

Received April 21, 2020, accepted May 4, 2020, date of publication May 12, 2020, date of current version May 26, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2994226

A Novel Multimodal Approach for Hybrid Brain–Computer Interface

ZHE SUN^{ID1}, ZIHAO HUANG^{ID2}, FENG DUAN^{ID2}, AND YU LIU³

¹Computational Engineering Applications Unit, Head Office for Information Systems and Cybersecurity, RIKEN, Wako 351-0198, Japan

²College of Artificial Intelligence, Nankai University, Tianjin 300350, China

³Key Laboratory of Exercise and Health Sciences of Ministry of Education, Shanghai University of Sport, Shanghai 200438, China

Corresponding author: Feng Duan (duanf@nankai.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFE0129700, in part by the National Natural Science Foundation of China (Key Program) under Grant 11932013, in part by the National Natural Science Foundation of China under Grant 61673224, and in part by the Tianjin Natural Science Foundation for Distinguished Young Scholars under Grant 18JCJQJC46100.

ABSTRACT Brain-computer interface (BCI) technologies have been widely used in many areas. In particular, non-invasive technologies such as electroencephalography (EEG) or near-infrared spectroscopy (NIRS) have been used to detect motor imagery, disease, or mental state. It has been already shown in literature that the hybrid of EEG and NIRS has better results than their respective individual signals. The fusion algorithm for EEG and NIRS sources is the key to implement them in real-life applications. In this research, we propose three fusion methods for the hybrid of the EEG and NIRS-based brain-computer interface system: linear fusion, tensor fusion, and p th-order polynomial fusion. Firstly, our results prove that the hybrid BCI system is more accurate, as expected. Secondly, the p th-order polynomial fusion has the best classification results out of the three methods, and also shows improvements compared with previous studies. For a motion imagery task and a mental arithmetic task, the best detection accuracy in previous papers were 74.20% and 88.1%, whereas our accuracy achieved was 77.53% and 90.19%. Furthermore, unlike complex artificial neural network methods, our proposed methods are not as computationally demanding.

INDEX TERMS Brain-computer interface, electroencephalography, near-infrared spectroscopy, multimodal signal, polynomial fusion.

I. INTRODUCTION

Brain-computer interface (BCI) is an important tool for the detection of signal patterns in brain activity. It has been used widely in many areas, such as in robotics control, workload detection or brain-disease detection [1]–[7]. BCI technology is usually divided into invasive BCI and non-invasive BCI. Invasive BCI requires the sensors to record brain activities from within the skull. Non-invasive BCI, on the other hand, records brain signals using sensors placed on the scalp, and it is undoubtedly a much safer technology and easier to use [8]. However, signals from non-invasive BCI sensors are usually full of noise, from the subjects' unconscious eye movement or ambient noise, for example. Finding a way to use non-invasive sensors to detect brain signals accurately is still a big challenge [9].

Transitional non-invasive BCI systems use either electroencephalography (EEG) sensors, near-infrared

spectroscopy (NIRS) sensors or Magnetic Resonance Imaging (MRI) to record brain activities [10]–[13]. Compared with the MRI equipment, EEG and NIRS equipments are low-cost and smaller in size. Both EEG and NIRS have been implemented in many real time BCI applications [12].

EEG equipment consists of metal electrodes placed on the scalp to record electrical signals [14]. The electrodes record the activity of the surrounding neurons. For the EEG-based BCI system, classical feature extraction methods such as common spatial patterns (CSP) [15], power spectrum density (PSD) and auto-regressive modeling (AR) [16] have been previously proposed to analyze and localize the EEG patterns and activated brain area of motor imagery (MI) and mental arithmetic (MA) [17]. Then, in regards to classifiers, several machine learning classifiers such as k-nearest neighbours (KNN) [18], support vector machine (SVM) [19], [20], and linear discriminant analysis (LDA) [21] are used in classifying these proposed EEG features. In recent years, with the wide application of deep learning, many studies have already shown that convolution neural network (CNN)

The associate editor coordinating the review of this manuscript and approving it for publication was Junhua Li^{ID}.

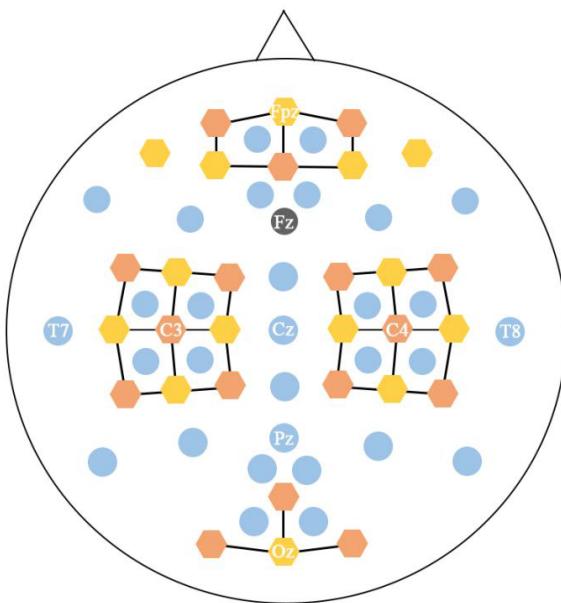


FIGURE 1. Locations of the EEG electrodes (solid circles), NIRS sources (yellow hexagons) and detectors (orange hexagons).

exhibits good performance for EEG processing or NIRS processing [22], [23]. CNN and CNN-based models, such as EEGNet [24] or convolutional neural networks-stacked autoencoders (CNN-SAE) [25], have obtained remarkable MI classification results. Moreover, some researchers have developed an LSTM-based framework [26] to extract essential features of time-varying EEG signals to classify motor imagery signals.

NIRS-based BCI systems, generally used to measure the hemodynamic signals from target regions of the brain [27], can also analyze oxy-hemoglobin NIRS (oxy-NIRS) or deoxy-hemoglobin NIRS (deoxy-NIRS) concentrations in order to localize and classify the brain activity [28]. Generally, when a specific brain area becomes more active, energy metabolism increases, leading to an increased oxygen consumption and increased levels of carbon dioxide in the area. Then, oxy-NIRS signals will decrease while deoxy-NIRS signals will increase. For the signal processing or pattern classification of NIRS, transitional algorithms such as SVM or artificial neural networks are used [29].

Instead of using EEG or NIRS methods individually, using them simultaneously provides a multimodal hybrid BCI. It has been repeatedly proven in literature that multimodal BCI systems have a better accuracy and stability than the BCIs based on a single modality [30]–[32]. For instance, Jaeyoung Shin *et al.* presented an open source benchmark for a hybrid EEG-NIRS fusion BCI system [12], [30] which also uses a linear discriminant analysis classifier for the multimodal data classifications.

Non-linear feature fusion is commonly used in deep learning, and there are many ways to achieve it. For instance, in [33] a bilinear method was used to fuse two feature vectors, which can be thought of as the 2-order tensor product fusion.

Even further, a trilinear method was used in [34] to fuse three modality features. In addition, some tensor-based methods have been proposed to solve multimodal fusion problems such as Visual Question Answering: for example, [35] proposed a Deep Attention Neural Tensor Network which can discover the joint correlations over images, questions and answers with tensor-based representations; [36] proposed a low-rank multimodal fusion method which performs multimodal fusion using low-rank tensors to improve efficiency; finally, [37] proposed a polynomial tensor pooling block for integrating multimodal features by considering high-order moments.

In this study, a novel approach for the hybrid BCI system is proposed. Firstly, CNN was used to get the feature vectors from the EEG, oxy-NIRS, and deoxy-NIRS signals. A single-modal data was calculated to compare with hybrid-system based classification results. Then, three fusion methods were used to process the multimodal feature vectors: linear fusion classification method, tensor fusion classification method and the p th-order polynomial fusion classification method. Jaeyoung Shin *et al.*'s benchmark dataset [30] was used for comparison purposes to validate the methods proposed in this paper.

The classification results for both the motor imagery (MI) experiment data and the mental arithmetic (MA) experiment data were compared with results from previous works. In both these two different tasks, the p th-order polynomial fusion classification shows the best classification performance. In the MI task, the accuracy achieved was 77.53%, while the accuracy achieved in the MA task was 90.19%. Furthermore, compared with the deep neural networks method, our algorithm is simple and uses less computing resources. It can be used for real-time hybrid BCI systems.

The rest of the paper is organized as follows: hybrid BCI benchmark dataset and fusion algorithm are introduced in Section II. Experiments and pre-processing, as well as their results, are presented and discussed in Section III. The discussion and conclusions are given in the Section IV and Section V respectively.

II. MATERIALS AND METHODS

A. DATASETS

1) SUBJECTS

29 subjects, consisting of 14 males and 15 females, participated in the data collection procedures. The average age was 28.5 ± 3.7 (mean \pm standard deviation).

2) DATA ACQUISITION

EEG data was collected by a thirty-channel BrainAmp EEG amplifier (Brain Products GmbH, Gilching, Germany), with a sampling rate of 1000 Hz. These electrodes were placed according to the international 10-5 system. NIRS data was collected by NIRScout (NIRx GmbH, Berlin, Germany) with a sampling rate of 12.5 Hz and containing thirty-six channels.

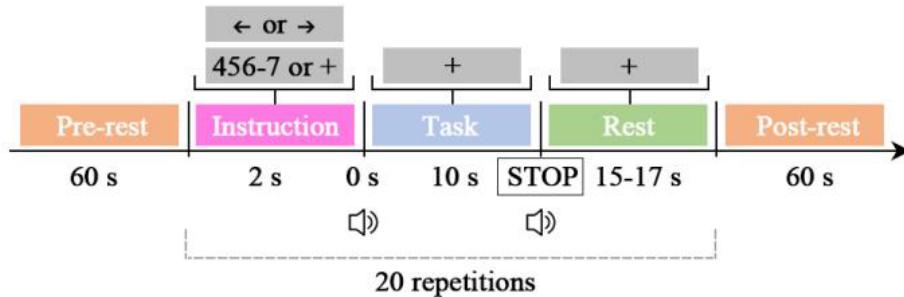


FIGURE 2. The flow of experiments.

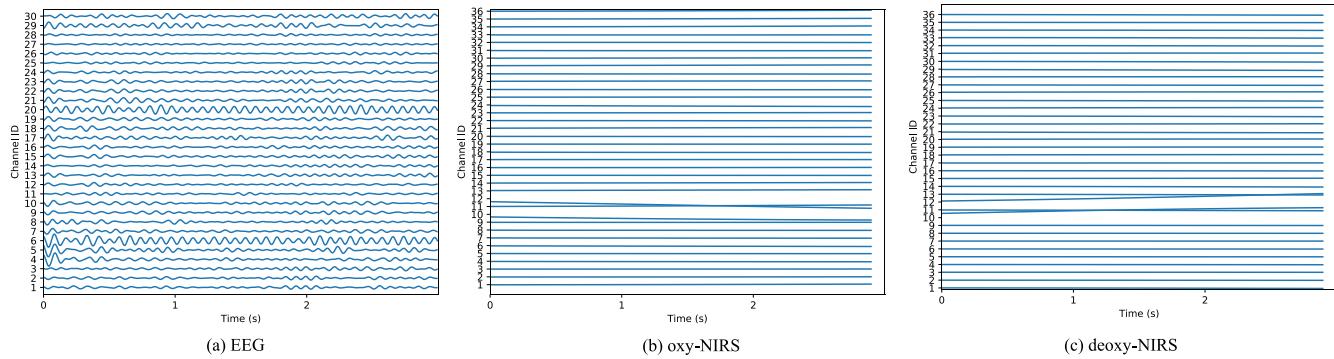


FIGURE 3. An example of EEG, oxy-NIRS and deoxy-NIRS data. Notice that the NIRS data are of low-frequency and are filtered by a bandpass filter from 0.01 Hz to 0.1 Hz.

More details about the electrodes' position can be found in [30].

3) MOTOR IMAGERY (MI) EXPERIMENT

Subjects were required to conduct kinesthetic MI. Namely, they needed to imagine opening or closing their hand while they were grabbing a ball, so that the MI is actual MI rather than visual. Each trial consisted of three parts: instruction (2s), task (10s) and rest (15-17s). Each session consisted of a pre-rest (60s), 20 repetitions of the aforementioned trial, and a post-rest (60s). The MI experiment dataset contains the data from 3 sessions.

4) MENTAL ARITHMETIC (MA) EXPERIMENTS

Subjects were required to conduct subtraction in the form of a three-digit number minus a one-digit number (e.g., 384-8). As in the MI experiment, the MA experiment dataset also contains data on 3 sessions where each session consists of 20 repetitions of the trial. Subjects were given an initial subtraction calculation, and completed this subtraction as their first trial of their total 20. For the remaining 19, the new subtraction was the result from the last calculation minus the initial one-digit number, which thus remained constant throughout the trials.

5) DATA PRE-PROCESSING

The EEG data was first re-referenced by using common average reference and filtered with the 4th-order Chebyshev type 2 filter with a bandpass of 0.5-50 Hz. Then, independent

component analysis (ICA)-based EOG rejection was used to remove artifacts. Then, the NIRS data was downsampled to 10 Hz. After that, the EEG data was downsampled to 200 Hz. For each trial, 35s of data was extracted, containing a segment of the last rest (10s), the task (10s) and the final rest (15s). Then, a 3s time window with 1s step was employed to collect data. Eventually, 33 segments were obtained from each trial. Each segment has a shape of 30×600 (channels \times times). For the NIRS data, the concentration changes of oxy- and deoxy-hemoglobin (oxy-NIRS and deoxy-NIRS) were first calculated by using the modified Beer-Lambert law. The oxy-NIRS and deoxy-NIRS were then filtered by a 6th-order zero-phase Butterworth filter with a bandpass of 0.01-0.1 Hz. After that, a 3s time window like the EEG was used to segment the NIRS data. After this process, a segment ends up with a shape of 36×30 . Thus, for both the MI and MA experiments, each subject completed 60 trials, and each trial has 33 segments. Fig. 3 shows the segments of one of the subjects for three modal data from 0s to 3s.

B. MULTIMODAL CLASSIFICATION METHOD

In this section, we will first show how the classification model is built by using single modal data. Following this, two common methods to fuse features in deep learning are introduced to allow for comparison with our method and a better understanding of it. After that, our p th-order polynomial fusion method is introduced. In the end, in order to tackle the unacceptably large amount of parameters, tensor decomposition is used.

TABLE 1. 1D-CNN neural network structure to extract the EEG signal features.

Layer	Input (C×T)	Operation	Filter size	Stride	Padding	Output
1	30×600	Conv1D	9	4	0	60×148
	60×148	BatchNorm+ReLU				60×148
2	60×148	Conv1D	3	1	0	60×146
	60×146	BatchNorm+ReLU				60×146
3	60×146	Conv1D	3	1	0	60×144
	60×144	BatchNorm+ReLU				60×144
4	60×144	Conv1D	9	4	0	120×32
	120×32	BatchNorm+ReLU				120×32
5	120×32	Conv1D	3	1	0	120×30
	120×30	BatchNorm+ReLU				120×30
6	120×30	Conv1D	3	1	0	120×28
	120×28	BatchNorm+ReLU				120×28
7	120×28	AvgPool1D				120
8	120	Linear	120			60
9	60	ReLU				60
10	60	Linear	60			2
11	2	Softmax				2

1) NOTATION

Focusing on the tri-modal task, assume x^1, x^2 and x^3 denote the EEG, oxy-hemoglobin NIRS (oxy-NIRS) and deoxy-hemoglobin NIRS (deoxy-NIRS), respectively. Also assume that z^1, z^2 and z^3 denote their respective feature vectors obtained from three CNNs. To express the formula concisely, Einstein notation [38] is applied to describe multiplication between tensors or between a tensor and a vector. In particular, if we assume x_i to denote a vector and \mathcal{W}_{ijk} to denote a 3rd-order tensor, their product can be written as $y_{jk} = x_i \mathcal{W}_{ijk}$, which means $y_{jk} = \sum_i x_i \mathcal{W}_{ijk}$. Also, when we concatenate z^1, z^2 and z^3 , the result is then written as $z^{1,2,3}$. Given a vector x_{i_1} , its first copy is written as x_{i_2} , and its $(N - 1)$ th copy is x_{i_N} .

2) SINGLE MODAL CLASSIFICATION

We have data from three different modalities, and each has a shape of channels×times. Specifically, these are EEG, oxy-NIRS and deoxy-NIRS, with shapes 30×600 , 36×30 and 36×30 , respectively. For the single-modal classification, we conduct 1D-CNN and then each convolutional layer is followed by a batch-norm layer and a ReLU layer [39]. Since oxy-NIRS and deoxy-NIRS have same shape, we use the same CNN structure for them. Table 1 and 2 show the CNNs used on EEG and NIRS in detail. Notice that the feature vectors z^1, z^2 and z^3 are the results of “AvgPool1d”.

Considering the triple modal fusion problem, there are linear and multilinear methods. Fig. 4, 5, and 6 show three ways to fuse multimodal data.

3) LINEAR FUSION (LF) CLASSIFICATION

It is commonly used to simply concatenate all feature vectors obtained from the output of “AvgPool1d”. Thus, the linear

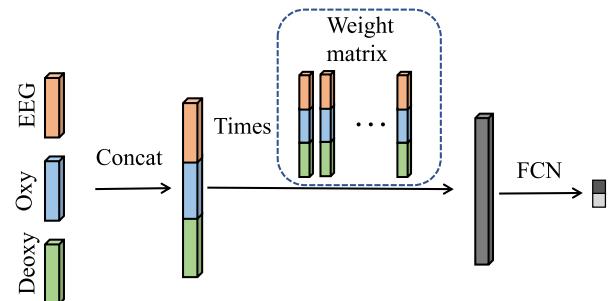


FIGURE 4. Diagram of the linear fusion (LF) method. Fused data is classified using a fully connected neural network (FCN).

fusion product can be written as

$$y_o = z_i^{1,2,3} W_{i,o}, \quad (1)$$

where $W_{i,o}$ is a fusion matrix, and y_o is the fused feature vector with a length of o . y_o is then put in a fully connected network (FCN) in order to generate a classification result. Fig. 4 shows the flow of LF classification.

4) TENSOR FUSION (LF) CLASSIFICATION

For the tensor fusion, the outer product is usually introduced. It is a natural way to obtain a feature containing the interaction amongst multiple feature vectors. In our task, the outer product result can be written as

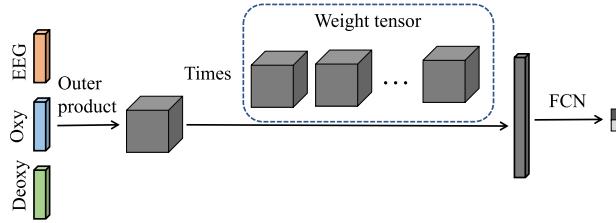
$$\mathcal{Z}_{abc} = z_a^1 z_b^2 z_c^3, \quad (2)$$

where \mathcal{Z}_{abc} is a 3rd-order feature tensor, and a, b, c are the lengths of three modal feature vectors. After that, a 4th-order weight tensor \mathcal{W}_{abco} is used to obtain the fused feature vector y_o :

$$y_o = \mathcal{Z}_{abc} \mathcal{W}_{abco}. \quad (3)$$

TABLE 2. 1D-CNN neural network structure to extract the NIRS signal features.

Layer	Input (C×T)	Operation	Filter size	Stride	Padding	Output
1	36×30	Conv1D	9	4	0	72×13
	72×13	BatchNorm+ReLU				72×13
2	72×13	Conv1D	3	1	0	72×11
	72×11	BatchNorm+ReLU				72×11
3	72×11	Conv1D	3	1	0	72×9
	72×9	BatchNorm+ReLU				72×9
4	72×9	Conv1D	9	4	0	144×7
	144×7	BatchNorm+ReLU				144×7
5	144×7	Conv1D	3	1	0	144×5
	144×5	BatchNorm+ReLU				144×5
6	144×5	Conv1D	3	1	0	144×3
	144×3	BatchNorm+ReLU				144×3
7	144×3	AvgPool1D				144
8	144	Linear	144			72
9	72	ReLU				72
10	72	Linear	72			2
11	2	Softmax				2

**FIGURE 5.** Diagram of a tensor fusion method. Fused data is classified using a fully connected neural network (FCN).

Then, L2 normalization is applied to y_o . Finally, FNC is used for classification. This operation is the same as the one shown above in the LF classification section. Fig. 5 shows the flowchart of TF classification.

5) p th-ORDER POLYNOMIAL FUSION (p th-PF) CLASSIFICATION

Tensor fusion considers the interaction between multiple feature vectors. However, the interactions within each feature vector or between two of the feature vectors are not present in the fusion. To tackle this problem, polynomial fusion is introduced. As in with linear fusion, we firstly obtain $z^{1,2,3}$ by concatenating all feature vectors. Then, $p - 1$ copies of $z^{1,2,3}$ are made and the outer product is calculated:

$$\mathcal{Z}_{i_1 i_2 \dots i_p}^{1,2,3} = z_{i_1}^{1,2,3} z_{i_2}^{1,2,3} \dots z_{i_p}^{1,2,3}. \quad (4)$$

Also, as in with tensor fusion, a $(p + 1)$ th-order weight tensor $\mathcal{W}_{i_1 i_2 \dots i_p o}$ is employed for fusion:

$$y_o = \mathcal{Z}_{i_1 i_2 \dots i_p}^{1,2,3} \mathcal{W}_{i_1 i_2 \dots i_p o}. \quad (5)$$

The subsequent operations are the same as shown in the TF classification section. Fig. 6 shows the flowchart of PF classification.

6) RETHINK THREE FUSION METHODS

As it turns out, the fusion weight matrix/tensor in these fusion methods can be “decomposed”. It should be noticed that the “decomposition” here is just to decompose the aforementioned formulas in order to make them easier to understand. For LF, Eq. 1 can be written as:

$$y_o = z_a^1 W_{ao}^1 + z_b^2 W_{bo}^2 + z_c^3 W_{co}^3, \quad (6)$$

where

$$W_{io} = \begin{bmatrix} W_{ao}^1 \\ W_{bo}^2 \\ W_{co}^3 \end{bmatrix}. \quad (7)$$

We can see that LF actually only considers the single modal feature and simply adds them together. For TF, the \mathcal{W}_{abco} in Eq. 3 cannot be “decomposed” since three modal vectors are entangled together. This may raise the problem of overfitting in the model, because the noise in the sample may be amplified during feature entanglement. For p th-PF, LF is a special case of p th-PF when $p = 1$. To be concise with the formula, if we let $p = 2$, then Eq. 5 can be written as:

$$\begin{aligned} y_o = & \mathcal{Z}_{a_1 a_2}^{1,1} \mathcal{W}_{a_1 a_2 o}^{1,1} + \mathcal{Z}_{a_1 b_2}^{1,2} \mathcal{W}_{a_1 b_2 o}^{1,2} + \mathcal{Z}_{a_1 c_2}^{1,3} \mathcal{W}_{a_1 c_2 o}^{1,3} \\ & + \mathcal{Z}_{b_1 a_2}^{2,1} \mathcal{W}_{b_1 a_2 o}^{2,1} + \mathcal{Z}_{b_1 b_2}^{2,2} \mathcal{W}_{b_1 b_2 o}^{2,2} + \mathcal{Z}_{b_1 c_2}^{2,3} \mathcal{W}_{b_1 c_2 o}^{2,3} \\ & + \mathcal{Z}_{c_1 a_2}^{3,1} \mathcal{W}_{c_1 a_2 o}^{3,1} + \mathcal{Z}_{c_1 b_2}^{3,2} \mathcal{W}_{c_1 b_2 o}^{3,2} + \mathcal{Z}_{c_1 c_2}^{3,3} \mathcal{W}_{c_1 c_2 o}^{3,3}, \end{aligned} \quad (8)$$

where

$$\mathcal{Z}_{i_1 i_2}^{1,2,3} = \begin{bmatrix} \mathcal{Z}_{a_1 a_2}^{1,1} & \mathcal{Z}_{a_1 b_2}^{1,2} & \mathcal{Z}_{a_1 c_2}^{1,3} \\ \mathcal{Z}_{b_1 a_2}^{2,1} & \mathcal{Z}_{b_1 b_2}^{2,2} & \mathcal{Z}_{b_1 c_2}^{2,3} \\ \mathcal{Z}_{c_1 a_2}^{3,1} & \mathcal{Z}_{c_1 b_2}^{3,2} & \mathcal{Z}_{c_1 c_2}^{3,3} \end{bmatrix}, \quad (9)$$

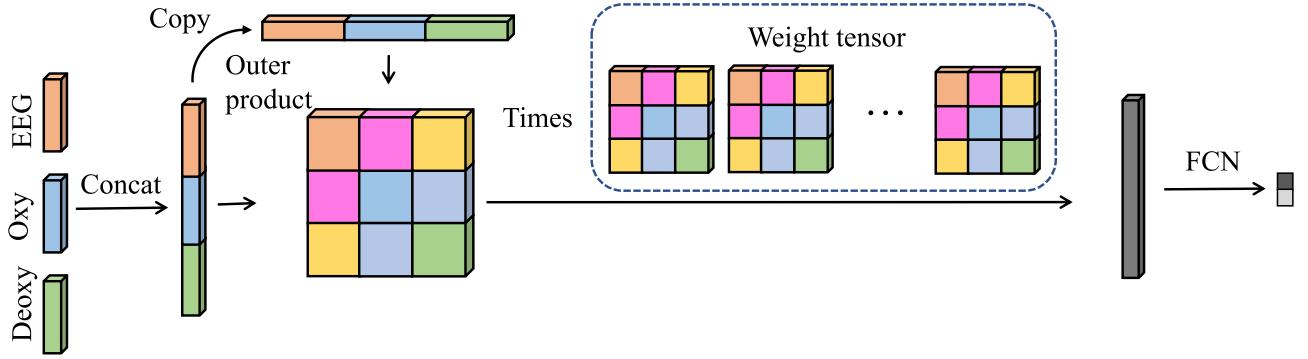


FIGURE 6. Diagram of the p th-order polynomial fusion method. Fused data is classified using a fully connected neural network (FCN).

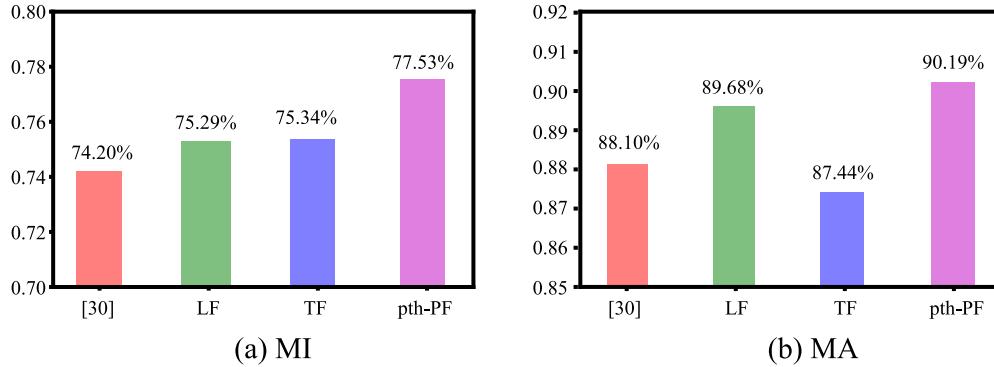


FIGURE 7. Classification accuracy for the MI (figure a) and MA (figure b) tasks. Comparison between our methods and the ones in [30]. MI: Motion Imagery task; MA: Mental arithmetic task.

and

$$\mathcal{W}_{i_1 i_2 \dots i_p o}^{1,2,3} = \begin{bmatrix} \mathcal{W}_{a_1 a_2 o}^{1,1} & \mathcal{W}_{a_1 b_2 o}^{1,2} & \mathcal{W}_{a_1 c_2 o}^{1,3} \\ \mathcal{W}_{b_1 a_2 o}^{2,1} & \mathcal{W}_{b_1 b_2 o}^{2,2} & \mathcal{W}_{b_1 c_2 o}^{2,3} \\ \mathcal{W}_{c_1 a_2 o}^{3,1} & \mathcal{W}_{c_1 b_2 o}^{3,2} & \mathcal{W}_{c_1 c_2 o}^{3,3} \end{bmatrix}. \quad (10)$$

We can therefore see that all interactions between modes are taken into account. Compared with TF, this kind of entanglement may also introduce more noise, but at the same time it also generates more fruitful features. It can be observed, from the experimental results, that this operation has indeed a positive effect.

7) DIMENSION REDUCTION BY TENSOR DECOMPOSITION

In deep learning, the usually large amount of parameters can lead to difficult training or overfitting [9], [40]. The number of parameters required by TF increases exponentially along with the increase of the mode, and the number of parameters required by PF increases exponentially along with the increase of the order. In order to tackle this issue, canonical/polyadic (CP) [41] decomposition is used. Its main idea is based on using several core tensors to represent the original weight tensor. Given a weight tensor $\mathcal{W}_{i_1 i_2 \dots i_p o}$, this tensor can be represented by:

$$\mathcal{W}_{i_1 i_2 \dots i_p o} = \mathcal{W}_{i_1 ro} \mathcal{W}_{i_2 ro} \dots \mathcal{W}_{i_p ro} \mathcal{W}_r, \quad (11)$$

where r is called *rank*, and is used to control the size of all core tensors. With the increase of r , the core tensors can better

approximate the weight tensor (i.e. increasing the ability of representation). Thus, Eq. 5 can be written as:

$$\begin{aligned} y_o &= z_{i_1 i_2 \dots i_p}^{1,2,3} \mathcal{W}_{i_1 ro} \mathcal{W}_{i_2 ro} \dots \mathcal{W}_{i_p ro} \mathcal{W}_r \\ &= (z_{i_1}^{1,2,3} \mathcal{W}_{i_1 ro})(z_{i_2}^{1,2,3} \mathcal{W}_{i_2 ro}) \dots (z_{i_p}^{1,2,3} \mathcal{W}_{i_p ro}) \mathcal{W}_r \end{aligned} \quad (12)$$

Considering that all of $z_{i_p}^{1,2,3}$ are the same, we can further reduce the parameters by assuming symmetric structure of the core tensors, namely, $\mathcal{W}_{i_m ro} = \mathcal{W}_{i_n ro}$, $m \neq n$.

8) COMPLEXITY ANALYSIS

Assume that three feature vectors have lengths of A , B and C , and the fusion vector has a length of O . Let CP rank be R . Then, the computational and storage complexity for LF, TF and p th-order PF will be $\mathcal{O}((A + B + C)O)$, $\mathcal{O}(ABCO)$ and $\mathcal{O}((ABC)^p O)$, respectively. By conducting CP decomposition, the computational and storage complexity for TF and p th-order PF will be $\mathcal{O}((A + B + C)RO)$ and $\mathcal{O}(p(A + B + C)RO)$. By assuming symmetric structure of the core tensors, the complexity for p th-order PF is reduced to $\mathcal{O}((A + B + C)RO)$. We can thus see that CP decomposition significantly reduces the amount of parameters, and also that symmetric structure can further reduce parameters on p th-order PF.

III. EXPERIMENTS

A. NETWORK CONFIGURATION

For the single-modal model, the network details are shown in Table 1 and Table 2. For the triple-modal model, the

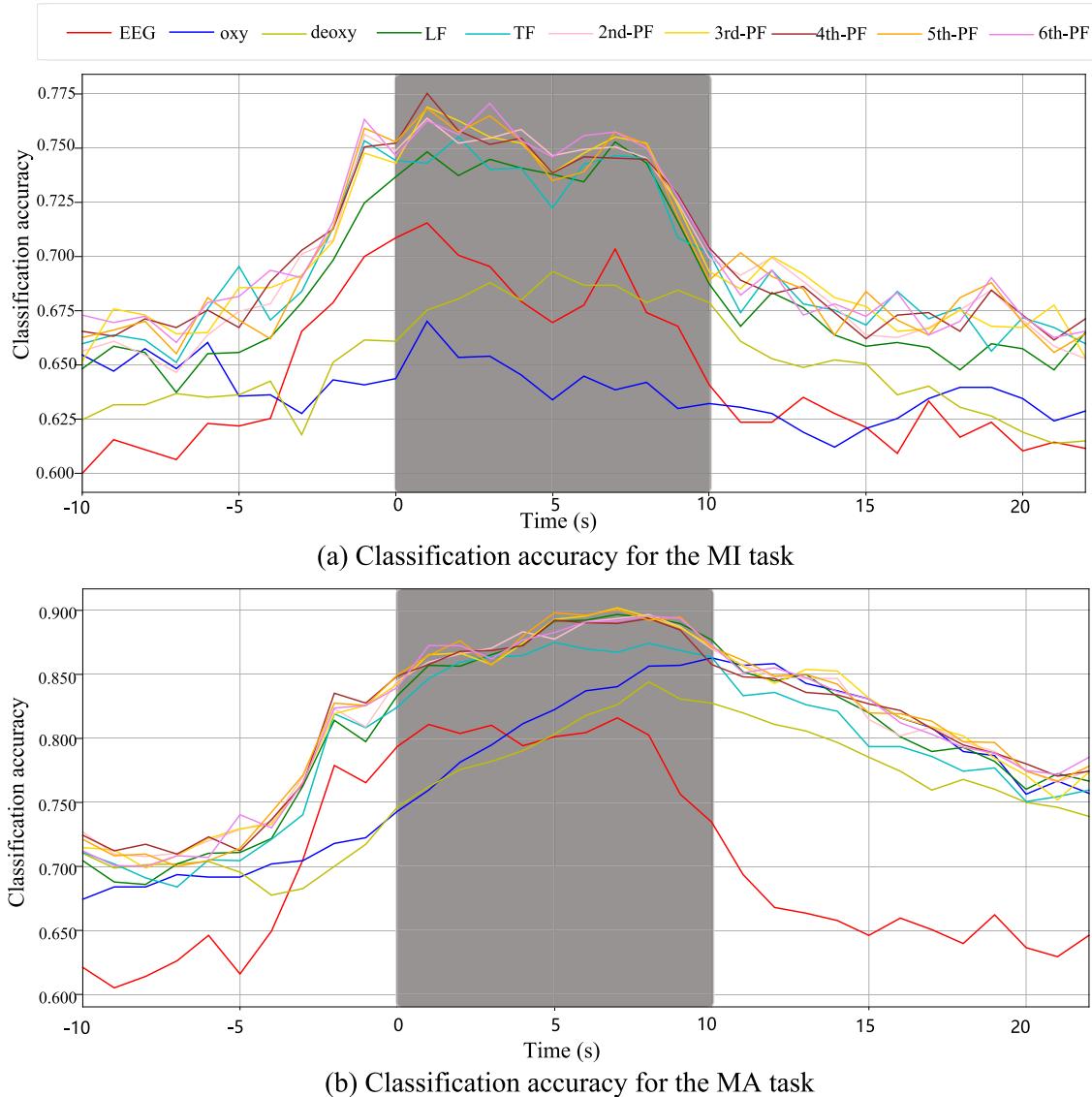


FIGURE 8. (a): Classification accuracy for the motion imagery (MI) task. (b): Classification accuracy for mental arithmetic (MA) task.

CNN model for extracting feature vectors is the same as the one before ‘‘Avgpool1d’’ in the Table 1 and Table 2. The fused feature vector has a length of 128. TF employs CP decomposition, and p th-order PF employs CP decomposition as well as symmetric structure. After that, a FCN with one layer is employed for classification.

In the training phase, we applied cross-entropy as the loss function, and the Adam optimizer to train the model, where the learning rate was set to 0.001 and other parameters were default. All models were trained by 300 epochs with a mini-batch size of 16.

For the triple-modal data, we have 29 subjects, with each subject carrying out 60 trials. A trial lasts 35 seconds, where resting goes from -10 s to 0s, the task goes from 0s to 10s, followed by another resting period from 10s to 25s. We only used 3s data as the input for the model. For each subject,

we conducted a 5-fold cross validation and took the average. Finally, all the subjects’ results were averaged.

B. RESULTS

We firstly compared the classification accuracy results with the best results from the following published paper [30]. Our methods show significant improvement.

The results of the classification accuracies of our methods are displayed in Fig. 8, where x-axis indicates the left edge of the moving time window and the y-axis shows the accuracy. We can see that all the tri-modal fusion models perform better than single-modal models. Furthermore, in both tasks, PF achieves better results than LF and TF. And the model complexities are shown in Table 3.

In the MI task, the best results for EEG, oxy-NIRS, deoxy-NIRS, LF, TF and PF are obtained by the time window 1

TABLE 3. Model complexity comparison.

Model	EEG	NIRS	LF	TF	2nd-PF	3rd-PF	4th-PF	5th-PF	6th-PF
Param	197,462	232,202	685,922	1,051,234	1,051,234	1,051,234	1,051,234	1,051,234	1,051,234
FLOPs	10,580,698	1,023,478	12,676,006	13,401,126	13,410,102	13,411,126	13,412,150	13,413,174	13,414,198

(71.55%), 1 (67.01%), 5 (69.31%), 7 (75.29%), -1 (75.34%) and 1 (77.53%)(5th-order), respectively.

In the MA task, these respective time windows for optimal results are 7 (81.60%), 10 (86.28%), 8 (84.42%), 7 (89.68%), 8 (87.44%) and 8 (90.19%)(3rd-order), respectively. The best results for LF, TF, PF and [30] are shown in Fig. 7.

IV. DISCUSSION

A. COMBINING EEG AND NIRS SIGNALS

In recent years, a large number of new technologies have been developed for the analysis of brain activity from single-modal BCI equipment. New algorithms are always being developed over time that improve the reliability of the current BCI system. However, a single-model BCI system still has many limitations. For instance, a multimodal BCI has better anti-noise capabilities than a single-model BCI [42]–[44]. Multimodality implies a multi-view, and thus a different representation for the same brain activity patterns [45]. It helps the algorithm to detect specific patterns from the signals. Many previous studies are in line with our results having also indicated the improvement by fusion of multi-signals [46].

We selected a NIRS and EEG-based hybrid BCI fusion system because both of these two systems are real-time and low-cost [31], [47]. Additionally, compared with MRI, these two technologies do not require a medical license. We believe EEG and NIRS-based hybrid BCI methods can and will be used as a critical tool in many applications.

B. HYBRID BCI SIGNAL PROCESSING BASED ON TENSOR FUSION

According to recent literature, deep learning methods or artificial neural networks have been applied for the detection of patterns in brain activity. Several artificial neural networks have also been applied to this dataset. However, almost all of the neural networks result in an over-fitting problem since the BCI dataset includes small samples. Although tuning the neural structures or parameters could slightly improve the accuracy, the method's stability leaves much to be desired [22], [48].

A tensor is a higher-order array that represents signals from different types of sensors [49]–[51]. For example, EEG data or NIRS data can be represented by time × frequency × electrode, and functional MRI data can be represented by voxels × scans × subjects [52]–[54]. In addition, tensor decomposition has been widely used to capture brain signal patterns. In this work, the classification accuracy of our proposed methods is much better than previous methods, and our system is also very robust.

V. CONCLUSIONS

In this paper we have focused on using tensor fusion methods to construct a hybrid BCI system. The hybrid BCI system

includes simultaneous EEG signals and NIRS signals. It also includes two tasks: a MI task and a MA task. We used a shallow CNN neural network to detect the EEG and NIRS feature vectors. Then, we used LF, TF and p th-order PF fusion methods to integrate the feature vectors. The results show that the p th-order PF fusion method has the best classification accuracy out of the three fusion methods and the single-modal classification. Furthermore, when comparing the performance of our method with published literature, p th-order shows better results. We believe our method could be useful for hybrid BCI systems.

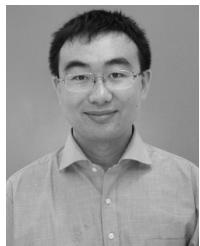
ACKNOWLEDGMENT

The authors would like to thank Dr. Kai Zhang for the fruitful advice and Pau Solé-Vilaró for the English review.

REFERENCES

- [1] C. Mühl, C. Jeunet, and F. Lotte, "EEG-based workload estimation across affective contexts," *Frontiers Neurosci.*, vol. 8, p. 114, Jun. 2014.
- [2] K. K. Ang, K. S. G. Chua, K. S. Phua, C. Wang, Z. Y. Chin, C. W. K. Kuah, W. Low, and C. Guan, "A randomized controlled trial of EEG-based motor imagery brain-computer interface robotic rehabilitation for stroke," *Clin. EEG Neurosci.*, vol. 46, no. 4, pp. 310–320, Oct. 2015.
- [3] C.-F.-V. Latchoumane, F.-B. Vialatte, J. Solé-Casals, M. Maurice, S. R. Wimalaratna, N. Hudson, J. Jeong, and A. Cichocki, "Multiway array decomposition analysis of EEGs in Alzheimer's disease," *J. Neurosci. Methods*, vol. 207, no. 1, pp. 41–50, May 2012.
- [4] B. Blankertz, M. Tangermann, C. Vidaurre, S. Fazli, C. Sannelli, S. Haufe, C. Maeder, L. E. Ramsey, I. Sturm, and G. Curio, "The Berlin brain-computer interface: Non-medical uses of BCI technology," *Frontiers Neurosci.*, vol. 4, p. 198, Dec. 2010.
- [5] Q. Liu, Y. Jiao, Y. Miao, C. Zuo, X. Wang, A. Cichocki, and J. Jin, "Efficient representations of EEG signals for SSVEP frequency recognition based on deep multiset CCA," *Neurocomputing*, vol. 378, pp. 36–44, Feb. 2020.
- [6] X. Mao, W. Li, C. Lei, J. Jin, F. Duan, and S. Chen, "A brain–robot interaction system by fusing human and machine intelligence," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 533–542, Mar. 2019.
- [7] W. Wu, "An electroencephalographic signature predicts antidepressant response in major depression," *Nature Biotechnol.*, vol. 38, no. 4, pp. 439–447, Apr. 2020.
- [8] S. Waldert, "Invasive vs. non-invasive neuronal signals for brain-machine interfaces: Will one prevail?" *Frontiers Neurosci.*, vol. 10, p. 295, Jun. 2016.
- [9] Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang, and A. Cichocki, "Sparse Bayesian classification of EEG for brain–computer interface," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 11, pp. 2256–2267, Nov. 2016.
- [10] Y. Zhang, H. Zhang, X. Chen, M. Liu, X. Zhu, S.-W. Lee, and D. Shen, "Strength and similarity guided group-level brain functional network construction for MCI diagnosis," *Pattern Recognit.*, vol. 88, pp. 421–430, Apr. 2019.
- [11] Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Frequency recognition in SSVEP-based BCI using multiset canonical correlation analysis," *Int. J. Neural Syst.*, vol. 24, no. 04, Jun. 2014, Art. no. 1450013.
- [12] S. Fazli, J. Mehnert, J. Steinbrink, G. Curio, A. Villringer, K.-R. Müller, and B. Blankertz, "Enhanced performance by a hybrid NIRS–EEG brain computer interface," *NeuroImage*, vol. 59, no. 1, pp. 519–529, Jan. 2012.
- [13] R. Sitaram, A. Caria, and N. Birbaumer, "Hemodynamic brain–computer interfaces for communication and rehabilitation," *Neuroetw.*, vol. 22, no. 9, pp. 1320–1328, Nov. 2009.

- [14] Y. Zheng, Y. Peng, G. Xu, L. Li, and J. Wang, "Using corticomuscular coherence to reflect function recovery of paretic upper limb after stroke: A case study," *Frontiers Neurol.*, vol. 8, p. 728, Jan. 2018.
- [15] Z. J. Koles and A. C. K. Soong, "EEG source localization: Implementing the spatio-temporal decomposition approach," *Electroencephalogr. Clin. Neurophysiol.*, vol. 107, no. 5, pp. 343–352, Nov. 1998.
- [16] Q. Wang and O. Sourina, "Real-time mental arithmetic task recognition from EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 2, pp. 225–232, Mar. 2013.
- [17] T. Inouye, K. Shinosaki, A. Iyama, and Y. Matsumoto, "Localization of activated areas and directional EEG patterns during mental arithmetic," *Electroencephalogr. Clin. Neurophysiol.*, vol. 86, no. 4, pp. 224–230, Apr. 1993.
- [18] S. K. Hadjidimitriou and L. J. Hadjileontiadis, "Toward an EEG-based recognition of music liking using time-frequency analysis," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 12, pp. 3498–3510, Dec. 2012.
- [19] Y. Liu, Q. Zhao, and L. Zhang, "Uncorrelated multiway discriminant analysis for motor imagery EEG classification," *Int. J. Neural Syst.*, vol. 25, no. 04, Jun. 2015, Art. no. 1550013.
- [20] A. Antoniades, L. Spyrou, D. Martin-Lopez, A. Valentin, G. Alarcon, S. Sanei, and C. C. Took, "Deep neural architectures for mapping scalp to intracranial EEG," *Int. J. Neural Syst.*, vol. 28, no. 08, Oct. 2018, Art. no. 1850009.
- [21] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput. Biol. Med.*, vol. 100, pp. 270–278, Sep. 2018.
- [22] Z. Zhang, F. Duan, J. Sole-Casals, J. Dinares-Ferran, A. Cichocki, Z. Yang, and Z. Sun, "A novel deep learning approach with data augmentation to classify motor imagery signals," *IEEE Access*, vol. 7, pp. 15945–15954, 2019.
- [23] T. Trakoolwilaiwan, B. Behboodi, J. Lee, K. Kim, and J.-W. Choi, "Convolutional neural network for high-accuracy functional near-infrared spectroscopy in a brain–computer interface: Three-class classification of rest, right-, and left-hand motor execution," *Neurophotonics*, vol. 5, no. 1, Sep. 2017, Art. no. 011008.
- [24] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.
- [25] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J. Neural Eng.*, vol. 14, no. 1, Feb. 2017, Art. no. 016003.
- [26] P. Wang, A. Jiang, X. Liu, J. Shang, and L. Zhang, "LSTM-based EEG classification in motor imagery tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 11, pp. 2086–2095, Nov. 2018.
- [27] B. Koo, H.-G. Lee, Y. Nam, H. Kang, C. S. Koh, H.-C. Shin, and S. Choi, "A hybrid NIRS-EEG system for self-paced brain computer interface with online motor imagery," *J. Neurosci. Methods*, vol. 244, pp. 26–32, Apr. 2015.
- [28] N. Tsujii, I. Otsuka, S. Okazaki, M. Yanagi, S. Numata, N. Yamaki, Y. Kawakubo, O. Shirakawa, and A. Hishimoto, "Mitochondrial DNA copy number raises the potential of left frontopolar hemodynamic response as a diagnostic marker for distinguishing bipolar disorder from major depressive disorder," *Frontiers Psychiatry*, vol. 10, p. 312, May 2019.
- [29] N. Naseer and K.-S. Hong, "FNIRS-based brain-computer interfaces: A review," *Frontiers Human Neurosci.*, vol. 9, p. 3, Jan. 2015.
- [30] J. Shin, A. von Luhmann, B. Blankertz, D.-W. Kim, J. Jeong, H.-J. Hwang, and K.-R. Muller, "Open access dataset for EEG+NIRS single-trial classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1735–1745, Oct. 2017.
- [31] J. Shin, J. Kwon, and C.-H. Im, "A ternary hybrid EEG-NIRS brain–computer interface for the classification of brain activation patterns during mental arithmetic, motor imagery, and idle state," *Frontiers Neuroinform.*, vol. 12, p. 5, Feb. 2018.
- [32] A. M. Chiarelli, P. Croce, A. Merla, and F. Zappasodi, "Deep learning for hybrid EEG-fNIRS brain–computer interface: Application to motor imagery classification," *J. Neural Eng.*, vol. 15, no. 3, Jun. 2018, Art. no. 036028.
- [33] T.-Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear CNN models for fine-grained visual recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1449–1457.
- [34] H. Ben-younes, R. Cadene, M. Cord, and N. Thome, "MUTAN: Multi-modal tucker fusion for visual question answering," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2612–2620.
- [35] Y. Bai, J. Fu, T. Zhao, and T. Mei, "Deep attention neural tensor network for visual question answering," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, pp. 20–35, 2018.
- [36] Z. Liu, Y. Shen, V. Bharadhwaj Lakshminarasimhan, P. P. Liang, A. Zadeh, and L.-P. Morency, "Efficient low-rank multimodal fusion with modality-specific factors," 2018, *arXiv:1806.00064*. [Online]. Available: <http://arxiv.org/abs/1806.00064>
- [37] M. Hou, J. Tang, J. Zhang, W. Kong, and Q. Zhao, "Deep multimodal multilinear fusion with high-order polynomial pooling," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 12113–12122.
- [38] A. P. Harrison and D. Joseph, "Numeric tensor framework: Exploiting and extending einstein notation," *J. Comput. Sci.*, vol. 16, pp. 128–139, Sep. 2016.
- [39] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017.
- [40] H. Shi, M. Xu, and R. Li, "Deep learning for household load Forecasting—A novel pooling deep RNN," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5271–5280, Sep. 2018.
- [41] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Aug. 2009.
- [42] J. Shin, D.-W. Kim, K.-R. Müller, and H.-J. Hwang, "Improvement of information transfer rates using a hybrid EEG-NIRS brain–computer interface with a short trial length: Offline and pseudo-online analyses," *Sensors*, vol. 18, no. 6, p. 1827, 2018.
- [43] M. J. Khan, K.-S. Hong, N. Naseer, and M. R. Bhutta, "Multi-decision detection using EEG-NIRS based hybrid brain–computer interface (BCI)," in *Proc. 20th Annu. Meeting Org. Hum. Brain Mapping (OHMB)*, 2014, pp. 1–5.
- [44] F. Wallois, M. Mahmoudzadeh, A. Patil, and R. Grebe, "Usefulness of simultaneous EEG–NIRS recording in language studies," *Brain Lang.*, vol. 121, no. 2, pp. 110–123, May 2012.
- [45] S. Fazli and S.-W. Lee, "Brain computer interfacing: A multi-modal perspective," *J. Comput. Sci. Eng.*, vol. 7, no. 2, pp. 132–138, Jun. 2013.
- [46] S. Ahn and S. C. Jun, "Multi-modal integration of EEG-fNIRS for brain–computer interfaces—Current limitations and future directions," *Frontiers Hum. Neurosci.*, vol. 11, p. 503, Oct. 2017.
- [47] G. Pei, G. Guo, D. Chen, R. Yang, Z. Shi, S. Wang, J. Zhang, J. Wu, and T. Yan, "BrainKilter: A real-time EEG analysis platform for neuro-feedback design and training," *IEEE Access*, vol. 8, pp. 57661–57673, 2020.
- [48] A. Appriou, A. Cichocki, and F. Lotte, "Modern machine learning algorithms to classify cognitive and affective states from electroencephalography signals," *IEEE Syst., Man Cybern. Mag.*, early access, Feb. 1, 2020. [Online]. Available: https://hal.inria.fr/hal-02483908/file/IEEE_system_man_cybernetics_new.pdf
- [49] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges, and prospects," *Proc. IEEE*, vol. 103, no. 9, pp. 1449–1477, Sep. 2015.
- [50] E. Acar, Y. Levin-Schwartz, V. D. Calhoun, and T. Adali, "Tensor-based fusion of EEG and fMRI to understand neurological changes in schizophrenia," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2017, pp. 1–4.
- [51] B. Hunyadi, P. Dupont, W. Van Paesschen, and S. Van Huffel, "Tensor decompositions and data fusion in epileptic electroencephalography and functional magnetic resonance imaging data," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 7, no. 1, p. e1197, Jan. 2017.
- [52] C. F. Beckmann and S. M. Smith, "Tensorial extensions of independent component analysis for multisubject fMRI analysis," *NeuroImage*, vol. 25, no. 1, pp. 294–311, Mar. 2005.
- [53] F. Miwakeichi, E. Martínez-Montes, P. A. Valdés-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchi, "Decomposing EEG data into space–time–frequency components using parallel factor analysis," *NeuroImage*, vol. 22, no. 3, pp. 1035–1045, Jul. 2004.
- [54] E. Acar, C. Aykut-Bingol, H. Bingol, R. Bro, and B. Yener, "Multiway analysis of epilepsy tensors," *Bioinformatics*, vol. 23, no. 13, pp. 10–18, Jul. 2007.



ZHE SUN received the Ph.D. degree from Yokohama City University, in 2017. He joined RIKEN, in 2015, as a Research Support Assistant. He has been a Research Scientist with RIKEN, since 2017. He is currently a Research Scientist with the Research and Development Group, Computational Engineering Applications Unit, Head Office for Information Systems and Cybersecurity, National Institute of RIKEN, Japan. From 2014 to 2017, his research topics were the development of spiking neuron model, and spiking neural network to understand and elucidate brain functions. His current research interests include large scale brain simulation and neuromorphic engineering.



FENG DUAN received the B.E. and M.E. degrees in mechanical engineering from Tianjin University, China, in 2002 and 2004, respectively, and the M.S. and Ph.D. degrees in precision engineering from The University of Tokyo, Japan, in 2006 and 2009, respectively. He is currently a Professor with Nankai University, China. His research interests include cellular manufacture systems, rehabilitation robots, and brain machine interfaces.



YU LIU is currently a Distinguished Professor and the Dean of the School of Kinesiology, Shanghai University of Sport, China. His research interests include biomechanics of injuries, footwear biomechanics, and neuromuscular control of human movement. He has published over 200 peer reviewed articles or book chapters, given over 130 international and domestic lectures. His research has been supported by large competitive grants from the Chinese Central Government, Shanghai Municipal Government, and Nike Global Research Partner. He is a Distinguished Professor of Yangtze River Scholar awarded by the China Ministry of Education, Outstanding Teacher in Ten Thousand Talent Program, and owned Special government allowances of the State Council. During three decades of teaching and researching in the field of biomechanics, he has served as the Vice President of the Asia Association of Coaching Science (AACCS), an Executive Council Member of International Society of Biomechanics (ISB), the Vice President of the Chinese Society of Biomechanics in Sports, and a member of the professional committee of biomechanics in the Chinese Society of Mechanics and Chinese Society of Biomedical Engineering. He has served on the editorial board of several national and international journals, including *Footwear Science*, the *Journal of Medical Biomechanics*, *China Sport Science*, and the *Chinese Journal of Sport Medicine*. He is also the Deputy Editor-in-Chief of the *Journal of Sport and Health Science*.



ZIHAO HUANG received the B.E. degree in automation from Nankai University, Tianjin, China, in 2018, where he is currently pursuing the master's degree with the College of Artificial Intelligence. His current research interests include EEG signal processing and robot research.