IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 25, 2019, accepted September 11, 2019, date of publication September 17, 2019,
date of current version September 30, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2941912

Dilated 3D Convolutional Neural Networks for Brain MRI Data Classification

ZIJIAN WANG[1], (Student Member, IEEE), YAORU SUN[1], QIANZI SHEN[1], AND LEI CAO[2], (Member, IEEE)

[1]Department of Computer Science, Tongji University, Shanghai 200000, China
[2]Department of Computer Science, College of Information Engineering, Shanghai Maritime University, Shanghai 200000, China

Corresponding author: Yaoru Sun (yaoru@tongji.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 91748122.

ABSTRACT Benefiting from the research of machine learning (ML) and deep learning(DL), multivariate methods based on ML and DL have been the mainstream and successful analysis methods in Neural Engineering or Neuroimaging research, for example, assisting diagnosis based on brain Magnetic Resonance Imaging (MRI). However, many existing methods based on traditional ML methods cannot sufficiently extract discriminative features, especially feature patterns across long-distance brain areas, resulting in unsatisfactory classification performance. Designing an effective and robust classifier for different MRI images remains a challenge. In this paper, we introduced dilated 3D CNN method for classifying 3D MRI images combining CNN structure and dilated convolution with a small number of feature maps. We also presented a methodology framework based on dilated 3D CNN method, which can classify both single MRI images and image sequences. Our method and framework were evaluated on the structural MRI images of ADHD-200 dataset and fMRI images of a Schizophrenia dataset, demonstrating better performances than some other state-of-the-art methods.

INDEX TERMS Biomedical image processing, magnetic resonance imaging, machine learning.

I. INTRODUCTION

In the last two decades, neuroscience and neuroimaging researchers relied on Univariate analysis, which compares patients against healthy subjects and finds anatomical or functional differences at a group level. However, these simple and interpretable methods have two defects. First, univariate analysis methods based on the assumption that activities within different brain regions or voxels are independent. However, this assumption is not in accord with present findings of brain function, which translate some brain function into network-level activities [4]–[6]. Second, univariate analysis can't be used to make statistical inferences at an individual level, which means these methods are failing to make diagnostic decisions.

In the past few years, multivariate analysis methods in neuroimaging research such as Magnetic Resonance Imaging (MRI) has become a major area of interest in the field of Neural Engineering and Artificial Intelligence [1]–[3]. Multivariate analysis methods introduced machine learning into the field of biomedical research as a powerful tool, which develops algorithms that mine the patterns of existing data and predict the patterns of new data. Machine learning algorithms employ multivariate and the relations among variates, which overcome the first defect of univariate analysis methods. They can also make statistical inferences at a subject or even image level, which overcome the second defect of univariate analysis methods [7].

Machine learning algorithms have been successfully applied to MRI data from cognitive experiments or patient scanning [7], [8]. Moreover, Support Vector Machine (SVM) is one of the most popular methods in this research field. Kernel function used in SVM resolves the classification problem where the two classes are not linearly separable in [10], [14]. SVM has been used in a large number of neuroimaging research, including classification of functional task states or resting states of patients and healthy subjects and estimation of subjects' traits, such as age and gender [15]–[17], [19].

There were also other kinds of machine learning algorithms were used in neuroimaging research. Chou et al. [20] presented a voxel (feature) selection algorithm based on

The associate editor coordinating the review of this manuscript and approving it for publication was Victor Hugo Albuquerque.

Mutual Information and Partial Least Squares for recognition of different types of fMRI Images. An information index for selecting voxels with more information was presented based on their relational degree with the experimental conditions. Khazaee *et al.* presented a machine learning method based on graph theory for recognizing different brain networks between healthy subjects and patients with Alzheimer's disease [21]. This method was used to extract optimal features from functional MRI (fMRI) connectivity matrix graph measures, and the features were feed into SVM for classification. Al-Zubaidi et al. had a similar result in classifying human with metabolic states (hunger/satiety) [22]. The extracted connectivity parameters from 90 brain regions and used sequential forward floating selection strategy with linear SVM to classify the human with the two states.

However, in recent years, success on computer vision, natural language processing of Deep learning (DL), an alternative algorithm family of machine learning algorithms, has been gaining attention in the other research fields, including Biomedical Engineering and Neuroscience [23]–[26]. Zhao *et al.* [30] presented a deep 3D Convolutional Neural Network (CNN) structure for recognizing a large number of sparse presentation of functional brain networks reconstructed by whole-brain fMRI signals. To reduce the influence of irrelevant part of fMRI images, the experimental results based on multiple fMRI data sets showed that the 3D CNN structure was effective and robust. Zou *et al.* presented a multi-modality CNN architecture combing fMRI and structural MRI (sMRI) for distinguishing neuroimages between healthy subjects and subjects with Attention deficit hyperactivity disorder (ADHD) [31]. In this method, refined features were extracted and combined from both fMRI and sMRI images and fed into the classifier. This structure achieved high accuracy of 69.15% in the data from ADHD-200 competition. Besides CNN structures, other types of deep learning algorithms have also been employed in fMRI classification. Xu *et al.* proposed a framework based on Tensor Neural Network that could avoid the curse of dimensionality problem and extract useful features from fMRI images. Their framework outperformed traditional machine learning methods in both binary-class and multi-class fMRI classification tasks [32]. Kasabov *et al.* proposed a new method based on NeuCube Spiking Neural Networks (SNN) for classifying subject's cognitive state, in which the fMRI images were encoded into spiking sequences [33].

Although the above mentioned methods have achieved outstanding performance for neuroimage, especially fMRI image classification, the following problems exist: (1) the sample size is usually small in neuroimage research. The small dataset could probably lead model overfitting or underfitting. (2) Some mental disorders, for example, ADHD, are concurrent with brain activity changes in patient's brain functional connectivity [34], [35]. However, the patterns of functional connectivity between two faraway brain regions are difficult to be detected by standard convolution layers. Currently, there is no machine learning or other classification methods used in brain image classification to conquer these problems.

To tackle these problems, we introduced effective dilated CNN classification algorithm into brain MRI images classification. Comparing with existing research for MRI image classification, several significant contributions that this study makes are described as follows: 1) We introduced a dilated 3D CNN algorithm with embedded dilated convolution layers. The embedded dilated convolution layers enlarge the receptive fields and detect the voxel patterns between long-distance brain areas. 2) We proposed a novel framework that can classify the subject's health state not only by a sequence of fMRI epipolar plane images (EPI) but also by a single EPI image. 3) Our proposed algorithm can outperform some state-of-the-art algorithms for classifying a single structural MRI (sMRI) image of different subjects. We performed two experiments, which classified MRI images of patients and healthy subjects. The first one used sMRI images from ADHD-200 datasets. The second experiment used fMRI images of healthy subjects and patients with schizophrenia. Proposed algorithm outperformed other algorithms in both experiments. We also achieved the highest accuracy in ADHD-200 structural MRI data in comparison with previous related research.

The rest of this paper proceeds as follows: we will introduce relevant literature and methods in Section 2. Section 3 details the presented algorithm and framework in this paper. Section 4 conducts two experiments and the experimental results are analyzed. We will draw conclusions in section.

II. RELATED WORK

Over the past few years, various approaches of neuroimaging classification have been proposed to increase classification performances. We review some CNN classification structures that have been used in neuroimaging, and also some methods based on dilated convolutional feature extraction.

A. CNN USED IN NEUROIMAGING

Multiple CNN structures that had performed well in image recognition has been introduced into neuroimaging classification. These structures also showed their power in neuroimaging classification. LeNet [36] was introduced for classifying Autism functional MRI images [37]. Convolution layers, pooling layers, and fully connected layers constitute the LeNet. The LeNet structure is shown in Fig. 1(a), the first four layers extract features that are fully connected to the output for image classification. ResNet proposed by He [38], which contains residual structures (shown in Fig. 1(b)) was also introduced in pancreatic tumor recognition based on MRI images [39]. The ResNet structure has performed well in tumor recognition of MRI images and achieved the accuracy of 91% in their experiments.

Previous successful application of classical CNN structures from image classification to neuroimaging classification enlightened us. We introduced a CNN structure into MRI images classification, which included layers that were all transformed into 3-dimensional layers.

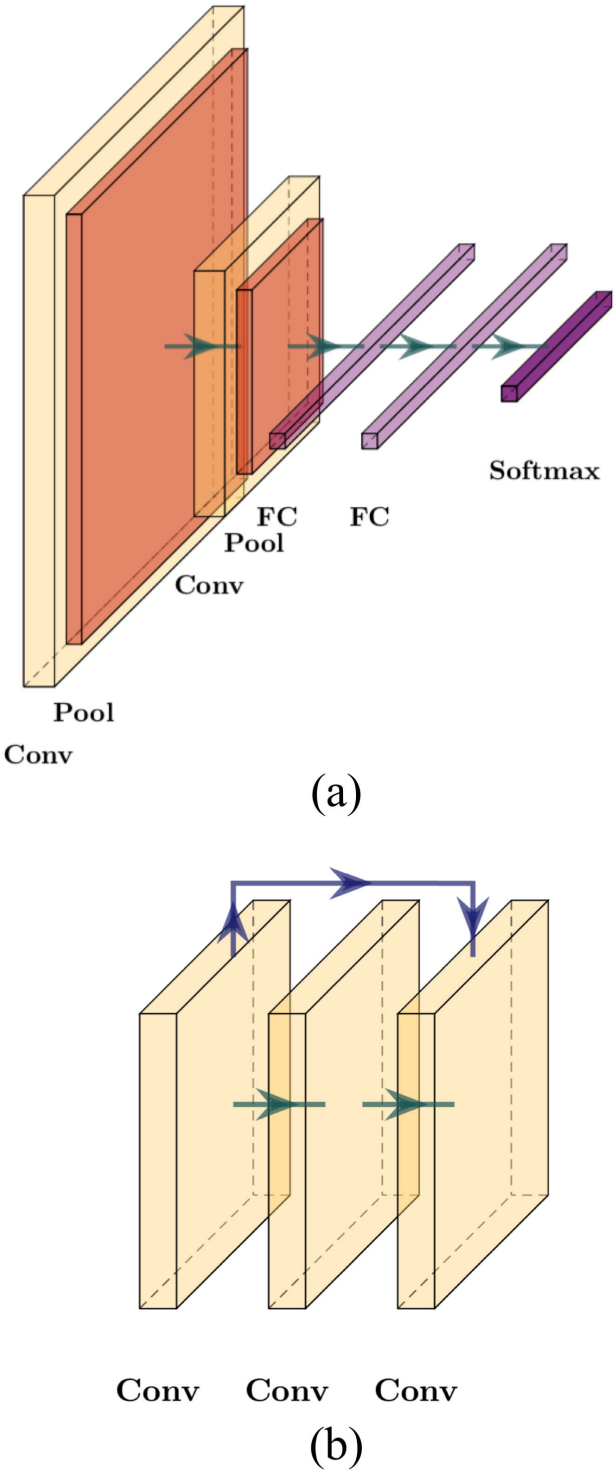

FIGURE 1. CNN structures that have been successfully used in neuroimaging classification (a) is the LeNet structure. (b) is a residual structure in ResNet.

B. DILATED CONVOLUTION IN DL

Some deficits exist in this the combination of standard convolution layers and pooling layers. Pooling layers are used to diminish the feature maps and to help extract features within far-distance regions. However, some parts of information could be lost in the subsampling operations, such as data structure or some spatial information.

The dilated convolution [40] resolved the problem. The dilated convolution is widely used in feature extraction with larger receptive fields, which could facilitate extracting the useful features within extended regions. The difference between standard convolution and dilated convolution is shown in Fig. 2. The dilated convolution is shown in Figure. 2(b) gets a dilated rate l more than the standard convolution. l means the gaps between the closest computing units of a convolution kernel. $l = 1$ means that the dilated convolution equals to standard convolution and $l = 3$ means it skips 3 pixels between closest computing units.

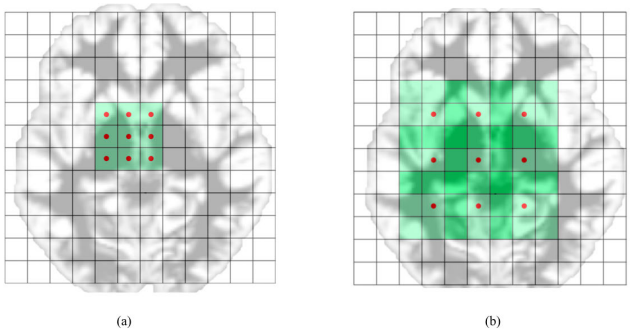

FIGURE 2. The comparison between standard convolution and dilated convolution. (a) Standard convolution. (b) Dilated convolution with $l = 2$.

Dilated convolution has been employed in classification or segmentation of medical MRI images. A 3D fully convolutional networks (FCN) with group dilated convolution was proposed by [41] to segment the prostate on MRI. This algorithm with dilated convolution outperformed other segmentation methods and was demonstrated its clinical feasibility. Huang presented a GlimpseNet with dilated convolution that got better performance than previous methods in classification and segmentation of mammogram MRI images [42].

For now, dilated convolution has not been used in the brain image classification. In this paper, the dilated convolution was embedded into the proposed CNN network structure to extract more features that are sensitive to brain activities or structural differences in large regions.

III. PROPOSED METHOD

Our framework consists of the following parts: data preprocessing, dataset balancing, classification of single images, and Aggregation results of an image sequence. A block diagram of our framework is presented in Fig. 3, and each part is described below.

A. DATA PREPROCESSING

Each MRI image, whether fMRI image or sMRI image, should be preprocessed into the same data coordinates before fed into classifiers. In experiments of a sequence of fMRI images, each subject has multiple EPI images. The preprocessing steps for all of the EPI images include slice timing,

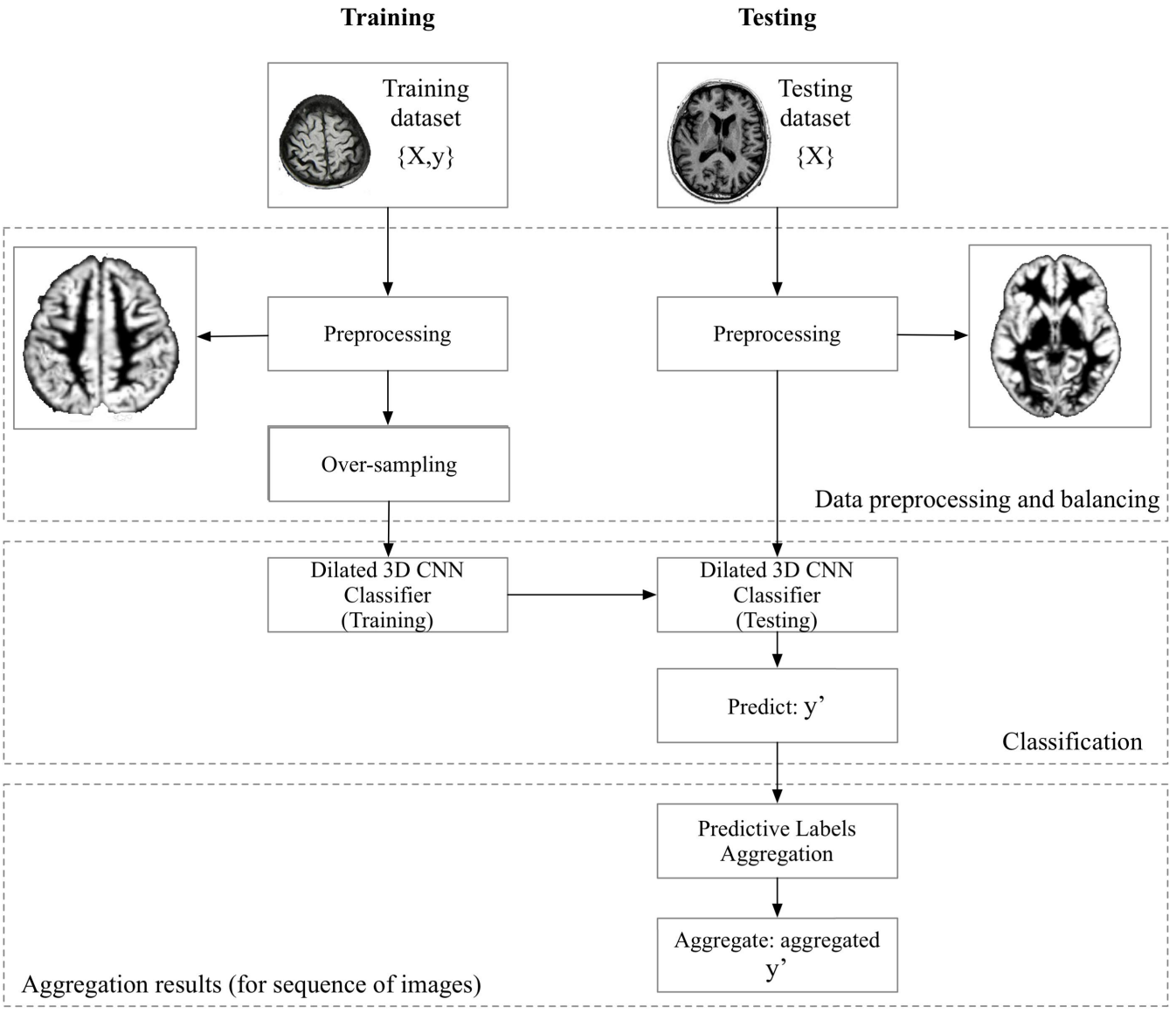

FIGURE 3. Flowchart of our framework based on Dilated 3D CNN. In the first step, images are preprocessed in both training and testing datasets. Random oversampling is employed on the imbalanced training dataset. In the next step, the preprocessed and balanced dataset is presented to dilated 3D CNN for training and testing. For a sequence of images, the final result is aggregated from previous classification results.

realign, normalization, and smooth. After the preprocessing steps, all raw functional EPI images should have been converted into the normalized images. Finally, all images are smoothed using a Gaussian kernel to increase the signal-to-noise ratio. Different from fMRI EPI images, in general conditions, one subject has only one structural MRI image. The structural MRI images should be skull-stripped, segmented and registered into the gray matter (GM), white matter (WM) and Cerebrospinal fluid (CSF) images are normalized in Montreal Neurological Institute (MNI) templates. The GM images are also finally smoothed using Gaussian kernel.

B. DATASET BALANCING

Imbalance of dataset may be one of the common issues found in classification in most of the biomedical application. Data imbalance often meant the classes within a dataset that are unequally distributed, and it may have a negative effect on the performance of a classifier by making the prediction results focus on majority class. Random oversampling or under-sampling methods are effective to counter this problem. We applied the oversampling method to produce a balanced dataset for training.

C. DILATED 3D CNN

Fig. 4 sketches our proposed dilated 3D CNN classifier architecture. All of the convolution and max-pooling layers are transformed from 2D to 3D, whose kernels are converted from a square to a cube. dilated 3D CNN consists of 6 dilated convolution and max-pooling layers. All of the dilated convolution layers have $3 \times 3 \times 3$ convolution kernels with the

TABLE 1. Details of the dilated 3D CNN architecture.

Layer	Feature maps	Stride	Kernel	Dilated Rate	Activation Structure
Dilated Convolution	4	$1 \times 1 \times 1$	$3 \times 3 \times 3$	$3 \times 3 \times 3$	BatchNormalization+LeakyRelu
Dilated Convolution	4	$1 \times 1 \times 1$	$3 \times 3 \times 3$	$3 \times 3 \times 3$	BatchNormalization+LeakyRelu
Max Pooling	4	$2 \times 2 \times 2$	$4 \times 4 \times 4$		
Dilated Convolution	8	$1 \times 1 \times 1$	$3 \times 3 \times 3$	$3 \times 3 \times 3$	BatchNormalization+LeakyRelu
Dilated Convolution	8	$1 \times 1 \times 1$	$3 \times 3 \times 3$	$3 \times 3 \times 3$	BatchNormalization+LeakyRelu
Max Pooling	8	$2 \times 2 \times 2$	$4 \times 4 \times 4$		
Dilated Convolution	16	$1 \times 1 \times 1$	$3 \times 3 \times 3$	$3 \times 3 \times 3$	BatchNormalization+LeakyRelu
Dilated Convolution	16	$1 \times 1 \times 1$	$3 \times 3 \times 3$	$3 \times 3 \times 3$	BatchNormalization+LeakyRelu
Max Pooling	16	$2 \times 2 \times 2$	$4 \times 4 \times 4$		
Fully Connected	128				LeakyRelu
Fully Connected	128				LeakyRelu
Fully Connected	2				Softmax
Classification Layer					Argmax

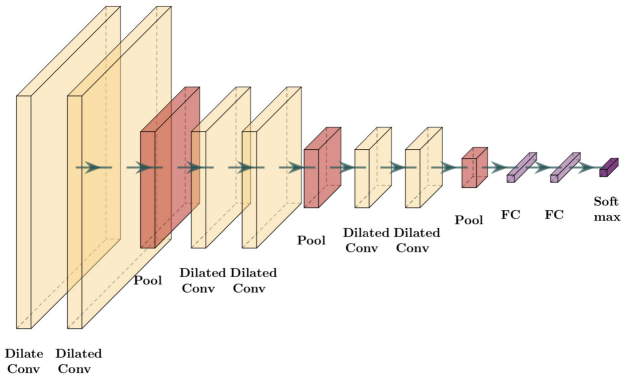

FIGURE 4. The network structure of our proposed dilated 3D CNN classifier.

dilated rate of 3 and stride of $1 \times 1 \times 1$. All of the max-pooling layers also have convolution kernels with a size of $4 \times 4 \times 4$ and stride of $2 \times 2 \times 2$, which are designed to diminish the size of feature maps. Each convolution layer is followed by a common activation module that consists of a batch normalization layer [43] and a Leaky Relu [44] activation function. The last max-pooling layer is followed by two fully connected layers and softmax classifier. Moreover, as we all know, the 3D CNN usually requires a large memory usage in GPU, which makes the classifier cannot work well in most GPUs and too complicated to train properly. We decided to use a small number of feature maps in dilated 3D CNN to resolve these problems. Detailed parameters of dilated 3D CNN are listed in Table 1.

This classifier is designed to be a 2-class classifier. The output of the softmax layer contains two units. One denotes the probability of classified into healthy class (images of healthy subjects), and the other one denotes the probability of classified into patient class (images of patients).

D. AGGREGATION RESULTS

For structural MRI image, which is the only one image matching one subject, the predictive label from dilated 3D CNN is the final predictive result. But a subject has a sequence of fMRI EPI images, which means that the sequence labels from EPI sequences should be aggregated.

We define a true label ratio r that is calculated as the following equation:

$$r = \frac{N_p}{N_p + N_h}. \tag{1}$$

N_p is the number of EPI images that are predicted to patient class. On the opposite, N_h is the number of EPI images that are predicted to healthy subject class.

The aggregated result g is defined as follow:

$$g = \begin{cases} 1 & r >= 0.5 \\ 0 & r < 0.5 \end{cases} \tag{2}$$

IV. EXPERIMENTS AND RESULTS

To evaluated the proposed dilated 3D CNN framework, we started two experiments respectively on sequences of fMRI EPI images and structural MRI images. In the first experiment, the structural images from public ADHD-200 datasets were used to evaluate the Dilated 3D CNN, which predicted sMRI images with two-class classification (typically developing children (TDC) vs. children with ADHD). And in the second experiment, we employed an fMRI dataset which was scanned from schizophrenia (SZ) and healthy subjects. The proposed framework was used to classify schizophrenia and healthy subjects.

In the two experiments, our algorithm and framework is compared with three frequently-used CNN models, ResNet [38], ResNetXt [45], VGG [46], SparseNet [52], AlexNet [23] and Inception-v3 [53]. All the images were pre-processed using SPM8 (https://www.fil.ion.ucl.ac.uk/) and MATLAB (Mathworks, Natic, MA) software. In this work, we employed TensorFlow [48] as the DNN framework to experiment with all methods in two experiments. Experiments were implemented in the software environment: Centos 7.5 64 bit, python 3.6, Tensorflow 1.11.0 and hardware environment: Intel i7-7820X, RAM 64G, and a NVIDIA GeForce GTX 2080Ti GPU.

A. EXPERIMENT 1: SMRI DATASETS OF ADHD-200

In the experiment, 587 sMRI images were obtained from ADHD-200 dataset with a size of $121 \times 145 \times 121$ voxels. We used 5-fold cross-validation to train and test models.

In each test, 470 images were used for training including 353 images of healthy subjects and 117 images of patients, and the remaining 117 images were used for testing including 88 images of healthy subjects and 29 images of patients. After over-sampling in training phase, the number of images of patients in training was 353.

With the random initialization of the network parameters in predefined ranges and random partition of dataset, we repeated training and testing for 10 times and selected the parameters in which the network performed in an average level. The initial weights are random, and the learning rate is 10^{-5}. Models used RMSprop [47] as optimizer and categorical cross-entropy as the loss function. The number of the batch is 32 in the training process.

B. RESULT 1

According to the average testing accuracy in the training process shown in Fig. 5, all of the four methods could converge at a stable level after iteration 60. The accuracy results are shown in Table 2. According to the accuracy comparison, it's intuitionistic that dilated 3D CNN is the most accurate model with the ResNetXt follows. The accuracy of the proposed method was 0.766, which outperformed others and ResNetXt performed most closely to it. But the imbalanced data might make the predictive labels more likely to be the majority class. The standard deviation of accuracy of SparseNet is the highest, which meant that SparseNet could be not stable in this experiment.

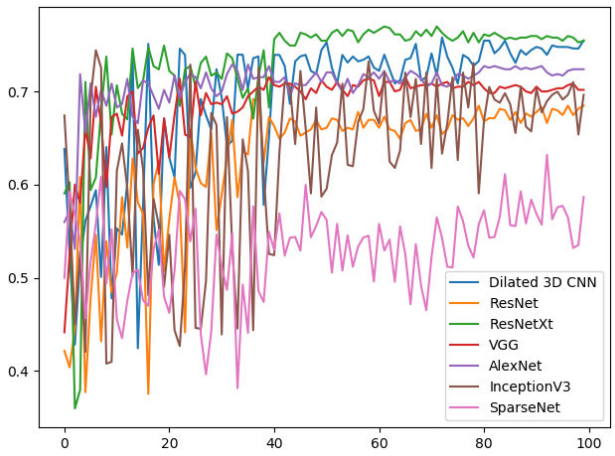

FIGURE 5. Average testing accuracy in the training processes of ADHD classification methods.

TABLE 2. Accuracy of cross-validation test of all models in experiment 1.

Model	CV1	CV2	CV3	CV4	CV5	Avg.	Std.
dilated 3D CNN	0.786	0.829	0.726	0.761	0.731	0.766	0.042
ResNet	0.684	0.769	0.692	0.624	0.655	0.685	0.054
ResNetXt	0.735	0.752	0.752	0.769	0.765	0.755	0.013
VGG	0.692	0.769	0.641	0.709	0.697	0.702	0.046
SparseNet	0.391	0.645	0.609	0.640	0.608	0.579	0.106
AlexNet	0.709	0.744	0.701	0.744	0.723	0.724	0.02
Inception-v3	0.658	0.675	0.701	0.684	0.765	0.697	0.041

Average sensitivity, specificity, areas under the curve (AUC) and F1 scores are shown in Table 3. Sensitivity and specificity are statistical measures of the performance of a binary classification test. Sensitivity is calculated according to the following equation:

$$Sensitivity = \frac{TruePositive}{TruePositive + FalseNegative}. \quad (3)$$

TABLE 3. Average sensitivity, specificity, AUC and F1 scores of all models in experiment 1.

Model	Sensitivity	Specificity	AUC	F1score
dilated 3D CNN	0.390	0.891	0.671	0.437
ResNet	0.377	0.789	0.638	0.373
ResNetXt	0.075	0.982	0.478	0.132
VGG	0.164	0.882	0.568	0.312
SparseNet	0.449	0.651	0.612	0.436
AlexNet	0.219	0.891	0.592	0.283
Inception-v3	0.301	0.828	0.587	0.331

Sensitivity means the probability of a patient being classified into patient class. Specificity is calculated according to the following equation:

$$Specificity = \frac{TrueNegative}{TrueNegative + FalsePositive}. \quad (4)$$

Specificity means the probability of a healthy subject being classified into the healthy subject class.

It's obvious that although ResNetXt got a high accuracy, the sensitivity and F1 score was extremely low comparing to other models. dilated 3D CNN got both high AUC and F1 score comparing to remaining models. F1 score of SparseNet is 0.436, close to F1 score of dilated 3D CNN. Based on the AUC shown in Table 3 and receiver operating characteristic (ROC) curves plotting in Fig. 6, dilated 3D CNN outperformed other methods in the most validation test.

In order to evaluate the dilated 3D CNN method, we compared our results with the results in previous results on sMRI images of ADHD-200 dataset. Authors of these research performed cross-validation in training and testing ADHD-200 dataset. The comparison is presented in Table 4. The accuracy achieved by our proposed methodology is better than the others.

TABLE 4. Comparison between our results and results of previous research.

Model	Accuracy	Data Combination
dilated 3D CNN	0.766	sMRI
Zou et al. 2017 [31]	0.6586	sMRI
Zou et al. 2017 [31]	0.6915	sMRI+fMRI
Sen et al. 2018 [49]	0.689	sMRI+fMRI
Sina et al. 2016 [50]	0.70	sMRI+fMRI

It is also a question that how it could be if employ a hybrid CNN structure with standard and dilated convolution layers. Whether hybrid CNN could improve the performance? We lunched a comparison between Dilated 3D CNN, hybrid CNN

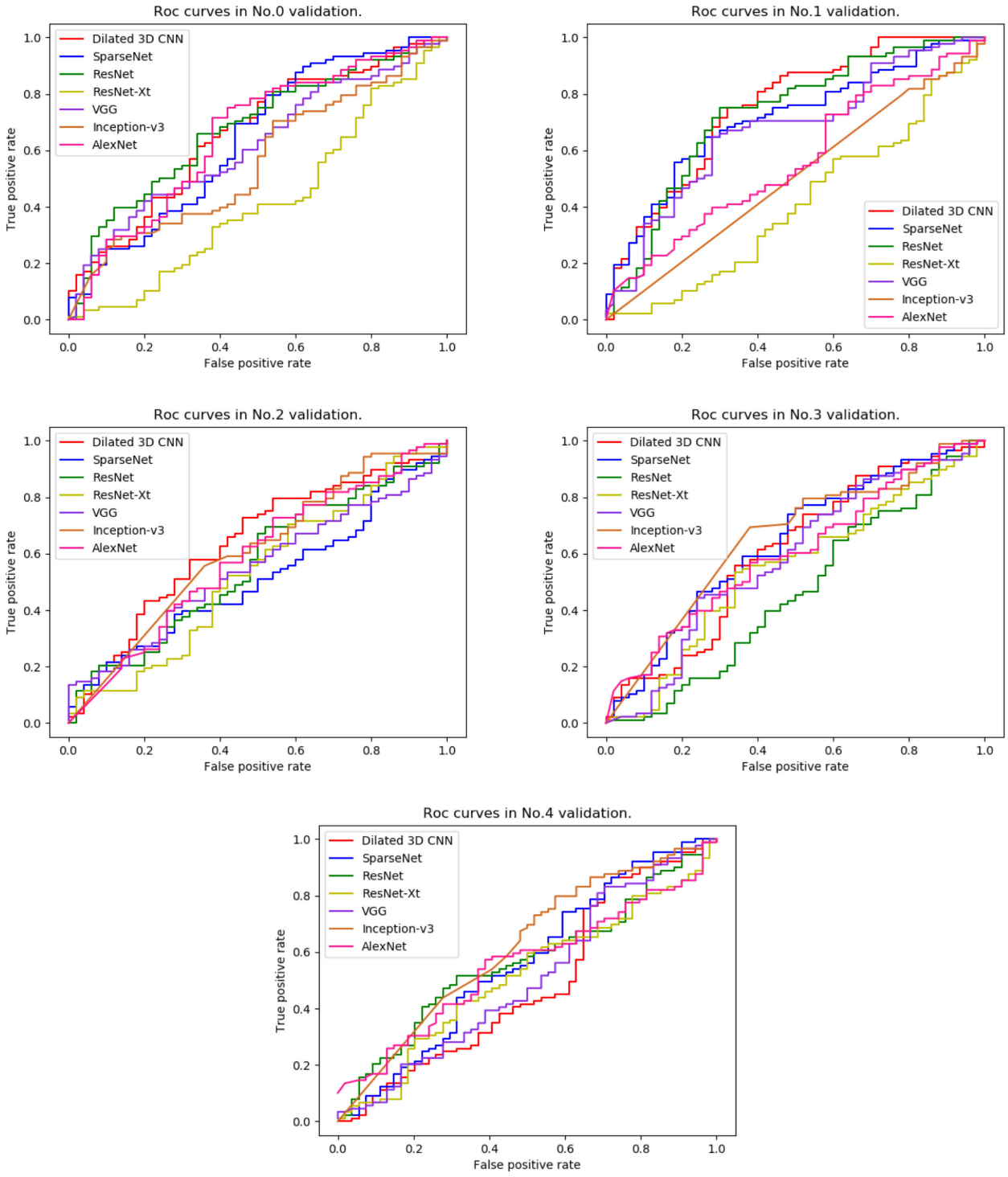

FIGURE 6. Seven methods' ROC curves in each cross-validation test in experiment 1.

with 2 standard convolution layers and 4 dilated convolution layers (hybrid CNN 2/4) and hybrid CNN with 4 standard convolution layers and 4 dilated convolution layers (hybrid CNN 4/2). Performance metrics are shown in Table 5.

Dilated 3D CNN got the highest average accuracy, specificity, AUC score and F1 score. Hybrid CNN 2/4 follows it. "N/A" which means an incalculable value was filled in F1 score of hybrid CNN 4/2, because this model has predicted all labels as 0 or 1 in each validation. It was obvious that hybrid CNN with four dilated convolution layers could not reach a balance between positive and negative classes. dilated 3D CNN outperformed both hybrid CNN models.

TABLE 5. Comparison of dilated 3D CNN and hybrid CNNs.

Model	Avg. Acc.	Sensitivity	Specificity	AUC	F1 score
Dilated 3D CNN	0.766	0.390	0.891	0.671	0.437
hybrid CNN 2/4	0.697	0.322	0.821	0.647	0.347
hybrid CNN 4/2	0.450	0.400	0.600	0.5	N/A

These results may be caused by the characteristics of ADHD. According to findings in [51], cortical thickness in bilateral frontal regions and the right cingulate cortex were found thinner for ADHD patients. And They had significantly decreased structural and functional connectivity in many different brain regions, including cingulate cortex, lateral prefrontal cortex, left precuneus and thalamus. But the functional connectivity was increased in bilateral posterior medial frontal cortex. Some of these regions of interest (ROIs) are faraway from each other and the changes in connectivity could be easier for dilated convolution to found. Although standard convolution could extract more features from near brain regions, these features perhaps didn't help recognize patterns.

C. EXPERIMENT 2: FMRI DATASETS OF SCHIZOPHRENIA

In the second experiment, we used clinical data scanned in local hospital under the permission of Ethics Committee. fMRI EPI images were scanned using a 3-T GE MRI scanner. The repetition time was 2000ms, and echo time was 30ms. Each image consisted of 50 slices. 28 healthy and 28 SZ patients participated in this experiment (age range: 15-44, healthy subjects: 17 females and 11 males, patients: 14 females and 14 males). 50 EPI images of each subject were scanned and used in classification. There were totally 2800 EPI images in our dataset for training and testing. During scanning, each subject was told to lie in the MRI scanner and keep in an awake resting state with head fixed by sponges.

In the experiment, fMRI EPI images were with a size of $61 \times 73 \times 61$ voxels. We used 5-fold cross-validation to train and test models. Dataset was split at subject level. In each test, 2300 images of 23 patients and 23 healthy subjects were used for training, and the 500 images of remaining subjects were used for testing. Healthy and patient class has the same number in training and testing phase.

With the random initialization of the network parameters in predefined ranges and random partition of dataset, we repeated training and testing for 10 times and selected the parameters in which the network performed in an average level. The initial weights are random, and the learning rate is 10^{-5}. Models used RMSprop [47] as optimizer and categorical cross-entropy as the loss function. The number of the batch is 32 in the training process.

Because a sequence of fMRI EPI images existed in this experiment, we would compare the results of classification as well as the aggregated results of our proposed framework.

D. RESULT 2

The accuracy results are shown in Table 6. The accuracy of the proposed dilated 3D CNN method was 0.822, which was higher than the performance of other methods. And like the accuracy results in experiment 1, the ResNetXt got the accuracy of 0.806, which was closest to dilated 3D CNN. Because of the difference of datasets, sensitivity, and specificity of dilated 3D CNN shown in Table 7 is respectively 0.793 and 0.855, which is higher than those in experiment 1. And the AUC and F1 score of dilated 3D CNN still got highest in among these methods. These results mean that our proposed method could recognize both SZ patients and healthy subjects with high accuracy, based on a single fMRI EPI image.

TABLE 6. Accuracy of cross-validation test of all models in experiment 2.

Model	CV1	CV2	CV3	CV4	CV5	Avg.	Std.
dilated 3D CNN	0.837	0.910	0.819	0.711	0.833	0.822	0.071
ResNet	0.335	0.559	0.438	0.820	0.551	0.608	0.181
ResNetXt	0.822	0.585	0.715	0.569	0.785	0.806	0.115
VGG	0.796	0.662	0.862	0.445	0.642	0.686	0.161
SparseNet	0.898	0.718	0.756	0.814	0.792	0.796	0.068
AlexNet	0.596	0.683	0.647	0.822	0.794	0.710	0.097
Inception-v3	0.650	0.635	0.597	0.597	0.419	0.579	0.093

TABLE 7. Average sensitivity, specificity, AUC and F1 scores of all models in experiment 2.

Model	Sensitivity	Specificity	AUC	F1 score
dilated 3D CNN	0.793	0.855	0.827	0.818
ResNet	0.454	0.618	0.665	0.495
ResNetXt	0.725	0.672	0.567	0.706
VGG	0.648	0.723	0.787	0.699
SparseNet	0.772	0.818	0.812	0.790
AlexNet	0.751	0.661	0.756	0.719
Inception-v3	0.506	0.651	0.729	0.546

The AUC scores and ROC curves are shown in Table 6 and Fig. 7. Although the ROC curves in experiment 2 are more disordered than those in experiment 1, the average AUC score showed dilated 3D CNN's domination among seven methods.

The previous results substantiated the accuracy and robustness of dilated 3D CNN on a single EPI image level. In the last part of our proposed framework, prediction labels on the single EPI image level were aggregated into prediction labels on subject level. The accuracy, sensitivity, specificity and F1 score results on a subject level are shown in Table 8. Measures of all models decreased slightly comparing to

TABLE 8. Average sensitivity, specificity and accuracy results aggregated on subject level.

Model	Sensitivity	Specificity	Accuracy	F1 score
dilated 3D CNN	0.774	0.862	0.818	0.808
ResNet	0.435	0.589	0.512	0.465
ResNetXt	0.632	0.597	0.639	0.688
VGG	0.644	0.742	0.692	0.677
SparsNet	0.768	0.835	0.801	0.784
AlexNet	0.727	0.662	0.695	0.694
Inception-v3	0.530	0.634	0.582	0.526

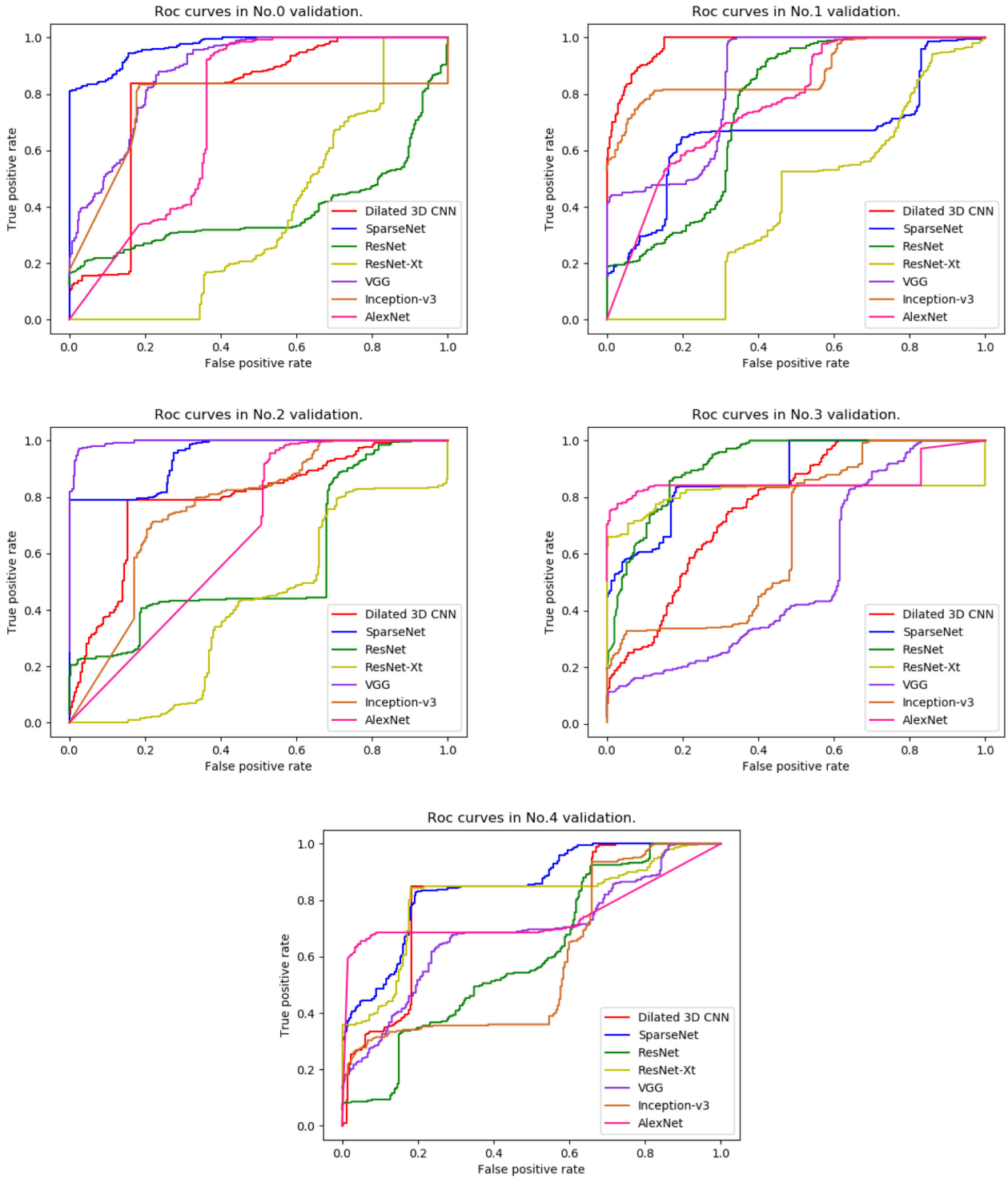

FIGURE 7. Seven methods' ROC curves in each cross-validation test in experiment 2.

results on the single EPI image level. But all of the measures of proposed dilated 3D CNN were still the highest, which means that our method and framework could classify SZ and healthy subjects using fMRI images at high accuracy of 81.8%, sensitivity of 0.774, specificity of 0.862 and F1 score of 0.805.

V. CONCLUSION

In this work, we introduced the dilated 3D CNN method for 3D brain MRI image classification. In this method, dilated convolution was used for recognizing patterns across long-distance brain areas in both structural and functional MRI images. To reduce the memory usage of 3D CNN method, we cut down the number of feature maps in convolution layers. Moreover, a methodology framework based on dilated 3D CNN method was also presented, including data pre-processing, data balancing, single images classifying, and prediction labels aggregating. This framework could classify both single MRI images and a sequence of EPI images. In the task of ADHD sMRI image classification, the proposed framework based on dilated 3D CNN method outperformed other methods. It achieved accuracy of 76.6%, which is higher than the results of previous related research. In the other task of SZ fMRI EPI image classification, the proposed method achieved accuracy of 82.7% at single EPI image level and 81.8% at subject level. These results confirmed that the dilated 3D CNN method and framework is effective and robust in some different kinds of MRI images. This method could also be used in Computer-Assisted Diagnosis and real-time fMRI analysis. It is worth exploring in future research how increasing/decreasing dilation rate affects performance metrics.

REFERENCES

[1] P. M. A. Sloot, A. Tirado-Ramos, I. Altintas, M. Bubak, and C. Boucher, "From molecule to man: Decision support in individualized e-health," *Computer*, vol. 39, no. 11, pp. 40–46, Nov. 2006.

[2] L. S. Ronga, S. Jayousi, E. Del Re, L. Colitta, G. Iannone, A. Scorpiniti, C. Aragno, and C. P. Neja, "TESHEALTH: An integrated satellite/terrestrial system for e-health services," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Ottawa, ON, Canada, Jun. 2012, pp. 2890–3286.

[3] J. Liang and T. Sahama, "Online multiple profile manager for eHealth information sharing," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Ottawa, ON, Canada, Jun. 2012, pp. 3461–3465.

[4] M. D. Fox, A. Z. Snyder, J. L. Vincent, M. Corbetta, D. C. Van Essen, and M. E. Raichle, "The human brain is intrinsically organized into dynamic, anticorrelated functional networks," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 27, pp. 9673–9678, 2005.

[5] B. B. Biswal, M. Mennes, X. N. Zuo, S. Gohel, C. Kelly, S. M. Smith, C. F. Beckmann, J. S. Adelstein, R. L. Buckner, S. Colcombe, and A. M. Dogonowski, "Toward discovery science of human brain function," *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 10, pp. 4734–4739, 2010.

[6] P. C. Mulders, P. F. van Eijndhoven, A. H. Schene, C. F. Beckmann, and I. Tendolkar, "Resting-state functional connectivity in major depressive disorder: A review," *Neurosci. Biobehavioral Rev.*, vol. 56, pp. 330–344, Sep. 2015.

[7] M. R. Arbabshirani, S. Plis, J. Sui, and V. D. Calhoun, "Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls," *NeuroImage*, vol. 145, pp. 137–165, Mar. 2016.

[8] T. Wolfers, J. K. Buitelaar, C. F. Beckmann, B. Franke, and A. F. Marquand, "From estimating activation locality to predicting disorder: A review of pattern recognition for neuroimaging-based psychiatric diagnostics," *Neurosci. Biobehavioral Rev.*, vol. 57, pp. 328–349, Oct. 2015.

[9] E. Z. Li, "The application of BOLD-fMRI in cognitive neuroscience," *J. Frontiers Comput. Sci. Technol.*, vol. 2, no. 6, pp. 589–600, 2008.

[10] F. Pereira, T. Mitchell, and M. Botvinick, "Machine learning classifiers and fMRI: A tutorial overview," *NeuroImage*, vol. 45, no. 1, pp. S199–S209, Mar. 2009.

[11] S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller, "Introduction to machine learning for brain imaging," *NeuroImage*, vol. 56, no. 2, pp. 387–399, 2011.

[12] S. Klöppel, C. M. Stonnington, J. Barnes, F. Chen, C. Chu, C. D. Good, I. Mader, L. A. Mitchell, A. C. Patel, and N. C. Fox, "Accuracy of dementia diagnosis—A direct comparison between radiologists and a computerized method," *Brain*, vol. 131, no. 11, pp. 2969–2974, Oct. 2008.

[13] C. Pinardi, O. Ortenzia, S. Gardini, R. Aldigeri, M. Micheli, V. Spigoni, A. De Cais, and C. Ghetti, "Automated classification of resting state fMRI networks using machine learning algorithms," *Phys. Medica*, vol. 56, pp. 219–220, Dec. 2018.

[14] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, Jun. 1999.

[15] K. Franke, G. Ziegler, S. Klöppel, and C. Gaser, "Estimating the age of healthy subjects from T_1-weighted MRI scans using kernel methods: Exploring the influence of various parameters," *NeuroImage*, vol. 50, no. 3, pp. 883–892, 2010.

[16] A. Ortiz, J. M. Górriz, J. Ramírez, and F. J. Martínez-Murcia, "LVQ-SVM based CAD tool applied to structural MRI for the diagnosis of the Alzheimer's disease," *Pattern Recognit. Lett.*, vol. 34, no. 14, 2013, pp. 1725–1733.

[17] K. Machhale, H. B. Nandpuru, V. Kapur, and L. Kosta, "MRI brain cancer classification using hybrid classifier (SVM-KNN)," in *Proc. Int. Conf. Ind. Instrum. Control (ICIC)*, May 2015, pp. 60–65.

[18] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Neural Networks for Perception*. New York, NY, USA: Academic, 1992, pp. 65–93.

[19] K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby, "Beyond mind-reading: Multi-voxel pattern analysis of fMRI data," *Trends Cogn. Sci.*, vol. 10, no. 9, pp. 424–430, Sep. 2006.

[20] C.-A. Chou, K. Kampa, S. H. Mehta, R. F. Tungaraza, W. A. Chaovalitwongse, and T. J. Grabowski, "Voxel selection framework in multi-voxel pattern analysis of fMRI data for prediction of neural response to visual stimuli," *IEEE Trans. Med. Imag.*, vol. 33, no. 4, pp. 925–934, Apr. 2014.

[21] A. Khazaee, A. Ebrahimzadeh, and 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 Imag. Behav.*, vol. 10, no. 3, pp. 799–817, 2016.

[22] A. Al-Zubaidi, A. Mertins, M. Heldmann, K. Jauch-Chara, and T. F. Münte, "Machine learning based classification of resting-state fMRI features exemplified by metabolic state (hunger/satiety)," *Frontiers Hum. Neurosci.*, vol. 13, May 2019, Art. no. 164. doi: 10.3389/fnhum.2019.00164.

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

[24] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649.

[25] M. N. I. Qureshi, J. Oh, and B. Lee, "3D-CNN based discrimination of schizophrenia using resting-state fMRI," *Artif. Intell. Med.*, vol. 98, pp. 10–17, Jul. 2019.

[26] S. Vieira, W. H. L. Pinaya, and A. Mechelli, "Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications," *Neurosci. Biobehavioral Rev.*, vol. 74, Mar. 2017, pp. 58–75.

[27] S. Sarraf and G. Tofighi, "Classification of Alzheimer's disease using fMRI data and deep learning convolutional neural networks," Mar. 2016, *arXiv:1603.08631*. [Online]. Available: https://arxiv.org/abs/1603.08631

[28] S. Sarraf, J. Anderson, and G. Tofighi, "DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI," *BioRxiv*, vol. 1, Jan. 2017, Art. no. 070441.

[29] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[30] Y. Zhao, Q. Dong, S. Zhang, W. Zhang, H. Chen, X. Jiang, L. Guo, X. Hu, J. Han, and T. Liu, "Automatic recognition of fMRI-derived functional networks using 3-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1975–1984, Sep. 2018.

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

[32] X. Xu, Q. Wu, S. Wang, J. Liu, J. Sun, and A. Cichocki, "Whole brain fMRI pattern analysis based on tensor neural network," *IEEE Access*, vol. 6, pp. 29297–29305, 2018.

[33] N. Kasabov, L. Zhou, M. G. Doborjeh, Z. G. Doborjeh, and J. Yang, "New algorithms for encoding, learning and classification of fMRI data in a spiking neural network architecture: A case on modeling and understanding of dynamic cognitive processes," *IEEE Trans. Cogn. Develop. Syst.*, vol. 9, no. 4, pp. 293–303, Dec. 2017.
[34] Z. Yu-Feng, H. Yong, Z. Chao-Zhe, C. Qing-Jiu, S. Man-Qiu, L. Meng, T. Li-Xia, J. Tian-Zi, and W. Yu-Feng, "Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI," *Brain Develop.*, vol. 29, no. 2, pp. 83–91, 2007.
[35] M.-E. Lynall, D. S. Bassett, R. Kerwin, P. J. McKenna, M. Kitzbichler, U. Müller, and E. Bullmore, "Functional connectivity and brain networks in schizophrenia," *J. Neurosci.*, vol. 30, no. 28, pp. 9477–9487, Jul. 2010.
[36] Y. LeCun. *LeNet-5, Convolutional Neural Networks*. Accessed: May 20, 2015. [Online]. Available: http://yann.lecun.com/exdb/lenet
[37] X. Yang, S. Sarraf, and N. Zhang, "Deep learning-based framework for autism functional MRI image classification," *J. Arkansas Acad. Sci.*, vol. 72, no. 1, pp. 47–52, 2018.
[38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[39] X. Chen, Y. Chen, C. Ma, X. Liu, and X. Tang, "Classification of pancreatic tumors based on MRI images using 3D convolutional neural networks," in *Proc. 2nd Int. Symp. Image Comput. Digit. Med.*, Oct. 2018, pp. 92–96.
[40] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*. [Online]. Available: https://arxiv.org/abs/1511.07122
[41] B. Wang, Y. Lei, S. Tian, T. Wang, Y. Liu, P. Patel, A. B. Jani, H. Mao, W. J. Curran, T. Liu, and X. Yang, "Deeply supervised 3D fully convolutional networks with group dilated convolution for automatic MRI prostate segmentation," *Med. Phys.*, vol. 46, no. 4, pp. 1707–1718, Apr. 2019.
[42] W. Hang, Z. Liu, and A. Hannun, "GlimpseNet: Attentional methods for full-image mammogram diagnosis," Stanford AI Lab, Stanford Univ., Stanford, CA, USA, Internal Rep., 2017. [Online]. Available: http://cs231n.stanford.edu/reports/2017/pdfs/517.pdf
[43] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: https://arxiv.org/abs/1502.03167
[44] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, Jun. 2013, vol. 30, no. 1, p. 3.
[45] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1492–1500.
[46] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556
[47] T. Tieleman and G. Hinton, "Lecture 6.5-RMSPROP: Divide the gradient by a running average of its recent magnitude," *COURSERA, Neural Netw. Mach. Learn.*, vol. 4, no. 2, pp. 26–31, 2012.
[48] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Operating Syst. Design Implement. (OSDI)*, 2016, pp. 265–283.
[49] B. Sen, N. C. Borle, R. Greiner, and M. R. G. Brown, "A general prediction model for the detection of ADHD and Autism using structural and functional MRI," *PLoS ONE*, vol. 13, no. 4, 2018, Art. no. e0194856.
[50] S. Ghiassian, R. Greiner, P. Jin, and M. R. G. Brown, "Using functional or structural magnetic resonance images and personal characteristic data to identify ADHD and autism," *PLoS ONE*, vol. 11, no. 12, 2016, Art. no. e0166934.
[51] M.-G. Qiu, Z. Ye, Q.-Y. Li, G.-J. Liu, B. Xie, and J. Wang, "Changes of brain structure and function in ADHD children," *Brain Topography*, vol. 24, nos. 3–4, pp. 243–252, Oct. 2011.
[52] W. Liu and K. Zeng, "SparseNet: A sparse densenet for image classification," 2018, *arXiv:1804.05340*. [Online]. Available: https://arxiv.org/abs/1804.05340
[53] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2818–2826.

ZIJIAN WANG received the B.Sc. degree in computer science from Tongji University, Shanghai, China, where he is currently pursuing the Ph.D. degree. He has authored or coauthored about ten journals and conference papers. His research interests include machine learning, medical image analysis, and cognitive neuroscience.

YAORU SUN received the Ph.D. degree in artificial intelligence from the University of Edinburgh. He is currently a Full Professor with the Department of Computer Science and Technology, Tongji University, China. His research interests include brain-like computation, machine intelligence, and cognitive neuroscience.

QIANZI SHEN received the B.Sc. degree in computer science from Shandong University, Shandong, China. She is currently pursuing the M.Sc. degree with Tongji University. Her research interests include machine learning, artificial intelligence, and data mining.

LEI CAO received the Ph.D. degree in computer software and theory from Tongji University, China, in 2016. He was a Lecturer with Shanghai Maritime University, China. He was a Visiting Researcher with the Institute of Medical Psychology and Behavioral Neurobiology, University of Tübingen, in 2013. N. Birbaumer was his mentor for academic advising. He has authored or coauthored about 20 journals and conference papers. His research interests include EEG signal processing, brain–computer interface, machine learning, and medical image analysis.