2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Radiogenomics of Alzheimer's disease: exploring gene related metabolic imaging markers

Yanru Huang, Lanlan Li, Jiehui Jiang, *Member, IEEE* and Alzheimer's Disease Neuroimaging Initiative

***Abstract*— Alzheimer's disease (AD) is the most prevalent neurodegenerative disorder and considerably determined by genetic factors. Fluorodeoxyglucose positron emission tomography (FDG-PET) can reflect the functional state of glucose metabolism in the brain, and radiomic features of FDG-PET were considered as important imaging markers in AD. However, radiomic features are not highly interpretable, especially lack of explanation of underlying biological and molecular mechanisms. Therefore, this study used radiogenomics analysis to explore prognostic metabolic imaging markers by associating radiomics features and genetic data. In the study, we used the FDG-PET images and genotype data of 389 subjects (Cohort B) enrolled in the ADNI, including 109 AD, 134 healthy controls (HCs), 72 MCI non-converters (MCI-nc) and 74 MCI converters (MCI-c). Firstly, we performed a Genome-wide association study (GWAS) on the genotype data of 998 subjects (Cohort A), including 632 AD and 366 HCs after quality control (QC) steps to identify susceptibility loci as the gene features. Secondly, radiomics features were extracted from the preprocessed PET images. Thirdly, two-sample t-test, rank sum test and F-score were regarded as the feature selection step to select effective radiomic features. Fourthly, a support vector machine (SVM) was used to test the ability of the radiomic features to classify HCs, MCI and AD patients. Finally, we performed the Spearman correlation analysis on the genetic data and radiomic features. As a result, we identified rs429358 and rs2075650 as genome-wide significant signals. The radiomic approach achieved good classification abilities. Two prognostic FDG-PET radiomic features in the amygdala were proven to be correlated with the genetic data.**

***Keywords*— Imaging genomics, Fluorodeoxyglucose positron emission tomography, Genome-wide association studies (GWAS)**

INTRODUCTION

Alzheimer's disease (AD) is a kind of nervous system degenerative disease characterized by cognitive dysfunction associated with age, so far affecting millions of people worldwide. AD is considerably determined by genetic factors. Therefore, elucidating the genetic architecture of AD-related phenotypic traits, ideally those linked to the underlying disease process, holds great promise in gaining deeper insights into the genetic basis of AD and in developing better clinical prediction models.

* This study was supported by grants received from the National Natural Science Foundation of China (82020108013); the Shanghai Municipal Science and Technology Major Project (grant number 2017SHZDZX01).

Yanru Huang, Lanlan Li and Jiehui Jiang are with the Institute of biomedical engineering, Shanghai University, Shanghai, China (corresponding author to provide phone: +86 139-1892-0926; e-mail: jiangjiehui@shu.edu.cn).

Fluorodeoxyglucose positron emission tomography (FDG-PET) is a recognized neuroimaging marker of functional degeneration. It can be used to increase the certainty of the pathophysiological process of AD in research and can also be used as a clinical diagnostic tool. The relationship between AD and glucose metabolism has always been a concern because brain activity is inseparable from the effects of glucose. Further, researches[1] have shown that radiomic features in cortical glucose metabolism can not only assess early changes in cognitive function, but also detect changes in disease progression. However, radiomic features are not highly interpretable, especially lack of explanation of underlying biological and molecular mechanisms. Little is known about the correlation between glucose metabolism and heredity.

Radiogenomics is an emerging field that holds great promise for a system biology of the brain to better understand complex neurobiological systems, from genetic determinants to cellular processes to the complex interplay of brain structure, function, behavior and cognition[2]. It can help physicians to gain insights into the underlying pathologic processes of AD.

Therefore, in this study, we aimed to explore prognostic radiomic markers for AD by associating radiomics features and Genetic data.

TABLE I. THE DEMOGRAPHIC INFORMATION OF DATA

Cohort A								
	AD (n = 632)				***HC (n = 366)***			
Gender(M/F)	368/264				177/189			
Age	74.5±7.4				74.0±5.7			
Education	15.5±2.9				16.4±2.7			
MMSE	22.3±4.0				29.0±1.1			
	Cohort B				**P value (Two sample t-test)**			
	AD (n=109)	***HC (n = 134)***	***MCI (n = 146)***		***AD vs HC***	***MCI vs HC***	***AD vs MCI***	***MCI-c vs MCI-nc***
			MCI-c (n = 74)	***MCI-nc (n = 72)***				
Gender (M/F)	69/40[a c]	68/66[a b]	49/25[d]	45/27[d]	-	-	-	-
Age	74.2 ±8.5	74.7 ±6.3	74.6 ±6.9	72.6 ±6.8	0.58	0.17	0.57	0.09
Education	15.8 ±2.8	16.4 ±2.5	15.9 ±2.6	16.1 ±2.8	0.11	0.19	0.70	0.74
MMSE	22.8 ±3.0[a c]	29.1 ± 1.1[a b]	26.5 ±2.2[d]	28.4 ±1.6[d]	<0.001,	<0.001,	<0.001	<0.001

Note: MCI-c = MCI-converters, MCI-nc = MCI non-converters, MMSE = Mini Mental State Examination, Age, Education, MMSE are given as mean ± standard deviation.

[a] $p < 0.05$: AD vs HC, [b] $p < 0.05$: MCI vs HC, [c] $p < 0.05$: AD vs MCI, [d] $p < 0.05$: MCI-c vs MCI-nc.

MATERIALS AND METHODS

A. Participants

Data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). One goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. For up-to-date information, see www.adni-info.org. The ADNI participants utilized for our analyses originated from ADNI1 and ADNIGo/2.

The genetic data for GWAS analysis was based on 998 subjects (Cohort A) enrolled in the ADNI, including 632 AD and 366 HCs. Then, 389 subjects (Cohort B) who own both genotype data and FDG-PET images were involved, including 109 AD, 134 healthy controls (HCs), 72 MCI non-converters (MCI-nc) and 74 MCI converters (MCI-c). As shown in Table 1, a chi-squared test was conducted to evaluate the gender effect, and analysis of variance (ANOVA) with Bonferroni correction and two sample t-test were used to evaluate the statistical differences in the age, MMSE, and Education.

B. Experimental framework

As shown in Fig. 1, the framework of the study consists of four parts. First, quality control procedures were performed to exclude unsatisfactory genotype data and the FDG-PET images ought to be preprocessed for further study. Next, we performed the genome-wide association studies (GWAS) analysis on the genotype data to identify susceptibility loci in AD. Then, some typical radiomics features extracted from FDG-PET were selected with two-sample t-test, rank sum test and F-score. After that, the support vector machine (SVM) classifiers were applied to verify the diagnostic capabilities of the features selected. Finally, we conducted the Spearman correlation analysis on the genetics data and radiomics features.

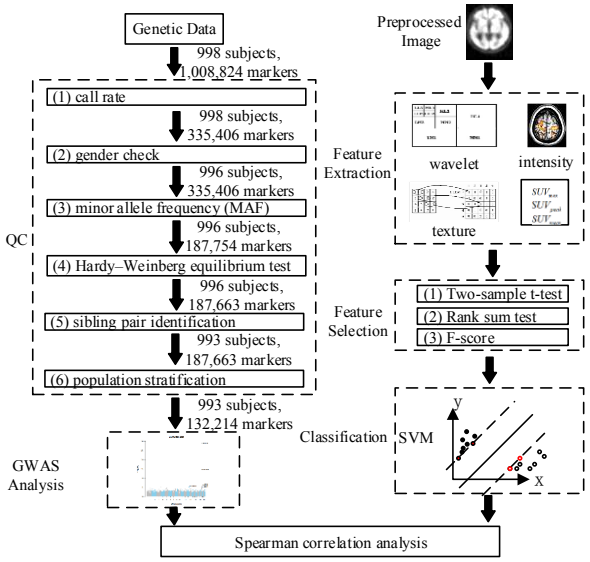

Figure 1. The framework of the study

C. Data preprocessing

The genotype data of the ADNI samples was performed using the Illumina 610-Quad BeadChip (Illumina, Inc., San Diego, CA, USA). The APOE genotype is an established risk factor in AD research[3]. Therefore, we used the APOE ε2/ε3/ε4 status information from the ADNI clinical database for each participant. Using the PLINK software package (http://pngu.mgh.harvard.edu/~purcell/plink/), release v1.9, the following quality control (QC) steps were performed on these genotype data. The QC criteria for the SNP data include (1) call rate per SNP≥ 90%, call rate per participant ≥ 90%, (2) gender check, (3) minor allele frequency (MAF) ≥ 5%, and (4) Hardy–Weinberg equilibrium test of $p \leq 10^{-6}$, (5) sibling pair identification, (6) population stratification.

Each FDG-PET image was preprocessed with Statistical Parametric Mapping V.12 (SPM12; https://www.fil.ion.ucl.ac.uk/spm/software/spm12)software package. Firstly, linear and nonlinear 3D transformations were used to normalize the scan space of each subject into the Montreal Neurological Institute (MNI) space. Then, the spatially normalized PET images were smoothed in three-dimensional space using an 8 mm full width at half maximum (FWHM) Gaussian kernel. Finally, taking account of the difference of FDG absorption of each individual, the smoothed images were normalized to a range of 0 to 255.

D. GWAS analysis

GWAS has been emerged as a popular tool to identify genetic variants that are associated with disease risk. The standard analysis of a case-control GWAS involves assessing the association between each individual genotyped SNP and disease risk[4]. A Manhattan plot and a quantile-quantile (Q–Q) plot were used to visualize GWAS results. All association results surviving the significance threshold of $p < 3.78 \times 10^{-7}$ (i.e., 0.05/132,214 markers) were saved and prepared for additional pattern analysis.

E. Radiomic feature extraction

In this study, the radiomics tool developed by Vallieres4 (https://github.com/mvallieres/radiomics) was used to exact features[5]. It has previously been proved that the extracted features can be used for the prediction and diagnosis of AD[6]. The features we used included the global characteristics, texture matrices features and standard uptake value (SUV) features. Table 2 provides the concreted feature names. Wavelet bandpass filtering 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 λ, and the values of λ were 1/2, 2/3, 1, 3/2 and 2. The number of gray levels was selected as 32 and 64. With the different choices of bandpass sub-bands and quantized gray levels, 470 features were extracted from each brain region in total.

TABLE II. EXTRACTED RADIOMIC FEATURES

Feature type	Feature name	Feature type	Feature name
Global	Variance		Run-length variance (RLV)
	Skewness	gray-level size zone matrix (GLSZM)	Small zone emphasis (SZE)
	Kurtosis		Large zone emphasis (LZE)
gray-level co-occurrence	Energy		Gray-level nonuniformity (GLN)

matrix (GLCM)	Contrast		Zone-size nonuniformity (ZSN)
	Correlation		Zone percentage (ZP)
	Homogeneity		Low-gray-level zone emphasis (LGZE)
	Variance		High-gray-level zone emphasis (HGZE)
	Sum average		Small zone low-gray-level emphasis (SZLGE)
	Entropy		Small zone high-gray-level emphasis (SZHGE)
	Auto correlation		Large zone low-gray-level emphasis (LZLGE)
	Dissimilarity		Large zone high-gray-level emphasis (LZHGE)
gray-level run-length matrix (GLRLM)	Short-run emphasis (SRE)		Gray-level variance (GLV)
	Long-run emphasis (LRE)		Zone-size variance (ZSV)
	Gray-level nonuniformity (GLN)	neighborhood gray-tone difference	Coarseness
	Run-length nonuniformity (RLN)	matrix (NGTDM)	Contrast
	Run percentage (RP)		Busyness
	Low-gray-level run emphasis (LGRE)		Complexity
	High-gray-level run emphasis (HGRE)		Strength
	Short-run low-gray-level emphasis (SRLGE)	SUV	Maximum SUV (SUVmax)
	Short-run high gray-level emphasis (SRHGE)		Peak SUV (SUVpeak)
	Long-run low-gray-level emphasis (LRLGE)		Mean SUV （SUVmean）
	Long-run high-gray-level emphasis (LRHGE)		Area under the curve of the cumulative SUV volume histogram （AUC-CSH）
	Gray-level variance (GLV)		

F. Radiomic feature selection

Feature selection helped to speed up the classification process by decreasing computational time and increases the performance of classification accuracy.

Firstly, we performed the significance test of difference (P<0.01) to select features that were effective for classification. Two-sample t-test was applied to the features that obeyed normal distribution and generality variances are equal. The rank sum test was utilized for the remaining features.

After that, F-score was employed for obtaining subset features. F-score is a simple feature selection filter method by evaluating the discrimination of two sets of real numbers. Given training vectors X_k, k=1,2,⋯,m, if the number of positive and negative instances are $n+$ and $n-$, respectively, then the F-score of the ith feature is defined as[7]:

$$F(\mathrm{i}) \equiv \frac{(\overline{x_i}^{(+)} - \overline{x_i})^2 + (\overline{x_i}^{(-)} - \overline{x_i})^2}{\frac{1}{n_+ - 1}\sum_{k=1}^{n_+}(x_{k,i}^{(+)} - \overline{x_i}^{(+)})^2 + \frac{1}{n_- - 1}\sum_{k=1}^{n_-}(x_{k,i}^{(-)} - \overline{x_i}^{(-)})^2}$$

Where $\overline{x_i}$, $\overline{x_i}^{(+)}$ and $\overline{x_i}^{(-)}$ are the averages of the ith feature of the whole, positive, and negative datasets. $x_{k,i}^{(+)}$ is the ith feature of the kth positive instance and $x_{k,i}^{(-)}$ is the ith feature of the kth negative instance.

In addition, the number of occurrences of extracted characteristics was measured. The top 20 characteristics with the highest number of appearances were recorded for further research.

G. Classification

To verify the diagnostic capabilities of the feature-set selected above, we performed four SVM classification experiments, including AD versus HCs, MCI versus HCs, AD versus MCI and MCI-c versus MCI-nc. SVM is a supervised learning method, which works by finding a hyperplane that best separates two data groups[7]. The SVM classifiers were trained by training data in n-dimensional training space after which test subjects are classified according to their position in n-dimensional feature space. To improve the performance of SVM, Grid Search[8] was applied to optimize the SVM parameters in this study. To evaluate the classification performance, we conducted 5-fold cross-validation 200 times for each SVM classification. In addition, we performed comparative trials. A clinical model including age, gender, education and MMSE score was also developed and compared to the PET imaging model, as well as the combined model of clinical and imaging information.

H. Spearman correlation analysis

Spearman correlation was applied as a criterion to estimate whether genes had a correlation with the imaging features of high frequency. Among all the Gene indicators, theta proved to have the most direct relationship with SNP changes. The theta value represents the normalized theta value of an SNP of the sample. Therefore, we chose theta values as the Genetic features. Finally, the high-frequency radiomic features were selected to conduct correlation analysis with gene features.

I . Statistical analysis

Demographic characteristics were compared based on two-sample t test or the chi-square t test. All statistical analyses were performed in SPSS Version 22.0 software (SPSS Inc., Chicago, IL). All P value < 0.05 was considered significant.

RESULTS

A. Outcomes of GWAS analysis

After the GWAS analysis, we observed two genome-wide significant signals on chromosome 19, APOE (rs429358, the epsilon 4 marker) and TOMM40 (rs2075650, OR=2.73, P=2.23E-15). Fig. 2 showed the Manhattan and Q–Q plots of the GWAS analysis.

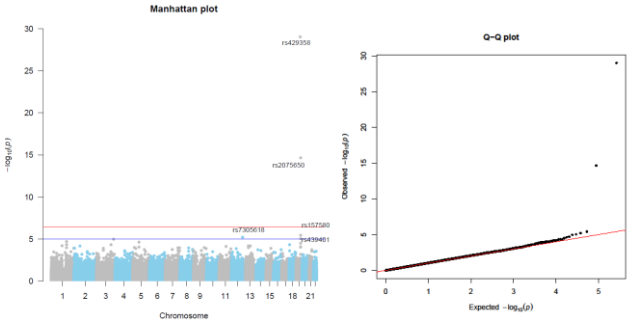

Figure 2. Manhattan and Q–Q plots of genome-wide association study (GWAS). The horizontal lines in the Manhattan plot display the

cutoffs for two significant levels: blue line for $p < 10^{-5}$, and red line for $p < 3.78 \times 10^{-7}$. Genomic inflation factor is 1.073.

B. classification performance

Table 3 provided the classification accuracy, sensitivity, specificity and area under curve (AUC) of the clinical model, PET imaging model and the combined model. The SVM classifiers using selected radiomic features could achieve accuracies of 92.9±0.8%, 83.7±0.6%, 87.9±1.1% and 88.0 ±1.9% in AD vs HC, MCI vs HC, AD vs MCI, and MCI-c vs MCI-nc, respectively. This indicated that radiomic approach used was stable, and achieved good classification abilities.

TABLE III. RESULTS OF EXPERIMENTS

	Model	ACC(%)	SEN(%)	SPEC(%)	AUC
AD vs HC	***Clinical***	86.8±4.6	83.9±6.5	89.2±4.9	0.83±0.05
	PET	92.9±0.8	84.8±1.6	99.5±0.5	0.89±0.02
	combine	97.4±0.5	98.8±0.5	95.7±0.9	**0.94±0.01**
MCI vs HC	***Clinical***	69.7±1.2	64.0±1.5	76.3±1.8	0.68±0.02
	PET	83.7±0.6	69.9±1.0	98.8±0.7	0.81±0.02
	combine	84.5±0.9	96.1±1.3	74.0±1.3	**0.82±0.02**
AD vs MCI	***Clinical***	82.9±0.8	77.7±1.5	87.0±0.9	0.80±0.02
	PET	87.9±1.1	82.9±1.6	91.8±1.4	0.84±0.02
	combine	94.2±0.6	95.5±0.8	92.5±0.9	**0.91±0.01**
MCI-c vs MCI-nc	***Clinical***	69.7±1.6	63.8±2.3	76.5±2.4	0.65±0.03
	PET	88.0±1.9	95.0±2.7	81.2±2.8	0.81±0.03
	combine	88.5±1.8	83.0±2.5	94.1±2.4	**0.82±0.02**

Note: the explicit results are given as mean ± standard deviation.

C. Results of correlation analysis,

The possible association of radiomic features with genes was explored based on the Spearman correlation. Setting the threshold for P<0.05 after FDR adjusting, two radiomic features with different correlation coefficients R were proved to be correlated to gene features. As shown in Table 4, coarseness of the amygdala with different λ values were proved to be associated with the significant SNP, rs2075650. Coarseness represented the level of spatial change rate of intensity. Therefore, a high value of coarseness meant that the gray level difference between the groups was small.

TABLE IV. RESULTS OF CORRELATION ANALYSIS

ROI	feature	λ	R	P
Amygdala_L	Coarseness	1	0.11	0.047
Amygdala_L	Coarseness	3/2	0.12	0.047

DISCUSSION

This study explored the Gene related metabolic imaging markers of AD using radiogenomics analysis. The outcomes of this study may help physicians to gain insights into the explanations of underlying biological and molecular mechanisms for FDG-PET radiomic features.

A. Outcomes of GWAS analysis

At the $p < 3.78 \times 10^{-7}$ significance level, two SNPs were identified in the GWAS analysis. As a well-established AD risk factor, the APOE SNP rs429358 was determined as the most prominent imaging genetics. Moreover, the second significant SNP, rs2075650 (TOMM40), supported the finding of TOMM40 as a gene adjacent to APOE and an additional contributor to AD [9]. In addition, SNPs from several other candidate genes showed less robust indications of possible association that nonetheless encourage further investigation. These results were consistent of previous studies.

B. Results of radiomic features

The amygdala is located in the medial temporal lobe region of the brain and has been implicated in emotional processes, survival instincts and aspects of memory, especially for emotional components. Amygdala is prominently related to AD and its progression[10] and has been used to assist the clinical diagnosis of AD. Studies on FDG-PET have demonstrated different usage patterns of glucose metabolism in the amygdala between AD and healthy control subjects. These results were also consistent with previous studies.

C. Limitation

The study is limited by several factors as below. (1) data we used were all from the ADNI database, which lack universality. The validation part using data from other sources needs to be studied in the future; (2) the number of samples participated in GWAS analysis needs to be augmented to obtain more significant SNPs.

CONCLUSION

In the study, we identified two Gene-influenced FDG-PET radiomic features (coarseness of the amygdala). This study showed that radiogenimic approach may be useful for the AD study.

REFERENCES

[1] M. Lehmann et al., "Diverging patterns of amyloid deposition and hypometabolism in clinical variants of probable Alzheimer's disease," Brain, vol. 136, no. 3, pp. 844-858, 2013.

[2] H. Wang et al., "Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning," Bioinformatics, vol. 28, no. 12, pp. i127-i136, 2012.

[3] L. A. Farrer et al., "Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: a meta-analysis," Jama, vol. 278, no. 16, pp. 1349-1356, 1997.

[4] M. C. Wu et al., "Powerful SNP-set analysis for case-control genome-wide association studies," The American Journal of Human Genetics, vol. 86, no. 6, pp. 929-942, 2010.

[5] M. Vallières, C. R. Freeman, S. R. Skamene, and I. El Naqa, "A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities," Physics in Medicine & Biology, vol. 60, no. 14, p. 5471, 2015.

[6] L. Sørensen, C. Igel, and M. Nielsen, "Hippocampal texture predicts conversion from MCI to Alzheimer's disease," Alzheimer's & Dementia, vol. 4, no. 9, p. P52, 2013.

[7] Y.-W. Chen and C.-J. Lin, "Combining SVMs with various feature selection strategies," in Feature extraction: Springer, 2006, pp. 315-324.

[8] I. Syarif, A. Prugel-Bennett, and G. Wills, "SVM parameter optimization using grid search and genetic algorithm to improve classification performance," Telkomnika, vol. 14, no. 4, p. 1502, 2016.

[9] L. Osherovich, "TOMMorrow's AD marker," Science-Business eXchange, vol. 2, no. 30, pp. 1165-1165, 2009.

[10] A. M. Fjell et al., "CSF biomarkers in prediction of cerebral and clinical change in mild cognitive impairment and Alzheimer's disease," Journal of Neuroscience, vol. 30, no. 6, pp. 2088-2101, 2010.