



Multiscale spatial gradient features for ^{18}F -FDG PET image-guided diagnosis of Alzheimer's disease

Xiaoxi Pan^{b,c}, Mouloud Adel^{a,c,*}, Caroline Fossati^{b,c}, Thierry Gaidon^{b,c}, Julien Wojak^{a,c}, Eric Guedj^{a,c,d}, for Alzheimer's Disease Neuroimaging Initiative¹

^a Aix Marseille Univ, Marseille 13013, France

^b Centrale Marseille, Marseille 13013, France

^c Institut Fresnel, 52 Avenue Escadrille Normandie Niemen, Marseille 13013, France

^d Centre Européen de Recherche en Imagerie Médicale, Marseille 13005, France

ARTICLE INFO

Article history:

Received 20 April 2019

Revised 7 August 2019

Accepted 9 August 2019

Keywords:

Multiscale spatial gradients

Ensemble classification

^{18}F -FDG PET

Alzheimer's disease

ABSTRACT

Background and Objective: ^{18}F -FluoroDeoxyGlucose Positron Emission Tomography (^{18}F -FDG PET) is one of the imaging biomarkers to diagnose Alzheimer's Disease (AD). In ^{18}F -FDG PET images, the changes of voxels' intensities reflect the differences of glucose rates, therefore voxel intensity is usually used as a feature to distinguish AD from Normal Control (NC), or at earlier stage to distinguish between progressive and stable Mild Cognitive Impairment (pMCI and sMCI). In this paper, ^{18}F -FDG PET images are characterized in an alternative way—the spatial gradient, which is motivated by the observation that the changes of ^{18}F -FDG rates also cause gradient changes.

Methods: We improve Histogram of Oriented Gradient (HOG) descriptor to quantify spatial gradients, thereby achieving the goal of diagnosing AD. First, the spatial gradient of ^{18}F -FDG PET image is computed, and then each subject is segmented into different regions by using an anatomical atlas. Second, two types of improved HOG features are extracted from each region, namely Small Scale HOG and Large Scale HOG, then some relevant regions are selected based on a classifier fed with spatial gradient features. Last, an ensemble classification framework is designed to make a decision, which considers the performance of both individual and concatenated selected regions.

Results: the evaluation is done on ADNI dataset. The proposed method outperforms other state-of-the-art ^{18}F -FDG PET-based algorithms for AD vs. NC with an accuracy, a sensitivity and a specificity values of 93.65%, 91.22% and 96.25%, respectively. For the case of pMCI vs. sMCI, the three metrics are 75.38%, 74.84% and 77.11%, which is significantly better than most existing methods. Besides, promising results are also achieved for multiple classifications under ^{18}F -FDG PET modality.

Conclusions: ^{18}F -FDG PET images can be characterized by spatial gradient features for diagnosing AD and its early stage, and the proposed ensemble framework can enhance the classification performance.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

There is a new case of dementia in the world every 3 s [1], and the most common form of dementia, which accounts for 60–70%, is Alzheimer's Disease (AD) [2]. It affects one in nine over 65 s [3] and one in three over 85 s [4]. It is estimated that 131 million

people will be living with the disease in 2050 [5]. At present, AD is incurable and irreversible, even if promising developments for treatments are expected to be achieved in the near future. Diagnosis at early stage of Mild Cognitive Impairment (MCI) is of importance in the therapeutic perspective to stop or slow down the neurodegenerative process.

^{18}F -FluoroDeoxyGlucose Positron Emission Tomography (^{18}F -FDG PET, referred to as FDG-PET hereafter) has proved to be one of the effective modalities for identifying AD or MCI since it can detect the subtle changes in the cortex [6–8]. Therefore, a variety of studies have been published in either predicting MCI conversion or differentiating AD from Normal Control (NC) using FDG-PET modality. Among them, considerable attention has been dedicated to developing Computer-Aided Diagnosis (CAD) approaches for AD. Their basic idea is to investigate characteristic patterns in FDG-PET

* Corresponding author at: Aix Marseille Univ, Marseille 13013, France.

E-mail address: mouloud.adel@univ-amu.fr (M. Adel).

¹ 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.

images through combining medical image processing and machine learning techniques, thereby achieving diagnosis goals.

A fundamental step in CAD methods is feature extraction from images. There have been many studies on feature extraction for Alzheimer Disease and its prodromal stage [9]. Feature extraction approaches from FDG-PET images could be roughly grouped into two main categories: voxel-wise and region-wise. Voxel-based methods could include the whole voxels of each brain image, as has been done by Hinrichs et al. [10] who utilized all the voxels to predict AD under a boosting framework. In this case, even though using the whole voxels preserves all the information, it can lead to feature redundancy, which is a burden for training a classifier. Voxel-based methods could also use a set of discriminative voxels in order to avoid feature redundancy. Cabral et al. [11] selected typical voxels through using mutual information, which obtained promising performance for predicting MCI conversion. Padilla et al. [12] applied non-negative matrix factorization to select and extract the most relevant voxels for the purpose of diagnosing AD. Region-based methods generally use a pre-defined anatomical atlas to segment the FDG-PET image into different Regions of Interest (ROIs) in order to compute statistical or other parameters that can represent glucose rates within ROIs or connections between them. Pagani et al. [13], Li et al. [14], Cheng et al. [15] and Gray et al. [16] used the mean intensity value of each ROI as the feature, whereas Garali et al. [17] used first order statistics (mean, variance, skewness and kurtosis) and entropy extracted from the grey-level histogram of each ROI as features to identify AD. Pan et al. [18] computed connectivities among ROIs and exploited these connections as features to tackle the problem of AD diagnosis.

Another important part of CAD is machine learning techniques. Traditional machine learning methods, such as Support Vector Machine (SVM) [19], random forests [20], are usually applied to do the classification task. Cabral et al. [11] and Padilla et al. [12] used SVM and FDG-PET images to predict MCI conversion and identify AD, respectively. Gray et al. [21] exploited random forests to derive the pairwise similarity measures from features and then made a classification by using random forests as well for multi-modality, in which FDG-PET was included. Recently, deep learning techniques have shown impressive performance in recognition and classification tasks [22,23]. Therefore, a great amount of research have turned to the use of neural networks to address the problem of AD diagnosis and obtained satisfactory results [24–26]. Even though deep learning models have made significant progress, their interpretability still needs to be investigated.

In this paper, the feature is represented from another point of view, the spatial gradient of FDG rates in PET brain images, instead of voxel-wise and ROI-wise features as many studies have done previously. This work is motivated by the observed differences of FDG rates between NC and AD subjects. The spatial gradient is quantified by a 2D histogram of orientation, which is similar to Histogram of Oriented Gradient (HOG) [27] that has been successfully applied for object detection in 2D images. Our contributions can be summarized into three aspects: 1) 1D HOG descriptor, used in natural scene images originally, is improved to 2D HOG to quantify spatial gradients, thereby characterizing 3D FDG-PET brain images. Moreover, 2D HOG is expressed as Small Scale HOG (SSH) and Large Scale HOG (LSH), which proves to be more effective than the commonly used features, voxel-wise and ROI-wise; 2) a region ranking method is proposed to select distinctive ROIs by using multiple SSH features; 3) an ensemble classification framework is designed through considering both small and large scales 2D HOG for individual and concatenated regions, which enhances the diagnosis accuracy for both binary and multiple classifications.

The rest of the paper is organized as follows. In Section 2, multiscale spatial gradient features for diagnosing AD is described. Section 3 presents and analyzes experimental results, and in

Section 4, a discussion is given. Finally, a conclusion of this work is given in Section 5.

2. Methods

In this paper, FDG-PET images are characterized by spatial gradients for AD diagnosis. First, the spatial gradient of FDG-PET image is computed and then 90 ROIs are extracted from the gradient image through Automated Anatomical Labeling (AAL) [28] atlas, in which the cerebellum is not considered. Next, some distinctive ROIs are selected through a proposed ROI ranking method which considers multiple SSH descriptors of each region. Finally, an ensemble classifier is trained under the selected ROIs by using SSH and LSH features. Fig. 1 illustrates the flowchart of the proposed diagnosis method.

2.1. Dataset

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD.

In ADNI dataset, there are four types of FDG-PET data,² which are 1) Co-registered Dynamic; 2) Co-registered, Averaged; 3) Co-reg, Avg, Standardized Image and Voxel Size; 4) Co-reg, Avg, Std Img and Vox Siz, Uniform Resolution. The proposed method is evaluated on FDG-PET images from type 3). After downloading the baseline data, images are spatially normalized to MNI template using SPM12 [29] with $2 \times 2 \times 2 \text{ mm}^3$ voxel size and $91 \times 109 \times 91$ tensor dimension. This normalization is based on two steps: a global affine transformation followed by a non-rigid spatial transformation. For each step, a minimization of the mean squared difference between the image to be normalized and a linear combination of one or more template images. The general affine transformation assumes a 12-parameter model whereas the non-rigid spatial transformation is based on the combination of the lowest frequency components of the three dimensional cosine transform. Next, the intensity normalization is performed through dividing each voxel intensity by the average value of global gray matter which is extracted under the aid of AAL template, and the data is with 32-bit depth. Thereafter, images are further smoothed by a Gaussian kernel with a full width at half maximum of 8 mm following many related studies [10,14,17,25,30]. Through the pre-processing pipeline, experimental data are converted to a standard space, among which 741 images are selected according to the diagnostic information³ (LONI → Download → Study Data → Assessments → Diagnosis → Diagnostic Summary). Specifically, there are 247 AD subjects who stay in AD stage without conversion, 246 NC subjects who stay in NC stage without progression, 123 progressive MCI (pMCI) subjects who have progressed from MCI stage to AD in the available scan time and 125 stable MCI (sMCI) subjects who have not converted to AD in the follow-up period which lasts at least 24 months. A full list of subject IDs used in this paper can be found in the Supplementary Material. The demographic and clinical information of subjects is provided in Table 1, in which MMSE stands for the Mini-Mental State Examination.

² <http://adni.loni.usc.edu/methods/pet-analysis-method/pet-analysis/>.

³ <https://ida.loni.usc.edu/>.

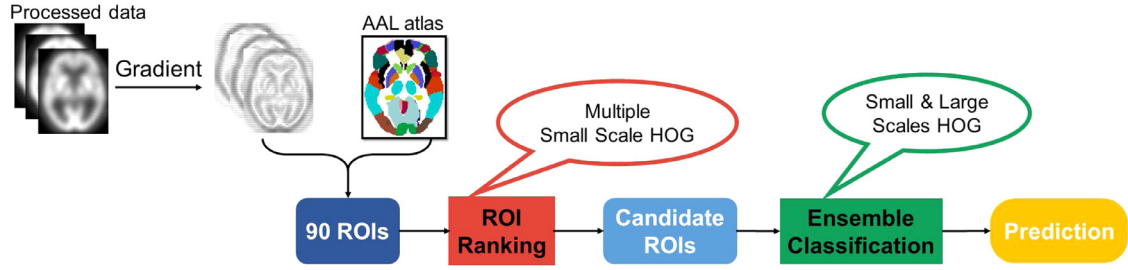


Fig. 1. The flowchart of the proposed method.

Table 1
Demographic and clinical information of subjects.

Characteristic	AD	NC	pMCI	sMCI
Number of subjects	247	246	123	125
Female/male	103/144	119/127	51/72	44/81
Age(Mean \pm SD*)	75.09 \pm 8.01	74.41 \pm 6.09	74.01 \pm 6.87	74.45 \pm 7.86
MMSE(Mean \pm SD)	23.17 \pm 2.13	29.04 \pm 1.18	26.81 \pm 1.64	27.66 \pm 1.69

* SD—Standard Deviation

2.2. Histogram of oriented gradient for 2D images

Histogram of Oriented Gradient (HOG) is a descriptor which was proposed in for human detection [27]. The fundamental idea is that the local object appearance and shape within an image can be represented by the distribution of intensity gradients or edge directions. Generally, the image is divided into small connected regions, and for the pixels within each region, a histogram of gradient directions is computed. The descriptor is the concatenation of these histograms. HOG feature is an effective hand-crafted descriptor for object detection since it can capture the edge or gradient structure which is discriminative for local shape [27].

2.3. 2D histogram of oriented gradient for FDG-PET images

Fig. 2 shows the difference between an NC (top row) and an AD (bottom) subjects. Fig. 2(a) displays one of the slices and the circled area belongs to region Parietal_Inf_R in AAL template. The enlarged areas are shown in Fig. 2(b) (different colors indicate different intensities). It can be clearly seen that the intensities are different between NC and AD PET scans. In addition, the gradients are also different, as shown in Fig. 2(c) (different colors indicate different gradient magnitudes), and the gradient change of the AD scan (bottom) is more obvious than that of NC (top). This observation drove us to investigate the effectiveness of spatial gradients in diagnosing AD. In this paper, we exploit a 2D histogram of orientation to quantify spatial gradients in order to characterize FDG-PET images.

2.3.1. Spatial gradient computation

For a 2D image, the gradient is computed from the horizontal and vertical directions, and the corresponding orientation is determined by one angle. Similarly, the spatial gradient of a 3D image is calculated in the x , y and z directions and the orientation is decided by two angles.

For a voxel with an intensity $f(x, y, z)$ at the position (x, y, z) , its numerical gradient can be computed as:

$$\begin{aligned} g_x &= 0.5 \times (f(x+1, y, z) - f(x-1, y, z)) \\ g_y &= 0.5 \times (f(x, y+1, z) - f(x, y-1, z)) \\ g_z &= 0.5 \times (f(x, y, z+1) - f(x, y, z-1)) \end{aligned} \quad (1)$$

where g_x , g_y and g_z are gradients in the x , y and z directions, respectively. The magnitude $\|g\|$ is obtained through:

$$\|g\| = \sqrt{g_x^2 + g_y^2 + g_z^2} \quad (2)$$

and the orientation is represented by the polar angle, ϕ , and the azimuth angle, θ , as shown in Fig. 3

$$\begin{aligned} \phi &= \arctan\left(\frac{g_y}{g_x}\right) \\ \theta &= \arccos\left(\frac{g_z}{\|g\|}\right) \end{aligned} \quad (3)$$

where ϕ is in the range $[-180^\circ, 180^\circ]$ and θ is in the range $[0^\circ, 180^\circ]$.

2.3.2. Spatial gradient quantification

In this step, a 2D histogram is constructed based on gradient orientations (ϕ and θ), and the magnitude is used to count the occurrence of a certain orientation. Specifically, ϕ and θ are viewed as two properties of the 2D histogram, and then are evenly divided into several intervals or bins, respectively. Last, if the gradient orientation is within a certain interval, the value for that interval is accumulated. According to [27], the value is computed via a function of the gradient magnitude. Considering the magnitude of each voxel is small, the exponential function of magnitude, g_e , is applied as the counting value in each interval,

$$g_e = \exp(\|g\|) \quad (4)$$

Consequently, the FDG-PET image can be represented by a 2D histogram, and meanwhile, the representation will vary with the number of bins in the histogram. Fig. 4 shows histograms with different numbers of bins for the same FDG-PET image, where the top row is for an NC subject and the bottom row is a subject with AD. It can be seen that for a subject, either NC or AD, a change in the number of bins can result in different representations (each row). Moreover, the difference between NC and AD is obvious under the same number of bins (each column), which also implies that it is reasonable to use 2D HOG as the feature to diagnose AD.

Compared with Global HOG which is computed on the whole subject without segmentation, calculating locally is also worth considering. To achieve this goal, each subject is divided into 90 regions by using AAL atlas, and a histogram can be constructed for each region. It should be noted that the 2D histogram is computed in irregular regions which contain different numbers of voxels. Since the number of bins in a histogram is adjustable, multiple

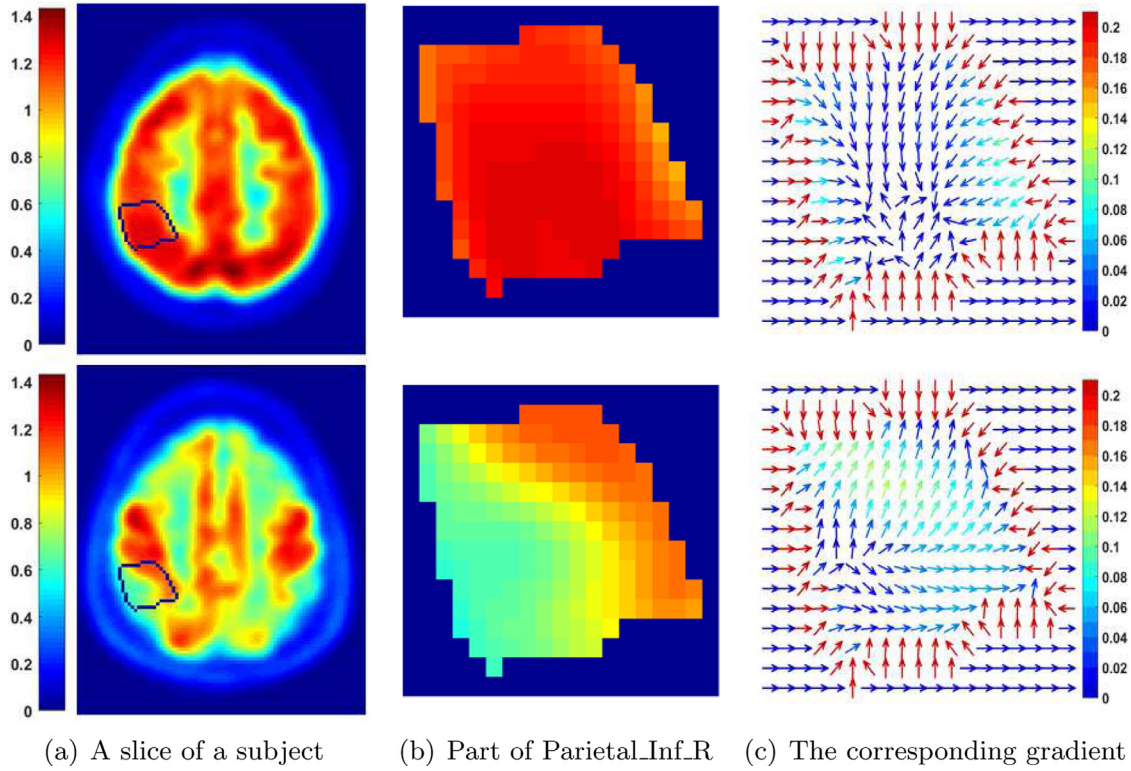


Fig. 2. An instance to differentiate between NC and AD via intensity and gradient, where the top row is for an NC subject and the bottom row is for a subject with AD.

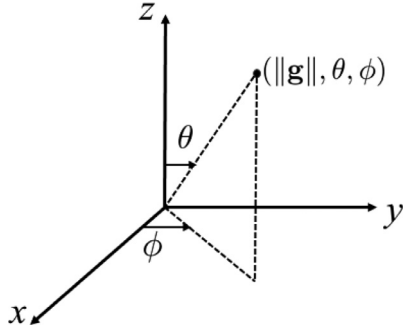


Fig. 3. An instance of the polar angle ϕ and azimuth angle θ .

scales of 2D HOG features are extracted from each ROI to make the features informative, including 18×18 ($\phi \times \theta$), 12×18 , 18×9 , 12×9 and 12×6 bins, which are denoted Small Scale HOG (SSH) because of the small interval used in gridding angles for constructing a histogram. Correspondingly, some Large Scale HOG (LSH) descriptors, in which a large interval is used to grid angles, are also extracted from each ROI, including 5×5 ($\phi \times \theta$), 1×1 bins. The features from different ROIs exhibit different ranges, so normalization is essential in order to achieve good performance. In this paper, L_2 norm is used to do the normalization.

$$\mathbf{h}' = \frac{\mathbf{h}}{\|\mathbf{h}\|} \quad (5)$$

where \mathbf{h} is 2D HOG vector for each scale of any ROI in a subject, \mathbf{h}' stands for 2D HOG descriptor after normalization.

2.4. ROI ranking

The probability of AD occurring in each ROI is not consistent, which means that some ROIs are more likely to be affected by AD,

while others are not. In practice, doctors pay more attention to key areas as well. Therefore, it is necessary to select typical ROIs which are more susceptible to AD. A region ranking method is developed to achieve the goal of region selection. Specifically, 2D HOG features are applied to characterize each ROI and then fed into a linear SVM to compare each ROI's classification accuracy, thereby ranking ROIs. In order to obtain a robust and reliable result, multiple SSH features are considered, and the average accuracy is used for ranking ROIs. The reason LSH is not taken into account is that the feature dimension of LSH is relatively low, for example, the dimension of 1×1 LSH is only 1. Fig. 5 shows the framework of the ranking method. A region with a higher classification accuracy implies its stronger ability to recognize AD. At last, some ROIs can be selected through the ranking order. The ranking results are analyzed in Section 3.3.

2.5. Ensemble classification

After ranking ROIs, the top N ROIs with higher performance are selected as candidate regions for the classification. SVM is a popular and effective classifier in AD diagnosis [12,16,31]. In this study, SVM is applied in an ensemble classification framework which considers both SSH and LSH features of selected ROIs. The motivation of designing an ensemble framework is inspired by the idea that weak classifiers can be combined into a strong classifier. Although SVM is usually treated as a strong classifier, we make it become a relatively weak model through controlling input features (ROI by ROI, scale by scale) and setting the parameter C which is introduced in Section 3.1. In addition, concatenating all the scales of 2D HOG and all the regions is not an efficient way to train a model due to too many inputs. Therefore, the selected regions guided by the ROI ranking method are utilized to train a set of weak classifiers, and then multiple classifiers are integrated to make a prediction. The framework of the ensemble classifier is briefly presented in Fig. 6. Specifically, five types of SSH

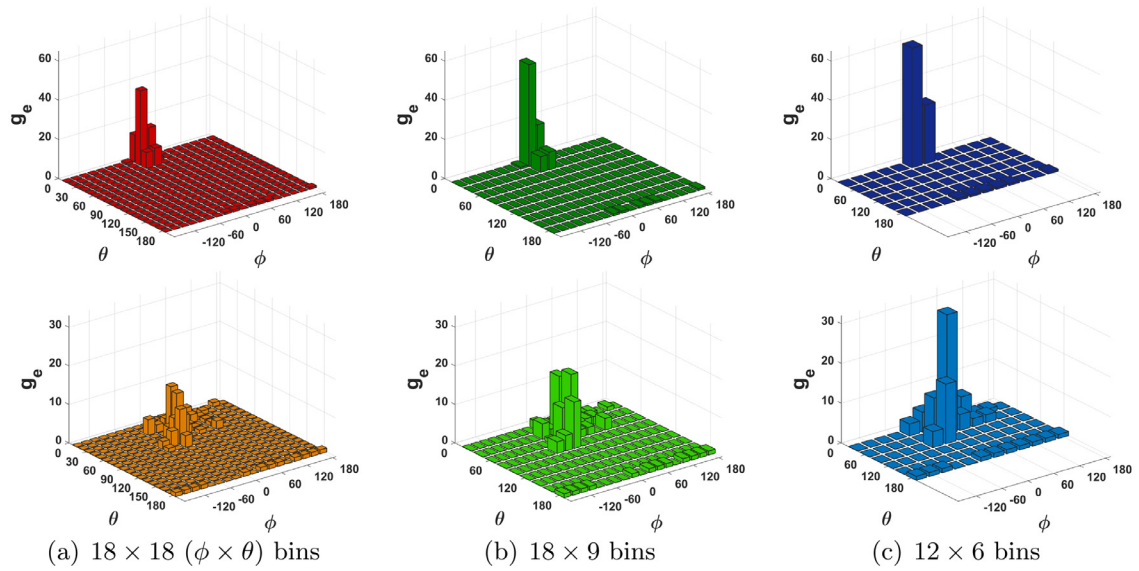


Fig. 4. Improved 2D Histogram of Oriented Gradient (HOG) for an NC subject (top row) and an AD subject (bottom row) without segmentation.

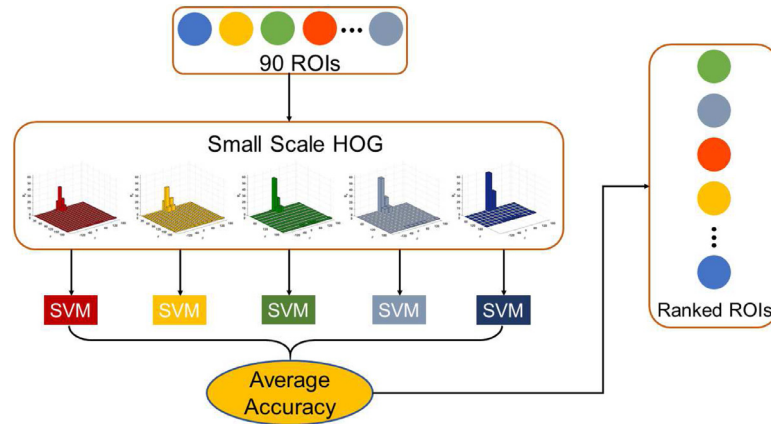


Fig. 5. The framework of region ranking method.

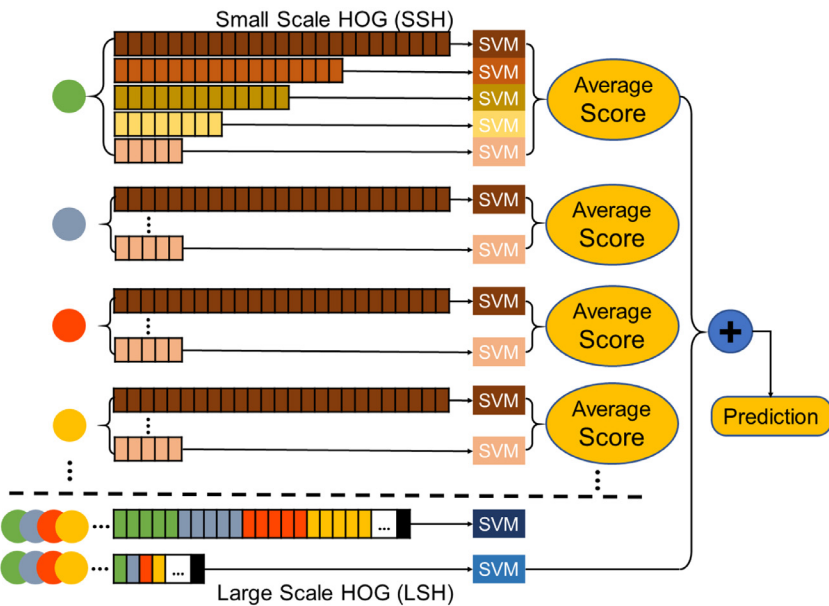


Fig. 6. The framework of the ensemble classification, circles with different colors indicate different regions.

Table 2
Performance of different type of feature for AD vs. NC(%).

Feature	Dimension	ACC	SEN	SPE	AUC
Voxel	160, 990	91.25	90.44	92.05	96.43
ROI	90	86.47	85.23	87.71	93.40
Global HOG	162	63.75	62.37	64.95	68.89
LSH	2250	92.64	91.45	93.95	97.31
SSH	14,580	93.37	91.77	95.10	97.49

features ($18 \times 18 (\phi \times \theta)$, 12×18 , 18×9 , 12×9 and 12×6 bins) are extracted from each selected ROI and then are used to train five classifiers, respectively. The average score of five classifiers is considered as the corresponding ROI's output, \mathbf{S} , which is expressed as:

$$\mathbf{S} = \frac{1}{T} \sum_{t=1}^T \mathbf{s}_t \quad (6)$$

where \mathbf{s}_t is the output score of SVM with t -th scale of SSH descriptor and T is the number of scales of SSH, here $T = 5$. For LSH, all the features of candidate ROIs are concatenated to feed into a classifier because of the low feature dimension. The final decision, \mathbf{Y} , is made through an addition strategy of two parts' outputs, small and large scales,

$$\mathbf{Y} = \text{sgn} \left(\sum_{i=1}^N \mathbf{S}_i + \sum_{j=1}^M \mathbf{L}_j \right) \quad (7)$$

where $\text{sgn}(\cdot)$ is a sign function, N is the number of candidate ROIs, M is the number of scales of LSH, $M = 2$, and \mathbf{L} is the score of LSH-based classifier. Therefore, the proposed ensemble classifier not only considers the performance of each individual ROI, but also considers the performance of cascade regions.

3. Results

3.1. Setup

In order to test the effectiveness of the proposed method, the experiments are mainly conducted on two classification tasks, AD vs. NC and pMCI vs. sMCI. We apply four metrics, namely classification accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under curve (AUC), to evaluate the corresponding performance. For all the metrics, a higher value indicates better performance. Specifically, ACC is the proportion of samples that are correctly predicted. SEN implies the proportion of correctly classified AD or pMCI samples. SPE represents the proportion of NC or sMCI samples that are correctly classified. Due to the limited number of subjects, we use a 10-fold cross-validation technique to assess the performance and repeat 10 times to reduce the possible bias. The margin parameter C of all the SVMs used in the ensemble classifier is set to 0.5 in order to construct relatively weak classifiers. In this paper, the SVM algorithm is implemented with the LIBSVM toolbox [32].

3.2. Evaluation on spatial gradient feature for FDG-PET

The spatial gradient feature is compared with commonly used voxel-wise and ROI-wise features. The comparison results of different representations are presented in Tables 2 and 3 for AD vs. NC and pMCI vs. sMCI, respectively. The terms 'Voxel' and 'ROI' stand for classification results which obtained by using voxel intensity and region's mean intensity, respectively. 'Global HOG' means 2D HOG descriptor is computed on the subject without parcellation. The results of Global HOG and SSH shown in the following two tables are achieved based on a histogram with 18×9 bins, while for LSH, its results are computed on a histogram with 5×5 bins.

Table 3
Performance of different type of feature for pMCI vs. sMCI(%).

Feature	Dimension	ACC	SEN	SPE	AUC
Voxel	160, 990	62.79	61.80	64.76	68.94
ROI	90	64.46	63.29	65.84	69.75
Global HOG	162	50.57	52.63	50.39	51.90
LSH	2250	64.74	64.55	65.74	69.83
SSH	14,580	66.53	65.24	68.18	72.80

It can be seen from Table 2 that voxel intensity is a kind of effective feature in classifying AD from NC, which achieves an accuracy of 91.25%. But the feature dimension is too large, which is a drawback for training a model. Even though the dimension of ROI-wise feature is small, its performance is not satisfactory and it is not as effective as voxel-wise feature, with an accuracy rate of 86.47%. Global HOG, with 63.75% accuracy, cannot achieve a noteworthy result. It can be explained that: 1) Global HOG is computed on the whole subject without considering local details; 2) the dimension, 162 (18×9), seems to be a satisfactory one, but the effective information is much less than that because of a lot of zero values, like Fig. 4(b), which is less informative. LSH and SSH descriptors can guarantee the performance and meanwhile tackle the problem of large dimension. Even though 14580 ($18 \times 9 \times 90$) is not an absolutely desirable dimension, it is still acceptable compared to the dimension of the voxel-wise feature. In addition, owing to the adjustable number of bins in a histogram, a smaller dimension can be obtained if setting an appropriate number of bins.

As for classifying pMCI from sMCI, which is reported in Table 3, voxel-wise feature is inferior to ROI-wise feature, which is different to their performance in AD classification. The reason could be that the dimension of voxel-wise feature is far greater than that of ROI-wise feature, which can cause the problem of feature redundancy. Such a problem would harm the classifier training, especially for the case that the classifier could not work well. As can be seen from Table 3, all the features (voxel-wise, ROI-wise, Global HOG, LSH and SSH) cannot perform as well as in AD diagnosis, thus prediction of pMCI is a challenging task and it is easily influenced by redundant features. Nevertheless, SSH descriptor still has dominant performance with an accuracy of 66.53%, which is 2.07% higher than the ROI-wise feature, and LSH has comparative performance with the ROI-wise features. The other three metrics also indicate SSH is superior to the other features. Therefore, characterizing FDG-PET images by 2D HOG locally is effective and feasible. The reason could be that the spatial gradient is calculated at voxel level and the 2D histogram is computed at ROI level, which makes the 2D HOG descriptor become a *bridge* to link voxel-wise feature and ROI-wise feature.

3.3. Evaluation on ROI ranking method

Since different regions have different abilities to diagnose AD, a simple ROI ranking method using multiple SSH features and SVM is proposed. Specifically, for each region, five scales of SSH features are extracted and then fed into five SVMs, respectively. The average accuracy of classifiers is considered as the ranking metric, and the higher the accuracy is, the stronger the ability of the region to distinguish AD from NC. Multiple SSH features are applied to ensure the reliability and robustness of the results. Fig. 7 shows the maximum difference of accuracy in five scales of SSH, Δ_{acc} , for each ROI. As can be seen, the difference is obvious, especially for identifying pMCI from sMCI, and the highest Δ_{acc} is 17.84%, which proves that it is rational to use multiple SSH features to rank ROIs. Moreover, the change of maximum difference in pMCI prediction is larger than that of AD classification, which is because that the later task is easier than the former one. For AD classification,

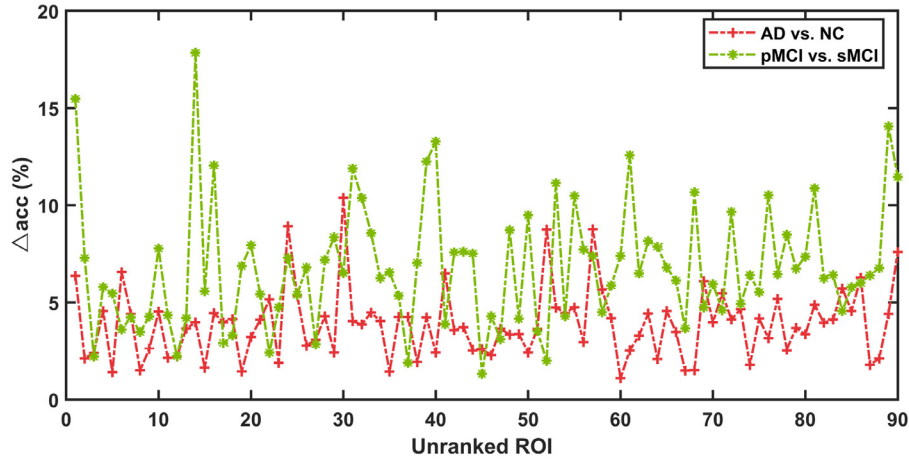


Fig. 7. Maximum difference of accuracy in five scales of SSH for each ROI.

Table 4

Top 20 ROIs for AD vs. NC and pMCI vs. sMCI.

Rank	AD vs. NC	pMCI vs. sMCI
1	Cingulum_Post_R	Cuneus_L
2	Cingulum_Post_L	Cingulum_Post_L
3	Hippocampus_L	Temporal_Inf_L
4	Hippocampus_R	Precuneus_L
5	Precuneus_R	Temporal_Mid_R
6	Cuneus_R	Rectus_L
7	Fusiform_R	Parietal_Inf_R
8	Temporal_Inf_L	Lingual_R
9	Cuneus_L	SupraMarginal_L
10	Parietal_Inf_R	Rectus_R
11	Precuneus_L	Temporal_Inf_R
12	Paracentral_Lobule_R	Cuneus_R
13	Paracentral_Lobule_L	Hippocampus_R
14	Temporal_Mid_L	SupraMarginal_R
15	Fusiform_L	Parietal_Inf_L
16	Parietal_Sup_R	Frontal_Sup_L
17	Occipital_Mid_R	Supp_Motor_Area_L
18	Temporal_Sup_L	Parietal_Sup_R
19	Occipital_Sup_R	Occipital_Mid_L
20	Angular_L	Angular_R

different scales of effective features usually achieve high-level accuracies but with a little differences. While for pMCI prediction, it is challenging, so different scales of features may not be that effective and may achieve unstable performance, which could cause a larger difference within different scales of features.

Table 4 presents the top 20 regions ranked by the proposed method for two tasks, AD vs. NC and pMCI vs. sMCI, which can be regarded as potential FDG-PET indicators for the subsequent classification tasks.

3.4. Evaluation on ensemble classification

An ensemble classification framework is designed by considering SSH and LSH descriptors together since these multiple scales of 2D HOG contain both specific (SSH) and general information (LSH). Fig. 8 shows the accuracies under different number of regions for AD vs. NC and pMCI vs. sMCI. It should be noted that the x axis denotes ROIs which have been ranked from most to least relevant to AD/pMCI according to the proposed ROI ranking method. As can be seen from Fig. 8(a), SSH achieves a higher ACC, 93.06%, with fewer regions (15 ROIs) than LSH whose best performance (ACC: 92.64%) is obtained by using 57 ROIs. Thus SSH descriptor performs better than LSH in AD diagnosis. In addition, the ensemble classification through integrating SSH and LSH improves the accu-

Table 5

Performance comparison for AD vs. NC(%).

Method	Subjects	ACC	SEN	SPE	AUC
Hinrichs et al. [10]	89AD + 94NC	84	84	82	87.16
Padilla et al. [12]	53AD + 52NC	86.59	87.50	85.36	–
Li et al. [14]	25AD + 30NC	89.1	92	86	97
Gray et al. [16]	50AD + 54NC	88.4	83.2	93.6	–
Pan et al. [18]	247AD + 246NC	90.55	90.46	90.92	95.22
Zhu et al. [30]	51AD + 52NC	93.3	–	–	–
Lu et al. [24]	226AD + 304NC	93.58	91.54	95.06	–
Liu et al. [25]	93AD + 100NC	91.2	91.4	91.0	95.3
Our method	247AD + 246NC	93.65	91.22	96.25	96.93

racy to 93.65% under top 15 ROIs, which achieves a slight increase for AD diagnosis (0.59%). In the case of pMCI vs. sMCI, as illustrated in Fig. 8(b), the performance of SSH is still superior to LSH and both of them reach the best accuracy with the top 10 regions, which are 73.92% and 68.81%. Furthermore, the ensemble classifier raises the accuracy to 75.38% with 10 ROIs, and can significantly improve performance under different numbers of ROIs. In summary, for the two tasks (AD vs. NC and pMCI vs. sMCI), LSH descriptor does not perform as well as SSH feature. It is because LSH contains more general features, which makes it less informative than SSH. Besides, the ensemble classifier is better than the SSH-based classifier, which proves that the integration strategy is effective. Moreover, the accuracy trend is a result of a trade-off between informative and redundant features. The highest accuracy is attributed to the optimal balance between effective features and a proper dimension, while for the decline of the accuracy trend, it is caused by the feature redundancy overwhelming its effectiveness.

3.5. Comparison with the state-of-the-art methods

The proposed method is compared with other methods which also address the problem of AD/pMCI diagnosis under the modality of FDG-PET, including methods of Hinrichs et al. [10], Padilla et al. [12], Li et al. [14], Gray et al. [16], Pan et al. [18], Zhu et al. [30], Lu et al. [24] and Liu et al. [25]. Lu's and Liu's methods applied the neural network technique, and other approaches were developed under the traditional classification framework. Noting that Pan's method is evaluated on the same dataset with the proposed method for the sake of a fairer comparison.

Tables 5 and 6 present the comparison results of two binary tasks, AD vs. NC and pMCI vs. sMCI. It can be seen that the proposed method outperforms others in terms of ACC and SPE for AD diagnosis, but it is inferior to Lu's method in respect of SEN and

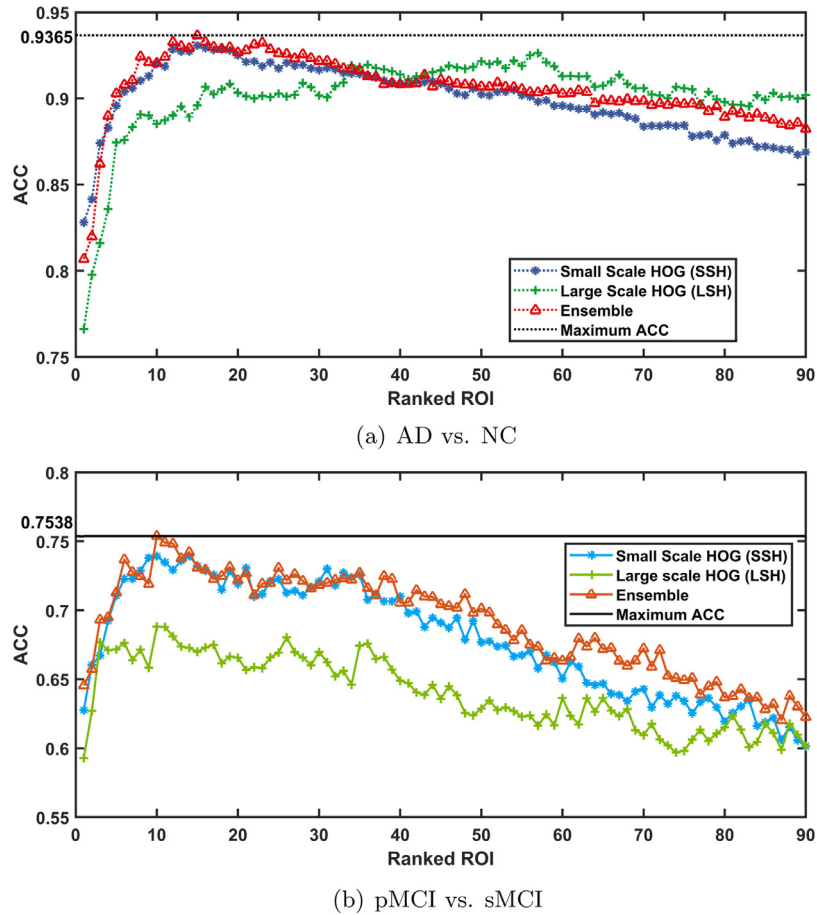


Fig. 8. Performance under different numbers of ROIs.

Table 6
Performance comparison for pMCI vs. sMCI(%).

Method	Subjects	ACC	SEN	SPE	AUC
Gray et al. [16]	53pMCI + 64sMCI	63.1	52.2	73.2	–
Pan et al. [18]	123pMCI + 125sMCI	71.10	68.04	74.38	74.78
Zhu et al. [30]	43pMCI + 56sMCI	69.9	–	–	–
Lu et al. [24]	112pMCI + 409sMCI	81.55	73.33	83.83	–
Our method	123pMCI + 125sMCI	75.38	74.84	77.11	80.70

Li's method in AUC, with a slight difference of 0.32% and 0.07%, respectively. For the case of distinguishing between pMCI and sMCI, our method is not as effective as Lu's method, which utilized multiscale deep neural networks in terms of ACC and SPE, but it should be noted that the data used in Lu's method for the task of pMCI vs. sMCI is unbalanced (pMCI: 112, sMCI: 409), which can enhance the accuracy to a certain degree.

Moreover, we also report the results of multiple classifications, AD vs. MCI (pMCI + sMCI) vs. NC and AD vs. pMCI vs. sMCI vs. NC, and compare them with the results of voxel-based, ROI-based and state-of-the-art methods, which are shown in Tables 7 and 8. It is worth noting that the number of subjects of AD and NC is reduced to 123 respectively in order to achieve balance data for the four-class classification task. Clearly, the proposed method still outperforms voxel-based and ROI-based approaches, especially for the task of four-class classification, where the improvement is significant. Besides, our method is also superior to the comparison methods, and the differences with the best result are 9.2% and 13.75% regarding three-class and four-class classifications, respectively. The results of multiple classifications are lower than those

Table 7
Performance comparison for AD vs. MCI vs. NC(%).

Method	Subjects	ACC
Voxel	247AD + 248MCI + 246NC	73.65
ROI	247AD + 248MCI + 246NC	64.73
Our method	247AD + 248MCI + 246NC	74.70
Zhu et al. [30]	51AD + 99MCI + 52NC	65.50
Zhou et al. [33]	190AD + 389MCI + 226NC	58

Table 8
Performance comparison for AD vs. pMCI vs. sMCI vs. NC(%).

Method	Subjects	ACC
Voxel	123AD + 123pMCI + 125sMCI + 123NC	60.57
ROI	123AD + 123pMCI + 125sMCI + 123NC	53.65
Our method	123AD + 123pMCI + 125sMCI + 123NC	70.04
Zhu et al. [30]	51AD + 43pMCI + 56sMCI + 52NC	56.29
Zhou et al. [33]	190AD + 157pMCI + 205sMCI + 226NC	48

of binary classifications, which indicates the multiple classification is a very challenging problem and more investigations need to be carried out in the future.

4. Discussion

This paper extends the feature in object detection, HOG, to FDG-PET brain images to aid AD diagnosis, which is effective as well. Compared to the classic representations (voxel-wise and ROI-wise), 2D HOG descriptor is more informative than the

ROI-wise feature and more sparse than the voxel-wise feature and meanwhile, can ensure effectiveness, which will motivate us to explore the effects of other descriptors used in natural scene images for FDG-PET images, such as LBP (Local Binary Pattern) [34], SIFT (Scale Invariant Feature Transform) [35]. Furthermore, 2D HOG can characterize different regions with the same feature dimension, even if each ROI contains a different number of voxels, which may be significant for some research and applications. As for the dataset, a key issue needs to be taken seriously. In fact, most state-of-the-art methods are not comparable because the experimental dataset is different. Even though most methods use the ADNI dataset, the image processing procedures and subject IDs are different, which greatly weakens the significance of comparison. Considering not all methods are open-source, evaluating methods over the same dataset is more practical.

5. Conclusion

We introduce 2D HOG to quantify spatial gradients thereby characterizing FDG-PET images instead of the common features, such as voxel-wise and ROI-wise features. Experimental results suggest that 2D HOG is more efficient and effective than voxel-wise and ROI-wise features, respectively. Besides, a ROI ranking method is proposed by applying multiple SSH descriptors and according to which, a set of candidate ROIs are selected to assist the diagnosis. Furthermore, an ensemble classification framework is designed over the selected regions through using SSH and LSH features. The ensemble classifier is effective and outperforms other methods according to the evaluation on the ADNI dataset. Meanwhile, we test the proposed method on multiple classification tasks as well, and improvements are also significant.

Declaration of Competing Interest

The authors have no conflicts of interest to declare.

Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

This work was partly supported by the Chinese Scholarship Council.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.cmpb.2019.105027](https://doi.org/10.1016/j.cmpb.2019.105027).

References

- [1] C. Patterson, World Alzheimer Report 2018: The State of the Art of Dementia Research: New Frontiers, Technical Report, Alzheimer's Disease International, London, UK, 2018.
- [2] A. Burns, S. Iliffe, Alzheimer's disease, *BMJ* 338 (2009).
- [3] Alzheimer's Association, 2015 Alzheimer's disease facts and figures, *Alzheimer's Dement* 11 (3) (2015) 332.
- [4] L.E. Hebert, J. Weuve, P.A. Scherr, D.A. Evans, Alzheimer disease in the United States (2010–2050) estimated using the 2010 census, *Neurology* 80 (19) (2013) 1778–1783.
- [5] M. Prince, A. Comas-Herrera, M. Knapp, M. Guerchet, M. Karagiannidou, World Alzheimer Report 2016: Improving Healthcare for People Living with Dementia: Coverage, Quality and Costs now and in the Future, Technical Report, Alzheimer's Disease International, London, UK, 2016.
- [6] C.M. Davison, J.T. O'Brien, A comparison of FDG-PET and blood flow SPECT in the diagnosis of neurodegenerative dementias: a systematic review, *Int. J. Geriatr. Psychiatry* 29 (6) (2014) 551–561.
- [7] L. Mosconi, V. Berti, L. Glodzik, A. Pupi, S.D. Santi, M.J. de Leon, Pre-clinical detection of Alzheimer's disease using FDG-PET, with or without amyloid imaging, *J. Alzheimer's Dis.* 20 (3) (2010) 843–854.
- [8] L. Rice, S. Bisdas, The diagnostic value of FDG and amyloid PET in Alzheimer's disease—a systematic review, *Eur. J. Radiol.* 94 (2017) 16–24.
- [9] 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* 155 (2017) 530–548.
- [10] C. Hinrichs, V. Singh, L. Mukherjee, Guo.Xu, M.K. Chung, S.C. Johnson, Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset, *NeuroImage* 48 (1) (2009) 138–149.
- [11] C. Cabral, P.M. Morgado, D.C. Costa, M. Silveira, Predicting conversion from MCI to AD with FDG-PET brain images at different prodromal stages, *Comput. Biol. Med.* 58 (2015) 101–109.
- [12] P. Padilla, M. López, J.M. Górriz, J. Ramirez, D. Salas-Gonzalez, I. Alvarez, NM-F-SVM based CAD tool applied to functional brain images for the diagnosis of Alzheimer's disease, *IEEE Trans. Med. Imaging* 31 (2) (2012) 207–216.
- [13] M. Pagani, F.D. Carli, S. Morbelli, J. Öberg, A. Chincari, G.B. Frisoni, S. Galuzzi, R. Perneczky, A. Drzezga, B.N.M. van Berckel, R. Ossenkoppele, M. Didic, E. Guedj, A. Brugnolo, A. Picco, D. Arnaldi, M. Ferrara, A. Buschiazzi, G. Sambucetti, F. Nobili, Volume of interest-based [¹⁸F] fluorodeoxyglucose PET discriminates MCI converting to Alzheimer's disease from healthy controls. A European Alzheimer's Disease Consortium (EADC) study, *NeuroImage* 7 (2015) 34–42.
- [14] R. Li, R. Perneczky, I. Yakushev, S. Förster, A. Kurz, A. Drzezga, S. Kramer, Gaussian mixture models and model selection for [¹⁸F] fluorodeoxyglucose positron emission tomography classification in Alzheimer's disease, *PLoS One* 10 (4) (2015).
- [15] B. Cheng, M. Liu, D. Zhang, B.C. Munsell, D. Shen, Domain transfer learning for MCI conversion prediction, *IEEE Trans. Biomed. Eng.* 62 (7) (2015) 1805–1817.
- [16] K.R. Gray, R. Wolz, R.A. Heckemann, P. Aljabar, A. Hammers, D. Rueckert, Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease, *NeuroImage* 60 (1) (2012) 221–229.
- [17] I. Garali, M. Adel, S. Bourenane, E. Guedj, Histogram-Based features selection and volume of interest ranking for brain PET image classification, *IEEE J. Transl. Eng. Health Med.* 6 (2018) 1–12.
- [18] X. Pan, M. Adel, C. Fossati, T. Gaidon, E. Guedj, Multi-level feature representation of FDG-PET brain images for diagnosing Alzheimer's disease, *IEEE J. Biomed. Health Inform.* (2018), doi:[10.1109/JBHI.2018.2857217](https://doi.org/10.1109/JBHI.2018.2857217).
- [19] C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.* 20 (3) (1995) 273–297.
- [20] L. Breiman, Random forests, *Mach. Learn.* 45 (1) (2001) 5–32.
- [21] K.R. Gray, P. Aljabar, R.A. Heckemann, A. Hammers, D. Rueckert, Random forest-based similarity measures for multi-modal classification of Alzheimer's disease, *NeuroImage* 65 (2013) 167–175.
- [22] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [23] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.
- [24] D. Lu, K. Popuri, G.W. Ding, R. Balachandrar, M. FaisalBeg, Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease, *Med. Image Anal.* 46 (2018) 26–34.
- [25] M. Liu, D. Cheng, W. Yan, Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images, *Front. Neuroinform.* 12:35 (2018).
- [26] Y. Ding, J.H. Sohn, M.G. Kawczynski, H. Trivedi, R. Harnish, N.W. Jenkins, D. Lit-tuiev, T.P. Copeland, M.S. Aboian, C.M. Aparici, S.C. Behr, R.R. Flavell, S.Y. Huang, K.A. Zalocusky, L. Nardo, Y. Seo, R.A. Hawkins, M.H. Pampaloni, D. Hadley,

- B.L. Franc, A deep learning model to predict a diagnosis of Alzheimer disease by using 18F-FDG PET of the brain, *Radiology* 290 (2) (2018) 456–464.
- [27] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.
- [28] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, M. Joliot, Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain, *NeuroImage* 15 (1) (2002) 273–289.
- [29] W.D. Penny, K.J. Friston, J.T. Ashburner, S.J. Kiebel, T.E. Nichols, *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, Elsevier, London, UK, 2011.
- [30] X. Zhu, H. Suk, S. Lee, D. Shen, Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification, *IEEE Trans. Biomed. Eng.* 63 (3) (2016) 607–618.
- [31] M. Liu, D. Zhang, D. Shen, Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment, *IEEE Trans. Med. Imaging* 35 (6) (2016) 1463–1474.
- [32] C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines, *ACM Trans. Intell. Syst. Technol.* 2 (3) (2011) 27.
- [33] T. Zhou, K.H. Thung, X. Zhu, D. Shen, Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis, *Hum. Brain Mapp.* 40 (3) (2019) 1001–1016.
- [34] T. Ojala, M. Pietikinen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, *Pattern Recognit.* 29 (1) (1996) 51–59.
- [35] D.G. Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vis.* 60 (2) (2004) 91–110.