Progress In Electromagnetics Research, Vol. 169, 73–86, 2020

Distinguishing Bipolar Depression from Major Depressive Disorder Using fNIRS and Deep Neural Network

Tengfei Ma[1, 2, #], Hailong Lyu[3, 4, #], Jingjing Liu[5], Yuting Xia[1], Chao Qian[5], Julian Evans[1], Weijuan Xu[3, 4], Jianbo Hu[3, 4], Shaohua Hu[3, 4, *], and Sailing He[1, 2, 6, *]

Abstract—A variety of psychological scales are utilized at present as the most important basis for clinical diagnosis of mood disorders. An experienced psychiatrist assesses and diagnoses mood disorders based on clinical symptoms and relevant assessment scores. This symptom based clinical criterion is limited by the psychiatrist's experience. In practice, it is difficult to distinguish the patients with bipolar disorder with depression episode (bipolar depression, BD) from those with major depressive disorder (MDD). The functional near-infrared spectroscopy (fNIRS) technology is commonly used to perceive the emotions of a human. It measures the hemodynamic parameters of the brain, which correlate with cerebral activation. Here, we propose a machine learning classification method based on deep neural network for the brain activations of mood disorders. Large time scale connectivity is determined using an attention long short term memory neural network and short-time feature information are considered using the InceptionTime neural network in this method. Our combined method is referred to as AttentionLSTM-InceptionTime (ALSTMIT). We collected fNIRS data of 36 MDD patients and 48 BD patients who were in the depressed state. All the patients were monitored by fNIRS during conducting the verbal fluency task (VFT). We trained the model with the ALSTMIT network. The algorithm can distinguish the two types of patients effectively: the average accuracy of classification on the test set can reach 96.2% stably. The classification can provide an objective diagnosis tool for clinicians, and this algorithm may be critical for the early detection and precise treatment for the patients with mood disorders.

1. INTRODUCTION

The main features of bipolar disorder (BD) are mood changes between episodes of depression and hypomania or mania. Because the pathophysiological mechanisms of the disorder are unclear and there is a lack of clear objective biological markers, the diagnosis of bipolar disorder depends on clinical observations using a diagnostic system [35]. The correct identification of BD in clinical practice is not sufficient. Among patients seeking improvement in their depressive symptoms, 60% of bipolar depression is misdiagnosed as major depressive disorder (MDD) [35]. Misdiagnosis will lead to poor medical management and prolonged symptoms, even wrong clinical prognosis. How to precisely distinguish the bipolar depression from MDD is a critical challenge in psychiatry.

Previous MRI studies have identified structural and functional changes in different brain regions in patients with BD and MDD [51]. Studies have found abnormal activation in frontal and temporal

---

Received 22 October 2020, Accepted 24 December 2020, Scheduled 31 December 2020

* Corresponding author: Sailing He (sailing@kth.se), Shaohua Hu (dorhushaohua@zju.edu.cn).

[1] Center for Optical and Electromagnetic Research, National Engineering Research Center for Optical Instruments, Zhejiang University, Hangzhou 310058, China. [2] Ningbo Research Institute, Zhejiang University, Ningbo 315100, China. [3] Department of Psychiatry, the First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. [4] The Key Laboratory of Mental Disorder Management in Zhejiang Province, Hangzhou, China. [5] Zhejiang University School of Medicine, Hangzhou, China. [6] Department of Electromagnetic Engineering, School of Electrical Engineering, Royal Institute of Technology (KTH), S-100 44 Stockholm, Sweden. # Tengfei Ma and Hailong Lyu contributed equally to this work.

brain regions in BD, which are associated with executive functions [5]. Functional Near-Infrared Spectroscopy (fNIRS) can be utilized to identify the activation of different brain regions in the task state by detecting changes in oxygen saturation. The fNIRS uses near infrared (NIR) light at 650–900 nm to measure cerebral blood oxygenation, hemodynamic parameters, and metabolic status in local brain regions [28, 29, 39]. The activation of brain's specific parts will lead to an increase in local cerebral blood flow, while deoxy-hemoglobin (HbR) in the local venous blood will decrease accordingly. The growth rate of local cerebral blood flow can exceed the metabolic rate of local cerebral blood oxygen [22, 40]. Therefore, the activation of the cerebral cortex causes the concentration of total-hemoglobin (HbT) and oxy-hemoglobin (HbO) to increase, while the concentration of HbR decreases. During activation and rest, the absorption of near-infrared (NIR) light by the blood in the brain changes as the concentration of HbO and HbR changes. According to the modified Beer-Lambert law (MBLL) [16, 17], the concentration changes of HbO and HbR can be reflected by measuring the attenuation of NIR light. Since fNIRS was first implemented almost 30 years ago [14, 19, 48], it has been widely used for studying brain function and related diseases [4]. At present, fNIRS is mainly used in the research fields of cognitive behavior development in infants and children [3, 8, 37], mental illness [10], epilepsy [30, 50], stroke and brain injury [31]. The Hitachi ETG-4000 fNIRS acquisition instrument is used in this research, it has a total of 52 channels, and each channel independently corresponds to a different region of the brain. The corresponding distribution is shown in Figure 1.

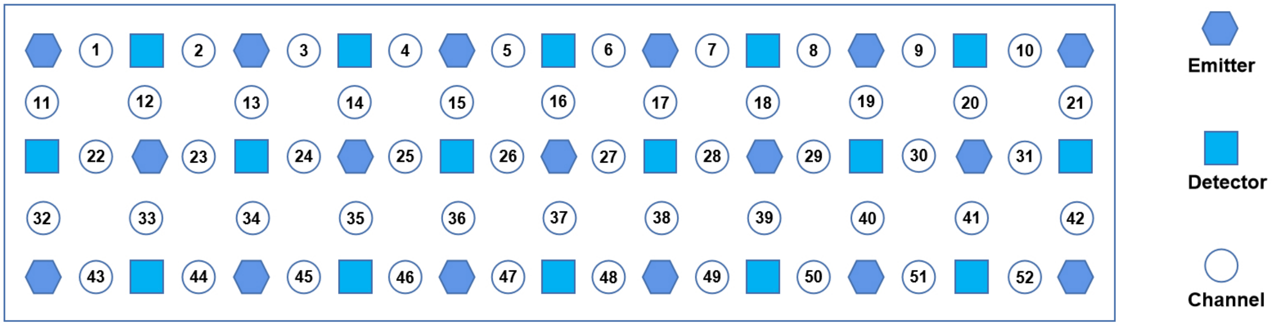


Figure 1. Channel distribution map for Hitachi fNIRS ETG-4000 acquisition instrument, 52 channels in total.

In order to distinguish BD patients who are in the depressed state from MDD patients, it is necessary to design or select appropriate activation tasks to induce a large difference in brain function activation between the two types of patients. The Verbal Fluency Task (VFT) is commonly used to assess patients verbal fluency, which reflects the patients ability to produce language [9]. Previous studies have identified reduced subfrontal activation during the VFT as a potential biological marker of depression [46]. It was demonstrated that the function of verbal fluency was worse in BD patients in comparison with MDD patients [38].

In recent years, with the development of machine learning, many methods in many different fields with excellent performance have been proposed. In the biomedical field, machine learning has been applied to areas such as pathological brain detection [49], breast cancer classification [32] and mood disorder classification [21]. In the field of inverse problem solving, deep learning has played an important role, which makes the traceability problem solved greatly [6]. In computer vision, the CNN [23] method enables the classifier to better extract features in the data, which significantly improves the accuracy of image classification. In natural language processing (NLP) field, RNN [34] method makes the artificial neural network have a strong ability to classify and predict time series signals. The fNIRS data activated by VFT have significant time series property and contain a lot of abstract feature information. Therefore, a combination of RNN based and CNN based methods will help to improve the classification.

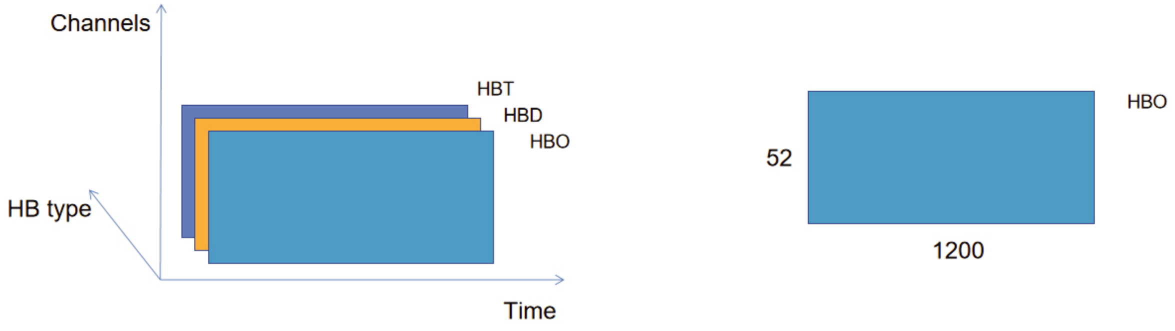
The remainder of the paper is organized as follows. Section 2 describes the data collection and organization methods. Section 3 describes the structure of AttentionLSTM-InceptionTime (ALSTMIT) which based on attention mechanism, CNN and RNN. Section 4 shows the experiments performed by applying the classification method to the collected fNIRS data. Section 5 shows and analyzes the classification results. Finally, conclusions are drawn in Section 6.

2. DATA COLLECTING AND ORGANIZATION

A large number of MDD and BD patients are diagnosed and treated in the mental health department of the First Affiliated Hospital of Zhejiang University. The fNIRS data activated by VFT of some of these patients are collected using a Hitachi fNIRS instrument. The study has been approved by the ethics committee of the First Affiliated Hospital of Zhejiang University, and written informed consent has been acquired from the patients. In this study, we performed the VFT and collected fNIRS data from 36 patients with MDD and 48 patients with BD, and all BD patients are in the depressed state during the experiment. The disease type of patients was determined by a diagnosis report provided by experienced psychiatrists. Demographic data on patients are shown in Table 1 where HAMD is for Hamilton Depression Scale. The performance of VFT is much better in MDD group than in BD group. One possible explanation is that cognitive impairment, especially verbal fluency, is not as severe in depressed patients as in patients with bipolar disorder, as shown in previous studies. Wolfe et al. [52] have found depressed bipolar patients were significantly more impaired than unipolar ones, i.e., unipolar individuals produced more correct words than bipolar patients. Depression and bipolar are two different disease spectra and the differences in cognitive function may be due to their different pathogenesis [27]. The data from each patient consist of 1200 time points from 52 channels with three different hemodynamic parameters considered. This data are represented as a three-dimensional matrix with size 1200 × 52 × 3 as shown in Figure 2(a). Sampling occurs 10 times per second over a 120 second VFT recording. The three hemodynamic parameters are HbO, HbR, and HbT.

Table 1. Demographic data on patients with bipolar disorder and depression.

	MDD (n = 36)	**BD (n = 48)**	**P value**
Age	27.06 (7.49)	24.00 (7.92)	0.075
Gender (F/M)	23/13	27/21	0.509
Education (years)	12.18 (2.99)	12.38 (2.86)	0.764
VFT	9.78 (4.50)	7.63 (4.79)	0.038
HAMD	21.70 (7.15)	19.96 (7.94)	0.312



(a) (b)

Figure 2. Structure of the fNIRS data for the hemoglobin levels. (a) A complete fNIRS sample data, including time, channel number, and hemoglobin types; (b) the HbO data matrix, its time dimension is 1200 (indicating 1200 samples), and the channel dimension is 52 (indicating 52 channels in total).

2.1. VFT Procedures

The Chinese version of the VFT was used in the study [36]. The task procedure consists of a pre-task baseline period (30 seconds), a task period (60 seconds), and a post-task baseline period (30 seconds). During the pre-task and the post-task baseline periods, participants were prompted to count from 1 to 5 repeatedly. During the 60-second task period, participants were prompted to form expression using the three Chinese characters “ 白”, “ 天”, “ 大” (pronounced as “bai”, “tian”, “da”, and meaning “white”, “sky”, and “big”, respectively). The prompts were computer-generated speech with a fixed frequency and duration. These three characters are very commonly used Chinese characters. Chinese people with different education levels can easily produce phrases or four-character idioms using these characters [24]. Participants were prompted to verbally generate as many phrases or four-character idioms beginning with each given character as possible. Note that VFT can be divided into two types of tests, Letter Fluency Task (LFT) and Category Fluency Task (CFT). In the present study we used the LFT task paradigm mainly in reference to previous studies conducted in psychiatric disorders where the same task paradigm was used [41, 47]. Onishi et al. found that brain activation measured using fNIRS was higher in the LFT paradigm than in the CFT paradigm [33]. Thus, we think for similar studies the LFT task may give a better performance than the CFT task.

2.2. NIRS Measurement

The instrument used in this study was a 52-channel multichannel fNIRS instrument (ETG-4000, Hitachi Medical Corporation, Japan). Two wavelengths of near-infrared light (695 nm and 830 nm) were used to measure the relative changes in oxyhemoglobin and deoxyhemoglobin levels, respectively. The instrument acquires data primarily using 17 optical emitters and 16 optical detectors. The emitter probes are 3.0 cm away from the detector probes and the measurement area between the probes is defined as a channel. The fNIRS probes are arranged in a 3×11 matrix, which creates 52 measurement channels. In accordance with the international 10–20 system, the lowest channel is positioned flush with the Fp1–Fp2 line [24]. The area covered by the probe allows measurement of the relative HbO and HbR signal changes in bilateral frontal and temporal cortices.

3. ATTENTIONLSTM-INCEPTIONTIME NETWORK

3.1. Background

The fNIRS signal has typical time series properties, and the classification of such signals is collectively referred to as a time series classification (TSC) problem.

Definition 1 *A time series can be expressed in the form of* $X = [x_1, x_2, ..., x_T]$*, where* T *refers to sampling* T *times in the time dimension, and* $x_t, t \in [1, T]$ *represents the result of the* t*-th sampling. Depending on the number of sensor sampling channels,* x_t *can be either univariate or multivariate.*

Definition 2 *A time series dataset can be expressed as* $D = \{(X_1, Y_1), (X_2, Y_2), ..., (X_N, Y_N)\}$*, where* $(X_i, Y_i), i \in [1, N]$ *indicates the* i*-th time series with its one-hot label. When the dataset* D *includes the* K *class, the one-hot label* Y_i *is a vector containing* K *elements, where only the element* $k \in [1, K]$ *is equal to 1 if the class index of* X_i *is* k*, and other elements are all equal to 0.*

The target of TSC is to train a model based on the dataset D, this model will map the time series data space to the corresponding one-hot label space.

3.1.1. Long Short Term Memory

Long Short Term Memory (LSTM) is a special form of RNN, which can avoid local minima and learn long term dependent information. LSTM which first proposed by Hochreiter and Schmidhuber (1997) [13], has been improved and promoted by Kawakami [20] in recent years. The single cell of LSTM is shown

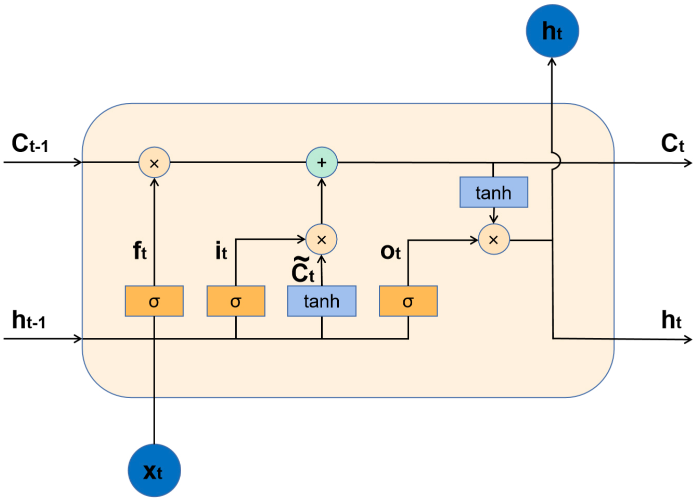


Figure 3. The structure of the LSTM cell.

in Figure 3. The LSTM cell is composed of three gates with different functions, namely the forgetting gate, updating gate and output gate.

$$\begin{aligned}
&forgetting\ gate:\\
&\quad f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)\\
&updating\ gate:\\
&\quad i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)\\
&\quad \widetilde{C_t} = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)\\
&\quad C_t = f_t C_{t-1} + i_t \widetilde{C_t}\\
&output\ gate:\\
&\quad o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)\\
&\quad h_t = o_t \tanh(C_t)
\end{aligned} \tag{1}$$

Each cell receives the hidden state h_{t-1}, cell state C_{t-1} at the previous time, and the data input x_t at the current time, then hidden state h_t and cell state C_t at the current time are calculated through the above three gate operations.

3.1.2. Attention Mechanism

The attention mechanism in machine learning is an effective method inspired by the attention model of human brain. It can select the information that is more critical to the current task goal, and then invest more attention resources in these key information, while suppressing less useful information. In Bahdanau et al.'s [1] presentation, the attention mechanism is typically used in the Encoder-Decoder model. In the original Encoder-Decoder model which proposed by Cho et al. [7], the Encoder encodes all input information into a vector C, and the Decoder applies vector C to the decoding process indifferently, that is:

$$\begin{aligned}
&Encoder: \quad && C = \mathscr{F}(x_1, x_2, \ldots, x_T)\\
&Decoder: \quad && y_i = \mathscr{G}(C, y_1, y_2, \ldots, y_{i-1})
\end{aligned}$$

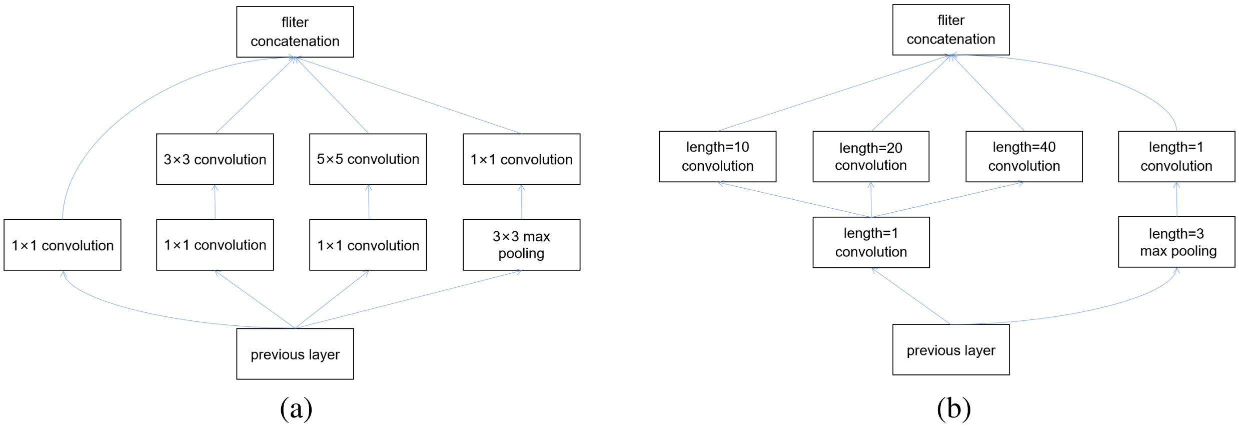
The attention mechanism implements different attention weights for different tasks y_i. Concretely, the input sequence $(x_1, x_2, \ldots, x_T)$ is mapped to a sequence of annotations $(h_1, h_2, \ldots, h_T)$ through an encoder mapping function. The context vector c_i is calculated by the formula $c_i = \sum_{j=1}^{T} \alpha_{ij} h_j$. This formula adjusts the contribution of h_j to c_i through the weighting factor α_{ij}. The α_{ij} indicates how much c_i pays attention to h_j, which can be calculated by the following formula:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T} \exp(e_{ik})} \tag{2}$$

where $e_{ij} = a(s_{i-1}, h_j)$ is an alignment model, which scores how well the input around position j and the output at position i match. The score is based on the RNN hidden state s_{i-1} (just before emitting y_i) and the j-th annotation h_j of the input time series [1].

3.1.3. InceptionTime Network

Inception [15, 43–45] is an algorithm proposed by Google for the field of computer vision. The main idea of the Inception architecture is based on finding out how an optimal local sparse structure in a convolutional vision network can be approximated and covered by readily available dense components [43]. The original 2D Inception module can be represented by Figure 4(a), the 2D Inception module is composed of four parallel flows, which are 1×1 convolution, 3×3 convolution, 5×5 convolution and a pooling layer. Finally, all parallel flows will be passed to the next layer through a concatenation filter. This crafted design allows for increasing the depth and width of the network while keeping the computational budget constant [43].



(a) (b)

Figure 4. Two-dimensional and one-dimensional Inception module. (a) 2D Inception module, (b) 1D Inception module.

Inspired by this, Fawaz et al. [11] introduced 1D Inception to the TSC problem, named it InceptionTime. As shown in Figure 4(b), the 1D Inception module is also composed of four parallel flows, which are three 1D convolution with kernel lengths equal to 10, 20, 40 respectively, as well as a pooling layer of 1D. Three convolutions share a 1D convolution output with kernel length equal to 1, which is called a bottleneck layer. This will significantly reduce the dimensionality of the time series as well as the model's complexity. The bottleneck layer can dramatically reduce overfitting problems for small datasets [11]. Similar to the 2D Inception module, the results of all parallel flows will be passed to the next layer through a concatenation filter. InceptionTime produced the current state-of-the-art results for TSC on the 85 datasets of the UCR archive [11].

3.2. Network Architecture

Inspired by Karim et al.'s LSTM-FCN [18], the network structure we proposed is shown in Figure 5. This neural network combines the attention mechanism, LSTM, and InceptionTime. The network consists of two branches, they are the Attention-LSTM branch and the InceptionTime branch. These two branches focus on different information. Attention-LSTM pays more attention to the relationship between the far-away moments in the time dimension. InceptionTime pays more attention to the extraction of key

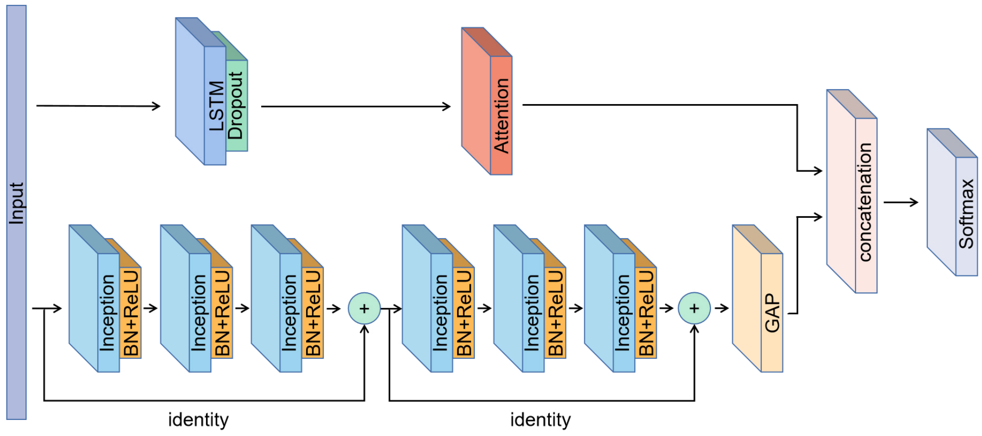


Figure 5. The structure of the AttentionLSTM-InceptionTime network.

features at close intervals. The two methods complement each other, making the extraction of hidden information in fNIRS data more comprehensive and stable.

The Attention-LSTM branch lets the input matrix pass an LSTM network with 256 cells, this will obtain a sequence of annotations $\overline{\mathbf{h}} = (\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_T)$ where the s-th element is recorded as $\mathbf{h}_s$.

The annotations will be input into a dropout layer in order to prevent the model from overfitting. For convenience of expression, there is no symbol change after the dropout layer. According to the description of global attention in Luong et al.'s paper [42], the attention mechanism is calculated by the following formulas to obtain attention hidden state $\widetilde{h}_T$:

Score function (Luong et al.'s [42] general style):

$$score(\mathbf{h}_T, \mathbf{h}_s) = \mathbf{h}_T^T \mathbf{W}_a \mathbf{h}_s \tag{3}$$

Attention weights:

$$\mathbf{a}_T(s) = align(\mathbf{h}_T, \mathbf{h}_s) = \frac{\exp(score(\mathbf{h}_T, \mathbf{h}_s))}{\sum_{s'} \exp(score(\mathbf{h}_T, \mathbf{h}_{s'}))} \tag{4}$$

Context vector:

$$\mathbf{c}_T = \sum_s \mathbf{a}_T(s) \mathbf{h}_s \tag{5}$$

Attention hidden state:

$$\widetilde{\mathbf{h}}_T = \tanh(\mathbf{W}_c[\mathbf{c}_T, \mathbf{h}_T]) \tag{6}$$

The $\widetilde{\mathbf{h}}_T$ calculated by the attention mechanism will be used as the final output of the Attention-LSTM branch. Both $\mathbf{W}_a$ and $\mathbf{W}_c$ in the above formula are parameters that the network needs to learn by training.

The InceptionTime branch is constructed according to the structure proposed by Fawaz et al. [11]. The structure of the 1D Inception module is shown in Figure 4(b). After each 1D Inception module in the InceptionTime branch, Batch Normalization [15] and a ReLU activation layer are introduced to prevent the model from overfitting. Every three 1D Inception modules form a block, and a residual connection [12] is added between the input and output of each block. The whole network is composed of two blocks including a total of six 1D Inception modules, and the final result of InceptionTime branch is output by a Global Average Pooling (GAP) layer. The results of the Attention-LSTM branch and the InceptionTime branch are finally concatenated, and the final result is output through a softmax layer.

4. EXPERIMENT

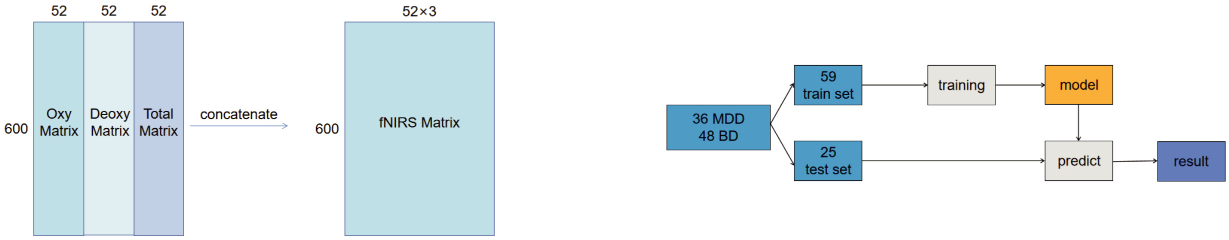
4.1. fNIRS Data Preprocessing

4.1.1. Intercept the Active Part of the VFT

Since the pre-task and post-task of the VFT contain less information, and in order to maintain reasonable length of time series, we process only the 60 s task period of the standard VFT experiment for processing. This interception method can significantly save training time.

4.1.2. Measurand Combination

Firstly, we consider HbO, HbR, and HbT separately to evaluate which hemodynamic data play a more important role in classification. Secondly, in order to utilize all the data as much as possible, we combine the three matrices which are HbO, HbR, and HbT along the channel direction, as shown in Figure 6(a). In this method, each patient can be described by a fNIRS matrix of size 600×126. Thirdly, in order to evaluate the contribution of single channel to the classification result, we take the individual channels of HbO, HbR, and HbT respectively and combine them along the channel direction. In this method, each channel for a given patient can be described by a fNIRS matrix of size 600×3, we have evaluated the classification results of each single channel combination.



(a) (b)

Figure 6. fNIRS data preprocessing. (a) The three matrices of HbO, HbR and HbT are combined along the channel direction. (b) Data partition. A total of 59 cases used as the training set, and the remaining 25 cases as the test set.

4.1.3. Normalization

As the contact for each LED light source or detector on the head of each patient may vary quite much (poor contact will give a much smaller signal), we perform zero-mean normalization on the fNIRS data by

$$x^* = (x - \mu)/std \tag{7}$$

where μ is the mean value of all sampling data in time dimension, and std is the standard deviation respectively.

4.2. Training and Validation

The algorithm reads all fNIRS data, including 36 cases of MDD and 48 cases of BD and randomly shuffle them, then select 70%, a total of 59 cases as the training set, and the remaining 30%, a total of 25 cases as the test set, as shown in Figure 6(b).

We trained the network model on Nvidia GPU RTX2080Ti and test it on a completely independent test set. We choose accuracy as evaluation indicators, and its definitions is as follows:

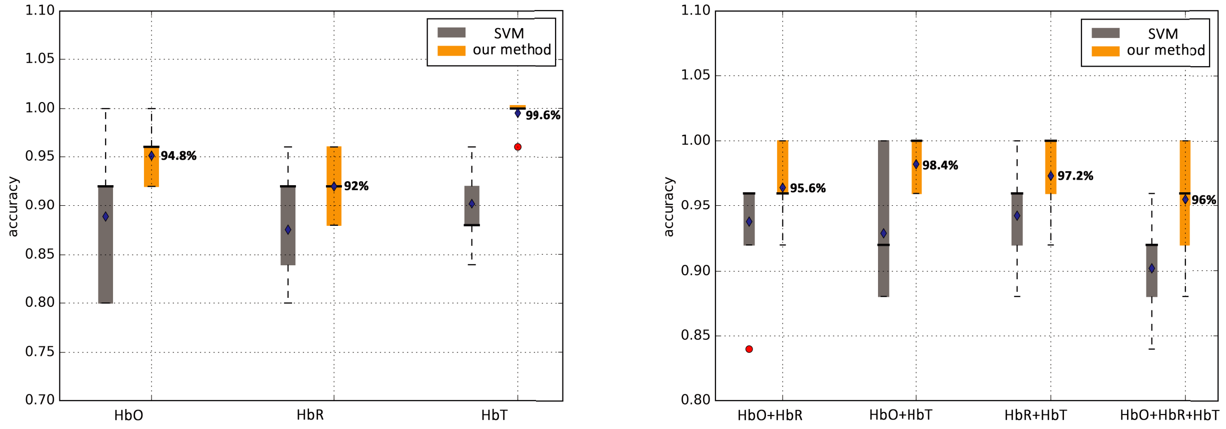
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively.

We removed the attention mechanism of ALSTMIT and named it LSTMIT as our baseline method. We denoised fNIRS data with PCA and median fliter, and then classify it by Li's support vector machine (SVM) [25]. InceptionTime, LSTM, FCN, LSTM-FCN, MLP are TSC methods with considerable performance. We also apply these methods on fNIRS data and conduct comparative experiments with ours.

5. RESULTS AND DISCUSSIONS

Figure 7(a) shows the accuracy box plot of the classification results using HbO, HbR, and HbT matrices, where each blue point represents the mean value of 10 experiments (results using our method are indicated in orange color). We compared this classification method with Li et al.'s SVM [25]. The average accuracy of the classification using our method on the HbT matrix is highest and can reach 99.6%. We can also find that the classification accuracy of our method is significantly higher than SVM. This is because our method has more powerful feature extraction capabilities. ALSTMIT is an end-to-end method, which requires no data denoising processing before starting it. Although Li's SVM performs PCA and median filter processing [25], it is still not as effective as our method. This shows that our method has the ability to suppress fNIRS noise.



(a) (b)

Figure 7. Classification results. (a) Classification results using HbO, HbR, and HbT data. Blue points represent mean values and the red points represents outliers. (b) Classification results using various combinations of HbO, HbR, and HbT along channel direction. Blue points represent mean values and the red points represents outliers.

We combined measurands HbO, HbR, and HbT along the channel direction to form four input matrices and input them into our algorithm as well as SVM. Those matrices are HbO+HbR, HbO+HbT, HbR+HbT, HbO+HbR+HbT. Figure 7(b) shows the average accuracy box plot of their classification results. The results show that our method on HbO+HbT matrix achieved the best classification performance, with average accuracy of 98.4%, which is slightly smaller than the average accuracy of using our method on HbT matrix in Figure 7(a). This shows that the measurand combination does not significantly help improve the classification performance. The channel dimension is too large after the measurand combination, which leads performance degradation.

Table 2 shows the results of fNIRS classification by different TSC methods. Our method achieved the best result on the average accuracy of seven different input fNIRS matrices, and can reach 96.2%. ALSTM-FCN took the second place, which was 0.9% smaller than ours.

Table 2. Comparison results of different deep learning TSC methods.

	HbO	**HbR**	**HbT**	**HbO+HbR**	**HbO+HbT**	**HbR+HbT**	**HbO+HbR+HbT**	**Average**
LSTM	0.88	**0.908**	0.844	0.9	0.92	0.832	0.876	0.88
MLP	0.84	0.76	0.92	0.88	0.88	0.92	0.88	0.869
FCN	0.916	0.9	0.96	0.928	0.96	0.956	0.94	0.937
InceptionTime	0.928	0.872	0.976	0.932	0.964	**0.968**	0.948	0.941
LSTM-FCN	0.916	0.904	0.976	0.932	0.952	0.96	**0.96**	0.943
ALSTM-FCN	**0.94**	0.904	**0.996**	0.928	**0.968**	0.96	**0.976**	**0.953**
LSTMIT (our baseline)	0.904	0.88	0.992	**0.94**	**0.968**	**0.968**	**0.96**	0.945
ALSTMIT (ours)	**0.948**	**0.92**	**0.996**	**0.956**	**0.984**	**0.972**	**0.96**	**0.962**
Average	0.909	0.881	0.958*	0.925	0.95*	0.942	0.938	

The first two highest scores in each column is shown in bold.

We calculated the average accuracy of the eight classification methods. HbT achieved the highest average accuracy, which is probably because HbT carries most of the useful features, and the channel dimension is not particularly high. HbT is the sum of HbO and HbR. Too many channels may cause overfitting problems. HbR had the lowest accuracy, which is 7.7% lower than that of HbT. As far as the measurand combination matrices are concerned, the accuracy of HbO+HbR is 1.6% and 4.4% higher than that of HbO and HbR, respectively. However, the average accuracies of HbO+HbT, HbR+HbT, HbO+HbR+HbT are 0.8%, 1.6%, 2% lower than that of using HbT alone, which shows that increasing the dimension simply may lead performance degradation due to overfitting.

Figure 8 shows the classification results which use our method on single channel combination data. It is found through experiments that the 32nd, 41st, 43rd, 51st, and 52nd channels have achieved considerable classification results. The 51st channel achieves 96% classification accuracy. Channels 32 and 43 are located in the left temporal lobe, while channels 41, 51 and 52 are located in the right temporal lobe. The temporal lobe is mainly related to hearing and memory, which is consistent with the functions required to perform VFT. This experimental result reflects that if patients with BD and MDD

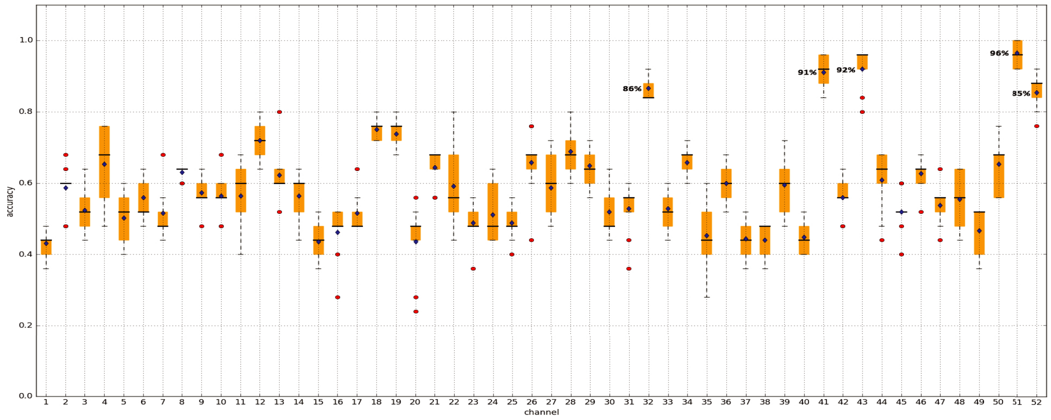
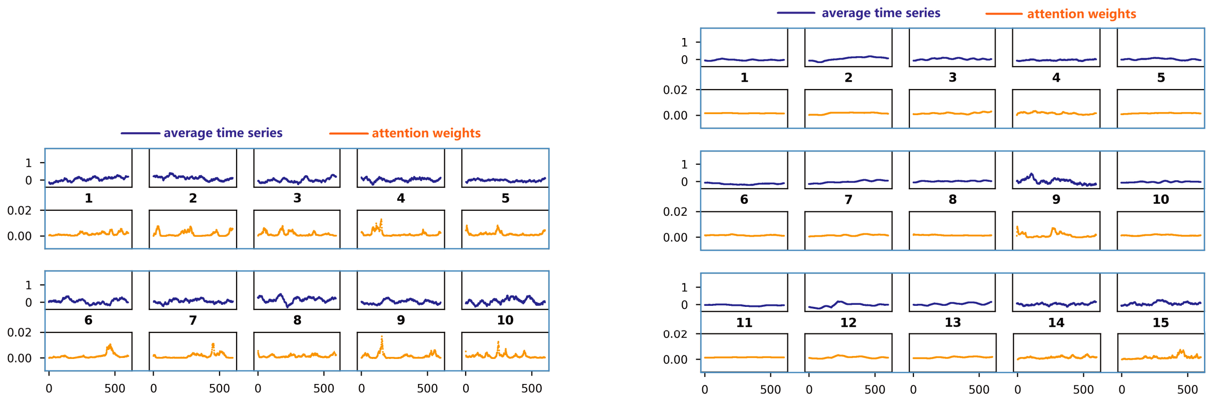


Figure 8. Individual channel classification results.



(a)

(b)

Figure 9. Average time series and attention weights. (a) Average time series (blue) and attention weights (orange) for 10 MDD cases. (b) Average time series (blue) and attention weights (orange) for 15 BD cases.

only focus on the VFT itself, there is a significant difference in the activation mode of their temporal lobe, while the prefrontal lobe related to emotion does not show a significant difference in activation mode. At the same time, experiments have shown that through VFT activation, it is possible to use fewer channels, specifically in the temporal lobe, or even one channel to distinguish BD and MDD very well, which suggests potential for the miniaturization and portability of the instruments.

The fNIRS data reflect the brain activation of the subjects under the VFT, and in machine learning, attention weights can be interpreted as the algorithm's attention distribution to the fNIRS data. Therefore, attention weights reflect the attention degree which is contributed to VFT by the subjects. In order to reveal the difference of attention distribution between BD and MDD patients under VFT, we drew the time series and attention weights of 25 cases in the test set, including 10 cases of MDD and 15 cases of BD. Specifically, we take the mean value of the HbO matrix in 52 channels to get the average time series, and we also take the attention weights in the Attention-LSTM branch. Figure 9(a) shows the average time series and their attention weights of 10 MDD patients in the test set. Similarly, Figure 9(b) shows the 15 BD patients in the test set. By comparing the average time series between BD and MDD patients, it is found that the brain activation of BD is dramatically lower than that of MDD, and the fluctuation of BD in VFT is smaller. Then, by comparing the attention weights of patients in two groups, it is found that the attention distribution of BD patients in VFT has no significant change, while the attention fluctuation of MDD patients in VFT is pronounced. This is consistent with attention deficit in BD. Previous studies have found persistent attentional deficits in BD, both in the euthymic state and depressed state. While attention of patients with unipolar depression is not significantly different from healthy controls [2, 26]. It is possible that bipolar depressed patients do not show the attentional fluctuations in the VFT that they should.

In short, the ALSTMIT network distinguishes two types of patients with a very high accuracy, and the classification algorithm is also very stable and reliable. The distribution of attention under the VFT differed between bipolar and depressed patients, suggesting that abnormalities in verbal fluency may be one of the characteristic manifestations of patients with affective disorders.

6. CONCLUSION

We have analyzed fNIRS data of 48 BD and 36 MDD patients under VFT activation. We have proposed a classification method AttentionLSTM-InceptionTime network for distinguishing BD from MDD patients. This deep learning method extracts useful BD characteristics in fNIRS data. The introduced attention mechanism allows the algorithm to pay more attention to critical characteristics, which greatly improves the accuracy and stability of classification. In 10 independent training tests, the average classification accuracy has reached 96.2%. This provides clinicians with an objective, accurate, and reliable technology for supplementary diagnosis. This is potentially helpful for the early screening for patients, and may promote the development of precision psychiatry, so that doctors can reasonably provide the precise managements in the early stage. We believe that this will improve the treatment effect significantly. However, the relative small sample size is a major limitation. The model still exhibits some overfitting due to limited sample size, although we have used many methods, including Batch Normalization, Dropout and other techniques to reduce network complexity. In the future, we hope to expand the sample size for model training in order to reduce the degree of overfitting and further improve the classification model.

ACKNOWLEDGMENT

This work was partially supported by the Key Research and Development Program of Zhejiang Province (2021C03107), the National Natural Science Foundation of China (No. 11621101), the National Key Research and Development Program of China (No. 2018YFC1407500), and the Fundamental Research Funds for the Central Universities (Zhejiang University NGICS Platform).

REFERENCES

1. Bahdanau, D., K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *ICLR 2015*, 2015.
2. Barreiros, A. R., I. A. Breukelaar, W. Chen, M. Erlinger, and M. S. Korgaonkar, "Neurophysiological markers of attention distinguish bipolar disorder and unipolar depression," *Journal of Affective Disorders*, Vol. 274, 411–419, 2020.
3. Benavides-Varela, S., D. M. Gómez, and J. Mehler, "Studying neonates' language and memory capacities with functional near-infrared spectroscopy," *Frontiers in Psychology*, Vol. 2, 64, 2011.
4. Boas, D. A., C. E. Elwell, M. Ferrari, and G. Taga, "Twenty years of functional near-infrared spectroscopy: Introduction for the special issue," *NeuroImage*, Vol. 85, 1–5, 2014.
5. Cerullo, M. A., et al., "Bipolar I disorder and major depressive disorder show similar brain activation during depression," *Bipolar Disorders*, Vol. 16, No. 7, 703–712, 2015.
6. Chen, X., Z. Wei, M. Li, and P. Rocca, "A review of deep learning approaches for inverse scattering problems (invited review)," *Progress In Electromagnetics Research*, Vol. 167, 67–81, 2020.
7. Cho, K., et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1724–1734, 2014.
8. Cristia, A., et al., "An online database of infant functional near infrared spectroscopy studies: A community-augmented systematic review," *PloS One*, Vol. 8, No. 3, e58906, 2013.
9. Dieler, A. C., S. V. Tupak, and A. J. Fallgatter, "Functional near-infrared spectroscopy for the assessment of speech related tasks," *Brain & Language*, Vol. 121, No. 2, 90–109, 2012.
10. Ehlis, A.-C., S. Schneider, T. Dresler, and A. J. Fallgatter, "Application of functional near-infrared spectroscopy in psychiatry," *NeuroImage*, Vol. 85, 478–488, 2014.
11. Fawaz, H. I., et al., "InceptionTime: Finding alexnet for time series classification," *Data Mining and Knowledge Discovery*, Vol. 34, 1936–1962, 2020.
12. He, K., X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778, 2016.
13. Hochreiter, S. and J. Schmidhuber, "Long short-term memory," *Neural Computation*, Vol. 9, No. 8, 1735–1780, 1997.
14. Hoshi, Y. and M. Tamura, "Detection of dynamic changes in cerebral oxygenation coupled to neuronal function during mental work in man," *Neuroscience Letters*, Vol. 150, No. 1, 5–8, 1993.
15. Ioffe, S. and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *Proceedings of the 32nd International Conference on Machine Learning, PMLR*, Vol. 37, 448–456, 2015.

16. Jobsis, F. F., "Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters," *Science*, Vol. 198, No. 4323, 1264–1267, 1977.
17. Jobsis-vander Vliet, F. F., "Discovery of the near-infrared window into the body and the early development of near-infrared spectroscopy," *Journal of Biomedical Optics*, Vol. 4, No. 4, 392–396, 1999.
18. Karim, F., S. Majumdar, H. Darabi, and S. Chen, "LSTM fully convolutional networks for time series classification," *IEEE Access*, Vol. 6, No. 99, 1662–1669, 2018.
19. Kato, T., A. Kamei, S. Takashima, and T. Ozaki, "Human visual cortical function during photic stimulation monitoring by means of near-infrared spectroscopy," *Journal of Cerebral Blood Flow & Metabolism*, Vol. 13, No. 3, 516–520, 1993.
20. Kawakami, K., "Supervised sequence labelling with recurrent neural networks," Ph.D. dissertation, 2008.
21. Kim, Y.-K. and K.-S. Na, "Application of machine learning classification for structural brain MRI in mood disorders: Critical review from a clinical perspective," *Progress in Neuropsychopharmacology and Biological Psychiatry*, Vol. 80, 71–80, 2018.
22. Kopton, I. M. and P. Kenning, "Near-infrared spectroscopy (NIRS) as a new tool for neuroeconomic research," *Frontiers in Human Neuroscience*, Vol. 8, 549, 2014.
23. LeCun, Y., Y. Bengio, et al., "Convolutional networks for images, speech, and time series," *The Handbook of Brain Theory and Neural Networks*, Vol. 3361, No. 10, 1995, 1995.
24. Li, Z., Y. Wang, W. Quan, T. Wu, and B. Lv, "Evaluation of different classification methods for the diagnosis of schizophrenia based on functional near-infrared spectroscopy," *Journal of Neuroence Methods*, Vol. 241, 101–110, 2015.
25. Li, Z., Y. Wang, W. Quan, T. Wu, and B. Lv, "Evaluation of different classification methods for the diagnosis of schizophrenia based on functional near-infrared spectroscopy," *Journal of Neuroscience Methods*, Vol. 241, 101–110, 2015.
26. Maalouf, F. T., et al., "Impaired sustained attention and executive dysfunction: Bipolar disorder versus depression-specific markers of affective disorders," *Neuropsychologia*, Vol. 48, No. 6, 1862–1868, 2010.
27. McIntyre, R. S., M. Berk, E. Brietzke, B. I. Goldstein, C. López-Jaramillo, L. V. Kessing, G. S. Malhi, A. A. Nierenberg, J. D. Rosenblat, A. Majeed, et al., "Bipolar disorders," *The Lancet*, Vol. 396, No. 10265, 1841–1856, 2020.
28. Molavi, B., L. May, J. Gervain, M. Carreiras, J. F. Werker, and G. A. Dumont, "Analyzing the resting state functional connectivity in the human language system using near infrared spectroscopy," *Frontiers in Human Neuroscience*, Vol. 7, 921, 2014.
29. Naseer, N. and K.-S. Hong, "fNIRS-based brain-computer interfaces: A review," *Frontiers in Human Neuroscience*, Vol. 9, 3, 2015.
30. Nguyen, D. K., et al., "Non-invasive continuous EEG-fNIRS recording of temporal lobe seizures," *Epilepsy Research*, Vol. 99, Nos. 1–2, 112–126, 2012.
31. Obrig, H., "Nirs in clinical neurology — A 'promising' tool?," *NeuroImage*, Vol. 85, 535–546, 2014.
32. O'Halloran, M., B. McGinley, R. C. Conceicao, F. Morgan, E. Jones, and M. Glavin, "Spiking neural networks for breast cancer classification in a dielectrically heterogeneous breast," *Progress In Electromagnetics Research*, Vol. 113, 413–428, 2011.
33. Onishi, A., H. Furutani, T. Hiroyasu, and S. Hiwa, "An fNIRS study of brain state during letter and category uency tasks," *Journal of Robotics, Networking and Artificial Life*, Vol. 5, No. 4, 228–231, 2019.
34. Pascanu, R., C. Gulcehre, K. Cho, and Y. Bengio, "How to construct deep recurrent neural networks," *Proceedings of the Second International Conference on Learning Representations (ICLR 2014)*, 2014.
35. Phillips, M. L. and D. J. Kupfer, "Bipolar disorder diagnosis: Challenges and future directions," *Lancet*, Vol. 381, No. 9878, 1663–1671, 2013.

36. Quan, W., T. Wu, Z. Li, Y. Wang, W. Dong, and B. Lv, "Reduced prefrontal activation during a verbal fluency task in chinese-speaking patients with schizophrenia as measured by near-infrared spectroscopy," *Progress in Neuropsychopharmacology and Biological Psychiatry*, Vol. 58, 51–58, 2015.
37. Quaresima, V., S. Bisconti, and M. Ferrari, "A brief review on the use of functional near-infrared spectroscopy (fNIRS) for language imaging studies in human newborns and adults," *Brain and Language*, Vol. 121, No. 2, 79–89, 2012.
38. Raucher-Chene, D., A. M. Achim, A. Kaladjian, and C. Besche-Richard, "Verbal uency in bipolar disorders: A systematic review and meta-analysis," *Journal of Affective Disorders*, Vol. 207, 359–366, 2017.
39. Santosa, H., M. J. Hong, and K.-S. Hong, "Lateralization of music processing with noises in the auditory cortex: An fNIRS study," *Frontiers in Behavioral Neuroscience*, Vol. 8, 418, 2014.
40. Sitaram, R., A. Caria, and N. Birbaumer, "Hemodynamic brain-computer interfaces for communication and rehabilitation," *Neural Networks*, Vol. 22, No. 9, 1320–1328, 2009.
41. Suto, T., M. Fukuda, M. Ito, T. Uehara, and M. Mikuni, "Multichannel near-infrared spectroscopy in depression and schizophrenia: Cognitive brain activation study," *Biological Psychiatry*, Vol. 55, No. 5, 501–511, 2004.
42. Luong, M.-T., H. Pham, C. D. Manning, "Effective approaches to attention-based neural machine translation," *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 1412–1421, Lisbon, Portugal, September 17–21, 2015.
43. Szegedy, C., et al., "Going deeper with convolutions," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1–9, 2015.
44. Szegedy, C., S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," *Thirty-first AAAI Conference on Artificial Intelligence*, 2017.
45. Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2818–2826, 2016.
46. Tomioka, H., et al., "A longitudinal functional neuroimaging study in medication-nave depression after antidepressant treatment," *PLoS One*, Vol. 10, No. 3, e0120828, 2015.
47. Tomioka, H., B. Yamagata, S. Kawasaki, S. Pu, A. Iwanami, J. Hirano, K. Nakagome, and M. Mimura, "A longitudinal functional neuroimaging study in medication-naive depression after antidepressant treatment," *PLoS One*, Vol. 10, No. 3, e0120828, 2015.
48. Villringer, A., J. Planck, C. Hock, L. Schleinkofer, and U. Dirnagl, "Near infrared spectroscopy (NIRS): A new tool to study hemodynamic changes during activation of brain function in human adults," *Neuroscience Letters*, Vol. 154, Nos. 1–2, 101–104, 1993.
49. Wang, S., Y. Zhang, T. Zhan, P. Phillips, Y.-D. Zhang, G. Liu, S. Lu, and X. Wu, "Pathological brain detection by artificial intelligence in magnetic resonance imaging scanning (invited review)," *Progress In Electromagnetics Research*, Vol. 156, 105–133, 2016.
50. Watanabe, E., Y. Nagahori, and Y. Mayanagi, "Focus diagnosis of epilepsy using near-infrared spectroscopy," *Epilepsia*, Vol. 43, 50–55, 2002.
51. Wise, T., J. Radua, G. Nortje, A. J. Cleare, A. H. Young, and D. Arnone, "Voxel-based meta-analytical evidence of structural disconnectivity in major depression and bipolar disorder," *Biological Psychiatry*, 2016.
52. Wolfe, J., E. Granholm, N. Butters, E. Saunders, and D. Janowsky, "Verbal memory deficits associated with major affective disorders: A comparison of unipolar and bipolar patients," *Journal of Affective Disorders*, Vol. 13, No. 1, 83–92, 1987.