Radiomics: a novel feature extraction method for brain neuron degeneration disease using ^{18}F-FDG PET imaging and its implementation for Alzheimer's disease and mild cognitive impairment

Yupeng Li, Jiehui Jiang [ID], Jiaying Lu, Juanjuan Jiang, Huiwei Zhang and Chuantao Zuo and the Alzheimer's Disease Neuroimaging Initiative

Abstract

Background: Alzheimer's disease (AD) is the most common form of progressive and irreversible dementia, and accurate diagnosis of AD at its prodromal stage is clinically important. Currently, computer-aided diagnosis of AD and mild cognitive impairment (MCI) using ^{18}F-fluorodeoxy-glucose positron emission tomography (^{18}F-FDG PET) imaging is usually based on low-level imaging features or deep learning methods, which have difficulties in achieving sufficient classification accuracy or lack clinical significance. This research therefore aimed to implement a new feature extraction method known as radiomics, to improve the classification accuracy and discover high-order features that can reveal pathological information.

Methods: In this study, ^{18}F-FDG PET and clinical assessments were collected in a cohort of 422 individuals [including 130 with AD, 130 with MCI, and 162 healthy controls (HCs)] from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and 44 individuals (including 22 with AD, and 22 HCs) from Huashan Hospital, Shanghai, China. First, we performed a group comparison using a two-sample Student's *t* test to determine the regions of interest (ROIs) based on 30 AD patients and 30 HCs from ADNI cohorts. Second, based on two time scans of 32 HCs from ADNI cohorts, we used Cronbach's alpha coefficient for radiomic feature stability analyses. Pearson's correlation coefficients were regarded as a feature selection criterion, to select effective features associated with the clinical cognitive scale [clinical dementia rating scale in its sum of boxes (CDRSB); Alzheimer's disease assessment scale (ADAS)] with 500-times cross-validation. Finally, a support vector machine (SVM) was used to test the ability of the radiomic features to classify HCs, MCI and AD patients.

Results: As a result, we identified brain regions which were mainly distributed in the temporal, occipital and frontal areas as ROIs. A total of 168 radiomic features of AD were stable (alpha > 0.8). The classification experiment led to maximal accuracies of 91.5%, 83.1% and 85.9% for classifying AD *versus* HC, MCI *versus* HCs and AD *versus* MCI.

Conclusion: The research in this paper proved that the novel approach based on high-order radiomic features extracted from ^{18}F-FDG PET brain images that can be used for AD and MCI computer-aided diagnosis.

Keywords: ^{18}F-FDG PET, Alzheimer's disease, mild cognitive impairment, radiomics

Correspondence to:
Jiehui Jiang
Shanghai Institute for Advanced Communication and Data Science, Shanghai University, T Building, 99 ShangDa Road, BaoShan District, Shanghai, 200444, China
jiangjiehui@shu.edu.cn

Chuantao Zuo
PET Center, Huashan Hospital, Fudan University, Shanghai, China
Institute of Functional and Molecular Medical Imaging, Fudan University, Shanghai, China
zuochuantao@fudan.edu.cm

Yupeng Li
Juanjuan Jiang
Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai, China

Jiaying Lu
Huiwei Zhang
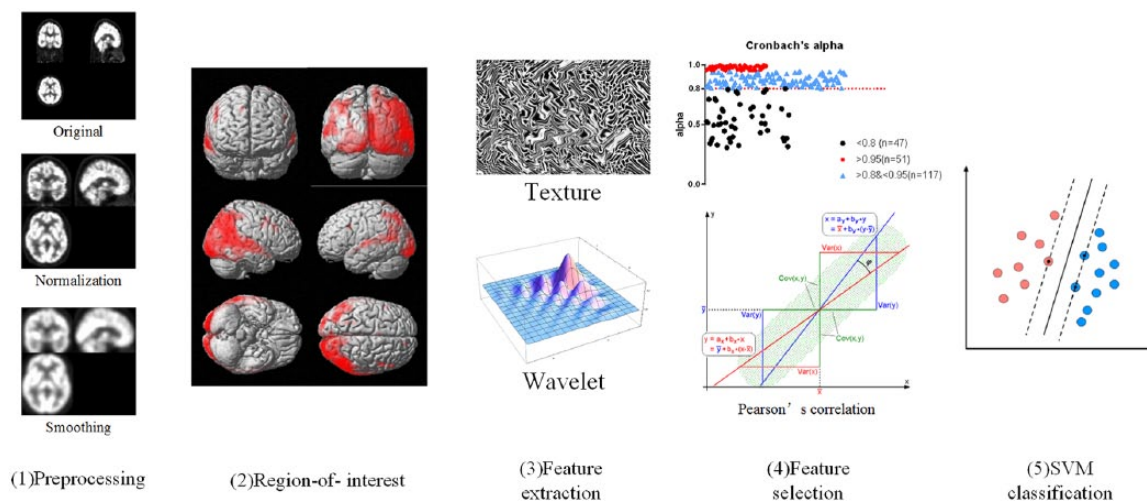PET Center, Huashan Hospital, Fudan University, Shanghai, China

Figure 1. Workflow of the analysis methods in this study, which comprised five steps: image preprocessing, image preprocessing, identification and extraction of regions of interest, feature extraction, feature selection, and SVM classification.
SVM, support vector machine.

Introduction

Alzheimer's disease (AD) is the most common form of progressive and irreversible dementia, with occurrences doubling approximately every 5 years after the age of 65 years. Presently, approximately 90 million people have been diagnosed with AD, and it is estimated that the number of AD patients will reach 300 million by 2050.[1] Mild cognitive impairment (MCI), which is considered a precursor of AD, has been an increasingly common target of potential therapeutic trials.[2] Currently, there is no effective cure solution for AD; thus, the early detection at its prodromal stage and accurate diagnosis of AD are important for patient care and developing future treatment.[3]

In recent years, positron emission tomography (PET) imaging technology has been widely used in the diagnosis, classification and rehabilitation evaluation of AD. In particular, ^{18}F-fluorodeoxy-glucose (^{18}F-FDG) PET, which is a functional molecular imaging modality performed utilizing glucose metabolic activity and distribution *via* imaging agents,[4] has been proven to be a valid tool to help doctors diagnose AD and MCI.[5]

Currently, there are two primary categories of computer-aided diagnoses of AD and MCI based on ^{18}F-FDG PET imaging: (1) the numerical distribution method of glucose uptake values, and its common imaging markers are usually low-order and simple image features[6]; (2) machine learning and deep learning techniques. Gary and colleagues[7] used multi-region PET image fusion structural magnetic resonance imaging (MRI) to extract the average signal intensity per cubic millimeter in each region as features to classify AD patients and healthy controls (HCs), achieving an accuracy of 82%. Silveira and colleagues[5] proposed a boosting classification technique based on simple classifier hybridization for the diagnosis of AD and MCI, with average accuracies of 90.97% and 79.63%, respectively. Liu and colleagues[3] invented a classification framework based on the combination of two-dimensional convolutional neural networks and recurrent neural networks, and the classification accuracies of AD or MCI *versus* HCs could reach 91.2% and 78.9%, respectively. However, the aforementioned low-order markers used in numerical methods were mostly hand-crafted, original, and low-level features that could not reveal the neuropathological heterogeneity of brain tissue and could hardly achieve high classification accuracy.[3–5] By comparison, although the deep learning frameworks could achieve better diagnostic accuracy, these methods only obtain the calculation values without clinical significance; thus, we cannot determine the association between intermediate values and the disease. For clinical practice, finding image features that can provide pathological information about diseases and using them for efficient classification diagnosis have great implications for physicians.

In this paper, to find high-level image features and develop efficient and accurate diagnostic tools, we proposed an emerging method, radiomics, for ^{18}F-FDG PET image feature extraction. The term 'radiomics' refers to the extraction and analysis of large amounts of advanced and high-order quantitative features with high-throughput from medical images.[8,9] These radiomic features could not only effectively diagnose disease and assist in treatment but also reveal the in-depth information hidden in the images that may help develop personalized and accurate medical plans.[8–10] Radiomics has been well developed in oncological studies.[11–14] Although current radiomics research studies have been mainly focused on oncology, considerations have been recently extended to numerous medical applications.[8] However, no application of radiomics has been described in AD and MCI. Therefore, the present study was aimed to determine whether radiomic features extracted from ^{18}F-FDG PET brain images could be used for AD and MCI computer-aided diagnoses, and then we proposed a novel computer-aided Alzheimer's diseases diagnosis approach based on radiomics.

Methods

As shown in the Figure 1, we first preprocessed the collected ^{18}F-FDG PET data using normalization and smoothing. Next, we performed statistical parametric mapping (SPM) analyses based on two-sample, Student's *t* tests of the preprocessed data to determine the regions of interest (ROIs). Subsequently, a few radiomic features were extracted from the ROIs. Thereafter, we applied the Cronbach's alpha coefficient and Pearson's correlation coefficient for feature selection. Finally, based on these selected radiomic features, we completed three classifications for AD *versus* HCs, MCI *versus* HCs and AD *versus* MCI using a support vector machine (SVM). Details on this approach are described in following sections.

Materials

The study was approved by the ethics committee of Huashan Hospital, Fudan University, Shanghai, China (permission number: KY2013-336). All patients of Huashan Hospital provided written informed consent. For ethical review information on Alzheimer's Disease Neuroimaging Initiative (ADNI) data, please refer to the website (adni.loni.usc.edu).

The data used in this study included two cohorts: (1) 422 cohorts from the ADNI database, including 162 HCs (32 HCs had two time-interval ^{18}F-FDG PET scans, while 130 HCs had only one ^{18}F-FDG PET scan), 130 MCI and 130 AD patients. We selected clinical variables [clinical dementia rating scale in its sum of boxes (CDRSB); Alzheimer's disease assessment scale (ADAS)] and ^{18}F-FDG PET images for these cohorts. The ADNI was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the United States Food and Drug Administration (US FDA), private pharmaceutical companies, and nonprofit organizations, as a $60 million, 5-year, public-private partnership. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of MCI and early AD. (2) A total of 44 cohorts from the PET Center, Huashan Hospital, Fudan University, Shanghai, China, including 22 HCs and 22 AD patients. We selected basic information (sex and age) and ^{18}F-FDG PET images for these cohorts. Table 1 lists the basic information of all the data.

Image acquisition

All patients who underwent ^{18}F-FDG PET brain scans at Huashan Hospital were in a resting state. A 222-296 MBq injection of ^{18}F-FDG was administered intravenously under standardized conditions (in a quiet, dimly lit room with the patient's eyes open). A 10-min three-dimensional brain emission scan was acquired at 45-min post injection with a state-of-the-art PET scanner (Siemens Biograph 64 HD PET/CT; Siemens, Germany). During the scanning procedure, the patient's head was immobilized using a head holder. Attenuation correction was performed using low-dose computed tomography (150 mAs, 120 kV, Acq. 64×0.6 mm) prior to the emission scan. Following corrections for scatter, dead time, and random coincidences, PET images were reconstructed by three-dimensional filtered back-projection and a Gaussian filter [full-width at half maximum (FWHM) 3.5 mm], providing 148 contiguous transaxial slices of 3-mm-thick spacing. For images downloaded from the ADNI database, detailed information regarding the data acquisition protocol is publicly available on the LONI website (https://ida.loni.usc.edu/login.jsp).

Table 1. Basic information of all the data.

Group		Sex (M/F)	Age (years)	ADAS	CDRSB
ADNI cohorts (422)	AD1 (*n* = 130)	70/60	71.3 ± 6.1	30.2 ± 7.1	4.5 ± 1.6
	MCI (*n* = 130)	66/64	70.7 ± 5.4	17.3 ± 6.4	1.7 ± 0.8
	HC1 (*n* = 32)	13/19	76.2 ± 6.8	Twice ^{18}F-FDG PET imaging acquisition time interval: 131.2 ± 80.8 days	
	HC2 (*n* = 130)	68/62	71.8 ± 5.9	8.8 ± 3.6	0
Huashan cohorts (44)	AD2 (*n* = 22)	16/6	57.3 ± 6.5	N/A	N/A
	HC3 (*n* = 22)	16/6	57.3 ± 6.5	N/A	N/A

^{18}F-FDG PET, ^{18}F-fluorodeoxy-glucose positron emission tomography; AD, Alzheimer's disease; ADAS, Alzheimer's disease assessment scale; ADNI, Alzheimer's Disease Neuroimaging Initiative; CDRSB, clinical dementia rating scale in its sum of boxes; F, female; HC, healthy control; M, male; MNI, Montreal Neurological Institute; N/A, not available.

Image preprocessing

Image data were processed using statistical parametric mapping (SPM8, www.fil.ion.ucl.ac.uk/spm/) implemented in MATLAB R2014b. The aim of preprocessing was to spatially normalize the images into a standard space defined by template images and to remove unwanted distortions such as low-frequency background noise. The image preprocessing consisted of two steps: normalization and smoothing. All the original Digital Imaging and Communications in Medicine (DICOM) data were converted into NIfTI-formatted files using DCM2NII (http://people.cas.sc.edu/rorden/mricron/index.html). For each patient, the PET image was first normalized to the Montreal Neurological Institute (MNI; McGill University, Montreal, Canada) space through the 'normalize: estimate and write' methodology. Next, the normalized images were smoothed using an isotropic Gaussian smoothing kernel with a FWHM value of $10 \times 10 \times 10 \, mm^3$. The normalized images had a spatial resolution of $91 \times 109 \times 91$ with voxel sizes of $2 \times 2 \times 2 \, mm^3$.

ROIs

In this study, we focused on brain areas that were relevant to AD pathology. To characterize morphological differences in AD patients compared with HCs, we performed a group comparison using a two-sample, Student's *t* test implemented in SPM8, based on 30 AD patients (from AD1 group) and 30 HCs (from HC2 group) randomly selected from ADNI cohorts. These 60 samples were not included in subsequent feature selection and classification experiments. We set the peak threshold at $p < 0.01$ with family-wise error correction over the entire brain regions with a threshold of 20 voxels. Significantly different brain areas were localized using the software xjView9.6 (www.alivelearn.net/xjview). Because MCI is an intermediate process of HC conversion to AD, we assume that these ROIs can also be used for feature extraction of MCI data. Thus, these regions were treated as ROIs in subsequent studies. To verify the reliability of these ROIs, we repeated this SPM analysis as a comparison in the Huashan cohort.

Feature extraction

Using morphological results from the last section, we extracted ROIs of the remaining samples (including 100 AD patients, 130 MCI patients and 132 HCs) in the ADNI cohorts for further analysis. In this section, the radiomics tool developed by Vallieres[4] (https://github.com/mvallieres/radiomics) was used. All steps were performed in MATLAB R2014b, including wavelet bandpass filtering, Lloyd–Max quantization and feature calculation. The first step was wavelet bandpass filtering. This step was carried out by

applying different weights [low (L) and high (H)] to bandpass sub-bands (LHL, LHH, LLH, HLL, HHL, and HLH) of the ROIs, compared with low- and high-frequency sub-bands (LLL and HHH) in the wavelet domain. The ratio of the weight was defined by R, and the values of R were 1/2, 2/3, 1 (no wavelet filtering), 3/2, and 2. Global features were immediately extracted before other steps. Second, because the feature extraction algorithm required that the image grayscale value should be a discrete value and voxels should be isotropic, the Lloyd–Max quantization algorithm was applied to normalize the ^{18}F-FDG PET images to 256 gray-level images. Finally, four types of texture matrices [gray-level co-occurrence matrix (GLCM) gray-level run-length matrix (GLRLM), gray-level size zone matrix (GLSZM), and neighborhood gray-tone difference matrix (NGTDM)] could be obtained from quantized PET images. According to these texture matrices, numerous high-order features could be calculated. In addition to texture features, we also calculated wavelet features throughout 16 orders of wavelet decomposition. In total, 215 radiomic features were extracted from each sample, including 43 texture features (The R parameter values for these features was 1) and 172 wavelet features that were defined as the features extracted after wavelet filtering (The R parameters values for these features were 1/2, 2/3, 3/2 and 2).[11] The detailed mathematical definition of the global features and four texture matrices were as follows[15–20]:

(1) Global texture (GT): let P define the histogram of a volume $V(x, y, z)$ with isotropic voxel size. $P(i)$ represents the number of voxels with gray-level i, and N_g represents the number of gray-level bins set for P. The ith entry of the normalized histogram is then defined as follows:

$$p(i) = \frac{P(i)}{\sum_{i=1}^{N_g} P(i)} \tag{1}$$

(2) GLCM: let P define the GLCM of a quantized volume $V(x, y, z)$ with isotropic voxel size. $P(i, j)$ represents the number of times voxels of gray-level i were neighbors with voxels of gray-level j in V, and N_g represents the predefined number of quantized gray-levels set in V. Only the GLCM of size $N_g \times N_g$ is computed per volume V by simultaneously adding up the frequency of co-occurrences of all voxels with their 26-connected neighbors in three-dimensional space,

with all voxels (including the peripheral region) considered once as a center voxel (according to Haralick,[17] $d = 1$). To account for discretization length differences, neighbors at a distance of $\sqrt{3}$ voxels around a center voxel increment the GLCM by a value of $\sqrt{3}$, neighbors at a distance of $\sqrt{2}$ voxels around a center voxel increment the GLCM by a value of $\sqrt{2}$, and neighbors at a distance of 1 voxel around a center voxel increment the GLCM by a value of 1. The entry (i, j) of the normalized GLCM is then defined as follows:

$$p(i, j) = \frac{P(i, j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i, j)} \tag{2}$$

The following quantities are also defined:

$$\mu_i = \sum_{i=1}^{N_g} i \sum_{j=1}^{N_g} p(i, j), \mu_j$$
$$= \sum_{j=1}^{N_g} j \sum_{i=1}^{N_g} p(i, j)$$
$$\sigma_i = \sum_{i=1}^{N_g} (i - \mu_i)^2 \sum_{j=1}^{N_g} p(i, j), \sigma_j \tag{3}$$
$$= \sum_{j=1}^{N_g} (j - \mu_j)^2 \sum_{i=1}^{N_g} p(i, j)$$

(3) GLRLM: let P define the GLRLM of a quantized volume $V(x, y, z)$ with an isotropic voxel size. $P(i, j)$ represents the number of runs of gray-level i and of length j in V, N_g represents the predefined number of quantized gray-level sets in V, and L_γ represents the length of the longest run (of any gray-level) in V. Only one GLRLM of size $N_g \times L_\gamma$ is computed per volume V by simultaneously adding up all possible longest run lengths in the 13 directions of three-dimensional space (one voxel can be part of multiple runs in different directions but can be part of only one run in a given direction). To account for discretization length differences, runs were constructed with voxels separated by a distance of $\sqrt{3}$ in the GLRLM; the results of an increment of $\sqrt{3}$ were similar to those with increments of $\sqrt{2}$ and 1. The entry (i, j) of the normalized GLRLM is then defined as follows:

$$p(i, j) = \frac{P(i, j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{L_\gamma} P(i, j)} \tag{4}$$

The following quantities are also defined:

$$\mu_i = \sum_{i=1}^{N_g} i \sum_{j=1}^{L_\gamma} p(i,j), \mu_j = \sum_{j=1}^{L_\gamma} j \sum_{i=1}^{N_g} p(i,j) \quad (5)$$

(4) GLSZM: let P define the GLSZM of a quantized volume $V(x,y,z)$ with isotropic voxel size. $P(i,j)$ represents the number of 3D zones of gray-levels i and of size j in V, N_g represents the predefined number of quantized gray-levels set in V, and L_z represents the size of the largest zone (of any gray-level) in V. A GLSZM of size $N_g \times L_z$ is computed per volume V by adding up all possible largest zone sizes, with zones constructed from 26 connected neighbors of the same gray level in three-dimensional space (one voxel can be part of only one zone). The entry (i,j) of the normalized GLSZM is then defined as follows:

$$p(i,j) = \frac{P(i,j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{L_z} P(i,j)} \quad (6)$$

The following quantities are also defined:

$$\mu_i = \sum_{i=1}^{N_g} i \sum_{j=1}^{L_z} p(i,j), \mu_j = \sum_{j=1}^{L_z} j \sum_{i=1}^{N_g} p(i,j) \quad (7)$$

(5) NGTDM: let P define the NGTDM of a quantized volume $V(x,y,z)$ with isotropic voxel size. $P(i)$ represents the summation of the gray-level differences between all voxels with gray-level i and the average gray-level of their 26-connected neighbors in three-dimensional space. N_g represents the predefined number of quantized gray-levels set in V, and $\left(N_g\right)_{eff}$ is the effective number of gray levels in V, with $\left(N_g\right)_{eff} < N_g$ (let the vector of gray-level values in V be denoted as $g = g(1), g(2), \ldots, g(N_g)$; some gray levels excluding $g(1)$ and $g(N_g)$ may not appear in V due to different quantization schemes). A NGTDM of size $N_g \times 1$ is computed per volume V. To account for discretization length differences, all averages around a center voxel located at position (j,k,l) in V are performed such that the neighbors at a distance of $\sqrt{3}$ voxels are given a weight of $1/\sqrt{3}$, similarly to $\sqrt{2}$ and 1. The ith entry of the NGTDM is then defined as follows:

$$P(i) = \begin{cases} \sum_{all\ voxels\ \in\{N_i\}} \left|i - \bar{A}_i\right| if\ N_i > 0 \\ 0\ if\ N_i = 0 \end{cases} \quad (8)$$

Where $\{N_i\}$ is the set of all voxels with gray level i in V (including the peripheral region), N_i is the number of voxels with gray level i in V, and $\bar{A}_i$ is the average gray level of the 26-connected neighbors around a center voxel with gray level i and located at position (j,k,l) in V such that:

$$\bar{A}_i = \bar{A}_i(j,k,l) = \frac{\sum_{m=-1}^{m=1}\sum_{n=-1}^{n=1}\sum_{o=-1}^{o=1} w_{m,n,o}}{\sum_{m=-1}^{m=1}\sum_{n=-1}^{n=1}\sum_{o=-1}^{o=1} w_{m,n,o}}$$

$$w_{m,n,o} = \begin{cases} 1\ if\ |j-m|+|k-n|+|l-o| = 1 \\ \frac{1}{\sqrt{2}}\ if\ |j-m|+|k-n|+|l-o| = 2 \\ \frac{1}{\sqrt{3}}\ if\ |j-m|+|k-n|+|l-o| = 3 \\ 0\ if\ V(j+m,k+n,l+o)\ is\ undefiend \end{cases} \quad (9)$$

The following quantities were also defined as follows:

$$n_i = \frac{N_i}{N} \quad (10)$$

Table 2 provides more details on the names, references and mathematical definitions of above texture features.

Feature selection

The feature selection step was done in ADNI cohorts. In this step, we first performed a stability analysis on the features mentioned above to eliminate the unstable invalid features. The radiomic features were separately extracted from the 32 HCs (HC1 group) in using their first-time and second-time ^{18}F-FDG PET images. The one type of radiomic feature extracted from these samples was divided into two feature vectors (first and second-time). The two vectors were then treated as inputs to calculate the Cronbach's alpha coefficient. We judged the stability of each feature by the value of alpha. The coefficient threshold was 0.8 and features above this value were considered stable and those features below this value were excluded from follow-up studies.

Second, to screen out the radiomic features associated with AD and MCI, we used Pearson's

Table 2. Details of radiomic texture features.

Texture matrices	References	Feature name	Formula
Global		Variance	$\sigma^2 = \sum_{i=1}^{N_g} (i - \mu)^2 p(i)$
		Skewness	$s = \sigma^{-3} \sum_{i=1}^{N_g} (i - \mu)^3 p(i)$
		Kurtosis	$k = \sigma^{-4} \sum_{i=1}^{N_g} [(i - \mu)^4 p(i)] - 3$
Gray-level co-occurrence matrix (GLCM)	Haralick and colleagues[1]	Energy	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [p(i,j)]^2$
		Contrast	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i-j)^2 p(i,j)$
		Correlation	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \dfrac{(i - \mu_i)(i - \mu_j) p(i,j)}{\sigma_i \sigma_j}$
		Homogeneity	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \dfrac{p(i,j)}{1 + |i-j|}$
		Variance	$\dfrac{1}{N_g \times N_g} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [(i - \mu_i)^2 p(i,j) + (j - \mu_i)^2 p(i,j)]$
		Sum average	$\dfrac{1}{N_g \times N_g} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [ip(i,j) + jp(i,j)]$
		Entropy	$-\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \left[ip(i,j) + jp(i,j) \right]$
		Auto correlation	refer to the references
		Dissimilarity	
Gray-level run-length matrix (GLRLM)	Galloway[2]	Short-run emphasis (SRE)	$\sum_{i=1}^{N_g} \sum_{j=1}^{L_r} \dfrac{p(i,j)}{j^2}$
		Long-run emphasis (LRE)	$\sum_{i=1}^{N_g} \sum_{j=1}^{L_r} j^2 p(i,j)$

(Continued)

Table 2. (Continued)

Texture matrices	References	Feature name	Formula
		Gray-level nonuniformity (GLN)	$\sum_{i=1}^{N_g}\left(\sum_{j=1}^{L_\gamma}p(i,j)\right)^2$
		Run-length nonuniformity (RLN)	$\sum_{j=1}^{L_\gamma}\left(\sum_{i=1}^{N_g}p(i,j)\right)^2$
		Run percentage (RP)	$\dfrac{\sum_{i=1}^{N_g}\sum_{j=1}^{L_\gamma}p(i,j)}{\sum_{j=1}^{L_\gamma}\sum_{i=1}^{N_g}p(i,j)}$
	Chu and colleagues[3]	Low-gray-level run emphasis (LGRE)	$\sum_{i=1}^{N_g}\sum_{j=1}^{L_\gamma}\dfrac{p(i,j)}{i^2}$
		High-gray-level run emphasis (HGRE)	$\sum_{i=1}^{N_g}\sum_{j=1}^{L_\gamma}i^2 p(i,j)$
	Dasarathy and Holder[4]	Short-run low-gray-level emphasis (SRLGE)	$\sum_{i=1}^{N_g}\sum_{j=1}^{L_\gamma}\dfrac{p(i,j)}{i^2 j^2}$
		Short-run high gray-level emphasis (SRHGE)	$\sum_{i=1}^{N_g}\sum_{j=1}^{L_\gamma}\dfrac{i^2 p(i,j)}{j^2}$
		Long-run low-gray-level emphasis (LRLGE)	$\sum_{i=1}^{N_g}\sum_{j=1}^{L_\gamma}\dfrac{j^2 p(i,j)}{i^2}$
		Long-run high-gray-level emphasis (LRHGE)	$\sum_{i=1}^{N_g}\sum_{j=1}^{L_\gamma}i^2 j^2 p(i,j)$
	Thibault and colleagues[5]	Gray-level variance (GLV)	$\dfrac{1}{N_g\times L_\gamma}\sum_{i=1}^{N_g}\sum_{j=1}^{L_\gamma}\left[ip(i,j)-\mu_i\right]^2$
		Run-length variance (RLV)	$\dfrac{1}{N_g\times L_\gamma}\sum_{i=1}^{N_g}\sum_{j=1}^{L_\gamma}\left[jp(i,j)-\mu_j\right]^2$
Gray-level size zone matrix (GLSZM)	Galloway[2]; Thibault and colleagues[5]	Small zone emphasis (SZE)	$\sum_{i=1}^{N_g}\sum_{j=1}^{L_z}\dfrac{p(i,j)}{j^2}$

Table 2. (Continued)

Texture matrices	References	Feature name	Formula
		Large zone emphasis (LZE)	$\sum_{i=1}^{N_g}\sum_{j=1}^{L_z} j^2 p(i,j)$
		Gray-level nonuniformity (GLN)	$\sum_{i=1}^{N_g}\left(\sum_{j=1}^{L_z} p(i,j)\right)^2$
		Zone-size nonuniformity (ZSN)	$\sum_{j=1}^{L_z}\left(\sum_{i=1}^{N_g} p(i,j)\right)^2$
		Zone percentage (ZP)	$\dfrac{\sum_{i=1}^{N_g}\sum_{j=1}^{L_z} p(i,j)}{\sum_{j=1}^{L_z}\sum_{i=1}^{N_g} p(i,j)}$
	Chu and colleagues[3]; Thibault and colleagues[5]	Low-gray-level zone emphasis (LGZE)	$\sum_{i=1}^{N_g}\sum_{j=1}^{L_z} \dfrac{p(i,j)}{i^2}$
		High-gray-level zone emphasis (HGZE)	$\sum_{i=1}^{N_g}\sum_{j=1}^{L_z} i^2 p(i,j)$
	Dasarathy and Holder[4]; Thibault and colleagues[5]	Small zone low-gray-level emphasis (SZLGE)	$\sum_{i=1}^{N_g}\sum_{j=1}^{L_z} \dfrac{p(i,j)}{i^2 j^2}$
		Small zone high-gray-level emphasis (SZHGE)	$\sum_{i=1}^{N_g}\sum_{j=1}^{L_z} \dfrac{i^2 p(i,j)}{j^2}$
		Large zone low-gray-level emphasis (LZLGE)	$\sum_{i=1}^{N_g}\sum_{j=1}^{L_z} \dfrac{j^2 p(i,j)}{i^2}$
		Large zone high-gray-level emphasis (LZHGE)	$\sum_{i=1}^{N_g}\sum_{j=1}^{L_z} i^2 j^2 p(i,j)$
	Thibault and colleagues[5]	Gray-level variance (GLV)	$\dfrac{1}{N_g \times L_z}\sum_{i=1}^{N_g}\sum_{j=1}^{L_z}\left[ip(i,j)-\mu_i\right]^2$
		Zone-size variance (ZSV)	$\dfrac{1}{N_g \times L_z}\sum_{i=1}^{N_g}\sum_{j=1}^{L_z}\left[jp(i,j)-\mu_j\right]^2$

(Continued)

Table 2. (Continued)

Texture matrices	References	Feature name	Formula
Neighborhood gray-tone difference matrix (NGTDM)	Amadasun and King[6]	Coarseness	$\left[\epsilon + \sum_{i=1}^{N_g} n_i P(i)\right]^{-1}$
		Contrast	$\left[\dfrac{1}{(N_g)_{eff}\left[(N_g)_{eff}-1\right]}\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} n_i n_j (i-j)^2\right]\left[\dfrac{1}{N}\sum_{i=1}^{N_g} P(i)\right]$
		Busyness	$\dfrac{\sum_{i=1}^{N_g} n_i P(i)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left(in_i - jn_j\right)},$ $n_i, n_j \neq 0$
		Complexity	$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\dfrac{|i-j|\left[n_i P(i) + n_j P(j)\right]}{N(n_i + n_j)}$ $n_i, n_j \neq 0$
		Strength	$\dfrac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left(n_i + n_j\right)(i-j)^2}{\left[\epsilon + \sum_{i=1}^{N_g} n_i P(i)\right]}$ $n_i, n_j \neq 0$

1. Haralick RM, Shanmugam K and Dinstein IH. Textural features for image classification. *IEEE Trans Syst Man Cybern* 1973; 3: 610–621.
2. Galloway MM. *Texture analysis using grey level run lengths*. Vol. 75. *NASA STI/Recon Technical Report N*. Linthicum Heights, MD, 1974.
3. Chu A, Sehgal CM and Greenleaf JF. Use of gray value distribution of run lengths for texture analysis. *Patt Recog Lett* 1990; 11: 415–419.
4. Dasarathy BV and Holder EB. Image characterizations based on joint gray level—run length distributions. *Patt Recog Lett* 1991; 12: 497–502.
5. Thibault G, Fertil B, Navarro C, et al. Texture indexes and gray level size zone matrix application to cell nuclei classification. In: *10th International conference on pattern recognition and information processing*, Minsk, Belarus, 2009, pp.140–145.
6. Amadasun M and King R. Textural features corresponding to textural properties. *IEEE Trans Syst Man Cybern* 1989; 19: 1264–1274.

correlation coefficients to evaluate whether there was a correlation between the clinical scale and radiomic features. If there was a correlation, we believe that this feature was effective for AD or MCI and would be used for classification. To ensure statistical reliability, we conducted a cross-validation test. Taking AD as an example, the cross-validation process was as follows: (1) we randomly chose 70 patients from the AD1 group (100 patients that were not used for SPM analysis described in section *Regions of interest*); (2) we extracted ROIs from the above 70 patients and got stable radiomic features; (3) we calculated the correlation between the ADAS scale and each feature using Pearson's correlation coefficients. As a result, those features which were related to ADAS ($p < 0.05$ with false discovery rate (FDR) correction) were considered as selected features; (4) the remaining 30 patients from the AD1 group were used as a test dataset for subsequent classification experiments. Finally, the above cross-validation process was repeated 500 times.

For the MCI group, we calculated the correlation between the CDRSB scale and each feature and did the same cross-validation process 500 times. In addition, we also did the same cross-validation process to explore radiomic features that were effective for both the AD group and MCI group by calculating the correlation between the ADAS scale and each feature. All calculations were implemented in MATLAB R2014b using raw values without normalization.

To further explore the consistency of the selected features in different conditions, we also extracted selected features using ROIs from the Huashan cohort and calculated the intraclass correlation coefficient (ICC) for those feature using samples which participated in feature selection cross-validation (CV) experiments (100 AD and 100 MCI patients).

SVM classification

After the feature selection step, we identified three selected radiomic feature sets (AD/MCI/HC) to distinguish AD/HC, MCI/HC and AD/MCI. To verify the diagnostic capabilities of these feature sets, we performed three SVM classification for AD *versus* HCs, MCI *versus* HCs and AD *versus* MCI. The SVM is one kind of supervised learning model with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a nonprobabilistic binary linear classifier.

In this step, we also conducted classification experiments 500 times for each SVM classification. Taking the AD *versus* HCs classification as an example, we used 70 patients from the AD1 group (the same as 70 patients selected in the section, *Feature selection*) and randomly selected 70 patients from the HC2 group (100 patients that were not used for SPM analysis described in the section, *Regions of interest*) to train the SVM classifier. Then, we used the remaining 30 patients from the AD1 group and 30 patients from the HC2 group to test the SVM classifier. Finally, the Huashan cohort was used as the second test dataset to further validate the classification results. We also did the above process for AD *versus* MCI, and MCI *versus* HCs 500 times. The mean and standard deviation values were also calculated 500 times in each experiment.

In the classification experiment, four kernel (linear, sigmoid, polynomial, and radial basis) functions were used to detect feature generalization ability and classification reliability. Before classification, all feature values were normalized in both the train dataset and test dataset by using a '*min-max normalization*' method separately. In addition, features, age and sex had also been considered as the inputs for SVM classification. The classifier version used in this experiment was LIBSVM 2.9.1 (www.csie.ntu.edu.tw/~cjlin/libsvm/) and was implemented in MATLAB R2014b. In addition, we repeated the above feature selection experimental process using ROIs obtained by the Huashan cohort as a comparison. The same classification experiments were also repeated.

Results

ROIs

Figure 2 shows the results of the voxel-based two-sample Student's *t* test of AD patients and HCs in both ADNI cohorts (30 AD patients *versus* 30 HCs) and Huashan cohorts (22 AD patients *versus* 22 HCs). These brain regions comprised brain tissue related to the AD pathology and are summarized in Tables 3 and 4. The results obtained from the two groups were consistent. Most of the ROIs were distributed in the temporal, occipital and frontal areas.

Cronbach's alpha coefficient

Figure 3 shows an overall scatter plot of the Cronbach's alpha coefficient for all radiomic features. Of these, 168 features were stable (alpha > 0.8), and 47 were unstable (alpha < 0.8). Among the 168 stable features, 51 features were extremely stable (alpha > 0.95). The results from the stability analysis proved that the radiomic features were stable and reliable in brain ^{18}F-FDG PET images. Table 5 shows all stable features.

Pearson's correlation coefficients

After each cross-validation in the feature selection step, it can be observed that there were about 50–70 kinds of features associated with AD, about 30–40 features associated with MCI and about 10–20 features associated with both AD and MCI. Table 6 lists the most frequently features, their occurring times and ICC values at 500 times cross-validation. In general, these
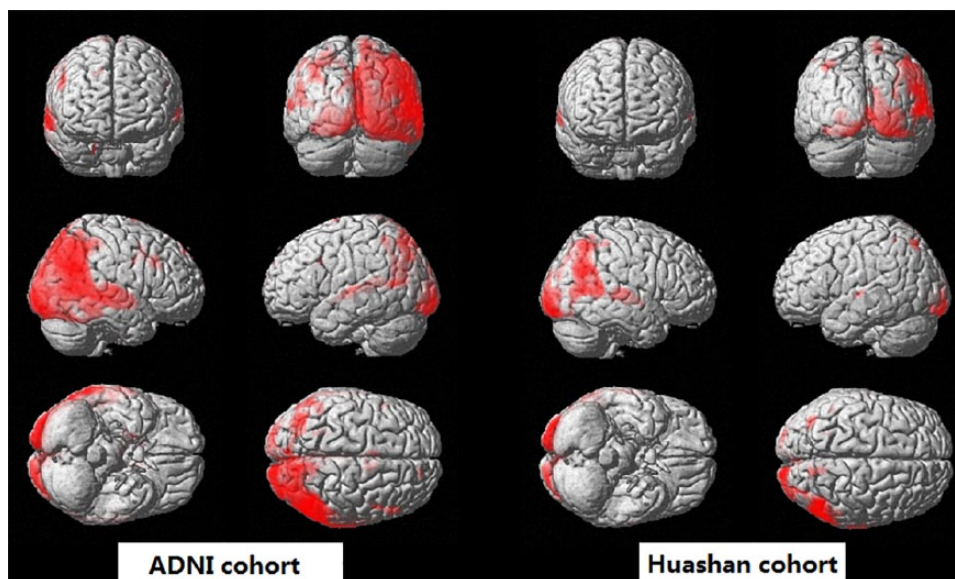
Figure 2. Results of the two-sample Student's *t* test brain 18F-FDG PET images conducted to assess differences between AD patients and HCs.
18F-FDG PET, 18F-fluorodeoxy-glucose positron emission tomography; AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; HC, healthy control.

Table 3. Brain regions with significant differences between AD patients and HCs based on ADNI cohorts.

MNI coordinate			Cluster location (standardized automated anatomical labeling template)	Brodmann area	Hemisphere	Cluster size	Z-score
14	−68	74	Temporal_Mid; Angular; Temporal_Sup; Calcarine; Occipital_Inf; Occipital_Sup; Parietal_Sup; Lingual; Occipital_Mid; Temporal_Inf; Precuneus; Cingulum_Mid; Cingulum_Post; Fusiform; Cuneus	18,39,19,40,21,22, 17,7,31,37,23,42,30	Right/left	13535	−3.84
−64	14	2	Temporal_Mid; Temporal_Sup; Angular; Occipital_Mid	22,39,19,21,42,41,40	Left	643	−3.28
8	−14	14	Thalamus	–	Right	99	−2.96
−12	2	16	Caudate	–	Left	68	−2.93
18	−4	20	Caudate	–	Right	22	−2.79
54	22	38	Frontal_Inf_Oper; Frontal_Mid; Frontal_Inf_Tri	9	Right	37	−2.8
−52	15	34	Precentral; Frontal_Inf_Oper; Frontal_Mid; Frontal_Inf_Tri	9	Left	31	−2.85
48	0	34	Precentral	6,9	Right	32	−2.87
−38	−72	54	Parietal_Inf; Angular; Parietal_Sup	7,40,19	Left	518	−3.25
16	52	44	Frontal_Sup; Frontal_Sup_Medial	8,9	Right	41	−2.92
52	8	48	Precentral; Frontal_Mid	6	Right	34	−2.79

AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; HC, healthy control; MNI, Montreal Neurological Institute.

Table 4. Brain regions with significant differences between AD and HC based on Huashan cohorts.

MNI coordinates			Cluster location (standardized automated anatomical labeling template)	Brodmann area	Hemisphere	Cluster size	Z-score
20	−100	−8	Temporal_Mid; Temporal_Inf; Temporal_Sup; Angular; Occipital_Mid; Occipital_Sup; Occipital_Inf; Calcarine; Lingual; Parietal_Inf; Parietal_Sup; Cuneus; SupraMarginal; Fusiform	7, 13, 18, 17, 19, 21, 22, 23, 30, 37, 39, 40	Right/left	5902	−5.22
70	−22	4	Temporal_Sup; Temporal_Mid	21, 22, 42	Right	317	−4.52
14	−72	28	Cuneus; Precuneus; Cingulum_Post; Cingulum_Mid	7, 23, 31	Right	180	−4.17
−38	72	54	Parietal_Sup; Angular; Parietal_Inf	7	Left	155	−4.49
−48	50	62	Parietal_Inf	–	Left	33	−4.17
14	−66	76	Parietal_Sup; Precuneus; Postcentral	–		353	−5.07

AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; HC, healthy control; MNI, Montreal Neurological Institute.
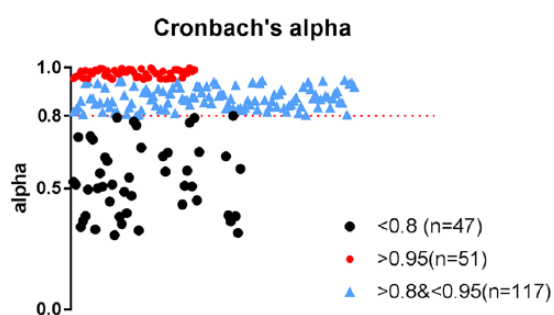
Figure 3. Scatter plot of all radiomic features in relation to Cronbach's alpha coefficient.

features showed excellent consistency in repeated experiments. For example, 'energy' and 'entropy' appeared in both AD and MCI experiments, indicating that these two features have good pathological revealing ability. As a comparison, the top relative features selected from Huashan cohort's ROIs were consistent with Table 6. Table 7 shows the details of these key features.

SVM classification results

As shown in Table 8, the use of selected radiomic features with linear, polynomial, radial basis, and sigmoid kernels could achieve average accuracies of 91.5%, 88.1%, 86.1%, and 86.3%, respectively, to distinguish AD patients and HCs, average accuracies of 85.9%, 83.4%, 85.0%, and 83.5%, respectively, to distinguish AD and MCI patients, and accuracies of 83.1%, 81.8%, 82.9%, and 81.5%, respectively, to distinguish MCI patients

and HCs. As a result, the linear kernel could achieve an optimal classification performance. The classification performance of Radial Basis Function (RBF) and the sigmoid kernel were slightly poor and relatively unstable, probably because these two kernels depend on parameter adjustments and in this study, we only used default parameters for classification. As a comparison, similar results were achieved by using ROIs from the Huashan cohorts. The best classification performances achieved average accuracies of 91.2%, 85.5%, 82.3% and 92.1%, in AD vs HC, AD vs MCI, MCI vs HC for ADNI cohorts, and AD vs HC for Huashan cohorts, respectively. Table 7 shows more details on classification accuracy, Area Under Curve (AUC), sensitivity, and specificity using Huashan cohort ROIs. The above results could prove the reliability of our research framework.

Discussion

This study utilized statistical analysis and SVM to investigate whether a radiomic method based on ^{18}F-FDG PET images could be used for AD and MCI computer-aided diagnosis. To prove the stability and generalization of the proposed radiomic method, we selected different samples by different PET scanners with different imaging properties, including samples from the ADNI database and Huashan Hospital. This kind of cross-dataset research approach was frequently used in the radiomic studies to avoid the over-fitting problem, as well as to test the generalization ability of the whole research framework.[8,11,12,21–23]

Table 5. Stable features.

R = 1/2	R = 2/3	R = 1	R = 3/2	R = 2
Skewness	Skewness	Skewness	Skewness	Skewness
Energy	Energy	Kurtosis	Energy	Contrast
Contrast	Contrast	Energy	Contrast	Entropy
Entropy	Entropy	Contrast	Entropy	Homogeneity
Homogeneity	Correlation	Entropy	Homogeneity	SumAverage
Correlation	SumAverage	Homogeneity	Correlation	Variance
SumAverage	Variance	Correlation	SumAverage	Dissimilarity
Variance	Dissimilarity	SumAverage	Variance	SRE
Dissimilarity	AutoCorrelation	Variance	Dissimilarity	GLN
AutoCorrelation	SRE	Dissimilarity	AutoCorrelation	RLN
SRE	LRE	AutoCorrelation	SRE	RP
LRE	GLN	SRE	GLN	LGRE
GLN	RLN	LRE	RLN	HGRE
RLN	RP	GLN	RP	SRLGE
RP	LGRE	RLN	HGRE	LRLGE
LGRE	HGRE	RP	LRLGE	LRHGE
HGRE	SRHGE	LGRE	LRHGE	GLV
SRLGE	LRLGE	HGRE	GLV	RLV
SRHGE	LRHGE	SRLGE	RLV	LZE
LRLGE	GLV	SRHGE	SZE	GLN
LRHGE	RLV	LRLGE	LZE	ZP
GLV	LZE	LRHGE	GLN	LGZE
RLV	GLN	GLV	ZP	HGZE
SZE	ZSN	RLV	LGZE	SZLGE
LZE	ZP	SZE	HGZE	SZHGE
GLN	LGZE	LZE	SZLGE	LZLGE
ZSN	HGZE	GLN	SZHGE	LZHGE
ZP	SZLGE	ZSN	LZHGE	GLV
LGZE	SZHGE	ZP	GLV	ZSV
HGZE	LZLGE	LGZE	ZSV	Coarseness
SZLGE	GLV	HGZE	Coarseness	

Table 5. (Continued)

R = 1/2	R = 2/3	R = 1	R = 3/2	R = 2
SZHGE	ZSV	SZLGE		
LZLGE	Coarseness	SZHGE		
LZHGE		LZLGE		
ZSV		LZHGE		
Coarseness		GLV		
		ZSV		
		Coarseness		

AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; GLN, gray-level nonuniformity; GLV, gray-level variance; HGRE, high-gray-level run emphasis; HGZE, high-gray-level zone emphasis; ICC, intraclass correlation coefficient; LGRE, low-gray-level run emphasis; LGZE, low-gray-level zone emphasis; LRE, long-run emphasis; LRHGE, long-run high-gray-level emphasis; LRLGE, long-run low-gray-level emphasis; LZE, large zone emphasis; LZHGE, large zone high-gray-level emphasis; LZLGE, large zone low-gray-level emphasis; MCI, mild cognitive impairment; RLN, run-length nonuniformity; RLV, run-length variance; RP, run percentage; SRE, short-run emphasis; SRHGE, short-run high gray-level emphasis; SRLGE, short-run low-gray-level emphasis; SZHGE, small zone high-gray-level emphasis; SZLGE, small zone low-gray-level emphasis; ZP, zone percentage; ZSN, zone-size nonuniformity; ZSV, zone-size variance.

Table 6. The key relative features selected by cross-validation, 500 times.

AD				MCI				AD + MCI			
Feature	R	times	ICC[a]	Feature	R	times	ICC[a]	Feature	R	times	ICC[a]
Energy	1	488	0.956	Entropy	1	475	0.949	Skewness	1	412	0.901
GLV	1/3	479	0.947	Homogeneity	1	471	0.941	Coarseness	1	387	0.875
Contrast	1	465	0.931	SRE	1/2	447	0.927	Correlation	3/2	331	0.821
Variance	2/3	461	0.933	RP	2/3	413	0.894	SZLGE	2/3	317	0.806
Entropy	1	443	0.915	Energy	1	397	0.881	Skewness	1/2	311	0.803
ZSV	3/2	410	0.899	GLV	2/3	374	0.836	–			
HGRE	1/2	393	0.878	RLN	1/2	363	0.829				
SZHGE	1	389	0.842	LZE	1	330	0.815				
LRHGE	2/3	354	0.827	Dissimilarity	1	314	0.803				
RLV	1	316	0.807	ZP	1/2	276	0.797				

[a]ICC > 0.9 means excellent consistency and ICC > 0.8 means statistically acceptable consistency.
AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; GLV, gray-level variance; HGRE, high-gray-level run emphasis; ICC, intraclass correlation coefficient; LRHGE, long-run high-gray-level emphasis; LZE, large zone emphasis; MCI, mild cognitive impairment; RLN, run-length nonuniformity; RLV, run-length variance; RP, run percentage; SRE, short-run emphasis; SZHGE, small zone high-gray-level emphasis; SZLGE, small zone low-gray-level emphasis; ZP, zone percentage; ZSV, zone-size variance.

To define ROIs, we performed a voxel-based, two-sample, Student's *t* test in a group of 30 HCs and 30 AD individuals from ADNI cohorts to determine the brain areas associated with AD pathology. In the subsequent experiments, we regarded the ROIs of AD as those of MCI because the patients with MCI who will be converted to AD have brain tissue lesions similar to those of AD.[24–27] As shown

Table 7. Results for ADNI cohorts using the Huashan cohort ROIs.
(a) The top relative features selected by cross-validation 500 times.

AD			MCI			AD + MCI		
Feature	**R**	**times**	**Feature**	**R**	**times**	**Feature**	**R**	**times**
GLV	1/2	471	Entropy	1	452	Skewness	1	383
Energy	1	463	SRE	1/2	449	Coarseness	1	361
Contrast	1	455	Homogeneity	1	431	Correlation	3/2	355
Entropy	1	449	RP	2/3	423	Skewness	1/2	328
Variance	2/3	447	Energy	1	396	SZLGE	2/3	291
HGRE	1/2	417	RLN	1/2	388	–		
ZSV	3/2	405	GLV	2/3	359			
SZHGE	1	383	LZE	1	341			
LRHGE	2/3	345	Dissimilarity	1	317			
RLV	1	324	ZP	1/2	280			

(b) Classification accuracy, AUC, sensitivity, and specificity.

Group		Accuracy / AUC / sensitivity / specificity (average)			
		Linear	**Polynomial**	**RBF**	**Sigmoid**
ADNI cohorts	AD *versus* HC	**91.2% ± 2.1%**	87.8% ± 2.4%	85.3% ± 2.7%	85.8% ± 2.5%
		0.91 ± 0.02	0.86 ± 0.03	0.85 ± 0.03	0.85 ± 0.03
		92.5% ± 1.9%	89.1% ± 2.3%	84.8% ± 2.9%	86.5% ± 2.5%
		90.1% ± 2.2%	86.9% ± 2.6%	87.1% ± 2.5%	83.1% ± 2.9%
	AD *versus* MCI	**85.5% ± 2.4%**	83.1% ± 2.8%	84.4% ± 2.7%	83.3% ± 2.7%
		0.84 ± 0.03	0.83 ± 0.03	0.84 ± 0.03	0.82 ± 0.04
		86.6% ± 2.5%	86.2% ± 2.7%	86.2% ± 2.6%	80.1% ± 2.9%
		85.9% ± 2.5%	79.9% ± 3.0%	82.7% ± 2.9%	86.2% ± 2.5%
	MCI *versus* HC	**82.3% ± 2.9%**	80.8 % ± 3.1%	81.9% ± 2.9%	80.4% ± 3.0%
		0.79 ± 0.04	0.76 ± 0.04	0.78 ± 0.04	0.76 ± 0.05
		83.1% ± 2.8%	82.7% ± 3.0%	81.1% ± 3.0%	82.3% ± 2.9%
		82.3% ± 2.9%	79.8% ± 3.2%	82.3% ± 2.7%	76.5% ± 3.2%
Huashan cohorts	AD *versus* HC	**92.1% ± 2.0%**	89.1% ± 2.1%	86.4% ± 2.4%	87.3% ± 2.5%
		0.93 ± 0.02	0.90 ± 0.02	0.86 ± 0.03	0.89 ± 0.03
		90.9% ± 2.1%	89.8% ± 2.2%	85.6% ± 2.7%	88.7% ± 2.6%
		91.1% ± 2.2%	88.2% ± 2.2%	88.1% ± 2.5%	82.5% ± 2.9%

AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; AUC, Area Under Curve; GLV, gray-level variance; HGRE, high-gray-level run emphasis; LRHGE, long-run high-gray-level emphasis; LZE, large zone emphasis; LZHGE, large zone high-gray-level emphasis; MCI, mild cognitive impairment; RBF, Radial Basis Function; RLN, run-length nonuniformity; RLV, run-length variance; RP, run percentage; SRE, short-run emphasis; SZHGE, small zone high-gray-level emphasis; SZLGE, small zone low-gray-level emphasis; ZP, zone percentage; ZSV, zone-size variance.

Table 8. Classification accuracy, AUC, sensitivity, and specificity.

Group		Accuracy / AUC / sensitivity / specificity (average)			
		Linear	**Polynomial**	**RBF**	**Sigmoid**
ADNI cohorts	AD *versus* HC	**91.5% ± 1.9%**	88.1% ± 2.3%	86.1% ± 2.4%	86.3% ± 2.5%
		0.92 ± 0.01	0.88 ± 0.03	0.87 ± 0.03	0.88 ± 0.03
		92.9% ± 2.4 %	89.5% ± 2.6%	85.3% ± 2.5%	87.1% ± 2.6%
		90.2% ± 2.1 %	87.1% ± 2.5%	87.5% ± 2.7%	83.2% ± 2.4%
	AD *versus* MCI	**85.9% ± 2.1 %**	83.4% ± 2.8%	85.0% ± 2.5%	83.5% ± 2.7%
		0.85 ± 0.02	0.84 ± 0.04	0.86 ± 0.03	0.82 ± 0.03
		87.3% ± 2.2%	86.5% ± 2.9%	86.7% ± 2.3%	80.4% ± 2.9%
		86.2% ± 2.3%	80.1% ± 2.8%	83.3% ± 2.7%	86.5% ± 2.8%
	MCI *versus* HC	**83.1% ± 2.8%**	81.8% ± 2.9%	82.9% ± 2.8%	81.5% ± 3.1%
		0.80 ± 0.04	0.79 ± 0.04	0.81 ± 0.03	0.78 ± 0.05
		83.8% ± 2.9%	83.4% ± 3.1%	83.1% ± 2.9%	84.5% ± 2.9%
		82.9% ± 2.6%	80.2% ± 2.9%	82.7% ± 2.8%	77.3% ± 3.2%
Huashan cohorts	AD *versus* HC	**91.9% ± 2.3%**	88.4% ± 2.5%	85.9% ± 2.6%	87.1% ± 2.5%
		0.93 ± 0.02	0.89 ± 0.03	0.85 ± 0.03	0.89 ± 0.03
		90.7% ± 2.4%	89.4% ± 2.8%	85.1% ± 2.8%	88.2% ± 2.7%
		90.5% ± 2.6%	87.8% ± 2.7%	87.7% ± 2.4%	82.4% ± 2.9%

AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; AUC, Area Under Curve; HC, healthy control; MCI, mild cognitive impairment; RBF, Radial Basis Function.

in Tables 3 and 4, most of the results are consistent with the typical results reported previously. Ferreira and colleagues found that the medial temporal region is the most consistent neurostructural biomarker for predicting AD,[28] and our results were consistent with this finding (AAL: Temporal_Mid, Temporal_Inf and Temporal_Sup). Occipital lobe regions (AAL: Occipital_Mid, Occipital_Sup, Occipital_Inf), anterior cingulate (AAL: Cingulum_Post, Cingulum_Mid), and the parietal cortex area (AAL: Parietal_Sup, Parietal_Inf, Parietal_Sup) were also identified in agreement with previous studies.[28–32] After *t* test analysis, an ROI template was obtained (Figure 2 and Table 3), and features were extracted from this ROI. In total, 215 radiomic features were extracted for each sample, including intensity, texture, and wavelet features.

Recent radiomics research studies usually involved shape features because the target ROIs with various shapes, such as the tumor areas, were segmented manually.[8,9,11,13,14,22] However, those shape features were not included in this study. For brain PET studies, preprocessing steps usually include spatial normalization and smoothing. In this study, we followed this preprocessing process, further reducing the effects of shape differences. Moreover, individual heterogeneity in tumors was much higher than that in normal brain tissue. Therefore, our feature set did not contain shape features.

In addition, the stability of 215 radiomic features in brain ^{18}F-FDG PET images was determined, revealing 168 stable features (alpha > 0.8). Stability analyses proved that more than half of

Table 9. Classification accuracy of existing literature.

References	Accuracy	
	AD *versus* HCs	MCI *versus* HCs
Silveira and Marques[5]	90.9%	79.6%
Gray and colleagues[7]	81.6%	70.2%
Liu, Manhua and colleagues[3]	91.2%	78.9%
Our method	**91.5%**	**83.1%**

AD, Alzheimer's disease; HC, healthy control; MCI, mild cognitive impairment.

the radiomic features would not be disturbed by random errors and imaging noise. Of the 168 stable features, 51 were extremely stable (alpha > 0.95). Stability score analysis was performed in HC samples, and subsequent analyses indicated that radiomic feature values were associated with clinical cognitive scales, which could reflect disease pathology in AD and MCI patients. Obviously, these connections were not caused by random factors. Hence, the stability results proved the reliability of subsequent analyses and indirectly proved that radiomic features contains a wealth of pathological information. The possible association of radiomic features with CDRSB and ADAS was explored based on Pearson's correlation. We found that many radiomic features that were not studied previously were relevant with cognitive scale values. Here, we studied two scales because ADAS is a scale specifically designed to measure cognitive performance in AD patients.[33] As described in the section *Feature selection*, two group analyses containing AD patients used ADAS values; however, for MCI patients, we used CDRSB because it is a scale specifically sensitive for MCI measurement. As a result of the feature selection, we found that about 50–70 features were associated with AD through the Pearson correlation coefficient, about 30–40 features were associated with MCI, and about 10–20 features were associated with AD and MCI mixed samples. Jaccard index values of frequently occurring features in Table 6 showed that selected features were consistent with the correlation with disease. These correlations suggested that high-order radiomic features extracted from ^{18}F-FDG PET brain imaging could be used not only for classification diagnosis but also contained rich information related to pathological processes for further study and mining.[8–10,22]

Importantly, on a 500-times cross-validation experiment based on both ADNI and Huashan cohorts using SVM as a classifier, we found that radiomic features had a distinct ability to classify AD *versus* HCs, MCI *versus* HCs and AD *versus* MCI with maximum average accuracies of 91.5%, 83.1% and 85.9%, respectively. By using the cross-validation method, the above results avoided the bias of test results, and presented the validity of the selected features, and SVM classification.[34] Table 9 lists the results of this paper compared with the results of previous studies, and these methods were complex image classification algorithms based on deep learning techniques. As shown in Table 9, the radiomic method could achieved a satisfied classification accuracy in both AD *versus* HCs (91.5%) and MCI *versus* HCs (83.1%) comparable with complex algorithms. Thus, radiomics may have application prospects in the field of AD and MCI diagnosis. Because the significance of the calculation values existing in the neural network framework are unknown, compared with the deep learning framework, another advantage of radiomics is that its features are correlated with AD and MCI clinical scales. Based on this, a future study can investigate the relationship between the radiomic feature values and the disease directly to help physicians carry out personalized and precise treatment.[8]

Limitations and further considerations
Although the potential application value of radiomics in the diagnosis of AD and MCI has been proven in the paper, some limitations persist that may influence the results of this study. First, for individuals from ADNI and the Huashan Hospital, their ages and races were not matched. The average

age of individuals from ADNI was approximately 72 years and mainly included Europeans and Americans. The average age of individuals from Huashan Hospital was 57 years and mainly included Asians. Whether the differences in age and race will influence the results has not been studied yet and can be explored in subsequent studies.

Second, the shape and some texture features of PET images were lost in the preprocessing step, events that may influence the results. In this study, we used a smooth step and selected the MNI template for image registration and lost some texture and morphology information during the registration process. However, this is a routine step in other AD studies; thus, we followed this preprocessing principle. The effect of these preprocessing methods on the results is unknown and requires follow-up studies.

Third, the pathobiological mechanism concerning the correlation between radiomic features and clinical scale was not explored in this paper. In our current research, we only confirmed that the radiomic features were related to the clinical scale, indicating that there is indeed a correlation between radiomic features and AD/MCI pathology; however, we did not further study the pathobiological mechanism. In future research, the mechanism between each feature and disease should be studied in detail.

Conclusion

In summary, the research in this paper proved that high-order radiomic features extracted from ^{18}F-FDG PET brain images can be used for AD and MCI computer-aided diagnoses. Radiomic features can reflect the pathological information of MCI and AD, and they can diagnose MCI and AD with increased accuracy. The simplicity of the acquisition of radiomic features and its high-throughput nature would constitute powerful tools for personalized precision medicine for the population affected by AD and MCI.

Author Contribution

Yupeng Li and Jiaying Lu have contributed equally to this work.

Funding

This study was supported by the National Science Foundation of China (grant numbers

Conflict of interest statement

The authors declare that there is no conflict of interest.

ORCID iD

Jiehui Jiang https://orcid.org/0000-0003-4948-3683

References

1. Hurd MD, Martorell P, Delavande A, *et al.* Monetary costs of dementia in the United States. *New Engl J Med* 2013; 368: 1326–1334.

2. Schneider JA, Arvanitakis Z, Leurgans SE, *et al.* The neuropathology of probable Alzheimer disease and mild cognitive impairment. *Ann Neurol* 2009; 66: 200–208.

3. Liu M, Cheng D and Yan W. Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG PET images. *Front Neuroinform* 2018; 12: 35.

4. Minati L, Edginton T, Bruzzone MG, *et al.* Reviews: current concepts in Alzheimer's disease: a multidisciplinary review. *Am J Alzheimers Dis Other Demen* 2009; 24: 95.

5. Silveira M and Marques J. Boosting Alzheimer disease diagnosis using PET images. In: *20th International conference on pattern recognition*, Istanbul, Turkey, 23–26 August 2010, pp.2556–2559.

6. Henriques AD, Benedet AL, Camargos EF, *et al.* Fluid and imaging biomarkers for Alzheimer's disease: where we stand and where to head to. *Exp Gerontol* 2018; 107: 169–177.

7. Gray KR, Wolz R, Keihaninejad S, *et al. Regional analysis of FDG PET for use in the classification of Alzheimer's disease.* Paper presented at IEEE International Symposium on Biomedical Imaging: From Nano To Macro, Chicago, IL, 2011, pp.1082–1085.

8. Gillies RJ, Kinahan PE and Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016; 278: 563–577.

9. Kumar V, Gu YH, Basu S, *et al.* Radiomics: the process and the challenges. *Magn Reson Imaging* 2012; 30: 1234–1248.

10. Panth KM, Leijenaar RT, Carvalho S, *et al.* Is there a causal relationship between genetic changes and radiomics-based image features? An in vivo preclinical experiment with doxycycline inducible GADD34 tumor cells. *Radiother Oncol* 2015; 116: 462–466.

11. Aerts HJ, Velazquez ER, Leijenaar RT, *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014; 5: 4006.

12. Cameron A, Khalvati F, Haider MA, *et al.* MAPS: A quantitative radiomics approach for prostate cancer detection. *IEEE Trans Biomed Eng* 2016; 63: 1145–1156.

13. Zhou H, Vallieres M, Bai HX, *et al.* MRI features predict survival and molecular markers in diffuse lower-grade gliomas. *Neuro Oncol* 2017; 19: 862–870.

14. Vallieres M, Freeman CR, Skamene SR, *et al.* A radiomics model from joint FDG PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol* 2015; 60: 5471–5496.

15. Dasarathy BV and Holder EB. Image characterizations based on joint gray level—run length distributions. *Pattern Recognition Letters* 1991; 12: 497–502.

16. Amadasun M and King R. Textural features corresponding to textural properties. *IEEE Trans Syst Man Cybern* 1989; 19: 1264–1274.

17. Haralick RM, Shanmugam K and Dinstein IH. Textural features for image classification. *IEEE Trans Syst Man Cybern* 1973; 3: 610–621.

18. Galloway MM. Texture analysis using gray level run lengths. *Computer Graphics & Image Processing* 1975; 4(2): 172–179.

19. Thibault G, Fertil B, Navarro C, *et al.* Texture indexes and gray level size zone matrix application to cell nuclei classification. In: *10th International conference on pattern recognition and information processing*, Minsk, Belarus, 2009, pp.140–145.

20. Chu A, Sehgal CM and Greenleaf JF. Use of gray value distribution of run lengths for texture analysis. *Patt Recog Lett* 1990; 11: 415–419.

21. Chung AG, Khalvati F, Shafiee MJ, *et al.* Prostate cancer detection via a quantitative radiomics-driven conditional random field framework. *IEEE Access* 2015; 3: 2531–2541.

22. Huang Y, Liu Z, He L, *et al.* Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. *Radiology* 2016; 281: 947–957.

23. Wu W, Parmar C, Grossmann P, *et al.* Exploratory study to identify radiomics classifiers for lung cancer histology. *Front Oncol* 2016; 6: 71.

24. Jack CR, Jr., Wiste HJ, Vemuri P, *et al*. Brain beta-amyloid measures and magnetic resonance imaging atrophy both predict time-to-progression from mild cognitive impairment to Alzheimer's disease. *Brain* 2010; 133: 3336–3348.

25. Julkunen V, Niskanen E, Muehlboeck S, *et al*. Cortical thickness analysis to detect progressive mild cognitive impairment: a reference to Alzheimer's disease. *Dement Geriatr Cogn* 2009; 28: 404–412.

26. Julkunen V, Niskanen E, Koikkalainen J, *et al*. Differences in cortical thickness in healthy controls, subjects with mild cognitive impairment, and Alzheimer's disease patients: a longitudinal study. *J Alzheimers Dis* 2010; 21: 1141–1151.

27. Singh V, Chertkow H, Lerch JP, *et al*. Spatial patterns of cortical thinning in mild cognitive impairment and Alzheimer's disease. *Brain* 2006; 129: 2885–2893.

28. Ferreira LK, Diniz BS, Forlenza OV, *et al*. Neurostructural predictors of Alzheimer's disease: a meta-analysis of VBM studies. *Neurobiol Aging* 2011; 32: 1733.

29. Lerch JP, Pruessner J, Zijdenbos AP, *et al*. Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. *Neurobiol Aging* 2008; 29: 23–30.

30. Plant C, Teipel SJ, Oswald A, *et al*. Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. *Neuroimage* 2010; 50: 162.

31. Lerch JP and Evans AC. Cortical thickness analysis examined through power analysis and a population simulation. *Neuroimage* 2005; 24: 163–173.

32. Wee CY, Yap PT and Shen D. Prediction of Alzheimer's disease and mild cognitive impairment using baseline cortical morphological abnormality patterns. *Hum Brain Mapp* 2013; 34: 3411–3425.

33. Zec RF, Landreth ES, Vicari SK, *et al*. Alzheimer Disease assessment scale: a subtest analysis. *Alzheimer Dis Assoc Disord* 1992; 6: 164–181.

34. Mathotaarachchi S, Pascoal TA, Shin M, *et al*. Identifying incipient dementia individuals using machine learning and amyloid imaging. *Neurobiol Aging* 2017; 59: 80.

Visit SAGE journals online
journals.sagepub.com/
home/tan

⑤SAGE journals