



Optimal PET-based radiomic signature construction based on the cross-combination method for predicting the survival of patients with diffuse large B-cell lymphoma

Chong Jiang¹ · Ang Li² · Yue Teng¹ · Xiangjun Huang² · Chongyang Ding³ · Jianxin Chen² · Jingyan Xu⁴ · Zhengyang Zhou¹

Received: 28 August 2021 / Accepted: 1 February 2022 / Published online: 11 February 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Purpose To develop and externally validate models incorporating a PET radiomics signature (R-signature) obtained by the cross-combination method for predicting the survival of patients with diffuse large B-cell lymphoma (DLBCL).

Methods A total of 383 patients with DLBCL from two medical centres between 2011 and 2019 were included. The cross-combination method was used on three types of PET radiomics features from the training cohort to generate 49 feature selection-classification candidates based on 7 different machine learning models. The R-signature was then built by selecting the optimal candidates based on their progression-free survival (PFS) and overall survival (OS). Cox regression analysis was used to develop the survival prediction models. The calibration, discrimination, and clinical utility of the models were assessed and externally validated.

Results The R-signatures determined by 12 and 31 radiomics features were significantly associated with PFS and OS, respectively ($P < 0.05$). The combined models that incorporated R-signatures, metabolic metrics, and clinical risk factors exhibited significant prognostic superiority over the clinical models, PET-based models, and the National Comprehensive Cancer Network International Prognostic Index in terms of both PFS (C-index: 0.801 vs. 0.732 vs. 0.785 vs. 0.720, respectively) and OS (C-index: 0.807 vs. 0.740 vs. 0.773 vs. 0.726, respectively). For external validation, the C-indices were 0.758 vs. 0.621 vs. 0.732 vs. 0.673 and 0.794 vs. 0.696 vs. 0.781 vs. 0.708 in the PFS and OS analyses, respectively. The calibration curves showed good consistency, and the decision curve analysis supported the clinical utility of the combined model.

Conclusion The R-signature could be used as a survival predictor for DLBCL, and its combination with clinical factors may allow for accurate risk stratification.

Keywords [18F]-FDG PET/CT · Diffuse large B-cell lymphoma · Prognosis · Machine learning · Radiomics

Chong Jiang and Ang Li are co-first authors. They contributed equally to the work.

This article is part of the Topical Collection on Hematology

✉ Jianxin Chen
chenjx@njupt.edu.cn

✉ Jingyan Xu
xjy1967@sina.com

✉ Zhengyang Zhou
zyzhou@nju.edu.cn

¹ Department of Nuclear Medicine, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, Nanjing 210000, China

² The Key Laboratory of Broadband Wireless Communication and Sensor Network Technology (Ministry of Education), Nanjing University of Posts and Telecommunications, Nanjing, China

³ Department of Nuclear Medicine, The First Affiliated Hospital of Nanjing Medical University, Jiangsu Province Hospital, Nanjing, China

⁴ Department of Hematology, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, Nanjing 210000, China

Introduction

Diffuse large B-cell lymphoma (DLBCL) is the most common type of lymphoma [1]. As a remarkably heterogeneous disease, DLBCL is known for its considerable variation in clinical behaviour, response to therapy, and prognosis [2]. Chemoimmunotherapy with R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone) therapy is now used as the standard therapy for DLBCL [3]. However, 30–50% of patients ultimately relapse and die as a result of the disease [4, 5]. Therefore, an accurate identification of patients who will not respond well to standard regimens at diagnosis is needed to guide new therapeutic strategies. The International Prognostic Index (IPI) and its successor (NCCN-IPI) are prognostic scoring systems that predict survival. However, it is difficult to accurately identify DLBCL patients who are likely to relapse, especially those with a poor long-term prognosis, to individualize therapy [6–8].

Volume metabolic parameters derived from [18F]-FDG PET/CT, such as metabolic tumour volume (MTV) and total lesion glycolysis (TLG), are powerful indices for assessing tumour burden and risk stratification in DLBCL patients [9–13]. However, these metabolic parameters are limited in their ability to reflect tumour heterogeneity, which ultimately limits their contribution to predicting cancer progression, therapeutic resistance, and survival outcomes [14, 15]. Machine learning advancements have enabled clinicians to use radiomics to extract implicit features from images to assess intratumoural biological heterogeneity and assist in clinical outcome prediction [16, 17]. Several studies [18–23] have explored the prognostic value of PET/CT radiomics in lymphoma, including Hodgkin lymphoma (HL), nasal-type extranodal natural killer/T cell lymphoma (ENKTL), mantle cell lymphoma (MCL), and DLBCL. In general, these studies used one machine learning model to select and develop radiomic signatures, e.g. the multilayer perceptron neural network combined with logistic regression analysis [19], the least absolute shrinkage and selection operator regression (LASSO) model [20, 22], and the logistic regression model [23]. However, there is no model that works well for all kinds of data. To address this limitation, cross-combinations of classification and selection methods based on radiomics features have recently appeared and have proven effective in improving classification and diagnostic performance [24, 25]. However, it is still unknown whether these methods can be applied to radiomic images for DLBCL. In this paper, we explored the PET radiomics features using a cross-combination of machine learning models to assess risk in DLBCL patients. Furthermore, we established and externally validated models for predicting survival outcome in routine clinical practice, allowing for its clinical translation.

Materials and methods

Patient data collection

A total of 273 patients from Nanjing Drum Tower Hospital, the Affiliated Hospital of Nanjing University Medical School, were included as the training cohort. To examine the general applicability of nomograms, an independent external validation cohort of 110 patients from Jiangsu Province Hospital, the First Affiliated Hospital of Nanjing Medical University, was included. Ethical approval was obtained for this retrospective analysis, and the requirement for written consent from patients was waived. The workflow of patient selection is shown in Fig. 1. The inclusion criteria were as follows: (I) pathologically confirmed DLBCL, (II) pretreatment [18F]-FDG PET/CT scan was performed, and (III) treatment with R-CHOP-like regimens with curative intent was carried out. Patients with incomplete medical data, those lost to follow-up, and those with a previous history of cancer were excluded. Clinical data (sex, age, B symptoms, Eastern Cooperative Oncology Group performance status (ECOG PS), NCCN-IPI index, lactate dehydrogenase (LDH) level, Ann Arbor stage, and immunohistochemistry result) were obtained from medical records. Participants were divided into two groups: those with a low/low-intermediate risk (NCCN-IPI of <4) and those with a high-intermediate/high risk (NCCN-IPI of ≥ 4) [8]. The follow-up data were obtained through electronic medical records and telephone interviews. Progression-free survival (PFS) and overall survival (OS) were chosen as endpoints for evaluating the prognosis of DLBCL patients. PFS was calculated as the time between diagnosis and the first relapse, progression, death from any cause, or last follow-up. OS was calculated as the time between diagnosis and death from any cause or last follow-up.

PET/CT Scanning Protocol

All patients underwent PET/CT scans with one of the following systems: Gemini GXL (Philips Corp, Netherlands) for the training cohort and Biograph 16 PET/CT scanner (Siemens Health care, Erlangen, Germany) for the external validation cohort, based on standard clinical scanning protocols. All patients fasted for at least 6 hours before scans, resulting in blood glucose levels under 8.7 mmol/L. After 6 hours of fasting (no oral or intravenous fluids containing sugar or dextrose), 185–370 MBq of [18F]-FDG (5.18 MBq/kg) was administered intravenously. Each patient was weighed to determine the standardized uptake value (SUV) prior to each scan. Whole-body PET/CT scans (from the base of the skull to the upper thigh) were performed 60 minutes after the radiopharmaceutical injection. Emission

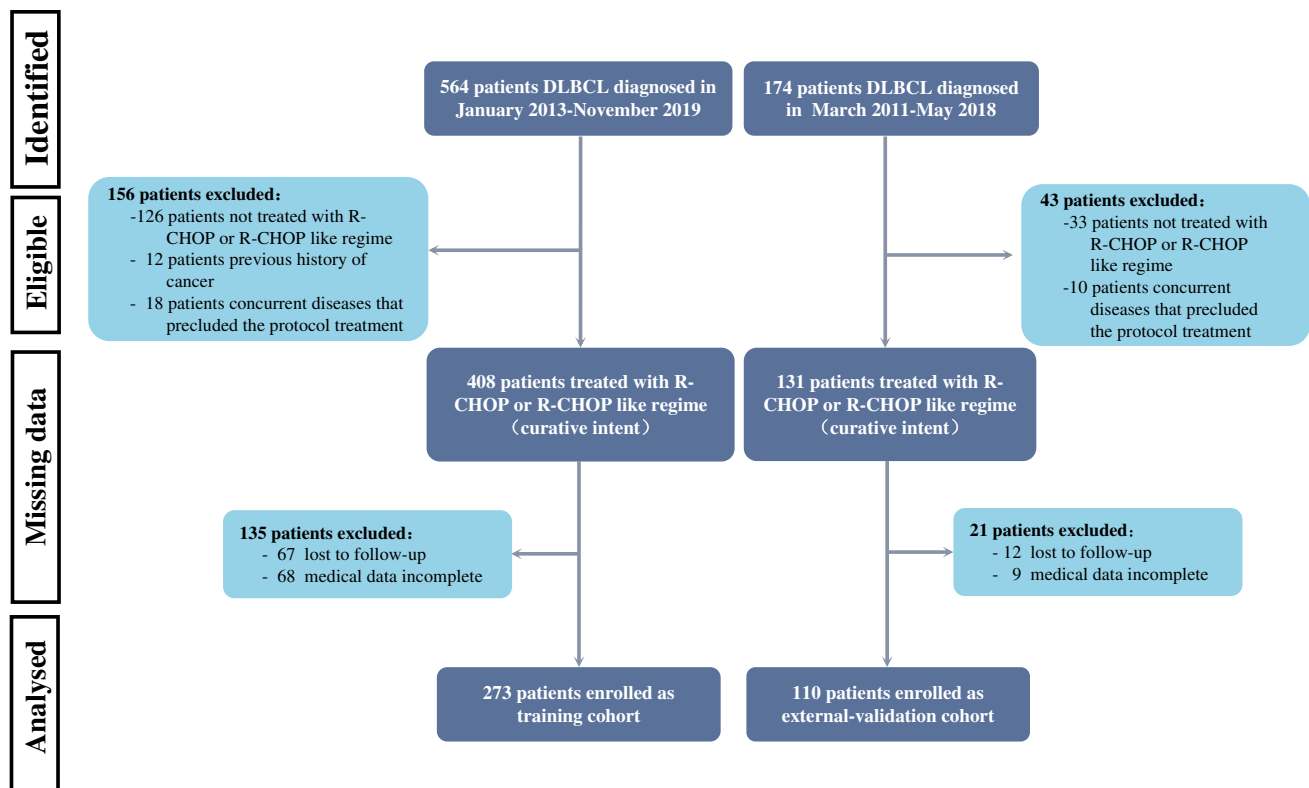


Fig. 1 Flow chart of participant selection

data were acquired for 2 minutes in each bed position. CT acquisition data were used for attenuation correction, and corrected PET images were reconstructed using ordered-subset expectation maximization (OSEM). Images were reconstructed with standard $4 \times 4 \times 4 \text{ mm}^3$ voxels using 3 iterations of 33 subsets.

VOI drawing and radiomics processing

PET images were uploaded to 3D Slicer software (version 4.8.0; <http://www.slicer.org>). The Grow-Cut algorithm in 3D Slicer was used to semiautomatically generate the volume of interest (VOI). Two physicians manually adjusted the VOI to ensure that the measurement was reliable. If there was a discrepancy, the VOIs were reviewed and determined by a senior nuclear medical scientist. The 3 largest lesions were chosen as the targets for radiomic feature extraction. For the three largest VOIs, we summed the radiomic size and shape features, while all other features were averaged for each VOI [26]. Feature extraction was performed based on PET images by the built-in open-source PyRadiomics package (<http://www.radiomics.io/pyradiomics.html>) in 3D Slicer. Before feature extraction, we preprocessed the PET images, including resampling all images to a voxel size of $1 \times 1 \times 1 \text{ mm}^3$ using bilinear interpolation and discretization with a

fixed bin count of 128, as recommended by [27]. To obtain the comprehensive radiomics features, we converted the original PET images into wavelet and Laplacian of Gaussian (LoG) images with wavelet transforms and LoG transforms, respectively. Then, three types of radiomics features were extracted from the original, wavelet and LoG images: (I) first-order statistics, (II) shape features and (III) texture features. The full list of characteristics is provided in Supplementary Table 1. The workflow of our radiomic analysis is shown in Fig. 2.

In addition, the conventional metabolic parameters, including the SUVmax, MTV, and TLG, were calculated by volume viewer software on a dedicated workstation (CompassView 5.0, Philips Corp., Amsterdam, the Netherlands). Regions of interest (ROIs) were manually placed to cover the lesion, and the maximal SUV (SUVmax) value was recorded for each lesion in the PET images of the patients. For each PET dataset, the SUVmax was defined as the highest SUV among all hypermetabolic tumour foci. The MTV was determined by drawing a circular ROI that included all involved lesions in the axial, coronal, and sagittal PET images. The voxel boundaries were produced automatically with the 41% SUVmax threshold method recommended by the European Association of Nuclear Medicine [28]. The total metabolic tumour volume (TMTV) was obtained by summing the

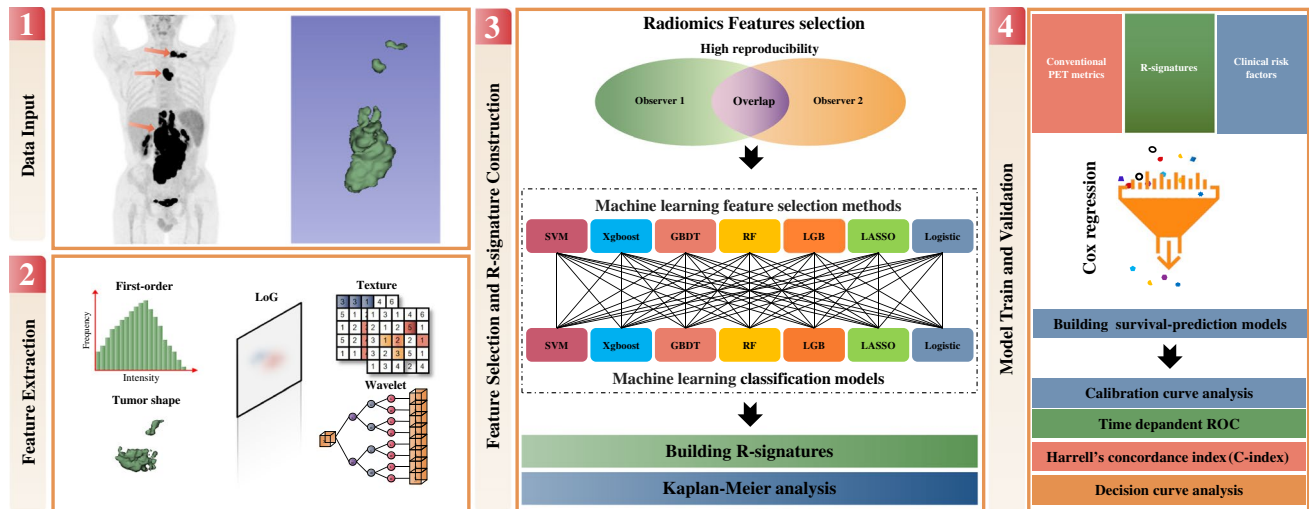


Fig. 2 Analysis workflow in this study. *SVM*: support vector machine, *GBDT*: gradient boosting decision tree, *XGBoost*: extreme gradient boosting, *RF*: random forest, *LGB*: light gradient boosting machine,

LASSO: least absolute shrinkage and selection operator, *Logistic*: logistic regression

MTV of all lesions. The TLG was calculated as the sum of the product of the SUVmean and MTV of all lesions (TLG = sum of [SUVmean × MTV]).

Radiomics feature selection and signature construction

First, all extracted radiomics features were harmonized using z scores. The interobserver repeatability of each radiomic feature was evaluated in 50 randomly chosen patients using the intraclass correlation coefficient (ICC) method, and features with an ICC greater than 0.75 were selected.

Then, we calculated the effective and robust R-signature based on the cross-combination method of seven machine learning models. Specifically, we choose seven classical machine learning models: support vector machines (SVMs), gradient boosting decision trees (GBDTs), extreme gradient boosting (XGBoost), random forest (RF), light gradient boosting machines (LGB), the least absolute shrinkage and selection operator (LASSO), and logistic regression (Logistic). The SVM [29] maps the input samples to points in high-dimensional space to maximize the gap between two categories. The GBDT [30] is a robust prediction model that consists of a series of weak decision trees. The XGBoost method extends GBDTs [31] in terms of structure and training strategy for more efficient predictions. Furthermore, the LGB [32] outperforms XGBoost in terms of accuracy and operation speed due to its tree structure and optimization algorithm. The RF method [33] consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the RF produces a class prediction, and the class with the most votes becomes the model's prediction.

LASSO [34] is a classical regression analysis method that performs both feature selection and regularization, with features with nonzero coefficients viewed as useful features. Logistic regression [35] combines linear regression and activation functions to estimate the probability of a certain class.

These seven models acted as feature selection and classification models for predicting the PFS and OS. The feature selection models were used to select the most significant radiomics features, and the classification models were used to construct an R-signature based on the selected features. In previous works, LASSO was generally used, as it produced a simpler and more interpretable format than the other models. However, the effectiveness of LASSO has not been thoroughly considered. In this paper, we generated a set of feature selection-classification candidates with cross-combinations of the seven machine learning models, such as LASSO-SVM, LGB-XGBoost and Logistic-RF. Then, the R-signature was built by the optimal candidate. The optimal candidate was determined with tenfold cross-validation in four steps (see Supplementary Fig. 1): (I) the samples were split into 10 groups of approximately equal size. One group was used for validation, and the remaining 9 groups were used for training. This process was repeated 10 times, with different groups used for validation each time. Considering that our samples came from two medical centres, in the split procedure, we ensured that the cohorts retained the same proportion of sample sources as the original dataset. (II) Given a training cohort, we trained a feature selection model with all PET radiomics features for predicting the PFS/OS to obtain the corresponding feature weights. Based on the feature weights, we trained a feature classification model by recursively considering the subsets of the radiomics features.

(III) We identified the feature subset with the largest AUC value as containing the most important features, and the corresponding feature selection-classification model was identified as the optimal candidate. The R-signature was the prediction result of the optimal classification model. (IV) Based on the optimal candidate, we calculated the AUC of each validation cohort and outputted the average AUC of all 10 validation cohorts.

Development and validation of the models

Univariate and multivariate Cox regression analyses were used to select the potential independent prognostic factors of PFS and OS in the training cohort. All the significant clinical factors (age, LDH, B symptoms, ECOG, Ann Arbor stage, extranodal involvement, bulky disease) in the univariate analysis were included in the multivariate analysis. Independent predictors (age, Ann Arbor stage, bulky disease) were used to build the clinical model. Then, all statistically significant PET factors (SUVmax, TMTV, TLG, R-signature) in the univariate analysis were included in the multivariate analysis. The independent predictors (SUVmax, TMTV, R-signature) were used to build a combined PET-based model. After that, we assembled the models by combining all the independent clinical and PET predictors to build the combined model. Calibration curves were used to assess the models. To quantify the discrimination performance of the models, the Harrell concordance index (C-index) and the time-dependent area under the ROC curve (tdROC) were calculated. The performance of the models was also externally tested in the independent validation cohort with the formula and cut-off values derived from the training cohort.

Clinical benefit analysis based on the models

To estimate the clinical utility of the models, decision curve analysis (DCA) was performed by calculating the net benefits for a range of threshold probabilities in the training cohort and validation cohort.

Statistical analysis

All statistical tests were performed using SPSS 22.0 (IBM Corp., Armonk, NY, USA) and R statistical software (version 4.0.2). A P value less than 0.05 was considered statistically significant. The difference in the related clinical information between the training and validation cohorts was assessed using χ^2 tests or Mann–Whitney U tests, as appropriate. We determined the optimal threshold of the R-signature based on the receiver operating characteristic (ROC) curve. Cox regression analysis was used to analyse the potential independent predictors and build the models. The survival conditions were evaluated by the Kaplan–Meier

method and compared by log-rank tests. The calibration curves, C-index, tdAUC and DCA were calculated for the models in the training and validation cohorts.

Results

Patient characteristics

The baseline characteristics of the included patients in the training and validation cohorts are summarized in Table 1. No statistically significant ($P < 0.05$) difference was observed between the two cohorts. A total of 14 participants had SUVmean values in the liver < 1.3 (training cohort: $n=3$; validation cohort: $n=2$) or ≥ 3.0 (training cohort: $n=5$; validation cohort: $n=4$). The median follow-up times for the training and validation cohorts were 35.0 and 23.5 months, respectively. Of the participants who experienced disease relapse or progression in the training cohort, 103 and 76 died, respectively, and the 1-year, 3-year, and 5-year PFS and 1-year, 3-year, and 5-year OS rates were 77.3%, 66.1%, and 54.6% and 85.7%, 76.3%, and 66.8%, respectively. Of the participants who experienced disease relapse or progression in the validation cohort, 33 and 25 died, respectively, and the 1-year, 3-year, and 5-year PFS and 1-year, 3-year, and 5-year OS rates were 83.5%, 67.9%, and 63.4% and 88.9%, 74.1%, and 71.9%, respectively.

Radiomics feature selection and R-signature construction

In the first step of reproducible feature selection, we excluded features with ICCs less than 0.75, and 274 features with high reproducibility were selected for further analysis. Based on the results of the feature extraction-classification models, we determined 12 radiomic features for constructing the R-signature_{PFS} for the SVM-SVM model (AUC=0.757) and 31 radiomic features for constructing the R-signature_{OS} for the XGBoost-XGBoost model (AUC=0.762) (Fig. 3 and Supplementary Tables 2 and 3). The optimal cut-off values of the R-signatures were 0.386 for PFS and 0.268 for OS. Using these cut-offs, the risk groups identified by the R-signatures showed distinct OS and PFS outcomes in both the training and validation cohorts (Fig. 3).

Univariate and multivariate analysis results

The differences in the clinical characteristics between the dichotomized R-signature_{PFS} and R-signature_{OS} groups are shown in Table 2.

For the clinical variables, we performed univariate and multivariate Cox regression analyses to identify the independent factors associated with survival. We identified age

Table 1 Demographics and clinical characteristics of the study population

Characteristic	Training cohort (n = 273)	External validation cohort (n = 110)	P value
Gender			
Female	124(45.4%)	55(50.0%)	0.331
Male	149(54.6%)	55(50.0%)	
Age (year)			
<60	156(57.1%)	59(53.6%)	0.440
≥60	117(42.9%)	51(46.4%)	
LDH			
normal	151(55.3%)	80(72.7%)	0.360
elevated (>271 U/L)	122(44.7%)	30(27.3%)	
B symptoms			
no	172(63.0%)	70(63.6%)	0.552
yes	101(37.0%)	40(36.4%)	
ECOG PS			
0-1	184(67.4%)	98(89.1%)	0.131
≥2	89(32.6%)	12(10.9%)	
Ann Arbor stage			
I-II	100 (36.6%)	65 (59.1%)	0.583
III-IV	173 (63.4%)	45 (40.9%)	
Extranodal involvement			
0-1	137 (50.2%)	80(72.7%)	0.726
≥2	136 (49.8%)	30(27.3%)	
NCCN-IPI			
0-3	138 (50.5%)	68(61.8%)	0.241
≥4	135 (49.5%)	42(38.2%)	
Bulky disease			
no	197 (72.2%)	88 (80.0%)	0.813
yes	76 (27.8%)	22 (20.0%)	
Pathological type			
GCB	150 (54.9%)	47 (42.7%)	0.170
non-GCB	123 (45.1%)	63 (57.3%)	
SUVmax [#]	19.6 (3.5-46.4)	19.9 (4.5-47.5)	0.871*
<22.5	182 (66.7%)	67 (60.9%)	0.836
≥22.5	91 (33.3%)	43 (39.1%)	
TMTV (cm ³) [#]	264.3 (25.0-2255.7)	224.0 (18.4-4323.6)	0.102*
<314.0	202 (74.0%)	87 (79.1%)	0.586
≥314.0	71 (26.0%)	23 (20.9%)	
TLG [#]	2293.9 (84.7-15787.4)	2362.9 (88.9-19958.4)	0.088*
<2088.3	181 (66.3%)	78 (70.9%)	0.655
≥2088.3	92 (33.7%)	32 (29.1%)	

F, female; M, male; LDH, lactate dehydrogenase; ECOG PS, Eastern Cooperative Oncology Group performance status; NCCN-IPI, National Comprehensive Cancer Network International Prognostic Index; GCB, germinal centre B cell; SUVmax, maximum standardized uptake value; TMTV, total metabolic tumour volume; TLG, total lesion glycolysis

[#]median (range)

*p value derived from Mann–Whitney U test. The other p value was derived from the χ^2 test

(HR=2.460, $P<0.001$; HR=2.810, $P<0.001$), Ann Arbor stage (HR=3.139, $P<0.001$; HR=3.105, $P<0.001$) and bulky disease (HR=1.518, $P=0.043$; HR=1.631, $P=0.041$) as independent risk factors for PFS and OS (Table 3).

For the PET variables, univariate and multivariate Cox regression analyses showed that the R-signature (HR=5.024, $P<0.001$; HR=4.054, $P<0.001$), SUVmax (HR=2.027, $P=0.001$; HR=2.005, $P=0.004$) and TMTV (HR=2.265,

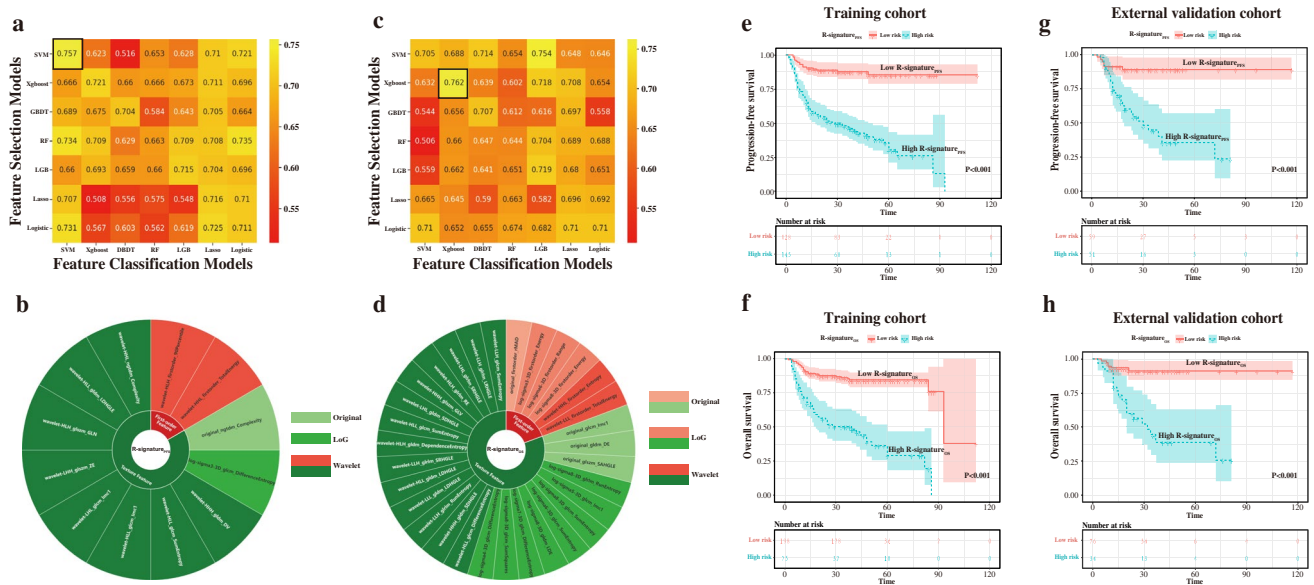


Fig. 3 Heatmaps indicate the AUC performance of the cross-combinations of the feature selection methods (columns) and classification models (rows) in predicting PFS (a) and OS (c). The spectra demonstrating the selected features to build R-signature_{PFS} (b) and R-signa-

ture_{OS} (d). Kaplan–Meier plots for patients stratified by risk classification according to R-signatures for patients in the training (e, f) and validation (g, h) cohorts

Table 2 Comparison of the R-signature with the patient clinical data and PET metabolic parameters in the training cohort

Patient data	R-signature _{PFS}			R-signature _{OS}		
	Low (n = 128)	High (n = 145)	P value*	Low (n = 198)	High (n = 75)	P value*
Sex, F/M	64/64	60/85	0.180	91/107	33/42	0.787
Age, <60/≥60	80/48	76/69	0.111	118/80	38/37	0.218
LDH level, normal/elevated	95/33	56/89	<0.001	131/67	20/55	<0.001
B symptoms, no/yes	92/36	80/65	0.006	138/60	34/41	<0.001
ECOG PS, 0–1/≥2	91/37	87/58	0.007	144/54	40/35	0.004
Ann Arbor stage, I-II/III-IV	71/57	29/116	<0.001	92/106	8/67	<0.001
Extranodal involvement	85/43	52/93	<0.001	113/85	24/51	<0.001
NCCN-IPI, 0-3/≥4	91/37	47/98	<0.001	124/74	14/61	<0.001
Bulky disease, no/yes	106/22	91/54	<0.001	153/45	44/31	0.004
Pathological type, GCB/non-GCB	68/60	82/63	0.626	111/87	39/36	0.587
SUVmax, low/high	87/41	95/50	0.701	133/65	50/25	1.000
TMTV, low/high	115/13	87/58	<0.001	173/25	29/46	<0.001
TLG, low/high	111/17	70/75	<0.001	161/37	20/55	<0.001

LDH, lactate dehydrogenase; ECOG PS, Eastern Cooperative Oncology Group performance status; NCCN-IPI, National Comprehensive Cancer Network International Prognostic Index; GCB, germinal centre B cell; SUVmax, maximum standardized uptake value; TMTV, total metabolic tumour volume; TLG, total lesion glycolysis

* P < 0.05

P<0.001; HR=1.789, P=0.036) were independent risk factors for PFS and OS (Table 3).

Assessment and validation of models built for predicting PFS and OS

The independent clinical (age, Ann Arbor stage, and bulky disease) and PET (R-signature, TMTV and SUVmax)

Table 3 Univariate and multivariate analyses of factors predictive of progression-free survival and overall survival in the training cohort

Category	Variables	Progression-free survival				Overall survival			
		Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
		HR(95% CI)	P value*	HR(95% CI)	P value*	HR(95% CI)	P value*	HR(95% CI)	P value*
Clinical predictors	Sex, F/M	0.765 (0.518-1.130)	0.179	-	-	0.738 (0.469-1.161)	0.189	-	-
	Age, <60/≥60	2.551 (1.709-3.808)	<0.001	2.460 (1.645-3.679)	<0.001	2.928 (1.809-4.740)	<0.001	2.810 (1.732-4.560)	<0.001
	LDH level, normal/elevated	1.644 (1.113-2.248)	0.012	-	-	1.729 (1.095-2.729)	0.019	-	-
	B symptoms, no/yes	1.556 (1.054-2.298)	0.026	-	-	1.508 (0.956-2.378)	0.077	-	-
	ECOG PS, 0-1/≥2	1.827 (1.234-2.706)	0.003	-	-	1.728 (1.092-2.735)	0.020	-	-
	Ann Arbor stage, I-III/III-IV	3.463 (2.054-5.840)	<0.001	3.139 (1.857-5.307)	<0.001	3.529 (1.900-6.553)	<0.001	3.105 (1.668-5.780)	<0.001
	Extranodal involvement	1.863 (1.243-2.792)	0.003	-	-	1.519 (0.956-2.414)	0.077	-	-
	Bulky disease, no/yes	1.733 (1.160-2.589)	0.007	1.518 (1.014-2.273)	0.043	1.786 (1.121-2.845)	0.015	1.631 (1.019-2.608)	0.041
	Pathological type, GCB/non-GCB	1.216 (0.826-1.792)	0.332	-	-	1.270 (0.809-1.996)	0.299	-	-
PET predictors	SUVmax, low/high	1.713 (1.154-2.543)	0.008	2.027 (1.349-3.044)	0.001	1.736 (1.099-2.744)	0.018	2.005 (1.256-3.200)	0.004
	TMTV, low/high	3.182 (2.153-4.703)	<0.001	2.265 (1.501-3.419)	<0.001	3.402 (2.161-5.357)	<0.001	1.789 (1.038-3.200)	0.036
	TLG, low/high	3.087 (2.088-4.563)	<0.001	-	-	3.795 (2.386-6.036)	<0.001	-	-
	R-signature _{PFS} , low/high	8.037 (3.304-19.551)	<0.001	5.024 (2.892-8.728)	<0.001	-	-	-	-
	R-signature _{OS} , low/high	-	-	-	-	5.347 (3.355-8.522)	<0.001	4.054 (2.337-7.031)	<0.001
NCCN-IPI	NCCN-IPI, 0-3/≥4	3.091 (2.009-4.756)	<0.001	-	-	2.969 (1.793-4.918)	<0.001	-	-

CI, confidence interval; SE, standard error; HR, hazard ratio; LDH, lactate dehydrogenase; ECOG PS, Eastern Cooperative Oncology Group performance status; NCCN-IPI, National Comprehensive Cancer Network International Prognostic Index; GCB, germinal centre B cell; SUVmax, maximum standardized uptake value; TMTV, total metabolic tumour volume; TLG, total lesion glycolysis

* $P < 0.05$

predictors that were significant in the multivariate analysis were used to build the combined models for predicting the PFS and OS rates (Fig. 4 and Table 4). Individual clinical models and PET-based models were also constructed (Table 4).

For PFS, the C-indices of the combined model in the training and validation cohorts were 0.801 (95% CI: 0.745-0.848) and 0.758 (95% CI: 0.618-0.795), respectively; these values were superior to those of the clinical model (C-index: 0.732, 95% CI: 0.679-0.785 in the training cohort, and C-index: 0.621, 95% CI: 0.512-0.731 in the validation cohort), PET-based model (C-index: 0.785, 95% CI:

0.734-0.835 in the training cohort, and C-index: 0.732, 95% CI: 0.642-0.823 in the validation cohort) and NCCN-IPI model (C-index: 0.720, 95% CI: 0.668-0.772 in the training cohort, and C-index: 0.673, 95% CI: 0.584-0.763 in the validation cohort) (Table 5).

For OS, the C-indices of the combined model in the training and validation cohorts were 0.807 (95% CI: 0.736-0.854) and 0.794 (95% CI: 0.668-0.881); these values were superior to those of the clinical model (C-index: 0.740, 95% CI: 0.680-0.800 in the training cohort, and C-index: 0.696, 95% CI: 0.584-0.807 in the validation cohort), PET-based model (C-index: 0.773, 95% CI: 0.713-0.833 in the

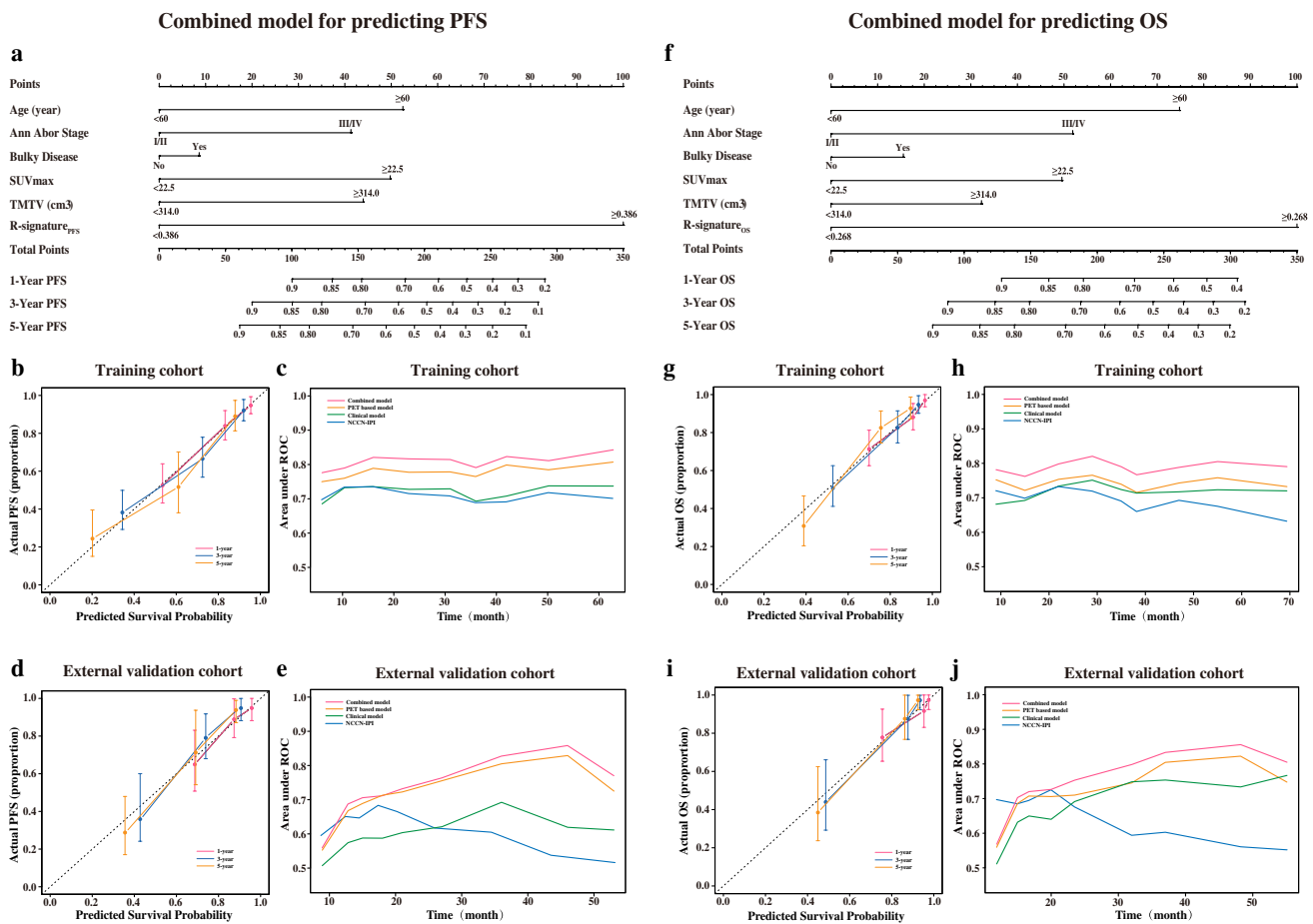


Fig. 4 **a** The combined model for predicting PFS. **b, d** Calibration curves of the model for predicting PFS in the training and validation cohorts at the 1-year, 3-year, and 5-year time points. **c, e** Time-dependent area under the ROC curve of the models for predicting PFS in the training and validation cohorts. **f** The combined model for

predicting OS. **g, i** Calibration curves of the models for predicting OS in the training and validation cohorts at the 1-year, 3-year, and 5-year time points. **h, j** Time-dependent area under the ROC curve of the models for predicting OS in the training and validation cohorts

Table 4 The survival-prediction models included in this study

Models	Included variables
Combined Model	Age, Ann Abor stage, Bulky disease, SUVmax, TMTV, R-signature
Clinical model	Age, Ann Abor stage, Bulky disease
PET based model	SUVmax, TMTV, R-signature
NCCN-IPI	NCCN-IPI

CI, confidence interval; HR, hazard ratio; NCCN-IPI, National Comprehensive Cancer Network International Prognostic Index; SUVmax, maximum standardized uptake value; TMTV, total metabolic tumour volume

training cohort, and C-index: 0.781, 95% CI: 0.697-0.902 in the validation cohort) and NCCN-IPI model (C-index: 0.726, 95% CI: 0.665-0.786 in the training cohort, and C-index: 0.708, 95% CI: 0.610-0.805 in the validation cohort) (Table 5).

Table 5 The Harrell’s C-index results in the training and validation cohorts

	Training cohort		External validation cohort	
	C-index	95% CI	C-index	95% CI
Progression-free survival				
Combined model	0.801	0.745-0.848	0.758	0.618-0.795
Clinical model	0.732	0.679-0.785	0.621	0.512-0.731
PET based model	0.785	0.734-0.835	0.732	0.642-0.823
NCCN-IPI	0.720	0.668-0.772	0.673	0.584-0.763
Overall survival				
Combined model	0.807	0.736-0.854	0.794	0.668-0.881
Clinical model	0.740	0.680-0.800	0.696	0.584-0.807
PET based model	0.773	0.713-0.833	0.781	0.697-0.902
NCCN-IPI	0.726	0.665-0.786	0.708	0.610-0.805

CI, confidence interval; HR, hazard ratio

The calibration curves for predicting the probability of PFS and OS at 1, 3, or 5 years for each model after 1000 bootstrap replicates are shown in Fig. 4. The calibration curves showed satisfactory agreement between the estimations and actual observations in the training and validation cohorts for the combined models.

Performance analysis of the combined models in clinical use

The decision curve analyses for the combined models are presented in Fig. 5. The decision curve analysis showed that the combined models had a higher overall net benefit than the clinical models, PET-based models, and NCCN-IPI across most of the risk thresholds in both the training and validation cohorts for PFS and OS.

Outcome according to R-signature in the subgroups of the NCCN-IPI model

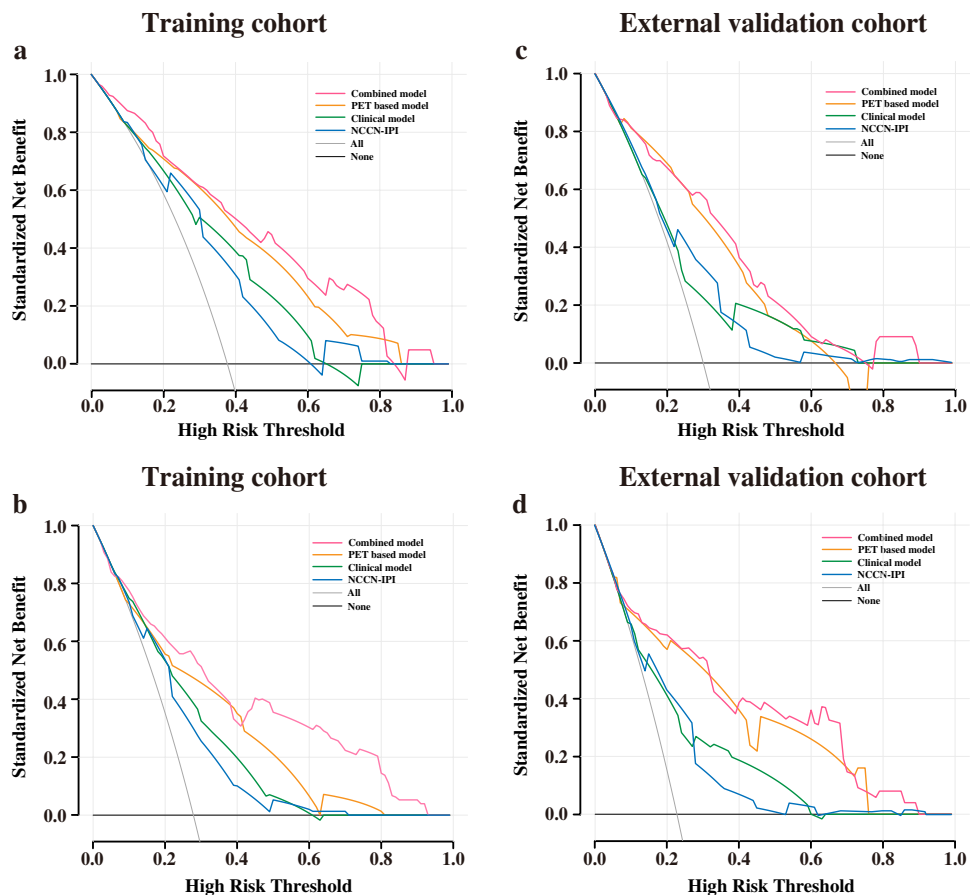
To confirm the added prognostic value of radiomics features to the NCCN-IPI, we evaluated the prognostic value of R-signatures in low/low-intermediate risk (NCCN-IPI

of <4) and high-intermediate/high (NCCN-IPI of ≥ 4) groups.

For the training cohort, in the low/low-intermediate risk group ($n=138$), patients with a low R-signature had a better survival rate (92.3% for PFS and 89.5% for OS) than those with a high R-signature (51.1% for PFS and 35.7% for OS), and the P values were both <0.001 . In the high-intermediate/high risk group ($n=135$), patients with a low R-signature had a better survival rate (75.7% for PFS and 77.0% for OS) than those with a high R-signature (34.7% for PFS and 39.3% for OS), and the P values were both 0.001 (Fig. 6 and Table 6).

For the validation cohort, in the low/low-intermediate risk group ($n=68$), patients with a low R-signature had a better survival rate (97.4% for PFS and 96.2% for OS) than those with a high R-signature (55.2% for PFS and 56.3% for OS), and the P values were both statistically significant (<0.001). In the high-intermediate/high risk group ($n=42$), patients with a low R-signature had a better survival rate (75.0% for PFS and 83.3% for OS) than those with a high R-signature (36.4% for PFS and 33.3% for OS), and the P values were 0.031 and 0.007 (Fig. 6 and Table 6).

Fig. 5 Decision curve analysis for the models in the training (a, b) and validation (c, d) cohorts



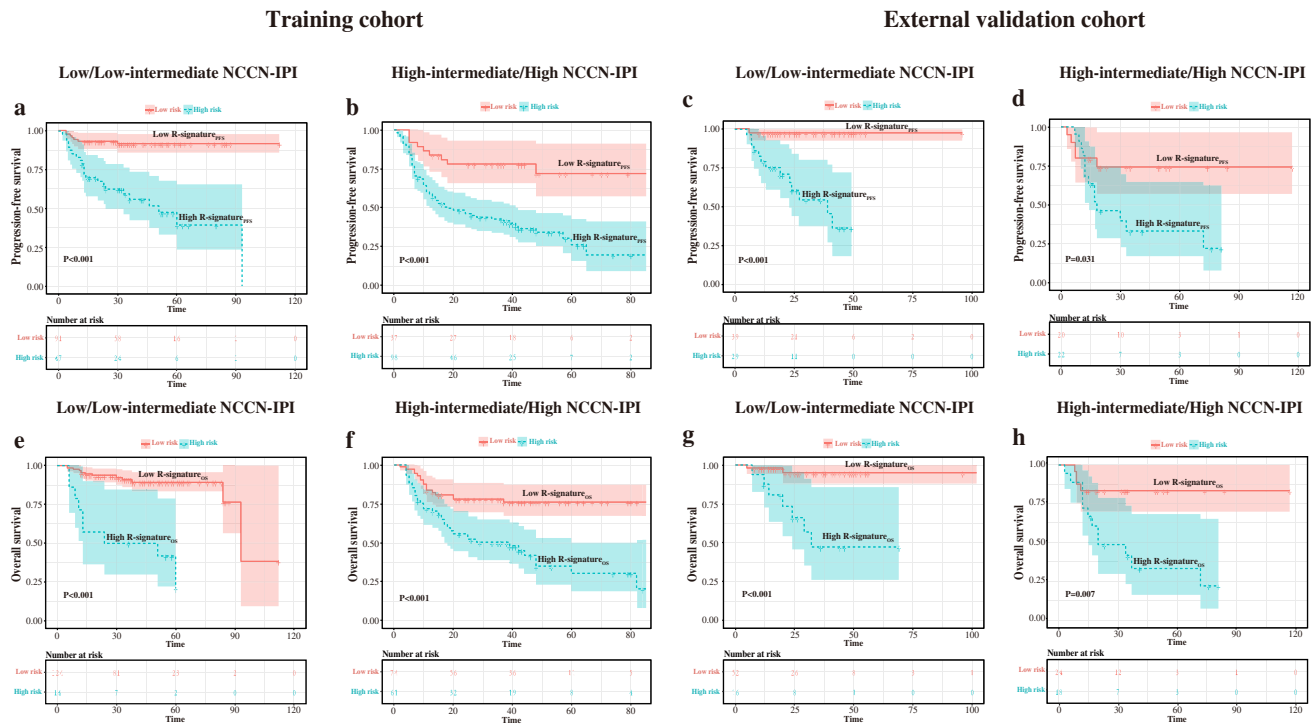


Fig. 6 Kaplan–Meier plots for PFS and OS in patients with different NCCN-IPI scores in relation to their R-signatures. **a–d** Low/low-intermediate NCCN-IPI (score: 0,1/2,3) and high-intermediate/high NCCN-IPI (score: 4,5/6–8) in relation to R-signature_{PFS} for PFS stratification in the training (**a, b**) and validation (**c, d**) cohorts. **e–h** Low/low-intermediate NCCN-IPI (score: 0,1/2,3) and high-intermediate/high NCCN-IPI (score: 4,5/6–8) in relation to R-signature_{OS} for OS stratification in the training (**e, f**) and validation (**g, h**) cohorts

Discussion

Our study analysed PET radiomics features to predict survival outcomes in DLBCL patients. The results of this study indicated that combined models that integrated R-signatures calculated by radiomics features extracted from pretreatment [18F]-FDG PET using machine learning methods could be used as a preoperative tool for individualized prognosis prediction treatment guidance in patients with DLBCL.

Accurately predicting the prognosis of patients is of great importance for optimizing DLBCL treatment strategies. Several studies have attempted to assess the predictive value of radiomics information from PET images in DLBCL [21–23]. Aide et al. recruited 132 patients with DLBCL to evaluate the ability of pretreatment PET radiomics features to predict prognosis and found that LZLGE was the only independent predictor of 2-year event-free survival (HR=2.84, $P=0.01$) with ROC analysis accuracy (area under curve=0.76) [21]. Lue KH et al. extracted 80 PET-based radiomics features from 171 patients with DLBCL and found that RLN_{GLRLM} was independently associated with PFS (HR=15.7, $P=0.007$) and OS (HR=8.64, $P=0.040$) [22]. Eertink et al. found that PET radiomics features extracted from the largest lesion of DLBCL patients were able to predict the 2-year time to progression (area under curve=0.67)

[23]. Consistent with previous studies, the R-signatures consisting of PET radiomics features identified with machine learning in the present study were independently associated with PFS (HR=4.15, $P<0.001$) and OS (HR=4.029, $P<0.001$). Both our study and previous studies support the viewpoint that pivotal radiomics feature data extracted from PET images could help predict clinical outcomes in patients with DLBCL. The underlying cause might be explained by the hypothesis that radiomics features are strongly correlated with tumour heterogeneity [36–38], which is a prognostic determinant of patient survival [39–41].

Recently, machine learning models have attracted great attention in both research and industry because they can automatically detect inherent patterns in data distribution [42]. With advancements in machine learning, different machine learning models, such as random forest (RF) and support vector machines (SVMs), have been utilized to extract data features. These methods differ from each other in terms of structure, optimization algorithm, scale, and performance in different datasets. For example, SVMs are more suitable than logistic regression for large datasets [43]. However, most existing works have only focused on a single method for feature selection and R-signature construction, and few comparisons have been performed with other machine learning methods to assess robustness and

Table 6 PFS and OS of patients based on their R-signatures in the low/low-intermediate risk and high-intermediate/high risk groups

Patients	NCCN-IPI Group	R-signature Group	No. of patients		Progression-free survival			Overall survival		
			No. of patients	No. of patients	Number of events	Survival (%)	χ^2 (P value*)	Number of events	Survival (%)	χ^2 (P value*)
Training cohort	Low/low-intermediate risk	Low	91	7	92.3	28.472 (<0.001)	124	13	89.5	32.176 (<0.001)
		High	47	23	51.1		14	9	35.7	
Validation cohort	High-intermediate/High risk	Low	37	9	75.7	16.555 (<0.001)	74	17	77.0	19.152 (<0.001)
		High	98	64	34.7		61	37	39.3	
	Low/low-intermediate risk	Low	39	1	97.4	17.690 (<0.001)	52	2	96.2	14.099 (<0.001)
		High	29	13	55.2		16	7	56.3	
High intermediate/high risk		Low	20	5	75.0	4.656 (0.031)	24	4	83.3	7.255 (0.007)
		High	22	14	36.4		18	12	33.3	

* $P < 0.05$

efficiency. The cross-combination of feature selection-classification models has shown great efficiency and robustness in clinical practice [24, 25]. Thus, we developed a cross-combination approach for determining the optimal feature selection-classification pair based on AUC results. As shown in Fig. 3, the AUC used by the LASSO method for predicting OS was poor (0.696). Therefore, we chose the most powerful selection-classification pair from 49 possible combinations. The SVM-SVM model had an AUC of 0.757 for predicting PFS, and the XGBoost-XGBoost model had an AUC of 0.762 for predicting OS. In addition, we found that the selected features (see Fig. 3) were mainly texture features with the wavelet and LoG filtering algorithms, indicating that these features have better discrimination than shape and first-order features.

[18F]-FDG PET results can be used as a measure of cellular glucose metabolism to provide biological information about tumours. Our study demonstrated that the metabolic parameter TMTV was an independent predictor of PFS (HR=1.90, $P=0.003$), and SUV_{max} was found to be an independent predictor of both PFS (HR=2.03, $P=0.001$) and OS (HR=1.77, $P=0.015$). These results were consistent with previous studies [44–47]. To apply these results in clinical practice, we developed user-friendly predictive combined models that incorporated the R-signature, metabolic metrics (SUV_{max} and TMTV), and clinical risk factors (age, stage and bulky disease), and we compared the predictive performance of the models with clinical, PET-based, and NCCN-IPI prognostic models. The study found that the combined models had satisfactory agreement with the calibration curve and exhibited significant prognostic superiority over other models in terms of C-indices and tAUCs.

To confirm the combined model’s risk discrimination, we compared the clinical utility of the combined model with that of other models (such as the clinical model, PET-based model and NCCN-IPI model). The DCA demonstrated that the combined models were superior to the other models across most of the range of reasonable threshold probabilities, thus suggesting that they are more suitable for estimating individual survival. Consistent with previous studies [6–8], our results showed that the NCCN-IPI might be improved for identifying high-risk groups for which novel treatment approaches are most needed (Supplementary Fig. 2 and Supplementary Table 4). The possible additional predictive value of the combined R-signature and the NCCN-IPI model was explored in this study. The results showed that the addition of the R-signature to the NCCN-IPI model yielded a better stratification for DLBCL patients in both high-intermediate/high risk (NCCN-IPI of ≥ 4) and low/low-intermediate risk (NCCN-IPI of < 4) groups, in both the training and validation cohorts. Interestingly, several studies have shown that the TMTV also had additive prognostic value in patients in the NCCN-IPI subgroup [48,

49]. In agreement with the work of Eertink et al. [23], our results support further exploration of combining radiomics features with clinical predictors to more accurately identify high-risk patients. To uncover the potential additional prognostic value of these radiomics features, these results may contribute to better understanding of the clinically beneficial effects of these parameters with machine learning methods.

One of the limitations of our study was its retrospective nature. Additionally, the biological features of DLBCL, including the double expressor, double hit, and genomic prognostic factors, were not investigated in this study. Furthermore, the SUVmean of the liver was outside the normal range in a small portion of our patients (n=14). According to the quality control (QC) criteria described by EANM guidelines, the SUVmean of the liver should be between 1.3 and 3.0. Moreover, given the low number of progression or death events in the NCCN-IPI categories, conclusions about the additional predictive value of the R-signature to the NCCN-IPI model should be interpreted with caution, and larger cohorts are needed in the future. In addition, we used only the 3 largest lesions as targets for radiomic feature calculation in the present study. To date, there is no consensus on the tumour segmentation method for radiomic feature calculation in patients with DLBCL. Due to the high distribution variability of nodal and/or extranodal lesions with heterogeneous volumes and variable metabolic activity, lymphoma segmentation is more challenging than that of primary tumour lesions. Therefore, we proposed a concise tumour segmentation approach for the assessment of radiomics features that may be easier to apply in clinical settings. Finally, the 3D slicer and the growth cut algorithm applied in the present study have not been clinically approved. Although the development of PyRadiomics is involved in the standardization effort of the IBSI (The Image Biomarker Standardization Initiative) team, there are still some differences between PyRadiomics-based feature extraction and IBSI-recommended feature extraction. The main differences are due to grey value discretization, resampling (grid alignment, grey value rounding, and mask resampling), resegmentation, and kurtosis calculation. All of these differences may alter the extrapolation of our results to other clinical practices. The main strengths of our study are the homogeneity of the included patients (de novo DLBCL histology and R-CHOP-like regimens as standard treatment) and the use of an external validation method that supports the general applicability of the models.

Conclusion

The R-signature obtained with the cross-combination method could be used as a survival predictor for patients with DLBCL. Furthermore, the combined models that

incorporate R-signature metabolic metrics and clinical risk factors may allow for accurate risk stratification and individualized management of DLBCL patients.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00259-022-05717-9>.

Authors' contributions Chong Jiang, Ang Li, Yue Teng, Xiangjun Huang, and Chongyang Ding collected data and analysed data; Jingyan Xu, Jianxin Chen, and Zhengyang Zhou participated in the research design; Chong Jiang and Ang Li contributed to the writing of the manuscript, discussed data, and supervised the study, and all authors performed data analysis and interpretation and read and approved the final version of the article.

Funding This work was partially supported by fundings for Clinical Trials from the Affiliated Drum Tower Hospital, Medical School of Nanjing University under Grant No. 2021-LCYJ-MS-04.

This work was also partially supported by fundings for the Key Project of Medical Science and Technology of Nanjing under Grant No.ZKX21011.

Data availability The datasets generated and analysed during the current study are available from the Nanjing Drum Tower Hospital, the Affiliated Hospital of Nanjing University Medical School, and the Jiangsu Province Hospital, the First Affiliated Hospital of Nanjing Medical University.

Declarations

Conflicts of interest/Competing interests The authors declare that they have no conflicts of interest

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

1. Liu Y, Barta SK. Diffuse large B-cell lymphoma: 2019 update on diagnosis, risk stratification, and treatment. *Am J Hematol*. 2019;94(5):604–16.
2. Bakhshi TJ, Georger PT. Genetic and epigenetic determinants of diffuse large B-cell lymphoma. *Blood Cancer J*. 2020;10(12):123.
3. Kwak JY. Treatment of diffuse large B cell lymphoma. *Korean J Intern Med*. 2012;27(4):369–77.
4. Coiffier B, Thieblemont C, Van Den Neste E, Lepage G, Plantier I, Castaigne S, Lefort S, Marit G, Macro M, Sebban C, et al. Long-term outcome of patients in the LNH-98.5 trial, the first randomized study comparing rituximab-CHOP to standard CHOP chemotherapy in DLBCL patients: a study by the Groupe d'Etudes des Lymphomes de l'Adulte. *Blood*. 2010;116(12):2040–2045.
5. Coiffier B, Sarkozy C. Diffuse large B-cell lymphoma: R-CHOP failure-what to do? *Hematol Am Soc Hematol Educ Program*. 2016;2016(1):366–78.
6. Sehn LH, Berry B, Chhanabhai M, Fitzgerald C, Gill K, Hoskins P, Klasa R, Savage KJ, Shenkier T, Sutherland J, et al. The revised International Prognostic Index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP. *Blood*. 2007;109(5):1857–61.

7. International Non-Hodgkin's Lymphoma Prognostic Factors P. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med.* 1993;329(14):987–994.
8. Zhou Z, Sehn LH, Rademaker AW, Gordon LI, Lacasce AS, Crosby-Thompson A, et al. An enhanced International Prognostic Index (NCCN-IPI) for patients with diffuse large B-cell lymphoma treated in the rituximab era[J]. *Blood.* 2014;123(6):837–42.
9. Gallicchio R, Mansueto G, Simeon V, Nardelli A, Guariglia R, Capacchione D, Soscia E, Pedicini P, Gattozzi D, Musto P, et al. F-18 FDG PET/CT quantization parameters as predictors of outcome in patients with diffuse large B-cell lymphoma. *Eur J Haematol.* 2014;92(5):382–9.
10. Jiang C, Teng Y, Chen J, Wang Z, Zhou Z, Ding C, Xu J. Value of (18)F-FDG PET/CT for prognostic stratification in patients with primary intestinal diffuse large B cell lymphoma treated with an R-CHOP-like regimen. *Ann Nucl Med.* 2020;34(12):911–9.
11. Xie M, Zhai W, Cheng S, Zhang H, Xie Y, He W. Predictive value of F-18 FDG PET/CT quantization parameters for progression-free survival in patients with diffuse large B-cell lymphoma. *Hematology.* 2016;21(2):99–105.
12. Xie M, Wu K, Liu Y, Jiang Q, Xie Y. Predictive value of F-18 FDG PET/CT quantization parameters in diffuse large B cell lymphoma: a meta-analysis with 702 participants. *Med Oncol.* 2015;32(1):446.
13. Vercellino L, Cottreau AS, Casasnovas O, Tilly H, Feugier P, Chartier L, Fruchart C, Roulin L, Oberic L, Pica GM, et al. High total metabolic tumor volume at baseline predicts survival independent of response to therapy. *Blood.* 2020;135(16):1396–405.
14. Stanta G, Bonin S. Overview on Clinical Relevance of Intra-Tumor Heterogeneity. *Front Med (Lausanne).* 2018;5:85.
15. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell.* 2017;168(4):613–28.
16. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue RTHM, Even AJG, Jochems A, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* 2017;14(12):749–62.
17. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* 2012;48(4):441–6.
18. Lue KH, Wu YF, Liu SH, Hsieh TC, Chuang KS, Lin HH, Chen YH. Prognostic Value of Pretreatment Radiomic Features of 18F-FDG PET in Patients With Hodgkin Lymphoma. *Clin Nucl Med.* 2019;44(10):e559–65.
19. Mayerhoefer ME, Riedl CC, Kumar A, Gibbs P, Weber M, Tal I, Schilksy J, Schöder H. Radiomic features of glucose metabolism enable prediction of outcome in mantle cell lymphoma. *Eur J Nucl Med Mol Imaging.* 2019;46(13):2760–9.
20. Wang H, Zhao S, Li L, Tian R. Development and validation of an (18)F-FDG PET radiomic model for prognosis prediction in patients with nasal-type extranodal natural killer/T cell lymphoma. *Eur Radiol.* 2020;30(10):5578–87.
21. Aide N, Fruchart C, Nganoa C, Gac AC, Lasnon C. Baseline (18) F-FDG PET radiomic features as predictors of 2-year event-free survival in diffuse large B cell lymphomas treated with immunotherapy. *Eur Radiol.* 2020;30(8):4623–32.
22. Lue KH, Wu YF, Lin HH, Hsieh TC, Liu SH, Chan SC, Chen YH. Prognostic Value of Baseline Radiomic Features of 18F-FDG PET in Patients with Diffuse Large B-Cell Lymphoma. *Diagnostics (Basel).* 2020;11(1):36.
23. Eertink JJ, van de Brug T, Wiegers SE, Zwezerijnen GJC, Pfähler EAG, Lugtenburg PJ, van der Holt B, de Vet HCW, Hoekstra OS, Boellaard R, et al. 18F-FDG PET baseline radiomics features improve the prediction of treatment outcome in diffuse large B-cell lymphoma. *Eur J Nucl Med Mol Imaging.* 2022;49(3):932–42.
24. Zhou Y, Ma XL, Zhang T, Wang J, Zhang T, Tian R. Use of radiomics based on (18)F-FDG PET/CT and machine learning methods to aid clinical decision-making in the classification of solitary pulmonary lesions: an innovative approach. *Eur J Nucl Med Mol Imaging.* 2021;48(9):2904–13.
25. Sun P, Wang D, Mok V C, Shi L. Comparison of feature selection methods and machine learning classifiers for radiomics analysis in glioma grading. *IEEE Access.* 2019;102010-102020.
26. Chang E, Joel MZ, Chang HY, Du J, Khanna O, Omuro A, et al. Comparison of radiomic feature aggregation methods for patients with multiple tumors. *Sci Rep.* 2021;11(1):9758.
27. Tixier F, Cheze-Le Rest C, Hatt M, Albarghach NM, Pradier O, Metges J-P, et al. Intratumor Heterogeneity Characterized by Textural Features on Baseline 18F-FDG PET Images Predicts Response to Concomitant Radiochemotherapy in Esophageal Cancer. *J Nucl Med.* 2011;52:369–78.
28. Boellaard R, Delgado-Bolton R, Oyen WJ, Giammarile F, Tatsch K, Eschner W. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging.* 2015;42(2):328–354.
29. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *Cancer Genomics Proteomics.* 2018;15(1):41–51.
30. Zhang Z, Jung C. GBDT-MO: Gradient-Boosted Decision Trees for Multiple Outputs. *IEEE Trans Neural Netw Learn Syst.* 2021;32(7):3156–67.
31. Hou N, Li M, He L, Xie B, Wang L, Zhang R, Yu Y, Sun X, Pan Z, Wang K. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. *J Transl Med.* 2020;18(1):462.
32. Zhang L, Liu M, Qin X, Liu G. Succinylation Site Prediction Based on Protein Sequences Using the IFS-LightGBM (BO) Model. *Comput Math Methods Med.* 2020;2020:8858489.
33. Sylvester EVA, Bentzen P, Bradbury IR, Clément M, Pearce J, Horne J, Beiko RG. Applications of random forest feature selection for fine-scale genetic population assignment. *Evol Appl.* 2018;11(2):153–65.
34. McEligot AJ, Poynor V, Sharma R, Panangadan A. Logistic LASSO Regression for Dietary Intakes and Breast Cancer. *Nutrients.* 2020;12(9):2652.
35. Nick TG, Campbell KM. Logistic regression. *Methods Mol Biol.* 2007;404:273–301.
36. Moon SH, Kim J, Joung JG, Cha H, Park WY, Ahn JS, Ahn MJ, Park K, Choi JY, Lee KH, et al. Correlations between metabolic texture features, genetic heterogeneity, and mutation burden in patients with lung cancer. *Eur J Nucl Med Mol Imaging.* 2019;46(2):446–54.
37. Choi ER, Lee HY, Jeong JY, Choi YL, Kim J, Bae J, Lee KS, Shim YM. Quantitative image variables reflect the intratumoral pathologic heterogeneity of lung adenocarcinoma. *Oncotarget.* 2016;7(41):67302–13.
38. Sala E, Mema E, Himoto Y, Veeraraghavan H, Brenton JD, Snyder A, Weigelt B, Vargas HA. Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging. *Clin Radiol.* 2017;72(1):3–10.
39. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature.* 2013;501(7467):338–45.
40. Liu J, Dang H, Wang XW. The significance of intertumor and intratumor heterogeneity in liver cancer. *Exp Mol Med.* 2018;50(1):e416.
41. Morris LG, Riaz N, Desrichard A, Şenbabaoğlu Y, Hakimi AA, Makarov V, Reis-Filho JS, Chan TA. Pan-cancer analysis of

- intratumor heterogeneity as a prognostic determinant of survival. *Oncotarget*. 2016;7(9):10051–63.
42. Sharma N, Gautam S K, Henry AA, Kumar A. Application of Big Data and Machine Learning. *Machine Learning and Big Data: Concepts, Algorithms, Tools and Applications*. 2020;305–333.
 43. Gondaliyaa C P, Patel A M, Parikh S M. A Comparative Study on Machine Learning Based Algorithms. *Proceedings of 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT)*. 2018;26–27.
 44. Sasanelli M, Meignan M, Haioun C, Berriolo-Riedinger A, Casanovas RO, Biggi A, Gallamini A, Siegel BA, Cashen AF, Véra P, et al. Pretherapy metabolic tumour volume is an independent predictor of outcome in patients with diffuse large B-cell lymphoma. *Eur J Nucl Med Mol Imaging*. 2014;41(11):2017–22.
 45. Mikhaeel NG, Smith D, Dunn JT, Phillips M, Møller H, Fields PA, Wrench D, Barrington SF. Combination of baseline metabolic tumour volume and early response on PET/CT improves progression-free survival prediction in DLBCL. *Eur J Nucl Med Mol Imaging*. 2016;43(7):1209–19.
 46. Zucca E, Cascione L, Ruberto T, Facchinelli D, Schär S, Hayoz S, Dirnhofer S, Giovanella L, Bargetzi M, Mamot C, et al. Prognostic models integrating quantitative parameters from baseline and interim positron emission computed tomography in patients with diffuse large B-cell lymphoma: post-hoc analysis from the SAKK38/07 clinical trial. *Hematol Oncol*. 2020;38(5):715–25.
 47. Miyazaki Y, Nawa Y, Miyagawa M, Kohashi S, Nakase K, Yasukawa M, Hara M. Maximum standard uptake value of 18F-fluorodeoxyglucose positron emission tomography is a prognostic factor for progression-free survival of newly diagnosed patients with diffuse large B cell lymphoma. *Ann Hematol*. 2013;92(2):239–44.
 48. Shagera QA, Cheon GJ, Koh Y, Yoo MY, Kang KW, Lee DS, Kim EE, Yoon SS, Chung JK. Prognostic value of metabolic tumour volume on baseline (18)F-FDG PET/CT in addition to NCCN-IPI in patients with diffuse large B-cell lymphoma: further stratification of the group with a high-risk NCCN-IPI. *Eur J Nucl Med Mol Imaging*. 2019;46(7):1417–27.
 49. Zhao P, Yu T, Pan Z. Prognostic value of the baseline 18F-FDG PET/CT metabolic tumour volume (MTV) and further stratification in low-intermediate (L-I) and high-intermediate (H-I) risk NCCNIPI subgroup by MTV in DLBCL MTV predict prognosis in DLBCL. *Ann Nucl Med*. 2021;35(1):24–30.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.