Received March 22, 2019, accepted April 19, 2019, date of publication April 30, 2019, date of current version May 28, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2913847

Deep Learning Framework for Alzheimer's Disease Diagnosis via 3D-CNN and FSBi-LSTM

CHIYU FENG[1], AHMED ELAZAB[1,2], PENG YANG[1], TIANFU WANG[1], FENG ZHOU[3], HUOYOU HU[4], XIAOHUA XIAO[4], AND BAIYING LEI[1], (Senior Member, IEEE)

[1]National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518060, China
[2]Computer Science Department, Misr Higher Institute of Commerce and Computers, Mansoura 35516, Egypt
[3]Industrial and Manufacturing Systems Engineering, University of Michigan–Dearborn, Dearborn, MI 48128, USA
[4]Affiliated Hospital of Shenzhen University, Health Science Center, Shenzhen University, Shenzhen Second People's Hospital, Shenzhen 530031, China

Corresponding authors: Xiaohua Xiao (tu_xi8888@163.com) and Baiying Lei (leiby@szu.edu.cn)

This work was supported partly by National Natural Science Foundation of China (Nos. 61871274, 61801305 and 81571758), National Natural Science Foundation of Guangdong Province (Nos. 2017A030313377 and 2016A030313047), Shenzhen Peacock Plan (No. KQTD2016053112051497 and KQTD2015033016104926), and Shenzhen Key Basic Research Project (Nos. JCYJ20170818142347251 and JCYJ20170818094109846) and Open Fund Project of Fujian Provincial Key Laboratory of Information Processing and Intelligent Control (Minjiang University) (No. MJUKF201711).

ABSTRACT Alzheimer's disease (AD) is an irreversible progressive neurodegenerative disorder. Mild cognitive impairment (MCI) is the prodromal state of AD, which is further classified into a progressive state (i.e., pMCI) and a stable state (i.e., sMCI). With the development of deep learning, the convolutional neural networks (CNNs) have made great progress in image recognition using magnetic resonance imaging (MRI) and positron emission tomography (PET) for AD diagnosis. However, due to the limited availability of these imaging data, it is still challenging to effectively use CNNs for AD diagnosis. Toward this end, we design a novel deep learning framework. Specifically, the virtues of 3D-CNN and fully stacked bidirectional long short-term memory (FSBi-LSTM) are exploited in our framework. First, we design a 3D-CNN architecture to derive deep feature representation from both MRI and PET. Then, we apply FSBi-LSTM on the hidden spatial information from deep feature maps to further improve its performance. Finally, we validate our method on the AD neuroimaging initiative (ADNI) dataset. Our method achieves average accuracies of 94.82%, 86.36%, and 65.35% for differentiating AD from normal control (NC), pMCI from NC, and sMCI from NC, respectively, and outperforms the related algorithms in the literature.

INDEX TERMS Alzheimer's disease, 3D-CNN, FSBi-LSTM, multi-modal fusion.

I. INTRODUCTION

Alzheimer's disease (AD) is an irreversible and progressive neurodegenerative disorder, which mainly occurs in the population of 65 and older. Mild cognitive impairment (MCI) is the prodromal state of AD and can be further categorized into progressive MCI (pMCI) and stable MCI (sMCI) [1]. Alzheimer's Disease International released that 50 million people worldwide were suffering from dementia in 2018 and the number will increase to 152 million by 2050 [2]. The total estimated worldwide expenses of AD in 2018 are 1 trillion dollars and will be doubled by 2030 [2]. To date, AD is incurable. However, we can prevent patients from deterioration effectively by early detection and diagnosis of AD. Medical imaging techniques, including magnetic resonance imaging (MRI) and positron emission tomography (PET), provide rich and complementary imaging information for diagnosis [3]–[9]. Early diagnosis of AD mainly depends on the doctor's experience, and such human visual inspection is often too subjective. Thus, computer aided diagnosis in evaluating the early stages of AD is highly desirable.

Many studies in literature focused on developing automatic algorithms for discovering the changes of functional and anatomical neural structures that are related to AD by using

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan.

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

traditional machine learning techniques [3]–[18]. Generally, traditional machine learning methods exploit two types of features in early diagnosis of AD, including region of interest (ROI) based features [4], [12]–[14], [16], [18] and voxel based features [7], [8], [10]. More specifically, the former relies heavily on specific assumptions about structural or functional abnormalities in the brain, such as regional cortical thickness [19], hippocampal volume [20], and gray matter volume [21]. However, these feature extraction methods are limited since they require complex preprocessing steps and advanced clinical domain knowledge. In addition, the brain is a huge interconnected network, and ROIs cannot sufficiently express these connections. Moreover, extraction of ROIs yields a large amount of information loss due to compression. The latter focuses on acquiring features by measuring tissue density, such as gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) without relying on any hypothesis on the brain structure. However, image volumes come with a huge number of voxels (in millions), while the number of samples is limited. Hence, the overfitting issue can occur. Traditional machine learning methods rely on manual feature extraction, which depends heavily on professional knowledge, repeated attempts, and tends to be time-consuming and subjective.

To solve these problems and to further improve performance, convolutional neural networks (CNNs) are an effective solution and have showed great success in AD diagnosis [22]–[30]. However, these studies were unable to build a deep 3D-CNN network, such as VGG [31], since it is difficult to obtain a huge amount of labeled clinical data. Hence, shallow 3D-CNN networks are preferred. However, in traditional CNN models, the fully connected (FC) layer can only input 1D data. The feature maps of 3D-CNN parts are always in 3D, thus the 3D spatial information in the feature maps will be lost in the flattening operation. As a result, many efforts have been devoted to solving the shortcomings of CNNs in this aspect [23], [24], [29] by using other layer to replace the fully-connected (FC) layer.

Recurrent neural network (RNN) is a powerful model in sequence analysis. Since RNN adopts a ''state'' vector in its hidden units, it implicitly contains information of the sequence's all history information [32]. Recent studies have shown that, RNN and related structure networks not only can analyze sequence information, but also can achieve good results in structure analysis [24], [33]. Compared with RNN, its improved version long short-term memory (LSTM) can effectively solve problems of gradient explosion or gradient disappearance by using several gates to control information flow [34]. Furthermore, bidirectional LSTM (Bi-LSTM) can have contextual information in both directions [35]. In fact, Bi-LSTM can get more information without choosing the scanning direction, which can be enhanced by stacking the LSTM to explore the spatial information of feature maps from 3D-CNN as well.

Therefore, we propose to use the stacked Bi-LSTM (SBi-LSTM) instead of the traditional Bi-LSTM [36] in this paper. In addition, each output of LSTM is related to historical input, but the current input has a greater impact. Hence, the FC layer can boost accuracy by enhancing the connection between different output nodes of SBi-LSTM. Motivated by this observation, we design a novel deep learning network that uses multimodal data for AD diagnosis via 3D-CNN and fully stacked bidirectional LSTM (FSBi-LSTM). Specifically, the image of each MRI or PET is transferred to the 3D-CNN network to extract features from a more macroscopic perspective. In addition, FSBi-LSTM is used to extract high-level semantic and spatial information instead of the traditional FC layer. By inputting one pixel of all features to the corresponding position at each step, FSBi-LSTM can preserve the spatial information of feature maps corresponding to different parts of the data. Because the output of LSTM is closely related to the neighboring input, we add a FC layer for feature extraction after the output of SBi-LSTM. Accordingly, the output of each step is closely related to other steps. Finally, the features from MRI and PET are fused and fed into the SoftMax classifier for disease diagnosis.

The rest of this paper is organized as follows. We briefly recall the relevant researches in Section II. Then, we give the detailed description of our method in Section III. Section IV describes the experimental results. The advantages and limitations of the proposed method are discussed in Section V. Finally, we summarize conclusions in Section VI.

II. RELTEAD WORK

A. TRADITIONAL MACHINE LEARNING BASED METHODS

Many studies in literature focused on developing automatic algorithms to observe the functional and anatomical neural lesions related to AD by traditional machine learning methods [3]–[18]. For example, Gray *et al.* introduced a multimodal classification method by using similarity measure generated from random forest classifier [10]. Zhang *et al.* proposed a multi-layer classifier, where the first layer is multi-view input and explores the complex correlation between the feature and the label by building a latent representation [15]. A discriminative sparse learning method was recommended by Lei *et al.* to predict the clinical score jointly with relational regularization and use multimodal features to classify AD stages [16]. However, these methods are computationally intensive and mainly depend on the handcrafted features, which are not appealing and difficult to obtain.

B. DEEP LEARNING BASED METHODS

To address the traditional machine learning problems, some studies have used the deep learning for AD diagnosis, recently [22]–[30]. For instance, Islam et al. presented an AD diagnosis method based on 2D DenseNet and sliced the MRI data in three directions. Three parallel 2D DenseNets were then used to analyze and fuse the final diagnosis results [28]. Liu *et al.* proposed a deep multitask multichannel learning configuration for clinical score regression and brain disease

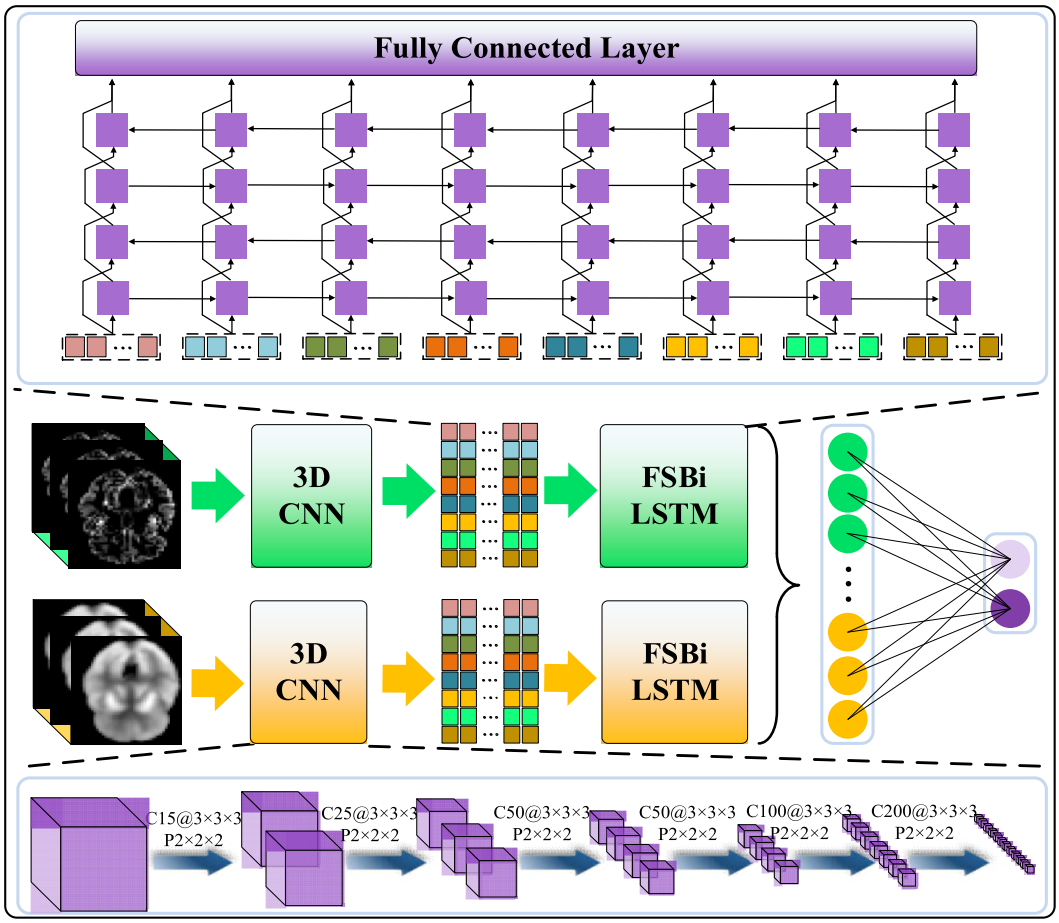

FIGURE 1. General framework of the proposed FSBi-LSTM method for AD diagnosis from MRI and PET neuroimages. C is convolutional layer, the P is mean pooling layer, @ is the number of filters such as 15@3× 3× 3 is 15 filters which size are 3× 3× 3 and P2 × 2 × 2 is pooling layers, which size are 2 × 0 × 2.

classification via MRI data and subjects' demographic information[24]. They identified the landmarks of discriminative anatomical from MRI data via data driven technique and then extracted multiple image patches from these detected landmarks. However, flattening operation before using the FC layer always ignores the spatial information in the feature map. In addition, Liu *et al.* recommended a framework based on 2D CNN and bidirectional gated recurrent unit (Bi-GRU) to alleviate the effect of this problem[30]. However, the way of converting 3D data into a series of 2D slices causes CNNs to completely ignore the characteristics of 3D data, and different slicing methods may lead to loss of different features. Moreover, existing CNN-based methods use the flattening layer after the CNN because the FC layer can only process 1D information. Using flattening layer will lead to the feature maps loss in all the 3D spatial information. In our work, we propose to utilize FSBi-LSTM to get rich spatial and semantic information from feature maps for efficient diagnosis of AD from MRI and PET.

III. METHODOLOGY

The framework of our proposed method is shown in Fig. 1. We employ 3D-CNN to extract the primary features of both MRI and PET inputs. Then, FSBi-LSTM is used to extract high level semantic and spatial information from the output of 3D-CNN instead of traditional FC layer. Finally, the learned features are concatenated and further passed to SoftMax classifier for disease diagnosis. In the following, a detailed description of the proposed method is presented.

A. DATA PREPROCESSING

For the MRI data, we conduct Anterior Commissure (AC)–Posterior Commissure (PC) reorientation and resample the data to 256 × 256 × 256 via MIPAV Software (https://mipav.cit.nih.gov/). Tissue intensities inhomogeneity is then corrected using N3 algorithm [37] followed by skull stripping and cerebellum removal. Afterwards, we segment the brain to GM, WM, and CSF using FSL package [38]. Existing research shows that compared with WM or CSF, GM demonstrated higher relatedness to AD/MCI [7]. Therefore, we choose the GM masks in this work. Finally, we use the hierarchical attribute matching mechanism for elastic registration (HAMMER) algorithm [39] to spatially register the GM masks to the MNI brain atlas coordinate space and extract the regional volumetric maps by image warping and tissue preserving method [40]. For the PET, firstly, we rigidly

align them to the MRI space. Then, we use Gaussian kernel with zero mean and unit standard deviation to handle MRI and PET. Finally, all the MRI and PET data are down-sampled to $64 \times 64 \times 64$ to save memory without compromising the classification as suggested by Suk *et al.* [7].

B. FEATURE LEARNING BASED 3D-CNN

The CNN is a powerful multilayer neural network in image analysis. However, 2D-CNN structures are designed for analyzing 2D images, which is inefficient to extract 3D medical images' spatial information. Therefore, we adopt the 3D convolution kernel instead of 2D one. For hierarchical learning the multi-level features, we use alternatively stacking convolutional layers and down-sampling layers to build the 3D convolutional kernel. Finally, we get feature maps from CNN model.

The input image is convolved with a list of kernel filters in convolutional layer. Then, a bias term is added between the activation function and convolutional layer. In this work, the rectified linear unit (ReLU) is chosen as the activation function. Finally, the CNN model can output a series of feature maps. We define the voxel positions for a given 3D image as x, y,and z,respectively, the j-th 3D kernel weight represents as $W_{kj}^{l}\left(\delta_x, \delta_y, \delta_z\right)$ connects the l-1 layer's k-th feature maps and the j-th feature maps of the l *layer*, the k-th feature maps of the l-1 layer as F_k^{l-1}, the kernel size corresponding to the x, y, and z is δ_x, δ_y, and δ_z, respectively. The convolutional response of the kernel filter is $u_{kj}^{l}(x, y, z)$. Then, the 3D convolutional layer is defined as

$$u_{kj}^{l}(x, y, z) = \sum_{\delta_x} \sum_{\delta_y} \sum_{\delta_z} F_k^{l-1}\left(x+\delta_x, y+\delta_y, z+\delta_z\right) \times W_{kj}^{l}\left(\delta_x, \delta_y, \delta_z\right), \quad (1)$$

After convolution, we add a ReLU to activate features:

$$F_j^{l}(x, y, z) = max\left(0, b_j^{l} + \sum_k u_{jk}^{l}(x, y, z)\right), \quad (2)$$

where b_j^{l} is bias term from the l-th layer's j-th feature map, $F_j^{l}(x, y, z)$ is obtained by summation of the response maps of the j-th 3D feature map's different convolution kernels.

After convolutional layer, a max-pooling layer is added to obtain more efficient and compact features. Besides, by using max-pooling layer, the features become more compact from low level to high level which can achieve the robustness against some variations.

In our model, the 3D-CNN architecture is adapted from Liu *et al.* [29]. However, since input data are not exactly the same, we modify the 3D-CNN structure of Liu *et al.* [29] by increasing the number of convolutional filters and layers. In addition, to avoid overfitting, we have appropriately reduced the number of filters in each convolution layer. More details can be found in Fig. 1. After 6 stacking convolutional and max-pooling layers, we add 2 FC layers and SoftMax classifier for training. All the features before the FC layers are flattened into an 1D vector. After training, we extract the feature before the last max-pooling layer as input to the FSBi-LSTM.

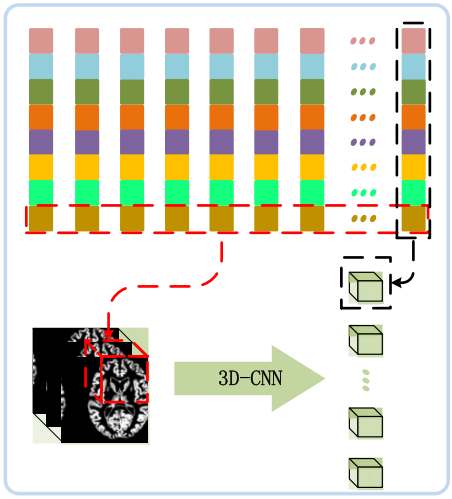

FIGURE 2. Relations among feature maps and the input data.

C. FSBI-LSTM BASED CLASSIFICATION

Typically, FC layers are used to do high level analysis in the CNN. However, the FC layers are unable to effectively extract all the spatial information from the feature map as it just simply connects all neurons. Fig. 2 shows the spatial information, which represents the relation between the feature maps and the input data. We can see each model's output of our 3D-CNN network contains 200 features with each dimension at $2 \times 2 \times 2$. In this figure, each column (i.e., the black box in Fig. 2) shows the whole features of the brain. Then, each row (*i.e.*, the red box in Fig. 2) shows all the features of a part of the brain. If we intercept a row-by-row feature maps like a red box, it is actually a part-by-part examination of the brain. Existing research shows that RNN and similar networks have the function of fusing different structures [24], [33]. Therefore, we design a FSBi-LSTM instead of FC layer to extract all the spatial information from the feature maps.

In traditional RNN, the output of the current cell state is defined as h_t and is expressed as

$$h_t = f\left(Ux_t + Wh_{t-1}\right), \quad (3)$$

where the x_t is input of the t-th unit, U is the weight from the input layer to the hidden layer, and W is the connection weight from the previous unit to the current unit. The $f(\cdot)$ is the *tanh* function. However, the traditional RNNs have problems of gradient explosion or gradient disappearance. To address this problem, we introduce the concept of gate and cell state using LSTM. In LSTM, the gate is an FC layer and the current cell state is defined as c_t. In order to control h_t and c_t, LSTM mainly uses three gates, including the input gate, the forget gate, and the output gate, respectively. The input gate is given as

$$i_t = \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + b_i\right), \quad (4)$$

where W_{xi} is the weight of input x_t, W_{hi} is the weight of the last cell output h_{t-1}, b_i is bias of the input gate, and σ is the

sigmoid function. Similarly, the forget gate is given as

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + b_f\right), \tag{5}$$

The weight of input x_t is defined as W_{xf}, the weight of the last cell output h_{t-1} is W_{hf} and b_{if} is bias of the forget gate.

In addition, we calculate the state of the input modulation (*i.e.*, short memory) based on the last output and the current input. It is computed as

$$g_t = \varphi\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right), \tag{6}$$

where W_{xc} is the weight of input x_t, W_{hc} is the weight of last cell output h_{t-1}, b_c is bias of the output gate, and φ is the *tanh* function. We calculate the cell state at the current time. It is generated by multiplying the element by the forget gate from the last unit state, multiplying the element by the input gate with the current input unit state, and adding the two products together. The long memory, c_t, is denoted as

$$c_t = i_t \odot g_t + f_t \odot c_{t-1}, \tag{7}$$

where $\odot$ denotes element-wise multiplication. In this way, we combine LSTM's current and long-term memory to form a new unit state. By controlling the forget gate, we can choose useful historical information. While controlling the input gate, we can avoid feeding insignificant current content into memory. The output gate controls the long-term memory on current output, which is represented as

$$o_t = \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + b_o\right), \tag{8}$$

where W_{xo} is the weight matrix of input x_t, W_{ho} is the weight matrix of the last cell output h_{t-1} and b_{io} is bias for the output gate. Finally, the output of LSTM is determined by the output gate and the unit state, which is formulated as

$$h_t = o_t \odot \varphi(c_t). \tag{9}$$

Eqs. 4 to 9 are the calculation process of single LSTM cell. We use $H(\cdot)$ to denote a LSTM cell. The idea of a Bi-LSTM assumes that each training sequence has to input forward and backward via two LSTMs. In output layer, two LSTMs are connected. This structure provides each point to complete future and previous contextual information in the output layer. Bi-LSTM computes the forward hidden sequence $\overrightarrow{h}$ and backward hidden sequence $\overleftarrow{h}$ as

$$\overrightarrow{h_t} = H\left(W_{x\overrightarrow{h}}x_t + W_{h\overrightarrow{h}}\overrightarrow{h_{t-1}} + \overrightarrow{b_h}\right), \tag{10}$$

$$\overleftarrow{h_t} = H\left(W_{x\overleftarrow{h}}x_t + W_{h\overleftarrow{h}}\overleftarrow{h_{t-1}} + \overleftarrow{b_h}\right), \tag{11}$$

where $W_{x\overrightarrow{h}}$ is the forward calculation of W_x, $W_{h\overrightarrow{h}}$ is the forward calculation of W_h, $\overrightarrow{b_h}$ is the parameter of forward calculation in $H(\cdot)$, and the $W_{x\overleftarrow{h}}$, $W_{h\overleftarrow{h}}$, and $\overleftarrow{b_h}$ are the parameters of backward calculations in $H(\cdot)$, respectively. We combine $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ to generate the final output y_t:

$$y_t = H\left(W_{\overrightarrow{h}y}\overrightarrow{h_t} + W_{\overleftarrow{h}y}\overleftarrow{h_t} + b_y\right), \tag{12}$$

where $W_{\overrightarrow{h}y}$ (forward hidden state weights), $W_{\overleftarrow{h}y}$ (backward hidden state weights), b is a bias vector for each layer.

By superimposing a basic LSTM cell in both forward and backward LSTM of Bi-LSTM and adding a FC layer in the output, we can get the FSBi-LSTM. For the hidden layer of FSBi-LSTM, the forward calculation is the same as LSTM except that the input sequence is opposite to the two hidden layers. FSBi-LSTM computes the forward hidden sequence $\overrightarrow{h^s}$ and backward hidden sequence $\overleftarrow{h^s}$, which is expressed as

$$\overrightarrow{h_t^s} = H\left(W_{x\overrightarrow{hs}}x_t\overrightarrow{h_t} + W_{h\overrightarrow{hs}}\overrightarrow{h_{t-1}^s} + \overrightarrow{b_{hs}}\right), \tag{13}$$

$$\overleftarrow{h_t^s} = H\left(W_{x\overleftarrow{hs}}\overleftarrow{h_t} + W_{h\overleftarrow{hs}}\overleftarrow{h_{t-1}^s} + \overleftarrow{b_{hs}}\right), \tag{14}$$

where $W_{\overrightarrow{hy}}$, $W_{h\overrightarrow{hs}}$, $\overrightarrow{b_{hs}}$, $W_{x\overleftarrow{hs}}$, $W_{h\overleftarrow{hs}}$, and $\overleftarrow{b_{hs}}$ are the same as Eqs. 10 and 11, respectively. Finally, the output of SBi-LSTM is denoted as

$$y_{st} = H\left(W_{\overrightarrow{h}y}\overrightarrow{h_t} + W_{\overleftarrow{h}y}\overleftarrow{h_t} + b_y\right). \tag{15}$$

In fact, y_{st} can be used as the final feature processing after fusion. However, the feature maps used are the brain structure feature maps, and all features are related to each other. Therefore, each output node is related to the current input. The outputs of every repeating cell can be of equal importance and shall be concatenated into a full connection layer. We may extract a common closely connected brain structure information from all the SBi-LSTM cells to represent constant "*trait*" information of each subject via the FC layer, instead of the part of the brain structure information. This is the whole computing process of FSBi-LSTM.

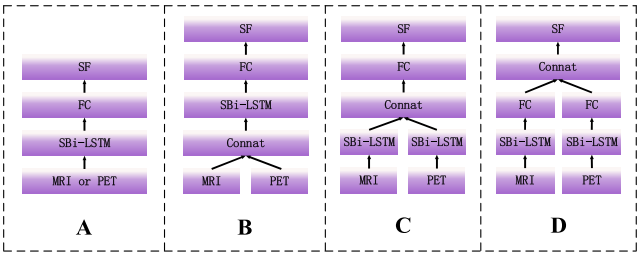

FIGURE 3. Different fusion strategies. A) Single modality without fusion, B) FSB, C) FFC, D) FFS.

Since MRI and PET modalities are used in this study, fusion is an effective way for performance boosting. Based on our data and network structure, we devise 3 fusion strategies: fusion before SBi-LSTM (FSB), fusion before FC layer (FFC) and fusion before final SoftMax classifier (FFS). Fig. 3 shows the structure of different fusion strategies. Since FSB needs to fuse forward LSTM and backward LSTM, we choose FFS as our integration strategy. As a result, we can avoid modal fusion influence on the performance of FSBi-LSTM. After the fusion, we use SoftMax classifier to get the final diagnosis result.

IV. EXPERIMENTAL RESULTS

A. EXPERIMENTAL SETUP

In this paper, our dataset is based on the Alzheimer's Disease Neuroimaging Initiative (ADNI) (http://adni.loni.usc.edu/).

TABLE 1. Performance evaluation of different classification tasks using different configurations (%).

Classifier	Method	ACC	SEN	SPEC	F1	BAC	AUC
AD vs. NC	FBL1	93.26±4.96	97.62±4.22	89.91±7.87	92.66±6.14	93.76±3.89	96.92±4.24
	BL1	92.23±7.01	94.32±6.07	90.48±9.31	91.71±8.20	92.40±6.42	**97.11±3.55**
	FSBL2	**94.82±4.90**	**97.70±4.22**	**92.45±7.31**	**94.44±5.97**	**95.08±3.96**	96.76±4.02
	SBL2	93.78±7.76	95.51±8.35	92.31±8.38	93.41±8.68	93.91±7.78	96.14±7.78
	FSBL3	92.75±7.55	93.41±8.23	92.16±8.41	92.39±8.45	92.78±7.60	94.28±6.21
	SBL3	93.26±7.84	95.45±8.41	91.43±7.99	92.82±8.70	93.44±7.86	95.62±5.36
	FSBL4	92.75±6.16	95.40±6.10	90.57±6.83	92.22±7.16	92.98±5.93	96.94±6.32
	SBL4	92.23±6.21	93.33±5.89	91.26±7.44	91.80±7.23	92.30±5.97	94.80±6.19
pMCI vs. NC	FBL1	83.52±9.32	84.06±16.34	83.18±6.81	80.00±10.87	83.62±10.36	90.80±7.36
	BL1	82.95±8.18	**84.85±15.14**	81.82±5.92	78.87±9.61	83.33±8.94	90.42±7.24
	FSBL2	**86.36±11.02**	83.33±16.32	**88.78±8.78**	**84.42±11.80**	**86.05±11.10**	**91.11±9.02**
	SBL2	84.09±12.57	81.58±18.06	86.00±10.36	81.58±14.69	83.79±13.08	89.84±8.87
	FSBL3	85.23±11.84	82.89±15.43	87.00±10.68	82.89±13.40	84.95±12.09	90.51±10.96
	SBL3	83.52±12.01	81.33±16.58	85.15±11.97	80.79±14.08	83.24±12.62	88.43±9.78
	FSBL4	84.66±11.67	81.82±16.89	86.87±10.60	82.35±13.49	84.34±12.34	85.18±10.19
	SBL4	83.52±13.00	80.52±18.46	85.86±11.90	81.05±15.09	83.19±13.72	85.36±10.60
sMCI vs. NC	FBL1	60.09±7.47	62.42±5.77	55.70±17.97	67.15±7.98	59.06±10.68	61.97±6.75
	BL1	62.72±10.92	67.48±10.43	57.14±11.83	66.14±10.13	62.31±11.00	65.84±11.38
	FSBL2	**65.35±10.22**	**70.59±9.30**	**59.63±11.58**	68.02±9.82	**65.11±10.21**	**69.17±10.70**
	SBL2	64.47±10.69	69.75±10.40	58.72±11.39	67.21±9.65	64.23±10.83	64.27±11.58
	FSBL3	63.60±9.97	68.29±10.60	58.10±10.27	66.93±8.75	63.19±10.26	65.62±10.96
	SBL3	62.72±10.03	67.77±11.07	57.01±10.32	65.86±9.27	62.39±10.37	67.21±10.52
	FSBL4	64.47±12.09	68.50±10.89	59.41±13.84	**68.24±10.80**	63.95±12.30	67.64±10.73
	SBL4	63.60±9.53	68.60±10.34	57.94±10.30	66.67±8.23	63.27±10.00	63.90±11.82

We choose the baseline MRI data and 18-Fluoro-DeoxyGlucose PET data acquired from 93 AD, 76 pMCI, 128 sMCI, and 100 normal control (NC). To verify the efficacy of our model, we set up the following experiments: AD vs. NC, pMCI vs. NC, and sMCI vs. NC instead of AD vs. NC, MCI vs. NC, and pMCI vs. sMCI. The main reason for this setup is that more meaningful pathological structures can be found by comparing with healthy people. We use two duals by alternatively stacking 6 convolutional and max-pooling layers to get the feature maps from MRI and PET, separately. In order to facilitate training, we add two FC layers and SoftMax as the classifier. To avoid overfitting, we randomly drop out ten percent of neurons during training in the 3D-CNN. In 3D-CNN part, we set all the convolutional layer stride to 2 and padding is the same with layer input, we choose *Adam* optimizer [41] for optimization and categorical cross entropy as the loss function. In addition, we set the batch size as 20, the number of epochs as 60, the learning rate as 10^{-4}, the fuzz factor as 10^{-9}, the first order exponential decay rates for the moment estimates as 0.9, and the second order exponential decay rates for the moment estimates as 0.999. After 3D-CNN training via two models, we extract the feature before the last max-pooling layer as input to the FSBi-LSTM. In FSBi-LSTM part, we choose *RMSprop* as the optimizer [42] for speeding up training and categorical cross entropy as the loss function. We set the batch size as 30, the number of epochs as 25, the learning rate as 10^{-3}, the fuzz factor as 10^{-8}, the rho as 0.9, and the learning rate decay as 10^{-5}.

For classification performance evaluation, we use different performance metrics, namely, accuracy (ACC), sensitivity (SEN), specificity (SPEC), F1 score (F1), balanced accuracy (BAC), and area under receive operation curve (AUC). A 10-fold cross-validation algorithm is adopted to assess both classification performance. Specifically, all samples are divided into 10 portions, then samples in one portion are successively used as the testing data while the rest are utilized as the training data. All the experiments are conducted on a Windows machine with NVIDIA TITAN Xt GPU and implemented using Keras library with Tensorflow as backend.

B. DEPTH EFFECTS ON SBI-LSTM AND FSBI-LSTM

For investigating the influence of depth on the FSBi-LSTM performance, we test our model with varied number of layers while fixing the other parameters. In order to compare the impact of the FC layer on the performance, we also set up different layers of SBi-LSTM for comparison. Namely, we test the following models: 1 layer FBi-LSTM (FBL1), 1 layer Bi-LSTM (BL1), 2 layers FSBi-LSTM (FSBL2), 2 layers SBi-LSTM (SBL2), 3 layers FSBi-LSTM (FSBL3), 3 layers SBi-LSTM (SBL3), 4 layers FSBi-LSTM (FSBL4), and 4 layers SBi-LSTM (SBL4). We use the same CNN output feature maps to test these configurations. We then use t-test statistical analysis to evaluate the significance of the obtained results. Our results show statistical significance with the confidence interval at 0.001.

Table 1 summarizes the influence of the number of LSTM layers and the FC layer on the FSBi-LSTM (boldfaces denote the best performance). Fig. 4 illustrates the corresponding results on different classification tasks under different configurations of the proposed method. As shown in Table 1 and Fig. 4, our model outperforms other LSTM-based models. We find the optimal number of LSTM layers by changing

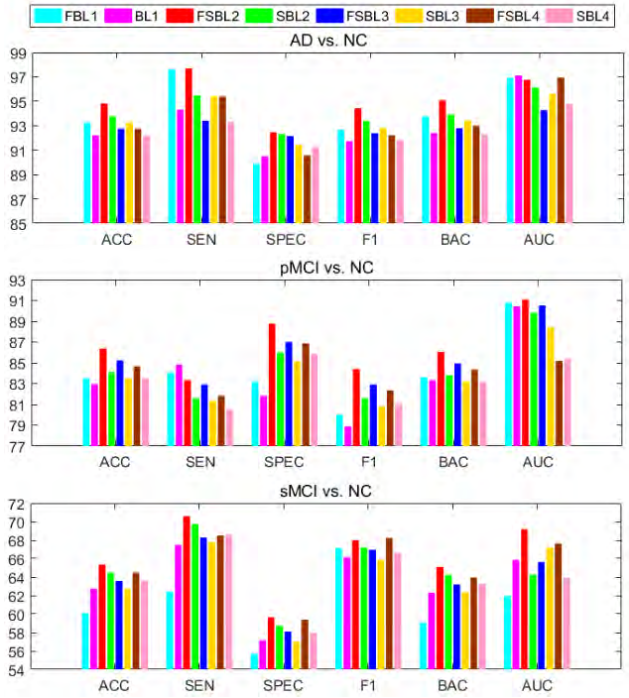

FIGURE 4. Effects of different layers on SBi-LSTM and FSBi-LSTM performances for different classification tasks.

the number of layers of LSTM in FSBi-LSTM. From these experiments, we have the following observations:

(1) Model performance can be improved by increasing LSTM layers, but this is not always true. In our experiment, the best performance is obtained when the number of LSTM layers is set to 2. The reason is that, if the number of layers is smaller than 2, the FSBi-LSTM model cannot extract enough deep features. On the other hand, if the layer number is bigger than 2, the performance decreases due to the gradient vanishing problem, which occurs when the network depth is increased. In addition, as layer number increases, the overfitting issue occurs with the increased number of parameter and fixed training data samples. Therefore, we choose 2 layers LSTM in our FSBi-LSTM model.

(2) FSBi-LSTM performs better than SBi-LSTM as the former can better preserve extracted features from the input since the feature maps represent the brain structure feature maps, and all the features are related to each other. Due to the characteristics of LSTM, the output of each node is the most relevant to the current input, and the outputs of every repeating cell can be of equal importance. Therefore, we can extract closely connected brain structure information from all the SBi-LSTM cells to represent constant "*trait*" information of each subject by FC layer. Thus, FSBi-LSTM performs better than SBi-LSTM.

C. EFFECT OF FUSION STRATEGIES

The previous studies demonstrated that different fusion strategies with modal choices can have varied classification results [7], [43]. Thus, we evaluate the effectiveness of our FSBi-LSTM using different feature fusion strategies. Fig. 3 shows the three different strategies and model structures. Furthermore, we use the t-test to evaluate the significance of the obtained results. Our results again show statistical significance with the confidence interval at 0.001. Likewise, we use the same CNN output feature maps to test different fusion strategies. Table 2 and Fig. 5 show the influence of fusion strategies on the FSBi-LSTM performance (boldfaces denote the best performance).

From the results in Fig. 5 and Table 2, it is clear that the multi-modal fusion can get better results than a single modality. The results also show that MRI has better performance than PET since MRI can sufficiently capture structural information of brain regions. In addition, comparisons of different fusion strategies show that FFS has the best performance. The main reasons are as follows. In case of FFC, forward LSTM, backward LSTM fusion and mode fusion will interfere with each other. On the other hand, in FSB, fusion can be either along the edge of the feature map, which causes the input size too long (400×8). If fusion is along the short edge of feature map, which makes the feature map size becomes 200×16. Accordingly, MRI is the forward input of SBi-LSTM, while PET is the backward input of SBi-LSTM, which will have an influence for fusion forward and backward.

TABLE 2. Performance evaluation of different classification tasks using different fusion strategies (%).

Classifier	Method	ACC	SEN	SPEC	F1	BAC	AUC
AD vs. NC	MRI	92.75±7.55	93.41±8.23	92.16±8.41	92.39±8.45	92.78±7.60	95.23±6.49
	PET	92.23±7.94	93.33±8.36	91.26±8.02	91.80±8.78	92.30±7.94	96.51±4.64
	FSB	94.30±7.63	95.56±8.25	**93.20±8.62**	93.99±8.55	94.38±7.61	97.19±4.60
	FFC	93.26±7.88	94.44±8.42	92.23±8.41	92.90±8.77	93.34±7.89	**97.35±4.36**
	FFS	**94.82±4.90**	**97.70±4.22**	92.45±7.31	**94.44±5.97**	**95.08±3.96**	96.76±4.02
pMCI vs. NC	MRI	80.68±11.59	77.63±16.76	83.00±11.72	77.63±14.34	80.32±12.06	89.83±10.00
	PET	83.52±12.84	80.52±16.35	85.86±11.13	81.05±14.42	83.19±12.98	89.50±9.60
	FSB	85.80±10.42	83.12±15.51	87.88±9.84	83.66±12.38	85.50±11.02	89.18±9.47
	FFC	84.66±10.23	81.82±16.06	86.87±7.61	82.35±11.46	84.34±10.60	**91.36±9.17**
	FFS	**86.36±11.02**	**83.33±16.32**	**88.78±8.78**	**84.42±11.80**	**86.05±11.10**	91.11±9.02
sMCI vs. NC	MRI	55.26±9.52	59.70±7.96	48.94±12.60	61.07±8.45	54.32±10.18	57.17±11.81
	PET	61.84±10.83	67.52±11.02	55.86±11.97	64.49±9.96	61.69±11.15	67.93±11.95
	FSB	64.04±10.84	69.49±10.44	58.18±11.85	66.67±9.83	63.84±11.01	67.51±10.56
	FFC	64.47±12.83	69.11±11.11	59.05±14.86	67.73±11.99	64.08±12.90	67.46±12.47
	FFS	**65.35±10.22**	**70.59±9.30**	**59.63±11.58**	**68.02±9.82**	**65.11±10.21**	**69.17±10.70**

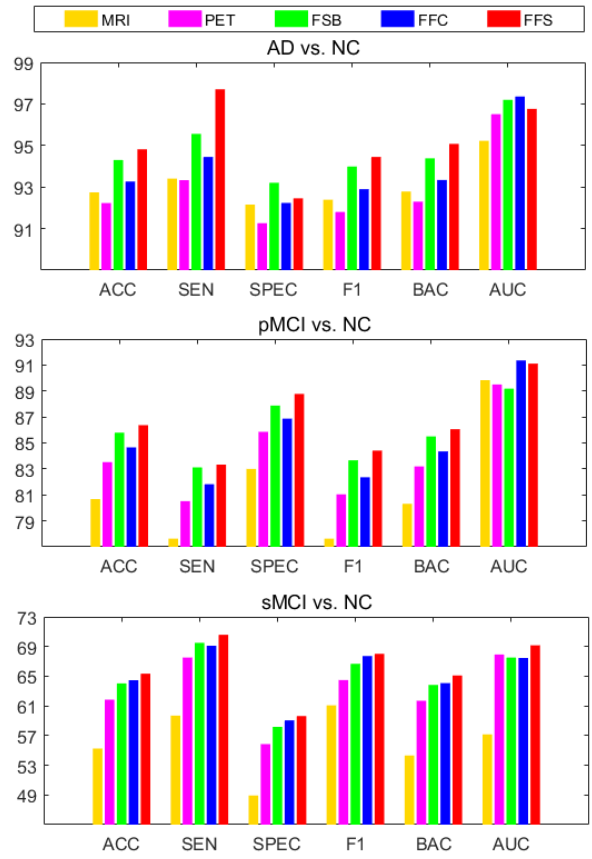

FIGURE 5. Effects of different fusion strategies on FSBi-LSTM performances for different classification tasks.

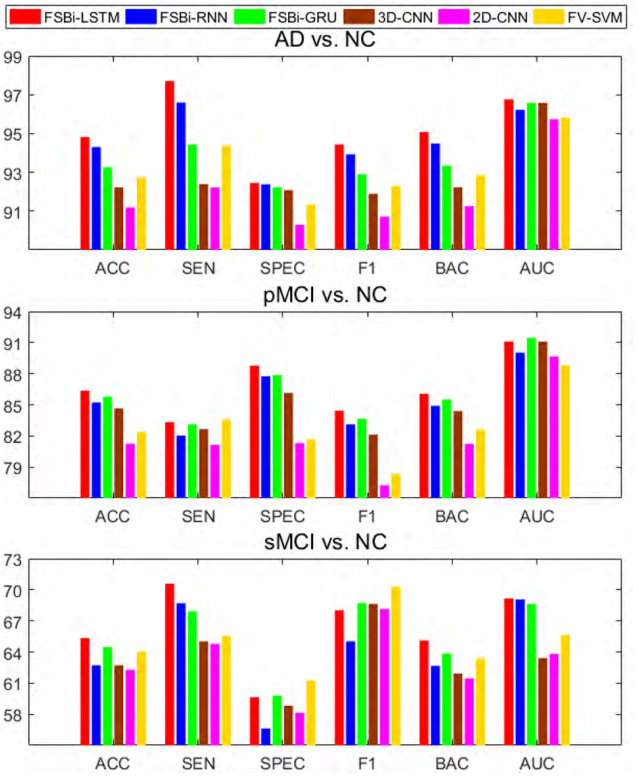

FIGURE 6. Classification performances of different information extraction methods for different classification tasks.

D. FUSION METHOD COMPARISON

In this sub-section, we compare FSBi-LSTM with other commonly used feature fusion methods. There are three main types of fusion methods. The first type is to use the other RNN cell instead of the LSTM cell such as simple RNN and GRU [44], we denote them as FSBi-RNN and FSBi-GRU. The second type is to use the CNN model instead of the RNN model including both 2D-CNN and 3D-CNN. The third type is the traditional machine learning method via Fisher vector (FV) [45] and support vector machine (SVM) [46]. We use the t-test to evaluate the significance of the results. When the confidence interval is set as 0.001, our results have statistical significance. We use the same CNN output feature map to test different feature fusion methods. Table 3 and Fig. 6 illustrate the influence of the number of feature fusion methods. In addition, Fig.7 shows the ROC curves of different models. From these results, we observe that the

TABLE 3. Classification accuracies of different feature fusion methods in different classification tasks (%).

Classifier	Method	ACC	SEN	SPEC	F1	BAC	AUC
AD vs. NC	FSBi-LSTM	**94.82±4.90**	**97.70±4.22**	**92.45±7.31**	**94.44±5.97**	**95.08±3.96**	**96.76±4.02**
	FSBi-RNN	94.30±4.55	96.59±4.83	92.38±7.31	93.92±5.62	94.49±3.62	96.23±4.10
	FSBi-GRU	93.26±6.07	94.44±5.97	92.23±7.82	92.90±7.13	93.34±5.79	96.58±4.12
	3D-CNN	92.23±6.19	92.39±7.98	92.08±7.82	91.89±7.10	92.24±5.83	96.58±4.50
	2D-CNN	91.19±7.70	92.22±8.56	90.29±8.33	90.71±8.40	91.26±7.56	95.74±7.70
	FV-SVM	92.75±6.64	94.38±8.95	91.35±7.05	92.31±7.06	92.86±5.95	95.83±4.64
pMCI vs. NC	FSBi-LSTM	**86.36±11.02**	**83.33±16.32**	**88.78±8.78**	**84.42±11.80**	**86.05±11.10**	91.11±9.02
	FSBi-RNN	85.23±9.88	82.05±13.49	87.76±10.18	83.12±11.57	84.90±10.37	90.03±10.04
	FSBi-GRU	85.80±10.05	83.12±14.65	87.88±9.84	83.66±11.96	85.50±10.61	**91.45±10.17**
	3D-CNN	84.66±9.23	82.67±13.55	86.14±8.07	82.12±10.88	84.40±9.69	91.09±9.39
	2D-CNN	81.25±9.46	81.16±16.73	81.31±6.97	77.24±10.43	81.23±10.44	89.66±7.76
	FV-SVM	82.39±7.42	83.58±13.94	81.65±6.20	78.32±9.13	82.62±8.49	88.82±7.79
sMCI vs. NC	FSBi-LSTM	**65.35±9.30**	**70.59±11.58**	59.63±9.82	68.02±10.21	**65.11±10.21**	**69.17±10.70**
	FSBi-RNN	62.72±7.90	68.70±9.61	56.64±8.59	65.02±6.69	62.67±8.58	69.06±10.52
	FSBi-GRU	64.47±12.25	67.94±10.77	59.79±14.35	68.73±10.99	63.87±12.46	68.64±10.98
	3D-CNN	62.72±7.87	65.03±6.46	58.82±12.38	68.63±7.30	61.93±8.79	63.41±7.87
	2D-CNN	62.28±10.27	64.79±7.28	58.14±15.53	68.15±9.78	61.46±11.29	63.83±5.93
	FV-SVM	64.04±9.30	65.54±7.18	**61.25±14.95**	**70.29±9.20**	63.40±10.48	65.64±11.90

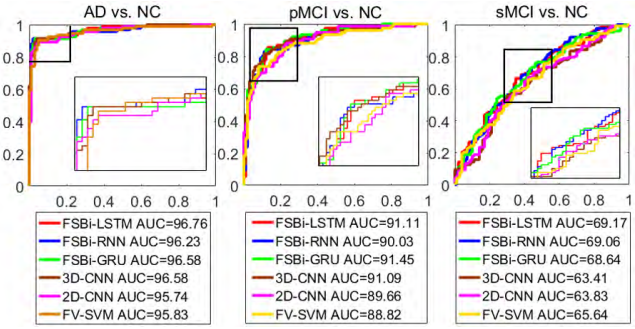

FIGURE 7. ROC curves of different information extraction models.

result of FSBi-LSTM is better than that of FSBi-RNN and FSBi-GRU.

The main explanations of the superior performance are as follows. Compared with the simple RNN cell, the LSTM cell can use three gates to solve the gradient exploding/gradient vanishing problem. In addition, as a variant of LSTM cell, GRU cell synthesizes a single update gate from the forget gate and the input gate, which can train the network faster with decreased accuracy. However, our network is insensitive to time but sensitive to accuracy. Hence, we choose the LSTM instead of the GRU. Compared with 2D-CNN and 3D-CNN, we use CNN to extract spatial information from feature maps. FSBi-LSTM can provide a better result than the FSBi-LSTM with progressive scans, which is more effective than direct convolution using convolution kernels to identify the informative features. Because CNN structure cannot output 1D feature, LSTM structure can output 1D feature to avoid space information loss caused by the flattening operation. Compared with 2D-CNN, 3D-CNN can get a better result, because 2D-CNN loses 3D information. Accordingly, the FSBi-LSTM can produce better results than traditional machine learning methods as our method can extract deep features.

E. VISUALIZATION ANALYSIS

For the visualization analysis, we have conducted two experiments: t-SNE feature visualization and disease-related region search. For the t-SNE feature visualization experiments, we visualize the feature maps before and after input FSBi-LSTM through the t-SNE model. We also visualize the feature maps obtained from different models and tasks. The results shown in Fig. 8 demonstrates FSBi-LSTM plays an important role in feature discrimination. However, the last subfigures show that the diagnostic performance of sMCI is still limited. From Fig. 8, we can see that intra-class differences are even greater than inter-class differences between sMCI and NC.

Brain consists of many regions which are responsible for many tasks and not all regions are closely related to AD. Hence, we attempt to utilize our proposed method to search for these relevant ROIs for understanding brain abnormalities. To achieve this aim, we exclude brain images' different local areas systematically with a 3D ROI grey box and monitor the classifier outputs. If the grey box covers the important area that is related to AD, the correct class' prediction will significantly drops. Here, we use the 93 brain ROIs proposed by Kabani *et al.* [47] as a reference. We use the whole brain as a control group. We then shield 93 brain regions in turn and get 93 classification results after shielding. The classification accuracy obtained by shielding different brain regions is shown in Table 4. We also choose the top 10 brain regions, which have the greatest impacts on the diagnosis of AD, pMCI, and sMCI, respectively, as illustrated in Fig. 9.

From the results, we conclude that the top 10 ROIs with the greatest impact on the diagnosis of AD are: uncus right, superior frontal gyrus right, parahippocampal gyrus left, superior temporal gyrus right, hippocampal formation right, subthalamic nucleus right, thalamus right, middle frontal gyrus left, precuneus left, and inferior temporal gyrus left. The top 10 ROIs with the greatest impact on the diagnosis of

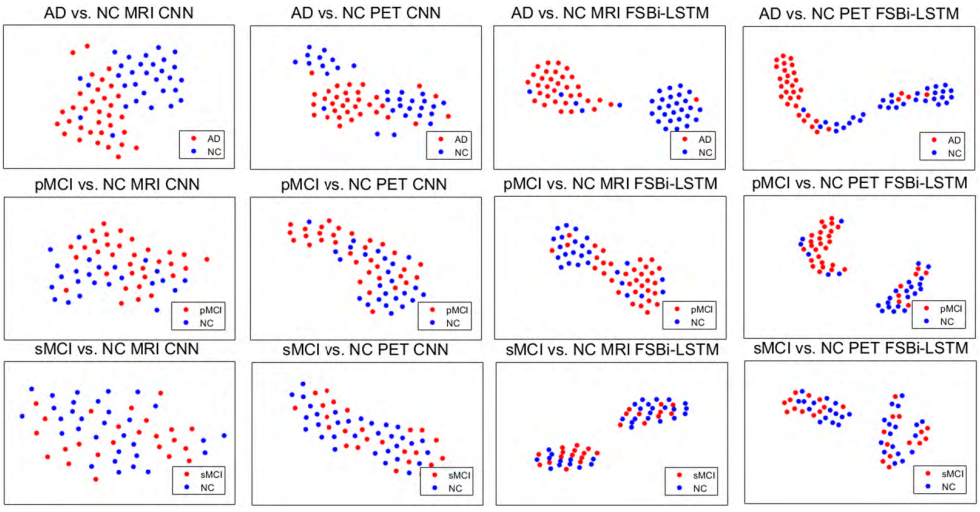

FIGURE 8. The t-SNE visualization result of feature maps from different models and different tasks.

TABLE 4. The 93 ROIs and their corresponding classification accuracy obtained from the proposed method (%).

No.	ROI	AD	pMCI	sMCI	No.	ROI	AD	pMCI	sMCI
0	unscreened brain area	94.82	86.36	65.35	47	middle occipital gyrus right	89.05	77.91	55.24
1	medial front-orbital gyrus right	86.97	74.44	50.45	48	middle temporal gyrus left	90.11	75.00	56.58
2	middle frontal gyrus right	90.63	75.59	53.50	49	lingual gyrus left	91.18	76.05	58.36
3	lateral ventricle left	89.55	74.97	53.06	50	superior frontal gyrus left	90.63	77.29	53.02
4	insula right	90.63	74.35	53.91	51	nucleus accumbens left	91.63	75.00	51.80
5	precentral gyrus right	88.55	75.59	55.65	52	occipital lobe WM left	89.55	76.63	55.32
6	lateral front-orbital gyrus right	89.58	73.82	55.67	53	postcentral gyrus left	90.13	73.33	55.28
7	cingulate region right	89.61	75.62	58.77	54	inferior frontal gyrus right	91.71	75.56	56.13
8	lateral ventricle right	89.08	77.94	51.36	55	precentral gyrus left	90.08	77.25	58.77
9	medial frontal gyrus left	90.61	70.92	52.67	56	temporal lobe WM left	89.61	74.44	53.48
10	superior frontal gyrus right	90.11	77.16	55.24	57	medial front-orbital gyrus left	92.21	73.30	52.23
11	globus palladus right	87.50	78.40	56.13	58	perirhinal cortex right	89.08	78.95	57.47
12	globus palladus left	92.18	77.25	57.41	59	superior parietal lobule right	89.08	74.44	55.71
13	putamen left	90.11	74.97	52.63	60	lateral front-orbital gyrus left	89.08	76.67	55.73
14	inferior frontal gyrus left	90.08	76.70	54.45	61	perirhinal cortex left	90.61	78.37	57.83
15	putamen right	91.66	75.03	55.63	62	inferior temporal gyrus left	88.53	76.14	56.19
16	frontal lobe WM right	93.24	75.56	57.00	63	temporal pole left	90.63	74.97	55.69
17	parahippocampal gyrus left	90.13	77.29	49.53	64	entorhinal cortex left	90.61	73.89	53.52
18	angular gyrus right	87.53	73.92	56.56	65	inferior occipital gyrus right	90.61	75.49	52.59
19	temporal pole right	89.61	76.67	57.02	66	superior occipital gyrus left	91.16	75.07	53.54
20	subthalamic nucleus right	89.55	77.29	54.35	67	lateral occipitotemporal gyrus right	89.55	76.80	56.09
21	nucleus accumbens right	88.03	75.62	53.08	68	entorhinal cortex right	91.66	75.00	57.00
22	uncus right	90.08	77.25	61.01	69	hippocampal formation left	90.63	74.97	53.50
23	cingulate region left	91.66	73.89	57.41	70	thalamus left	91.18	78.92	55.73
24	fornix left	90.13	73.86	53.10	71	parietal lobe WM right	89.58	76.80	54.86
25	frontal lobe WM left	89.55	75.00	55.22	72	insula left	90.13	77.81	54.84
26	precuneus right	89.63	76.83	53.58	73	postcentral gyrus right	90.61	73.24	55.71
27	subthalamic nucleus left	90.63	78.40	57.02	74	lingual gyrus right	89.08	76.08	58.26
28	posterior limb of internal capsule inc. cerebral peduncle left	88.58	76.11	53.44	75	medial frontal gyrus right	89.58	80.65	53.95
29	posterior limb of internal capsule inc.cerebral peduncle right	88.55	77.84	57.49	76	amygdala left	90.11	78.46	52.19
30	hippocampal formation right	88.03	78.99	55.32	77	medial occipitotemporal gyrus left	92.18	75.03	55.30
31	inferior occipital gyrus left	90.08	77.91	50.93	78	parahippocampal gyrus right	90.11	78.89	53.93
32	superior occipital gyrus right	89.61	76.11	56.19	79	anterior limb of internal capsule right	89.08	76.73	51.72
33	caudate nucleus left	90.63	76.73	54.37	80	middle temporal gyrus right	89.11	76.18	52.15
34	supramarginal gyrus left	90.63	74.35	59.23	81	occipital pole right	90.63	78.40	57.00
35	anterior limb of internal capsule left	90.11	76.67	56.54	82	corpus callosum	89.05	74.97	54.35
36	occipital lobe WM right	92.18	79.02	50.87	83	amygdala right	90.63	76.60	57.85
37	middle frontal gyrus left	88.53	73.43	56.21	84	inferior temporal gyrus right	90.63	74.97	55.22
38	superior parietal lobule left	89.08	76.14	54.37	85	superior temporal gyrus right	88.00	75.13	49.05
39	caudate nucleus right	90.63	73.86	49.98	86	middle occipital gyrus left	91.13	75.56	53.16
40	cuneus left	90.08	74.48	58.34	87	angular gyrus left	91.68	74.97	54.86
41	precuneus left	88.53	75.62	58.34	88	medial occipitotemporal gyrus right	90.08	79.44	57.85
42	parietal lobe WM left	91.66	75.62	58.36	89	cuneus right	90.11	79.61	57.43
43	temporal lobe WM right	90.61	75.62	57.09	90	lateral occipitotemporal gyrus left	90.11	77.91	55.24
44	supramarginal gyrus right	91.16	78.92	53.56	91	thalamus right	88.08	76.11	57.94
45	superior temporal gyrus left	89.58	75.03	55.26	92	occipital pole left	89.58	77.22	52.25
46	uncus left	89.03	75.03	53.99	93	fornix right	90.11	76.11	56.56

pMCI are: medial frontal gyrus left, postcentral gyrus right, medial front-orbital gyrus left, postcentral gyrus left, middle frontal gyrus left, lateral front-orbital gyrus right, fornix left, caudate nucleus right, a cingulate region left, and entorhinal cortex left. Finally, the top 10 ROIs with greatest impact on the diagnosis of sMCI are: superior temporal gyrus right, parahippocampal gyrus left, caudate nucleus right, medial front-orbital gyrus right, occipital lobe WM right, inferior occipital gyrus left, lateral ventricle right, anterior limb of internal capsule right, nucleus accumbens left, and middle temporal gyrus right.

F. COMPARISION WITH OTHER DEEP LEARNING MODEL

In this subsection, we compare the performance of the proposed method with other related deep learning models. To guarantee fair comparison, we choose the models that used the same dataset. Tables 5, 6, and 7 show that our method can obtain higher accuracies than existing methods. In [30], Liu *et al.* used 2D-CNN to capture the features of image slices, then the SBi-GRU was cascaded to learn and integrate the inter-slice features for image classification. Our method outperforms this method due to the following reasons. The 3D-CNN can preserve more space information than 2D slices without information loss although GRU can replenish 3D information. Compared with Liu *et al* [29], the FSBi-LSTM with progressive scans is more effective than direct convolution using a 2D convolution kernels to identify the informative features. Because flattening may lead to a large loss of 3D spatial information when feature maps output from 3D-CNN is input into 2D-CNN network

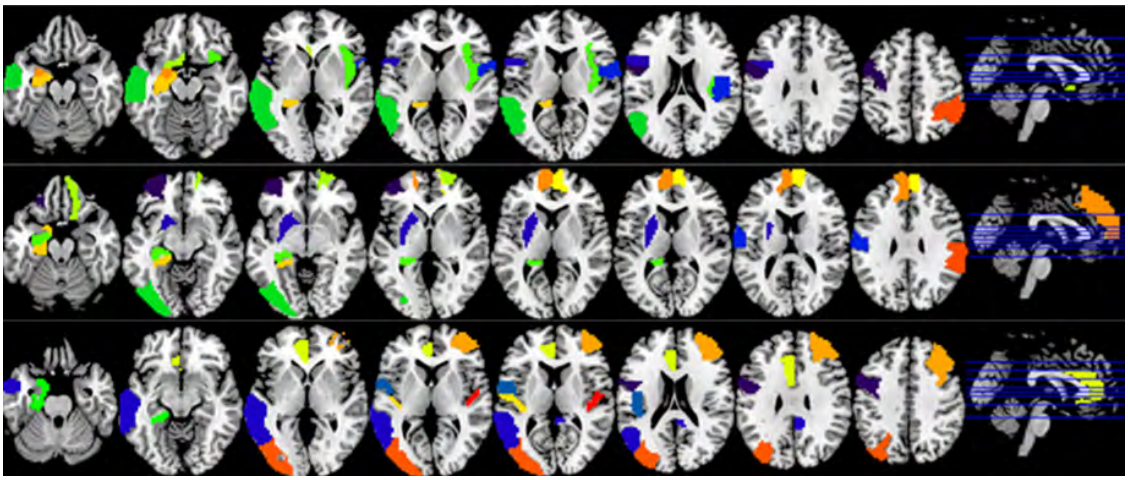

FIGURE 9. Top 10 brain regions that have the greatest impact on the diagnosis of AD, pMCI, and sMCI shown in top, middle, and bottom row, respectively.

TABLE 5. Algorithm comparisons for AD vs. NC classification(%).

Algorithm	Subject	Modality	ACC	SEN	SPE	AUC
Liu *et al.* 2018 [30]	93AD+100NC	PET	91.20	91.40	91.00	95.30
Liu *et al.* 2018 [29]	93AD+100NC	PET+MRI	93.26	92.55	**93.94**	95.68
Feng *et al.* 2018 [36]	93AD+100NC	PET+MRI	94.29	96.59	92.38	96.23
Ours	93AD+100NC	PET+MRI	**94.82**	**97.70**	92.45	**96.76**

TABLE 6. Algorithm comparisons for pMCI vs. NC classification(%).

Algorithm	Subject	Modality	ACC	SEN	SPE	AUC
Liu *et al.* 2018 [29]	76pMCI+100NC	PET+MRI	82.95	81.08	84.31	88.43
Feng *et al.* 2018 [36]	76pMCI+100NC	PET+MRI	84.66	**83.56**	85.44	89.63
Ours	76pMCI+100NC	PET+MRI	**86.36**	83.33	**88.78**	**91.11**

TABLE 7. Algorithm comparisons for sMCI vs. NC classification(%).

Algorithm	Subject	Modality	ACC	SEN	SPE	AUC
Liu *et al.* 2018[29]	128sMCI+100NC	PET+MRI	64.04	63.07	**67.31**	67.05
Feng *et al.* 2018 [36]	128sMCI+100NC	PET+MRI	64.47	70.43	48.41	67.14
Ours	128sMCI+100NC	PET+MRI	**65.35**	**70.59**	59.63	**69.17**

to extract high level semantic features. Comparing with the input data that is cut into several blocks to train several relatively independent 3D-CNN modules and then fusion, our method of down-sampling can greatly reduce the demand for data as we have less training parameters. Compared with our previous work [36], LSTM can effectively alleviate the gradient vanishing problem by controlling information flow with several gates. In addition, if we use the FC layer, we can further extract and sort out the output of SBi-LSTM. Thus, our method can achieve the best result.

V. DISCUSSIONS AND LIMITATIONS

In this study, the proposed framework efficiently shows good performance on the three binary classification tasks (*i.e.* AD vs. NC, pMCI vs. NC, and sMCI vs. NC). In our proposed framework, there are 3D-CNN and FSBi-LSTM. Compared with the traditional computational algorithms (such as strength and similarity guided group-level brain functional network), our method does not need prior knowledge to extract features manually, which can avoid the subjectivity.

It is important to know whether further processing of extracted features in the CNN network can further improve the early diagnosis of AD. Hence, we provide a comprehensive investigation about the influence of the model performance. In fact, the brain structural and functional information extracted using CNN are regarded as sequences and then further analyzed by FSBi-LSTM to get the high-level space information. In addition, FSBi-LSTM has fewer parameters than FC layers (27938 vs. 57578), which can make convergence faster.

Also, we enhance the relationship between features extracted by SBi-LSTM to further boost AD early diagnosis performance. From the experimental results, it is clear that our method is statistically superior to the related algorithms (i.e., SBi-LSTM and SBi-GRU) and previous studies. The primary explanation is that the input feature maps are the brain structure feature maps, and all the features are related to each other. Therefore, the outputs of every repeating cells are equally important. Besides, each output node of LSTM is more relevant to the current input node. Hence, with this layer, we may extract common closely connected brain structure information from all the SBi-LSTM cells, which may represent constant "*trait*" information of each subject, instead of the part of the brain structure information. Finally, it is noteworthy that the experimental results of our method are consistent with the related previous studies [36].

Despite the promising performance achieved by the proposed method, it still suffers from few limitations. First, despite the good accuracies obtained for AD vs. NC and pMCI vs. NC tasks, the diagnostic performance of sMCI is still limited. This is probably due to that the anatomical changes of sMCI are very subtle and cannot be well observed. Second, our method cannot directly find the brain lesion structure by up-sampling or deconvolution due to the cascade of CNN and LSTM. Instead, it is only found by shielding the brain area. Third, in this study, we do not utilize the longitudinal MRI data, which can further provide complementary information about disease evolution.

Since we only focus on the voxel features currently, it might be beneficial to use the state-of-the-art methods to integrate the visual features computer vision techniques as well. For processing, we can maximize the preservation of structural information in the brain. Besides, the sharing and common information can be uncovered by us among different features to facilitate the prognosis and diagnosis in the clinical application. These insights and limitations are yet to be explored in our future work.

VI. CONCLUSIONS

In this paper, a novel framework composed of 3D-CNN and FSBi-LSTM is proposed for diagnosing AD. Specifically, we propose a new LSTM network framework instead of the FC layer in 3D-CNN. Our method can preserve space information from feature map as much as possible. Compared with traditional SBi-LSTM, FSBi-LSTM extracts common closely connected brain structure information from all the SBi-LSTM cells, which can represent constant "*trait*" information of each subject via the FC layer, instead of the part of the brain structure information. We perform extensive experiments based on the ADNI dataset and demonstrate the effectiveness of our method. Our method also outperforms the other competitive methods by using CNN for label identification. Furthermore, we enhance the clinical explanation of in-depth learning in clinical diagnosis through brain shielding experiments.

REFERENCES

[1] L. Minati, T. Edginton, M. G. Bruzzone, and G. Giaccone, "Current concepts in Alzheimer's disease: A multidisciplinary review," *Amer. J. Alzheimers Disease, Other Dementias*, vol. 24, no. 2, pp. 95–121, 2009.

[2] C. Patterson, *World Alzheimer Report 2018-the State of the Art of Dementia Research: New Frontiers*. London, U.K.: Alzheimer's Disease International, 2018.

[3] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 856–867, 2011.

[4] D. Zhang, D. Shen, and The Alzheimer's Disease Neuroimaging Initiative, "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease," *NeuroImage*, vol. 59, no. 2, pp. 895–907, 2012.

[5] X. Zhu, H.-I. Suk, and D. Shen, "A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis," *NeuroImage*, vol. 100, pp. 91–105, Oct. 2014.

[6] X. Zhu *et al.*, "A novel relational regularization feature selection method for joint regression and classification in ad diagnosis," *Med. Image Anal.*, vol. 38, pp. 205–214, 2017.

[7] H. I. Suk, S. W. Lee, D. Shen, and The Alzheimer's Disease Neuroimaging Initiative, "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis," *NeuroImage*, vol. 101, pp. 569–682, 2014.

[8] B. Lei, S. Chen, D. Ni, and T. Wang, "Discriminative learning for Alzheimer's disease diagnosis via canonical correlation analysis and multimodal fusion," *Frontiers Aging Neurosci.*, vol. 8, pp. 77–94, May 2016.

[9] L. Huang *et al.*, "Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based random forest," *Neurobiol. Aging*, vol. 46, pp. 180–191, Oct. 2016.

[10] K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, D. Rueckert, and The Alzheimer's Disease Neuroimaging Initiative, "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease," *NeuroImage*, vol. 65, pp. 167–175, Jan. 2013.

[11] I. Garali, M. Adel, S. Bourennane, and E. Guedj, "Region-based brain selection and classification on PET images for Alzheimer's disease computer aided diagnosis," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 1473–1477.

[12] S. Liu *et al.*, "Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, pp. 1132–1140, Apr. 2015.

[13] T. Tong, K. Gray, Q. Gao, L. Chen, D. Rueckert, and The Alzheimer's Disease Neuroimaging Initiative, "Multi-modal classification of Alzheimer's disease using nonlinear graph fusion," *Pattern Recognit.*, vol. 63, pp. 171–181, Mar. 2017.

[14] B. Jie, M. Liu, J. Liu, D. Zhang, and D. Shen, "Temporally constrained group sparse learning for longitudinal data analysis in Alzheimer's disease," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 238–249, Jan. 2017.

[15] C. Zhang, E. Adeli, T. Zhou, X. Chen, and D. Shen, "Multi-layer multi-view classification for Alzheimer's disease diagnosis," in *Proc. Assoc. Adv. Artif. Intell.*, 2018, pp. 4406–4413.

[16] B. Lei, P. Yang, T. Wang, S. Chen, and D. Ni, "Relational-regularized discriminative sparse learning for Alzheimer's disease diagnosis," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1102–1113, Apr. 2017.

[17] D. Shen, G. Wu, and H. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.

[18] I. Garali, M. Adel, S. Bourennane, and E. Guedj, "Histogram-based features selection and volume of interest ranking for brain PET image classification," *IEEE J. Transl. Eng. Health Med.*, vol. 6, 2018, Art. no. 2100212.

[19] R. Cuingnet *et al.*, "Automatic classification of patients with alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," *NeuroImage*, vol. 56, no. 2, pp. 766–781, May 2011.

[20] B. Dubois *et al.*, "Donepezil decreases annual rate of hippocampal atrophy in suspected prodromal Alzheimer's disease," *Alzheimers Dement*, vol. 11, no. 9, pp. 1041–1049, 2015.

[21] M. Liu, J. Zhang, P.-T. Yap, and D. Shen, "View-aligned hypergraph learning for Alzheimer's disease diagnosis with incomplete multi-modality data," *Med. Image Anal.*, vol. 36, pp. 123–134, Feb. 2017.

[22] R. Li *et al.*, "Deep learning based imaging data completion for improved brain disease diagnosis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervent*, 2014, pp. 305–312.

[23] D. Cheng and M. Liu, "Classification of Alzheimer's disease by cascaded convolutional neural networks using PET images," in *Proc. Int. Workshop Mach. Learn. Med. Imag.*, 2017, pp. 106–113.

[24] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis," *IEEE Trans. Biomed. Eng.*, to be published.

[25] H. I. Suk, S. W. Lee, D. Shen, and The Alzheimer's Disease Neuroimaging Initiative, "Deep ensemble learning of sparse regression models for brain disease diagnosis," *Med. Image Anal.*, vol. 37, pp. 101–113, Apr. 2017.

[26] T. Zhou, K. H. Thung, X. Zhu, and D. Shen, "Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis," *Hum. Brain Mapping*, vol. 44, no. 3, pp. 1001–1016, 2018.

[27] W. Yan, H. Zhang, J. Sui, and D. Shen, "Deep chronnectome learning via full bidirectional long short-term memory networks for MCI diagnosis," in *Proc. 21st Int. Conternational Conf. Med. Image Comput., Comput. Assist. Intervent.*, 2018, pp. 249–257.

[28] J. Islam and Y. Zhang, "Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks," *Brain Informat.*, vol. 5, no. 2, pp. 1–14, 2018. [Online]. Available: https://link.springer.com/content/pdf/10.1186%2Fs40708-018-0080-3.pdf

[29] M. Liu, D. Cheng, K. Wang, Y. Wang, and The Alzheimer's Disease Neuroimaging Initiative, "Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis," *Neuroinformatics*, vol. 16, nos. 3–4, pp. 295–308, Oct. 2018.
[30] M. Liu, D. Cheng, W. Yan, and The Alzheimer's Disease Neuroimaging Initiative, "Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images," *Front Neuroinform*, vol. 12, no. 35, p. 2, 2018.
[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Comput. Vis. Pattern Recognit.*, 2014, pp. 1–14.
[32] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[33] S. H. Lee, C. S. Chan, and P. Remagnino, "Multi-organ plant classification based on convolutional and recurrent neural networks," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4287–4301, Sep. 2018.
[34] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[35] A. G. J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM networks," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2005, pp. 2047–2052.
[36] C. Feng, A. Elazab, P. Yang, T. Wang, B. Lei, and X. Xiao, "3D convolutional neural network and stacked bidirectional recurrent neural network for Alzheimer's disease diagnosis," in *Proc. 1st Int. Workshop Predictive Intell. Med.*, 2018, pp. 138–146.
[37] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE Trans. Med. Imag.*, vol. 17, no. 1, pp. 87–97, Feb. 1998.
[38] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
[39] D. Shen and C. Davatzikos, "HAMMER: Hierarchical attribute matching mechanism for elastic registration," *IEEE Trans. Med. Imag.*, vol. 21, no. 11, pp. 1421–1439, Nov. 2002.
[40] C. Davatzikos, A. Genc, D. Xu, and S. M. Resnick, "Voxel-based morphometry using the RAVENS maps: Methods and validation using simulated longitudinal atrophy," *NeuroImage*, vol. 14, no. 6, pp. 1361–1369, 2001.
[41] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Comput. Vis. Pattern Recognit.*, 2014, pp. 1–15.
[42] A. Graves, "Generating sequences with recurrent neural networks," *Comput. Sci.*, 2013, pp. 1–43. [Online]. Available: https://arxiv.org/pdf/1308.0850.pdf
[43] B. Shi, Y. Chen, P. Zhang, C. D. Smith, and J. Liu, "Nonlinear feature transformation and deep fusion for Alzheimer's disease staging analysis," *Pattern Recognit.*, vol. 63, pp. 487–498, Mar. 2017.
[44] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *Proc. Deep Learn. Represent. Workshop (NIPS)*, 2014, pp. 1–9. [Online]. Available: https://arxiv.org/pdf/1412.3555v1.pdf
[45] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek, "Image classification with the fisher vector: Theory and practice," *Int. J. Comput. Vis.*, vol. 105, no. 3, pp. 222–245, 2013.
[46] M. A. Hearst, S. T. Dumais, E. Osman, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, Jul./Aug. 2008.
[47] N. J. Kabani, D. J. MacDonald, C. J. Holmes, and A. C. Evans, "3D anatomical atlas of the human brain," *NeuroImage*, vol. 7, no. 4, p. S717, 1998.

CHIYU FENG received the B.E. degree in communication engineering from Shenzhen University, in 2017, where he is currently pursuing the M.S. degree. His research interests include machine learning, pattern recognition, and medical image analysis.

AHMED ELAZAB received the Ph.D. degree in pattern recognition and intelligent system from the Shenzhen Institutes of Advanced Technology, University of Chinese Academy of Sciences, China, in 2017. He is currently an Assistant Professor with the Computer Science Department, Misr Higher Institute for Commerce and Computers, Mansoura, Egypt. He is also a Postdoctoral Fellow with the School of Biomedical Engineering, Shenzhen University, Shenzhen, China. He has authored or coauthored more than 30 peer-reviewed papers. He received the Best Paper Award of the 7th Cairo International Biomedical Engineering Conference 2014 (IEEE/EMB). He received two Outstanding Student Awards from the Shenzhen Institutes of Advanced Technology. He served as a Reviewer for prestigious peer-reviewed international journals. His current research interests include pattern recognition, medical image analysis, and computer-aided diagnosis.

PENG YANG received the B.E. degree in biomedical engineering from Northeastern University, in 2016. He is currently pursuing the Ph.D. degree with Shenzhen University. His research interests include machine learning, pattern recognition, and medical image analysis.

TIANFU WANG received the Ph.D. degree in biomedical engineering from Sichuan University, in 1997. He is currently a Professor with the School of Biomedical Engineering, and the Associate Chair of the Health Science Center, Shenzhen University, China. His research interests include ultrasound image analysis, medical image processing, pattern recognition, and medical imaging.

FENG ZHOU received the Ph.D. degree in engineering design from the G.W. Woodruff School of Mechanical Engineering, Georgia Tech, in 2014, under the supervision of Dr. R. Jiao. He joined the Department of Industrial and Manufacturing Systems Engineering, University of Michigan–Dearborn, in 2017. He is currently involved in a MCity project to understand the influence of various factors (traffic, age, trust, non-driving related tasks, takeover lead time, situational awareness, and warning effectiveness) on the takeover performance in highly automated driving. His research interests include physiological computing, sentiment analysis, and human factors issues of takeover control in highly automated driving.

HUOYOU HU is currently the Deputy Chief Physician of the Neurology Department, Shenzhen Second People's Hospital. He has been engaged in clinical work of neurology department for 11 years. He was a member of the China Stroke Society, Sleep Psychology Committee of Guangdong Society of Integrated Traditional Chinese and Western Medicine, Guangdong Society of Brain Development and Encephalopathy Prevention and Treatment, and Shenzhen Brain Neuroelectrophysiology Association, a Member and Secretary.

XIAOHUA XIAO received the Ph.D. degree from Zhongshan Medical University, in 2000. He has been engaged in clinical work of Neurology for more than 20 years and is good at diagnosis and treatment of epilepsy, vertigo, and cerebrovascular diseases. He is proficient in neuroelectrophysiological examination and is good at diagnosis and treatment of epilepsy, motor neuron disease, and other neuromuscular diseases by using neuroelectrophysiological examination technology. He has deep experience in diagnosis and treatment of benign paroxysmal positional vertigo and is good at using manual reduction to treat benign paroxysmal positional vertigo.

BAIYING LEI received the M.Eng. degree in electronics science and technology from Zhejiang University, China, in 2007, and the Ph.D. degree from Nanyang Technological University (NTU), Singapore, in 2013. She is currently an Associate Professor with the School of Biomedical Engineering, Shenzhen University, China. Her current research interests include medical image analysis, machine learning, digital watermarking, and signal processing.