Youth Depression Diagnosis Algorithm Based on 3D-WGMobileNet and Transfer Learning

Yu WANG [1], Zhaohui GUO [1], Ke SUN [1], Hongbing XIAO [1], Wenmin WANG [2]

[1] School of Computer and Artificial Intelligence, Beijing Technology and Business University, 100048, Beijing, China
[2] School of Computer Science and Engineering, MaCau University of Science and Technology, 999078, Macao, China

wangyu@btbu.edu.cn, 15701573421@163.com, kesun105@126.com, x.hb@163.com, wmwang@must.edu.mo

Submitted December 20, 2024 / Accepted February 8, 2025 / Online first March 6, 2025

Abstract. *Depression is a common mental illness that not only profoundly infests the psychological state of patients, but also tends to cause damage to the functioning of patients' brain areas. To construct a comprehensive and detailed framework for a supporting diagnostic network that will help physicians make accurate and timely diagnoses when dealing with patients at different stages of depression, a network model based on three-dimensional (3D) weight group MobileNet (3D-WGMobileNet) and transfer learning is proposed. Firstly, fMRI data is preprocessed, and regional homogeneity analysis is used to reduce the dimension of the image. Then, the characteristics of Alzheimer's disease are learned by transfer learning and transferred to the proposed model. Next, the dynamic group convolution was used to construct the expert weight matrix of the convolution kernel, and the sliding window group convolution was used to compress the parameters of the model to improve the expression ability and computing power of the model. By using 5-fold cross-validation, we conducted experiments using data from HCP and REST-meta-MDD. The experiment results show that the proposed model gives a superior performance compared with other state-of-the-art methods, especially on the classification of the healthy group with major depression groups, where the two datasets achieve 88% and 91% accuracy, respectively, which verifies the feasibility and effectiveness of our model.*

Keywords

Depression, functional magnetic resonance imaging, transfer learning, MobileNet, dynamic group convolution

1. Introduction

Depression usually refers to a mood disorder, a syndrome characterized by a depressed state of mind [1], whose clinical manifestations mainly include low mood, interest, cognition, thinking, volitional activities, and physiological function disorders, etc. And some patients commit suicide, even engage in aggressive behavior. The diagnosis of early depression is based on clinical symptoms, medical history, course of illness, and physical examination, as well as laboratory tests, but quantitative physiological indicators are lacked [2]. Therefore, it is of great significance to explore the imaging neurobiological markers of depression diagnosis in depth, and to break through the bottlenecks faced by modern medical technology in the clinical diagnosis of depression, such as too much subjectivity and lacking quantitative indexes.

With the rapid development of medical imaging technologies and deep learning in recent years, some new ideas are provided for the study of brain diseases [3], [4]. Functional magnetic resonance imaging (fMRI) [5] is a popular neuroimaging technology at present, whose principle is to use magnetic resonance imaging (MRI) to reflect the changes on blood oxygen level dependent (BOLD) in the brain. fMRI images have a high temporal resolution, and can dynamically reflect the changes of signal intensity in the brain areas. Deep learning is a new field motivated for building neural networks, which mimics the human brain's mechanisms for interpreting data such as images, sounds, and text. Recently, more and more people are combining deep learning with medical imaging as a research hotspot to assist doctors for diagnosing medical diseases [6–9]. Wang et al. [10] used convolutional neural network (CNN) to classify major depressive disorder (MDD) for resolving the problems of insufficiently raw image data which easily leads to overfitting and poor generalization ability of common classification models. But CNN itself has a simple structure, and cannot simultaneously extract shallow and deep information of the image. Hamid et al. [11] proposed a deep learning method based on bi-directional long short-term memory (Bi-LSTM), which combined electroencephalograph (EEG) data and facial features to detect depression. But the EEG data is two-dimensional (2D) data, and the extent of brain damage is not detectable. Jan et al. [12] proposed a dynamic visual motion feature extraction algorithm, which could predict the degree of depression based on an individual's visual and acoustic features. Finally, partial least squares regression model was used to obtain the correlation between visual features and depression, but the method based on dynamic visual features is unstable. Therefore, an accurate diagnosis of depression is unable to make.

DOI: 10.13164/re.2025.0018

Melo et al. [13] proposed a distributed learning method based on ResNet-50 to determine whether a subject has depression by recognizing facial expressions, but its parameters and computational quantities were huge, and it was not easy to train. Ahmed et al. [14] proposed an fMRI-S4 lightweight network model based on long short-term memory and one-dimensional convolution to classify the functional connectivity maps of brain regions in MDD patients. However, one-dimensional convolution is not sufficient for medical image feature extraction, and the classification accuracy is required to improve. Daegil et al. [15] proposed a depression diagnosis algorithm based on 2-stream CNN, which combined ResNet and SeNet to extract feature information in the image, and to classify patients. However, this network has high complexity and requires a large number of samples as support.

Because transfer learning can acquire knowledge from different types of images in different domains, it is widely used to solve the problems like insufficient data for the target task. Tao et al. [16] used a ViT-Transformer coding network combined with EEG signal data to classify MDD patients, and achieved better classification results. Jazaery and Guo et al. [17] used RNN-C3D network to extract useful feature information from continuous facial expressions, and to get prediction results for depression.

In all of the above work, only healthy controls and MDD are categorized, and the course of depression, such as mild, moderate and major, failed to accurately be diagnosed. As inspired by the above ideas, in this paper a deep learning model based on the designed 3D-WGMobileNet and transfer learning is proposed to accurately classify fMRI images of depression patients. Main ideas include that four-dimensional (4D) images are converted into three-dimensional (3D) images using the regional homogeneity (ReHo) analysis method, which facilitates the effective processing of the deep learning model at later stage. In addition, transfer learning is used to solve the problem of poor generalization ability due to the lack of medical data, and in the proposed 3D-WGMobileNet, dynamic group convolution is utilized to balance weight distribution, to extract detailed features of the image, and to reduce the number of parameters. Furthermore, the training speed of the model is accelerated by sliding window group convolution. The proposed model can achieve better accuracy while improving real-time performance, and can finally realize the correct classification of depression patients.

The remainder of this paper is organized as follows. In Sec. 2, the experimental dataset and the image preprocessing method are presented, respectively. In Sec. 3, the proposed method in detail is described. Section 4 gives the experimental results and analysis. Finally, a conclusion is drawn in Sec. 5.

2. Data

The current feature extraction algorithms are difficult to directly extract the features of 4D fMRI data, so in this paper, fMRI images are subjected to ReHo analysis. By analyzing the blood oxygen content of the patient's brain region over a period of time, the local activity information of the brain functional region can be extracted, and the high-dimensional data can be converted into low-dimensional data by statistical methods. In the first part of this section, we introduce the data information and preprocessing methods. In the second part data dimension reduction method is presented.

2.1 Experimental Dataset

In this paper, fMRI images of 144 subjects were used, including 25 cases of mild depression (MID), 41 cases of moderate depression (MOD), 8 cases of MDD, and 70 cases of healthy control group. All of depression data came from the HCP database (https://humanconnectome.org/). The fMRI data of each subject are 33 layers of head images scanned from left to right. A total of 1200 time point image data are scanned, and the diagnostic and statistical manual of mental disorders (DSM) values of the patients were all above 50 points. The specific information is shown in Tab. 1. In addition, in order to verify the in scalability of our model on a larger dataset, we also selected data from the REST-meta-MDD Consortium [18], which consists of 830 MDD and 771 HC. For mitigating data bias and ensuring data integrity, subject selection was governed by the following criteria: (1) Removal of low-quality images due to inadequate coverage, poor spatial normalization, or significant head motion. (2) Exclusion of sites contributing fewer than 10 participants or a disproportionately high number of elderly participants. (3) Discard of zero signal images in target graph detection.

2.2 Data Preprocessing

During the collection of medical images, due to the influence of imaging equipments, imaging principles, and individual differences, the original images generally contain degradation phenomena such as uneven brightness and noise etc. [19]. Therefore, preprocessing of raw fMRI data, such as head movement correction, normalization, and smoothing etc., is needed to reduce errors, and to improve image quality. In this paper, the preprocessing of fMRI data is realized using FMRIB's software library (FSL) and statistical parametric mapping (SPM) software. fMRI preprocessing flowchart is shown in Fig. 1.

Dataset	Number	Age	Sex (M/F)	DSM values	Mean	Standard Deviation
HC	70	28	35/35	45.6	0.4	0.02
MID	25	28	5/20	54.8	0.35	0.05
MOD	41	29	18/23	63	0.3	0.08
MDD	8	29.8	3/5	73.3	0.2	0.12

Tab. 1. Statistical analysis of subject information.

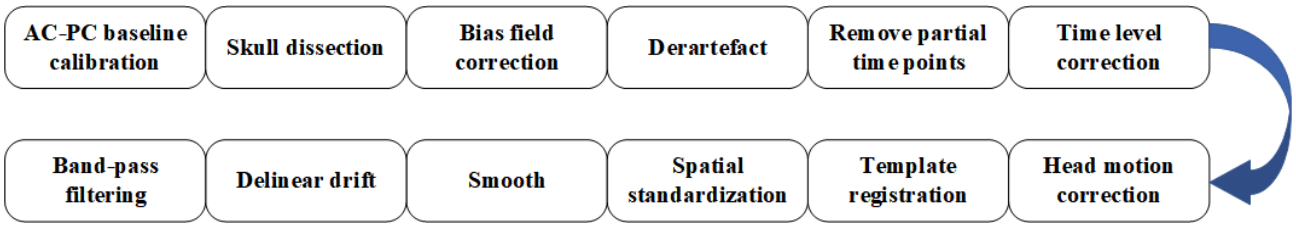

Fig. 1. Preprocessing of fMRI images.

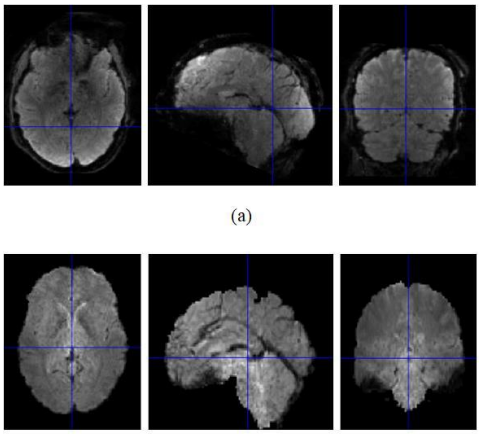

Fig. 2. Comparison of fMRI images (a) before and (b) after preprocessing.

Due to the instability of the initial fMRI signal, the first 10 time points of each fMRI data were deleted, and the rest of points were made timing correction, realigning, and normalization. The images are registered to the template proposed by the Montreal Neurological Institute (MNI). The comparison plots of fMRI images before and after preprocessing are shown in Fig. 2.

2.3 ReHo Transformation Analysis Method

ReHo analysis method is used to characterize the consistency of the BOLD signals (time series) of a voxel and its nearby voxels, which is measured by the Kendall consistency coefficient (KCC) [21]. ReHo method first proposed by Zang et al. [20] is used to calculate the area of fMRI time series in the process of blood oxygen level synchronization.

Suppose that an fMRI data is represented by $F(X, Y, Z, M)$, where X is the sagittal plane (the number of rows), Y is the coronal plane (the number of columns), Z is the transverse plane (the number of layers), M is the number of time points of the current voxel (the length of the BOLD signal), and the data contains $X_o \times Y_o \times Z_o$ voxel points. T_m represents the time series of the m-th voxel $V_m(x, y. z)$ $(1 \le x_m \le X_o,\ 1 \le y_m \le Y_o,\ 1 \le z_m \le Y_o,)$, the ReHo procedure for calculating the BOLD signal sequences of the T_m voxel and the K_m (usually is 6, 18, and 26) voxels in the nearest neighboring domain is as follows.

(1) The time series of the T_m and K_m voxels are arranged into a matrix $\mathbf{C}_{m,k+1}(i,j)$ of $m \times (k + 1)$, where $C(i, j)$ represents the i_m-th time point of the j_m-th voxel.

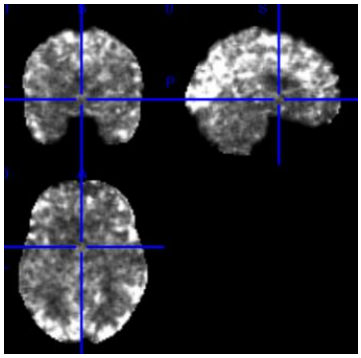

Fig. 3. An example of ReHo transformation results.

(2) The elements of the j_m-th column is filled with the ordering of the values in the columns, and the matrix $\mathbf{S}_{m,k+1}(i, j)$ of the size $m \times (k + 1)$ is obtained, where $S(i, j)$ represents the ordering of the data value F at the i_m-th time point of the j_m-th voxel among the m time points data in this column, where $1 \le i \le m$, $1 \le j \le k + 1$. The characteristic of the matrix $\mathbf{S}_{m,k+1}(i, j)$ is that the elements of each column are positive integers from 1 to m without repetition.

(3) The KCC of the BOLD signal sequence of T_m and neighboring $K_m$voxels is calculated as shown in (1).

$$W = \frac{\sum_{i=1}^{m} (S_i)^2 - m(\bar{S})^2}{\frac{1}{12}(K+1)^2(m^3 - m)} \tag{1}$$

where m denotes the number of time points, K denotes the size of the neighborhood selected in the calculation. S_i denotes the sum of row i in the matrix $\mathbf{S}_{m,k+1}(i, j)$, and $\bar{S} = (m + 1)(k + 1)/2$ denotes the mean value of S_i. W denotes the KCC value of the voxel which is also called the ReHo value, and the value of W ranges from 0 to 1. The larger the value of W, the higher regional homogeneity of this voxel $V_m(x, y. z)$ is, and vice versa.

The larger the KCC in the ReHo analysis method is, the more similar the time series of these neighboring voxels are, and the mean-averaged ReHo image is obtained by the normalization method. The transformed image using ReHo analysis method is shown in Fig. 3.

3. Methods

Specifically, first, the fMRI image is preprocessed and ReHo transformed. Then, transfer learning was used to obtain the basic feature information of the image from

other data and transfer it to the proposed 3D-WGMobileNet. Then, the improved 3D-WGMobileNet was used to extract the features of the transformed image. Finally, patients with different stages of depression were classified. The preprocessing and transformation steps have been presented in the previous section and other detailed steps are explained in the following subsections.

3.1 3D-WGMobileNet

3D-WGMobileNet is designed by changing convolution kernel into dynamic group convolution (DGConv) based on the original 2D-MobileNet [22] and by adding sliding window group convolution (SGConv) layer behind the depth-separable convolution Block, which is used to fuse the features, and finally the fusion features are outputted through the fully connected layer. The input of the model is N ReHo transformed images of size $C \times X \times Y \times Z$, and the global information of the image is extracted using the dynamic convolution [23] module and the depthwise separable convolution group in turn. The expert weight measurement matrix is constructed, and the weight information of each convolution kernel is more carefully allocated to improve the efficiency of the convolution kernel. Subsequently, the extracted feature information was spliced and fused by sliding window group convolution to refine the extracted features and to reduce information redundancy. Finally, the depression features are classified into four categories such as MID, MOD, MDD, and HC through the fully connected layer, and the classification results are obtained.

The 3D-WGMobileNet model structure consists of 16 modules, including 2 3D-DGConv modules, 11 3D-WGBlock modules, 1 3D-Avg_pool module, and 2 fully connected (FC) layers modules. The overall network structural framework is shown in Fig. 4.

Among them, the 3D-DGConv module is the 3D dynamic group convolution module, and the 3D-Avg_pool module is the 3D average pooling layer. In addition, the FC module is the fully connected layer, and the 3D-WGBlock module is the improved Block group whose specific structure is shown in Fig. 5. BN is the batch normalization, and SE denotes the squeeze excitation [24] model, RE represents ReLU6 activation function, HS is H-Switch activation function, and SGConv denotes sliding window group convolution.

Dynamic Group Convolution. The convolution kernels of traditional deep learning network models are 2D static convolution kernels. By setting a fixed convolution kernel size, the convolution is carried out according to the size of each input image, and the features of the image are extracted using pooling and activation functions. However, the weights of the static convolution kernel are shared, and it is easy to extract duplicate features. Therefore, in this paper a new dynamic group convolution is proposed to replace the original static convolution. The squeeze excitation principle is used to construct expert weight measurement matrix, and the static convolution kernel group is transformed into a dynamic convolution kernel for improving the expression ability of the model. The input feature map and output feature map of the module are $N \times C \times X \times Y \times Z$ dimensions, and the image size remains unchanged after dynamic group convolution. The specific calculation method process is as follows.

(1) An expert weight restructuring function is defined including global average pooling layer, fully connected layer, and RE activation function, where α_i is the final weight coefficient, R is the expert weight coefficient, GAP is the global average pooling, and σ is the activation function, as shown in (2).

$$\begin{aligned} &\alpha_i = r_i(x), \\ &r(x) = \sigma\left(\mathrm{GAP}(x) R\right). \end{aligned} \tag{2}$$

(2) A dynamic convolution on the expert weight reorganization is constructed whose structure consists of a global average pooling layer α_i, two fully connected layers W_{fc1} and W_{fc2}, and an activation function, where w is the weight of the convolution kernel, l is the number of experts, α_{out} is the output convolution kernel weight, and σ is the activation function, as shown in (3).

$$\begin{aligned} &\alpha_{\mathrm{out}} = \sigma\left(W_{\mathrm{fc1}} \times W_{\mathrm{fc2}} \times \frac{1}{ijk} \sum_{i,j,k} X_{c,i,j,k}\right), \\ &W_{\mathrm{fc1}} = \alpha_1 w_1 + \alpha_2 w_2 + \ldots \alpha_l w_l. \end{aligned} \tag{3}$$

(3) A dynamic group convolution is constructed, connecting a grouped fully connected layer behind the dynamic convolution, and the convolution kernel weights for convolution are grouped, where α_{Gout} is a $C \times C \times K_{\mathrm{x}} \times K_{\mathrm{y}} \times K_{\mathrm{z}}$ dimensional vector, and W_{fc3} is the grouped fully connected layer. W_{fc3} is divided into $G \times C$ groups at operation time, where G is the number of groups, which is used to adjust the model to achieve the optimal effect, as shown in (4).

$$\alpha_{\mathrm{Gout}} = W_{fc3} \times \alpha_{\mathrm{out}} \tag{4}$$

DGConv saves memory space occupation in the kernel space by recalculating the weights of each convolution kernel in each layer. Furthermore, the model is easier to train, and has a strong generalization ability. The overall structural module of DGConv is shown in Fig. 6.

Sliding window Group Convolution. 2D-MobileNet contains depth separable convolution module, which can reduce the computation and number of parameters of the network. In this paper, the improved 3D-WGMobileNet is proposed. Although it is easier to extract the features of 3D images, but due to the increase of the network dimensions, the number of parameters of the overall network structure is very large. And due to the sharing of the output weight, slow training is generated. Therefore, a sliding window group convolution layer behind the depth separable convolution is added for fusing features. The SGConv uses a convolution kernel with a step size of s to slide on the input to extract features, to sparsify the connection between the input and output, and to reduce the number of parameters

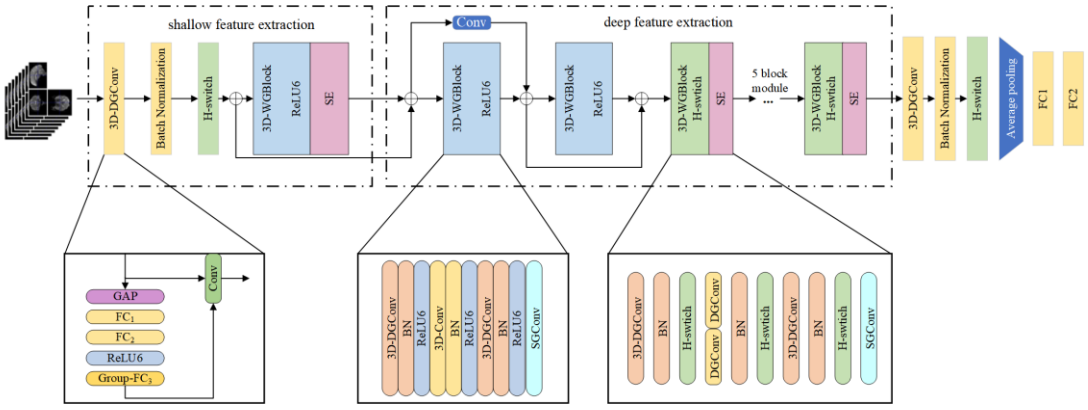

Fig. 4. The network structure of the designed 3D-WGMobileNet.

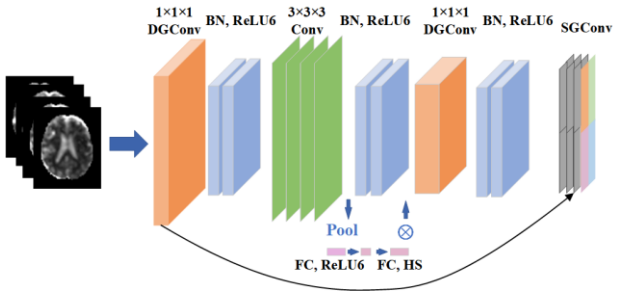

Fig. 5. The structure of the 3D-WGBlock module.

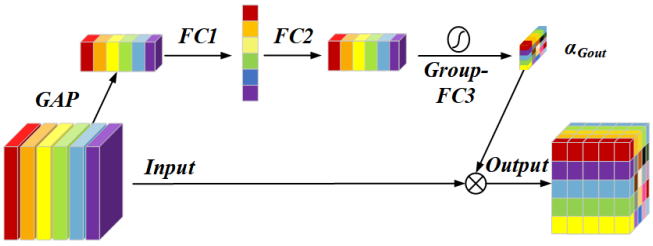

Fig. 6. The structure of DGConv.

while ensuring the extraction of a certain degree of information from neighboring channels. The principle structure is shown in Fig. 7.

Assuming that the size of the convolution kernel for SGConv is d_k, the input image is of size $X \times Y \times Z$ voxels and dimension M, and the number of channels of the output is h. Dividing the input into g groups and using the convolution kernel with step size s to convolve the input image will produce d_c windows with parameter number $d_c \times d_k \times d_k \times d_k$, and computation amount is $d_c \times d_k \times d_k \times d_k \times x \times y \times z \times h$. Compared with ordinary convolution, the number of parameters of this method is reduced by $m \times h$ orders of magnitude, and the amount of calculation is reduced by m orders of magnitude. The calculation method of d_c is consistent with the one of ordinary convolution, where P is the number of fillers and s is the speed of the sliding window, and the calculation in X dimension is shown in (5):

$$d_c = \frac{X - d_k + 2P}{s} + 1. \tag{5}$$

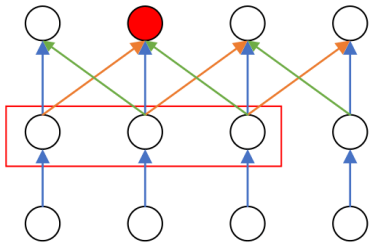

Fig. 7. Schematic diagram of SGConv.

On this basis, for facilitating the model training, we add skip connections to the depthwise separable convolutional Blocks to generate the group features with the global information.

3.2 Transfer Learning

With the wild application of artificial intelligence and deep learning in image processing, supervised learning and unsupervised learning have developed rapidly, such as RNN and DC-Gan [25] networks, etc. However, these methods require extremely high quality and a large number of labeled datasets for training. Transfer learning gives an effective solution to the problem of model overfitting due to the lack of data and other reasons. Tajbakhsh et al. [26] concluded in their published paper that in the methods using deep learning for diagnosis in medical images, it is better to obtain initial image information using transfer learning, then fine-tuning the network compared with zero initialization to train the network. Medical imaging is difficult to obtain a lot of labeled image data due to problems such as experimental equipment, rare specialists etc. Therefore, in the above cases, direct use of deep learning for feature extraction and classification often does not perform well. Verified by predecessors, transfer learning can be used to pre-train the model to solve the problem of insufficient data for depression, and then improve the generalization ability of 3D-WGMobileNet network.

In this paper, the large public Alzheimer's disease neuroimaging initiative (ADNI) database is used as the pre-training data, and the model is pre-trained by transfer learning method. The steps of transfer learning pre-training

method based on ADNI are as follows. Firstly, fMRI raw images of 670 AD subjects without other diseases were screened out from the ADNI database, containing three categories of AD patients, mild cognitive impairment, and healthy controls. And the same preprocessing steps as those for depression subjects are carried out. Secondly, the 3D-WGMobileNet model was trained with the preprocessed ADNI data, and the training weight files of the model were saved. Then, the training weights of the backbone part of the model were transferred to the assisted model 3D-WGMobileNet. Next, the depression data was used to fine-tuned the model. Finally, the output features from the model were classified.

4. Experiment and Result Analysis

4.1 Experimental Environment

In this paper, all network models use cross-entropy loss function and Adam optimization algorithm, the data is divided into training, validation and test sets according to the ratio of 8:1:1, and five-fold cross validation is used, and the epoch value is set to 80. The batch is set to 64 when training the 2D network model and 8 when training the 3D network model. The initial learning rate is set to 0.01, and the initial learning rate of the network model after transfer learning is 0.0001. For experimental fairness, the parameters of each model are adjusted, and the best results are selected for comparison.

4.2 Model Evaluation Index

For better evaluating the performance of the proposed model, in this paper, Accurary (ACC), F1 score, and the area under curve (AUC) covered by receiver operating characteristic (ROC) curve are used as the evaluation indexes of the model [26]. The calculation is shown in (6), (7) and (8):

$$\mathrm{ACC}=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}, \tag{6}$$

$$\begin{cases} \mathrm{Pre}=\dfrac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \\ \mathrm{Sen}=\dfrac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \\ \mathrm{F1}=\dfrac{2\times\mathrm{Pre}\times\mathrm{Sen}}{(\mathrm{Pre}+\mathrm{Sen})}, \end{cases} \tag{7}$$

$$\mathrm{AUC}=\frac{\sum_{ins_i\in positiveclass}\mathrm{rank}_{ins_i}-\dfrac{U\times(U+1)}{2}}{U\times V} \tag{8}$$

where TP and TN respectively represent the number of subjects in the true positive and negative category. FP and FN respectively represent the number of subjects in the false positive and negative category. The ACC is obtained by calculating the ratio of predicted positive samples to all samples, which reflects the quality of neural network classification results. The F1 score is a comprehensive evaluation index, which combines precision (Pre) and sensitivity (Sen). rank_{ins_i} represents the number of the *i*-th sample, and *U* and *V* respectively represent the number of positive and negative samples. $\sum_{ins_i\in positiveclass}$ denotes the summation of only the serial number of positive samples, and the larger the AUC value indicates, the better classifier performance is. The above three indexes range from 0 to 1. The larger the result is, the better the performance of the method shows.

4.3 Results and Analysis

In this paper, three groups of controlling experiments are designed to verify the feasibility and effectiveness of the improved 3D-WGMobileNet for the diagnosis of depression.

Controlling experiment 1: The classical 2D and 3D lightweight deep learning networks were used to extract and classify the depression images after ReHo transformation, and compared with the proposed 3D-MobileNet the ability of 3D network to extract and classify medical images is explored. The experimental results are shown in Tab. 2.

The experimental results show that in the above lightweight model network, with the increase of network model depth, the classification results of the model are improved. MobileNetV1 adds a depthwise separable convolution architecture, which reduces the memory occupation and computational complexity of the convolutional layers. While MobileNetV3 [28] uses neural architecture search (NAS) relative to V1 version, and adds the SE module to the depthwise separable convolution to obtain more feature information, which changes the calculation method of the model and enhances the performance of the model.

However, compared with the 2D convolutional network, the 3D network can read the inter-layer information of the image, and has good adaptability to the fMRI data of medical images. In the comparison experiment between 3D and 2D-MobileNetV3, 3D-MobileNetV3 can directly encode the features of three dimensions when extracting fMRI image features, which enhances the feature extraction ability of spatial information. The recognition accuracy of depression patients on HCP dataset reaches 78.02%. The importance of inter-layer information is further demonstrated.

Controlling experiment 2: In the Block module of the MobileNetV3 network, the number of static convolutions replaced by DGConv is compared. The first group experiment is to replace the original static convolutions with two or three DGConv modules only in the Block module with ReLU6 activation function. The second group one is based on the first group experiment, which adds again 2 or 3 DGConv respectively replacing the original static convolutions in the Block (ReLU6 and HS activation function)

module, means that in one MobileNetV3 network, Block (ReLU6) and Block (HS) are simultaneously contained, and 2 DGConv are used respectively. While in another MobileNetV3 network, Block (ReLU6) and Block (HS) are also simultaneously contained, and 3 DGConv are used respectively. The number of DGConv modules is used to verify the feature extraction ability of the DGConv idea in the deep feature extraction part of the model, and the advantages of achieving the best results are discussed. The experimental results are shown in Tab. 3.

The experimental results from Tab. 3 show that replacing static convolution with DGConv helps to improve the classification accuracy of the model, but adding too many DGConv modules does not achieve better results, and can lead to an increase on the computational complexity of the model. Adding DGConv to the Block module using ReLU6 and HS functions achieves a 2.68% improvement over that of not applying DGConv, which indicates that using the weight adjustable function on the convolution kernel is helpful to improve the ability of convolution kernel, and to learn deep features.

Controlling experiment 3: For verifying the effectiveness of the transfer learning modules, DGConv and SGConv added to the improved 3D-WGMobileNet, an ablation experiment on the network is conducted, and the results are shown in Tab. 4.

Methods	Criteria	HC vs MID	MID vs MOD	MOD vs MDD	HC vs MDD
2D-Vgg16	ACC	62.22%	61.11%	63.00%	65.60%
	F1	65.85%	51.79%	63.52%	77.14%
	AUC	61.25%	53.88%	62.10%	52.80%
2D-Resnet50	ACC	62.78%	60.00%	62.50%	63.20%
	F1	63.13%	40.08%	63.61%	71.21%
	AUC	46.63%	56.25%	64.50%	58.00%
2D-MobileNetV1	ACC	67.56%	62.17%	60.50%	70.04%
	F1	69.76%	61.17%	50.93%	72.56%
	AUC	69.25%	63.88%	54.20%	74.59%
3D-MobileNetV1	ACC	73.11%	65.56%	59.00%	73.61%
	F1	72.59%	66.30%	47.29%	74.50%
	AUC	74.88%	68.00%	48.50%	77.92%
2D-MobileNetV3	ACC	70.17%	66.71%	62.34%	72.82%
	F1	72.37%	65.21%	59.53%	75.63%
	AUC	72.48%	76.20%	55.88%	76.70%
3D-MobileNetV3	ACC	73.33%	71.89%	70.12%	78.02%
	F1	74.35%	70.50%	71.91%	77.78%
	AUC	75.25%	77.75%	70.01%	78.33%

Tab. 2. Experimental results of 2D deep learning network.

Methods	DGConv	Precision	Recall
3D-MobileNetV3	-	81.50%	74.32%
+Block (ReLU6)	2	83.85%	75.88%
	3	83.82%	75.67%
+Block (ReLU6, HS)	4	84.18%	78.78%
	6	84.01%	78.34%

Tab. 3. Results on the addition effect of the idea of DGConv in 3D-MobileNet.

The experimental results in Tab. 4 show that the use of transfer learning method can obtain more basic features of the original data, supplement the data volume, reduce the training time of the model, and improve the generalization ability of the model. Therefore, the accuracy of using transfer learning is 4.38% higher than that of directly using 3D-MobileNet for the classification of depression and HC. Because the state grouped convolution module can dynamically adjust the weight of the convolution kernel using the self-attention mechanism and the input features, and reduce the repeated calculation by grouping, improve the feature extraction ability of the convolution kernel, and decrease the model parameters. The classification of HC and MID is improved by 1.21% by replacing the DGConv module compared with only adding transfer learning. Because the SGConv uses the feature fusion of adjacent channels for convolution, the reuse of global features is avoided, the amount of network parameters is reduced, and the computational efficiency of the model is improved. Therefore, compared with only using transfer learning, the classification accuracy of using SGConv in HC and depression increases by 3.11%, and compared with adding transfer learning and DGConv, the classification accuracy of HC and depression increases by 2.49%. In particular, the experimental results of HC vs MID using ROC are visualized, as shown in Fig. 8. The horizontal axis represents False Positive Rate (FPR) and the vertical axis represents True Positive Rate (TPR). As can be seen from this figure, the curve is significantly closer to the upper left corner, indicating that the model performs well in distinguishing between the healthy and depressed groups. The diagonal line represents the baseline for random classification, and the further the curve moves away from this baseline, the better the model's performance is.

The accuracy of the method proposed in this paper reaches 88.00%, 87.9% on the recognition accuracy of HC vs. depression, and MID vs. MOD, respectively. It is effectively proved that adding transfer learning, subdividing the weight of the convolution module, and changing the weight sharing of the fully connected layer help to capture more small feature information, and to improve the accuracy of the model. And the main architecture of the model uses a lightweight network, which is more conducive to running on the Central Processing Unit (CPU). In addition, to validate the in scalability of our model on larger datasets, we

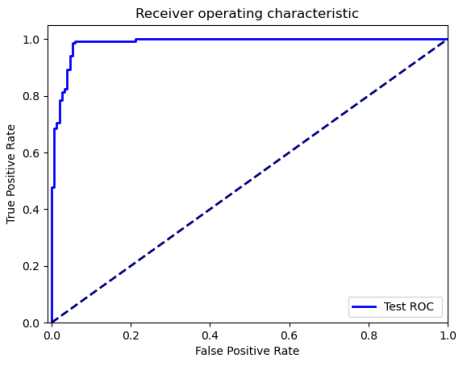

Fig. 8. Schematic diagram of ROC.

also selected data from the REST-meta-MDD Consortium. The experimental results are shown in Tab. 5. On this larger dataset, it can be found that the performance of the model shows an increase compared to the performance of the smaller dataset, proving that the model can still run effectively when the data size increases. This indicates that our model has good scalability for larger datasets and has strong potential for practical applications.

Controlling experiment 4: In this paper, the proposed method is compared with the latest methods such as 3-D CNN proposed by Zhao et al. [29], P-mRMR proposed by Bhaskar et al. [30], Constrained PARAFAC redictive sub-network proposed by Bhaskar et al. [31], GAE-FCNN proposed by Fuad et al. [32]. The relevant results are shown in Tab. 6. It can be seen from Tab. 6 that the holistic results obtained by the proposed method in the small data set are better, and got the specificity of 92%, compared with the other methods. Because in this paper transfer learning is used as the pre-training basis of the model, and the ideas of dynamic grouped convolution and sliding window grouped convolution effectively avoid the overfitting problem caused by using small sample data to train model. Therefore, the specificity of image classification is improved, and the best experimental results are obtained.

Methods	Criteria	HC vs MID	MID vs MOD	MOD vs MAD	HC vs MDD
3D-MobileNet+Transfer learning	ACC	82.83%	81.80%	80.55%	82.40%
	F1	83.75%	81.94%	78.62%	85.31%
	AUC	80.42%	79.93%	84.00%	78.47%
3D-MobileNet+ Transfer learning + DGConv	ACC	84.04%	81.20%	80.56%	83.60%
	F1	81.56%	82.86%	81.76%	85.71%
	AUC	80.59%	79.20%	86.25%	79.67%
3D-MobileNet+ Transfer learning+ SGConv	ACC	84.73%	82.60%	81.11%	85.51%
	F1	80.72%	81.69%	82.59%	85.25%
	AUC	84.67%	78.53%	81.88%	79.07%
The proposed method	ACC	85.15%	87.90%	82.57%	88.00%
	F1	82.37%	85.81%	81.79%	87.78%
	AUC	84.59%	82.26%	83.28%	84.07%

Tab. 4. Experimental results of different improved ideas of 3D-WGMobileNet.

Methods	Criteria	HC vs MID	MID vs MOD	MOD vs MAD	HC vs MDD
The proposed Method	ACC	87.35%	89.60%	90.38%	91.00%
	F1	86.24%	88.46%	89.21%	90.34%
	AUC	85.74%	84.38%	84.26%	85.22%

Tab. 5. Experimental results in REST-meta-MDD.

Methods	Depression/HC	ACC	Spe	Sen
Zhao et al. [29]	40/37	90%	10%	79%
Bhaskar et al. [30]	49/33	78%	55%	90%
Bhaskar et al. [31]	49/33	82%	79%	84%
Fuadet al. [32]	250/227	65.07%	60.00%	69.74%
The proposed method	74/70	88%	92%	83.88%

Tab. 6. Comparison results of different methods.

5. Discussion

In addition to the technological advances demonstrated in this study, a number of challenges exist which must be addressed before our proposed data processing approach can be widely applied in clinical settings. In a clinical setting, computational cost can be a limiting factor, especially if large amounts of real-time data need to be processed. High-performance computing equipment may be required to ensure timely analysis, but not all healthcare organizations can be equipped with such resources. In addition, the sheer size of fMRI images requires significant data storage and processing power. To mitigate these awkward problems, in the future we will further optimize the computational efficiency of the model. In terms of clinical applications, although the proposed data processing method performs well in experimental settings, there are still some limitations. For example, in a clinical setting, it may be more difficult for patients to remain still, resulting in poorer data quality. Individual differences may also affect the effectiveness of the method. In addition, real-time data processing and immediate feedback are critical in clinical practice. Therefore, for successfully applying the method to the clinic, we need to do the following work in the future. Firstly, the real-time processing requirements is considered, and we will continue to optimize the algorithm in the future to achieve fast computation while ensuring diagnostic accuracy. Secondly, large-scale clinical trials are needed in the future to fully validate the accuracy, reliability and stability of the model in order to ensure the validity of the model in a clinical setting. This will help to evaluate the performance of the method in different patient populations, especially in different disease stages and different age groups. Finally, any clinical decision support system based on deep learning requires the involvement of physicians, with which we combine clinical experience to make rational decisions. Finally, considering the complexity of depression data, we will further explore multimodal fusion to integrate information from different data sources in order to extract more representative and discriminative features, and to improve classification accuracy and model robustness.

6. Conclusion

In this paper a depression assisted diagnosis algorithm is proposed based on a lightweight deep learning network in order to fully extract the local and global feature information of MDD images, and to improve the network's classification performance on 3D medical data. Firstly, the fMRI image data are preprocessed, and ReHo analysis is used to reduce the dimension of the original images.

Secondly, the fMRI image data are transferred to the proposed model 3D-WGMobileNet as the pre-training data using the transfer learning method. In addition, a DGConv module is designed to improve the computing power of the

convolution kernel by dynamically calculating the weight of the convolution kernel, and the weight matrix of each channel is divided in groups to compress the computation of the convolution kernel, which effectively enhances the feature extraction ability of the convolution kernel. Furthermore, SGConv and skip connection are added to the network to extract the local and global information of the features, and to avoid the redundancy of information in the kernel space and the feature space. Finally, the depression data is inputted into the pre-trained 3D-WGMobileNet to fine-tune the model, and the classification results are obtained.

The results of the classification experiments show that the classification accuracy of depression and HC, MID and HC, MID and MOD, MOD and MDD reaches 88.00%, 85.15%, 87.90% and 82.57%, respectively, which verified that using the proposed method in this paper, patients with different stages of depression can be effectively classified, and some theoretical basis for the adjuvant treatment of depression is provided.

Acknowledgments

This research is supported by Joint Project of Beijing Natural Science Foundation and Beijing Municipal Education Commission (Grant No. KZ202110011015).

References

[1] BAI, R., XIAO, L., GUO, Y., et al. Tracking and monitoring mood stability of patients with major depressive disorder by machine learning models using passive digital data (preprint). *JMIR mHealth and uHealth*, 2021, vol. 9, no. 3, p. e24365. DOI: 10.2196/24365

[2] LIU, H.-J., QIU, J. Depression brain dysfunction: evidence from a local functional connectivity method (in Chinese). *Journal of Psychological Science*, 2016, vol. 39, no. 1, p. 239–244. DOI: 10.16719/j.cnki.1671-6981.20160135

[3] CHEN, Z., ZHAO, R., WANG, Q., et al. Functional connectivity changes of the visual cortex in the cervical spondylotic myelopathy patients: A resting-state fMRI study. *Spine*, 2020, vol. 45, no. 5, p. E272–E279. DOI:10.1097/BRS.0000000000003245

[4] JU, R., HU, C., LI, Q., et al. Early diagnosis of Alzheimer's disease based on resting-state brain networks and deep learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2017, vol. 16, no. 1, p. 244–257. DOI: 10.1109/TCBB.2017.2776910

[5] JUN, E., NA, K.-S., KANG, W., et al. Identifying resting-state effective connectivity abnormalities in drug-naïve major depressive disorder diagnosis via graph convolutional networks. *Human Brain Mapping*, 2020, vol. 41, no. 17, p. 4997–5014. DOI: 10.1002/hbm.25175

[6] HUANG, H., HUANG, Y., KAGGIE, J. D., et al. Multiparametric MRI-based deep learning radiomics model for assessing 5-year recurrence risk in non-muscle invasive bladder cancer. *Journal of Magnetic Resonance Imaging*, 2025, vol. 61, no. 3, 1442–1456. DOI: 10.1002/jmri.29574

[7] HOSSEINI-ASL, E., GHAZAL, M., MAHMOUD, A., et al. Alzheimer's disease diagnostics by a deeply supervised adaptable 3D convolutional network. *Frontiers in Bioscience-Landmark*, 2018, vol. 23, no. 3, p. 584–596. DOI: 10.2741/4606

[8] LAO, H., ZHANG, X., TANG, Y., et al. Alzheimer's disease diagnosis based on the visual attention model and equal-distance ring shape context features. *IET Image Processing*, 2021, vol. 15, no. 10, p. 2351–2362. DOI: 10.1049/IPR2.12218

[9] CHEN, E., WU, X., WANG, C., et al. Application of improved convolutional neural network in lung image segmentation. In *2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*. Taiyuan (China), 2019, p. 1–6. DOI: 10.1109/MLBDBI48998.2019.00027

[10] WANG, Y., ZHENG, G., JIE, R., et al. A major depression auxiliary diagnosis method based on convolution neural network (in Chinese). *Journal of Lanzhou University (Medical Edition)*, 2022, vol. 48, no. 08, p. 5–10. DOI: 10.13885/j.issn.1000-2812.2022.08.002

[11] HAMID, D. S. B. A., GOYAL, S. B., BEDI, P. Integration of deep learning for improved diagnosis of depression using EEG and facial features. *Materials Today: Proceedings*, 2023, vol. 80, p. 1965–1969. DOI: 10.1016/J.MATPR.2021.05.659

[12] JAN, A., MENG, H., GAUS, Y. F. A., et al. Artificial intelligent system for automatic depression level analysis through visual and vocal expressions. *IEEE Transactions on Cognitive and Developmental Systems*, 2017, vol. 10, no. 3, p. 668–680. DOI: 10.1109/TCDS.2017.2721552

[13] DE MELO, W. C., GRANGER, E., HADID, A. Depression detection based on deep distribution learning. In *2019 IEEE International Conference on Image Processing (ICIP)*. Taipei (Taiwan), 2019, p. 4544–4548. DOI: 10.1109/ICIP.2019.8803467

[14] EL-GAZZAR, A., THOMAS, R. M., VAN WINGEN, G. fMRI-S4: Learning short-and long-range dynamic fMRI dependencies using 1D convolutions and state space models. In *International Workshop on Machine Learning in Clinical Neuroimaging*, 2022, p. 158–168. DOI: 10.48550/arXiv.2208.04166

[15] CHOI, D., ZHANG, G., JUNG, J., et al. Depression diagnosis algorithm based on 2-stream CNN using facial image. In *2023 IEEE/ACIS 23rd International Conference on Computer and Information Science (ICIS)*. Wuxi (China), 2023, p. 43–47. DOI: 10.1109/ICIS57766.2023.10210234

[16] FU, Y., ALDRICH, C. Flotation froth image analysis by use of a dynamic feature extraction algorithm. *IFAC-PapersOnLine*, 2016, vol. 49, no. 20, p. 84–89. DOI: 10.1016/J.IFACOL.2016.10.101

[17] AL JAZAERY, M., GUO, G. Video-based depression level analysis by encoding deep spatiotemporal features. *IEEE Transactions on Affective Computing*, 2018, vol. 12, no. 1, p. 262 to 268. DOI: 10.1109/TAFFC.2018.2870884

[18] CHEN, X., LU, B., LI, H.-X., et al. The DIRECT consortium and the REST-meta-MDD project: Towards neuroimaging biomarkers of major depressive disorder. *Psychoradiology*, 2022, vol. 2, no. 1, p. 32–42. DOI: 10.1093/psyrad/kkac005

[19] SARRAF, S., TOFIGHI, G. Classification of Alzheimer's disease using fMRI data and deep learning convolutional neural networks. *arXiv*, 2016, p. 1–5. DOI: 10.48550/arXiv.1603.08631

[20] ZANG, Y., JIANG, T., LU, Y., et al. Regional homogeneity approach to fMRI data analysis. *NeuroImage*, 2004, vol. 22, no. 1, p. 394–400. DOI: 10.1016/j.neuroimage.2003.12.030

[21] HARADA, A., BOLLEGALA, D., CHANDRASIRI, N. P. Discrimination of human-written and human and machine written sentences using text consistency. In *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*. Greater Noida (India), 2021, p. 41–47. DOI: 10.1109/ICCCIS51004.2021.9397237

[22] ZHANG, C., YANG, T., YANG, J. Image recognition of wind turbine blade defects using attention-based MobileNetv1-YOLOv4

and transfer learning. *Sensors*, 2022, vol. 22, no. 16, p. 1–18. DOI: 10.3390/s22166009

[23] LUO, Y., ZHU, K., WANG, W., et al. A speaker recognition method based on dynamic convolution with dual attention mechanism. *Engineering Letters*, 2023, vol. 31, no. 2, p. 1–8.

[24] ROY, S. K., DUBEY, S. R., CHATTERJEE, S., et al. FuSENet: Fused squeeze-and-excitation network for spectral-spatial hyperspectral image classification. *IET Image Processing*, 2020, vol. 14, no. 8, p. 1653–1661. DOI: 10.1049/iet-ipr.2019.1462

[25] RADFORD, A., METZ, L., CHINTALA, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv*, 2015, p. 1–16. DOI: 10.48550/arXiv.1511.06434

[26] TAJBAKHSH, N., SHIN, J. Y., GURUDU, S. R., et al. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging*, 2016, vol. 35, no. 5, p. 1299–1312. DOI: 10.1109/TMI.2016.2535302

[27] JIA, H., WANG, Y., DUAN, Y., et al. Alzheimer's disease classification based on image transformation and features fusion. *Computational and Mathematical Methods in Medicine*, 2021, no. 1, p. 1–11. DOI: 10.1155/2021/9624269

[28] CHANG, J., GUAN, S. Q., SHI, H. Y. Strip defect classification based on improved generative adversarial networks and MobileNetV3 (in Chinese). *Laser Optoelectronics Progress*, 2021, vol. 58, no. 4, p. 1–6. DOI: 10.3788/LOP202158.0410016

[29] ZHAO, Y., DONG, Q., ZHANG, S., et al. Automatic recognition of fMRI-derived functional networks using 3-D convolutional neural networks. *IEEE Transactions on Biomedical Engineering*, 2017, vol. 65, no. 9, p. 1975–1984. DOI: 10.1109/TBME.2017.2715281

[30] BHASKAR SEN, MUELLER, B., KLIMES-DOUGAN, B., et al. Classification of major depressive disorder from resting-state fMRI. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Berlin (Germany), 2019, p. 3511–3514. DOI: 10.1109/EMBC.2019.8856453

[31] BHASKAR SEN, CULLEN, K. R., PARHI, K. K. Classification of adolescent major depressive disorder via static and dynamic connectivity. *IEEE Journal of Biomedical and Health Informatics*, 2021, vol. 25, no. 7, p. 2604–2614. DOI: 10.1109/JBHI.2020.3043427

[32] NOMAN, F., TING, M. C., KANG, H., et al. Graph autoencoders for embedding learning in brain networks and major depressive disorder identification. *IEEE Journal of Biomedical and Health Informatics*, 2024, vol. 28, no. 3, p. 1644-1655. DOI: 10.1109/JBHI.2024.3351177

About the Authors...

Yu WANG (corresponding author) received her Ph.D. degree from the University of Science and Technology Beijing in 2009. She was engaged in scientific research as a post-doctoral in the Beijing Key Laboratory of Multidimensional and Multiscale Computing Photography, Tsinghua University from 2009 to 2011. She is now a Professor and doctoral supervisor of the Beijing Technology and Business University. Her research interests include pattern recognition, medical image processing and computer vision.

Zhaohui GUO received her master's degree from the School of Artificial Intelligence, Beijing Technology and Business University, Beijing, China, in 2024. Her research interests include pattern recognition, depression detection.

Ke SUN is currently a PhD student at the School of Artificial Intelligence, Beijing Technology and Business University, where her research interests include depression detection, and medical image segmentation.

Hongbing XIAO received his Ph.D. degree from Beijing Institute of Technology, Beijing, China, in 2008. He is currently an Associate Professor and graduate supervisor in Beijing Technology and Business University. His research interests include artificial intelligence, food safety testing, and transient dynamic testing.

Wenmin WANG received Ph.D. degree from Harbin Institute of Technology in 1989. Thereafter, he worked as an Associate Professor until 1991. He then gained overseas industrial experience for 18 years. Invited to return to China, he served from 2009 as a Professor at the School of Electronic and Computer Engineering, Peking University. Since 2019, he has been a Professor at the School of Computer Science and Engineering, Macau University of Science and Technology. His research interests include computer vision and multimedia processing.