



# Development and validation of an $^{18}\text{F}$ -FDG PET radiomic model for prognosis prediction in patients with nasal-type extranodal natural killer/T cell lymphoma

Hongxi Wang<sup>1</sup> · Shengnan Zhao<sup>2</sup> · Li Li<sup>1</sup> · Rong Tian<sup>1</sup>

Received: 14 January 2020 / Revised: 2 April 2020 / Accepted: 7 May 2020 / Published online: 20 May 2020  
© European Society of Radiology 2020

## Abstract

**Objectives** To identify an  $^{18}\text{F}$ -fluorodeoxyglucose ( $^{18}\text{F}$ -FDG) positron emission tomography (PET) radiomics-based model for predicting progression-free survival (PFS) and overall survival (OS) of nasal-type extranodal natural killer/T cell lymphoma (ENKTL).

**Methods** In this retrospective study, a total of 110 ENKTL patients were divided into a training cohort ( $n = 82$ ) and a validation cohort ( $n = 28$ ). Forty-one features were extracted from pretreatment PET images of the patients. Least absolute shrinkage and selection operator (LASSO) regression was used to develop the radiomic signatures (R-signatures). A radiomics-based model was built and validated in the two cohorts and compared with a metabolism-based model.

**Results** The R-signatures were constructed with moderate predictive ability in the training and validation cohorts (R-signature<sub>PFS</sub>: AUC = 0.788 and 0.473; R-signature<sub>OS</sub>: AUC = 0.637 and 0.730). For PFS, the radiomics-based model showed better discrimination than the metabolism-based model in the training cohort (C-index = 0.811 vs. 0.751) but poorer discrimination in the validation cohort (C-index = 0.588 vs. 0.693). The calibration of the radiomics-based model was poorer than that of the metabolism-based model (training cohort:  $p = 0.415$  vs. 0.428, validation cohort:  $p = 0.228$  vs. 0.652). For OS, the performance of the radiomics-based model was poorer (training cohort: C-index = 0.818 vs. 0.828,  $p = 0.853$  vs. 0.885; validation cohort: C-index = 0.628 vs. 0.753,  $p < 0.05$  vs. 0.913).

**Conclusions** Radiomic features derived from PET images can predict the outcomes of patients with ENKTL, but the performance of the radiomics-based model was inferior to that of the metabolism-based model.

## Key Points

- The R-signatures calculated by using  $^{18}\text{F}$ -FDG PET radiomic features can predict the survival of patients with ENKTL.
- The radiomics-based models integrating the R-signatures and clinical factors achieved good predictive values.
- The performance of the radiomics-based model was inferior to that of the metabolism-based model in the two cohorts.

**Keywords** Lymphoma · Positron emission tomography · Prognosis

## Abbreviations

AUC	Area under curves
CI	Confidence interval
C-index	The Harrell concordance index
DICOM	Digital Imaging and Communications in Medicine
ECOG PS	Eastern Cooperative Oncology Group performance status
ICC	The intraclass correlation coefficients
IPI	International Prognostic Index
LASSO	The least absolute shrinkage and selection operator
LDH	Lactate dehydrogenase
MTV	Metabolic tumour volume

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00330-020-06943-1>) contains supplementary material, which is available to authorized users.

✉ Rong Tian  
rongtiannuclear@126.com

<sup>1</sup> Department of Nuclear Medicine, West China Hospital, Sichuan University, Chengdu, Sichuan, China

<sup>2</sup> Department of Oncology, West China Hospital, Sichuan University, No. 37, Lane Guoxue Wuhou District, Chengdu City, Sichuan Province, China

OS	Overall survival
PFS	Progression-free survival
ROIs	Regions of interest
SUV <sub>max</sub>	Standardised uptake value
TLG	Total lesion glycolysis

## Introduction

Extranodal natural killer/T cell lymphoma (ENKTL) is an aggressive malignancy of putative NK cell origin, with a minority deriving from the T cell lineage [1, 2]. ENKTL is much more common in Asia and Latin America, comprising approximately 3 to 10% of all lymphomas in East Asia, but less than 1% in Western countries. Establishing an optimal treatment strategy for ENKTL has yet to be determined, as accurate assessments are crucial for prognosis prediction and individualised treatment strategy decisions [3, 4].

<sup>18</sup>F-Fluorodeoxyglucose (<sup>18</sup>F-FDG) positron emission tomography-computed tomography (PET/CT) is routinely used for lesion detection, response monitoring, and prognostication assessments in ENKTL patients. The maximum standardised uptake value (SUV<sub>max</sub>) is the most widely used PET imaging parameter for prediction; moreover, metabolic tumour volume (MTV) and total lesion glycolysis (TLG), which are based on both SUV and tumour volume, have also been reported to be significant prognostic biomarkers for ENKTL [5, 6]. However, these metabolic parameters do not fully reflect the spatial distribution of a tracer, which has been suggested to correlate with intra- and intertumoural heterogeneity, and even worse prognosis and survival [7]. Over the past decade, an emerging and promising field, radiomics, has been implemented to extract and analyse a large amount of advanced quantitative imaging features with high throughput from medical images to provide abundant tumour heterogeneity information [8, 9]. The most widely used medical imaging modalities in radiomic research are computed tomography (CT) and magnetic resonance imaging (MRI). Currently, radiomic features extracted from PET are considered to contain large amounts of underlying information distinct from that provided by CT and MRI [10, 11]. A continuously increasing number of studies have reported promising results regarding the value of PET-based radiomic features for various types of solid tumours [12–16]. Nevertheless, few published studies have reported the diagnostic, staging, or prognostic value of PET-based radiomic features in lymphoma [17–21]. In addition, combining multiple imaging biomarkers as a predictive signature using radiomic methods, rather than individual analyses, is a promising and useful approach for prognosis prediction and clinical management [22].

Thus, in this retrospective study, we aimed to develop the radiomic signatures (R-signatures) using <sup>18</sup>F-FDG PET radiomic features for the prediction of progression-free

survival (PFS) and overall survival (OS) in patients with ENKTL. Subsequently, the radiomics-based model integrating the R-signatures and clinical factors was established and then validated and compared with the metabolism-based model integrating metabolic parameters and clinical factors. The models are visualised as nomograms.

## Materials and methods

### Patients and follow-up

Ethical approval was obtained, and the requirement for informed consent from patients was waived. The inclusion criteria included as follows: (a) pathologically diagnosed ENKTL between January 2011 and January 2017, (b) pretreatment <sup>18</sup>F-FDG PET/CT, and (c) stage I to II patients who received 2 cycles of chemotherapy with LVP (L-asparaginase, vincristine, prednisone) and concurrent radiotherapy with 2 cycles of cisplatin chemotherapy, followed by 2 cycles of LVP; stage III to IV patients who received chemotherapy with LVP with or without radiotherapy if a response was observed, and otherwise received second-line therapy. In total, 110 consecutive patients were enrolled and randomly allocated to two cohorts (82 and 28 patients in the training and validation cohorts, respectively). Follow-up was performed every 3 months after the completion of treatment. The last follow-up was conducted in April 2018. The end-points of this study were PFS and OS. PFS was defined as the interval between the date of diagnosis and the date of the first relapse, progression, or death. OS was defined as the interval between the date of diagnosis and the time of death.

### PET scanner and acquisition parameters

<sup>18</sup>F-FDG PET/CT examinations were performed on a Gemini GXL PET/CT scanner equipped with a 16-slice CT (Philips Medical System). After at least 6 h of fasting, 190–375 MBq of <sup>18</sup>F-FDG was administered intravenously. The blood glucose level was controlled to be lower than 8.0 mmol/L. Whole-body PET/CT scans were started 60 min after radiopharmaceutical injection. Emission data were acquired for 2 min per bed position. PET emission acquisition was performed in 3D mode (3D-RAMLA); the dimensions of the in-plane matrix were 4 mm × 4 mm, and the slice thickness was 4 mm. All examinations were reconstructed using an OSEM algorithm, and the CT acquisition data were used for attenuation correction.

### Segmentation and feature extraction

Two nuclear medicine physicians independently performed segmentation using the Local Image Features Extraction

(LifeX) package (version 4.00, <http://www.lifexsoft.org>) [23]. First, the lymphoma lesions were delineated manually. Then, the regions of interest (ROIs) were defined based on a threshold of 40% of the  $SUV_{max}$  of the defined lesions [24], and spatial resampling ( $2 \times 2 \times 2$  mm), absolute intensity resampling (0–20), and intensity discretisation (number of grey levels = 64, size of bins = 0.3125) were performed [25]. A total of 41 features (Supplementary Materials) of the ROIs were extracted as follows: first-order metrics extracted from the histogram and shape; features derived from the grey-level co-occurrence matrix (GLCM), the neighbourhood grey-level different matrix (NGLDM), the grey-level run length matrix (GLRLM), and the grey-level zone length matrix (GLZLM); and conventional metabolic parameters, including the  $SUV_{max}$ , MTV, and TLG.

### Radiomic feature selection and model building

First, the interobserver repeatability of the segmentation was evaluated using the intraclass correlation coefficient (ICC) method, and the features with an ICC greater than 0.70 were selected [26]. Then, the least absolute shrinkage and selection operator (LASSO) Cox regression algorithm was applied to the selected features [27]. Cross-validation was applied to optimise the value of  $\lambda$ , the coefficients of indistinctive covariates were reduced to zero, and the remaining nonzero coefficients were selected. The nonzero coefficients of the selected features were defined as radiomic scores (R-scores). We calculated the combination of R-scores for all selected radiomic features, defined as the R-signature. We determined the optimal threshold value of the R-signature by the receiver operating characteristic (ROC) curve and divided patients into high- and low-risk groups. The potential association of the R-signature with PFS and OS was evaluated using the Kaplan-Meier analysis and log-rank test.

We used the univariate Cox regression to select the significant prognostic factors of PFS and OS, respectively. R-signature and clinical variables were entered into a multivariate Cox regression to build the radiomics-based model for PFS and OS prediction. Likewise, metabolic parameters and clinical variables were entered into a multivariate Cox regression to build the metabolism-based model. Models were then visualised as nomograms. The flowchart of model building is presented in Fig. 1.

### Model validation

The discrimination of the models was assessed using the Harrell concordance index (C-index) [28, 29]. Bootstrap analyses with 1000 resamples were used to obtain a corrected C-index. The calibration of the models was assessed by Hosmer-Lemeshow tests and calibration

curves, and  $p > 0.05$  accounted for a nonsignificant deviance from the theoretical perfect calibration [30].

### Statistical analysis

Statistical analyses were performed using IBM SPSS Statistics (version 19.0, IBM Corp) and R software (version 3.4.2; <http://www.R-project.org>). All tests were two-sided, and  $p$  values of  $< 0.05$  were considered significant.

### Results

The clinical characteristics of the patients are summarised in Table 1. No differences were found between the training and validation cohorts ( $p = 0.149$ – $0.945$ ). The median follow-up time was 33 months (2–90 months). As of the final follow-up, the numbers of survivals and deaths were 63 (57.3%) and 25 (22.7%), respectively.

### R-signature construction and assessment

Six radiomic features had low ICCs ( $ICC < 0.70$ ), and twenty-seven radiomic features had good repeatability ( $ICC \geq 0.85$ ). The thirty-two features with  $ICCs \geq 0.70$  extracted in the second round were selected for further analysis. According to the LASSO results (Fig. 2), we obtained 4 and 3 radiomic features with nonzero coefficients for PFS and OS, respectively, in order to calculate the  $R\text{-signature}_{PFS}$  and the  $signature_{OS}$ . The results of the ICC analysis and the formulas for the R-signatures can be found in the Supplementary Materials.

The ROC-AUCs of the  $R\text{-signature}_{PFS}$  were 0.788 (95% CI = 0.682–0.895) and 0.473 ( $p = 0.803$ ) in the two cohorts. When analysing the association of the  $R\text{-signature}_{PFS}$  with PFS, the results of the log-rank test indicated significant discrimination between the high- and low-risk groups in the training cohort (Fig. 3a), but no discrimination in the validation cohort (Fig. 3c). The ROC-AUCs of the  $R\text{-signature}_{OS}$  were 0.637 (95% CI = 0.488–0.786) and 0.730 (95% CI = 0.548–0.912) in the two cohorts. We also observed that the  $R\text{-signature}_{OS}$  was significant for classifying the patients into high- and low-risk groups in the two cohorts (Fig. 3b, d).

### Building of the radiomics-based model

The results of the univariate and multivariate analyses are listed in Tables 2 and 3. In the univariate analysis, the Eastern Cooperative Oncology Group performance status (ECOG PS), Ann Arbor stage, lactate dehydrogenase (LDH), International Prognostic Index (IPI),  $SUV_{max}$ , MTV, TLG, and the  $R\text{-signature}_{PFS}$  were associated with

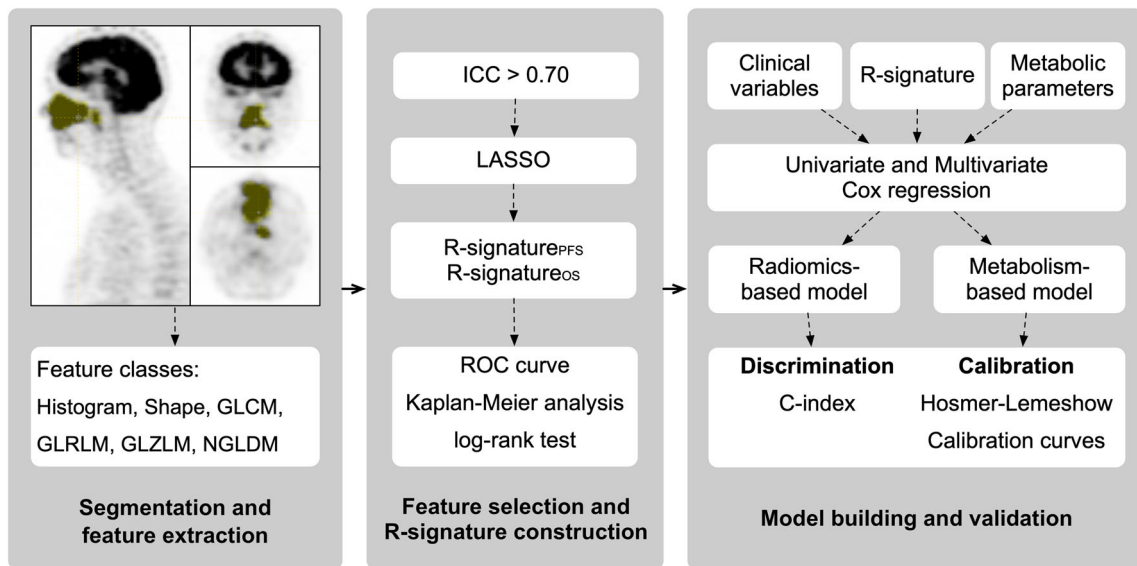


Fig. 1 Flowchart showing the development of the models

PFS; bone marrow (BM), ECOG PS, Ann Arbor stage, LDH, IPI,  $SUV_{max}$ , MTV, TLG, and the  $R\text{-signature}_{OS}$  were associated with OS.

For PFS, the  $R\text{-signature}_{PFS}$  and significant clinical variables in the univariate analysis were selected for inclusion in the multivariate Cox regression. The  $R\text{-}$

$signature_{PFS}$  and IPI remained as prognostic factors in the multivariate analysis and were used to build the radiomics-based model (Fig. 4a). For OS, the radiomics-based model was built using the  $R\text{-signature}_{OS}$  and ECOG PS (Fig. 4b), which were independent prognostic factors of OS identified by the multivariate analysis.

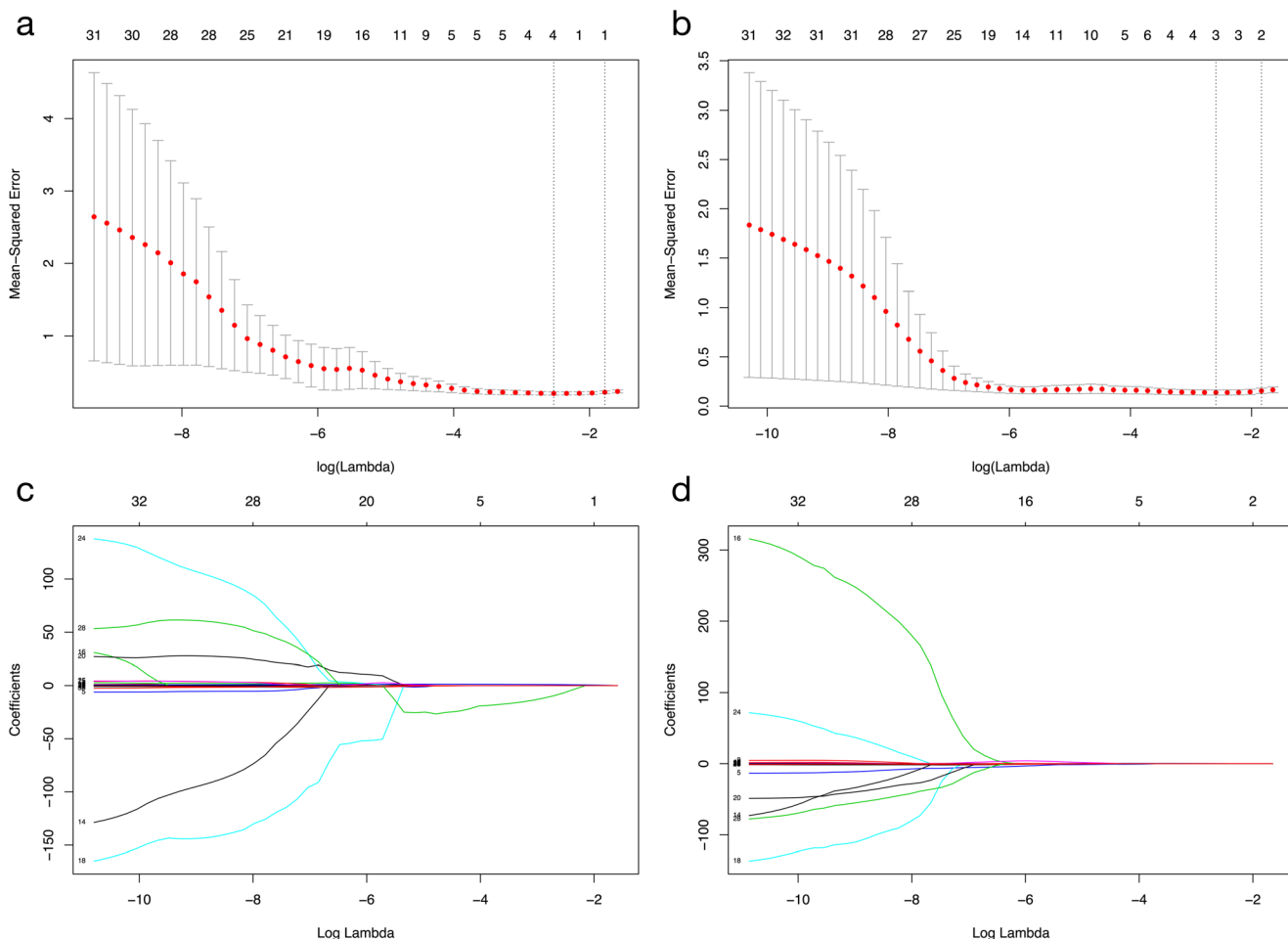
Table 1 Baseline characteristics of the patients in the training and the validation cohorts

Characteristic	No. of patients			p value
	Overall (n = 110)	Training (n = 82)	Validation (n = 28)	
Age (years) <sup>a</sup>	45.59 ± 14.55	45.25 ± 15.19	46.33 ± 12.64	0.420 <sup>b</sup>
Male/female	69 (62.7)/31 (37.3)	52 (63.4)/30 (36.6)	17 (60.7)/11 (39.3)	0.799
BS	58 (52.7)	45 (54.9)	13 (46.4)	0.439
BM	5 (4.5)	4 (4.9)	1 (3.6)	1.000
ECOG PS				0.149
0–1	98 (89.1)	71 (86.6)	27 (96.4)	
≥ 2	12 (10.9)	11 (13.4)	1 (3.6)	
Ann Arbor stage				0.925
I–II	90 (81.8)	66 (80.5)	24 (85.7)	
III–IV	20 (18.2)	16 (19.5)	4 (14.3)	
LDH > 245 U/L	45 (40.9)	32 (39)	13 (46.4)	0.491
IPI score				0.945
0–1	86 (78.2)	64 (78)	22 (78.6)	
≥ 2	24 (21.8)	18 (22)	6 (21.4)	
EBV positive	84 (76.4)	61 (74.4)	23 (82.1)	0.750
Radiotherapy	96 (87.3)	70 (85.4)	26 (92.9)	0.880

BS B symptoms, BM bone marrow infiltration, ECOG PS Eastern Cooperative Oncology Group performance status, IPI International Prognostic Index, LDH lactate dehydrogenase, EBV Epstein-Barr virus. Values in parentheses are percentages unless indicated otherwise

<sup>a</sup> Values are mean ± SD

<sup>b</sup> p value derived from two-sample t test. The other p value was derived from the  $\chi^2$  test



**Fig. 2** Feature selection for the prediction using the LASSO model, tuning parameter ( $\lambda$ ) selection in the LASSO model involved the use of tenfold cross-validation with minimum criteria ((a) PFS and (b) OS); coefficient profiles of the radiomics features ((c) PFS and (d) OS)

### Building of the metabolism-based model

For PFS, the significant clinical variables and metabolic parameters in the univariate analysis were entered into the multivariate Cox regression, and MTV and IPI were identified as independent prognostic factors and were used to build the metabolism-based model. Likewise, the metabolism-based model for OS prediction was built using the  $SUV_{max}$ , MTV, and ECOG PS.

### Validation and comparison of the model

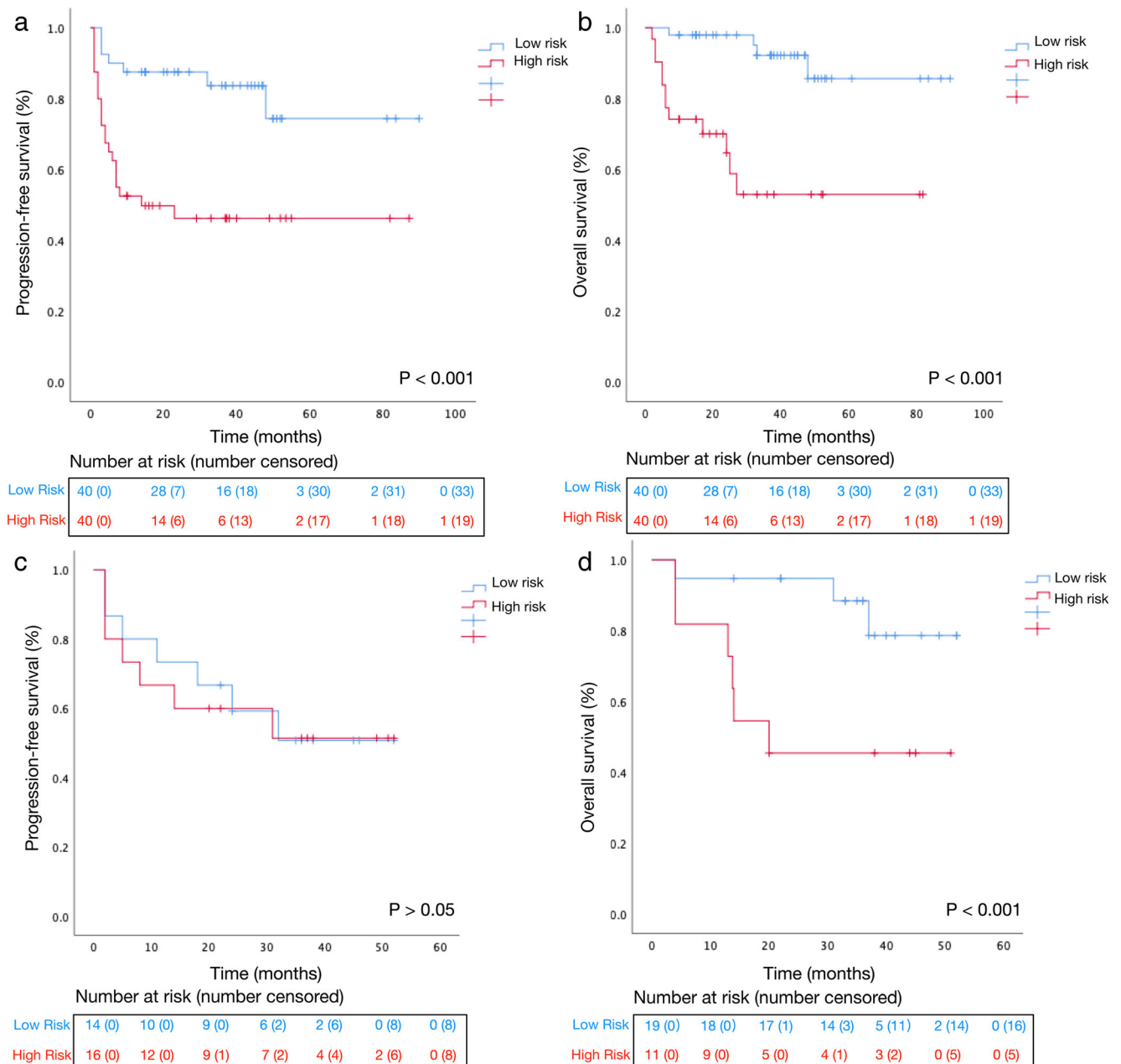
The results of the C-index and the Hosmer-Lemeshow test are shown in Table 4. The calibration curves of the models are shown in the [Supplementary Materials](#). For PFS prediction, the radiomics-based model showed better discrimination than the metabolism-based model in the training cohort (C-index = 0.811 vs. 0.751) but poorer discrimination in the validation cohort (C-index = 0.588 vs. 0.693). The Hosmer-Lemeshow test showed that the calibration of the radiomics-based model was poorer than that of the metabolism-based model (training

cohort:  $p = 0.415$  vs. 0.428; validation cohort:  $p = 0.228$  vs. 0.652). The calibration curves showed that the calibration of the two models was better in the training cohort than that in the validation cohort (Supplementary Figs. S1 and S2).

For OS prediction, the discrimination of the radiomics-based model was poorer than that of the metabolism-based model in the training cohort (C-index = 0.818 vs. 0.828) and the validation cohort (C-index = 0.628 vs. 0.753). The Hosmer-Lemeshow test indicated that the calibration of the radiomics-based model was poorer than that of the metabolism-based model (training cohort:  $p = 0.853$  vs. 0.885; validation cohort:  $p < 0.05$  vs. 0.913). According to the calibration curves, the calibration of the two models was better in the training cohort than that in the validation cohort (Supplementary Figs. S3 and S4).

### Discussion

In the present study, we developed R-signatures with moderate predictive ability in a training cohort and a validation



**Fig. 3** The Kaplan-Meier survival curves (**a** R-signature<sub>PFS</sub>, training cohort; **b** R-signature<sub>OS</sub>, training cohort; **c** R-signature<sub>PFS</sub>, validation cohort; and **d** R-signature<sub>OS</sub>, validation cohort)

cohort (R-signature<sub>PFS</sub>: AUC = 0.788 and 0.473, respectively; R-signature<sub>OS</sub>: AUC = 0.637 and 0.730, respectively). Although no significant association was found between the R-signature<sub>PFS</sub> and PFS in the validation cohort, the R-signatures were associated with PFS in the training cohort and with OS in both cohorts. These results provide evidence that radiomic features extracted from pretreatment <sup>18</sup>F-FDG PET images can predict lymphoma outcomes. We further developed radiomics-based models combining R-signatures and clinical variables to predict PFS and OS among patients with ENKTL, and the models were then visualised via nomograms with the aim of identifying patients at a high risk of early

progression and death who could be offered an alternative treatment strategy. The radiomics-based models achieved good predictive values (PFS: C-index = 0.811 and 0.588, *p* = 0.415 and 0.228; OS: C-index = 0.818 and 0.628, *p* = 0.853 and < 0.05). However, the performance of the radiomics-based model was inferior to that of the metabolism-based model in the two cohorts.

The main goal of radiomics is to build a prediction model for clinical outcomes using selected radiomic features, and integrating radiomic features with traditional prognostic indicators (clinical indicators) in one model can improve the prediction performance of a single prognostic indicator [8, 9, 31,

**Table 2** The results of the univariate Cox regression analysis

Variables	PFS		OS	
	HR (95% CI)	<i>p</i> value	HR (95% CI)	<i>p</i> value
<b>Clinical variables</b>				
Age ( $\leq 60$ / $> 60$ )	NS	0.587	NS	0.465
Gender (male/female)	NS	0.474	NS	0.517
BS (no/yes)	NS	0.142	NS	0.099
BM (no/yes)	NS	0.152	5.648 (1.264–25.238)	0.023
ECOG PS (0–1/ $\geq 2$ )	4.674 (2.013–10.853)	$< 0.001$	4.645 (1.722–12.531)	0.002
Ann Arbor stage (I–II/III–IV)	1.950 (1.466–2.595)	$< 0.001$	8.422 (2.871–24.704)	$< 0.001$
LDH ( $\leq 245$ / $> 245$ U/L)	3.736 (1.717–8.130)	0.001	4.226 (1.463–12.203)	0.008
IPI (0–1/ $\geq 2$ )	5.231 (2.459–11.128)	$< 0.001$	1.800 (1.353–2.394)	$< 0.001$
EB DNA (negative/positive)	NS	0.538	NS	0.350
<b>Metabolic parameters</b>				
SUV <sub>max</sub> ( $\leq 11.05$ / $> 11.05$ )	2.366 (1.089–5.137)	0.030	2.924 (1.013–8.442)	0.047
MTV ( $\leq 22.4$ / $> 22.4$ mL)	4.333 (1.949–9.623)	$< 0.001$	6.437 (2.055–20.163)	0.001
TLG ( $\leq 134.5$ / $> 134.5$ )	4.685 (2.046–10.730)	$< 0.001$	5.927 (1.890–18.589)	0.002
<b>R-signature</b>				
R-signature <sub>PFS</sub> ( $\leq 0.34$ / $> 0.34$ )	3.971 (1.683–9.368)	0.002	NA	NA
R-signature <sub>OS</sub> ( $\leq 0.06$ / $> 0.06$ )	NA	NA	6.119 (1.727–20.163)	0.005

NS not significant, NA not apply

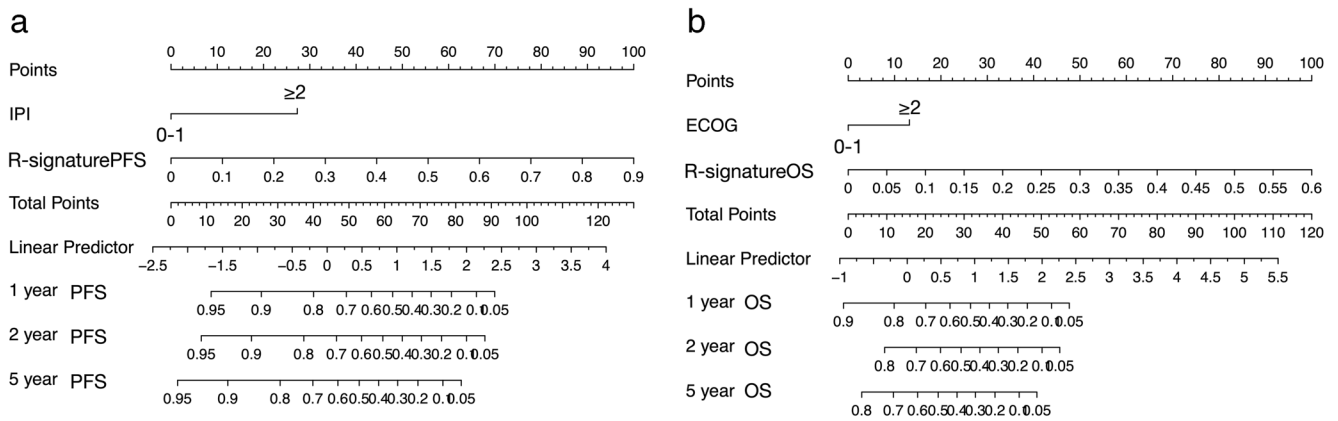
32]. Traditional PET metabolic parameters have been confirmed to be significant prognostic indicators for outcomes of patients with ENKTL and are widely used in clinical management [4, 5]. However, most radiomic indices extracted from PET images have been considered to be significantly correlated with metabolic parameters, especially MTV and TLG [33, 34], and such correlations are postulated to be inherent to the definitions of the features as opposed to being variable depending on the tumour type or acquisition/reconstruction protocols [11], which were also reported to differ in susceptibility to methodological, biological, and

metabolic features [35, 36]. Based on these considerations, we built radiomics-based models to investigate the potential added prognostic value of radiomic features in ENKTL patients by comparing the models with metabolism-based models. Our results suggested that the radiomics-based models may provide limited prognostic information for ENKTL patients compared with the metabolism-based model.

Our findings are in line with those of recent studies indicating that the performance of PET-derived radiomic features for tumour prognosis prediction is poor compared with that of PET-derived metabolic parameters. In a cohort of 82 patients

**Table 3** The results of the multivariate Cox regression analysis

Variables	PFS				OS			
	Metabolism-based model		Radiomics-based model		Metabolism-based model		Radiomics-based model	
	HR (95% CI)	<i>p</i> value	HR (95% CI)	<i>p</i> value	HR (95% CI)	<i>p</i> value	HR (95% CI)	<i>p</i> value
Ann Arbor stage		0.286		0.444		0.151		0.061
ECOG PS		0.539		0.632	3.612 (1.118–11.664)	0.032	4.930 (1.652–14.956)	0.005
LDH		0.167		0.133		0.183		0.266
IPI	4.667 (2.152–10.112)	$< 0.001$	4.199 (1.729–9.815)	$< 0.001$		0.936		0.574
BM	NA	NA	NA	NA		0.046		0.228
SUV <sub>max</sub>		0.200	NA	NA	3.231 (1.067–9.785)	0.038	NA	NA
MTV	3.885 (1.722–8.762)	0.001	NA	NA	5.161 (1.427–18.666)	0.012	NA	NA
TLG		0.259	NA	NA		0.891	NA	NA
R-signature	NA	NA	5.505 (2.536–11.951)	0.001	NA	NA	4.279 (1.133–16.167)	0.032



**Fig. 4** The nomograms of the radiomics-based models (a) for PFS and (b) for OS

with aggressive B cell lymphoma, MTV was correlated with the response to therapy, but texture features could not predict the therapy response, although several features were correlated with the presence of a residual mass and outcomes [18]. Rogasch et al reported that the asphericity feature could predict the response after chemotherapy in 50 children with paediatric Hodgkin’s lymphoma (HL), although MTV showed a better performance [35]. In contrast, several studies have reported promising results for the utilisation of radiomic features derived from PET to risk-stratify patients with lymphoma. Lue et al found that the intensity nonuniformity of pretreatment PET was a prognostic indicator in 42 patients with HL and may outperform MTV [31]. In a cohort of 17 patients with ENKTL, texture features (dissimilarity and LISZE) extracted from pretreatment PET images were independent predictors of PFS, whereas the  $SUV_{max}$ , MTV, and TLG were not associated with PFS [37]. Wang et al demonstrated a relationship between PET radiomic features and OS in 19 patients with renal/adrenal lymphoma, and MTV was not an independent factor [38].

These differences may be attributed to many sources. First, as mentioned above, several studies incorporated radiomic and metabolic features into one model, which may result in an underlying risk of redundancy and underestimation of the performance of both types of features. Second, a sample size of at least 10 to 15 patients per predictor variable has been

proposed to be required to produce valid estimates for multiple regression models [39]. We reduced the number of radiomic features to 4 and 3, which is reasonable for minimising false detection rates. Previous studies with small sample sizes and datasets may have a risk of bias. Additionally, many variables may affect the stability and prognostic value of PET-based radiomic features, such as the scanner, segmentation, reconstruction parameters, formulas used to define the radiomic features, and software [40, 41]. Heterogeneity among different studies may have contributed to controversial results. Lastly, validation analysis has been regarded as an indispensable step in radiomics research to show the potential value of a radiomics model for clinical application [8, 42]. In this study, the performance of the models in the internal validation cohort was inferior to that in the training cohort, suggesting that the stability of the models should be considered with caution. Comparing previous findings directly is difficult considering the limited number of lymphoma radiomics studies that have performed internal and/or external validation.

Segmentation is an important step in radiomics research as edges can substantially affect feature values. No consensus on optimal segmentation is available for lymphoma because lymphoma lesions usually have heterogeneous sizes, shapes, and locations. On PET images, some normal organs with high uptake and physiological FDG uptake and excretion may

**Table 4** The results of the C-index and the Hosmer-Lemeshow test

	Training cohort		Validation cohort	
	C-index (95% CI)	<i>p</i> value <sup>a</sup>	C-index (95% CI)	<i>p</i> value <sup>a</sup>
<b>PFS</b>				
Radiomics-based model	0.811 (0.733–0.888)	0.415	0.588 (0.363–0.812)	0.228
Metabolism-based model	0.751 (0.650–0.851)	0.428	0.693 (0.495–0.892)	0.652
<b>OS</b>				
Radiomics-based model	0.818 (0.693–0.944)	0.853	0.628 (0.462–0.796)	< 0.05
Metabolism-based model	0.828 (0.735–0.921)	0.885	0.753 (0.598–0.907)	0.913

<sup>a</sup>*p* value for the Hosmer-Lemeshow test



cause confusion [43]. Therefore, identifying accurate and robust segmentation methods for lymphoma is important. We defined ROIs using a semiautomatic threshold-based method (the 40% threshold segmentation method) and assessed the interobserver repeatability of the segmentation, and the results of the ICC analysis indicated that 27 of 38 radiomic features had good repeatability ( $ICC \geq 0.85$ ), which is consistent with previous studies demonstrating that most PET image features exhibited high stability in test-retest and interobserver analyses [40, 44]. Several different segmentation methods have been devised for lymphoma [43, 45, 46]. Hu et al proposed an entropy-based optimisation strategy to detect and segment lymphoma in PET images and reported a good performance [43]. Hu et al proposed an automatic approach for ENKTL segmentation that was more stable than traditional deep-learning segmentation [46].

The limitations of the present study are as follows. First, this was a single-centre, retrospective study, and external validation was not performed, which may have impacted patient selection, the examination protocol, and the radiomic quantification results. Second, we used only  $^{18}\text{F}$ -FDG PET images to extract radiomic features. PET images have relatively low spatial resolution and high noise, which may influence lesion identification [9]. A combination of PET and CT images may expand the feature pool and lead to the discovery of more predictive radiomic features. In addition, we extracted only traditional radiomic features, which may contain limited information. Other types of features (model- or transform-based features and deep-learning features) have gained popularity because they are more specific to data and clinical outcomes [47], and such features should be further investigated to better understand the predictive power of radiomics.

## Conclusions

In conclusion, a pretreatment  $^{18}\text{F}$ -FDG PET radiomics-based model was designed and showed significant stratification power in predicting PFS and OS in ENKTL, but the performance was inferior to that of the metabolism-based model. Therefore, further multi-centre, prospective studies with external validation are required to ensure that the results are reproducible and do not require refinement to achieve a higher level of evidence.

**Funding information** This study has received funding by the Key Projects of the Ministry of Science and Technology (grant 2017YFC0113304).

## Compliance with ethical standards

**Guarantor** The scientific guarantor of this publication is Rong Tian, PhD.

**Conflict of interest** The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

**Statistics and biometry** One of the authors has significant statistical expertise. No complex statistical methods were necessary for this paper.

**Informed consent** Written informed consent was waived by the Institutional Review Board because this was a retrospective study.

**Ethical approval** The study was approved by the institutional ethics review board of the West China Hospital, Sichuan University.

## Methodology

- Retrospective
- Diagnostic or prognostic study
- Performed at one institution

## References

1. Tse E, Kwong YL (2017) The diagnosis and management of NK/T-cell lymphomas. *J Hematol Oncol* 10:85
2. William BM, Armitage JO (2013) International analysis of the frequency and outcomes of NK/T-cell lymphomas. *Best Pract Res Clin Haematol* 26:23–32
3. Yamaguchi M, Miyazaki K (2017) Current treatment approaches for NK/T-cell lymphoma. *J Clin Exp Hematop* 57:98–108
4. Ding JJ, Chen YL, Zhou SH, Zhao K (2018) Positron emission tomography/computed tomography in the diagnosis, staging, and prognostic evaluation of natural killer/T-cell lymphoma. *J Int Med Res* 46:4920–4929
5. Wang H, Shen G, Jiang C, Li L, Cui F, Tian R (2018) Prognostic value of baseline, interim and end-of-treatment  $^{18}\text{F}$ -FDG PET/CT parameters in extranodal natural killer/T-cell lymphoma: a meta-analysis. *PLoS One* 13:1–14
6. Jiang C, Zhang X, Jiang M et al (2015) Assessment of the prognostic capacity of pretreatment, interim, and post-therapy ( $^{18}\text{F}$ -FDG PET/CT in extranodal natural killer/T-cell lymphoma, nasal type. *Ann Nucl Med* 29:442–451
7. Yan J, Chu-Shern JL, Loi HY et al (2015) Impact of image reconstruction settings on texture features in  $^{18}\text{F}$ -FDG PET. *J Nucl Med* 56:1667–1673
8. Lambin P, Leijenaar RTH, Deist TM et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 14:749–762
9. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D (2017) Characterization of PET/CT images using texture analysis: the past, the present... any future? *Eur J Nucl Med Mol Imaging* 44: 151–165
10. Davnall F, Yip CSP, Ljungqvist G et al (2012) Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? *Insights Imaging* 3:573–589
11. Orlhac F, Soussan M, Maisonobe JA, Garcia CA, Vanderlinden B, Buvat I (2014) Tumor texture analysis in  $^{18}\text{F}$ -FDG PET: relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. *J Nucl Med* 55:414–422
12. Lee JW, Lee SM (2018) Radiomics in oncological PET/CT: clinical applications. *Nucl Med Mol Imaging* 52:170–189
13. Carvalho S, Leijenaar RTH, Troost EGC et al (2018)  $^{18}\text{F}$ -fluorodeoxyglucose positron-emission tomography (FDG-PET)-Radiomics of metastatic lymph nodes and primary tumor in non-

- small cell lung cancer (NSCLC) - a prospective externally validated study. *PLoS One* 13:e0192859
14. Mena E, Sheikhabaehi S, Taghipour M et al (2017) 18F-FDG PET/CT metabolic tumor volume and intratumoral heterogeneity in pancreatic adenocarcinomas: impact of dual-time point and segmentation methods. *Clin Nucl Med* 42:e16–e21
  15. Nakajo M, Jinguji M, Nakabeppu Y et al (2017) Texture analysis of 18F-FDG PET/CT to predict tumour response and prognosis of patients with esophageal cancer treated by chemoradiotherapy. *Eur J Nucl Med Mol Imaging* 44:206–214
  16. Feliciani G, Fioroni F, Grassi E et al (2018) Radiomic profiling of head and neck cancer: 18F-FDG PET texture analysis as predictor of patient survival. *Contrast Media Mol Imaging* 2018:3574310
  17. Aide N, Talbot M, Fruchart C, Damaj G, Lasnon C (2018) Diagnostic and prognostic value of baseline FDG PET/CT skeletal textural features in diffuse large B cell lymphoma. *Eur J Nucl Med Mol Imaging* 45:699–711
  18. Parvez A, Tau N, Hussey D, Maganti M, Metser U (2018) 18F-FDG PET/CT metabolic tumor parameters and radiomics features in aggressive non-Hodgkin's lymphoma as predictors of treatment outcome and survival. *Ann Nucl Med* 32:410–416
  19. Zhou Y, Ma XL, Pu LT, Zhou RF, Ou XJ, Tian R (2019) Prediction of overall survival and progression-free survival by the 18F-FDG PET/CT radiomic features in patients with primary gastric diffuse large B-cell lymphoma. *Contrast Media Mol Imaging* 2019:5963607
  20. Ganeshan B, Miles KA, Babikir S et al (2017) CT-based texture analysis potentially provides prognostic information complementary to interim FDG-PET for patients with Hodgkin's and aggressive non-Hodgkin's lymphomas. *Eur Radiol* 27:1012–1020
  21. Kang D, Park JE, Kim YH et al (2018) Diffusion radiomics as a diagnostic modal for atypical manifestation of primary central nervous system lymphoma: development and multicenter external validation. *Neuro Oncol* 20:1251–1261
  22. Wang L, Dong T, Xin B et al (2019) Integrative nomogram of CT imaging, clinical, and hematological features for survival prediction of patients with locally advanced non-small cell lung cancer. *Eur Radiol* 29:2958–2967
  23. Nioche C, Orhac F, Boughdad S et al (2018) Lifex: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res* 78:4786–4789
  24. Parmar C, Velazquez ER, Leijenaar R et al (2014) Robust radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS One* 9:e012107
  25. Orhac F, Nioche C, Soussan M, Buvat I (2017) Understanding changes in tumor texture indices in PET: a comparison between visual assessment and index values in simulated and patient data. *J Nucl Med* 58:387–392
  26. Weir JP (2005) Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 19:231
  27. Sauerbrei W, Royston P, Binder H (2007) Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med* 26:5512–5528
  28. Alba AC, Agoritsas T, Walsh M et al (2017) Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *JAMA* 318:1377–1384
  29. Zieffler AS, Harring JR, Long JD (2011) Comparing groups. Wiley, Hoboken
  30. Pencina MJ, D'Agostino RB, Steyerberg EW (2011) Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 30:11–21
  31. Lue KH, Wu YF, Liu SH et al (2019) Prognostic value of pretreatment radiomic features of 18F-FDG PET in patients with Hodgkin lymphoma. *Clin Nucl Med* 44:e559–e565
  32. Desseroit MC, Visvikis D, Tixier F et al (2016) Development of a nomogram combining clinical staging with 18F-FDG PET/CT image features in non-small-cell lung cancer stage I–III. *Eur J Nucl Med Mol Imaging* 43:1477–1485
  33. Brooks FJ, Grigsby PW (2014) The effect of small tumor volumes upon intra-tumoral tracer uptake heterogeneity studies. *J Nucl Med* 55:37–42
  34. Ha S, Choi H, Paeng JC, Cheon GJ (2019) Radiomics in oncological PET/CT: a methodological overview. *Nucl Med Mol Imaging* 53:14–29
  35. Rogasch JMM, Hundsdoerfer P, Hofheinz F et al (2018) Pretherapeutic FDG-PET total metabolic tumor volume predicts response to induction therapy in pediatric Hodgkin's lymphoma. *BMC Cancer* 18:1–9
  36. Ben Bouallègue F, Tabaa YA, Kafrouni M, Cartron G, Vauchot F, Mariano-Goulart D (2017) Association between textural and morphological tumor indices on baseline PET-CT and early metabolic response on interim PET-CT in bulky malignant lymphomas. *Med Phys* 44:4608–4619
  37. Ko KY, Liu CJ, Ko CL, Yen RF (2016) Intratumoral heterogeneity of pretreatment 18F-FDG PET images predict disease progression in patients with nasal type extranodal natural killer/T-cell lymphoma. *Clin Nucl Med* 41:922–926
  38. Wang M, Xu H, Xiao L, Song W, Zhu S, Ma X (2019) Prognostic value of functional parameters of 18 F-FDG-PET images in patients with primary renal/adrenal lymphoma. *Contrast Media Mol Imaging* 2019:1–9
  39. Chalkidou A, O'Doherty MJ, Marsden PK (2015) False discovery rates in PET and CT studies with texture features: a systematic review. *PLoS One* 10:1–18
  40. Leijenaar RT, Carvalho S, Velazquez ER et al (2015) Stability of FDG-PET radiomics features: an integrated analysis of test-retest and inter-observer variability. *Acta Oncol* 52:1391–1397
  41. Zwanenburg A (2019) Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. *Eur J Nucl Med Mol Imaging* 46:2638–2655
  42. Liu Z, Wang S, Dong D et al (2019) The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. *Theranostics* 9:1303–1322
  43. Hu H, Decazes P, Vera P, Li H, Ruan S (2019) Detection and segmentation of lymphomas in 3D PET images via clustering with entropy-based optimization strategy. *Int J Comput Assist Radiol Surg* 14:1715–1724
  44. Van Velden FHP, Nissen IA, Jongsma F et al (2014) Test-retest variability of various quantitative measures to characterize tracer uptake and/or tracer uptake heterogeneity in metastasized liver for patients with colorectal carcinoma. *Mol Imaging Biol* 16:13–18
  45. Camacho MR, Etchebehere E, Tardelli N, et al (2020) Validation of a multi-foci segmentation method for measuring metabolic tumor volume in Hodgkin's lymphoma. *J Nucl Med Technol* 48:30–35
  46. Hu X, Guo R, Chen J et al (2020) Coarse-to-fine adversarial networks and zone-based uncertainty analysis for NK/T-cell lymphoma segmentation in CT/PET images. *IEEE J Biomed Health Inform.* <https://doi.org/10.1109/JBHI.2020.2972694>
  47. Koçak B, Durmaz EŞ, Ateş E, Kılıçkesmez Ö (2019) Radiomics with artificial intelligence: a practical guide for beginners. *Diagn Interv Radiol* 25:485–495