ELSEVIER

Seminars in
NUCLEAR
MEDICINE

Role of Artificial Intelligence in PET/CT Imaging for Management of Lymphoma

Eren M. Veziroglu, MS,[†] Faraz Farhadi, BS,[*,†] Navid Hasani, BS,[*] Moozhan Nikpanah, MD,[*,‡] Mark Roschewski, MD,[§] Ronald M. Summers, MD, PhD,[*,‖] and Babak Saboury, MD,MPH[*]

Our review shows that AI-based analysis of lymphoma whole-body FDG-PET/CT can inform all phases of clinical management including staging, prognostication, treatment planning, and treatment response evaluation. We highlight advancements in the role of neural networks for performing automated image segmentation to calculate PET-based imaging biomarkers such as the total metabolic tumor volume (TMTV). AI-based image segmentation methods are at levels where they can be semi-automatically implemented with minimal human inputs and nearing the level of a second-opinion radiologist. Advances in automated segmentation methods are particularly apparent in the discrimination of lymphomatous vs non-lymphomatous FDG-avid regions, which carries through to automated staging. Automated TMTV calculators, in addition to automated calculation of measures such as Dmax are informing robust models of progression-free survival which can then feed into improved treatment planning.

Semin Nucl Med 53:426-448 © 2022 Published by Elsevier Inc.

Introduction

Lymphomas are pathologic proliferations of lymphocytic immune cells within the lymphatic system.[1,2] Lymphomas are clinically grouped into Hodgkin (HL) and non-Hodgkin (NHL) types, then further sub-classified by immunophenotype, histopathology, and cytogenetic profiles.[2-4] Common origins include B, T and NK cell types.[5] Lymphoid neoplasms in the United States will account for approximately 5% of all new cancer cases in 2022, placing lymphoma at the sixth highest cancer incidence rate overall (~89,000/y).[6,7] Overall, approximately one million Americans currently experience lymphoma (~760,000 NHL and ~220,000 HL).

Clinical manifestations are diverse and presumably result from metabolic stress (aggressive growth), mass effects (the tumor pressing on surrounding organs), and hematopoietic dysregulation or dysfunction. The clinical manifestation of lymphoma differs depending on histologic subtype and site of involvement. NHL can manifest subacutely or acutely with a rapidly expanding mass, constitutional symptoms, and tumor lysis syndrome, or it can develop insidiously with slowly rising lymphadenopathy that waxes and wanes over years (eg, follicular lymphoma). Most commonly, HL manifests as asymptomatic lymphadenopathy, constitutional symptoms (present in 40% of patients), or a mass on a chest radiograph.[8]

Since *1990s*, 18F-FDG-PET in combination with CT has been deemed the state-of-the-art imaging technology used in the staging workup and therapy response assessment of HL and numerous forms of NHL.[9,10,11] Subsequently, Deauville criteria was proposed to standardize lymphoma interpretation using

*Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda, MD.
[†]Geisel School of Medicine at Dartmouth, Hanover, NH.
[‡]Department of Radiology, University of Alabama at Birmingham, AL.
[§]Lymphoid Malignancies Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD.
[‖]Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences, National Institutes of Health Clinical Center, Bethesda, MD.
Financial disclosure: This research was supported by the Intramural Research Program of the NIH, Clinical Center. The opinions expressed in this publication are the author's own and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government.
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
Address reprint requests to Babak Saboury, MD,MPH, Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, 10 Center Drive, Bethesda, MD 20892. E-mail: Babak.Saboury@nih.org

 https://doi.org/10.1053/j.semnuclmed.2022.11.003
0001-2998/© 2022 Published by Elsevier Inc.

the degree of FDG avidity in relation to the mediastinal blood pool and liver. The Lugano classification, proposed in 2014, sought to simplify and standardize baseline and response assessment in FDG-PET/CT in lymphoma. 18F-FDG-PET/CT imaging provides a clear picture of the metabolic activity and anatomical structure of lymphomas enabling improved diagnosis/differentiation, staging, and prognosis capabilities. Considering the diversity of the lymphomas and their manifestations, functional imaging prior to or during therapy may be used to guide management decisions based on geographic heterogeneity as well as the biological, pathological, and metabolic condition of the tumor.

Artificial intelligence (AI) is uniquely positioned to fundamentally alter medicine, possibly enhancing both physicians' and patients' experiences. AI-based solutions have enhanced workflow in online appointment scheduling, online check-ins at medical centers, the digitization of medical records, reminder calls for follow-up visits and immunization dates, and drug adverse effect warnings when several prescriptions are prescribed.[12,13] AI models have demonstrated a potential to facilitate mammography interpretation,[14,15] echocardiogram interpretation,[16] and cancer screening,[17] prediction and prognosis.[18]

In medical imaging, AI systems can further facilitate many aspects such as clinical decision-making, enhance image acquisition, quality assessment, and post-processing techniques (such as tumor delineation, registration, and quantification), and dose estimation.[19,20,21] Specifically in FDG-PET imaging of lymphoma, recent studies have demonstrated this potential by developing models able to automatically recognize the location of the lymphoma, segment the lesion, summarize lesions characteristics and heterogeneity with appropriate radiomics, and assess disease transformations over several time points.[22,23,24,25,26]

Various pathologic subtypes of lymphoma have different patterns of 18F-FDG avidity. For instance, HD and DLBC NHL lesions often have high FDG avidity and therefore can be staged and evaluated with 18F FDG-PET/CT. Meanwhile, other histological subtypes of NHL such as MALT, marginal zone lymphoma, and small lymphocytic lymphoma may have lymphomatous involvement without high FDG avidity.[27] In such conditions, AI-based FDG-PET/CT imaging may identify patterns of FDG uptake and notify the physicians that the evaluated subtype has limited FDG avidity and may not be further assessed by AI-based tools.

In a previous publication, our team conducted a systematic scoping review of the applications of AI in 18F-FDG-PET/CT lymphoma based on studies published prior to September 2021.[22] There, we explored the various structures of proposed AI systems, tasks performed by the models (eg, classification, characterization, detection, segmentation), and applications of radiomic analysis in lymphoma prognostication and management.

To perform this review, the PubMed database was queried using the search terms "artificial intelligence," "lymphoma," and "positron emission tomography." English articles on pertinent clinical research were included. Publications focusing on several diseases, conference articles, and literature reviews, articles written in a language other than English, and articles that were inaccessible were eliminated. This search yielded 17 articles, the majority of which developed AI-based lymphoma segmentation models or used AI based PET/CT radiomics to predict prognosis in various lymphoma subtypes. Key studies are highlighted in Table 1, and Table 2.

In this article, while reviewing novel uses of AI in lymphoma PET/CT reported since September 2021, we also focus on automatic segmentation and the utility of radiomic features with discussion tailored toward clinicians. AI based segmentation can enable prognostication and radiomic analysis of lymphoma PET/CT studies to obtain useful insights regarding therapy enhancement, remission, and recurrence prediction.

Automatic segmentation of PET images in management of lymphoma

On Disease Prognostication and Radiologic Biomarkers

In this section, we discuss quantifiable radiologic features as proxies of disease burden that are otherwise impractical to implement without AI methods.

Similar to many other malignancies, lymphomas go through progressively worsening clinical stages as the cancer cells spread to distant parts of the body. Quantifying the spread of the disease is part of many of the important prognostic factors in lymphoma management. Imaging, particularly PET-based imaging, is key for determining the extent of lymphomatous disease, treatment responsiveness, and ultimately defining and determining disease remission. Most clinical radiologists apply a very basic level of image analysis, commenting for example on the number, anatomic locations, and longest axes of the regions of FDG avidity. Experienced radiologists can accurately integrate PET/CT data, identify pathologic from non-pathologic regions, and draw their boundaries to create a set of all FDG signals related to tumor activity; however, this process is not routinely done. We call this set of all FDG-avid volumes of interest the total metabolic tumor volume (TMTV).

Our current clinical models of disease prognostication comparatively input a very basic level of image information. Notably, the Lugano classification system stratifies disease into four stages: a single node region, two or more ipsilateral regions, involvement both superior and inferior to the diaphragm, and diffuse/disseminated involvement of one or more extra lymphatic organs. These stages are presumably a proxy for the state of someone's disease along a timeline, and although there is some quantitative element in this staging system, it is highly discretized and therefore may not as accurately represent total disease burden as a continuous variable such as TMTV. We have substantial evidence to support a hypothesis that TMTV is a high-quality proxy for disease burden and therefore stage along a timeline, particularly for lymphomas[28,61](Fig. 1). However, its implementation in

Table 1 Summary of AI Studies of Lymphoma Modified From: Hasani N, Paravastu SS, Farhadi F, Yousefirizi F, Morris MA, Rahmim A, et al

CNN Models					
Author (Year)	**Tasks Performed**	**Task Specific Input/ Out**	**FoM**	**Details (Model-Related)**	**Details (Ground Truth, Sample Size)**
Pinochet et al (2021)	Classification (Radiophenomics)	Input: 2D; WB; axial/sagittal/coronal 18F-FDG-PET slice Output: slice-level III-category classification (Benign, Malignant, Equivocal lymph nodes)	AUC = 0.62	Evaluate PET Assisted Reporting System (PARS-PET) by Siemens on DLBCL patients	Ground Truth: 2 NM physicians segmented DLBCL lesions
	Segmentation	Input1: 2D; WB; axial/sagittal/coronal 18F-FDG-PET slices Output: 2D; segmented lesions with borders masked on PET slice	Dice = 0.65 (research cohort) Dice = 0.48 (routine cohort)	CNN model: PET Assisted Reporting System (PARS)	Test sample: 119 patients (research cohort) + 430 patients (routine cohort)
	Statistical Classification (Prediction/ Prognosis)	Input: TMTV value from 2D; WB; axial/sagittal/coronal 18F-FDG-PET slices Output: Patient-level II-class prognostication (low-TMTV and high-TMTV groups)	ICC = 0.68 (research cohort) ICC = 0.61 (routine cohort)		
Sadik et al (2021)	Classification (Radiophenomics)	Input: 2D; WB; sagittal/coronal/axial; 18F-FDG-PET slice and CT slice (2 channel input) Output: Pt-level 4-class classification [high vs low diffuse bone marrow uptake] x [presence vs absence of focal lesion])	PA = 0.85 Kappa = 0.41	Alert for focal skeleton/ bone marrow uptake in Hodgkin's lymphoma patients CNN model: based on RECOMIA prototype	Ground Truth: 10 independent NM physicians with 2-12 y of experience with PET/CT classified lesions Training sample size: 156 Test sample size: 49
Guo et al (2021)	Radiophenomics-part1 (Characterization/Radiomics)	Input: 2D; WB; axial; 18F-FDG-PET and CT slice (1 channel) Output: 128 features grouped into feature maps of 16 $\times$ 8 strips		Extraction of feature maps surrogates for prognosis prediction in nasal ENKTL. Proposes PSI to be predictor of PFS; PSI is the ratio of the PPV to NPV	Ground Truth: 1 NM physician (15 yrs. experience) segmented nasal ENKTL lesions

Table 1 (*Continued*)

CNN Models Author (Year)	Tasks Performed	Task Specific Input/ Out	FoM	Details (Model-Related)	Details (Ground Truth, Sample Size)
	Radiophenomics-part 2 (statistical classification (Prediction/Prognosis))	Input: Prediction similarity index (PSI) derived from image features Output: Relapsed vs non-relapsed classes for ENKTL	AUC = 0.88 (for PSI) PSI-based PFS prediction: Spec = 0.80, Sens = 0.83, Accuracy = 0.85	Model: Weakly supervised deep learning (WSDL) based on Residual Network-18 (ResNet-18) and PNU classifier.	Training sample: 64 Testing sample: 20
Yuan et al (2021)	Detection	Input: 2D; axial; neck/chest/abdomen; 18F-FDG-PET slice and CT slices (2 channel input) Output: 2D; axial; detection map with lesions in rectangular boxes	Sens (chest) = 83.2%, Spec (chest) = 99.75%, Accuracy = 99.5%	Hybrid Learning for feature fusion of DLBCL Segmentation Hybrid CNN model can create feature fusion maps and quantify spatial contributions of each modality.	Ground Truth: 1 physician manually segmented DLBCL lesions
	Segmentation	Input: 2D; axial; neck/chest/abdomen; 18F-FDG-PET slice and CT slices (2 channel input) Output: lesions border segmentation map	Dice = 0.73, MHD= 4.38mm	PET and CT image feature-based hybrid learning CNN model architecture	Training and Validation: Randomly divided total 1242 PET/CT slice pairs from 45 PET/CT samples into 15 distinct training and test sets for a 15-fold cross-validation experiment.
Blanc-Durand et al (2020)	Segmentation	Input: 3D; coronal; WB; 18F-FDG-PET and 3D CT two separate channels Output: Mask of segmented lesions with calculated TMTV on 18F-FDG-PET/CT	Jaccard = 0.60, Dice = 0.73 Predicted TMTV R = 0.88, 0.82 in first cohort, second cohort respectively	Fully automatic segmentation of DLBCL lesions for total MTV prediction - 3D FDG-PET/CT	Ground truth masks were manually obtained after a 41% SUVmax adaptive thresholding of lesions. TMTV protocol from LIFEx used for VOI semi-automatically segmented. The resulting clusters were reviewed by 2 experienced physicians to remove physiological uptakes and segment lesions.

Table 1 (*Continued*)

CNN Models Author (Year)	Tasks Performed	Task Specific Input/ Out	FoM	Details (Model-Related)	Details (Ground Truth, Sample Size)
Weisman et al (2020)	Segmentation	Input: 3D; coronal; WB; 18F-FDG-PET and CT image (2 channels) Output: Map of masked segmented lesions	Dice = 0.86	Fully automatic measurement of PET imaging features in PET/CT images of pediatric lymphoma	Ground Truth: 1 NM physician with 11yrs of experience segmented and determined malignancy status at lymph nodes
	Characterization (Radiomics)	Input: 3D; coronal; WB; PET/CT slices with segmented lesions Output: SUVmax, MTV, TLG, SA/MTV, measure of disease spread (Dmaxpatient)	R = 0.95	Model: ensemble of 3 DeepMedics	Training/validation: 80 patients Testing: 20 patients
Weisman et al (w/ Kieler) (2020)	Detection	Input: 2D; coronal; WB; 18F-FDG-PET slice (from PET/CT) Output: Lymph nodes probability map contoured	TPR = 0.85 4 false positives/patient	Automated detection of diseased lymph node Burden in lymphoma patients - PET/CT Model: ensemble of 3 DeepMedics	Ground Truth: 1 NM physician with 11y of experience segmented and determined malignancy status at lymph nodes Training: 58 patients Test: 90 patients
Sibille et al (2020)	Detection (localization + 4-category classification suspicious vs nonsuspicious for lung cancer or lymphoma)	Input: 2D; coronal; WB; 18F-FDG-PET slice fused with CT, MIP, anatomical atlas Output: Map of detected lesions classified under [suspicious or non-suspicious] x [lung cancer or lymphoma]	For Localization: Sens = 0.81, Spec = 0.97, accuracy = 0.96 (for body parts), 0.87 (for region), 0.81 (for subregion) For Classification: false positive = 1.47 (96/65), false negative = 1.76 (115/65), AUC = 0.98	18F-FDG Uptake Classification in Lymphoma and Lung Cancer - using CT, PET, MIP, and atlas information	Ground Truth: 2 NM physicians annotated and segmented foci with increased F-FDG uptake specified the anatomic location and classified.
Li et al (2019)	Segmentation	Input: 2D; axial; WB; 18F-FDG-PET and CT slices (6 channels) Output: segmentation map of lymphoma	Dice = 0.73, Precision = 0.70, Recall = 0.81	End-to-End lymphoma segmentation - WB PET/CT DenseX-Net	Ground Truth: 3 clinicians delineated images, then verified and revised by 1 nuclear medicine expert

Table 1 (*Continued*)

Author (Year)	Tasks Performed	Task Specific Input/ Out	FoM	Details (Model-Related)	Details (Ground Truth, Sample Size)
CNN Models					
Sadik (2019)	Segmentation	Input: 2D axial/coronal 18F-FDG-PET and CT images Output: Segmentation of liver and the mediastinal blood pool (aorta)	Dice = 0.95	Automated quantification of reference levels in liver and mediastinal blood pool for therapy response classification in HL and NHL - FDG-PET/CT	Ground Truth: 2 radiologists segmented images Training: 80 Validation: 6
Bi et al. (2017)	Detection	Input: 3D; WB; coronal; 18F-FDG-PET with CT slices (2 channels) Output: 3D; WB; coronal; map of sFEPU regions (ie, Left, and right kidneys, bladder, brain, heart)	F1 Score: 0.92	Automatic detection of superpixel regions of FDG uptake of lymphoma regions Model details: MSE + CFSC	Ground Truth: 1 experienced operator manually identified ROI using PERCIST thresholding and the diagnostic report of PET/CT scan Trained: 1.5 million non-medical images, validated: 50,000 non-medical images Testing: 11 lymphoma patients
Classic Machine Learning					
Annunziata et al (2021)	Statistical Classification (Prediction/ Prognosis)	Input: Deauville Score, qPET, MTV0, slope (slope of linear function of MTV) features from 3D; axial/coronal end-of-treatment compared to beginning-of-treatment 18F-FDG-PET and CT slices Output: Patient-level II-class prediction (relapse vs progression)	PPV = 0.55, NPV = 0.83 (for DS 4-5) PPV = 0.89, NPV = 0.82 (for positive qPET) R = 0.63 (for ANN)	Assess the prognostic role of end-of-treatment FDG-PET/CT in DLBCL patients Model details: multi-regression model (MM), ANN	Ground Truth: 2 NM physicians independently evaluated using a dedicated fusion and display software Training: 26 patients Testing: 11 patients K-fold cross validation

Table 1 (*Continued*)

CNN Models Author (Year)	Tasks Performed	Task Specific Input/ Out	FoM	Details (Model-Related)	Details (Ground Truth, Sample Size)
Lippi et al (2020)	Classification (Radiophenomic)	Input: 3D; WB; coronal 18F-FDG-PET slices Output: Patient-level IV-class classification of malignant lymphoma (DLBCL, follicular lymphoma, HL, and mantle cell lymphoma)	Sens = 0.97, PPV = 0.94	Texture analysis and classification of malignant lymphoma Model details: SVM + RF	Ground Truth: 1 NM physician with 5 yrs experience extracted VOIs using a 40%-threshold of SUVmax Evaluation: leave-one-out (LOO) procedure, where each patient was used, in turn, as the test set, and all the other patients constituted the training set.
Mayerhoefer et al (2019)	Statistical Classification (Prediction/ Prognosis)	Input: TMTVs, SUV, TLG, 16 others textural radiomic features Output: Patient-level III-category metabolic risk (low, intermediate, high) of progression	AUC = 0.72,	Radiomic features for prediction of outcome in mantle cell lymphoma International prognostic indices for MCL = MIPI and MIPI-b Model details: Multilayer perceptron (MLP) feed-forward ANN	Ground Truth: TMTV protocol used to semi-automatically construct with 41% SUVmax threshold Samples size: Training: 75, Testing: 32 patients
Hu et al (2019)	Detection	Input: 3D; WB; coronal/ axial/sagittal 18F FDG-PET slices and CT slices (2 Channels) Output: 3D probability map of the segmented lesion (normal organ and tumors)	Sens = 0.80, F1-Score = 0.59	Entropy-based optimization strategy for clustering by integrating physical spatial attributes and prior knowledge	Ground Truth: Segmentation ground truth obtained by 41% SUV max thresholding, no information on the physicians
	Segmentation	Input1: 3D; WB; coronal/ axial/sagittal 18F FDG-PET slices and CT slices (2 Channels) Input2: Detection results Output: 3D, coronal/axial/sagittal slice with segmented normal organ and tumor lesions.	Dref = 0.74, Dglobal = 0.50, Volumesup = 0.39	CNN: Density-based spatial clustering of applications with noise (DBSCAN)	Testing sample: 48 patients

Table 1 (*Continued*)

CNN Models Author (Year)	Tasks Performed	Task Specific Input/ Out	FoM	Details (Model-Related)	Details (Ground Truth, Sample Size)
Yu et al (2018)	Detection	Input: 2D; WB; coronal/ axial/sagittal 18F-FDG-PET slices and CT slices (2 Channels) Output: probability map of detected lymphoma lesions	Sens = 1.0	Semi-automatic lymphoma detection and segmentation	Ground Truth: 1 physician contoured images
	Segmentation	Input1: PET/CT images with physiologic hypermetabolic organs removed. input2: Detection results Output: Border mask segmentation visualized on software on axial, sagittal, and coronal	Dice = 0.84	Model details: FC CRF	Training/validation sample size: 11 patients
Grossiord et al (2017)	Classification (Radiophenomics)	Input: 2D; coronal; 18F-FDG-PET and CT slices (2 channel) Output: slice-level III-class classification (Organ, tumor, non-relevant)	Sens = 0.65, Spec = 0.92, Accuracy = 0.86	Automated 3D lymphoma lesion segmentation - PET/CT	Ground Truth: 1 expert manually expert segmentation at 41% SUVmax
	Segmentation	Input: 2D; WB; coronal 18F-FDG-PET and CT slices (2 channel) Output: 2D, WB, coronal segmentation map	Dice = 0.75		Training/validation sample size: 43 patients Leave-one patient-out (LOPO) cross-validation for classification task
Desbordes et al (2016)	Segmentation	Input: 2D; WB; coronal/ axial/sagittal 18F FDG-PET slices and CT slices (2 Channels) Output: 2D, WB, lesion segmentation map	Dice = 0.80	Cellular automata define tumor seed within ROI to obtain final segmentation by iterative growth. Model details: auto-initialization cellular automata (CA)	Ground Truth: 1 NM physician manually selected and segmented the ROI Testing sample size: 12 patients

Artificial Intelligence in Lymphoma PET Imaging: A Scoping Review (Current Trends and Future Directions). PET Clin. 2022 Jan;17(1):145-74.

Table 2 Summary of AI Studies of Lymphoma Modified From: Hasani N, Paravastu SS, Farhadi F, Yousefirizi F, Morris MA, Rahmim A, et al

Authors (Year)	Lymphoma Subtypes	Aim of Study		Radiomic Feature Information			Discriminator Used	Figures of Merit
		Input	Goal	Extraction Method	Features Used	Notable Features		
Studies in which radiomic features were used for the prognosis/prediction of lymphoma								
Rodriguez Taroco et al,[78] 2021	HL	Segmented tumor VOIs on 18F-FDG-PET	Prediction of PFS from 18F-FDG-PET radiomic features in HL and DLBCL	Not specified	8 first-order features, 23 features from GLCM, 11 features from GLRLM, 5 features from NGLM, 3 features from the neighborhood grey-tone difference	PFS in patients with Deauville scores of 1, 2, 3, and X at initial PET was higher than that in patients with a Deauville score of 4	Univariate and multivariate Cox regression analysis	Average PFS, for patients with Deauville 4 score, of 1120 d (95% CI, 229-672)
Eertink et al.[79] 2021	DLBCL	Segmented tumor VOIs on 18F-FDG-PET	Prediction of treatment outcome with first-line treatment of DLBCL from baseline 18F-FDG-PET radiomic features	RaCat	Large number of morphologic and texture features were extracted	Five models were created based on radiomic features as well as clinical predictors; combination of clinical and radiomics predictors was best	ROC analysis	Combined model: HR = 4.6 (95% CI, 2.6-7.9)
Wang et al.[80] 2020	ENKTL	Segmented tumor VOIs on 18F-FDG-PET	Identify a 18F-FDG-PET radiomics-based model for predicting PFS and OS in ENKTL	LifeX	41 features	Radiomics and metabolism-based models were combined to predict both PFS and OS	Univariate and multivariate Cox regression analysis	PFS: 0.788 (95% CI = 0.682-0.895) and 0.473 (P = 0.803) OS: 0.637 (95% CI = 0.488-0.786) and 0.730 (95% CI = 0.548-0.912)

Table 2 (*Continued*)

Authors (Year)	Lymphoma Subtypes	Aim of Study		Radiomic Feature Information			Discriminator Used	Figures of Merit
		Input	Goal	Extraction Method	Features Used	Notable Features		
Sun et al,[81] 2020	Primary gastric DLBCL	Segmented tumor VOIs on 18F-FDG-PET	Texture analysis of 18F-PET/CT scans to predict interim response after 3-4 rounds of chemotherapy in primary gastric DLBCL	In-house software	First and second-order features	Combination of SUVmax, volume, and entropy in one model best predicted treatment response	Mann-Whitney U	AUC = 0.915
Aide et al,[82] 2020	DLBCL	Segmented tumor VOIs on 18F-FDG-PET	Prognosticate DLBCL treated with first-line immunotherapy using radiomic features from baseline 18F-FDG-PET	LifeX	19 features	18F-FDG-PET heterogeneity of the largest lymphoma lesion is associated with 2y-event free survival (EFS)	Univariate and multivariate Cox regression analysis	EFS: HR = 7.47 (95% CI = 0.83-66.99)
Wu et al,[83] 2019	DLBCL	18F-FDG-PET/CT pre and posttreatment	Radiomics-based treatment outcome prediction model	MATLAB	GLCM, GLRLM, GLSZM	Belief-function theory-based outcome prediction outperformed than other studies	EK-NN and SVM	Therapy response: NS
Tatsumi et al.[84] 2019	FL	Segmented tumor VOIs on 18F-FDG-PET	Predict response and recurrence after therapy in FL	PETSTAT	6 texture features	low gray-level zone emphasis (LGZE) in texture features predicted complete response	Logistic regression	Therapy response: AUC = 0.720; PFS: NS
Lue et al.[85] 2019	HL	Segmented tumor VOIs on 18F-FDG-PET	18F-FDG-PET was analyzed using radiomics to predict/prognose HL	OsiriX, CGITA, MATLAB	11 first-order, 39 higher-order, 400 wavelet features	Ann Arbor stage, GLRLM and SUV kurtosis were associated with PFS	Univariate and multivariate Cox regression analysis	PFS: HR = 6.640 (95% CI, 1.261-34.96; $P = 0.026$); OS: HR = 14.54 (95% CI, 1.808-117.0; $P = 0.012$)

Table 2 (*Continued*)

Authors (Year)	Lymphoma Subtypes	Aim of Study		Radiomic Feature Information			Discriminator Used	Figures of Merit
		Input	Goal	Extraction Method	Features Used	Notable Features		
Lue et al.[86] 2019	HL	Segmented tumor VOIs on 18F-FDG-PET	Radiomic intratumor heterogeneity in 18F-FDG-PET to predict treatment response and survival outcomes in patients with HL	OsiriX, CGITA, MATLAB	7 SUV and HU, 78 second- and higher-order, 624 wavelet features	Treatment response was associated with high-intensity run emphasis (HIR) was performed on PET images and run-length nonuniformity (RLNU) of CT extracted from gray-level run-length matrix (GLRM) in high-frequency wavelets PFS was independently associate with intensity nonuniformity (INU) of PET and wavelet short-run emphasis (SRE) of CT from GLRM and Ann Arbor stage. OS was associated with zone-size nonuniformity (ZSNU) of PET from gray-level size zone matrix (GLSZM)	Cox proportional hazards model, ROC curve, logistic regression	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

Table 2 (*Continued*)

Authors (Year)	Lymphoma Subtypes	Aim of Study		Radiomic Feature Information			Discriminator Used	Figures of Merit
		Input	Goal	Extraction Method	Features Used	Notable Features		
Zhou et al,[87] 2019	Primary gastric DLBCL	Segmented tumor VOIs on 18F-FDG-PET	Prediction of OS and PFS from 18F-FDG-PET radiomic features in primary gastric DLBCL	LifeX	44 texture features	Kurtosis, TMTV, GLNU, and HGZE were identified as independent prognostic factors	Univariate and multivariate Cox regression analysis	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; $P = 0.016$)
Milgrom et al.[88] 2019	Mediastinal HL	Segmented nodal disease on 18F-FDG-PET/CT	Predict response to therapy in mediastinal HL	MIM, IBEX	GLCM, intensity histogram, shape	A combination model of 5 most predictive features accomplished the highest AUC (SUVmax, TMTV, inverse variance, and 2 measures of tumor heterogeneity)	ROC analysis	AUC = 0.952
Wang et al,[89] 2019	Renal/adrenal lymphoma	Segmented tumor VOIs on 18F-FDG-PET	Prognose patients with primary renal lymphoma and primary adrenal lymphoma using texture features	LifeX	37 texture features	GLRLM_RLNU (gray-level co-occurrence matrix run-length nonuniformity) was most predictive of OS.	Univariate and multivariate Cox regression analysis	OS: HR = 9.016 (95% CI, 1.041-78.112; $P = 0.046$)

Table 2 (*Continued*)

Authors (Year)	Lymphoma Subtypes	Aim of Study		Radiomic Feature Information			Discriminator Used	Figures of Merit
		Input	Goal	Extraction Method	Features Used	Notable Features		
Parvez et al.[90] 2018	NHL	TMTV using thresholding and radiomic features	Predict response to therapy and outcome in NHL using radiomic features extracted from 18F-FDG-PET/CT	LifeX	GLCM, NGLDM, GLRLM, GLZLM, indices from sphericity and histogram	GLNU correlated to DFS, and kurtosis correlated with OS	Univariate Cox regression analysis	Therapy response: NS; DFS: $P = 0.013$; OS: $P = 0.035$
Aide et al,[77] 2018	DLBCL	Axial skeleton segmented on 18F-FDG-PET	Determine prognostic value of skeletal textural features in DLBCL	LifeX	4 first-order, 6 second-order and 11 third-order texture features	The only independent predictor of PFS was SkewnessH	ROC analysis	PFS: HR = 3.17 (95% CI, 1.00-10.04; $P = 0.032$)
Ben Bouallègue et al. 91 2017	Bulky HL and NHL	Segmented tumor VOIs on 18F-FDG-PET	Predict response to therapy in bulky HL and NHL	In-house software	Shape, texture features	SVM accounting for both texture and shape features achieved the highest ROC AUC	ROC analysis	AUC = 0.820
Coskun et al.[65]	DLBCL	Segmented tumor VOIs on 18F-FDG-PET	Assessment of baseline 18F-FDG-PET/CT using radiomics to predict response to chemotherapy (R-CHOP) in DLBCL.	Recursive feature elimination (RFE) algorithm	14 textural features	SUVmax, Gray level co-occurance matrix	Multivariate analysis and ROC analysis	Model for predicting incomplete response had 0.87 accuracy and 0.81 AUC.
Yuan et al.[66]	DLBCL	Hidden image features	Interim [18F] FDG-PET/CT for predicting therapy failure in DLBCL	Conv-LSTM	High-level semantic features reflecting intraslice spatial structures and interslice contextual correlations		ROC analysis	In the test cohort and external datasets, the hybrid learning model attained AUCs of 0.926 and 0.925, respectively.

Table 2 (*Continued*)

Authors (Year)	Lymphoma Subtypes	Aim of Study		Radiomic Feature Information			Discriminator Used	Figures of Merit
		Input	Goal	Extraction Method	Features Used	Notable Features		
Jiang et al[40]	DLBCL	Segmented tumor VOIs on 18F-FDG-PET	Using automatic segmentation radiomics for prognostication of DLBCL	Not specified	gtTMTV, pTMTV. pTMTV	pTMTVwas determined as an independent prognostic factor of survival in DLBCL	Significant univariate pTMTV and clinical variables were included into a Cox proportional hazards model multivariate analysis. ROC curves were used to determine the appropriate pTMTV cut-off values.	pTMTV PFS: HR = 3.097; OS: HR = 6.601
Frood et al[67]	cHL	Segmented tumor VOIs on 18F-FDG-PET	Predicting outcome in cHL patient with pretreatment FDG-PET/CT	PyRadiomics with ComBat harmonisation	MTV, TMTV		ROC analysis	Mean SUV of Training: 0.82 Validation: 0.79 Test: 0.81
Jiang et al[68]	Primary Gastrointestinal DLBCL	Semi-automatically segmented VOI on 18FDG-PET	Investigate the prognostic value of PET radiomics feature in the prognosis	PyRadiomics	1421 total features extracted	Metabolic parameters, and clinical characteristics outperformed clinical models and NCCN-IPI (National Comprehensive Cancer Network International Prognostic Index).	Multivariate analysis and multivariate Cox regression analysis	Combined model (radiomics signatures, metabolic, clinical factors) exhibited strong C-indices training: PFS: 0.825, OS: 0.834; validation set PFS: 0.831, OS: 0.877

Table 2 (Continued)

Authors (Year)	Lymphoma Subtypes	Aim of Study		Radiomic Feature Information			Discriminator Used	Figures of Merit
		Input	Goal	Extraction Method	Features Used	Notable Features		
Jiang et al[69]	DLBCL	Segmented tumor VOIs on 18F-FDG-PET	To create and externally evaluate OS and PFS prediction models using a PET radiomics signature (R-signature)	PyRadiomics	49 feature selection-classification candidates were extracted	R-signature of 12 and 31 radiomics features were significantly associated with PFS and OS	Univariate and multivariate Cox regression	Combined model: Training: PFS:0.80, OS:0.807 External validation: PFS:0.758, OS:0.696
Ritter et al[70]	DLBCL	Segmented tumor VOIs on pre-treatment 18F-FDG-PET	To study value of image radiomics and clinical parameters in determining 2-y event free survival	IBSI (Imaging Biomarker Standardization Initiative) radiomic features	30 features were selected	Max diameter, neighbor gray tone difference matrix (NGTDM) busyness, TLG, TMTV, and NGTDM coarseness.	ROC analysis	AUC: 0.85

Artificial Intelligence in Lymphoma PET Imaging: A Scoping Review (Current Trends and Future Directions). PET Clin. 2022 Jan;17(1):145-74.

patient care is challenged by the time and labor-intensive nature of the process causing it to be underutilized.[29] The ability to automatically segment PET images can enable numerous additional analytical tools for improved understanding and management of lymphoma beyond TMTV as well. Thus far, we have shown that TMTV proxies for total disease burden and therefore prognostication, and that automated segmentation methods are key to enable the widespread clinical implementation of these measures.

On AI-Based Automated TMTV Calculation

AI methods have been used for TMTV calculation. In this section we provide an outlined description of automated segmentation for calculating TMTV tailored toward clinicians. For more detail on study design, data collection and labeling, see our collaborators' previous work with the SNMMI task force on AI.[30] A workflow presented by the task force is presented here (Fig. 2).

Data Labeling and Challenges Therein

The determination of TMTV in current practice is typically based on a semi-automatic segmentation process by using a thresholding algorithm or region growing to define volumes of interest around the tumor. Then, a trained radiologist makes manual modifications (boundary adjustment) and annotations (tumor or not) to these regions, as well as adding new regions or removing erroneous regions to define the ground truth. In addition to labor and time-intensity of the task, challenges in ground truth determination include: (1) intra- and inter-reader variability meaning that each human reader will have a different "truth" and (2) that annotation with one hundred percent accuracy is essentially impossible, meaning there will always be some number of false positives and false negatives with both human and computer analysis. For example, Weisman et al. trained a CNN with 5 fold cross-validation and compared the algorithm's performance with agreement between physicians. The authors showed the algorithm could automatically detect diseased lymph nodes in PET/CT with an error rate comparable to inter-physician variance. In 20 patient scans read by two radiologists, the second reader identified 210 of 219 of the nodes while the CNN identified 197 of 219 of the nodes.[24] The relevance of this study is underlying in the idea of our inability to exactly know the ground truth of what constitutes the set of diseased lymph nodes vs benign ones. Therefore, we cannot grade the CNN model necessarily on a true-positive rate, but rather how well it can identify the set of all lymph nodes that were shared in their annotation between trained physicians. The AI task force of the SNMMI has agreed that the ground truth is rarely known in clinical studies, as it requires a biopsy or post-mortem evaluation to truly determine. While many clinical studies designate a "ground truth," they are in fact referring to adjudication by an expert labeler or labelers.

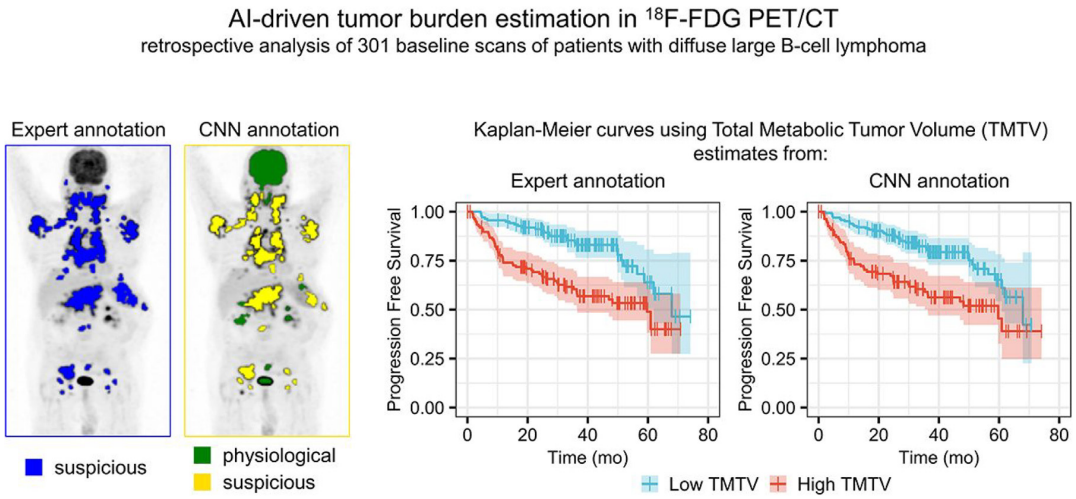

Figure 1 Stratification of progression free survival by TMTV level. This research was originally published in JNM. Capobianco et al. Deep-Learning 18F-FDG Uptake Classification Enables Total Metabolic Tumor Volume Estimation in Diffuse Large B-Cell Lymphoma. J Nucl Med. 2021 Jan;62(1):30-6.

Utilizing deep learning for the identification and segmentation of locations of disease activity in lymphoma enables global and site-specific evaluation of disease burden, giving vital prognostic data to already-in-use clinical risk scores.[31]

Overcoming challenges of automated segmentation (sFEPU)

One of the challenges for both radiologists and automated methods for pathologic FDG-avid region identification are sites of FDG excretion and physiologic uptake (sFEPU). sFEPU include the bladder (excretory pool) and highly metabolic tissues such as the kidneys, brain and heart (Fig. 1). In a study Bi et al. approached the problem of identifying sites of disease involvement in lymphoma in a reverse manner by removal of normal structures. The authors used a multiscale superpixel-based encoding method and class-driven feature selection and classification model that was applied to 40 whole-body lymphoma PET/CT scans and demonstrated the ability to classify entire sFEPU regions.[32]

Using CT information in automatic detection of sites of involvement in lymphoma is an opportunity for differentiating normal vs abnormal uptake. In a study Lartizien et al. tested performance of a support vector machine (SVM) based on texture features to evaluate a database of regions of interest (ROI) from PET/CT scans of patients with lymphoma consisting of 156 lymphomatous and 32 suspicious but non lymphomatous ROIs. They included training databases of PET only or PET and CT features with or without feature selection to assess the added value of multimodality approach to classification performance. The series of combined PET and CT features resulted in the greatest classification performance, highlighting the potential benefits of combining different complementary imaging modalities.[33] While most publications on the topic report on combined information from both modalities, differentiating abnormal vs normal but suspicious regions remains a challenge in

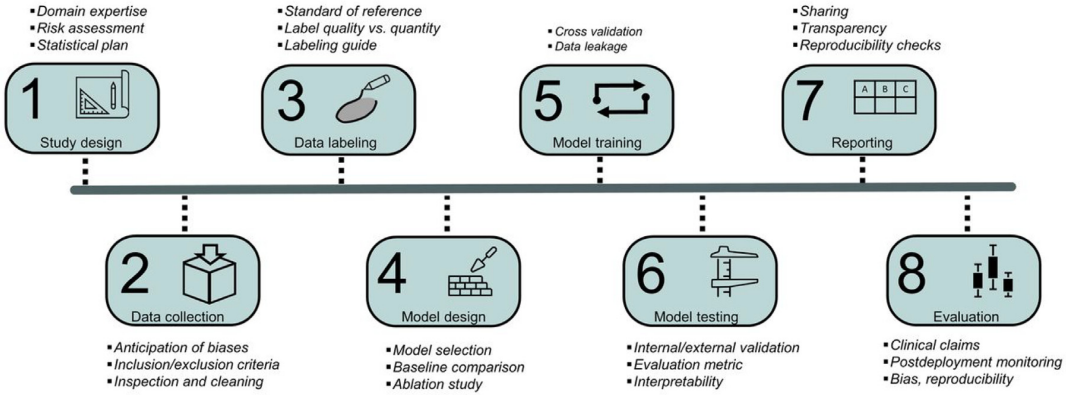

Figure 2 A general paradigm for development of automated image segmentation algorithms. This research was originally published in JNM. Bradshaw TJ, Boellaard R, Dutta J, Jha AK, Jacobs P, Li Q, et al. Nuclear Medicine and Artificial Intelligence: Best Practices for Algorithm Development. J Nucl Med. 2022 Apr;63(4):500-10.

detection of sites of disease involvement. Weisman et al. reported FP findings were found most often in areas of higher physiologic uptake for instance in several cases the network was unable to differentiate brown adipose tissue uptake from lymph node uptake despite having CT as an input.[24]

We have seen the challenges of automated segmentation methods in determining sFEPU and expect that further developments focused on uncertainty management will dramatically improve model outcomes. In this subsection, we have discussed how regions of interest are identified and the challenges in automated tumor region detection.

Model Development, Training, and Challenges Therein

Next, a specific model for artificial intelligence must be defined. The current state of the art is a convolutional neural network (CNN).[34] CNNs work well for image segmentation tasks and they are inspired by visual systems in animals.[35] CNNs apply matrix operations to learn about image features at different levels, for example, edge finding, loco regional features, and macro-level shapes. Best practices for model development are to have multiple radiologists for annotating training datasets and evaluating segmentation. It has also been recommended to consider algorithms that are data-efficient and can input unlabeled data for semi-supervised learning given the bottleneck of human radiologists available for data labeling/annotation.[30]

Computation power is a challenge in training automatic segmentation algorithms particularly with the large size of whole-body PET/CT studies. Jema et al. proposed a method to rapidly identify and segment tumors and gather metabolic data using 2D and 3D convolutional neural networks.[36] The architecture was created to accommodate the large size of whole-body scans, and significant imbalance between the volume of healthy tissue and the tumor load. Their technique achieves computational efficiency by dividing the body into three parts. It performs 2D axial and sagittal slice-by-slice segmentations, and then refines the 2D predictions with region-specific 3D CNNs. The authors reported a mean Dice score of 88.6% using a training dataset of 2266 scans from patients with diffuse large B cell lymphoma (DLBCL) and testing on 1124 scans from patients with follicular lymphoma (FL). In addition to compute resource efficiency, these results suggest that a model trained on one type of lymphoma has the potential to be expanded to other types. Diao et al. also developed an approach toward reducing compute resource needs. By separately training a UNet-based model on PET and CT data, then fusing the evidence together, Diao et al. reduced GPU needs by 50% of models training on PET/CT data together.[37] They also achieved a high level of accuracy with automatic segmentation. In 70% of the test instances they achieved a Dice similarity coefficient greater than 0.85. The authors argue that this approach takes better account of the uncertainty involved in different imaging methods. In this method, they define the evidence of PET and CT as a set of ordered triple probability values. Each voxel is assigned a probability of either belonging to tumor, background, or indeterminate. The contribution to the field is a proposal of the evidence loss function, which is an operation on the ordered triple of probability belonging to background, tumor, and uncertainty. Uncertainty is output directly through the network and is then minimized with the combination of PET and CT information. These two studies exemplify challenges in model development and training particularly related to compute resources and the potential relevance of transfer learning for lymphoma models.

Validation and Testing the Model

There are different approaches to model testing, depending on the level of data availability. Ideally, one would test the model on a fully external data set (eg, different institution, different scanner); however, most of the time, this is not the case since most studies involve working with limited datasets within one institution. A popular way to overcome this limitation in data availability is by cross-validation, either testing exhaustively or non-exhaustively. Exhaustive cross-validation techniques often take a leave-p-out approach, where a number of observations, p, is used to test the model, and all other data is used to train the model. Then, the sets designated for training and testing are changed until all possible combinations of training and testing sets are performed, hence an "exhaustive" cross-validation. Since a new model is trained for each combination, leave-p-out approaches can quickly become computationally intense as the number of combinations scales non-linearly. Non-exhaustive cross-validation techniques approximate exhaustive cross-validation by applying randomness to the process of assigning samples toward testing or training, and by reducing the number of combinations tested. K-fold cross-validation is a common approach, whereby the samples are randomly ordered and divided into a number of sets, k, of equal size; then, the model is trained on k-1 sets and tested on one set. Finally, the process is repeated such that k models are trained and tested. Other approaches to overcome limited data availability for model training and testing include data creation or augmentation that create new samples by image manipulation (eg, image mirroring, rotation). Importantly, the model's success or failure should be carefully measured by proper selection of the figures of merit. Common figures of merit include Dice similarity score and area under receiver operating characteristic curve (AUC); however, the SNMMI AI task force has evidenced a need for even more task-specific measures.[38]

Examples

Technical Evaluation

The first stage of model implementation beyond proof-of-concept is technical evaluation, meaning the "evaluation on specific clinically relevant tasks such as those of detection and quantification using figures of merit that quantify aspects such as accuracy

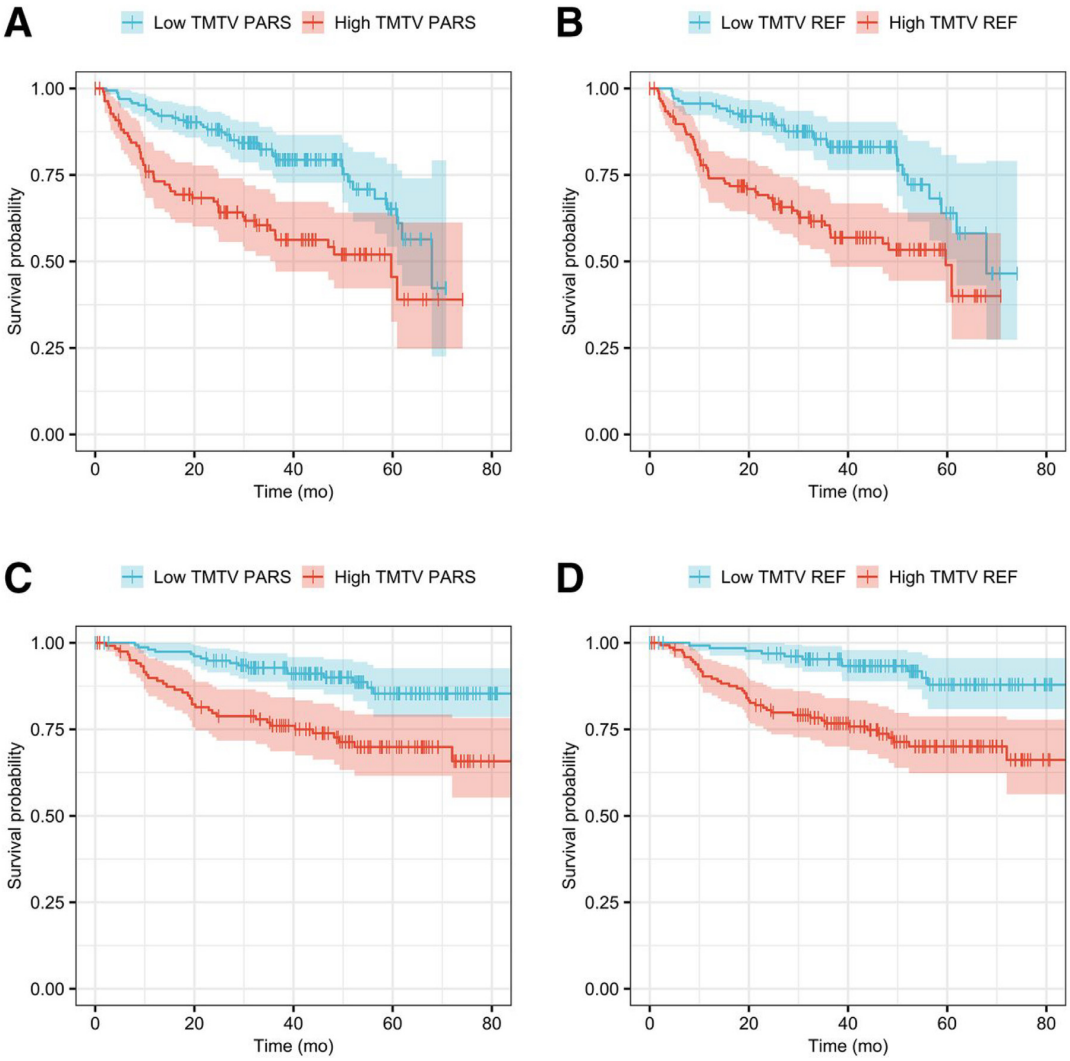

Figure 3 Kaplan-Meier survival curves for PFS (A and B) and OS (C and D). Automated TMTV calculation is predictive of disease prognosis. This research was originally published in JNM. Capobianco et al. Deep-Learning 18F-FDG Uptake Classification Enables Total Metabolic Tumor Volume Estimation in Diffuse Large B-Cell Lymphoma. J Nucl Med. 2021 Jan;62(1):30-6.

and precision.[38]" In one notable example, a study by Capobianco et al., a Siemens-developed TMTV calculator (designated as TMTV PARS) was compared with a reference standard constructed by two nuclear medicine physicians. All else equal, the people with high TMTV had a worse prognosis measured by progression-free survival and overall survival compared with those having a low TMTV. This statement held true in both the reference standard and automatic calculated groups (Fig. 3).[39]

Clinical Evaluation

The best practices for evaluation of clinical tasks as defined by the RELAINCE guidelines state that we should "evaluate the impact of the AI algorithm on making clinical decisions, including diagnostic, prognostic, predictive, and therapeutic decisions for primary endpoints.[38]" Jiang et al. employed a deep learning approach for segmenting diffuse large B-cell lymphoma and used an external validation cohort to demonstrate its efficacy in prognosis. In this study, the authors constructed a 3D U-Net based model trained on randomly sampled patches of PET scans from 297 patients from a local center and tested the model on 117 patients from an external center. Two expert readers used semiautomatic segmentation on PET/CT images to construct the ground truth using adaptive SUVmax thresholding. Based on the available histopathology report and review of other modalities, sites of suspicious uptake such as high intensity normal uptake, inflammation, and infection were verified and eliminated. Only when focal uptake was present were measurements of the spleen, liver, and bone marrow performed.[40] Authors used intensity normalization using nnUNet as proposed by Isensee et al. that normalizes each image by subtracting its mean and then dividing by its standard deviation to give a value of 0 to non-lymphoma voxels.[41] All of the training and testing was performed on PET images and this study did

not use CT images. By comparing the automatic segmentation generated from this algorithm with manual segmentation in training and validation datasets, authors reported a median dice similarity coefficient (DSC) of 0.88 (IQR: 0.77-0.94) and Hausdroff distance of median: 4.12, IQR: 1.00-21.00 in training cohort and median: 0.88, IQR: 0.74-0.93 and median: 2.83, IQR: 1.00-38.10 in validation cohort. No significant difference existed in similarity metrics between training and validation cohorts. The TMTV calculated based on automatic segmentation highly agreed with TMTV from manual segmentation with Bland-Altman biases of -3.0 (-209.2, 203.2 ± 1.96 SD) and -5.4 (-176.8, 166.1 ± 1.96 SD) with 95% CIs in the training and validation cohorts, respectively. Therefore, the authors argue that calculated TMTV held up as a prognostic factor of progression free survival and overall survival.

Current State and Potential Future Role of AI

Diagnosis and Staging

Diagnostic criteria for most hematopoietic and lymphoid tissue malignancies is largely based on histopathologic (morphology, immunohistochemistry and flow cytometry) evaluation of surgically excised specimens.[42] While fine-needle aspiration cytology and core needle biopsy (FNAC/CNB) are often part of the initial evaluation of lymphadenopathy, it has been shown that these methods reach definitive diagnosis in 65%-75% of cases. Moreover, in cases where immediate diagnosis is necessary, FNAC/CNB have shown to fail to provide an actionable diagnosis in a quarter of cases, which further postpones optimal therapy. In regards to FDG avidity, as a result of glucose uptake by proliferating lymphoma cells or accompanying inflammatory cell infiltrates, most lymphomas are FDG-avid. In comparison to anatomical modalities, FDG-PET provides a high diagnostic sensitivity in both HL and NHL for initial detection of disease, as well as detecting extra nodal sites of involvement used for initial staging.[43] While FDG-PET/CT is not currently used as a tool for diagnosis and differentiation of patients with Lymphoma, there has been ongoing research on its potential applications. Sasaki et al. and Yammamoto et al. reported sensitivities of 92-100 and 74%-78% for detection of lymphoma nodal and extra-nodal lesions, respectively.[44,45] Histology type of lymphoma is determinant for degree of FDG uptake, with more aggressive types showing higher avidity.[46] Weiler-Sagi et al. in a study on 766 lymphoma patients classified the lesions as FDG avid and non-avid.[47] FDG-PET showed sensitivity of 100% for Hodgkin's disease, lymphoblastic lymphoma, nodal marginal zone lymphoma, Burkitt's lymphoma, mantle cell lymphoma, and sensitivity of 97% for DLBC and 95% for follicular lymphoma.[47] With regards to staging, the current practices for both HL and NHL are primarily performed according to the Lugano classification.[48] As a modification of the Ann Arbor staging system introduced in 1971, Lugano criteria incorporates FDG-PET/CT results to determine staging of the lymphoma.[48] Isasi et al. in a meta-analysis on staging and restaging of lymphoma using FDG-PET, reported median sensitivity of 90% and 97% for per patient unit and per lesion unit, respectively.[49] FDG-PET has particularly shown higher sensitivity compared to conventional imaging methods for detection of bone marrow and extra-nodal lesions in patients with Hodgkin's lymphoma, resulting in upstaging of these lesions in 15%-25% of cases.[46] Non-FDG avid subtypes of lymphoma are staged primarily based on disease symptoms and anatomical imaging modalities such as CT.[46] While combined FDG-PET/CT provides a great picture of anatomical and functional characteristics of lymphoma that is used with high sensitivity for identifying sites of involvements, interpreting this large amount of functional and anatomical information provides a challenging task for radiologists. Automatic detection algorithms using artificial intelligence can augment radiologist work and greatly contribute to accuracy and speed of analyzing images.

A number of studies have investigated the application of AI-based models applied to FDG-PET images for differentiating lymphoma from other malignancies or even from other subtypes of lymphoma. Yang et al. investigated the utility of ML in assessing pathologic origin of an enlarged lymph node. Using radiomics features identified by a CNN trained on non-medical images, 165 enlarged cervical lymph nodes were classified into lymphomatous and metastatic with an AUC of 0.9. Such classification algorithms can be useful during the management of patients with confirmed diagnoses.[50] For instance, when applied to interim PET imaging, these algorithms could be used to monitor disease histopathological transformation. In a study by de Jesus et al. the authors trained a machine learning-based classifier using radiomic features extracted from PET/CT scans of FL and DLBCL lesions to differentiate between these two subtypes. The radiomics based classifier achieved a high level of accuracy and discriminatory capability, implying that PET/CT can provide useful information beyond staging alone. This type of classifier could potentially aid in the diagnosis and differentiation of NHL subtypes at presentation. This application could also be used at presentation or for monitoring lymphoma over time. For instance, in some cases, FL can transform into DLBCL. This highlights the importance of noninvasive methods to distinguish between these two NHL subtypes in initial stages of assessment.

Radiomic features have differentiated bulky mediastinal lymphomas that were previously undifferentiable by imaging methods.[51] Heterogeneous FDG avidity can manifest in FDG-PET as textural image features that in this study differentiated between classical hodgkin lymphoma (cHL), primary mediastinal B cell lymphoma, and gray zone lymphoma, which historically were not differentiable using traditional image analysis. In this retrospective study comparing lymphoma and sarcoidosis, a radiomics-based machine learning model discriminated between sarcoidosis and lymphoma with very high accuracy equivalent to that of trained radiologists.[52] Similar methods have differentiated a number of different tumor types including: breast carcinoma from breast lymphoma,[53] primary CNS lymphoma from glioblastoma,[54] and SCC from non-hodgkin's lymphoma of

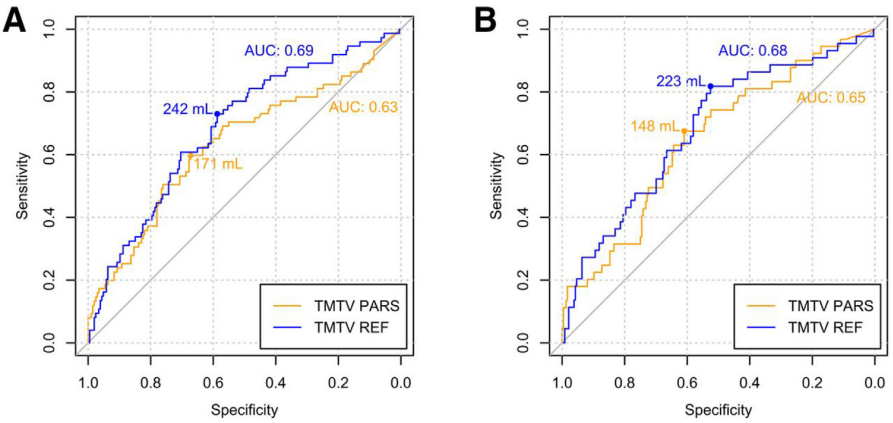

Figure 4 Performance of automated TMTV calculation vs human readers. This research was originally published in JNM. Capobianco et al. Deep-Learning 18F-FDG Uptake Classification Enables Total Metabolic Tumor Volume Estimation in Diffuse Large B-Cell Lymphoma. J Nucl Med. 2021 Jan;62(1):30-6.

oropharynx.[55] These studies highlight the capability and value of AI-based image analysis tools in the physician's diagnostic workflow.

Prognosis

Great heterogeneity exists in prognosis of patients newly diagnosed with lymphomas. Lymphomas are the optimal tumor for which to create imaging-adapted therapy methods due to their FDG avidity and treatment sensitivity. As the primary imaging method for the assessment of FDG-avid lymphomas, FDG-PET/CT imaging enables the pathophysiologic characterization of tumor metabolic activity in the continuous monitoring of lymphoma patients.[56] FDG-PET/CT has shown substantial promise in lymphoma prognostication from early trials. This sparked a multitude of research into interim FDG-PET/CT for informed treatment approaches to lymphomas, notably in HL and DLBCL.[57,58] For FDG-PET/CT-based response evaluation, the Deauville five-point scale introduced in 2009, a reproducible visual scale in the setting of early response assessment, has demonstrated reliability in terms of inter-observer agreement.[27,59,60] Other prognostic tools developed to stratify patient management in lymphoma include the International Prognostic Index (IPI) and its modified forms (R-IPI, NCCN-IPI), TMTV, and Dmax.[61,62] The use of artificial intelligence to the prognosis of lymphoma patients has been investigated in combination with currently available prognostic tools, to facilitate their implementation, and as a stand-alone prognostic marker.

In a study by Capobianco et.al. including 301 individuals from the REMARC trial on DLBCL, the authors compared AI based TMTV with manually estimated TMTV to test whether an automatically calculated TMTV would be relevant to prognosis (Fig. 1). They used a CNN to first identify all FDG-avid regions of interest and classify each region as either suspicious (ie, malignant) or non suspicious (ie, sFEPU). Then, they summed the volumes of all suspicious regions to calculate TMTV. Reference TMTV was manually determined by 2 experienced readers using an independent semi-automated method. Automated TMTV performed well in classification accuracy (85%) (Fig. 4) and in prognosticating progression-free survival and overall survival (Fig. 3). For progression-free survival the hazard ratios were 2.3 and 2.6 for the automated and semi-automated methods, respectively. For overall survival the hazard ratios were 2.8 and 3.7 for the automated and semi-automated methods, respectively.[39] Although immunochemotherapy frequently yields encouraging outcomes in DLBCL patients, studies have indicated that more than 30% of individual's relapse or develop treatment resistance. Such studies, which enable the use of lymphoma prognostic tools, can help patient management identify high-risk individuals early on who may benefit from intensive or novel therapy.

In addition to aiding in the measurement of baseline prognostic indicators that use clinical parameters such as metabolic tumor volume, lactate dehydrogenase, and number of sites of involvement, algorithms may be taught to predict prognosis based on only input from imaging biomarkers or combination of imaging biomarkers and other clinical parameters. In a study Eertink et al. investigated various models including combinations of radiomics properties and existing clinical models including the international prognostic index (IPI)'s ability to predict outcomes following first-line therapy. This study included 317 individuals with DLBCL enrolled in the HOVON-84 trial, where 490 radiomic features were extracted using semi-automated methods, compared and combined with clinical models to form a prognostic model for predicting 2-year time to progression. The authors found that adding radiomics features to IPI score prediction significantly improved identification of patients who are at risk for recurrence compared to using IPI score alone.[63] Despite these improvements, radiomic models have yet to achieve a performance level that is clinically significant.

In a more recent work the same group investigated variation in lesion and feature selection approaches in predicting progression after 2 years. The authors compared different lesion selection (eg,: hottest, largest, aggregate of all lesions)

and feature selection (principal component analysis, factor analysis, univariate selection) methods and compared the predictive value of all models using five-fold cross-validation approach. They concluded that lesion selection and PET feature reduction methods don't impact the model performance significantly and reported that patient level conventional PET features and dissemination features have the highest predictive value among the tested features.[64]

Conclusion

The application of AI methods to PET imaging of lymphoma has significantly increased over the past several years and applies to patient care at all stages, especially prognostication. We have highlighted key developments in AI methods toward improving the clinical care of patients with lymphoma. TMTV quantifies disease burden and is directly related to lymphoma prognosis; however, it is not directly part of the current clinical prognostication models. Our review shows that there is untapped value in TMTV-based disease models. Automatic tumor segmentation has high potential for impacting the lymphoma clinical workflow and is a focus of future development, especially with regard to implementing TMTV measurements.

Box [Terminology]

Training stage	**To train the AI system by matching inputs with predicted outputs using a ground-truth dataset. At this stage, the AI will learn to identify patterns and make predictions using gold-standard data.**
Verification stage	**To evaluate the system's performance at each stage of development to ensure it fulfills all of its defined requirements.**
Validation stage	**To evaluate the performance of the trained models with a validation dataset. Unlike verification testing, validation often occurs after the program has been entirely developed to identify the model with optimal performance.**
Testing stage	**To evaluate the performance of an AI model using an external testing dataset, separate from the training or validation datasets. Provided that the testing dataset is representative of the population, the testing stage aims to produce an unbiased estimate of model performance in the general population.**
Test reliability	**The reliability or consistency with which a test assesses a property of the AI system.**
Test validity	**Validity relates to what trait the test measures and how well it reflects that feature of the AI. Test validity offers information on whether or not the trait being assessed by a test is relevant to the task performed by the AI.**

References

1. Shankland KR, Armitage JO, Hancock BW: Non-Hodgkin lymphoma. Lancet 380:848-857, 2012
2. Connors JM, Cozen W, Steidl C, et al: Hodgkin lymphoma. Nat Rev Dis Primers 6:61, 2020
3. Elenitoba-Johnson KSJ, Lim MS: New insights into lymphoma pathogenesis. Annu Rev Pathol 13:193-217, 2018
4. Armitage JO, Gascoyne RD, Lunning MA, et al: Non-hodgkin lymphoma. Lancet 390:298-310, 2017
5. Lu P: Staging and classification of lymphoma. Semin Nucl Med 35:160-164, 2005
6. Non-Hodgkin Lymphoma - Cancer Stat Facts [Internet]. SEER. Available from: https://seer.cancer.gov/statfacts/html/nhl.html. Accessed October 5, 2022
7. Hodgkin lymphoma - cancer stat facts [Internet]. SEER. [cited Available from: https://seer.cancer.gov/statfacts/html/hodg.html. Accessed October 10, 2022
8. Shimabukuro-Vornhagen A, Haverkamp H, Engert A, et al: Lymphocyte-rich classical Hodgkin's lymphoma: Clinical presentation and treatment outcome in 100 patients treated within German Hodgkin's Study Group trials. J Clin Oncol 23:5739-5745, 2005
9. Meignan M, Itti E, Gallamini A, Younes A: FDG-PET/CT imaging as a biomarker in lymphoma. Eur J Nucl Med Mol Imaging 42:623-633, 2015
10. Miller E, Metser U, Avrahami G, et al: Role of 18F-FDG-PET/CT in staging and follow-up of lymphoma in pediatric and young adult patients. J Comput Assist Tomogr 30:689-694, 2006
11. Rizzo A, Triumbari EKA, Gatta R, et al. The role of 18F-FDG-PET/CT radiomics in lymphoma. Clin Translati Imaging [Internet] 20, 2021. https://doi.org/10.1007/s40336-021-00451-y
12. Toosi A, Bottino AG, Saboury B, et al: A brief history of ai: How to prevent another winter (a critical review). PET Clin 16:449-469, 2021
13. Beegle C, Hasani N, Maass-Moreno R, et al: Artificial intelligence and positron emission tomography imaging workflow: Technologists' perspective. PET Clin 17:31-39, 2022
14. Wu N, Phang J, Park J, et al: Deep neural networks improve radiologists' performance in breast cancer screening. IEEE Trans Med Imaging 39:1184-1194, 2020
15. McKinney SM, Sieniek M, Godbole V, et al: International evaluation of an AI system for breast cancer screening. Nature 577:89-94, 2020
16. Ghorbani A, Ouyang D, Abid A, et al: Deep learning interpretation of echocardiograms. NPJ Digit Med 3:10, 2020
17. Ardila D, Kiraly AP, Bharadwaj S, et al: End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 25:954-961, 2019
18. Huang P, Lin CT, Li Y, et al: Prediction of lung cancer risk at follow-up screening with low-dose CT: A training and validation study of a deep learning method. Lancet Digit Health 1:e353-e362, 2019
19. Shad R, Cunningham JP, Ashley EA, et al: Designing clinically translatable artificial intelligence systems for high-dimensional medical imaging. Nat Machine Intell 3:929-935, 2021
20. Domingues I, Pereira G, Martins P, et al: Using deep learning techniques in medical imaging: A systematic review of applications on CT and PET. Artific Intell Review 53:4093-4160, 2020
21. Oren O, Gersh BJ, Bhatt DL: Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. Lancet Digit Health 2:e486-e488, 2020
22. Hasani N, Paravastu SS, Farhadi F, et al: Artificial intelligence in lymphoma PET imaging: A scoping review (current trends and future directions). PET Clin 17:145-174, 2022
23. Blanc-Durand P, Jégou S, Kanoun S, et al: Fully automatic segmentation of diffuse large B cell lymphoma lesions on 3D FDG-PET/CT for total metabolic tumor volume prediction using a convolutional neural network. Eur J Nucl Med Mol Imaging 48:1362-1370, 2021
24. Weisman AJ, Kieler MW, Perlman SB, et al: Convolutional neural networks for automated PET/CT detection of diseased lymph node burden in patients with lymphoma. Radiol: Artificial Intell 2: e200016, 2020
25. Sadik M, López-Urdaneta J, Ulén J, et al. Artificial intelligence could alert for focal skeleton/bone marrow uptake in Hodgkiń́s lymphoma

patients staged with FDG-PET/CT [Internet]. Research Square. Research Square; 2021. Available from: https://www.researchsquare.com/article/rs-143352/latest.pdf
26. Liu L, Liu J, Nag MK, et al: Improved multi-modal patch based lymphoma segmentation with negative sample augmentation and label guidance on PET/CT Scans. Multiscale Multimodal Med Imag. Springer Nat Switzerland: 121-129, 2022
27. Barrington SF, Mikhaeel 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 32:3048-3058, 2014
28. Meignan M, Cottereau AS, Specht L, et al: Total tumor burden in lymphoma - An evolving strong prognostic parameter. Br J Radiol 94:20210448, 2021
29. Burggraaff CN, Rahman F, Kaßner I, et al: Optimizing workflows for fast and reliable metabolic tumor volume measurements in diffuse large b cell lymphoma. Mol Imaging Biol 22:1102-1110, 2020
30. Bradshaw TJ, Boellaard R, Dutta J, et al: Nuclear medicine and artificial intelligence: Best practices for algorithm development. J Nucl Med 63:500-510, 2022
31. Jemaa S, Paulson JN, Hutchings M, et al: Full automation of total metabolic tumor volume from FDG-PET/CT in DLBCL for baseline risk assessments. Cancer Imaging 22:39, 2022
32. Bi L, Kim J, Kumar A, et al: Automatic detection and classification of regions of FDG uptake in whole-body PET/CT lymphoma studies. Comput Med Imaging Graph 60:3-10, 2017
33. Lartizien C, Rogez M, Niaf E, et al: Computer-aided staging of lymphoma patients with FDG-PET/CT imaging based on textural information. IEEE J Biomed Health Inform 18:946-955, 2014
34. Gu J, Wang Z, Kuen J, et al: Recent advances in convolutional neural networks. Pattern Recognit 77:354-377, 2018
35. Hubel DH, Wiesel TN: Receptive fields and functional architecture of monkey striate cortex. J Physiol 195:215-243, 1968
36. Jemaa S, Fredrickson J, Carano RAD: Tumor segmentation and feature extraction from whole-body FDG-PET/CT using cascaded 2D and 3D convolutional neural networks. J Digit Imaging 33:888-894, 2020
37. Diao Z, Jiang H, Han XH, et al. EFNet: evidence fusion network for tumor segmentation from PET/CT volumes. Phys Med Biol [Internet] 66(20), 2021. https://doi.org/10.1088/1361-6560/ac299a
38. Jha A, Bradshaw T, Buvat I, et al: Best practices for evaluation of artificial intelligence-based algorithms for nuclear medicine: The RELIANCE guidelines. J Nucl Med 63(2):2725, 2022. supplement-2725
39. Capobianco N, Meignan M, Cottereau AS, et al: Deep-learning 18F-FDG uptake classification enables total metabolic tumor volume estimation in diffuse large b-cell lymphoma. J Nucl Med 62:30-36, 2021
40. Jiang C, Chen K, Teng Y, et al: Deep learning-based tumor segmentation and total metabolic tumor volume prediction in the prognosis of diffuse large B-cell lymphoma patients in 3D FDG-PET images. Eur Radiol 32:4801-4812, 2022
41. Isensee F, Jaeger PF, Kohl SAA, et al: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 18:203-211, 2021
42. Campo E, Harris NL: WHO classification of tumours of hematopoietic and lymphoid tissues. Int Agency Res Cancer 586, 2017
43. El-Galaly TC, Gormsen LC, Hutchings M: PET/CT for staging; past, present, and future. Semin Nucl Med 48:4-16, 2018
44. Sasaki M, Kuwabara Y, Koga H, et al: Clinical impact of whole body FDG-PET on the staging and therapeutic decision making for malignant lymphoma. Ann Nucl Med 16:337-345, 2002
45. Yamamoto F, Tsukamoto E, Nakada K, et al: 18p-FDG-PET is superior to67Ga SPECT in the staging of non-Hodgkin's lymphoma [Internet]. Ann Nucl Med 18:519-526, 2004
46. Baba S, Abe K, Isoda T, et al: Impact of FDG-PET/CT in the management of lymphoma. Ann Nucl Med 25:701-716, 2011
47. Weiler-Sagie M, Bushelev O, Epelbaum R, et al: 18F-FDG avidity in lymphoma readdressed: A study of 766 patients. J Nucl Med 51:25-30, 2010
48. Cheson BD, Fisher RI, Barrington SF, et al: Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: The Lugano classification. J Clin Oncol 32:3059-3068, 2014
49. Isasi CR, Lu P, Blaufox MD: A metaanalysis of 18F-2-deoxy-2-fluoro-D-glucose positron emission tomography in the staging and restaging of patients with lymphoma. Cancer 104:1066-1074, 2005
50. Yang Y, Zheng B, Li Y, et al. Computer-aided diagnostic models to classify lymph node metastasis and lymphoma involvement in enlarged cervical lymph nodes using PET/CT. Med Phys [Internet]. 4, 2022. https://doi.org/10.1002/mp.15901
51. Abenavoli EM, Linguanti F, Barbetti M, et al: Machine-Learning approach using FDG-PET-based radiomics in the characterization of mediastinal bulky lymphomas [Internet]. Research Square 2022. Available from https://www.researchsquare.com/article/rs-1330600/latest.pdf. November 19, 2022
52. Lovinfosse P, Ferreira M, Withofs N, et al: Distinction of lymphoma from sarcoidosis at FDG-PET/CT - evaluation of radiomic-feature guided machine learning versus human reader performance. J Nucl Med [Internet] 2022. https://doi.org/10.2967/jnumed.121.263598
53. Ou X, Zhang J, Wang J, et al: Radiomics based on 18 F-FDG-PET/CT could differentiate breast carcinoma from breast lymphoma using machine-learning approach: A preliminary study. Cancer Med 9:496-506, 2020
54. Zhou W, Wen J, Hua F, et al: 18F-FDG-PET/CT in immunocompetent patients with primary central nervous system lymphoma: Differentiation from glioblastoma and correlation with DWI. Eur J Radiol 104:26-32, 2018
55. Mitamura K, Norikane T, Yamamoto Y, et al: Texture indices of 18F-FDG-PET/CT for differentiating squamous cell carcinoma and non-hodgkin's lymphoma of the oropharynx. Acta Med Okayama 75:351-356, 2021
56. Van Heertum RL, Scarimbolo R, Wolodzko JG, et al: Lugano 2014 criteria for assessing FDG-PET/CT in lymphoma: an operational approach for clinical trials. Drug Des Devel Ther 11:1719-1728, 2017
57. Hutchings M, Loft A, Hansen M, et al: FDG-PET after two cycles of chemotherapy predicts treatment failure and progression-free survival in Hodgkin lymphoma. Blood 107:52-59, 2006
58. Gallamini A, Hutchings M, Rigacci L, et al: Early interim 2-[^{18}f]fluoro-2-deoxy-d-glucose positron emission tomography is prognostically superior to international prognostic score in advanced-stage hodgkin's lymphoma: a report from a joint italian-danish study [internet]. J Clin Oncol 25:3746-3752, 2007
59. Meignan M, Gallamini A, Meignan M, et al: Report on the First International Workshop on Interim-PET-Scan in Lymphoma. Leuk Lymphoma 50:1257-1260, 2009
60. Barrington SF, Kirkwood AA, Franceschetto A, et al: PET/CT for staging and early response: results from the response-adapted therapy in advanced hodgkin lymphoma study. Blood 127:1531-1538, 2016
61. Ruppert AS, Dixon JG, Salles G, et al: International prognostic indices in diffuse large B-cell lymphoma: a comparison of IPI, R-IPI, and NCCN-IPI. Blood 135:2041-2048, 2020
62. Cottereau AS, Nioche C, Dirand AS, et al: 18F-FDG-PET dissemination features in diffuse large b-cell lymphoma are predictive of outcome. J Nucl Med 61:40-45, 2020
63. 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 Imaging 49:932-942, 2022
64. 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 Imaging [Internet] 2022. https://doi.org/10.1007/s00259-022-05916-4
65. Coskun N, Okudan B, Uncu D, et al: Baseline 18F-FDG-PET textural features as predictors of response to chemotherapy in diffuse large B-cell lymphoma. Nucl Med Commun 42:1227-1232, 2021
66. Yuan C, Shi Q, Huang X,], et al: Multimodal deep learning model on interim [18F]FDG-PET/CT for predicting primary treatment failure in diffuse large B-cell lymphoma. Eur Radiol [Internet] 2022. https://doi.org/10.1007/s00330-022-09031-8

67. Frood R, Clark M, Burton C, et al: Utility of pre-treatment FDG-PET/CT-derived machine learning models for outcome prediction in classical Hodgkin lymphoma. Eur Radiol 32:7237-7247, 2022
68. Jiang C, Huang X, Li A, et al: Radiomics signature from [18F] FDG-PET images for prognosis predication of primary gastrointestinal diffuse large B cell lymphoma. Eur Radiol 32:5730-5741, 2022
69. 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 Imaging 49:2902-2916, 2022
70. Ritter Z, Papp L, Zámbó K, et al: Two-year event-free survival prediction in DLBCL patients based on in vivo radiomics and clinical parameters. Front Oncol 12:820136, 2022