Research article

A 3D multiscale view convolutional neural network with attention for mental disease diagnosis on MRI images

Zijian Wang[1], Yaqin Zhu[1], Haibo Shi[2,*], Yanting Zhang[1] and Cairong Yan[1]

[1] School of Computer Science and Technology, Donghua University, Shanghai 200000, China
[2] School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200000, China

* **Correspondence:** Email: shihaibo@sufe.edu.cn; Tel: +8602135073183.

Abstract: Computer Assisted Diagnosis (CAD) based on brain Magnetic Resonance Imaging (MRI) is a popular research field for the computer science and medical engineering. Traditional machine learning and deep learning methods were employed in the classification of brain MRI images in the previous studies. However, the current algorithms rarely take into consideration the influence of multi-scale brain connectivity disorders on some mental diseases. To improve this defect, a deep learning structure was proposed based on MRI images, which was designed to consider the brain's connections at different sizes and the attention of connections. In this work, a Multiscale View (MV) module was proposed, which was designed to detect multi-scale brain network disorders. On the basis of the MV module, the path attention module was also proposed to simulate the attention selection of the parallel paths in the MV module. Based on the two modules, we proposed a 3D Multiscale View Convolutional Neural Network with Attention (3D MVA-CNN) for classification of MRI images for mental disease. The proposed method outperformed the previous 3D CNN structures in the structural MRI data of ADHD-200 and the functional MRI data of schizophrenia. Finally, we also proposed a preliminary framework for clinical application using 3D CNN, and discussed its limitations on data accessing and reliability. This work promoted the assisted diagnosis of mental diseases based on deep learning and provided a novel 3D CNN method based on MRI data.

Keywords: convolutional neural netwoks; attention; MRI; deep learning; ADHD; schizophrenia

1. Introduction

In the past few decades, Computer Aided Diagnosis (CAD) for mental disease has become an evolving research field of computer science and medical engineering. However, MRI-based machine learning diagnostic methods are still being explored, especially for mental disorders, such as attention deficit hyperactivity disorder (ADHD) and schizophrenia. ADHD is a neurodevelopmental disorder characterized by attention deficits, impulsiveness, and executive dysfunction. Schizophrenia (SZ) is a serious chronic mental illness. Now, the diagnosis of these diseases is often based on interviews, history and clinical symptoms. However, early and accurate diagnosis of them can facilitate treatment planning and improve disease outcomes. Improvement of machine learning based on medical markers, such as neuroimages, can facilitate the diagnosis of individuals [1].

Magnetic Resonance Imaging (MRI) is a kind of neuroimaging method that accurately measures hemodynamic changes caused by neural activities in the brain, and generates three-dimensional neural images [2]. The researches of cognitive neuroscience based on MRI mainly focus on using statistical analysis to locate brain function and analyze brain networks [3]. CAD generally classifies neuroimages through supervised machine learning methods to determine whether subjects have certain neurological diseases. However, due to the high dimension and low sample size of MRI images, the performance of some traditional machine learning algorithms in MRI image data classification are poor [4]. Some studies improved the performance of MRI data classification by improving the machine learning algorithm [5,6]. These methods can be generally divided into two kinds: traditional machine learning methods and deep learning methods.

Traditional machine learning methods, such as support vector machine (SVM) [7] and back propagation neural network (BPNN) [8], can be used to diagnose mental disease by classifying designed features in MRI images. Since a large number of mental diseases have been found to be closely related to brain network disorders, some studies have proposed the automatic diagnosis methods of machine learning based on brain network. For example, Khazaee et al. proposed a graph-theory-based machine learning approach for identifying different brain networks in healthy subjects and Alzheimer's disease patients [9,10]. The method extracts the optimal features from the MRI connection matrix graph measurements, and inputs them to support vector machine for classification. Similarly, Al-Zubaidi et al. extracted connection parameters from 90 brain regions, and used linear support vector machine sequence with forward floating selection strategy to classify the hunger and satiety states of subjects [9].

Traditional machine learning methods have some limitations due to their inability to correlate the relationships among local voxels in MRI images. However, deep learning, especially, the Convolutional Neural Network (CNN) integrates the location correlation between features, feature extraction and feature selection with machine learning algorithms, and its classification performance in medical images is also outstanding. Deep learning methods have been used in the diagnosis of chronic kidney disease [10], coronary cardiac disease [11] and mental diseases [12,13], chronic myocardial infarction [14], prostate cancer [15] and even the COVID-19 [16]. These applications include extensive automated segmentation of lesion sites. For example, a deep learning framework for the diagnosis of chronic myocardial infarction was proposed by Zhang et al., which extracts local and global motion features, and relates them in late gadolinium enhancement MRI images [14]. This framework got excellent performance, but only with nonenhanced cardiac cine MRI images.

Some mental diseases are difficult to be directly judged from neuroimages, but the convolution

operation in CNN can extract local features of the adjacent voxels, so as to combine the complex features applied to disease classification. The CNN-based approach has been shown to be effective in classifying mental disorders in a number of studies. Saman et al. classified the brain MRI images of subjects by using CNN with LeNet-5 structure [17,18] to judge whether the subjects suffered from Alzheimer's disease. A convolutional neural network was proposed by Zhao et al. [19] based on 3D convolutional kernels, which can accurately classify brain functional networks reconstructed by whole-brain fMRI signals, and this network has a good classification performance in the application of fMRI images from the Human Connectome Project (HCP). To reduce the influence of irrelevant part of MRI images, Zou presented a multi-modality CNN architecture combing fMRI and structural MRI (sMRI) for distinguishing the neuroimages between healthy subjects and subjects with Attention deficit hyperactivity disorder (ADHD) [20]. The network extracts useful features from sMRI and fMRI, and assists in automating ADHD diagnosis. A 3D CNN structure with multiple dilated convolution kernels and its associated computational framework were also proposed, which can be applied to structural MRI (sMRI) and functional MRI (fMRI) classification [21]. The proposed method has a good performance in the classification tasks of ADHD and schizophrenia.

However, most deep learning algorithms based on neuroimaging for diagnosing mental diseases did not consider the influence of brain connectivity disorders on some mental illnesses. There are evidences that some psychiatric disorders result in the changes or disruptions in the structure or functional connections. For example, it has been found that SZ often resulted in a disorder of connectivity between the areas of large-scale brain networks, such as the medial parietal, the premotor, and the cingulate regions [22]. Similarly, ADHD has also been found to have disorders of brain functional connections in the frontal lobe, insula and sensorimotor systems [23,24]. Therefore, the deep learning model's ability to learn long-distance and short-distance structural and functional connections of a brain determines the performance of its recognition of mental diseases.

In this study, the proposed method is mainly applied to the automatic diagnosis of cognitive disorders, such as schizophrenia (SZ) and ADHD.

In this paper, a deep learning structure was proposed based on MRI images, which was designed to consider the brain's connections at different sizes and the attention of connections. We proposed a 3D Multiscale View Convolutional Neural Network with Attention (3D MVA-CNN) based on ResNetXt [25,26] and Sqeeze-and-Excitation (SE) [27]. ResNetXt is a transformation of ResNet [28]. ResNetXt uses a number of parallel residual blocks, which makes the network deeper and wider. This wide structure is suitable for analyzing the whole-brain MRI images that have a mass of features (voxels). We first modified ResNetXt into a 3D structure to make it suitable for fMRI data. Then, the proposed Multiscale View (MV) module, on the one hand, uses the multi-scale convolutional kernel to extract the effective features of coarse-grained and fine-grained voxel activities in MRI images, so as to improve the sensitivity of the algorithm to the disorders of long-distance and short-distance brain connectivity. On the other hand, SE determines the importance of the features extracted by convolutional kernels in different scales dynamically. Similarly, we proposed a path attention module to dynamically assign weights for each parallel path in MV module. We tested the proposed model on a self-scanned fMRI dataset for schizophrenia, and a MRI dataset from ADHD-200 [29]. The results showed that the proposed 3D MVA-CNN has a better performance than some other 3D CNN structures in both two datasets. The main contributions of this study are list as follows:

1) A 3D Multiscale View Convolutional Neural Network with Attention was proposed, which considered the brain's connections at different sizes and the attention of them.

2) The proposed method was tested on an fMRI dataset for schizophrenia and an MRI dataset for ADHD, and outperformed other commonly-used 3D CNN methods.

3) We also presented a framework for clinical application based on the proposed model and discussed the feasibility and limitations of the framework.

The remainder of this paper is organized as follows. In Section II, we will first describe the research works related to the method proposed in this paper. In Section III, we will introduce our proposed 3D MVA-CNN. In Section IV, we will describe the experiments, and discuss the results. Finally, in Section V, we will give the conclusion of this study.

2.   Related works

The proposed 3D MV-CNN is based on ResNetXt and Sqeeze-and-Excitation (SE). In this section, the details of ResNetXt and SE will be introduced.

2.1. ResNetXt

The traditional methods make CNN networks deeper or wider to improve the performance of a CNN for classification task. But usually, the number of hyperparameters will increase along with the deepening or widening. The increasing number of hyperparameters often leads to the complex design and computation. However, the ResNetXt was designed to make networks wider without much extra cost.
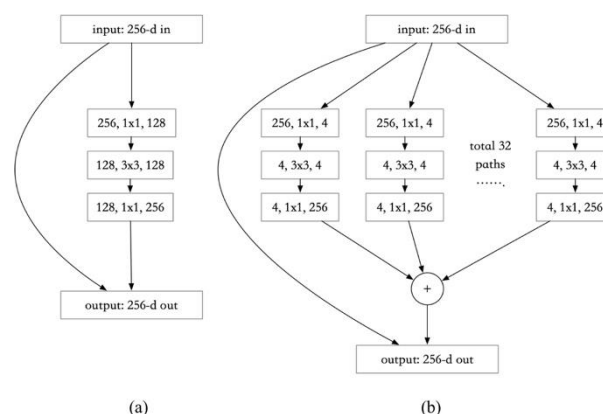

Figure 1. The comparison between ResNet block and ResNetXt block. (a) A ResNet block with three convolutional layers. (b) A ResNetXt block with 32 paralleled paths. Each path in ResNetXt block has the same number of convolutional layers as ResNet block. These layers generate smaller feature maps. All the small feature maps are finally concentrated in the last concentrating layer.

The structure of ResNetXt block is shown in Figure 1(b). This structure consists of several parallel blocks of residuals. The concept of Cardinality is introduced to represent the number of parallel and independent residual paths.

The ResNetXt block is equivalent to the ResNet block in Figure 1(a). Resnet residual paths are divided into multiple residual paths in ResnetXt, in which the number of channels in ResNet residual

paths is averaged among all residual paths in Resnet. Compared with ResNet, the structure of ResNetXT can improve the efficiency of feature extraction while keeping the number of network parameters similar.

2.2. Squeeze-and-Excitation

Attention mechanisms have been widely used in the field of deep learning automated diagnosis. For example, the dilated attention network in the left atrium anatomy and scar segmentations was proposed by Yang et al., which learns the feature maps for left atrial scars [30]. And the feature pyramid attention was applied by Liu et al. in a fully convolutional network for automatic prostate zonal segmentation, which was combined with a modified ResNet50, Feature Pyramid Attention and Decoder [31]. In this work, attention mechanism is applied to the weight assignment of parallel paths in recognition networks by using the Squeeze-and-Excitation.

The Squeeze-and-Excitation is one of the classic implementations for the feature-channel-based attention mechanism implementation, as shown in Figure 2.

SE module first squeezes the feature maps obtained from the convolution layer to obtain the channel-level global features. Then the excitation operation is done to the global features, which learns the relationship between various channels, and obtains the weights of different channels. Finally, the weights of different channels are multiplied by the feature maps from original channels to get the final features. In essence, the SE module does attention or gating operation on channel dimension. This attention operation enables the model to pay more attention to channel features with more information, while suppressing those features that are not important. SE modules are generic, which means they can be embedded into existing network architectures.
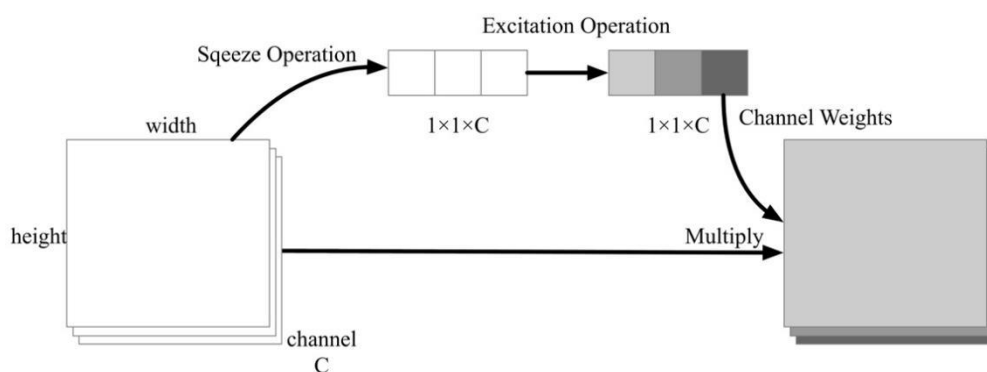

Figure 2. The Squeeze-and-Excitation module.

3. Methods

In this paper, a 3D MVA-CNN model for MRI data classification is proposed. It is also proposed that the classification model can be applied to the classification of multiple fMRI images generated by a single fMRI scan. Therefore, the method proposed in this paper can not only identify the mental diseases related to brain structure, but also identify the mental diseases caused by the changes of brain function.

There are two main improvements in the application of 3D MVA-CNN in automatic diagnostic classification of MRI data proposed in this paper. Firstly, the network structure named Multiscale View similar as ResNetXt was proposed, which applies convolutional kernels with different scales to enhance the sensitivity of the network model to the disorders of brain functional networks of different scales. On the other hand, an attention mechanism similar to the SE module was applied to scale the feature maps generated by the pathways of convolutional layers at different scales, so as to enhance the attention degree of network model to a feature.

3.1. Multiscale view module

The concept of multi-view was used in the artery-specific calcification analysis [32], quantification of coronary artery stenosis [33], left ventricle detection [34] and echocardiographic sequences segmentation [35]. Multi-view in these works mostly means that multiple different types of two-dimensional images are input into deep neural networks at one time, and fused. For example, Zhang et al. used axial, coronal and sagittal views in a multi-task learning network for artery-specific calcification analysis [32].

In this work, we designed a Multiscale View module (MV module) to make neural networks more sensitive to the correlations between regions with different distances. The multiscale views didn't mean the multi input to the network, but were the parallel convolutional layers with different kernels. Different sizes of convolution kernel could extract features of different fields of view. MV module was shown in Figure 3. This module parallels multi paths similar to the regular ResNetXt block.
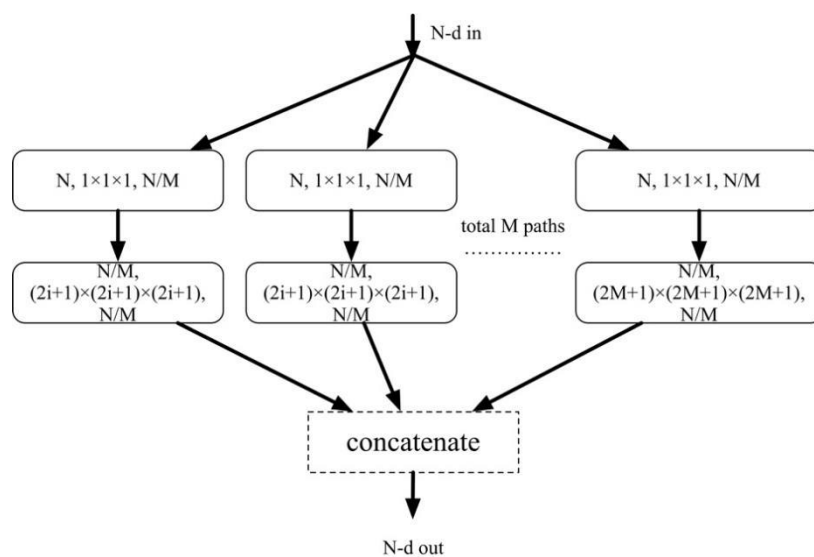

Figure 3. Multiscale View module. The Multiscale View module is the extension of simple ResNetXt block, which parallels multi paths with convolutional layers with different size of kernels. These paths with different kernels could help the 3D CNN find more disordered patterns of brain networks.

This module separates multiple parallel paths from the feature maps output from the previous convolutional neural layer, in which each path has the same number of convolutional layers and each

convolutional layer has the same feature maps, but the sizes of the convolutional kernels on different paths are different. There are two convolutional layers on each path, in which the first convolutional layers map the input images with N channels to images with N/M channels, where N is the number of channels of the output results of the previous layer, and M is the number of paths. The second convolutional layers convolve the output image of the previous layer with the same number of channels. Finally, the output results on all paths are 3D images with N/M channels and the same size. The images on each path are concatenated along the channel dimension and combined into a 3D image with N channels. In the first convolutional layers, strides = $2 \times 2 \times 2$, meaning that the sizes of input images are reduced by half. The sizes of the convolutional kernel in the second convolutional layer are $2i + 1$, where i is the order number of the parallel path, so each path uses $3 \times 3 \times 3$, $5 \times 5 \times 5$ and other sizes of the convolutional kernels to achieve multi-scale feature extraction.
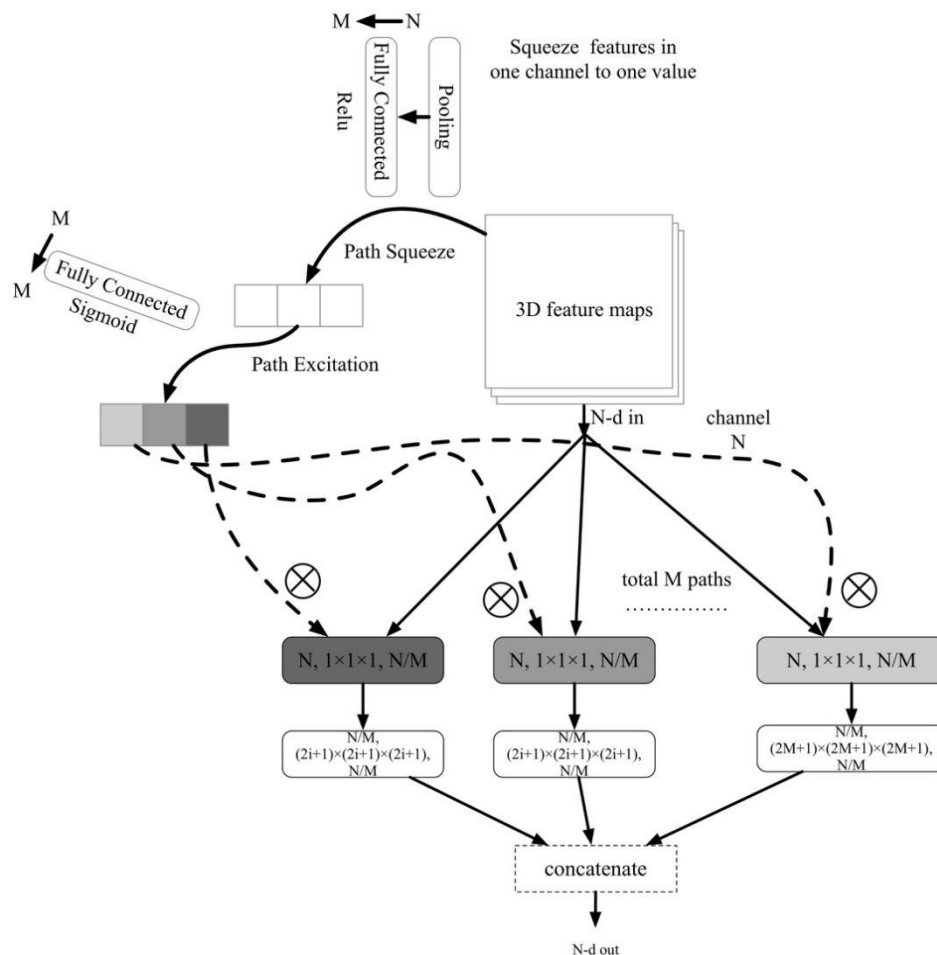
3.2. Path attention

Figure 4. The proposed MV module with path attention. Three steps, Path Squeeze and Path Excitation, are included in path attention. The final M weights are used to scale features in M paths. Finally, the output in each path concatenates together.

We designed an attention mechanism, called Path Attention, for channel weight selection of MV

module by imitating the attention mechanism of SE module.

The design of path attention is shown in Figure 4. This module is used in conjunction with the MV module to dynamically assign weights to parallel paths in the MV module. This module has two parts: Path Squeeze and Path Excitation.

The purpose of the Path Squeeze operation is to gather the global features of each channel resulted from the previous convolution operation. The Path Squeeze operation first squeezes the 3D feature map of each channel with a pooling layer and obtains an average feature. A feature graph set with N channels can be pooled to obtain N feature values, each of which represents the global feature of the corresponding channel. Then, a full connection layer with a ReLU activation is used to transform the global features of N channels into M values, where M is the number of parallel paths in the MV module.

The purpose of Path Excitation is to convert M features to the attention levels (weights) of M paths. There is a full connection layer with a Sigmoid activation. The full connection layer converts M features to M weights between 0–1.

Finally, M weights will be provided to M paths of the MV module, and multiplied by the first convolutional output on each path to scale the output features. The MV module with path attention is called MVA module in this paper.

4. Experiments and results

Two experiments were conducted to compare several deep learning methods with proposed MVA-CNN in the classification of fMRI images for SZ and sMRI images for ADHD. The results indicated that the proposed method could outperform other deep learning structures in mental disease automatic diagnosis based on fMRI and sMRI data.

4.1. Datasets

In Experiment 1, we used ADHD-200 dataset to test the proposed method. The ADHD-200 data set includes sMRI and fMRI images of about 800 subjects, which were spontaneously provided by eight scientific research institutions, including Brown University, University of Pittsburgh, New York University Medical Center, Peking University, etc. In the experiment, 587 sMRI images provided by some institutions were selected for training and testing, involving 441 healthy subjects' images and 146 patients' images, each with the size of $121 \times 145 \times 121$ voxels.

In Experiment 2, we tested the proposed method in a self-collected fMRI dataset of SZ. We scanned the EPI resting fMRI images using a 3-T GE MRI scanner. The repetition time (TR) was 2000 ms, and the echo time (TE) was 30 ms. Each image was scanned 50 in slices. 28 healthy and 28 patient subjects participated in the experiment (age range: 15–44, healthy subjects: 17 females and 11 males, patients: 14 females and 14 males), 50 resting MRI images of each subject were used in classification. Totally 2800 fMRI images were used in classification. During MRI scanning, the subjects were asked to lie in the MRI scanner and stay awake without any action. The subjects' heads were fixed by four sponges.

The images were processed using SPM8 (https://www.fil.ion.ucl.ac.uk/) and MATLAB (Mathworks, Natic, MA) software. Images were realigned, co-registered and normalized on the Montreal Neurological Institute (MNI) template. All images are normalized and reshaped into the size of $61 \times 73 \times 61$ voxels.
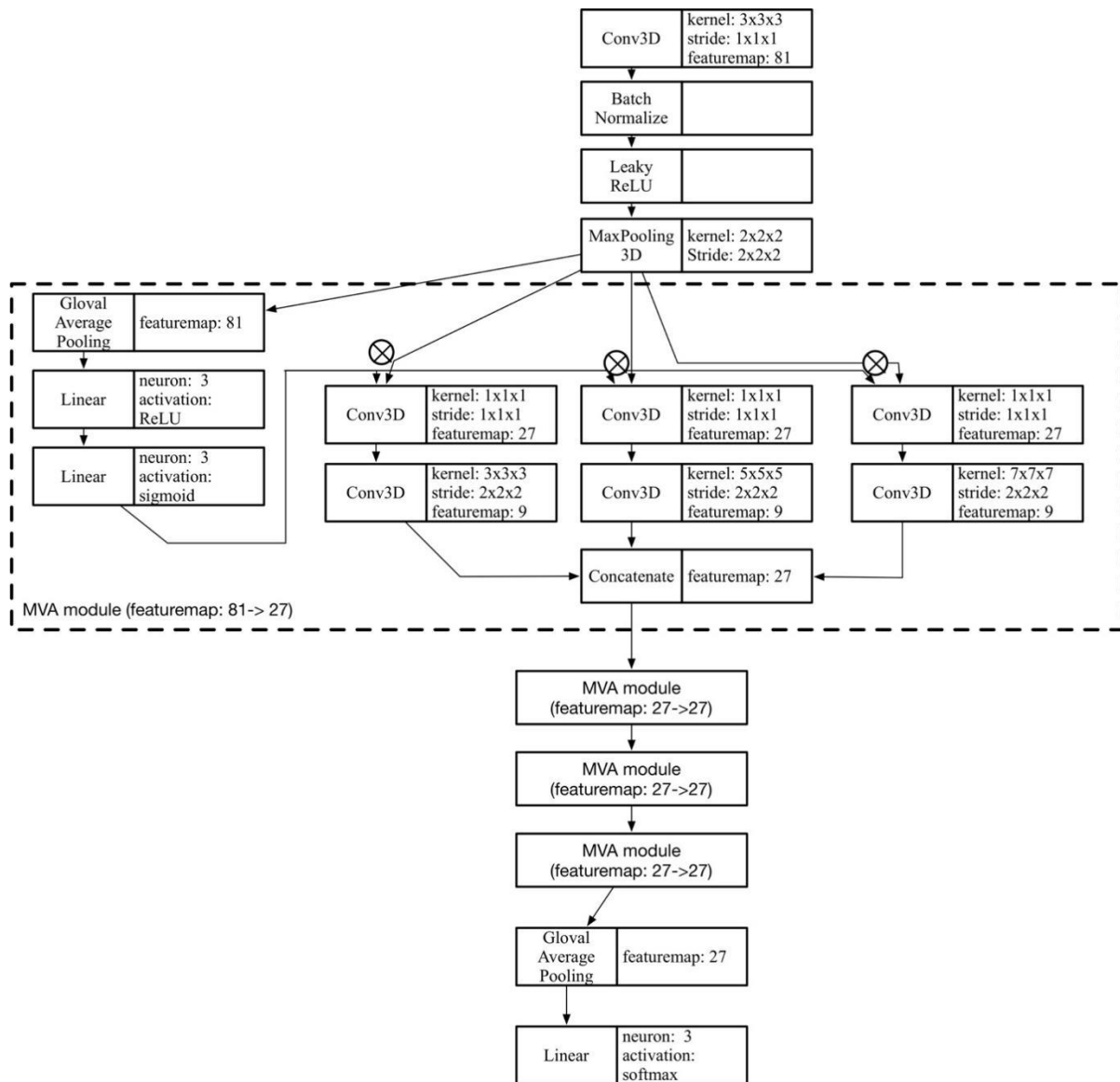
4.2. Experiment setup

Figure 5. The 3D MVA-CNN structure used in the two experiments. Here, the details of the first MVA module are expanded and drawn out. Due to space, we did not expand the following three MVA modules. In order to reduce overfitting and the memory usage, a small number of feature graphs were used in the structure.

In both experiments, the 3D MVA-CNN structure we used is shown in Figure 5, in which four continuous MVA modules are used, and the number of parallel paths in each Module M is 3. Because the sample size used in the two experiments is not very large, the complex model is easy to cause overfitting. We reduced the number of feature maps in 3D MVA-CNN and other models in the experiments to reduce model complexity and memory consumption. Only 160k trainable parameters were used in the 3D MVA-CNN structure. The experiment will calculate the classification accuracy, sensitivity, specificity and AUC of the proposed method, and compare with the performances of ResNet, ResNetXt, VGGNet, AlexNet, SparseNet [36] and Inception-V3 [37]. The paired t test was

also employed to check whether the AUC score of the proposed method was significantly different from those of other models. We also compared the performance of 3D MVA-CNN with 2 times of feature maps to test the effect of model complexity on classification performance.

All models were adjusted to accommodate the preprocessed sizes of MRI images in experiments. All the convolutional layers were transformed into convolutional layers with 3D cores, and all the pooling layers were transformed into pooling layers with 3D convolutional kernels. Similar to 3D MVA-CNN, the feature maps in these models were reduced to avoid overfitting. The layers in ResNet and ResNetXt were reduced to fit the size of MRI images.

Five-fold cross validation was used in both experiments. Experiments were conducted to compare the average measures of cross-validation results. In the ADHD-200 dataset, 470 images were used for training and 117 for testing in each validation. 353 samples of healthy subjects and 117 samples of patients were used in training. And 88 samples of healthy subjects and 29 samples of patients were used in the testing phase. Due to the small number of patient samples, the patient samples from the training set were oversampled prior to training. In Experiment 2, datasets were divided according to subjects. In each validation, 2,300 fMRI images of 23 patients and 23 healthy subjects were used for training, and the remaining 500 images of 5 healthy subjects and 5 patients were used for testing.

A normal distribution was used to initialize the weights in the networks, and the learning rate was set to 10–5. RMSProp was used as the optimizer, and the classification cross entropy was used as the loss function for all models. Each training iteration contains 100 epochs, and the batch size is 16. Experiments were performed in software environment: Centos 7.5 64bit, python 3.6, pytorch 1.9.0 and hardware environment: Intel Xeon E5 2680 V3, RAM 64G, and a NVIDIA GeForce RTX 3090 GPU.

4.3. Results

4.3.1.   Experiment 1

In Experiment 1, the classification results of seven deep learning models are shown in Table 1. And the confusion matrix of all the models is shown in Figure 6. As shown in Table 1 and Figure 6, it can be seen that although we have achieved the balance of the sample size by over-sampling the samples of the patient class, the predicted results were still more biased to the healthy subject class. This might be because that although oversampling increased the number of samples, it did not extend the diversity of samples, so it did not completely eliminate the bias of the classifier in training. Compared with the other models, the 3D MVA-CNN method achieved the highest average accuracy of 78.8%, the highest average sensitivity of 38.1% and the highest average AUC of 69.7%. Although the specificity of ResNetXt is higher than that of 3D MVA-CNN, it might be the cause of over-fitting, which output healthy labels in almost all the predictions. The bias made its specificity approach to 100% but the sensitivity too low. This conclusion can also be easily seen in Figure 6(e). According to the paired t test results, it could be found that the AUC of 3D MVA-CNN was significantly higher than all other models except for ResNet (T (4) = 2.765, p = 0.051). Although compared with ResNet, the P value of the proposed model was not significantly less than 0.05, it was very close to 0.05. Overall, 3D MVA-CNN has the best performance among the seven models in Experiment 1 with the acceptable evaluation time.
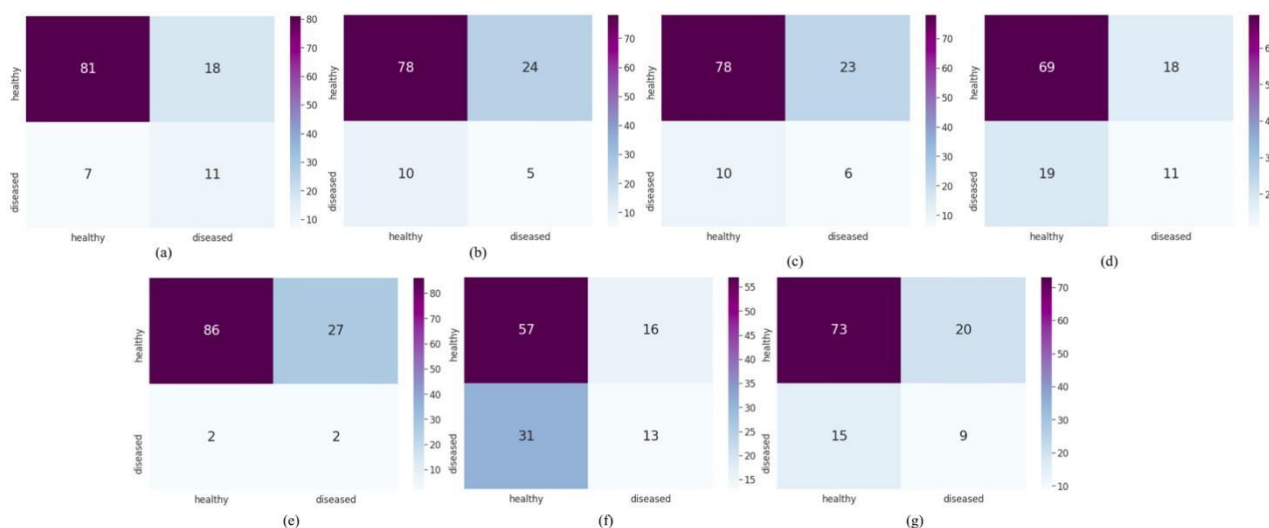
Figure 6. The confusion matrices of seven models in Experiment 1. (a) The confusion matrix of the 3D MVA-CNN. (b) The confusion matrix of the VGGNet. (c) The confusion matrix of the AlexNet. (d) The confusion matrix of the ResNet. (e) The confusion matrix of the ResNetXt. (f) The confusion matrix of the SparseNet. (g) The confusion matrix of the Inception-V3.

Table 1. The results of seven models in Experiment 1.

Model	Avg. Accuracy	Avg. Sentivity	Avg. Specificity	Avg. AUC	Running time (per epoch)	Inference time (per sample)	t value	p value
VGGNet	70.2%	16.4%	88.2%	56.8%	42 s	0.13 s	9.333	< 0.001*
AlexNet	72.4%	21.9%	89.1%	59.2%	46 s	0.14 s	6.823	0.002*
ResNet	68.5%	37.7%	78.9%	63.8%	76 s	0.24 s	2.765	0.051
ResNetXt	75.5%	7.5%	98.2%	47.8%	62 s	0.20 s	7.011	0.002*
SparseNet	57.9%	44.9%	65.1%	61.2%	92 s	0.29 s	6.383	0.003*
Inception-V3	72.4%	30.1%	82.8%	33.1%	85 s	0.27 s	9.513	< 0.001*
3D MVA-CNN	78.8%	38.1%	92.3%	69.7%	81 s	0.26 s	/	/
3D MVA-CNN with 2x feature maps	71.1%	31.0%	81.8%	60.9%	119 s	0.38 s	/	/

We also compared the results of 3D MVA-CNN on the ADHD dataset with those reported in the previous studies, as shown in Table II. All these studies used deep learning method to train and classify the selected data from ADHD-200 dataset. Among them, only sMRI data from ADHD-200 dataset was used in the study of Zou [38] and Wang [21], while sMRI and fMRI data were used in the study of Sen [22] and Sina [39]. The sample size used in this study was similar to that used in these studies. We used the same data as [40] did. It was found that our proposed 3D MVA-CNN was able to achieve the highest classification accuracy using only sMRI data, even though some of the methods proposed in the previous studies were able to integrate the features of sMRI and fMRI.

Table 2. Comparison between 3D MVA-CNN and previous methods.

Model	Accuracy	Data Combination
Zou et al. 2017 [22]	65.9%	sMRI
Sen et al. 2018 [21]	68.9%	sMRI + fMRI
Sina et al. 2016 [39]	70.0%	sMRI + fMRI
Wang et al. 2019 [40]	76.6%	sMRI
3D MVA-CNN	78.8%	sMRI

4.3.2.   Experiment 2

Table 3. The results of seven models in Experiment 2.

Model	Avg. Accuracy	Avg. Sentivity	Avg. Specificity	Avg. AUC	Running time (per epoch)	Inference time (per sample)	t value	p value
VGGNet	68.6%	64.8%	72.3%	78.7%	82 s	0.05 s	2.803	0.049*
AlexNet	71.0%	75.1%	66.1%	75.6%	84 s	0.05 s	4.531	0.011*
ResNet	54.1%	45.4%	61.8%	56.7%	149 s	0.10 s	21.063	< 0.001*
ResNetXt	69.5%	72.5%	67.2%	66.5%	122 s	0.08 s	8.826	< 0.001*
SparseNet	79.6%	77.2%	81.8%	81.2%	182 s	0.12 s	3.279	0.031*
Inception-V3	57.9%	50.6%	65.1%	54.6%	168 s	0.11 s	17.576	0.001*
3D MVA-CNN	84.2%	79.4%	88.6%	84.3%	164 s	0.11 s	/	/
3D MVA-CNN with 2x feature maps	77.8%	80.0%	75.6%	79.1%	225 s	0.15 s	/	/

The comparison results of seven models in Experiment 2 are shown in Table 3 and Figure 7. The scores were at the sample level and not aggregated to the subject level. It can be seen that for fMRI data of schizophrenia, the proposed 3D MVA-CNN is far superior to other models in all measurements, with an average accuracy of 88.2%, an average sensitivity of 79.4%, an average specificity of 88.6%, and an AUC of 84.3%. The AUC of 3D MVA-CNN was significantly higher than those of all other models, which means a big improvement of the proposed method. In Figure 7, it is also easy to find from the confusion matrix that the proposed 3D MVA-CNN performed best for schizophrenia fMRI classification. While ResNet and Inception-V3 performed worst, although both of them were reported to have excellent performance for 2D image classification. In addition, in both experiments, the 3D MVA-CNN models with fewer features performed better than the models with more features. This indicates that the proposed model may perform better when the structure is relatively simple. The results of the proposed 3D MVA-CNN in the two experiments are higher than those of other commonly used CNN networks. This result indicates that the 3D MVA-CNN proposed in this paper has a good performance in the application of mental disease classification based on MRI data. Both ADHD and schizophrenia were involved in our study, which have been found to be strongly associated with brain network disorders. The experimental results also illustrate that the multi-size path mechanism of the multiscale view module and the attention selection mechanism for multi-size paths of the path attention module designed in this study are helpful to improve the sensitivity of deep convolutional neural
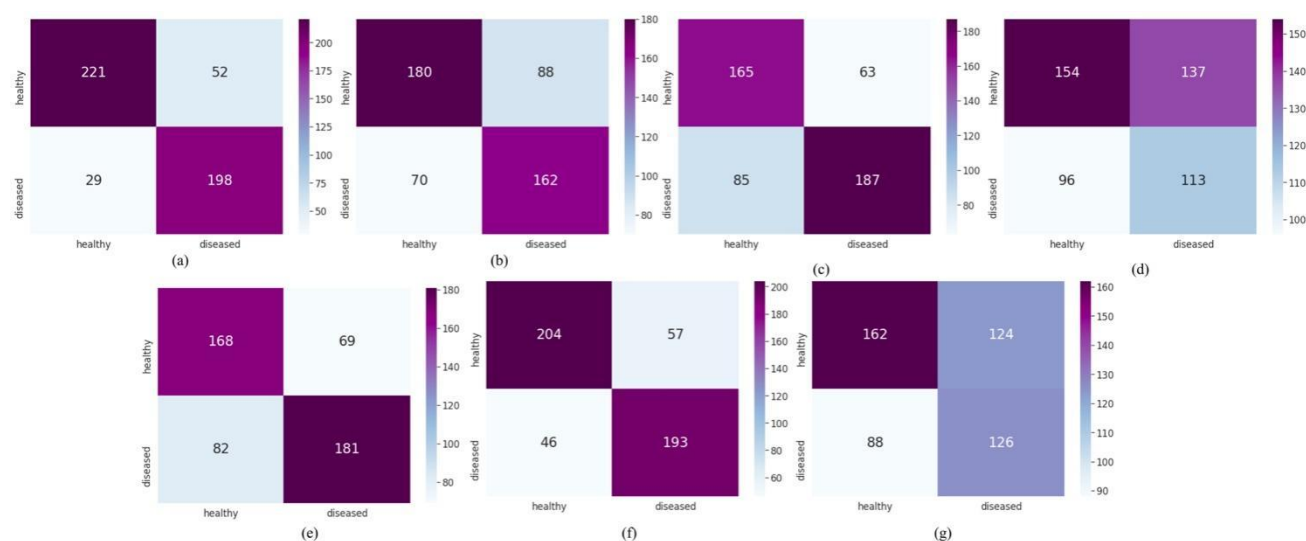
network to brain network changes.

Figure 7. The confusion matrices of seven models in Experiment 2. (a) The confusion matrix of the 3D MVA-CNN. (b) The confusion matrix of the VGGNet. (c) The confusion matrix of the AlexNet. (d) The confusion matrix of the ResNet. (e) The confusion matrix of the ResNetXt. (f) The confusion matrix of the SparseNet. (g) The confusion matrix of the Inception-V3.

The proposed CNN method can also be applied in practical clinical application and research under certain conditions. To this end, we also designed a preliminary clinical application framework, as shown in Figure 8. However, there are three limitations to this framework needing to be improved in the future:

1) Although the ideal data access method is to obtain the scanned MRI images directly through a data interface, it is necessary to get the data interface from the MRI scanner manufacturer, which might be difficult (① in Figure 8). Therefore, a feasible solution is that the doctors manually download MRI images and upload them to a server running the CNN model (② in Figure 8).

2) As the accuracy of such algorithms is far from 100%, the output results could only be used as a reference. Doctors still need to make a final judgment based on the results of the model and other assessments. Further improving the diagnostic performance and the interpretability of the model would realize more accurate computer-aided diagnosis in the future.

3) At present, the deep learning automated diagnosis method based on multi-modal or multi-sequence MRI input has become an important research direction in the interdisciplinary field of artificial intelligence and medical engineering [41]. However, the input and fusion of multi-modal MRI data have not been considered in the proposed method. The proposed approach may have the potential to process multimodal MRI data due to its inherent parallel feature selection paths. In the future, we would carry out further research on this issue.
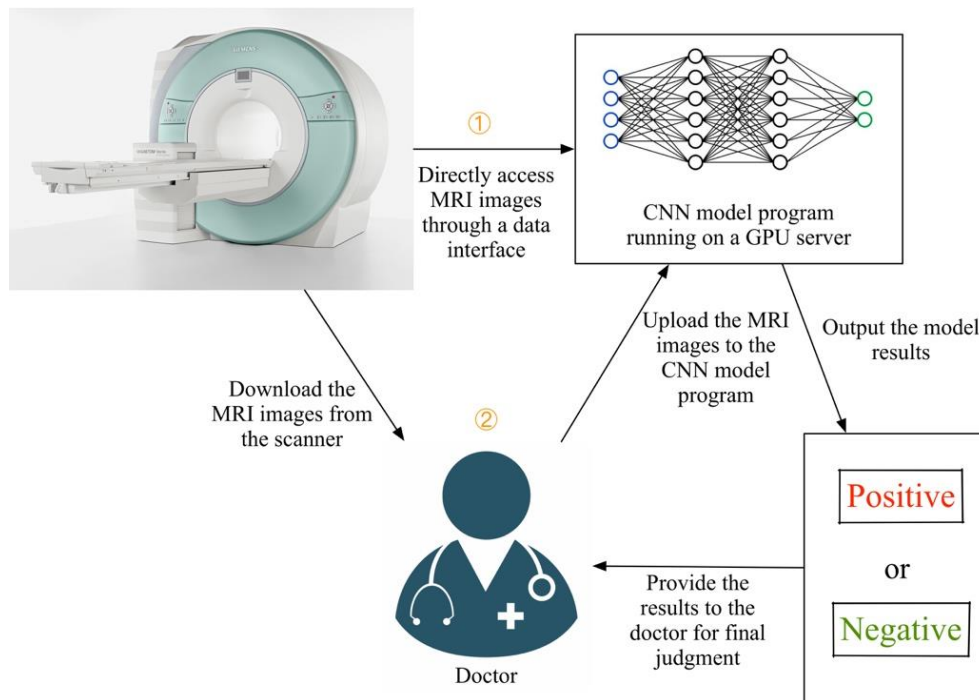
Figure 8. The preliminary clinical application framework for the proposed CNN method.

5. Conclusions

In this paper, we proposed a 3D Multiscal View Convolutional Neural Network with Attention (3D MVA-CNN) for automatic diagnosis of mental disease based on brain MRI data. 3D MVA-CNN is based on RexNetXt, with the addition of our proposed Multiscale View module and path attention, which is transformed into 3D convolutional network. The Multiscale View module is similar to the ResNetXt block. However, it can detect brain functional connections or activity associations of brain regions at different scales by allocating convolutional layers with different sizes of convolutional kernels to each parallel path. Path attention for MV module implements Path selective attention mechanism, and the design concept is similar to Squeeze-and-Excitation. It provides dynamic weights for the parallel paths of the MV modules through the path squeeze and path excitation operations. It increases the network sensitivity to the connection between brain regions at a certain scale. Finally, we tested 3D MVA-CNN on the sMRI dataset of ADHD-200 and the fMRI dataset of schizophrenia collected by ourselves. The results show that the application of our proposed 3D MVA-CNN has excellent performance in the structural MRI and functional MRI data classification of mental diseases related to brain network disorders. This work further advances the application of deep learning methods based on neuroimaging in the diagnosis of neurological diseases.

Acknowledgments

References

1. S. G. Shamay-Tsoory, J. Aharon-Peretz, Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study, *Neuropsychologia*, **45** (2007), 3054–3067.

2. M. Hu, K. Sim, J. H. Zhou, X. Jiang, C. Guan, Brain MRI-based 3D Convolutional Neural Networks for Classification of Schizophrenia and Controls, *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, (2020), 1742–1745

3. E. Li, The application of BOLD-fMRI in cognitive neuroscience, *J. Frontiers Comput. Sci. Technol.*, **2** (2008), 589–600.

4. K. J. Friston, L. Harrison, W. Penny, Dynamic causal modelling, *Neuroimage*, **19** (2003), 1273–1302.

5. F. Pereira, T. Mitchell, M. Botvinick, Machine learning classifiers and fMRI: a tutorial overview, *Neuroimage*, **45** (2009), S199–S209.

6. S. Lemm, B. Blankertz, T. Dickhaus, K. R. Müller, Introduction to machine learning for brain imaging, *Neuroimage*, **56** (2011), 387–399.

7. J. A. Suykens, J. Vandewalle, Least squares support vector machine classifiers, *Neural Process. Lett.*, **9** (1999), 293–300.

8. R. Hecht-Nielsen, *Neural Networks for Perception*, Academic Press, 1992.

9. A. Khazaee, A. Ebrahimzadeh, A. Babajani-Feremi, Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer's disease, *Brain Imaging Behav.*, **10** (2016), 799–817.

10. A. Al-Zubaidi, A. Mertins, M. Heldmann, K. Jauch-Chara, T. F. Münte, Machine learning based classification of resting-state fMRI features exemplified by metabolic state (hunger/satiety), *Front. Hum. Neurosci.*, **13** (2019), 164.

11. S. Patil, S. Choudhary, Deep convolutional neural network for chronic kidney disease prediction using ultrasound imaging, *Bio-Algorithms Med. Syst.*, **17** (2021), 137–163

12. A. Dutta, T. Batabyal, M. Basu, S. T. Acton, An efficient convolutional neural network for coronary heart disease prediction, *Expert Syst. Appl.*, **159** (2020), 113408.

13. Y. Cao, Z. Wang, Z. Liu, Y. Li, X. Xiao, L. Sun, et al., Multiparameter synchronous measurement with IVUS images for intelligently diagnosing coronary cardiac disease, *IEEE Trans. Instrum. Meas.*, (2020), 1–1

14. N. Zhang, G. Yang, Z. Gao, C. Xu, Y. Zhang, R. Shi, et al., Deep learning for diagnosis of chronic myocardial infarction on nonenhanced cardiac cine MRI, *Radiology*, **291** (2019), 606–617.

15. Y. Jin, G. Yang, Y. Fang, R. Li, X. Xu, Y. Liu, et al., 3D PBV-Net: an automated prostate MRI data segmentation method, *Comput. Biol. Med.*, **128** (2021), 104160.

16. D. Driggs, I. Selby, M. Roberts, E. Gkrania-Klotsas, J. H. Rudd, G. Yang, et al., Machine learning for COVID-19 diagnosis and prognostication: lessons for amplifying the signal while reducing the noise, *Radiol. Artif. Intell.*, **3** (2021), e210011.

17. S. Sarraf, G. Tofighi, Classification of alzheimer's disease using fmri data and deep learning convolutional neural networks, preprint, arXiv:1603.08631

18. S. Sarraf, D. D. DeSouza, J. Anderson, G. Tofighi, DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI, preprint, BioRxiv:070441.

19. Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE*, **86** (1998), 2278–2324.

20. Y. Zhao, Q. Dong, S. Zhang, W. Zhang, H. Chen, X. Jiang, et al., Automatic recognition of fMRI-derived functional networks using 3-D convolutional neural networks, *IEEE Trans. Med. Imaging*, **65** (2017), 1975–1984.

21. L. Zou, J. Zheng, C. Miao, M. J. Mckeown, Z. J. Wang, 3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI, *IEEE Access*, **5** (2017), 23626–23636.

22. Z. Wang, Y. Sun, Q. Shen, L. Cao, Dilated 3D convolutional neural networks for brain MRI data classification, *IEEE Access*, **7** (2019), 134388–134398.

23. A. G. Garrity, G. D. Pearlson, K. McKiernan, D. Lloyd, K. A. Kiehl, V. D. Calhoun, Aberrant "default mode" functional connectivity in schizophrenia, *Am. J. Psychiatry*, **164** (2007), 450–457.

24. M.-E. Lynall, D. S. Bassett, R. Kerwin, P. J. McKenna, M. Kitzbichler, U. Muller, et al., Functional connectivity and brain networks in schizophrenia, *J. Neurosci. Res.*, **30** (2010), 9477–9487.

25. M. Murias, J. M. Swanson, R. Srinivasan, Functional connectivity of frontal cortex in healthy and ADHD children reflected in EEG coherence, *Cereb. Cortex*, **17** (2007), 1788–1799.

26. D. Fair, J. T. Nigg, S. Iyer, D. Bathula, K. L. Mills, N. U. Dosenbach, et al., Distinct neural signatures detected for ADHD subtypes after controlling for micro-movements in resting state functional connectivity MRI data, *Front. Syst. Neurosci.*, **6** (2013), 80.

27. S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, *IEEE Comput. Conf. Comput. Vis. Pattern Recogn.*, (2017), 1492–1500

28. J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, *IEEE Comput. Conf. Comput. Vis. Pattern Recogn.*, (2018), 7132–7141

29. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, *IEEE Comput. Conf. Comput. Vis. Pattern Recogn.*, (2016), 770–778

30. G. Yang, J. Chen, Z. Gao, S. Li, H. Ni, E. Angelini, et al., Simultaneous left atrium anatomy and scar segmentations via deep learning in multiview information with attention, *Future Gener. Comput. Syst.*, **107** (2020), 215–228.

31. Y. Liu, G. Yang, S. A. Mirak, M. Hosseiny, A. Azadikhah, X. Zhong, et al., Automatic prostate zonal segmentation using fully convolutional network with feature pyramid attention, *IEEE Access*, **7**(2019), 163626–163632.

32. W. Zhang, G. Yang, N. Zhang, L. Xu, X. Wang, Y. Zhang, et al., Multi-task learning with multi-view weighted fusion attention for artery-specific calcification analysis, *Inf. Fusion*, **71** (2021), 64–76.

33. D. Zhang, G. Yang, S. Zhao, Y. Zhang, D. Ghista, H. Zhang, et al., Direct quantification of coronary artery stenosis through hierarchical attentive multi-view learning, *IEEE Trans. Med. Imaging*, **39** (2020), 4322–4334.

34. M. Yang, X. Xiao, Z. Liu, L. Sun, W. Guo, L. Cui, et al., Deep retinaNet for dynamic left ventricle detection in multiview echocardiography classification, *Sci. Program*, **2020** (2020), 7025403

35. M. Li, C. Wang, H. Zhang, G. Yang, MV-RAN: Multiview recurrent aggregation network for echocardiographic sequences segmentation and full cardiac cycle analysis, *Comput. Biol. Med.*, **120** (2020), 103728.

36. M. R. Brown, G. S. Sidhu, R. Greiner, N. Asgarian, M. Bastani, P. H. Silverstone, et al., ADHD-200 global competition: diagnosing ADHD using personal characteristic data can outperform resting state fMRI measurements, *Front. Syst. Neurosci.*, **6** (2012), 69.

37. W. Liu, K. Zeng, SparseNet: A sparse denseNet for image classification, preprint, arXiv:1804.05340

38. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, *IEEE Comput. Conf. Comput. Vis. Pattern Recogn.*, (2016), 2818–2826

39. B. Sen, N. C. Borle, R. Greiner, M. R. Brown, A general prediction model for the detection of ADHD and Autism using structural and functional MRI, *PloS One*, **13** (2018), e0194856.

40. S. Ghiassian, R. Greiner, P. Jin, M. R. Brown, Using functional or structural magnetic resonance images and personal characteristic data to identify ADHD and autism, *PloS One*, **11** (2016), e0166934.

41. F. Raschke, T. R. Barrick, T. L. Jones, G. Yang, X. Ye, F. A. Howe, Tissue-type mapping of gliomas, *NeuroImage: Clin.*, **21** (2019), 101648.