Journal of Cancer Research and Clinical Oncology (2024) 150:452
https://doi.org/10.1007/s00432-024-05905-0

RESEARCH

Survival prediction in diffuse large B-cell lymphoma patients: multimodal PET/CT deep features radiomic model utilizing automated machine learning

Jianxin Chen[1] · Fengyi Lin[1] · Zhaoyan Dai[1] · Yu Chen[1] · Yawen Fan[1] · Ang Li[1] · Chenyu Zhao[1]

Received: 12 June 2024 / Accepted: 21 July 2024 / Published online: 9 October 2024

Abstract

Purpose We sought to develop an effective combined model for predicting the survival of patients with diffuse large B-cell lymphoma (DLBCL) based on the multimodal PET-CT deep features radiomics signature (DFR-signature).

Methods 369 DLBCL patients from two medical centers were included in this study. Their PET and CT images were fused to construct the multimodal PET-CT images using a deep learning fusion network. Then the deep features were extracted from those fused PET-CT images, and the DFR-signature was constructed through an Automated machine learning (AutoML) model. Combined with clinical indexes from the Cox regression analysis, we constructed a combined model to predict the progression-free survival (PFS) and the overall survival (OS) of patients. In addition, the combined model was evaluated in the concordance index (C-index) and the time-dependent area under the ROC curve (tdAUC).

Results A total of 1000 deep features were extracted to build a DFR-signature. Besides the DFR-signature, the combined model integrating metabolic and clinical factors performed best in terms of PFS and OS. For PFS, the C-indices are 0.784 and 0.739 in the training cohort and internal validation cohort, respectively. For OS, the C-indices are 0.831 and 0.782 in the training cohort and internal validation cohort.

Conclusions DFR-signature constructed from multimodal images improved the classification accuracy of prognosis for DLBCL patients. Moreover, the constructed DFR-signature combined with NCCN-IPI exhibited excellent potential for risk stratification of DLBCL patients.

Keywords Multimodal fusion · Deep features · Automated machine learning · Radiomics signature · Diffuse large B-cell lymphoma

Introduction

Diffuse large B-cell lymphoma (DLBCL), a prevalent form of non-Hodgkin lymphoma (NHL), exhibits significant genetic and phenotypic heterogeneity, which results in notable variations in treatment response and survival prognosis (Li et al. 2018; Wright et al. 2020). Despite the standard immunochemotherapy with rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP), about 30–40% of cases experienced refractory disease or death, which ultimately resulted in poor disease outcomes (Ruppert et al. 2020). Therefore, identifying factors influencing DLBCL prognosis and establishing an effective prognostic model to guide clinical decision-making are very significant in current lymphoma diagnosis and treatment (Cunningham et al. 2017). Currently for DLBCL clinical prognosis, the primary indicators include International Prognostic Index (IPI), Revised International Prognostic Index (R-IPI), and NCCN-IPI. However, these indicators fall short of accurately identifying patients with relapsed disease or poor long-term prognosis (Afonso et al. 2021; Sehn et al. 2005; Xu et al. 2023b).

Using [^{18}F]-FDG PET/CT images to evaluate DLBCL patients before treatment has been proven to be promising.

✉ Jianxin Chen
chenjx@njupt.edu.cn

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

Several studies (Chen et al. 2019; Shagera et al. 2019) demonstrated that semiquantitative metabolic parameters from PET/CT images, including total lesion glycolysis (TLG), baseline metabolic tumor volume (MTV), and standardized uptake value (SUV), can be utilized to assess tumor heterogeneity and risk stratification for DLBCL patients, especially in high-risk cases. However, these parameters cannot capture the nuances of the internal structure of the tumor.

The emergence of radiomics made it possible to alleviate this issue. By exploring CT, MR, and PET, other medical images (Eertink et al. 2022) to extract valuable information, radiomics enables a non-invasive and quantitative observation of spatiotemporal heterogeneity in tumors. This provides a new prognostic solution for various diseases, especially lymphomas (Eertink, van de Brug, Eertink et al. 2022a, b; Fan et al. 2022; Li et al. 2023). However, the majority of existing studies (Eertink, van de Brug, Eertink et al. 2022a, b; Jiang et al. 2022a; Jiang et al. 2022b) focused on traditional radiomics, which faced challenges in extracting deep features of medical images. The appearing deep learning techniques are expected to overcome this shortcoming. The deep learning-based radiomics method has shown significant advantages for prognosis of various cancers (Li et al. 2017; Peng et al. 2019; Zhong et al. 2021). However, these models focused on one modality image, ignoring the complementary features of multimodal images. In addition, in these models, manual parameter tuning is rather time-consuming during the feature classification predictions (de Jesus et al. 2022; Fan et al. 2022; Shuilin et al. 2023). Recently, the Automated machine learning (AutoML) framework AutoGluon (Erickson et al. 2020) made it possible to solve this problem. In this study, we tried to construct a multimodal PET/CT deep features-based radiomic model and used the AutoML framework to further enhance the prognostic accuracy of DLBCL patients.

Materials and methods

Patients data collection

Ethical approvals were obtained for this retrospective analysis and the requirements for written consent from patients were waived. This retrospective study enrolled 369 DLBCL patients. Among them, 225 patients were from Nanjing Drum Tower Hospital, affiliated with the Medical School of Nanjing University and 144 patients were from West China Hospital of Sichuan University. All patients received 6 cycles of R-CHOP followed by 2xR. Radiotherapy for PET-CT positive remaining lesions was excluded from our study. Similarly, the Pola-R-CHP regimen was not included. Patients with IPI 0 and no bulk did not receive 4 cycles of R-CHOP followed by 2 cycles of R according to the FLYER trial collective. For patients with a high CNS-IPI, we administered prophylactic intrathecal injections with methotrexate (MTX), but did not administer high-dose MTX prophylactically. Therefore, high-dose MTX was not included in this study. These patients were randomly divided into a training cohort ($n=258$) and an internal validation cohort ($n=111$) with a 7:3 ratio. The detailed patient selection workflow was depicted in Fig. 1.

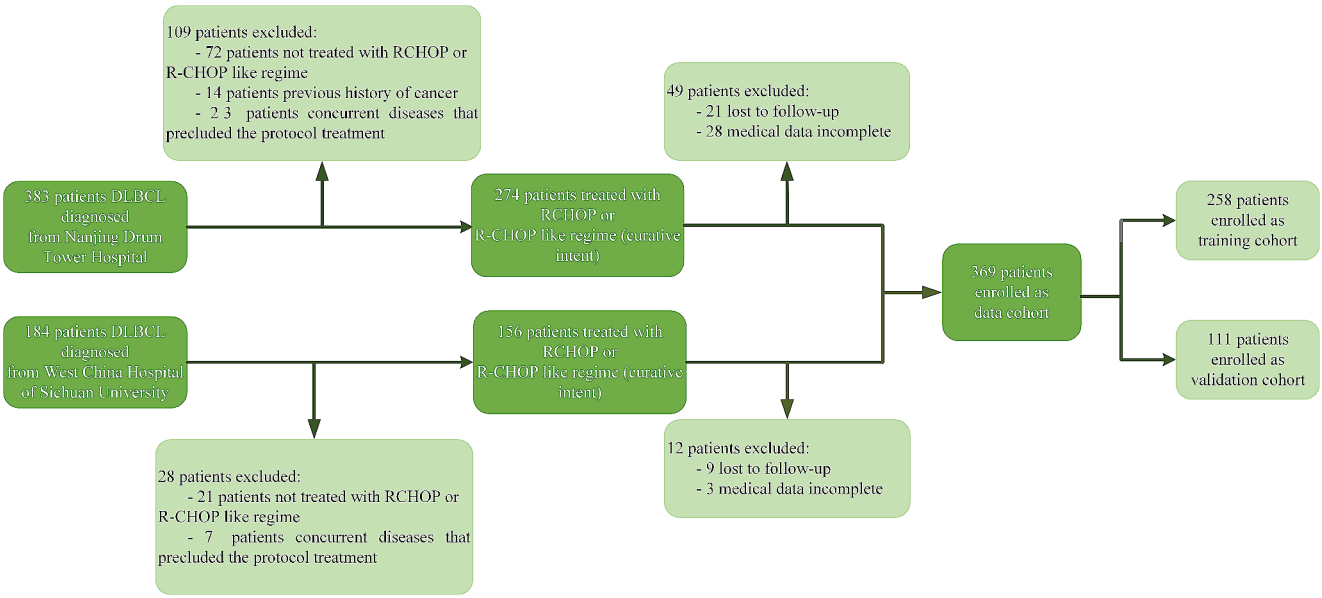

Fig. 1 Patient selection workflow chart

PET/CT scanning protocol

Patients from Nanjing Drum Tower Hospital, affiliated with the Medical School of Nanjing University were scanned using the Gemini GXL PET/CT scanner. Patients from West China Hospital of Sichuan University were scanned using the Gemini GXL and UM780 PET/CT scanner. The main procedure includes: (I) Patients fasted for at least 6 h before the scan to have a blood glucose level below 11.1 mmol/L; (II) After 6 h of fasting (no oral or intravenous fluids containing sugar or glucose), 185 to 370 MBq of [^{18}F]-FDG (5.18 MBq/kg) was administered intravenously; (III) The PET/CT scans were started 60 min after the injection of radiopharmaceuticals. The scans were performed from the base of the skull to the upper thighs and lasted 2 min in each bed position. The scanning parameters were described in the Supplementary Text 1.

Data preprocessing and PET-CT image fusion

Volume of interest (VOI) delineations were completed on PET images with the 3D Slicer tool (version 5.2.0; http://www.slicer.org). Firstly, VOIs were delineated semi-automatically through the Grow-Cut algorithm. Subsequently, the generated VOIs were manually adjusted by two physicians. If there were discrepancies, they were reviewed and determined by a senior nuclear medicine scientist. On CT and multimodal PET-CT images, we did not perform VOI delineations manually. We used the VOI of PET images to do registration on CT images, and in the fused multimodal PET/CT images, which obtained the final VOIs. This approach significantly reduced the work for VOI delineations CT and fused PET/CT images.

To construct the fused PET/CT images, we preprocessed the PET and CT images independently. For CT images, the window width and window level were adjusted to 350HU and 50HU to optimize the contrast (Cui et al. 2022). For PET images, they underwent smoothing and denoising. After that, Z-score normalization was applied to pixel values of PET and CT images to ensure a consistent pixel range. Then, the CT images were performed registration to PET images with the SimpleITK library (https://simpleitk.readthedocs.io/en/master/gettingStarted.html, version 2.0.0). It facilitated the correlations between PET and CT images in terms of the bio-metabolic information and structure information.

After the registration of CT images to PET images, a multi-scale feature-weighted fusion module was utilized to build the fused images. Firstly, PET and CT images underwent down-sampling and pooling to obtain PET/CT feature maps with smooth filtering at multiple scales. After that, they were concatenated to obtain the spatial fusion maps. This process tried to comprehensively capture the multi-level feature information of PET and CT images. Meanwhile, given that PET images emphasized bio-metabolic information in lesion regions, an attention mechanism was added to capture the semantic information contained in PET images, which generated PET attention maps. Then, they were concatenated to CT images to form hybrid attentional maps. Finally, the hybrid attentional maps were combined with the spatial fusion maps to generate multimodal PET-CT maps. These maps were up-sampled to generate the final multimodal fused images.

Radiomics feature extraction

To extract deep features of lesion regions, we utilized several transfer learning and pre-trained deep learning models (such as ResNet-50, VGG-16, VGG-19, Densenet-121, Densenet-169, Xception, and NASNet). This approach aimed to reduce training costs, minimize training time, and enhance model performance. For these models, the ImageNet dataset was widely utilized for training (Morid et al. 2021). The pre-training weights of models on the ImageNet dataset were generally used as the initial weights of the models. Initially, 224 224 pixel-sized images containing tumors were cropped. Subsequently, cubic spline interpolation was used to convert them from grayscale images to RGB images. Finally, there were 1000 deep features extracted through the fully connected layer for each patient. These models were implemented using the Python Keras library (https://github.com/fchollet/keras). The process of deep features-based radiomics analysis is illustrated in Fig. 2. For comparison, we also extracted the radiomics features with the traditional method (see Supplementary Text 2 and Text 3).

Feature selection and DFR-signature construction

All extracted features were normalized through Z-scores. Then the intraclass correlation coefficient (ICC) was calculated to determine the interobserver repeatability of each radiomic feature. Those features with ICCs above 0.8 were kept for further analysis. On these features, we selected the key features with the Mutual Information (MI) algorithm. All features were queued according to their significance to the survival events, and finally, the first 10 features were selected as the key features to construct the DFR-signature (Fan et al. 2022). In addition, the AutoGluon framework (version 0.7.0, https://github.com/autogluon/autogluon) was utilized to optimize the construction of the DFR-signature.

AutoGluon simplified end-to-end machine learning tasks by automating several time-consuming processes such as data preprocessing, optimal algorithm selection, and hyperparameter tuning. The framework included two layers. The

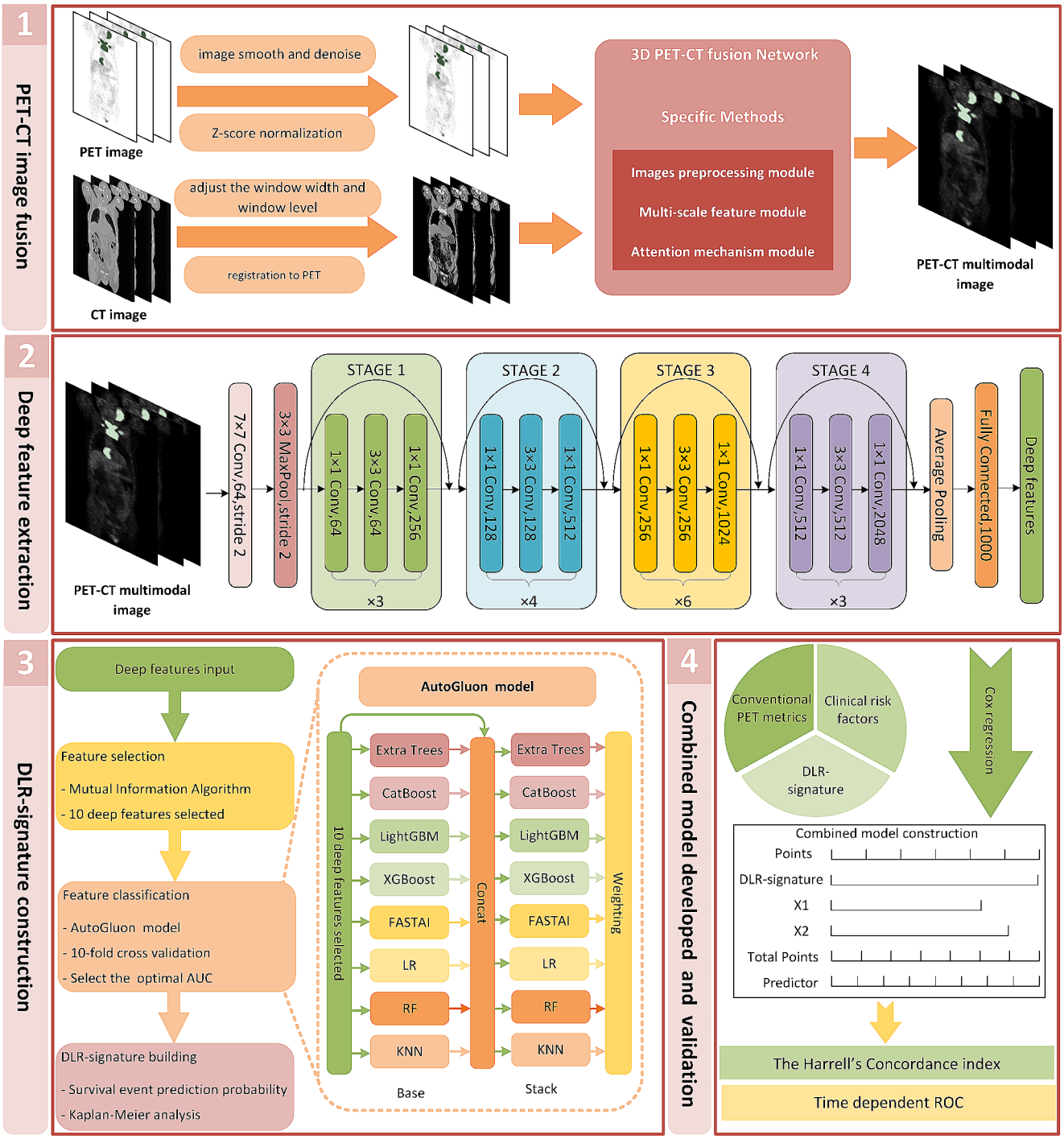

Fig. 2 Flowchart of deep learning radiomics analysis

first layer as the base learner consisted of multiple machine learning models for prediction. The prediction results of the first layer combined with the input data were fed into the second layer, which consisted of multiple stacker models. The machine learning models used in AutoGluon are described in the Supplementary Text 4.

Development and validation of the models

During the training process, a 10-fold cross-validation approach was utilized to mitigate the risk of overfitting. In the training cohort, univariate and multivariate Cox regression analyses were utilized to identify the potential independent clinical factors and metabolic parameters. Then these

prognostic factors were combined with DFR-signature to construct the prognosis model. To validate the combined model based on the multimodal DFR-signature, we also constructed a PET model, a CT model, a NCCN-IPI model, and the conventional radiomics model. The performance of each model was quantified by computing the Harrell consistency index (C-index) and the time-dependent area under the ROC curve (tdAUC).

Statistical analysis

All statistical tests were implemented using SPSS 22.0 and R statistical software (version 4.0.2). Progression-free survival (PFS) and overall survival (OS) were selected as endpoints for assessing the prognosis of DLBCL patients. Here PFS is defined as the time between diagnosis and the first relapse, progression, or death from any cause, or the last follow-up visit, while OS is defined as the time between diagnosis and death from any cause or the last follow-up visit. The difference in clinical information between the training and validation cohorts was assessed using the $\chi 2$ test or the Mann-Whitney U test. $P < 0.05$ was considered statistically significant. In the training cohort, the receiver operating characteristic (ROC) curve was utilized to determine the optimal threshold for SUVmax, TMTV, TLG, and DFR-signature. Potential independent predictors were analyzed using Cox regression analysis. Survival conditions were evaluated by the Kaplan-Meier method, which were compared with log-rank tests. Models were validated in C-index and tdAUC method.

Results

Patient characteristics

The characteristics of all patients were summarized in Supplementary Table 1. There was no statistical difference between the baseline characteristics in both training and validation cohorts ($P > 0.05$). The median follow-up of them was 31 months. In the training cohort, disease relapse occurred in 93 participants, and 66 participants died. Their PFS rates are 80.8%, 42.0% and 7.6% at 1, 3, and 5 years, respectively, and OS rates are 86.7%, 43.1%, and 10.6% at 1, 3, and 5 years, respectively.

Table 1 AUC of different deep learning models for predicting OS

model	Training cohort	95%CI	Validation cohort	95%CI
VGG16	0.807	0.744–0.863	0.713	0.582–0.819
VGG19	0.780	0.718–0.835	0.669	0.543–0.789
RseNet50	0.863	0.812–0.907	0.767	0.657–0.873
DenseNet121	0.785	0.718–0.846	0.740	0.622–0.853
DenseNet169	0.839	0.778–0.895	0.698	0.556–0.836
Xception	0.768	0.699–0.825	0.707	0.606–0.798
NASNet	0.781	0.704–0.844	0.693	0.572–0.784

Predictive performance of the deep learning models and DFR-signatures

We compared several CNN models for feature extraction of DLBCL patient images, including ResNet-50, VGG-16, VGG-19, Densenet-121, Densenet-169, Xception, and NASNet. Results showed that ResNet-50 achieved the biggest value of AUC (see in Table 1). In the training cohort and internal validation cohort, it was 0.863 and 0.767, respectively.

We also performed ResNet-50 to extract features of CT and PET images independently. The ROC results, as presented in Fig. 3, revealed that the R-signature model with the deep features from the fused PET/CT images outperformed that from one modality images and traditional R-signature model in the classification performance.

Kaplan-Meier survival analysis

The cutoff values for DFR-signature of the multimodal PET/CT images in the training cohort were 0.2839 and 0.1911 for PFS and OS (see in Supplementary Table 2). These values were applied to stratify patients into low-risk (below cutoff) and high-risk (equal to or higher than cutoff) groups, which were mapped to the internal validation cohort. To further refine the risk stratification, we combined the traditional NCCN-IPI risk classification and DFR-signature. On the basis of the low-risk and high-risk mentioned above, the group with NCCN-IPI < 4 was subdivided into a low/low-intermediate-risk group, while a high-intermediate/high-risk group was identified within the NCCN-IPI ≥ 4 group. Kaplan-Meier survival analysis was conducted to evaluate the prognostic status of patients using PFS and OS as dependent variables. The Log-rank test was employed as the statistical test. Both training and validation cohorts underwent a significance test with $P < 0.05$ (see in Fig. 4). The results indicated that the DFR-signature constructed in this study combined with the traditional NCCN-IPI could subdivide the patient risk class into four categories accurately. Thereby, it effectively refined the risk assessment in high-risk populations, addressing the limitations of the NCCN-IPI in identifying individuals at high risk.

Univariate and multivariate analysis results

Univariate analysis was carried out using clinical variables, metabolic parameters and DFR-signature. Due to

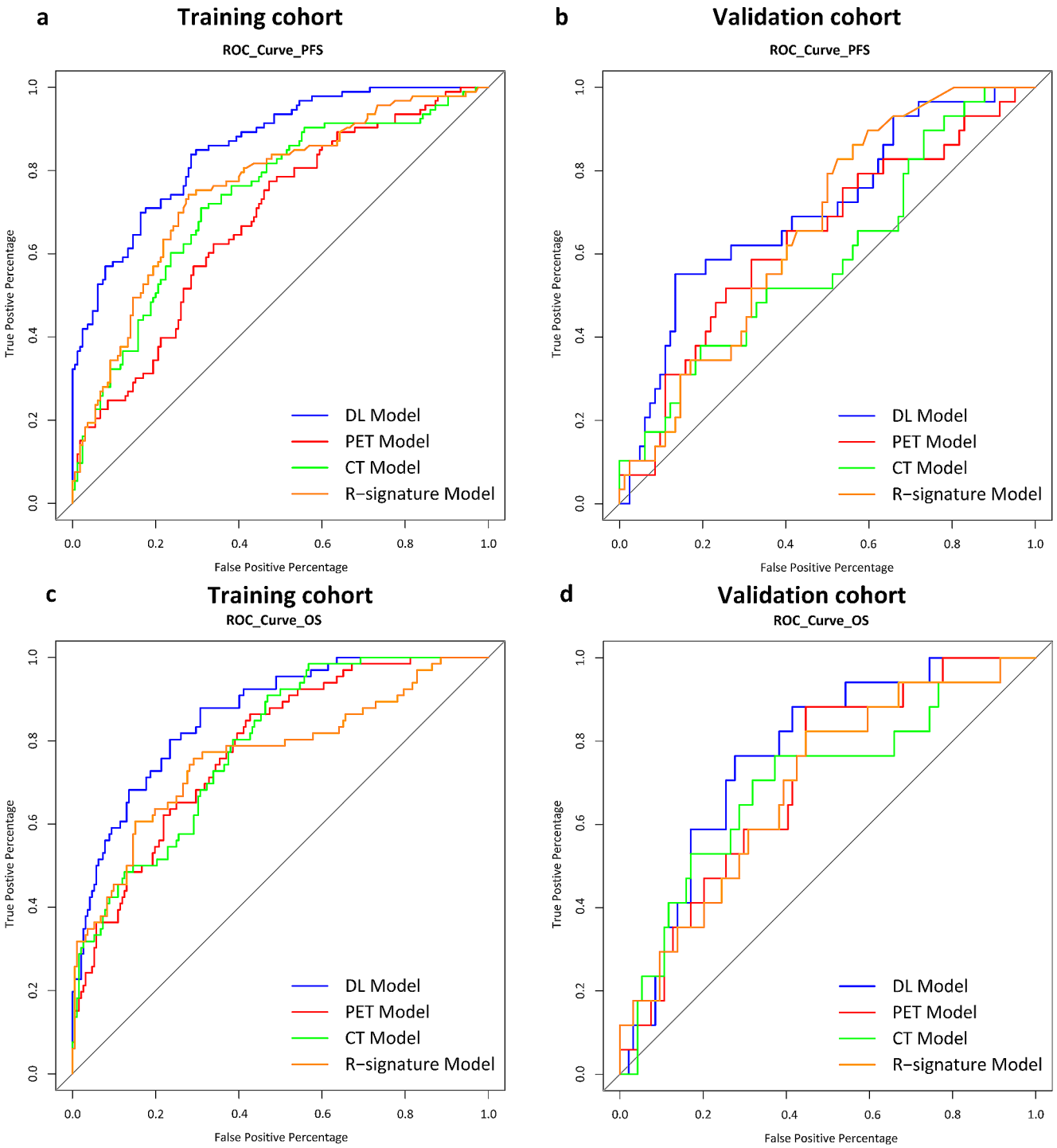

Fig. 3 The ROC curve analysis results for predicting PFS and OS in the training (**a, c**) and internal validation cohorts (**b, d**)

the strong correlation between TMTV and TLG, only TMTV was included in the multivariate analysis. The results were detailed in Table 2. Risk factors associated with PFS included stage (HR = 1.703, $P = 0.044$), NCCN (HR = 2.351, $P < 0.001$), SUV (HR = 1.575, $P = 0.032$), TMTV (HR = 2.535, $P < 0.001$), and DFR-signature$_{PFS}$ (HR = 6.048, $P < 0.001$), while the those associated with OS included subtype (HR = 1.986, $P = 0.006$), NCCN (HR = 3.608, $P < 0.001$), SUV (HR = 1.931, $P = 0.009$), TMTV (HR = 2.280, $P < 0.001$) and DFR-signature$_{OS}$ (HR = 8.058, $P < 0.001$).

3.5 Assessment and validation of models.

According to the results of Cox regression analysis, the combined model, the CT model, the PET model, the

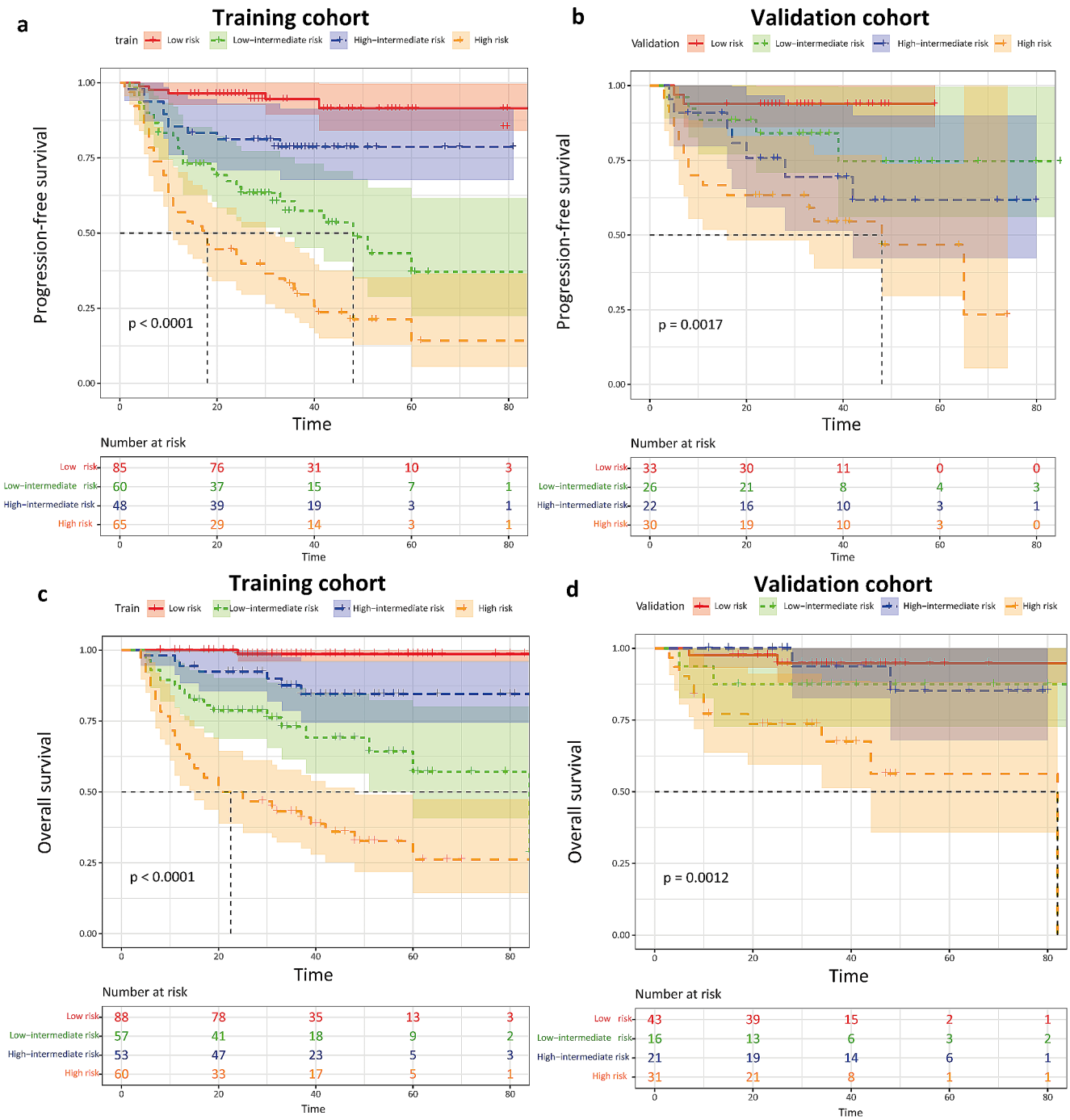

Fig. 4 Kaplan–Meier survival analyses for patients according to the DFR-signature and NCCN-IPI in the training (**a, c**) and internal validation cohorts (**b, d**)

NCCN-IPI model, and the radiomics model (see in Supplementary Table 3) were constructed for predicting PFS and OS. Based on the combined model, individualized prediction nomograms for predicting PFS and OS were built (see in Fig. 5).

To assess the consistency of the classification results, Harrell's C-index was employed for these five models. For PFS, the C-indices of the combined model in the training and internal validation cohorts were 0.784 and 0.739. For OS, the C-indices of the combined model were 0.831 and 0.782. These values exceeded those of competing models (see in Table 3). Furthermore, the tdAUC indicated that the combined model achieved a larger AUC than other models in both the prediction of PFS and OS (see in Fig. 5). The calibration curves for 1-year, 3-year, and 5-year probabilities, as illustrated in Supplementary Fig. 1. They demonstrated

Table 2 Univariate and multivariate analyses of factors predictive of PFS and OS 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	Gender, Female/Male	1.016 (0.677–1.526)	0.938	-	-	1.140(0.703–1.848)	0.597	-	-
	Age, ≤60/>60	1.316 (0.873–1.985)	0.190	-	-	2.013(1.221–3.320)	0.006	-	-
	Ann Arbor stage, I-II/III-IV	2.316 (1.423–3.767)	< 0.001	1.703(1.014–2.862)	0.044	2.191(1.231–3.899)	< 0.001	-	-
	B symptoms, no/yes	1.815 (1.204–2.736)	0.004	-	-	1.870(1.151–3.039)	0.012	-	-
	LDH level, normal/elevated	1.576 (1.048–2.368)	0.029	-	-	1.675(1.031–2.720)	0.037	-	-
	ECOG PS, 0–1/≥2	1.712 (1.097–2.670)	0.018	-	-	1.984(1.187–3.318)	0.009	-	-
	NCCN-IPI, 0–3/≥4	2.818 (1.835–4.327)	< 0.001	2.351(1.493–3.704)	< 0.001	3.518(2.063–5.998)	< 0.001	3.608 (2.107–6.179)	< 0.001
	Bulky disease, no/yes	2.166 (1.399–3.352)	< 0.001	-	-	2.178(1.300–3.650)	0.003	-	-
	Extranodal involvement, no/yes	1.488 (0.962–2.301)	0.074	-	-	1.333(0.802–2.214)	0.267	-	-
	Pathological type, non-GCB/GCB	1.298 (0.864–1.950)	0.209	-	-	1.647(1.012–2.679)	0.044	1.986 (1.215–3.245)	0.006
PET-CT predictors	SUVmax	1.932 (1.282–2.911)	0.002	1.575(1.040–2.385)	0.032	2.470(1.510–4.043)	< 0.001	1.931 (1.175–3.174)	0.009
	TMTV	3.195 (2.125–4.804)	< 0.001	2.535(1.681–3.822)	< 0.001	3.867(2.370–6.309)	< 0.001	2.280 (1.382–3.763)	< 0.001
	TLG	3.272 (2.170–4.934)	< 0.001	-	-	3.564(2.173–5.845)	< 0.001	-	-
	DFR-signature$_{PFS}$	7.248 (4.164–12.610)	< 0.001	6.084(3.480-10.636)	< 0.001	-	-	-	-
	DFR-signature$_{OS}$	-	-	-	-	9.105(4.158–19.94)	< 0.001	8.058 (3.788–17.143)	< 0.001

Abbreviations: CI, confidence interval; 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 center B cell; SUVmax, maximum standardized uptake value; TMTV, total metabolic tumor volume; TLG, total lesion glycolysis

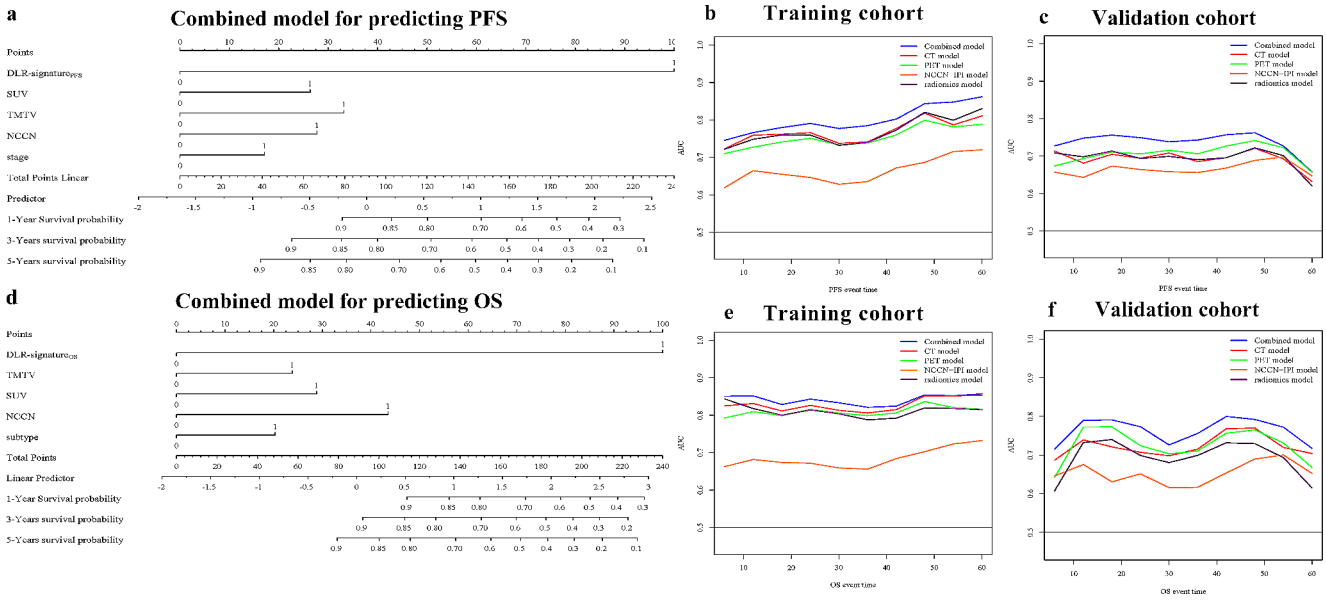

Fig. 5 **a** The combined model for predicting PFS in the training cohort. **b**, **c** Time-dependent area under the ROC curve of the models for predicting PFS in the training and validation cohorts. **d** The combined model for predicting OS. **e**, **f** Time-dependent area under the ROC curve of the models for predicting OS in the training and validation cohorts

Table 3 Harrell's C-index results for PFS and OS in the training and validation cohorts

classify	model	Training cohort		Validation cohort	
		C-index	95%CI	C-index	95%CI
PFS	Combined model	0.784	0.741–0.827	0.739	0.649–0.829
	CT model	0.732	0.685–0.779	0.718	0.632–0.804
	PET model	0.724	0.675–0.773	0.717	0.627–0.807
	NCCN-IPI model	0.632	0.581–0.683	0.654	0.566–0.742
	Radiomics model	0.740	0.693–0.787	0.721	0.635–0.807
OS	Combined model	0.831	0.788–0.874	0.782	0.674–0.890
	CT model	0.799	0.748–0.850	0.721	0.594–0.848
	PET model	0.791	0.736–0.846	0.745	0.631–0.859
	NCCN-IPI model	0.659	0.600-0.718	0.636	0.518–0.754
	Radiomics model	0.787	0.736–0.838	0.734	0.601–0.867

Table 4 AUC of DFR-signatures-based model for predicting PFS and OS

classify	Training cohort	95%CI	Test cohort	95%CI	Validation cohort	95%CI
PFS	0.758	0.674–0.834	0.703	0.575–0.823	0.660	0.556–0.769
OS	0.753	0.672–0.829	0.701	0.550–0.826	0.686	0.504–0.841

that the estimates predicted by the combined model aligned closely with the actual observations in both the training and internal validation cohorts. These results suggested that the combined model has clinical utility in guiding the treatment and prognostic of DLBCL patients.

To further validate the generalizability of our model, we used dataset from Nanjing Drum Tower Hospital for training and testing, the dataset from West China Hospital of Sichuan University as a completely independent external validation cohort. Table 4 lists the PFS and OS under our method. We can find the results are promising on the independent dataset, which convinced the model generalization.

Discussion

In this study, we developed a combined model based on multimodal DFR-signatures to predict and validate the PFS and OS of DLBCL patients at 1, 3, and 5 years. The experimental results showed that the DFR-signature constructed by multimodal images could be served as an independent factor for the prognosis of DLBCL patients. Furthermore, our model was more instructive in individualized prognosis of DLBCL patients compared with the unimodal PET/CT model, NCCN-IPI model, and conventional radiomics model.

Clinically, imaging bio-metabolic information extracted from PET images have proven valuable for the prognosis of DLBCL patients (Cottereau et al. 2021; Eertink et al. 2022a; Fan et al. 2022; Kostakoglu et al. 2022). A recent study by Mikhaeel et al. demonstrated that Ann Arbor stage, age and MTV were effective in predicting the probability of progression-free survival at 3 years (Mikhaeel et al. 2022). In addition, Eertink et al. (Eertink et al. 2022) developed a model that combined PET-extracted features (including MTV and $Dmax_{bulk}$) with clinical parameters to obtain a time-to-progression model capable of predicting 2 years (AUC of 0.77). Meanwhile, the analysis of tumor regions from CT images has been shown to be an effective discriminative technique for predicting primary treatment failure (PTF) (Seidler et al. 2019; Xu et al. 2021). However, the information provided by unimodal PET and CT images was limited, which restricted further enhancements in prognostic accuracy. Consequently, in recent years, researchers have started to integrate PET and CT images to create multimodal images in the medical field (Li et al. 2021; Lv et al. 2020; Xu et al. 2023a, 2023b). Zhang et al. (Zhang et al. 2020) demonstrated a multimodal MRI-based radiomics model could be used to identify benign or malignant breast lesions. Moreover, Amini et al. (Amini et al. 2021) further indicated that the multimodal radiomics fusion model outperformed the unimodal model in predicting OS of non-small-cell lung cancer patients. Drawing inspiration from their conclusions, we constructed multimodal PET/CT images for the prognosis of DLBCL patients. Multimodal images took full advantage of the semantic complementarity between PET and CT images, which combined the advantages of anatomical information of CT images and metabolic information from PET images. Compared to PET/CT images, the results proved that the predictive performance based on multimodal PET-CT images surpassed that of PET/CT images alone of both PFS (C-indices: 0.784, 0.732 and 0.724 in the training cohort, 0.739, 0.718 and 0.717 in the internal validation cohort) and OS (C-indices: 0.831, 0.799 and 0.791 in the training cohort, 0.782, 0.721 and 0.745 in the internal validation cohort).

The studies mentioned above have shown that traditional radiomics methods appeared promising in various aspects. However, there are still limitations such as restricted automation, standardization, and efficiency in feature extraction. Particularly, these methods faced challenges in capturing advanced and nonlinear features from images. To overcome these issues, recent studies have progressively integrated deep learning networks into the feature extraction process, replacing the conventional manual methods (Li et al. 2017; Zheng et al. 2020). For instance, Liu et al. (Liu et al. 2022) developed a multi-task 3D U-Net for lymphoma segmentation and PFS prediction, achieving commendable results (AUC of 0.827) in comparison to radiomics-based models and single-task CNNs. Jiang et al. (Jiang et al. 2021) trained a deep neural network (S-net) with CT images to predict the prognosis in gastric cancer patients. In this study, we constructed PET/CT multimodal DFR-signatures through deep feature extraction. They were independently associated with PFS (HR = 6.084, $P = 0.001$) and OS (HR = 8.058, $P = 0.001$). To summarize, both this study and our previous studies have validated that DFR were strongly correlated with tumor heterogeneity and might be helpful for the prognosis of DLBCL patients.

However, there still existed limited research on the application of CNNs in lymphoma prognosis. Few studies have explored comparisons between various network models. It was noteworthy that we compared the feature extraction capabilities of seven different network models and ultimately selected the ResNet-50 model for its optimal performance. The DFR-signature, derived from features extracted by the ResNet-50, exhibited an impressive AUC of 0.857 and 0.863 in predicting PFS and OS. This underscored the substantial advantage of the DFR-signature based on the ResNet-50 for prognosis of DLBCL patients. Our study not only offered a more comprehensive and reliable method for model selection in DLBCL prognosis, but also served as a reference for further research in this field. Furthermore, compared with the traditional radiomics model, our results demonstrated that the combined model based on DFR-signature had a significant predictive advantage in assessment metrics, including C-indices (PFS: 0.784 and 0.740, OS: 0.831 and 0.787 in the training cohort, PFS: 0.739 and 0.721, OS: 0.782 and 0.734 in the internal validation cohort) and tdAUC analysis. While the traditional radiomics method extracted first-order statistical features and second-order texture features based on the grey-scale histogram of a lesion, deep learning networks extracted high-dimensional features from tumor regions.

Jiang et al. (Jiang et al. 2023) employed deep convolutional neural network architectures, VGG19 and DenseNet121, to predict the survival of DLBCL patients using prognostic imaging markers generated from PET images alone. Our study extended this by exploring the predictive value of DFR-signature from multimodal PET/CT images in DLBCL. We achieved more accurate predictive results compared to the unimodal approach, overcoming the limitations of unimodal images. Additionally, Hao et al. (Peng et al. 2019) only utilized features extracted from PET and CT images to predict disease-free survival in patients with advanced nasopharyngeal carcinoma. Zhao et al. (Cho et al. 2021) employed a deep learning model to extract features from CT images for prognostic stratification of lung adenocarcinoma prior to treatment. These studies relied on features extracted by deep learning to predict prognosis,

overlooking the clinical characteristics of the patients. In contrast, our approach combined the clinical features, metabolic indicators, and DFR-signature to construct a combined model, which significantly improved the accuracy of prognosis. Furthermore, compared to the NCCN-IPI model, our combined model also exhibited superior performance in terms of C-indices (PFS: 0.784 and 0.632, OS: 0.831 and 0.659 in the training cohort, PFS: 0.739 and 0.654, OS: 0.782 and 0.636 in the internal validation cohort) and tdAUC.

Additionally, to our knowledge, this study represented the first attempt to evaluate the prognostic impact of DFR on DLBCL patients using an AutoML stacked integration approach. In recent studies, Jiang et al. (Jiang et al. 2022) proposed a cross-combination of seven machine learning models for feature selection, and predicted survival events of DLBCL patients (AUC = 0.757). Zhao et al. (Shuilin et al. 2023) introduced a stacked integration approach based on four machine learning classifiers, demonstrating a more accurate classification performance compared to the single machine learning classifier (AUC = 0.771). In our study, with AutoGluon framework, we obtained the predicted AUCs of 0.857 and 0.863 for PFS and OS in the training cohort. This was explained by the fact that AutoGluon employs multiple model fusion strategies, including voting-based and stacking integration, and consistently achieves classification performance even across different datasets (Erickson et al. 2020).

On central nervous system (CNS) relapse, Kang et al. (Kang et al., 2018) extracted 1,618 radiomic features from patients with glioblastoma and primary CNS lymphoma. The prognostic accuracy of radiomics analysis (AUC = 0.944) was superior to conventional genomics (AUC = 0.819). Nie et al. (Nie et al. 2019) achieved 90.66% accuracy in predicting the survival of high-grade gliomas using a radiomics-based deep learning approach. Our study indicated that DFR feature-based model outperformed conventional radiomics in predicting outcomes for patients with DLBCL. Therefore, it was reasonable to believe that DFR feature-based model could also be used as a new low-cost tool to predict CNS relapse, improve treatment decisions for CNS diseases, and play an important role in the routine clinical management of these conditions.

Despite the strengths of this paper, certain limitations and challenges persisted. The novel therapies of DLBCL were not considered in our study. These therapies included small molecule drugs, antibody–drug conjugates (ADC), CD3xCD20 bispecific antibodies, and chimeric antigen receptor (CAR) T cell therapies (Abramson et al. 2020). Notably, the CAR-T cell therapy fundamentally changed the management of patients with relapsed and refractory large B-cell lymphomas (Nastoupil 2023). However, studies have indicated that more than 50% patients did not benefit from this therapy due to non-response or further relapse (Caballero et al. 2022). Therefore, it was essential to continue researching potential biomarkers to identify individuals most likely to benefit from these new therapies. Future studies could be further expanded to include these biological features of novel therapies. This expansion aimed to more accurately assess the response in patients with DLBCL and improve patient survival.

Conclusions

The DFR-signature extracted from multimodal PET-CT images can be used as an independent survival predictor for DLBCL patients. The accuracy in prognosis significantly surpasses that of unimodal models and the conventional radiomics model. Furthermore, the combination with NCCN-IPI have the potential to achieve more refined survival risk stratification of patients.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00432-024-05905-0.

Author contributions F.L., Z.D. and Y.C.: Software, Validation, Formal analysis, Investigation, and Data Curation; J.C.: Conceptualization, Methodology, Resources, Writing - Review & Editing, Project administration, and Funding acquisition; F.L.: Writing – original draft, Visualization, Writing - Review & Editing and Supervision; Y.F.: Investigation and Writing - Review & Editing; A.L. and C.Z.: Software and Investigation. All authors reviewed the manuscript.

Funding This work was supported by the Clinical Trials from the Affiliated Drum Tower Hospital, Medical School of Nanjing University (Grant No.2021-LCYJ-MS-04 and 2022-LCYJ-PY-44); and the Key Project of Medical Science and Technology of Nanjing (Grant No. ZKX21011).

Data availability No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate 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.

Competing interests The authors declare no competing interests.

References

Abramson JS, Ghosh N, Smith SM (2020) ADCs, BiTEs, CARs, and small molecules: a new era of targeted therapy in Non-hodgkin Lymphoma. Am Soc Clin Oncol Educational Book 40302–313. https://doi.org/10.1200/EDBK_279043

Afonso C, Duarte S, Marques BA, Lima CB, Neves D, Lai AC, Monteiro M, Saraiva T, Roque A, Ruzickova L, Carda JP, Gomes M, Julião MJ, Costa G, Espadana A (2021) Prognostic Value of Baseline PET/CT Imaging in diffuse large B-Cell lymphoma: does the Largest Distance between two lesions play a role in prognosis? Blood 138:4577. https://doi.org/10.1182/blood-2021-153103

Amini M, Nazari M, Shiri I, Hajianfar G, Deevband MR, Abdollahi H, Arabi H, Rahmim A, Zaidi H (2021) Multi-level multi-modality (PET and CT) fusion radiomics: prognostic modeling for non-small cell lung carcinoma. Phys Med Biol 66(20):205017. https://doi.org/10.1088/1361-6560/ac287d

Caballero AC, Escribà-Garcia L, Alvarez-Fernández C, Briones J (2022) CAR T-Cell therapy predictive response markers in diffuse large B-Cell lymphoma and therapeutic options after CART19 failure. Front Immunol 13:904497. https://doi.org/10.3389/fimmu.2022.904497

Chen S, He K, Feng F, Wang S, Yin Y, Fu H, Wang H (2019) Metabolic tumor burden on baseline 18F-FDG PET/CT improves risk stratification in pediatric patients with mature B-cell lymphoma. Eur J Nucl Med Mol Imaging 46(9):1830–1839. https://doi.org/10.1007/s00259-019-04363-y

Cho H-h, Lee HY, Kim E, Lee G, Kim J, Kwon J, Park H (2021) Radiomics-guided deep neural networks stratify lung adenocarcinoma prognosis from CT scans. Commun Biology 4(1):1286. https://doi.org/10.1038/s42003-021-02814-7

Cottereau AS, Meignan M, Nioche C, Capobianco N, Clerc J, Chartier L, Vercellino L, Casasnovas O, Thieblemont C, Buvat I (2021) Risk stratification in diffuse large B-cell lymphoma using lesion dissemination and metabolic tumor burden calculated from baseline PET/CT†. Ann Oncol 32(3):404–411. https://doi.org/10.1016/j.annonc.2020.11.019

Cui Y, Zhang J, Li Z, Wei K, Lei Y, Ren J, Wu L, Shi Z, Meng X, Yang X, Gao X (2022) A CT-based deep learning radiomics nomogram for predicting the response to neoadjuvant chemotherapy in patients with locally advanced gastric cancer: a multicenter cohort study. eClinicalMedicine 46. https://doi.org/10.1016/j.eclinm.2022.101348

Cunningham J, Iyengar S, Sharma B (2017) Evolution of lymphoma staging and response evaluation: current limitations and future directions. Nat Rev Clin Oncol 14(10):631–645. https://doi.org/10.1038/nrclinonc.2017.78

de Jesus FM, Yin Y, Mantzorou-Kyriaki E, Kahle XU, de Haas RJ, Yakar D, Glaudemans AWJM, Noordzij W, Kwee TC, Nijland M (2022) Machine learning in the differentiation of follicular lymphoma from diffuse large B-cell lymphoma with radiomic [18F] FDG PET/CT features. Eur J Nucl Med Mol Imaging 49(5):1535–1543. https://doi.org/10.1007/s00259-021-05626-3

Eertink JJ, van de Brug T, Wiegers SE, Zwezerijnen GJC, Pfaehler EAG, Lugtenburg PJ, van der Holt B, de Vet HCW, Hoekstra OS, Boellaard R, Zijlstra JM (2022a) 18F-FDG PET baseline radiomics features improve the prediction of treatment outcome in diffuse large B-cell lymphoma. Eur J Nucl Med Mol Imaging 49(3):932–942. https://doi.org/10.1007/s00259-021-05480-3

Eertink JJ, Zwezerijnen GJC, Heymans MW, Pieplenbosch S, Wiegers SE, Dührsen U, Hüttmann A, Kurch L, Hanoun C, Lugtenburg P, Barrington SF, Mikhaeel G, Ceriani L, Zucca E, Czibor S, Györke T, Chamuleau MED, Hoekstra OS, de Vet HCW, Boellaard R, Zijlstra JM (2022b) External validation shows that baseline PET Radiomics Outperform the IPI risk score for prediction of Outcome in DLBCL. Blood 140(Supplement 1):777–778. https://doi.org/10.1182/blood-2022-163454

Erickson N, Mueller J, Shirkov A, Zhang H, Larroy P, Li M, Smola AJ (2020) a. e.-p. AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data, arXiv:2003.06505. Retrieved from https://ui.adsabs.harvard.edu/abs/2020arXiv200306505Ehttps://doi.org/10.48550/arXiv.2003.06505

Fan S, Jiang C, Teng Y, Li A, Chen J, Xu J (2022) Predictive value of Multi-parameter Model incorporating PET-Based Radiomics features for Survival Prediction of patients with diffuse large B-Cell lymphoma in older patients. Blood 140(Supplement 1):3764–3764. https://doi.org/10.1182/blood-2022-165599

Jiang Y, Jin C, Yu H, Wu J, Chen C, Yuan Q, Huang W, Hu Y, Xu Y, Zhou Z, Fisher GA Jr., Li G, Li R (2021) Development and validation of a Deep Learning CT Signature to Predict Survival and Chemotherapy Benefit in Gastric Cancer: a Multicenter, Retrospective Study. Ann Surg 274(6). https://doi.org/10.1097/SLA.0000000000003778. nie

Jiang C, Li A, Teng Y, Huang X, Ding C, Chen J, Xu J, Zhou Z (2022) Optimal PET-based radiomic signature construction based on the cross-combination method for predicting the survival of patients with diffuse large B-cell lymphoma. Eur J Nucl Med Mol Imaging 49(8):2902–2916. https://doi.org/10.1007/s00259-022-05717-9

Jiang C, Huang X, Li A, Teng Y, Ding C, Chen J, Xu J, Zhou Z (2022a) Radiomics signature from [18F]FDG PET images for prognosis predication of primary gastrointestinal diffuse large B cell lymphoma. Eur Radiol 32(8):5730–5741. https://doi.org/10.1007/s00330-022-08668-9

Jiang C, Qian C, Jiang Z, Teng Y, Lai R, Sun Y, Ni X, Ding C, Xu Y, Tian R (2023) Robust deep learning-based PET prognostic imaging biomarker for DLBCL patients: a multicenter study. Eur J Nucl Med Mol Imaging 50(13):3949–3960. https://doi.org/10.1007/s00259-023-06405-y

Kang, D., Park, J. E., Kim, Y.-H., Kim, J. H., Oh, J. Y., Kim, J., Kim, Y., Kim, S. T., & Kim, H. S. (2018).Diffusion radiomics as a diagnostic model for atypical manifestation of primary central nervous system lymphoma: development and multicenter external validation Neuro-Oncology 20(9) 1251-1261 https://doi.org/10.1093/neuonc/noy021

Kostakoglu L, Dalmasso F, Berchialla P, Pierce LA, Vitolo U, Martelli M, Sehn LH, Trněný M, Nielsen TG, Bolen CR, Sahin D, Lee C, El-Galaly TC, Mattiello F, Kinahan PE, Chauvie S (2022) A prognostic model integrating PET-derived metrics and image texture analyses with clinical risk factors from GOYA. *EJHaem*. 3(2):406–414. https://doi.org/10.1002/jha2.421

Li Z, Wang Y, Yu J, Guo Y, Cao W (2017) Deep learning based Radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. Sci Rep 7(1):5467. https://doi.org/10.1038/s41598-017-05848-2

Li S, Young KH, Medeiros LJ (2018) Diffuse large B-cell lymphoma. Pathology 50(1):74–87. https://doi.org/10.1016/j.pathol.2017.09.006

Li Y, Zhang Y, Fang Q, Zhang X, Hou P, Wu H, Wang X (2021) Radiomics analysis of [18F]FDG PET/CT for microvascular invasion and prognosis prediction in very-early- and early-stage hepatocellular

carcinoma. Eur J Nucl Med Mol Imaging 48(8):2599–2614. https://doi.org/10.1007/s00259-020-05119-9

Li M, Yao H, Zhang P, Zhang L, Liu W, Jiang Z, Li W, Zhao S, Wang K (2023) Development and validation of a [18F]FDG PET/CT-based radiomics nomogram to predict the prognostic risk of pretreatment diffuse large B cell lymphoma patients. Eur Radiol 33(5):3354–3365. https://doi.org/10.1007/s00330-022-09301-5

Liu P, Zhang M, Gao X, Li B, Zheng G (2022) Joint Lymphoma Lesion Segmentation and prognosis prediction from baseline FDG-PET images via Multitask Convolutional neural networks. IEEE Access 10:81612–81623. https://doi.org/10.1109/ACCESS.2022.3195906

Lv W, Ashrafinia S, Ma J, Lu L, Rahmim A (2020) Multi-level Multi-modality Fusion Radiomics: application to PET and CT imaging for prognostication of Head and Neck Cancer. IEEE J Biomedical Health Inf 24(8):2268–2277. https://doi.org/10.1109/JBHI.2019.2956354

Mikhaeel NG, Heymans MW, Eertink JJ, de Vet HC, Boellaard R, Dührsen U, Ceriani L, Schmitz C, Wiegers SE, Hüttmann AJJ-oCO (2022) Proposed new dynamic prognostic index for diffuse large B-cell lymphoma: international metabolic prognostic index. 40(21):2352

Morid MA, Borjali A, Del Fiol G (2021) A scoping review of transfer learning research on medical image analysis using ImageNet. Comput Biol Med 128:104115. https://doi.org/10.1016/j.compbiomed.2020.104115

Nastoupil L (2023) Will CAR T-Cell Therapy be the Preferred modality in Frontline Treatment of large B-Cell lymphoma? Hematologist 20. https://doi.org/10.1182/hem.V20.4.2023414

Nie D, Lu J, Zhang H, Adeli E, Wang J, Yu Z, Liu L, Wang Q, Wu J, Shen D (2019) Multi-channel 3D deep feature learning for Survival Time Prediction of Brain Tumor patients using multi-modal neuroimages. Sci Rep 9(1):1103. https://doi.org/10.1038/s41598-018-37387-9

Peng H, Dong D, Fang M-J, Li L, Tang L-L, Chen L, Li W-F, Mao Y-P, Fan W, Liu L-Z, Tian L, Lin A-H, Sun Y, Tian J, Ma J (2019) Prognostic Value of Deep Learning PET/CT-Based Radiomics: potential role for future individual induction chemotherapy in Advanced Nasopharyngeal Carcinoma. Clin Cancer Res 25(14):4271–4279. https://doi.org/10.1158/1078-0432.CCR-18-3065

Ruppert AS, Dixon JG, Salles G, Wall A, Cunningham D, Poeschel V, Haioun C, Tilly H, Ghesquieres H, Ziepert M, Flament J, Flowers C, Shi Q, Schmitz N (2020) International prognostic indices in diffuse large B-cell lymphoma: a comparison of IPI, R-IPI, and NCCN-IPI. Blood 135(23):2041–2048. https://doi.org/10.1182/blood.2019002729

Sehn LH, Chhanabhai M, Fitzgerald C, Gill K, Hoskins P, Klasa R, Savage K, Shenkier T, Sutherland J, Wilson K, Gascoyne RD, Connors JM (2005) Revised International Prognostic Index (R-IPI) is a Better Predictor of Outcome Than the Standard IPI for patients with diffuse large B-Cell lymphoma (DLBCL) treated with Rituximab and CHOP (R-CHOP). Blood 106(11):492. https://doi.org/10.1182/blood.V106.11.492.492

Seidler M, Forghani B, Reinhold C, Pérez-Lara A, Romero-Sanchez G, Muthukrishnan N, Wichmann JL, Melki G, Yu E, Forghani R (2019) Dual-energy CT texture analysis with machine learning for the evaluation and characterization of cervical Lymphadenopathy. Comput Struct Biotechnol J 17:1009–1015. https://doi.org/10.1016/j.csbj.2019.07.004

Shagera QA, Cheon GJ, Koh Y, Yoo MY, Kang KW, Lee DS, Kim EE, Yoon S-S, Chung J-K (2019) Prognostic value of metabolic tumour volume on baseline 18F-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 46(7):1417–1427. https://doi.org/10.1007/s00259-019-04309-4

Shuilin Z, Jing W, Chentao J, Xiang Z, Chenxi X, Rui Z, Yan Z, Yuwei L, Xuexin H, Youyou Z, Caiyun X, Lixia Z, Wenbin Q, Hong Z, Xiaohui Z, Mei T (2023) Stacking ensemble learning–based [18F]FDG PET radiomics for outcome prediction in diffuse large B-Cell lymphoma. J Nucl Med 64(10):1603. https://doi.org/10.2967/jnumed.122.265244

Wright GW, Huang DW, Phelan JD, Coulibaly ZA, Roulland S, Young RM, Wang JQ, Schmitz R, Morin RD, Tang J, Jiang A, Bagaev A, Plotnikova O, Kotlov N, Johnson CA, Wilson WH, Scott DW, Staudt LM (2020) A probabilistic classification Tool for genetic subtypes of diffuse large B cell lymphoma with therapeutic implications. Cancer Cell 37(4):551–568e514. https://doi.org/10.1016/j.ccell.2020.03.015

Xu Q, Sun Z, Li X, Ye C, Zhou C, Zhang L, Lu G (2021) Advanced gastric cancer: CT radiomics prediction and early detection of downstaging with neoadjuvant chemotherapy. Eur Radiol 31(11):8765–8774. https://doi.org/10.1007/s00330-021-07962-2

Xu H, Lv W, Zhang H, Yuan Q, Wang Q, Wu Y, Lu L (2023) Multimodality radiomics analysis based on [18F]FDG PET/CT imaging and multisequence MRI: application to nasopharyngeal carcinoma prognosis. Eur Radiol 33(10):6677–6688. https://doi.org/10.1007/s00330-023-09606-z

Xu H, Ma J, Yang G, Xiao S, Li W, Sun Y, Sun Y, Wang Z, Zhao H (2023b) Prognostic value of metabolic tumor volume and lesion dissemination from baseline PET/CT in patients with diffuse large B-cell lymphoma: further risk stratification of the group with low-risk and high-risk NCCN-IPI. Eur J Radiol 163:110798. https://doi.org/10.1016/j.ejrad.2023.110798

Zhang Q, Peng Y, Liu W, Bai J, Zheng J, Yang X, Zhou L (2020) Radiomics based on Multimodal MRI for the Differential diagnosis of Benign and malignant breast lesions. J Magn Reson Imaging 52(2):596–607. https://doi.org/10.1002/jmri.27098

Zheng X, Yao Z, Huang Y, Yu Y, Wang Y, Liu Y, Mao R, Li F, Xiao Y, Wang Y, Hu Y, Yu J, Zhou J (2020) Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. Nat Commun 11(1):1236. https://doi.org/10.1038/s41467-020-15027-z

Zhong L, Dong D, Fang X, Zhang F, Zhang N, Zhang L, Fang M, Jiang W, Liang S, Li C, Liu Y, Zhao X, Cao R, Shan H, Hu Z, Ma J, Tang L, Tian J (2021) A deep learning-based radiomic nomogram for prognosis and treatment decision in advanced nasopharyngeal carcinoma: a multicentre study. EBioMedicine 70:103522. https://doi.org/10.1016/j.ebiom.2021.103522