WCCI 2012 IEEE World Congress on Computational Intelligence
June, 10-15, 2012 - Brisbane, Australia IJCNN

Robust EEG Channel Selection across Sessions in Brain-Computer Interface Involving Stroke Patients

Mahnaz Arvaneh*†, Cuntai Guan*, Kai Keng Ang*, Chai Quek†
*Institute for Infocomm Research - Agency for Science Technology And Research, Singapore
Email: {stuma, ctguan, kkang}@i2r.a-star.edu.sg
†School of Computer Engineering, Nanyang Technological University, Singapore

***Abstract*—Brain-computer interface (BCI) technology has shown the capability of improving the quality of life for people with severe motor disabilities. To improve the portability and practicability of BCI systems, it is crucial to reduce the number of EEG channels as well as to have a good reliability. However, a relatively neglected issue in the EEG channel selection studies is the robustness of selected channels across sessions. This paper investigates whether the selected channels from first session is also useful for subsequent sessions on other days for a stroke patient. For this purpose, a new robust sparse common spatial pattern (RSCSP) algorithm is proposed for optimal EEG channel selection. Thereafter, the robustness of the proposed algorithm as well as 5 existing channel selection algorithms is investigated across 12 sessions data from 11 stroke patients who performed motor imagery based-BCI rehabilitation. The experimental results show that the proposed RSCSP channel selection algorithm significantly outperforms the other channel selection algorithms, when the 8 channels selected from the first session are evaluated on the 11 subsequent sessions. Moreover, there is no significant difference between the classification results of 8 channels selected by the proposed RSCSP algorithm from the first session and the classification results of 8 optimal channels selected from the same session as the test session.**

I. Introduction

Brain-computer interface (BCI) provides a direct communication pathway between a human brain and an external device [1]. Thus, BCI technology enables people with sever motor disabilities to use their brain signals for communication and control [2]. Furthermore, BCI as a rehabilitation tool has been effectively used in restoring motor functions in patients with moderate to sever stroke impairments [3].

In majority of current BCI systems, the brain signals are measured by EEG, due to its low cost and high time resolution compared to other modalities, such as fMRI, fNIRS, etc [4]. To achieve good performance, most EEG based BCIs require signals from multiple sites of the scalp [5]. However, applying a large number of EEG channels may include noisy and redundant signals that degrade the BCI performance [6], [7]. Moreover, using a large number of channels involves a prolonged preparation time that directly impacts the convenience in the use of the BCI. The convenience of users is more crucial when a BCI application is going to be used as a clinical device by a patient in multiple sessions. Therefore, selecting the least number of channels that yield the best or required accuracy over subsequent sessions can balance both needs for performance and convenience.

Various channel selection methods have been proposed in the literature. In [7]-[9], channel selection is embedded in a classifier such as support vector machine (SVM), which recursively eliminates the least-contributed channels in the classification accuracy. In [10] and [11], the channels are ranked based on the mutual information (MI) between the channels and the class labels. The common spatial pattern (CSP) algorithm is also used for channel selection [12], whereby the channels are directly selected according to their CSP coefficients. Recently, a sparse common spatial pattern (SCSP) algorithm was proposed to select the least number of channels within a constraint of classification accuracy [13]. It was shown that the SCSP channel selection significantly reduced the number of channels, and outperformed several existing channel selection algorithms.

Despite extensive works, few channel selection studies focused on stroke patients who are potential users for BCI [14], [15]. Furthermore, the robustness and stability of the channel selection algorithms across different sessions have been relatively neglected. Although many channel selection algorithms are effective in selecting a subset of channels for the class prediction in a same session, they may not be necessarily reliable to identify channels for subsequent sessions on other days. EEG patterns vary from session to session due to several subject's preconditions [16]. For instance, the physical properties of the electrodes such as position and conductivity can change over sessions. Besides, mental conditions such as attention, awakeness and task involvement can display a large variability between sessions. Therefore, the question arises whether a set of channels selected from one session is also effective for other subsequent sessions.

To address this research question, in this study 12 sessions of motor imagery-based BCI data collected from 11 stroke patients [3] are used. First, a new robust sparse common spatial pattern (RSCSP) algorithm is proposed for optimal EEG channel selection, whereby the estimates of the covariance matrices are replaced with the robust minimum covariance determinant (MCD) estimates [17]. Thereafter, the robustness of the proposed algorithm, as well as several existing channel selection algorithms, is investigated across the 12 sessions of EEG data. For this purpose, the proposed RSCSP channel selection algorithm is compared with SCSP [13], CSP [12], fisher criterion (FC) [7], MI [10], and SVM [7] based channel selection algorithms in terms of the classification accuracies of the subsequent sessions using the optimal eight channels selected from the first session. The results are also compared with the results obtained from the optimal eight channels selected from the same session as the test session.

The remainder of this paper is organized as follows: Section II briefly describes 5 popular EEG channel selection algorithms as well as the proposed RSCSP channel selection algorithm. The applied dataset and the performed experiments are explained in Section III. Section IV presents the experimental results and finally Section V concludes the paper.

II. METHOD

In this section, 5 successful EEG channel selection algorithms based on SCSP [13], CSP [12], FC [7], MI [10], and SVM [7] are briefly introduced. Furthermore, a new EEG channel selection algorithm based on a robust sparse common spatial pattern (RSCSP) algorithm is proposed.

A. SCSP based channel selection

The CSP algorithm [18] is an effective technique in discriminating two classes of EEG data. The CSP algorithm linearly transforms the EEG data to a spatially filtered space, such that the variance of one class is maximized while the variance of the other class is minimized. The CSP transformation matrix, $\mathbf{W}$, is generally computed by solving the eigenvalue decomposition problem:

$$\mathbf{C}_1\mathbf{W} = (\mathbf{C}_1 + \mathbf{C}_2)\,\mathbf{W}\mathbf{D}, \tag{1}$$

where $\mathbf{C}_1$ and $\mathbf{C}_2$ are respectively estimates of the covariance matrices of the band-passed EEG measurements of each class; $\mathbf{D}$ is the diagonal matrix that contains the eigenvalues of $\mathbf{C}_1$. The rows of the CSP matrix, $\mathbf{W}$, are the spatial filters and the columns of $\mathbf{W}^{-1}$ are the spatial patterns.

Since band-pass EEG measurements have approximately zero mean values, the covariance matrices are estimated by

$$\hat{\mathbf{C}}_\omega = \frac{1}{(t \times n_\omega) - 1}\mathbf{E}_\omega\mathbf{E}_\omega^T, \tag{2}$$

where $\omega = 1, 2$; $\mathbf{E}_\omega \in \mathbf{R}^{c\times(t*n_\omega)}$ denotes the concatenated EEG measurements of all the trials in the training data for the motor imagery action of class ω; c and t denote the number of EEG channels and EEG samples per channel respectively, and n_ω denotes the number of trials in the training data that belong to class ω.

The rows of the CSP projection matrix give nonuniform weights to channels, so that the differences between two classes of the EEG signals are maximized. Hence, the CSP spatial filters can be seen as source distribution vectors.

The use of the CSP algorithm for EEG channel selection was proposed by Wang *et al.* [12]. In the proposed method, four channels corresponding to the maximal CSP vector coefficients were selected as the optimal channels. However, the weights of the CSP are dense (not sparse). Thus, by eliminating other channels, the remaining signals can no longer be projected onto the direction that best discriminates the two classes of EEG signals. Moreover, since EEG measurements are generally contaminated by artifacts and noise, the CSP algorithm that is based on the covariance matrices of EEG trials, can be distorted by these contaminants [19].

These issues motivated the approach to sparsify the CSP spatial filters to emphasize on a limited number of channels with high variances between the classes, and to discard the rest of the channels with low or irregular variances that may be due to noise or artifacts. To sparsify the CSP spatial filters, first we formulate the CSP algorithm as an optimization problem, thereafter the sparsity is induced in the CSP algorithm by adding an l_1/l_2 norm regularization term to the optimization problem as presented in [13]. The proposed SCSP algorithm is then formulated as:

$$\begin{aligned}
&\min_{\mathbf{w}_i} \quad (1-r)\Big(\sum_{i=1}^{i=m}\mathbf{w}_i\mathbf{C}_2\mathbf{w}_i^T+\sum_{i=m+1}^{i=2m}\mathbf{w}_i\mathbf{C}_1\mathbf{w}_i^T\Big)+r\sum_{i=1}^{i=2m}\frac{\|\mathbf{w}_i\|_1}{\|\mathbf{w}_i\|_2}\\
&\text{Subject to:}\quad \mathbf{w}_i(\mathbf{C}_1+\mathbf{C}_2)\mathbf{w}_i^{\mathrm{T}}=1 \quad i=\{1,2,...,2m\}\\
&\qquad \mathbf{w}_i(\mathbf{C}_1+\mathbf{C}_2)\mathbf{w}_j^{\mathrm{T}}=0 \quad i,j=\{1,2,...,2m\}\;\; i\neq j,
\end{aligned} \tag{3}$$

where r $(0 \leq r \leq 1)$, is a regularization parameter that controls the sparsity (number of removed channels) and the classification accuracy. When $r = 0$, the solution is essentially the same as the CSP algorithm. In this study, for $r \neq 0$, spatial filters obtained from the CSP algorithm are used as the initial point.

To select channels using the proposed method, first two sparse spatial filters corresponding to two motor imagery tasks are obtained by solving the optimization problem given in (3) with $m = 1$. Since the value r controls the number of selected channels and the achieved classification accuracy, it should be carefully chosen to fulfil the application needs. After obtaining the sparse filters, channels corresponding to the zero elements in both of the spatial filters are discarded, and the rest are defined as the selected channels.

B. New RSCSP based channel selection

Many multivariate datasets contain data points that deviate from the pattern suggested by the majority of the data [19], [20]. These data points are called outliers. The classical multivariate method to estimate the covariance matrix of EEG measurements can be strongly affected by even a few outliers [20]. On the other hand, robust statistics provides alternatives to the classical statistical estimates that are less affected by outliers [17]. A useful measure of robustness of an estimator is the breakdown value, which states the smallest amount of outlier contamination that can have an arbitrarily large effect on the estimator [20].

1) Minimum Covariance Determinant estimate: The non-robust CSP algorithm has an inherent breakdown value of 0. The SCSP algorithm, which uses the classical multivariate estimate in (2), also has an inherent breakdown value of 0. Therefore, the robust sparse common spatial pattern (RSCSP) is proposed to replace (2) with the MCD estimator [17]:

$$\hat{\mathbf{C}}_\omega = \frac{1}{(\alpha \times t \times n_\omega) - 1}\hat{\mathbf{E}}_\omega\hat{\mathbf{E}}_\omega^T, \tag{4}$$

where
$$\hat{\mathbf{E}}_\omega = \arg\min_{\hat{\varepsilon} \subset \varepsilon} |\hat{\mathbf{C}}_\omega|; \tag{5}$$
$(1-\alpha)$ is the fraction of outliers to resist, $\alpha = [0.5, 1]$; ε is the set of $t \times n_\omega$ c-dimensional elements of $\mathbf{E}_\omega \in \mathbf{R}^{c\times(t*n_\omega)}$; $\hat{\varepsilon}$ is the subset of ε containing $\alpha * t * n_\omega$ c-dimensional elements of $\mathbf{E}_\omega \in \mathbf{R}^{c\times(\alpha*t*n_\omega)}$; and $|.|$ denotes determinant.

The MCD estimator, given in (5), thus computes a defined fraction α of the data such that the determinant of the estimate of the covariance matrix is minimized. Therefore, the MCD covariance estimation is the mean of the covariance of those α points. The MCD is a robust method in the sense that the estimates are not unduly influenced by outliers in the data, even if there are many outliers. The issues of the MCD estimator are that it depends on the initial estimates and it is iterative [21]. The FASTMCD algorithm resolves these issues by drawing multiple random subsets of the data and iteratively approximates towards a lower determinant [21]. The implementation of FASTMCD is available as the MATLAB function "mcdcov" in the LIBRA Toolbox from [22].

C. CSP based channel selection

In the CSP based channel selection [12], [23] optimal channels for each motor imagery task are determined through the maximums of the absolute value of the corresponding spatial pattern. Let $\mathrm{ch}(|\mathbf{SP}_{1,i}|)$ and $\mathrm{ch}(|\mathbf{SP}_{2,i}|)$ respectively denote the i^{th} best channel of the first and second motor imagery tasks, with corresponding absolute spatial pattern coefficients $|\mathbf{SP}_{1,i}|$ and $|\mathbf{SP}_{2,i}|$. Therefore (6) is calculated to obtain overall ranking:

$$\begin{aligned} \mathrm{CH}_{2i-1} &= \mathrm{ch}(\max(|\mathbf{SP}_{1,i}|, |\mathbf{SP}_{2,i}|)) \\ \mathrm{CH}_{2i} &= \mathrm{ch}(\min(|\mathbf{SP}_{1,i}|, |\mathbf{SP}_{2,i}|)), \end{aligned} \tag{6}$$

where i varies from 1 to the total number of channels. Finally since each channel has been iterated twice in CH, the lower rank is discarded. As shown in (6), in this method channels are pair-wisely selected from both performed motor imagery areas.

D. MI based channel selection

In this algorithm, the channels that their corresponding features have maximum MI with the class labels are ranked as the best channels [10]. In this study, the power of each channel was used as the corresponding feature. The MI of feature vector $\mathbf{X}$ with the class $\omega = \{1, 2\}$ is computed using:

$$I(\mathbf{X}; \omega) = H(\omega) - H(\omega|\mathbf{X}), \tag{7}$$

where $H(\omega)$ is the entropy of the class label defined as:

$$H(\omega) = -\sum_{\omega=1}^{2} P(\omega) \log_2 P(\Omega); \tag{8}$$

and the conditional entropy is

$$H(\omega|\mathbf{X}) = -\sum_{\omega=1}^{2}\sum_{k=1}^{n_t} P(\omega|X_k) \log_2 P(\omega|X_k), \tag{9}$$

where X_k is the feature value of the k^{th} trial from $\mathbf{X}$, and P is the probability function. The conditional probability $P(\omega|X_k)$ can be computed using Bayes rule given in (10) and (11).

$$P(\omega|X_k) = (P(X_k|\omega)\, P(\omega))/P(X_k), \tag{10}$$

$$P(X_k) = \sum_{\omega=1}^{2} P(X_k|\omega)\, P(\omega). \tag{11}$$

The conditional probability $P(X_k|\omega)$ can be estimated using the Parzen Window algorithm [24].

E. FC based channel selection

The FC determines how strongly a feature is correlated with the labels [7], whereby the score R_j of feature j is defined as

$$R_j(\mathbf{X}) = \frac{(\mu(\mathbf{X}_j^1) - \mu(\mathbf{X}_j^2))^2}{V(\mathbf{X}_j^1) + V(\mathbf{X}_j^2)}, \tag{12}$$

where $\mathbf{X}_j^1$ and $\mathbf{X}_j^2$ denote the feature vector of feature j in two different classes; μ and V respectively denote mean and variance. The rank of a feature is simply set to the mean score of the corresponding features. In this study, the power of each channel was used as the feature.

F. SVM based channel selection

The SVM is a classification technique [7] which performs efficiently in a number of real-world problems. In SVM based channel selection, channels are selected according to a recursive feature elimination (RFE) method. RFE method was proposed by Guyon *et al.* [25] and is based on the concept of margin maximization. RFE algorithm is started with all the features and eliminates them backward. In each iteration the SVM classifier is trained on the current subset of features. For each remaining feature $\mathbf{X}_i$, without retraining the classifier, the change in the classification accuracy from the removal of $\mathbf{X}_i$ is estimated. Thereafter the feature that results in improving or least degrading in the classification accuracy is removed. This algorithm is iterated till only a specified number of features remains.

III. Experiments

A. Data description

In this study, the EEG data from 11 hemiparetic stroke patients who underwent motor imagery-based BCI with robotic feedback neuro-rehabilitation were used (refer NCT00955838 in ClinincalTrials.gov) [3]. 27 channels of EEG measurements shown in Fig. 1 were acquired using Nuamps acquisition hardware (http://www.neuroscan.com) with unipolar Ag/AgCl electrodes channels, digitally sampled at 250 Hz with a resolution of 22 bits for voltage ranges of 130 mV. EEG measurements from all the channels were band-pass filtered from 0.05 to 40 Hz by the acquisition hardware.

In the rehabilitation phase, the patient's impaired limb was strapped to the MIT-Manus robot. In each trial the patient was first prepared with a visual cue for 2 s, then a go cue would instruct the patient to perform motor imagery of the impaired

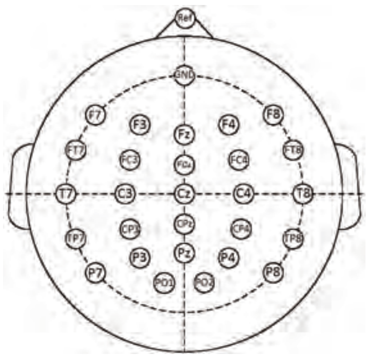

Fig. 1. Position of EEG electrodes used for data acquisition.

hand. If the voluntary motor intent was detected within the 4 s action period, the MIT-Manus robot would assist the patient in moving the impaired limb towards the goal. Finally the patient was asked to rest for 6 s. Each patient underwent 12 neuro-rehabilitation sessions, 3 sessions per week. There was a total of 160 repeats in each session (1 repeat means a complete run from preparation cue to the rest stage). There was a dedicated calibration phase before the rehabilitation phase to train the online classifier.

B. Evaluation of performance across sessions

In our study, the classification problem involved distinguishing between the motor imagery stage and the rest stage. Therefore, each session comprised 160 motor imagery actions of the affected hand and 160 rest conditions. There were a total of 12 sessions of BCI data recorded on different days for each patient. The EEG data were band-pass filtered using elliptic filters from 8 to 35Hz, since this frequency band included the range of frequencies that are mainly involved in performing motor imagery.

An overview of the performed experiments to evaluate different channel selection algorithms over the sessions are shown in Fig. 2 and Fig. 3.

In the first experiment, we aimed to evaluate the performance of the selected channels from the first session on the 11 subsequent sessions recorded on other days. Therefore, as shown in Fig. 2, all the trials of the first session were hired to select subsets of 8 optimal channels using the different channel selection algorithms. The performance of the selected channels was evaluated on the subsequent sessions, such that the first part of each session was used for training a model and the second half was used for testing. Thus, the CSP spatial filters were trained over the selected channels. Then, the EEG signals were spatially filtered using the two first and the two last CSP filters. Finally, the variance of the spatially filtered signals were applied as the inputs of the SVM classifier.

To investigate how well the selected channels from the first session performed on the 11 subsequent sessions, the obtained classification accuracies were compared with the results of the selected channels from each session. Thus, in the second experiment, as shown in Fig. 3, the 8 optimal channels of each session were selected from the first half of the corresponding session, and the performance of the selected channels was evaluated on the second part of it. Thus, the results from the new selected channels were not affected by inter-session nonstationarities and would be proper baselines to compare with the previous results coming from the first experiment.

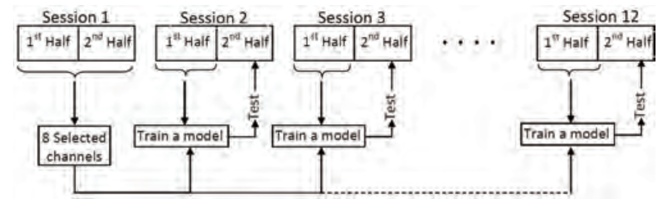

Fig. 2. First experiment: Diagram presenting method for evaluating multi-session channel selection performance.

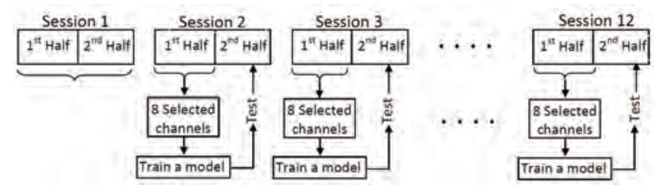

Fig. 3. Second Experiment: Diagram presenting method for evaluating single session channel selection performance.

C. Comparison between different algorithms

In this work, the performance of six different channel selection algorithms were evaluated across the sessions, namely RSCSP, SCSP [13], FC [7], MI [10], SVM [7], CSP [12]. Beside the channels selected from these algorithms, 2 fixed electrode layouts were also considered. Since for the fixed mental tasks of motor imagery, the EEG literature suggested the channels C3 and C4 and adjacent electrodes [4], [9] , our guesses at a generally good subgroup from the 27 recorded EEG channels (see Fig. 1) were as follows:

- The set of ”FC3, C3, CP3, FCz, CPz, FC4, C4, CP4” that is referred as ”Manual1” in this study.
- The set of ”FC3, C3, CP3, T7, T8, FC4, C4, CP4” that is referred as ”Manual2” in this study.

It is noted that this study focused on the subject-dependent channel selection. The power of each channel was used as the feature in FC, MI and SVM based channel selection algorithms. In the SCSP and RSCSP channel selection algorithms, the optimal r value was selected by testing a set of small r values in (3), such that only 8 channels were selected. Moreover, the MCD estimator in the RSCSP algorithm was configured with $\alpha = 0.75$.

IV. Results and Discussion

This section analyzes how transferable the selected channels are from one session to new sessions on other days.

The results of the first experiment (see Fig. 2) are presented in Table I, such that the averaged classification accuracies (averaged test results from session 2 to 12) of 8 channels

TABLE I

AVERAGED CLASSIFICATION ACCURACIES OF 11 SESSIONS USING 8 CHANNELS SELECTED FROM THE FIRST SESSIONS. THE CHANNEL SELECTION WAS PERFORMED USING THE PROPOSED RSCSP, SCSP, CSP, MI, FC, SVM, MANUAL1 AND MANUAL2 ALGORITHMS. "FULL CH." DENOTES THE RESULTS USING ALL THE CHANNELS.

Patient's code	P003	P005	P007	P010	P012	P029	P034	P037	P044	P047	P050	Mean
Full Ch.	59.09	73.71	90.53	66.39	59.94	61.39	56.95	72.96	75.17	73.98	78.29	69.85
RSCSP	59.71	**72.5**	87.68	69.04	**59.52**	**62.07**	**61.59**	**74.61**	**73.29**	70.74	**76.65**	69.76
SCSP	58.98	71.97	**89.88**	**69.78**	58.94	59.74	59.71	72.45	70.05	70.74	75.51	68.88
CSP	57.84	72.12	84.33	63.81	56.22	58.55	60.39	72.11	71.47	67.72	71.42	66.91
MI	**60.625**	72.16	84.47	63.29	55.92	57.24	60.67	69.22	70.62	**72.16**	71.42	67.07
FC	57.61	69.47	85.87	63.61	54.64	58.32	58.35	68.31	63.46	64.49	69.88	64.91
SVM	59.89	72.13	83.11	60.68	55.66	58.95	59.94	66.38	64.22	62.60	73.41	65.18
Manual1	57.33	67.41	85.45	61.55	55.53	59.40	57.53	68.76	65.00	60.56	71.76	64.57
Manual2	58.07	65.80	81.13	64.85	55.08	62.13	59.65	69.28	72.33	66.25	75.22	66.34

selected from the first session using the proposed RCSP, SCSP, CSP, MI, FC, SVM, Manual1 and Manual2 algorithms are compared. The results showed that in 7 patients transferring the 8 channels selected by the proposed RSCSP to the 11 subsequent sessions yielded the highest averaged accuracy compared to the other channel selection algorithms. In 2 other patients the SCSP channel selection algorithm outperformed the other algorithms. Finally in 2 remaining patients the MI-based algorithm achieved the highest accuracy across the sessions. However, averaged results over the 11 stroke patients showed that the proposed RSCSP channel selection algorithm outperformed the SCSP, CSP, MI, FC, SVM, Manual1 and Manual2 channel selection algorithms by an average accuracy of 0.88%, 2.85%, 2.69%, 4.85%, 4.58%, 5.19%,and 3.42% respectively.

Interestingly, the statistical t-test on the entire results across the 11 sessions and the 11 patients showed that the proposed RSCSP algorithm significantly outperformed all the other algorithms ($p = 0.04, 1.7 \times 10^{-8}, 2.8 \times 10^{-6}, 10^{-13}, 5 \times 10^{-12}, 3 \times 10^{-16}$ and 1.5×10^{-9} for the comparison with SCSP, CSP, MI, FC, SVM, Manual1 and Manual2 respectively).

The statistical results also showed that SCSP significantly outperformed CSP, MI, FC, SVM, Manual1 and Manual2 in terms of the classification accuracy across the sessions. This is in the same line with the findings of our previous study [13], although data from healthy subjects were used on that study.

Table I also shows that the 8 channels selected by the proposed RSCSP algorithm from the first session yielded almost the same averaged classification accuracy as all the 27 recorded channels. The statistical t-test on the entire results proved that there is no significant difference between these two groups ($p = 0.845$). This suggests that the 8 channels selected by the RSCSP algorithm from the first session could not only enhance the patient's convenience in the subsequent sessions from the use of lesser channels, but also yield almost the same accuracy as using all the 27 channels.

Table II presents the results of the second experiment (see Fig. 3). In this experiment, the optimal 8 channels were selected from the train data of each session and were tested on the test data of the same session. Thus, the classification results were not affected by inter-session nonstationarities and would be proper baselines to compare with the previous results coming from the first experiment. Since it was shown that the SCSP channel selection algorithm outperformed the CSP, MI, FC and SVM algorithms [13], in this experiment the optimal channels were only selected using the SCSP and RSCP algorithms.

The results in Table II shows that when the selected channels were tested on the same session, the RSCSP and SCSP channel selection algorithms averagely yielded the same classification accuracies. The statistical t-test on the entire results also showed that in this experiment there is no significant difference between the RSCSP and SCSP results ($p = 0.9$).

Comparison between Table I and Table II reveals that the results of 8 channels selected from the same session are superior to the results of channels selected by RSCSP from the first session by an average of less than 1%. Interestingly, the statistical t-test on the entire results showed that there is no significant difference between the results of 8 channels selected by RSCSP from the first session and the results of 8 channels selected by RSCSP and SCSP from the same session ($p = 0.185$ and 0.10 respectively).

V. CONCLUSION

Using a 12 sessions motor imagery-based BCI dataset recorded from 11 stroke patients, this paper investigated how transferable the selected channels are from first session to subsequent sessions on other days. For this purpose, first a new RSCSP algorithm was proposed for optimal EEG channel selection, whereby the estimates of the covariance matrices were replaced with the robust MCD estimates. Thereafter, the proposed RSCSP channel selection algorithm was compared with 5 popular channel selection algorithms in terms of the classification accuracies across the subsequent sessions. Beside the channels selected from those algorithms, two fixed electrode layouts containing channels near motor imagery areas were also considered. These two electrode layouts were abbreviated as Manual1 and Manual2.

Experimental results showed that the proposed RSCSP channel selection algorithm significantly outperformed the SCSP, CSP, MI, FC, SVM, Manual1 and Manual2 channel selection algorithms, in terms of classification accuracy over

TABLE II
AVERAGED CLASSIFICATION ACCURACIES OF 11 SESSIONS USING 8 CHANNELS SELECTED FROM THE FIRST HALF OF EACH SESSION. THE CHANNEL SELECTION WAS PERFORMED USING THE PROPOSED RSCSP AND SCSP ALGORITHMS. "FULL CH." DENOTES THE RESULTS USING ALL THE CHANNELS.

Patient's code	P003	P005	P007	P010	P012	P029	P034	P037	P044	P047	P050	Mean
Full Ch.	59.09	73.71	90.53	66.39	59.94	61.39	56.95	72.96	75.17	73.98	78.29	69.85
RSCSP	59.09	74.875	89.52	67.32	59.86	65.76	62.20	73.13	74.32	73.41	73.41	70.47
SCSP	59.26	73.36	88.45	69.02	57.76	65.93	62.09	74.17	74.00	73.81	77.27	70.46

the 11 subsequent sessions, by an average of 0.88%, 2.85%, 2.69%, 4.85%, 4.58%, 5.19%, and 3.42% respectively. Moreover, the results showed that the 8 channels selected by the proposed RSCSP algorithm from the first session yielded almost the same averaged classification accuracy as all the 27 recorded channels over the 11 sessions.

Finally, the classification results of the optimal eight channels selected from the same session as the tested session were obtained using the RSCSP and SCSP algorithms. Compared to the 8 channels selected by RSCSP from the first session, the selected channels from the same session yielded an average less than 1% improvement in the classification accuracy, while there was no significant difference between these two groups of results. In summary, the results suggested that 8 channels selected by the proposed RSCSP algorithm from one session could be efficiently used on 11 subsequent sessions on different days for stroke patients.

ACKNOWLEDGMENT

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

REFERENCES

[1] N. Birbaumer, "Brain-computer-interface research: Coming of age," *Clin. Neurophysiol.*, vol. 117, pp. 479- 483, 2006.

[2] 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, pp. 767-791, 2002.

[3] K. K. Ang, C. Guan, K. S. G. Chua, B. T. Ang et al, "A Large Clinical Study on the Ability of Stroke Patients in Using EEG-Based Motor Imagery Brain-Computer Interface", *Clinical EEG and Neuroscience*, vol. 42, no. 4, pp. 253-258,Oct. 2011.

[4] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 6, pp. 642-651, 1997.

[5] B. Blankertz, F. Losch, M. Krauledat, G. Dornhege, G. Curio, and K. R. Müller, "The Berlin brain-computer interface: Accurate performance from first-session in BCI-naive subjects," *IEEE Trans. Biomed. Eng.*, vol. 55, pp. 2452-2462, 2008.

[6] F. Popescu, S. Fazli, B. Yakob, B. Blankertz, and K. R. Müller, "Single trial classification of motor imagination using 6 dry EEG electrodes," *PLoS ONE*, vol. 2, 2007.

[7] T. N. Lal, M. Schröder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Schölkopf, "Support vector channel selection in BCI," *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 1003-1010, 2004.

[8] M. Schröder, T. N. Lal, T. Hinterberger, M. Bogdan, N. J. Hill, N. Birbaumer, W. Rosenstiel, and B. Schölkopf, "Robust EEG channel selection across subjects for brain-computer interfaces," *EURASIP JASP*, vol. 19, pp. 3103-3112, 2005.

[9] C. Sannelli, T. Dickhaus, S. Halder, E. M. Hammer, K. R. Müller, and B. Blankertz, "On optimal channel configurations for SMR-based brain-computer interfaces," *Brain Topogr.*, vol. 23, pp. 186-193, 2010.

[10] T. Lan, D. Erdogmus, A. Adami, M. Pavel, and S. Mathan, "Salient EEG channel selection in brain computer interfaces by mutual information maximization," in *Proc. 27th Annu. Int. Conf. Engineering in Medicine and Biology Society*, 2005, pp. 7064-7067.

[11] T. Lan, D. Erdogmus, A. Adami, S. Mathan, and M. Pavel, "Channel selection and feature projection for cognitive load estimation using ambulatory EEG," *Comput. Intell. Neurosci.*, vol. 2007, pp. 8-20, 2007.

[12] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selelction in motor imagery based brain-computer interface," in *proc. 27th Annu. Int. Conf. Engineering in Medicine and Biology Society*, 2005, pp. 5392-5395.

[13] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing the Channel Selection and Classification Accuracy in EEG-Based BCI," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1865-1873, Jun. 2011.

[14] J. Zhou, J. Yao, J. Deng, and J. P. A. Dewald, "EEG-based classification for elbow versus shoulder torque intentions involving stroke subjects," *Computers Biol. Medicine*, vol. 39, pp. 443452, May 2009.

[15] W. Tam, K. Tong, F. Meng, and S. K. Gao, "A Minimal Set of Electrodes for Motor Imagery BCI to Control an Assistive Device in Chronic Stroke Subjects: A Multi-Session Study," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, pp. 617-627, 2011.

[16] P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao, and K. R. Müller, "Towards adaptive classification for BCI," *J. Neural Eng.*,vol. 3, pp. R13R23, 2006.

[17] C. Croux and G. Haesbroeck, "Influence Function and Efficiency of the Minimum Covariance Determinant Scatter Matrix Estimator," *Journal of Multivariate Analysis*, vol. 71, pp. 161-190, 1999.

[18] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, pp. 41-56, 2008.

[19] M. Krauledat, G. Dornhege, B. Blankertz, and K. R. Müller, "Robustifying EEG data analysis by removing outliers," *Chaos and Complexity Letters*, vol. 2, pp. 259-274, 2007.

[20] M. Hubert, P. J. Rousseeuw, and S. Van Aelst, "High-Breakdown Robust Multivariate Methods," *Stat. Sci.*, vol. 23, no. 1, pp. 92-119, 2008.

[21] P. J. Rousseeuw and K. V. Driessen, "A Fast algorithm for the Minimum Covariance Determinant Estimator," *Technometrics*, vol. 41, no. 3, pp. 212-223, 1999.

[22] M. Hubert, "LIBRA: A Matlab Libray for Robust Analysis", Katholieke Universiteit Leuven, http://wis.kuleuven.be/stat/robust/LIBRA.html, 2009.

[23] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing EEG channel selection by regularized spatial filtering and multi band signal decomposition," in *IASTED Int. Conf. Biomedical Engineering*, Innsbruck, 2010, pp. 86-90.

[24] N. Kwak and C. H. Choi,"Input feature selection by mutual information based on Parzen window" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, pp. 1667-71, 2002.

[25] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157-1182, 2003.