BMC Medicine

RESEARCH

Open Access

Virtual biopsy for non-invasive identification of follicular lymphoma histologic transformation using radiomics-based imaging biomarker from PET/CT

Chong Jiang[1†], Chunjun Qian[2,3,4†], Qiuhui Jiang[5], Hang Zhou[6], Zekun Jiang[7], Yue Teng[8], Bing Xu[5], Xin Li[6*], Chongyang Ding[9*] and Rong Tian[1*]

Abstract

Background This study aimed to construct a radiomics-based imaging biomarker for the non-invasive identification of transformed follicular lymphoma (t-FL) using PET/CT images.

Methods A total of 784 follicular lymphoma (FL), diffuse large B-cell lymphoma, and t-FL patients from 5 independent medical centers were included. The unsupervised EMFusion method was applied to fuse PET and CT images. Deep-based radiomic features were extracted from the fusion images using a deep learning model (ResNet18). These features, along with handcrafted radiomics, were utilized to construct a radiomic signature (R-signature) using automatic machine learning in the training and internal validation cohort. The R-signature was then tested for its predictive ability in the t-FL test cohort. Subsequently, this R-signature was combined with clinical parameters and SUVmax to develop a t-FL scoring system.

Results The R-signature demonstrated high accuracy, with mean AUC values as 0.994 in the training cohort and 0.976 in the internal validation cohort. In the t-FL test cohort, the R-signature achieved an AUC of 0.749, with an accuracy of 75.2%, sensitivity of 68.0%, and specificity of 77.5%. Furthermore, the t-FL scoring system, incorporating the R-signature along with clinical parameters (age, LDH, and ECOG PS) and SUVmax, achieved an AUC of 0.820, facilitating the stratification of patients into low, medium, and high transformation risk groups.

Conclusions This study offers a promising approach for identifying t-FL non-invasively by radiomics analysis on PET/CT images. The developed t-FL scoring system provides a valuable tool for clinical decision-making, potentially improving patient management and outcomes.

Keywords Follicular lymphoma, Histologic transformation, Scoring system, Radiomics, PET/CT

†Chong Jiang and Chunjun Qian are cofirst authors. They contributed equally to the work.

*Correspondence:
Xin Li
lixin16@sdu.edu.cn
Chongyang Ding
chongyangding@163.com
Rong Tian
rongtiannuclear@126.com
Full list of author information is available at the end of the article

Background

Follicular lymphoma (FL) is the most common form of indolent lymphoma originating from germinal center B cells [1]. It typically follows an indolent course characterized by chronic relapsing disease, yet generally favorable outcomes. However, approximately 2 to 3% of FL patients experience histologic transformation to an aggressive lymphoma, predominantly diffuse large B-cell lymphoma with germinal center phenotype (GCB-DLBCL) [2]. Transformed follicular lymphoma (t-FL) poses a clinical challenge due to its rapid progression, resistance to treatment, and poor prognosis, resulting in a median survival of less than 2 years [3, 4]. Active treatment, such as the R-CHOP regimen (rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone), is recommended and considered effective [5]. Although biopsy with immunohistochemical examination is the standard for confirming t-FL [6], it is limited by factors such as sampling bias and the challenge of obtaining specimens from lesions in inaccessible locations. Moreover, biopsy only indicates whether the sampled lesion has undergone transformation, disregarding the intertumoral heterogeneity within the same patient [3]. Therefore, a more comprehensive and non-invasive method is needed to accurately identify t-FL and aid in clinical decision-making.

FL almost always exhibits FDG avidity, irrespective of its grade [7]. 18F-FDG PET/CT is currently considered the standard technique for staging, restaging, and evaluating responses in FL [8]. Previous studies have demonstrated that an SUVmax cutoff of 10 to 13 can differentiate between low-grade and aggressive lymphomas [9, 10]. Other studies have suggested a correlation between transformed and non-transformed low-grade lymphoma on SUVmax, including FL and t-FL [11–13]. However, these studies included a heterogeneous mix of various lymphomas rather than focusing solely on FL and were limited by small sample sizes. Additionally, metabolic parameters like SUVmax have limited ability to reflect tumor heterogeneity, a key feature of histologic transformation.

Artificial intelligence (AI) technology applied in medical imaging facilitates the extraction and analysis of numerous objective and quantifiable features from CT, PET, MRI, and ultrasound images in a high-throughput manner. Certain features may reveal connections between subtle radiological phenotypes and specific aspects of the underlying pathobiology [14], such as the Richter transformation of indolent chronic lymphocytic leukemia into DLBCL [15, 16]. Research conducted at a single center by de Jesus and colleagues demonstrated that machine learning analysis of handcrafted radiomics from [^{18}F]FDG PET/CT scans can effectively differentiate between follicular lymphoma (FL) and DLBCL [17]. Deep learning (DL)-based radiomics has emerged as an advanced quantitative tool, with recent studies highlighting its potential to predict pathological features from medical images [18–20]. Therefore, we conducted this multi-center study, involving five independent medical centers, to leverage radiomics from PET/CT images and construct an R-signature for accurately identifying FL patients with histologic transformation.

Methods

Study population

This study was approved by the institutional review boards of the involved centers (IRB No. 2024–1010). Informed consent was waived due to the retrospective nature of this study. In this multicenter study, 784 patients from five independent medical centers were included: West China Hospital, Sichuan University (center I), Jiangsu Province Hospital, the First Affiliated Hospital of Nanjing Medical University (center II), Nanjing Drum Tower Hospital, the Affiliated Hospital of Nanjing University Medical School (center III), the First Affiliated Hospital of Xiamen University (center IV), and Qilu Hospital of Shandong University (center V). Inclusion criteria were as follows: (1) confirmed histopathological diagnosis of primary FL (I-II or IIIa grades), de novo GCB-DLBCL, and t-FL according to the World Health Organization's classification [21]; (2) no history of other tumors; (3) availability of comprehensive medical records. Exclusion criteria included (1) prior history of other tumors; (2) incomplete medical records; (3) histopathological diagnosis of FL (IIIb grade); and (4) transformations from FL to other histological types (Burkitt or Hodgkin lymphoma). The workflow of patient selection is shown in Fig. 1.

PET/CT scanning protocol

All patients underwent PET/CT scans with one of the following systems: Gemini GXL, UM780PET/CT, Biograph 16 PET/CT, GE discovery PET/CT clarity 710, GE Discovery MI, GE Discovery STE. Patients fasted for at least 6 h before scans, resulting in blood glucose levels under 8.7 mmol/L. Then, 185–370 MBq of [^{18}F]FDG (5.18 MBq/kg) was administered intravenously. The PET/CT scans (from the base of the skull to the upper thigh) were performed 60 min after the radiopharmaceutical injection. Emission data were acquired for 1–2 min in each bed position.

Delineation of target lesions

The lymph node was selected for delineation based on the patient's biopsy records. Semi-automatic delineation was performed using LIFEx-7.3.0 software (https://www.lifexsoft.org/) [22]. The tumor volume of interest

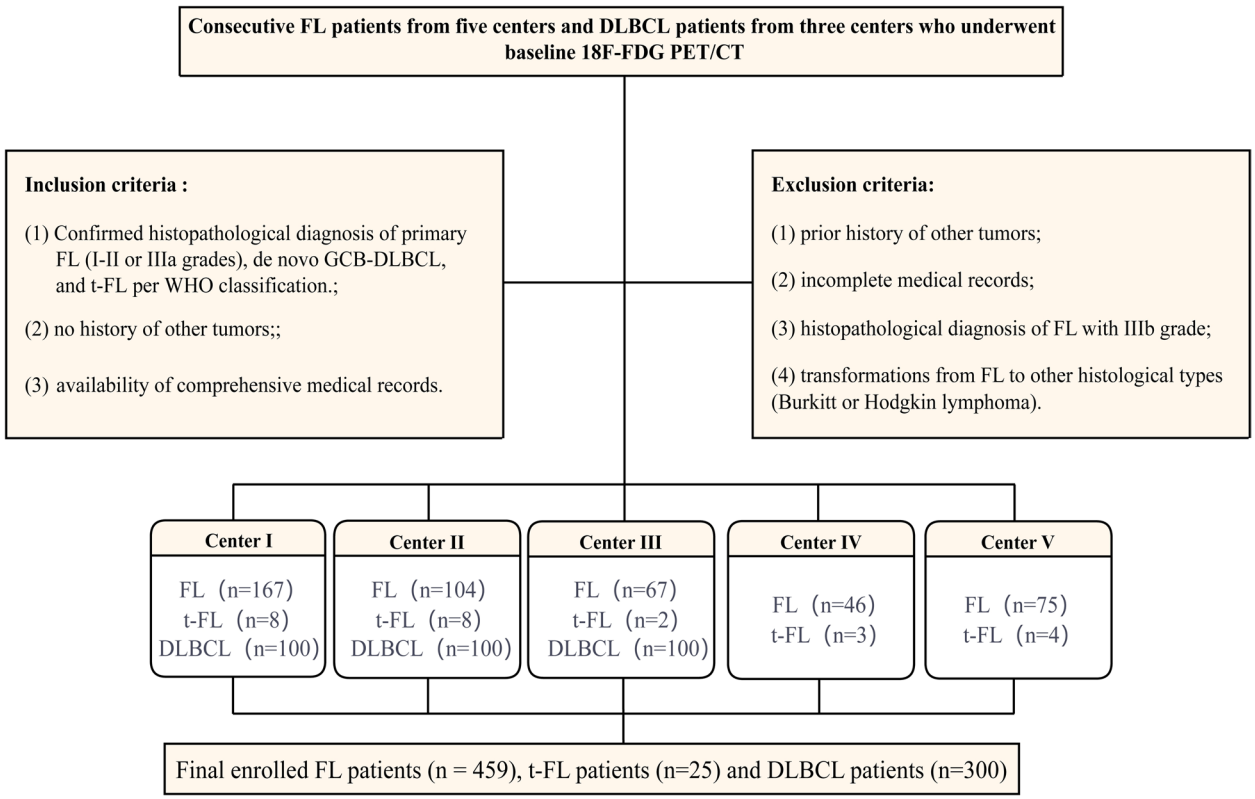


Fig. 1 Flow chart of participant selection

(VOI) was delineated on PET images using a 41% SUVmax threshold method, which allowed for the calculation of the tumor's SUVmax. All PET/CT images were jointly reviewed by two attending nuclear medicine specialists with 8 years of experience. In cases of disagreement, a senior nuclear medicine physician participated and made the final decision. To minimize the impact of PET/CT scanners from different manufacturers on radiomic features, we applied ComBat harmonization (https://github.com/Jfortin1/ComBatHarmonization) to harmonize the PET/CT images.

PET/CT image registration

We employed a CT-to-PET registration approach, with PET images serving as the reference and CT images as the moving images. The registration process was implemented using the open-source library SimpleElastix (https://simpleelastix.github.io/), which is integrated into Python and supports 3D medical image registration. During the registration, the "rigid" registration mode was selected. Once the CT images were successfully registered, we applied nearest-neighbor interpolation to resample the registered CT images, aligning their spatial resolution with that of the PET images.

EMFusion-based PET/CT image fusion

After registration, the CT image is in the same spatial resolution as the PET image. To enhance lymphoma classification performance based on images and to reduce the redundancy of information extracted from the two modalities, we explored the fusion of PET and CT images. Here, we applied the unsupervised EMFusion method proposed by Xu et al. for PET and CT image fusion [23]. The backbone network in the EMFusion model consists of eight convolutional layers, with dense connection layers in the first four. Short connections were established between each layer in the first four, and between the first four and the remaining layers in a feed-forward manner. This architecture alleviates vanishing gradients, enhances feature propagation, and significantly reduces parameters. Lastly, features extracted by the first four layers are fed into subsequent convolutional layers to gradually reduce the number of channels and produce the final fusion image. Before using the trained EMFusion model for PET and CT image fusion, the single-channel grayscale PET and CT slices are converted into RGB three-channel images to meet the input requirement of EMFusion. The image size is uniformly scaled to $180 \times 180 \times 3$ as input, and the output image size is 180×180 single-channel grayscale image. Considering

that feature extraction will be implemented using the lesion region delineated by physicians, the fused gray-scale image is resized to the original PET image size.

Feature extraction and R-signature construction

We constructed a 2D image classification model based on annotated lesion areas after obtaining the fused image. During model training, we used top-down 2D views of the fused images in the training set. ResNet18 was chosen as the backbone network for our classification model. ResNet18 is a well-established architecture known for its effectiveness in preventing overfitting and gradient explosion, while also having a relatively modest number of parameters. Only the biopsied lesion was used to extract radiomic features and deep features. The input images for the classification model are cropped images based on the lesion annotation (lesion ROI), and are resized to 128 × 128 pixels. We used the pyradiomics library (https://pyradiomics.readthedocs.io/en/latest/) in Python for radiomic feature extraction. Each 3D fused lesion ROI yielded 1967 radiomics features. We employed the ResNet18 to extract deep features from 2D lesion areas from each 3D fused lesion ROI, followed by weighted feature averaging. Here, averaging features is to reduce the dimension of the feature vector, and the number of slices is not fixed for different subjects. This process allowed us to extract 16,379 deep features for each 3D fused lesion ROI. Radiomic feature selection utilized ICC (intraclass correlation coefficient) consistency analysis, while LASSO regression analysis was employed for deep feature selection. By comparing the similarity of features extracted from lesion ROIs delineated by two experts, we set a feature selection threshold (e.g., ICC > 0.8) and selected 520 radiomic features. LASSO regression analysis was employed to calculate the importance of deep features, selecting those with an importance greater than 0.

After selecting features, the radiomic and deep features were input into the AutoGluon (a highly integrated auto-machine-learning library, version 0.7.0, available at https://auto.gluon.ai) for classification model training. Thirteen general classifiers (including KneighborsUnif, KNeighborsDist, LightGBMLarge, XGBoost, ExtraTreesGini, ExtraTreesEntr, RandomForestEntr, LightGBM, CatBoost, NeuralNetTorch, LightGBMXT, RandomForestGini, and NeuralNetFastAI) were trained with the AutoGluon library, then integrated into a strong classifier through a weighted ensemble strategy to generate the prediction probability for constructing the signature.

We pooled data from the five centers, comprising a total of 459 cases of FL. We randomly selected 80 FL cases, along with 25 t-FL cases, to constitute the t-FL test cohort. The remaining 379 grade FL cases were combined with 300 GCB-DLBCL cases to form the training and internal validation group. The training and internal validation cohort underwent fivefold cross-validation, producing five models. Each of these models was then applied to the t-FL test cohort. The classification results for the t-FL test cohort were averaged predicted probabilities from the five models, which were used to form the R-signature. The workflow of this study is shown in Fig. 2.

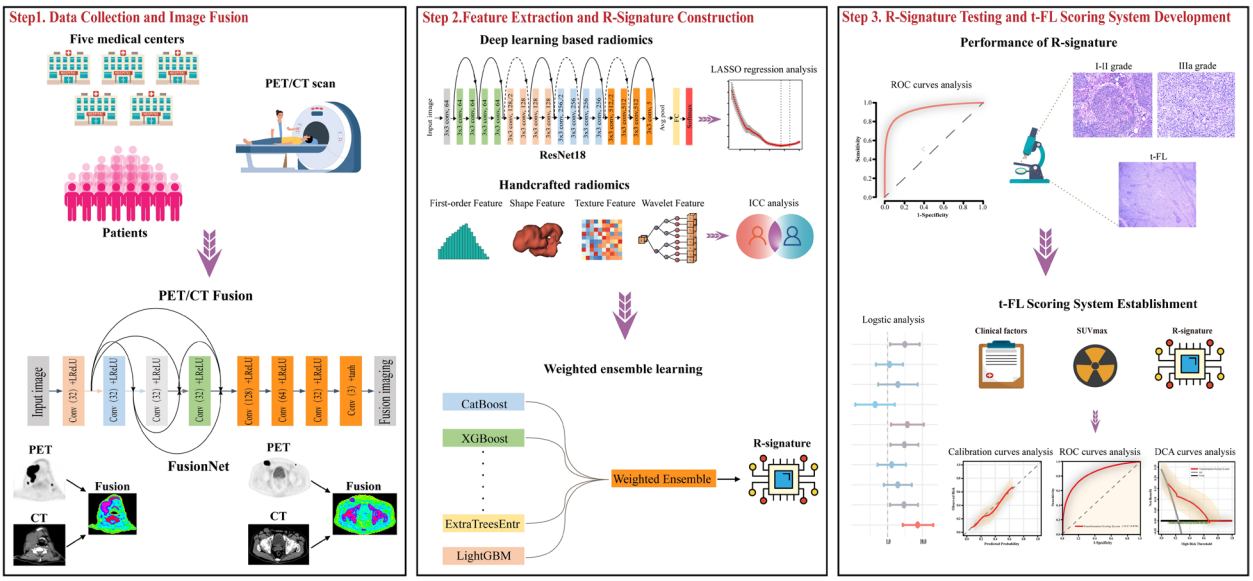


Fig. 2 Analysis workflow in this study

Establishment of t-FL scoring system

Univariate and multivariate logistic regression analyses were used to identify significant risk factors in the t-FL test cohort. These factors were then combined to develop the t-FL scoring system. Each risk factor was assigned 1 point, and the total score for each patient was calculated. Based on the total score, patients were classified into three risk groups: low risk (0–1 points), medium risk (2 points), and high risk (3–5 points). To assess the incremental predictive value of the t-FL scoring system, we developed two competing systems. The first system, referred to as the clinical scoring system, included clinical factors (age, LDH, and ECOG PS). The second system, referred to as the metabolic scoring system, incorporated both clinical factors and SUVmax. Calibration curves and ROC curves were generated for both systems. Additionally, decision curve analysis (DCA) was performed to estimate the false-positive rate for the systems.

Statistical analysis

Data analysis was conducted using IBM SPSS 25 and R software (version 4.2.2, www.R-project.org). Differences in clinical features, pathological characteristics, and SUVmax between the training and validation groups were assessed using the chi-squared test. Discrimination capability was evaluated by estimating the area under the receiver operating characteristic (ROC) curve. Logistic regression analysis was used to explore the predictive value of potential independent response predictors and to construct models. Progression-free survival (PFS) and overall survival (OS) were used as endpoints to assess patient prognosis. Survival analysis was performed using the Kaplan–Meier method, with comparisons made using the log-rank test. A p-value of less than 0.05 was considered statistically significant for all analyses.

Results

Patient characteristics

The baseline characteristics of patients included in the training, internal validation, and t-FL test cohorts are summarized in Table 1. In the t-FL test cohort, the prevalence of t-FL patients was 23.8%. Statistically significant differences were observed between FL and t-FL in age, LDH level, and ECOG PS (all $P < 0.05$).

Performance of R-signature in predicting t-FL

The R-signature in the fivefold cross-validation distinguished between DLBCL and FL, with mean AUC values of 0.994 (ranging from 0.992 to 0.998) in the training cohort and 0.976 (ranging from 0.961 to 0.991) in the internal validation cohort (see Fig. 3). For the t-FL test cohort, the mean R-signature from the fivefold cross-validation achieved an AUC of 0.749 (95% CI 0.635 to 0.863) with optimal cutoff values of 0.217 (see Figs. 3 and 4).

Comparison between SUVmax and R-signature in predicting t-FL

The optimal cutoff value of SUVmax was 9.65, with an AUC of 0.683 (95% CI 0.574 to 0.793). Sensitivity (SE), specificity (SP), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (PLR), negative likelihood ratio (NLR), and accuracy in predicting t-FL are listed in Table 2. The differences in the clinical characteristics between the dichotomized R-signature and SUVmax groups are shown in Table 3.

Predictive performance of R-signature

The R-signature demonstrated strong predictive performance, significantly distinguishing between PFS and OS outcomes in the t-FL test cohorts (see in Fig. 5). Statistically significant differences were observed between the high and low R-signature groups, with P values of 0.029 for PFS and 0.023 for OS. Table 4 provides a detailed comparison between the prognostic stratification based on pathological results and model predictions.

Multivariable analysis for t-FL prediction

The univariate and multivariate analysis results revealed that clinical variables, including age (OR = 3.502, 95% CI 1.181–10.380; $P = 0.024$), LDH (OR = 5.171, 95% CI 1.529–17.490; $P = 0.008$), and ECOG PS (OR = 3.231, 95% CI 1.174–8.891; $P = 0.023$), were identified as independent predictors of t-FL. Furthermore, SUVmax (OR = 3.252, 95% CI 1.136–9.309; $P = 0.028$) and R-signature (OR = 6.069, 95% CI 2.187–16.845; $P = 0.001$) were also identified as independent predictors of t-FL (see in Fig. 6).

Establishment and assessment of t-FL scoring system

The t-FL scoring system, incorporating R-signature, SUVmax, age, LDH, and ECOG PS status, was developed using the t-FL test cohort (see in Fig. 7). The patients were divided into three risk groups: low-risk group (55 participants), medium-risk group (29 participants), and high-risk group (21 participants). In a subanalysis, the high-risk group (14 t-FL in 21 patients, 66.7%) had a relatively higher rate than those in the low-risk group (4 t-FL in 55 patients, 4.5%) and the medium-risk group (7 t-FL in 29 patients, 24.1%) (see in Fig. 7). Calibration curves for the t-FL scoring system indicated good agreement between predictions and actual observations (see in Fig. 7). The t-FL scoring system demonstrated strong performance in t-FL prediction with an AUC of 0.820 (95% CI 0.725 to 0.914) (see in Fig. 7). It outperformed

Table 1 Characteristics of the study population

Characteristic	Training and internal validation cohort		t-FL test cohort		*P* value*
	FnL (*n* = 379)	DLBCL (*n* = 300)	FL (*n* = 80)	t-FL (*n* = 25)	
Gender					
Male	171	167	36	12	0.822
Female	208	133	44	13	
Age (years)					
Median#	55	51	49	57	0.019
Q1–Q3	45–67	42–59	41–58	48–68	
Elevate LDH					
No	307	185	71	17	0.026
Yes	72	115	9	8	
ECOG PS					
0–1	356	271	59	12	0.026
≥ 2	23	29	21	13	
Ann arbor staging					
I–II	81	143	17	3	0.391
III–IV	298	157	63	22	
B symptoms					
No	317	223	61	22	0.268
Yes	62	77	19	3	
Hemoglobin < 120 g/L					
No	292	-	56	16	0.625
Yes	87	-	24	9	
Elevate Serum β2-MG					
No	189	-	42	9	0.174
Yes	190	-	38	16	

Abbreviations: *DLBCL* diffuse large B-cell lymphoma, *FL* follicular lymphoma, *t-FL* transformed follicular lymphoma, *LDH* lactate dehydrogenase, *ECOG PS* Eastern Cooperative Oncology Group performance status, *LDH* lactate dehydrogenase, *β2-MG* β2-microglobulin

Median (range)

* *P* value derived from the Mann–Whitney U test or χ^2 test in t-FL test cohort

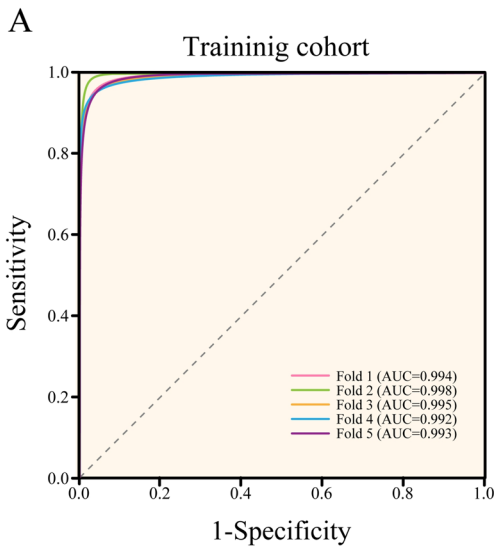
A *P*-value of < 0.05 indicates statistical significance

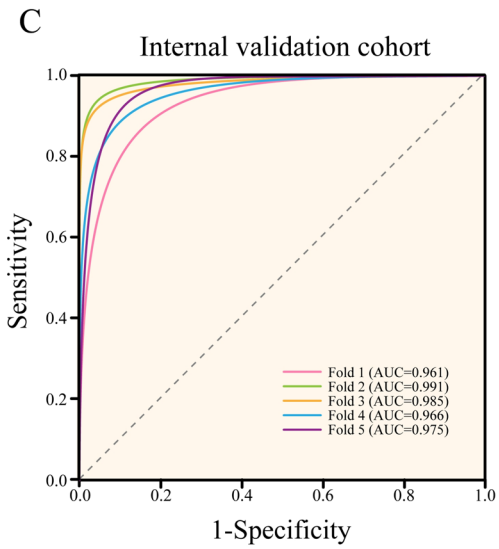
the clinical scoring system based solely on ECOG PS and LDH, which had an AUC of 0.739 (95% CI 0.641 to 0.836), and the metabolic scoring system incorporating ECOG, LDH, and SUVmax, which had an AUC of 0.741 (95% CI 0.642 to 0.839). Furthermore, DCA indicated that the t-FL scoring system provided a greater overall net benefit compared to competing scoring systems across most risk thresholds (see in Fig. 7).

Discussion

This study aimed to develop a non-invasive method using PET/CT imaging to identify histological transformation in follicular lymphoma (t-FL). Our results showed that the R-signature, derived from radiomics analysis, exhibited strong predictive power for identifying t-FL. By combining the R-signature with clinical parameters and SUVmax, we developed a t-FL scoring system that effectively stratifies patients based on their risk of transformation, thereby aiding clinical decision-making.

FDG-PET imaging for non-Hodgkin lymphoma (NHL) often requires time-consuming manual analysis to accurately quantify tumor burden for risk assessments. The integration of artificial intelligence (AI) in medical imaging, however, has significantly advanced automated segmentation in PET imaging for NHL. Deep learning, for instance, can now automatically segment DLBCL lesions and compute tumor volume, greatly reducing the time and effort required by clinicians [24–26]. Additionally, AI-based PET imaging has shown exceptional performance in virtual biopsy applications. For example, assessing bone marrow involvement in NHL is crucial for clinical decision-making [27, 28]. Machine learning-selected radiomic features in AI-based PET imaging can offer a promising non-invasive method for evaluating bone marrow involvement in DLBCL and FL [29, 30]. Moreover, radiomics, which extracts relevant imaging information, is considered capable of effectively assessing the three-dimensional





B

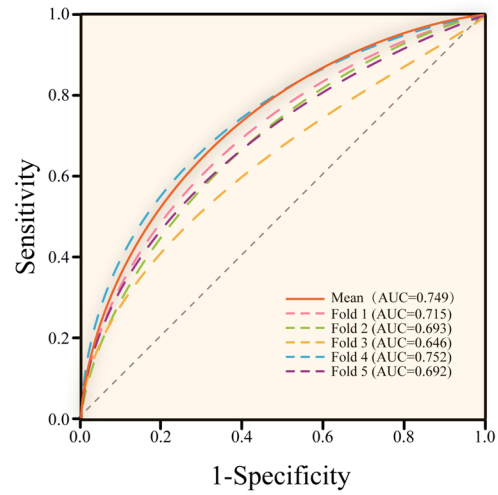
The performance of R-signature in training cohort

R-signature	AUC (95% CI)	ACC (95% CI)	SEN (95% CI)	SPE (95% CI)
Fold 1	0.994 (0.990-0.998)	97.5% (97.5%-97.5%)	97.9% (96.1%-99.7%)	97.2% (95.4%-99.0%)
Fold 2	0.998 (0.996-1.000)	98.4% (98.4%-98.4%)	99.2% (98.0%-100.0%)	97.8% (96.2%-99.4%)
Fold 3	0.995 (0.990-1.000)	98.6% (98.6%-98.6%)	97.9% (96.1%-99.7%)	99.1% (98.0%-100.0%)
Fold 4	0.992 (0.986-0.998)	97.7% (97.7%-97.7%)	96.7% (94.4%-98.9%)	98.4 % (97.1%-99.8%)
Fold 5	0.993 (0.988-0.997)	96.6% (96.6%-96.6%)	93.3% (90.2%-96.5%)	99.1% (98.0%-100.0%)

D

The performance of R-signature in internal validation cohort

R-signature	AUC (95% CI)	ACC (95% CI)	SEN (95% CI)	SPE (95% CI)
Fold 1	0.961 (0.922-1.000)	95.0% (94.9%-95.1%)	95.0% (89.5%-100.0%)	95.0% (90.2%-99.8%)
Fold 2	0.991 (0.982-1.000)	95.7% (95.7%-95.8%)	95.0% (89.5%-100.0%)	96.3% (92.1%-100.0%)
Fold 3	0.985 (0.967-1.000)	97.1% (97.1%-97.2%)	96.7% (92.1%-100.0%)	97.5% (94.1%-100.0%)
Fold 4	0.966 (0.937-0.995)	94.3% (94.2%-94.4%)	91.7% (84.7%-98.7%)	96.3% (92.1%-100.0%)
Fold 5	0.975 (0.949-1.000)	95.0% (94.9%-95.1%)	91.7% (84.7%-98.7%)	97.5% (94.1%-100.0%)



F

The performance of R-signature in t-FL test cohort

R-signature	AUC (95% CI)	ACC (95% CI)	SEN (95% CI)	SPE (95% CI)
Fold 1	0.715 (0.595-0.834)	63.8% (63.4%-64.2%)	80.0% (64.3%-95.7%)	58.8% (48.0%-69.5%)
Fold 2	0.693 (0.570-0.816)	75.2% (74.9%-75.6%)	64.0% (45.2%-82.8%)	78.7% (69.8%-87.7%)
Fold 3	0.646 (0.506-0.787)	73.3% (73.0%-73.7%)	60.0% (40.8%-79.2%)	77.5% (68.3%-86.7%)
Fold 4	0.752 (0.637-0.867)	80.0% (79.7%-80.3%)	52.0% (32.4%-71.6%)	88.7% (81.8%-95.7%)
Fold 5	0.692 (0.561-0.822)	72.4% (72.0%-72.8%)	68.0% (49.7%-86.3%)	73.8% (64.1%-83.4%)
Mean	0.749 (0.635-0.863)	75.2% (74.9%-75.6%)	68.0% (49.7%-86.3%)	77.5% (68.3%-86.7%)

Fig. 3 Performance of the R-signature in distinguishing between DLBCL and FL, and in t-FL. **A** The ROC curves for each fold in the fivefold cross-validation in the training cohort. **B** Performance metrics in the training cohort. **C** The ROC curves for each fold in the fivefold cross-validation in the internal validation cohort. **D** Performance metrics in the internal validation cohort. **E** The ROC curves for each fold in the fivefold cross-validation and the mean in the t-FL test cohort. **F** Performance metrics in the t-FL test cohort

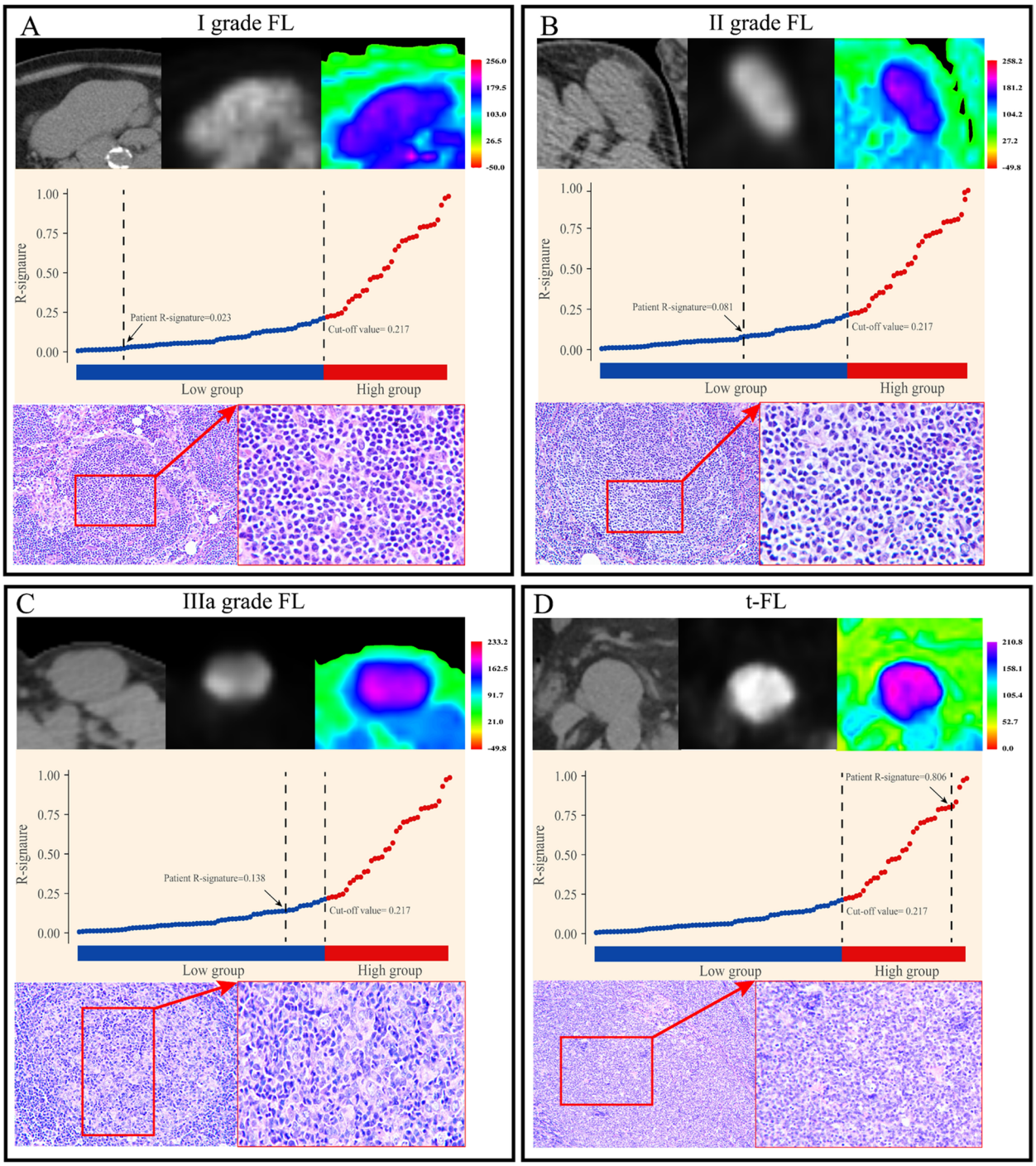


Fig. 4 Comparison of PET/CT fusion imaging, R-signature, and histopathology for FL (I, II, and IIIa grades) and t-FL. **A** I Grade FL: CT, PET, and fusion images; R-signature score 0.023 (low risk); HE staining shows predominantly centrocytes within the tumor follicles with slightly irregular nuclei. **B** II Grade FL: CT, PET, and fusion images; R-signature score 0.081 (low risk); HE staining shows tumor follicles containing both centrocytes and centroblasts, with a predominance of centrocytes. **C** IIIa Grade FL: CT, PET, and fusion images; R-signature score 0.138 (low risk); HE staining shows numerous centroblasts scattered among centrocytes. **D** t-FL: CT, PET, and fusion images; R-signature score 0.806 (high risk); HE staining shows sheets of aggregated centroblasts

Table 2 Prediction of t-FL with SUVmax and R-signature

Parameters	SE	SP	PPV	NPV	PLR	NLR	ACC	AUC
SUVmax	64.0%	62.5%	34.8%	84.7%	1.71	0.576	62.9%	0.683
R-signature	68.0%	77.5%	48.6%	88.6%	3.02	0.413	75.2%	0.749

Prediction of survival based on evaluation with SUVmax and R-signature

SE sensitivity, *SP* specificity, *PPV* positive predictive value, *NPV* negative predictive value, *PLR* positive likelihood ratio, *NLR* negative likelihood ratio, *ACC* accuracy, *AUC* area under the curve

Table 3 Comparison of the R-signature with the patient clinical data in the t-FL test cohort

Characteristic		R-signature		*P* value*	SUVmax		*P* value*
		Low ($n=70$)	High ($n=35$)		Low ($n=59$)	High ($n=46$)	
Gender	Male	28	20	0.103	32	25	1.000
	Female	42	15		27	21	
Age (years)	< 60	60	21	0.006	48	33	0.254
	≥ 60	10	14		11	13	
Elevate LDH	no	65	23	0.001	52	36	0.192
	yes	5	12		7	10	
ECOG PS	0–1	59	12	0.026	41	30	0.687
	≥ 2	21	13		18	16	
Ann arbor staging	I–II	11	9	0.292	13	7	0.457
	III–IV	59	26		46	39	
B symptoms	No	54	29	0.615	47	36	1.000
	Yes	16	6		12	10	
Hemoglobin < 120 g/L	No	45	27	0.265	39	33	0.672
	Yes	25	8		20	13	
Elevate Serum β2-MG	No	38	13	0.147	32	19	0.293
	Yes	32	22		27	27	

Abbreviations: *LDH* lactate dehydrogenase, *ECOG PS* Eastern Cooperative Oncology Group performance status, *LDH* lactate dehydrogenase, *β2-MG* β2-microglobulin

* *P* value derived from χ^2 test in t-FL test cohort

A *P*-value of <0.05 indicates statistical significance

tumor landscape and decoding the phenotypes of various histological structures [31]. In our study, the R-signature demonstrated high accuracy and AUC values, ranging from 0.992 to 0.998 in the training cohort and 0.961 to 0.991 in the internal validation cohort, effectively distinguishing between FL and DLBCL. These results are consistent with previous studies, such as those by de Jesus et al. [17], which highlighted the potential of machine learning and handcrafted radiomic features from [18F]FDG PET/CT scans in differentiating FL and DLBCL. This indicates that the R-signature constructed by radiomics effectively captures the complex imaging patterns associated with t-FL and demonstrates good transportability. Furthermore, the R-signature outperformed SUVmax in terms of sensitivity, specificity, PPV, NPV, and accuracy. SUVmax, a conventional metabolic parameter, can be influenced by inflammatory cell uptake. Studies have shown that high FDG avidity in FL is linked to inflammatory cell uptake in the tumor microenvironment, rather than tumor cells [32].

Our study demonstrates that the R-signature has significant prognostic predictive value. Previous research has confirmed that t-FL exhibits a markedly poorer prognosis compared to FL [3, 4]. A multicenter study in the USA reported that patients with t-FL have an estimated 5-year overall survival (OS) rate of only 49%, which is even lower than that of de novo DLBCL, with an OS rate of 57% [33]. As shown in Fig. 5 and Table 4, while the R-signature demonstrates slightly lower performance in prognostic stratification compared to pathological results, it represents a promising non-invasive biomarker that could serve as a viable alternative for assessing prognosis in patients.

To enhance image fusion, we employed the unsupervised EMFusion method, proposed by Xu et al., for PET/CT image integration [23]. Unlike traditional fusion methods such as DDcGAN and CNN, EMFusion

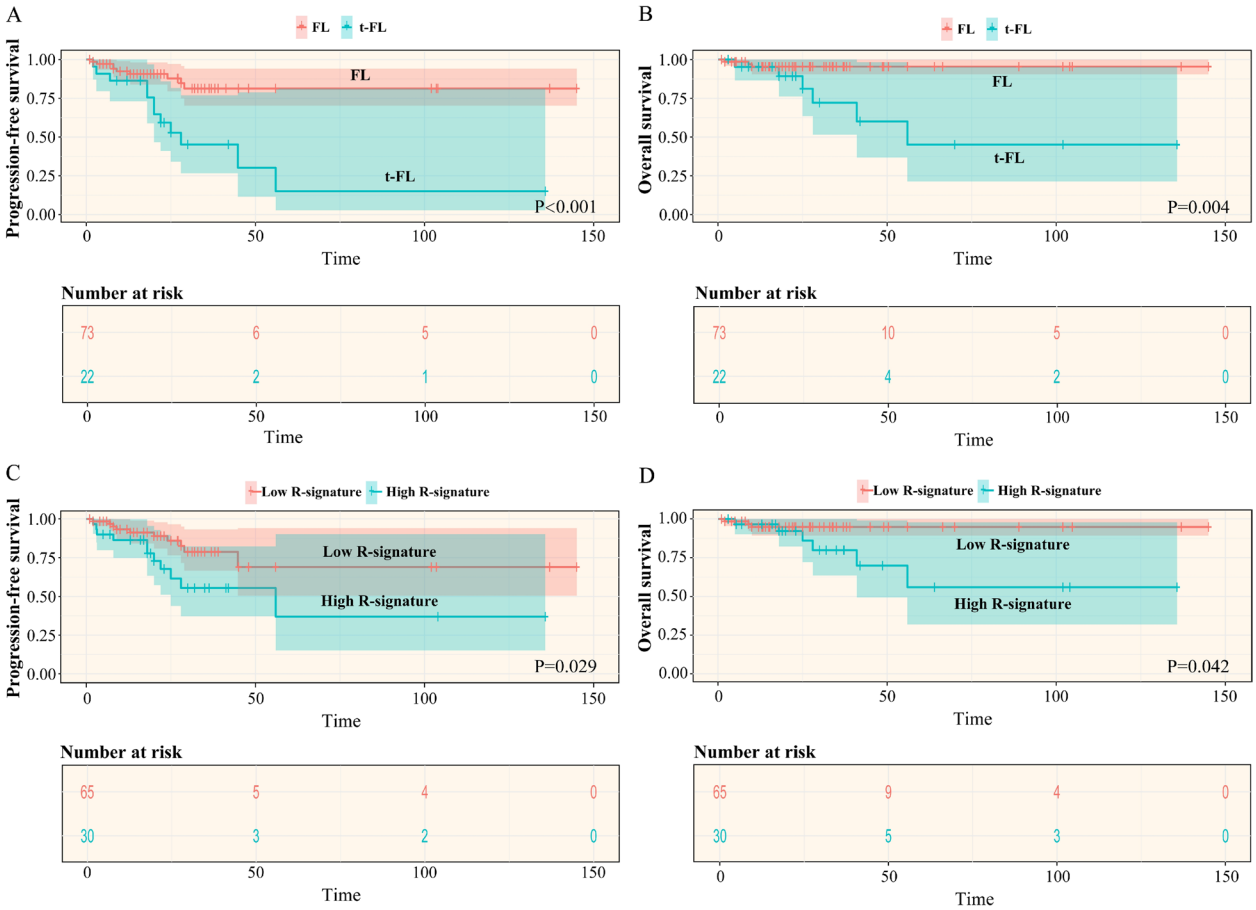


Fig. 5 Kaplan–Meier survival analyses of PFS and OS according to the pathological result (**A**, **B**) and R-signature (**C**, **D**)

accounts for structural information from the source images (e.g., constraints based on salience and abundance) [34–39]. It preserves and enhances distinct information from each modality by introducing depth-level constraints and considering chromatic information in PET images. This results in fused images that combine the structural details of CT with the functional information of PET. Furthermore, we utilized deep learning networks for feature extraction, enabling more robust and comprehensive analysis of the imaging data. In the classification stage, we applied the automated machine learning model AutoGluon, which trains multiple weak classifiers and generates a strong ensemble classifier using weighted methods [40]. This ensemble approach helps mitigate overfitting, provides a more comprehensive analysis of the features, and ensures more stable and reliable predictive performance.

In the present study, age, LDH, and ECOG PS were identified as independent clinical predictors of t-FL. Additionally, an SUVmax of 9.65 was determined as the optimal cut-off value for distinguishing between FL and t-FL, with multivariate logistic analysis indicating that SUVmax is also an independent predictor (OR = 3.252, 95% CI 1.136–9.309; $P = 0.028$). These findings are consistent with previous studies [9–13, 41–43]. The t-FL scoring system we developed incorporates the R-signature, along with SUVmax, age, LDH, and ECOG PS. This combined approach achieved an AUC of 0.820, outperforming other scoring systems, thereby highlighting the added value of integrating radiomics-based imaging biomarkers into conventional assessment protocols. The scoring system stratified patients into low-risk (4 t-FL in 55 patients, 4.5%), medium-risk (7 t-FL in 29 patients, 24.1%), and high-risk (14 t-FL in 21 patients, 66.7%) groups, with significantly different proportions of t-FL, facilitating more tailored treatment strategies.

Our study has several limitations. First, due to its retrospective design, there may be inherent selection bias. Second, we observed a reduction in diagnostic performance when the R-signature was applied to the t-FL test cohort. Following the approach of de Jesus et al. [17], DLBCL was selected as the positive control in the training and internal validation sets. Despite the clinical and pathological similarities between DLBCL and t-FL, including

Table 4 Prognostic performance of classification results using R-signature and Pathological result in the test cohorts

Progression-free survival						**Overall survival**					
R-signature			Pathological result			R-signature			Pathological result		
Classification	Survival (%)	χ^2	Classification	Survival (%)	χ^2	Classification	Survival (%)	χ^2	Classification	Survival (%)	χ^2
Low	84.6	4.752	FL	87.7	13.244	Low	95.4	4.118	FL	95.9	8.287
High	63.3		t-FL	45.5		High	80.0		t-FL	72.7	

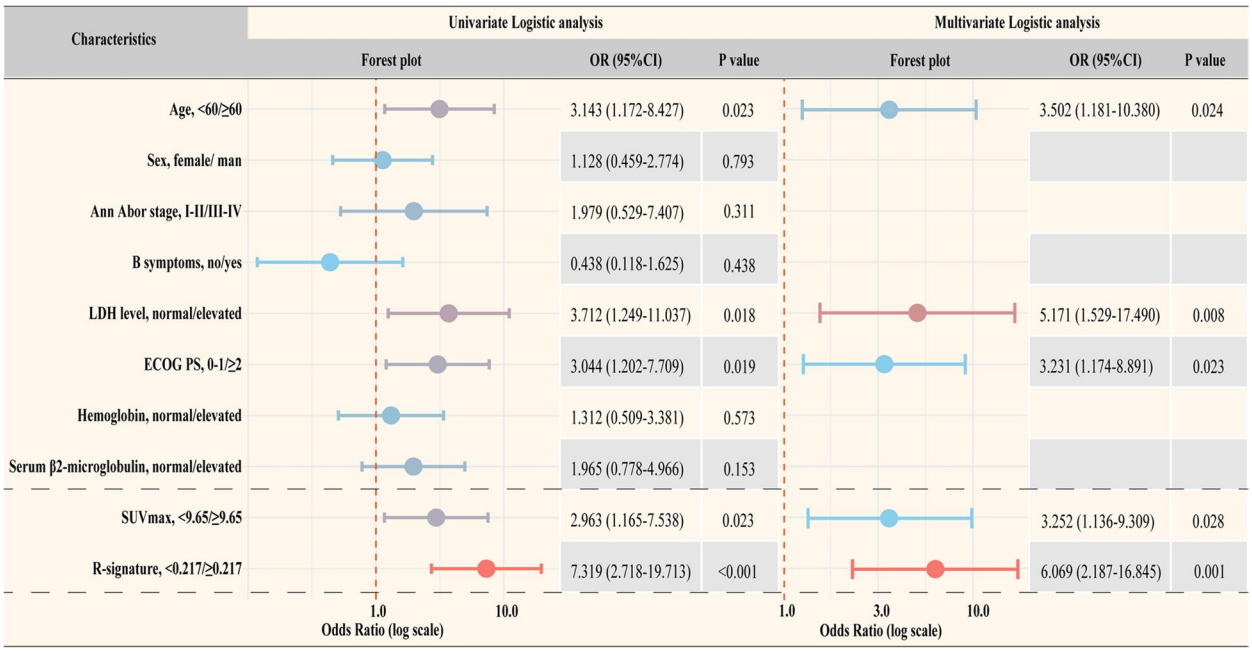


Fig. 6 Forest plot showing the underlying predictors of t-FL based on the results of univariate and multivariate logistic analyses

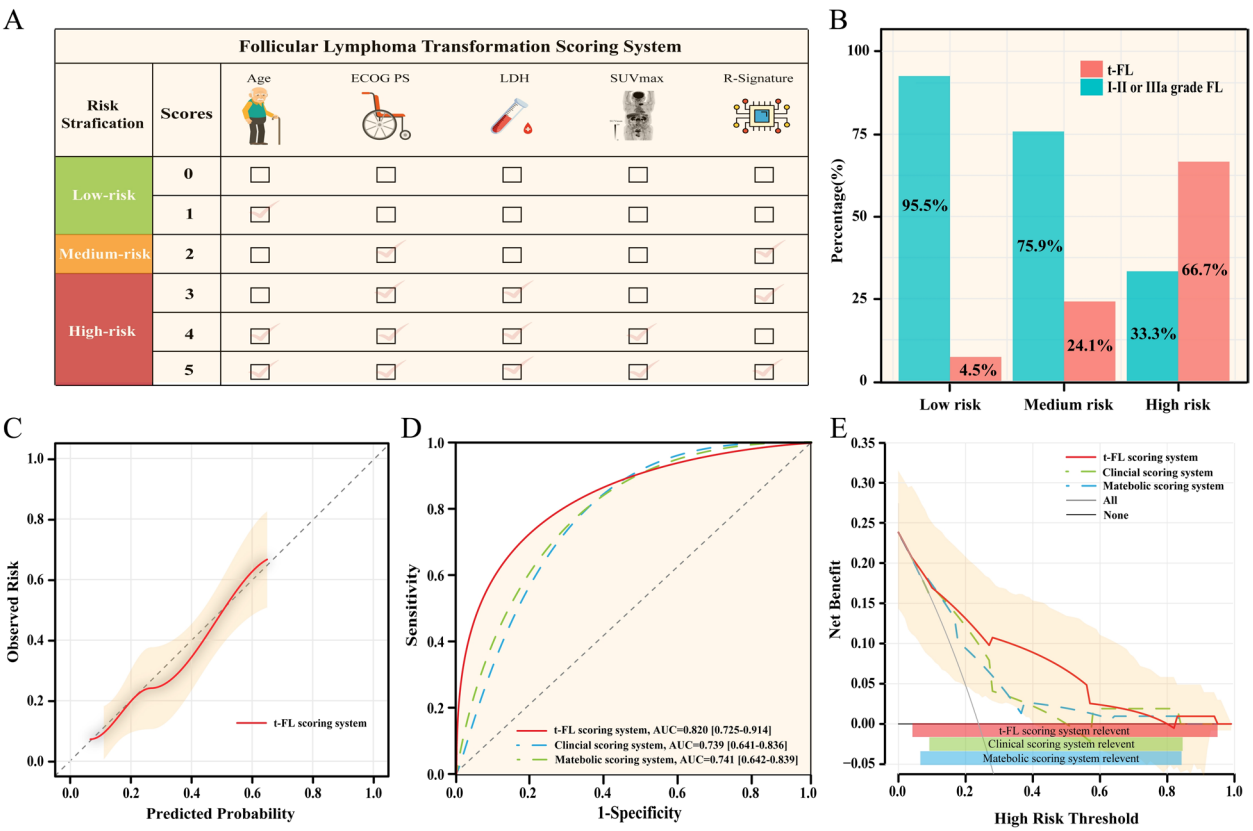


Fig. 7 The t-FL scoring system and its evaluation. **A** The t-FL scoring system classifies patients into low, medium, and high-risk groups based on their scores. **B** Bar chart showing the percentage of t-FL patients in the low, medium, and high-risk groups. **C** Calibration curve showing the observed risk versus predicted probability for the t-FL scoring system. **D** ROC curves for the t-FL scoring system, clinical scoring system, and metabolic scoring system. **E** Decision curve analysis comparing the net benefit of different scoring systems across various high-risk thresholds

comparable morphology [44], there are inherent differences in their genetic mutation patterns and biological characteristics [45, 46], which may explain the observed variations in model performance between the internal validation and t-FL test cohorts. Moreover, due to the low incidence of t-FL, only 25 t-FL samples were included in the test cohort. This limited sample size may have further constrained the model's performance when evaluated as an independent test set. Future studies should include a larger t-FL dataset and establish comprehensive databases to improve model performance and generalizability. Additionally, data augmentation techniques, such as Generative Adversarial Networks (GANs) and Diffusion models, should be explored to expand sample sizes for rare diseases and enhance model robustness. Third, there was heterogeneity in PET/CT image acquisition across the patient cohorts from five independent medical centers, which may have influenced the extracted features and, consequently, the performance of the model.

Conclusions

This study demonstrates the feasibility and effectiveness of using radiomics analysis of PET/CT images for the non-invasive identification of t-FL. The t-FL scoring system developed in this study provides a valuable tool for clinical decision-making, with the potential to improve patient management and outcomes.

Abbreviations

FL	Follicular lymphoma
t-FL	Transformed follicular lymphoma
GCB-DLBCL	Diffuse large B-cell lymphoma with germinal center phenotype
R-signature	Radiomic signature
AI	Artificial intelligence
DL	Deep learning
VOI	Volume of interest
DCA	Decision curve analysis
ROC	Receiver operating characteristic
PFS	Progression-free survival
OS	Overall survival

Acknowledgements

Not applicable.

Authors' contributions

Study concept and design: R.T., C.D., X.L; Acquisition of data: Q.J., H.Z., Z.J., Y.T., X.L., B.X.; Analysis and interpretation of data: Q.J., H.Z., Z.J., Y.T., X.L., B.X.; Drafting of the manuscript: C.J., C.Q.; Critical revision of the manuscript: R.T., C.D., X.L; Statistical analysis: C.J., C.Q.; Obtained funding: R.T.; Administrative support: R.T., C.D., X.L; Study supervision: R.T., C.D., X.L. All authors read and approved the final manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (grant 81971653), the 1·3·5 project for disciplines of excellence, West China Hospital, Sichuan University (grant ZYAI24014), and China Postdoctoral Science Foundation (grant 2024M762244).

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

The study protocol was conducted in accordance with ethical guidelines (Declaration of Helsinki and Istanbul) and approved by the Institutional Ethics Committee of involved centers (IRB No. 2024–1010). Informed consent was waived due to the retrospective nature of this study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Nuclear Medicine, West China Hospital, Sichuan University, Guoxue Alley, Address: No.37, Chengdu City, Sichuan 610041, China. [2]School of Electrical and Information Engineering, Changzhou Institute of Technology, Changzhou, Jiangsu, China. [3]The Affiliated Changzhou NO.2 People's Hospital of Nanjing Medical University, Changzhou, China. [4]Center of Medical Physics, Nanjing Medical University, Changzhou, China. [5]Department of Hematology, School of Medicine, The First Affiliated Hospital of Xiamen University and Institute of Hematology, Xiamen University, Xiamen, China. [6]Department of Nuclear Medicine, Qilu Hospital of Shandong University, No.107, Wenhua Xilu, Lixia District, Jinan City, Shandong 250012, China. [7]Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu, China. [8]Department of Nuclear Medicine, Nanjing Drum Tower Hospital, the Affiliated Hospital of Nanjing University Medical School, Nanjing, China. [9]Department of Nuclear Medicine, The First Affiliated Hospital of Nanjing Medical University, Jiangsu Province Hospital, No.321, Zhongshan Road, Nanjing City, Jiangsu Province 210008, China.

Received: 29 September 2024 Accepted: 22 January 2025
Published online: 29 January 2025

References

1. Dada R. Diagnosis and management of follicular lymphoma: a comprehensive review. Eur J Haematol. 2019;103:152–63.
2. Maeshima AM. Histologic transformation of follicular lymphoma: pathologists' viewpoint. J Clin Exp Hematop. 2023;63:12–8.
3. Fischer T, Zing NPC, Chiattone CS, Federico M, Luminari S. Transformed follicular lymphoma. Ann Hematol. 2018;97:17–29.
4. Mendez M, Torrente M, Provencio M. Follicular lymphomas and their transformation: past and current research. Expert Rev Hematol. 2017;10:515–24.
5. Montoto S. Treatment of patients with transformed lymphoma. Hematology Am Soc Hematol Educ Program. 2015;2015:625–30.
6. Casulo C, Burack WR, Friedberg JW. Transformed follicular non-Hodgkin lymphoma. Blood. 2015;125:40–7.
7. Wöhrer S, Jaeger U, Kletter K, Becherer A, Hauswirth A, Turetschek K, et al. 18F-fluoro-deoxy-glucose positron emission tomography (18F-FDG-PET) visualizes follicular lymphoma irrespective of grading. Ann Oncol. 2006;17:780–4.
8. Díaz-Silván A, Otón-Sánchez LF, Caresia-Aróztegui AP, Del Puig C-S, Orcajo-Rincón J, de Arcocha-Torres M, et al. Clinical application of (18F) FDG PET/CT in follicular lymphoma. Rev Esp Med Nucl Imagen Mol (Engl Ed). 2022;41:202–12.
9. Schöder H, Noy A, Gönen M, Weng L, Green D, Erdi YE, et al. Intensity of 18fluorodeoxyglucose uptake in positron emission tomography distinguishes between indolent and aggressive non-Hodgkin's lymphoma. J Clin Oncol. 2005;23:4643–51.
10. Ngeow JYY, Quek RHH, Ng DCE, Hee SW, Tao M, Lim LC, et al. High SUV uptake on FDG-PET/CT predicts for an aggressive B-cell lymphoma in a prospective study of primary FDG-PET/CT staging in lymphoma. Ann Oncol. 2009;20:1543–7.
11. Noy A, Schöder H, Gönen M, Weissler M, Ertelt K, Cohler C, et al. The majority of transformed lymphomas have high standardized uptake

values (SUVs) on positron emission tomography (PET) scanning similar to diffuse large B-cell lymphoma (DLBCL). Ann Oncol. 2009;20:508–12.
12. Bodet-Milin C, Kraeber-Bodéré F, Moreau P, Campion L, Dupas B, Le Gouill S. Investigation of FDG-PET/CT imaging to guide biopsies in the detection of histological transformation of indolent lymphoma. Haematologica. 2008;93:471–2.
13. Wondergem MJ, Rizvi SN, Jauw Y, Hoekstra OS, Hoetjes N, van de Ven PM, et al. 18F-FDG or 3'-deoxy-3'-18F-fluorothymidine to detect transformation of follicular lymphoma. J Nucl Med. 2015;56:216–21.
14. Zhang C, Gu J, Zhu Y, Meng Z, Tong T, Li D, et al. AI in spotting high-risk characteristics of medical imaging and molecular pathology. Precis Clin Med. 2021;4:271–86.
15. Reinert CP, Federmann B, Hofmann J, Bösmüller H, Wirths S, Fritz J, et al. Computed tomography textural analysis for the differentiation of chronic lymphocytic leukemia and diffuse large B cell lymphoma of Richter syndrome. Eur Radiol. 2019;29:6911–21.
16. Albano D, Calabrò A, Dondi F, Bertagna F. 2-[18F]-FDG PET/CT Semiquantitative and radiomics predictive parameters of Richter's transformation in CLL patients. Medicina (Kaunas). 2024;60(2):203.
17. de Jesus FM, Yin Y, Mantzorou-Kyriaki E, Kahle XU, de Haas RJ, Yakar D, et al. 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. 2022;49:1535–43.
18. Li M, Fan Y, You H, Li C, Luo M, Zhou J, et al. Dual-energy CT deep learning radiomics to predict macrotrabecular-massive hepatocellular carcinoma. Radiology. 2023;308:e230255.
19. Mu W, Jiang L, Shi Y, Tunali I, Gray JE, Katsoulakis E, et al. Non-invasive measurement of PD-L1 status and prediction of immunotherapy response using deep learning of PET/CT images. J Immunother Cancer. 2021;9:e002118.
20. Jiang M, Zhang D, Tang SC, Luo XM, Chuan ZR, Lv WZ, et al. Deep learning with convolutional neural network in the assessment of breast cancer molecular subtypes based on US images: a multicenter retrospective study. Eur Radiol. 2021;31:3673–82.
21. Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. Blood. 2016;127:2375–90.
22. Nioche C, Orlhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, et al. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. Cancer Res. 2018;78:4786–9.
23. Xu H, Liang P, Yu W, Jiang J, Ma J. Learning a generative model for fusing infrared and visible images via conditional generative adversarial network with dual discriminators. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19). 2019:3954–3960.
24. Capobianco N, Meignan M, Cottereau AS, Vercellino L, Sibille L, Spottiswoode B, et al. Deep-learning 18F-FDG uptake classification enables total metabolic tumor volume estimation in diffuse large B-cell lymphoma. J Nucl Med. 2021;62(1):30–6.
25. Jemaa S, Fredrickson J, Carano RAD, Nielsen T, de Crespigny A, Bengtsson T. Tumor segmentation and feature extraction from whole-body FDG-PET/CT using cascaded 2D and 3D convolutional neural networks. J Digit Imaging. 2020;33(4):888–94.
26. Jemaa S, Paulson JN, Hutchings M, Kostakoglu L, Trotman J, Tracy S, et al. Full automation of total metabolic tumor volume from FDG-PET/CT in DLBCL for baseline risk assessments. Cancer Imaging. 2022;22(1):39.
27. Khan AB, Barrington SF, Mikhaeel NG, Hunt AA, Cameron L, Morris T, et al. PET-CT staging of DLBCL accurately identifies and provides new insight into the clinical significance of bone marrow involvement. Blood. 2013;122(1):61–7.
28. Cerci JJ, Györke T, Fanti S, Paez D, Meneghetti JC, Redondo F, et al. Combined PET and biopsy evidence of marrow involvement improves prognostic prediction in diffuse large B-cell lymphoma. J Nucl Med. 2014;55(10):1591–7.
29. Aide N, Talbot M, Fruchart C, Damaj G, Lasnon C. Diagnostic and prognostic value of baseline FDG PET/CT skeletal textural features in diffuse large B cell lymphoma. Eur J Nucl Med Mol Imaging. 2018;45(5):699–711.
30. Faudemer J, Aide N, Gac AC, Damaj G, Vilque JP, Lasnon C. Diagnostic value of baseline 18FDG PET/CT skeletal textural features in follicular lymphoma. Sci Rep. 2021;11(1):23812.
31. Avanzo M, Wei L, Stancanello J, Vallières M, Rao A, Morin O, et al. Machine and deep learning methods for radiomics. Med Phys. 2020;47:e185–202.
32. Péricart S, Martin-Blondel G, Franchet C, Hitzel A, Brousset P. 18F-fluorodeoxyglucose positron emission tomography computed tomography detection threshold in follicular lymphoma: a case report. Medicine (Baltimore). 2017;96(47):e8705.
33. Vaughn JL, Epperla N. Survival of patients with transformed follicular lymphoma in the United States: a multiple cohort study. Biomark Res. 2023;11(1):84.
34. Han Xu, Ma J. EMFusion: an unsupervised enhanced medical image fusion network. Information Fusion. 2021;76:177–86.
35. Ma J, Han Xu, Jiang J, Mei X, Zhang XP. DDcGAN: a dual-discriminator conditional generative adversarial network for multi-resolution image fusion. IEEE Transactions on Image Processing. 2020;29:4980–95.
36. Huang J, Le Z, Ma Y, Fan F, Zhang H, Yang L. MGMDcGAN: medical image fusion using multi-generator multi-discriminator conditional generative adversarial network. IEEE Access. 2020;8:55145–57.
37. Liu Y, Chen X, Cheng J, Peng H. A medical image fusion method based on convolutional neural networks. In: Proc. 20th International Conference on Information Fusion (Fusion). Xi'an: Conference; 2017. p. 1–7.
38. Yu Zhang Yu, Liu PS, Yan H, Zhao X, Zhang Li. IFCNN: a general image fusion framework based on convolutional neural network. Information Fusion. 2020;54:99–118.
39. Liu R, Liu J, Jiang Z, Fan X, Luo Z. A bilevel integrated model with data-driven layer ensemble for multi-modality image fusion. IEEE Trans Image Process. 2020;30:1261–74.
40. Qi W, Chong Xu, Xiwei Xu. AutoGluon: a revolutionary framework for landslide hazard analysis. Natural Hazards Research. 2021;1:103–8.
41. Madsen C, Plesner TL, Bentzen HH, Jørgensen J, Sillesen IB, Himmelstrup BM, et al. Real world data on histological transformation in patients with follicular lymphoma: incidence, clinico-pathological risk factors and outcome in a nationwide Danish cohort. Leuk Lymphoma. 2020;61:2584–94.
42. Casulo C, Herold M, Hiddemann W, Iyengar S, Marcus RE, Seymour JF, et al. Risk factors for and outcomes of follicular lymphoma histological transformation at first progression in the GALLIUM study. Clin Lymphoma Myeloma Leuk. 2023;23:40–8.
43. Méndez M, Torrente M, Sánchez-Beato M, González-Rincón J, Royuela A, Gómez-Codina J, et al. Transformed follicular lymphoma in the rituximab era: a report from the Spanish Lymphoma Oncology Group. Hematol Oncol. 2019;37:143–50.
44. Sun C, Shen Z, Bian Z, Ma Y, Li J, Ma Y et al (2022) Disparity analysis of clinicopathologic hallmarks between transformed DLBCL and primary DLBCL. Clin Lab 68:1394–1401
45. Uhl B, Prochazka KT, Pansy K, Wenzl K, Strobl J, Baumgartner C, et al. Distinct chemokine receptor expression profiles in de novo DLBCL, transformed follicular lymphoma, Richter's trans-formed DLBCL and germinal center B-cells. Int J Mol Sci. 2022;23(14):7874.
46. Dreval K, Hilton LK, Cruz M, Shaalan H, Ben-Neriah S, Boyle M, et al. Genetic subdivisions of follicular lymphoma defined by distinct coding and noncoding mutation patterns. Blood. 2023;142(6):561–73.

Publisher's Note