Machine learning in the differentiation of follicular lymphoma from diffuse large B-cell lymphoma with radiomic [F-18]FDG PET/CT features

de Jesus, F Montes; Yin, Y; Mantzorou-Kyriaki, E; Kahle, X U; de Haas, R J; Yakar, D; Glaudemans, A W J M; Noordzij, W; Kwee, T C; Nijland, M

Published in:
European Journal of Nuclear Medicine and Molecular Imaging

DOI:
10.1007/s00259-021-05626-3

Publication date:
2022

European Journal of Nuclear Medicine and Molecular Imaging
https://doi.org/10.1007/s00259-021-05626-3

ORIGINAL ARTICLE

Machine learning in the differentiation of follicular lymphoma from diffuse large B-cell lymphoma with radiomic [^{18}F]FDG PET/CT features

F. Montes de Jesus[1] · Y. Yin[1] · E. Mantzorou-Kyriaki[1] · X. U. Kahle[1] · R. J. de Haas[1] · D. Yakar[1] · A. W. J. M. Glaudemans[1] · W. Noordzij[1] · T. C. Kwee[1] · M. Nijland[1]

Received: 18 September 2021 / Accepted: 16 November 2021

Abstract

Background One of the challenges in the management of patients with follicular lymphoma (FL) is the identification of individuals with histological transformation, most commonly into diffuse large B-cell lymphoma (DLBCL). [^{18}F]FDG-PET/CT is used for staging of patients with lymphoma, but visual interpretation cannot reliably discern FL from DLBCL. This study evaluated whether radiomic features extracted from clinical baseline [^{18}F]FDG PET/CT and analyzed by machine learning algorithms may help discriminate FL from DLBCL.

Materials and methods Patients were selected based on confirmed histopathological diagnosis of primary FL (n=44) or DLBCL (n=76) and available [^{18}F]FDG PET/CT with EARL reconstruction parameters within 6 months of diagnosis. Radiomic features were extracted from the volume of interest on co-registered [^{18}F]FDG PET and CT images. Analysis of selected radiomic features was performed with machine learning classifiers based on logistic regression and tree-based ensemble classifiers (AdaBoosting, Gradient Boosting, and XG Boosting). The performance of radiomic features was compared with a SUV_{max}-based logistic regression model.

Results From the segmented lesions, 121 FL and 227 DLBCL lesions were included for radiomic feature extraction. In total, 79 radiomic features were extracted from the SUVmap, 51 from CT, and 6 shape features. Machine learning classifier Gradient Boosting achieved the best discrimination performance using 136 radiomic features (AUC of 0.86 and accuracy of 80%). SUV_{max}-based logistic regression model achieved an AUC of 0.79 and an accuracy of 70%. Gradient Boosting classifier had a significantly greater AUC and accuracy compared to the SUV_{max}-based logistic regression ($p \leq 0.01$).

Conclusion Machine learning analysis of radiomic features may be of diagnostic value for discriminating FL from DLBCL tumor lesions, beyond that of the SUV_{max} alone.

Keywords [^{18}F]FDG-PET/CT, machine learning, radiomic features, follicular lymphoma, diffuse large B-cell lymphoma

Introduction

One of the challenges in the management of patients with follicular lymphoma (FL) is the identification of individuals with histological transformation, most commonly into diffuse large B-cell lymphoma (DLBCL). The annual rate of histological transformation is 1–2% and unlike patients with low-grade follicular lymphoma, those with transformed follicular lymphoma have an adverse prognosis (median overall survival of one to two years) requiring prompt identification and initiation of chemoimmunotherapy [1–5].

Identification of transformation remains a clinical challenge, and several studies have attempted to identify various biological parameters [1, 5–8]. Although no single factor has been identified, commonly reported biological parameters include altered Eastern Cooperative Oncology Group performance status, elevated lactate dehydrogenase, and low hemoglobin [5, 7].

Newly diagnosed FL patients undergo a combination of clinical assessment and imaging, frequently with 2-[^{18}F]

F. Montes de Jesus and Y. Yin contributed equally to this work.

This article is part of the Topical Collection on Advanced Image Analyses (Radiomics and Artificial Intelligence).

✉ F. Montes de Jesus
f.m.montes.de.jesus@umcg.nl

[1] Universitair Medisch Centrum Groningen, Groningen, Netherlands

Published online: 01 December 2021

fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography ([^{18}F]FDG PET/CT) [9]. Current visual interpretation of [^{18}F]FDG PET/CT can identify fast-growing lesions with high focal intense uptake but cannot reliably discern between FL, transformed FL, and DLBCL [10]. Ultimately, upon clinical suspicion of histological transformation, patients will undergo an invasive biopsy, which may be limited by sampling errors or lead to complications.

Several studies have evaluated the role of [^{18}F]FDG PET/CT standardized uptake values (SUVs) in identifying unfavorable prognosis and histological transformation of FL [11, 12]. However, while some studies found SUV valuable to identify transformation of FL and predict survival, others presented contradictory results [11–13]. Despite being the most frequently investigated parameter, SUVs remain limited in their capability to characterize microenvironment and tumor phenotype [14]. Such parameters omit intratumoral [^{18}F]FDG spatial distribution, and in particular SUV_{max} is susceptible to noise artifacts as it relies on a single voxel representation of the lesion [15].

In radiomics, large amounts of quantitative data are extracted from medical images, which may more accurately represent in vivo tumor conditions than SUV alone [16]. Moreover, the application of machine learning methods to radiomics is suited for classification modeling and capable of handling its inherent high-dimensionality [17]. Preliminary data on [^{18}F]FDG PET radiomic feature analysis suggests that machine learning algorithms may distinguish between DLBCL, FL, mantle cell lymphoma, and Hodgkin's lymphoma [18, 19]. Furthermore, a support vector machine classifier based on 105 features obtained from [^{18}F]FDG PET/CT has been shown to discriminate aggressive lymphoma lesions from hypermetabolic inflammatory/physiological uptake (brown fat, inflammation, infection, physiologic thymic uptake) [20]. Regarding therapy response and outcome prediction, analysis of [^{18}F]FDG PET radiomic features has been shown to correlate with survival makers in DLBCL, FL, and marginal zone lymphoma [21–23].

The aim of this study was to determine whether radiomic features extracted from clinical baseline [^{18}F]FDG PET/CT and analyzed by machine learning algorithms can be used to discriminate between FL and DLBCL tumor lesions in patients with a histopathologically confirmed diagnosis. If [^{18}F]FDG PET/CT radiomic features allow for the differentiation between FL and DLBCL tumor lesions, it may help identify patients with transformation of FL during the course of disease.

Materials and methods

Study design and patient selection

A retrospective cohort was selected from the electronic patient records at the University Medical Center Groningen (UMCG) between the years 2010 and 2020. Patients were selected based on the confirmed histopathological diagnosis of primary FL or DLBCL and available [^{18}F]FDG PET/CT with European Association of Nuclear Medicine Research (EARL) program reconstruction parameters within 6 months of diagnosis. Patients with co-existing disease (i.e., infection, other malignancy, autoimmune disease), grade III FL, and previously treated lymphoma and those where accurate segmentation was not possible (i.e., central nervous system, adjacent to areas of high physiologic uptake) were excluded. Histopathological diagnosis was based on core biopsy or lymph node excision. The study was conducted in accordance with the ethical principles of the Declaration of Helsinki and with the approval of the Medical Ethics Review Board of the UMCG (202100165, 02-03-2021). The Medical Ethics Review Board waived informed consent of all study participants.

[^{18}F]FDG PET/CT acquisition

[^{18}F]FDG PET/CT scans were performed using a Siemens Biograph mCT or Siemens Biograph Vision (Siemens Healthcare, Knoxville, Tennessee, USA) according to the European Association of Nuclear Medicine (EANM) procedure guidelines for tumor imaging [24]. Imaging protocol included a minimum fasting of 6 h with glucose levels < 11mmol/L. A weight-based [^{18}F]FDG dose of 3MBq [^{18}F]FDG per kg body weight was administered, followed by 60 min of rest. Low-dose CT was performed with 100kV and 30mAs. Subsequently, a whole-body (vertex to mid-thigh) PET scan with 2–3 min per bed position was acquired. Integrated [^{18}F]FDG PET/CT images were corrected for scattering and attenuation based on CT information. For standardized calculation purposes, all scans were reconstructed according to the specifications of the EARL program [25–27].

[^{18}F]FDG PET/CT image preprocessing and lesion segmentation

EARL compliant [^{18}F]FDG PET images and low-dose CT images were used for the extraction of radiomic features. Images from both modalities were resampled to the same reconstruction slice spacing. Images with larger reconstruction slice spacing were downsized to avoid generation of extra image information. Lesions were segmented in two phases. In a first phase, lesions were preselected in line with PERCIST recommendations with a semiautomatic segmentation tool on Hermes Hybrid 3D software (Hermes Medical Solutions AB, Stockholm, Sweden) by FMJ and blinded for histopathologic diagnosis

(Figure 1) [28]. Lesions not automatically segmented but suspected of malignancy were manually added, while any metabolically active foci interpreted as physiological were removed. In a second phase, lesions smaller than 3ml were excluded (to avoid partial volume effects), and the five lesions with the highest SUV_{max} (highest uptake voxel within the volume of interest) were selected for the radiomic feature extraction.

Radiomic feature extraction and selection

Radiomics workflow is shown in Figure 2. [^{18}F]FDG PET images were first converted to SUVmaps. SUVmaps and CT images were discretized using a fixed bin size of 0.25 SUV and five Hounsfield units, respectively. A total of 93 features were extracted from [^{18}F]FDG PET and 93 features from CT according to the image biomarker standardization initiative (IBSI) [29]. Radiomic features were extracted from the volume of interest on co-registered [^{18}F]FDG PET and CT images. A total of 18 were first-order statistic features (including SUV_{max} and SUV_{mean}), 24 gray-level co-occurrence matrix features, 16 gray-level size zone matrix features, 16 gray-level run-length matrix features, five neighboring gray-tone difference matrix features, and 14 gray-level dependence matrix features [29, 30]. Seventeen 3D shape-based features were extracted from the segmented tumor lesions. After univariate feature selection, features shown to have a statistically significant variation ($p \leq 0.05$) between the labels (FL and DLBCL) were retained in the selected radiomic feature subset. Feature extraction and

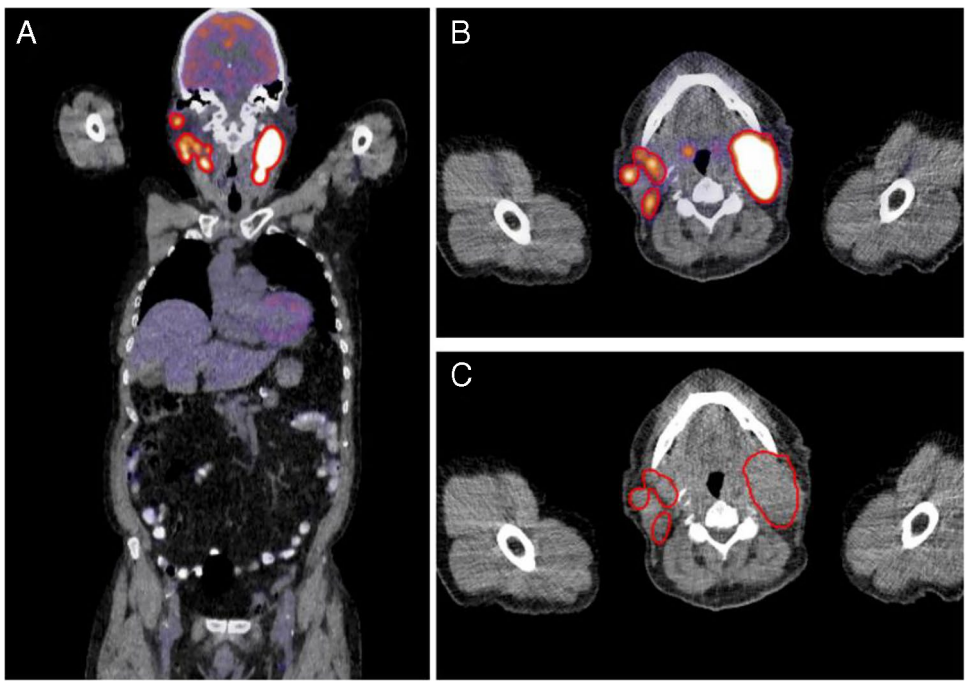

Fig. 1 Example of lesion segmentation in a patient with DLBCL. **A** Fused [^{18}F]FDG PET/CT shows in coronal view the delineation of segmented lymph nodes on both sides of the neck, as performed with the "Tumor Finder" application in Hermes Hybrid 3D. Fused [^{18}F]FDG PET/CT (**B**) and CT (**C**) show corresponding delineation in axial view

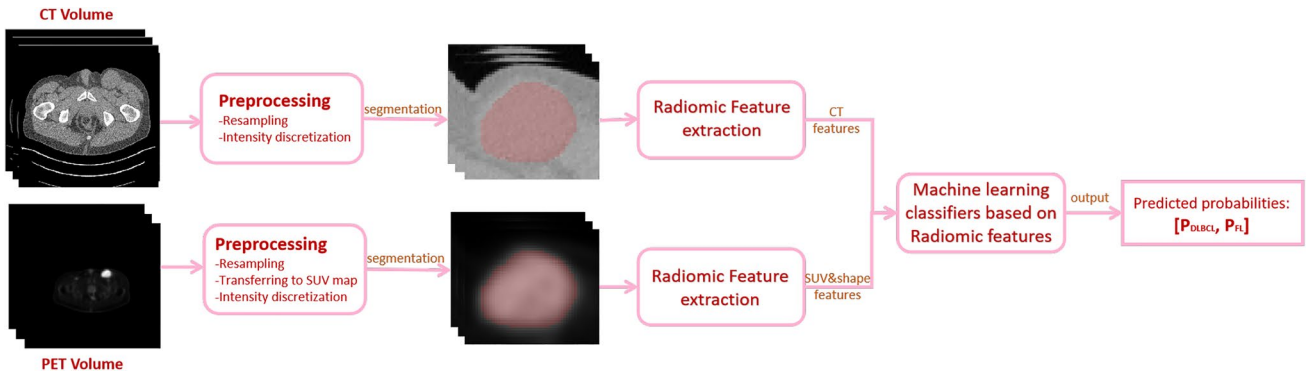

Fig. 2 Radiomics workflow. [^{18}F]FDG PET and CT images were first preprocessed. Then radiomic features were extracted from the segmented lesions. Machine learning classifiers were trained by the combination of radiomic features from [^{18}F]FDG PET and CT as well as shape features

selection were performed with Python 3.7.9 in the open-source library with PyRadiomics 3.0 and library scikit-learn 0.23.2, respectively [31, 32].

Machine learning classifiers

Machine learning classifiers were trained on a per lesion basis. A random training set was selected from the study cohort containing 80% of tumor lesions and classifier features modulated through a 5-fold cross-validation. Machine learning classifiers were based on logistic regression and tree-based ensemble classifiers. L1 (lasso regression) and L2 (ridge regression) regularization were used in the logistic regression classifier to avoid overfitting on the training set. Tree-based classifiers were trained with three ensemble methods, including AdaBoosting, Gradient Boosting, and XG Boosting [33–35]. Feature importance for the tree-based models was measured with Gini importance (mean decrease in impurity). Discrimination performance was evaluated based on the area under the receiver-operating characteristic (ROC) curve (AUC) in a test set containing 20% of the tumor lesions, independent of the training set. Confidence intervals of machine learning classifier outputs were calculated using bootstrap analysis [36]. Classifiers were built with Python 3.7.9 and library scikit-learn 0.23 [32].

SUV_{max}-based classification

As SUV_{max} is the most recognized [^{18}F]FDG PET/CT semi-quantitative parameter and has been the focus of previous studies on this topic, we evaluated a SUV_{max}-based logistic regression model on the test set [12, 13, 37]. The AUC and accuracy of the SUV_{max}-based logistic regression model were compared with those of the machine learning classifiers utilizing bootstrapping of 1000 repetitions [36]. Statistical significance was set at $p \leq 0.05$. Analysis was performed with Python 3.7.9 and library scikit-learn 0.23 [32].

Results

Demographic characteristics

The study cohort included 76 patients with DLBCL and 44 patients with FL. We excluded 25 patients with an unclear or inconclusive pathology report, 17 with co-existing disease, 16 where accurate segmentation was not possible, nine with grade III FL, and eight with previously treated lymphoma. There were 71 (59%) males and 49 (41%) females, with a median age at diagnosis of 65 years (Table 1). Patients with DLBCL had a low-, intermediate-, and high-risk international prognostic index (IPI) in 7%, 43%, and 50% of cases, respectively. FL patients were evenly distributed across the follicular lymphoma prognostic index (FLIPI) risk groups, with 34%, 25%, and 41% of patients included in risk group 0–1, 2, and ≥3, respectively. Counts and percentages regarding each variable included in the calculation of the IPI and FLIPI can be found in Table 1.

Radiomic model

Radiomic features were extracted from 121 FL and 227 DLBCL segmented tumor lesions. In total, 79 features were extracted from the SUVmap, 51 from CT, and six 3D shape

Table 1 Demographic characteristics (n=120)

	DLBCL n=76	FL n=44
Gender		
Male	45 (59%)	18 (41%)
Female	31 (41%)	26 (59%)
Age (years)		
Median	67	63
Q1–Q3	57–74	55–71
Ann arbor staging		
I–II	23 (30%)	14 (32%)
III–IV	53 (70%)	30 (68%)
Elevated LDH		
Yes	42 (55%)	5 (12%)
No	34 (45%)	38 (86%)
NA		1 (2%)
IPI		
0	5 (7%)	
1–2	33 (43%)	
3–5	38 (50%)	
ECOG performance status		
0–1	68 (89%)	
≥2	8 (11%)	
>1 extra-nodal site		
Yes	33 (43%)	
No	43 (57%)	
FLIPI		
0–1		15 (34%)
2		11 (25%)
≥3		18 (41%)
>4 nodal sites		
Yes		23 (52%)
No		21 (48%)
Hemoglobin <120 g/L		
Yes		5 (12%)
No		38 (86%)
NA		1 (2%)

Abbreviations: DLBCL, diffuse large B-cell lymphoma; ECOG, Eastern Cooperative Oncology Group performance status; FL, follicular lymphoma; FLIPI, follicular lymphoma prognostic index; IPI, international prognostic index; LDH, lactate dehydrogenase

features. The sensitivity, specificity, accuracy, and AUC of different classifiers are listed in Table 2 with corresponding ROC curves plotted in Figure 3. The best performing model was achieved with Gradient Boosting, reaching an AUC of 0.86 and an accuracy 80%. The SUV_{max}-based logistic regression model had an AUC of 0.79 and an accuracy of 70%. Gradient Boosting and XG Boosting achieved a significantly greater AUC and accuracy than the SUV_{max}-based logistic regression model (AUC comparison $p \leq 0.01$, accuracy comparison $p \leq 0.01$).

The top five radiomic features selected based on weight ranking for each machine learning classifier are listed in Tables 3 and 4. In the best performing machine learning classifiers (Gradient Boosting and XG Boosting), all top five radiomic features were SUV dependent (Table 4). The entire radiomic feature list and corresponding weight ranking for each classifier can be found in Supplementary Table 1.

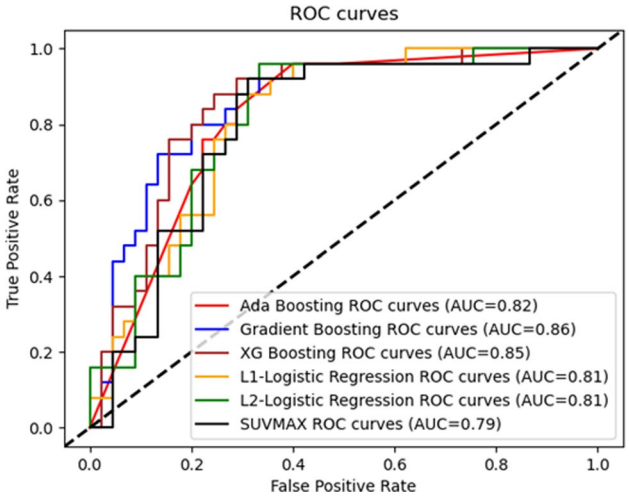

Fig. 3 ROC curves of different classifiers. Graphical representation of ROC with respective AUC obtained from tree-based classifiers (Ada-Boosting (red), Gradient Boosting (blue) and XG Boosting (brown)), L1 (lasso regression) (yellow) and L2 (ridge regression) (green) logistic regression classifiers, and SUV_{max}-based logistic regression model (black)

Discussion

The present study aimed to determine whether radiomic features extracted from clinical baseline [^{18}F]FDG PET/CT and analyzed by machine learning algorithms may be used to discriminate between FL and DLBCL tumor lesions in patients with a histopathologically confirmed diagnosis. Machine learning classifier Gradient Boosting achieved the best AUC of 0.86 and an accuracy of 80%, outperforming SUV_{max}-based logistic regression model (AUC of 0.79 and accuracy of 70%). Gradient Boosting model achieved a significantly greater AUC and accuracy compared to the SUV_{max}-based logistic regression.

Unlike biopsy, [^{18}F]FDG PET/CT semiquantification is not limited by geographical sampling, allowing for simultaneous visualization and semiquantification of tumor lesions. Should [^{18}F]FDG PET/CT facilitate clinical identification of FL from DLBCL tumor lesions, it may more readily and more accurately identify patients with FL transformation than current standards. Early [^{18}F]FDG PET/CT semiquantification studies reported higher SUVs in aggressive non-Hodgkin's lymphoma than in indolent subtypes [38–40]. Subsequently, a similar rationale was implemented in studies evaluating SUV differences in FL compared to DLBCL and prediction of histological transformation in FL [12, 13, 37, 41–43]. SUV_{max} cutoffs of 10, 14, and 17 have been proposed to identify patients at higher risk for transformation and previous studies reported AUCs between 0.80 and 0.85 for SUV_{max} to distinguish aggressive from indolent non-Hodgkin's lymphoma [37, 38, 42, 44]. Nevertheless, heterogenous methodologies, different camera systems, small study cohorts, and contradictory results have prevented clinical translation of SUV_{max} as a tool to differentiate FL from DLBCL and potentially identify patients with FL transformation.

Previous literature largely focused on determining SUV_{max} cutoff values to distinguish aggressive from indolent non-Hodgkin's lymphoma (while combining various

Table 2 Performance of different machine learning classifiers and SUV_{max} logistic regression

Classifiers	Sensitivity [95%CI]	Specificity [95%CI]	Accuracy [95%CI]	*p* value accuracy	AUC [95%CI]	*p* value AUC
Lasso regression	0.68 [0.62,0.76]	0.76 [0.71,0.81]	0.73 [0.68,0.77]	0.26	0.81 [0.78, 0.85]	0.16
Ridge regression	0.72 [0.65,0.81]	0.76 [0.71,0.81]	0.74 [0.7,0.78]	0.15	0.81 [0.78, 0.86]	0.23
Ada boosting	0.76 [0.7,0.84]	0.76 [0.71,0.81]	0.76 [0.72,0.8]	0.04	0.82 [0.78, 0.86]	0.10
Gradient boosting	0.68 [0.61,0.76]	0.87 [0.84,0.92]	0.8 [0.77,0.83]	<0.01	0.86 [0.83, 0.89]	<0.01
XG boosting	0.76 [0.7,0.85]	0.82 [078,0.87]	0.8 [0.77,0.83]	<0.01	0.85 [0.82, 0.89]	0.01
SUV_{max}	0.56 [0.48,0.65]	0.78 [0.74,0.83]	0.7 [0.67,0.75]	-	0.79 [0.76,0.85]	-

Area under the curve and accuracy of each machine learning classifier are compared with SUV_{max}-based logistic regression model and statistically tested by bootstrapping method. Abbreviations: AUC, area under the curve; CI, confidence interval; SUV, standardized uptake value

Table 3 Top five radiomic features in logistic regression classifier

Weight ranking	Radiomic features											
	Lasso regression						Ridge regression					
	FL			DLBCL			FL			DLBCL		
	Source	*Feature class*	*Feature name*	*Source*	*Feature class*	*Feature name*	*Source*	*Feature class*	*Feature name*	*Source*	*Feature class*	*Feature name*
1st		Shape	Surface volume ratio	SUV	First order	Mean	SUV	GLSZM	Gray-level non-uniformity	CT	First-order	Energy
2nd	SUV	GLSZM	Gray-level non-uniformity	SUV	First order	Median		Shape	Surface volume ratio	SUV	GLRLM	Gray-level variance
3rd	SUV	GLSZM	Gray-level non-uniformity normalized	SUV	GLRLM	Gray-level variance	SUV	NGTDM	Coarseness	CT	First-order	10 percentile
4th	SUV	GLSZM	Small area emphasis	CT	First order	Energy	SUV	NGTDM	Strength	SUV	First-order	Median
5th	CT	GLCM	Difference entropy		Shape	Maximum 2D-diameter row	SUV	GLSZM	Small area emphasis	SUV	NGTDM	Busyness
Total *n*° features	32			58			57			79		

Abbreviations: CT, computed tomography; GLCM, gray-level co-occurrence matrix; GLSZM, gray-level size zone matrix; NGTDM, neighboring gray tone difference matrix; SUV, standard uptake value

Table 4 Top five radiomic features in tree-based ensemble classifiers

Weight ranking	Radiomic features								
	AdaBoosting			Gradient Boosting			XG Boosting		
	Source	*Feature class*	*Feature name*	*Source*	*Feature class*	*Feature name*	*Source*	*Feature class*	*Feature name*
1st	SUV	GLSZM	Low gray level zone emphasis	SUV	First order	Range	SUV	First–order	Minimum
2nd	SUV	First–order	Median	SUV	First order	Median	SUV	First–order	Median
3rd	SUV	First–order	Minimum	SUV	GLCM	Correlation	SUV	First–order	Total energy
4th	SUV	First–order	Range	SUV	First order	Kurtosis	SUV	GLSZM	Low gray-level zone emphasis
5th				SUV	First order	Minimum	SUV	First–order	Range
Total *n*° features	4			136			117		

Abbreviations: GLCM, gray-level co-occurrence matrix; GLSZM, gray-level size zone matrix; SUV, standard uptake value

lymphoma histological subtypes into the aggressive and indolent groups) or identifying patients with indolent lymphoma at high risk for transformation. As the aim and methodologies significantly differ from that of our study, we cannot directly compare the results obtained [12, 37, 38, 42, 43]. Nevertheless, a study by Wu and colleagues reported a SUV_{max} AUC of 0.80 to distinguish FL from DLBCL, similar to the AUC of 0.79 obtained with our SUV_{max}-based logistic regression model [44].

Our study is innovative in its application of radiomic feature extraction and machine learning algorithms to discriminate between patients with FL from DLBCL using [^{18}F]FDG PET/CT. Additionally, it combines EARL compliant [^{18}F]FDG PET images and low-dose CT images for extraction of radiomic features. Previous studies have relied on SUV obtained from [^{18}F]FDG PET for lesion semiquantification. Yet, commonly used SUV parameters (SUV_{max}, SUV_{mean}, SUV_{peak}) are based exclusively on the distribution of individual voxel values without concern for spatial relationships. Particulary SUV_{max}, which is the most frequently studied [^{18}F]FDG PET/CT semiquantitative parameter, is susceptible to noise artifacts as it is calculated from a single voxel representation within the volume of interest [14, 15, 45]. Compared to second-order or higher order radiomics, which take into consideration the interrelationships between neighboring voxels, SUVs are limited in their capability to characterize microenvironment and tumor phenotype [45–47].

A limited number of previous studies have reported improved model discrimination of [^{18}F]FDG PET/CT radiomic features as compared to [^{18}F]FDG PET/CT SUV semiquantification in lymphoma. Addition of [^{18}F]FDG PET/CT radiomic features to SUV improved the assessment of bone marrow involvement in patients with mantle cell lymphoma, with a reported AUC increase from 0.66 to 0.81 [48]. Similarly, selected [^{18}F]FDG PET radiomic features improved discriminatory performance of central nervous system lymphomas from glioblastoma multiforme as compared to SUV_{max}, with AUCs of 0.99 and 0.94 respectively [48].

Our study highlights improvements of a radiomic approach compared to SUV_{max}-based semiquantification when applied to the differentiation of patients with FL from DLBCL. We present a selection of radiomic features which may be explored to identify FL patients with histological transformation. We observe that all top five features obtained with our best performing model (Gradient Boosting) are SUV dependent and predominately first-order radiomics. Therefore, it may be suggested that metabolic differences (as measured by SUV) regardless of spatial relationships are essential for the differentiation of FL from DLBCL tumor lesions. This observation is consistent with previous studies, reporting higher SUVs in aggressive non-Hodgkin's lymphoma than in indolent subtypes [38–40]. Nevertheless, we must be aware that our methodology for lesion selection may have influenced our results. As we based our lesion selection on the five lesions with the highest SUV_{max}, we may have biased our model towards SUV-based parameters. Currently, there are no lymphoma-specific guidelines for radiomic analysis, which may aid researchers in post-acquisition methodology pertaining to lesion selection and lesion segmentation. Although our exploratory study may serve as the comparison for future studies, the optimal methodology for lesion selection and lesion segmentation remains undefined. Future studies may also focus on expanding the clinical validity of our methods. In the present study, in order to avoid introducing bias to model, we performed machine learning training on a per lesion basis and did not include any transformed FL patients due to the limited number of cases. These limitations can be mitigated by applying current methods to larger, multicenter cohorts which include a greater number of patients, including cases of transformed FL.

Other improvements to future radiomic analysis on this topic may include the inclusion of clinical parameters to the training of our machine learning classifier. The combination of [^{18}F]FDG PET/CT radiomic features with clinical features has been shown to improve discriminatory performance of models in other clinical scenarios [48, 49]. As routine clinical parameters such as altered Eastern Cooperative Oncology Group performance status, elevated lactate dehydrogenase and low hemoglobin have been suggested as risk factors for transformation, they may further improve the model, and may be investigated in future studies [5, 7]. Finally, researchers may consider including diagnostic CT images, instead of low-dose CT, in an attempt to improve tumor characterization.

Conclusion

Machine learning analysis of radiomic features may be of diagnostic value for discriminating FL from DLBCL tumor lesions, beyond that of the SUV_{max} alone.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00259-021-05626-3.

Acknowledgements Not applicable

Availability of data and materials The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author contribution F. Montes de Jesus and Y. Yin, conception and design, acquisition of data, literature search, analysis/interpretation of data and writing of the manuscript. M. Nijland, W. Noordzij, A.WJM Glaudemans, T.C. Kwee, conception and design, analysis/interpretation of data, supervision, and writing of the manuscript. E. Mantzorou-Kyriaki and XU Kahle, acquisition of data. RJ de Haas, D. Yakar, conception and design, acquisition of data, interpretation of data, writing of the manuscript.

Declarations

Ethics approval and consent to participate The study was conducted in accordance with the ethical principles of the Declaration of Helsinki and with the approval of the Medical Ethics Review Board of the UMCG (202100165, 02-03-2021). The Medical Ethics Review Board waived informed consent of all study participants.

Consent for publication Not applicable

Competing interests All authors declare that they have no conflict of interest.

References

1. Al-Tourah AJ, Gill KK, Chhanabhai M, Hoskins PJ, Klasa RJ, Savage KJ, et al. Population-based analysis of incidence and outcome of transformed Non-Hodgkin's lymphoma. J Clin Oncol. 2008;26:5165–9.
2. Conconi A, Ponzio C, Lobetti-Bodoni C, Motta M, Rancoita PMV, Stathis A, et al. Incidence, risk factors and outcome of histological transformation in follicular lymphoma. Br J Haematol. 2012;157:188–96.
3. Freeman CL, Kridel R, Moccia AA, Savage KJ, Villa DR, Scott DW, et al. Early progression after bendamustine-rituximab is associated with high risk of transformation in advanced stage follicular lymphoma. Blood. American Society of Hematology. 2019;134:761–4.
4. Casulo C, Byrtek M, Dawson KL, Zhou X, Farber CM, Flowers CR, et al. Early relapse of follicular lymphoma after rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone defines patients at high risk for death: an analysis from the National LymphoCare Study. J Clin Oncol. American Society of Clinical Oncology. 2015;33:2516–22.
5. Montoto S, Davies AJ, Matthews J, Calaminici M, Norton AJ, Amess J, et al. Risk and clinical implications of transformation of follicular lymphoma to diffuse large B-cell lymphoma. J Clin Oncol. 2007;25:2426–33.
6. Link BK, Maurer MJ, Nowakowski GS, Ansell SM, MacOn WR, Syrbu SI, et al. Rates and outcomes of follicular lymphoma transformation in the immunochemotherapy era: a report from the university of Iowa/mayo clinic specialized program of research excellence molecular epidemiology resource. J Clin Oncol. 2013;31:3272–8.
7. Sarkozy C, Trneny M, Xerri L, Wickham N, Feugier P, Leppa S, et al. Risk factors and outcomes for patients with follicular lymphoma who had histologic transformation after response to first-line immunochemotherapy in the PRIMA trial. J Clin Oncol. 2016;34:2575–82.
8. Wagner-Johnston ND, Link BK, Byrtek M, Dawson KL, Hainsworth J, Flowers CR, et al. Outcomes of transformed follicular lymphoma in the modern era: a report from the national LymphoCare study (NLCS). Blood. 2015;126:851–7.
9. Dreyling M, Ghielmini M, Rule S, Salles G, Ladetto M, Tonino SH, et al. Newly diagnosed and relapsed follicular lymphoma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol Off J Eur Soc Med Oncol. Elsevier Ltd.; 2021;32:298–308.
10. Papajík T, Mysliveček M, Šedová Z, Buriánková E, Procházka V, Koranda P, et al. Standardised uptake value of 18F-FDG on staging PET/CT in newly diagnosed patients with different subtypes of non-Hodgkin's lymphoma. Eur J Haematol. 2011;86:32–7.
11. Novelli S, Briones J, Flotats A, Sierra J. PET/CT assessment of follicular lymphoma and high grade B cell lymphoma - good correlation with clinical and histological features at diagnosis. Adv Clin Exp Med. 2015;24:325–30.
12. Mir F, Barrington SF, Brown H, Nielsen T, Sahin D, Meignan M, et al. Baseline SUVmax did not predict histological transformation in follicular lymphoma in the phase 3 GALLIUM study. Blood. 2020;135:1214–8.
13. 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.
14. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they Are data. Radiology. 2016;278:563–77.

15. Lodge MA, Chaudhry MA, Wahl RL. Noise considerations for PET quantification using maximum and peak standardized uptake value. J Nucl Med. 2012;53:1041–7.
16. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RGPM, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48:441–6.
17. Koçak B, Durmaz EŞ, Ateş E, Kılıçkesmez Ö. Radiomics with artificial intelligence: a practical guide for beginners. Diagnostic Interv Radiol. 2019;25:485–95.
18. Wang H, Zhou Y, Li L, Hou W, Ma X, Tian R. Current status and quality of radiomics studies in lymphoma: a systematic review. Eur Radiol. 2020;30:6228–40.
19. Lippi M, Gianotti S, Fama A, Casali M, Barbolini E, Ferrari A, et al. Texture analysis and multiple-instance learning for the classification of malignant lymphomas. Comput Methods Programs Biomed. Elsevier B.V.; 2020;185:105153.
20. Lartizien C, Rogez M, Niaf E, Ricard F. Computer-aided staging of lymphoma patients with FDG PET/CT imaging based on textural information. IEEE J Biomed Heal Informatics. IEEE. 2014;18:946–55.
21. Zhou Y, Ma X-L, Pu L-T, Zhou R-F, Ou X-J, Tian R. Prediction of overall survival and progression-free survival by the 18 F-FDG PET/CT radiomic features in patients with primary gastric diffuse large B-cell lymphoma. Contrast Media Mol Imaging. 2019;2019:1–9.
22. Lue K-H, Wu Y-F, Liu S-H, Hsieh T-C, Chuang K-S, Lin H-H, et al. Intratumor heterogeneity assessed by 18F-FDG PET/CT predicts treatment response and survival outcomes in patients with Hodgkin lymphoma. Acad Radiol. 2020;27:e183–92.
23. Mayerhoefer ME, Riedl CC, Kumar A, Gibbs P, Weber M, Tal I, et al. Radiomic features of glucose metabolism enable prediction of outcome in mantle cell lymphoma. Eur J Nucl Med Mol Imaging. 2019;46:2760–9.
24. Boellaard R, Tatsch K, Pike LC, Testanera G, Kotzerke J, Graham MM, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2015;42:328–54.
25. Boellaard R, Oyen WJG, Hoekstra CJ, Hoekstra OS, Visser EP, Willemsen AT, et al. The Netherlands protocol for standardisation and quantification of FDG whole body PET studies in multi-centre trials. Eur J Nucl Med Mol Imaging. 2008;35:2320–33.
26. Boellaard R, Willemsen A, Arends B, Visser EP. EARL procedure for assessing PET/CT system specific patient FDG activity preparations for quantitative FDG PET/CT studies. April. 2013;2010:1–3.
27. Pfaehler E, van Sluis J, Merema BBJ, van Ooijen P, Berendsen RCM, van Velden FHP, et al. Experimental multicenter and multivendor evaluation of the performance of PET radiomic features using 3-dimensionally printed phantom inserts. J Nucl Med. 2020;61:469–76.
28. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. J Nucl Med. 2009;50:122S–50S.
29. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology. 2020;295:328–38.
30. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology. 2020;295:328–38.
31. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res. 2017;77:e104–7.
32. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. J Mach Learn Res. 2011;12:2825–30.
33. Schapire RE. Explaining AdaBoost. Empir Inference. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. p. 37–52.
34. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001;29.
35. Chen T, Guestrin C. XGBoost. Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min. New York, NY, USA: ACM; 2016. p. 785–94.
36. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.
37. 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.
38. 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.
39. Rodriguez M, Rehn S, Ahlström H, Sundström C, Glimelius B. Predicting malignancy grade with PET in non-Hodgkin's lymphoma. J Nucl Med. 1995;36:1790–6.
40. Lapela M, Leskinen S, Minn HR, Lindholm P, Klemi PJ, Söderström KO, et al. Increased glucose metabolism in untreated non-Hodgkin's lymphoma: a study with positron emission tomography and fluorine-18-fluorodeoxyglucose. Blood. 1995;86:3522–7.
41. Strati P, Ahmed MA, Fowler NH, Nastoupil LJ, Samaniego F, Fayad LE, et al. Pre-treatment maximum standardized uptake value predicts outcome after frontline therapy in patients with advanced stage follicular lymphoma. Haematologica. 2020;105:1907–13.
42. 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.
43. Karam M, Feustel PJ, Vera CD, Nazeer T. Features of large cell transformation of indolent lymphomas as observed on sequential PET/CT. Nucl Med Commun. 2011;32:177–85.
44. Wu X, Pertovaara H, Korkola P, Dastidar P, Järvenpää R, Eskola H, et al. Correlations between functional imaging markers derived from PET/CT and diffusion-weighted MRI in diffuse large B-cell lymphoma and follicular lymphoma. PLoS One. 2014;9:e84999.
45. Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, et al. Introduction to radiomics. J Nucl Med. 2020;61:488–95.
46. Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, et al. Radiomics: the facts and the challenges of image analysis. Eur Radiol Exp. 2018;2.
47. Ha S, Choi H, Paeng JC, Cheon GJ. Radiomics in oncological PET/CT: a methodological overview. Nucl Med Mol Imaging (2010). 2019;53:14–29.
48. Mayerhoefer ME, Riedl CC, Kumar A, Dogan A, Gibbs P, Weber M, et al. Marrow involvement in mantle cell lymphoma : a retrospective study in 97 patients. Cancers (Basel). 2020;1–13.
49. Zhang M, Bao Y, Rui W, Shangguan C, Liu J, Xu J, et al. Performance of 18F-FDG PET/CT radiomics for predicting EGFR mutation status in patients with Non-small cell lung cancer. Front Oncol. 2020;10:1–8.