

Diagnostic Value of ^{18}F -FDG PET/CT Radiomics in Lymphoma: A Systematic Review and Meta-Analysis

Chaoying Liu, MD^{1,2,3,4} , Jun Zhao, PhD⁵, Heng Zhang, PhD^{2,3,4,6} 
 and Xinye Ni, PhD^{2,3,4,6} 

Abstract

Introduction: Various machine learning models and features have been proposed for lymphoma diagnosis using ^{18}F -fluorodeoxyglucose (^{18}F -FDG) PET/CT radiomics. This research aimed to systematically evaluate the diagnostic value of ^{18}F -FDG PET/CT radiomics in lymphoma by conducting a meta-analysis.

Methods: Data from published studies regarding the diagnosis of lymphoma using ^{18}F -FDG PET/CT radiomics, from January 2010 to July 2024, were gathered from PubMed, Web of Science, and the Cochrane Library. Following their separate searches and screenings of the literature, two researchers extracted data and assessed the caliber of all the included studies. The quality assessment involved the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2), the Radiomics Quality Score (RQS), and the METHodological RadiomIcs Score (METRICS). The meta-analysis was conducted by using RevMan 5.4.1, R 4.4.0, and Stata 17.0 software. Six meta-regressions were conducted on study performance, considering sample size, image modality, region of interest (ROI) selection, ROI segmentation, radiomics mode, and algorithms.

Results: In total, 20 studies classified as type 2a or above according to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement were included for this systematic review and meta-analysis. The studies achieved an average RQS of 13 (ranging from 10 to 17), accounting for 36.1% of the total points. The average METRICS score was 69.3% (ranging from 54.8% to 80.9%). The quality category of the studies is mainly “good”. The results of our meta-analysis showed that the pooled sensitivity (SEN), specificity (SPE), positive likelihood ratio, negative likelihood ratio and diagnostic odds ratio with 95% confidence interval (CI) were 0.82 (0.78, 0.88), 0.83 (0.76, 0.87), 4.7 (3.4, 6.6), 0.20 (0.15, 0.28) and 23 (13, 42), respectively. The area under the curve of the summary receiver operating characteristic curve was 0.90 (0.87, 0.92). The results of Spearman correlation analysis revealed no threshold effect among the studies ($P = .423$). Significant heterogeneity was observed among the studies (overall $I^2 = 83.7\%$; 95% CI: 76.0, 88.9; $P < .01$). Meta-regressions indicated that sample size and ROI selection contributed to the heterogeneity in SEN, while algorithms affected the heterogeneity in SPE ($P < .05$). Deeks’ test confirmed there was no significant publication bias in all the included studies. The Fagan nomogram showed an absolute increase of 34% in the post-test probability following a positive test result.

Conclusion: The results supported that ^{18}F -FDG PET/CT radiomics has high diagnostic value for lymphoma. However, there is high heterogeneity among different studies. In the future, clinical practicality needs to be substantiated by more prospective studies with rigorous adherence to existing guidelines and multicentric validation.

¹Department of Medical Equipment, the Third Affiliated Hospital of Nanjing Medical University, The Second People's Hospital of Changzhou, Changzhou, Jiangsu, China

²Jiangsu Province Engineering Research Center of Medical Physics, Changzhou, Jiangsu, China

³Medical Physics Research Center, Nanjing Medical University, Changzhou, Jiangsu, China

⁴Key Laboratory of Medical Physics in Changzhou, Changzhou, Jiangsu, China

⁵Department of Nuclear Medicine, the Third Affiliated Hospital of Nanjing Medical University, The Second People's Hospital of Changzhou, Changzhou, Jiangsu, China

⁶Department of Radiotherapy, the Third Affiliated Hospital of Nanjing Medical University, The Second People's Hospital of Changzhou, Changzhou, Jiangsu, China

Corresponding Author:

Xinye Ni, Department of Radiotherapy, the Third Affiliated Hospital of Nanjing Medical University, The Second People's Hospital of Changzhou, Changzhou, Jiangsu, 213003, China; Jiangsu Province Engineering Research Center of Medical Physics, Changzhou, Jiangsu 213003, China; Medical Physics Research Center, Nanjing Medical University, Changzhou, Jiangsu 213003, China; Key Laboratory of Medical Physics in Changzhou, Changzhou Jiangsu 213003, China.

Email: nxy@njmu.edu.cn



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and

distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Keywords

radiomics, ^{18}F -FDG, diagnosis, lymphoma, meta-analysis

Received: 11 February 2025; revised: 15 April 2025; accepted: 28 April 2025

Introduction

Lymphoma is a highly heterogeneous hematological malignancy that can affect any organ in the body. Biopsy remains the gold standard for lymphoma diagnosis, although it cannot provide the whole situation of the tumor and only yields limited information. Therefore, efficient and non-invasive diagnostic indicators still have important clinical significance in evaluating individual disease risk.^{1,2} ^{18}F -fluorodeoxyglucose (^{18}F -FDG) PET/CT is an important imaging evaluation technique in the diagnosis of lymphoma, which helps identify the anatomy and quantify the metabolic characteristics of highly invasive lymphomas.^{3,4} The clinical application of ^{18}F -FDG PET/CT is growing daily as living standards rise, but its limitations are also gradually exposed: routine examinations often reveal the same disease, but images differ because of the high tumor heterogeneity, which lowers the diagnostic accuracy of lymphoma.

In recent years, radiomics has emerged with the evolution of computer hardware and image analysis technology. Its advantage lies in obtaining a large number of image features that are difficult to distinguish by the naked eyes through automated algorithms and quantifying them into numerical data to increase the accuracy and efficiency of image diagnosis. This technology is expected to reduce subjective errors caused by differences in physician experience and assist in achieving precision medicine.^{5,6} In previous studies, ^{18}F -FDG PET/CT radiomics performed well in the diagnosis and prognosis prediction of lymphoma, offering direction for treatment plan development and prognosis evaluation.⁷⁻⁹ However, the results differ significantly because varying study methodologies were used. Consequently, this study aimed to systematically evaluate the diagnostic value of ^{18}F -FDG PET/CT radiomics for lymphoma, and our results can be used as reference for subsequent research.

Materials and Methods

Protocol and Registration

This study was conducted in accordance with the Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies guidelines (PRISMA-DTA).¹⁰ The study protocol was prospectively registered with PROSPERO (CRD42025640732) <https://www.crd.york.ac.uk/prosperto/>.

Literature Search

Computer searches were performed on PubMed, Web of Science, and the Cochrane Library databases to find relevant research on the diagnosis or differential diagnosis of lymphoma

based on ^{18}F -FDG PET/CT radiomics features. The search period was from January 1, 2010 to July 30, 2024. The literature search adopted a combination of subject words and free words, modifying its approach based on the characteristics of each database. The English search terms included PET/CT, radiomics, machine learning, lymphoma, and diagnosis. The search strategy was as follows: (“radiomics” OR “machine learning” OR “artificial intelligence” OR “deep learning” OR “texture analysis”) AND (lymphoma OR DLBCL) AND (PET/CT OR “positron emission tomography”) AND (diagnose OR diagnosis OR diagnosing OR diagnostic) AND (distinguish OR classification OR differentiation OR differential). Additionally, the references included in the research were simultaneously retrieved to supplement relevant information.

Literature Screening

Inclusion criteria: (1) Research on the diagnosis or differential diagnosis of lymphoma based on ^{18}F -FDG PET/CT radiomics; (2) diagnostic outcomes, including true positive (TP), false positive (FP), false negative (FN), and true negative (TN), can be acquired directly or indirectly from scholarly literature; and (3) researches classified as type 2a or above according to Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement.¹¹

Exclusion criteria: (1) Non-English literature; (2) the number of lymphoma cases or lesions included in the literature is less than 10; (3) literature lacking sufficient data or unable to extract data; and (4) studies in the form of letters, conference abstracts, editorials, reviews, case reports and duplicates.

Two researchers conducted independent literature screening, data extractions, and cross-checks. For any discrepancies, they were worked out through discussion or consultation with a senior physician. When screening literature, researchers read the article title first and then read the abstract and entire article to decide whether to include them after removing material that was blatantly unrelated.

Data Extraction

Data extraction was performed systematically using a data extraction form for each full-text publication: (1) fundamental study information, such as first author, publication year, diagnostic subjects, and sample size. (2) Radiomics analysis: region of interest (ROI) used for feature extraction, machine learning models, and texture features with the best diagnostic performance. If the research subjects were split into training and validation sets, only the validation set was included as the outcome indicator to maintain data integrity, reduce bias, and provide a

more accurate assessment of model performance.^{12,13} When more than one model was employed concurrently in a study, only the model's outcome metric with the highest area under the curve (AUC) was selected. If the AUC was not provided, the model with the highest Youden index was chosen. For research convenience, the same literature could be split into several separate studies for data extraction if it simultaneously separated lymphoma from numerous other lesions. (3) Directly obtained or indirectly calculated outcome measures: TP, FP, FN, and TN.¹⁴

Quality Assessment

The quality of the literature included in this study was assessed independently by two academics, who also cross-checked the findings. The risk of bias and concerns regarding applicability questions of all included studies were assessed with the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2).¹⁵ This tool includes four domains in terms of risk of bias (patient selection, index test, reference standard, as well as flow and timing) and three domains in terms of applicability concerns (patient selection, index test, and reference standard). Each domain of risk of bias and concerns about applicability is rated as low, high, or unclear.

Additionally, the radiomics quality score (RQS)⁶ and METHodological RadiomICS Score (METRICS),¹⁶ instruments developed specifically for radiomics studies, were applied to assess the methodological quality of the included studies. The RQS comprises 16 essential components, with each research allocated a certain number of points per component, yielding a total score ranging from -8 to 36, where -8 to 0 is defined as 0% and 36 is defined as 100%. The METRICS is a checklist-based framework that includes 30 items over 9 categories, and each item is assigned a different weight based on the completeness of reporting and methodological rigor (<https://metricsscore.github.io/metrics/METRICS.html>).

Statistical Analysis

This meta-analysis was conducted by using RevMan 5.4.1, R 4.4.0, and Stata 17.0 software. Heterogeneity among the studies was evaluated using Cochran's Q test. A random effects model was used if $P < .05$ and $I^2 > 50\%$, which indicated significant heterogeneity; otherwise, a fixed effects model was adopted. The diagnostic value of ¹⁸F-FDG PET/CT radiomics in lymphoma was evaluated by integrating the summary receiver operating characteristic (SROC) curve, and Spearman correlation analysis was conducted to determine whether heterogeneity caused by threshold effects exists. If there was no threshold effect, we used TP, FP, FN, and TN equivalent strain to calculate the pooled sensitivity (SEN), specificity (SPE), positive likelihood ratio (PLR), negative likelihood ratio (NLR), and diagnostic odds ratio (DOR), and the results were visualized through forest plots.

Subgroup analyses of various relevant covariates were performed to investigate the influencing factors of the diagnostic

value of ¹⁸F-FDG PET/CT radiomics. The covariates included the sample size, image modality, global or non-global ROI selection, manual or automatic ROI segmentation, radiomics mode, and deep learning (DL) or traditional machine learning (TML) algorithms. In addition, the publication bias and clinical utility was evaluated with Deeks' funnel plot and Fagan nomogram respectively. $P < .05$ was regarded as statistically significant.

Results

Literature Screening Process and Results

Our search strategy yielded a total of 1473 pertinent publications during the initial search. Following the removal of duplicate records, 415 papers were evaluated based on their titles and abstracts, leading to the deletion of 352 non-original researches or those irrelevant to this study. The remaining 63 full texts were independently evaluated by the same investigators. Among them, 43 studies were excluded due to data mismatch or inability to extract data. Ultimately, 20 studies were incorporated into meta-analysis after progressive screening. The search and selection flowchart complying with the guidance of PRISMA-DTA is shown in Figure 1.

Characteristics of the Included Studies

This research encompassed 20 studies that reported a cumulative total of 19 087 samples. All included studies were designed retrospectively; four were multicentric, while the remainder were monocentric. Nine studies were performed in China, seven in Europe, two in Japan, one in America, and one in Egypt. Various methodologies were employed in ROI segmentation; eight studies implemented a threshold based on the maximum standardized uptake value (SUVmax), two studies applied a threshold value of CT density, two studies utilized maximum intensity projection (MIP) images, seven studies designated all malignant lesions within a specific organ as ROI, and one study identified five lesions with the highest SUVmax. The predominant feature reduction algorithm was the least absolute shrinkage and selection operator (LASSO; $n = 7$). Logistic regression (LR) was the most frequently used approach for developing radiomics signatures ($n = 8$). The basic characteristics of the included studies were shown in Table 1.

Risk of Bias Assessment

The quality evaluation findings of the included literature by QUADAS-2 are shown in Figure 2. In the patient selection domain, we noted unclear risks of bias and unclear applicability concerns for two studies because the inclusion/exclusion criteria were not specified. For the index test, there was an unclear risk of bias in four studies because cross-validation methods were not used for modeling. However, no applicability issue was found because they matched research questions. Two studies in the reference standard area were found to have a high risk

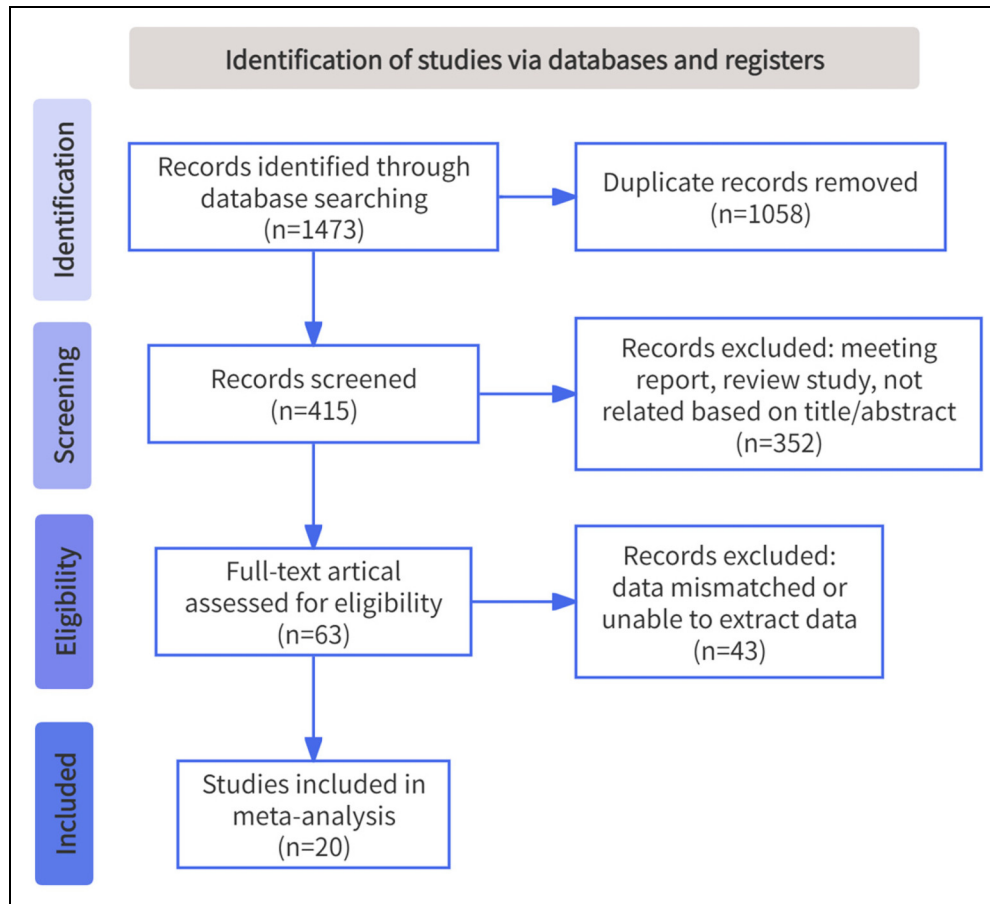


Figure 1. Flow Diagram Shows the Literature Screening Process and Results.

of bias because they obtained their specimens through biopsy. Finally, we observed an unclear risk of bias in five studies because they did not mention the interval between imaging acquisition and histopathological examination. Overall, the quality of the included literature was high.

RQS and METRICS

The single score of each RQS component and the total score of the whole study were provided in Table 2. The included studies achieved an average RQS of 13 (ranging from 10 to 17), accounting for 36.1% of the total points. Among the six domains with 16 items of RQS, there are four items, such as phantom studies on all scanners, biological correlates discussion, prospective design, and cost-effectiveness analysis, were not performed in the 20 included studies, and all the studies achieved 0 points in these items. Feature reduction, cutoff analyses, and discrimination statistics were performed in every included study.

The total METRICS score and quality category for each study were shown in Table 3. The average METRICS score across the 20 studies is 69.3% (ranging from 54.8% to 80.9%). One study is classified as excellent, sixteen as good, and three as moderate in quality category. Among numerous

significant weight items, all studies used eligibility criteria to describe a representative study population and high-quality reference standard with a clear definition; fifteen studies implemented image preprocessing techniques with transparent description and proper data partitioning process, yet only four studies tested the model with independent data from other institution.

Meta-Analysis

Diagnostic Performance of ^{18}F -FDG PET/CT Radiomics in Lymphoma. The pooled SEN, SPE, PLR, NLR, and DOR with 95% CI were 0.82 (0.78, 0.88), 0.83 (0.76, 0.87), 4.7 (3.4, 6.6), 0.20 (0.15, 0.28), and 23 (13, 42), respectively. The coupled forest plot showing SEN, SPE, and heterogeneity indicators (I^2 and Cochran's Q test) is depicted in Figure 3. The SROC curve with pooled AUC value of 0.90 (0.87, 0.92) is depicted in Figure 4, which suggested high diagnostic value of ^{18}F -FDG PET/CT radiomics in lymphoma.

Heterogeneity Assessment. The Higgins' I^2 value of Cochran's Q test with 95% CI was 83.7% (76.0, 88.9), and $P < .01$ showed significant heterogeneity among the included studies. Spearman's correlation coefficient (r) was 0.116 ($P = .423$),

Table 1. Characteristics of the Included Studies.

Study ID	Diagnostic Subject	Institutional Setting	Sample Size	Image Modality	ROI	Best Model/ Features	TP	FP	FN	TN
Abenavoli 2023 ¹⁷	mediastinal bulky disease due to cHL, PMBCL and GZL	Monocentric	117	PET	absolute threshold SUVmax ≥ 2.5	RF	30	30	5	52
Aide 2017 ¹⁸	BMI in DLBCL	Monocentric	82	PET	CT densities of Hu > 150	SkewnessH	27	9	6	40
Albano 2024 ¹⁹	RT in CLL	Monocentric	157	PET	41% from attenuation corrected PET images	L-L SUV R	97	6	40	14
Aoki 2023 ²⁰	sarcoidosis/ML	Multicentric	118	PET	reshaped frontal and lateral MIP images	CNN	50	6	6	56
Chen 2023 ²¹	breast DLBCL/breast IDC	Multicentric	324	PET + CT	breast nodules	AACNN	24	6	15	30
Cui 2023 ²²	PCNSL/Metastases	Monocentric	126	PET	intracranial tumors	RF	57	5	6	58
Domenico 2024 ²³	DLBCL/MALT	Monocentric	91	PET	41% from attenuation corrected PET images	TLG	49	8	5	29
Faudemer 2021 ²⁴	BMI in FL	Multicentric	66	PET	C3 vertebra to the upper third of femurs	PET pred.score	21	14	9	22
Hägström 2024 ²⁵	hypermetabolic tumour	Multicentric	16 583	PET + CT	reshaped frontal and lateral MIP images	LARS-avg	483	22	71	424
Jesus 2021 ²⁶	DLBCL/FL	Monocentric	348	PET + CT	five lesions with the highest SUVmax	Gradient Boosting	131	26	43	148
Katsuya 2021 ²⁷	SCC/NHL	Monocentric	52	PET + CT	a threshold of 40% of the SUVmax	LGZE	15	3	12	22
Kenawy 2020 ²⁸	BMI in lymphoma	Monocentric	44	PET	absolute threshold SUVmax ≥ 3	LRE	10	4	8	22
Kong 2019 ²⁹	PCNSL/GBM	Monocentric	77	PET	intracranial tumors	GLRLM	51	0	2	24
Li 2019 ³⁰	BMI in relapsed AL	Monocentric	41	PET + CT	CT densities of Hu > 130	LIME	3	0	1	2
Lovinfosse 2022 ⁸	SCC/lymphoma	Monocentric	420	PET	absolute threshold SUVmax ≥ 3	RF	188	32	12	187
Ou 2019 ³¹	breast carcinoma/ lymphoma	Monocentric	67	PET + CT	breast nodules	PET + CT pred.score	20	15	4	28
Sun 2021 ³²	gastric cancer/ lymphoma	Monocentric	79	PET	gastric lesions	Inertia_GLCM	35	12	10	22
Wang 2024 ³³	pancreatic carcinoma/ lymphoma	Monocentric	86	PET + CT	cystic and necrotic regions within the tumor	LR	63	7	6	10
Xu 2019 ³⁴	HCC/HL	Monocentric	99	PET + CT	40% threshold of SUVmax in liver lesions	Combined model	69	5	7	18
Yang 2024 ³⁵	T-Cell lymphoma in Children	Monocentric	110	PET + CT	absolute threshold SUVmax ≥ 4	LR	42	9	9	50

Abbreviations: cHL, classical Hodgkin lymphoma; PMBCL, primary mediastinal B cell lymphoma; GZL, grey zone lymphoma; RF, random forest; DLBCL, diffuse large B cell lymphoma; BMI, bone marrow involvement; Hu, Hounsfield unit; RT, Richter transformation; CLL, chronic lymphocytic leukemia; L-L SUV R, lesion to liver SUVmax ratio; ML, malignant lymphoma; CNN, convolutional neural network; IDC, invasive ductal carcinoma; PCNSL, primary central nervous system lymphoma; TLG, total lesion glycolysis; MALT, mucosa-associated lymphoid tissue; FL, follicular lymphoma; LARS, lymphoma artificial reader system; SCC, squamous cell carcinoma; NHL, non-Hodgkin's lymphoma; LGZE, low gray-level zone emphasis; LRE, long-run emphasis; GBM, glioblastoma multiforme; GLRLM, gray-level run-length matrix; AL, acute leukemia; LIME, local interpretable model-agnostic explanation; GLCM, grey-level co-occurrence matrix; LR, logistic regression; HCC, hepatocellular carcinoma; HL, hepatic lymphoma.

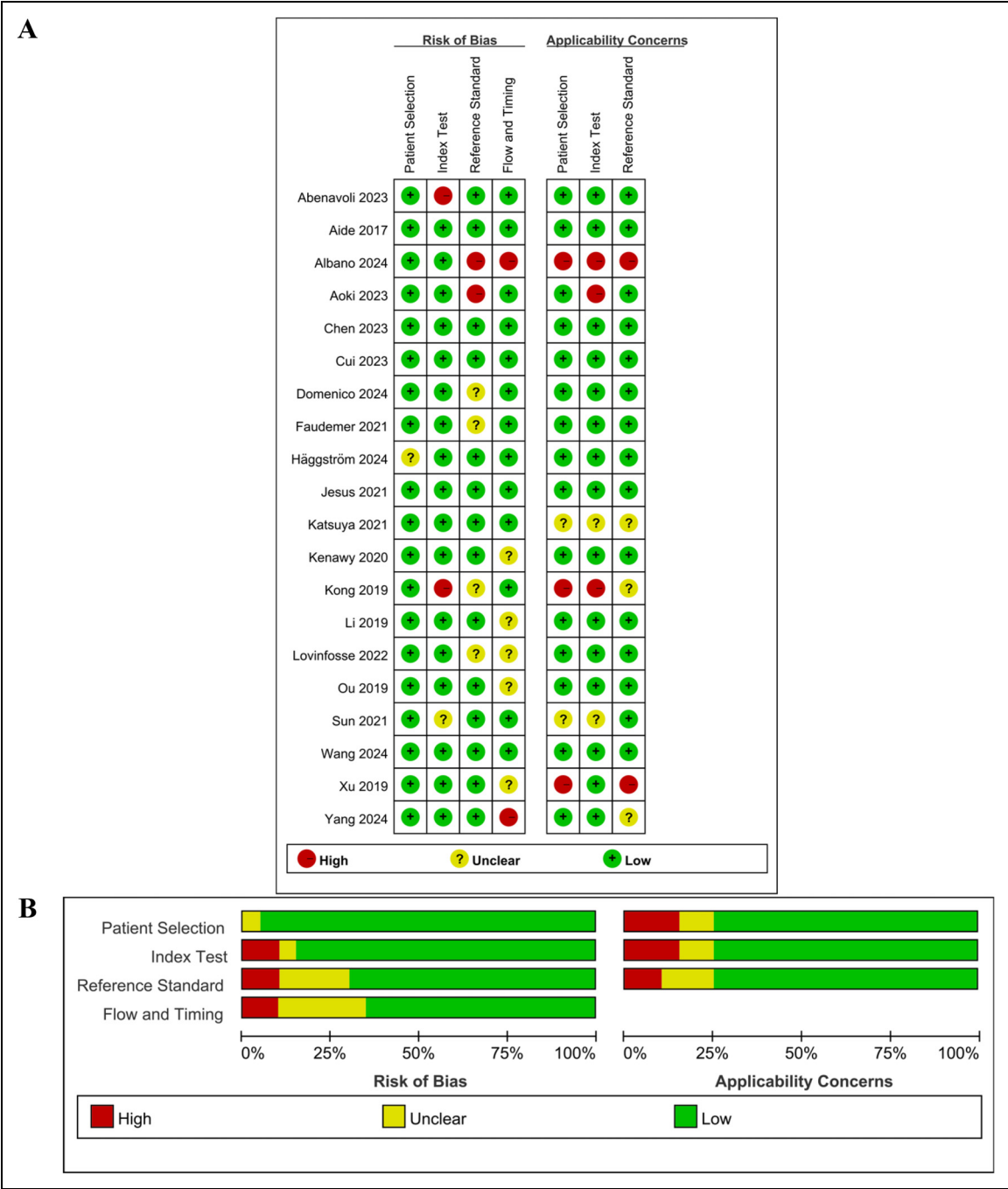


Figure 2. Stacked Bar Charts of the QUADAS-2 Scale of Methodological Quality Evaluation. Risk of Bias and Applicability Concerns of Each Included Study. (A) Individual Studies and (B) Summary. For Each Quality Domain, the Proportions of the Included Studies that Suggested Low, High, or Unclear Risk of Bias and Applicability Concerns are Displayed in Green, Red, and Yellow, Respectively.

suggesting the absence of a threshold effect as the possible cause of heterogeneity.

Meta-Regression and Subgroup Analyses. Considerable heterogeneity existed among the studies (overall I^2 83.7% [76.0, 88.9]; $P < .01$). Similarly, the forest plots indicated high heterogeneity with $I^2 > 50\%$ for SEN (I^2 82.1%; 95% CI: 75.0, 89.2; $P < .001$) and SPE (I^2 84.5%; 95% CI: 78.6, 90.4; $P < .001$).

To identify the sources of heterogeneity, we performed univariable meta-regression and subgroup analyses to investigate the impact of sample size, radiomics features, imaging modality, ROI, and algorithms on the heterogeneity. The results were presented in Table 4. Fully automatic segmentation was defined as a process executed entirely by algorithms without any human intervention. In contrast, semi-automatic segmentation referred to approaches that required manual initialization, correction, or

Table 2. Radiomics Quality Score of the Included Studies.

Study ID	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	Total Score
Abenavoli 2023 ¹⁷	2	1	0	1	3	0	0	1	2	1	0	2	2	2	0	0	17
Aide 2017 ¹⁸	1	0	0	0	3	1	0	1	1	0	0	0	2	1	0	0	10
Albano 2024 ¹⁹	2	0	0	1	3	1	0	1	2	0	0	4	0	1	0	0	15
Aoki 2023 ²⁰	1	0	0	1	3	0	0	1	2	0	0	2	2	1	0	0	13
Chen 2023 ²¹	1	0	0	1	3	0	0	1	1	0	0	2	2	2	0	0	13
Cui 2023 ²²	1	0	0	1	3	0	0	1	2	1	0	2	2	1	0	0	14
Domenico 2024 ²³	2	1	0	1	3	0	0	1	2	0	0	0	0	1	0	0	11
Faudemer 2021 ²⁴	2	0	0	1	3	1	0	1	2	0	0	2	2	1	0	0	15
Häggström 2024 ²⁵	0	0	0	1	3	0	0	1	2	0	0	4	0	2	0	0	13
Jesus 2021 ²⁶	0	0	0	1	3	0	0	1	1	0	0	2	2	1	0	0	11
Katsuya 2021 ²⁷	1	0	0	0	3	0	0	1	2	0	0	2	2	1	0	0	12
Kenawy 2020 ²⁸	0	0	0	0	3	1	0	1	1	0	0	2	2	1	0	0	11
Kong 2019 ²⁹	1	0	0	0	3	0	0	1	1	0	0	2	2	1	0	0	11
Li 2019 ³⁰	1	0	0	0	3	0	0	1	2	0	0	2	2	2	0	0	13
Lovinfosse 2022 ⁸	1	0	0	1	3	0	0	1	2	0	0	2	2	1	0	0	13
Ou 2019 ³¹	1	0	0	1	3	1	0	1	2	0	0	2	2	1	0	1	15
Sun 2021 ³²	1	0	0	1	3	0	0	1	1	0	0	2	2	1	0	0	12
Wang 2024 ³³	1	1	0	1	3	0	0	1	2	0	0	4	2	2	0	0	17
Xu 2019 ³⁴	1	0	0	1	3	0	0	1	2	0	0	2	2	1	0	1	14
Yang 2024 ³⁵	2	0	0	1	3	0	0	1	2	1	0	2	2	2	0	1	17

#1, Image protocol quality; #2, Multiple segmentations; #3, Phantom study on all scanners; #4, Imaging at multiple time points; #5, Feature reduction or adjustment for multiple testing; #6, Multivariable analysis with non-radiomics features; #7, Detect and discuss biological correlates; #8, Cutoff analyses; #9, Discrimination statistics; #10, Calibration statistics; #11, Prospective study registered in a trial database; #12, Validation validation; #13, Comparison to “gold standard”; #14, Potential clinical utility; #15, Cost-effectiveness analysis; #16, Open science and data.

supervision during or after the segmentation process.^{36,37} The findings of univariable meta-regression and subgroup analyses revealed that sample size and ROI selection considerably impacted the heterogeneity in SEN ($P < .01$), while algorithms significantly influenced the heterogeneity in SPE ($P < .01$); other assessed parameters showed no statistical significance regarding overall heterogeneity ($P > .05$).

In terms of sample size, studies with 100 participants or more ($n = 9$) had a higher SEN (83%; 95% CI: 76, 89) and SPE (85%; 95% CI: 78, 91) than studies with fewer than 100 participants ($n = 11$; SEN, 82% [95% CI: 72, 89]; SPE, 74% [95% CI: 67, 81]). Regarding different imaging modalities, PET alone ($n = 11$) had a higher SEN (84%; 95% CI: 80, 86) than PET combined with CT ($n = 9$; 80%; 95% CI: 71, 87), but SPE (80%; 95% CI: 72, 86 vs 83%; 95% CI: 73, 90) showed the opposite trend.

Regardless of whether ROI was segmented in the global images, non-global ROI ($n = 12$) had higher SEN (84%; 95% CI: 77, 90) and SPE (80%; 95% CI: 74, 85) than global ROI ($n = 8$; SEN, 81% [95% CI: 75, 86]; SPE, 79% [95% CI: 69, 87]). In terms of ROI segmentation, semi-automatic segmentation ($n = 12$) had the highest SEN (83%; 95% CI: 75, 89) than manual segmentation ($n = 4$; 81%; 95% CI: 69, 89) and fully automatic segmentation ($n = 4$; 78%; 95% CI: 69, 85), while fully automatic segmentation had the highest SPE (84%; 95% CI: 82, 85) than manual segmentation (63%; 95% CI: 55, 71) and semi-automatic segmentation (81%; 95% CI: 84, 86).

Regarding radiomics mode, texture features ($n = 5$) had the highest SEN (86%; 95% CI: 69, 97) than histogram ($n = 4$; 80%; 95% CI: 66, 88) and other features ($n = 11$; 83%; 95% CI: 77, 88), whereas other features had the highest SPE

(82%; 95% CI: 75, 87) than texture features (79%; 95% CI: 61, 90) and histogram (75%; 95% CI: 67, 82). In terms of algorithms, TML ($n = 16$) shown a superior SEN (83%; 95% CI: 76, 87) compared to DL ($n = 4$; 78%; 95% CI: 69, 85), although SPE (77%; 95% CI: 70, 82 vs 84%; 95% CI: 82, 85) demonstrated an inverse trend.

Publication Bias. According to Deeks' asymmetry test, quantitative analysis results showed no statistical significance among the included studies ($P = .14$), indicating that publication bias was not significant. Deeks' funnel plot is shown in Figure 5.

Clinical Utility. The Fagan nomogram in Figure 6 showed that the prior probability in our study was 20%. This result revealed the clinical utility of ¹⁸F-FDG PET/CT radiomics by quantifying its ability to significantly modify the probability of lymphoma diagnosis from a specified baseline pre-test estimate. A PLR of 5 elevated the post-test probability to 54% for a positive test result, with an absolute increase of 34%. Likewise, an NLR of 0.20 diminished the post-test probability to 5% for a negative test outcome. The notable alterations in posterior probabilities of positive and negative outcomes indicated that the ¹⁸F-FDG PET/CT radiomics model may function as an efficacious clinical instrument for lymphoma diagnosis.

Discussion

The current gold standard for lymphoma diagnosis is histopathological biopsy. Surgery or biopsy is invasive, and the overall lesion condition cannot be obtained. PET/CT

Table 3. METHodological RadiomiCs Score of the Included Studies.

Study ID	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17	#18	#19	#20	#21	#22	#23	#24	#25	#26	#27	#28	#29	#30	Total score	Quality category			
Abenavoli 2023 ¹⁷	N	Y	Y	N	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	N	78.70%	Good			
Aide 2017 ¹⁸	N	Y	Y	N	Y	Y	Y	N	N	Y	N	Y	Y	N	Y	Y	N	N	Y	Y	N	N	N	Y	N	N	Y	Y	N	N	N	54.80%	Moderate		
Albano 2024 ¹⁹	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	N	N	68.50%	Good	
Aoki 2023 ²⁰	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	N	N	77.20%	Good	
Chen 2023 ²¹	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	N	N	71.20%	Good	
Cui 2023 ²²	N	Y	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	N	N	75.80%	Good	
Domenico 2024 ²³	N	Y	Y	N	Y	Y	Y	N	N	Y	N	Y	N	Y	Y	Y	Y	Y	N	Y	Y	N	Y	N	N	N	Y	Y	Y	Y	N	N	59.30%	Moderate	
Faudemer 2021 ²⁴	N	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	74.50%	Good	
Häggström 2024 ²⁵	N	Y	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	Y	Y	Y	Y	Y	Y	N	N	80.90%	Excellent	
Jesus 2021 ²⁶	N	Y	Y	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	N	N	N	Y	Y	Y	N	Y	Y	Y	N	N	67.70%	Good	
Katsuya 2021 ²⁷	N	Y	Y	N	Y	Y	Y	N	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	N	Y	Y	Y	N	N	64.50%	Good	
Kenawy 2020 ²⁸	N	Y	Y	N	Y	Y	Y	N	Y	N	N	Y	Y	Y	Y	Y	N	Y	N	Y	N	N	N	Y	Y	Y	N	Y	Y	Y	N	N	57.50%	Moderate	
Kong 2019 ²⁹	N	Y	Y	N	Y	Y	Y	N	Y	Y	N	Y	Y	Y	Y	Y	N	Y	N	Y	N	N	N	Y	Y	Y	N	Y	Y	Y	N	N	64.80%	Good	
Li 2019 ³⁰	N	Y	Y	N	Y	Y	Y	N	Y	Y	N	Y	N	Y	Y	Y	N	Y	N	Y	N	N	N	Y	N	Y	N	Y	Y	Y	N	N	62.80%	Good	
Lovinfosse 2022 ⁸	N	Y	Y	N	Y	Y	N	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	N	Y	Y	Y	N	N	61.80%	Good	
Ou 2019 ³¹	N	Y	Y	N	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	N	N	N	Y	Y	Y	N	Y	Y	Y	N	N	68.60%	Good	
Sun 2021 ³²	N	Y	Y	N	Y	Y	Y	N	Y	N	Y	Y	N	Y	Y	Y	N	N	Y	Y	N	N	Y	Y	Y	Y	N	Y	Y	Y	N	N	67.70%	Good	
Wang 2024 ³³	N	Y	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N	Y	Y	Y	N	N	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	N	N	79.40%	Good	
Xu 2019 ³⁴	N	Y	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	N	Y	N	N	Y	Y	Y	Y	N	Y	Y	Y	N	N	71.50%	Good	
Yang 2024 ³⁵	N	Y	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	N	N	79.50%	Good

#1, Adherence to radiomics and/or machine learning-specific checklists or guidelines; #2, Eligibility criteria that describe a representative study population; #3, High-quality reference standard with a clear definition; #4, Multi-center; #5, Clinical translatability of the imaging data source for radiomics analysis; #6, Imaging protocol with acquisition parameters; #7, The interval between imaging used and reference standard; #8, Transparent description of segmentation methodology; #9, Formal evaluation of fully automated segmentation; #10, Test set segmentation masks produced by a single reader or automated tool; #11, Appropriate use of image preprocessing techniques with transparent description; #12, Use of standardized feature extraction software; #13, Transparent reporting of feature extraction parameters, otherwise providing a default configuration statement; #14, Removal of non-robust features; #15, Removal of redundant features; #16, Appropriateness of dimensionality compared to data size; #17, Robustness assessment of end-to-end deep learning pipelines; #18, Proper data partitioning process; #19, Handling of confounding factors; #20, Use of appropriate performance evaluation metrics for task; #21, Consideration of uncertainty; #22, Calibration assessment; #23, Use of uni-parametric imaging or proof of its inferiority; #24, Comparison with a non-radiomics approach or proof of added clinical value; #25, Comparison with simple or classical statistical models; #26, Internal testing; #27, External testing; #28, Data availability; #29, Code availability; #30, Model availability; Y, Yes; N, No.

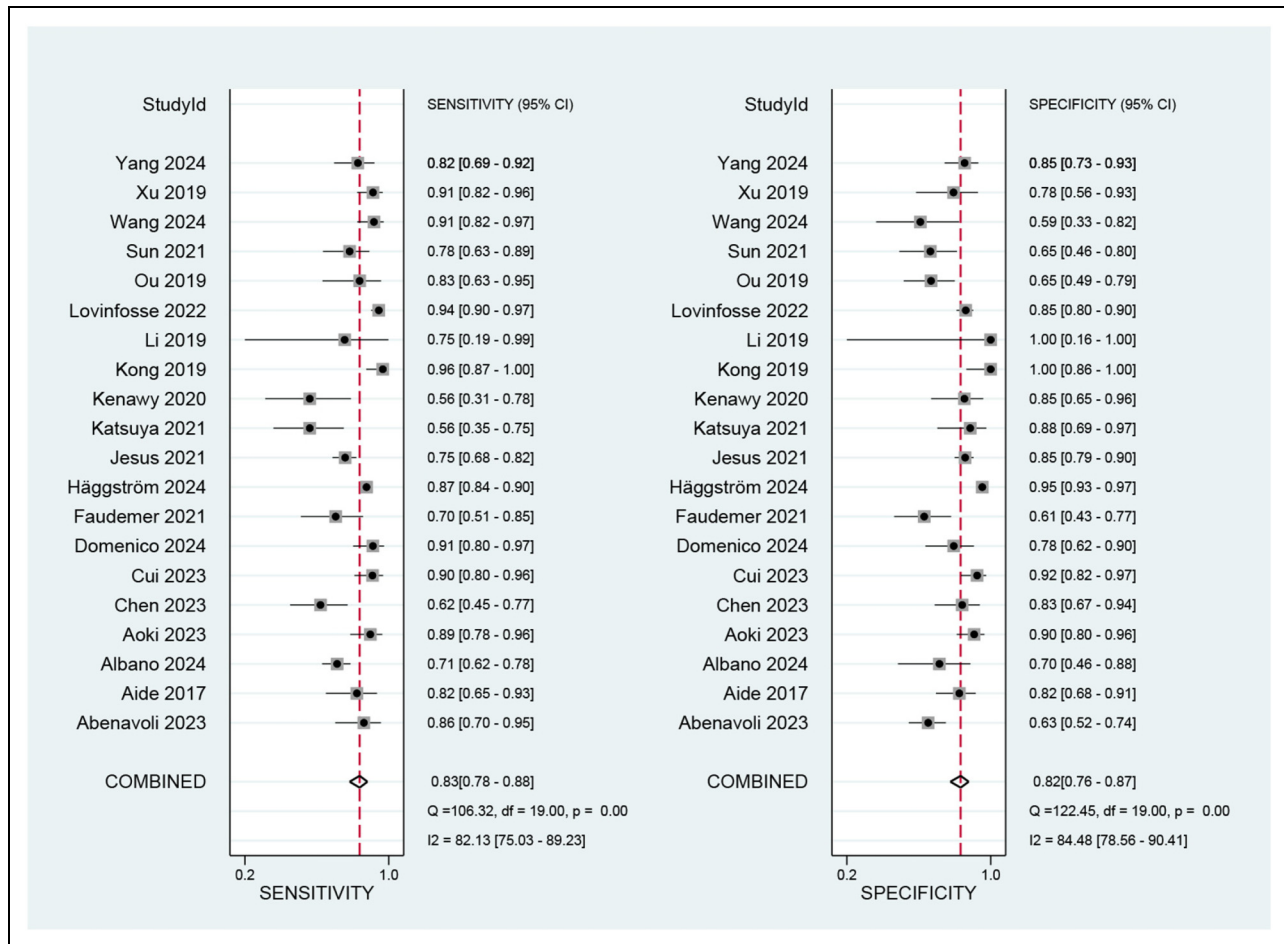


Figure 3. Forest Plots Show the Performance Estimates (SEN and SPE) of Each Study Based on ^{18}F -FDG PET/CT Radiomics in Lymphoma Diagnosis. the Pooled SEN and SPE are Displayed as Vertical Lines in the Forest Plots. Significant Variability in the Diagnostic Parameters Between Trials is Indicated by $I^2 > 50\%$.

examination can reflect overall lesion information and avoid unnecessary surgery. Since 2014, international experts have reached a consensus on the core role of PET/CT in the diagnosis and prognosis assessment of lymphoma.³ In recent years, with the development of medical imaging technology, the application of ^{18}F -FDG PET/CT radiomics in lymphoma diagnosis and prognosis prediction has become increasingly widespread.³⁸⁻⁴⁰ This present systematic review and meta-analysis investigated the diagnostic value of ^{18}F -FDG PET/CT radiomics models in lymphoma and evaluated the quality and risk of bias of included studies utilizing QUADAS-2, RQS and METRICS tools for radiomics studies. For the first time, our meta-analysis amalgamated and statistically analyzed data from distinct independent studies, which may potentially offer vital clues for its clinical utility and future research. Twenty studies were included in the systematic review and meta-analysis. The results showed that the pooled SEN and SPE with 95% CI of ^{18}F -FDG PET/CT radiomics in diagnosing lymphoma were 0.82 (0.78, 0.88) and 0.83 (0.76, 0.87), respectively. The AUC with 95% CI of the SROC curve was 0.90 (0.87, 0.92), indicating that ^{18}F -FDG PET/CT radiomics has

high diagnostic efficiency for lymphoma. Deeks' funnel plot showed no publication bias, indicating that the results of this study were reliable.

Upon observing the high heterogeneity among studies (overall I^2 83.7% [76.0, 88.9]; $P < .01$), we observed $I^2 > 50\%$ for SEN (I^2 82.1%; 95% CI: 75.0, 89.2; $P < .001$) and SPE (I^2 84.5%; 95% CI: 78.6, 90.4; $P < .001$). We conducted subgroup analysis on the potential parameters contributing to high heterogeneity, and the results suggested that sample size, ROI and algorithms were the primary contributors of heterogeneity. When the sample size was greater than 100, the diagnostic performance of the model was considerable, with a combined SEN and SPE of 83% (95% CI: 76, 89) and 85% (95% CI: 78, 91), respectively. This result may be due to the Neg correlation between sample deviation and sample size,⁴¹ which further indicated that ^{18}F -FDG PET/CT radiomics is a scalable and reliable diagnostic tool for lymphoma. In addition, four studies^{20,21,25,26} have adopted DL models. The sample sizes of these studies are all greater than 100 or even greater than 1000. DL radiomics models require high-quality datasets to ensure accuracy and generalization. These datasets should be

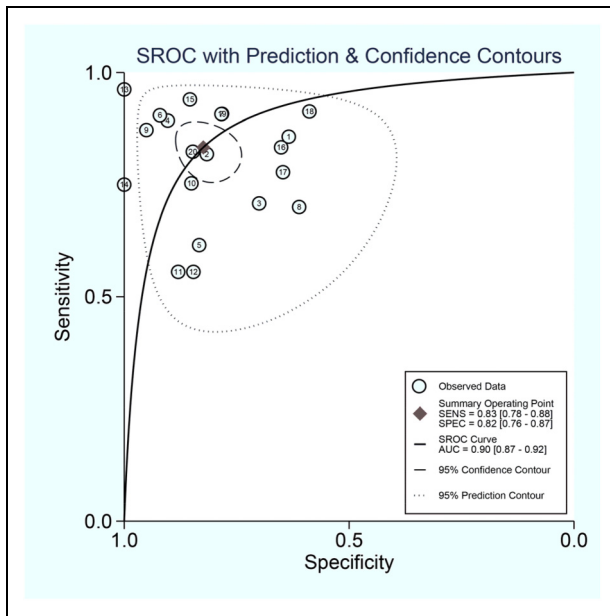


Figure 4. SROC Plots of ^{18}F -FDG PET/CT Radiomics in Lymphoma Diagnosis. Each Circle Indicates One Included Study. Values in Brackets are 95% CIs.

large and diverse to capture the variability in medical imaging. However, acquiring such datasets is challenging due to factors like imaging costs, annotation efforts, and patient privacy concerns.^{42,43} To address these issues, data augmentation techniques, such as geometric transformations and generative adversarial networks (GANs), are employed to artificially expand dataset size and diversity.⁴⁴ Additionally, rigorous data preprocessing, including image acquisition standardization and feature extraction, is essential to minimize variability and enhance model robustness. The dependability and usability of DL models in clinical contexts are further guaranteed by data standardization, the creation of strong models, and thorough statistical analysis.⁴⁵

ROI is the most critical and controversial part of radiomics research, and the accuracy of data analysis and the reproducibility of features are closely related to the selection of ROI. One of the primary sources of heterogeneity arises from variations in ROI delineation methods, which can be manual, semi-automatic, or fully automatic.^{46,47} Although manual segmentation is widely employed, it is very dependent on the radiologist's skill and experience, which can cause variability both within and between observers. Additionally, the defining of ROI is difficult since many tumors have hazy or diffuse borders. The performance of the model can be greatly affected by variations in feature extraction, which can result from different segmentation techniques, such as concentrating only on the tumor core or incorporating peritumoral regions.⁴⁸ The subgroup analyses of our study indicated that ROI Segmentation has no statistically effect on heterogeneity ($P > .05$), while ROI Selection is one of the primary factors influencing it ($P < .01$). In recent years, semi-automatic and DL-based segmentation methods have been proposed to improve reproducibility, yet they still

require careful validation to ensure consistency across different datasets and imaging modalities.^{49,50} Furthermore, variations in ROI selection criteria, such as tumor size, shape, and intensity thresholds, contribute to dataset heterogeneity. Previous studies have found that semi-automatic segmentation approaches based on 3D Slicer software can significantly improve the accuracy and reproducibility of radiomics feature quantification, with low heterogeneity.⁵¹ Therefore, semi-automatic segmentation is a potential method for constructing strong robust radiomics models. Standardizing ROI segmentation through harmonized protocols, automated algorithms, and quality control measures is essential to minimize variability. Robust cross-validation and external validation using independent datasets are crucial steps to assess the impact of ROI heterogeneity on radiomics models and improve their clinical applicability.⁵² By adhering to these stringent data requirements, PET/CT-based radiomics models can be effectively developed and translated into clinical practice.

Meanwhile, all 20 included studies utilized machine learning models. Machine learning typically divides datasets into training and validation sets, and it improves our decision-making process by classifying, predicting, and evaluating known datasets. However, there were two machine learning models included in the literature that have not been validated on other datasets, which may lead to overfitting of the models.^{26,53} This is a primary factor contributing to the considerable heterogeneity shown in this work. In radiomics research, TML models typically rely on handcrafted features extracted from medical images, necessitating domain expertise for feature selection and engineering. In contrast, DL models automatically learn hierarchical features directly from raw image data through neural networks, reducing the need for manual intervention. Comparative study indicated that DL models often outperform traditional radiomics approaches, achieving higher diagnostic accuracy in various clinical scenarios.^{54,55} There are four studies^{20,21,25,26} used DL models in this meta-analysis, all of which have diagnostic accuracy above 85%. Specifically, a DL-based diagnostic platform²⁵ for DLBCL achieved an accuracy of 95%, highlighting its potential in clinical applications. These findings suggest that DL models may offer significant advantages over TML models in radiomics, particularly in enhancing diagnostic accuracy for lymphoma. Our meta-analysis revealed that the SPE of DL was markedly superior to that of TML, which serves as a source for examining heterogeneity ($P < .01$). This finding is consistent with the results of previous research. However, DL requires a larger amount of data compared with traditional radiomics approaches, and the generalization ability of models on independent datasets is still lacking.^{56,57} Therefore, the research on DL is still far from complete.

Multiple studies have shown that radiomics has promising prospects in diagnosing highly heterogeneous tumors like those with the same disease but different images and those with different diseases but the same image. Compared with conventional diagnostic methods, radiomics has stronger diagnostic capabilities, and further consideration should be given to standardize its application

Table 4. Univariable meta-Regression and Subgroup Analyses.

Parameter	Category	Number of Studies	SEN (%)	P_1	SPE (%)	P_2
Sample Size	<100	11	82 (72,89)	<0.01	74 (67,81)	0.14
	≥100	9	83 (76,89)		85 (78,91)	
Image Modality	PET	11	84 (80,86)	0.30	80 (72,86)	1.70
	PET + CT	9	80 (71,87)		83 (73,90)	
ROI Selection	global	8	81 (75,86)	<0.01	79 (69,87)	0.57
	non-global	12	84 (77,90)		80 (74,85)	
ROI Segmentation	manual	4	81 (69,89)	5.78	63 (55,71)	1.83
	semi-automatic	12	83 (75,89)		81 (74,86)	
	fully automatic	4	78 (69,85)		84 (82,85)	
Radiomics Mode	texture features	5	86 (69,97)	6.47	79 (61,90)	5.51
	histogram	4	80 (66,88)		75 (67,82)	
	other features	11	83 (77,88)		82 (75,87)	
Algorithms	DL	4	78 (69,85)	9.23	84 (82,85)	<0.01
	TML	16	83 (76,87)		77 (70,82)	

Abbreviations: SEN, sensitivity; SPE, specificity; ROI, region of interest; DL, deep learning; TML, traditional machine learning.

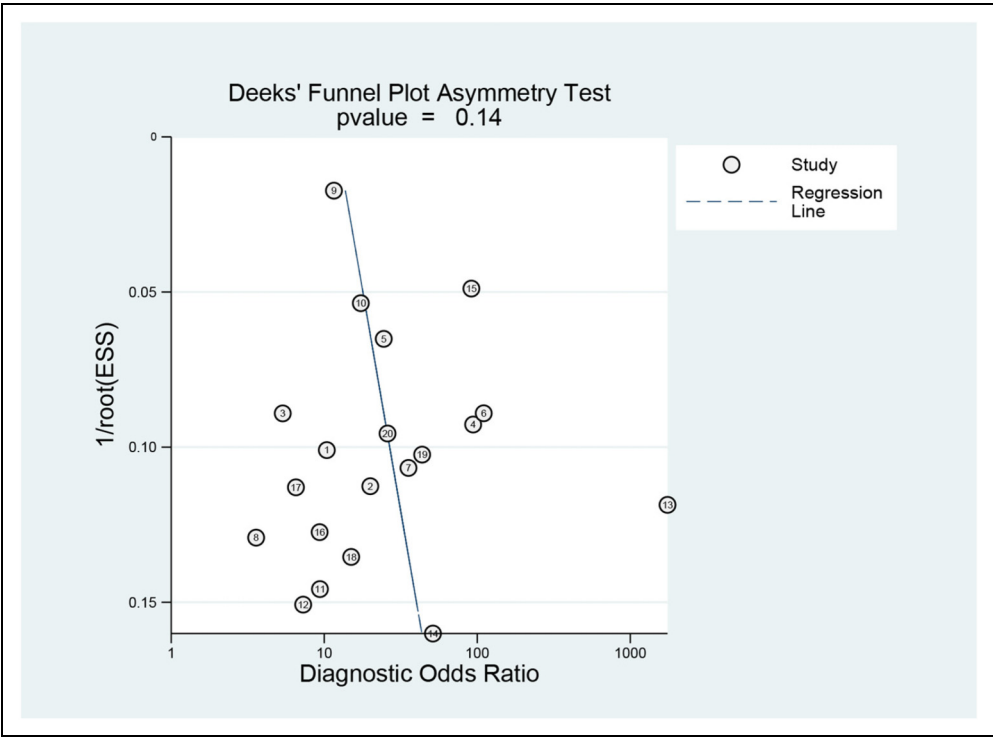


Figure 5. Deeks' Funnel Plot Shows no Asymmetry and the Presence of Publication Bias. Numbers in Circles Refer to the Study ID. ESS, Effective Sample Size.

in clinical practice.^{8,20–23,26,27,29,31–34,58} ¹⁸F-FDG metabolic parameters and texture features are the most commonly used features for constructing lymphoma diagnostic models. They indicate tumor heterogeneity by reflecting the metabolic differences of tumor cells and the local spatial arrangement of voxel intensity.^{47,59} However, the reproducibility of radiomics features is poor due to different implementation plans of various studies, so seeking universal ¹⁸F-FDG PET/CT radiomics features to reflect the biological characteristics of tumors is challenging. The subgroup analysis of this study showed that the use of

different radiomics methods to construct models had no statistical significance on heterogeneity ($P > .05$), suggesting that the variability in model performance was probably affected by extraneous factors, such as datasets or patient characteristics, rather than the radiomics methodology itself. This finding supports the robustness of radiomics-based models across different feature extraction or analysis techniques.

In clinical practice, the successful implementation of ¹⁸F-FDG PET/CT radiomics in lymphoma diagnosis necessitates strict adherence to standardized image acquisition

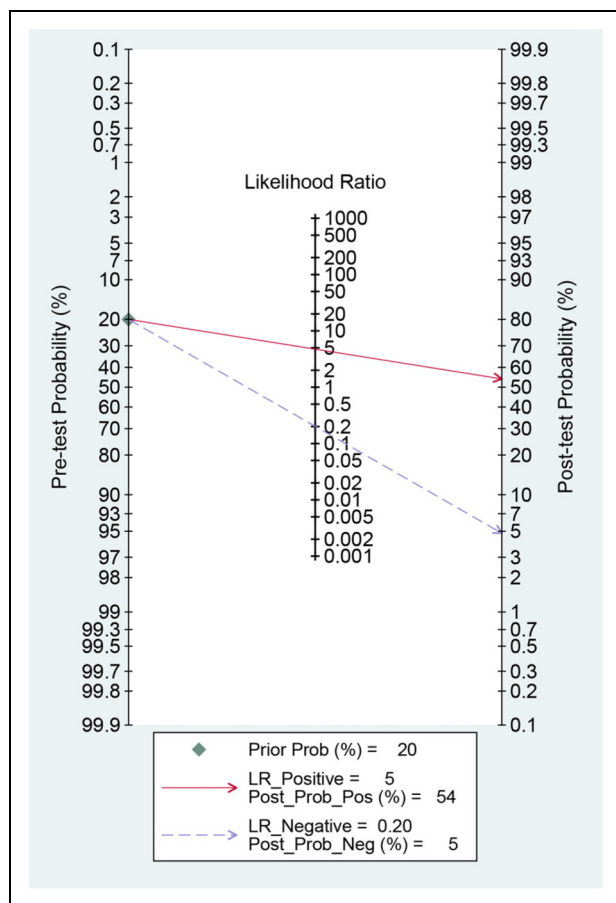


Figure 6. Fagan Nomogram of ^{18}F -FDG PET/CT Radiomics in Lymphoma Diagnosis. LR, Likelihood Ratio.

protocols to ensure data consistency and reproducibility. This entails the use of uniform scanning parameters, radiotracer injection doses, and acquisition timing to minimize artifacts and individual variability. Additionally, image preprocessing techniques such as normalization, filtering, and segmentation should follow established methodologies to optimize image quality and enhance the reliability of radiomics feature extraction.^{38,60,61} For robust feature extraction, validated computational algorithms should be employed to quantify key radiomics features, such as texture, shape, and metabolic parameters. The integration of TML or DL algorithms can further facilitate automated analysis, improving diagnostic accuracy and efficiency.^{56,62,63} Multicenter collaborations play a critical role in constructing large-scale, diverse datasets and enabling the integration of multimodal data, such as radiomics with genomic or biomarker information, to enhance model generalizability.^{53,64} Moreover, the development of open-access databases and data-sharing platforms can foster research transparency and reproducibility, thereby accelerating the clinical adoption of radiomics in lymphoma diagnostics.^{56,65,66} From a clinical integration perspective, radiomics analysis should be incorporated into existing diagnostic workflows, complementing traditional image interpretation and pathological

assessments.⁶⁷ Developing clinical decision support systems can aid physicians in optimizing treatment strategies, such as precise staging, therapy response prediction, and relapse monitoring.^{46,68,69} By addressing these challenges, ^{18}F -FDG PET/CT radiomics can be effectively translated into clinical practice, ultimately improving diagnostic accuracy and patient outcomes in lymphoma.

Limitations

There are several limitations of this study that warrant consideration. First, all included studies were retrospective, potentially impacting the generalizability of our findings. Compared with prospective studies, retrospective designs are susceptible to unquantifiable confounders, such as discrepancies in image acquisition protocols, variances in patient management, or undocumented clinical variables. Additionally, imbalanced datasets, such as uneven representation of lymphoma subtypes or stages, may impair model training and validation in radiomics studies, hence limiting the applicability of findings in a wider range of clinical settings. Future studies should more adopt prospective design to improve research quality. Second, as a result of incomplete data or poor research implementation, some original studies were excluded from this meta-analysis, which may lead to selection bias and affect the accuracy of conclusions. Lastly, the main radiomics features reported across studies showed notable variability and limited reproducibility, likely due to differences in segmentation methods, image preprocessing (such as resampling and discretization), feature extraction, and model construction. To overcome the limitations of low repeatability, researchers may adhere to established protocols for radiomics study,^{69,70} integrate the quality evaluation framework of radiomics,⁶ and conduct radiomics research with standardization.

Conclusion

In conclusion, this meta-analysis of 20 studies based on ^{18}F -FDG PET/CT radiomics demonstrated promising diagnostic performance in lymphoma, indicating the potential clinical utility of radiomics models. However, there is high heterogeneity among different studies. Future clinical practicality needs to be confirmed by large-scale, prospective, and externally validated studies with harmonized imaging protocols and radiomics methodologies.

Abbreviations

^{18}F -FDG	^{18}F -fluorodeoxyglucose
AL	acute leukemia
AUC	area under the curve
BMI	bone marrow involvement
CI	confidence interval
CLL	chronic lymphocytic leukemia
CNN	convolutional neural network
DL	deep learning

DLBCL	diffuse large B cell lymphoma
DOR	diagnostic odds ratio
FL	follicular lymphoma
FN	false negative
FP	false positive
GANs	generative adversarial networks
GBM	glioblastoma multiforme
GLCM	grey-level co-occurrence matrix
LR	logistic regression
GLRLM	gray-level run-length matrix
GZL	grey zone lymphoma
HCC	hepatocellular carcinoma
HL	hepatic lymphoma
IDC	invasive ductal carcinoma
LARS	lymphoma artificial reader system
LASSO	least absolute shrinkage and selection operator
LGZE	low gray-level zone emphasis
LIME	local interpretable model-agnostic explanation
LR	logistic regression
LRE	long-run emphasis
MALT	mucosa-associated lymphoid tissue
METRICS	METHodological Radiomics Score
MIP	maximum intensity projection
ML	malignant lymphoma
NHL	non-Hodgkin's lymphoma
NLR	negative likelihood ratio
PCNSL	primary central nervous system lymphoma
PLR	positive likelihood ratio
PMBCL	primary mediastinal B cell lymphoma
QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies 2
RF	random forest
ROI	region of interest
RQS	radiomics quality score
RT	Richter transformation
SCC	squamous cell carcinoma
SEN	sensitivity
SPE	specificity
SROC	summary receiver operating characteristic
SUVmax	maximum standardized uptake value
TML	traditional machine learning
TLG	total lesion glycolysis
TN	true negative
TP	true positive
TRIPOD	Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis.

ORCID iDs

Chaoying Liu  <https://orcid.org/0009-0004-1872-5058>

Heng Zhang  <https://orcid.org/0000-0002-2019-0038>

Xinye Ni  <https://orcid.org/0000-0002-2402-9719>

Ethical Considerations

This study does not involve research on animals or humans; hence, no ethical approval was necessary for this meta-analysis and secondary research utilizing data from public sources. Consequently, ethical statements are not applicable.

Author Contributions

Chaoying Liu: Data curation; formal analysis; investigation; writing – original draft. Jun Zhao: Data curation; formal analysis; investigation. Heng Zhang: formal analysis; project administration; writing – review

and editing. Xinye Ni: formal analysis; project administration; writing – review and editing.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by the National Natural Science Foundation of China (No. 62371243), Jiangsu Provincial Key Research and Development Program Social Development Project (No. BE2022720), the National Natural Science Foundation of Jiangsu (No. BK20231190), Jiangsu Provincial Medical Key Discipline Cultivation Unit of Oncology Therapeutics (Radiotherapy) (No. JSDW202237), General Program of Jiangsu Provincial Health Commission (No. M2020006), Changzhou Social Development Program (No. CE20235063 and CJ20244020).

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

Data are available upon reasonable request.

References

1. Miller KD, Nogueira L, Devasia T, et al. Cancer treatment and survivorship statistics, 2022. *CA Cancer J Clin.* 2022;72(5): 409-436. DOI:10.3322/caac.21731
2. Alaggio R, Amador C, Anagnostopoulos I, et al. The 5th edition of the world health organization classification of haematolymphoid tumours: Lymphoid neoplasms. *Leukemia.* 2022;36(7): 1720-1748. DOI: 10.1038/s41375-022-01620-2
3. Barrington SF, Mikhael NG, Kostakoglu L, et al. Role of imaging in the staging and response assessment of lymphoma: Consensus of the international conference on malignant lymphomas imaging working group. *J Clin Oncol.* 2014;32(27):3048-3058. DOI: 10.1200/jco.2013.53.5229
4. Pepper NB, Oertel M, Rehn S, et al. Modern PET-guided radiotherapy planning and treatment for malignant lymphoma. *Semin Nucl Med.* 2023;53(3):389-399. DOI: 10.1053/j.semnuclmed.2022.09.001
5. Mayerhoefer ME, Materka A, Langs G, et al. Introduction to radiomics. *J Nucl Med.* 2020;61(4):488-495. DOI: 10.2967/jnumed.118.222893
6. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* 2017;14(12):749-762. DOI: 10.1038/nrclinonc.2017.141
7. Chong J, Yue T, Ang L, et al. Predictive value of multi-parameter model incorporating PET-based radiomics features for survival of older patients (≥ 60 years) with diffuse large B-cell lymphoma. *Clin J Nucl Med Mol Imaging.* 2023;43(05):257-262. DOI:10.3760/cma.j.cn321828-202111119-00408
8. Lovinfosse P, Ferreira M, Withofs N, et al. Distinction of lymphoma from sarcoidosis on ^{18}F -FDG PET/CT: Evaluation of radiomics-feature-guided machine learning versus human reader

- performance. *J Necl Med.* 2022;63(12):1933-1940. DOI: 10.2967/jnumed.121.263598
9. Li M, Yao H, Zhang P, et al. Development and validation of a [18F]FDG PET/CT-based radiomics nomogram to predict the prognostic risk of pretreatment diffuse large B cell lymphoma patients. *Eur Radiol.* 2022;33(5):3354-3365. DOI: 10.1007/s00330-022-09301-5
 10. McInnes MDF, Moher D, Thombs BD, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: The PRISMA-DTA statement. *JAMA.* 2018; 319(4):388-396. DOI: 10.1001/jama.2017.19163
 11. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Br Med J.* 2015; 122(3): 434-443. DOI: 10.1136/bmj.g7594
 12. Hou R, Lo JY, Marks JR, et al. Classification performance bias between training and test sets in a limited mammography dataset. *PLoS One.* 2024;19(2):e0282402. DOI: 10.1371/journal.pone.0282402
 13. Luo WA-O, Phung DA-O, Tran TA-O, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. *J Med Internet Res.* 2016;18(12):e323. DOI: 10.2196/jmir.5870
 14. Guesné SJJ, Hanser T, Werner S, et al. Mind your prevalence!. *J Cheminform.* 2024;16(1):43. DOI: 10.1186/s13321-024-00837-w
 15. Whiting PF, Rutjes AWS, QUADAS-2 Group, et al. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Inter Med.* 2011; 155(8): 529-536. DOI: 10.7326/0003-4819-155-8-201110180-00009.
 16. Kocak B, Akinci D'Antonoli T, Mercaldo N, et al. METHodological RadiomIcs score (METRICS): A quality scoring tool for radiomics research endorsed by EuSoMII. *Insights Imaging.* 2024;15(1):8. DOI: 10.1186/s13244-023-01572-w
 17. Abenavoli EM, Barbetti M, Linguanti F, et al. Characterization of mediastinal bulky lymphomas with FDG-PET-based radiomics and machine learning techniques. *Cancers (Basel).* 2023;15(7) 1931. DOI: 10.3390/cancers15071931
 18. Aide N, Talbot M, Fruchart C, et al. Diagnostic and prognostic value of baseline FDG PET/CT skeletal textural features in diffuse large B cell lymphoma. *Eur J Nucl Med Mol I.* 2017; 45(5):699-711. DOI: 10.1007/s00259-017-3899-6
 19. Albano D, Calabrò A, Dondi F, et al. 2-[18F]-FDG PET/CT semi-quantitative and radiomics predictive parameters of richter's transformation in CLL Patients. *Medicina (B Aires).* 2024; 60(2):203. DOI: 10.3390/medicina60020203
 20. Aoki H, Miyazaki Y, Anzai T, et al. Deep convolutional neural network for differentiating between sarcoidosis and lymphoma based on [18F]FDG maximum-intensity projection images. *Eur Radiol.* 2023;34(1):374-383. DOI: 10.1007/s00330-023-09937-x
 21. Chen W, Liu F, Wang R, et al. End-to-end deep learning radiomics: Development and validation of a novel attention-based aggregate convolutional neural network to distinguish breast diffuse large B-cell lymphoma from breast invasive ductal carcinoma. *Quant Imaging Med Surg.* 2023;13(10):6598-6614. DOI: 10.21037/qims-22-1333
 22. Cui C, Yao X, Xu L, et al. Improving the classification of PCNSL and brain metastases by developing a machine learning model based on ¹⁸F-FDG PET. *J Pers Med.* 2023;13(3):539. DOI: 10.3390/jpm13030539
 23. Albano D, Calabrò A, Dondi F, et al. The role of baseline 2-[18F]-FDG-PET/CT metrics and radiomics features in predicting primary gastric lymphoma diagnosis. *Hematol Oncol.* 2024; 42(2):1-8. DOI: 10.1002/hon.3266
 24. Faudemer J, Aide N, Gac A-C, et al. Diagnostic value of baseline 18FDG PET/CT skeletal textural features in follicular lymphoma. *Sci Rep.* 2021;11(1):23812. DOI: 10.1038/s41598-021-03278-9
 25. Häggström I, Leithner D, Alvéén J, et al. Deep learning for [18F] fluorodeoxyglucose-PET-CT classification in patients with lymphoma: A dual-centre retrospective analysis. *Lancet Digit Health.* 2024;6(2):e114-e125. DOI: 10.1016/s2589-7500(23) 00203-0
 26. de Jesus FM, Yin Y, Mantzorou-Kyriaki E, 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 I.* 2021;49(5):1535-1543. DOI: 10.1007/s00259-021-05626-3
 27. Katsuya Mitamura TN, Yamamotoa Y, Ihara-Nishishitaa A. Texture indices of ¹⁸F-FDG PET/CT for differentiating squamous cell carcinoma and non-Hodgkin's lymphoma of the oropharynx. *Acta Med Okayama.* 2021;75(3):351-356.
 28. Kenawy MA, Khalil MM, Abdelgawad MH, et al. Correlation of texture feature analysis with bone marrow infiltration in initial staging of patients with lymphoma using ¹⁸F-fluorodeoxyglucose positron emission tomography combined with computed tomography. *Pol J Radiol.* 2020;85(1):586-594. DOI: 10.5114/pjr.2020.99833
 29. Kong Z, Jiang C, Zhu R, et al. ¹⁸F-FDG-PET-based radiomics features to distinguish primary central nervous system lymphoma from glioblastoma. *NeuroImage-Clin.* 2019;23:101912. DOI: 10.1016/j.nicl.2019.101912
 30. Li H, Xu C, Xin B, et al. ¹⁸F-FDG PET/CT radiomic analysis with machine learning for identifying bone marrow involvement in the patients with suspected relapsed acute leukemia. *Theranostics.* 2019;9(16):4730-4739. DOI: 10.7150/thno.33841
 31. Ou X, Wang J, Zhou R, et al. Ability of ¹⁸F-FDG PET/CT radiomic features to distinguish breast carcinoma from breast lymphoma. *Contrast Media Mol Imaging.* 2019;2019:4507694. DOI: 10.1155/2019/4507694
 32. Sun Y-W, Ji C-F, Wang H, et al. Differentiating gastric cancer and gastric lymphoma using texture analysis (TA) of positron emission tomography (PET). *Chin Med J.* 2021;134(4): 439-447. DOI: 10.1097/cm9.0000000000001206
 33. Wang J, Zhou Y, Zhou J, et al. Preliminary study on the ability of the machine learning models based on ¹⁸F-FDG PET/CT to differentiate between mass-forming pancreatic lymphoma and pancreatic carcinoma. *Eur J Radiol.* 2024;176:111531. DOI: 10.1016/j.ejrad.2024.111531
 34. Xu H, Guo W, Cui X, et al. Three-dimensional texture analysis based on PET/CT Images to distinguish hepatocellular carcinoma and hepatic lymphoma. *Front Oncol.* 2019;9:844. DOI: 10.3389/fonc.2019.00844
 35. Yang T, Liu D, Zhang Z, et al. Predicting T-cell lymphoma in children from ¹⁸F-FDG PET-CT imaging with multiple machine

- learning models. *J Imaging Inform Med.* 2024;37(3):952-964. DOI: 10.1007/s10278-024-01007-y
36. Gau KA-O, Schmidt CSM, Urbach H, et al. Accuracy and practical aspects of semi- and fully automatic segmentation methods for resected brain areas. *Neuroradiology.* 2020;62(12):1637-1648. DOI: 10.1007/s00234-020-02481-1
 37. Ogura T, Hida K, Masuzuka T, et al. An automated ROI setting method using NEUROSTAT on cerebral blood flow SPECT images. *Ann Nucl Med.* 2009;23(1):33-41. DOI: 10.1007/s12149-008-0203-7
 38. Jiang H, Li A, Ji Z, et al. Role of radiomics-based baseline PET/CT imaging in lymphoma: Diagnosis, prognosis, and response assessment. *Mol Imaging Biol.* 2022;24(4):537-549. DOI: 10.1007/s11307-022-01703-7
 39. Jiang C, Li A, Teng Y, et al. Optimal PET-based radiomic signature construction based on the cross-combination method for predicting the survival of patients with diffuse large B-cell lymphoma. *Eur J Nucl Med Mol I.* 2022;49(8):2902-2916. DOI: 10.1007/s00259-022-05717-9
 40. Eertink JJ, van de Brug T, Wiegers SE, et al. ¹⁸F-FDG PET baseline radiomics features improve the prediction of treatment outcome in diffuse large B-cell lymphoma. *Eur J Nucl Med Mol I.* 2021;49(3):932-942. DOI: 10.1007/s00259-021-05480-3
 41. Schulz KF, Grimes DA. Sample size calculations in randomised trials: Mandatory and mystical. *Lancet.* 2005;365(9467):1348-1353. DOI: 10.1016/S0140-6736(05)61034-3
 42. Kebaili A, Lapuyade-Lahorgue J, Ruan S. Deep learning approaches for data augmentation in medical imaging: A review. *J Imaging.* 2023;9(4):81. DOI: 10.3390/jimaging9040081
 43. Lo Iacono F, Maragna R, Pontone G, et al. A novel data augmentation method for radiomics analysis using image perturbations. *J Imaging Inform Med.* 2024;37(5):2401-2414. DOI: 10.1007/s10278-024-01013-0
 44. Sun Y, Yuan P, Sun Y. MM-GAN: 3D MRI data augmentation for medical image segmentation via generative adversarial networks. In *2020 IEEE International Conference on Knowledge Graph (ICKG)*, 9-11 August 2020; pp. 227-234. DOI: 10.1109/ICKG50248.2020.00041.
 45. Parmar C, Barry JD, Hosny A, et al. Data analysis strategies in medical imaging. *Clin Cancer Res.* 2018;24(15):3492-3499. DOI: 10.1158/1078-0432.Ccr-18-0385
 46. Zwanenburg A. Radiomics in nuclear medicine: Robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. *Eur J Nucl Med Mol I.* 2019;46(13):2638-2655. DOI: 10.1007/s00259-019-04391-8
 47. Eertink JJ, Pfäehler EAG, Wiegers SE, et al. Quantitative radiomics features in diffuse large B-cell lymphoma: Does segmentation method matter? *J Nucl Med.* 2022;63(3):389-395. DOI: 10.2967/jnumed.121.262117
 48. Goceri E. Medical image data augmentation: Techniques, comparisons and interpretations. *Artif Intell Rev.* 2023;56(11):12561-12605. DOI: 10.1007/s10462-023-10453-z
 49. Venkataraman V, Browning T, Pedrosa I, et al. Implementing shared, standardized imaging protocols to improve cross-enterprise workflow and quality. *J Digit Imaging.* 2019;32(5):880-887. DOI: 10.1007/s10278-019-00185-4
 50. Trojani V, Bassi MC, Verzellesi L, et al. Impact of preprocessing parameters in medical imaging-based radiomic studies: A systematic review. *Cancers (Basel).* 2024;16(15):2668. DOI: 10.3390/cancers16152668
 51. Velazquez ER, Parmar C, Jermoumi M, et al. Volumetric CT-based segmentation of NSCLC using 3D-Slicer. *Sci Rep.* 2013;3(1):3529. DOI: 10.1038/srep03529
 52. Eertink JJ, Heymans MW, Zwezerijnen GJC, et al. External validation: A simulation study to compare cross-validation versus hold-out or external testing to assess the performance of clinical prediction models using PET data from DLBCL patients. *EJNMMI Res.* 2022;12(1):58. DOI: 10.1186/s13550-022-00931-w
 53. Zhang Y-P, Zhang X-Y, Cheng Y-T, et al. Artificial intelligence-driven radiomics study in cancer: the role of feature engineering and modeling. *Military Med Res.* 2023;10(1):22. DOI: 10.1186/s40779-023-00458-8
 54. Hussain S, Lafarga-Osuna Y, Ali M, et al. Deep learning, radiomics and radiogenomics applications in the digital breast tomosynthesis: A systematic review. *BMC Bioinform.* 2023;24(1):401. DOI: 10.1186/s12859-023-05515-6
 55. Afshar P, Mohammadi A, Plataniotis KN, et al. From handcrafted to deep-learning-based cancer radiomics: Challenges and opportunities. *IEEE Signal Process Mag.* 2019;36(4):132-160. DOI: 10.1109/MSP.2019.2900993
 56. Zhang X, Zhang Y, Zhang G, et al. Deep learning with radiomics for disease diagnosis and treatment: Challenges and potential. *Front Oncol.* 2022;12:773840. DOI: 10.3389/fonc.2022.773840
 57. Steinbuss G, Kriegsmann M, Zgorzelski C, et al. Deep learning for the classification of non-Hodgkin lymphoma on histopathological images. *Cancers (Basel).* 2021;13(10):2419. DOI: 10.3390/cancers13102419
 58. Ou X, Zhang J, Wang J, et al. Radiomics based on ¹⁸F-FDG PET/CT could differentiate breast carcinoma from breast lymphoma using machine-learning approach: A preliminary study. *Cancer Med.* 2019;9(2):496-506. DOI: 10.1002/cam4.2711
 59. Eertink JJ, Zwezerijnen GJC, Cysouw MCF, et al. Comparing lesion and feature selections to predict progression in newly diagnosed DLBCL patients with FDG PET/CT radiomics features. *Eur J Nucl Med Mol I.* 2022;49(13):4642-4651. DOI: 10.1007/s00259-022-05916-4
 60. Rizzo A, Triumbari EKA, Gatta R, et al. The role of ¹⁸F-FDG PET/CT radiomics in lymphoma. *Clin Transl Imaging.* 2021;9(6):589-598. DOI: 10.1007/s40336-021-00451-y
 61. Hasanabadi S, Aghamiri SMR, Abin AA, et al. Enhancing lymphoma diagnosis, treatment, and follow-up using ¹⁸F-FDG PET/CT imaging: Contribution of artificial intelligence and radiomics analysis. *Cancers (Basel).* 2024;16(20):3511. DOI: 10.3390/cancers16203511
 62. Rundo L, Militello C. Image biomarkers and explainable AI: Handcrafted features versus deep learned features. *Eur Radiol Exp.* 2024;8(1):130. DOI: 10.1186/s41747-024-00529-y
 63. Raptis S, Ilioudis C, Theodorou K. Uncovering the diagnostic power of radiomic feature significance in automated lung cancer detection: An integrative analysis of texture, shape, and intensity contributions. *BioMedInform.* 2024;4(4):2400-2425. DOI: 10.3390/biomedinformatics4040129

64. Bleker J, Yakar D, van Noort B, et al. Single-center versus multi-center biparametric MRI radiomics approach for clinically significant peripheral zone prostate cancer. *Insights Imaging*. 2021;12(1):150. DOI: 10.1186/s13244-021-01099-y
65. Akinci D'Antonoli T, Cuocolo R, Baessler B, et al. Towards reproducible radiomics research: Introduction of a database for radiomics studies. *Eur Radiol*. 2024;34(1):436-443. DOI: 10.1007/s00330-023-10095-3
66. Horvat N, Papanikolaou N, Koh DM. Radiomics beyond the hype: A critical evaluation toward oncologic clinical use. *Radiol Artif Intell*. 2024;6(4):e230437. DOI: 10.1148/ryai.230437
67. Zhang W, Guo Q, Zhu Y, et al. Cross-institutional evaluation of deep learning and radiomics models in predicting microvascular invasion in hepatocellular carcinoma: Validity, robustness, and ultrasound modality efficacy comparison. *Cancer Imaging*. 2024;24(1):142. DOI: 10.1186/s40644-024-00790-9
68. Ibrahim A, Vallières M, Woodruff H, et al. Radiomics analysis for clinical decision support in nuclear medicine. *Semin Nucl Med*. 2019;49(5):438-449. DOI: 10.1053/j.semnuclmed.2019.06.005
69. van Timmeren JE, Cester D, Tanadini-Lang S, et al. Radiomics in medical imaging—"how-to" guide and critical reflection. *Insights Imaging*. 2020;11(1):91. DOI: 10.1186/s13244-020-00887-2
70. Shur JD, Doran SJ, Kumar S, et al. Radiomics in oncology: A practical guide. *Radiographics*. 2021;41(6):1717-1732. DOI: 10.1148/rg.2021210037