A multi-omics framework based on machine learning as a predictor of cognitive impairment progression in early Parkinson's disease

Yang Luo [1], YaQin Xiang [1], JiaBin Liu [1], YuXuan Hu [1], JiFeng Guo [1,2,3,4*]

[1] Department of Neurology, XiangYa Hospital, Central South University, Changsha, China.

[2] National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, Hunan, China.

[3] Key Laboratory of Hunan Province in Neurodegenerative Disorders, Central South University, Changsha, Hunan, China.

[4] Centre for Medical Genetics & Hunan Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan, China.

* Corresponding Author: JiFeng Guo, guojifeng@csu.edu.cn

Correspondence address: Department of Neurology, XiangYa Hospital, No. 87 Xiangya Road, Changsha, Hunan, China, 410008

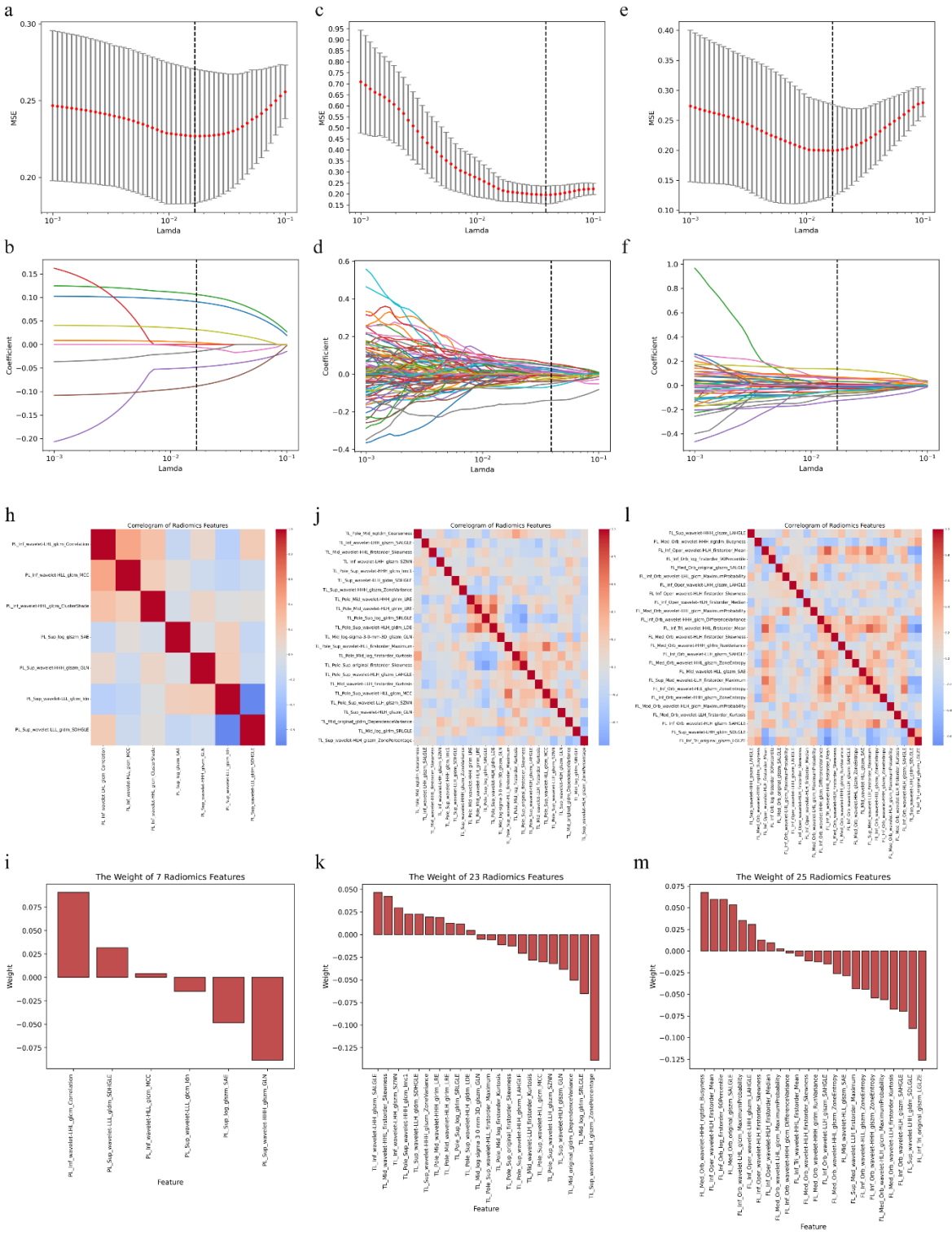

Figure S1. The radiomics signatures selection. a-f, the Least absolute shrinkage and selection operator(LASSO) analysis of radiomics signatures in parietal lobe (a, b), temporal lobe (c, d), frontal lobe(e, f). h-m, the heatmap of Pearson correlation coefficient and weight for significantly different radiomics signatures in parietal lobe (h, i), temporal lobe (j, k), frontal lobe(l, m).

Table S1. Information of 42 including genes

Gene	Chrome	Protein coding
ADAMTS4	1	ADAM Metallopeptidase With Thrombospondin Type 1 Motif 4
CR1	1	Complement C3b/C4b Receptor 1
GBA	1	Glucosylceramidase Beta 1
BIN1	2	Bridging Integrator 1
INPP5D	2	Inositol Polyphosphate-5-Phosphatase D
NCK2	2	NCK Adaptor Protein 2
SPRED2	2	Sprouty Related EVH1 Domain Containing 2
TMEM163	2	Transmembrane Protein 163
CLNK	4	Cytokine Dependent Hematopoietic Cell Linker
SNCA	4	Synuclein Alpha
EPHA1	7	EPH Receptor A1
SPATA48	7	spermatogenesis associated 48
TMEM106B	7	Transmembrane Protein 106B
ZCWPW1/NYAP1	7	Zinc Finger CW-Type And PWWP Domain Containing 1
CCDC6	10	Coiled-Coil Domain Containing 6
ECHDC3	10	Enoyl-CoA Hydratase Domain Containing 3
TSPAN14	10	Tetraspanin 14
BDNF	11	Brain Derived Neurotrophic Factor
MADD/SPI1	11	MAP Kinase Activating Death Domain
MS4A4A	11	Membrane Spanning 4-Domains A4A
PICALM	11	Phosphatidylinositol Binding Clathrin Assembly Protein
SORL1	11	Sortilin Related Receptor 1
LRRK2	12	Leucine Rich Repeat Kinase 2
FERMT2	14	FERM Domain Containing Kindlin 2
RIN3	14	Ras And Rab Interactor 3
SLC24A4	14	Solute Carrier Family 24 Member 4
ADAM10	15	ADAM Metallopeptidase Domain 10
APH1B	15	Aph-1 Homolog B, Gamma-Secretase Subunit
SPPL2A	15	Signal Peptide Peptidase Like 2A
KAT8	16	Lysine Acetyltransferase 8
PLCG2	16	Phospholipase C Gamma 2
ABI3	17	ABI Family Member 3
ACE	17	Angiotensin I Converting Enzyme
GRN	17	Granulin Precursor
MAPT	17	Microtubule Associated Protein Tau
SCIMP/RABEP1	17	SLP Adaptor And CSK Interacting Membrane Protein
TSPOAP1-AS1	17	TSPOAP1, SUPT4H1 And RNF43 Antisense RNA 1
ABCA7	19	ATP Binding Cassette Subfamily A Member 7
APOE	19	Apolipoprotein E
CD33	19	CD33 Molecule
APP	21	Amyloid Beta Precursor Protein
COMT	22	Catechol-O-Methyltransferase

Table S2. Selected differential radiomics signatures from the parietal lobe (PL).

	Parietal lobe	nCI	CI	LASSO-coef	P value
1	PL Inf wavelet-LHL glcm Correlation	4.358E-01	4.486E-01	0.0811	0.016
2	PL Inf wavelet-HHL glcm ClusterShade	-0.0004	0.0022	0.0961	0.010
3	PL Sup wavelet-LLL gldm SDHGLE	0.0004	0.0005	0.0242	0.037
4	PL Inf wavelet-HLL glcm MCC	0.4562	0.4736	0.0000	0.047
5	PL Sup wavelet-LLL glcm Idn	0.9565	0.9552	-0.0060	0.047
6	PL Sup log glszm SAE	0.2680	0.2529	-0.0428	0.013
7	PL Sup wavelet-HHH glszm GLN	2.109E+01	1.984E+01	-0.0754	0.044

Abbreviations were provided in Table S5

Table S3. Selected differential radiomics signatures from the temporal lobe (TL).

	Temporal lobe	nCI	CI	LASSO-coef	P value
1	TL Pole Mid ngtdm Coarseness	6.592E-04	6.756E-04	0.0560	0.017
2	TL Inf wavelet-LHH glszm SALGLE	0.1343	0.1816	0.0467	0.005
3	TL Mid wavelet-HHL firstorder Skewness	-0.1210	-0.0255	0.0422	0.037
4	TL Inf wavelet-LHH glszm SZNN	0.3544	0.3751	0.0295	0.041
5	TL Pole Sup wavelet-HHH glcm Imc1	-0.0032	-0.0021	0.0226	0.010
6	TL Sup wavelet-LLH gldm SDHGLE	0.0009	0.0011	0.0225	0.031
7	TL Sup wavelet-HHH glszm ZoneVariance	4.757E+07	4.998E+07	0.0195	0.041
8	TL Pole Mid wavelet-HHH glrlm LRE	4.368	4.389	0.0191	0.048
9	TL Pole Mid wavelet-HLH glrlm LRE	5.6025	5.6796	0.0127	0.023
10	TL Pole Sup log glrlm SRLGLE	0.0282	0.0377	0.0117	0.004
11	TL Pole Sup wavelet-HLH gldm LDE	194.5817	196.6367	0.0048	0.038
12	TL Mid log-sigma-3-0-mm-3D glszm GLN	50.7196	47.7722	-0.0051	0.025
13	TL Pole Sup wavelet-HLL firstorder Maximum	171.4687	129.4312	-0.0057	0.001
14	TL Pole Mid log firstorder Kurtosis	3.0801	2.7379	-0.0113	0.041
15	TL Pole Sup original firstorder Skewness	0.0615	-0.1096	-0.0127	4.191E-03
16	TL Pole Sup wavelet-HLH glszm LAHGLE	5.185E+07	4.081E+07	-0.0207	0.020
17	TL Mid wavelet-LLH firstorder Kurtosis	6.6796	6.4199	-0.0283	0.045
18	TL Pole Sup wavelet-HLL glcm MCC	0.5513	0.4951	-0.0301	0.004
19	TL Pole Sup wavelet-LLH glszm SZNN	0.2291	0.2136	-0.0319	0.002
20	TL Sup wavelet-HLH glszm GLN	26.8952	24.9611	-0.0383	0.014
21	TL Mid original gldm DependenceVariance	31.4291	30.3642	-0.0504	0.045
22	TL Mid log glrlm SRLGLE	0.0309	0.0276	-0.0653	0.040
23	TL Sup wavelet-HLH glszm ZonePercentage	8.627E-04	7.475E-04	-0.1390	4.058E-05

Abbreviations were provided in Table S5.

Table S4. Selected differential radiomics signatures from the frontal lobe (FL).

	Frontal lobe	nCI	CI	LASSO-coef	P value
1	FL Sup wavelet-HHH glszm LAHGLE	4.027E+07	4.584E+07	0.1311	0.0485
2	FL Med Orb wavelet-HHH ngtdm Busyness	0.0805	0.3029	0.0643	0.0466
3	FL Inf Oper wavelet-HLH firstorder Mean	-0.0029	-0.0002	0.0592	0.0331
4	FL Inf Orb log firstorder 90 Percentile	0.3510	0.3580	0.0570	0.0473
5	FL Med Orb original glszm SALGLE	2.498E-01	2.496E-01	0.0530	0.0378
6	FL Inf Orb wavelet-LHL glcm Maximum Probability	2.1686	2.0835	0.0337	0.0291
7	FL Inf Oper wavelet-LHH glszm LAHGLE	1.934E-02	-5.548E-02	0.0291	0.0349
8	FL Inf Oper wavelet-HLH firstorder Skewness	1.4090	1.3976	0.0114	0.0264
9	FL Inf Oper wavelet-HLH firstorder Median	10.7631	8.7056	0.0089	0.0191
10	FL Med Orb wavelet-LHL glcm Maximum Probability	3.5373	3.3751	0.0009	0.0328
11	FL Inf Orb wavelet-HHH glcm Difference Variance	0.4373	0.4283	-0.0012	0.0230
12	FL Inf Tri wavelet-HHL firstorder Mean	85.4594	76.4503	-0.0059	0.0459
13	FL Med Orb wavelet-HLH firstorder Skewness	5.7516	5.6502	-0.0115	0.0430
14	FL Med Orb wavelet-HHH glrlm Run Variance	3.0556	2.9437	-0.0125	0.0425
15	FL Inf Orb wavelet-LLH glszm SAHGLE	0.2811	0.2793	-0.0157	0.0459
16	FL Med Orb wavelet-HHL glszm Zone Entropy	7.5565	6.6814	-0.0251	0.0283
17	FL Mid wavelet-HLL glszm SAE	8.2639	6.7976	-0.0280	0.0393
18	FL Sup Med wavelet-LLH firstorder Maximum	0.0012	0.0010	-0.0402	0.0257
19	FL Inf Orb wavelet-HLL glszm Zone Entropy	0.1459	0.1341	-0.0422	0.0346
20	FL Inf Orb wavelet-HHH glszm Zone Entropy	3.838E+08	4.341E+08	-0.0526	0.0247
21	FL Med Orb wavelet-HLH glcm Maximum Probability	5440.7702	6419.7249	-0.0535	0.0486
22	FL Med Orb wavelet-LLH firstorder Kurtosis	1.6059	1.5908	-0.0651	0.0283
23	FL Inf Orb wavelet-HLH glszm SAHGLE	9.6749	11.4737	-0.0666	0.0389
24	FL Sup wavelet-LHH gldm SDLGLE	0.0442	0.0530	-0.0860	0.0107
25	FL Inf Tri original glszm LGLZE	0.3714	0.3794	-0.1205	0.0291

Abbreviations were provided in Table S5.

Table S5. Abbreviation.

Abbreviation	Meaning
Ada	AdaBoost
AD	Alzheimer's disease
AUC	Area Under the ROC curve
Bayes	Bayesian Classification
CSF	cerebrospinal fluid
CI	cognitive impairment
TL	temporal lobe
FL	frontal lobe
GLCM	Gray Level Co-occurrence Matrix
GLDM	Gray Level Dependence Matrix
GLRLM	Gray Level Run Length Matrix
GLSZM	Gray Level Size Zone Matrix
GLN	GrayLevelNonUniformity
GLN	GrayLevelNonUniformity
Imc	Informational Measure of Correlation
Idn	Inverse Difference Normalized
LAHGLE	LargeAreaHighGrayLevelEmphasis
LAHGLE	LargeAreaHighGrayLevelEmphasis
LDE	LargeDependenceEmphasis
LASSO	Least absolute shrinkage and selection operator
LR	Logistic Regression
LRE	LongRunEmphasis
LGLZE	LowGrayLevelZoneEmphasis
MCC	Maximal Correlation Coefficient
MCI	mild cognitive impairment
NGTDM	Neighbouring Gray Tone Difference Matrix
PL	parietal lobe
PPMI	Parkinson’s Progression Markers Initiative
PD	Parkinson's disease
PDD	Parkinson's disease dementia
PET	positron emission tomography
PCA	Principal Component Analysis
QOL	quality of life
RF	Random Forest
ROC	Receiver Operating Characteristic
SRLGLE	ShortRunLowGrayLevelEmphasis
SNP	Single Nucleotide Polymorphism
SZNN	SizeZoneNonUniformityNormalized
SAE	SmallAreaEmphasis
SAHGLE	SmallAreaHighGrayLevelEmphasis
SALGLE	SmallAreaLowGrayLevelEmphasis
SDLGLE	SmallDependenceLowGrayLevelEmphasis
SDHGLE	SmallDependenceLowGrayLevelEmphasis
SVM	Support Vector Machine
WGS	whole-genome sequencing
α-syn	α-synuclein
Aβ42	β-amyloid 42