each stimulus in the task. The symbol * stands for the target number during each trial. Five brain regions: Fronto-Central (FCentral)-FP1, FP2, F7, F3, Fz, F4 and F8, Left Sensorimotor (LSM)-C3 and P3, Central- Cz and Pz, Right Sensorimotor (RSM) - C4 and P4, and Occipital-O1 and O2.



3.2. Recognition of Mental Task EEG

We proposed a mental task-based BCI approach, as illustrated in Figure 2.

Figure 2. The basic diagram of mental task recognition.



WPE can characterize the physiological state changes of the subjects during different mental tasks [12]. It provides an effective way for studying the complexity of EEG signals that can decode specific function states from the brain activity, and the obtained model from the time sequence and the estimated parameters can reveal the mechanism of EEG.

Brain electrical activity mapping (BEAM) summarizes the EEG data as color maps. Li *et al.* [12] studied the BEAM at 2.9 Hz, 6.1 Hz, 10 Hz, and 22 Hz of five different mental states and found that different mental tasks had different energy distributions in the brain area. They correspond to specific frequency bands in the EEG, such as 0–4 Hz (delta), 4–8 Hz (theta), 8–12 Hz (alpha), and 12–32 Hz (beta). The wavelet packet transform (WPT) can be viewed as a generalization of the classical wavelet transform, which provides a multi-resolution and time-frequency analysis for nonstationary EEG signals with a binary tree structure. The entropy provides a measurement of the signal uncertainty.

As mentioned above, we extracted four frequency band data (*i.e.*, delta, theta, alpha, and beta) of mental task EEG signals through the WPT decomposition, then calculated their entropies to analyze locally nonlinear features that can reflect the complexity distribution of the signal in each frequency band. A wavelet packet is represented as a function [25]:

$$\psi_{j,k}^i(t) = 2^{-\frac{j}{2}}\psi^i(2^{-j}t - k), i = 1, 2, \dots, j^n \quad (6)$$

where i is the modulation parameter, j is the dilation parameter, k is the translation parameter, n is the level of the decomposition in the wavelet packet tree, and ψ^i is called as a mother wavelet.

The wavelet packet coefficients $c_{j,k}^i$ corresponding to the signal $f(t)$ can be obtained as:

$$c_{j,k}^i(t) = \int_{-\infty}^{\infty} f(t) \psi_{j,k}^i(t) dt \quad (7)$$

The wavelet packet component of the signal at a particular node can be obtained as:

$$f_j^i(t) = \sum_{k=-\infty}^{\infty} c_{j,k}^i \psi_{j,k}^i(t) dt \quad (8)$$

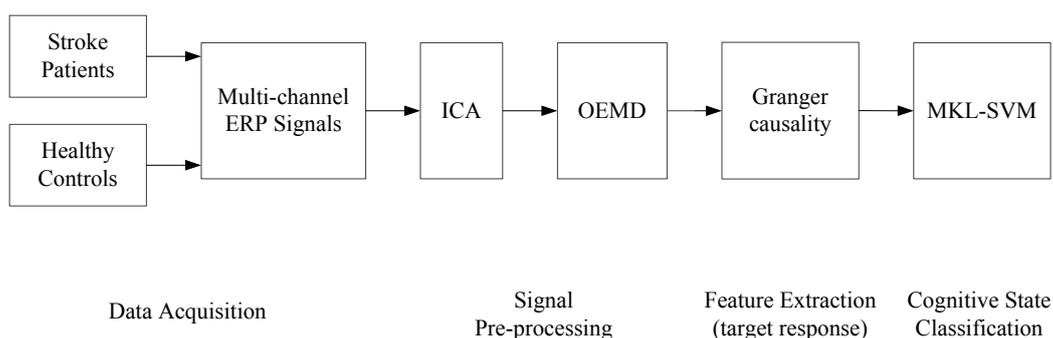
The Shannon entropy $SE_{f_j^i}$ can be defined as the entropy of the signal component f_j^i , $SE_{f_j^i}$ is written as:

$$SE_{f_j^i} = - \sum_i f_j^{i2} \log(f_j^{i2}) \quad (9)$$

3.3. Recognition of Cognitive Task P300

We proposed a classification approach of stroke patients and healthy controls, as illustrated in Figure 3.

Figure 3. The general diagram of cognitive task recognition.



3.3.1. Pre-Processing

Independent component analysis (ICA) is a kind of blind source separation technique that extracts statistically independent sources called independent components (ICs) from a set of recorded signals. Orthogonal empirical model decomposition (OEMD) is a self-adaptive signal processing and data driven method. Compared with classical time-frequency analysis methods, such as short time Fourier transform (STFT) and wavelet, it is based on the local characteristic time scales of a signal and could decompose the signal into a set of complete orthogonal components called intrinsic mode functions (IMFs) which are determined by the signal itself without prior knowledge about the signal [35,36].

In this study, ICA and OEMD were combined to extract more informative features, which can be used as inputs to a classifier to improve the classification accuracy of task-related activity. We first extracted statistically independent sources from the given ERP signals by ICA, and then decomposed them into spectrally independent modes using OEMD algorithm.

3.3.2. Feature Extraction

Understanding and modeling the brain function is based not only on the correct identification of the active brain regions, but also on the functional interactions among the neural assemblies distributed across different brain regions. Correlation, coherence, Granger causality and information entropy have been widely used for the estimation of effective connectivity from the EEG data in the space-frequency and space-time-frequency domains. In particular, Granger causality is one of the prototypical data-driven effective connectivity techniques. It can describe the direct or indirect information flow that one neural system exerts another neural system and quantify causal interactions between brain sources [26,37]. According to the principle of Granger causality, X_2 causes X_1 ($F_{2 \rightarrow 1}$) if the inclusion of past observations of X_2 reduces the prediction error of X_1 in a linear regression model of X_1 and X_2 , as compared to a model which includes only previous observations of X_1 . Two time series $X_1(t)$ and $X_2(t)$ can be defined as [38]:

$$\begin{aligned} X_1(t) &= \sum_{j=1}^p A_{11,j} X_1(t-j) + \sum_{j=1}^p A_{12,j} X_2(t-j) + e_1(t) \\ X_2(t) &= \sum_{j=1}^p A_{21,j} X_1(t-j) + \sum_{j=1}^p A_{22,j} X_2(t-j) + e_2(t) \end{aligned} \quad (10)$$

where p is the maximum number of lagged observations included in the autoregressive (AR) model, A represents the AR coefficients, $e_1(t)$ and $e_2(t)$ are the residuals.

The magnitude of $F_{2 \rightarrow 1}$ in the time domain can be obtained by the log ratio of the prediction error variances, as follows:

$$F_{2 \rightarrow 1} = \ln \frac{\text{var}(e_{1R(12)})}{\text{var}(e_{1U})} \quad (11)$$

where $e_{1R(12)}$ is derived from the AR model omitting the $A_{12,j}$ coefficients and e_{1U} is derived from the full AR model.

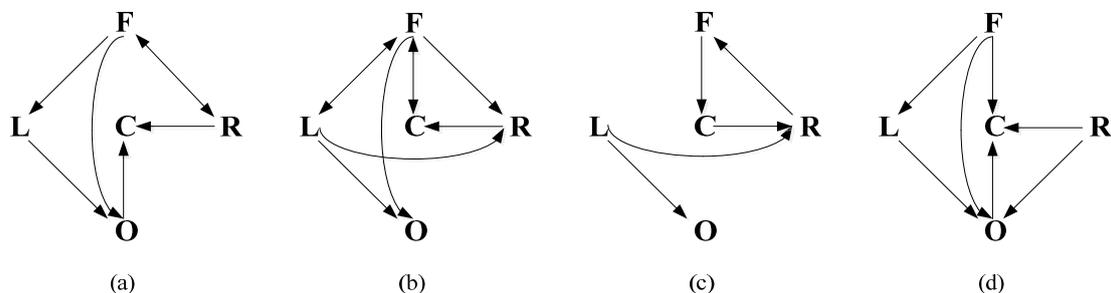
The spectral Granger causality from X_2 to X_1 can be defined as:

$$I_{2 \rightarrow 1}(f) = -\ln \left(1 - \frac{(\Sigma_{22} - (\Sigma_{12}^2 / \Sigma_{11})) |H_{12}(f)|^2}{S_{11}(f)} \right) \quad (12)$$

where Σ is the noise covariance matrix, H is the transfer matrix and S is the power spectrum of X_1 at frequency f [38].

In this study, we analyzed the spectral Granger causality in the theta band. To simplify the analysis, the EEG signals were divided into five regions. They were FCentral (F), LSM (L), Central (C), RSM (R) and Occipital (O), respectively. The raw signals of the electrodes within each region were averaged as the overall activity of the region. Figure 4 shows the average causal network between brain sources for stroke patients and healthy controls in 0-back and 1-back tasks. The connectivity pattern difference between stroke patient group and healthy control group was obvious ($t = 3.8621$, $p < 0.05$). Compared with healthy controls, stroke patients showed a connectivity reduction, especially for the left cortex. There were some functional disconnections among different cortical regions for stroke patients, such as $F \leftrightarrow C$ and $L \rightarrow R$ in 0-back task, $F \rightarrow L$, $O \rightarrow C$ and $F \rightarrow O$ in 1-back task. From an anatomical standpoint, the cognitive disturbances associated with the information transmission may not solely be due to the loss of neurons, but also due to the impairment of distributed neuronal activity [37–40].

Figure 4. Brain connectivity patterns of stroke patients and healthy controls. (a) Stroke patients in 0-back task; (b) Healthy controls in 0-back task; (c) Stroke patients in 1-back task. (d) Healthy controls in 1-back task.



3.4. Parameter Settings

In this study, all EEG signals were classified via a 5-fold cross-validation. 80% of all samples were used for training and 20% for testing.

When WPE was used, the EEG signals in each channel were decomposed into 4 levels through the WPT decomposition with the db4 wavelet. We used D40, D41, D42 and (D43, D21) node to reconstruct the signal, which corresponded to delta, theta, alpha and beta frequency bands respectively. WPE feature was computed for each frequency band.

For the proposed MKL-SVM algorithm, three Poly kernels and ten RBF kernels were chosen. The Poly kernel is a global kernel which has a good generalization ability and a low learning capacity, while the RBF kernel is a local kernel which has a poor generalization ability and a high learning capacity. The overall performance of the kernel function may be improved by their linear combination. In the Poly kernel, if the freedom degree is too high, the generalization ability will decrease, and the

over-fitting problem may occur. The degrees (d) of Poly kernels were chosen as [1, 2, 3]. In the RBF kernel, the small Gaussian width is available for severe changeable samples. The large Gaussian width is available for mild changeable samples. The widths (δ) of RBF kernels were chosen as [0.5, 1, 2, 5, 7, 10, 12, 15, 17, 20]. The kernel parameters chosen in the following single kernel SVM also came from these parameters.

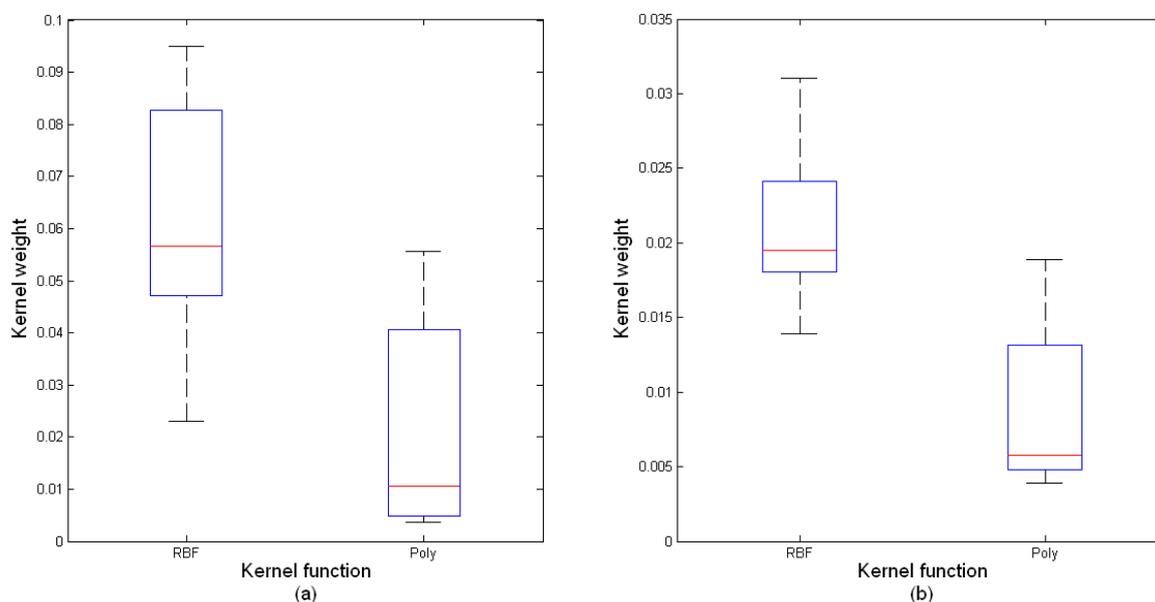
SVM was originally designed for binary classification, but can be extended to the multiclass classification cases. Several approaches have been suggested for multiclass classification using SVM, and here we adopted the one-versus-rest approach. In this approach, a set of binary classifiers, each of which is trained to separate one class from the rest, are undertaken and each test sample is allocated to the class for which the largest decision value is determined [21].

4. Results

4.1. Kernel Weights

The MKL-SVM algorithm combines several possible kernels through optimizing kernel weights. The weight values vary with different classification tasks. Figure 5 shows the boxplots of kernel weights for mental task and cognitive task classifications. In general, the weight values of RBF kernels were greater than the ones of Poly kernels and more RBF kernels were favorable.

Figure 5. Boxplots of kernel weights for mental task and cognitive task classifications. (a) Mental task classification; (b) Cognitive task classification.



4.2. Mental Task Classification

The mental task combinations used for the classification are shown in Table 1. Corresponding to the WPE feature, the accuracies of 2-class, 3-class, 4-class, and 5-class classification are presented in Table 2. The average accuracies over all seven subjects were 99.20%, 81.25%, 76.76%, and 75.25% respectively.

Table 1. The classification problems for different mental task combinations. The five mental tasks are represented by B, C, L, M and R, as described in Section 3.1.1. Here the notation “BC” is used to represent the 2-class classification problem of mental tasks B and C (baseline, visual counting). The notation “BCL” is used to represent the 3-class classification problem of mental tasks B, C and L (baseline, visual counting and letter composing). The notation “BCLM” is used to represent the 4-class classification problem of mental tasks B, C, L and M (baseline, visual counting, letter composing and multiplication). The other notations are similarly defined.

2-Class Classification	3-Class Classification	4-Class/5-Class Classification
BC	BCL	BCLM
BL	BCM	BCLR
BM	BCR	BCMR
BR	BLM	BLMR
CL	BLR	CLMR
CM	BMR	BCLMR
CR	CLM	
LM	CLR	
LR	CMR	
MR	LMR	

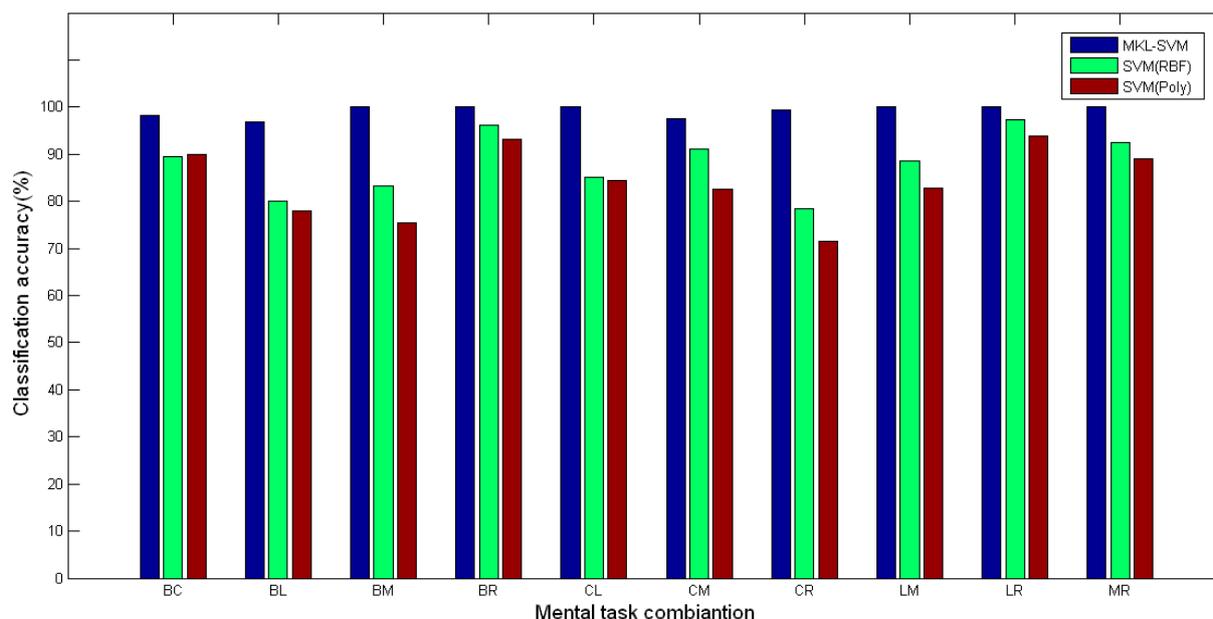
Table 2. Classification accuracies for mental tasks over all seven subjects. Here S1 represents the first subject. The other notations are similarly defined.

Subject	Average Classification Accuracies (%)			
	2-Class	3-Class	4-Class	5-Class
S1	98.24	73.95	73.14	66.15
S2	100	83.78	78.46	72.22
S3	100	80.89	69.62	75.00
S4	100	84.89	88.17	85.71
S5	96.61	79.72	78.64	67.64
S6	99.58	87.17	70.99	85.04
S7	100	78.33	78.33	75.00
Mean	99.20	81.25	76.76	75.25

For 2-class classification, all subjects can completely recognize BM, BR, CL, LM, LR and MR pair tasks. Their classification accuracies were all 100%. The classification accuracy of the BL classification was the lowest, 96.94%. Therefore, it was easy for all subjects to recognize mathematical multiplication and geometric figure rotation tasks, with respect to the baseline state. In addition, the combination of mathematical multiplication and geometric figure rotation tasks was also easy to be distinguished. This supports the neuroscience knowledge that the brain can make fine distinction between calculation and visual tasks, because the left hemisphere of the brain is dominant for the calculation task, while the right hemisphere of the brain is dominant for the visual task [41]. The accuracy of 2-class classification using MKL-SVM was further compared with the single kernel SVM when using same features. The comparison results are shown in Figure 6. As can be seen, the classification results based on MKL-SVM were better than those based on SVM with the Poly kernel ($t = 6.306$, $p < 0.05$) and the

RBF kernel ($t = 5.308$, $p < 0.05$). Besides, RBF had better results than Poly. These results indicated that the overall classification accuracy of MKL-SVM was much higher than the accuracy of the SVM algorithm based on a single kernel.

Figure 6. Accuracy comparisons of 2-class classification cases. In the legend, the notation “SVM (RBF)” represents the SVM algorithm with the RBF kernel. The notation “SVM (Poly)” represents the SVM algorithm with the Poly kernel.



In all ten 3-class mental task classification cases, the average accuracy for BMR classification was the highest, 89.93%, while CLM was the lowest, 74.31%.

In all five 4-class mental task classifications, the average accuracy for BLMR classification was the highest, 83.14%. The average accuracy for BCLM classification was the lowest, 71.09%.

The sensitivity degrees for certain mental tasks in subjects varied, and there were individual differences for the adaptability for the mental task experiment. We noted that the most discriminatory tasks differed among the subjects, which was consistent with the reported studies [42]. None of the subjects had the perfect classification performance for CLM and BCLM. It could be noted from the results that most of the subjects may well recognize BMR tasks due to the high difference in the baseline, mathematical multiplication and geometric figure rotation tasks. From a physiological point of view, the region and level of the excitation in the cerebral cortex are different during different mental tasks. Compared with other regions, the EEG generated from the active area of the cortex is obvious.

Zhang *et al.* [43] studied the multiclass classification from four subjects using the frequency band powers and asymmetry ratios from the frequency range of 0 to 100Hz with the Fisher discriminant analysis, in which the average accuracies for 3-class, 4-class, and 5-class classifications were 65.88%, 58.18%, and 52.75% respectively (different sizes of training set and test set). Comparatively, the performance of the proposed algorithm was better. As mentioned above, SVM with the RBF kernel performed better than that with the Poly kernel. We also compared the average accuracy using MKL-SVM with SVM based on single RBF kernel. The comparisons of 3-class classification are shown in Figure 7 ($t = 2.108$, $p < 0.05$). The comparisons of 4-class and 5-class classifications are

shown in Figure 8 ($t = 2.719, p < 0.05$). We note that MKL-SVM performed consistently better. From Figures 7 and 8, we also noted that the same best mental task combinations were observed for both MKL-SVM and the single kernel SVM, *i.e.*, BMR and BLMR. This observation suggested that the most easily recognized tasks were rather independent of the classification methods in most cases.

Figure 7. Accuracy comparisons of 3-class classification cases.

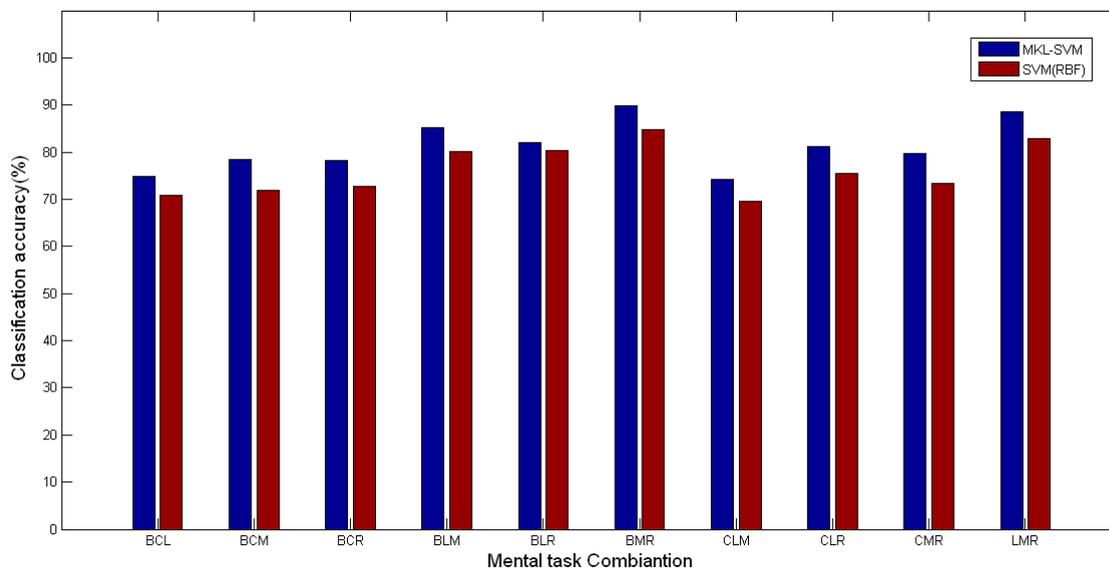
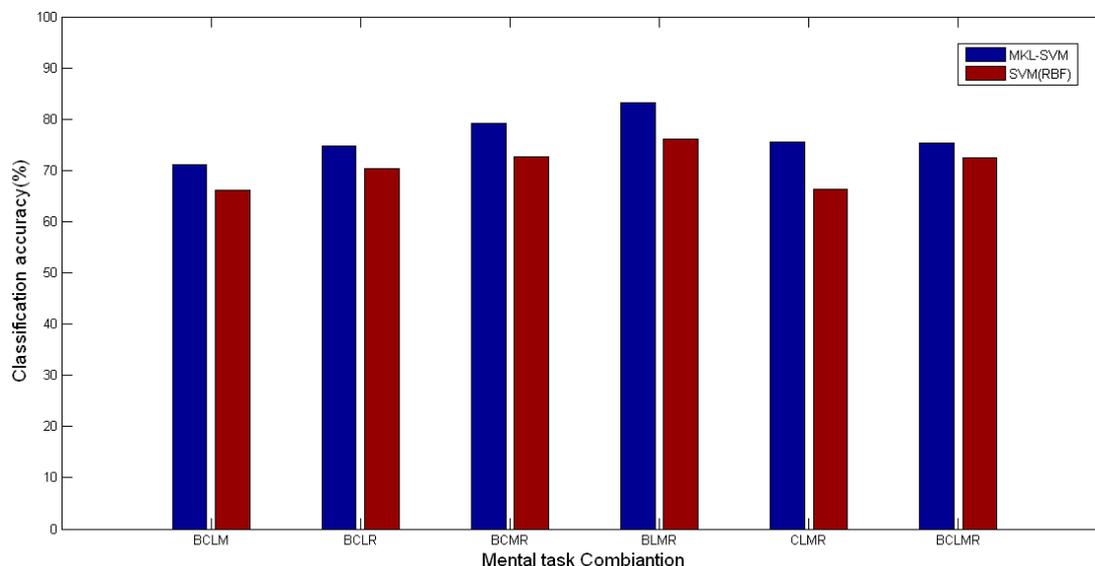


Figure 8. Accuracy comparisons of 4-class and 5-class classification cases.



4.3. Cognitive Task Classification

Table 3 shows the classification comparison results of 0-back and 1-back tasks distinguishing stroke patients *versus* healthy controls between MKL-SVM and another two single kernel SVM algorithms with the spectral causality flow feature. The average classification accuracy for 0-back task was 89.24%. The average classification accuracy for 1-back task was 80.33%. As has been shown, the classification results based on MKL-SVM were better than those based on Poly kernel ($t = 6.221, p < 0.05$) and RBF kernel ($t = 4.187, p < 0.05$). The general classification accuracy of 0-back task was higher than that of

1-back task. These results provide theoretical and experimental basis of the quantity diagnosis for cognitive impairment. It is helpful for the intelligent identification of cognitive function and appropriate rehabilitation training.

Table 3. Classification accuracies for 0-back and 1-back tasks with different classifiers (%). Here SP represents stroke patient, HC represents healthy control.

Classification Tasks	Classifiers	Classification Accuracies (%)
0-back (SP-HC)	MKL-SVM	89.24
	SVM(RBF)	81.61
	SVM(Poly)	77.14
1-back (SP-HC)	MKL-SVM	80.33
	SVM(RBF)	75.28
	SVM(Poly)	70.81

4.4. Computational Cost

The proposed MKL-SVM algorithm was implemented in MATLAB 7.12.0 on a personal computer (using an Intel(R) Core(TM) i5-2430M CPU@2.40 GHz, 2.91 GB RAM). The average runtime varied from 6.86 s to 25.78 s. It included the cross validation and the optimization process. The average runtime with SVM was from 5.44 s to 20.47 s.

5. Conclusions

In this paper, we have presented the MKL-SVM algorithm for the classifications of mental tasks and working memory tasks with WPE and Granger causality features. The proposed method yielded better performance than the single kernel SVM. It can achieve 99.20% average accuracy for 2-class classification and above 75% for multiclass classification. It achieved above 89% accuracy for 0-back task and above 80% for 1-back task. The average CPU runtime was 15.39 s for the training and testing. Therefore, it can be a practical method for EEG-based BCI. Meanwhile, the classification results showed that the most discriminatory tasks varied among different subjects. In general, BMR and BLMR were the most suitable mental tasks to be distinguished. The Granger causality results showed that the effective connections among different cortical regions were reduced in stroke patients.

The physiological phenomena whereby different types of mental activities can activate distinct areas of the cortex and evoke different EEG rhythms made it possible to classify mental tasks using EEG signals and to realize a BCI system based on the conversion of different mental tasks. For example, invalids can express their demands to the carers and people can operate home computers using only their mind through the BCI. It is also an effective method to implement working memory task-based BCI based on cognitive impairment. First, we can perform the accurate identification and assessment of cognitive function by the EEG signal detection. Then, the rehabilitation training for cognitive impairment may be employed by the BCI with the appropriate apparatus.

Acknowledgments

Many thanks to NATION Corporation and Huadong Hospital for providing the EEG instrument and clinical patients. This study was funded by Shanghai Municipal Education Commission Project (No. 12YZ194), Shanghai Science Technology Research Project (No. 11441902400) and Shanghai Natural Science Foundation (No. 14ZR1440100).

Author Contributions

Xiaoou Li performed the algorithm research and data analysis under the supervision of Z. Jane Wang and Xun Chen. Yuning Yan and Wenshi Wei provided the clinical guidance.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Wolpaw, J.R.; Birbaumer, N.; McFarland, D.J.; Pfurtscheller, G.; Vaughan, T.M. Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* **2002**, *6*, 767–791.
2. Leuthardt, E.C.; Schalk, G.; Wolpaw, J.R.; Ojemann, J.G.; Moran, D.W. A brain computer interface using electrocorticographic signals in humans. *J. Neural Eng.* **2004**, *1*, 63–71.
3. Kronegg, J.; Chanel, G.; Voloshynovskiy, S.; Pun, T. EEG-based synchronized brain-computer interfaces: A model for optimizing the number of mental tasks. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2007**, *1*, 50–58.
4. Kachenoura, A.; Albera, L.; Senhadji, L.; Comon, P. ICA: A potential tool for BCI systems. *IEEE Signal Process. Mag.* **2008**, *1*, 57–68.
5. Artusi, X.; Niazi, I.K.; Lucas, M.F.; Farina, D. Performance of a simulated adaptive BCI based on experimental classification of movement-related and error potentials. *IEEE J. Emerg. Sel. Top. Circuit Syst.* **2011**, *1*, 480–488.
6. Furdea, A.; Ruf, C.A.; Halder, S.; Massari, D.D.; Bogdan, M.; Rosenstiel, W.; Matuz, T.; Birbaumer, N. A new (semantic) reflexive brain-computer interface: In search for a suitable classifier. *J. Neurosci. Methods* **2012**, *203*, 233–240.
7. Palaniappan, R.; Raveendran, P.; Nishida, S.; Saiwaki, N. A new brain-computer interface design using fuzzy ARTMAP. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2002**, *10*, 140–148.
8. Donchin, E.; Spencer, K.M.; Wijesinghe, R. The mental prosthesis: Assessing the speed of a P300-based brain-computer interface. *IEEE Trans. Rehabil. Eng.* **2000**, *2*, 174–179.
9. Keirn, Z.A.; Aunon, J.I. A new mode of communication between man and his surroundings. *IEEE Trans. Biomed. Eng.* **1990**, *37*, 1209–1214.
10. Palaniappan, R. Brain computer interface design using band powers extracted during mental tasks. In Proceedings of the 2nd International IEEE EMBS Conference on Neural Engineering, Arlington, VA, USA, 16–19 March 2005; pp. 321–324.

11. Garrett, D.; Peterson, D.A.; Anderson, C.W.; Thaut, M.H. Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2003**, *11*, 141–144.
12. Li, Z.W.; Shen, M.F. Classification of mental task EEG signals using wavelet packet entropy and SVM. In Proceedings of the 8th International Conference on Electronic Measurement and Instruments, Xian, China, 16–18 August 2007; pp. 906–909.
13. Gupta, A. Agrawal, R.K. Relevant feature selection from EEG signal for mental task classification. In Proceedings of the 16th Pacific-Asia Conference Knowledge Discovery and Data Mining, Kuala Lumpur, Malaysia, 29 May–1 June 2012; pp. 431–442.
14. Lehmann, C.; Koenig, T.; Jelic, V.; Prichep, L.; John, R.E.; Wahlund, L.-O.; Dodge, Y.; Dierks, T. Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG). *J. Neurosci. Methods* **2007**, *161*, 342–350.
15. Dauwels, J.; Vialatte, F.; Latchoumane, C.; Jeong, J.; Cichocki, A. EEG synchrony analysis for early diagnosis of Alzheimer's disease: A study with several synchrony measures and EEG data sets. In Proceedings of the 31st Annual International Conference of the IEEE EMBS, Minneapolis, MN, USA, 3–6 September 2009; pp. 2224–2227.
16. Gallego-Jutglà, E.; Elgendi, M.; Vialatte, F.; Sole-Casals, J.; Cichocki, A.; Latchoumane, C.; Jeong, J.; Dauwels, J. Diagnosis of Alzheimer's disease from EEG by means of synchrony measures in optimized frequency bands. In Proceedings of the 34th Annual International Conference of the IEEE EMBS, San Diego, CA, USA, 28 August–1 September 2012; pp. 4266–4270.
17. Ianez, E.; Azorin, J.M.; Ubeda, A.; Fernandez, E.; Sirvent, J.L. LDA-based classifiers for a mental tasks-based brain-computer interface. In Proceedings of IEEE International Conference on System, Man and Cybernetics, Seoul, South Korea, 14–17 October 2012; pp. 546–551.
18. Guo, L.; Wu, Y.X.; Zhao, L.; Cao, T.; Yan, W.L.; Shen, X.Q. Classification of mental task from EEG signals using immune feature weighted support vector machines. *IEEE Trans. Magn.* **2011**, *5*, 866–869.
19. Abdi, M.J.; Hosseini, S.M.; Rezghi, M. A novel weighted support vector machine based on particle swarm optimization for gene selection and tumor classification. *Comput. Math. Method* **2012**, Article ID 320698.
20. Gonen, M.; Alpaydm, E. Multiple kernel learning algorithms. *J. Mach. Learn. Res.* **2011**, *12*, 2211–2268.
21. Chen, X.; Wang, Z.J. Pattern recognition of number gestures based on a wireless surface EMG system. *Biomed. Signal. Proces. Contr.* **2013**, *2*, 184–192.
22. Lanckriet, G.R.G.; Cristianini, N.; Bartlett, P.; Ghaoui, L.E.; Jordan, M.I. Learning the kernel matrix with semidefinite programming. *J. Mach. Learn. Res.* **2004**, *1*, 27–72.
23. Sonnenburg, S.; Ratsch, G.; Schafer, C.; Scholkopf, B. Large scale multiple kernel learning. *J. Mach. Learn. Res.* **2006**, *7*, 1531–1565.
24. Samek, W.; Binder, A.; Müller, K.R. Multiple Kernel Learning for Brain-Computer Interfacing. In Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Osaka, Japan, 3–7 July 2013; pp. 7048–7051.

25. Wang, D.; Miao, D.Q.; Xie, C. Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection. *Expert. Syst. Appl.* **2011**, *38*, 14314–14320.
26. Greenblatt, R.E.; Pflieger, M.E.; Ossadtchi, A.E. Connectivity measures applied to human brain electrophysiological data. *J. Neurosci. Methods* **2012**, *207*, 1–16.
27. Deepa, V.B.; Thangaraj, P.; Chitra, S. Investigation of support vector machine classification system for brain computer interface. *Eur. J. Sci. Res.* **2011**, *4*, 555–560.
28. Guler, I.; Vbeyli, E.D. Multiclass support vector machines for EEG signal classification. *IEEE Trans. Inf. Technol. Biomed.* **2007**, *2*, 117–126.
29. Muller, K.R.; Mika, S.; Ratsch, G.; Tsuda, K.; Scholkopf, B. An introduction to kernel-based learning algorithms. *IEEE Trans. Neural Net.* **2001**, *2*, 181–201.
30. Rakotomamonjy, A.; Bach, F.R.; Canu, S.; Grandvalet, Y. Simple MKL. *J. Mach. Learn. Res.* **2008**, *11*, 2491–2521.
31. Bach, F.R.; Lanckriet, G.R.G.; Jordan, M.I. Multiple kernel learning, conic duality, and the SMO algorithm. In Proceedings of the 21st International Conference on Machine Learning, Banff, AL, Canada, 4–8 July 2004; pp. 41–48.
32. Sonnenburg, S.; Ratsch, G.; Schafer, C. A general and efficient multiple kernel learning algorithm. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 5–8 December 2005; pp. 1273–1280.
33. Rakotomamonjy, A.; Bach, F.R.; Canu, S.; Grandvalet, Y. More efficiency in multiple kernel learning. In Proceedings of the 24th International Conference on Machine Learning, Corvallis, OR, USA, 20–24 June 2007; pp. 775–782.
34. Missonnier, P.; Deiber, M.P.; Gold, G.; Herrmann, F.R.; Millet, P.; Michon, A.; Fazio-Costa, L.; Ibanez, V.; Giannakopoulos, P. Working memory load-related electroencephalographic parameters can differentiate progressive from stable mild cognitive impairment. *Neuroscience* **2007**, *150*, 346–356.
35. Kong, W.Z.; Zhou, Z.P.; Hu, S.Q.; Zhang, J.H.; Babiloni, F.; Dai, G.J. Automatic and direct identification of blink components from scalp EEG. *Sensors* **2013**, *13*, 10783–10801.
36. Mijovic, B.; Vos, M.D.; Gligorijevic, I.; Taelman, J.; Huffel S.V. Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis. *IEEE Trans. Biomed. Eng.* **2010**, *9*, 2188–2196.
37. Sakkalis, V. Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. *Comput. Biol. Med.* **2011**, *41*, 1110–1117.
38. Knyazeva, M.G.; Carmeli, C.; Khadivi, A.; Ghika, J.; Meuli, R.; Frackowiak, R.S. Evolution of source EEG synchronization in early Alzheimer’s disease. *Neurobiol. Aging* **2013**, *34*, 694–705.
39. Seth, A.K. A matlab toolbox for Granger causal connectivity analysis. *J. Neurosci. Methods* **2010**, *186*, 262–273.
40. Dubovik, S.; Bouzerda-Wahlen, A.; Nahum, L.; Gold, G.; Schnider, A.; Guggisberg, A.G. Adaptive reorganization of cortical networks in Alzheimer’s disease. *Clin. Neurophysiol.* **2013**, *124*, 35–43.
41. Palaniappan, R. Utilizing gamma band to improve mental task based brain-computer interface design. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2006**, *3*, 299–303.

42. Faradji, F.; Ward, R.K.; Birch, G.E. Plausibility assessment of a 2-state self-paced mental task-based BCI using the no-control performance analysis. *J. Neurosci. Methods* **2009**, *180*, 330–339.
43. Zhang, L.; He, W.; He, C.; Wang, P. Improving mental task classification by adding high frequency band information. *J. Med. Syst.* **2010**, *34*, 51–60.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).