Extracting and selecting discriminative features from high density NIRS-based BCI for numerical cognition

Kai Keng Ang, Juanhong Yu, Cuntai Guan
Institute for Infocomm Research – Agency for Science and Technology And Research, Singapore
Email: {kkang, jyu, ctguan}@i2r.a-star.edu.sg

***Abstract*—Near-Infrared Spectroscopy (NIRS)–based Brain-Computer Interface (BCI) was recently studied for numerical cognition. This study presents a study using high density 348 channels NIRS-based BCI from 8 healthy subjects while solving mental arithmetic problems with two difficulty levels and the rest condition. The existing feature extraction and selection methods on the existing study were presented only for low density 16 channels NIRS-based BCI, and required the specification on the number of features to select to yield desirable performance. This paper presents a method of extracting discriminative features from high density single-trial NIRS data using common average reference spatial filtering and single-trial baseline reference, and a method of automatically selecting a set of discriminative and non-redundant features using the Mutual Information-based Rough Set Reduction (MIRSR) and Supervised Pseudo Self-Evolving Cerebellar (SPSEC) algorithms. The performance of the proposed method is evaluated using 5×5-fold cross-validations on the single-trial NIRS data collected using the support vector machine classifier. The results yielded an overall average accuracy of 71.4% and 91.0% in classifying hard versus easy tasks and hard versus rest tasks respectively using the proposed method, compared to 46.1% and 62.2% respectively using existing methods. The results demonstrated the effectiveness of using the proposed feature extraction and selection method in high density NIRS-based BCI for assessing numerical cognition.**

Keywords–Brain-Computer interface; near-infrared spectroscopy, mental arithmetic; feature extraction; feature selection.

I. INTRODUCTION

Studies have investigated the neural correlates of arithmetic and numeric processing using functional magnetic resonance imaging (fMRI) [1]. However, fMRI-based studies are costly, and the physical restrictions of the fMRI scanner limit its usability to laboratory and clinical facilities. Alternatively, Near-Infrared Spectroscopy (NIRS) is a non-invasive optical neural imaging technique that measures concentration changes of oxyhemoglobin (HbO_2) and deoxyhemoglobin (Hb) in the cerebral vessels by means of different absorption spectra in the near infrared range [2]. Compared to fMRI, NIRS instrumentation is relatively smaller, more portable, and less expensive to purchase and maintain [3]. For these reasons, NIRS is ideally suited for the development of portable real-time brain signal measuring device, known as a brain-computer interface (BCI), which allows the direct translation of brain signals into commands for controlling an external device [4].

The feasibility of using NIRS-based BCI to discriminate left and right motor imagery from hemodynamic responses was first demonstrated in [5], and later in [6], [7]. Subsequently, studies have also shown that other cognitive tasks, such as performing mental arithmetic, generally cause an increase of oxyhemoglobin associated with decreases of deoxyhemoglobin in the prefrontal cortex [8]. Recently, the feasibility of using a low density 16 channels NIRS-based BCI for assessing level of numerical cognition had been demonstrated in [9].

The previous study in [9] collected NIRS data from 20 healthy subjects who performed 3 difficulty levels of mental arithmetic tasks. The subjects performed two 1-digit additions for the easy tasks, 1-digit and 2-digits additions for the medium tasks, and two 2-digits additions for the hard tasks. A total of 75 trials of mental arithmetic tasks evenly distributed into the 3 difficulty levels were collected from each subject. However, the data were collected such that 5 trials of the same difficulty level formed a block, and a total of 15 randomized blocks were collected from each subject in consideration of the slow hemodynamic responses. This experimental design inherently included correlated single-trials in each block. The previous study also presented a simple method of extracting the features by taking the averaged changes in oxyhemoglobin and deoxyhemoglobin across 12 s of NIRS data recorded for a single trial, and results were presented using 5×5-fold cross-validations on the 75 single-trials of NIRS data collected. In addition, the feature selection method in the previous study required the specification on the number of features to select in order to yield desirable performance.

This paper presents a study of high density 348 channel NIRS-based BCI for assessing the level of numerical cognition in subjects performing mental arithmetic tasks. The motivations of conducting this study compared to the previous study [9] are: Firstly, to address the issue on correlated single-trials of mental arithmetic tasks executed in blocks of 5 trials in the previous study. Secondly, to investigate the effectiveness of a high density NIRS-based BCI for assessing numerical cognition compared to low density NIRS-based BCI in the previous study. Thirdly, to investigate the effectiveness using the simple feature extraction method from the previous study in a high density NIRS-based BCI. Last but not least, to investigate improved feature extraction and computational intelligent feature selection method to extract and select discriminative features in high density NIRS-based BCI, as well as to address the issue on the need to specify the number of features to select in the existing feature selection method.

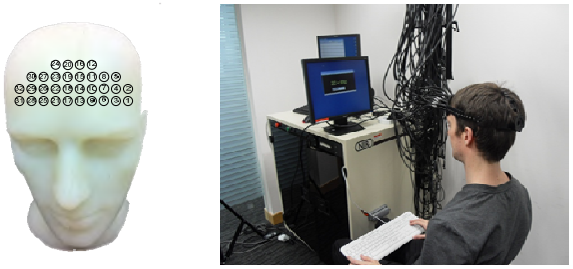
II. METHOD

This section describes the experiment that used high density NIRS-based BCI for assessing numerical cognition, the computation of the hemodynamic responses from the data, the feature extraction and selection method used in the previous study [9], and the proposed feature extraction and selection method used in this study.

A. NIRS data collection

The NIRS data were collected from 8 healthy subjects (3 females, mean age 28.6±8.38) recruited from staffs and students of the Brain-Computer Interface laboratory in the Institute for Infocomm Research, A*STAR. All subjects were fully informed, and consented to participate in the study.

The NIRS data were collected using the DYnamic Near-Infrared Optical Tomography (DYNOT) Imaging System (NIRx Medizintechnik GmbH, Berlin, Germany) with two wavelengths (λ = 760 & 830 nm) using 32 co-located optodes, each serving as source and detector, on the prefrontal cortex of the subject's head as shown in Fig. 1(a). The optodes were fixed on the prefrontal cortex using an open scaffolding structure with individually spring-loaded fibers to ensure stable optical contact. The setup measured 32 channels from 32 detectors for each source for each wavelength, which yielded a total of 1024 channels for each wavelength. Since not all channels contained useful data, only those channels with source and detector distances between 1.5 to 3.5 cm measured using the Xensor digitizer were used, yielding a total of 348 data channels for each wavelength.



(a) (b)

Fig. 1.(a) NIRS data were collected using a fiber grid with 32 co-located sources and detectors over the prefrontal cortex. (b) NIRS data collection setup whereby the numerical arithmetic question was presented to the subject on the screen and the answer was captured using a keyboard.

B. Experimental Protocol

During the NIRS data collection experiment, the subjects were seated in a comfortable chair in a room with normal lighting. They were asked to relax before the collection. They were also asked to minimize movement and to respond as quickly and as correctly as possible during data collection.

The subjects underwent a total of 40 trials of mental arithmetic tasks that were evenly distributed into 2 difficulty levels of easy and hard. Each trial comprised of 3 numerical arithmetic questions from the same difficulty level. The subjects performed mental additions of 3-digits number with 2-digits number without carryover for the easy tasks (eg. 543 + 12), and additions of 4-digits numbers with 3-digits with at least 1 carry over (eg. 5432 + 612) for the hard tasks. At the start of each trial, a question was displayed for a maximum of 12 s. The next question was displayed immediately if the subject responded within 12 s, or at the end of 12 s. A period of 20 s of rest was given between each trial. If all the 3 questions from a trial were correctly answered, then the trial was considered correct.

C. NIRS data preprocessing

Let the optical density for wavelength λ from a source and detector channel c be denoted as OD_c^{λ}, the normalized change in optical density denoted as $\Delta\overline{\mathrm{OD}}_c^{\lambda}$ was first computed by dividing each time sample with the mean of the optical signal acquired for the entire session. Next, $\Delta\overline{\mathrm{OD}}_c^{\lambda}$ was low-pass filtered using Chebychev type II filter with a cut-off frequency of 0.14 Hz and pass-band attenuation of 0.02 dB. Subsequently, linear-detrending was performed to remove the drift (low frequency bias) in the NIRS data due to various reasons, such as subject movement, blood pressure variation, and instrumental instability [10]. After filtering and detrending, unity was added to bring the mean of the optical density to unity instead of zero. The optical density changes were denoted as $\Delta\mathrm{OD}_c^{\lambda}$ after these preprocessing steps.

D. Computing hemodynamic responses

The optical density changes $\Delta\mathrm{OD}_c$ were expressed as a linear combination of the changes in oxyhemoglobin $\Delta[\mathrm{HbO}_2]_c$ and deoxyhemoglobin $\Delta[\mathrm{Hb}]_c$ using the modified Beer-Lambert law (MBLL) [2], [11] given by

$$\Delta\mathrm{OD}_c^{\lambda} = L^{\lambda}\mathrm{DPF}^{\lambda}\left(\varepsilon_{\mathrm{Hb}}^{\lambda}\Delta\left[\mathrm{Hb}\right]_c + \varepsilon_{\mathrm{HbO}_2}^{\lambda}\Delta\left[\mathrm{HbO}_2\right]_c\right), \quad (1)$$

where ε^{λ} is the wavelength-dependent extinction coefficient, L^{λ} is the path length from source to detector, and DPF^{λ} is the differential path-length. In this study, the values of ε^{λ} are obtained from [12], and DPF^{λ} = 6.3 and 6.0 are used for λ = 760 and 830 nm respectively.

The optical density changes from the two wavelength were converted to changes in HbO_2 and HB by solving [6]

$$\begin{bmatrix}\Delta\left[\mathrm{HbO}_2\right]_c \\ \Delta\left[\mathrm{Hb}\right]_c\end{bmatrix} = \left(\mathbf{E}^T\mathbf{E}\right)^{-1}\mathbf{E}^T\begin{bmatrix}\Delta\mathrm{OD}_c^{\lambda_1}/L^{\lambda_1}\mathrm{DPF}^{\lambda_1} \\ \Delta\mathrm{OD}_c^{\lambda_2}/L^{\lambda_2}\mathrm{DPF}^{\lambda_2}\end{bmatrix}, \quad (2)$$

where

$$\mathbf{E} = \begin{bmatrix}\varepsilon_{HbO_2}^{\lambda_1} & \varepsilon_{Hb}^{\lambda_1} \\ \varepsilon_{HbO_2}^{\lambda_2} & \varepsilon_{Hb}^{\lambda_2}\end{bmatrix}. \quad (3)$$

E. Previous feature extraction method

The feature extraction method in the previous study [9] was performed by taking the average $\Delta[HBO_2]_c$ across a time segment T of NIRS data recorded for a single trial given by

$$\Delta\left[\mathrm{HbO}_2\right]_c = \frac{1}{T}\int_0^T \Delta\left[\mathrm{HbO}_2\right]_c(t), \quad (4)$$

and the average of $\Delta[HB]_c$ can be similarly performed. The extracted feature vector for the i^{th} trial was then formed using

$$\mathbf{x}_i = \left[\Delta\left[\mathrm{HbO}_2\right]_1 \cdots \Delta\left[\mathrm{HbO}_2\right]_{n_c} \quad \Delta\left[\mathrm{Hb}\right]_1 \cdots \Delta\left[\mathrm{Hb}\right]_{n_c}\right], \quad (5)$$

where $\mathbf{x}_i \in \mathbb{R}^{1\times n_t}$, i=1,2,…,n_t; n_t denotes the total number of trials in the training data, and n_c denotes the total number of channels for each wavelength.

The feature matrix for the training data was then formed using $\mathbf{X} = \left[\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_{n_t}\right]^T$.

F. Previous feature selection method

Feature selection is a process defined as: given a set of d features, select a subset of size k that leads to the smallest classification errors [13]. In the previous study [9], the feature selection is performed using the mutual information criteria defined as: given an initial set $\mathcal{F}$ with d features, find the subset $\mathcal{S}\subset\mathcal{F}$ with k features that maximizes Mutual Information $I(\mathcal{S};\mathcal{C})$ [14]. However, the computation of the mutual information on the features selected is often computationally prohibitive. Hence a suboptimal and computationally efficient method, such as Mutual Information based Best Individual Feature (MIBIF) algorithm [15], [16], is used in [9]. The use of the MIBIF algorithm, which is described in the following, required the specification of k, the number of features to select.

MIBIF Algorithm

- Step 1: Initialization
  Initialize set of d features $\mathcal{F} = \{\mathbf{f}_1, \mathbf{f}_2, \ldots \mathbf{f}_d\}$, set of selected features $\mathcal{S} = \varnothing$.
- Step 2: Compute the MI of features with the output class
  Compute $I(\mathbf{f}_i;\mathcal{C})\ \forall i = 1..d, \mathbf{f}_i \in \mathcal{F}$.
- Step 3: Select the best k features
  Repeat
  Select the feature $\mathbf{f}_i$ that maximizes $I(\mathbf{f}_i;\mathcal{C})$ using

$$\mathcal{F} = \mathcal{F}\setminus\{\mathbf{f}_i\}, \mathcal{S} = \mathcal{S}\cup\{\mathbf{f}_i\} \mid I(\mathbf{f}_i;\mathcal{C}) = \max_{j=1..d, \mathbf{f}_j\in\mathcal{F}} I(\mathbf{f}_j;\mathcal{C}). \quad (6)$$

  Until $|\mathcal{S}| = k$

Given a d-dimensional feature data $\mathcal{F} = \{\mathbf{f}_1, \mathbf{f}_2, \ldots \mathbf{f}_d\}$, whereby $\mathbf{f}_j$=[$f_{j,1}$, $f_{j,2}$, … $f_{j,n}$], n is the number of training data samples; the Mutual Information (MI) between the features $\mathcal{F}$ that are continuous and the class labels $\mathcal{C}$ that are discrete is given as

$$I(\mathcal{F};\mathcal{C}) = H(\mathcal{C}) - H(\mathcal{C}\mid\mathcal{F}), \quad (7)$$

where $\omega\in\mathcal{C}$={1,2…,n_ω}. The entropy of the class $\mathcal{C}$ is

$$H(\mathcal{C}) = -\sum_{\omega=1}^{n_\omega} P(\omega)\log_2 P(\omega), \quad (8)$$

and the conditional entropy of a feature $\mathbf{f}_j$ can be estimated using [17]

$$\hat{H}(\mathcal{C}\mid\mathbf{f}_j) = -\sum_{\omega=1}^{n_\omega}\frac{1}{n_t}\sum_{i=1}^{n_t}\hat{p}(\omega\mid f_{j,i})\log_2\hat{p}(\omega\mid f_{j,i}), \quad (9)$$

where $f_{j,i}$ is the i^{th} trial sample of the j^{th} column of $\mathcal{F}$, and the probability distribution function $\hat{p}(\omega\mid f_{j,i})$ can be estimated using Parzen Window [17], [18]. The conditional entropy of more than one features $\mathbf{s}$ can be estimated using Equation (9) and a multivariate estimate of $\hat{p}(\omega\mid\mathbf{s}_i)$.

G. Existing feature selection method

A more optimal and less computationally efficient feature selection method exists, such as, the Mutual Information-based Sequential Feature Selection (MISFS) algorithm [14]. The MISFS algorithm is described in the following:

MISFS Algorithm

- Steps 1 & 2: Initialize and compute MI of features
  Same as steps 1 & 2 of the MIBIF algorithm.
- Step 3: Select the first feature
  Select the feature $\mathbf{f}_i$ that maximizes $I(\mathbf{f}_i;\mathcal{C})$ using (6).
- Step 4: Greedy selection
  Repeat
  a) Compute $I(\mathbf{f}_i\cup\mathcal{S};\mathcal{C})\ \forall i = 1..d, \mathbf{f}_i\in\mathcal{F}$, the joint MI between the feature i and selected features with the output class.
  b) Select next feature using

$$\mathcal{F} = \mathcal{F}\setminus\{\mathbf{f}_i\}, \mathcal{S} = \mathcal{S}\cup\{\mathbf{f}_i\} \mid I(\mathbf{f}_i\cup\mathcal{S};\mathcal{C}) = \max_{j=1..d, \mathbf{f}_j\in\mathcal{F}} I(\mathbf{f}_j\cup\mathcal{S};\mathcal{C}). \quad (10)$$

  Until $(|\mathcal{S}| = k)\vee(I(\mathcal{S};\mathcal{C}) = 1)$

H. Proposed CAR-SBR feature extraction method

NIRS signals are often contaminated by noise and artifacts of both physical and physiological origin, such as subject's movement, heartbeat, respiration effects and other trends [19]. The filtering and detrending performed in section II.C might not be sufficient to remove all the noise and artifacts. Therefore, the Common Average Reference (CAR) Spatial Filtering, commonly used in EEG-based BCI [20], is proposed to reduce noise and artifacts that are common in all the channels. The CAR method was performed on $\Delta[HBO_2]_c$ using

$$\Delta\overline{[\mathrm{HbO}_2]}_c(t) = \Delta[\mathrm{HbO}_2]_c(t) - \frac{1}{n_c}\sum_{j=1}^{n_c}\Delta[\mathrm{HbO}_2]_c(t), \quad (11)$$

and CAR was similarly performed on $\Delta[HB]_c$.

After performing CAR, Single-trial Baseline Reference (SBR) is proposed to reduce noise and artifacts in each specific channel. The SBR method was performed on $\Delta[HBO_2]_c$ using

$$\Delta[\mathrm{HbO}_2]_c = \frac{2}{T}\left(\int_{T/2}^{T}\Delta\overline{[\mathrm{HbO}_2]}_c(t) - \int_{0}^{T/2}\Delta\overline{[\mathrm{HbO}_2]}_c(t)\right), \quad (12)$$

and SBR on $\Delta[HB]_c$ was similarly performed. The proposed SBR method first computes the baseline reference for a single trial from the average of the first half of the time segment T, then subtracts this baseline reference from the next half of the time segment for $\Delta[HBO_2]_c$ and $\Delta[HB]_c$ respectively. The extracted feature vector for the i^{th} trial is then formed using equation (4) whereby $\Delta[HBO_2]_c$ and $\Delta[HB]_c$ are computed from equation (12).

I. Proposed MIRSR feature selection method

In order to select discriminative and non-redundant features, as well as to address the requirement on the number of features to select, the following Mutual Information-based Rough Set Reduction (MIRSR) algorithm [16] is used.

MIRSR Algorithm

- Step 1: Generation of fuzzy membership functions
  Given n data $\mathcal{F} = \{\mathbf{f}_1, \ldots \mathbf{f}_i, \ldots \mathbf{f}_d\}$ with d features, generate fuzzy membership functions of feature $\mathbf{f}_i$ using the Supervised Pseudo Self-Evolving Cerebellar (SPSEC) algorithm [21] $\forall i$=1..d.
- Step 2: Compute the MI of features $\mathcal{F}$ with the class Ω
  For i=1..d:
  a) Given features $\mathbf{f}_i = \{x_{i,1}, \ldots x_{i,k}, \ldots x_{i,n}\}$ with n data samples, perform classification of each data $x_{i,k}$ using

$$\omega_{i,k} = p_{i,j} \mid \mu_{i,j} = \max_{j'=1..J_i}\mu_{i,j}(x_{i,k}), \quad (13)$$

where $p_{i,j}$ is the class associated with the membership function $\mu_{i,j}$; J_i is the number of membership functions generated for feature i.

b) Estimate $p(\omega \mid \mathbf{f}_i)$ using

$$\hat{p}(\omega \mid \mathbf{f}_i) = \hat{n}_i / n_i, \quad (14)$$

where $\hat{n}_i$ is the number of correct classification $\hat{\omega}_{i,k} = \omega$ $\forall k$=1..n.

c) Compute the conditional entropy using (15) and subsequently $I(\mathbf{f}_i; \mathcal{C})$ using (7) and

$$H(\mathcal{C} \mid \mathbf{f}_i) = -\sum_{\omega=1}^{N_c}\hat{p}(\omega \mid \mathbf{f}_i)\log_2\hat{p}(\omega \mid \mathbf{f}_i). \quad (15)$$

End for

- Step 3: Select best k features
  Same as MIBIF step 3 using k=2 $\log_2 d$ in (6).
- Step 4: Remove redundant attributes
  Remove membership functions that are not selected from step 3 and perform reduction using RSPOP step 2 [22].

In step 1 of the MIRSR algorithm, the SPSEC algorithm [21] is used to first generate the fuzzy membership functions of each individual features. In step 2, the fuzzy membership functions generated are then used to perform a classification of the data to estimate the conditional entropy and subsequently the mutual information of each individual feature. In contrast, the MIBIF and MISFS algorithms estimate the conditional entropy and subsequently the mutual information of the features from the probability distribution functions estimated using Parzen Window. In step 3, the MIBIF algorithm is used to select a small subset of best k features, and then the RSPOP algorithm is used in step 4 to perform reduction on these selected best features (see RSPOP Attribute Reduction in [22] for a more detailed description).

J. Classification

After feature selection was performed to select discriminative features on the training data, the Support Vector Machine (SVM) classifier [23] was used to classify the selected features.

III. Experimental Results

This section presents the experimental results on the performance of the NIRS-based BCI evaluated using 5×5-fold cross-validations in the classification of the single-trial high density NIRS data on easy versus hard (EvH) tasks, easy versus rest (EvR) tasks, and hard versus rest (HvR) tasks. The time segment T for classifying the EvH tasks was computed for each subject based on the average time taken to answer all the 3 arithmetic questions in a single trial. The time segment of a fixed T=14 s was used for classifying EvR and HvR tasks.

Table 1. Experimental results on the number of correct trials answered by the subjects, and 5×5-fold cross-validations accuracies in classifying the single-trial high density NIRS data on Easy versus Hard (EvH), Easy versus Rest (EvR), and Hard versus Rest (HvR) tasks using the Support Vector Machine (SVM) classifier. The classification accuracies of using previous feature extraction method [9] and MIBIF to select 10 out of 686 extracted features are compared to using the proposed Common Average Reference and Single-trial Baseline Reference (CAR-SBR) feature extraction method and the MIBIF to select 10 features. The results of using the proposed feature extraction method are also presented using MIBIF to select 12 features, using existing MISFS algorithm, and using the proposed MIRSR algorithm.

		Previous FE [9] & MIBIF10 FS			CAR-SBR FE & MIBIF10 FS			CAR-SBR FE & MIBIF12 FS			CAR-SBR FE & MISFS FS			CAR-SBR FE & MIRSR FS		
Subjects	Correct Trials	EvH	EvR	HvR	EvH	EvR	HvR	EvH	EvR	HvR	EvH	EvR	HvR	EvH	EvR	HvR
1	32	34.0	37.0	39.5	59.5	59.5	84.0	57.5	57.0	73.5	57.5	64.5	79.5	69.0	66.5	81.5
2	23	55.0	64.5	71.0	71.0	97.5	95.0	68.0	94.0	96.0	70.5	86.5	92.0	70.5	92.0	92.5
3	28	60.0	84.5	82.0	74.0	100.0	96.0	73.0	100.0	97.5	72.5	100.0	97.0	76.5	99.5	96.5
4	33	57.5	45.0	78.5	73.5	80.5	97.0	76.5	77.5	98.0	69.5	71.5	95.5	70.5	76.0	97.0
5	24	36.0	45.0	54.0	78.5	59.5	84.0	77.5	59.5	83.0	70.5	50.0	79.5	69.0	57.0	83.5
6	30	32.0	34.0	42.0	72.0	61.0	96.0	66.0	59.5	94.0	64.0	56.0	89.5	65.0	60.0	92.5
7	35	44.5	41.5	69.0	61.0	74.0	85.5	43.0	78.0	84.0	52.5	72.0	81.0	59.5	74.0	85.5
8	31	50.0	37.0	61.5	91.5	83.5	100.0	91.0	86.0	100.0	89.0	79.5	99.5	91.0	83.5	99.0
Average		46.1	48.6	62.2	72.6	76.9	92.2	69.1	76.4	90.8	68.3	72.5	89.2	71.4	76.1	91.0

Table 1 shows the number of trials whereby all 3 questions were correctly answered by the subjects, and the 5×5-fold cross-validation classification accuracies obtained using:

1. the previous feature extraction method in [9] described in section II.E and the previous MIBIF feature selection method described in section II.F to select k=10 features (denoted 'Previous FE & MIBIF 10 FS'),
2. the proposed CAR-SBR feature extraction method described in section II.H and MIBIF to select 10 features (denoted 'CAR-SBR FE & MIBIF 10 FS'),
3. the proposed CAR-SBR feature extraction method and MIBIF to select an increased number of 12 features (denoted 'CAR-SBR FE & MIBIF 12 FS'),
4. the proposed CAR-SBR feature extraction method and the existing MISFS algorithm described in section II.G to automatically select a set of features (denoted 'CAR-SBR FE & MISFS FS'),
5. and the proposed CAR-SBR feature extraction and the proposed MIRSR algorithm described in section II.I to automatically select a set of discriminative features (denoted 'CAR-SBR FE & MIRSR FS').

The results showed that the classification of the EvH tasks using the feature extraction method in [9] yielded an averaged accuracy of 46.1% across the 8 subjects. This is in contrast with the result of 69.8% presented in [9] because the study in [9] was performed on 16 channel low density NIRS-based BCI. Thus the previous feature extraction method in [9] was not effective in high density NIRS-based BCI, and that the NIRS data from trials in the same block in [9] were correlated.

The results also showed that the classification of EvH tasks on features extracted using the proposed CAR-SBR method yielded significantly improved averaged accuracies of 72.6% compared to 46.1% using the method in [9] (p=0.0006 using paired two-tail t-test). Furthermore, the results on the classification of the HvR tasks and the EvR tasks on features extracted using the proposed CAR-SBR method also showed significantly improved averaged accuracies of 92.2% and 76.9% compared to 62.2% and 48.6% respectively using the method in [9] (p=0.0006, 0.0001). Hence the result showed that the proposed CAR-SBR feature extraction method yielded significant improvements over the previous feature extraction method.

However, the results showed that the classification of the EvH tasks using the proposed CAR-SBR feature extraction method and MIBIF to select 12 features yielded a deteriorated accuracy of 69.1% compared to the use of MIBIF to select 10 features, although it was not statistically significant (p=0.078 using paired one-tail t-test). Hence the result showed that the use of the previous MIBIF feature selection algorithm required the specification on the number of features to select in order to yield good classification performance. This motivated the use of other feature selection algorithms that do not require the specification on the number of features to select.

On the other hand, the results showed that the classification of the EvH tasks using the proposed CAR-SBR feature extraction method and MISFS to select a set of features also yielded significantly deteriorated accuracy of 68.3% compared to the use of MIBIF to select 10 features (p=0.0035 using paired one-tail t-test). Hence the results showed that although the use of existing MISFS feature selection algorithm did not require the specification on the number of features to select, it did not yield good performance compared to the use of the MIBIF feature selection algorithm. This motivated the investigation of more advanced feature selection algorithms.

Last but not least, the results showed that the classification of the EvH tasks using the proposed CAR-SBR feature extraction method and the proposed MIRSR algorithm to select a set of discriminative and non-redundant features yielded accuracy of 71.4%, which is significantly better than the use of the MISFS algorithm (p=0.038 using paired one-tail t-test). Although this results is not significantly different from the use of the existing MIBIF feature selection algorithm to select 10

features (p=0.561 using paired two-tail t-test), the use of the MIRSR algorithm does not require the specification on the number of features to select. Thus the unbiased result of using the MIRSR algorithm is comparable to biased results of using the MIBIF algorithm whereby the better classification accuracy could be determined from various selection on the number of features to select.

IV. Conclusions

This paper presents a method of extracting discriminative features from high density NIRS-based BCI using Common Average Reference spatial filtering and Single-trial Baseline Reference (CAR-SBR), and a method of selecting a set of discriminative features using the Mutual Information-based Rough Set Reduction (MIRSR) algorithm. A study was performed to collect a high density 348 channels of NIRS data for each wavelength from the prefrontal cortex of 8 subjects in performing two difficulty levels of mental arithmetic tasks and the rest condition.

The results showed that previous feature extraction method for low density NIRS-based BCI was not effective for the high density NIRS-based BCI for assessing numerical cognition. Although the use of the proposed CAR-SBR feature extraction method improved upon the previous method, the use of the previous Mutual Information-based Best Individual Features (MIBIF) feature selection algorithm required the specification on the number of features to select. Hence the classification results were biased since the number of features that yielded better results was presented. However, the results showed that the use of existing Mutual Information-based Sequential Feature Selection (MISFS) algorithm that could automatically select optimal number of features did not yield comparable performance compared to the MIBIF algorithm. Nevertheless, the results showed that the use of the proposed Mutual Information-based Rough Set Reduction (MIRSR) algorithm yielded improved results compared to the MISFS algorithm, as well as comparable results to the biased MIBIF algorithm that selected 10 features.

Hence the results demonstrated the effectiveness of using the proposed feature extraction and selection methods in high density NIRS-based BCI, as well as the application of such a BCI to provide a feedback on the level of numerical cognition.

References

[1] M. H. Ashcraft, M. M. Guillaume, and H. R. Brian, "Chapter 4 Mathematical Cognition and the Problem Size Effect," in *Psychology of Learning and Motivation*: Academic Press, 2009, vol. Volume 51, pp. 121-151.

[2] A. Kubler, V. K. Mushahwar, L. R. Hochberg, and J. P. Donoghue, "BCI meeting 2005-workshop on clinical issues and applications," *IEEE Trans. Neural Syst. Rehabil. Eng.,* vol. 14, no. 2, pp. 131-134, Jun 2006.

[3] M. I. Núñez-Peña, M. L. Honrubia-Serrano, and C. Escera, "Problem size effect in additions and subtractions: an event-related potential study," *Neurosci. Lett.,* vol. 373, no. 1, pp. 21-25, 2004.

[4] 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, Jun. 2002.

[5] S. Coyle, T. Ward, C. Markham, and G. McDarby, "On the suitability of near-infrared (NIR) systems for next-generation brain-computer interfaces," *Physiol. Meas.,* vol. 25, no. 4, p. 815, Aug. 2004.

[6] R. Sitaram, H. Zhang, C. Guan, M. Thulasidas, Y. Hoshi, A. Ishikawa, K. Shimizu, and N. Birbaumer, "Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface," *NeuroImage,* vol. 34, no. 4, pp. 1416-1427, Feb. 2007.

[7] S. M. Coyle, T. E. Ward, and C. Markham, M., "Brain-computer interface using a simplified functional near-infrared spectroscopy system," *J. Neural Eng.,* vol. 4, no. 3, p. 219, Sep. 2007.

[8] M. Tanida, K. Sakatani, R. Takano, and K. Tagai, "Relation between asymmetry of prefrontal cortex activities and the autonomic nervous system during a mental arithmetic task: near infrared spectroscopy study," *Neurosci. Lett.,* vol. 369, no. 1, pp. 69-74, 2004.

[9] K. K. Ang, C. Guan, K. Lee, J. Q. Lee, S. Nioka, and B. Chance, "A Brain-Computer Interface for Mental Arithmetic Task from Single-Trial Near-Infrared Spectroscopy Brain Signals," in *Proc. 20th Int. Conf. on Pattern Recogn.*, 2010, pp. 3764-3767.

[10] N. Thakor, "In the Spotlight: Neuroengineering," *IEEE Reviews in Biomedical Engineering,* vol. 2, pp. 18-20, 2009.

[11] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, "Brain-Computer Communication: Motivation, Aim, and Impact of Exploring a Virtual Apartment," *IEEE Trans. Neural Syst. Rehabil. Eng.,* vol. 15, no. 4, pp. 473-482, Dec. 2007.

[12] M. Cope, "The application of near infrared spectroscopy to non invasive monitoring of cerebral oxygenation in the newborn infant, Ph.D Thesis," University College London, London, 1991.

[13] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 22, no. 1, pp. 4-37, 2000 2000.

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

[15] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Mutual information-based selection of optimal spatial-temporal patterns for single-trial EEG-based BCIs," *Pattern Recogn.,* vol. 45, no. 6, pp. 2137-2144, Jun. 2012.

[16] K. K. Ang and C. Quek, "Rough Set-based Neuro-Fuzzy System," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2006, pp. 742-749.

[17] N. Kwak and C.-H. Choi, "Input feature selection by mutual information based on Parzen window," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 24, no. 12, pp. 1667-1671, Dec. 2002.

[18] E. Parzen, "On Estimation of a Probability Density Function and Mode," *Annals Math. Statist.,* vol. 33, no. 3, pp. 1065-1076, Sep. 1962.

[19] F. Matthews, B. A. Pearlmutter, T. E. Ward, C. Soraghan, and C. Markham, "Hemodynamics for Brain-Computer Interfaces," *IEEE Signal Process. Mag.,* vol. 25, no. 1, pp. 87-94, 2008.

[20] Y. Li, K. K. Ang, and C. Guan, "Digital Signal Processing and Machine Learning," in *Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction*. B. Graimann, B. Allison, and G. Pfurtscheller, Eds. Berlin: Springer, 2011, pp. 305-330.

[21] K. K. Ang and C. Quek, "Supervised Pseudo Self-Evolving Cerebellar algorithm for generating fuzzy membership functions," *Expert Syst. Appl.,* vol. 39, no. 3, pp. 2279-2287, 2012.

[22] K. K. Ang and C. Quek, "RSPOP: Rough Set-Based Pseudo Outer-Product Fuzzy Rule Identification Algorithm," *Neural Comput.,* vol. 17, no. 1, pp. 205-243, Jan. 2005.

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