

Major depression disorder diagnosis and analysis based on structural magnetic resonance imaging and deep learning

Yu Wang^{1,*}, Ning Gong¹, Changyang Fu¹

¹Beijing Key Laboratory of Big Data Technology for Food Safety, School of Artificial Intelligence, Beijing Technology and Business University, 100048 Beijing, China

*Correspondence: wangyu@btbu.edu.cn (Yu Wang)

DOI: [10.31083/j.jin2004098](https://doi.org/10.31083/j.jin2004098)

This is an open access article under the CC BY 4.0 license (<https://creativecommons.org/licenses/by/4.0/>).

Submitted: 27 October 2021 Revised: 2 December 2021 Accepted: 9 December 2021 Published: 30 December 2021

Major depression disorder is one of the diseases with the highest rate of disability and morbidity and is associated with numerous structural and functional differences in neural systems. However, it is difficult to analyze digital medical imaging data without computational intervention. A voxel-wise densely connected convolutional neural network, Three-dimensional Densenet (3D-DenseNet), is proposed to mine the feature differences. In addition, a novel transfer learning method, called Alzheimer's Disease Neuroimaging Initiative Transfer (ADNI-Transfer), is designed and combined with the proposed 3D-DenseNet. The experimental results on a database that contains 174 subjects, including 99 patients with major depression disorder and 75 healthy controls, show that large changes in brain structures between major depressive disorder patients and healthy controls mainly are located in the regions including superior frontal gyrus, dorsolateral, middle temporal gyrus, middle frontal gyrus, postcentral gyrus, inferior temporal gyrus. In addition, the proposed deep learning network can better extract different features of brain structures between major depressive disorder patients and healthy controls and achieve excellent classification results of major depressive disorder. At the same time, the designed transfer learning method can further improve classification performance. These results verify that our proposed method is feasible and valid for diagnosing and analyzing major depression disorder.

Keywords

Major depression disorder; Machine learning algorithm; Structural magnetic resonance imaging; 3D-DenseNet; ADNI-transfer; Computational neuroscience

1. Introduction

Major depression disorder (MDD) is one of the most common mental disorders whose causes and pathological mechanism are the most complicated and is seriously harmful to society today. Therefore, accurate and rapid diagnosis of MDD is extremely important for patients. However, in affective disorders, the intrinsic complexity of brain neuroanatomy and its functional connectivity is further complicated by the considerable heterogeneity of these conditions and the effects of treatment on the brain, which makes making and analysis of MDD particularly challenging [1]. Neuroimaging, like structural magnetic resonance imaging (sMRI), is a popular medical imaging method nowadays which has many ad-

vantages such as non-invasiveness and high contrast and is widely used in the diagnosis and research of depression [2–4]. So far, researchers have found that the brain differences in functions and structures between MDD patients and healthy people exist [4]. In particular, the connectivity between brain areas such as the hippocampus, frontal lobe, cerebellum, and other parts is changed.

Machine learning methods are used to diagnose mental illness [5]. However, with the arrival of deep learning [6] in the field of image processing [7–9], the application of deep learning methods in the medical images field [10, 11] led to the convolutional neural networks (CNN) is a common deep learning algorithm, in which the backpropagation algorithm is used to adjust its internal parameters and stack multiple layers of neurons to find deeper features on large data sets. Previous research [12] evidenced that the depth of the network had a crucial impact on the final performance of networks. It is supposed that the deeper the network is, the better its generalization ability tends to be. With this basic criterion, CNN [7] has developed from 7 layers to 16 layers, even 19 layers of Visual Geometry Group (VGG) [13]. With the increase in layers, the computing power and time cost required for network training also increase. However, the results are not always improved only by simply increasing the depth of the network. When the number of layers of the network reaches a certain amount, the network will converge more slowly, and classification accuracy will gradually saturate. And if the network continues to go deeper, the accuracy will even decrease. This phenomenon is known as the degradation problem. He *et al.* [14] proposed deep residual networks (ResNet) to solve this problem. Using skip connections and after-addition activation, ResNet allows signals to be directly propagated from one block to other blocks, which is beneficial to the backpropagation of gradients during training. Thus, the depth of ResNet can be above 152 or more, which solves the problems of gradient disappearance and network degradation to a certain extent. Subsequently, Huang *et al.* [15] proposed a densely connected network (DenseNet) whose basic idea is the same as ResNet's. Still, it establishes

connections between layers in one block to achieve feature reuse. In this case, the number of parameters and the calculation cost of DenseNet are less than those of ResNet. And DenseNet shows better performance on many public large data sets [16].

However, most current deep learning networks can only process two-dimensional (2D) natural image data and rarely deal with three-dimensional (3D) data. Especially in 3D sMRI data of depression, deep learning-related research has not appeared at present [17]. Nevertheless, several studies have shown that using 3D networks to process 3D data can get better results than using 2D networks. For instance, Chen *et al.* [18] extended the 2DResNet into a 3D variant to automatically segment brain structures from 3D MR images. This 3D method achieved much better performance compared to the 2D CNNs method. Hosseini-Asl *et al.* [19] proposed a 3D-CNN classifier, which can predict Alzheimer's disease on sMRI data more accurately than several other state-of-the-art 2D networks. Thus, using 3D networks to classify depressions MRI data has great potential.

Moreover, training a deep learning network usually requires a huge amount of annotated data which is hard to achieve in medical imaging, where data is often expensive and protected. CNN's are trained using a backpropagation algorithm in which the unknown weights of each layer are continuously updated during iterations to minimize specific loss functions. Normally, those weights are initialized with random values before training. However, the increase of network layers will increase network parameters, which requires more training data to make the backpropagation algorithm converge better. A limited amount of data is easy to cause the problem of overfitting, which makes the algorithm get stuck at a local minimum value. Then suboptimal classification performance will happen. To solve this problem, a feasible way is transfer learning in which the initial values of network weights are not random but copied from a network that has been trained and fine-tuned on a larger data set.

Tajbakhsh *et al.* [20] discussed and compared the results of training from scratch and transfer learning in the field of medical imaging. It shows that transfer learning and fine-tuning are better than training networks from scratch in most cases. So far, transfer learning has been applied to medical image classification or segmentation of diseases such as Alzheimer's disease [21], brain tumors [22], and pulmonary nodules [23] and has shown excellent results. Chen *et al.* [24] collected 8 different datasets of 3D medical image segmentation tasks, including liver, heart, etc., 8 datasets shared one encoder during the training process, and 8 decoders were used, respectively. Finally, only the common encoder part is transferred for the next segmentation and classification tasks.

A novel method based on deep learning, called 3D-DenseNet, is proposed for classifying and predicting MDD in terms of sMRI. A transfer learning method, called ADNI-Transfer, is designed and combined with the proposed 3D-DenseNet to improve the classification results. Our main

contributions are as follows: (1) a three-dimensional (3D) densely connected convolutional network is proposed, which borrows the spirit of a two-dimensional (2D) densely connected convolutional network, and extends the 2D network into a 3D form. The proposed deep learning network can fully mine the spatial information in the 3D sMRI data. Finally, accurate classification of patients with MDD and healthy controls (HC) is obtained; (2) a novel transfer learning workflow is designed. The networks are initialized with pre-trained weights from a similar larger dataset and are fine-tuned to solve the problem of overfitting caused by insufficient data; (3) comparative experiments with multiple groups of advanced 2D and 3D networks have been done to prove the superiority and effectiveness of the proposed method for the classification task of MDD based on magnetic resonance imaging.

2. Data

2.1 Database

There are 174 subjects, including 99 patients with MDD and 75 age-, sex-, and education-matched healthy controls (HC). Patients are recruited from Beijing Anding Hospital Affiliated with Capital Medical University, and the HC group is recruited through newspaper advertisements. All the patients in MDD met the DSM-IV diagnostic criteria of depression, and all the HC were interviewed using the non-patient edition of DSM-IV. Before the experiment, all of the subjects signed informed consent. The clinical characteristics of MDD and HC are shown in Table 1. *p*-value stands for the two-sample *t*-test of MDD and HC, HAMD denotes the Hamilton depression rating scale, and HAMA expresses for the Hamilton anxiety rating scale. The data we used and the data Zheng *et al.* [25] used were collected from the same group of subjects.

Table 1. Demographic and clinical characteristics of subjects.

Variables	MDD	HC	<i>p</i> -value
Gender (M:F)	43:56	33:42	0.941
Age (years)	34.57 ± 12.18	35.65 ± 12.63	0.57
Education level (years)	13.75 ± 3.01	12.93 ± 2.40	0.61
Age range	18–65	19–60	-
Duration of illness (years)	7.88 ± 7.87	-	-
Number of depressive episodes	2.63 ± 1.26	-	-
HAMD	21.44 ± 3.97	-	-
HAMA	16.00 ± 9.61	-	-

HAMD, hamilton depression; HAMA, hamilton anxiety.

Using SPM software (version 12, University of London, London, UK), two-sample *t*-tests (*p* = 0.05) were performed on brain sMRI of 99 MDD patients and 75 HC normal. The significance level of tissue voxel values difference can be observed. Therefore, MDD's lesion areas can be obtained, and the disease reasons can be explored. To display the lesion areas more intuitively, all images were activated. The le-

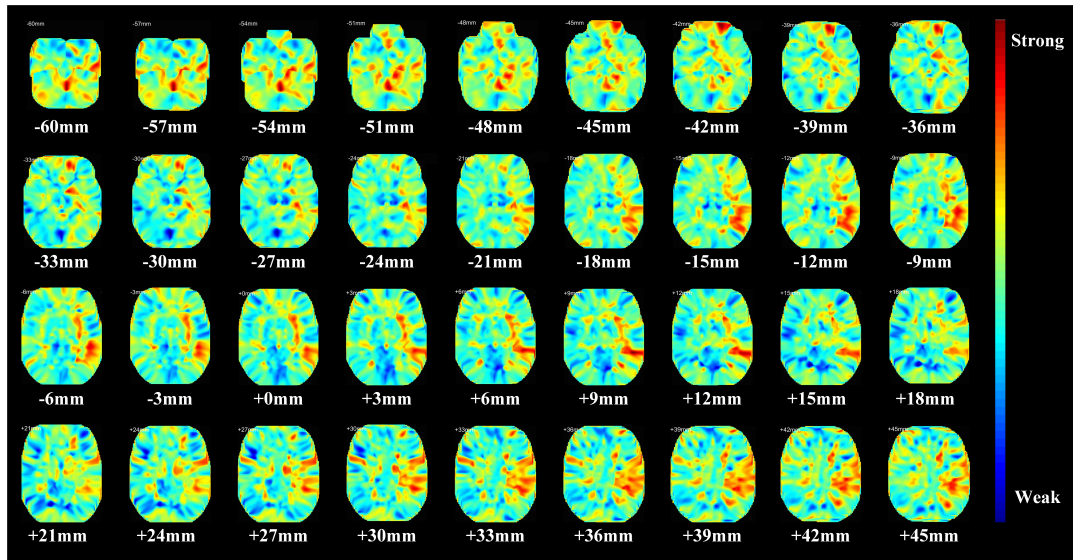


Fig. 1. Activation slices in the whole brain. By analysis of FDR (false discovery rate), it can be found that the common brain lesion areas of the MDD patients where the changes of brain structure are large include the Superior frontal gyrus, dorsolateral (SFGdor), Middle temporal gyrus (MTG), Middle frontal gyrus (MFG), Postcentral gyrus (PoCG), inferior temporal gyrus (ITG), Precuneus (PCUN), Precentral gyrus (PreCG), Middle occipital gyrus (MOG), Temporal pole: superior temporal gyrus (TPOsup) and Superior frontal gyrus, medial (SFGmed) according to the extent of pathological injury of brain structures.

sion areas were stratified, as shown in Fig. 1. The red parts meant that large changes in brain structures happened between MDD patients and HC.

2.2 sMRI data acquisition

All the sMRI scans were acquired using a MAGNETOM Trio, A TimSystem3.0-Tesla scanner (Siemens, Erlangen, Germany) in the National Key Laboratory for Cognitive Neuroscience and Learning, Beijing Normal University, using magnetization prepared rapid gradient echo (MPRAGE). The scanning parameters are as follows: repetition time (TR) = 2530 ms, echo time (TE) = 3.39 ms, flip angle (FA) = 7°, field of view (FOV) = 256 mm × 256 mm, voxel size = 1 mm × 1 mm × 1.33 mm, slice thickness = 1.33 mm, slices number = 128.

2.3 Data preprocessing

The data preprocessing is realized using SPM12¹ toolkit based on MATLAB R2013b. Considering the important influence of the gray matter area on the diagnosis of MDD [26], only the gray matter (GM) part is used for the next experiments.

The specific preprocessing steps are shown in Fig. 2. The size of each subject's sMRI data after preprocessing is 121 × 145 × 121 pixels.

3. Methods

3.1 3D-DenseNet for 3D image

Although 2D DenseNet has achieved remarkable results on many 2D natural image datasets, it has few achievements in medical image analysis. The reason is that the convolution

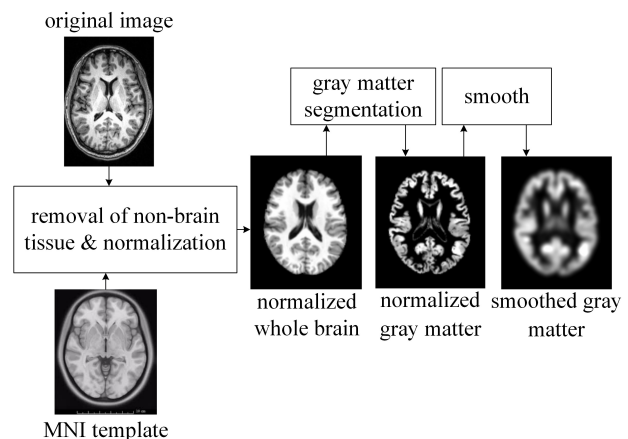


Fig. 2. Data preprocessing flowchart. The interference of non-brain tissue was removed by overall cleaning. GM segmentation is registered by non-linear warping to Montreal neurological institute (MNI) template generated using diffeomorphic anatomical registration through exponentiated lie algebra (DARTEL) which can obtain a high-dimensional normalization including 60 mm full bias width at half maximum (FWHM) cut-off, warping regularization of 4, spatial-adaptive non-local means (SANLM) denoising filter, and Markov random field (MRF) weighting of 0.15. The voxel size of sMRI data after normalization is 1.5 × 1.5 × 1.5. A Gaussian kernel smoothed all the data with the size 8 × 8 × 8 mm and 3 FWHM.

kernel and pooling kernel in 2D networks like DenseNet are two-dimensional matrices, which can only move in two directions of image height and width of the 2D images. Thus, only two-dimensional features can be extracted. However, most medical image data such as sMRI are 3D data, which can only be input into 2D networks hierarchically, or one of the dimensions must be regarded as channel dimension. And

¹ Available: <https://www.fil.ion.ucl.ac.uk/spm/software/spm12>.

neither of the two methods can make good use of the spatial information between slices of the sMRI data. We added a depth dimension to filters such as convolution kernel and pooling kernel, which extended these kernels to the 3D matrix. In this way, the filters can move in all three directions of sMRI data, and the spatial information of data is fully mined. The output of each filter is also 3D data. If the size of one of the 3D convolution kernels is $k \times k \times k \times \text{channel}$, the number is n , and the input data size is $h \times w \times d$. Because the sMRI data used in this paper is similar to the grayscale image and the channel dimension is 1, the output size of the convolution kernel is described by

$$(h - k + 1) \times (w - k + 1) \times (d - k + 1) \times n \quad (1)$$

Similar methods can extend the pooling layer and batch normalization layer in DenseNet. The 3D-DenseNet is constructed, which can better extract representative features from 3D sMRI data and improve the classification accuracy of MDD-HC MRI data. A 121-layer 3D-DenseNet structure is shown in Fig. 3.

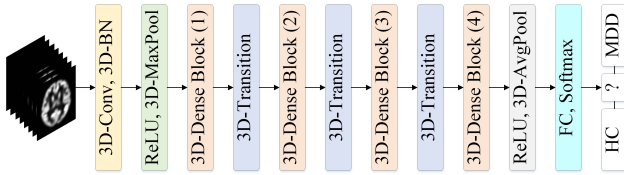


Fig. 3. The structure of 3D-DenseNet121. 3D-DenseBlock (1) contains 6 layers. 3D-DenseBlock (2) contains 12 layers. 3D-DenseBlock (3) contains 24 layers, and 3D-DenseBlock (4) contains 16 layers.

Each of these layers includes a $1 \times 1 \times 1$ convolutional layer, a $3 \times 3 \times 3$ convolutional layer, two batch normalization (BN) [27] layers, and two rectified linear unit (ReLU) [28] layers. The structure of a 6-layer 3D-DenseBlock is shown in Fig. 4.

The dense connectivity of each of these layers can be expressed as follows.

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]), \quad (2)$$

where x_l refers to the feature map received by layer l , and $[x_0, x_1, \dots, x_{l-1}]$ denotes the concatenation of the feature maps produced in layers. $0, \dots, l-1$ $H_l(\cdot)$ is defined as a composite function of three consecutive operations including a BN, a ReLU, and a $3 \times 3 \times 3$ convolution (3D-Conv). If each $H_l(\cdot)$ produces k feature maps, the total number of input feature maps of the l -th layer is $k_0 + k \times (l-1)$, where k_0 represents the number of channels in the input layer.

The dense connection operation in Eqn. 2 is not feasible when the size of the feature maps is inconsistent, so a 3D-Transition module is added between each 3D dense block. Each 3D-Transition module contains a BN layer, a ReLU

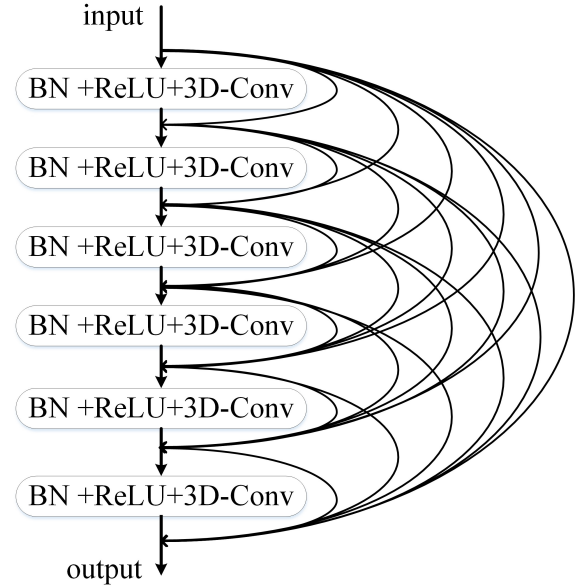


Fig. 4. The structure of a 6-layer 3D-Dense block in which each arrow junction represents dense connectivity. For each layer, the feature maps of all previous layers are used as input of this layer, and the feature map of this layer is used as input of all subsequent layers.

layer, a $1 \times 1 \times 1$ convolutional layer, and an average pooling layer (AvgPool) for reducing the dimension of the feature maps. The last 3D-Dense Block is connected with a ReLU layer, an AvgPool layer, a fully connected layer (FC), and a Softmax layer for implementing the final feature reduction and classification. The specific parameters and architecture of a 121-layer 3D-DenseNet are shown in Table 2, in which each conv represents a BN-ReLU-Conv sequence, and (*) denotes different part of 3D-DenseNet as shown in Fig. 3,5.

3.2 Transfer learning for 3D data

In the medical field, the amount of data is often limited, leading to bad results. The motivation of using transfer learning is to train a model with a relatively large 3D medical dataset, which can be used as the backbone pre-trained model to boost the target task with insufficient training data. In this way, we can mine more knowledge and information of the small sample data by using the related other data and transfer learning. Inspired by Chen *et al.* [24], we designed a novel transfer learning framework for 3D sMRI data. When it comes to the data selection, only the same part (brain) and the same type (sMRI) are collected for pre-training, and only classification tasks are considered. Because it is particularly challenging to get MDD and HC data from hospitals or labs due to privacy and there's no open-source MDD-HC dataset on the internet, we chose to use the Alzheimer's disease dataset (ADNI, <https://ida.loni.usc.edu>) as the pre-training dataset. A four-step processing workflow is designed to achieve our transfer learning model, as shown in Fig. 5.

The reason why we only select data from brain sMRI datasets for classification is that if the similarity between the

Table 2. The parameters and architecture of 3D-DenseNet121.

Layers	Output size	Parameters
Input layer	$1 \times 121 \times 145 \times 121$	-
3D-Conv	$64 \times 121 \times 73 \times 61$	kernel size: (7, 7, 7), stride: (1, 2, 2)
3D-BN	$64 \times 121 \times 73 \times 61$	eps: 1e-5, momentum: 0.1
ReLU	$64 \times 121 \times 73 \times 61$	-
3D-MaxPool	$64 \times 61 \times 37 \times 31$	kernel size: (3, 3, 3), stride: (1, 1, 1)
Dense Block (1)	$256 \times 61 \times 37 \times 31$	$\begin{bmatrix} 1 \times 1 \times 1 \text{ conv} \\ 3 \times 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
3D-Transition	$128 \times 61 \times 37 \times 31$	$1 \times 1 \times 1 \text{ conv}$
	$128 \times 30 \times 18 \times 15$	$2 \times 2 \times 2 \text{ average pool, stride: 2}$
Dense Block (2)	$512 \times 30 \times 18 \times 15$	$\begin{bmatrix} 1 \times 1 \times 1 \text{ conv} \\ 3 \times 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
3D-Transition	$256 \times 30 \times 18 \times 15$	$1 \times 1 \times 1 \text{ conv}$
	$256 \times 15 \times 9 \times 7$	$2 \times 2 \times 2 \text{ average pool, stride: 2}$
Dense Block (3)	$1024 \times 15 \times 9 \times 7$	$\begin{bmatrix} 1 \times 1 \times 1 \text{ conv} \\ 3 \times 3 \times 3 \text{ conv} \end{bmatrix} \times 24$
3D-Transition	$512 \times 15 \times 9 \times 7$	$1 \times 1 \times 1 \text{ conv}$
	$512 \times 7 \times 4 \times 3$	$2 \times 2 \times 2 \text{ average pool, stride: 2}$
Dense Block (4)	$1024 \times 7 \times 4 \times 3$	$\begin{bmatrix} 1 \times 1 \times 1 \text{ conv} \\ 3 \times 3 \times 3 \text{ conv} \end{bmatrix} \times 16$
ReLU	$1024 \times 7 \times 4 \times 3$	-
3D-AvgPool	$1024 \times 1 \times 1 \times 1$	kernel size: (7, 4, 3), stride: (1, 1, 1)
Fully Connected & Softmax Layer	2	-

selected source domain and the target domain is too small, it is likely to cause negative transfer, which will lead to worse performance, i.e., not increase but decrease of classification accuracy rate. On the contrary, the more similar the two data sets are, the more similar the high-level features of the two datasets will be, which will result in better representative features and a more suitable pre-training model for the target domain to improve classification performance. A 3D-ResNet is also trained with the same process and the same data in the third step for doing a contrast experiment. We use a small learning rate to fine-tune the backbone and a relatively large learning rate to train the classification layer. The transferred network extracts new features from our MDD data and boosts the classification performance.

4. Results

4.1 Evaluation metrics

The classification in this paper is a binary classification problem, that is, samples are divided into two categories, including MDD patients and HC. We specify that MDD pa-

tients are considered as positive and HC as negative. So the classification algorithm has the right or wrong predictions for the test data set, including the prediction of positive classes as positive ones (true positive, TP), the prediction of positive classes as negative ones (false negative, FN), the prediction of negative classes as positive ones (false positive, FP), and the prediction of negative classes as negative ones (true negative, TN). We select accuracy and recall as metrics to evaluate the model's classification performance. The accuracy rate is defined as $\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FN} + \text{FP} + \text{TN})$, which reflects the ability of the classifier to judge all samples. The recall rate is defined as $\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$, reflecting the proportion of MDD patients correctly judged in the total number of patients. The AUC is defined as the value of the area under the receiver operating characteristic (ROC) curve.

4.2 Training configuration

All the networks are trained using the Adam optimization algorithm [29] with a weight decay of 0.001 and cross-entropy loss function. All the data is divided into training-validation-test sets according to the 80%-10%-10% ratio, and

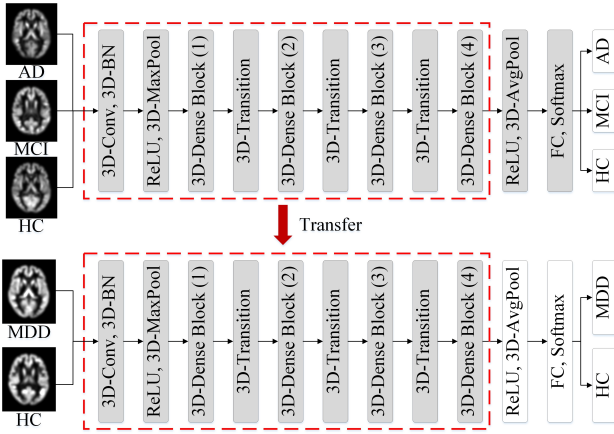


Fig. 5. The processing workflow of the proposed transfer learning model. The proposed transfer learning method includes the following four steps. Firstly, appropriate sMRI data from the ADNI database is selected, including Alzheimer's disease (AD), mild cognitive impairment (MCI), and healthy control (HC), a total of 656 subjects. Secondly, these data are pre-processed using the same preprocessing steps as in section 2.3. Then a 3D-DenseNet is trained with the preprocessed ADNI dataset to let the network learn the features of the sMRI data. Finally, the trained network's backbone (the red box part) is transferred to the classification task of MDD sMRI data, and a classification layer (includes a ReLU, a 3D-AvgPool, an FC, and a Softmax) is added.

5-fold cross-validation is used for 100 epochs. The data are randomly selected according to the proportion of MDD:HC in the original data set. Due to the limited memory capacity of GPU, the batch size is set to 64 when training 2D networks and to 8 when training 3D networks. When transfer learning is not used, the learning rate is set to 0.01 initially. When transfer learning is used, the initial learning rate of the non-transferred part remains the original parameter 0.01, and the learning rate for the transferred part is 0.001 times that of the original one. The learning rate will be lowered 10 times when the loss value of the validation set does not decrease for 10 consecutive epochs. All training is performed on a server with an NVIDIA TITAN Xp GPU.

4.3 Comparison experiments of 2D networks with different depths

During the 2D network experiments, traditional 2D DenseNet [15] is compared with 2D AlexNet [7], 2D VGG [13], and 2D ResNet [14]. The preprocessed sMRI data were hierarchically inputted into the network with an input size of 121×145 , and a voting algorithm was used. That is, for each subject, if more than half of its layers' test results are positive, it will be determined as a positive class. Otherwise, it will be judged as a negative class. The experimental results are shown in Table 3.

It is observed that with the increase of network layers, the classification accuracy and recall rate of the networks increase gradually, which shows that the deepening of the network can provide better non-linear expression ability, can enable the network to learn more complex knowledge, and can fit more complex input feature. Also, DenseNet performs bet-

Table 3. Experiment results of 2D networks.

Method	Accuracy (%)	Recall (%)	AUC
2D AlexNet	58.45	65.68	0.63
2D VGG19	60.32	67.39	0.64
2D ResNet34	63.33	68.26	0.66
2D ResNet50	63.88	69.34	0.66
2D ResNet101	65.59	72.23	0.69
2D ResNet152	66.06	72.09	0.69
2D ResNet200	67.94	74.82	0.70
2D DenseNet121	67.38	74.65	0.71
2D DenseNet169	68.20	74.91	0.71
2D DenseNet201	68.84	75.35	0.72
2D DenseNet264	69.96	76.32	0.73

ter than other convolutional networks such as ResNet when the number of layers is approximately the same, which shows that Densenet's dense connection idea is better than ResNet's residual learning idea in this task. Therefore, the subsequent experiments are mainly based on DenseNet.

4.3 Comparison experiments of 2D networks and 3D networks

For proving the superiority of the 3D network, our proposed 3D DenseNet is compared with 2D DensNet [15], 2D ResNet [14], 3D ResNet [18] with different layers, and the channel dimension method. We also compare our method with some traditional machine learning methods like local binary pattern (LBP) combined with support vector machine SVM (LBP + SVM) method in which the neighbor is 8 and radius is 1, and radial basis kernel function is selected. The experimental results are shown in Table 4.

Table 4. Comparison of experiment results between 2D and 3D networks.

Method	Accuracy (%)	Recall (%)	AUC
LBP + SVM	65.67	68.67	0.70
Channel dimension	62.50	65.88	0.68
2D ResNet101	65.59	72.23	0.69
3D ResNet101	73.26	78.46	0.75
2D ResNet152	66.06	72.09	0.69
3D ResNet152	73.47	79.33	0.75
2D ResNet200	67.94	74.82	0.70
3D ResNet200	74.81	80.66	0.76
2D DenseNet121	67.38	74.65	0.71
3D DenseNet121	74.26	80.20	0.76
2D DenseNet169	68.20	74.91	0.71
3D DenseNet169	75.38	81.26	0.77
2D DenseNet201	68.84	75.35	0.72
3D DenseNet201	76.53	82.59	0.79
2D DenseNet264	69.96	76.32	0.73
3D DenseNet264	77.42	83.72	0.80

From the data in Table 4, it can be seen that the classification accuracy, the recall rate and the AUC have been significantly improved after the network is expanded to 3D (e.g., the classification accuracy of 3D-DenseNet264 is 77.42% which is

higher than that of DenseNet264 69.96%). And 3D-DenseNet with a similar number of layers performs better than 3D-ResNet (e.g., the classification accuracy of 3D-DenseNet201 is 76.53%, and that of 3D-ResNet200 is 74.81%). This result indicates that the hierarchical information of MDD-MRI data is very rich, and the 3D network can mine this information effectively and provide more useful features than the 2D network. Therefore, the classification performance is improved. In addition, experimental results on Accuracy, Recall and AUC show that our proposed deep learning method can mine more rich, robust and complete features on data. It is superior to the channel dimension and traditional machine learning methods such as LBP + SVM.

4.4 Comparison experiments of transfer learning

For proving the positive role of transfer learning, we pre-trained the best performance on our 3D DenseNet264 model using the ADNI database and performed transfer learning (denoted by ADNI-Transfer) compared with training from scratch with MDD data, i.e., no transfer learning involved (denoted by None). Because the transfer learning method used by Chen *et al.* [24] (denoted by Med3D-Transfer) has only been performed on 3D ResNet series networks, and only the pre-trained model is opened. At the same time, training data cannot be provided. Therefore, to prove the superiority of our ADNI-Transfer method, we also performed the ADNI-Transfer on 3D ResNet200 [18] (denoted by None), denoted by ADNI-Transfer, and compared it with the 3D ResNet200 network with Med3D-Transfer (denoted by Med3D-Transfer). Please see the results of Table 5.

Table 5. Comparison of experimental results of transfer learning.

Method	Pretrain	Accuracy (%)	Recall (%)	AUC
3D ResNet200	None	74.81	80.66	0.76
	Med3D-Transfer	78.62	84.37	0.81
	ADNI-Transfer	81.45	86.52	0.84
3D DenseNet264	None	77.42	83.72	0.80
	ADNI-Transfer	84.37	87.26	0.86

It can be seen from the data in Table 5 that the classification performance of the networks has been improved significantly after transfer learning is used (e.g., after 3D-DenseNet264 has undergone ADNI-Transfer, the classification accuracy has been increased by 6.95%). It proves that transfer learning can introduce knowledge from other fields into the classification task of MDD and HC sMRI data. To some extent, it can solve the problem of insufficient samples. At the same time, the efficiency of model training is speeded up, and the final generalization ability of the model is improved. Compared with the Med3D-Transfer method, our proposed ADNI-Transfer method has better performance, which indicates that the information extracted from source domain data with the same position and the same type as

the target domain data is more valuable for the target task. Therefore, our method can improve the classification accuracy and recall rate of MDD and HC sMRI data.

The above results indicate that the classification performances are not good using 2D networks because these 2D methods ignore the information between sMRI layers. After the networks are extended to 3D, the classification accuracies are improved from 6.87% to 7.69%. And our proposed 3D-DenseNet achieved a very competitive accuracy of 77.42%. Compared to training from scratch, our proposed transfer learning method ADNI-Transfer improves the accuracy by 9.95%, which is also 2.83% higher than the existing Med3D-Transfer method. Consequently, we believe that transfer learning is of great significance in medical image classification due to the general lack of data. And it seems that the more similar the pre-training data to the target domain data is, the higher the improvement of classification performance is.

5. Conclusions

MDD is a highly prevalent psychiatric disorder that can cause a persistent feeling of sadness and loss of interest and seriously affect life quality. To early diagnose and treat MDD, a 3D deep learning network, called 3D-DenseNet, is proposed and first applied to the classification task for MDD and HC based on sMRI data in this paper. Our method extends the 2D densely connected network to a 3D version for fully mining the feature differences of brain structure between MDD patients and HC. Furthermore, a transfer learning workflow, called ADNI-Transfer, is designed to solve the problem of insufficient data. Experimental results show that the common brain lesion areas of the MDD patients where brain structure changes are large include SFGdor, MTG, MFG, PoCG, ITG etc. And our network performs better than many other advanced ones. In addition, our proposed transfer learning method can also further improve the generalization ability of the proposed network and achieve superior results. The classification accuracy and recall for MDD patients and HC can reach 84.37% and 87.26%, respectively, which verifies our method has feasibility and validity.

Author contributions

YW and CF designed the research study. NG performed the research. YW analyzed the data. NG and CF wrote the manuscript. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Acknowledgment

This article is extended from our abstract included in the International Conference on NLP & Big Data (NLPD 2020). Large amounts of experiments and comparative analysis are added. At the same time, the algorithm is further discussed.

Funding

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

Conflict of interest

The authors declare no conflict of interest.

References

- [1] Wise T, Cleare AJ, Herane A, Young AH, Arnone D. Diagnostic and therapeutic utility of neuroimaging in depression: an overview. *Neuropsychiatric Disease and Treatment*. 2014; 10: 1509–1522.
- [2] Rubin-Falcone H, Zanderigo F, Thapa-Chhetry B, Lan M, Miller JM, Sublette ME, *et al.* Pattern recognition of magnetic resonance imaging-based gray matter volume measurements classifies bipolar disorder and major depressive disorder. *Journal of Affective Disorders*. 2017; 227: 498–505.
- [3] Hilbert K, Lueken U, Muehlhan M, Beesdo-Baum K. Separating generalized anxiety disorder from major depression using clinical, hormonal, and structural MRI data: a multimodal machine learning study. *Brain and Behavior*. 2017; 7: e00633.
- [4] Sankar A, Zhang T, Gaonkar B, Doshi J, Erus G, Costafreda SG, *et al.* Diagnostic potential of structural neuroimaging for depression from a multi-ethnic community sample. *BJPsych Open*. 2016; 2: 247–254.
- [5] Stoyanov D, Kandilarova S, Aryutova K, Paunova R, Todeva-Radneva A, Latypova A, *et al.* Multivariate Analysis of Structural and Functional Neuroimaging Can Inform Psychiatric Differential Diagnosis. *Diagnostics*. 2020; 11: 19.
- [6] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015; 521: 436–444.
- [7] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In *Advances In Neural Information Processing Systems*. 2012; 25: 1097–1105.
- [8] Dash SR, Cacha LA, Poznanski RR, Parida S, Parida Shantipriya. Epileptic seizure detection: a comparative study between deep and traditional machine learning techniques. *Journal of Integrative Neuroscience*. 2020; 19: 1–9.
- [9] Donahue J, Hendricks LA, Rohrbach M, Venugopalan S, Guadarrama S, Saenko K, *et al.* Long-Term Recurrent Convolutional Networks for Visual Recognition and Description. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI / PAMI)*. 2017; 39: 677–691.
- [10] Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, *et al.* Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell*. 2018; 172: 1122–1131.
- [11] Yu W, Na Z, Fengxia Y, Yanping G. Magnetic resonance imaging study of gray matter in schizophrenia based on XGBoost. *Journal of Integrative Neuroscience*. 2018; 17: 331–336.
- [12] Szegedy C, Wei Liu, Yangqing Jia, Sermanet P, Reed S, Anguelov D, *et al.* 'Going deeper with convolutions', In *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE: Boston, 2015. IEEE: Boston, America. 2015.
- [13] Simonyan K, Zisserman A. 'Very Deep Convolutional Networks for Large-Scale Image Recognition', *ICLR*. San Diego, 2015. Computer Science: California, America. 2014.
- [14] He K, Zhang X, Ren S, Sun J. 'Deep Residual Learning for Image Recognition,' *IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, 2016. IEEE: Las Vegas, Nevada, America. 2016.
- [15] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. 'Densely Connected Convolutional Networks'. *IEEE Conference on Computer Vision and Pattern Recognition*. Hawaii, 2017. IEEE: Hawaii America. 2017.
- [16] Feng Y, Qiu D, Cao H, Zhang J, Xin Z, Liu J. Research on coronavirus disease 2019 (COVID-19) detection method based on depthwise separable DenseNet in chest X-ray images. *Journal of Biomedical Engineering*. 2020; 37: 557–565. (In Chinese)
- [17] Gao S, Calhoun VD, Sui J. Machine learning in major depression: from classification to treatment outcome prediction. *CNS Neuroscience and Therapeutics*. 2018; 24: 1037–1052.
- [18] Chen H, Dou Q, Yu L, Qin J, Heng P. VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage*. 2018; 170: 446–455.
- [19] Hosseini-Asl E, Keynton R, El-Baz A. 'Alzheimer's disease diagnostics by adaptation of 3D convolutional network', 2016 *IEEE International Conference on Image Processing (ICIP)*. Phoenix, 2016. IEEE International Conference on Image Processing: Phoenix, USA. 2016.
- [20] Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, *et al.* Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? *IEEE Transactions on Medical Imaging*. 2016; 35: 1299–1312.
- [21] Hon M, Khan NM. 'Towards Alzheimer's disease classification through transfer learning', 2017 *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Kansas City, 2017. IEEE: Kansas City, America. 2017.
- [22] Liu R, Hall LO, Goldgof DB, Zhou M, Gatenby, RA. 'Exploring deep features from brain tumor magnetic resonance images via transfer learning', 2016 *International Joint Conference on Neural Networks (IJCNN)*. Vancouver, 2016. IEEE: Vancouver, Canada. 2016.
- [23] Da Nóbrega RVM, Peixoto SA, da Silva SPP, Rebouças Filho PP. 'Lung Nodule Classification via Deep Transfer Learning in CT Lung Images', 2018 *IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*. Los Alamitos, 2018. IEEE Computer Society: Los Alamitos, USA. 2018.
- [24] Chen S, Ma K, Zheng Y. Med3D: Transfer Learning for 3D Medical Image Analysis. *arXiv preprint*. (in press)
- [25] Zheng H, Xu L, Xie F, Guo X, Zhang J, Yao L, *et al.* The Altered Triple Networks Interaction in Depression under Resting State Based on Graph Theory. *BioMed Research International*. 2015; 2015: 1–8.
- [26] Arnone D, McKie S, Elliott R, Juhasz G, Thomas EJ, Downey D, *et al.* State-dependent changes in hippocampal grey matter in depression. *Molecular Psychiatry*. 2013; 18: 1265–1272.
- [27] Sergey I, Christian S. 'Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift', *International Conference on Machine Learning*. Lille, 2015. JMLR.org: Lille, France. 2015.
- [28] Glorot X, Bordes A, Bengio Y. 'Deep Sparse Rectifier Neural Networks', *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. Journal of Machine Learning Research. 2011; 15: 315–323.
- [29] Kingma D, Ba J. 'Adam: A Method for Stochastic Optimization. Computer Science', *ICLR*. San Diego, 2015. Computer Science: California, America. 2014.