EUROPEAN SOCIETY OF RADIOLOGY
European Radiology

IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE

Using interpretable deep learning radiomics model to diagnose and predict progression of early AD disease spectrum: a preliminary [^{18}F]FDG PET study

Jiehui Jiang[1*], Chenyang Li[1], Jiaying Lu[2], Jie Sun[1], Xiaoming Sun[3], Jiacheng Yang[1], Luyao Wang[1], Chuantao Zuo[2,4*], Kuangyu Shi[5] for the Alzheimer's Disease Neuroimaging Initiative

Abstract

Objectives In this study, we propose an interpretable deep learning radiomics (IDLR) model based on [^{18}F]FDG PET images to diagnose the clinical spectrum of Alzheimer's disease (AD) and predict the progression from mild cognitive impairment (MCI) to AD.

Methods This multicentre study included 1962 subjects from two ethnically diverse, independent cohorts (a Caucasian cohort from ADNI and an Asian cohort merged from two hospitals in China). The IDLR model involved feature extraction, feature selection, and classification/prediction. We evaluated the IDLR model's ability to distinguish between subjects with different cognitive statuses and MCI trajectories (sMCI and pMCI) and compared results with radiomic and deep learning (DL) models. A Cox model tested the IDLR signature's predictive capability for MCI to AD progression. Correlation analyses identified critical IDLR features and verified their clinical diagnostic value.

Results The IDLR model achieved the best classification results for subjects with different cognitive statuses as well as in those with MCI with distinct trajectories, with an accuracy of 76.51% [72.88%, 79.60%], (95% confidence interval, CI) while those of radiomic and DL models were 69.13% [66.28%, 73.12%] and 73.89% [68.99%, 77.89%], respectively. According to the Cox model, the hazard ratio (HR) of the IDLR model was 1.465 (95% CI: 1.236–1.737, $p < 0.001$). Moreover, three crucial IDLR features were significantly different across cognitive stages and were significantly correlated with cognitive scale scores ($p < 0.01$).

Conclusions Preliminary results demonstrated that the IDLR model based on [^{18}F]FDG PET images enhanced accuracy in diagnosing the clinical spectrum of AD.

A list of members and their affiliations appears in the Supplementary Information.

Jiehui Jiang, Chenyang Li and Jiaying Lu contributed equally to this work.

*Correspondence:
Jiehui Jiang
jiangjiehui@shu.edu.cn
Chuantao Zuo
zuochuantao@fudan.edu.cn
Full list of author information is available at the end of the article

Springer

Key Points

Question *The study addresses the lack of interpretability in existing DL classification models for diagnosing the AD spectrum.*

Findings *The proposed interpretable DL radiomics model, using radiomics-supervised DL features, enhances interpretability from traditional DL models and improves classification accuracy.*

Clinical relevance *The IDLR model interprets DL features through radiomics supervision, potentially advancing the application of DL in clinical classification tasks.*

Keywords Interpretable deep learning radiomics model, Alzheimer's disease, Mild cognitive impairment, Radiomics, Cox model

Introduction

Alzheimer's disease (AD) is the most prevalent neurodegenerative disorder and is primarily characterised by the abnormal accumulation of pathological proteins, including β-amyloid and tau-neurofibrillary tangles, leading to subsequent neurodegeneration [1, 2]. Early identification of the clinical spectrum of AD, which mainly includes mild cognitive impairment (MCI) due to AD and AD dementia, is crucial due to its irreversible congenital features, subtle yet progressive nature, and the existence of various subtypes. Although both clinical and research efforts are presently concentrated on MCI as a precursor to AD, the neuropathological underpinnings of MCI are diverse, with some MCI patients remaining stable (sMCI) or even reverting to a normal state, while others progress to AD (pMCI) [3, 4]. Hence, it holds immense clinical significance to identify biomarkers and develop predictive models encompassing the entire spectrum of AD and forecasting the progression from MCI to AD dementia.

Positron emission tomography (PET) is a noninvasive neuroimaging technique that plays a pivotal role in tracking pathophysiological changes within the AD brain through real-time visualisation and various molecular correlations. Among the array of available PET tracers, ^{18}F-Fluorodeoxyglucose ([^{18}F]FDG) PET has emerged as a clinical powerhouse, surpassing magnetic resonance imaging (MRI) in utility due to its potential for early detection of neurodegenerative abnormalities [5] and its superior accessibility to newly developed amyloid- and tau-PET [6–9]. In contrast to other imaging markers, such as amyloid-PET or tau-PET ligands, [^{18}F]FDG is the most widely used PET ligand worldwide and can be used at any institution that offers PET examinations [10–14]. The clinical diagnostic value of [^{18}F]FDG PET imaging in distinguishing various stages of the clinical spectrum of AD, as well as distinguishing between sMCI and pMCI has been demonstrated [15]. Notably, the advent of artificial intelligence has led to the widespread application of radiomics. In a study by Jiang et al [16], the use of [^{18}F]FDG PET for predicting MCI conversion yielded a Harrell's consistency coefficient of 70.3%, another in a study by Li et al [17], the [^{18}F]FDG PET achieved an accuracy of 91.5% in distinguishing between AD and health controls (HC). Deep learning (DL) methods also offer significant advantages in this field, Zhao et al [18] and Jiang et al [19] used DL models to distinguish the AD disease spectrum on FDG PET and Tau-PET images, respectively. Radiological image analysis and, thus, radiomics benefit greatly from the current enormous developments in machine learning, especially in DL. Furthermore, effective and efficient integration of different clinical, imaging and molecular profile data is often also necessary to understand complex diseases and make accurate diagnoses to determine the best possible treatment [20–22]. Nevertheless, both models have unique advantages and drawbacks. Radiomics methods can extract interpretable image markers; however, they require prior knowledge and manual coding for defining regions of interest (ROI) and feature extraction. In contrast, DL methods excel in achieving high classification and prediction accuracy, but their complexity often makes them difficult to interpret. As DL models become more intricate, their explainability diminishes significantly. Nevertheless, analysing the progression patterns of AD necessitates not only models with strong generalisation capabilities but also effective interpretation techniques [23]. Consequently, a compelling need exists for an alternative approach that can bridge the gap between interpretability and repeatability to balance the two conflicting requirements [24].

Building upon this foundation, our objective is to introduce a novel interpretable deep learning radiomics (IDLR) model. The innovative aspect of this model lies in its utilisation of radiomic features to guide and supervise the selection of DL features, thus maintaining the classification and prediction accuracy inherent to DL methods while rendering DL features interpretable. Consequently, the primary objectives of this study are to propose the IDLR model and validate its clinical diagnostic utility in the classification and prediction of the Alzheimer's clinical spectrum.

Materials and methods

Subjects

This multicentre study involved a total of 1962 subjects drawn from three distinct cohorts. All participants had undergone both MRI and [^{18}F]FDG PET scans, along with neuropsychological assessments, including the mini-mental state examination (MMSE).

Cohort 1, from the Alzheimer's disease neuroimaging initiative (ADNI), encompassing subjects from the ADNI-1, ADNI-2, and ADNI-3, was used as the training and validation set for constructing the IDLR model. Cohort 1 comprised 1347 individuals, categorised as follows: 502 healthy controls (HCs), 108 subjects with subjective cognitive decline (SCD), 187 with sMCI, 168 with pMCI, and 382 patients diagnosed with AD.

Cohort 2, from ADNI GO, was reserved as an internal independent test set and included 120 subjects, further divided into 81 individuals with sMCI and 39 with pMCI.

Cohort 3, from two hospitals in China, namely, Huashan Hospital and Xuanwu Hospital, constituted the external independent test set, which included 495 subjects. This cohort was distributed across different diagnostic groups, comprising 223 HCs, 150 subjects with SCD, 41 subjects with MCI, and 81 subjects with AD.

Additionally, a subset of subjects within Cohort 3 also underwent ^{18}F-AV45 amyloid-PET scans, totalling 238 of the 495 individuals in this cohort. The study received approval from the institutional review boards of ADNI, Huashan Hospital, and Xuanwu Hospital, ensuring compliance with ethical standards. Moreover, written informed consent was diligently obtained from all participants or their authorised representatives, underscoring the commitment to ethical research practices [25].

The detailed inclusion process for screening, acquisition protocols, and image preprocessing, including a detailed definition of all three cohorts, are described in Supplementary Materials and Fig. S1.

IDLR workflow

The proposed IDLR model included the following three steps (Fig. 1): (a) feature extraction, (b) feature selection, and (c) classification and prediction. The detailed information is as follows:

Feature extraction

The feature extraction block consists of two components: a radiomics extractor, described in detail in Supplementary Material 1 following the work of Wu et al [26], and a DL extractor, which is responsible for extracting radiomics and DL features from the input images. By comparing the results of the pre-training models, we ultimately selected ResNet50 as the DL extractor. Specifically, to obtain DL features, the feature map was first extracted from the last convolutional layer of the selected convolutional network, and then we converted the maximum value of each feature map into the original value through global max pooling. Afterwards, these extracted features (defined as DL features) were supervised and screened through radiomics features and combined with clinical parameters (sex and age) as input data for classification (defined as IDLR features) [15]. Please refer to the Supplementary Materials for the specific DL extractor

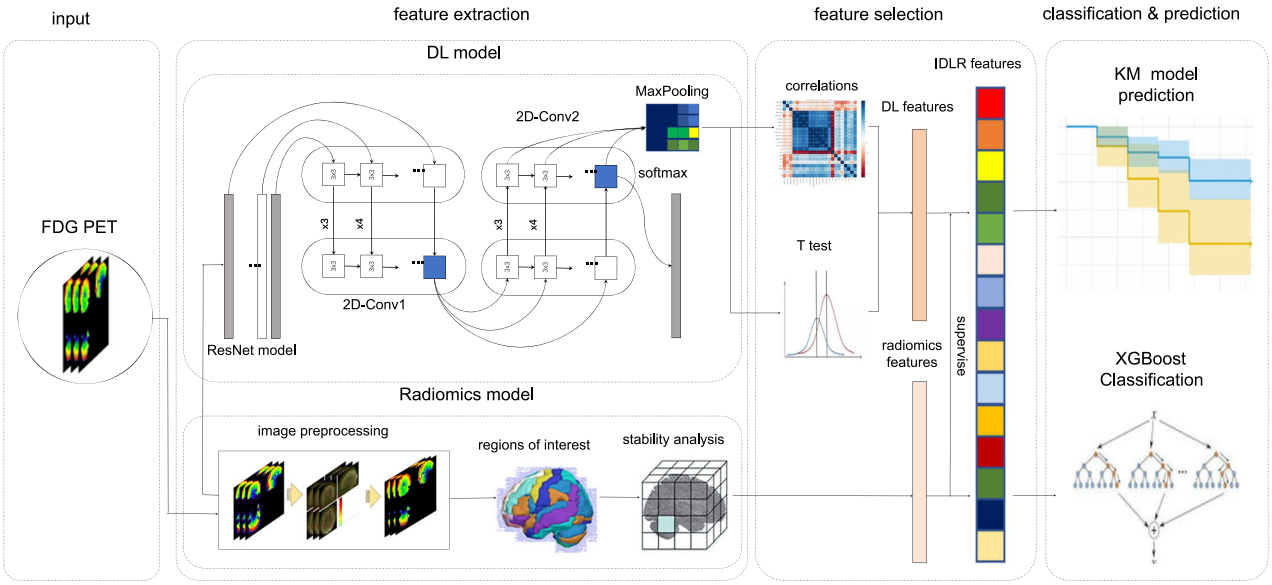

Fig. 1 Workflow of IDLR method

selection, model pretraining and model parameters (Fig. S2 and Table S2).

Feature selection

To address the challenges of high-dimensional features and overfitting, we employed three distinct feature selection steps, utilising both supervised and unsupervised screening methods: (a) feature autocorrelation analysis: we conducted feature autocorrelation analysis to mitigate redundancy within the high-dimensional feature set. For each feature, we calculated the average absolute correlation through pairwise comparisons, defined by the following formula:

$$C_i = \frac{1}{n}\sum_{j=i}^{n} C_{ij} \quad (1)$$

In the pairwise correlation analysis, when the C_i value exceeded 0.8, we systematically removed the feature with the highest average absolute correlation for redundant feature screening. (b) Statistical testing: we utilised a two-sample *t*-test to discern features with statistically significant differences to distinguish different diseases ($p < 0.01$) [27]. (c) Feature supervision: furthermore, we calculated correlation coefficients between DL features and a set of 11 key radiomic features, which were computed based on our prior study [16] and are available in the Supplementary Materials (Table S1).

Classification and prediction

The classification model incorporated retained IDLR features, radiomic features, and clinical information (age, education, sex, and MMSE score) and utilised eXtreme Gradient Boosting (XGBoost) for the task. Objective was set to "multi: softmax" the learning rate was set to 0.025, gamma was set to 0.1, max_depth was set to 6, and lambda was set to 2. We compared the IDLR model with four other models, including radiomic features with clinical information, DL features with clinical information, DL features without clinical information, and DL and radiomic features without clinical information. We performed both tri-classification and pairwise classification simultaneously. Tri-classification utilises the AUC, confusion matrix, recall accuracy curve, and corresponding metrics such as AUC score, recall, F1 score, and precision to evaluate model efficiency. Binary classification was assessed using accuracy, sensitivity, and specificity.

A Cox model was used to test the prediction performance from MCI to AD with the IDLR features. The Kaplan–Meier (KM) curve was used for conversion predictions.

Model explainability

To enhance the interpretability of the DL model's learned representations, we employed two well-known techniques: Shapley additive explanations (SHAP) [28] and a gradient-weighted class activation map (Grad-CAM) [29]. Additional SHAP analysis was conducted on two cohorts to diagnose the AD disease spectrum using the XGBoost classifier to comprehensively comprehend how IDLR characteristics influence decision-making at all stages. The SHAP analysis was carried out in three main areas: (1) identifying the most critical input features that predominantly influence model decisions across all categories, (2) evaluating the influence of these key features on model decisions for the final decision of each class, and (3) revealing the contribution of feature importance to a specific sample. Please refer to the Supplementary Materials for descriptions of the two methods.

Validation of conserved IDLR features

To further investigate the relationships between the retained IDLR features and clinical information across various cognitive stages, we calculated partial correlation coefficients. These coefficients were computed considering features and clinical scales (MMSE scores) within individual subjects in Cohort 3 while simultaneously accounting for the effects of age, sex, and education.

Similar correlation analyses were conducted to assess associations between conserved IDLR features and global standard uptake value ratios (SUVRs). Furthermore, to gain insight into the distribution of these conserved IDLR features across the entire clinical spectrum of AD, we evaluated variations in these features across different groups.

Statistical analysis

We evaluated group differences in clinical characteristics through a range of statistical tests, including two-sample *t*-tests, $\chi 2$ tests, one-sample *t*-tests, or Tukey's tests, as appropriate for the specific analysis. To assess survival disparities between different risk groups, we utilised the log-rank test. In all cases, *p* values were calculated as two-tailed, with statistical significance set at $p < 0.05$.

All statistical analyses were conducted using SPSS 24.0. Cox models were constructed using R (http://www.R-project.org/), and the 'glmnet' and 'survival' packages were used to facilitate the analysis.

Results

Subjects

Table 1 presents the demographic details of the participants from each cohort. As expected, cognitive disease severity, as measured by MMSE scores, exhibited significant differences between the different groups.

Table 1 The clinical characteristics of all cohorts

		Group		Gender, (M/F)	Age, (years)	Education, (year)	MMSE	Conversion time, (months)
Caucasian Cohort	Cohort 1	HC ($n = 502$)		284/218	75.5 ± 5.4	16.2 ± 2.9	29.1 ± 1.0	/
		SCD ($n = 108$)		45/63	72.0 ± 5.9	16.7 ± 2.5	29.0 ± 1.2	/
		MCI	sMCI ($n = 187$)	109/78	72.1 ± 7.5	16.0 ± 2.6	28.0 ± 1.6	/
			pMCI ($n = 168$)	95/73	74.0 ± 7.1	16.0 ± 2.6	26.5 ± 2.2	14.1 ± 8.9
		AD ($n = 382$)		225/157	75.5 ± 7.5	15.0 ± 3.1	21.7 ± 4.0	/
		p value		0.635[a]	< 0.001[c]	0.465[c]	< 0.001[c]	/
	Cohort 2	MCI	sMCI ($n = 81$)	44/37	72.2 ± 7.8	15.6 ± 2.8	28.2 ± 1.4	/
			pMCI ($n = 39$)	21/18	72.1 ± 6.9	16.3 ± 2.7	26.7 ± 1.9	16.4 ± 7.3
		p value		0.96[a]	0.68[b]	0.35[b]	< 0.001[b]	/
Asian Cohort	Cohort 3	HC ($n = 223$)		104/119	58.9 ± 10.5	13.1 ± 3.1	29.1 ± 1.2	/
		SCD ($n = 150$)		28/122	66.1 ± 5.1	12.9 ± 2.8	28.9 ± 0.9	/
		MCI ($n = 41$)		20/21	66.5 ± 7.4	12.6 ± 3.6	24.0 ± 5.4	/
		AD ($n = 81$)		30/51	61.7 ± 10.3	10.1 ± 2.3	19.3 ± 7.2	/
		p value		< 0.001[a]	< 0.001[c]	< 0.001[c]	< 0.001[c]	/

HC healthy control, *SCD* subjective cognitive decline, *MCI* mild cognitive impairment, *AD* Alzheimer's disease, *sMCI* stable mild cognitive impairment, *pMCI* progressive mild cognitive impairment, *MMSE* mini-mental state examination
[a] Chi-squared tests
[b] Two samples *t*-test
[c] ANOVA test

Specifically, significant differences were detected between HCs and patients with AD, between patients with SCD and patients with AD, and between patients with MCI and patients with AD ($p < 0.001$). Furthermore, significant differences were evident between pMCI patients and sMCI patients in terms of cognitive disease severity (MMSE) in both Cohort 1 and Cohort 2 (all $p < 0.001$). In Cohort 3, similar trends were observed, with significant differences in cognitive disease severity (MMSE) between AD patients and HCs ($p < 0.001$) and between MCI patients and HCs ($p < 0.001$). However, it is worth noting that MMSE scores did not significantly differ between SCD patients and HCs ($p = 0.21$). In addition, the conversion time periods were 14.1 ± 8.9 and 16.4 ± 7.3 months in Cohort 1 and Cohort 2, respectively.

Conserved IDLR features

From the last convolution layer of ResNet50, a total of 256 DL features were initially extracted. However, after these features were subjected to feature autocorrelation analysis and two-sample *t*-tests, only 108 features remained. Subsequently, we performed correlation analyses between these 108 DL features and 11 key radiomic features. As a result, 71 IDLR features were identified as conserved features, demonstrating their robust relevance. Finally, these conserved IDLR features, in conjunction with clinical scales and key radiomic features, were input into the classifiers for comprehensive analysis. For detailed descriptions of all the conserved IDLR features, please refer to Table S3 in the Supplementary Materials.

Classification performance of the IDLR model in three cohorts

Figure 2 shows the three-classification efficiency of the IDLR model. IDLR achieved the best performance, with an accuracy of 79.30% [76.23%, 82.42%] in Cohort 1 and 88.87% [84.01%, 92.32%] in Cohort 3.

Table 2 summarises the results of the three classification comparison experiments involving five different methods. Table 3 shows the superior classification and diagnostic performance of the IDLR method in distinguishing between all groups within the three cohorts. Refer to the Supplementary Materials for additional triclassification evaluation metrics for the other four methods (Figs. S3 and S4).

SHAP and Grad-CAM

Figure 3a–h summarises the results from two cohorts that were utilised to diagnose the AD spectrum using SHAP analysis with the XGBoost classifier. Notably, ResNet_latent_26 and ResNet_latent_223 exhibited superior taxonomic contributions in the Asian and Caucasian cohorts, respectively. Additionally, ResNet_latent_37 and ResNet_latent_109, along with two key radiomic features (correlation and ZP), play a crucial role in distinguishing NC categories across various racial groups.

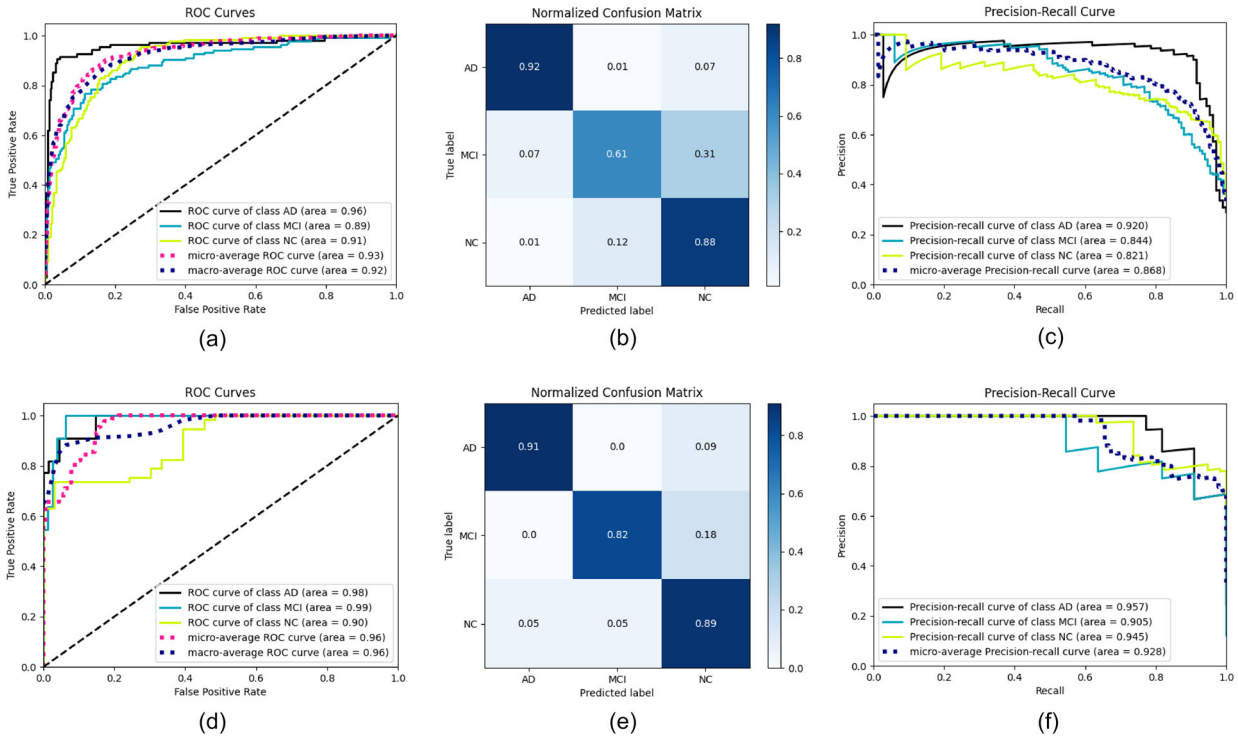

Fig. 2 AUC, confusion matrix, and recall-precision curves. **a**–**c** Three classification performances of the IDLR model in Cohort 1 included AUC, confusion matrix, and recall-precision curves. **d**–**f** Three classification performances of the IDLR model in Cohort 3 included AUC, confusion matrix, and recall-precision curves

Figure 4a–f shows a typical transverse cross-section (typical transaxial slice) of the 40th layer in two-dimensional space. Representative cases from both Caucasian and Asian populations were chosen for visual presentation. These visualisations provided valuable insights (Fig. 4a–d) showing that the brain regions and high-dimensional image features closely aligned with the positions of key radiomic features in the brain when the model distinguished between sMCI and pMCI patients. Figure 4e, f, focusing on the AD group, revealed a heightened emphasis on brain regions well documented for their association with AD. Notably, areas such as the inferior and middle temporal gyrus, as well as the hippocampus, were at the centre of the heatmap. These observations align with relevant studies that have contributed to our understanding of the complex dynamics underlying the transition from MCI to AD [5].

Survival analysis

Figure 5 shows the Kaplan–Meier (KM) survival curves for the high-risk and low-risk groups predicted by all conserved IDLR features. These KM survival curves effectively illustrate the distinction between groups with high and low risks of conversion to AD, as evidenced by a highly significant difference ($p < 0.001$ in the log-rank test). Notably, the KM curve generated based on other methods (radiomics features with clinical information with p value of 0.024; DL features without clinical information with p value of 0.029; and DL and radiomics features without clinical information with p value of 0.021) yielded comparatively poorer results. The results generated by the KM curve based on DL features with the clinical information method were the worst, with a p value of 0.041.

Within the Cox model, ResNet50_Latent_130, representing the zone percentage of the left hippocampus, emerged as the most potent predictor. This feature exhibited robust predictive power, with a hazard ratio (HR) of 1.465 (95% confidence interval, 95% CI: 1.236–1.737, $p < 0.001$).

Validation of conserved IDLR features in Cohort 3

For further validation, the model's performance was examined in Cohort 3 (the external independent test set). Figure S5 provides a comprehensive overview of the differences observed in the 71 conserved IDLR features across the HC, SCD, MCI, and AD groups within Cohort 3.

Table 2 Three-classification performance of each method

Cohort	Method	Accuracy, (%)	AUC score, (%)	F1 score, (%)	Recall, (%)	Precision (%)
Cohort1	Radiomic features with clinical information	76.13 [72.87, 80.06]	88.13 [84.42, 92.17]	76.13 [73.23, 79.34]	76.14 [72.87, 80.52]	76.04 [73.17, 79.18]
	DL features with clinical information	79.12 [74.17, 83.87]	93.27 [90.12, 96.41]	79.13 [75.41, 82.27]	79.01 [75.42, 82.67]	78.92 [75.63, 82.01]
	DL features without clinical information	77.12 [73.51, 80.82]	92.36 [88.24, 95.62]	76.21 [72.89, 80.14]	77.32 [73.24, 80.97]	76.95 [72.35, 81.62]
	DL and radiomic features without clinical information	75.26 [71.74, 78.98]	90.38 [77.13, 93.84]	75.58 [71.64, 78.89]	75.58 [71.23, 78.58]	75.60 [71.09, 78.91]
	IDLR	79.30 [76.23, 82.42]	91.94 [88.64, 95.23]	79.68 [77.32, 81.81]	80.18 [77.21, 83.73]	81.06 [78.04, 84.77]
Cohort3	Radiomic features with clinical information	66.17 [63.12, 70.43]	82.41 [79.26, 86.14]	63.26 [60.04, 67.17]	68.06 [64.12, 71.43]	62.04 [58.17, 66.24]
	DL features with clinical information	82.22 [78.67, 86.28]	88.52 [84.13, 91.89]	75.31 [72.31, 77.77]	70.97 [67.24, 74.52]	82.97 [78.29, 85.41]
	DL features without clinical information	68.89 [64.13, 71.92]	84.04 [81.34, 87.48]	69.70 [66.14, 73.12]	70.32 [66.10, 75.83]	69.77 [66.03, 72.71]
	DL and radiomic features without clinical information	84.45 [81.43, 87.62]	93.35 [90.67, 96.84]	82.13 [79.29, 84.52]	84.13 [81.27, 87.37]	80.52 [77.62, 83.29]
	IDLR	88.87 [84.01, 92.32]	95.48 [92.23, 97.60]	86.07 [83.50, 88.58]	87.40 [84.45, 90.40]	84.89 [81.69, 87.04]

DL deep learning, *IDLR* interpretable deep learning radiomics

Remarkably, among these features, three stand out which exhibited statistically significant differences among the HC, SCD, MCI, and AD groups: ResNet50_Latent_130, ResNet50_Latent_139, and ResNet50_Latent_225.

Figure 6d, g, and j further reveal the distributions of these three pivotal features at various cognitive stages. ResNet50_Latent_225, ResNet50_Latent_130, and ResNet50_Latent_139 exhibited significant correlations with MMSE scores ($r = 0.5151$, 0.4173, and −0.4667, respectively) and global SUVRs (AV45 level) ($r = -0.1818$, −0.2434, and 0.2097, respectively), all with $p < 0.0001$.

Discussion

In this study, we introduced a novel IDLR model utilising [^{18}F]FDG PET images to predict the progression from MCI to AD by assessing its clinical applicability in classification and prediction. Our model leverages the ability of ResNet50 to extract features and, crucially, selects IDLR features with guidance from 11 key radiomic features. This approach not only improves the accuracy of classification and prediction but also ensures the interpretability of DL features.

In this study, we identified three key IDLR features that exhibit close associations with clinical scales and display specific trends within the clinical spectrum of AD. These findings suggest the potential of these IDLR features to serve as valuable imaging markers in clinical practice. It is worth noting that this research encompasses a multicentre approach, allowing us to validate the model's performance in diverse cohorts, including Caucasian and Asian populations. Based on our results, the IDLR model demonstrated the highest classification accuracy when distinguishing between sMCI and pMCI, surpassing traditional radiomic and DL models [16, 30]. Several studies have investigated the use of various machine learning and DL techniques to diagnose the spectrum of AD disease. Jiang et al [16] reported a classification accuracy of 83.8% using an RPM (radiomics) model. Zhao et al [18] used a novel DLR model to differentiate patients with AD and achieved an accuracy of 90.8%. In comparison with prior research, our IDLR model achieved a commendable level of classification accuracy (92.6%). The IDLR model achieves superior performance for the following reasons. First, the IDLR model capitalises on the strengths of two high-dimensional feature extraction methods—radiomics and DL—synergizing their capabilities to enhance overall performance. Second, the model eliminates redundant information by employing key radiomic features as a guiding framework, ensuring that the most relevant and informative data are retained. Finally, by integrating IDLR features, clinical scales, and radiomic features into the classification process, the IDLR model also effectively

Table 3 Binary classification performance of each method

Cohort		Method	Accuracy, (%)	Sensitivity, (%)	Specificity, (%)
Cohort1	HC vs AD	Radiomic features with clinical information	83.41 [80.05, 86.27]	82.61 [79.46, 85.24]	84.72 [81.33, 87.89]
		DL features with clinical information	90.13 [87.22, 93.47]	91.31 [87.96, 94.52]	88.93 [85.88, 92.17]
		DL features without clinical information	80.81 [77.23, 83.48]	83.52 [80.44, 87.17]	77.27 [73.17, 81.22]
		DL and radiomic features without clinical information	82.19 [78.13, 85.67]	84.31 [81.21, 87.63]	80.14 [77.10, 83.91]
		IDLR	92.60 [89.15, 95.41]	93.31 [90.12, 96.22]	92.04 [88.77, 94.14]
	HC vs MCI	Radiomic features with clinical information	70.23 [67.11, 73.46]	73.14 [69.73, 77.42]	71.39 [68.81, 74.88]
		DL features with clinical information	82.28 [78.67, 85.77]	84.15 [81.04, 87.22]	80.41 [77.01, 83.38]
		DL features without clinical information	73.08 [70.24, 76.77]	75.21 [72.23, 78.20]	71.83 [68.23, 74.54]
		DL and radiomic features without clinical information	74.43 [70.81, 78.07]	73.91 [70.15, 77.18]	75.63 [71.23, 79.55]
		IDLR	82.34 [79.24, 85.87]	84.61 [81.12, 88.69]	80.82 [76.61, 83.91]
	MCI vs AD	Radiomic features with clinical information	72.63 [69.12, 75.61]	71.10 [68.23, 74.15]	74.13 [71.42, 78.28]
		DL features with clinical information	79.23 [75.82, 82.88]	82.08 [78.19, 85.92]	77.29 [73.32, 81.23]
		DL features without clinical information	74.08 [71.32, 77.42]	75.21 [71.53, 78.60]	73.42 [70.15, 77.21]
		DL and radiomic features without clinical information	72.23 [68.42, 76.53]	74.81 [70.89, 78.95]	69.78 [66.21, 73.65]
		IDLR	80.51 [76.81, 84.27]	82.23 [78.45, 86.29]	78.07 [75.13, 82.32]
Cohort2	sMCI vs pMCI	Radiomic features with clinical information	71.21 [68.17, 74.52]	66.43 [61.24, 71.22]	74.58 [70.35, 79.19]
		DL features with clinical information	74.21 [70.18, 78.44]	75.46 [71.58, 79.48]	72.97 [68.94, 76.27]
		DL features without clinical information	68.62 [64.12, 72.87]	70.61 [66.53, 74.69]	67.67 [63.19, 72.01]
		DL and radiomic features without clinical information	75.13 [71.19, 78.95]	78.26 [74.15, 83.42]	72.91 [69.12, 75.87]
		IDLR	78.23 [75.01, 81.19]	76.32 [73.29, 80.17]	80.54 [77.25, 83.78]
Cohort3	sMCI vs pMCI	Radiomic features with clinical information	69.13 [66.28, 73.12]	74.27 [70.14, 78.67]	63.91 [59.27, 68.28]
		DL features with clinical information	73.89 [68.99, 77.89]	72.21 [68.10, 76.85]	75.78 [71.23, 78.91]
		DL features without clinical information	67.42 [63.37, 71.42]	68.23 [65.17, 71.86]	65.71 [61.49, 69.27]
		DL and radiomic features without clinical information	70.01 [66.41, 74.10]	71.29 [67.68, 75.52]	68.24 [64.41, 72.22]
		IDLR	76.51 [72.88, 79.60]	77.09 [73.91, 80.99]	75.97 [71.23, 79.05]
	HC vs AD	Radiomic features with clinical information	90.12 [87.24, 93.73]	88.37 [85.29, 91.33]	92.21 [89.51, 95.05]
		DL features with clinical information	89.17 [86.51, 93.18]	85.98 [81.47, 89.63]	92.67 [88.62, 95.41]
		DL features without clinical information	86.21 [82.44, 89.91]	85.27 [82.15, 88.65]	88.08 [85.01, 91.42]
		DL and radiomic features without clinical information	88.28 [84.64, 92.57]	84.67 [80.89, 88.59]	92.31 [88.72, 95.17]
		IDLR	92.17 [89.20, 95.36]	88.41 [85.20, 91.87]	95.88 [93.12, 97.66]
	HC vs MCI	Radiomic features with clinical information	79.21 [76.55, 82.47]	75.07 [71.25, 79.15]	84.33 [80.21, 88.24]
		DL features with clinical information	84.17 [80.66, 87.75]	82.36 [78.89, 85.91]	86.29 [82.11, 90.08]
		DL features without clinical information	77.01 [73.44, 80.59]	73.42 [70.51, 78.62]	81.27 [77.47, 85.12]
		DL and radiomic features without clinical information	80.36 [76.58, 83.38]	77.12 [74.42, 80.18]	83.16 [78.34, 88.62]
		IDLR	87.21 [83.80, 90.88]	85.32 [81.27, 88.67]	89.22 [85.34, 92.74]
	MCI vs AD	Radiomic features with clinical information	78.08 [75.61, 81.18]	74.06 [70.58, 78.82]	81.22 [77.47, 75.25]
		DL features with clinical information	85.14 [81.46, 88.87]	84.67 [81.92, 89.22]	86.77 [83.39, 90.46]
		DL features without clinical information	77.19 [74.12, 80.95]	75.41 [71.11, 78.57]	79.53 [75.62, 83.37]
		DL and radiomic features without clinical information	76.20 [72.51, 80.11]	78.01 [74.49, 81.79]	74.64 [71.23, 78.05]
		IDLR	88.11 [85.07, 91.26]	85.42 [82.14, 88.91]	90.63 [86.59, 93.47]

DL deep learning, *IDLR* interpretable deep learning radiomics

solves the problem of inadequate fitting by increasing interpretability.

Importantly, most of the anatomical brain regions in which the contribution rate conservative features were selected by SHAP analysis in the current study were located in regions consistent with previous studies, such as the hippocampus and parahippocampal gyrus in the temporal cortex, the precuneus and superior marginal gyrus in the parietal cortex, and the medial and accessory cingulate gyrus in the frontal cortex [30]. On the one hand, the medial temporal region, precuneus and cingulate gyrus are the brain regions affected by early

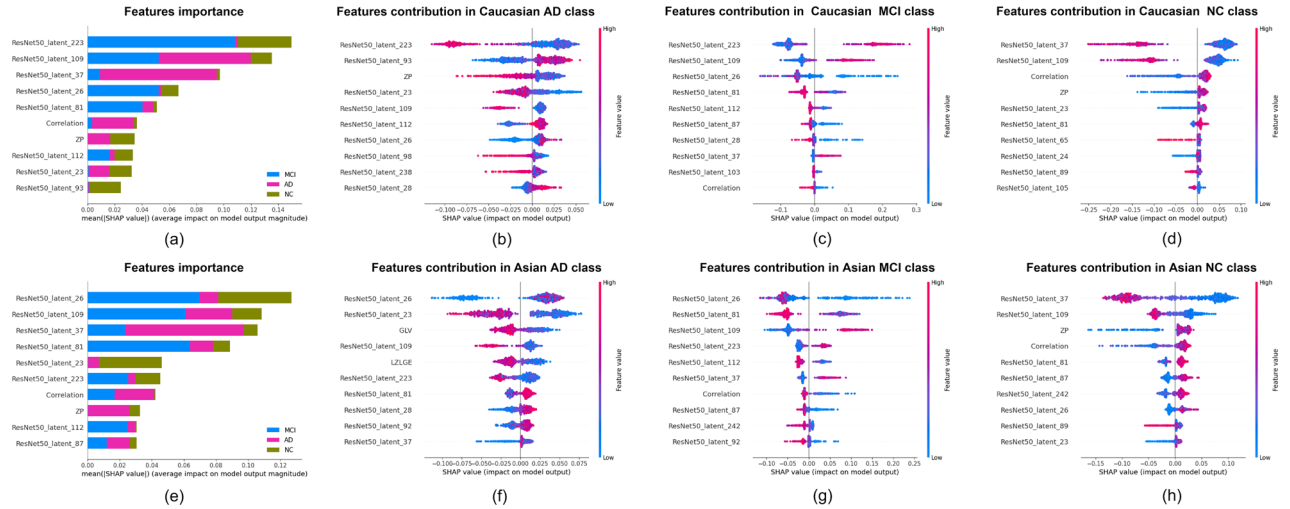

Fig. 3 Combined SHAP results for AD, MCI and NC classifications of XGBoost classifier. **a** The most important input features in Cohort 1 that mainly affect the decision of the XGBoost classifier for AD, MCI, and NC classes. **b** The effect of each feature in Cohort 1 on the final decision of the AD class in the classification. **c** The influence of the characteristics of Cohort 1 on the final decision of the MCI class in the classification. **d** The influence of the characteristics of Cohort 1 on the final decision of class NC in the classification. **e** The most important input features in Cohort 3 primarily affect XGBoost classifiers' decisions for AD, MCI, and NC classes. **f** The effect of each feature in Cohort 3 on the final decision of the AD class in the classification. **g** The influence of the characteristics of Cohort 3 on the final decision of the MCI class in the classification. **h** The influence of the characteristics of Cohort 3 on the final decision of class NC in the classification

AD-related pathological protein (Aβ and hyperphosphorylated tau) deposition and are also the areas affected by early atrophy, reduced thickness or reduced glucose metabolism [5, 31–33]. On the other hand, this finding also shows that IDLR features may comprehensively capture various clinically valuable information from multiple brain regions, thereby mitigating the overreliance on traditional radiomic features extracted solely from specific ROI brain regions. IDLR features solve the shortcomings of traditional radiomic methods in terms of category imbalance and parameter heterogeneity.

In particular, five Cox models were constructed based on feature sources in our study, and the IDLR model performed better than the other models, which may be related to our high-dimensional imaging features. According to the survival analysis, the results of the KM curve analysis based on the DL features and clinical information methods were the worst, which may be related to the low sensitivity of the clinical data and the high subjectivity of the MMSE scores. Longitudinal analysis further demonstrated the great potential of IDLR features in predicting MCI progression.

Furthermore, our study conducted a cross-sectional evaluation of the alterations in IDLR features across various cognitive stages. Notably, three features displayed significant differences among the HC, SCD, MCI, and AD groups. This finding suggests that these features may harbour unique information pertaining to microscopic-level changes that precede macroscopic-level alterations, aligning with results from prior research [34]. Moreover, our correlation analyses revealed significant associations between these features and MMSE scores, as well as AV45 levels, indicating that the key IDLR features we found can be used as pathological markers to characterise cognitive decline in the AD disease spectrum. This underscores the enduring relevance of these features, as they continue to exhibit a substantial relationship with conventional assessment scales. This further bolsters their role as predictive indicators of disease progression.

Finally, we investigated the interpretability of the model and validated the stability of the IDLR model. Intriguingly, the regions highlighted in warm colours within the Grad-CAM visual images were consistent with the learning regions identified through SHAP-based input image interpretation. These regions are recognised as significant factors influencing the conversion of MCI to AD, thereby offering valuable insights into potential pathological mechanisms within the AD continuum. In essence, these two methods of explaining a DL model mutually reinforce each other, forming the foundation for the model's output judgement results [35]. After interpretability is achieved, DL features can also be used to explore more effective and deeper imaging biomarkers, promoting the treatment of related targets. Our SHAP feature importance analysis shows that the IDLR model has highly consistent feature contributions in the analysis of multicentre cohorts.

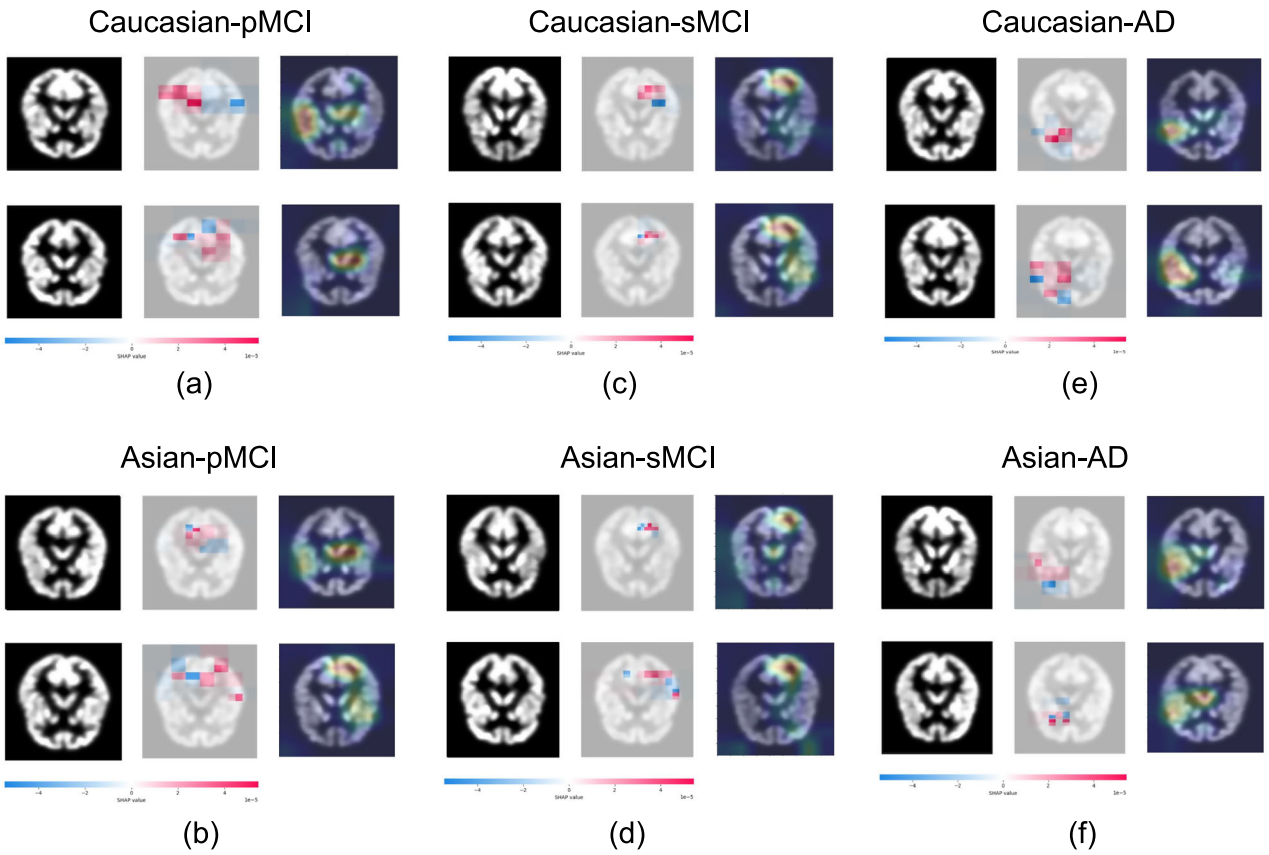

Fig. 4 SHAP and Grad-CAM. **a**–**f** Shows heat mapping of our IDLR model. Visual mapping of SHAP and Grad-CAM features in randomly selected cohorts of different ethnic groups (Caucasian: (**a**, **c**, and **e**) and Asian: (**b**, **d**, and **f**)) and populations with different disease status (pMCI: **a**, **b**, sMCI: **c**, **d** and AD: **e**, **f**). For each combination, the corresponding two-layer slices are randomly selected, and the visualisation of the original slice, SHAP and Grad-CAM in the random two-layer slices is shown from left to right

This finding corroborates the reliability and consistency of our model's interpretability and demonstrates its ability to align with established research in the field.

Nonetheless, it is essential to acknowledge the limitations of this study. First, the IDLR method is primarily based on a single modality, [^{18}F]FDG PET. Exploring whether incorporating characteristics from other modalities, such as MRI, or combining multiple modalities can enhance the classification accuracy of the method remains a potential avenue for future research. Additional neuropsychological tests for different cognitive domains were performed only in some subjects. Whether the addition of other clinical scales can improve the accuracy of the IDLR method and whether a model with higher interpretability after the fusion of features can maintain high performance necessitate further research in the future. Second, although we employed two independent cohorts as test sets, each cohort was subjected to distinct tasks. Therefore, additional tests are warranted to thoroughly assess the reproducibility and robustness of the model across different datasets and tasks. This will contribute to a more comprehensive evaluation of the generalizability and applicability of the model. Third, the newly proposed biological definition of AD was not fully considered in the current study. The biggest pitfall is the adaptation of a non-specific tracer for dementia, which can be supported by poor prediction accuracy. Nowadays, a specific technique like Tau-PET is more predictive. We plan to validate our model with other AD imaging markers, such as tau-PET or amyloid-PET images. Fourth, direct comparisons between the IDLR method and visual assessment method were not performed in this study. The clinical practical value of our model needs to be verified in the future.

Conclusion

Currently, cooperation between AI and human intelligence has become necessary to understand complex diseases and make accurate diagnoses to provide the best possible treatment [22]. In this comprehensive multicentre study,

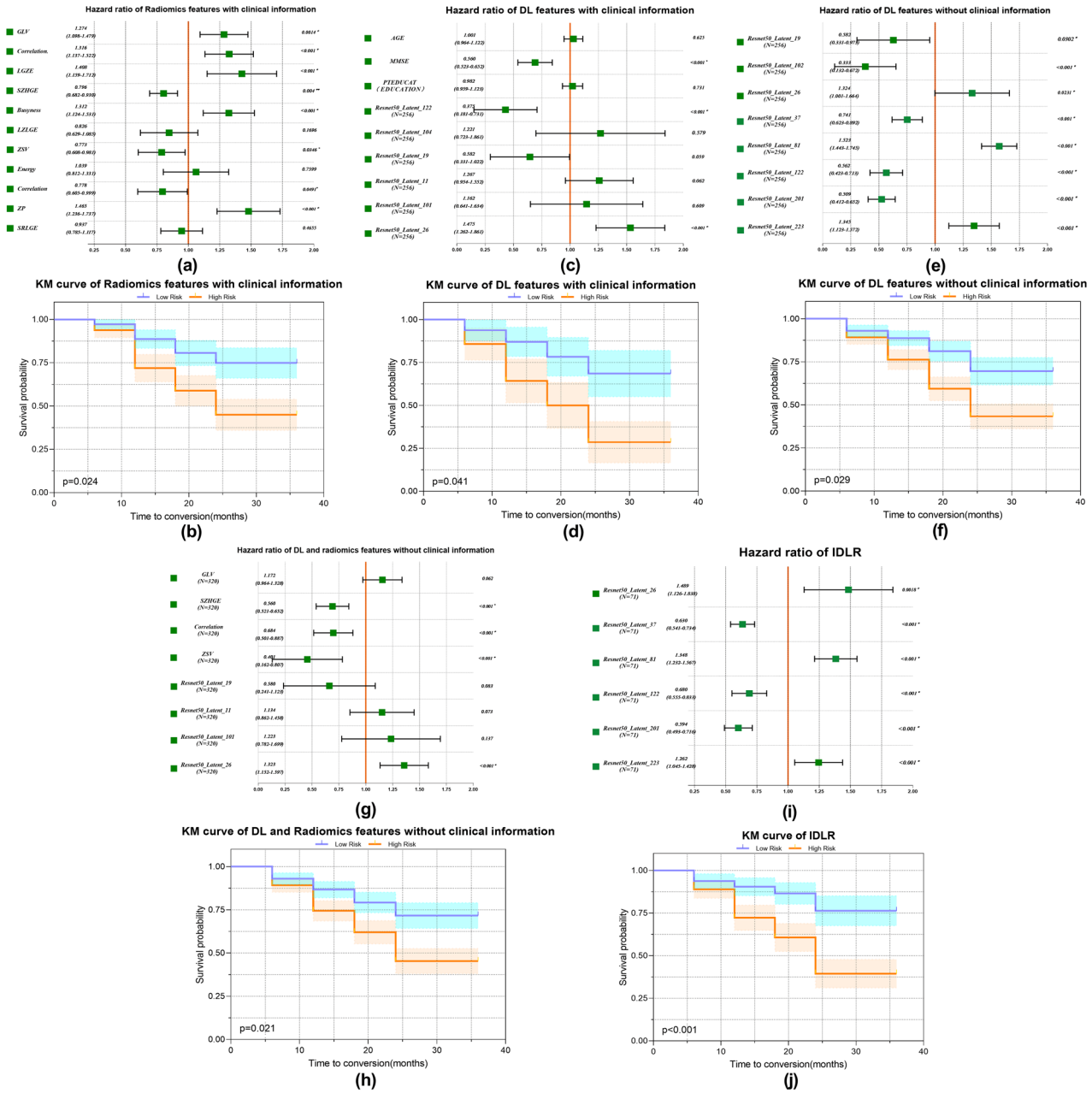

Fig. 5 COX model with hazard ratio (HR) and KM survival curves. **a** HR based on radiomics features with clinical information method. **b** Risk stratification based on radiomics features with clinical information method (log-rank test, $p = 0.024$). **c** HR based on DL features with clinical information method. **d** Risk stratification based on DL features with clinical information method (log-rank test, $p = 0.041$). **e** HR based on DL features without clinical information method. **f** Risk stratification based on DL features without clinical information method (log-rank test, $p = 0.029$). **g** R based on DL and radiomics features without clinical information method. **h** Risk stratification based on DL and radiomics features without clinical information method (log-rank test, $p = 0.021$). **i** HR based on the IDLR model. **j** Risk stratification based on IDLR (log-rank test, $p < 0.001$)

we introduced an innovative IDLR-based predictive modelling approach using [^{18}F]FDG PET images. Our findings showcased the superior predictive capabilities of IDLR features in accurately forecasting the entire AD disease spectrum and forecasting the progression from MCI to AD, highlighting their robustness and reliability in aiding disease stratification and management. Furthermore, we revealed significant associations between the identified IDLR features and cognitive dysfunction, highlighting their capacity to capture distinctions across various cognitive

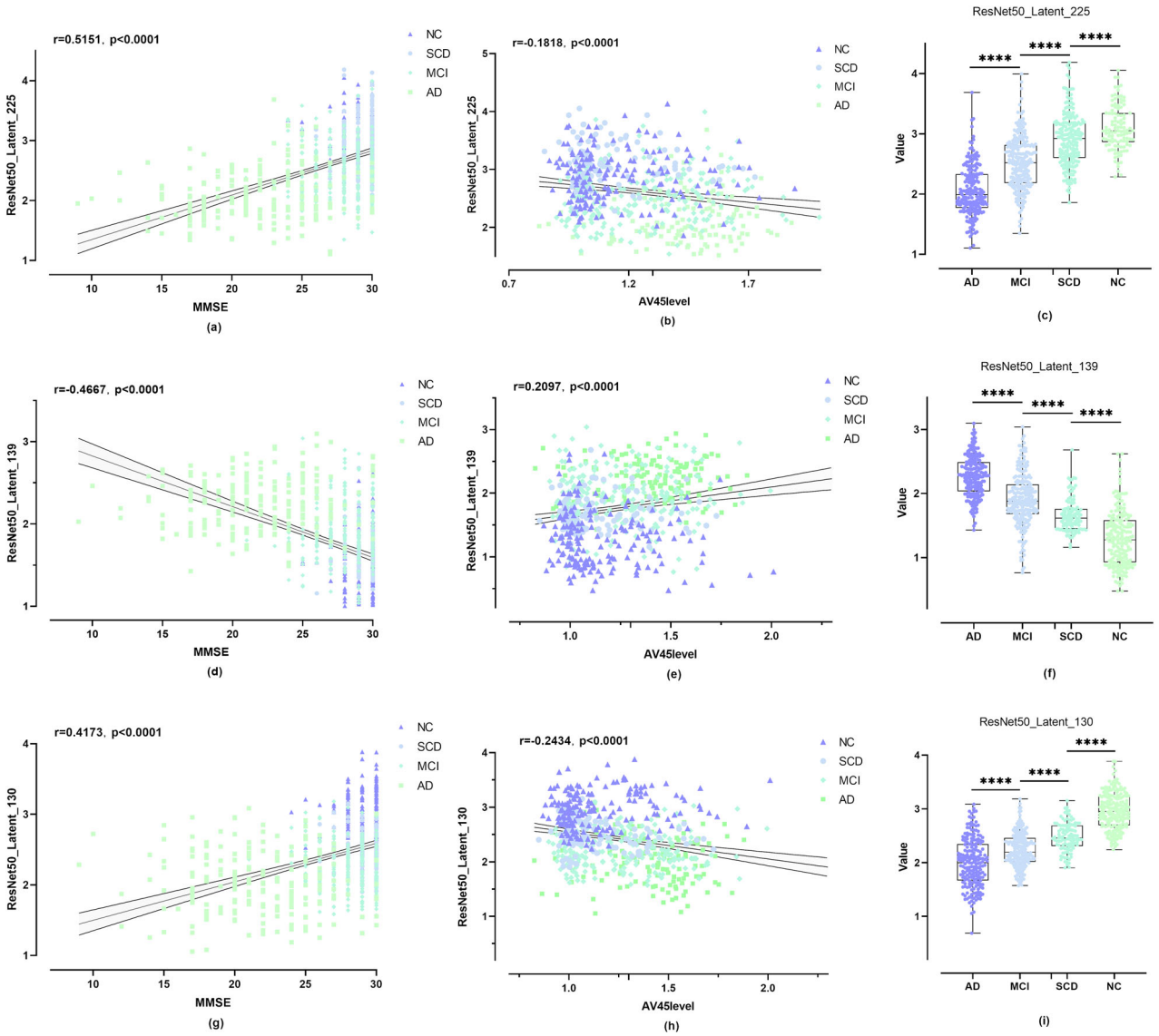

Fig. 6 **a**–**i** Shows group differences of a subset of IDLR features at different cognitive stages. The *p* values were derived using the ANOVA test and two-sample *t*-test. Note: the *p* value correlation between any two groups was less than 0.05

stages. The IDLR model can be explained by the consistency of the anatomical position and is explained by anatomy and physiology. These preliminary assessments underscore the promising potential of the IDLR method as a valuable clinical tool for enhancing MCI conversion prediction and aiding in clinical decision-making.

Abbreviations

AD	Alzheimer's disease
ADNI	Alzheimer's disease neuroimaging initiative
DL	Deep learning
FDG	Fluorodeoxyglucose
Grad-CAM	Gradient-weighted class activation map
HC	Health controls
HR	Hazard ratio
IDLR	Interpretable deep learning radiomics
KM	Kaplan–Meier
MCI	Mild cognitive impairment
MMSE	Mini-mental state examination
MRI	Magnetic resonance imaging
PET	Positron emission tomography
ROI	Regions of interest
SCD	Subjective cognitive decline
SHAP	Shapley additive explanations
SUVR	Standard uptake value ratio
XGBoost	eXtreme gradient boosting

Supplementary information

The online version contains supplementary material available at https://doi.org/10.1007/s00330-024-11158-9.

Author contributions

J.J.: conceptualisation, formal analysis, writing—original draft, writing—review and editing, validation, and funding acquisition. C.L.: conceptualisation, formal analysis, software, and writing an original draft. J.L.: formal analysis, methodology, and writing an original draft. J.S.: formal analysis and methodology. X.S.: data curation. J.Y.: data curation. L.W.: visualisation. C.Z.: conceptualisation, writing—review and editing, validation, and funding acquisition. K.S.: conceptualisation, writing—review and editing, validation, and funding acquisition. Data for this research were obtained from the ADNI database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and provided data but did not participate in the analysis or writing of this report. A complete list of ADNI investigators can be found at: https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Funding

This research was supported by grants from the National Natural Science Foundation of China (no. 62376150), and the Science and Technology Innovation 2030—major projects (no. 2022ZD021606). National Natural Science Foundation of China (82020108013), Shanghai Industrial Collaborative Innovation Project (no. XTCX-KJ-2023-37) and Discipline Construction of Pudong New Area Health Commission (PWZxk2022-03). Data collection and sharing for this project were funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI; National Institutes of Health Grant U01 AG024904) and DODADNI (Department of Defence, award number W81XWH-12–2-0012). ADNI is funded by the National Institute of Aging and the National Institute of Biomedical Imaging and Bioengineering and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai, Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F.Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson &Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer, Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada.

Data availability

The datasets generated during and/or analysed during the current study are available in the ADNI database (adni.loni.usc.edu).

Compliance with ethical standards

Guarantor

The scientific guarantor of this publication is J.J.

Conflict of interest

The authors declare no potential conflicts of interest relevant to this article.

Statistics and biometry

Fan Yang and J.S. kindly provided statistical advice for this manuscript.

Informed consent

Written informed consent was obtained from all subjects (patients) in this study.

Ethical approval

The research involved human participants. Data used in this project was approved by institutional review boards of ADNI and written informed consent was obtained from all participants or authorised representatives.

Study subjects or cohorts overlap

Some study subjects or cohorts have been previously reported by Jiang et al [16].

Methodology

- Prospective
- Experimental
- Multicentre study

Author details

[1]Institute of Biomedical Engineering, School of Life Sciences, Shanghai University, Shanghai, China. [2]Department of Nuclear Medicine & PET Center, Huashan Hospital, Fudan University, Shanghai, China. [3]School of Communication and Information Engineering, Shanghai University, Shanghai, China. [4]Human Phenome Institute, Fudan University, Shanghai, China. [5]Department of Nuclear Medicine, University Hospital Bern, Bern, Switzerland.

Received: 4 January 2024 Revised: 12 August 2024 Accepted: 17 September 2024
Published online: 31 October 2024

References

1. Huang K, Lin Y, Yang L et al (2020) A multipredictor model to predict the conversion of mild cognitive impairment to Alzheimer's disease by using a predictive nomogram. Neuropsychopharmacology 45:358–366. https://doi.org/10.1038/s41386-019-0551-0
2. Fan D-Y, Wang Y-J (2020) Early intervention in Alzheimer's disease: How early is early enough? Neurosci Bull 36:195–197. https://doi.org/10.1007/s12264-019-00429-x
3. Schneider JA, Arvanitakis Z, Leurgans SE, Bennett DA (2009) The neuropathology of probable Alzheimer disease and mild cognitive impairment. Ann Neurol 66:200–208. https://doi.org/10.1002/ana.21706
4. Petersen RC, Roberts RO, Knopman DS et al (2009) Mild cognitive impairment: ten years later. Arch Neurol 66:1447–1455. https://doi.org/10.1001/archneurol.2009.266
5. Yang L, Yan Y, Wang Y et al (2018) Gradual disturbances of the amplitude of low-frequency fluctuations (ALFF) and fFractional ALFF in Alzheimer spectrum. Front Neurosci 12:975. https://doi.org/10.3389/fnins.2018.00975
6. Jack CR, Bennett DA, Blennow K et al (2018) NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimers Dement 14:535–562. https://doi.org/10.1016/j.jalz.2018.02.018
7. Blazhenets G, Ma Y, Sörensen A et al (2019) Principal components analysis of brain metabolism predicts development of Alzheimer dementia. J Nucl Med 60:837–843. https://doi.org/10.2967/jnumed.118.219097
8. Chaddad A, Desrosiers C, Hassan L, Tanougast C (2017) Hippocampus and amygdala radiomic biomarkers for the study of autism spectrum disorder. BMC Neurosci 18:52. https://doi.org/10.1186/s12868-017-0373-0
9. Dubois B, Feldman HH, Jacova C et al (2014) Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. Lancet Neurol 13:614–629. https://doi.org/10.1016/S1474-4422(14)70090-0
10. Tian M, Zuo C, Civelek AC et al (2023) International nuclear medicine consensus on the clinical use of amyloid positron emission tomography in Alzheimer's disease. Phenomics 3:375–389. https://doi.org/10.1007/s43657-022-00068-9
11. Lan L, Feng K, Wu Y et al (2023) Phenomic imaging. Phenomics 3:597–612. https://doi.org/10.1007/s43657-023-00128-8
12. Tian M, Civelek AC, Carrio I et al (2022) International consensus on the use of tau PET imaging agent 18F-flortaucipir in Alzheimer's disease. Eur J Nucl Med Mol Imaging 49:895–904. https://doi.org/10.1007/s00259-021-05673-w
13. Nobili F, Arbizu J, Bouwman F et al (2018) European Association of Nuclear Medicine and European Academy of Neurology recommendations for the use of brain 18F-fluorodeoxyglucose positron emission tomography in neurodegenerative cognitive impairment and dementia: Delphi consensus. Eur J Neurol 25:1201–1217. https://doi.org/10.1111/ene.13728

14. Boccardi M, Festari C, Altomare D et al (2018) Assessing FDG-PET diagnostic accuracy studies to develop recommendations for clinical use in dementia. Eur J Nucl Med Mol Imaging 45:1470–1486. https://doi.org/10.1007/s00259-018-4024-1
15. Arbizu J, Festari C, Altomare D et al (2018) Clinical utility of FDG-PET for the clinical diagnosis in MCI. Eur J Nucl Med Mol Imaging 45:1497–1508. https://doi.org/10.1007/s00259-018-4039-7
16. Jiang J, Wang M, Alberts I et al (2022) Using radiomics-based modelling to predict individual progression from mild cognitive impairment to Alzheimer's disease. Eur J Nucl Med Mol Imaging 49:2163–2173. https://doi.org/10.1007/s00259-022-05687-y
17. Li Y, Jiang J, Lu J et al (2019) Radiomics: a novel feature extraction method for brain neuron degeneration disease using 18F-FDG PET imaging and its implementation for Alzheimer's disease and mild cognitive impairment. Ther Adv Neurol Disord 12:1756286419838682. https://doi.org/10.1177/1756286419838682
18. Zhao Y, Zhang J, Chen Y, Jiang J (2022) A novel deep learning radiomics model to discriminate AD, MCI and NC: an exploratory study based on tau PET scans from ADNI. Brain Sci 12:1067. https://doi.org/10.3390/brainsci12081067
19. Jiang J, Zhang J, Li C et al (2022) Development of a machine learning model to discriminate mild cognitive impairment subjects from normal controls in community screening. Brain Sci 12:1149. https://doi.org/10.3390/brainsci12091149
20. De Santi LA, Pasini E, Santarelli MF et al (2023) An explainable convolutional neural network for the early diagnosis of Alzheimer's disease from 18F-FDG PET. J Digit Imaging 36:189–203. https://doi.org/10.1007/s10278-022-00719-3
21. Sun X, Ge J, Li L et al (2022) Use of deep learning-based radiomics to differentiate Parkinson's disease patients from normal controls: a study based on [18F]FDG PET imaging. Eur Radiol 32:8008–8018. https://doi.org/10.1007/s00330-022-08799-z
22. Sorantin E, Grasser MG, Hemmelmayr A et al (2022) The augmented radiologist: artificial intelligence in the practice of radiology. Pediatr Radiol 52:2074–2086. https://doi.org/10.1007/s00247-021-05177-7
23. Holzinger A, Haibe-Kains B, Jurisica I (2019) Why imaging data alone is not enough: AI-based integration of imaging, omics, and clinical data. Eur J Nucl Med Mol Imaging 46:2722–2730. https://doi.org/10.1007/s00259-019-04382-9
24. Pang T, Wong JHD, Ng WL, Chan CS (2020) Deep learning radiomics in breast cancer with different modalities: overview and future. Expert Syst Appl 158:113501
25. Tan M-S, Yang Y-X, Wang H-F et al (2020) PET amyloid and tau status are differently affected by patient features. J Alzheimers Dis 78:1129–1136. https://doi.org/10.3233/JAD-200124
26. Wu Y, Jiang J-H, Chen L et al (2019) Use of radiomic features and support vector machine to distinguish Parkinson's disease cases from normal controls. Ann Transl Med 7:773. https://doi.org/10.21037/atm.2019.11.26
27. Nie D, Lu J, Zhang H et al (2019) Multi-channel 3D deep feature learning for survival time prediction of brain tumor patients using multi-modal neuroimages. Sci Rep 9:1103. https://doi.org/10.1038/s41598-018-37387-9
28. Aldughayfiq B, Ashfaq F, Jhanjhi NZ, Humayun M (2023) Explainable AI for retinoblastoma diagnosis: interpreting deep learning models with LIME and SHAP. Diagnostics 13:1932. https://doi.org/10.3390/diagnostics13111932
29. Liu S-C, Lai J, Huang J-Y et al (2021) Predicting microvascular invasion in hepatocellular carcinoma: a deep learning model validated across hospitals. Cancer Imaging 21:56. https://doi.org/10.1186/s40644-021-00425-3
30. Babu G, Sundaram S, Mahanand BS (2014) A novel PBL-McRBFN-RFE approach for identification of critical brain regions responsible for Parkinson's disease. Expert Syst Appl 41:478–488. https://doi.org/10.1016/j.eswa.2013.07.073
31. Risacher SL, Saykin AJ (2019) Neuroimaging in aging and neurologic diseases. Handb Clin Neurol 167:191–227. https://doi.org/10.1016/B978-0-12-804766-8.00012-1
32. Pini L, Pievani M, Bocchetta M et al (2016) Brain atrophy in Alzheimer's disease and aging. Ageing Res Rev 30:25–48. https://doi.org/10.1016/j.arr.2016.01.002
33. Lotan E, Friedman KP, Davidson T, Shepherd TM (2020) Brain 18F-FDG-PET: utility in the diagnosis of dementia and epilepsy. Isr Med Assoc J 22:178–184
34. Somayah A, Maha T, Asim A et al (2021) Machine learning and deep learning methods that use omics data for metastasis prediction. Comput Struct Biotechnol. https://pubmed.ncbi.nlm.nih.gov/34589181/. Accessed 6 Mar 2024
35. Tang X, Zhang J, He Y et al (2023) Explainable multi-task learning for multi-modality biological data analysis. Nat Commun 14:2546. https://doi.org/10.1038/s41467-023-37477-x

Publisher's Note