# Depression Analysis and Recognition Based on Functional Near-Infrared Spectroscopy

Rui Wang, Yixue Hao, Qiao Yu, Min Chen, *Fellow, IEEE*, Iztok Humar, *Senior Member, IEEE*, and Giancarlo Fortino, *Senior Member, IEEE*

*Abstract*—Depression is the result of a complex interaction of social, psychological and physiological elements. Research into the brain disorders of patients suffering from depression can help doctors to understand the pathogenesis of depression and facilitate its diagnosis and treatment. Functional near-infrared spectroscopy (fNIRS) is a non-invasive approach to the detection of brain functions and activities. In this paper, a comprehensive fNIRS-based depression-processing architecture, including the layers of source, feature and model, is first established to guide the deep modeling for fNIRS. In view of the complexity of depression, we propose a methodology in the time and frequency domains for feature extraction and deep neural networks for depression recognition combined with current research. It is found that compared to non-depression people, patients with depression have a weaker encephalic area connectivity and lower level of activation in the prefrontal lobe during brain activity. Finally, based on raw data, manual features and channel correlations, the AlexNet model shows the best performance, especially in terms of the correlation features and presents an accuracy rate of 0.90 and a precision rate of 0.91, which is higher than ResNet18 and machine-learning algorithms on other data. Therefore, the correlation of brain regions can effectively recognize depression (from cases of non-depression), making it significant for the recognition of brain functions in the clinical diagnosis and treatment of depression.

*Index Terms*—Depression recognition, fNIRS, Feature extraction, Channel correlation.

## I. INTRODUCTION

Depression is a common condition throughout the world that can severely endanger physical and psychological health. People with depression always have a high morbidity and suicide rate, a low recognition rate, medical treatment rate and rate of being cured. It can bring serious adverse effects to patients and families [1]–[3]. Furthermore, the patient's stigma also worsens the conditions for the diagnosis and treatment of depression [4].

In clinical diagnosis, depression manifests itself with physical symptoms like anorexia, sleep disorders, pains, and psychological features such as black mood, sadness and waning interests. At present, most researchers are devoted to studying behavior information (e.g., audio, video and text data) [5]–[7] and electrophysiological data (e.g., medical examination, nuclear magnetic and electrophysiological signals) [8]–[10] based on deep-learning and machine-learning technologies to perform a deep cognition for different data [11]. Yang *et al.* put forward a multi-modal fusion framework for depression recognition based on deep neural networks by considering audio, video and text data [12]. They took the results of the Patient Health Questionnaire-9 (PHQ-9) as the ground truth and estimated the scores for the severity of depression. Some researches have shown that depression is related to the distribution characteristics of cerebral cortex signals [13]. Li *et al.* adopted a graph theory and analyzed the fundamental mechanism of the functional brain tissue of patients with depression based on EEG. They found that the brain-function network of mildly depressed people has a longer feature path and a lower convergence factor when compared to normal people. The conclusion can be used to identify mildly depression people with an accuracy of 80.74% [14], [15]. It is better to research depression with a brain-function imaging technology to reveal the pathogenesis serving the clinical diagnosis and treatment as well as deep analysis and application [16]. Functional near-infrared spectroscopy (fNIRS), as a non-invasive brain-function imaging technology with prominent merits including ease of operation, safe performance and strong resistance to noises, has been widely used in research on cognitive neuroscience for the purpose of recognizing patients with depression [17].

By using the penetrability of near-infrared (NIR) light and the absorbing ability of blood to light of different wavelengths, fNIRS reflects the relative changes in the concentration of

oxyhemoglobin (HbO) and deoxygenated hemoglobin (HbR) in the brain tissue [18]. At present, fNIRS is frequently used to analyze the conditions in the cerebral cortex, emotional problems, and the correlation between simple movements and the brain-related cognitive functions of patients suffering from mental diseases [19]. Hiwa *et al.* analyzed the ability of fNIRS in recognizing the sex of a respondent based on a convolution neural network (CNN) [20]. Similarly, Trakoolwilaiwan *et al.* recognized different movements, including rest, motion of the left and right hands with a support vector machine (SVM) and an artificial neural network (ANN) based on features including the mean value, the peak value, and the slope, and the result proved that CNN exhibited better performance in classification tasks [21]. However, it is difficult to fully describe a characteristic by merely using the fundamental features, since abundant information is contained in the hemoglobin data acquired from a variety of channels in different periods. As for the research on mental diseases, Li *et al.* detected schizophrenia with fNIRS and trained traditional machine-learning models with the features of the three kinds of NIR data obtained in each channel. Lei *et al.* extracted hematological data with the method of generalized linear regression (GLM) and trained the SVM classifier using an eigenvector to recognize depression with an accuracy of 89.71% [22]. Although machine-learning algorithms have achieved satisfactory recognition performance for mental diseases, these models have many disadvantages when fitting multi-channel multi-period data. Akiyama *et al.* studied functional activities in the prefrontal encephalic region of depressed patients while they were carrying out verbal fluency tasks. They found that depressed people had a reduced metabolism in the bilateral prefrontal encephalic region when compared to normal people. A lower metabolism in the left-lateral prefrontal cortex can reflect the recovery degree of depression with respect to black moods and lower levels of interest [23]. Therefore, it is important to be able to recognize depression with fNIRS when further exploring the pathological and physiological changes in the brain.

Currently, there are still many problems in research with fNIRS. On the one hand, in most cases, the initial data of HbO, HbR, and total hemoglobin (tHb, the sum of HbO and HbR) concentration changes were calculated as the basic features. Manual features have strong localization and high accuracy, but in the existing research the number of features is very small and the features are very simple. It is difficult to fully describe the characteristics of the data. On the other hand, due to simple features, the recognition models are generally restricted to machine-learning algorithms and shallow neural networks, leading to the result that most of the extracted features are extremely superficial. Furthermore, most of the researches focused on improving the accuracy rate of classification, paid little attention to the deep mechanism of brain disorders of depressed patients and made little use of the features of blood changes in the brain over time. As a result, such researches do not encourage future clinical studies. For these reasons, we propose a multi-dimensional recognition and modeling method based on fNIRS to identify patients with depression.

In view of the complexity of the encephalic region and the multi-dimensionality of NIR data, it is fundamental to extract multiple features from all channels to analyze the performance of the NIR data comprehensively. Moreover, further interpretation of the extracted features will help to determine the pathogenesis of brain diseases in patients with depression. Therefore, it is very critical to train the depression-recognition model with multi-feature and multi-dimensional data. The contributions of this paper are as follows:

- **A comprehensive fNIRS-based depression-processing architecture has been established.** The architecture includes three layers for data preparation, channel selection and feature extraction, as well as model construction. It provides complete flow control for processing the NIR data and using the data for depression recognition.
- **Multi-view feature analysis has been made for distinguishing between depression and non-depression.** By means of manual features from the time and frequency domains, we made a comparison between the brain functions of depression and non-depression. The results of the difference are beneficial for clinical diagnosis and treatment.
- **A deep-recognition framework has been implemented for depression classification.** Deep-learning models are built around original hemoglobin data, extracted manual features, and channel correlation to realize the depression recognition. The classification results can be used to assist doctors in assessing the condition of patients.

The rest of this paper is organized as follows: Section II reviews related work associated with fNIRS and depression recognition. Section III builds a complete processing architecture based on NIR data. Section IV introduces a methodology for the feature extraction and recognition model. Section V shows the experiments and results of the feature analysis and classification. Section VI concludes the paper and presents future work.

## II. RELATED WORK

As a non-invasive approach to brain-function detection, fNIRS has the ability to reflect the activity level of the cerebral cortex directly. At present, many researchers use fNIRS to conduct research into brain nerves.

fNIRS is used to measure the changes in hemoglobin concentration in the cerebral cortex with different activities. It has been proven that the blood-oxygen concentration in the brain will change significantly when a brain nerve receives an external stimulation. Under the stimulation of activity, the concentration of oxyhemoglobin will increase, while the concentration of deoxyhemoglobin will decrease, which indicates that fNIRS can be used to detect the changes in brain hemodynamics. Therefore, fNIRS is widely used in research on the cognitive function of the brain. Nasi *et al.* proved that the concentration of both the HbR and the HbO change during visual stimulation [24]. Nagai *et al.* verified with fNIRS that the left prefrontal cortex of the human brain will be active when a person tries to capture and hold something in the memory [25]. By means of the brain's hemodynamic differences under different experimental conditions, fNIRS can be used for the identification of cognitive functions
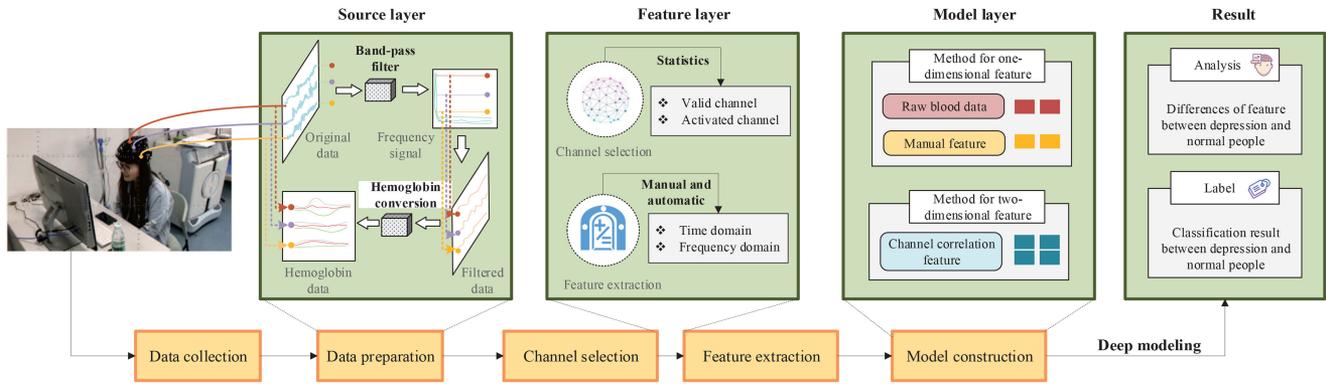
Fig. 1. Comprehensive fNIRS-based depression-processing architecture.

and activities [20] [21] [26]. Antonio *et al.* used fNIRS and EEG to recognize the left-hand and right-hand motion imaging tasks based on a deep network, which is of great significance for improving the performance of the brain-computer interface (BCI) [25]. Johannes *et al.* employed NIR data to establish a deep network model to classify brain-activation patterns. Among these researches, a variety of features are extracted after the pre-processing for fNIRS, and the classifications are executed with machine-learning and deep-learning methods [27]. Feature-extraction methods include basic features (e.g., mean, variance, kurtosis and skewness) and feature reduction based on principal component analysis, or automatic feature extraction based on CNN.

For the differences in brain function between patients with depression and normal people, many studies have explored the physiological characteristics of patients with depression using fNIRS, especially their brain function and frontal lobe abnormalities. Researches have proved that the prefrontal lobe of depressed people is less activated than that of healthy people [28] [29] [30] [31] [32] and the inter-regional functional connectivity is weaker when conducting word-fluency-based tasks [33]. It can be concluded that since the change in the hemoglobin concentration in patients with depression may be different from that in healthy people under the stimulation of activity, fNIRS has the potential to be a effective tool in the diagnosis of depression.

When fNIRS is used to recognize depression, researchers generally use the methods of feature extraction, including GLM, fast independent component analysis (FastICA) and wavelet packet decomposition (WPD) based energy for manual feature extraction [34] [35] [36]. When performing classification tasks, SVM and fisher linear discriminant analysis (FLDA) both showed good performance. Song *et al.* put forward a fNIRS-based model to recognize people with depression in the period of activity task where the features extracted manually were classified with SVM. The general accuracy rate of classification reached 86.76%, and the accuracy rate of aiming at patient classification reached 90.74% [34]. Zhu *et al.* used the discriminative model of multivariate pattern classification to recognize senile patients with depression, and the accuracy of classification reached 88% [35]. In these researches, features extracted with GLM and WPD-based energy feature extraction are generally the mean

value and the quadratic sum of hemoglobin concentration in the task period and deep features can be difficult to extract. After realizing the feature extraction, classification algorithms are always traditional machine-learning algorithms or shallow CNN, like SVM and FLDA, which are both relatively simple, making it difficult to classify effectively. We extracted more features to identify depression with an in-depth recognition model based on existing researches.

## III. A COMPREHENSIVE FNIRS-BASED DEPRESSION PROCESSING ARCHITECTURE

fNIRS relies on the different capabilities of major blood compositions in absorbing and scattering NIR light of 600–900 nm to learn about the changes of HbO and HbR in the brain for different periods. And when applying fNIRS to depression recognition, it is necessary to conduct a series of processing activities on optical data. In this section, a comprehensive processing framework is established for NIR data. To obtain clean data on hemoglobin-concentration changes for a better in-depth analysis, the fNIRS data will be pre-processed before it is formally used for feature extraction and classification.

### A. Architecture Overview

When identifying and diagnosing depression based on fNIRS data, there are five steps, i.e., data collection, data preparation, channel selection, feature extraction and model construction, as shown in Fig. 1. Among them, in several key steps, the data need to be processed step by step from the data source to feature extraction and finally to modeling. These are mainly divided into the following three layers:

- **Source layer.** At the source layer the data preparation for fNIRS is executed. Since the NIR light source data obtained from the NIR device contains a lot of noise, it is necessary to filter them first. After filtering out the noisy signal, it is transformed into an optical signal in the time domain. In order to further analyze the changes in blood flow in the brain, the changes of HbO, HbR and tHb in the same channel will be calculated through the light-intensity values for different time slots.
- **Feature layer.** At the feature layer, the channel will be selected, and then the necessary features in the selected

(a) FNIRS data before and after filtering ($\lambda$ = 690nm)

(b) Spectrum before filtering ($\lambda$ = 690nm)

(c) FNIRS data before and after filtering ($\lambda$ = 830nm)

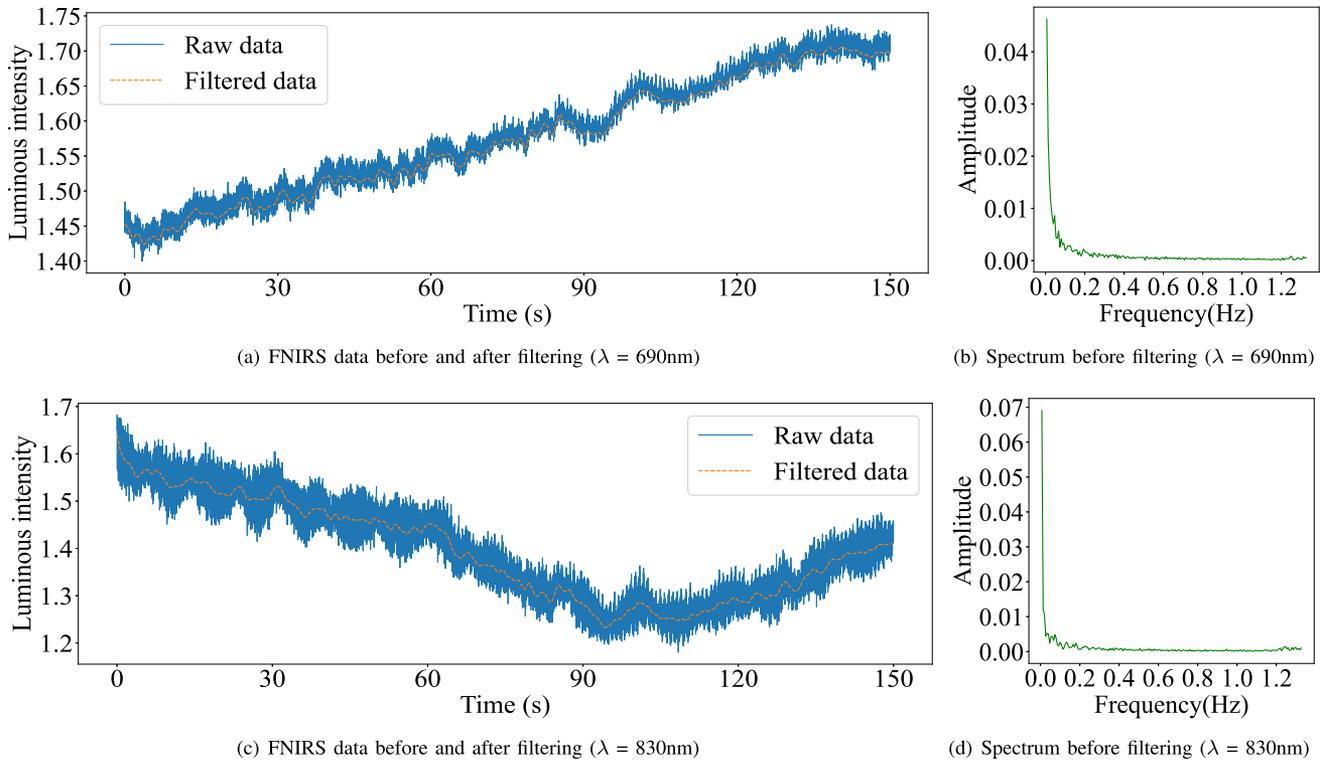(d) Spectrum before filtering ($\lambda$ = 830nm)

Fig. 2.    Band-pass filtering of fNIRS data on one channel for the two wavelengths of 690 nm and 830 nm.

channel will be extracted. When selecting the channels based on statistical theory, on the one hand, the effective channels that can collect effective signals are screened. On the other hand, the activation channels corresponding to the brain regions that are activated during the task are selected. When extracting features, the manual features at the time-domain level and the energy features at the frequency-domain level are extracted manually.

- **Model layer.** At the model layer, the depression-recognition model will be constructed. Since the features from different views are extracted in the feature layer, they will be processed to one-dimensional features and two-dimensional features, respectively. One-dimensional features include raw blood data and manual features, and two-dimensional features are mainly channel-correlation features. Combining feature dimensions and characteristics, a variety of adaptive machine-learning and deep-network methods will be used to participate in the construction of the model, so as to attain a high-precision depression-classification model.

Finally, based on deep modeling we make an analysis of the difference in characteristics and the result of the binary classification between depressed patients and normal people. By using the architecture, a complete analysis program, from fNIRS data acquisition to analysis, is implemented.

### B. Preparation for fNIRS Data

Before the data flows into the feature and model layers, it is necessary to perform pre-processing operations on the collected optical signals to complete the data preparation. The data pre-processing includes two steps: firstly, the signal-to-noise ratio of the original signal is improved by a filtering operation, and then the light-intensity data is transformed into the hemoglobin-concentration change [20][21] [26] [34].

*1) Filtering:* In the process of data acquisition with fNIRS, the physiological noises of the respondent, the external motion artifact, and the inter-optrode contact can bring lots of noise, making it necessary to remove the noise by filtering. Most of the original time-domain signals have a low-frequency band, and thus a band-pass filter was used. Signals with frequency band ranging between 0.0 and 0.6 Hz were selected, and the high-frequency signals were filtered out. Fig. 2 shows the signals and frequency spectra of the original light-intensity data with wavelengths of 690 nm and 830 nm, respectively, on the same channel, before and after the filtering, which present outstanding de-noising results.

*2) Conversion of Hemoglobin:* When the light-intensity data incident on the brain tissue using the NIR light are acquired by a specific device, it is necessary to convert the original data into the changes of hemoglobin concentration manually. The intensities of the NIR light before and after entering the cerebral cortex are set to $I_0$ and $I$, according to Beer-Lambert's law:

$$I = I_0 e^{-U(\lambda) \cdot L + g} \tag{1}$$

The output optical density (OD) is formulated as:

$$\ln \frac{I_0}{I} = U(\lambda) \cdot L + g \tag{2}$$

where $U(\lambda)$ is the absorption coefficient of wavelength $\lambda$, $L$ is the photon path length defined as $\eta \cdot \kappa$ and $g$ is the loss of light intensity, denoted as a constant. $\eta$ is the differential path length factor (DPF) and $\kappa$ is the thickness of the cerebral cortex tissue, which can be set by experience. By eliminating the baseline value of the hemoglobin at different times, we can obtain the change in the hemoglobin concentration, as follows:

$$\Delta OD = [\varepsilon_{HbO}(\lambda) \cdot \Delta c_{HbO} + \varepsilon_{HbR}(\lambda) \cdot \Delta c_{HbR}] \cdot L \quad (3)$$

where $\Delta OD$ can be known from the measured output optical intensity, both the extinction coefficient $\varepsilon$ and the photon path length $L$ are constants [37] [38]. By selecting NIR lights of different wavelengths, the concentration changes of the HbO and HbR can be obtained. To minimize the error, the mean value in a specific beginning period was chosen and defined as the initial concentration and the later concentrations in whichever moment was formulated as the initial concentration adding the mean value. Finally, the changed values of HbO, HbR and tHb are obtained for use in subsequent research.

## IV. METHODOLOGY FOR THE EFFECTIVE FEATURE AND IDENTIFIED MODEL

In order to obtain effective features, channel selection and key feature extraction are important steps. Based on the critical features of effective channels, a recognition model can be established to make a classification for a patient with depression and for non-depressed people.

### A. Channel Selection

There are two main reasons for the poor activation of brain regions corresponding to some channels when respondents are going through tests with the NIR device on their heads. Firstly, the contact points of different channels of the NIR equipment do not fit the scalp well, leading to low signal-to-noise ratio and the large curve fluctuation of channels. These channels are called the invalid channels. When wearing an NIR device, it will ensure that the number of effective channels is greater than a certain number to start the experiment. Secondly, when the respondents take part in the testing, some regions of their brains will be activated. However, some other brain regions will not participate in the task, and thus the overall extent of the activation is poor. And an intuitive result is the lack of obvious contrast between the task period and the silent period for both the non-depressed people and depressed patients. Channels located in the activated brain region are mentioned as activated channels.

When building the depression-recognition model with machine-learning algorithms, if features are extracted manually after picking out the valid and activated channels, the value of the extracted features will be improved significantly. The *T-test* is utilized to analyze the statistical differences for the valid and activated channels in the task period and the silent period of the hemoglobin concentration when compared to the HbR concentration, so as to extract valid and activated channels. The *T-test* is the most essential hypothetical test method that is capable of determining whether two datasets with the same variance have the same mean value, which is featured with easy

operation and rapid computation. The data from the silent period and the task period of each test stage are selected as two groups of samples, and the *T-test* method is adopted to judge whether there was an obvious difference in the hemoglobin concentration between the silent period and the task period. Next, the HbR will be taken as an example to explain how to calculate the value of $T$.

Firstly, the data from the silent period $t_1$ and the task period $t_2$ of each channel are selected and marked as $HbR(t_1)$ and $HbR(t_2)$ respectively. Next, we define $HbR(t_1) \sim N(\mu_1, \sigma^2)$, and $HbR(t_2) \sim N(\mu_2, \sigma^2)$, then the combined variance $S_c{}^2$ is formulated as follows:

$$S_c{}^2 = \frac{(n_1 - 1)S_1{}^2 + (n_2 - 1)S_2{}^2}{n_1 + n_2 - 2} \quad (4)$$

where $n_1$ and $n_2$ are the number of samples of $HbR(t_1)$ and $HbR(t_2)$, respectively, while $S_1{}^2$ and $S_2{}^2$ are the variances of $HbR(t_1)$ and $HbR(t_2)$. Thus, the difference $T(t_1, t_2)$ between HbR in the silent period $t_1$ and the task period $t_2$ can be obtained with a calculation:

$$T(t_1, t_2) = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_c{}^2(\frac{1}{n_1} + \frac{1}{n_2})}} \quad (5)$$

where $\bar{X}_1$ and $\bar{X}_2$ are the sample means of $HbR(t_1)$ and $HbR(t_2)$, respectively. The value of $T(t_1, t_2)^{HbR}$ and the activation difference between the silent period and the task period are proportional. According to the critical value table of the *T-test*, when the significance level is set to 0.05 and the degree of freedom is 90, the two-sided critical value is 1.987. When the absolute value of $T(t_1, t_2)$ is greater than 1.987, the HbR concentration in the silent period and the task period shows a statistical difference, indicating that the channels are activated.

Similarly, the value of $T(t_1, t_2)^{HbO}$ can be determined. Compared to the silent period, the concentration of HbO and HbR in the task period increases and decreases, respectively. Therefore, $T(t_1, t_2)^{HbO}$ will be negative, while $T(t_1, t_2)^{HbR}$ will be positive. For the purposes of convenience, we define the $T(t_1, t_2)^{HbO}$ to be the opposite number of itself. Based on this, the $T$ value of each sample on HbR and HbO can be calculated. If the value is greater than 1.987, it is considered to be an active channel. For a single sample, the final activated channel depends on the intersection of the activated channels on HbR and HbO, so that the activated channels for each sample are obtained.

### B. Key Feature Extraction

The brain's hemodynamic signals acquired by fNIRS contain lots of physiological information, but direct data observation cannot reflect the difference between depressed patients and non-depressed people. Therefore, we need to perform a key feature extraction to explore the differences. As shown in Table I, a total of 67 features, including ten kinds of features corresponding to the three indicators of HbO, HbR and tHb, are acquired.

When initiating an experiment with a fNIRS device, the tester will go through three periods of before-task-silent, on-task and after-task-silent. And the completion of three periods can also be regarded as a whole period. The feature of *total* means the

TABLE I

MANUAL FEATURE EXTRACTION FOR CHANNELS

| Feature | Description | Number |
|---|---|---|
| Total | Sum of hemoglobin concentration changes in the whole process. | 1 |
| Peak | Peak value of hemoglobin concentration changes in four periods. | 4 |
| Valley | Valley value of hemoglobin concentration changes in four periods. | 4 |
| Average | Mean value of hemoglobin concentration changes in four periods. | 4 |
| Variance | Variance of hemoglobin concentration changes in four periods. | 4 |
| Integral | Integral of hemoglobin concentration changes in four periods. | 4 |
| Linear | Slope and intercept of hemoglobin concentration changes in three periods. | 6 |
| Quadratic term | Quadratic-term fit coefficient of hemoglobin concentration changes in three periods. | 9 |
| Power spectrum | Power spectrum amplitudes of hemoglobin concentration changes at five frequency bands. | 5 |
| Wavelet coefficient | Related entropy value of coefficients after wavelet decomposition. | 25 |

sum of hemoglobin (HbO, HbR and tHb) concentration changes in the whole process. Similarly, the features of *peak*, *valley*, *average*, *variance*, *integral* mean the peak, valley, average, variance and integral of hemoglobin concentration changes in the four periods (before-task-silent, on-task, after-task-silent and the whole period). In the meantime, linear and quadratic fittings are applied to the curve of hemoglobin changes in three periods. Considering the complex changes through the whole period, they are not fitted to the whole period. The feature of *linear* includes two values of slope and intercept in one period. The feature of *quadratic term* includes three values of quadratic coefficient, the first coefficient and the constant in one period. The above are all features in the time domain. In the frequency domain, the features of *power spectrum* and *wavelet coefficient* are extracted. The *power spectrum* contains five power features in the frequency band between 0 and 0.6 Hz. First, the maximum power values of the signals in the two frequency bands 0.01–0.25 Hz and 0.25–0.5 Hz are obtained as two features $P_\nu^1$ and $P_\nu^2$. Then, using the maximum power value minus the signal power values represented as $P_{\nu 0}^1$ and $P_{\nu 0}^2$ at the initial frequency values of 0.01 Hz and 0.25 Hz, respectively, as the other two features, denoted as $P_\nu^1 - P_{\nu 0}^1$ and $P_\nu^2 - P_{\nu 0}^2$. Finally, the ratio of the two features after subtracting the initial power spectrum is expressed as the fifth feature $(P_\nu^2 - P_{\nu 0}^2)/(P_\nu^1 - P_{\nu 0}^1)$. Since fNIRS signals are non-stable random signals with a time-varying characteristic, the entropy features can be extracted. The wavelet *db6* was adopted to decompose the signals and five entropies in five frequency bands were obtained, including *shannon entropy*, *energy entropy*, *paradigm entropy*, *threshold entropy*, and *sure entropy*, which are denoted as $W_{sh}$, $W_e$, $W_p$, $W_t$, $W_{su}$ and defined as follows:

$$W_{sh} = -\sum_x \gamma_x^2 \cdot log\gamma_x^2 \qquad (6)$$

$$W_e = \sum_x \gamma_x \qquad (7)$$

$$W_p = \sum_x |\gamma_x|^q - \|\gamma\|_q^q \qquad (8)$$

$$W_t = \sum_x \begin{cases} 1 & |\gamma_x > q| \\ 0 & otherwise \end{cases} \qquad (9)$$

$$W_{su} = m - \sum_x \begin{cases} 1 & |\gamma_x \le q| \\ 0 & otherwise \end{cases} + min(\gamma_x^2, q^2) \qquad (10)$$

where $\gamma$ is the value of signal, $q$ is the threshold, and $m$ is the number of signals. Therefore, there are a total of 25 features extracted from the wavelet decomposition. It should be noted that the above 67 features can be obtained in the signal generated by each channel of the device.

As for the analysis on channel correlation, the whole periods were divided into three parts: before-task-silent, on-task, and after-task-silent, and *Pearson's correlation coefficient* was used to calculate the channel correlation between depressed patients and non-depressed people during different periods. If the value of $p$ is tested to be smaller than 0.05, obvious correlation exists between the two channels of the NIR device mapping to the brain. Thus, the value of $r$ was included into the calculation of the correlation to the corresponding group (non-depression and depression). After obtaining the sum of the correlation results for the samples meeting significant relevance, the mean correlation was calculated to find the channel where the non-depression group and the depression group showed strong a correlation in different periods. The mean correlation $\mathcal{R}_{i,j}^{\alpha,\tau}$ of group $\alpha \in$ {Depression, Non-depression} in the period $\tau \in$ {before-task-silent, on-task, after-task-silent} at the channel $i$ and channel $j$ is:

$$\mathcal{R}_{i,j}^{\alpha,\tau} = \frac{\sum_{m=1}^M r_{m,i,j}^{\alpha,\tau}}{N_\alpha}, M \in \{p_{i,j}^{\alpha,\tau} < 0.05\} \qquad (11)$$

where $N_\alpha$ is the number of samples in group $\alpha$. In this paper, the threshold value of the correlation strength (strong or weak) was defined manually. Highly correlated channels will be obtained to illustrate the gap between the depressed patients and the non-depressed people. At the same time, the value of the channel correlation can be used to identify patients with depression.

### C. Construction of the Recognition Model

Since different feature data have different characteristics, differentiated models will be used for the training and learning. As shown in Fig. 3, it is a deep-recognition framework for depression classification. According to the results of the feature extraction, three kinds of data, i.e., raw blood data,
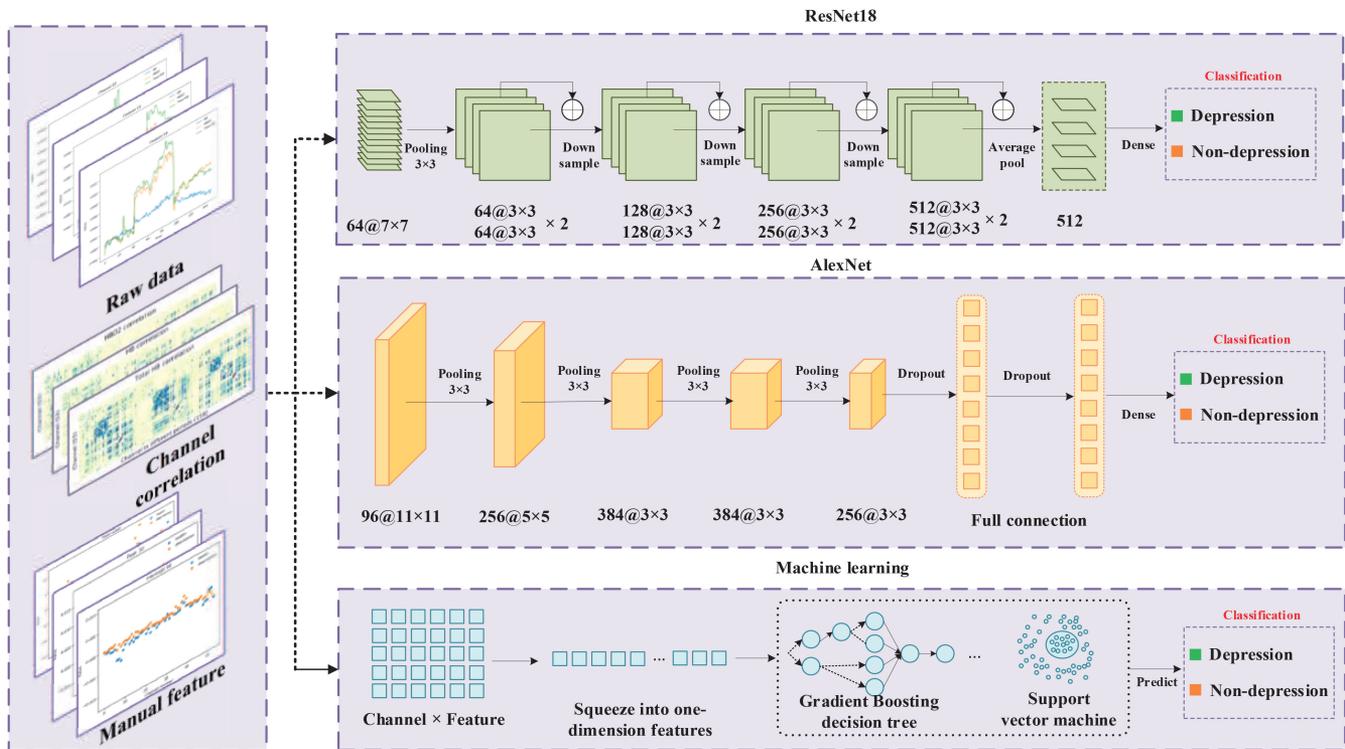
Fig. 3. Deep-recognition framework for depression classification.

manual features and channel correlation, can be the input to the classification model.

We selected three models, i.e., ResNet18 [39], AlexNet [40] and machine-learning algorithms [41] like the gradient-boosting decision tree (GBDT) and SVM. The ResNet is a network composed of a series of residual blocks. It was proposed to solve the degradation problem of deep networks after increasing the number of network layers. Compared with ordinary networks, it forms residual learning by increasing the depth of the network and adding a short-circuit mechanism between every two layers, thereby greatly improving the performance. ResNet18, containing 18 network layers, is the shallowest network in the existing ResNet networks. The settings of the specific network parameters are shown in Fig. 3. At the end, the fully connected layer containing 512 neurons will be mapped to the binary classification result. AlexNet is a shallow classic convolutional neural network model, which is an efficient GPU convolution operation structure. It solves the over-fitting problem of deep neural networks through many skills, such as dropout and ReLU, and uses a local response normalization layer to improve the accuracy. The AlexNet network contains eight weight layers, including five convolutional layers and three fully connected layers. At the end, the result of the fully connected layer is also mapped to the result of the classification. With regard to machine-learning algorithms, they are good at handling the small features of data, and have achieved good results in many fields due to their simplicity and efficiency. Typical algorithms like SVM and GBDT are used for depression classification.

The raw data keeps the most original features of the blood changes in the brain regions of the sample, which reflects directly the changes in the hemoglobin concentration of depressed patients and normal people in different periods of NIR testing. The raw data contains changes of HbO, HbR and tHb in different channels (corresponding to different brain regions) during a period. By extracting the important physiological features from raw data, we can further identify the differences between depressed patients and normal people. Thus, feature models can be established with physiological features to recognize depression. The feature data contains 67 features of changes in HbO, HbR, and tHb concentrations in different channels. As for different periods of NIR data acquisition, we found significant differences between depressed people and non-depressed people in the channel correlation of the corresponding brain regions. Thus, channel correlation can be defined as an important feature to recognize depression. Correlation data includes the inter-channel correlation coefficients of HbO, HbR, and tHb in different periods. Based on the data characteristics, three modeling methods can be adaptively applied to different data.

## V. EXPERIMENTS AND RESULTS

Related experiments were performed based on the dataset obtained from a hospital to verify the effectiveness of the established architecture and methods.

### A. Dataset

The data comes from the Renmin Hospital of Wuhan University, Wuhan, China from March to October 2018. The NIR device has 16 NIR emitting probes and receiving probes, embedding 53 channels. The probe emits NIR light with two wavelengths:

TABLE II
NUMERICAL VALUE OF THE EXTINCTION COEFFICIENT AND THE PHOTON
PATH LENGTH

| $\lambda$ | $\eta$ | $\kappa$ | $L$ | $\varepsilon_{HbO}$ | $\varepsilon_{HbR}$ |
|---|---|---|---|---|---|
| 690nm | 6.51 | 3 | 19.53 | 0.3123 | 2.1382 |
| 832nm | 5.86 | 3 | 17.38 | 1.0621 | 0.7804 |

690 nm and 830 nm. For NIR light of two wavelengths and the 53 channels, digital gains that are to regulate the impulse amplitude of the digital-to-analogue conversion input and analog gains, which is to regulate the signal strength of linear amplified input, are available to guarantee receiving NIR signals of higher quality.

When collecting the NIR data, the doctor firstly helped the respondent put on the NIR device to make sure the probe fitted the scalp tightly, until the passing rate of the channel reaches 80%. The respondent was arranged to sit in front of a computer with the device on the head. The testing lasted for 150 seconds (s), including the before-task-silent period (30 s), the on-task period (60 s), and the after-task-silent period (60 s). In the silent period, the tester was required to sit upright in front of the computer and remain calm. In the task period, three questions appearing on the computer screen needed an answer from the respondent. It required the tester to speak the names of electric apparatuses, fruits and vegetables that they can associate and the answering time for each question is 15 s. During this period, the NIR equipment was collecting the intensity of the emitted light of two wavelengths at a sampling rate of 100 Hz. At the end of testing, a data size of $15000 \times 53 \times 2$ was generated for a respondent, representing the strength of the two kinds of near-infrared light in the 53 channels at 15 000 time-points, respectively.

In total, 96 samples were acquired, including 79 samples from the depression group and 17 samples from non-depression group. The respondents were determined as having depression based on their results from the scale PHQ-9. According to the instructions of PHQ-9 and suggestions from a professional doctor, testers with scores from zero to nine would be admitted to the non-depression group, while patients with scores of from ten to twenty-seven would be admitted to the depression group. All the personal information in these samples in the dataset were desensitized.

### B. Experimental Setting

The experimental parameters focus on the basic conversion of the source layer for data pre-processing and the parameters and training environment of the recognition model in the model layer. When converting optical data into hemoglobin-concentration-change data, the conversion parameters that need to be used at the two wavelengths of 690 nm and 832 nm are shown in Table II. Meanwhile, for the deep-network model, it is necessary to set a proper learning rate and loss function to guarantee the convergence of the network so that the back-propagation of the network can finish updating the parameters. When training ResNet18 and AlexNet on the model layer, the loss function is Cross Entropy Loss. The Adam Optimizer is used to adaptively adjust different

learning rates according to different parameters to complete the parameter updates. The learning rate is $1e-3$, the coefficient for calculating the running average of the gradient and the square of the gradient is $(0.9, 0.99)$, and the weight attenuation is 0 by default. At the same time, when the machine-learning algorithm is used for classification, the SVM algorithm adjusts its kernel to RBF, and the remaining hyperparameters are tuned by using the grid-search method. As for the evaluation on recognition models, the *Accuracy*, *Precision*, *Recall* and *F1-score* are used as evaluation indexes for the performance of the model.

### C. Results on the Feature Layer

In the feature layer, channel selection and feature extraction are performed, and the differences between depression and non-depression are compared in terms of three aspects: brain activation, manually extracted features, and channel correlations.

*1) Difference in Brain Activation:* We compared the brain activation of depressed patients and non-depressed people based on HbO and HbR, respectively. By comparing and analyzing the activation of 53 channels, we found that: 1) the number of activated channels in depressed patients is less than that of non-depressed people; 2) In the activated channels, the $T$ values of depressed patients are generally lower than those of the non-depressed, i.e., among the activated channels, the activation level of depressed people is lower; 3) the value of $T(t_1, t_2)^{HbO}$ is generally higher than the value of $T(t_1, t_2)^{HbR}$, and the difference between non-depressed and depressed is greater, indicating that the change degree of HbO concentration during the task is higher than that of HbR.

*2) Difference in Manual Features:* In order to analyze the differences of physiological characteristics between patients with depression and non-depressed people, 67 features were compared in HbR, HbO and tHb. We found that the four features *total*, *average*, *variance* and *paradigm entropy* have no difference between the depression and the non-depression. Some features of non-depressed people like *peak*, *linear*, *quadratic term*, *power spectrum* and *shannon entropy* are higher than those of depressed patients, while *valley*, *threshold entropy*, *energy entropy* and *sure entropy* are lower. In addition, for the feature *integral*, some periods are higher and some periods are lower. As shown in Fig. 4, the violin figures of some features are presented. It can be seen that for the *valley* during four periods and *energy entropy*, the distribution of non-depressed people is much more concentrated than that of depressed patients. For *shannon entropy*, the median value of non-depression is always greater than that of depression. Therefore, it can be concluded that there are obvious differences in the direct and indirect characteristics of hemodynamics between patients with depression and normal people, which may be caused by the low degree of brain activation in patients with depression.

*3) Difference of Channel Correlations:* Table III has statistics on the quantity of channels at which the correlation of the depression group and the non-depression group was greater than 0.6. It can be seen that in the periods of before-task-silent, on-task, and after-task-silent, the quantity of correlation channels of the non-depression group was greater than that of
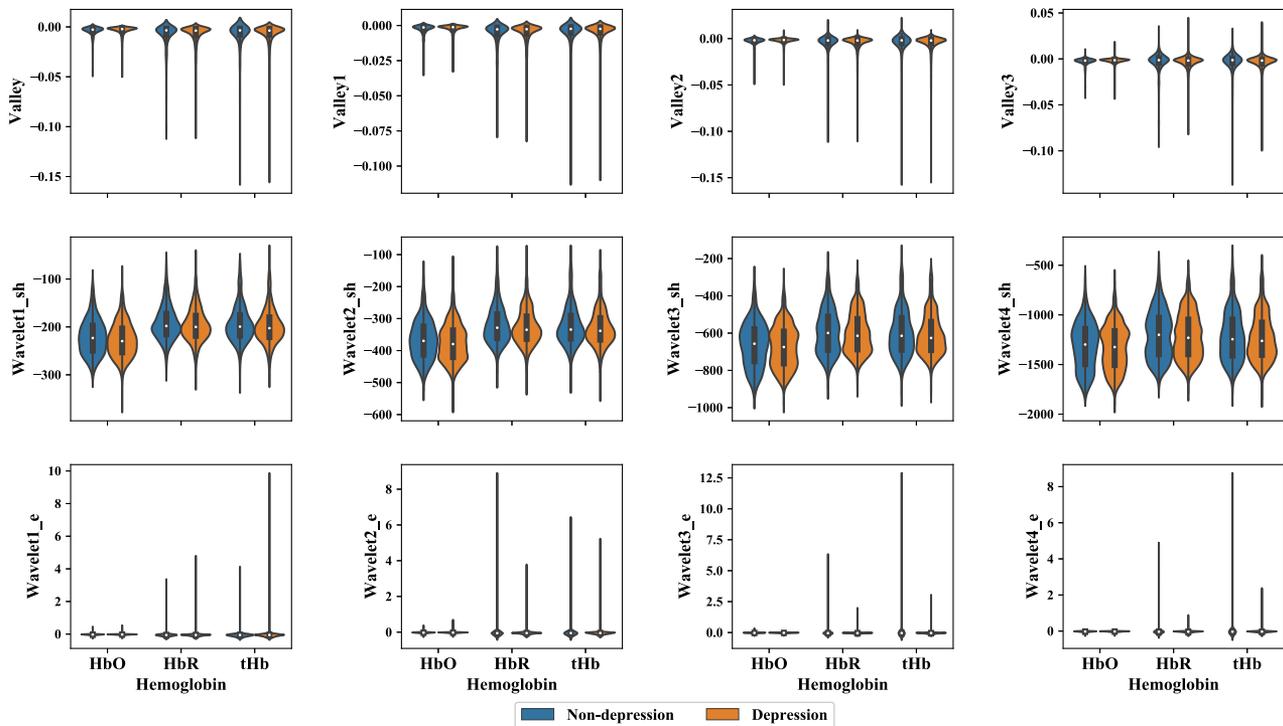
Fig. 4.    Comparison of typical features of HbO, HbR and tHb between depression and non-depression. a). The First row is the valley in the different four periods (whole process, before-task-silent, on-task and after-task-silent). b). The second and third rows are the two wavelet features (*shannon entropy* and *energy entropy*) in the four different frequency bands.

TABLE III
STATISTICS FOR THE NUMBER OF CHANNELS WITH STRONG CORRELATION IN THE DEPRESSION GROUP AND NON-DEPRESSION GROUP DURING THE THREE PERIODS

| Group | Before-task-silent | On-task | After-task-silent |
|---|---|---|---|
| Depression | 17 | 24 | 20 |
| Non-depression | 30 | 93 | 29 |

TABLE IV
RESULTS OF EVALUATION WITH THE IN-DEPTH DEPRESSION-RECOGNIZING MODEL OF MULTI-DIMENSIONAL DATA

| Input | Model | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|---|
| Raw data | AlexNet | 0.72 | **0.80** | 0.72 | **0.75** |
| | ResNet18 | **0.76** | 0.67 | **0.76** | 0.71 |
| Feature | AlexNet | **0.83** | **0.79** | **0.83** | **0.80** |
| | ResNet18 | 0.72 | 0.67 | 0.72 | 0.70 |
| | SVM | **0.83** | 0.68 | **0.83** | 0.75 |
| | GBDT | **0.83** | 0.68 | **0.83** | 0.75 |
| Correlation | AlexNet | **0.90** | **0.91** | **0.90** | **0.88** |
| | ResNet18 | 0.83 | 0.68 | 0.83 | 0.75 |

the depression group, especially in the task period. Fig. 5 shows the distribution of the 53 channels of the NIR device and the channel correlation during three periods. In each grid, the bottom parts are fitted to the forehead of the respondent, while the upper parts are close to the back of the head of the respondent. It is clear that the depression group had sparse channel correlation, while channels 38, 39 and 40 near to the rear left of the head showed strong correlation. As for the non-depression group, the channel correlation was strong, especially in the forehead area in the task period but weak in the others and, similarly, the correlation for channels 38, 39, and 40 was weaker than that of the depression group. Therefore, it can be concluded that obvious differences exist between the non-depression group and the depression group in terms of channel correlation.

## D. Results on Model Layer

The model layer deploys recognition models to identify depression and non-depression using the features of 96 sets of data to train the model. Training sets and testing sets were classified at the rate of 7:3 and for the purpose of ensuring the balance of

different types of sample proportion between the training set and the testing set, the principle of hierarchical sampling is adopted.

*1) Results of Raw Data:* For a sample, the dimension of raw data is $3 \times 15000 \times 53$. During the training of a deep network, it was deemed as a matrix with a length and width of 15 000 and 53, respectively, for 3 channels. The hemoglobin model is trained by the Alexnet and ResNet18 networks, respectively, and the iteration number is set to 500, and the models have all converged. The results of the evaluation on the testing set after the end of the training are shown in Table IV. The accuracy rate and recall rate of ResNet18 was 0.76, higher than those of AlexNet, while the *Accuracy* and *F1-score* of AlexNet were 0.80 and 0.75. In short, ResNet18 showed better performance than AlexNet in recognizing depression on the raw data.

*2) Results of Manual Features:* For a sample the dimension of the manual feature is $3 \times 53 \times 67$. When training the deep
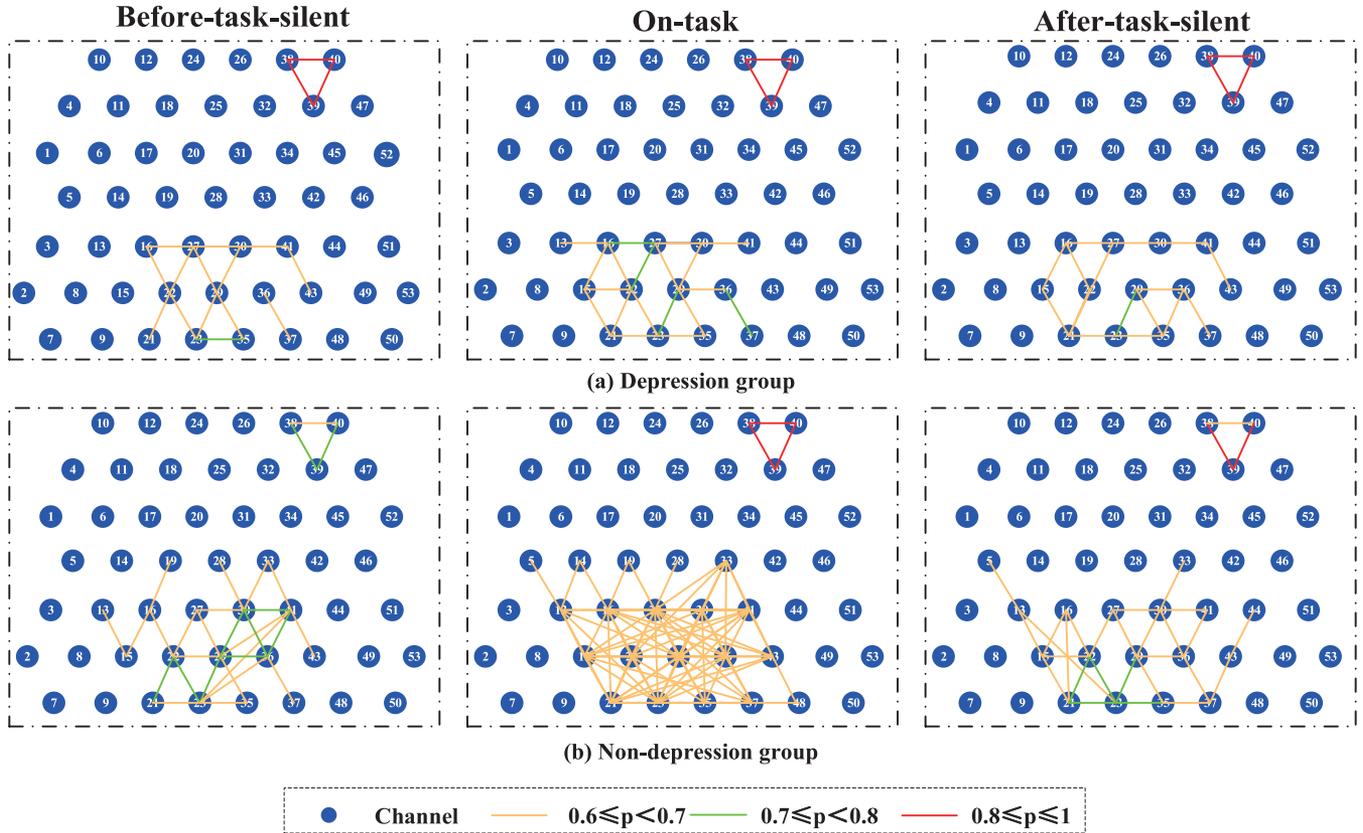
Fig. 5.    The channel correlations of the depressed and non-depression group in the three periods.

network, it was also treated as a matrix with a length of 53 and a width of 67 in three channels. The hemoglobin model is trained by the AlexNet and ResNet18 networks, respectively, and the iteration number is set to 500, and the models all converged. Furthermore, the machine-learning model SVM is used for the training, and the feature of each sample is compressed to 10 653. In the training of the SVM, according to the activated channel and important features, an additional channel and feature-extraction work is carried out. The results of the evaluation on the testing dataset after the end of training are shown in Table IV. The accuracy rate of Alexnet and SVM is 0.83, and the *Precision*, *Recall* and *F1-score* of AlexNet are the highest, i.e., 0.79, 0.83 and 0.80 respectively. Therefore, AlexNet showed the best performance among the recognizing models in the aspect of feature data.

*3) Results of Correlation Features:* For a sample the dimension of the correlation features is $3 \times 53 \times 159$, where 159 is the correlation coefficient between each two of the 53 channels in periods of before-task-silent, on-task, and after-task-silent. The *Accuracy*, *Precision*, *Recall* and *F1-score* of AlexNet on the testing dataset are 0.90, 0.91, 0.90 and 0.88, which are better than ResNet18.

According to these classification results of different models for three kinds of data, it is proved that Alexnet, on the correlation data, has the best performance, which can be used to assist doctors in judging the condition of patients.

## VI. CONCLUSIONS AND FUTURE WORK

fNIRS is an important tool used for research on brain function. Due to the complexity of fNIRS data, this paper describes a comprehensive depression-processing architecture to realize the pre-processing, feature extraction and recognition with NIR data. By conducting a comparative study on the features of patients with depression and non-depressed people, we found that the patients have a weaker brain functional connectivity in the prefrontal cortex and poor activation of brain functions during a task. It can also be seen that significant differences exist between the depressed patients and the non-depressed people in multi-dimensional features, such as the peak, valley and entropy values of hemoglobin-concentration changes. Furthermore, by using the extracted different features, including raw data, manual features, and channel correlation, a deep-recognition framework for depression classification was established, and three models ResNet18, SVM as well as AlexNet, and AlexNet showed the best performance for the three kinds of data with the recognition *Accuracy* of 0.76, 0.83 and 0.90. The channel-correlation feature showed the highest accuracy rate with 0.90. In short, there are great differences between depressed patients and non-depressed people in terms of brain-region correlation. The findings and established recognition models will be of great significance for a clinical diagnosis and treatment of depression.

Some limitations exist in this research, which need to be further study in future work. There are only 96 pairs of samples used in the analysis and recognition, including 79 depression and 17 non-depression. Since the imbalance of the dataset may cause an inaccuracy in the statistical analysis the over-fitting of the training model, it is necessary to collect more data to carry out the experiments. At the same time, through the analysis of this article, the difference in channel correlation between the depressed patients and the non-depressed people is obtained. In the future, we can fully consider the spatial relationship of the channel and use a model that is sensitive to spatial distribution, like a graphic neural network, to carry out further research.

## REFERENCES

[1] M. Marcus, M. T. Yasamy, M. v. van Ommeren, D. Chisholm, and S. Saxena, "Depression: A global public health concern," *World Health Organization Paper Depression*, vol. 1, pp. 6–8, 2012.

[2] J. LeMoult and I. H. Gotlib, "Depression: A cognitive perspective," *Clin. Psychol. Rev.*, vol. 69, pp. 51–66, 2019.

[3] P. Pace, G. Aloi, R. Gravina, G. Caliciuri, G. Fortino, and A. Liotta, "An edge-based architecture to support efficient applications for healthcare industry 4.0," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 481–489, Jan. 2019.

[4] M. Fava, "Diagnosis and definition of treatment-resistant depression," *Biol. Psychiatry*, vol. 53, no. 8, pp. 649–659, 2003.

[5] S. Yin, C. Liang, H. Ding, and S. Wang, "A multi-modal hierarchical recurrent neural network for depression detection," in *Proc. 9th Int. Audio/Vis. Emotion Challenge Workshop*, 2019, pp. 65–71.

[6] X. Ma, H. Yang, Q. Chen, D. Huang, and Y. Wang, "DepAudioNet: An efficient deep model for audio based depression classification," in *Proc. 6th Int. Workshop Audio/Vis. Emotion Challenge*, 2016, pp. 35–42.

[7] G. Aloi, G. Fortino, R. Gravina, P. Pace, and C. Savaglio, "Simulation-driven platform for edge-based aal systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 2, pp. 446–462, Feb. 2021.

[8] Y. Jiang *et al.*, "Recognition of epileptic eeg signals using a novel multiview tsk fuzzy system," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 1, pp. 3–20, Feb. 2017.

[9] E. Brown, N. D. Forkert, L. Marcil, A. S. Talai, and R. Ramasubbu, "S116. the use of arterial spin labeling perfusion MRI for automated classification of major depression disorder," *Biol. Psychiatry*, vol. 83, no. 9, 2018, Art no. S392.

[10] G. Fortino, D. Parisi, V. Pirrone, and G. Di Fatta, "BodyCloud: A SaaS approach for community body sensor networks," *Future Gener. Comput. Syst.*, vol. 35, pp. 62–79, 2014.

[11] M. Chen, Y. Hao, H. Gharavi, and V. C. Leung, "Cognitive information measurements: A new perspective," *Inf. Sci.*, vol. 505, pp. 487–497, 2019.

[12] L. Yang, D. Jiang, X. Xia, E. Pei, M. C. Oveneke, and H. Sahli, "Multimodal measurement of depression using deep learning models," in *Proc. 7th Annu. Workshop Audio/Vis. Emotion Challenge*, 2017, pp. 53–59.

[13] R. J. Davidson, "What does the prefrontal cortex "do" in affect: Perspectives on frontal EEG asymmetry research," *Biol. Psychol.*, vol. 67, no. 1-2, pp. 219–234, 2004.

[14] X. Li, R. La, Y. Wang, B. Hu, and X. Zhang, "A deep learning approach for mild depression recognition based on functional connectivity using electroencephalography," *Front. Neurosci.*, vol. 14, p. 192, 2020.

[15] M. Fraschini, M. Demuru, A. Crobe, F. Marrosu, C. J. Stam, and A. Hillebrand, "The effect of epoch length on estimated EEG functional connectivity and brain network organisation," *J. Neural Eng.*, vol. 13, no. 3, 2016, Art. no. 036015.

[16] M. Chen *et al.*, "Living with i-fabric: Smart living powered by intelligent fabric and deep analytics," *IEEE Netw.*, vol. 34, no. 5, pp. 156–163, Sep./Oct. 2020.

[17] A. Villringer, 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," *Neurosci. Lett.*, vol. 154, no. 1/2, pp. 101–104, 1993.

[18] F. N. Jobsis, "Infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters," *Science*, vol. 198, pp. 1264–1267, 1977.

[19] Y. Hoshi and M. Tamura, "Detection of dynamic changes in cerebral oxygenation coupled to neuronal function during mental work in man," *Neurosci. Lett.*, vol. 150, no. 1, pp. 5–8, 1993.

[20] S. Hiwa, K. Hanawa, R. Tamura, K. Hachisuka, and T. Hiroyasu, "Analyzing brain functions by subject classification of functional near-infrared spectroscopy data using convolutional neural networks analysis," *Comput. Intell. Neurosci.*, vol. 2016, 2016, Art. no. 1841945.

[21] 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, 2017, Art. no. 011008.

[22] H. Song *et al.*, "Automatic schizophrenic discrimination on fNIRS by using complex brain network analysis and SVM," *BMC Med. Inf. Decis. Making*, vol. 17, no. 3, pp. 1–9, 2017.

[23] T. Akiyama, M. Koeda, Y. Okubo, and M. Kimura, "Hypofunction of left dorsolateral prefrontal cortex in depression during verbal fluency task: A multi-channel near-infrared spectroscopy study," *J. Affect. Disord.*, vol. 231, pp. 83–90, 2018.

[24] T. Näsi, K. Kotilahti, T. Noponen, I. Nissilä, L. Lipiäinen, and P. Meriläinen, "Correlation of visual-evoked hemodynamic responses and potentials in human brain," *Exp. Brain Res.*, vol. 202, no. 3, pp. 561–570, 2010.

[25] M. Nagai, N. Endo, and T. Kumada, "Measuring brain activities related to understanding using near-infrared spectroscopy (NIRS)," in *Symp. Hum. Interface. Manage. Inf.*. Springer, 2007, pp. 884–893.

[26] Z. Li, 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," *J. Neurosci. Methods*, vol. 241, pp. 101–110, 2015.

[27] A. Amini, W. Chen, G. Fortino, Y. Li, Y. Pan, and M. D. Wang, "Editorial special issue on AI-driven informatics, sensing, imaging and big data analytics for fighting the COVID-19 pandemic," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 10, pp. 2731–2732, Oct. 2020.

[28] K. Matsuo, T. Kato, M. Fukuda, and N. Kato, "Alteration of hemoglobin oxygenation in the frontal region in elderly depressed patients as measured by near-infrared spectroscopy," *J. Neuropsychiatry Clin. Neurosci.*, vol. 12, no. 4, pp. 465–471, 2000.

[29] K. Matsuo, Y. Onodera, T. Hamamoto, K. Muraki, N. Kato, and T. Kato, "Hypofrontality and microvascular dysregulation in remitted late-onset depression assessed by functional near-infrared spectroscopy," *Neuroimage*, vol. 26, no. 1, pp. 234–242, 2005.

[30] T. Suto, M. Fukuda, M. Ito, T. Uehara, and M. Mikuni, "Multichannel near-infrared spectroscopy in depression and schizophrenia: Cognitive brain activation study," *Biol. Psychiatry*, vol. 55, no. 5, pp. 501–511, 2004.

[31] M. Herrmann, A.-C. Ehlis, and A. Fallgatter, "Bilaterally reduced frontal activation during a verbal fluency task in depressed patients as measured by near-infrared spectroscopy," *J. Neuropsychiatry Clin. Neurosci.*, vol. 16, no. 2, pp. 170–175, 2004.

[32] S. Y. Baik, J.-Y. Kim, and S.-H. Lee, "Prefrontal asymmetry during cognitive tasks in depression and its relationship to suicide ideation: A functional near-infrared spectroscopy (fNIRS) study," *IBRO Rep.*, vol. 6, 2019, Art no. S187.

[33] H. Zhu *et al.*, "Decreased functional connectivity and disrupted neural network in the prefrontal cortex of affective disorders: A resting-state fnirs study," *J. Affect. Disord.*, vol. 221, pp. 132–144, 2017.

[34] H. Song *et al.*, "Automatic depression discrimination on fNIRS by using general linear model and SVM," in *Proc. 7th Int. Conf. Biomed. Eng. Informat.*, 2014, pp. 278–282.

[35] Y. Zhu, T. Jiang, Y. Zhou, and L. Zhao, "Discriminative analysis of functional near-infrared spectroscopy signals for development of neuroimaging biomarkers of elderly depression," *J. Innov. Opt. Health Sci.*, vol. 3, no. 01, pp. 69–74, 2010.

[36] H. Song, W. Du, and Q. Zhao, "Automatic depression discrimination on fnirs by using fastica/WPD and SVM," in *Proc. Chin. Intell. Automat. Conf.*, 2015, pp. 257–265.

[37] A. Duncan *et al.*, "Optical pathlength measurements on adult head, calf and forearm and the head of the newborn infant using phase resolved optical spectroscopy," *Phys. Med. Biol.*, vol. 40, no. 2, pp. 295–304, 1995.

[38] S. Matcher, C. Elwell, C. Cooper, M. Cope, and D. Delpy, "Performance comparison of several published tissue near-infrared spectroscopy algorithms," *Anal. Biochem.*, vol. 227, no. 1, pp. 54–68, 1995.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[40] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[41] M. Wang, W. Fu, X. He, S. Hao, and X. Wu, "A survey on large-scale machine learning," *IEEE Trans. Knowl. Data Eng.*, 2020.