



Current status and quality of radiomics studies in lymphoma: a systematic review

Hongxi Wang¹ · Yi Zhou¹ · Li Li¹ · Wenxiu Hou¹ · Xuelei Ma² · Rong Tian¹

Received: 18 December 2019 / Revised: 25 March 2020 / Accepted: 28 April 2020
© European Society of Radiology 2020

Abstract

Objectives To perform a systematic review regarding the developments in the field of radiomics in lymphoma. To evaluate the quality of included articles by the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2), the phases classification criteria for image mining studies, and the radiomics quality scoring (RQS) tool.

Methods We searched for eligible articles in the MEDLINE/PubMed and EMBASE databases using the terms “radiomics”, “texture” and “lymphoma”. The included studies were divided into two categories: diagnosis-, therapy response- and outcome-related studies. The diagnosis-related studies were evaluated using the QUADAS-2; all studies were evaluated using the phases classification criteria for image mining studies and the RQS tool by two reviewers.

Results Forty-five studies were included; thirteen papers (28.9%) focused on the differential diagnosis of primary central nervous system lymphoma (PCNSL) and glioblastoma (GBM). Thirty-two (71.1%) studies were classified as discovery science according to the phase classification criteria for image mining studies. The mean RQS score of all studies was 14.2% (ranging from 0.0 to 40.3%), and 23 studies (51.1%) were given a score of < 10%.

Conclusion The radiomics features could serve as diagnostic and prognostic indicators in lymphoma. However, the current conclusions should be interpreted with caution due to the suboptimal quality of the studies. In order to introduce radiomics into lymphoma clinical settings, the lesion segmentation and selection, the influence of the pathological pattern and the extraction of multiple modalities and multiple time points features need to be further studied.

Key Points

- *The radiomics approach may provide useful information for diagnosis, prediction of the therapy response, and outcome of lymphoma.*
- *The quality of published radiomics studies in lymphoma has been suboptimal to date.*
- *More studies are needed to examine lesion selection and segmentation, the influence of pathological patterns, and the extraction of multiple modalities and multiple time point features.*

Keywords Lymphoma · Multidetector computed tomography · Magnetic resonance imaging · Positron emission tomography, computed tomography

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00330-020-06927-1>) contains supplementary material, which is available to authorized users.

✉ Xuelei Ma
dmaxuelei@gmail.com

✉ Rong Tian
rongtianneuclear@126.com

¹ Department of Nuclear Medicine, West China Hospital, Sichuan University, No. 37, Guoxue Alley, Chengdu 610041, Sichuan, People's Republic of China

² Department of Biotherapy, West China Hospital and State Key Laboratory of Biotherapy, Sichuan University, Chengdu, Sichuan, People's Republic of China

Abbreviations

AUC	Area under the curve
CT	Computed tomography
DLBCL	Diffuse large B cell lymphoma
GBM	Glioblastoma
HL	Hodgkin's lymphoma
MRI	Magnetic resonance imaging
NHL	Non-Hodgkin's lymphoma
PCNSL	Primary central nervous system lymphoma
PET	Positron emission tomography
QUADAS-2	The Quality Assessment of Diagnostic Accuracy Studies-2
RQS	Radiomics quality scoring

Introduction

For decades, medical imaging has transformed from a simple diagnostic tool to an enormous source of clinical data. The emergence of new technologies and the requirements of precision medicine have given rise to a new promising field of research called “radiomics” [1, 2]. The main goal of radiomics is to extract quantitative imaging features from medical images and to analyse tumour heterogeneity as a whole [3, 4]. As one of the important characteristics of malignant tumours, heterogeneity is related to malignant biological behaviour and can reflect the development, therapy responses, and clinical outcomes of tumours [5, 6]. Noninvasive detection of tumour heterogeneity by quantitative imaging has attracted considerable attention, and recent radiomics research has shown a promising association between imaging heterogeneity and solid tumour characteristics [7–11].

Lymphoma presents serious challenges for diagnosis and treatment because of its complex heterogeneity. Given the many types of lymphoma, the intertumoural heterogeneity of lymphoma is common. In addition, one type of lymphoma can show different intratumoural heterogeneity in a single patient [12, 13]. An intuitive manifestation of intratumoural heterogeneity is lymphoma transformation, namely the evolution of low-grade lymphoma into high-grade lymphoma during the disease course; this transformation is related to disease progression and poor prognosis [13]. In this sense, it is essential to explore the heterogeneity of lymphoma and to identify the clinical-image-genome model of highly invasive lymphoma for treatment selection. A few previous studies have explored the potential role of the radiomics approach as a diagnostic, classification and prognostic predictor tool in lymphoma. However, the results are controversial. Recently, several evaluation criteria and guidelines have been proposed to aid the assessment of radiomics research [14, 15]. Thus, we sought to systematically review the lymphoma radiomics literatures and evaluate the quality of studies according to the above criteria and guidelines.

Materials and methods

This systematic review was performed according to the PRISMA statement [16]. The PRISMA checklist is provided in Supplemental Table S1.

Literature search strategy

We performed a comprehensive literature search to identify English language studies published in MEDLINE/PubMed and EMBASE (Ovid) without a start date

limitation; the search result was last updated in January 2020. We used a search string containing free-text and/or Medical Subject Headings (MeSH) search of 3 key search terms: “radiomics”, “texture” and “lymphoma”. The references cited in the retrieved studies were also explored, and duplicate findings were discarded to ensure that no data overlap occurred.

Eligibility criteria and data extraction

The criteria for including studies were as follows: (a) articles were full-text and written in English; (b) radiomics investigations aimed at relevant objectives in clinical application for lymphoma; and (c) imaging modalities such as ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), scintigraphy and positron emission tomography (PET) or PET/CT were applied. The exclusion criteria were as follows: (a) preclinical and animal studies; (b) studies not within the field of interest; (c) studies focused on methodological aspects, or test-retest studies; (d) testing data (not medical images) as input for algorithms; and (e) case reports, reviews, poster presentations, conference abstracts and expert opinion papers. All articles were identified by two reviewers according to the aforementioned criteria. The extracted data included the following: study title, author, publication time, study population, patient characteristics, measurement characteristics, results and conclusions.

Quality assessment and data analysis

The quality of the studies was independently assessed by two reviewers. The criteria included the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) [17], the phase classification criteria for image mining studies [15] and the RQS tool [14]. The QUADAS-2 is used for diagnostic accuracy studies as detailed in Supplemental Table S2; any disagreement was resolved by consensus. The phase classification criteria assign image mining studies to the discovery science and phases 0–IV (Supplemental Table S3); any disagreement was resolved by consensus. The RQS tool evaluates the validity and bias of the studies (Supplemental Table S4); a radiomics study can achieve a total of 36 points maximum, and scores are presented in the form of a percentage. Agreement between the reviewers was assessed by means of a weighted kappa statistic in SPSS version 25.0 for Mac. If all ratings are the same for at least one reviewer, the weighted kappa value cannot be calculated [18, 19].

Results

A total of 332 records were identified until 25 January 2020. After full-text review, 45 studies were included in

this systematic review, and the entire process is shown in Fig. 1. For the sake of simplicity, we condensed the included studies into two categories by intended use: diagnosis-related studies ($n = 29$) [20–42, 44–49] and therapy response and outcome-related studies ($n = 15$) [50–64]. One study explored both the diagnostic and prognostic value of radiomics features in diffuse large B cell lymphoma (DLBCL) [43], so we analysed it in both categories. All the studies were retrospective. The sample size ranged from 9 to 251, and 9 of 45 (20%) studies enrolled more than 100 patients. A total of 15 studies (33.3%) performed validation analysis, including 12 diagnosis-related studies [22–24, 26–28, 31, 33, 37, 41, 44, 48] and 3 therapy response- and outcome-related studies [58–60]. The most commonly used radiomics software included MATLAB and the MATLAB-based packages ($n = 14$), LifeX ($n = 10$), MaZda ($n = 6$) and Pyradiomics ($n = 5$). A brief introduction of the software or the platforms used in the included studies is shown in Supplemental Table S5. There were 23, 16 and 6 studies that performed manual, semiautomatic and automatic segmentation, respectively.

Radiomics for diagnosis and classification of lymphoma

Thirty studies (66.7%) [20–49] explored the diagnostic feasibility of radiomics features in lymphoma (Table 1). The most common modality was MRI (16 of 30). Thirteen papers

[21–32, 42] focused on the differential diagnosis of primary central nervous system lymphoma (PCNSL) and glioblastoma (GBM). A majority of results showed that radiomics features can be used to effectively differentiate lymphoma from another disease, and the area under the curve (AUC) values of all the studies ranged from 0.730 to 1.000. Several studies [27, 28, 35, 42, 45–47] compared the diagnostic performance of radiomics features with that of other diagnostic methods (e.g., visual evaluation, the ADC value, PET metabolic parameters); the results indicated that radiomics features can diagnose or classify lymphoma more accurately. Only one study [32] reported that textural features are noninferior to expert human evaluation in the differentiation of PCNSL and enhancing glioma.

Radiomics for prediction of therapy responses and outcome of lymphoma

Sixteen studies (35.6%) [43, 50–64] examined the association between radiomics features and the therapy response or outcome of lymphoma (Table 2); of these studies, ^{18}F -FDG-PET and PET/CT were the most common modalities (11 of 16 studies). Nine studies demonstrated a promising relationship between radiomics features and therapy response in patients with Hodgkin's lymphoma (HL) [54, 58, 59, 63], non-Hodgkin's lymphoma (NHL) [50, 51, 57, 59], DLBCL [60] and follicular lymphoma (FL) [61]. Two studies [57, 60] reported that texture features could not evaluate the therapy

Fig. 1 Flowchart of study selection. n = number of records

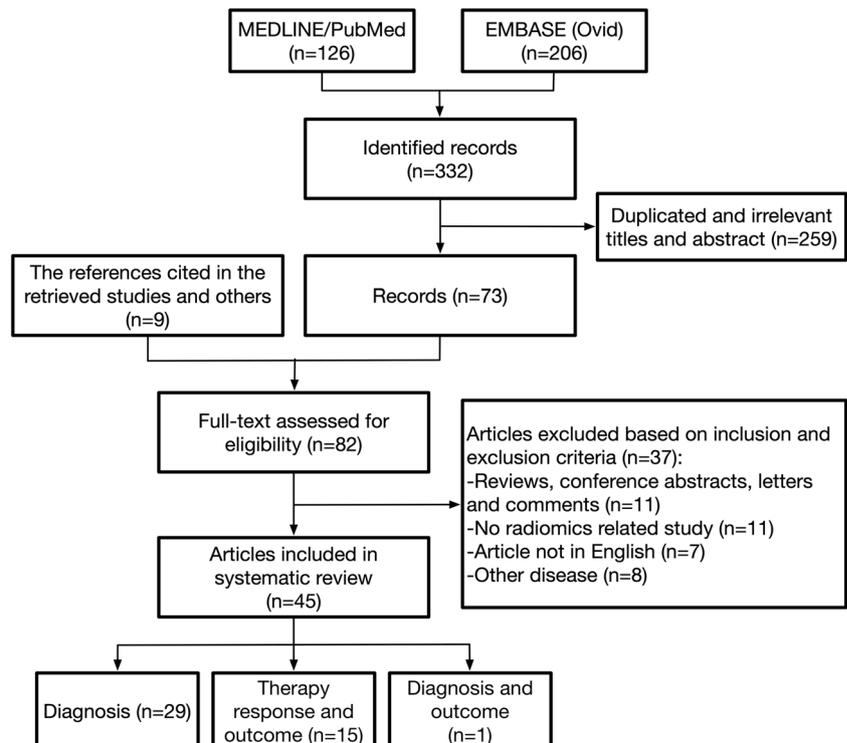


Table 1 Characteristics of the diagnosis-related studies

Reference	Disease	Number of patients	Modality	Software	Features	Segmentation	Main results ^a	Phase
MRI								
[20]	PCNSL and PNG	9 (5 vs. 4)	T1WI, T2WI, FLAIR, DWI, ADC, CE-T1WI	TexRAD	Statistical, histogram shape, distribution	M	AUC = 1.000	Discovery science
[21]	PCNSL and GBM	109 (28 vs. 81)	T2WI	ImageJ	Angular second moment, contrast, correlation, inverse difference moment, entropy	M	AUC = 0.917	Discovery science
[22]	PCNSL and GBM	102 (32 vs. 70)	T2WI, CE-T1WI	MATLAB	Deep learning features	A + S	ACC = 98.51%; SEN = 100%; SPE = 97.78%; AUC (T) = 0.945; AUC (IV) = 0.991; AUC (EV) = 0.947	Discovery science
[23]	PCNSL and GBM	195 (76 vs. 119)	CE-T1WI, DWI	MITK, MATLAB	936 RF (first-order, texture, wavelet)	S	AUC (T) = 0.945; AUC (IV) = 0.991; AUC (EV) = 0.947	II
[24]	PCNSL and GBM	76 (21 vs. 55)	CE-T1WI	ImageJ, MATLAB	67 TF	M	AUC = 0.990	0
[25]	PCNSL and GBM	60 (16 vs. 44)	CE-T1WI	ImageJ, MATLAB	67 TF	M	NR	Discovery science
[26]	PCNSL and GBM	143 (65 vs. 78)	CE-T1WI, T2WI, DWI	3D-slicer, ANTs, MATLAB	127 features (shape-based, histogram-based, TF)	S	AUC (T) = 0.979; AUC (EV) = 0.956	0
[27]	PCNSL and GBM	77 (54 vs. 23)	CE-T1WI, T2WI, FLAIR	Pyradiomics	6366 RF (shape, volume, first-order, texture, wavelet)	S	AUC = 0.921	0
[28]	PCNSL and GBM	154 (56 vs. 98)	CE-T1WI, ADC	MITK, MATLAB	1618 RF (first-order, volume and shape, texture, wavelet)	S	AUC (T) = 0.983; AUC (IV) = 0.984; AUC (EV) = 0.944	0
[29]	PCNSL and GBM	70 (25 vs. 45)	T2WI, ADC, rCBV, CE-T1WI	LifeX, Python	12 histograms and TF	M	AUC = 0.980	Discovery science
[30]	PCNSL and GBM	82 (22 vs. 60)	T1WI, CE-T1WI	ITK-SNAP, Pyradiomics, Python, Weka	105 TF and geometric features	M	AUC = 0.900	Discovery science
[31]	PCNSL and GBM	96 (30 vs. 66)	CE-T1WI	NR	Features using SIFT	A	AUC = 0.991	0
[32]	PCNSL and glioma	106 (35 vs. 71)	CE-T1WI	MATLAB	11 first-order, 142 second-order features	M	AUC = 0.877	Discovery science
[33]	Ocular adnexal lymphoma and inflammation	157 (84 vs. 73)	T2WI, CE-T1WI	ITK-SNAP, MATLAB	First-order statistics, TF	M	AUC (T) = 0.740; AUC (IV) = 0.730	0
[34]	HNSCC and malignant lymphoma	67 (57 vs. 10)	T2WI	MATLAB	Histogram and GLCM features	M	NR	Discovery science
[35]	DLBCL and FL	41 (30 vs. 11); 66 lesions (49 vs. 17)	T1WI, CE-T1WI, T2WI	Mazda	The RUN and COC	M	AUC = 0.800	Discovery science
CT								
[36]	Gastric lymphoma, AC, GIST	95 (13 vs. 50 vs. 32)	Enhanced arterial and venous phase	Mazda	Histogram, COC, RUN, gradient, AG model, wavelet	M	Arterial phase: misclassification rates = 0–3.1%; venous phase: misclassification rates = 4.4–10%	Discovery science
[37]	Gastric lymphoma and cancer	70 (30 vs. 40)	Portal venous phase	MATLAB	485 RF (first order, shape and size, TF, wavelet)	M	AUC = 0.903	Discovery science
[38]	Pancreatic lymphoma and adenocarcinoma	45 (15 vs. 30)	Pre-contrast, arterial, portal vein phase	Mazda	Histogram, gradient, RUN, COC, AG model, wavelet	M	AUC = 0.821	Discovery science

Table 1 (continued)

Reference	Disease	Number of patients	Modality	Software	Features	Segmentation	Main results ^a	Phase
[39]	CLL and DLBCL	52 (34 vs. 18)	Enhanced portal-venous phases	Pyradiomics	First, second and higher-order TF	M	AUC = 0.850	Discovery science
[40]	HNSCC, lymphoma, inflammatory, normal nodes	50; 412 lesions	Enhanced phases	TexRAD	6 first-order statistics	M	AUC (T) = 0.990; AUC (IV) = 0.960	0
[41]	Splenic infiltration of lymphoma, splenomegaly, normal spleens	74 (29 vs. 30 vs. 15)	Before and after treatment CT; enhanced portal-venous phases	Prototype	Heterogeneity, intensity, deviation, skewness, entropy, difference variance, NGLDM features	S	AUC = 0.729–0.857	Discovery science
¹⁸ F-FDG-PET or PET/CT								
[42]	PCNSL and GBM	77 (24 vs. 53)	PET	ITK-SNAP, SimpleITK, Pydicom, Pyradiomics	107 RF	M	AUC (T) = 0.971–0.998; AUC (IV) = 0.784–0.969	0
[43]	BMI (+) vs. BMI (-) in DLBCL	82 (22 vs. 60)	PET	LifeX	4 first-order; 6 second-order and 11 third-order features	S	AUC = 0.821	Discovery science
[44]	DLBCL, FL, HL, MCL	60	PET	MATLAB, CGITA	Shape-based, first-order, TF, higher-order	A	NR	0
[45]	Hepatic lymphoma and HCC	99 (23 vs. 76)	PET	LifeX	45 RF	S	AUC = 0.898	Discovery science
[46]	Renal lymphoma and RCC	38 (20 vs. 18)	PET	LifeX	RF	S	AUC = 1.000	Discovery science
[47]	Breast lymphoma and carcinoma	44 (19 vs. 25)	PET/CT	LifeX	First- and second-order RF	M	PET: AUC = 0.751; CT: AUC = 0.729; Combination of CT and PET: AUC = 0.771	Discovery science
[48]	Breast lymphoma and carcinoma	44 (19 vs. 25); 65 lesions (31 vs. 34)	PET/CT	LifeX	RF	M	PET: AUC (T) = 0.867; AUC (IV) = 0.806; CT: AUC (T) = 0.891; AUC (IV) = 0.759	0
[49]	Lymphoma and non-cancer	25; 188 lesions (156 vs. 32)	PET/CT	NR	105 features (first-order; GLCM, GLRLM, GLISZ and GLDM)	M	Combination of CT and PET: AUC = 0.910	Discovery science

PCNSL primary central nervous system lymphoma, PNG parenchymal neurosarcooidosis granulomas, GBM glioblastoma, HNSCC head and neck squamous cell carcinoma, DLBCL diffuse large B cell lymphoma, FL follicular lymphoma, AC adenocarcinoma, GIST gastrointestinal stromal tumours, CLL chronic lymphocytic leukaemia, BMJ bone marrow involvement, HL Hodgkin's lymphomas, MCL mantle cell lymphoma, HCC hepatocellular carcinoma, RCC renal cell carcinoma, T1WI T1-weighted image, T2WI T2-weighted image, FLAIR fluid-attenuated inversion recovery, DWI diffusion-weighted image, ADC apparent diffusion coefficient, rCBV relative cerebral blood volume, CE-T1WI contrast-enhanced T1-weighted image, RF radiomics features, TF texture features, SIFT scale invariant feature transform, RUN run-length matrix, COC co-occurrence, AG autoregressive, GLCM grey-level co-occurrence matrix, NGLDM neighbourhood grey-level different matrix, GLRLM grey-level run-length matrix, GLISZ grey-level intensity size zone matrix, GLDM grey-level difference matrix, M manual, S semiautomatic, A automatic, T training, IV internal validation, EV external validation, AUC area under the curve, ACC accuracy, SEN sensitivity, SPE specificity, NR not reported

^a All AUC values are the highest values in the studies

Table 2 Characteristics of therapy response- and outcome-related studies

Reference	Lymphoma type	Number of patients	Modality	Software	Features	Segmentation	Main results	Phase
MRI [50]	NHL	19	T1WI, T2WI	MaZda	Histogram, gradient, RUN, COC, AG model, wavelet	M	Correct classification, 96%	Discovery science
[51]	NHL	10	T1WI, T2WI	MaZda	Histogram, gradient, RUN, COC, AG model, wavelet	M	Misclassification rates, 5–13%	Discovery science
[52]	PCNSL	52	CE-T1WI	LifeX	10 TF	M	OS: HR = 3.075 (95% CI, 1.188–7.957; $p = 0.021$)	Discovery science
CT [53]	HL and high-grade NHL	45	Non-enhanced phase	TexRAD	Filtration-histogram-based TF	S	PFS: HR = 25.5 (95% CI, 5.4–120; $p < 0.001$)	Discovery science
[54]	HL	29 (48 lesions)	Enhanced venous phase	MaZda	Histogram, COC, RUN, AG model, gradient	S	Therapy response: ACC = 83.3%; SEN = 86.2%; SPE = 78.9%	Discovery science
¹⁸ F-FDG-PET or PET/CT [55]	Renal/adrenal lymphoma	19	Pretreatment PET	LifeX	37 TF	M	OS: HR = 9.016 (95% CI, 1.041–78.112; $p = 0.046$)	Discovery science
[56]	MCL	107	Pretherapy PET	Pyradiomics	16 TF	S	PFS: HR = 4.884 (95% CI, 1.647–14.607; $p = 0.004$)	Discovery science
[57]	NHL	82	Baseline PET	LifeX	GLCM, NGLDM, GLRLM, GLZLM, indices from sphericity and histogram	A	Therapy response: NS; DFS: $p = 0.013$; OS: $p = 0.035$	Discovery science
[58]	Mediastinal HL	251	Baseline PET	MIM, IBEX	GLCM, intensity histogram, shape	S	AUC = 0.952	0
[59] [43]	Bulky HL and NHL DLBCL	57 82	Baseline PET Baseline PET	In-house software LifeX	Shape, TF 4 first-order, 6 second-order and 11 third-order features	S S	AUC = 0.820 PFS: HR = 3.17 (95% CI, 1.00–10.04; $p = 0.032$)	0 Discovery science
[60] [61]	DLBCL FL	45 45	Before and after therapy PET Pretreatment PET	MATLAB PETSTAT	GLCM, GLRLM, GLSZM 6 TF	A A + M	Therapy response: NS Therapy response: AUC = 0.720;	Discovery science Discovery science
[62]	HL	42	Pretreatment PET	OsiriX, CGITA, MATLAB	11 first-order, 39 higher-order, 400 wavelet features	S	PFS: NS PFS: HR = 6.640 (95% CI, 1.261–34.96; $p = 0.026$); OS: HR = 14.54 (95% CI, 1.808–117.0; $p = 0.012$)	Discovery science

Table 2 (continued)

Reference	Lymphoma type	Number of patients	Modality	Software	Features	Segmentation	Main results	Phase
[63]	HL	35	Before therapy PET/CT	OsiriX, CGITA, MATLAB	7 SUV and HU, 78 second-order and higher-order, 624 wavelet features	S	PET: Therapy response: OR = 36.4 (95% CI, 2.060–642.0; $p = 0.014$); PFS: HR = 9.286 (95% CI, 1.341–66.28; $p = 0.023$); OS: HR = 41.02 (95% CI, 4.206–400.1; $p = 0.001$) CT: Therapy response: OR = 30.4 (95% CI, 1.700–545.0; $p = 0.014$); PFS: HR = 18.480 (95% CI, 1.918–178.1; $p = 0.012$); OS: NS PET: PFS: HR = 14.642 (95% CI, 2.661–80.549; $p = 0.002$); OS: HR = 28.685 (95% CI, 2.067–398.152; $p = 0.012$) CT: PFS: HR = 11.504 (95% CI, 1.921–68.888; $p = 0.007$); OS: HR = 11.791 (95% CI, 1.583–87.808; $p = 0.016$)	Discovery science
[64]	Primary gastric DLBCL	35	Pretreatment PET/CT	LifeX	44 TF	S		Discovery science

NHL non-Hodgkin's lymphomas, *PCNSL* primary central nervous system lymphoma, *HL* Hodgkin's lymphomas, *MCL* mantle cell lymphoma, *DLBCL* diffuse large B cell lymphoma, *FL* follicular lymphoma, *T1WI* T1-weighted image, *T2WI* T2-weighted image, *CE-T1WI* contrast-enhanced T1-weighted image, *TF* texture features, *RUN* run-length matrix, *COC* co-occurrence, *AG* autoregressive, *GLCM* grey-level co-occurrence matrix, *NGLDM* neighbourhood grey-level different matrix, *GLRLM* grey-level run length matrix, *GLZLM* grey-level zone length matrix, *GLSZM* grey-level size-zone matrix, *HU* Hounsfield unit, *M* manual, *S* semiautomatic, *A* automatic, *PFS* progression-free survival, *DFS* disease-free survival, *HR* hazard ratio, *OR* odds ratio, *AUC* area under the curve, *ACC* accuracy, *SEN* sensitivity, *SPE* specificity, *NS* not significant

response of NHL, although several features were correlated with the presence of a residual mass at the end of therapy [57].

In addition to treatment response, ten studies focused on the outcome of lymphoma and found that radiomics features are prognostic predictors for the outcome of patients with HL [53, 62, 63], DLBCL [43, 64] and other types of lymphoma [52, 53, 55–57, 61]. Several studies [43, 53, 55, 56, 63, 64] reported a better performance of radiomics features than that of clinical parameters (e.g., Ann Arbor stage) and conventional parameters. Only one study [61] reported that texture features were not significant predictors of progression-free survival (PFS) in FL.

Quality analysis of the included studies

The risk of bias and applicability concerns according to the QUADAS-2 for 30 diagnosis-related studies are shown in Fig. 2 (Supplemental Table S6). Regarding patient selection, 20 studies were considered to have a high or an unclear bias risk because of the unclearness of the detailed exclusion criteria [20–23, 31, 38, 39, 44, 49] and/or the limitation of

case-control studies [20, 27, 33, 34, 39, 45, 47–49]. Regarding the index test, 10 studies were considered to have an unclear bias risk, because no threshold was used and the blinding between the index test and reference standard was unclear [22, 29–31, 36–38, 43, 44, 49]. The applicability concern was high for 19 studies [20–22, 24, 25, 28, 31, 32, 34–36, 38, 41, 43–47, 49] because the variability in the parameters of the radiomics approach was not detected. No study was considered as high risk in the domain of reference standard because the reference standard of all studies is pathology.

According to the phase classification criteria for image mining studies, 32 of 45 (71.1%) were classified as discovery science and 12 of 45 (26.7%) were classified into phase 0 (Fig. 3a). The details of the RQS of all included studies and the mean RQS between reviewers are provided in Supplemental Table S7. The mean RQS of all studies were 14.2% (ranging from 0.0 to 40.3%), and 23 of 45 (51.1%) studies were given a quality score of < 10% (Fig. 3b, c). The inter-observer agreement is presented in Supplemental Table S8. According to the weighted kappa coefficient, the inter-reviewer reliability of 11 domains was moderate to good

Fig. 2 Grouped bar charts of the risk of bias and concern for applicability of the included diagnosis-related studies using QUADAS-2

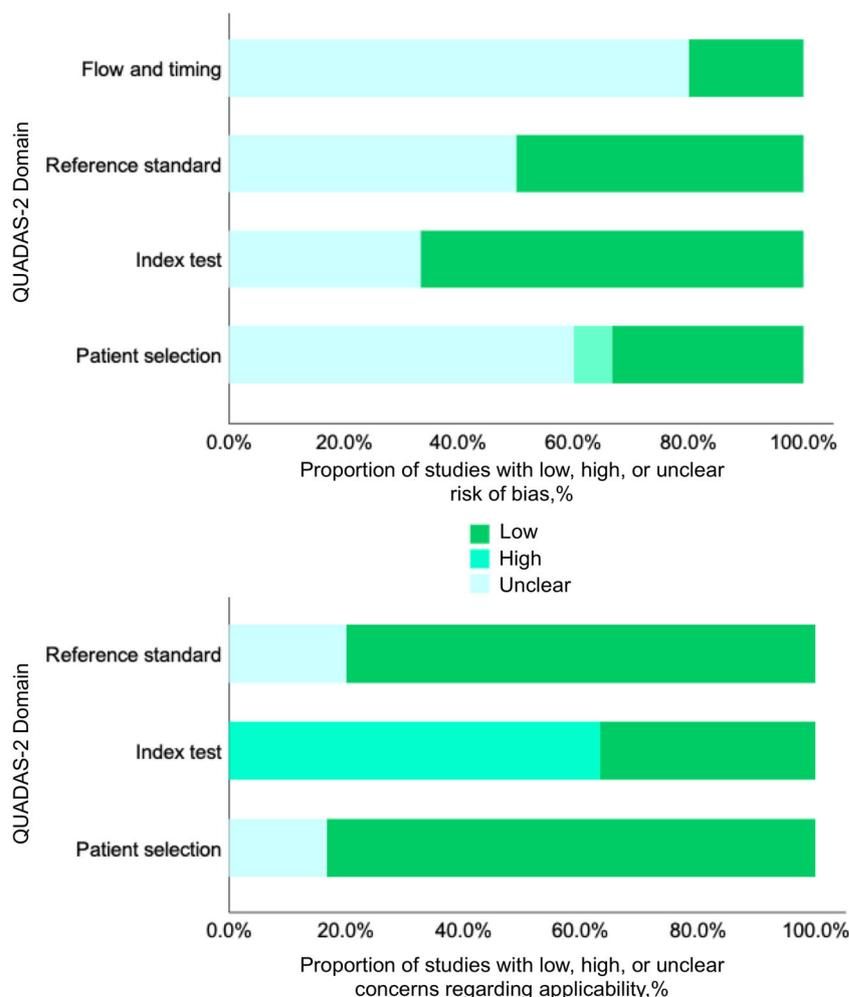
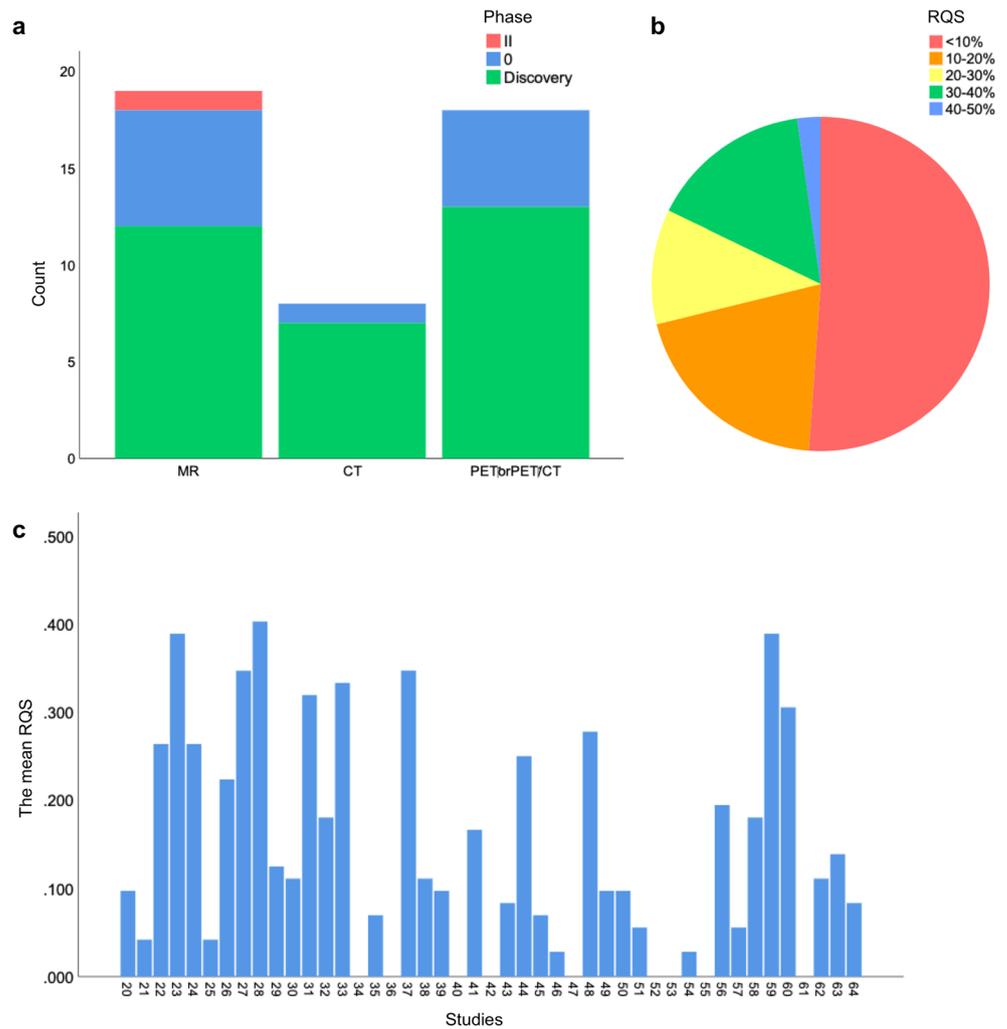


Fig. 3 **a** Histogram of the phase of studies according to the modality; **b** pie chart of the mean RQS score of studies; **c** histogram of the mean RQS score of each study, where the mean RQS scores are listed from left to right in order of citation



(0.60–1.00). Nevertheless, the inter-rater reliability was poor in the “Comparison to gold standard” (0.394) and the “Potential clinical utility” (0.036). In addition, the scores on the “Image protocol quality” and the “Prospective study registered in a trial database” are completely consistent between the two reviewers. The scores on the “Cost-effectiveness analysis” are discrepant (one reviewer gave all the studies a 0), and the inter-reviewer agreement of the above domains could not be assessed by the weighted kappa coefficient.

Discussion

The present systematic review explored whether radiomics could provide information about the diagnosis, therapy response and outcome of lymphoma, and evaluated the quality of included studies using the QUADAS-2, the phase classification criteria for image mining studies, and the RQS. Despite promising results, those studies were far from providing definitive conclusions for clinical implementation and

widespread use of radiomics with regard to immature phases and relatively poor quality.

Notably, 44 of 45 (97.8%) studies were classified as the discovery science or phase 0, because of a lack of validation analysis, sample size limitation (< 100 patients) and/or retrospective design. For example, one study with 77 patients only reported a high diagnostic accuracy (AUC = 0.921) of radiomics features in the training cohort, but this result is less reliable and needs to be further validated [27]. It has been proposed that at least 10 to 15 observations per variable will produce stable estimates for a linear model [65], and at least 50 patients are suggested to be included in radiomics studies [15]. Additionally, a limited number of studies performed external validation analysis [23, 26, 28]. Ideally, the performance of the prediction model should be validated on a dataset that is independent from the training dataset. Another available option is the cross-validation if independent validation cohorts are not available [66]. A systematic review [15] analysed 141 studies on radiomics and found that 84% of studies were classified

as phase II–III, but they only included high-quality studies and excluded studies that lacked validation or conventional metrics for the report of validation results. Thus, the current results of lymphoma radiomics research are still immature.

The lack of a rigorous procedure largely contributed to the low RQS scores of the included studies. Some studies only reported the association of radiomics features and lymphoma characteristics without the awareness of establishing a prediction model [21, 25, 34, 50, 51, 54]. Few studies have analysed feature robustness considering differences across machines, segmentations or temporal variability. In addition, the association and combination of radiomics and genomics, known as “radiogenomics”, were absent in the current studies. Further radiogenomic studies are needed to reveal predictive radiomic signatures and gene expression patterns and to broaden clinical imaging into genomic and molecular imaging. These results are similar to those of a previous systematic review [67], which reported that a majority of included studies scored below 50%. Another article [68] reported a mean RQS score of 26.1% in 77 radiomics studies with a high-impact factor, and the RQS scores were low in the domains of test–retest analysis, prospective study, clinical utility, and open science. In the last few years, several guidelines have been proposed to encourage radiomics workflows to be described in detail and provide suggestions for research related to the construction of models [69–71]. These guidelines need to be pursued for radiomics to become a viable tool in clinical situations.

There were some discrepancies in the inter-reviewer interpretation of the RQS tool. Eleven studies [26, 27, 29, 32, 33, 38, 43, 47, 54, 57, 59] received different scores in “Comparison to gold standard” by two reviewers, which may be explained by the unclear definition of the current gold standard method, and no gold standard exists for the assessment of the outcome (e.g., changes in NHL tissue during chemotherapy). Additionally, the contents of some domains may be subjectively interpreted by reviewers. For example, the domain of “Potential clinical utility” suggested reporting the application of the model in a clinical setting. One reviewer regarded decision curve analysis as the only criterion, while another also considered statements in these studies. Finally, some components may require a professional background and sufficient knowledge of statistics, which limits their widespread use and promotion, such as “Calibration statistics” (report calibration-in-the-large or calibration plots). The complex radiomics workflow contributed to the complex evaluation components of the RQS tool and posed challenges to inter-reviewer agreement. As radiomics is still a work-in-progress biomarker and is developing constantly, the RQS tool should be revised as well to become a widely accepted tool for radiological research methodologies [67, 68].

Segmentation methods of lymphoma are contentious due to the multiplicity and multifocality of lymphoma, especially

in lesions with indistinct borders. A total of 23 of 45 (51.1%) studies performed manual delineation methods. Although good inter-observer repeatability of manual segmentation was reported [25, 32, 35], it is a subjective and labourious method related to high intra- and interoperator variations [57, 72]. Additionally, different segmentation methods may include and exclude different tumour areas [73, 74], and further studies should explore what type of segmentation is more suitable for lymphoma. For instance, it was found that in PCNSL and GBM, the number of features increased after expanding the edge of the ROI, suggesting that the peritumour area may contain valuable information [42]. Several studies performed multiple segmentations [26, 42, 53, 57, 59] and reported better segmentation methods with respect to outcome prediction [57, 59].

The values of radiomics features may be significantly influenced by tumour volumes, particularly in tumours with relatively small size [75–77]. This is why previous studies recommended cutoff values of tumour volumes for radiomics analysis, for instance, 10 cm³ [75], 45 cm³ [76] and 60 cm³ [77]. Most of the included studies did not report the volume of target lesions or selected lesions according to the abovementioned criterion, thereby affecting the robustness of the results. For PET, some radiomics features also depend on metabolism [78]. Several studies defined conventional PET metabolic parameters (e.g., SUV-based parameters) as radiomics features and/or considered them independently of each other [55, 56, 61–64]. This may contribute to feature redundancy and underestimated performance of radiomics. Therefore, selecting lesions with caution and investigating the interactions between features are recommended for further research.

Inter- and intratumoural heterogeneities vary with the pathological pattern of lymphoma [13]. Lymphoma lesions located in different organs or tissues can provide different radiomics information, considering the different contrast and signal-to-noise ratio of images [35]. Published studies have mainly classified lymphoma according to affected organs [20–32, 36–38, 42, 45–48], and some studies did not present a detailed pathology pattern [33, 50, 53], which may contribute to confusing results. One study with 60 patients [44] classified lymphoma subtypes with learning algorithms. For HL, the sensitivity and predictive positive value of this approach were 97.0% and 94.1%, respectively. However, the small number of patients within each pathological type subgroup could have biased the results. It is unclear whether the change in features due to the pathology pattern is larger than that of the location or not, which should be evaluated in future work [15].

¹⁸F-FDG-PET/CT plays an indispensable role in the diagnosis, staging, therapeutic evaluation and follow-up management of lymphoma [79]. Eighteen of the included studies

(40.0%) extracted radiomics features from PET images. While a limited number of the included studies compared or integrated radiomics features extracted from PET and CT images [47–49, 63, 64], it was found that combined modalities achieved a better performance than a single modality [80]. Therefore, the performance improvement by use of multimodality features should be investigated. Additionally, among 11 studies focused on the predictive and prognostic value of PET, ten studies (90.9%) derived features from pre-treatment images. Only one study [60] explored the predictive value of features derived from pre- and posttreatment images in solid tumours and lymphoma, and the results showed that the change in radiomics features may be associated with treatment response in solid tumours. Longitudinal features derived from images at multiple time points can provide complementary information and improve prediction performance [60]. Thus, analysing features from images at different time points (pretreatment, during treatment and posttreatment) is useful to illuminate temporal variabilities of radiomics features and to deepen the understanding of radiomics potential in lymphoma treatment decisions.

The limitations of this systematic review are as follows. First, only published English language articles were included. Second, the body of literature is small when considering the number of lymphoma subtypes, treatments and imaging techniques. Third, we did not validate extracted study characteristics.

Conclusions

In summary, the radiomics approach shows potential utility for diagnosis, therapy response evaluation and outcome prediction in lymphoma. Nevertheless, the immature phases and low quality of the studies imply that current conclusions are not robust. Before radiomics can be successfully introduced into lymphoma clinical settings, more research is needed to examine lesion segmentation and selection, the influence of pathological pattern and the extraction of multiple modalities and multiple time point features.

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

Compliance with ethical standards

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

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

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was not required for this study because it is a systematic review.

Ethical approval Institutional review board approval was not required because it is a systematic review.

Methodology

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

References

1. Lambin P, Rios-Velazquez E, Leijenaar R et al (2012) Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 48:441–446
2. Patyk M, Silicki J, Mazur R, Kręcichwost R, Sokołowska-Dąbek D, Zaleska-Dorobisz U (2018) Radiomics – the value of the numbers in present and future radiology. *Pol J Radiol* 83:e171–e174
3. Acharya UR, Hagiwara Y, Sudarshan VK, Chan WY, Ng KH (2018) Towards precision medicine: from quantitative imaging to radiomics. *J Zhejiang Univ Sci B* 19:6–24
4. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D (2017) Characterization of PET/CT images using texture analysis: the past, the present... any future? *Eur J Nucl Med Mol Imaging* 44: 151–165
5. McGranahan N, Swanton C (2017) Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell* 168:613–628
6. Schwarz RF, Ng CKY, Cooke SL et al (2015) Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. *PLoS Med* 12:e1001789
7. Wilson R, Devaraj A (2017) Radiomics of pulmonary nodules and lung cancer. *Transl Lung Cancer Res* 6:86–91
8. Tandel GS, Biswas M, Kakde OG et al (2019) A review on a deep learning perspective in brain cancer classification. *Cancers (Basel)* 11
9. Wong AJ, Kanwar A, Mohamed AS, CD Fuller (2016) Radiomics in head and neck cancer: from exploration to application. *Transl Cancer Res* 5:371–382
10. Jeong WK, Jamshidi N, Felker ER, SS Raman, Lu DS (2019) Radiomics and radiogenomics of primary liver cancers. *Clin Mol Hepatol* 25:21–29
11. Horvat N, Bates DDB, Petkovska I (2019) Novel imaging techniques of rectal cancer: what do radiomics and radiogenomics have to offer? A literature review. *Abdom Radiol (NY)* 44:3764–3774
12. Swerdlow SH, Campo E, Pileri SA et al (2016) The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* 127:2375–2390
13. Schürch CM, Federmann B, Quintanilla-Martinez L, Fend F (2018) Tumor heterogeneity in lymphomas: a different breed. *Pathobiology* 85:130–145
14. Lambin P, Leijenaar RTH, Deist TM et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 14:749–762
15. Sollini M, Antunovic L, Chiti A, Kirienko M (2019) Towards clinical application of image mining: a systematic review on artificial intelligence and radiomics. *Eur J Nucl Med Mol Imaging* 46:2656–2672

16. Zorzela L, Loke YK, Ioannidis JP et al (2016) PRISMA harms checklist: improving harms reporting in systematic reviews. *BMJ* 352:i157
17. Whiting PF, Rutjes AWS, Westwood ME et al (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 155:529–536
18. Sheskin D (2007) Handbook of parametric and nonparametric statistical procedures, 4th edn. Chapman & Hall/CRC, Web, Boca Raton
19. Joseph LF, Bruce L, Myunghee Cho P (2004) Wiley series in probability and statistics. John Wiley & Sons, Web, Hoboken
20. Bathla G, Soni N, Endozo R, Ganeshan B (2019) Magnetic resonance texture analysis utility in differentiating intraparenchymal neurosarcoidosis from primary central nervous system lymphoma: a preliminary analysis. *Neuroradiol J* 32:203–209
21. Wang B, Liu M, Chen Z (2019) Differential diagnostic value of texture feature analysis of magnetic resonance T2 weighted imaging between glioblastoma and primary central neural system lymphoma. *Chin Med Sci J* 34:10–17
22. Wu G, Chen Y, Wang Y et al (2018) Sparse representation-based radiomics for the diagnosis of brain tumors. *IEEE Trans Med Imaging* 37:893–905
23. Yun J, Park JE, Lee H, Ham S, Kim N, Kim HS (2019) Radiomic features and multilayer perceptron network classifier: a robust MRI classification strategy for distinguishing glioblastoma from primary central nervous system lymphoma. *Sci Rep* 9:5746
24. Kunimatsu A, Kunimatsu N, Yasaka K et al (2019) Machine learning-based texture analysis of contrast-enhanced mr imaging to differentiate between glioblastoma and primary central nervous system lymphoma. *Magn Reson Med Sci* 18:44–52
25. Kunimatsu A, Kunimatsu N, Kamiya K, Watadani T, Mori H, Abe O (2018) Comparison between glioblastoma and primary central nervous system lymphoma using MR image-based texture analysis. *Magn Reson Med Sci* 17:50–57
26. Kim Y, Cho H h, Kim ST, Park H, Nam D, Kong DS (2018) Radiomics features to distinguish glioblastoma from primary central nervous system lymphoma on multi-parametric MRI. *Neuroradiology* 60:1297–1305
27. Suh HB, Choi YS, Bae S et al (2018) Primary central nervous system lymphoma and atypical glioblastoma: differentiation using radiomics approach. *Eur Radiol* 28:3832–3839
28. Kang D, Park JE, Kim YH et al (2018) Diffusion radiomics as a diagnostic modal for atypical manifestation of primary central nervous system lymphoma: development and multicenter external validation. *Neuro Oncol* 20:1251–1261
29. Nakagawa M, Nakaura T, Namimoto T et al (2018) Machine learning based on multi-parametric magnetic resonance imaging to differentiate glioblastoma multiforme from primary cerebral nervous system lymphoma. *Eur J Radiol* 108:147–154
30. Xiao DD, Yan PF, Wang YX, Osman MS, Zhao HY (2018) Glioblastoma and primary central nervous system lymphoma: pre-operative differentiation by using MRI-based 3D texture analysis. *Clin Neurol Neurosurg* 173:84–90
31. Chen Y, Li Z, Wu G et al (2018) Primary central nervous system lymphoma and glioblastoma differentiation based on conventional magnetic resonance imaging by high-throughput SIFT features. *Int J Neurosci* 128:608–618
32. Alcaide-Leon P, Dufort P, Geraldo AF et al (2017) Differentiation of enhancing glioma and primary central nervous system lymphoma by texture-based machine learning. *AJNR Am J Neuroradiol* 38:1145–1150
33. Guo J, Liu Z, Shen C et al (2018) MR-based radiomics signature in differentiating ocular adnexal lymphoma from idiopathic orbital inflammation. *Eur Radiol* 28:3872–3881
34. Fujima N, Homma A, Harada T et al (2019) The utility of MRI histogram and texture analysis for the prediction of histological diagnosis in head and neck malignancies. *Cancer Imaging* 19:5
35. Wu X, Sikiö M, Pertovaara H et al (2016) Differentiation of diffuse large B-cell lymphoma from follicular lymphoma using texture analysis on conventional MR images at 3.0 tesla. *Acad Radiol* 23:696–703
36. Ba-Ssalamah A, Muin D, Scherthaner R et al (2013) Texture-based classification of different gastric tumors at contrast-enhanced CT. *Eur J Radiol* 82:e537–e543
37. Ma Z, Fang M, Huang Y et al (2017) CT-based radiomics signature for differentiating Borrmann type IV gastric cancer from primary gastric lymphoma. *Eur J Radiol* 91:142–147
38. Huang Z, Li M, He D et al (2019) Two-dimensional texture analysis based on CT images to differentiate pancreatic lymphoma and pancreatic adenocarcinoma: a preliminary study. *Acad Radiol* 26:e189–e195
39. Reinert CP, Federmann B, Hofmann J et al (2019) Computed tomography textural analysis for the differentiation of chronic lymphocytic leukemia and diffuse large B cell lymphoma of Richter syndrome. *Eur Radiol* 29:6911–6921
40. Seidler M, Forghani B, Reinhold C et al (2019) Dual-energy CT texture analysis with machine learning for the evaluation and characterization of cervical lymphadenopathy. *Comput Struct Biotechnol J* 17:1009–1015
41. Reinert CP, Kloth C, Fritz J, Nikolaou K, Horger M (2018) Discriminatory CT-textural features in splenic infiltration of lymphoma versus splenomegaly in liver cirrhosis versus normal spleens in controls and evaluation of their role for longitudinal lymphoma monitoring. *Eur J Radiol* 104:129–135
42. Kong Z, Jiang C, Zhu R et al (2019) 18F-FDG-PET-based radiomics features to distinguish primary central nervous system lymphoma from glioblastoma. *Neuroimage Clin* 23:101912
43. Aide N, Talbot M, Fruchart C, Damaj G, Lasnon C (2018) Diagnostic and prognostic value of baseline FDG PET/CT skeletal textural features in diffuse large B cell lymphoma. *Eur J Nucl Med Mol Imaging* 45:699–711
44. Lippi M, Gianotti S, Fama A et al (2019) Texture analysis and multiple-instance learning for the classification of malignant lymphomas. *Comput Methods Prog Biomed* 185:105153
45. Xu H, Guo W, Cui X et al (2019) Three-dimensional texture analysis based on PET/CT images to distinguish hepatocellular carcinoma and hepatic lymphoma. *Front Oncol* 9:844
46. Zhu S, Xu H, Shen C et al (2019) Differential diagnostic ability of 18F-FDG PET/CT radiomics features between renal cell carcinoma and renal lymphoma. *Q J Nucl Med Mol Imaging*. <https://doi.org/10.23736/S1824-4785.19.03137-6>
47. Ou X, Wang J, Zhou R et al (2019) Ability of 18 F-FDG PET / CT radiomic features to distinguish breast carcinoma from breast lymphoma. *Contrast Media Mol Imaging* 2019:4507694
48. Ou X, Zhang J, Wang J et al (2019) Radiomics based on 18F-FDG PET/CT could differentiate breast carcinoma from breast lymphoma using machine-learning approach: a preliminary study. *Cancer Med* 9:496–506
49. Lartzien C, Rogez M, Niaf E, Ricard F (2014) Computer-aided staging of lymphoma patients with FDG PET/CT imaging based on textural information. *IEEE J Biomed Health Inform* 18:946–955
50. Harrison LC, Luukkaala T, Pertovaara H et al (2009) Non-Hodgkin lymphoma response evaluation with MRI texture classification. *J Exp Clin Cancer Res* 28:87
51. Harrison L, Dastidar P, Eskola H et al (2008) Texture analysis on MRI images of non-Hodgkin lymphoma. *Comput Biol Med* 38:519–524
52. Chen C, Zhuo H, Wei X, Ma X (2019) Contrast-enhanced MRI texture parameters as potential prognostic factors for primary central nervous system lymphoma patients receiving high-dose

- methotrexate-based chemotherapy. *Contrast Media Mol Imaging* 2019:5481491
53. Ganeshan B, Miles KA, Babikir S et al (2017) CT-based texture analysis potentially provides prognostic information complementary to interim FDG-pet for patients with Hodgkin's and aggressive non-Hodgkin's lymphomas. *Eur Radiol* 27:1012–1020
 54. Knogler T, El-Rabadi K, Weber M, Karanikas G, Mayerhoefer ME (2014) Three-dimensional texture analysis of contrast enhanced CT images for treatment response assessment in Hodgkin lymphoma: comparison with F-18-FDG PET. *Med Phys* 41:121904
 55. Wang M, Xu H, Xiao L, Song W, Zhu S, Ma X (2019) Prognostic value of functional parameters of 18 F-FDG-PET images in patients with primary renal/adrenal lymphoma. *Contrast Media Mol Imaging* 2019:2641627
 56. Mayerhoefer ME, Riedl CC, Kumar A et al (2019) Radiomic features of glucose metabolism enable prediction of outcome in mantle cell lymphoma. *Eur J Nucl Med Mol Imaging* 46:2760–2769
 57. Parvez A, Tau N, Hussey D, Maganti M, Metser U (2018) 18F-FDG PET/CT metabolic tumor parameters and radiomics features in aggressive non-Hodgkin's lymphoma as predictors of treatment outcome and survival. *Ann Nucl Med* 32:410–416
 58. Milgrom SA, Elhalawani H, Lee J et al (2019) A PET radiomics model to predict refractory mediastinal Hodgkin lymphoma. *Sci Rep* 9:1322
 59. Ben Bouallègue F, Tabaa YA, Kafrouni M, Cartron G, Vauchot F, Mariano-Goulard D (2017) Association between textural and morphological tumor indices on baseline PET-CT and early metabolic response on interim PET-CT in bulky malignant lymphomas. *Med Phys* 44:4608–4619
 60. Wu J, Lian C, Ruan S et al (2018) Treatment outcome prediction for cancer patients based on radiomics and belief function theory. *IEEE Trans Radiat Plasma Med Sci* 3:216–224
 61. Tatsumi M, Isohashi K, Matsunaga K et al (2019) Volumetric and texture analysis on FDG PET in evaluating and predicting treatment response and recurrence after chemotherapy in follicular lymphoma. *Int J Clin Oncol* 24:1292–1300
 62. Lue KH, Wu YF, Liu SH et al (2019) Prognostic value of pretreatment radiomic features of 18F-FDG PET in patients with Hodgkin lymphoma. *Clin Nucl Med* 44:e559–e565
 63. Lue KH, Wu YF, Liu SH et al (2019) Intratumor heterogeneity assessed by 18F-FDG PET/CT predicts treatment response and survival outcomes in patients with Hodgkin lymphoma. *Acad Radiol*. <https://doi.org/10.1016/j.acra.2019.10.015>
 64. Zhou Y, Ma XL, Pu LT, Zhou RF, Ou XJ, Tian R (2019) Prediction of overall survival and progression-free survival by the 18F-FDG PET/CT radiomic features in patients with primary gastric diffuse large B-cell lymphoma. *Contrast Media Mol Imaging* 2019:5963607
 65. Chalkidou A, O'Doherty MJ, Marsden PK (2015) False discovery rates in PET and CT studies with texture features: a systematic review. *PLoS One* 10:e0124165
 66. Forghani R, Savadjiev P, Chatterjee A, Muthukrishnan N, Reinhold C, Forghani B (2019) Radiomics and artificial intelligence for biomarker and prediction model development in oncology. *Comput Struct Biotechnol J* 17:995–1008
 67. Sanduleanu S, Woodruff HC, de Jong EEC et al (2018) Tracking tumor biology with radiomics: a systematic review utilizing a radiomics quality score. *Radiother Oncol* 127:349–360
 68. Park JE, Kim D, Kim HS et al (2019) Quality of science and reporting of radiomics in oncologic studies: room for improvement according to radiomics quality score and TRIPOD statement. *Eur Radiol* 30:523–536
 69. Liu Z, Wang S, Dong D et al (2019) The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. *Theranostics* 9:1303–1322
 70. Kalpathy-Cramer J, Freymann JB, Kirby JS, Kinahan PE, Prior FW (2014) Quantitative imaging network: data sharing and competitive algorithm validation leveraging the cancer imaging archive. *Transl Oncol* 7:147–152
 71. Collins GS, Reitsma JB, Altman DG, Moons KGM, (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 350:g7594
 72. Pavic M, Bogowicz M, Würms X et al (2018) Influence of inter-observer delineation variability on radiomics stability in different tumor sites. *Acta Oncol* 57:1070–1074
 73. Geets X, Lee JA, Bol A, Lonneux M, Grégoire V (2007) A gradient-based method for segmenting FDG-PET images: methodology and validation. *Eur J Nucl Med Mol Imaging* 34:1427–1438
 74. Hatt M, Cheze le Rest C, Descourt P et al (2010) Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications. *Int J Radiat Oncol Biol Phys* 77:301–308
 75. Orhac F, Soussan M, Chouahnia K, Martinod E, Buvat I (2015) 18F-FDG PET-derived textural indices reflect tissue-specific uptake pattern in non-small cell lung cancer. *PLoS One* 10:e0145063
 76. Brooks FJ, Grigsby PW (2014) The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. *J Nucl Med* 55:37–42
 77. Hatt M, Majdoub M, Vallieres M et al (2015) 18F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. *J Nucl Med* 56:38–44
 78. Ha S, Choi H, Paeng JC et al (2019) Radiomics in oncological PET/CT: a methodological overview. *Nucl Med Mol Imaging* 53:14–29
 79. Wang H, Shen G, Jiang C, Li L, Cui F, Tian R (2018) Prognostic value of baseline, interim and end-of-treatment 18F-FDG PET/CT parameters in extranodal natural killer/T-cell lymphoma: a meta-analysis. *PLoS One* 13:e0194435
 80. Sollini M, Cozzi L, Ninatti G et al (2020) PET/CT radiomics in breast cancer: mind the step. *Methods*. <https://doi.org/10.1016/j.ymeth.2020.01.007>

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