Application of rough set-based neuro-fuzzy system in NIRS-based BCI for assessing numerical cognition in classroom

Kai Keng Ang, Cuntai Guan, Kerry Lee, Jie Qi Lee, Shoko Nioka, Britton Chance

Abstract—**Near-infrared spectroscopy (NIRS) studies have revealed that performing mental arithmetic tasks have associated event-related hemodynamic responses that are detectable. Thus NIRS-based Brain Computer Interface (BCI) has the potential for investigating how to best teach mathematics in a classroom setting. This paper presents a novel computational intelligent method of applying rough set-based neuro-fuzzy system (RNFS) in NIRS-based BCI for assessing numerical cognition. A study is performed on 20 healthy subjects to measure 32 channels of hemoglobin responses in performing three difficulty levels of mental arithmetic. The accuracy is then presented using 5×5-fold cross-validations on the data collected. The results of applying RNFS and its Mutual Information-based Rough Set Reduction (MIRSR) for feature selection is then compared against the Naïve Bayesian Parzen Window classifier and other MI-based feature selection algorithms. The results of applying RNFS yielded significantly better accuracy of 75.7% compared to the other methods, thus demonstrating the potential of RNFS in NIRS-based BCI for assessing numerical cognition.**

I. Introduction

STUDIES using functional magnetic resonance imaging (fMRI) have investigated the neural correlates of arithmetic and numeric processing [1]. These studies could be used to discover ways of improving the teaching of mathematics in a classroom setting. However, the obstacles in fMRI-based studies are the cost and restrictions of the fMRI scanner, which limit its usability to lab 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 smaller, more portable, and less expensive to purchase and maintain [3]. For these reasons, NIRS is ideally suited for the development of portable and potentially clinical real-time systems. Such a portable, real-time brain signal measuring device, known as a brain-computer interface (BCI), allows the direct translation of brain signals into commands for controlling an external device [4]. Thus the development of a portable NIRS-based BCI that assesses the level of numerical cognition has the potential for investigating how to best teach mathematics in a classroom setting.

In the development of NIRS-based BCI, the suitability of recognizing left and right motor imagery from hemodynamic responses was first demonstrated in [5], and later in [6], [7]. NIRS studies have shown that other cognitive tasks, such mental arithmetic, generally cause an increase of HbO_2 associated with a decreases of Hb in the prefrontal cortex [8]. There are recent seminar papers that investigated hemodynamic response using NIRS on mental arithmetic tasks [9], [10], and the recognition of mental workload from NIRS signals [11], [12]. To the best of the authors' knowledge, there is currently no NIRS-based BCI designed to assess the numerical cognition.

This paper presents a study of a NIRS-based BCI for recognizing the problem size effect of mental arithmetic task from 20 healthy subjects. The motivation behind studying the problem size effect [13] is that the amplitude of an event-related brain potentials has been shown to be modulated by the size of the mental arithmetic task [14]. However, whether a NIRS instrument is sensitive to differences in activation induced by the problem size effect remains unknown. Therefore, the objective of this study is to investigate the feasibility of recognizing the problem size effect using NIRS-based BCI from the recorded brain signals, but explicit feedback to the user is not provided. In addition, this paper presents a novel computational intelligent approach of applying a Rough-set based Neuro-Fuzzy System (RNFS) with its Mutual Information-based Rough Set Reduction (MIRSR) feature selection algorithm for recognizing the problem size effect from the single-trial NIRS data collected.

The remainder of this paper is organized as follows: Section II gives a brief background on the mutual information-based feature selection algorithms. Section III describes the RNFS and its MIRSR algorithms. Section IV describes the experiment of collecting NIRS data, the methodology on how to compute the hemoglobin responses, and how to extract the data features for feature selection and classification. The results of applying RNFS in the NIRS-based BCI are then compared against a non-parametric Naïve Bayesian Parzen Window classifier with other Mutual-Information based feature selection algorithms in Section V. Finally Section VI concludes this paper.

K. K. Ang and C. Guan are with the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #21-01 Connexis, Singapore 138632. (email: {kkang, ctguan}@i2r.a-star.edu.sg).

K. Lee and J. Q. Lee are with the Centre for Research in Pedagogy and Practice, National Institute of Education, Singapore. (email: {jieqi.lee, kerry.lee}@nie.edu.sg).

S. Nioka and B. Chance are with Department of Biochemistry and Biophysics, University of Pennsylvania, PA, USA. (email: {nioka, chance}@ mail.med.upenn.edu).

978-1-4244-8126-2/10/$26.00 ©2010 IEEE

II. MUTUAL INFORMATION-BASED FEATURE SELECTION

A. Entropy and Mutual Information

Entropy is a measure of uncertainty of random variables. The entropy of a random variable **X** is defined as

$$H(\mathbf{X}) = -\sum_{x \in \mathbf{X}} p(x) \log_2 p(x), \quad (1)$$

where $p(x)$ is the probability density function of x.

Given a d-dimensional feature data $\mathcal{F} = \{\mathbf{f}_1, \mathbf{f}_2, \ldots \mathbf{f}_d\}$, whereby $\mathbf{f}_j = [f_{j,1}, f_{j,2}, \ldots f_{j,n}]$, 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

$$I(\mathcal{F}; \mathcal{C}) = H(\mathcal{C}) - H(\mathcal{C} \mid \mathcal{F}), \quad (2)$$

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

$$H(\mathcal{C}) = -\sum_{\omega=1}^{n_\omega} P(\omega) \log_2 P(\omega), \quad (3)$$

and the conditional entropy of a feature $\mathbf{f}_j$ can be estimated using [15]

$$\hat{H}(\mathcal{C} \mid \mathbf{f}_j) = -\sum_{\omega=1}^{n_\omega} \frac{1}{n_t} \sum_{i=1}^{n_t} \hat{p}(\omega \mid f_{j,i}) \log_2 \hat{p}(\omega \mid f_{j,i}), \quad (4)$$

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 [15], [16]. The conditional entropy of more than one features **s** can be estimated using Equation (4) and a multivariate estimate of $\hat{p}(\omega \mid \mathbf{s}_i)$.

B. Feature selection

The problem of feature selection is defined as follows: given a set of d features, select a subset of size k that leads to the smallest classification errors [17]. In mutual information-based feature selection, the problem is 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 $I(\mathcal{S};\mathcal{C})$ [18]. However, the application to classification problems is often computationally prohibitive. Hence a suboptimal and computationally efficient method, such as *Mutual Information based Best Individual Feature* (MIBIF) algorithm [19], is used.

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

$$\mathcal{F} = \mathcal{F} \setminus \{\mathbf{f}_i\}, \mathcal{S} = \mathcal{S} \cup \{\mathbf{f}_i\} \mid I(\mathbf{f}_i; \mathcal{C}) = \max_{j=1..d, \mathbf{f}_j \in \mathcal{F}} I(\mathbf{f}_j; \mathcal{C}). \quad (5)$$

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

A more optimal and less computationally efficient method, such as *Mutual Information-based sequential feature selection* (MISFS) algorithm that adopts a Sequential Forward Selection method [18], is described as follows:

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 (5).
- 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

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

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

III. ROUGH SET-BASED NEURO-FUZZY SYSTEM

Neural networks and fuzzy systems are very popular techniques in soft computing [20]. Neuro-Fuzzy Systems is a popular hybrid intelligent system that synergies these two techniques by combining the human-like reasoning style of fuzzy systems with the learning and connectionist structure of neural networks. In 2006, Rough-set based Neuro-Fuzzy System (RNFS) was proposed to synergize the concept of knowledge reduction in rough set theory with neuro-fuzzy systems [19]. Fundamentally, RNFS employed the Rough Set-based Pseudo Outer Product (RSPOP) algorithm [21] that applied the concept of knowledge reduction of rough set theory to solicit of IF-THEN fuzzy rules in neuro-fuzzy systems. The RNFS also incorporates a feature selection method that employs the mutual information maximization scheme that selects attributes with high relevance and the concept of knowledge reduction in rough set theory that selects attributes with low redundancy.

The structure of the RNFS is shown in Fig. 1. Each layer of the RNFS performs a specific fuzzy operation. The d-dimensional input features are represented as non-fuzzy vector $\mathbf{X} = [x_1, \ldots x_i, \ldots x_{n_1}]$ where $n_1 = d$ and the class output ω is represented as non-fuzzy variable Y. Each input node I_i^I represents an input linguistic variable χ_i of the corresponding input x_i. Each input-label node $IL_{i,j}^{II}$ represents the j^{th} linguistic label of the i^{th} input node from the input layer. The input-label nodes constitute the antecedent of the fuzzy rules

and they are represented by a membership function $\mu_{i,j}(x)$. Each rule node R_k^{III} represents an IF-THEN fuzzy rule. The output-label node OL_l^{IV} represents the l^{th} linguistic label of the output y which constitutes the conclusion of the IF-THEN fuzzy rules. Each output-label node is represented by a singleton membership function $\mu_l(x)$ that represents the class ω=l, $l \in \{1,\ldots N_\omega\}$. The output node O^V represents the output linguistic variable ϖ of the output y.

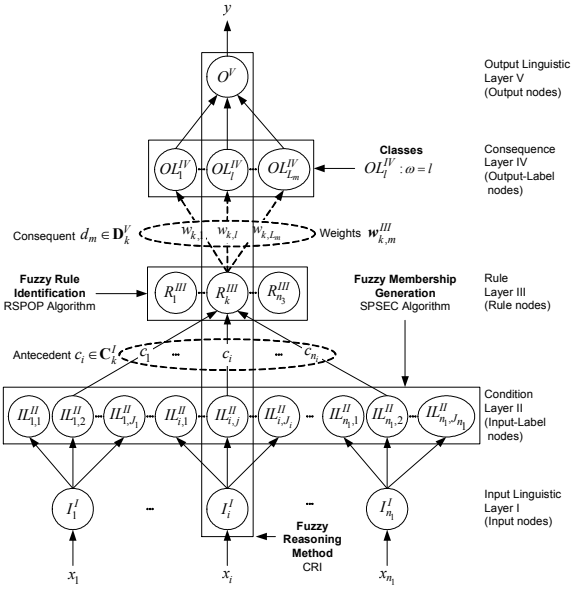

Fig. 1. Rough set-based neuro-fuzzy system architecture

The learning process of RNFS consists of mainly two phases; namely: the generation of membership functions using the Supervised Pseudo-Self Evolving Cerebellar (SPSEC) algorithm [19]; and identification of rules using the Rough set-based Pseudo Outer-Product (RSPOP) algorithm [21]. The following describes the Mutual Information-based Rough Set Reduction (MIRSR) algorithm that is used to perform feature selection [19].

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 SPSEC algorithm $\forall i$=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
  $$\omega_{i,k} = p_{i,j} \mid \mu_{i,j} = \max_{j'=1..J_i} \mu_{i,j}(x_{i,k}), \quad (7)$$
  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
  $$\hat{p}(\omega \mid \mathbf{f}_i) = \hat{n}_i / n_i, \quad (8)$$
  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 (9) and subsequently $I(\mathbf{f}_i;\mathcal{C})$ using (2) and
  $$H(\mathcal{C} \mid \mathbf{f}_i) = -\sum_{\omega=1}^{N_\omega} \hat{p}(\omega \mid \mathbf{f}_i) \log_2 \hat{p}(\omega \mid \mathbf{f}_i). \quad (9)$$
  End for
- Step 3: Select best k features
  Same as MIBIF step 3 using k=2 $\log_2 d$ in (5).
- Step 4: Remove redundant attributes
  Remove membership functions that are not selected from step 3 and perform reduction using RSPOP step 2 [21].

IV. EXPERIMENT

This section describes the data collection of NIRS brain signal for assessing numerical cognition, the methodology to compute the hemodynamic responses, and the methodology to assess the performance of the NIRS-based BCI to classify the problem size effect of mental arithmetic from the hemodynamic responses.

A. Data collection protocol

The data is collected from 20 healthy, right-handed participants (12 male, 8 female; mean age=24.7 years, range 19–30). All subjects had no neurological injury, completed at least 12 years of education, were fully informed, and consented to participate in the study. The subjects were seated in a comfortable chair in a room with normal lighting and were asked to relax before the data collection. They were also asked to minimize movement and to respond as quickly and correctly as possible during the data collection. The recording session for each subject was less than 30 mins.

During the data collection, the subjects underwent a total of 75 trials of arithmetical tasks that were evenly distributed into 3 difficulty levels of easy, medium and hard. The subjects performed two single-digit additions (e.g. 5 + 4) for the easy tasks, single-digit and double-digits additions (e.g. 5 + 34) for the medium tasks, and two double-digits additions (e.g. 53 + 34) for the hard tasks. Arithmetical tasks that involved the carry-over operation is one of the contributing factors of problem size effect [13]. Hence these tasks were excluded. 5 trials of the same difficulty level formed a block and a total of 15 randomized blocks were presented to the subjects. The duration of each trial was 12 s. At the start of each trial, the arithmetic task was presented in the center of a display screen and remained for at most 9 s until the subject responded. After the subject responded, a fixation cross

would appear for the remaining of the 12 s. After completing 1 block, a fixation cross would appear for 30 s before the next block begins.

The data is collected using 16 channels continuous-wave near infrared spectroscopy (NIRS) illustrated in Fig. 2 to record the hemodynamic changes of prefrontal cortex while performing the mental arithmetic tasks. The probes comprised 4 tri-wavelength (730nm 805nm, and 850nm) LEDs and 10 detectors held on the subject's forehead using a Velcro band. The 805nm wavelength is used to detect the dark current or the fixed-pattern noise. A total of 16 channels are collected for each wavelength using a sampling frequency of 3 Hz. This yielded a total of 32 channels of concentration changes in deoxyhemoglobin (HB) and oxyhemoglobin (HBO_2).

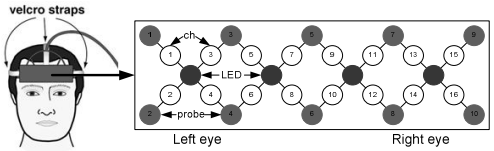

Fig. 2 Multichannel near infrared optodes arrangement on the frontal cortex for the study. A pair of LED and probe yields 1 channel of HB and 1 channel of HBO2 signal. A total of 16 channels of HB and 16 channels of HBO2 signals are thus collected.

B. Preprocessing

Let the optical density for wavelength λ acquired from a set of source and detector labeled as channel c be denoted as OD_c^{λ}. First, the optical densities for wavelength λ_1=730nm and λ_2=850nm are subtracted with wavelength 805nm to remove the ambient. Next, the relative changes in optical density are computed by dividing each time sample by the mean of the optical signal acquired for the session [6] given by

$$\Delta\overline{\mathrm{OD}}_c^{\lambda}(t) = \mathrm{OD}_c^{\lambda}(t) \Big/ \frac{1}{T}\sum_{\tau=1}^{T} \mathrm{OD}_c^{\lambda}(\tau), \qquad (10)$$

where $\Delta\overline{\mathrm{OD}}_c^{\lambda}$ is the normalized change in optical density, and T is the total number of time samples acquired for the session.

$\Delta\overline{\mathrm{OD}}_c^{\lambda}$ is then 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. Next, linear-detrending is perfomed to remove the drift (low frequency bias) in the NIRS measurements due to various reasons, such as subject movement, blood pressure variation, and instrumental instability [22]. These two preprocessing operations are on par with a study that performed low-pass filtering with cut-off frequency of 0.5 Hz and linear-detrending [5]. After filtering and detrending, unity is added to bring the mean of the optical density to unity instead of zero. The change in optical density $\Delta\mathrm{OD}_c^{\lambda}$ is then computed as the negative logarithm from the resultant given by

$$\Delta\mathrm{OD}_c^{\lambda} = -\log\left(\Delta\widetilde{\mathrm{OD}}_c^{\lambda}\right), \qquad (11)$$

where $\Delta\widetilde{\mathrm{OD}}_c^{\lambda}$ denotes the filtered, linearly-detrended optical density with unity added.

C. Computing hemodynamic responses

In NIRS studies, optical density changes $\Delta\mathrm{OD}_c$ can be expressed as a linear combination of the changes in oxyhemoglobin $\Delta[\mathrm{HbO}_2]_c$ and deoxyhemoglobin $\Delta[\mathrm{Hb}]_c$. This equation, referred to as the modified Beer-Lambert law (MBLL) [2], [23], is given by

$$\Delta\mathrm{OD}_c^{\lambda} = L^{\lambda}\mathrm{DPF}^{\lambda}\left(\varepsilon_{\mathrm{Hb}}^{\lambda}\Delta[\mathrm{Hb}]_c + \varepsilon_{\mathrm{HbO}_2}^{\lambda}\Delta[\mathrm{HbO}_2]_c\right), \qquad (12)$$

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.

Typically, dual wavelength measurements of optical absorption are often converted to changes in HbO_2 and HB by solving the series of linear equations [6]

$$\begin{bmatrix}\Delta[\mathrm{HbO}_2]_c \\ \Delta[\mathrm{Hb}]_c\end{bmatrix} = \left(\mathbf{E}^T\mathbf{E}\right)^{-1}\mathbf{E}^T\begin{bmatrix}\Delta\mathrm{OD}_c^{\lambda_1}/L^{\lambda_1}\mathrm{DPF}^{\lambda_1} \\ \Delta\mathrm{OD}_c^{\lambda_2}/L^{\lambda_2}\mathrm{DPF}^{\lambda_2}\end{bmatrix}, \qquad (13)$$

where

$$\mathbf{E} = \begin{bmatrix}\varepsilon_{HbO_2}^{\lambda_1} & \varepsilon_{Hb}^{\lambda_1} \\ \varepsilon_{HbO_2}^{\lambda_2} & \varepsilon_{Hb}^{\lambda_2}\end{bmatrix}. \qquad (14)$$

D. Feature extraction

Feature extraction is performed by taking the average $\Delta[\mathrm{HB}]_c$ and $\Delta[\mathrm{HBO}_2]_c$ across 12s of NIRS data recorded for a single trial. Since there are 16 channels, this resulted in a total of 32 features for a single trial.

The feature data $\mathcal{F}$ is then formed using

$$\begin{aligned}&\left[\mathbf{f}_1 \quad \mathbf{f}_2 \quad \cdots \quad \mathbf{f}_{32}\right] \\ &= \left[\Delta[\mathrm{HbO}_2]_1 \cdots \quad \Delta[\mathrm{HbO}_2]_{16} \quad \Delta[\mathrm{Hb}]_1 \cdots \quad \Delta[\mathrm{Hb}]_{16}\right].\end{aligned} \qquad (15)$$

V. Experimental Results

In this section, the performance of the NIRS-based BCI is evaluated using 5×5-fold cross-validations on the single-trial NIRS data collected. Feature selection and training of the classifier is performed only on the training data in each fold. The classification performance is then evaluated on the test data in each fold. The performance of using the RNFS with MIRSR is compared against the Naïve Bayesian Parzen Window (NBPW) classifier [19] with MIBIF and MISFS algorithms described in section II.

The NBPW is a non-parametric classifier that estimates conditional probability $p(\mathbf{x}|\omega)$ using Parzen Window and prior probability $P(\omega)$ from training data samples, then predicts the class ω with the highest posterior probability $p(\omega|\mathbf{x})$ given a test sample $\mathbf{x}$ using Bayes rule and the naïve assumption that all the features are conditionally independent.

TABLE I
Results for Easy versus Hard (*EvH*), Easy versus Medium (*EvM*), Medium versus Hard (*MvH*);
using NBPW with MIBIF k=4, NBPW with MIBIF k=5, NBPW with MIBIF k=6, NBPW with MISFS, and RNFS with MIRSR

Method	*NBPW, MIBIF k=4*			*NBPW, MIBIF k=5*			*NBPW, MIBIF k=6*			*NBPW, MISFS*			*RNFS, MIRSR*		
Subject	*EvH*	*EvM*	*MvH*	*EvH*	*EvM*	*MvH*	*EvH*	*EvM*	*MvH*	*EvH*	*EvM*	*MvH*	*EvH*	*EvM*	*MvH*
1	70.0	69.6	73.6	72.0	68.8	75.6	72.8	77.2	74.4	73.6	80.4	71.6	73.2	82.4	75.6
2	76.8	84.8	79.6	76.4	86.8	80.0	77.2	87.6	82.8	76.0	87.2	84.4	78.8	87.6	87.6
3	67.6	66.4	70.4	70.8	69.2	72.8	71.2	70.8	73.6	65.2	66.0	69.2	72.4	71.6	78.8
4	68.0	76.8	67.6	68.8	74.8	70.8	68.8	74.0	72.0	74.4	87.6	74.8	72.0	79.2	78.8
5	68.8	76.4	72.0	67.2	74.8	73.2	66.8	75.6	71.6	66.8	75.2	73.6	79.6	78.8	79.6
6	69.2	73.6	70.8	69.2	73.2	71.2	70.8	77.2	70.4	71.2	77.6	88.0	78.0	79.2	89.2
7	53.6	64.0	55.6	54.0	63.2	60.0	52.8	61.2	58.4	56.0	60.4	56.8	56.0	62.0	67.2
8	67.6	64.4	63.2	64.8	71.2	64.4	67.2	71.6	67.2	69.2	73.2	64.4	70.4	75.2	75.6
9	81.6	78.4	78.4	79.6	79.2	76.4	78.8	80.8	78.4	72.4	83.6	79.6	84.4	86.0	80.0
10	70.4	72.0	73.2	71.6	72.8	72.4	70.8	68.0	72.0	72.8	74.0	68.8	80.4	68.8	74.8
11	61.6	68.0	78.8	62.8	71.2	78.4	64.8	71.6	78.4	64.0	67.2	72.0	67.2	83.6	78.8
12	66.0	62.0	70.4	67.2	64.4	70.8	66.8	66.8	67.6	64.4	71.2	65.6	69.6	70.4	68.4
13	66.0	60.8	61.2	66.8	64.8	62.8	66.4	62.0	64.8	68.0	76.8	70.0	66.0	68.4	70.4
14	62.8	68.0	67.2	61.2	70.4	66.8	62.4	68.8	65.2	58.0	82.4	68.8	74.0	82.0	72.4
15	71.6	65.2	67.6	70.8	62.8	68.4	70.0	66.0	67.6	62.4	71.2	62.4	78.4	70.8	71.2
16	69.6	72.4	72.0	71.2	72.4	74.8	70.4	71.6	79.6	66.8	76.4	71.6	75.2	70.8	80.4
17	65.6	77.6	56.4	66.8	74.4	55.2	65.6	74.0	56.8	67.6	73.6	59.6	70.4	74.0	63.2
18	74.8	82.4	70.4	81.2	82.4	68.4	84.8	83.6	69.6	76.8	82.0	66.8	73.2	86.0	69.6
19	64.8	68.4	68.4	60.0	70.0	70.0	57.6	69.6	71.6	61.6	74.8	63.6	62.8	78.4	73.2
20	91.6	92.4	84.0	92.8	91.6	84.4	92.0	91.6	83.6	90.4	88.8	84.8	89.6	93.6	87.6
Avg	69.4	72.2	70.0	69.8	72.9	70.8	69.9	73.5	71.3	68.9	76.5	70.8	73.6	77.4	76.1
		70.5			71.2			71.6			72.1			75.7	

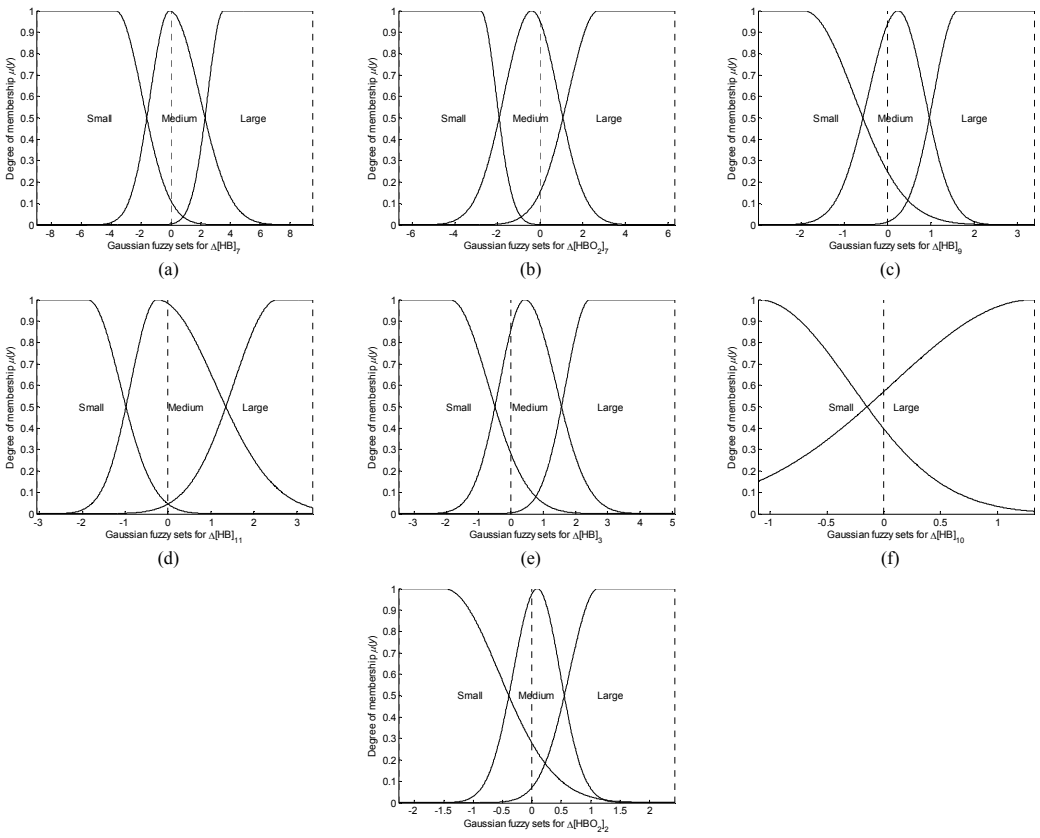

Fig. 3. Membership functions generated using RNFS on single trial NIRS data from subject 20

TABLE II
Interpretable if-then fuzzy rules identified using RNFS with features selected using MIRSR on single trial NIRS data from subject 20. The linguistic labels are abbreviated as Small (S), Medium (M) and Large (L)

R_1	if $\Delta[HB]_7$ is M	and $\Delta[HBO_2]_7$ is M	and $\Delta[HB]_9$ is M	and $\Delta[HB]_3$ is M	and $\Delta[HB]_{10}$ is L	and $\Delta[HBO_2]_2$ is S	
R_7	if $\Delta[HB]_7$ is M	and $\Delta[HBO_2]_7$ is M	and $\Delta[HB]_9$ is M	and $\Delta[HB]_{11}$ is M	and $\Delta[HB]_3$ is M	and $\Delta[HBO_2]_2$ is M	
R_{10}	if $\Delta[HB]_7$ is S	and $\Delta[HBO_2]_7$ is L	and $\Delta[HB]_9$ is S	and $\Delta[HB]_{11}$ is M	and $\Delta[HB]_3$ is M	and $\Delta[HB]_{10}$ is L	and $\Delta[HBO_2]_2$ is M
R_{12}	if $\Delta[HB]_7$ is M	and $\Delta[HBO_2]_7$ is M	and $\Delta[HB]_9$ is M	and $\Delta[HB]_{11}$ is M	and $\Delta[HB]_{10}$ is S	and $\Delta[HBO_2]_2$ is L	
R_{13}	if $\Delta[HB]_7$ is M then task is Easy	and $\Delta[HBO_2]_7$ is M	and $\Delta[HB]_9$ is M	and $\Delta[HB]_{11}$ is M	and $\Delta[HBO_2]_2$ is L		
R_2	if $\Delta[HB]_7$ is L	and $\Delta[HBO_2]_7$ is M	and $\Delta[HB]_9$ is M	and $\Delta[HB]_{11}$ is M	and $\Delta[HB]_3$ is L	and $\Delta[HB]_{10}$ is L	and $\Delta[HBO_2]_2$ is S
R_3	if $\Delta[HB]_7$ is M	and $\Delta[HBO_2]_7$ is M	and $\Delta[HB]_9$ is S	and $\Delta[HB]_{11}$ is S	and $\Delta[HB]_3$ is S	and $\Delta[HB]_{10}$ is S	and $\Delta[HBO_2]_2$ is M
R_4	if $\Delta[HB]_7$ is S	and $\Delta[HBO_2]_7$ is L	and $\Delta[HB]_9$ is S	and $\Delta[HB]_{11}$ is S	and $\Delta[HB]_3$ is S	and $\Delta[HBO_2]_2$ is M	
R_5	if $\Delta[HB]_7$ is S	and $\Delta[HBO_2]_7$ is L	and $\Delta[HB]_9$ is M	and $\Delta[HB]_{11}$ is M	and $\Delta[HB]_3$ is S	and $\Delta[HB]_{10}$ is S	and $\Delta[HBO_2]_2$ is M
R_6	if $\Delta[HB]_7$ is S	and $\Delta[HBO_2]_7$ is L	and $\Delta[HB]_9$ is S	and $\Delta[HB]_{11}$ is M	and $\Delta[HB]_3$ is M	and $\Delta[HB]_{10}$ is S	and $\Delta[HBO_2]_2$ is M
R_8	if $\Delta[HB]_7$ is L	and $\Delta[HBO_2]_7$ is S	and $\Delta[HB]_9$ is L	and $\Delta[HB]_{11}$ is M	and $\Delta[HB]_3$ is M	and $\Delta[HBO_2]_2$ is M	
R_9	if $\Delta[HB]_7$ is M	and $\Delta[HBO_2]_7$ is M	and $\Delta[HB]_9$ is S	and $\Delta[HB]_{11}$ is M	and $\Delta[HB]_3$ is M	and $\Delta[HB]_{10}$ is L	and $\Delta[HBO_2]_2$ is M
R_{11}	if $\Delta[HB]_7$ is L then task is Hard	and $\Delta[HBO_2]_7$ is S	and $\Delta[HB]_9$ is L	and $\Delta[HB]_{11}$ is L	and $\Delta[HB]_3$ is L	and $\Delta[HB]_{10}$ is L	and $\Delta[HBO_2]_2$ is M

Since the MIBIF algorithm described in section II.B requires the definition of k best features to select, the performance of the NIRS-based BCI is evaluated using NBPW with MIBIF using a range of k=4 to k=6. The results evaluated on two of three difficulty levels are presented in TABLE I. The averaged performance of each feature extraction technique is presented in the last row.

Statistical analysis using 1-way ANOVA on the results of the NIRS-based BCI using RNFS with MIRSR compared against NBPW with MIBIF and MISFS in TABLE I revealed significant differences (p=0.0033). Using NBPW, with MIBIF, selecting 5 best individual features yielded significantly better results than selecting 4 best features from t-test (μ=71.2% versus 70.5%, p=0.0266), but selecting 6 best individual features yielded no significantly different results than selecting 5 best individual features (μ=71.6% versus 71.2%, p=0.1648). Furthermore, using NBPW with the more optimal and less computationally efficient MISFS did not yield significantly different results compared to MIBIF selecting 5 best individual features (μ=72.1% versus 71.2%, p=0.2032). In contrast, using the computationally intelligent RNFS with MIRSR yielded significantly better results than using NBPW with MIBIF selecting 5 best individual features or with MISFS (μ=75.7%, p=3.219×10^{-10} and p=1.943×10^{-6} respectively).

Fig. 3 and TABLE II show the fuzzy membership functions generated and the if-then fuzzy rules identified using RNFS on the single trial NIRS data collected from subject 20. The results show that the RNFS with MIRSR not only outperformed the NBPW with MIBIF and MISFS in terms of classification accuracies, it also generated a human interpretable model.

VI. Conclusions

This paper presents a NIRS-based BCI for detecting the problem size effect of mental arithmetic task for assessing numerical cognition. A study is performed on 20 healthy subjects to measure changes in the concentration of oxyhemoglobin (ΔHBO_2) and deoxyhemoglobin (ΔHB) responses in performing three difficulty levels of mental arithmetic. A novel computational intelligent approach of applying Rough-set based Neuro-Fuzzy System (RNFS) with Mutual Information-based Rough Set Reduction (MIRSR) feature selection algorithm is proposed in the NIRS-based BCI for recognizing the problem size of mental arithmetic tasks from the single-trial NIRS data collected. The performance of the proposed NIRS-based BCI using the proposed RNFS and MIRSR is then evaluated using 5×5-fold cross-validations on the data collected. Its performance is compared with the Naïve Bayesian Parzen Window (NBPW) classifier with the Mutual Information-based Best Individual Features (MIBIF) and the Mutual Information-based Sequential Feature Selection (MISFS) algorithms.

The results have shown that RNFS with MIRSR not only outperformed NBPW with MIBIF and MISFS algorithms, it also generated humanly interpretable models that could facilitate the easy of interpreting the resultant model. The results have also demonstrated the potential of applying the RNFS in NIRS-based BCI for recognizing the problem size in mental arithmetic task. The development of such a NIRS-based BCI can potentially provide a feedback on the subject to assess the level of numerical cognition to investigate how to best teach mathematics in a classroom setting.

REFERENCES

[1] L. Zamarian, A. Ischebeck, and M. Delazer, "Neuroscience of learning arithmetic--Evidence from brain imaging studies," *Neuroscience & Biobehavioral Reviews,* vol. 33, no. 6, pp. 909-925, 2009.

[2] A. Villringer and B. Chance, "Non-invasive optical spectroscopy and imaging of human brain function," *Trends Neurosci.,* vol. 20, no. 10, pp. 435-442, 1997.

[3] A. F. Abdelnour and T. Huppert, "Real-time imaging of human brain function by near-infrared spectroscopy using an adaptive general linear model," *NeuroImage,* vol. 46, no. 1, pp. 133-143, 2009.

[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] E. Unlu, H. Bolay, and A. Akin, "Hemodynamic correlates of mental arithmetic task in migraine," in *Proc. BIYOMUT*, 2009, pp. 1-4.

[10] K. Ciftci, B. Sankur, Y. Kahya, and A. Akin, "Complexity and functional clusters of the brain during mental arithmetic," in *Proc. SIU*, 2008, pp. 1-4.

[11] L. Hirshfield, K. Chauncey, R. Gulotta, A. Girouard, E. Solovey, R. Jacob, A. Sassaroli, and S. Fantini, "Combining Electroencephalograph and Functional Near Infrared Spectroscopy to Explore Users' Mental Workload," in *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience*, 2009, pp. 239-247.

[12] L. M. Hirshfield, E. T. Solovey, A. Girouard, J. Kebinger, R. J. K. Jacob, A. Sassaroli, and S. Fantini, "Brain measurement for usability testing and adaptive interfaces: an example of uncovering syntactic workload with functional near infrared spectroscopy," in *Proc. ACM CHI*, 2009, pp. 2185-2194.

[13] 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.

[14] 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.

[15] 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.

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

[17] 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.

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

[19] K. K. Ang and C. Quek, "Rough Set-based Neuro-Fuzzy System," in *Proc. IJCNN'06*, 2006, pp. 742-749.

[20] L. A. Zadeh, "Fuzzy logic, neural networks, and soft computing," *Communications of the ACM,* vol. 37, no. 3, 1994 1994.

[21] 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.

[22] K. E. Jang, S. Tak, J. Jung, J. Jang, Y. Jeong, and J. C. Ye, "Wavelet minimum description length detrending for near-infrared spectroscopy," *Journal of Biomedical Optics,* vol. 14, no. 3, pp. 1-13, May 2009.

[23] T. J. Huppert, S. G. Diamond, M. A. Franceschini, and D. A. Boas, "HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain," *Appl. Opt.,* vol. 48, no. 10, pp. D280-D298, 2009.