



Novel method to classify hemodynamic response obtained using multi-channel fNIRS measurements into two groups: exploring the combinations of channels

Hiroko Ichikawa^{1,2,3*}, Jun Kitazono⁴, Kenji Nagata⁴, Akira Manda⁴, Keiichi Shimamura^{5,6}, Ryoichi Sakuta^{5,6}, Masato Okada^{4,7}, Masami K. Yamaguchi^{1,2}, So Kanazawa⁸ and Ryusuke Kakigi⁹

¹ Department of Psychology, Chuo University, Tokyo, Japan

² Research and Development Initiative, Chuo University, Tokyo, Japan

³ Japan Society for the Promotion of Sciences, Tokyo, Japan

⁴ Department of Complexity Science and Engineering, The University of Tokyo, Kashiwa, Japan

⁵ Department of Pediatrics, Dokkyo Medical University Koshigaya Hospital, Koshigaya, Japan

⁶ Center for Child Development and Psychosomatic Medicine, Dokkyo Medical University Koshigaya Hospital, Koshigaya, Japan

⁷ RIKEN Brain Science Institute, Wako, Japan

⁸ Department of Psychology, Japan Women's University, Kawasaki, Japan

⁹ Department of Integrative Physiology, National Institute for Physiological Sciences, Okazaki, Japan

Edited by:

Aina Puce, Indiana University, USA

Reviewed by:

Giancarlo Zito, 'S. Giovanni Calibita'

Fatebefratelli Hospital, Italy

Yosuke Kita, National Center of

Neurology and Psychiatry, Japan

*Correspondence:

Hiroko Ichikawa, Department of Psychology, Chuo University, 742-1, Higashi-Nakano, Hachioji, Tokyo 192-0393, Japan

e-mail: ichihiro@tamacc.chuo-u.ac.jp

Near-infrared spectroscopy (NIRS) in psychiatric studies has widely demonstrated that cerebral hemodynamics differs among psychiatric patients. Recently we found that children with attention-deficit/hyperactivity disorder (ADHD) and children with autism spectrum disorders (ASD) showed different hemodynamic responses to their own mother's face. Based on this finding, we may be able to classify the hemodynamic data into two those groups and predict to which diagnostic group an unknown participant belongs. In the present study, we proposed a novel statistical method for classifying the hemodynamic data of these two groups. By applying a support vector machine (SVM), we searched the combination of measurement channels at which the hemodynamic response differed between the ADHD and the ASD children. The SVM found the optimal subset of channels in each data set and successfully classified the ADHD data from the ASD data. For the 24-dimensional hemodynamic data, two optimal subsets classified the hemodynamic data with 84% classification accuracy, while the subset contained all 24 channels classified with 62% classification accuracy. These results indicate the potential application of our novel method for classifying the hemodynamic data into two groups and revealing the combinations of channels that efficiently differentiate the two groups.

Keywords: hemodynamic data, near-infrared spectroscopy (NIRS), support vector machine (SVM), sparse modeling, attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorders (ASD)

INTRODUCTION

Near-infrared spectroscopy (NIRS) has been utilized to measure brain activity in humans (for review, Ferrari and Quaresima, 2012). Because NIRS is non-invasive and requires less stabilization of participants than other neuroimaging techniques, NIRS is highly suitable for studies with infants, children (for review, Lloyd-Fox et al., 2010; Gervain et al., 2011) and patients with psychiatric symptoms or disorders such as schizophrenia, depression, anxiety disorder, and attention-deficit/hyperactivity disorder (ADHD) (for review, Fukuda, 2009; Ernst et al., 2012). ADHD is characterized by major symptoms of hyperactivity, impulsivity and inattention (American Psychiatric Association, 2000). Children with ADHD (ADHD children) show atypical hemodynamic response in the prefrontal region associated with attention and working memory deficits (Weber et al., 2005; Negoro et al., 2010; Monden et al., 2012, but see also Schecklmann et al., 2010).

The neural response of ADHD children in face processing is different from that of typically developing children (TD children). Tye et al. (2013) used ERP techniques to examine the face-inversion effect and gaze processing in ADHD children, children with ASD (ASD children), children with comorbid ASD+ADHD, and TD children. They found that children with ADHD (ADHD/ADHD+ASD) showed atypical response that reflect early attentional stage of face processing, while children with ASD (ASD/ASD+ADHD) showed atypical response in gaze processing and atypical neural specialization, which are likely to be more relevant to the characteristic social deficits of autism. As far as we know, Tye et al. (2013) was the first study that found the basic face processing in ADHD children different from that of ASD children and no previous studies has yet investigated familiar face processing in children with ADHD. Dawson et al. (2002) demonstrated that ASD children showed no differential ERPs when viewing their mother's face and viewing an unfamiliar

female face, while TD children did. Although the hemodynamic response to one's mothers' face has not been tested with ADHD children, we may suppose that an atypical response might also be observed in ADHD children.

Recently our group found that boys with ADHD showed a different hemodynamic response to their own mother's face than typically developing boys (Shimamura et al., 2012, under review). Thorell et al. (2012) tested non-clinical 8.5 year-old children and reported that attachment disorganization and executive functioning were independently related to ADHD symptoms. Furthermore, Carlsson et al. (2008) investigated the endogenous and exogenous factors that predict inattentiveness and hyperactivity in middle childhood and demonstrated that the quality of the caregiving more powerfully predicted inattention and hyperactivity than did early biological or temperamental factors. The quality of the relationship between the child and the caregiver might increase or reduce the amount of communication between the two, even in TD children, which may possibly affect the development of the children's neural basis for processing his/her mother's face. (Incidentally, all adults participated as "caregivers" in Shimamura et al.'s study were the mother's of the children.)

Furthermore, we found that boys with ADHD showed a different hemodynamic response to their own mother's face than boys with ASD (Shimamura et al., 2012, under review). Although ASD is characterized by difficulties with social interaction, restricted interests, and repetitive behaviors (American Psychiatric Association, 2000), there is a large overlap of symptoms in ADHD and ASD patients such as hyperactivity, restlessness, and impairments in social cognitive abilities (Yerys et al., 2009; Taurines et al., 2012; van der Meer et al., 2012). It is often difficult in actual clinical practice to distinguish between these patients due to their overlapping symptoms (Yoshida and Uchiyama, 2004). Using fNIRS, we presented boys of both groups with images of their mother's face and measured their cerebral hemodynamics in the bilateral temporal area. Only children with ADHD showed a significantly greater concentration of oxy-hemoglobin (oxy-Hb) in the bilateral temporal area than the pre-task baseline; children with ASD showed a decrease of oxy-Hb concentration in the left temporal area.

These findings suggest the possibility of distinguishing the cerebral hemodynamic data of the ADHD participants from those of the ASD participants and the possibility of classifying the data into two separate diagnostic groups. With such classification, we might be able to predict to which diagnostic group an unknown participant belongs by analyzing his/her hemodynamic data.

To classify the hemodynamic data into two groups, a promising method involves the Support Vector Machine (SVM; Vapnik, 1982). The SVM is a multivariate method for binary classification and has recently been introduced to NIRS studies aimed at the development of BCI (Sitaram et al., 2007; Cui et al., 2010).

When using multivariate pattern analysis methods such as SVM, we can improve classification accuracy by selecting informative variables and eliminating uninformative ones (Weston et al., 2001; Guyon and Elisseeff, 2003; Liu and Yu, 2005). This process is called feature selection or variable selection and has been increasingly applied in recent BCI studies (Yamashita et al.,

2008; Gottemukkula and Derakhshani, 2011). A number of feature selection methods have been proposed in past studies (Guyon and Elisseeff, 2003; Liu and Yu, 2005). These methods, however, do not necessarily select the subset of variables that gives the best classification accuracy because these methods are intended to be applied to high-dimensional data such as genetic data, and are designed to decrease computation time by using approximate treatment. Here, without approximate treatment, we exhaustively evaluated classification accuracy using all $2^4 - 1 = 16,777,215$ subsets of channels to find the best one, using 5-fold cross validation.

In this study, we applied SVM to real hemodynamic data obtained in Shimamura et al. (2012, under review) and tried to classify the data into two groups: ADHD participants and ASD participants. We exhaustively searched the optimal subset among all subsets of channels and evaluated the classification accuracy using 5-fold cross validation. To compare the effectiveness of SVM with a standard method, we applied two feature selection methods: Lasso (Tibshirani, 1996) and sparse logistic regression (SLR) (Yamashita et al., 2008), and a channel-wise *t*-test, which is one of the most popular statistical analysis to find an activated channel in NIRS measurement.

METHOD AND ANALYSIS

PARTICIPANTS

The participants in this study were nine boys with ADHD (three ADHD-only and six with comorbid ADHD+ASD; mean age = 9y9m, *SD* = 1y7m) and eight boys with ASD (two ASD-only and six with comorbid of ASD+ADHD; mean age = 9y9m, *SD* = 1y4m). Six ADHD boys received methylphenidate, and one ADHD boy received atomoxetine. The mean score of ADHD-Rating Scale was 34.2 (range = 11–52; *SD* = 13.9) in the ADHD group, 26.6 (range = 16–38; *SD* = 7.6) in ASD group. The two-tailed two-sample *t*-test did not show significant difference in their ADHD-RS scores [$t_{(15)} = 1.375$, *ns*]. These two diagnostic groups were not different in their age or sex. All diagnoses were based on the DSM-IV-TR (American Psychiatric Association, 2000) and were made by a pediatric neurologist (Ryoichi Sakuta).

PROTOCOL

During the measurement, participants observed the visual stimuli presented on the monitor. A single trial was comprised of a baseline period and test period. Each trial started with a baseline period during which the participants fixated on a black dot displayed on the monitor at the rate of 1 Hz. The duration of the baseline period was at least 20 s. Following this, a test period began. During the test period, the image of the child's mother's face or that of an unknown female face was presented. Either the mothers' face or the unknown female face appeared successively 10 times at the rate of 1 Hz. The duration of test trial was 10 s. The mothers' face and the unknown face were presented alternatively. Each participant performed five trials. In this study, we analyzed the hemodynamic data only from when the boys were passively looking at their mother's face.

The study was approved by the Ethical Committee of the Dokkyo Medical University Koshigaya Hospital and by the Ethical Committee of Chuo University. Written informed consent was

obtained from the participants and their parent. The experiments were conducted according to the Declaration of Helsinki.

FUNCTIONAL NIRS RECORDING

We used a Hitachi ETG-4000 system (Hitachi Medical, Chiba, Japan), which can record from 24 channels simultaneously, with 12 channels for the right temporal area and 12 for the left. This instrument generates two wavelengths of NIR (695 and 830 nm) and measures the time courses of the levels of oxyhemoglobin (oxy-Hb), deoxyhemoglobin (deoxy-Hb), and their sum (total hemoglobin: total-Hb) with 0.1 s time resolution. Based on the previous study, which showed that the oxy-Hb change reflects the task-related neural activity more reliably than deoxy-Hb or total-Hb (Strangman et al., 2003; Cui et al., 2011), we analyzed only the oxy-Hb concentration.

The probes were set on the child's scalp at the bilateral temporal area centered at T5 and T6 according to the International 10–20 system (Jasper, 1958) (**Figure 1**). When the probes were positioned, the experimenter checked to see if the fibers were touching the child's scalp correctly. The Hitachi ETG-4000 system automatically detects whether the contact can adequately measure the emerging photons for each channel. All the trials were rejected from the analysis if adequate contact between the fibers and the child's scalp couldn't be achieved because of hair interference.

PREPROCESSING OF DATA

The raw data on oxy-Hb concentrations from each channel were digitally band-pass-filtered at 0.02–1.0 Hz to remove any noise due to heartbeat pulsations and any longitudinal signal drift.

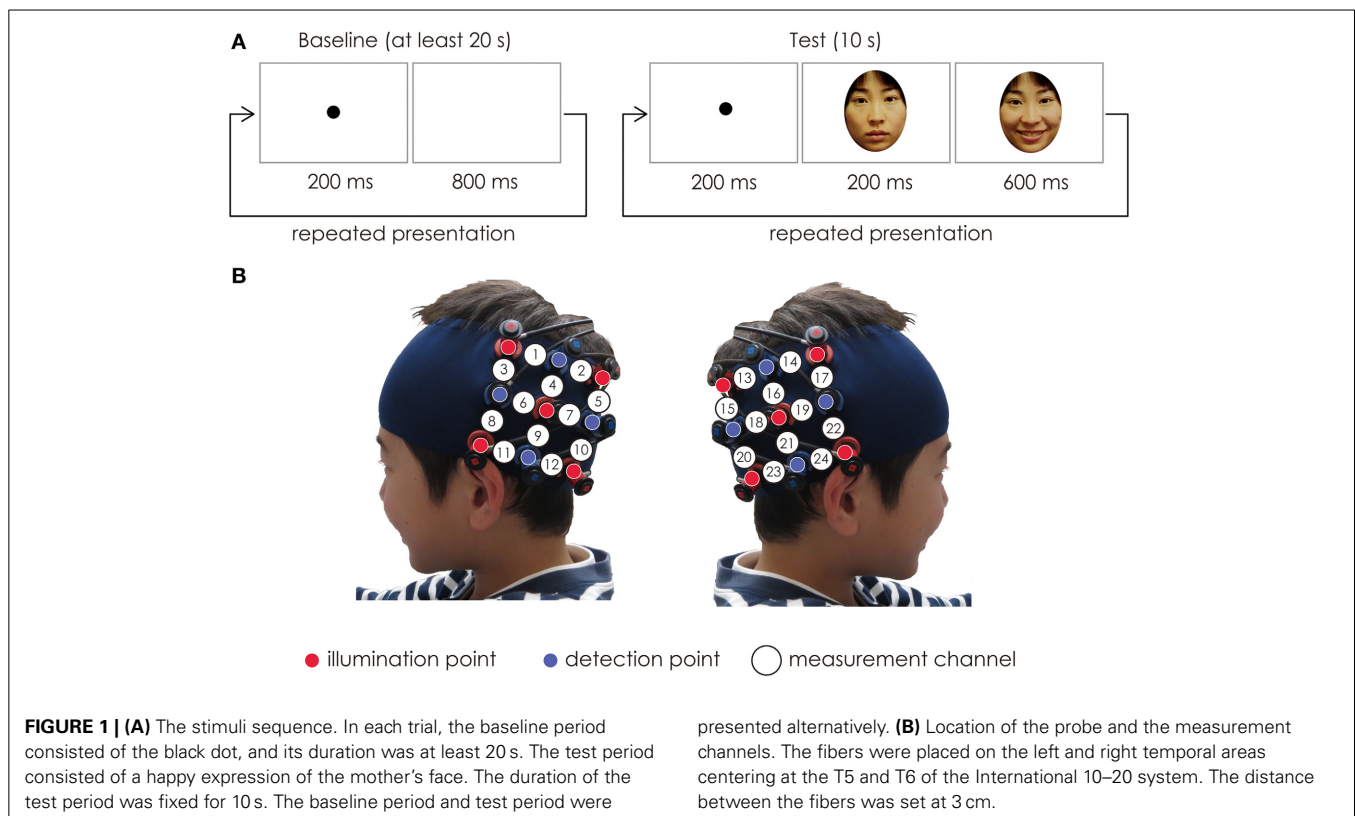
Although the raw NIRS data were originally relative values, and could not be compared directly across subjects or channels, the normalized data such as the Z-score could be averaged regardless of the unit (Schroeter et al., 2003; Shimada and Hiraki, 2006). This calculation of the Z-score is a reliable analysis for changes in concentration in the children's brains since the analysis is independent of the differential path length factor (DPF).

In order to show the relative change of oxy-Hb concentration during the task period, we standardized the raw data into Z-scores based on the baseline period ahead of the task period. We calculated the Z-scores of oxy-Hb in a time series of 0.1 s time resolutions from 3 s before the test period onset to the test period offset (Ichikawa et al., 2010; Kobayashi et al., 2011, 2012). The Z-score at each time point can indicate the deviation of hemodynamic response during the presentation of faces from the "baseline." The "baseline" for calculating the Z-score was a period of 3 s immediately before the beginning of the each test period, which reflects the activation during the observation of the blank and fixation points. The Z-scores were calculated using the following formula:

$$d = (x_{test} - m_{baseline}) / s \quad (1)$$

x_{test} represents the raw data [mM mm] at each time point during the test period and $m_{baseline}$ represents the mean of the raw data during the baseline period. s represents the SD of the raw data during the baseline period.

Consistent with a previous study (Boynton et al., 1996) and our previous studies using NIRS (Ichikawa et al., 2010; Nakato et al., 2011), we found that a response peak lags a few seconds



behind stimulus onset. Therefore, we performed the following analyses against the mean Z-scores from 3 to 10 s after the face stimulus onset.

The Z-scores were calculated separately for each trial. In this study, we eliminated the trials from further analysis if all of 24 channels were not completely recorded due to hair interference or motion artifact. Finally, the number of data we obtained was 50: 25 data from 6 participants of the ADHD group and 25 data from 8 participants of the ASD group. The mean number of valid trials was 4.3 ($SD = 2.7$) per child with ADHD and 3.1 ($SD = 1.3$) per child with ASD.

SUPPORT VECTOR MACHINE (SVM)

We trained the SVM to discriminate data by considering to which diagnosis group the owner of the data belongs. We used the mean Z-scores of hemodynamic activities and the diagnosis group of participants as inputs and outputs of the SVM, respectively.

SVMs are state-of-the-art models for classification with a high generalization capability (Vapnik, 1982; Bennett and Mangasarian, 1992; Cortes and Vapnik, 1995). Given input data, an SVM classifies them into two classes. An SVM learns the relationship between the input data and their classes from the training samples. It also predicts the class of unknown data.

$$\{(x_i, t_i) \mid x_i \in \mathbb{R}^D, t_i \in \{+1, -1\}\}_{i=1}^N, \quad (2)$$

where x_i is a D-dimensional feature vector, and t_i is a class label of x_i . In this study, the mean Z-scores were calculated for 24 channels (or 12 for hemispheric analysis) for each trial, and were used as a 24-dimensional (or 12-dimensional) feature vector. We had 50 valid trial data sets and 50 feature vectors input to the SVM. We related ADHD participants to $t_i = +1$ and ASD participants to $t_i = -1$.

Given this data set, an SVM finds a hyperplane in the feature vector space that separates the samples into two groups. The obtained hyperplane is called a decision boundary. New samples are then classified according to which side of the boundary they belong. Classifying samples using a decision boundary is a common practice for other linear classifiers such as logistic regression. What characterizes SVMs is margin maximizing. The margin is the distance between the decision boundary and the closest sample to the decision boundary. By maximizing the margin, SVMs are capable of accurately predicting the classes of new samples. The decision boundary is expressed as a linear equation as follows:

$$y(x) = w^T x + b = 0 \quad (3)$$

Here w is a weight vector. We want to find w and b that satisfy $y(x_i) > 0$ for $t_i = 1$ and $y(x_i) < 0$ for $t_i = -1$, that is, $ty(x) > 0$ for all samples (Vapnik, 1982). The larger value of w indicates that the more corresponding x contributes to classify data into two groups. A positive sign of w indicates that the larger value of a corresponding x raises the possibility of $t_i = 1$, while a negative sign of w indicates that the larger value of a corresponding x raises the possibility of $t_i = -1$.

CROSS VALIDATION

Cross validation (CV) is a method for evaluating how the ability of a learning machine such as an SVM is generalized to the unknown data that are not used in the training (Kohavi, 1995). In the CV, the data set is divided into two parts. One part is used for the training of the machine, and the other part is used for testing the machine's ability. This training and testing procedure is repeated using different partitioning. The CV is effective when the number of the data is limited.

We explain the K -fold CV for the SVM below. First, we divide the data set into K parts C_1, \dots, C_K . For each $k = 1, \dots, K$, we train the SVM using the data other than the k -th part C_k . We denote by $y \setminus_k(x)$ the decision boundary. Then, by using this boundary $y \setminus_k(x)$, we predict the classes of the data in C_k , and compare them against the true class labels t . We repeat this procedure for every $k = 1, \dots, K$, and calculate the following cross validation error (CVE):

$$\text{CVE} = \frac{1}{N} \sum_{k=1}^K \sum_{i \in C_k} L(t_i, y \setminus_k(x_i)), \quad (4)$$

$$L(t, y(x)) = \begin{cases} 1 & (ty(x) > 0) \\ 0 & (ty(x) < 0) \end{cases} \quad (5)$$

$L(t, y(x))$ indicates whether the prediction is correct or not. When the prediction is correct, $L(t, y(x)) = 0$, and when the prediction is incorrect, $L(t, y(x)) = 1$. CVE represents the ratio of the number of incorrectly predicted data to the total number of data. If the CVE is small, the generalization capability of the SVM is high. The classification accuracy is defined using the following formula:

$$(1 - \text{CVE}) \times 100 \quad (6)$$

We used a 5-fold CV. In this study, we used three data sets; (1) $D = 12$ (12 channels placed on the right temporal area), (2) $D = 12$ (12 channels placed on the left temporal area), and (3) $D = 24$ (24 channels placed on the bilateral temporal areas). The above sets of (1) and (2) were subsets of (3).

We selected a subset A of D features and set $x_{i_A} := (x_{i_d})_{d \in A} \in \mathbb{R}^{|A|}$. We then applied the 5-fold CV to the data set $\{(x_{i_S}, t_i)\}_{i=1}^N$ and calculated the CVE. We carried out this process for all $(2^D - 1)$ subsets.

Our aim was to find the optimal subsets that classify the data into two groups most correctly. In other words, we wanted to find the combination of NIRS measurement channels that distinctively respond to the experimental stimuli depending on which diagnostic groups the data belong.

RESULTS

We used the mean Z-scores of hemodynamic response obtained from each channel and the diagnostic group of participants as inputs and outputs for the SVM, respectively. We trained the SVM to find the optimal subset in the data set of channels in order to classify the data into two diagnostic groups. For each subset, we evaluated its classification accuracy using 5-fold CV.

We conducted the exhaustive search of three data sets of Z-scores of hemodynamic data: (1) the 12-dimension dataset

obtained from the right 12 channels, (2) the 12-dimension dataset obtained from the left 12 channels, and (3) the 24-dimension dataset obtained from the bilateral 24 channels. For the 12-dimension datasets [(1) and (2)], SVM training and calculation of the CVE was repeated with $2^{12} - 1 = 4095$ subsets. For the 24-dimension dataset (3), SVM training and calculation of the CVE was repeated with $2^{24} - 1 = 16,777,215$ subsets.

RESULTS OF SVM CLASSIFICATION APPLIED ON THE DATA OBTAINED FROM THE RIGHT HEMISPHERE

We found nine subsets to classify the data more correctly into two groups among the 4095 subsets. The measurement channels which comprised those subsets are listed in **Table 1**. The best classification accuracy for these subsets was 70%. **Figure 2A** shows the classification accuracy for all 4095 subsets.

Figure 2B shows the feature subsets corresponding to the classification accuracy represented in **Figure 2A**. **Figure 2B** shows that Ch. 15, 16, and 18 repeatedly appear in the subsets represented on the left side, which have relatively higher classification accuracy. These channels are filled with darker color, indicating these channels have a stronger coefficient than other channels. Furthermore, the colors of these three channels were consistent among all the subsets, respectively. This indicates that their weight vectors have the same sign commonly in all the subsets.

Moreover, **Figure 2B** shows that the red colored cells and blue colored cells are separated. The lower-numbered channels (Ch. 13–17) are constantly red and the higher-numbered channels (Ch. 18, 19, and 23) are constantly blue. This suggests that in the right hemisphere the brain area contributing to classification are separated. On the other hand, other channels were constantly filled with yellowish color. These channels do not have strong weight vectors, thus, they do not contribute to the classification in all of the 4095 subsets.

Figure 3 shows the classification accuracy of the best 50 subsets (**Figure 3A**) and the corresponding feature subsets (**Figure 3B**).

Table 1 | The measurement channels those comprised the subsets with best classification accuracy^a.

Right (70%)	Left (74%)	Bilateral (84%)
15, 16, 18, 24	1, 6, 7, 9, 11	1, 3, 4, 5, 6, 8, 9, 10, 13, 14, 18, 20
15, 16, 17, 18, 24		1, 3, 4, 5, 6, 8, 9, 13, 14, 15, 18, 20, 22, 23
13, 15, 16, 17, 18, 19		
15, 16, 18, 20, 22, 24		
15, 16, 18, 20, 23, 24		
14, 15, 16, 17, 18, 20, 24		
15, 16, 17, 18, 20, 23, 24		
13, 15, 16, 17, 18, 20, 23, 24		
14, 15, 16, 17, 18, 20, 23, 24		

^a The numbers indicated the channel numbers. In the right hemisphere, the number of the best subset was nine. The percentage represented in the parenthesis indicating the best classification accuracy for each data set.

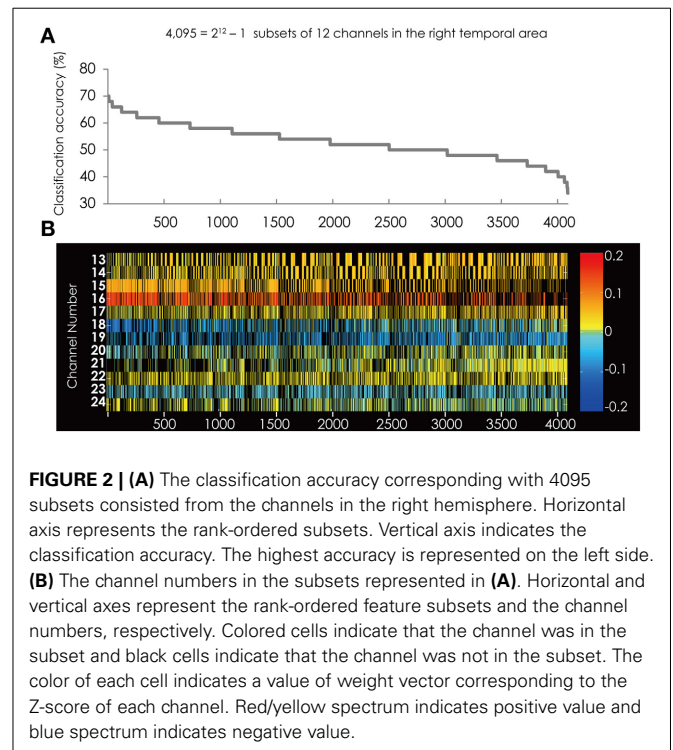
Each cell indicates the weight vector of each channel in the feature subset. Ch. 15, 16, and 18 appear in most of subsets. Ch. 15 was in 49 of the 50 subsets. Ch. 16 was in 45 subsets and Ch. 18 was in 40 subsets, respectively. These three channels were more often used in the best 50 subsets and more effectively contributed to the classification. The positions of these channels are illustrated in **Figure 3C**.

The color of the cells corresponding to Ch. 15 and 16 are red, but Ch. 18 is blue. This tendency is consistent through the best 50 subsets. This result indicated that the greater hemodynamic response in Ch. 15 and 16 more often occurred in ADHD participants than in ASD participants, and those at Ch. 18 were more often found in ASD participants than in ADHD participants.

RESULTS OF SVM CLASSIFICATION APPLIED ON THE DATA FROM THE LEFT HEMISPHERE

The best subset for classifying the data into two groups contained the five channels listed in **Table 1**. This subset had the best classification accuracy of 74%. **Figure 4** shows the classification accuracy for all 4095 subsets (**Figure 4A**) and the corresponding feature subsets (**Figure 4B**). Ch. 5 and 6 constantly appear in the subsets with relative higher classification accuracy and had greater weight value than the other channels. Furthermore, the color of their cells is consistent through all the subsets. This indicates the sign of the weight vector is common in all the subsets.

Compared with the result from the right hemisphere (**Figure 3B**), the red colored cells and blue colored cell are scattered among the channels. This suggests that in the left hemisphere the brain areas contributing to classification are not clearly separated.



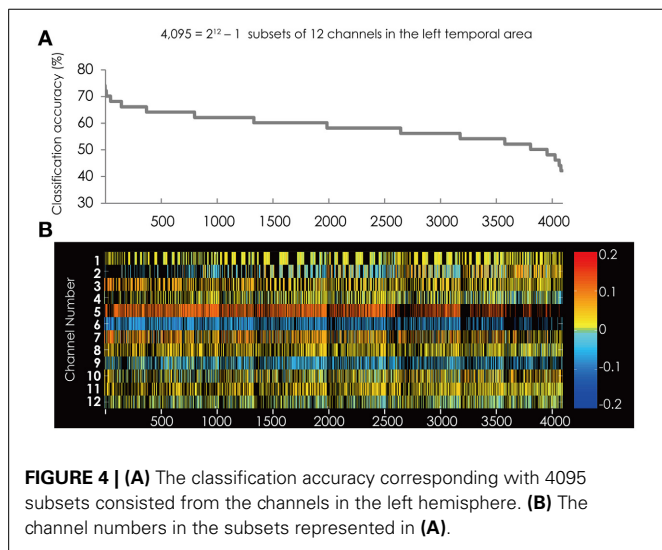
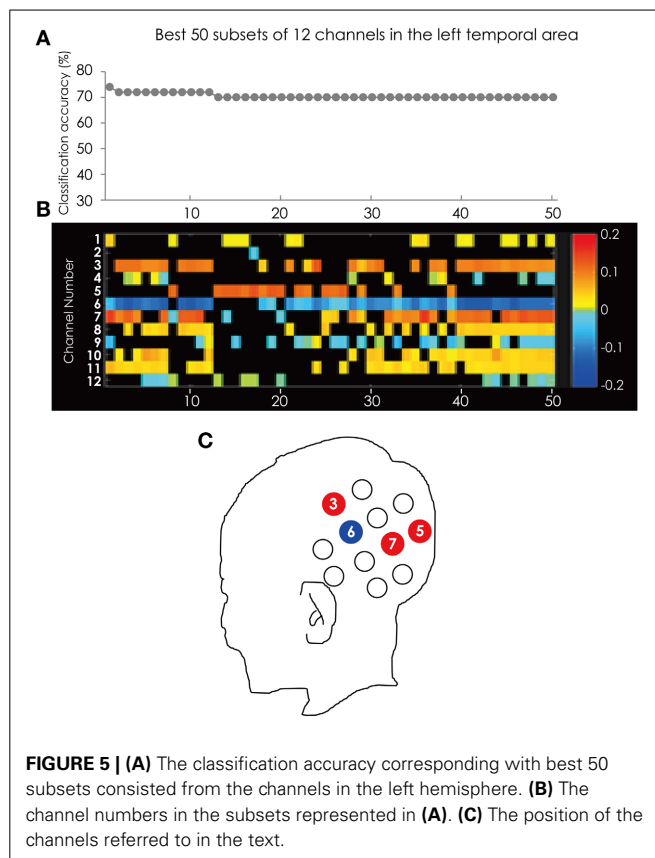
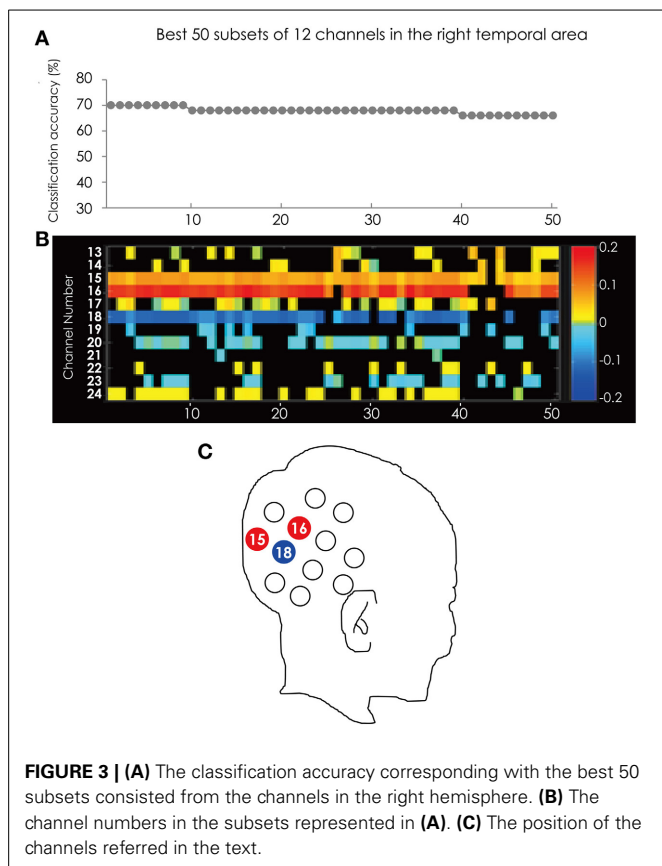


Figure 5 shows the classification accuracy for all 4095 subsets (Figure 5A) and the corresponding feature subsets (Figure 5B). Ch. 6 was most often used in the best 50 subsets. The channel was in 43 of the 50 subsets. Following Ch. 6, 3, and 7 were in 32 subsets. It is worth mentioning that Ch. 5 and 6 appear alternately in the top 17 subsets. The position of these channels is illustrated in Figure 5C.

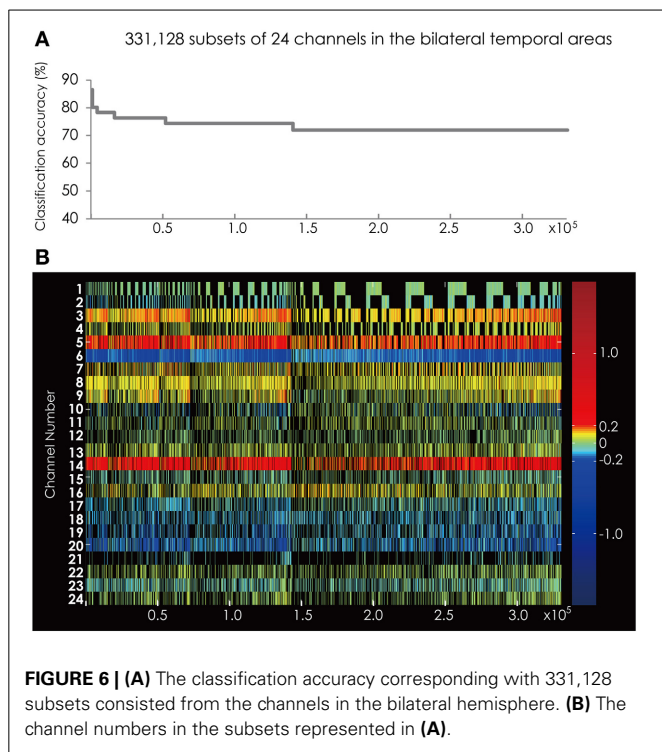
Ch. 3, 5, and 7 are red and only Ch. 6 is blue. This tendency is consistent through the best 50 subsets. This result indicated that at Ch. 3, 5, and 7 ADHD participants showed greater hemodynamic response than ASD participants, and only at Ch. 6 did ASD participants show greater hemodynamic response than ADHD participants.

RESULTS OF SVM CLASSIFICATION APPLIED ON THE DATA FROM BILATERAL HEMISPHERES

We found two subsets for classifying the data more correctly into two groups among the 16,777,215 subsets. The measurement channels that comprised those subsets listed in Table 1. These subsets had the best classification accuracy of 84%. Figure 6A shows the classification accuracy of the subsets with accuracy above 70%. The classification accuracy was higher when using the data set from the bilateral temporal areas than when using data sets from the one-side temporal area. This result indicated that the bilateral temporal areas interacted with each other and that the combined data analysis increased the amount of information about the classification.

Figure 6B shows the feature subsets corresponding to the classification accuracy represented in Figure 6A. As in the analysis on the data sets from each hemisphere, Ch. 6 appeared constantly and was colored blue.

Figure 7 shows the classification accuracy for all 4095 subsets (Figure 7A) and the corresponding feature subsets (Figure 7B). Ch. 6 was in all of the 50 subsets and was most often used in the



best 50 subsets. Ch. 3 and 14 were in 49 subsets. The position of these channels is illustrated in **Figure 7C**.

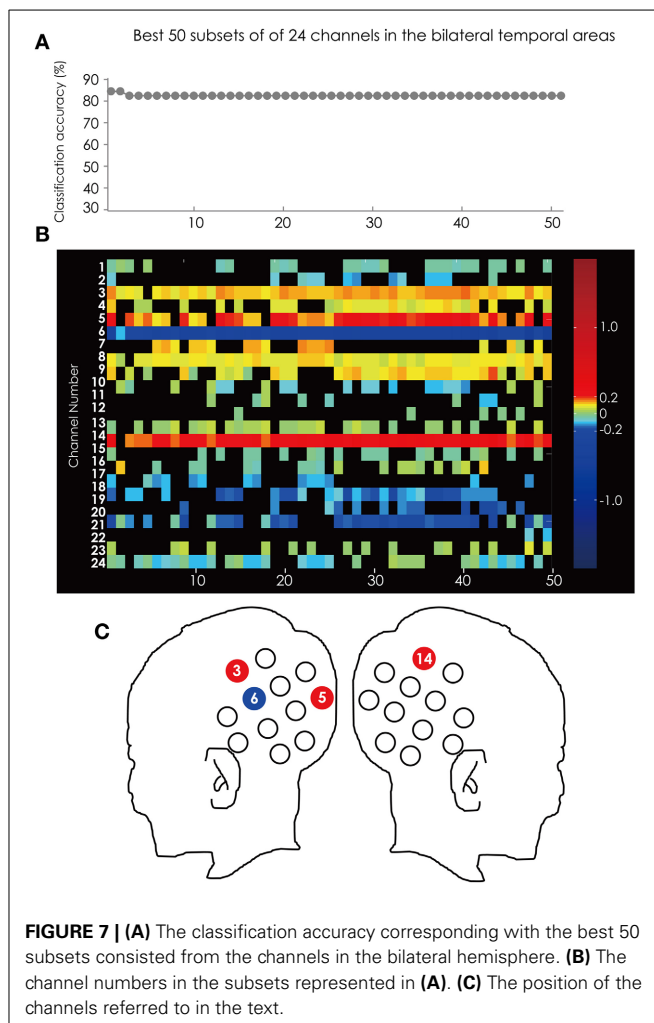
Ch. 3, 5, and 14 are red and only Ch. 6 is blue. This tendency is consistent through the top 50 subsets. This result indicated that the greater hemodynamic response in Ch. 3, 5, and 14 more often occurred in ADHD participants than in ASD participants, and that those at Ch. 6 are more often found in ASD participants than in ADHD participants.

LASSO AND SPARSE LOGISTIC REGRESSION

To demonstrate the effectiveness of the exhaustive search, we compared the exhaustive search with two existing methods, least absolute shrinkage and selection operator (LASSO, Tibshirani, 1996) and SLR (Yamashita et al., 2008). These are classification methods that incorporate feature selection as a part of the process of classifier training. We applied LASSO and SLR to datasets (1)–(3), and evaluated classification accuracies by using 5-fold CV.

By using LASSO, the classification accuracies were (1)57.5 and (2)66%. These accuracies rank (1)1106th (in the top 27.0%) and (2)147th (in the top 3.59%) of the 4095 accuracies obtained by the exhaustive search. The classification accuracy in dataset (3) was 70%, and this ranks 139,815th (in the top 0.833%) of the 16 million accuracies obtained by the exhaustive search.

By using SLR, the classification accuracies were (1)52 and (2)64%. These accuracies rank (1)1978th (in the top 48.3%) and (2)372th (in the top 9.08%) of the 4095 accuracies obtained by the exhaustive search. The classification accuracy in dataset (3) was 66%, and this ranks 698,955th (in the top 4.17%) of the 16 million accuracies obtained by the exhaustive search.



These results showed that, in datasets (2) and (3), the classification accuracies using LASSO and SLR were high and were located near the top of all the accuracies obtained by the exhaustive search. In comparison, in dataset (1) the classification accuracies were low compared to those in datasets (2) and (3), and were in the middle of all the accuracies obtained by the exhaustive search. These findings highlighted the fact that by using LASSO and SLR, the classification accuracy is not always high compared to the best result obtained by the exhaustive search. This fact was not revealed unless performing the exhaustive search. This means that neither LASSO nor SLR could uncover the latent structures relevant to discrimination between ASD and ADHD in the high-dimensional fNIRS data.

CHANNEL-WISE ANALYSIS OF 24 CHANNELS

With the aim of comparing the effectiveness of SVM with a standard method, we also performed a channel-wise *t*-test on each channel. A two-tailed two-sample *t*-test was conducted for the difference of Z-scores between the ADHD participants and ASD participants during the 3–10 s of the test trials. For all 24 channels, a *t*-test was performed. To reduce the risk of a Type I error,

we performed the corrections using the false discovery rate (FDR) (Singh and Dan, 2006).

We did not find any significant difference in the mean Z-scores between ADHD participants and ASD participants. To compare the hemodynamic change with the baseline activation, we conducted a two-tailed one-sample *t*-test against the baseline. The *t*-test was repeated for each of 24 channels by applying the FDR procedure (Singh and Dan, 2006). However, we did not find any significant hemodynamic change from the baseline either in ADHD participants or ASD participants.

DISCUSSION

In this study, we applied SVM to real hemodynamic data obtained in Shimamura et al. (2012, under review) and tried to classify the data into two groups: ADHD participants and ASD participants. We exhaustively searched the optimal subset of measurement channels and evaluated each classification accuracy using 5-fold CV. We showed that the classification accuracy when using the best subset of channels was 84%, while that using all 24 channels was 62%. Additionally, we applied two feature selection methods, LASSO (Tibshirani, 1996) and SLR (Yamashita et al., 2008), which are intended to be applied to high-dimensional data, and confirmed that the best subset of channels was not selected when using these methods. Furthermore, the channel-wise analysis did not find any significant channels that showed distinctive activation for groups.

In the right hemisphere, SVM classification indicated that the greater weight value was consistently in three channels: Ch. 15, 16, and 18. These channels contribute to the classification in opposite directions because of their opposite signs of weight vectors in constricting the decision boundary. The sign of weight vector corresponding to Ch. 15 and 16 were positive, while that corresponding to Ch. 18 was negative. This result indicates that hemodynamic data obtained from Ch. 15 and 16 might increase for ADHD children and those obtained from Ch. 18 might decrease. To classify the input data into an ADHD group or an ASD group, we can focus on only the activity of these three channels, rather than all the measurement channels. (Using only these channels, we can classify the hemodynamic data into an ADHD group and an ASD group with 68% classification accuracy.) Ch. 15 and 16 correspond to the temporal-parietal junction (TPJ). The TPJ is the area responsive to theory of mind (ToM) (Saxe and Kanwisher, 2003) and is involved in familiar face recognition (Gobbini et al., 2004; Gobbini and Haxby, 2007). Although the neural response to a maternal face has not been investigated with school-aged children, previous study with adults (Ramasubbu et al., 2007) demonstrated that a maternal face evoked a stronger and broader hemodynamic activation than did unfamiliar faces. On the other hand, Bartels and Zeki (2004) demonstrated that maternal attachment and romantic love commonly activated the brain's reward system, yet deactivated regions associated with negative emotions, social judgment, and ToM. Based on these previous studies, we can assume that increased hemodynamic response in the TPJ of ADHD participants might be related to the atypical attachment of ADHD children to their mother (Shimamura et al., 2012).

In the left hemisphere, though no channels appeared consistently through the best 50 subsets, some channels appeared

alternately. In the best 17 subsets Ch. 3, 5, 7 appeared, while in the 18th to 28th subsets only Ch. 5 appeared dominantly. Below the 29th subsets, Ch. 6 appeared again and various unstable channels appeared alternately. These channels behave like small "patches." These small-scale activities by some patches reflect the sparse expression of the information in/from the left hemisphere (Kitazono, 2013).

The imbalance of channels that contributes to the classification between the hemispheres might reflect the differential stage of face processing. Previous studies demonstrated that the left hemisphere predominates when faces are processed featurally, whereas the right hemisphere predominates when faces are processed configurally (Koenig and Hillger, 1991; Rossion et al., 2000; Scott and Nelson, 2006). Based on these findings, we can suppose that the greater number of channels in the left hemisphere might input raw (lower level) information and process featural properties such as eyes and mouth independently, whereas the lesser number of channels in the right hemisphere might input the pre-processed information and process the configuration of faces.

The best classification accuracy was obtained by the SVM classification on all 24 channels. The classification accuracy was 84%. We had 50 data, and 42 out of those 50 were correctly classified into the diagnosis group to which the participants belonged. Let us compare with the results of the SVM classification in each hemisphere and in the bilateral hemisphere. Three channels (Ch. 15, 16, and 18) are important in the SVM classification in the right hemisphere, while four channels (Ch. 3, 5, 6, and 7) are important in the left hemisphere. Also, four channels (Ch. 3, 5, 6, and 14) are important in the SVM classification in the bilateral hemisphere. Only three (Ch. 3, 5, 6) out of the seven channels (Ch. 3, 5, 6, 7, 15, 16, 18) are common between the one-sided hemisphere and the bilateral hemisphere analysis. Moreover, one channel (Ch. 14) is not important in the right hemisphere analysis. We found that the discrepancies in the results of the SVM classification in the bilateral hemisphere that the channels used in the optimal subset were not consistent with those used in the optimal subset in each hemisphere. It would be logical that different input results in different output (Kitazono, 2013).

In this study, we aimed to classify the hemodynamic data into two distinct participant groups: ADHD participants and ASD participants. We exhaustively searched the optimal subset of fNIRS measurement channel and evaluated the classification accuracy using 5-fold CV. We successfully found the optimal subset for the classification of the real hemodynamic data with 84% accuracy. We can conclude that SVM and exhaustive search provides an effective method for hemodynamic data classification obtained from multichannel NIRS measurement.

ACKNOWLEDGMENTS

This study was supported by Grant-in-Aid for Scientific Research on Innovative Areas, "Face perception and recognition" from MEXT KAKENHI (20119002 to Masami K. Yamaguchi, 23119708 to Masato Okada); Grant-in-Aid for Scientific Research on Innovative Areas, "Sparse Modeling" from JSPS KAKENHI (25120009 to Masato Okada and Kenji Nagata, 26120529 to Hiroko Ichikawa); Grant-in-Aid for Scientific Research on Innovative Areas, "Exploring the Limits of Computation (ELC)"

from JSPS KAKENHI (25106506 to Kenji Nagata); Grant-in-Aid for Scientific Research (A) from JSPS KAKENHI (20240020 to Masato Okada); Grant-in-Aid for Young Scientists (B) from JSPS KAKENHI (22700230 to Kenji Nagata) and a Grant-in-Aid for JSPS Fellows from JSPS KAKENHI (10J06155 to Jun Kitazono, 24 7809 to Hiroko Ichikawa).

REFERENCES

- American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders, 4th Edn.* Text Revision. Arlington, VA: American Psychiatric Association.
- Bartels, A., and Zeki, S. (2004). The neural correlates of maternal and romantic love. *Neuroimage* 21, 1155–1166. doi: 10.1016/j.neuroimage.2003.11.003
- Bennett, K. P., and Mangasarian, O. L. (1992). Robust linear programming discrimination of two linearly separable sets. *Optim. Methods Softw.* 1, 23–34. doi: 10.1080/10556789208805504
- Boynton, G., Engel, S., Glover, G., and Heeger, D. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.* 16, 4207–4221.
- Carlsson, J., Lagercrantz, H., Olson, L., Printz, G., Bartocci, M., and Jakobcarlssonkic, J. C. (2008). Activation of the right fronto-temporal cortex during maternal facial recognition in young infants. *Acta Paediatr.* 97, 1221–1225. doi: 10.1111/j.1651-2227.2008.00886.x
- Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018
- Cui, X., Bray, S., Bryant, D. M., Glover, G. H., and Reiss, A. L. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *Neuroimage* 54, 2808–2821. doi: 10.1016/j.neuroimage.2010.10.069
- Cui, X., Bray, S., and Reiss, A. L. (2010). Speeded Near Infrared Spectroscopy (NIRS) response detection. *PLoS ONE* 5:e15474. doi: 10.1371/journal.pone.0015474
- Dawson, G., Carver, L., Meltzoff, A. N., Panagiotides, H., McPartland, J., and Webb, S. J. (2002). Neural correlates of face and object recognition in young children with autism spectrum disorder, developmental delay, and typical development. *Child Dev.* 73, 700–717. doi: 10.1111/1467-8624.00433
- Ernst, L., Schneider, S., Ehlis, A.-C., and Fallgatter, A. (2012). Review: functional near infrared spectroscopy in psychiatry: a critical review. *J. Near Infrared Spectrosc.* 20, 93–105. doi: 10.1255/jnirs.970
- Ferrari, M., and Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *Neuroimage* 62, 921–935. doi: 10.1016/j.neuroimage.2012.03.049
- Fukuda, M. (2009). *Psychiatric Disorders and NIRS*. Bunkyo-ku: Nakayama Shoten Co. Ltd.
- Gervain, J., Mehler, J., Werker, J. F., Nelson, C. A., Csibra, G., Lloyd-Fox, S., et al. (2011). Near-infrared spectroscopy: a report from the McDonnell infant methodology consortium. *Dev. Cogn. Neurosci.* 1, 22–46. doi: 10.1016/j.dcn.2010.07.004
- Gobbini, M. I., and Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia* 45, 32–41. doi: 10.1016/j.neuropsychologia.2006.04.015
- Gobbini, M. I., Leibenluft, E., Santiago, N., and Haxby, J. V. (2004). Social and emotional attachment in the neural representation of faces. *Neuroimage* 22, 1628–1635. doi: 10.1016/j.neuroimage.2004.03.049
- Gottemukkula, V., and Derakhshani, R. (2011). “Classification-guided feature selection for NIRS-based BCI,” in *5th International IEEE/EMBS Conference on Neural Engineering*, 72–75. Available online at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5910491>
- Guyon, I., and Elisseeff, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res.* 3, 1157–1182.
- Ichikawa, H., Kanazawa, S., Yamaguchi, M. K., and Kakigi, R. (2010). Infant brain activity while viewing facial movement of point-light displays as measured by near-infrared spectroscopy (NIRS). *Neurosci. Lett.* 482, 90–94. doi: 10.1016/j.neulet.2010.06.086
- Jasper, H. H. (1958). The ten twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol.* 10, 371–375.
- Kitazono, J. (2013). *A Theoretical Study on Information Processing in Inferior Temporal Cortex*. Unpublished doctoral dissertation, The University of Tokyo, Tokyo.
- Kobayashi, M., Otsuka, Y., Kanazawa, S., Yamaguchi, M. K., and Kakigi, R. (2012). Size-invariant representation of face in infant brain: an fNIRS-adaptation study. *Neuroreport* 23, 984–988. doi: 10.1097/WNR.0b013e32835a4b86
- Kobayashi, M., Otsuka, Y., Nakato, E., Kanazawa, S., Yamaguchi, M. K., and Kakigi, R. (2011). Do infants represent the face in a viewpoint-invariant manner? Neural adaptation study as measured by near-infrared spectroscopy. *Front. Hum. Neurosci.* 5:153. doi: 10.3389/fnhum.2011.00153
- Koenig, O., and Hillger, L. A. (1991). Separable mechanisms in face processing: evidence from hemispheric specialization. *J. Cogn. Neurosci.* 3, 42–58. doi: 10.1162/jocn.1991.3.1.42
- Kohavi, R. (1995). “A Study of cross-validation and bootstrap for accuracy estimation and model selection,” in *Proceedings of the 14th International Joint Conference on Artificial Intelligence 2 (Montreal)*, 1137–1143.
- Liu, H., and Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Trans. Knowl. Data Eng.* 17, 491–502. doi: 10.1109/TKDE.2005.66
- Lloyd-Fox, S., Blasi, A., and Elwell, C. E. (2010). Illuminating the developing brain: the past, present and future of functional near infrared spectroscopy. *Neurosci. Biobehav. Rev.* 34, 269–284. doi: 10.1016/j.neubiorev.2009.07.008
- Monden, Y., Dan, H., Nagashima, M., Dan, I., Tsuzuki, D., Kyutoku, Y., et al. (2012). Right prefrontal activation as a neuro-functional biomarker for monitoring acute effects of methylphenidate in ADHD children: an fNIRS study. *Neuroimage Clin.* 1, 131–140. doi: 10.1016/j.nicl.2012.10.001
- Nakato, E., Otsuka, Y., Kanazawa, S., Yamaguchi, M. K., and Kakigi, R. (2011). Distinct differences in the pattern of hemodynamic response to happy and angry facial expressions in infants—a near-infrared spectroscopic study. *Neuroimage* 54, 1600–1606. doi: 10.1016/j.neuroimage.2010.09.021
- Negoro, H., Sawada, M., Iida, J., Ota, T., Tanaka, S., and Kishimoto, T. (2010). Prefrontal dysfunction in attention-deficit/hyperactivity disorder as measured by near-infrared spectroscopy. *Child Psychiatry Hum. Dev.* 41, 193–203. doi: 10.1007/s10578-009-0160-y
- Ramasubbu, R., Masalovich, S., Peltier, S., Holtzheimer, P. E., Heim, C., and Mayberg, H. S. (2007). Neural representation of maternal face processing: a functional magnetic resonance imaging study. *Can. J. Psychiatry* 52, 726–734.
- Rossion, B., Dricot, L., Devolder, A., Bodart, J. M., Crommelinck, M., de Gelder, B., et al. (2000). Hemispheric asymmetries for whole-based and part-based face processing in the human fusiform gyrus. *J. Cogn. Neurosci.* 12, 793–802. doi: 10.1162/089892900562606
- Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in theory of mind. *Neuroimage* 19, 1835–1842. doi: 10.1016/S1053-8119(03)00230-1
- Schecklmann, M., Romanos, M., Bretscher, F., Plichta, M. M., Warnke, A., and Fallgatter, A. J. (2010). Prefrontal oxygenation during working memory in ADHD. *J. Psychiatr. Res.* 44, 621–628. doi: 10.1016/j.jpsychires.2009.11.018
- Schroeter, M. L., Zysset, S., Kruggel, F., and von Cramon, D. Y. (2003). Age dependency of the hemodynamic response as measured by functional near-infrared spectroscopy. *Neuroimage* 19, 555–564. doi: 10.1016/S1053-8119(03)00155-1
- Scott, L. S., and Nelson, C. A. (2006). Featural and configural face processing in adults and infants: a behavioral and electrophysiological investigation. *Perception* 35, 1107–1128. doi: 10.1068/p5493
- Shimada, S., and Hiraki, K. (2006). Infant’s brain responses to live and televised action. *Neuroimage* 32, 930–939. doi: 10.1016/j.neuroimage.2006.03.044
- Shimamura, K., Ichikawa, H., Nakato, E., Yamaguchi, M. K., Kanazawa, S., and Sakuta, R. (2012). “NIRS response to visual stimulation by facial images to assess autistic aspects in AD/HD,” in *Proceedings of the 54th Annual Meeting of the Japanese Society of Child Neurology*, S206 (Shinjuku).
- Singh, A. K., and Dan, I. (2006). Exploring the false discovery rate in multichannel NIRS. *Neuroimage* 33, 542–549. doi: 10.1016/j.neuroimage.2006.06.047
- Sitaram, R., Zhang, H., Guan, C., Thulasidas, M., Hoshi, Y., Ishikawa, A., et al. (2007). Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain—computer interface. *Neuroimage* 34, 1416–1427. doi: 10.1016/j.neuroimage.2006.11.005
- Strangman, G., Franceschini, M. A., and Boas, D. A. (2003). Factors affecting the accuracy of near-infrared spectroscopy concentration calculations for focal changes in oxygenation parameters. *Neuroimage* 18, 865–879. doi: 10.1016/S1053-8119(03)00021-1
- Taurines, R., Schwenck, C., Westerwald, E., Sachse, M., Siniatchkin, M., and Freitag, C. (2012). ADHD and autism: differential diagnosis or overlapping

- traits? A selective review. *ADHD Atten. Defic. Hyperact. Disord.* 4, 115–139. doi: 10.1007/s12402-012-0086-2
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* 58, 267–288.
- Thorell, L. B., Rydell, A., and Bohlin, G. (2012). Parent-child attachment and executive functioning in relation to ADHD symptoms in middle childhood. *Attach. Hum. Dev.* 14, 517–532. doi: 10.1080/14616734.2012.706396
- Tye, C., Mercure, E., Ashwood, K. L., Azadi, B., Asherson, P., Johnson, M. H., et al. (2013). Neurophysiological responses to faces and gaze direction differentiate children with ASD, ADHD and ASD+ADHD. *Dev. Cogn. Neurosci.* 5C, 71–85. doi: 10.1016/j.dcn.2013.01.001
- van der Meer, J. M., Oerlemans, A. M., van Steijn, D. J., Lappenschaar, M. G., de Sonnevill, L. M., and Buitelaar, J. K. (2012). Are autism spectrum disorder and attention-deficit/hyperactivity disorder different manifestations of one overarching disorder? Cognitive and symptom evidence from a clinical and population-based sample. *J. Am. Acad. Child Adolesc. Psychiatry* 51, 1160–1172. doi: 10.1016/j.jaac.2012.08.024
- Vapnik, V. N. (1982). *Estimation of Dependences Based on Empirical Data*. New York, NY: Springer.
- Weber, P., Lütsch, J., and Fahrenstich, H. (2005). Cerebral hemodynamic changes in response to an executive function task in children with attention-deficit hyperactivity disorder measured by near-infrared spectroscopy. *Dev. Behav. Pediatr.* 26, 105–111. doi: 10.1097/00004703-200504000-00005
- Weston, J., Mukherjee, S., Chapelle, O., Pontil, M., Poggio, T., and Vapnik, V. (2001). “Feature Selection for SVMs,” in *Advances in Neural Information Processing Systems, 13*, eds T. Leen, T. Dietterich, and V. Tresp (Cambridge, MA: MIT Press), 668–674.
- Yamashita, O., Sato, M., Yoshioka, T., Tong, F., and Kamitani, Y. (2008). Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns. *Neuroimage* 42, 1414–1429. doi: 10.1016/j.neuroimage.2008.05.050
- Yerys, B. E., Wallace, G. L., Sokoloff, J. L., Shook, D. A., James, J. D., and Kenworthy, L. (2009). Attention deficit/hyperactivity disorder symptoms moderate cognition and behavior in children with autism spectrum disorders. *Autism Res.* 2, 322–333. doi: 10.1002/aur.103.
- Yoshida, Y., and Uchiyama, T. (2004). The clinical necessity for assessing Attention Deficit/Hyperactivity Disorder (AD/HD) symptoms in children with high-functioning Pervasive Developmental Disorder (PDD). *Eur. Child Adolesc. Psychiatry* 13, 307–314. doi: 10.1007/s00787-004-0391-1

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 05 March 2014; accepted: 13 June 2014; published online: 02 July 2014.

Citation: Ichikawa H, Kitazono J, Nagata K, Manda A, Shimamura K, Sakuta R, Okada M, Yamaguchi MK, Kanazawa S and Kakigi R (2014) Novel method to classify hemodynamic response obtained using multi-channel fNIRS measurements into two groups: exploring the combinations of channels. *Front. Hum. Neurosci.* 8:480. doi: 10.3389/fnhum.2014.00480

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2014 Ichikawa, Kitazono, Nagata, Manda, Shimamura, Sakuta, Okada, Yamaguchi, Kanazawa and Kakigi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.