Progress In Electromagnetics Research, Vol. 171, 21–34, 2021

Multimodal 2.5D Convolutional Neural Network for Diagnosis of Alzheimer's Disease with Magnetic Resonance Imaging and Positron Emission Tomography

Xuyang Zhang[1], Weiming Lin[1, 2, *], Min Xiao[1, 2], and Huazhi Ji[1]

Abstract—Alzheimer's disease (AD) is a degenerative disease of the nervous system that often occurs in the elderly. As magnetic resonance imaging (MRI) and positron emission tomography (PET) reflect the brain's anatomical changes and functional changes caused by AD, they are often used to diagnose AD. Multimodal fusion based on these two types of images can effectively utilize complementary information and improve diagnostic performance. To avoid the computational complexity of the 3D image and expand training samples, this study designed an AD diagnosis framework based on a 2.5D convolutional neural network (CNN) to fuse multimodal data. First, MRI and PET were preprocessed with skull stripping and registration. After that, multiple 2.5D patches were extracted within the hippocampus regions from both MRI and PET. Then, we constructed a multimodal 2.5D CNN to integrate the multimodal information from MRI and PET patches. We also utilized a training strategy called branches pre-training to enhance the feature extraction ability of the 2.5D CNN by pre-training two branches with corresponding modalities individually. Finally, the results of patches are used to diagnose AD and progressive mild cognitive impairment (pMCI) patients from normal controls (NC). The experiments were conducted with the ADNI dataset, and accuracies of 92.89% and 84.07% were achieved in the AD vs. NC and pMCI vs. NC tasks. The results are much better than using single modality and indicate that the proposed multimodal 2.5D CNN could effectively integrate complementary information from multi-modality and yield a promising AD diagnosis performance.

1. INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disease that often occurs in the elderly. Patients generally suffer from memory loss, agnosia, aphasia, and other symptoms. The disease will eventually lead to the death of patients. The deteriorating from normal people to AD is slow and gradual. The AD process may last for 3 to 11 years [1]. The cause of AD is unknown, and treatments do not have much effect in the late stage. However, if AD can be detected at an early stage, timely treatments would delay the AD progression. Therefore, early diagnosis of AD is crucial to delay the deterioration or even avoid the disease. Mild cognitive impairment (MCI) is considered a transitional stage between normal control (NC) and dementia [2]. MCI is usually partitioned into stable MCI (sMCI) and progressive MCI (pMCI). Among them, pMCI refers to MCI patients who convert to AD within three years, and sMCI refers to MCI patients who do not convert [3]. Therefore, the diagnosis of pMCI is also of great interest for the early detection of patients at high risk of AD and effective treatments to prevent further deterioration of AD.

AD could cause cerebral atrophy, particularly in the hippocampus. Magnetic Resonance Imaging (MRI) is non-invasive structural imaging and can detect hippocampus atrophy [4, 5]. In addition to

Received 11 May 2021, Accepted 3 August 2021, Scheduled 12 August 2021

* Corresponding author: Weiming Lin (linwm@xmut.edu.cn).

[1] School of Opto-Electronic and Communication Engineering, Xiamen University of Technology, Xiamen 361024, China. [2] Fujian Key Laboratory of Communication Network and Information Processing, Xiamen University of Technology, Xiamen 361024, China.

the brain's structure, AD also affects the brain's glucose metabolism. Fluorodeoxyglucose Positron Emission Tomography (FDG-PET) is functional imaging that can discover the abnormal brain glucose metabolism [6, 7]. Combining complementary information from MRI and PET images can better detect brain lesions caused by AD [8].

In recent years, deep learning applications in the medical image have received more and more attention. Convolutional neural network (CNN) is a network model of deep learning. It can capture features at different levels during the training process without manually extracting features based on prior knowledge. Previous studies have combined CNN with MRI or PET images used for disease diagnosis, tissue segmentation, cell or tumor detection and classification, and achieved excellent performances. For example, Mzoughi et al. proposed a multi-scale 3D CNN to classify glioma brain tumors into low-grade gliomas (LGG) and high-grade gliomas (HGG) using MRI which achieved an accuracy of 96.49% and outperformed the supervised and unsupervised state-of-the-art approaches [9]. Gao et al. implemented a triple cascaded CNN for automatic rat brain image segmentation of MRI [10]. Their proposed method achieved a mean dice score coefficient (DSC) of 0.965 and 0.927 in the whole brain and cerebellum segmentation, which improved performances compared with atlas-based methods. Zhang et al. proposed a deep learning framework based on symmetricity-driven CNN of PET for epilepsy diagnosis [11]. The dice coefficient of the framework in detecting epileptic foci is higher than visual assessment and statistical parametric mapping (SPM) analysis. Zhang et al. proposed a multi-scale Mask Region-Based Convolutional Neural Network (Mask R-CNN) combined with PET to detect lung tumors effectively and precisely [12]. Besides 3D network, 2.5D network is also a good way to analyze 3D images. Zheng et al. designed a novel 2.5D CNN in an automatic pancreas segmentation framework for capturing the cross-slice information of multiple feature levels [13]. The novel 2.5D CNN helps enhance the dice similarity coefficient, sensitivity, and specificity to 86.21%, 87.49%, and 85.11%, which are state-of-the-art performances validated in 3D CT scans. Kitrungrotsakul et al. cascaded a 2.5D faster region-based convolutional neural network (Faster R-CNN) and a convolutional long short-term memory (CLSTM) network to detect mitotic cells in microscopic images, and yielded better performance than other deep learning methods [14].

In addition, numerous studies have used CNN to diagnose AD. Li et al. combined 3D CNN and 2D CNN to form a multi-channel cascaded CNN to learn the combined hierarchical feature of hippocampal shape and asymmetry from preprocessed MRI for AD and MCI diagnosis [15]. Jo et al. proposed a deep learning framework by combining 3D CNN and LRP algorithms [16]. They used the deep learning framework with PET to detect informative features for AD classification. Gao et al. explored the significance and impact of the application of CNN to the CT brain image classification [17]. Combining multimodal medical data for joint classification can effectively improve the model's performance [18, 19]. Huang et al. utilized 3D CNN to fuse MRI and PET information for AD diagnosis or prognosis [20]. Cheng and Liu constructed multi-level CNNs based on 3D CNNs combined with 2D CNNs to gradually learn and fuse the multi-modality features from MRI and PET [21]. Zhou et al. proposed a sparse-response deep belief network (SR-DBN) combined with an extreme learning machine (ELM) to distinguish NC, MCI, and AD [22]. The SR-DBN extracted features from MRI and PET images, and the ELM performed as a classifier. Zhu et al. proposed a novel descending dimension method to utilize the variability, sparsity, and low-rankness of MRI and PET data for AD classification [23]. Lin et al. took MRI, PET, cerebrospinal fluid (CSF) biomarkers, and gene data as input [24]. Then they implemented an ELM-based grading method to fuse multi-modality features for the prediction of MCI-to-AD conversion.

Although previous studies have developed multimodal 3D CNN models for AD diagnosis, there are still some problems: training the 3D CNN directly on the 3D images requires high calculation and memory costs, particularly for the high-resolution MRI images, and the training sets are usually small, which hinders the efficient utilization of CNN. Extracting 2.5D patches from 3D images can significantly expand the number of samples. In this study, we propose a framework based on the multimodal 2.5D convolutional neural network (CNN) to fuse multimodal information from MRI and PET for AD and pMCI diagnosis. The contributions of our work are as follows:

(1) The counts of 3D MRI and PET are usually small. The multiple 2.5D patches strategy can greatly expand training samples and improve the training efficiency of CNN.

(2) In this study, we established a multimodal 2.5D CNN, which can fuse the complementary

information from MRI and PET 2.5D patches.

(3) We used a branches pre-training strategy to enhance the feature extraction ability of CNN. Compared with the network without branches pre-training, the proposed method has better classification performance.

The rest of this paper is summarized as follows. Section 2 describes the MRI and PET data process of the proposed framework based on multimodal 2.5D CNN. Section 3 presents the experiment results and comparisons with other studies. Section 4 concludes this paper.

2. MATERIAL AND METHOD

The flowchart of the proposed framework is shown in Figure 1, which includes four steps: 1) All MRI and PET images are preprocessed with skull stripping, non-rigid registration, and age correction. 2) Then, 151 points in the hippocampus regions are chosen. For each point, 2.5D patches, which is consisted of 2D patches of three axes, are extracted from both MRI and PET images to form pairwise patches. 3) Pairwise patches are fed into the multimodal 2.5D CNN, in which there are two branches at the front of CNN to process MRI and PET patches respectively, followed by a fusion network with the ResNet as the main structure, and fully connected layers and the softmax layer as a classifier. To improve performance, the branches of the multimodal CNN have been individually pre-trained with corresponding modal data before the training of the whole multimodal CNN. 4) With the multimodal network outputs classification probability for each pairwise patches, a final diagnosis is made based on the mean classification probabilities of 151 pairwise patches from one subject.

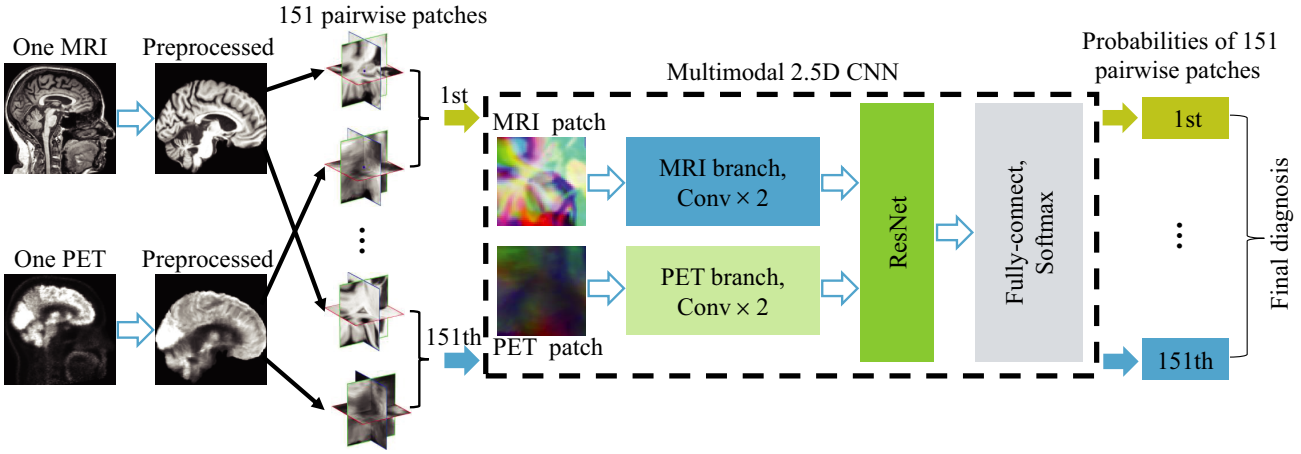

Figure 1. The flowchart of diagnosis.

2.1. Data Preprocessing

The MRI and PET data used in this work were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, publicly available on the website (www.loni.ucla.edu/ADNI). As there are many inter-subjects variables among individuals, which would interfere with the diagnosis, it is necessary to perform skull stripping and registration for both MRI and PET. Firstly, we used the BrainExtraction Tool (BET) in the FSL library (https://fsl.fmrib.ox.ac.uk/fslwiki/FSL) for skull stripping. Then, we used the Image Registration Toolkit (IRTK) of Imperial College London for registration. All MRI images were registered to the MNI152 template, and PET images were registered to corresponding MRI images. Before preprocessing, MRI images had different sizes, such as (176, 240, 256), (160, 192, 192), (180, 256, 256), after registration, all preprocessed MRI and PET images had the size of (182, 218, 182), which was the same as template MNI152. The images before and after skull stripping and registration are shown in Figure 2.

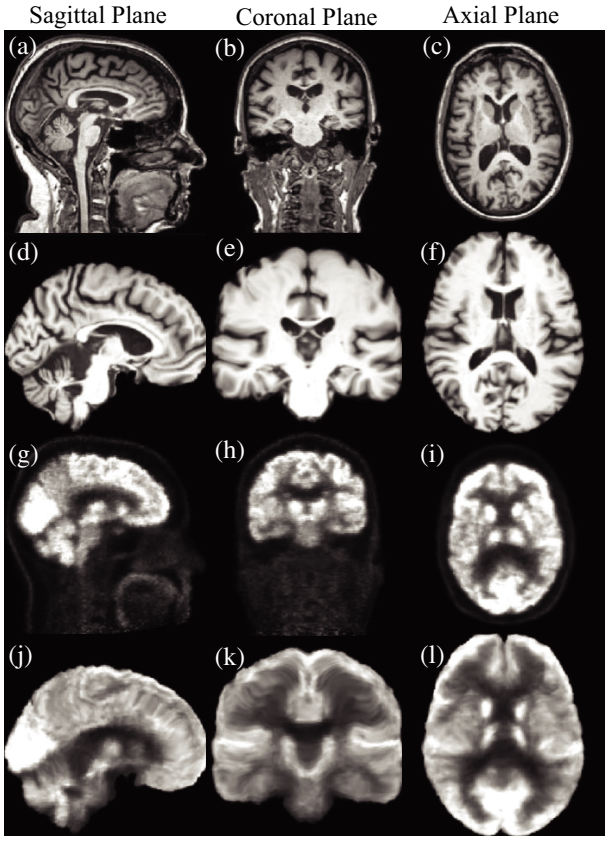

Figure 2. Comparison of the original image and preprocessed image. (a)–(c) Original MRI. (d)–(f) Preprocessed MRI. (g)–(i) Original PET. (j)–(l) Preprocessed PET.

2.2. Age Correction for MRI

As natural brain shrinkage occurs in normal aging people [25], natural age-related atrophy is easily confused with the pathological atrophy caused by AD. In order to eliminate age-related atrophy interference, it is necessary to calculate the effects of age on brain structure and reduce the interference [26]. As PET mainly reflects function information but not structural atrophy, we only performed the age correction on MRI.

Suppose that there are N MRI images belonging to the NC group, and the value of mth voxel of the ith image is denoted as y_{im}, then the vector consisting of N values from the mth voxel from N images is $\mathbf{y}_m = [y_{1m}, y_{2m}, ..., y_{Nm}]^T$, and the vector consisting of ages of N images is $\mathbf{x} = [x_1, x_2, ..., x_n, ..., x_N]^T$. After performing linear regression on $\mathbf{y}_m$ and $\mathbf{x}$, the fitted linear model can be obtained as $\mathbf{y}_m = a_m\mathbf{x} + b_m$. The coefficient a_m can be seen as the effect of normal aging on the mth voxel. We performed the process mentioned above on all voxels of MRI images and obtained the aging effect coefficients for all voxels. Finally, we used formula $y_{nm}^{new} = a_m(C - x_n) + y_{nm}$ to align the aging effect to the same age C [27]. In the AD vs. NC and pMCI vs. NC classification experiments, the average ages of all samples are 75.44 and 74.67. So all samples were aligned to the age of 75 years old.

2.3. Extraction of 2.5D Patches

To improve the efficiency of the convolutional neural network (CNN) and expand the number of training samples, we extracted 2.5D patches from the 3D images. As shown in Figure 3, a point in the 3D image was selected as the center to intercept three 2D grayscale patches of 48×48 from three axial directions, corresponding to the transverse plane, coronal plane, and sagittal plane. These three 48×48

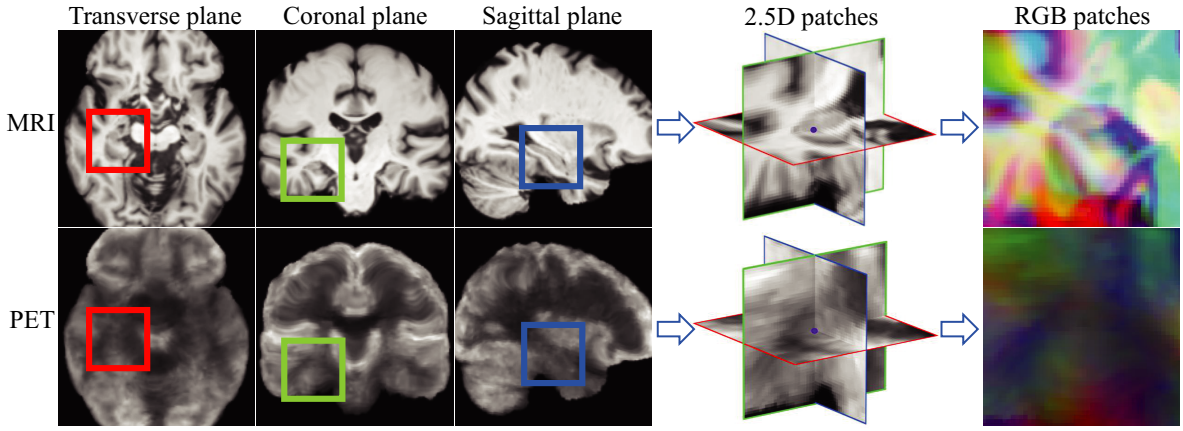

Figure 3. The demonstration of 2.5D patch extraction in the hippocampus region.

patches were assigned to three channels to form a $48 \times 48 \times 3$ patch [28], which we called a 2.5D patch. Obviously, a single 2.5D patch cannot obtain enough information from 3D MRI and PET images. Therefore, we extracted multiple 2.5D patches from different points within hippocampus regions, and the final diagnosis of AD could be made with the joint diagnosis of these 2.5D patches. This strategy could increase the number of training samples and improve the training efficiency of CNN. Besides, the multiple patches joint diagnosis can improve the diagnosis stability and reliability.

The selection of these center points of patches has three constraints: (1) The points must be within the left or right hippocampus regions because hippocampi are highly correlated with AD [29]. (2) The distance between two points should be greater than two voxels to ensure that all points are distributed evenly. (3) All points were chosen randomly.

With these selection constraints, we got about 150 to 160 points, then we chose an odd number 151. Figure 4 shows the distribution of a part of these points. We extracted these 151 2.5D patches from one 3D image. Although these points were mainly distributed inside the hippocampus regions, the size of patches was 48×48, which ensured that these patches could cover the hippocampus and a large surrounding area containing enough information.

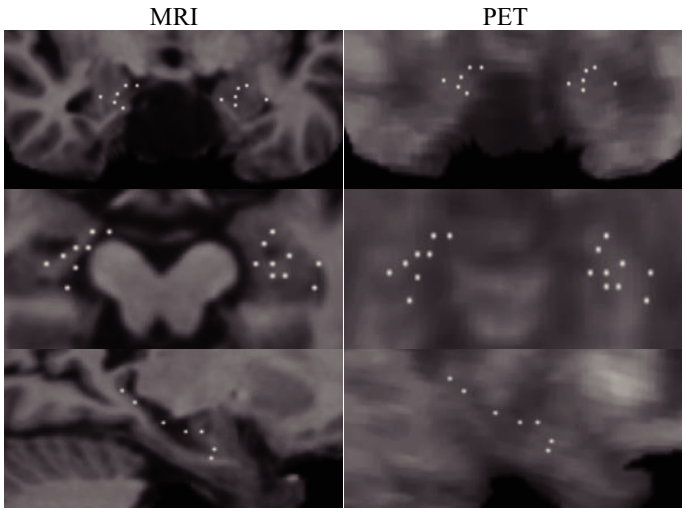

Figure 4. Distribution of part points (white color points) in MRI and PET images.

2.4. Multimodal 2.5D Convolutional Neural Network

The multimodal 2.5D convolutional neural network (CNN) built in this study is shown in Figure 5. The network had two branches in the front section, which input 2.5D patches of MRI and PET, respectively. Each branch had five layers consisting of a zero-padding layer, two convolution layers, and two max-pooling layers. These five layers were used to capture the features of MRI and PET. Each branch output 256 feature maps and a total of 512 feature maps of the two branches were concatenated. To enable the multimodal network to effectively integrate the multimodal features without increasing the learning parameters too much, we used 256 convolution kernels with the size of 1×1 in the fusion layer to reduce the number of feature maps to 256 [30, 31].

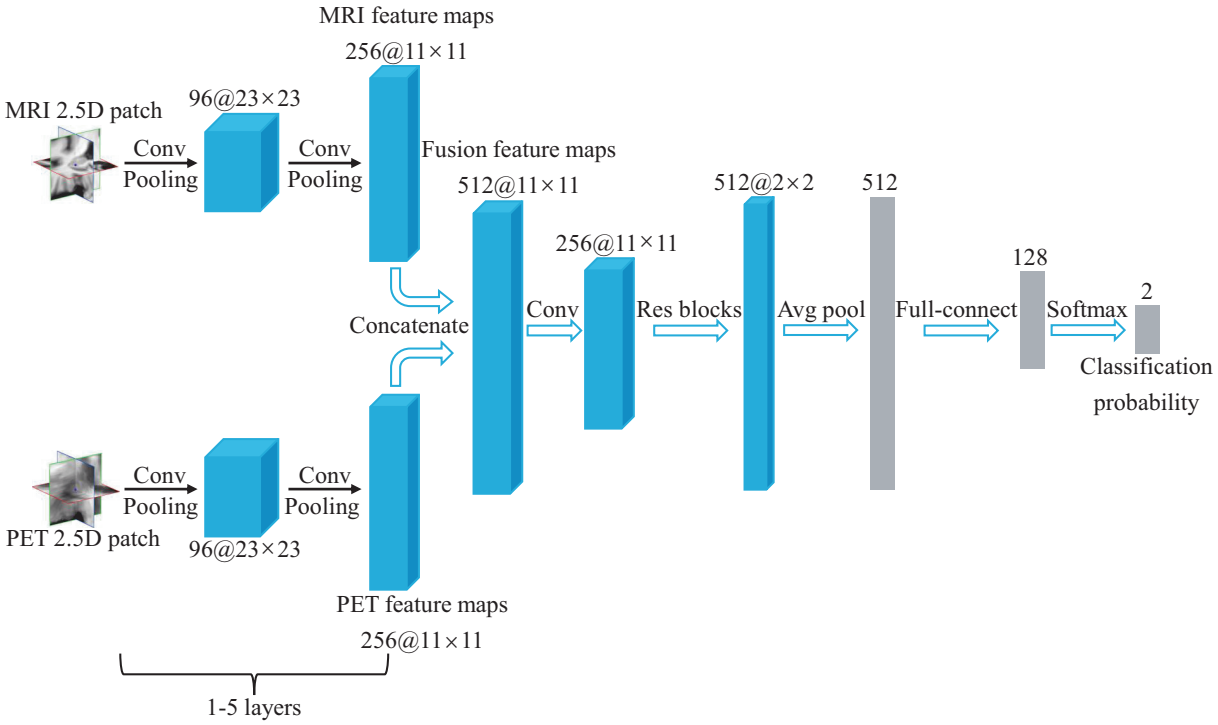

Figure 5. The flowchart of multimodal 2.5D CNN.

The subsequent layers take the residual blocks as the main structure, as shown in Figure 6. The ResNet composed of residual blocks can solve the difficulty of training the deep network [32], and the Batch Normalization in ResNet can effectively alleviate the gradient vanishing and gradient exploding problems [33]. The residual blocks in multimodal CNN outputted 512 feature maps with the size of 2×2. Then, the global average pooling layer was followed by the three full connection layers and dropout layers, and the final layer is the softmax layer. For each pairwise patches of MRI and PET,

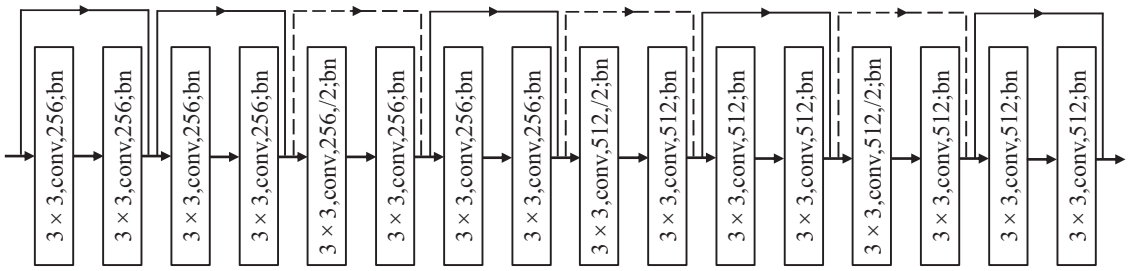

Figure 6. Residual blocks in multimodal 2.5D CNN.

the network output a classification probability, so we obtained 151 probabilities for one subject. All 151 probabilities were averaged, and then the final label of the subject was determined.

2.5. Branches Pre-Training Strategy

The multimodal CNN utilized a training strategy that is similar to transfer learning. Before training the multimodal network, we pre-trained the two branches individually. Two single modal networks for MRI and PET patches were built based on AlexNet. These two single modal networks were trained with MRI and PET patches individually. Then the trained coefficients of the first five layers of single modal networks were assigned to the branches of the multimodal 2.5D CNN. After that, the parameters of these two branches were frozen and fixed. They were not updated during the training of the multimodal network. This training strategy could improve the feature extraction ability of the two branches, which is conducive to the network to integrate complementary information and improve classification performance.

3. EXPERIMENTAL RESULTS AND ANALYSIS

3.1. Experiment Dataset

After data preprocessing and removing some incorrectly segmented images, we got MRI and PET images of 376 AD, 170 pMCI, and 499 NC samples. The demographic details of these samples used in AD vs. NC and pMCI vs. NC experiments are listed in Table 1 and Table 2. In Table 2, as the number of pMCI samples was 170, to avoid sample bias in the classification of pMCI vs. NC, we selected only the baseline images and got 243 NC samples and 239 AD samples.

Table 1. Summary of the dataset used in AD vs. NC.

Disgnosis	Number	Age	Gender (M/F)	MMSE	CDR-SB
NC	499	75.58 ± 5.91	272/227	29.07 ± 1.17	0.12 ± 0.36
AD	376	75.25 ± 7.32	174/202	22.66 ± 3.85	5.06 ± 2.29

Table 2. Summary of the dataset used in the pMCI vs. NC.

Disgnosis	Number	Age	Gender (M/F)	MMSE	CDR-SB
NC	243	74.65 ± 6.12	123/120	29.04 ± 1.17	0.10 ± 0.28
pMCI	170	73.74 ± 7.41	95/75	26.69 ± 2.23	2.17 ± 1.02
AD	239	74.56 ± 7.60	122/117	23.10 ± 3.58	4.65 ± 2.25

3.2. Experiment Settings

All the networks mentioned above were implemented with TensorFlow 2.0 and Keras 2.3.1. The experiments were conducted with NVIDIA GTX1080 GPU. We used the ADAM optimizer to train networks, and the learning rate was set to 0.0002. When we trained single modal networks, the batch size was set to 302. When we trained multimodal networks, the batch size was set to 151. The number of epochs was set to 40. Performance was evaluated by the following indicators: accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the receiver operating characteristic curve (AUC). In order to avoid sampling bias, all experiments used the five-fold cross-validation and repeated 30 times to calculate the average performance. During validation, the datasets were divided into training and testing sets according to the patient IDs instead of image IDs to ensure that images of the same person only appear in one set. We used 20% of subjects' images as the testing set in each fold and used the others as the training set. We evaluated the proposed algorithm on the classification tasks of AD vs.

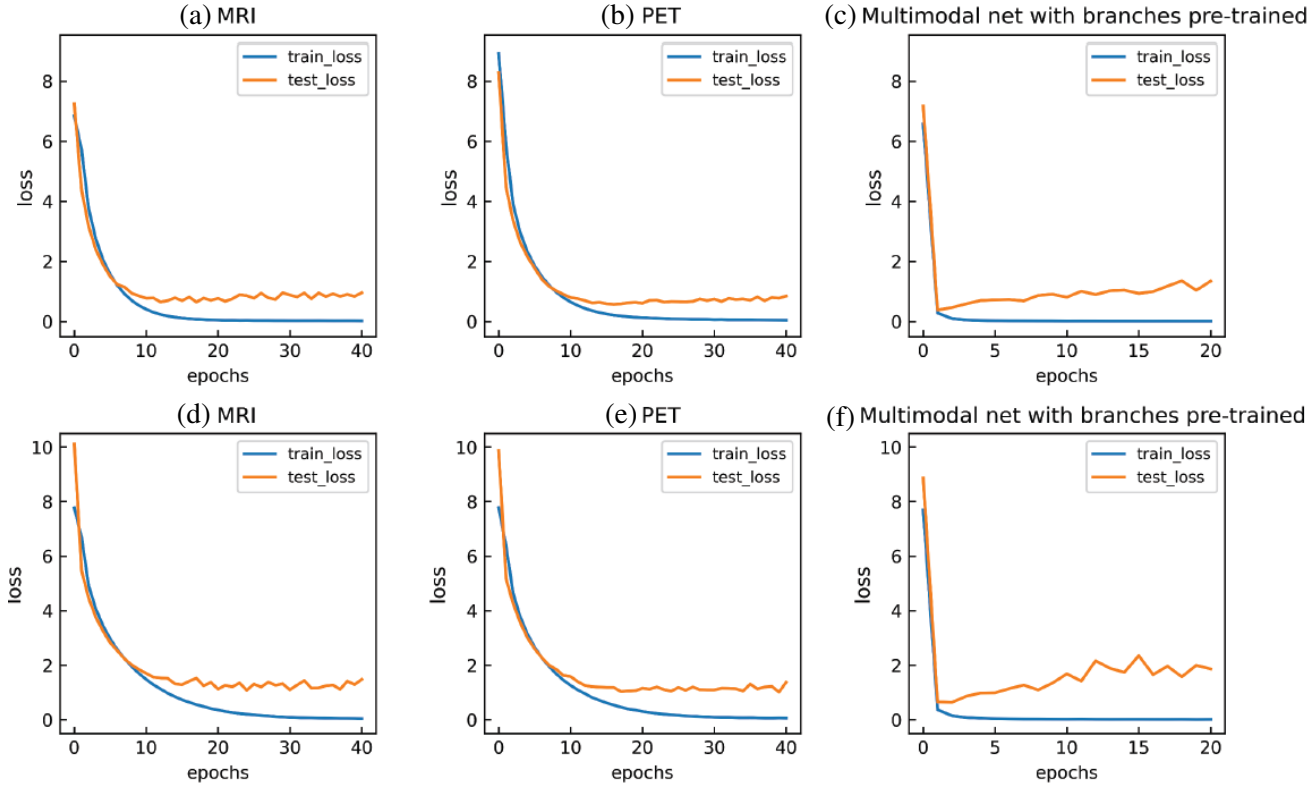

Figure 7. Loss curves. (a)–(c) show the loss curves for AD vs. NC classification. (d)–(f) show the loss curves for pMCI vs. NC classification. (a) and (d) are the loss curves of MRI single modal network. (b) and (e) are the loss curves of PET single modal network. (c) and (f) are the loss curves of multimodal network.

NC and pMCI vs. NC. The training and validation loss curves with the above parameter settings are shown in Figure 7.

In Figure 7, we can find that, as the branches were pre-trained, the loss curves of the multimodal network descended rapidly. The reason may be that the pre-trained branches already had the abilities to extract AD-related features, which can accelerate the training of the multimodal network. Therefore, the number of epochs of the multimodal network training was set to 20.

3.3. Performance of AD vs. NC Classification

To verify the improvements of the multimodal CNN and the branches pre-training strategy, we compared the proposed method with two single modal CNNs (for MRI and PET respectively) and the multimodal CNN without pre-trained branches. The performances are listed in Table 3, and box plots are shown in Figure 8.

As we can find from Table 3 and Figure 8, compared with the single modal CNNs using MRI or PET, the accuracy of multimodal CNN increased by 3.52% and 3.23%. In the multimodal CNN training process, the training strategy using branches pre-training improved the accuracy by 2.73% compared to direct training, which did not pre-train branches. It can be seen that the pre-trained branches can improve the feature extraction ability and help the subsequent fusion network to integrate the complementary information more effectively and improve the diagnostic performance.

3.4. Performance of pMCI vs. NC Classification

Firstly, we trained and validated models with pMCI vs. NC samples, similar to the AD vs. NC task. The results are listed in Table 4. According to the work of Huang et al. [20], training the pMCI vs.

Table 3. Comparison of classification performances on AD vs. NC.

Methods	ACC	SEN	SPE	AUC
MRI	$89.37 \pm 0.69\%$	$85.85 \pm 1.47\%$	$92.15 \pm 1.17\%$	$95.45 \pm 0.35\%$
PET	$89.66 \pm 0.66\%$	$83.93 \pm 1.68\%$	$94.13 \pm 1.31\%$	$95.09 \pm 0.41\%$
Without Branches Pre-trained	$90.16 \pm 0.62\%$	$88.05 \pm 2.28\%$	$91.79 \pm 2.44\%$	$96.39 \pm 0.34\%$
Branches Pre-trained	$\mathbf{92.89 \pm 0.62}\%$	$\mathbf{88.55 \pm 1.09}\%$	$\mathbf{96.27 \pm 0.92}\%$	$\mathbf{97.15 \pm 0.28}\%$

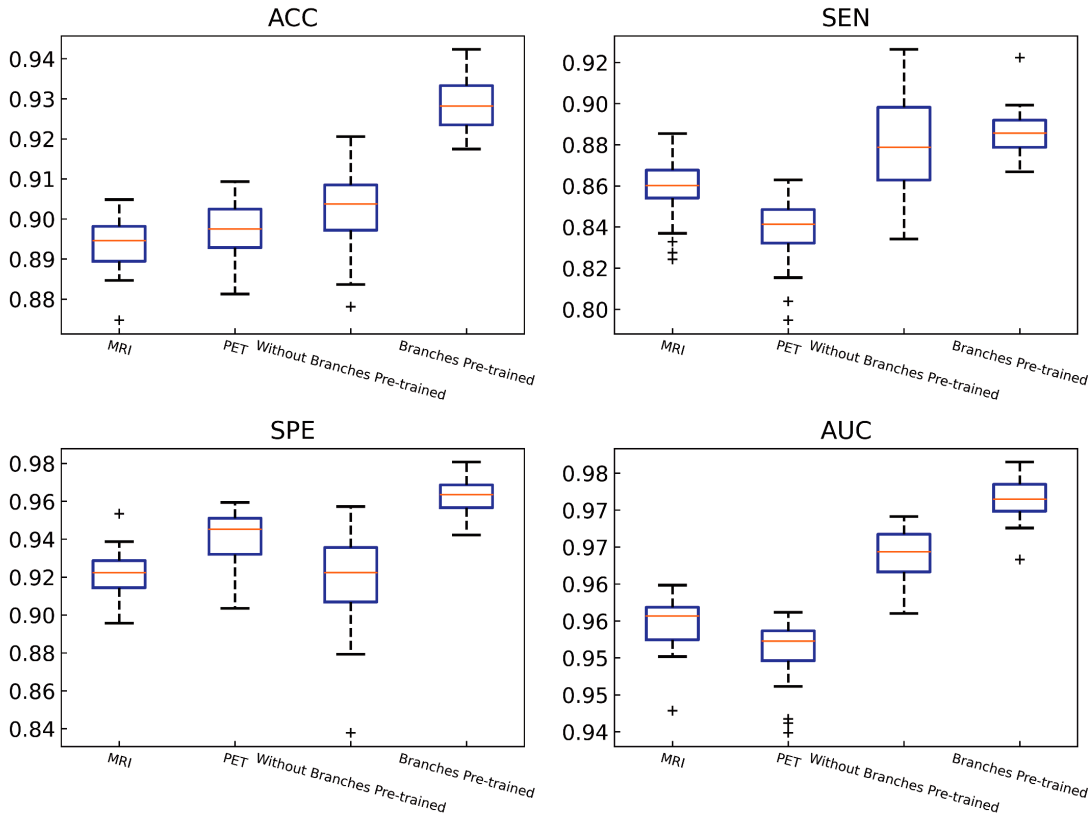

Figure 8. Performances of AD vs. NC classification. In each box, the central mark is the median, the edge of the box are 25th and 75th percentiles, outliers are denoted by +.

NC model with AD and NC could bring higher performance. Therefore, we used data of AD and NC to train the networks. Then we validated the networks with pMCI vs. NC. The results are listed in Table 4, and box plots are shown in Figure 9. The results showed that when AD and NC were used as the training set, the accuracy increased by 2.64% compared to that of pMCI and NC. From Figure 8 and Figure 9, we can find that the range of box plots decreased when the pre-trained branches were used, indicating that the branches pre-training strategy can not only improve the performances but also have more stable performances.

3.5. Comparison with Other Methods

In this section, we compared the proposed method with those published in previous literature. We first compared our method with multi-modality models based on 3D CNN [20]. Huang et al. directly used 3D CNN for training and testing with the hippocampus 3D patches of MRI and PET. They tried different

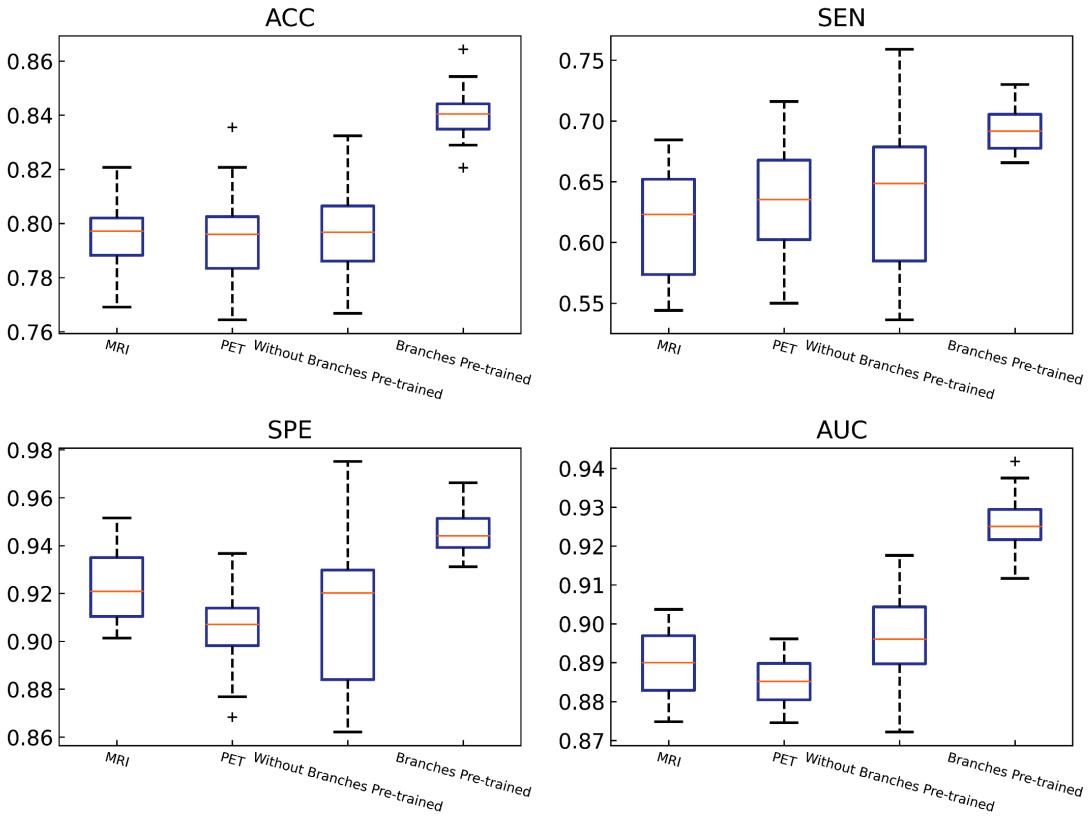

Figure 9. Performances of pMCI vs. NC classification.

training strategies to improve the diagnostic performance of the model. Table 5 shows the comparison between the proposed method and Huang et al. in single modality and multi-modality. Compared with the multi-modality 3D CNN in the AD vs. NC task, the accuracy of our model is higher by 2.79%, and the AUC is higher by 6.31%.

Moreover, Li et al. [15] used MRI single modality CNN to analyze hippocampus shape and asymmetry to assist AD diagnosis. Zhou et al. [22] extracted features based on a deep belief network and used an extreme learning machine for AD classification. Zhu et al. [23] used a low-rank dimensionality reduction method to select AD-related features in MRI and PET for identification and diagnosis. As

Table 4. Comparison of classification performances on pMCI vs. NC.

Training Set	Testing Set	Methods	ACC	SEN	SPE	AUC
pMCI/NC	pMCI/NC	MRI	$79.85 \pm 1.23\%$	$59.95 \pm 2.74\%$	$93.69 \pm 1.66\%$	$87.99 \pm 1.11\%$
		PET	$76.28 \pm 1.65\%$	$56.83 \pm 4.35\%$	$89.87 \pm 1.82\%$	$84.30 \pm 1.24\%$
		Without Branches Pre-trained	$73.34 \pm 2.11\%$	$54.20 \pm 9.83\%$	$86.40 \pm 5.92\%$	$81.26 \pm 2.04\%$
		Branches Pre-trained	$81.43 \pm 1.21\%$	$63.65 \pm 2.95\%$	$93.64 \pm 1.35\%$	$88.76 \pm 1.12\%$
AD/NC	pMCI/NC	MRI	$79.46 \pm 1.23\%$	$61.35 \pm 4.24\%$	$92.32 \pm 1.50\%$	$89.04 \pm 0.78\%$
		PET	$79.55 \pm 1.58\%$	$63.98 \pm 4.72\%$	$90.61 \pm 1.66\%$	$88.56 \pm 0.61\%$
		Without Branches Pre-trained	$79.67 \pm 1.81\%$	$63.46 \pm 6.29\%$	$91.07 \pm 3.12\%$	$89.62 \pm 1.17\%$
		Branches Pre-trained	$\mathbf{84.07 \pm 0.97\%}$	$\mathbf{69.28 \pm 1.75\%}$	$\mathbf{94.64 \pm 0.92\%}$	$\mathbf{92.61 \pm 0.71\%}$

Table 5. Comparison of the proposed method and Huang's multi-modality method.

Methods	Modality	AD vs. NC		pMCI vs. NC	
		ACC	AUC	ACC	AUC
Huang et al. [20]	MRI	81.19%	83.67%	–	–
	PET	89.11%	92.69%	–	–
	Both	90.10%	90.84%	**87.46%**	87.61%
Proposed method	MRI	89.37%	95.45%	79.85%	87.99%
	PET	89.66%	95.09%	76.28%	84.40%
	Both	**92.89%**	**97.15%**	84.07%	**92.61%**

Table 6. Comparison of the proposed method and published diagnosis methods.

Studies	Modality	AD vs. NC		pMCI vs. NC	
		ACC	AUC	ACC	AUC
Li et al. [15]	MRI	85.90%	88.40%	–	–
Elahifasaee et al. [34]	MRI	90.41%	93.83%	84.29%	83.54%
Oh et al. [35]	MRI	86.60%	–	77.37%	–
Salvatore et al. [36]	MRI	76.00%	–	72.00%	–
Liu et al. [7]	PET	91.20%	95.30%	–	–
Jo et al. [16]	PET	90.80%	–	–	–
Huang et al. [20]	MRI + PET	90.10%	90.84%	**87.46%**	87.61%
Zhou et al. [22]	MRI + PET	91.68%	87.00%	–	–
Zhu et al. [23]	MRI + PET	91.70%	94.30%	–	–
Liu et al. [37]	MRI + PET	**93.26%**	95.68%	82.95%	88.43%
Proposed method	MRI + PET	92.89%	**97.15%**	84.07%	**92.61%**

shown in Table 6, in the pMCI vs. NC task, the accuracy of the proposed method was slightly lower than that of Farzaneh et al. [34] and Huang et al. [20]; however, the AUC of the proposed method was the highest. Besides, the accuracy of the proposed method was higher than these two methods in the AD vs. NC task. In both tasks, the proposed method achieved favorable values in ACC, and highest values in AUC, showing promising diagnostic performance.

4. DISCUSSION

In this work, we studied the multimodal 2.5D CNN-based approach by extracting multiple 2.5D patches from MRI and PET images to train networks for AD and pMCI diagnosis. Because the change of pMCI relative to NC is more subtle than that of AD, it is more challenging to distinguish pMCI from NC. However, the pMCI is the patients at high risk of converting into AD. The detection of pMCI is more meaningful than AD diagnosis and has a higher reference value and clinical significance.

As shown in Table 3, in the AD vs. NC task, the diagnostic accuracies of the multimodal network without branches pre-trained were only 0.79% and 0.50% higher than that of two single modality networks. In order to extract multimodal features more effectively, we adopted the branches pre-training strategy by pre-training two single modal CNNs and assigning the parameters of 1–5 layers to the two branches of the multimodal 2.5D CNN. With the branches that could capture features more effectively from MRI and PET, the accuracy was raised to 92.89%, which was 3.52% and 3.23% higher than that of single modal networks and 2.73% higher than the multimodal network without branches pre-trained. We supposed that the reason might be that during training the multimodal networks without pre-trained

branches, the error needs to be equally divided into MRI and PET branches during the backpropagation process. However, the error caused by different modalities may be different or unbalance, so it cannot train the parameters of each branch well. The branches were trained more targeting when the branches were trained individually, so that the branches could better extract the features from the corresponding modalities and made it more effective to integrate multimodal information in the subsequent layers.

In the pMCI vs. NC task, when we used pMCI and NC to train the model, the accuracy was only 81.43%. Inspired by the work in Huang et al. [20], which used the AD and NC trained model to distinguish pMCI and NC, we used the model trained with AD and NC data to diagnose pMCI vs. NC. The accuracy reached 84.07%, which was improved by 2.64% compared to the model trained with pMCI and NC. We supposed that the reason might be that the anatomical and functional changes between the pMCI and NC are less noticeable than those between AD and NC. Therefore, it is more difficult for the model to learn valuable features about anatomical and functional changes from pMCI and NC than AD and NC.

5. CONCLUSION

In this study, we propose a framework based on multimodal 2.5D CNN for AD diagnosis and prognosis using MRI and PET images. The accuracy of models reached 92.89% for classification of AD vs. NC and 84.07% for classification of pMCI vs. NC. The results indicate that the multimodal 2.5D CNN, which utilized the training strategy of branches pre-training, can effectively integrate the features and complementary information among modalities. Compared with the single modal CNNs and multimodal CNN without branches pre-trained, the proposed model can better diagnose AD and pMCI. However, the proposed method requires preprocessing, such as segmentation and registration, which limits the end-to-end application of the 2.5D CNN model. In our future work, we will optimize this model to remove these preprocessing steps and further improve the performance.

ACKNOWLEDGMENT

This work was supported by the Foundation of Educational and Scientific Research Projects for Young and Middle-aged Teachers of Fujian Province [grant numbers JAT200471, JAT190674, JAT190677]; and the High-level Talent Project of Xiamen University of Technology [grant numbers YKJ20013R, YKJ17021R].

REFERENCES

1. Todd, S., S. Barr, M. Roberts, and A. P. Passmore, "Survival in dementia and predictors of mortality: A review," *International Journal of Geriatric Psychiatry*, Vol. 28, No. 11, 1109–1124, 2013.
2. Markesbery, W. R. and M. A. Lovell, "Neuropathologic alterations in mild cognitive impairment: A review," *Journal of Alzheimer's Disease*, Vol. 19, No. 1, 221–228, 2010.
3. Tong, T., Q. Gao, R. Guerrero, C. Ledig, and D. Rueckert, "A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease," *IEEE Transactions on Biomedical Engineering*, Vol. 64, No. 1, 1–1, 2016.
4. Eskilden, S. F., P. Coupé, D. García-Lorenzo, V. Fonov, J. C. Pruessner, and D. L. Collins, "Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning," *Neuroimage*, Vol. 65, 511–521, 2013.
5. Tong, T., R. Wolz, Q. Gao, and R. Guerrero, "Multiple instance learning for classification of dementia in brain MRI," *Medical Image Analysis*, Vol. 18, No. 5, 808–818, 2014.
6. Drzezga, A., D. Altomare, C. Festari, J. Arbizu, S. Orini, K. Herholz, P. Nestor, F. Agosta, F. Bouwman, and F. Nobili, "Diagnostic utility of 18F-Fluorodeoxyglucose positron emission tomography (FDG-PET) in asymptomatic subjects at increased risk for Alzheimer's disease," *European Journal of Nuclear Medicine and Molecular Imaging*, Vol. 45, No. 9, 1487–1496, 2018.

7. Liu, M., D. Cheng, and W. Yan, "Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images," *Frontiers in Neuroinformatics*, Vol. 12, 35, 2018.
8. Alberdi, A., A. Aztiria, and A. Basarab, "On the early diagnosis of Alzheimer's disease from multimodal signals: A survey," *Artificial Intelligence in Medicine*, Vol. 71, 1–29, 2016.
9. Mzoughi, H., I. Njeh, A. Wali, et al., "Deep multi-scale 3D convolutional neural network (CNN) for MRI gliomas brain tumor classification," *Journal of Digital Imaging*, Vol. 33, 903–915, 2020.
10. Gao, Y., Z. Li, C. Song, et al., "Automatic rat brain image segmentation using triple cascaded convolutional neural networks in a clinical PET/MR," *Physics in Medicine and Biology*, Vol. 66, No. 4, 04NT01, 2021.
11. Zhang, Q., Y. Liao, X. Wang, et al., "A deep learning framework for 18F-FDG PET imaging diagnosis in pediatric patients with temporal lobe epilepsy," *European Journal of Nuclear Medicine and Molecular Imaging*, Vol. 48, 2476–2485, 2021.
12. Zhang, R., C. Cheng, X. Zhao, and X. Li, "Multiscale mask R-CNN-based lung tumor detection using PET imaging," *Molecular Imaging*, Vol. 18, 1–8, 2019.
13. Zheng, H., L. Qian, Y. Qin, et al., "Improving the slice interaction of 2.5D CNN for automatic pancreas segmentation," *Medical Physics*, Vol. 47, 5543–5554, 2020.
14. Kitrungrotsakul, T., X. Han, Y. Iwamoto, et al., "A cascade of 2.5D CNN and bidirectional CLSTM network for mitotic cell detection in 4D microscopy image," *IEEE-Acm Transactions on Computational Biology and Bioinformatics*, Vol. 18, 396–404, 2021.
15. Li, A., F. Li, F. Elahifasaee, M. Liu, and L. Zhang, "Hippocampal shape and asymmetry analysis by cascaded convolutional neural networks for Alzheimer's disease diagnosis," *Brain Imaging and Behavior*, 1–10, 2021.
16. Jo. T., K. Nho, S. L. Risacher, and A. J. Saykin, "Deep learning detection of informative features in tau PET for Alzheimer's disease classification," *BMC Bioinformatics*, Vol. 21, No. Suppl 21, 496, 2020.
17. Gao, X., R. Hui, Z. Tian, et al., "Classification of CT brain images based on deep learning networks," *Computer Methods and Programs in Biomedicine*, Vol. 138, 49–56, 2017.
18. Suk, H. I., S. W. Lee, and D. Shen, "Latent feature representation with stacked auto-encoder for AD/MCI diagnosis," *Brain Structure and Function*, Vol. 220, No. 2, 841–859, 2015.
19. Zhang, D., Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, Vol. 55, No. 3, 856–867, 2011.
20. Huang, Y., J. Xu, Y. Zhou, and T. Tong, "Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network," *Frontiers in Neuroscience*, Vol. 13, 509, 2019.
21. Cheng, D. and M. Liu, "CNNs based multi-modality classification for AD diagnosis," *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 1–5, 2017.
22. Zhou, P., S. Jiang, L. Yu, Y. Feng, C. Chen, and F. Li, "Use of a sparse-response deep belief network and extreme learning machine to discriminate Alzheimer's disease, mild cognitive impairment, and normal controls based on amyloid PET/MRI images," *Frontiers in Medicine*, Vol. 7, 987, 2021.
23. Zhu, X., H. Suk, and D. Shen, "Low-rank dimensionality reduction for multi-modality neurodegenerative disease identification," *World Wide Web*, Vol. 22, No. 2, 907–925, 2019.
24. Lin, W., Q. Gao, J. Yuan, Z. Chen, and C. Feng, "Predicting Alzheimer's disease conversion from mild cognitive impairment using an extreme learning machine-based grading method with multimodal data," *Frontiers in Aging Neuroscience*, Vol. 12, 77, 2020.
25. Giorgio, A., L. Santelli, V. Tomassini, R. Bosnell, S. Smith, N. D. Stefano, and H. Johansen-Berg, "Age-related changes in grey and white matter structure throughout adulthood," *NeuroImage*, Vol. 51, No. 3, 943–951, 2010.
26. Dukart, J., M. L. Schroeter, and K. Muller, "Age correction in dementia-matching to a healthy brain," *PloS One*, Vol. 6, No. 7, e22193, 2011.

27. Lin, W., T. Tong, Q. Gao, D. Guo, X. Du, Y. Yang, G. Guo, M. Xiao, M. Du, and X. Qu, "Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment," *Frontiers in Neuroscience*, Vol. 12, 777, 2018.
28. Roth, H. R., L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, and R. M. Summers, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Transactions on Medical Imaging*, Vol. 35, No. 5, 1170–1181, 2016.
29. Han, X., J. Jovicich, D. Salat, A. Kouwe, B. Quinn, S. Czanner, et al., "Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer," *NeuroImage*, Vol. 32, No. 1, 180–194, 2006.
30. Lin, M., Q. Chen, and S. Yan, "Network in network," *Proceedings of the IEEE International Conference on Learning Representations*, 2014.
31. Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., "Going deeper with convolutions" *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1–9, 2015.
32. He, K., X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778, 2016.
33. Ioffe, S. and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, Vol. 37, 448–456, 2015.
34. Elahifasaee, F., F. Li, and M. Yang, "A classification algorithm by combination of feature decomposition and kernel discriminant analysis (KDA) for automatic MR brain image classification and AD diagnosis," *Computational and Mathematical Methods in Medicine*, Vol. 2019, 1–14, 2019.
35. Oh, K., Y. C. Chung, K. W. Kim, and I. S. Oh, "Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning," *Scientific Reports*, Vol. 9, No. 1, 18150–18165, 2019.
36. Salvatore, C., A. Cerasa, P. Battista, M. Gilardi, A. Quattrone, and I. Castiglioni, "Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: A machine learning approach," *Frontiers in Neuroscience*, Vol. 9, 307–319, 2015.
37. Liu, M., D. Cheng, K. Wang, et al., "Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis," *Neuroinformatics*, Vol. 16, Nos. 3–4, 295–308, 2018.