Multi-View Separable Pyramid Network for AD Prediction at MCI Stage by 18 F-FDG Brain PET Imaging

Xiaoxi Pan, Trong-Le Phan, Mouloud Adel, Caroline Fossati, Thierry Gaidon, Julien Wojak, Eric Guedj

Multi-view Separable Pyramid Network for AD Prediction at MCI Stage by ^{18}F-FDG Brain PET Imaging

Xiaoxi Pan, Trong Le-Phan, Mouloud Adel, Caroline Fossati, Thierry Gaidon, Julien Wojak, and Eric Guedj, for Alzheimer's Disease Neuroimaging Initiative

***Abstract*—Alzheimer's Disease (AD), one of the main causes of death in elderly people, is characterize by Mild Cognitive Impairment (MCI) at prodromal stage. Nevertheless, only part of MCI subjects could progress to AD. The main objective of this paper is thus to identify those who will develop a dementia of AD type among MCI patients. ^{18}F-FluoroDeoxyGlucose Positron Emission Tomography (^{18}F-FDG PET) serves as a neuroimaging modality for early diagnosis as it can reflect neural activity via measuring glucose consumption at resting-state. In this paper, we design a deep network on ^{18}F-FDG PET modality to address the problem of AD identification at early MCI stage. To this end, a Multi-view Separable Pyramid Network (MiSePyNet) is proposed, in which representations are learned from axial, coronal and sagittal views of PET scans and then combined to make a decision jointly. Different from the widely and naturally used 3D convolution operations for 3D images, the proposed architecture is deployed with separable convolution from slice-wise to spatial-wise successively, which can retain the spatial information and reduce training parameters compared to 2D and 3D networks, respectively. Experiments on ADNI dataset show that the proposed method is comparable to other state-of-the-art algorithms for classifying AD from Normal Control (NC). For predicting the progression of Mild Cognitive Impairment, our method can yield better performance than both traditional and deep learning-based algorithms, with a classification accuracy of 83.05%.**

***Index Terms*—Separable Convolution, Slice-wise CNN, Spatial-wise CNN, ^{18}F-FDG PET, Mild Cognitive Impairment**

I. Introduction

ALZHEIMER'S Disease (AD) is a dominant degenerative brain disease among elderly people, and it will get worse as time goes by. It is believed that at least 50 million people worldwide suffer from AD or other dementias, and the number will exceed 152 million by 2050 if this situation continues [1]. The main reason why AD dementia becomes the leading cause of death is that no cure or treatment exists. It is expected to identify those subjects who could develop AD dementia at their early stage, i.e., Mild Cognitive Impairment (MCI), in order to deliver a more reliable diagnosis and prognosis to them. Then some interventions can be applied to delay the onset and/or mitigate risks of converting to AD dementia. However, the prediction of MCI transition to AD is quite difficult and challenging since typical changes are subtle. A variety of neuroimaging modalities, such as Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), are able to identify early changes occurring in brain. ^{18}F-FluoroDeoxyGlucose PET (^{18}F-FDG PET) is seen as a powerful tool for early detection of AD since it is able to capture cerebral glucose metabolic rate at resting-state and reveal metabolic aberrations before structural brain changes [2]–[4].

Computer-aided diagnosis (CAD) techniques based on machine learning approaches have been widely applied to tackle AD detection and prediction problems at early stage [5]. These techniques are typically structured into two steps: feature extraction and classification. Feature reduction or selection methods could be utilized prior to classification if extracted features are redundant. Several methods have been dedicated to either discriminating between AD subjects and Normal Controls (NC) or identifying those who will develop AD dementia among MCI subjects. Such methods include voxel-based approaches where voxels are used as features [6]–[8] and ROI (region of interest)-based methods where an atlas is used to segment a subject into different anatomical regions from which information is then extracted as regional features [9]–[16].

Deep learning methods, as a kind of emerging techniques, have gained impressive performance in recognition and classification tasks [17]–[20]. Such techniques are also involved in CAD methods and have been successfully applied to various medical tasks [21]–[24]. Different from previous described conventional CAD algorithms, deep learning methods manage to integrate feature extraction and classification, which means the typical feature engineering involved in classical machine learning methods, including feature design, extraction, and selection or reduction, is no longer required. Recently, an in-

This work has been carried out in the framework of DHU-Imaging thanks to the support of the A*MIDEX project (no. ANR-11-IDEX-0001-02) ("Investissements d'Avenir" French Government program, managed by the French National Research Agency (ANR)) and partly supported by Chinese Scholarship Council.

Xiaoxi Pan, Caroline Fossati and Thierry Gaidon are with Ecole Centrale de Marseille and Institut Fresnel UMR 7249, Marseille, France.(e-mails: xiaoxi.pan@fresnel.fr, caroline.fossati@fresnel.fr, thierry.gaidon@centrale-marseille.fr)

Trong Le-Phan, Mouloud Adel, Julien Wojak and Eric Guedj are with Aix-Marseille Université and Institut Fresnel UMR 7249, Marseille, France.(e-mails: trong.le-phan@fresnel.fr, mouloud.adel@univ-amu.fr, julien.wojak@fresnel.fr, eric.guedj@univ-amu.fr).

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

creasing number of advanced deep learning-based CAD methods have been developed for neuroimaging data to address the problem of AD prediction at early stage. Such methods could be roughly categorized into three groups according to different inputs, including ROI/patch value, 2D slice/patch, and 3D subject/patch.

ROI/patch value This group of methods usually segments subjects into different regions using either an existing predefined atlas template or a customized one, and mean or voxel values of such regions are then fed into a deep architecture. Zhou *et al.* [25] deploy a novel three-stage deep feature learning and fusion framework on MRI, PET and genetic data to capture individual and joint distinct patterns at the same time, where the average intensity of each ROI is taken as an input. Lu *et al.* [26] design a multi-scale stacked-autoencoder to learn latent representations for ^{18}F-FDG PET modality in order to diagnose AD from NC and predict the progression of MCI subjects towards AD. PET images are segmented into a variety of patches with different numbers of voxels and these patches' mean values are fed into the proposed framework. Suk *et al.* [27] feed voxels within a candidate patch to Deep Boltzmann Machine (DBM) [28] to find a latent feature representation for such a patch.

2D slice/patch This kind of approach attempts to exploit a 2D Convolutional Neural Network (CNN) to solve the diagnosis and prediction problems. To this end, 3D neuroimaging data is generally transformed into a 2D one through decomposing into 2D slices [29], [30] or reconstructing a 2D image [31] so as to fit the CNN model. Benefiting from the success of 2D CNN in natural scene images, this group of methods can take advantage of the existing CNN models which have been pre-trained in a large-scale dataset and then fine-tuned with their local data [31]. It could save a lot of computing resources with achieving good results and moreover, this strategy could mitigate overfitting caused by the limited number of medical data. However, since slices are processed independently, spatial relation information existing in 3D data might be lost.

3D subject/patch Methods belonging to this category usually apply 3D CNN as the basic architecture to diagnose AD. Accordingly, the whole 3D brain image [32] or a sub-image, a 3D patch [33], is taken as the input for a 3D model. Yee *et al.* [32] propose a 3D CNN with residual connections for ^{18}F-FDG PET images, which follows the idea of ResNet [20], and such a method can achieve an accuracy of 93.5% for AD classification from NC and 74.7% for the prediction of MCI conversion to AD. Huang *et al.* [33] input a sub-image that covers the highly relevant region to AD, i.e. hippocampal area, to a 3D VGG-based [18] architecture and obtain notable performance. The main advantage of these methods is that the spatial information is fully considered, therefore such approaches have become a trend, especially for MRI [34]–[36].

Considering 3D architectures specialized for ^{18}F-FDG PET are with a limited number at present, we propose a novel deep method by using such a modality to predict AD at MCI stage as well as classify AD from NC. To this end, a Multi-view Separable Pyramid Network (MiSePyNet), which processes axial, coronal and sagittal views via a factorized convolution manner, is designed. The main contributions of our study are summarized as three folds,

- The separable convolution, slice-wise CNN followed by spatial-wise CNN, is applied for the first time for 3D neuroimaging data, which enables the three views (axial, coronal and sagittal) to be considered jointly without losing spatial information.
- The architecture of each view is designed in a multi-scale manner, from coarse to fine, to capture subtle differences and obtain diversity in the receptive field.
- The developed model can achieve promising performance and generalization ability with fewer parameters, particularly for the prediction of MCI conversion to AD.

The remaining of this paper is structured as follows. Next section is devoted to the description of dataset used for evaluation and the proposed method. Section III presents the experimental results with comparisons to baseline and state-of-the-art methods. A discussion is then given in Section IV followed by a conclusion in Section V.

II. Method

In this paper, we present a novel CNN-based architecture, Multi-view Separable Pyramid Network (MiSePyNet), to tackle the problem of the prediction of MCI conversion to AD along with AD diagnosis among NC subjects, in which axial, coronal and sagittal views can be taken into account jointly owing to the separable convolution strategy, as shown in Fig. 1. Briefly, slice-wise CNN is first performed on each view at the starting layer in a multi-scale manner to learn representations among slices thereby combining them. The outputs are then fed into spatial-wise CNN, which is also with different scales of convolutional kernels, to yield distinguishing spatial patterns for prediction tasks. Benefiting from an accurate design of kernel sizes, the feature maps from different scale streams within the same view have an identical dimension, which makes it possible to combine feature maps through element-wise addition. Afterwards, different views are concatenated and then fed into fully-connected layers followed by a softmax function to give a result.

A. Dataset

1) Data Selection: ^{18}F-FDG PET data downloaded from Alzheimer's Disease Neuroimaging Initiative (ADNI) is studied in this paper. Participants generally take several scans at different time points so as to track their health states. Since the main objective in the paper is to predict MCI conversion to AD, therefore only the baseline subjects are taken into consideration. Accordingly, data that meets the following criteria is selected:

- AD: subjects diagnosed as AD dementia at the baseline time point and do not change within the follow-up time[1].
- NC: subjects diagnosed as NC at the baseline and do not change within the follow-up time.

[1]Diagnosis information is avaiable at https://ida.loni.usc.edu/login.jsp

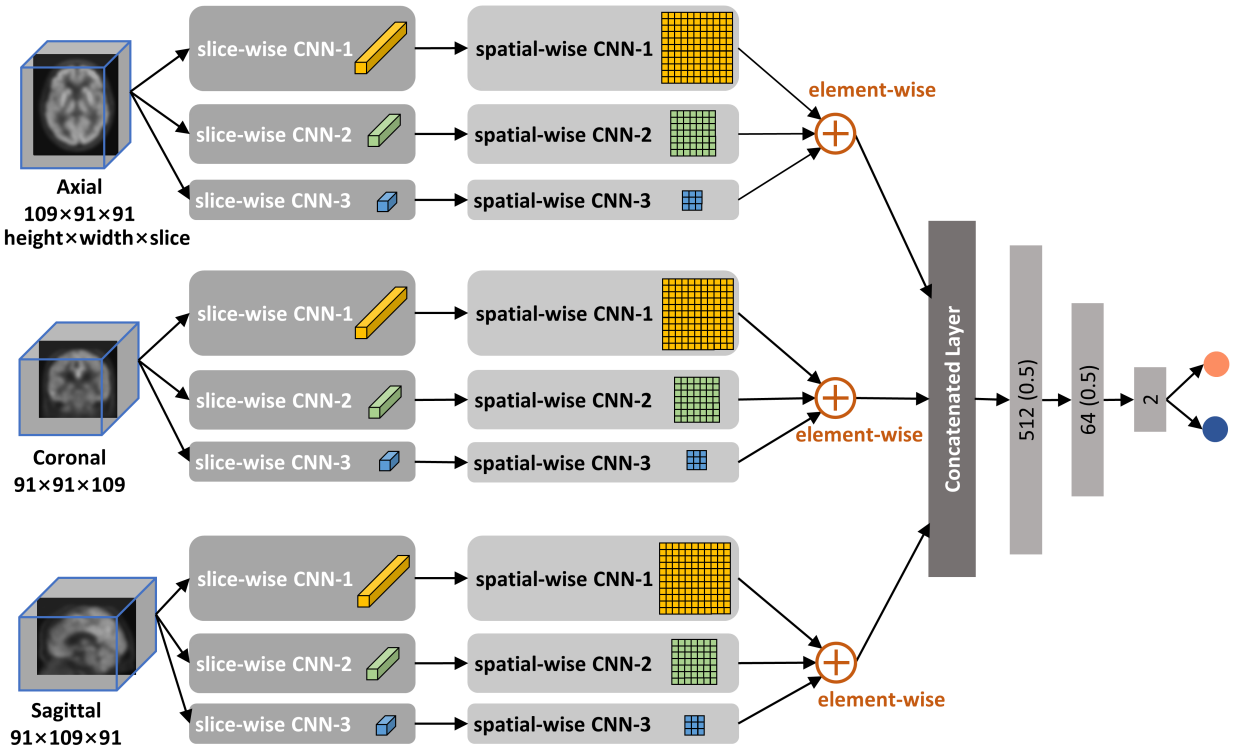


Fig. 1. Architecture of the proposed Multi-view Separable Pyramid Network (MiSePyNet).

- pMCI (progressive MCI): subjects diagnosed as MCI at the baseline and progress to AD and stay with AD within 36 months.
- sMCI (stable MCI): subject diagnosed as MCI at the baseline and stay in the phase of MCI or revert to NC within the available scan time and the visit time is not less than 24 months.

2) Data Pre-processing: The selected data is then pre-processed by performing spatial normalization, intensity normalization and smoothing. Specifically, the spatial normalization is to warp the image into MNI space and make the image have the same resolution, $91 \times 109 \times 91$ with a voxel size of $2 \times 2 \times 2\ mm^3$. The intensity normalization is then performed through dividing each voxel intensity by the global average value. Thereafter, images are further smoothed by a Gaussian kernel with a full width at half maximum of 8 mm. All the procedures are implemented with SPM12 [37]. These steps can ensure images are in the same standard space, making the subsequent analysis and comparison significant. After pre-processing, PET scans are checked manually in order to remove those subjects that failed in the processing procedure. At last, 1005 baseline ^{18}F-FDG PET images constitute the experimental dataset, among which 237 subjects are with AD, 242 subjects are under NC and 526 subjects are with MCI, including 166 pMCI and 360 sMCI cases. The demographic and clinical information of subjects is provided in Table I, in which MMSE stands for the Mini-Mental State Examination.

B. Architecture

In the following section, we introduce the proposed MiSePyNet method in detail, mainly including slice- and spatial-wise CNNs. Stacked convolutions with cuboïd kernels are firstly used for each view along the corresponding slice direction to compress the input and theoretical 2D convolutions (referred to 3D convolutions with a third dimension set to 1) are then performed spatially, rather than applying the commonly used 3D convolutions for neuroimaging data, as the work in [32], [34], [35]. The basic idea behind this method follows the convolution factorization implemented in Inception V3 [38] in which an $n \times n$ convolution is replaced by a $1 \times n$ convolution combined with an $n \times 1$ convolution. Thus, the proposed method that takes a 3D image as the input applies a $1 \times 1 \times n$ convolution[2] at first, namely, slice-wise CNN, which is followed by an $n \times n \times 1$ convolution, referred to as spatial-wise CNN.

1) Slice-wise CNN: The slice-wise CNN learns the representation among slices by exploiting cuboïd kernels whose size is $1 \times 1 \times n$ as shown in Fig. 2. The slice-CNN aims to compress the 3D input into a 2D feature map (regardless of channels) via slice-wise 1D convolution and such operations enable the network to pay more attention to significant slices. For this purpose, the first two dimensions of convolution kernels are set to 1, and the third dimension, n, depends on the number of times that convolutions are expected to be stacked before a

[2]Here n is decomposed into different scales, which will be discussed in the following.

TABLE I
DEMOGRAPHIC AND CLINICAL INFORMATION OF SUBJECTS.

Characteristic	AD	NC	pMCI	sMCI
Number of subjects	237	242	166	360
Female/male	97/140	122/120	70/96	153/207
Age (Mean ± Standard Deviation)	75.00 ± 7.91	73.66 ± 5.66	73.91 ± 6.74	71.73 ± 7.66
MMSE (Mean ± Standard Deviation)	23.19 ± 2.12	29.03 ± 1.20	26.99 ± 1.73	28.20 ± 1.59

2D feature map can be derived,

$$n = \frac{D-1}{t} + 1 \tag{1}$$

where D indicates the third dimension of an input view, in this paper, $D \in \{91, 109, 91\}$, and t stands for the times that convolutions will be stacked. Accordingly, the kernel size of slice-wise CNN is decided by the stacked convolution times rather than being set empirically. It should be noted that the corresponding stride involved in Eq. 1 is set to 1.

Taking the axial view ($109 \times 91 \times 91$) as an example, in this case, D is 91. If t is fixed to 1, which means after one convolution, the 3D input can be compressed into a 2D feature map (regardless of channels), consequently, n will equal 91. If t is set to 2, two convolutions are stacked, the corresponding n will be 46. Likewise, $t = 3$ suggests three convolutions with a kernel size of $1 \times 1 \times 31$ are stacked, as displayed in the block of slice-wise CNN-3 in Fig. 2. Accordingly, the slice-wise CNN takes advantage of three scales of cuboïd kernels to capture different changes among slices, thereby delivering several 2D outputs for each scale. Specifically, for the first scale, slice-wise CNN-1, 8 kernels with a size of $1 \times 1 \times 91$, denoted $8@1 \times 1 \times 91$ in Fig. 2, are applied, which is literally a weighted sum of slices. The slice-wise CNN-2 utilizes two stacked convolutions with 8 kernels each in which the kernel size is set to $1 \times 1 \times 46$. Three groups of 8 kernels are stacked in the third scale, with a kernel size of $1 \times 1 \times 31$. Each convolution layer is followed by batch normalization (BN) [39] and rectified linear unit (ReLu). The sagittal view has a similar kernel setting with the axial view, while for the coronal view, its kernel size is $1 \times 1 \times 109$, $1 \times 1 \times 55$ and $1 \times 1 \times 37$ for three scales of slice-wise CNNs. The outputs of different slice-wise CNNs within each view are with the same size, specifically, $109 \times 91 \times 1 \times 8$ (height×width×slice×channels) for the axial view, $91 \times 91 \times 1 \times 8$ for the coronal view and $91 \times 109 \times 1 \times 8$ for the sagittal view. The slice-wise CNN compresses a 3D image into a 2D representation, consequently, the parameters involved in the following spatial-wise CNN are relatively reduced.

2) Spatial-wise CNN: The spatial-wise CNN also utilizes multiple scales of kernels, from coarse to fine, to characterize the inputs in order to retain distinctive patterns, as presented in Fig. 3 which shows architecture details for the axial view. The first scale, spatial-wise CNN-1, begins with an $11 \times 11 \times 1$ convolution with a stride of 2, and followed by a $3 \times 3 \times 1$ max pooling layer as well as another group of convolution-max pooling operations but with a stride of 1 for the convolution. Afterwords, a $1 \times 1 \times 1$ convolution is applied at the end to increase the output dimension so that the number of channels is identical with the other two scales. The number of kernels

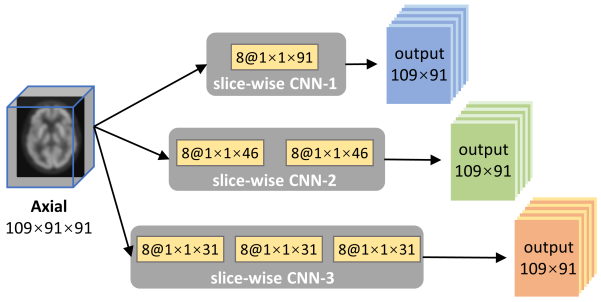


Fig. 2. Details of slice-wise CNN involved in MiSePyNet, taking the axial view as an instance.

for the three convolution layers (indicated by green color in the block) is set to 16, 32 and 64. For the second scale of spatial-wise CNN, it can be seen from Fig. 3 that the convolutional kernel size is uniformly fixed to $7 \times 7 \times 1$ together with a $2 \times 2 \times 1$ max pooling operation. Likewise, the spatial-wise CNN-3 also exploits a uniform kernel size, $3 \times 3 \times 1$, and is deployed with four convolution layers whose kernel number is 16, 32, 64 and 64, respectively. Moreover, max pooling layers involved in the *second and third* scales of spatial-wise CNNs are with padding. Similar to slice-wise CNN, convolution layers in spatial-wise CNN are also followed by BN and ReLu. As a result, different scales of networks within each view can yield outputs with the same dimension, $2 \times 1 \times 1 \times 64$ for the axial view, $1 \times 1 \times 1 \times 64$ for the coronal view and $1 \times 2 \times 1 \times 64$ for the sagittal view, which is attributed to the accurately designed architecture and enables the subsequent cross-scale combination. It is worth noting that the third dimension, i.e., slice, in each view is fixed to 1 among all the operations involved in spatial-wise CNN.

3) Combination and Fully-connected Layers: As can be seen from Fig. 1, the feature maps obtained from different scales within the same view are merged via the element-wise addition to strengthen distinctive patterns,

$$\mathbf{y}^v = \mathcal{F}^v\left(\mathbf{x}, \{\mathbf{W}_i\}_1\right) \oplus \mathcal{F}^v\left(\mathbf{x}, \{\mathbf{W}_i\}_2\right) \oplus \mathcal{F}^v\left(\mathbf{x}, \{\mathbf{W}_i\}_3\right) \tag{2}$$

where $\mathbf{x} \in \mathbb{R}^{h \times w \times s}$ stands for the input scan, $\mathbf{y}^v \in \mathbb{R}^{h \times w \times 1 \times c}$ is the output of each view after addition over multiple scales and $v \in$ {axial, coronal, sagittal}. $\mathcal{F}^v(\mathbf{x}, \{\mathbf{W}_i\}_j)$ represents a function to be learned in an effort to transform the input, $\mathbf{x}$, to various feature maps, in which $j \in \{1, 2, 3\}$ indicates different scales. Such an operation yields cross-scale combined feature maps, which can accumulate the distinctiveness derived from each scale and enhance the classification performance. Three views, i.e., axial, coronal and sagittal, are then concatenated after a flatten operation and fed into three fully-connected (FC)

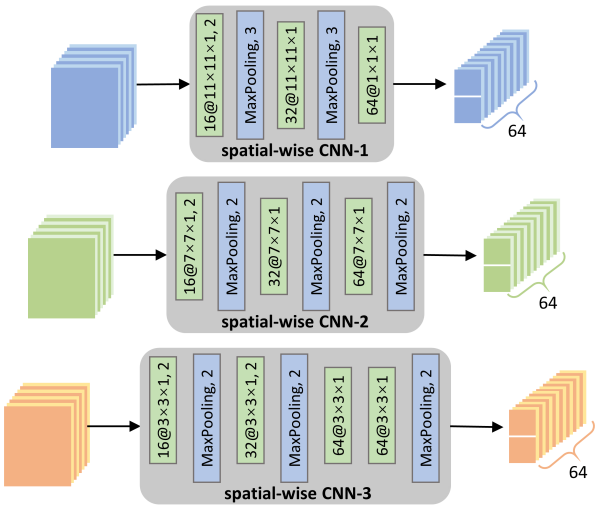


Fig. 3. Details of spatial-wise CNN involved in MiSePyNet, taking the axial view as an instance.

layers specified with 512, 64 and 2 neurons, respectively. In addition, the first two FC layers are followed by BN and ReLu, as well as a dropout strategy [40] with a rate of 0.5 to reduce the potential overfitting risk. The final result is delivered by a softmax function.

C. Implementations

All experiments are conducted by using Python 3.6 on a Linux machine equipped with an Nvidia Quadro P5000 graphics card with 16 GB of memory. The proposed architecture is implemented with Keras library [41] using TensorFlow [42] as the backend. The initialization method for all the convolution and dense layers follows 'he_uniform' [43]. The networks are trained for 40 epochs and the batch size is set to 8. The cross-entropy loss is applied as the objective function, which is minimized by a stochastic gradient descent (SGD) algorithm [44] with a step-wise learning rate, specifically, 10^{-3} is set as the initialization rate for epoch 1 to epoch 5, then it is decreased by 10 times, 10^{-4}, for epoch 6–20, followed by another 10 times decreasing, 10^{-5}, for the remaining 20 epochs, and the momentum coefficient is empirically set to 0.9. Each epoch takes about 80s, and the training of the proposed network takes less than 1 hour.

The dataset is randomly split into training, validation and testing sets with a percent of 60%, 20% and 20%, respectively. The model with the best performance on the validation set is then tested on the testing set. The prediction of MCI conversion to AD task is more challenging than classifying AD from NC as changes in MCI subjects could be quite subtle. According to [26], [36], information learned from AD classification can enrich the feature pool of MCI subjects thereby boosting the performance since AD is characterized by MCI at the prodromal phase. Consequently, the model for the prediction of MCI conversion to AD task is trained on MCI data as well as AD and NC data, and naturally, AD and pMCI are with the same label, while the other two groups, i.e., NC and sMCI, share the same one.

III. Experiments

A. Setup

The proposed MiSePyNet method is mainly tested on the prediction of MCI conversion to AD (pMCI vs. sMCI), as well as AD diagnosis among MC subjects (AD vs. NC). The performance evaluation is achieved by four metrics, namely, ACCuracy, the percent of correctly predicted samples, SENsitivity, the proportion of correctly classified positive (diseased) subjects, SPEcificity, the proportion of correctly identified NC or sMCI samples, and Area Under Curve (receiver operating characteristic curve determined by SEN and 1-SPE). Each of them is computed as,

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$
$$SEN = \frac{TP}{TP + FN}, \quad SPE = \frac{TN}{TN + FP} \tag{3}$$

where TP, TN, FP and FN stands for true positive, true negative, false positive and true negative, respectively. A higher value indicates better performance.

B. Performance of Single View Pyramid Network

In the proposed MiSePyNet architecture, convolutions are performed along the axial, coronal and sagittal views, and results are given by considering three views jointly. In order to compare each view's performance and meanwhile, validate the effectiveness of the integration strategy, the classification results (under the same data partition) obtained from a single view and multi-view networks are reported in Table II and Table III for AD vs. NC and pMCI vs. sMCI, respectively. It should be noted that multi-scale convolutions are retained in this group of experiments.

It can be seen from Table II that axial view can give relatively better overall performance than the other two views in AD classification, while coronal view is inferior, with differences of 2.09%, 4.25%, 0 and 1.47% regarding ACC, SEN, SPE and AUC to the best view. Moreover, the proposed multi-view network can further improve the diagnostic accuracy, with an increase of 2.08%, and also achieve a notable improvement concerning SEN, 4.26%. Surprisingly, axial, coronal and sagittal views, as well as their combination MiSePyNet, have the same SPE value, which implies that they have an identical ability in recognizing healthy subjects. Therefore, view integration can enhance the ability of a model to diagnose AD from NC since multiple views could offer complementary information, which also suggests that applying a multi-view strategy in the network is reasonable.

TABLE II
Performance of Different Views of Networks for AD vs. NC(%)

Vies	ACC	SEN	SPE	AUC
Axial view network	**91.67**	**87.23**	**95.92**	**96.61**
Coronal view network	89.58	82.98	**95.92**	95.14
Sagittal view network	90.63	85.11	**95.92**	96.57
MiSePyNet	**93.75**	**91.49**	**95.92**	**96.87**

The results of the prediction of MCI conversion to AD are shown in Table III, which are with different observations

in contrast to AD vs. NC. In general, among the three views, axial view has slightly better performance in terms of ACC and AUC, with differences of 0.95% and 0.62% to the corresponding slightly inferior metric (80.95% and 86.45%), respectively. However, coronal and sagittal views yield comparable performance, specifically, sagittal view network has advantages regarding SEN and AUC, whereas coronal view could give better ACC and SPE results. So it is uneasy to rank single view performance in this task. Nevertheless, axial view network outperforms the other two views, which is consistent with the fact that physicians take the brain scan in axial view as a reference to diagnose in practice. Furthermore, the multi-view network can still achieve satisfactory improvements, especially for pMCI prediction indicated by SEN, which has been increased by 6.06%. Similar to AD diagnosis among NC subjects, benefiting from multiple view combination, pMCI subjects are particularly concerned by the developed model, thereby resulting in performance enhancement.

TABLE III
PERFORMANCE OF DIFFERENT VIEWS OF NETWORKS FOR PMCI VS. SMCI(%)

Views	ACC	SEN	SPE	AUC
Axial view network	**81.90**	**69.70**	**87.50**	**87.07**
Coronal view network	80.95	66.67	**87.50**	86.32
Sagittal view network	80.00	**69.70**	84.72	86.45
MiSePyNet	**83.81**	**75.76**	**87.50**	**88.89**

C. Performance of Multi-scale Networks

Convolutions involved in the MiSePyNet are implemented in a multi-scale manner and kernel sizes are designed in an inverted pyramid fashion, from coarse to fine, along the scales. Feature maps within the same view are then integrated through element-wise addition across scale. In order to analyze the effects of multi-scale strategy, we conduct experiments on different scales of networks for both tasks and still under the same data partition. The multi-view operation is kept unchanged in such experiments. As can be seen from Fig. 4(a), i.e., AD vs. NC, the classification performance is monotonically increasing as the scale changes, from a single to three scales, in terms of ACC, SEN, AUC, while the value of SPE, 95.92%, is not affected by different scales of networks. Significant improvements lie in ACC and SEN, especially for SEN which is increased by 8.51%. For the case of pMCI vs. sMCI illustrated in Fig. 4(b), we can also observe that the performance is clearly enhanced concerning ACC, SEN and AUC, while SPE remains unchanged in single scale network, indicated by 's1', and MiSePyNet, indicated by 's1+s2+s3'. SEN is largely increased from 66.67% to 75.75% and nearly achieves an improvement of 10%. It implies that the MiSePyNet model is dedicated to positive samples (AD and pMCI) thereby enabling overall performance enhancement, which is potentially attributed to the multi-scale operation since it can gain various receptive fields and obtain diverse feature maps.

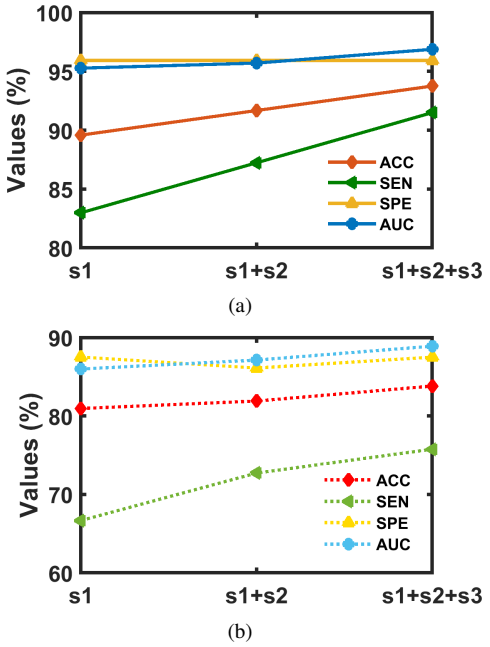


Fig. 4. Performance of multi-scale networks, 's1' indicates a model incorporating only slice- and spatial-wise CNN-1 along each view, 's1+s2' is a model with the first two scales, slice- and spatial-wise CNN-1,2, while 's1+s2+s3' represents the proposed architecture, MiSePyNet. (a) AD vs. NC. (b) pMCI vs. sMCI

D. Effectiveness of Cross-task Guidance

The trained model for pMCI vs. sMCI task is guided by AD classification for promoting its performance. In order to access the effects of cross-task guidance, we compare the results obtained without guidance with those given by a model trained with guidance (under the same data partition), as displayed in Fig. 5. With the assistance of AD and NC data, the accuracy has been increased nearly by 2% and the metric AUC also achieves an improvement of 1%. In spite of a decrease of 4.17% regarding SPE, SEN is drastically increased with an improvement of 15.15%, which appears that the model with guidance also has a stronger ability to identify diseased or positive subjects. It is reasonable since subjects with MCI could develop AD dementia, thus scans with relatively severe disease (AD) are able to contribute to the training. In summary, features learned from AD classification could yield a positive influence on the prediction of MCI conversion to AD.

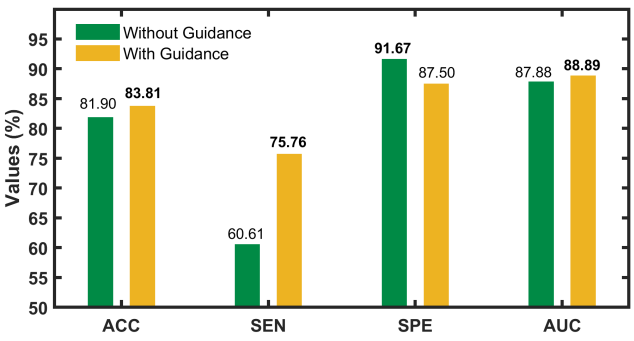
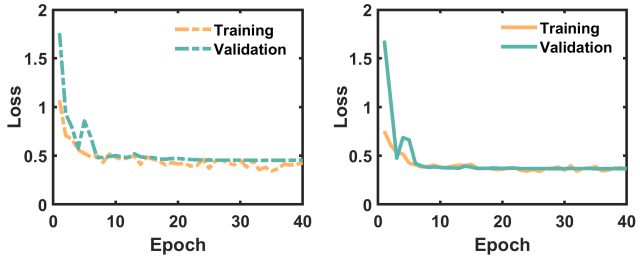


Fig. 5. Performance comparison between models without and with guidance.

Moreover, loss curves of training and validation sets, without and with guidance from AD vs. NC task, are presented in Fig. 6. It can be seen that, for both cases, the proposed MiSePyNet converges quickly within 10 epochs, in particular for the case with cross-task guidance (Fig. 6(b)). In addition, its loss differences between training and validation sets are very slight, and it is a little bit smaller than those derived from the model without guided information shown in Fig. 6(a). It implies the potential overfitting has been mitigated. Therefore, the cross-task data guidance could be seen as an alternative way of data augmentation that is dedicated to easing overfitting.



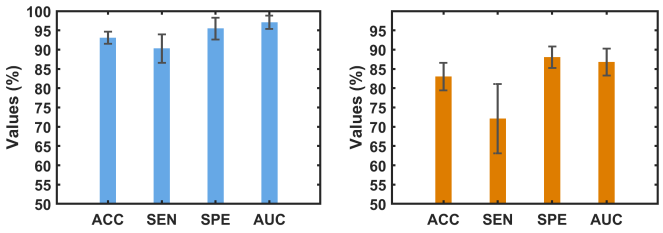
(a) (b)

Fig. 6. Changes of losses for training and validation sets. (a) Without guidance. (b) With guidance.

E. Influence of Data Partition

In the above experiments, the model performance is validated on fixed testing set for the sake of fair comparison. However, due to a limited number of samples, different data partitions could lead to differences in performance. Thus we evaluate the MiSePyNet model on different data splits to test its generalization ability. For this purpose, the classification of each task is repeated 10 times and the average results are used to evaluate the corresponding overall performance.

Figure 7(a) presents the results of AD diagnosis among NC subjects, which are 93.13%, 90.32%, 95.49% and 97.11% in terms of ACC, SEN, SPE and AUC, and the associated standard deviation indicated by the error band in the figure is 1.57, 3.69, 2.86 and 1.74, for each metric. While for the prediction of MCI conversion to AD, as shown in Fig. 7(b), the average results and standard deviation indicated in brackets are $83.05\%(3.56)$, $72.12\%(9.01)$, $88.06\%(2.80)$ and $86.80\%(3.49)$ regarding the four metrics. It implies the MiSePyNet method still achieves notable performance in the prediction of MCI conversion to AD despite its standard deviation is higher than that of AD classification, in particular for SEN. This is mainly due to the more challenging task of identifying pMCI. In summary, the proposed MiSePyNet method is with a good generalization ability in classifying AD from NC subjects. In order to reduce the possible effects caused by data partition and meanwhile show the overall performance in AD vs. NC and pMCI vs. sMCI, we use average results for the comparison with other methods in the remaining parts.



(a) (b)

Fig. 7. Overall performance evaluation in which the error band upon each bar indicates the corresponding standard deviation. (a) AD vs. NC. (b) pMCI vs. sMCI.

F. Comparison with Baseline Methods

We then compare the proposed network with baseline methods on the same experimental dataset. For the traditional baseline methods, two types of commonly used features are included in this paper, voxel- and ROI-wise features, as well as the classifier, Support Vector Machine (SVM) [45] with a linear kernel. Voxel-wise method takes intensities of all the voxels in cortical as the features, while ROI-wise method utilizes the mean value within each anatomical region as a feature. As to the comparison with deep learning baseline methods, we adopt VGG-16 as the backbone, specifically, 2D CNN model takes the third dimension of a 3D PET image as the channel and the corresponding number of convolution kernels are decreased by 4 times due to a limited number of samples, which is set to 16, 16, 32, 32, 64, 64, 64, 128, 128, 128, 256, 256, 256. Likewise, the baseline 3D CNN method also follows VGG-16 architecture only by extending the kernel size to a 3D shape, and the kernel number is kept the same with 2D CNN but only the first 10 convolution layers are considered according to experiments. The dense layers for both 2D and 3D CNNs remain identical with the proposed MiSePyNet method.

The comparison results are reported in Table IV and Table V for AD vs. NC and pMCI vs. sMCI, respectively. It can be seen from Table IV that the developed MiSePyNet method outperforms the traditional method which applies regional features, and the performance has been increased by 4.57%, 3.24%, 5.49 and 2.2% in terms of ACC, SEN, SPE and AUC. However, compared to the voxel-wise method, the improvements achieved by the MiSePyNet are not that notable, which is probably caused by a limited number of training data. Due to the same reason, baseline CNN-based methods (2D and 3D CNNs) do not perform as well as the voxel-wise method either. Nevertheless, the proposed MiSePyNet architecture can outperform the two baseline CNN methods with obvious advantages, specifically, compared to the 3D CNN model (the better baseline network), the metrics are improved by 6.57%, 9.91%, 2.91% and 3.03%. The considerable performance difference between 2D and 3D CNN models indicates that the spatial relationship information is so crucial that it cannot be neglected. Moreover, the parameters involved in the proposed network is dramatically reduced compared to both baseline CNN models. For the prediction of MCI conversion to AD, the improvement is more significant, as can be seen from Table V. The MiSePyNet model has dominant performance

among comparison methods, with an accuracy of 83.05% and a metric AUC of 86.80%, which are 8.05% and 8.43% higher than ROI-wise method (the better one in traditional methods), and 4.38% and 4.89% than the 3D CNN (the better one in CNN-based approaches). It implies that the proposed method is more capable to handle a quite challenging task. Therefore, the proposed lightweight model has its superiority on AD diagnosis among NC and particularly the prediction of MCI conversion to AD compared to baseline methods.

TABLE IV
COMPARISON WITH BASELINE METHODS FOR AD VS. NC(%)

Methods	ACC	SEN	SPE	AUC	Parameters
Voxel-wise	92.83	**91.90**	93.71	**97.17**	–
ROI-wise	88.56	87.08	90.00	94.91	–
2D CNN	80.31	70.41	90.07	87.30	2.79 M
3D CNN	86.56	80.41	92.58	94.08	11.30 M
MiSePyNet	**93.13**	90.32	**95.49**	97.11	**1.05 M**

TABLE V
COMPARISON WITH BASELINE METHODS FOR FOR PMCI VS. SMCI(%)

Methods	ACC	SEN	SPE	AUC	Parameters
Voxel-wise	74.38	54.59	83.67	78.11	–
ROI-wise	75.00	55.83	83.77	78.37	–
2D CNN	72.29	48.79	83.06	76.37	2.79 M
3D CNN	78.67	55.45	**89.31**	81.91	11.30 M
MiSePyNet	**83.05**	**72.12**	88.06	**86.80**	**1.05 M**

G. Comparison with State-of-the-art Methods

We also compare the proposed MiSePyNet method with state-of-the-art approaches, including methods that follow the conventional classification pipeline, such as work have been investigated by Hinrichs *et al.* [6], Padilla *et al.* [7], Li *et al.* [10], Gray *et al.* [9], Zhu *et al.* [11], [12], Cheng *et al.* [14] and Pan *et al.* [15], [16], as well as emerging techniques that employ deep learning methods, like Suk's [27], Lu's [26], Liu's [29] and Yee's methods [32]. Some of them focus on multiple modalities, e.g. MRI and ^{18}F-FDG PET, but we only compare with those results achieved on ^{18}F-FDG PET modality for a fair comparison. The comparison results are briefly summarized in Table VI and Table VII for AD vs. NC and pMCI vs. sMCI, respectively.

As can be seen, deep learning-based methods are superior to the traditional methods applying hand-crafted features for both tasks, especially for the challenging one, pMCI vs. sMCI, which is also the reason why deep learning has gained more attention in medical image analysis. In addition, our proposed method is comparable to the other methods in the task of AD vs. NC and has better overall performance in the prediction of MCI conversion to AD, which indicates the designed architecture considering axial, coronal and sagittal views jointly is effective and reasonable. It is worth noting that due to the potential differences involved in data selection, pre-processing and even dataset partition, results obtained from different methods are actually incomparable. The comparison just aims to provide an overview of other results and show the baseline of existing methods.

IV. DISCUSSION

The prediction of AD at MCI stage is of more clinical significance than identifying AD from NC subjects since it could not only delay the onset but also provide insights into such a disease. However, the acquisition of pMCI and sMCI subjects is not as easy as the way of getting AD and NC data, since they are labeled on the basis of longitudinal data, while such data usually takes tens of months, e.g. 24, 36 or even more, to be obtained. Considering we have incorporated AD and NC subjects into MCI training and validated the effectiveness of such a strategy in performance improvement, an extra experiment is conducted in which only AD and NC data are used as training set, while all the MCI subjects, totally 526, are for testing in order to further verify the effects of the proposed MiSePyNet method on an independent dataset. The comparison results are shown in Table VIII, it can be seen that despite the testing set is larger and more challenging, the proposed method, denoted 'MiSePyNet-ad', can still obtain favorable results with an accuracy of 80.16% and an AUC of 86.46%, which is superior to the baseline CNN methods, indicated '2/3D CNN-ad'. It suggests that the MiSePyNet is able to work even without direct information and its generalization ability is further verified on an independent testing set.

Although the MiSePyNet is able to largely enhance the performance of prediction of MCI conversion to AD, there are still some limitations, which cannot be ignored. *First,* our proposed network is not absolutely end-to-end since some pre-processing operations have to be applied. In fact, such an issue, whether pre-processing steps are necessary, still remains unclear [46]. On one hand, a standard space resulted from the pre-processing procedure could make a model focus on the problem-specific patterns, which is beneficial for training. On the other hand, diversity is reduced at the same time, which could hinder the generalization ability of a model. This issue should be investigated and considered carefully in our future work. *Second,* our method can achieve a notable overall accuracy, but the sensitivity is not that satisfactory in contrast to SPE, which could be addressed by Focal Loss [47]. *Third,* the developed model can give a prediction of the future state for MCI subjects, but it cannot decide when is the future. So it could be interesting to include longitudinal data to grade the severity of the baseline scan in our following work.

V. CONCLUSION

In this paper, we have proposed a novel CNN model, MiSePyNet, for ^{18}F-FDG PET modality in an effort to cope with the task of AD prediction at MCI stage as well as AD classification among NC subjects. The MiSePyNet follows the idea of factorized convolution and is deployed with separable CNNs, slice- and spatial-wise CNNs, for each view. Benefiting from such a design, MiSePyNet is able to consider axial, coronal and sagittal views jointly without losing spatial information. Furthermore, each view is characterized by multi-scale networks in order to capture different changes and expand the range of receptive field, thereby enhancing the discriminative feature maps. Experiments on ADNI data show that the proposed architecture can achieve satisfactory diagnosis results,

TABLE VI
PERFORMANCE COMPARISON WITH STATE-OF-THE-ART METHODS FOR AD VS. NC(%)

Category	Method	Data type	Subjects	ACC	SEN	SPE	AUC
Conventional methods	Hinrichs *et al.* [6]	MRI, ^{18}F-FDG PET	89AD + 94NC	84	84	82	87.16
	Padilla *et al.* [7]	^{18}F-FDG PET	53AD + 52NC	86.59	87.50	85.36	−−
	Gray *et al.* [9]	^{18}F-FDG PET	50AD + 54NC	88.4	83.2	93.6	−−
	Li *et al.* [10]	^{18}F-FDG PET	25AD + 30NC	89.1	92	86	97
	Zhu *et al.* 2014 [11]	MRI, ^{18}F-FDG PET, CSF*	51AD + 52NC	92.3	**92.3**	93.9	96.6
	Zhu *et al.* 2016 [12]	MRI, ^{18}F-FDG PET	51AD + 52NC	93.3	−−	−−	−−
	Pan *et al.* 2019a [15]	^{18}F-FDG PET	237AD + 242NC	92.57	90.89	94.42	96.83
	Pan *et al.* 2019b [16]	^{18}F-FDG PET	237AD + 242NC	**94.20**	91.45	**96.76**	**97.42**
Emerging methods	Lu *et al.* [26]	^{18}F-FDG PET	226AD + 304NC	**93.58**	91.54	95.06	−−
	Suk *et al.* [27]	MRI, ^{18}F-FDG PET	93AD + 101NC	92.20	88.04	**96.33**	**97.98**
	Liu *et al.* [29]	^{18}F-FDG PET	93AD + 100NC	91.2	91.4	91.0	95.3
	Yee *et al.* [32]	^{18}F-FDG PET	237AD + 359NC	93.5	**92.3**	94.2	97.6
	Huang *et al.* [33]	MRI, ^{18}F-FDG PET	465AD + 480NC	89.11	90.24	87.77	92.69
	MiSePyNet (Ours)	^{18}F-FDG PET	237AD + 242NC	93.13	90.32	95.49	97.11

*CSF = Cerebrospinal fluid

TABLE VII
PERFORMANCE COMPARISON WITH STATE-OF-THE-ART METHODS FOR PMCI VS. SMCI(%)

Category	Method	Data type	Subjects	ACC	SEN	SPE	AUC
Conventional methods	Gray *et al.* [9]	^{18}F-FDG PET	53pMCI + 64sMCI	63.1	52.2	73.2	−−
	Zhu *et al.* 2014 [11]	MRI, ^{18}F-FDG PET, CSF	43pMCI + 56sMCI	70.9	42.7	94.1	77.4
	Zhu *et al.* 2016 [12]	MRI, ^{18}F-FDG PET	43pMCI + 56sMCI	69.9	−−	−−	−−
	Cheng *et al.* [14]	MRI, ^{18}F-FDG PET, CSF	43pMCI + 56sMCI	71.6	76.4	67.9	74.1
	Pan *et al.* 2019a [15]	^{18}F-FDG PET	166pMCI + 360sMCI	79.43	69.14	84.16	83.88
	Pan *et al.* 2019b [16]	^{18}F-FDG PET	166pMCI + 360sMCI	**80.48**	65.04	**87.95**	**85.67**
Emerging methods	Lu *et al.* [26]	^{18}F-FDG PET	112pMCI + 409sMCI	82.51	**81.36**	82.85	−−
	Suk *et al.* [27]	MRI, ^{18}F-FDG PET	76pMCI + 128sMCI	70.75	25.45	**96.55**	72.15
	Yee *et al.* [32]	^{18}F-FDG PET	210pMCI + 427sMCI	74.7	74.0	75.0	81.1
	MiSePyNet (Ours)	^{18}F-FDG PET	166pMCI + 360sMCI	**83.05**	72.12	88.06	**86.80**

TABLE VIII
PERFORMANCE EVALUATION ON INDEPENDENT DATASET (%)

Methods	ACC	SEN	SPE	AUC
2D CNN–ad	73.00	50.60	83.33	76.24
3D CNN–ad	76.81	59.64	**84.72**	81.98
MiSePyNet–ad	**80.61**	**71.69**	**84.72**	**86.46**

particularly for pMCI vs. sMCI, with an average accuracy of 83.05% and an average AUC of 86.80%. We also test MiSePyNet on a more challenging testing set, the results reveal that the proposed method is with a strong generalization ability as well as good performance in predicting the conversion of MCI to AD even without direct information.

REFERENCES

[1] C. Patterson. "World Alzheimer Report 2018: the state of the art of dementia research: new frontiers." *Alzheimer's Disease International*, London, UK, 2018.

[2] W. Jagust, A. Gitcho, F. Sun, B. Kuczynski, D. Mungas, and M. Haan. "Brain imaging evidence of preclinical Alzheimer's disease in normal aging," *Ann. Neurol.*, vol. 59, no. 4, pp. 673-681, 2006.

[3] L. Mosconi, V. Berti, L. Glodzik, A. Pupi, S. De Santi, and M. J. Leon. "Pre-clinical detection of Alzheimer's disease using FDG-PET, with or without amyloid imaging." *J. Alzheimer's Dis.*, vol. 20, no. 3, pp. 843-854, 2010.

[4] C. R. Jack, D. S. Knopman, W. J. Jagust, L. M. Shaw, P. S. Aisen, M. W. Weiner, *et al*. "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade," *Lancet Neurol.*, vol. 9, no. 1, pp. 119–128, 2010.

[5] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, C. Davatzikos. "A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages," *NeuroImage*, vol. 155, pp. 530-548, 2017.

[6] C. Hinrichs, V. Singh, L. Mukherjee, G. Xu, M. K. Chung, S. C. Johnson . "Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset," *NeuroImage*, vol. 48, no. 1, pp. 138-149, 2009.

[7] P. Padilla, M. López, J. M. Górriz, J. Ramirez, D. Salas-Gonzalez and I. Alvarez. "NMF-SVM based CAD tool applied to functional brain images for the diagnosis of Alzheimer's disease," *IEEE Trans. Med. Imag.*, vol. 31, no. 2, pp. 207-216, 2012.

[8] C. Cabral, P. M. Morgado, D. C. Costa, and M. Silveira. "Predicting conversion from MCI to AD with FDG-PET brain images at different prodromal stages," *Computers in Biology and Medicine*, vol. 58, pp. 101-109, 2015.

[9] K. R. Gray, R. Wolz, R. A. Heckemann, P. Aljabar, A. Hammers and Daniel Rueckert . "Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease," *NeuroImage*, vol. 60, no. 1, pp. 221-229, 2012.

[10] R. Li, R. Perneczky, I. Yakushev, S. Förster, A. Kurz, A. Drzezga, *et al*. "Gaussian mixture models and model selection for [18F] fluorodeoxyglucose positron emission tomography classification in Alzheimer's disease," *PloS One*, vol. 10, no. 4, 2015.

[11] 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, 2014.

[12] X. Zhu, H. I. Suk, S. Lee, and D. Shen. "Subspace regularized sparse

multitask learning for multiclass neurodegenerative disease identification," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 607-618, Mar. 2016.

[13] I. Garali, M. Adel, S. Bourennane and E. Guedj. "Brain region ranking for 18FDG-PET computer-aided diagnosis of Alzheimer's disease," *Biomedical Signal Processing and Control*, vol. 27, pp. 15-23, 2016.

[14] B. Cheng, M. Liu, D. Zhang, B. C. Munsell and D. Shen. "Domain transfer learning for MCI conversion prediction," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 7, pp. 1805-1817, 2015.

[15] X. Pan, M. Adel, C. Fossati, T. Gaidon,and E. Guedj. "Multi-level feature representation of FDG-PET brain images for diagnosing Alzheimer's disease," *IEEE J. Biomed. Health Informatics*, vol. 23, no. 4, pp. 1499-1506, 2019.

[16] X. Pan, M. Adel, C. Fossati, T. Gaidon, J. Wojak and E. Guedj. "Multiscale spatial gradient features for ^{18}F-FDG PET image-guided diagnosis of Alzheimer's disease," *Comput. Meth. Prog. Bio.*, vol. 180, 105027, 2019.

[17] A. Krizhevsky, I. Sutskever I and G. E. Hinton. "Imagenet classification with deep convolutional neural networks," in*Proc. NeurIPS*, pp. 1097-1105, 2012.

[18] K. Simonyan and A. Zisserman. "Very deep convolutional networks for large-scale image recognition", in *Proc. ICLR*, 2015.

[19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, *et al*. "Going deeper with convolutions," in *Proc. CVPR*, pp. 1-9, 2015.

[20] K. He, X. Zhang, S. Ren and J. Sun. "Deep residual learning for image recognition," in *Proc. CVPR*, pp. 770-778, 2016.

[21] H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, *et al*. "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285-1298, 2016.

[22] F. C. Ghesu, B. Georgescu, Y. Zheng, S. Grbic, A. Maier, J. Hornegger, *et al*, "Multi-scale deep reinforcement learning for real-time 3D-landmark detection in CT scans," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 176-189, 2017.

[23] S. Graham, H. Chen, J. Gamper, Q. Dou, P. A. Heng, D. Snead *et al*. "MILD-Net: minimal information loss dilated network for gland instance segmentation in colon histology images," *Med. Image Anal.*, vol. 52, pp. 199-211, 2019

[24] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, *et al*. "Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1182-1195, 2016.

[25] 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 Mapp.* vol. 40, no. 3, pp. 1001-1016, 2019.

[26] D. Lu, K. Popuri, G. W. Ding, R. Balachandar and M. F. Beg. "Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease," *Med. Image Anal.*, vol. 46, pp. 26-34, 2018.

[27] H. I. Suk, S. W. Lee and D. Shen D. "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis," *NeuroImage*, vol. 101, pp. 569-582, 2014.

[28] R. Salakhutdinov and G. Hinton. "Deep boltzmann machines," *Proc. AISTATS*, pp. 448-455, 2009.

[29] M. Liu, D. Cheng and W. Yan. "Classification of Alzheimer's Disease by combination of convolutional and recurrent neural networks using FDG-PET images," *Front. Neuroinform.*, vol. 12, article 35, 2018.

[30] A. Gupta, M. Ayhan and A. Maida. "Natural image bases to represent neuroimaging data," in *Proc. ICML*, pp. 987-994, 2013.

[31] Y. Ding, J. H. Sohn, M. G. Kawczynski, H. Trivedi, R. Harnish, N. W. Jenkins, *et al*. "A deep learning model to predict a diagnosis of Alzheimer disease by using 18F-FDG PET of the brain," *Radiology*, vol. 290, no. 2, pp. 456-464, 2019.

[32] E. Yee, K. Popuri and M. F. Beg. "Quantifying brain metabolism from FDG-PET images into a probability of Alzheimer's dementia score," *Hum. Brain Mapp.*, 2019.

[33] Y. Huang, J. Xu, Y. Zhou, T. Tong and X. Zhuang. "Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network," *Front. Neurosci.*, vol. 13, article 509, 2019.

[34] C. Lian, M. Liu, J. Zhang and D. Shen. "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's Disease diagnosis using structural MRI," *IEEE Trans. IEEE Trans. Pattern Anal. Mach. Intell.*, 2018.

[35] S. Spasov, L. Passamonti, A. Duggento, P. Liò and N. Toschi. "A parameter-efficient deep learning approach to predict conversion from Mild Cognitive Impairment to Alzheimer's Disease," *NeuroImage*, vol. 189, pp. 276–287, 2019.

[36] M. Liu, J. Zhang, E. Adeli and D. Shen. "Landmark-based deep multi-instance learning for brain disease diagnosis," *Med. Image Anal.*, vol. 43, pp. 157-168, 2018.

[37] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel and T. E. Nichols. "Statistical parametric mapping: the analysis of functional brain images," London: UK, Elsevier, 2011.

[38] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna. "Rethinking the inception architecture for computer vision," in *Proc. CVPR*, pp. 2818-2826, 2016.

[39] S. Ioffe and C. Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint*, 1502.03167, 2015.

[40] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov. "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929-1958, 2014.

[41] F. Chollet. "Keras,", 2015.

[42] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, *et al*. "Tensorflow: A system for large-scale machine learning," in *Proc. OSDI*, pp. 265–283, 2016.

[43] K. He, X. Zhang, S. Ren, and J. Sun. "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. ICCV*, pp. 1026-1034, 2015.

[44] S. Boyd and L. Vandenberghe. "Convex optimization," *Cambridge university press*, 2004.

[45] C. Cortes and V. Vapnik. "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273-297, 1995.

[46] J. Wen J, E. Thibeau-Sutre, J. Samper-Gonzalez, . "Convolutional Neural Networks for Classification of Alzheimer's Disease: Overview and Reproducible Evaluation," *arXiv preprint*, arXiv:1904.07773, 2019.

[47] T. Y. Lin, P. Goyal, R. Girshick, K. He and P Dollár. "Focal loss for dense object detection," in *ICCV*, pp. 2980-2988, 2017.