ELSEVIER

Seminars in
NUCLEAR
MEDICINE

The Role of AI in Lymphoma: An Update

James Cairns, BMBS, MSc, FRCR,[*,†] Russell Frood, MBChB, PhD, FRCR,[*,†] Chirag Patel, MBBS, FRCR,[†] and Andrew Scarsbrook, BMBS, PhD, FRCR[*,†]

Malignant lymphomas encompass a range of malignancies with incidence rising globally, particularly with age. In younger populations, Hodgkin and Burkitt lymphomas predominate, while older populations more commonly experience subtypes such as diffuse large B-cell, follicular, marginal zone, and mantle cell lymphomas. Positron emission tomography/computed tomography (PET/CT) using [^{18}F] fluorodeoxyglucose (FDG) is the gold standard for staging, treatment response assessment, and prognostication in lymphoma. However, interpretation of PET/CT is complex, time-consuming, and reliant on expert imaging specialists, exacerbating challenges associated with workforce shortages worldwide. Artificial intelligence (AI) offers transformative potential across multiple aspects of PET/CT imaging in this setting.

AI applications in appointment planning have demonstrated utility in reducing nonattendance rates and improving departmental efficiency. Advanced reconstruction techniques leveraging convolutional neural networks (CNNs) enable reduced injected activities of radiopharmaceutical and patient dose whilst maintaining diagnostic accuracy, particularly benefiting younger patients requiring multiple scans. Automated segmentation tools, predominantly using 3D U-Net architectures, have improved quantification of metrics such as total metabolic tumour volume (TMTV) and total lesion glycolysis (TLG), facilitating prognostication and treatment stratification. Despite these advancements, challenges remain, including variability in segmentation performance, impact on Deauville Score interpretation, and standardization of TMTV/TLG measurements. Emerging large language models (LLMs) also show promise in enhancing PET/CT reporting, converting free-text reports into structured formats, and improving patient communication.

Further research is required to address limitations such as AI-induced errors, physiological uptake differentiation, and the integration of AI models into clinical workflows. With robust validation and harmonization, AI integration could significantly enhance lymphoma care, improving diagnostic precision, workflow efficiency, and patient outcomes.

Introduction

Lymphoma encompasses a heterogenous group of malignancies with variable prevalence depending on patient age. In younger patients Hodgkin and Burkitt lymphomas predominate. However, with increasing age, other subtypes such as diffuse large B-cell (DLBCL), follicular, marginal zone and mantle cell lymphomas predominate. The incidence of lymphoma is increasing worldwide.[1] Although the prevalence is higher in high-income countries, lymphoma-associated mortality is disproportionately greater in low-income countries. In the UK, there are around 4900 deaths from non-Hodgkin lymphoma and 310 deaths from Hodgkin lymphoma each year.[2,3]

[^{18}F] fluorodeoxyglucose positron emission tomography/computed tomography (FDG PET/CT) is the gold standard for staging patients with high-grade lymphoma, being superior to CT alone. The Lugano classification is a widely used and validated method of staging lymphoma with stage I disease representing the mildest form involving one nodal group and stage IV the most severe, representing extranodal disease.[4] FDG PET/CT is also widely used for interim and end

*Faculty of Medicine, University of Leeds, Leeds LS2 9JT, England.
†Department of Radiology, St James's University Hospital, Leeds Teaching Hospitals NHS Trust, Leeds, LS9 7TF, England.
Address reprint requests to Andrew Scarsbrook, BMBS, PhD, FRCR, Department of Radiology, St James's University Hospital, Leeds Teaching Hospitals NHS Trust, Leeds, LS9 7TF, England. E-mail: a.f.scarsbrook@leeds.ac.uk

https://doi.org/10.1053/j.semnuclmed.2025.02.007

of treatment response assessment in lymphoma[5] in conjunction with Deauville Score (DS), a widely validated response assessment reporting scale.[6]

In the context of increasing demand for medical imaging with aging populations, there is a worldwide shortage of radiologists and nuclear medicine physicians.[7] This workforce challenge has contributed to increased rates of burnout which can negatively impact patient care. There has been an explosion of research using artificial intelligence (AI) methods in medical imaging over the past few years, with imaging experiencing the greatest increase of AI-enabled medical device regulatory submissions.[8] There are a range of commercially available radiology AI products, however, relatively few target nuclear medicine imaging.

Interpretation of FDG PET/CT scans in patients with lymphoma can be complex, with heterogenous disease throughout the body and potential pitfalls related to physiological and false-positive activity, which requires experience for accurate interpretation. In addition, there is increasing evidence regarding the utility of quantitative measurements derived from FDG PET imaging to predict prognosis and response to treatment in lymphoma. Widespread clinical use of quantitative metrics is not yet established, in part because until recently this required additional, potentially time-consuming evaluation. AI has the potential to transform the way in which PET/CT is performed and interpreted. If used effectively AI could significantly reduce the burden of reporting PET/CT studies for patients with lymphoma while providing additional information which could directly enhance patient management. Figure 1 highlights the areas in which AI could be implemented to improve care of patients with lymphoma undergoing PET/CT investigations.

Appointment Scheduling

AI tools have been developed to facilitate efficient planning of hospital appointments. Several studies have demonstrated the benefits of utilizing machine learning to interrogate the electronic medical record in patients referred for diagnostic investigations and predict the likelihood of non-attendance.[9–11] This insight could prompt additional intervention for those more at risk of not attending, resulting in more effective scheduling and optimized departmental throughput delivering care for patients with lymphoma in a more-efficient manner. Published studies predicting likelihood of appointment attendance have largely employed relatively simple machine learning techniques; the performance of these models could be improved by applying more complex multi-modal AI techniques. Lymphoma is prone to a delay in diagnosis due to the variability in its presentation[12] and increasing efficiency of the investigative pathway particularly FDG PET/CT staging, could improve patient outcome.

Image Reconstruction

PET/CT image reconstruction is an area of intense research activity, with increasing evidence demonstrating real-world utility of more efficient scan acquisition, reduced dose and improved resolution for detection of small lesions. Until relatively recently, ordered-subset expectation maximization (OSEM) reconstruction was the standard for PET/CT image reconstruction, and quantitative and semi-quantitative analysis of conventionally reconstructed PET images were used to validate scoring and grading of treatment response in patients with lymphoma used routinely. More recent reconstruction techniques such as point-spread function (PSF)[13] and Bayesian penalized likelihood (BPL) reconstruction have been developed.[14] The latest generation of silicon photomultiplier enabled scanners also incorporate AI-assisted image reconstruction algorithms.[15] Whilst these developments provide several benefits including increasing sensitivity particularly due to superior signal recovery within small lesions increasing detectability, caution is required because of the nonspecific nature of FDG activity and potential reduced specificity.[16] These

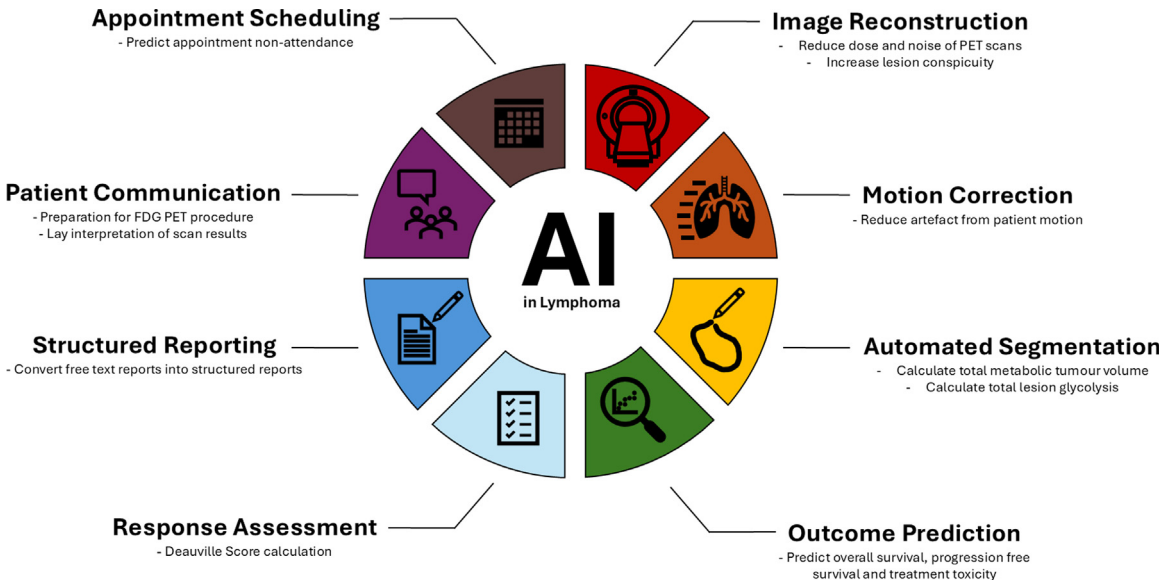

Figure 1 Steps in the PET imaging pathway which can be targeted by AI tools in lymphoma patients.

reconstructions can also augment benign/inflammatory tracer activity and in some circumstances, this could lead to upstaging of Deauville Score in the assessment of treatment response in patients which might impact management decisions.

AI can be utilized to aid dose reduction, improve image quality and result in more efficient scanning and subsequent cost savings, enhancing patient throughput by scanning more patients on a single scanner. It may also have benefits for patients who cannot tolerate the standard duration of scanning acquisition and reduce motion artefacts. A reduction in dose is particularly important for younger patients presenting with lymphoma who may undergo multiple PET scans over their lifetime. Dose reduction requires an efficient solution to deal with increased noise. Convolutional neural networks (CNN) can be trained on matching high and low noise PET/CT images to consistently produce low noise images from noisy data.[17]

Theruvath et al demonstrated in a prospective study that by simulating lower dose FDG PET/MRI scans (by reducing counts) in 20 children and young adults with lymphoma, a CNN could be developed to reduce dose associated with the investigation by 50% with identical sensitivity and specificity of reporting.[18] Similarly, Chaudhari et al. utilised low count whole body PET combined with deep learning to reduce the dose of a PET scan by four-fold. In a multicentre study involving 50 patients from 3 institutions, they demonstrated maintained diagnostic image quality in a variety of conditions including lymphoma.[19] Earlier attempts at low dose PET/CT imaging had demonstrated reduced conspicuity for hypermetabolic lesions and altered SUV measurements, but by utilizing a CNN their model demonstrated equivalence to full dose studies. This model was externally validated simulating real-world situations utilizing a variety of scanners. Katsari et al. utilized the FDA-cleared SubtlePET (Subtle Medical, subtlemedical.com) AI algorithm to reduce image noise and radiation exposure associated with PET imaging in a prospective study with 61 patients and demonstrated noninferiority of AI-processed FDG PET/CT examinations with 66% of the standard dose in a variety of malignancies including lymphoma. Reviewers could tell whether they were viewing the AI generated datasets, but the mean image quality score subjectively evaluated by reviewers for datasets was not significantly different. Only 9 (14.8%) patients in this study had lymphoma and the implications of this algorithm on quantitative metrics was not evaluated. This study incorporated cost analysis in which the departments were able to save 25% on gross annual radiopharmaceutical costs.[20]

Data-driven motion correction (as opposed to hardware motion correction) in PET/CT acquisition has facilitated the use of AI to reduce motion artefact, particularly associated with respiration and subsequently improve image quality.[21] To date, no studies have reported the impact of this technology on patients undergoing PET/CT for lymphoma.

Automated Segmentation

There is an increasing drive to perform quantitative analysis of PET/CT images in lymphoma for prognostication and treatment response assessment. Previously, simple quantitative measurements such as the diameter of the largest tumor lesion or the furthest distance between lesions have been utilized and shown to correlate with prognosis.[22] However, there is growing evidence supporting measurements of total metabolic tumor volume (TMTV) and total lesion glycolysis (TLG) as potentially more accurate measures for predicting prognosis across various malignancies including lymphoma.[23] In lymphoma, these measurements require segmentation of all radiologically evident lymphoma lesions in the body. There is no consensus currently on the most accurate technique for selection/segmentation of lymphoma lesions on PET/CT scans and currently manual verification of all lesions detected and segmented is advised for clinical use.

Manual segmentation of lymphoma lesions through delineation of regions of interest (ROI) by expert clinicians is often clinically impractical due to its time-consuming nature. This has necessitated development of semi-automated and automated segmentation techniques. Developing automated methods for lymphoma detection and segmentation is challenging due to the highly variable distribution, shape and volume of lymphoma lesions. Additionally, normal physiological FDG uptake and clearance can result in SUV measurements similar to lymphomatous lesions and have to be manually excluded during the segmentation process. Challenges may also arise with automated segmentation techniques when there is diffuse lymphomatous involvement of the liver and/or spleen.

Given the complexity and time required for segmentation of lesions to accurately determine TMTV and TLG, surrogate measures have been proposed. Girum et al. utilized a CNN with only sagittal and coronal PET maximum-intensity projection reconstruction images as a surrogate to full lesion segmentation in the calculation of TMTV in 382 patients with DLBCL.[24] This tool demonstrated efficacy as a prognostic biomarker being correlated to TMTV calculated from manual segmentation and demonstrating similar hazard ratios to TMTV from manual segmentation in predicting progression free survival. Similarly, Yousefirizi et al. effectively utilized MIP images to assess risk of relapse/disease progression using end of treatment scans in 31 patients with primary mediastinal large B cell lymphoma (PMBCL).[25]

Multiple semi-automated and automated segmentation techniques have evolved for TMTV and TLG measurement including threshold-based methods (eg. SUV_{max} >2.5 or 4), region-growing or ROI dependent methods (e.g. segment using 40% or 50% of SUV_{max} in ROI). This is an area where the introduction of convolutional neural networks (CNNs) has produced clear tangible benefit. CNNs use deep learning methods to learn the hierarchy of relevant features from provided training data. The open-source 3D U-Net architecture has been the most utilized class of CNN evaluated for segmenting lymphoma lesions on PET/CT so far. This has a distinctive U-shaped design including an encoder path for feature extraction and a decoder path for localization (Fig. 2). U-Nets enable pixel-level classification by incorporating skip connections that link corresponding encoder and decoder layers, preserving spatial details. Performance can be

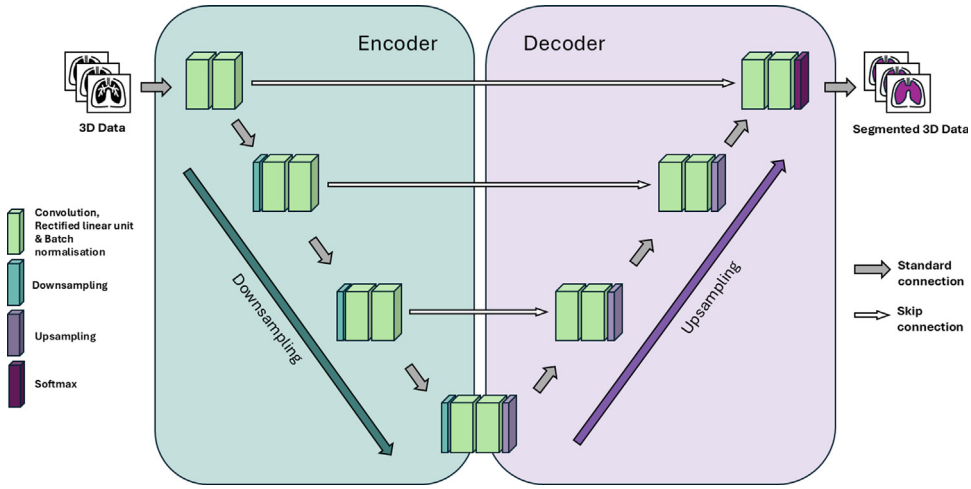

Figure 2 3D U-Net Architecture utilized for automated segmentation.

enhanced by modifying the architecture with additional layers and this has demonstrated impressive results.[26]

The dice similarity coefficient (DSC) is a commonly utilized measure of segmentation performance when compared to a defined ground truth (typically semi-automated/manual segmentation of lesions by an expert reader). Several CNNs such as that reported by Sibille et al. have been developed to identify suspicious lesions in patients with lymphoma.[27] Subsequent efforts have focused on using these techniques to calculate TMTV and TLG. Capobianco et al utilized a CNN for TMTV estimation in DLBCL with similar performance (DSC 0.73 (Interquartile range (IQR) 0.33-0.86)). This CNN-predicted TMTV correlated with progression-free survival (PFS) and overall survival (OS), with hazard ratios of 2.3 (95% Confidence Interval (CI): 1.5-3.6) and 2.8 (95% CI: 1.6-5.1) respectively in 301 patients.[28]

Blanc Durand et al utilized a 3D U-Net architecture to create an automated segmentation tool for TMTV calculation in DLBCL, trained on 639 patients and validated on 94 patients with a DSC of 0.73 (Standard deviation (SD) ± 0.2).[29] It incorporated an additional processing layer in which there is concatenation of the PET and CT segmentation data to improve the model performance. However, this model underestimated TMTV by 20.8% in the validation cohort and this was statistically significant. This underestimation has emerged as a common problem with AI-based segmentation demonstrating poor precision in small lymphoma lesions with limited stage disease and/or small lesions. Huang et al also utilized a 3D U-net architecture with concatenated PET/CT data in 173 patients with lymphoma, but utilized Dempster-Shafer theory to deal with discrepancies between the PET and CT data regions of interest and this improved the segmentation performance of the 3D U-net architecture, particularly for segmentation of small lymphoma lesions (DSC 0.64, SD ± 0.02).[30] By adding a squeeze and excitation model to the 3D U-net architecture, Yousefirizi et al. achieved a DSC of 0.77 (SD ± 0.08) in a multicentre study of 194 DLBCL and PMBCL patients.[31]

Yousefirizi et al. subsequently developed TMTV-Net, a fully automated TMTV segmentation tool utilizing 3D U-Net architecture with additional cascaded refinement. They added a test-time augmentation cascade to the 3D U-Net architecture to enhance prediction robustness and a soft voting cascade to better manage model uncertainty. Compared to the previously described Bland Durand et al and Huang et al 3D[32] U-net models these cascades improved performance in a multi-site cohort of 517 patients with DLBCL and primary mediastinal B cell lymphoma with an overall DSC 0.66 (SD ± 0.16).[33] The downside of the model was that the training data included a variety of different malignancies in addition to lymphoma, which may have impacted its effectiveness in accurately identifying lymphoma lesions.

Karimdjee et al. evaluated a commercial CNN-based segmentation tool (AI lesion detector) implemented within Syngo.via (Siemens Healthineers, siemens-healthineers.com) for automated TMTV and TLG measurement in 51 patients with DLBCL, evaluating inter-observer agreement and impact on reporting time in clinical practice.[34] They reported a possible reduction in time to calculate TMTV and TLG in clinical practice when compared to semi-automated threshold methods of segmentation with excellent inter-observer agreement.

Outcome Prediction

Accurate outcome prediction in lymphoma provides valuable prognostic information for clinicians and patients. There is also the future potential to stratify patient treatment and follow up with the goal being to improve outcomes and quality of life. The use of AI segmentation, as described previously, allows extraction of quantitative metrics relating to spread of disease, metabolic activity of lesions, density and texture of lesions and underlying body composition of the patient.[35] These models can be formed as part of a single neural network, outputted as a readable result for analysis and interpretation by the clinician or combined with other modelling/machine learning techniques.

One of the largest studies assessing semi-quantitative analysis as part of outcome prediction, performed by Mikhaeel et al., assessed the predictive ability of TMTV in DLBCL in 1241 patients derived from five different published studies.[36]

They demonstrated TMTV, segmented with a >4 SUV threshold, to be superior to the currently used clinically based International Prognostic Index (IPI) in predicting PFS and OS. The study created an International Metabolic Prognostic Index model consisting of TMTV, age and stage which also outperformed IPI and was superior at defining both PFS and OS high-risk groups. This model can be applied to real-world data with a simple CSV tool provided as part of their research output allowing computation of model coefficients. Cottereau et al. reported that MTV derived from >2.5 SUV thresholding in 258 patients was significant in predicting 5-year PFS and OS in early-stage HL.[37] Frood et al. in a mix of 289 early and advanced stage HL patients demonstrated the ability of TMTV derived using two different segmentation thresholds (>4.0 SUV and >1.5 times mean liver SUV) to predict 3-year PFS and OS.[38]

The largest study looking at distribution of disease based on distance between the two farthest PET-avid lesions (Dmax) was performed by Girum et al. in 382 patients.[24] Their study demonstrated the significant ability of Dmax calculated by CNN based segmentation from coronal and sagittal PET MIP images to predict PFS and OS in DLBCL, although confidence intervals were relatively large and the area under the curve (AUC) of time-dependent receiver operating characteristics (ROC) for OS of the test dataset was only 0.5. Durmo et al. demonstrated the ability of Dmax to predict PFS in 155 HL patients, using 20 cm as a cutoff value, to be independently associated with PFS (HR = 2.70, 95% CI 1.1-6.63, *P*-value = 0.03).[39] The use of radiomics based modelling has also been demonstrated as being able to predict PFS and OS in both DLBCL and HL. However, there is still currently a lack of repeatability. Carlier et al demonstrated an ROC AUC of 0.62 ± 0.07 for 2-year PFS in 545 patients with DLBCL using a model derived from radiomic features (RFs) extracted from the largest lesion, although RFs did not outperform clinical features alone.[40] The study by Frood et al. showed a ridge regression model using RFs derived from a 1.5x mean liver SUV segmentation threshold had the highest test AUC for 2-year event free survival (EFS) in HL at 0.81 ± 0.12.[38] However, there was no significant difference when compared to a logistic regression model derived from MTV alone.

Body composition analysis (Fig. 3) offers the opportunity to extract additional information from PET/CT indicating a patient's general frailty and may provide an indication of prognosis.[41] Presence of sarcopenia is one of the most reported body composition metrics in the literature. The largest study in lymphoma to date by Xiao et al demonstrated that the presence of baseline sarcopenia was associated with increased neutropenia hospitalization adjusted Odds Ratio 1.64 (95% CI 1.01-2.65) and inability to complete the standard number of treatment cycles (1.49, 95% CI 1.02-2.16).[42] Guo et al. demonstrated that the skeletal muscle gauge, a metric calculated from skeletal muscle area adjusted for height and muscle density was a predictor of toxicity (HR = 1.11, 95% CI, 1.04-1.18), PFS (2.889; 95% CI, 1.401-5.959) and OS (2.655; 95% CI, 1.218-5.787) in 201 patients with DLBCL.[43] Besutti et al. reported that reduced skeletal muscle density/increased intramuscular fat rather than skeletal area adjusted for height was associated with poor OS and PFS, HR = 1.35; CI = 1.03-1.7; *P* = 0.03, and HR = 1.30; CI = 1.04-1.64; *P* = 0.024, respectively.[44] These studies all used measurements taken at the level of L3.

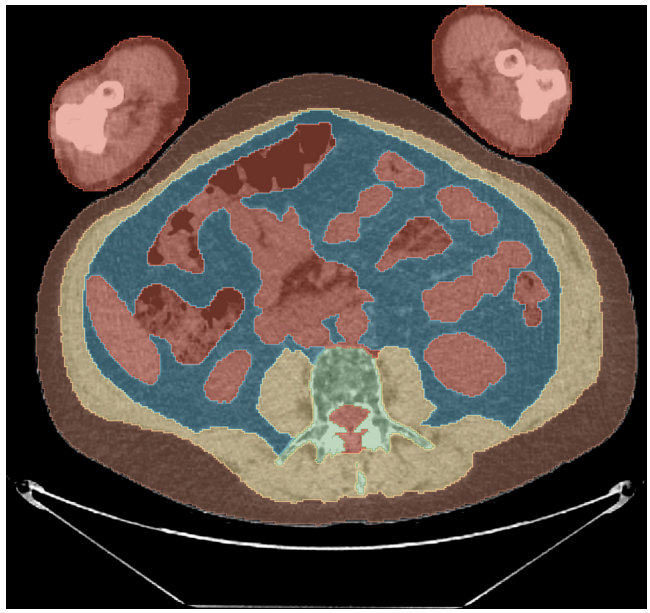

Figure 3 Example of AI-driven body composition analysis. Lymphoma PET-CT (CT component) at the level of the L3 vertebral body. Labels show adipose tissue (subcutaneous light brown; visceral blue), muscle (yellow), bone (green) and miscellaneous soft tissue (red) volumes. Figure provided by Chris Winder, PhD Fellow, University of Leeds.

One of the limitations when trying to evaluate the literature and for the community to build a body of evidence surrounding outcome prediction is the heterogeneity in data. There are variations in outcome measures assessed, the stages of disease included in analysis, definitions for metrics and thresholds used to segment lesions, treatment and follow up regimes used and differences in the populations studied. A consensus on definitions of thresholds, metrics and appropriate outcome measures needs to be reached and a large well planned multicentre, ideally federated, study needs to be designed.

Response Assessment

Accurate response assessment is vital in lymphoma as it frequently influences patient management and helps guide treatment adaption. Use of the Deauville Score is the current standard of care, evaluating any residual lymphomatous disease activity relative to background liver and mediastinal blood pool uptake to determine the degree of metabolic response[6]. End-of-treatment scans have routinely been performed for most lymphoma subtypes. More recently, the use of interim PET/CT, specifically during chemotherapy for Hodgkin's lymphoma has been formally incorporated into treatment guidelines.

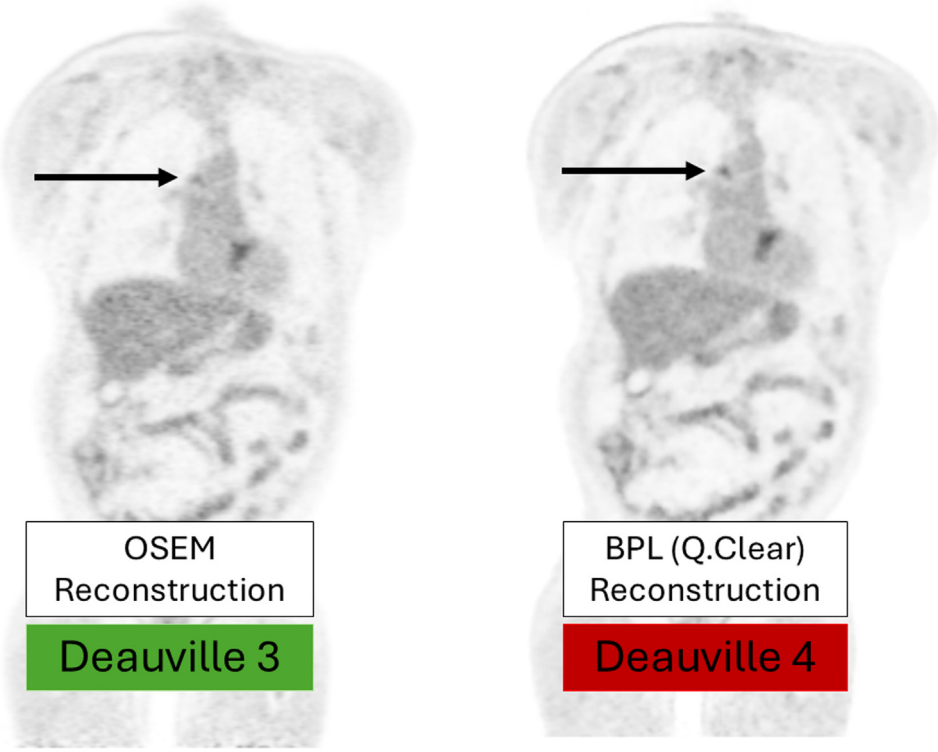

Figure 4 A 67-year-old man with relapsed Hodgkin's lymphoma treated with 3 cycles of salvage chemotherapy. 18F-FDG PET images in the coronal plane highlighting a residual right paratracheal lymph node (arrow) which changed from Deauville score 3 to 4 between ordered subset expectation maximization (OSEM) (left) and Bayesian penalized likelihood (BPL) Q.Clear (right) reconstructions.

During the validation of Deauville Score guidelines, images were acquired using OSEM image reconstruction. However, new image reconstruction technologies, including AI methods, may cause interpretative challenges for application of the Deauville Score. For example, the Q.Clear (GE Healthcare) reconstruction algorithm has been shown in some studies to upgrade the Deauville Score from 3 to 4 (Fig. 4), potentially leading to treatment escalation with increased risk of treatment toxicity without clear evidence of benefit.[45,46] If AI derived techniques of image reconstruction are to be utilized in clinical practice their impact on response assessment must be carefully evaluated.

Development of AI methods have largely focused on tumor segmentation and prognosis prediction at a single time point but have not addressed the evolution of disease or treatment response that can lead to immediate treatment decisions. Few have utilized CNN based automated TMTV calculation for pretreatment, interim and end of treatment scans to provide a more accurate assessment of treatment response. Sadik et al. demonstrated the efficacy of CNN-based automated mediastinal blood pool and liver SUV_{max} measurement to facilitate more efficient Deauville Score calculation in 80 patients with lymphoma. This model demonstrated good agreement with manual measurements by experienced imaging specialists.[47] Jemaa et al. demonstrated the efficacy of an end-to-end model to calculate TMTV and treatment response (utilizing Deauville Score) from pretreatment, interim and end of treatment PET/CT scans in patients with FDG-avid non-Hodgkin Lymphoma. This retrospective multicentre study validated this CNN-based model on 678 patients from 3 clinical trials involving over 90 centers. It demonstrated excellent observer agreement (85%-93%) compared to adjudicated responses from an independent review committee.[48] Further research is required to refine automated response assessment methods and demonstrate their clinical utility.

Enhanced Reporting

AI has the potential to improve the accuracy and efficiency of PET/CT interpretation in lymphoma allowing the reporter to focus on placing imaging findings in context for clinical colleagues.

By utilizing automated segmentation techniques described earlier, CNNs can be utilized to predict the likelihood that a detected lesion is suspicious for cancer or not. For example, Sibille et al. developed a CNN with automated lesion detection for use in lymphoma and lung cancer, which in tandem predicted the likelihood that a detected lesion was malignant and achieved excellent performance in lymphoma with an AUROC curve >0.95 in a 600-patient cohort.[27]

Combining automated segmentation with TMTV and TLG calculation should allow reporting clinicians to confer important additional information regarding the extent of disease beyond conventional staging using the Lugano classification. Frood et al demonstrated that the use of an automated FDG

PET/CT segmentation tool in high-grade lymphoma reduced the time taken to report studies without reducing report quality in reporters with a range of experience.[49]

Psychological implications of implementing AI tools in the interpretation of complex scans needs to be considered. Like humans, AI is prone to error and the extra layer of explainability associated with some AI tools can hinder and persuade physicians to follow incorrect suggestions.[50]

Large Language Models (LLMs) and Multi-Modal Models

LLMs particularly generative pretrained transformer (GPT)-based models have gained popularity for their ability to provide fast and accurate written answers to diverse questions. They have potential to be utilized in a variety of medical subfields and given that diagnostic radiology exists in predominantly a digital format, it lends itself well to the implementation and evaluation of LLM solutions.

Structured reporting in radiology is emerging as a necessity, with societies such as the European Society of Radiology (ESR) and the Radiological Society of North America (RSNA) promoting its adoption in clinical settings. It may improve the radiological workflow, allowing more efficient communication among physicians.[51] There may be resistance to the implementation of structured reporting due to the time-consuming nature of completing a structured report when compared to free text reports. Additionally, some express concerns regarding decreased diagnostic accuracy and completeness due to constraints of structured reports.[52]

Despite guidelines issued by radiological societies on the desired information to be issued in PET/CT reports, a study in 2010 demonstrated that a minority of information deemed desirable was included in clinical reports.[53] With increasing complexity of treatment regimens, accurate staging and conveying this information appropriately is vital. By utilizing structured reports in primary staging of diffuse large B-cell lymphoma Schoeppe et al. demonstrated a perceived significant improvement across many domains by hematologists such as comprehensibility, quicker information extraction and easier classification of staging, with all four of the hematologists involved in this study preferring structured reports over free text reports.[52] By leveraging LLMs free-text reports can effectively be converted into structured reports with the intended benefit of improving the quality of reports while alleviating the concerns regarding the time-consuming nature of structured reports.[51] Multiple studies have demonstrated the effectiveness of LLMs in summarizing radiology reports[54-56] including a study which utilized this for PSMA PET/CT studies.[57] Huemann et al. utilized a bidirectional encoder representation from transformers (BERT) language model which showed promise in accurately predicting Deauville Scores from 4542 PET/CT reports.[58]

Utilizing LLMs to improve the interaction of patients with the radiology department is another area of promise. LLMs can generate layperson or colloquial radiology reports to enhance understanding by patients[59] and Rogasch et al. demonstrated that ChatGPT (Open AI, chatgpt.com) could assist in preparing patients for FDG PET/CT studies and explain the reports of these scans. However, in this study, the LLM showed potential to cause confusion and/or harm through incorrect responses and hallucinations. Further improvement in consistency of responses is required to implement this into a clinical setting.[60] The recent evolution of LLMs into more generalist AI models has enabled them to process multi-modal inputs, such as text, images, and laboratory results, by incorporating additional neural networks. This could have transformative potential for medical imaging, including in patients with lymphoma, with the potential to streamline the reporting workflow by automating comparison with previous studies, summarizing pertinent clinical data from electronic health records and producing an interactive AI-assisted report for review and evaluation by clinicians.[61,62] However, there are several hurdles to achieving this, including poor performance in interpretation of 3D data, challenges combining image and text data and establishing business cases for their use.[63] There is a question of liability with large language and multi-modal models which needs to be addressed. Whilst these show considerable promise and there is extensive commercial interest, no healthcare-related LLMs or multi-modal models have been granted regulatory approval at the time of writing.

Current Limitations of AI in Lymphoma

Deployment of AI within the nuclear medicine department and to assist the management of patients with lymphoma has clear potential. However, there are several key areas which must be addressed through further research. Utilizing AI-driven PET/CT image reconstruction warrants more robust evaluation of the impact on patient management/outcome, due to divergence from established clinical practice based on high quality prospective trials employing older reconstruction algorithms from which Deauville Score assessment criteria were derived and utilized extensively in segmentation methods. All AI models (predominantly CNNs using the U-Net framework) performing automated segmentation still rely on end user validation for regions of interest and the removal of physiological uptake. Further work should be undertaken to implement a model which can more extensively target differentiation between physiological and pathological uptake. Additionally, the development of validation data for automated segmentation is vital to facilitating clinical implementation of measurements such as TMTV and TLG which could improve the prognostication and treatment stratification of patients with lymphoma. LLMs also require refinement with a reduction in inaccuracies and hallucination episodes to ensure reliability and clinical translation.

Conclusion

The integration of artificial intelligence (AI) into PET/CT imaging for lymphoma offers significant potential to enhance diagnostic accuracy, streamline workflows, and improve

patient outcomes. AI tools can optimize appointment scheduling, image reconstruction, segmentation, and reporting, addressing workforce shortages and increasing efficiency. However, challenges such as variability in segmentation performance, impact on established frameworks like the Deauville Score, and the reliability of large language models should be addressed. Future efforts should focus on robust validation and harmonization to ensure AI's safe and effective implementation, paving the way for more precise and personalized care in lymphoma management.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Professor Andrew Scarsbrook reports financial support was provided by Cancer Research UK. Professor Andrew Scarsbrook reports financial support was provided by Leeds Biomedical Research Centre. Dr Russell Frood reports financial support was provided by Cancer Research UK. Dr James Cairns reports financial support was provided by National Institute for Health and Care Research. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

James Cairns: Writing – review & editing, Writing – original draft. **Russell Frood:** Writing – review & editing, Writing – original draft. **Chirag Patel:** Writing – review & editing. **Andrew Scarsbrook:** Writing – review & editing.

Acknowledgments

A.S. is part-supported by Cancer Research UK (C19942/A28832) and NIHR Leeds Biomedical Research Centre (NIHR203331). J.C is funded by an NIHR Academic Clinical Fellowship. R.F is funded by a Cancer Research UK postdoctoral fellowship (RCCCTF-Oct22/100002). The views expressed are those of the authors and not necessarily those of Cancer Research UK, the NIHR Leeds BRC, or the Department of Health and Social Care. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Huang J, Pang WS, Lok V, et al: Incidence, mortality, risk factors, and trends for Hodgkin lymphoma: A global data analysis. J Hematol Oncol 15. https://doi.org/10.1186/s13045-022-01281-9, 2022
2. Cancer Research UK. Hodgkin lymphoma incidence statistics 2014. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/hodgkin-lymphoma (accessed January 29, 2025).
3. Cancer Research UK. Non-Hodgkin lymphoma incidence statistics 2010. http://www.cancerresearchuk.org/cancer-info/cancerstats/types/nhl/incidence/uk-nonhodgkin-lymphoma-incidence-statistics#world (Accessed January 29, 2025).
4. Razek AAKA, Shamaa S, Lattif MA, et al: Inter-Observer agreement of whole-body computed tomography in staging and response assessment in lymphoma: The Lugano classification. Pol J Radiol 82:441-447, 2018. https://doi.org/10.12659/PJR.902370
5. Zeman MN, Akin EA, Merryman RW, et al: Interim FDG-PET/CT for response assessment of lymphoma. Semin Nucl Med 53:371-388, 2023. https://doi.org/10.1053/j.semnuclmed.2022.10.004
6. 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. https://doi.org/10.1200/JCO.2013.53.5229
7. Jeganathan S: The growing problem of radiologist shortages: Australia and New Zealand's perspective. Korean J Radiol 24:1043-1045, 2023. https://doi.org/10.3348/kjr.2023.0831
8. U.S. Food & Drug Administration. Recently-approved devices 2024. https://www.fda.gov/medical-devices/device-approvals-and-clearances/recently-approved-devices (Accessed January 29, 2025).
9. Nelson A, Herron D, Rees G, et al: Predicting scheduled hospital attendance with artificial intelligence. NPJ Digit Med 2. https://doi.org/10.1038/s41746-019-0103-3, 2019
10. Chong LR, Tsai KT, Lee LL, et al: Artificial intelligence predictive analytics in the management of outpatient MRI appointment No-shows. Am J Roentgenol 215:1155-1162, 2020. https://doi.org/10.2214/AJR.19.22594
11. Srinivas S, Ravindran AR: Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: A prescriptive analytics framework. Expert Syst Appl 102:245-261, 2018. https://doi.org/10.1016/j.eswa.2018.02.022
12. Howell DA, Smith AG, Roman E: Lymphoma: variations in time to diagnosis and treatment. Eur J Cancer Care 15:272-278, 2006. https://doi.org/10.1111/j.1365-2354.2006.00651.x
13. Mawlawi O, Townsend DW: Multimodality imaging: An update on PET/CT technology. Eur J Nucl Med Mol Imaging 36:15-29, 2009. https://doi.org/10.1007/s00259-008-1016-6
14. Aide N, Lasnon C, Kesner A, et al: New PET technologies – embracing progress and pushing the limits. Eur J Nucl Med Mol Imaging 48:2711-2726, 2021. https://doi.org/10.1007/s00259-021-05390-4
15. Reader AJ, Pan B: AI for PET image reconstruction. Br J Radiol 96. https://doi.org/10.1259/bjr.20230292, 2023
16. Rogasch JMM, Boellaard R, Pike L, Borchmann P, Johnson P, Wolf J, et al: Moving the goalposts while scoring-the dilemma posed by new PET technologies. Eur J Nucl Med Mol Imaging 48:2696-2710, 2021. https://doi.org/10.1007/s00259-021-05403-2
17. Kaplan S, Zhu YM: Full-dose PET image estimation from low-dose PET image using deep learning: A pilot study. J Digit Imaging 32:773-778, 2019. https://doi.org/10.1007/s10278-018-0150-3
18. Theruvath AJ, Siedek F, Yerneni K, et al: Validation of deep learning–based augmentation for reduced18F-FDG dose for PET/MRI in children and young adults with lymphoma. Radiol Artif Intell 3. https://doi.org/10.1148/ryai.2021200232, 2021
19. Chaudhari AS, Mittra E, Davidzon GA, et al: Low-count whole-body PET with deep learning in a multicenter and externally validated study. NPJ Digit Med 4:1-11, 2021. https://doi.org/10.1038/s41746-021-00497-2
20. Katsari K, Penna D, Arena V, et al: Artificial intelligence for reduced dose 18F-FDG PET examinations: A real-world deployment through a standardized framework and business case assessment. EJNMMI Phys 8. https://doi.org/10.1186/s40658-021-00374-7, 2021
21. Lu Y, Kang F, Zhang D, et al: Deep learning-aided respiratory motion compensation in PET/CT: Addressing motion induced resolution loss,

attenuation correction artifacts and PET-CT misalignment. Eur J Nucl Med Mol Imaging 52:62-73, 2024. https://doi.org/10.1007/s00259-024-06872-x
22. Oh MY, Oh SB, Seoung HG, et al: Clinical significance of standardized uptake value and maximum tumor diameter in patients with primary extranodal diffuse large B cell lymphoma. Kor J Hematol 47:207-212, 2012. https://doi.org/10.5045/kjh.2012.47.3.207
23. Cottereau A-S, Lanic H, Mareschal S, et al: Molecular profile and FDG-PET/CT total metabolic tumor volume improve risk classification at diagnosis for patients with diffuse large B-cell lymphoma. Clin Cancer Res 22:3801-3809, 2016. https://doi.org/10.1158/1078-0432.CCR-15-2825
24. Girum KB, Rebaud L, Cottereau AS, et al: 18F-FDG PET maximum-intensity projections and artificial intelligence: A win-win combination to easily measure prognostic biomarkers in DLBCL patients. J Nucl Med 63:1925-1932, 2022. https://doi.org/10.2967/jnumed.121.263501
25. Yousefirizi F, Savage K, Sehn L, et al: Prognostic role of end-of-treatment FDG PET in primary mediastinal large B-cell lymphoma: Application of deep neural network for segmentation-free progression prediction. J Nucl Med 65(Supplement 2), 2024
26. Hellwig D, Hellwig NC, Boehner S, et al: Artificial intelligence and deep learning for advancing PET image reconstruction: State-of-the-art and future directions. Nuklearmedizin - NuclearMedicine 62:334-342, 2023. https://doi.org/10.1055/a-2198-0358
27. Sibille L, Seifert R, Avramovic N, et al: 18F-FDG PET/CT uptake classification in lymphoma and lung cancer by using deep convolutional neural networks. Radiology 294:445-452, 2020. https://doi.org/10.1148/radiol.2019191114
28. 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. https://doi.org/10.2967/jnumed.120.242412
29. 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 tumour volume prediction using a convolutional neural network. Eur J Nucl Med Mol Imaging 48:1362-1370, 2021. https://doi.org/10.1007/s00259-020-05080-7
30. Huang L, Denoeux T, Tonnelet D et al: Deep PET/CT fusion with Dempster-Shafer Theory for Lymphoma Segmentation. In: LIan C, Cao X, Rekik I, Xu X, Yan P (eds). Machine Learning in Medical Imaging. MLMI 2021. Lecture Notes in Computer Science, Vol 12966. Springer, Cham. https://doi.org/10.1007/978-3-030-87589-3_4
31. Yousefirizi F, Dubljevic N, Ahamed S, et al: Convolutional neural network with a hybrid loss function for fully automated segmentation of lymphoma lesions in FDG PET images. In: Proceedings SPIE 12032, Medical Imaging; 2022. https://doi.org/10.1117/12.26126752022
32. Huang L, Ruan S, Decazes P, Denúux T: Lymphoma segmentation from 3D PET-CT images using a deep evidential network. Int J Approx Reasoning 149:39-60, 2022. https://doi.org/10.1016/j.ijar.2022.06.007
33. Yousefirizi F, Klyuzhin IS, JH O, et al: TMTV-Net: fully automated total metabolic tumor volume segmentation in lymphoma PET/CT images — A multi-center generalizability analysis. Eur J Nucl Med Mol Imaging 51:1937-1954, 2024. https://doi.org/10.1007/s00259-024-06616-x
34. Karimdjee M, Delaby G, Huglo D, et al: Evaluation of a convolution neural network for baseline total tumor metabolic volume on [18F]FDG PET in diffuse large B cell lymphoma. Eur Radiol 33:3386-3395, 2023. https://doi.org/10.1007/s00330-022-09375-1
35. Albano D, Ravanelli M, Durmo R, et al: Semiquantitative 2-[18F]FDG PET/CT-based parameters role in lymphoma. Front Med 11. https://doi.org/10.3389/fmed.2024.1515040, 2024
36. Mikhaeel NG, Heymans MW, Eertink JJ, et al: Proposed new dynamic prognostic index for diffuse large B-cell lymphoma: International metabolic prognostic Index. J Clin Oncol 40:2352-2360, 2022. https://doi.org/10.1200/JCO.21
37. Cottereau A, Versari A, Loft A, et al: Prognostic value of baseline metabolic tumor volume in early-stage Hodgkin lymphoma in the standard arm of the H10 trial. Blood 131:1456-1463, 2018. https://doi.org/10.1182/blood-2017-07-795476
38. 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. https://doi.org/10.1007/s00330-022-09039-0
39. Durmo R, Donati B, Rebaud L, et al: Prognostic value of lesion dissemination in doxorubicin, bleomycin, vinblastine, and dacarbazine-treated, interimPET-negative classical Hodgkin Lymphoma patients: A radio-genomic study. Hematol Oncol 40:645-657, 2022. https://doi.org/10.1002/hon.3025
40. Carlier T, Frécon G, Mateus D, et al: Prognostic value of 18F-FDG PET radiomics features at baseline in PET-guided consolidation strategy in diffuse large B-cell lymphoma: A machine-learning analysis from the GAINED study. J Nucl Med 65:156-162, 2024. https://doi.org/10.2967/jnumed.123.265872
41. Shah UA, Ballinger TJ, Bhandari R, et al: Imaging modalities for measuring body composition in patients with cancer: Opportunities and challenges. J Natl Cancer Inst Monogr 61:56-67, 2023. https://doi.org/10.1093/jncimonographs/lgad001
42. Xiao DY, Luo S, O'Brian K, et al: Impact of sarcopenia on treatment tolerance in United States veterans with diffuse large B-cell lymphoma treated with CHOP-based chemotherapy. Am J Hematol 91:1002-1007, 2016. https://doi.org/10.1002/ajh.24465
43. Guo J, Cai P, Li P, et al: Body composition as a predictor of toxicity and prognosis in patients with diffuse large B-cell lymphoma receiving R-CHOP immunochemotherapy. Cur Oncol 28:1325-1337, 2021. https://doi.org/10.3390/curroncol28020126
44. Besutti G, Massaro F, Bonelli E, et al: Prognostic impact of muscle quantity and quality and fat distribution in diffuse large B-cell lymphoma patients. Front Nutr 8. https://doi.org/10.3389/fnut.2021.620696, 2021
45. Genc M, Yildirim N, Coskun N, et al: The variation of quantitative parameters and deauville scores with different reconstruction algorithms in FDG PET/CT imaging of lymphoma patients. Revista Española de Medicina Nuclear e Imagen Molecular (English Edition) 42:388-392, 2023. https://doi.org/10.1016/j.remnie.2023.07.006
46. Korsholm K, Overbeck N, Dias AH, et al: Impact of reduced image noise on Deauville scores in patients with lymphoma scanned on a long-axial field-of-view PET/CT-scanner. Diagnostics 13:947, 2023. https://doi.org/10.3390/diagnostics13050947
47. Sadik M, Lind E, Polymeri E, et al: Automated quantification of reference levels in liver and mediastinal blood pool for the Deauville therapy response classification using FDG-PET/CT in Hodgkin and non-hodgkin lymphomas. Clin Physiol Funct Imaging 39:78-84, 2019. https://doi.org/10.1111/cpf.12546
48. Jemaa S, Ounadjela S, Wang X, et al: Automated Lugano metabolic response assessment in 18F-fluorodeoxyglucose—Avid non-hodgkin lymphoma with deep learning on 18F-fluorodeoxyglucose—Positron emission tomography. J Clin Oncol 42:2966-2977, 2024. https://doi.org/10.1200/JCO.23.01978
49. Frood R, Willaime JMY, Miles B, et al: Comparative effectiveness of standard vs. AI-assisted PET/CT reading workflow for pre-treatment lymphoma staging: A multi-institutional reader study evaluation. Front Nucl Med 3:1-11, 2023. https://doi.org/10.3389/fnume.2023.1327186
50. Rahmim A, Bradshaw TJ, Buvat I, et al: The Bethesda Report (AI Summit 2022). https://doi.org/10.48550/arXiv.2211.03783, 2022 ArXiv
51. Mallio CA, Sertorio AC, Bernetti C, et al: Large language models for structured reporting in radiology: Performance of GPT-4, ChatGPT-3.5, perplexity and Bing. Radiol Med 128:808-812, 2023. https://doi.org/10.1007/s11547-023-01651-4
52. Schoeppe F, Sommer WH, Nörenberg D, et al: Structured reporting adds clinical value in primary CT staging of diffuse large B-cell lymphoma. Eur Radiol 28:3702-3709, 2018. https://doi.org/10.1007/s00330-018-5340-3
53. Coleman RE, Hillner BE, Shields AF, et al: PET and PET/CT reports: observations from the National oncologic PET Registry. J Nucl Med 51:158-163, 2010. https://doi.org/10.2967/jnumed.109.066399
54. Zhang L, Liu M, Wang L, et al: Constructing a large language model to generate impressions from findings in radiology Reports. Radiol: 312, 2024. https://doi.org/10.1148/radiol.240885
55. Bhayana R: Chatbots and large language models in radiology: A practical primer for clinical and research applications. Radiol: 310, 2024. https://doi.org/10.1148/radiol.232756

56. Doshi R, Amin KS, Khosla P, et al: Quantitative evaluation of large language models to streamline radiology report impressions: A multimodal retrospective analysis. Radiol 310. https://doi.org/10.1148/radiol.231593, 2024
57. Bülbül O, Bülbül HM, Kaba E: Assessing ChatGPT's summarization of 68Ga PSMA PET/CT reports for patients. Abdom Radiol 2024. https://doi.org/10.1007/s00261-024-04619-8
58. Huemann Z, Lee C, Hu J, et al: Domain-adapted large language models for classifying nuclear medicine reports. Radiol Artif Intell 5. https://doi.org/10.1148/ryai.220281, 2023
59. Tang CC, Nagesh S, Fussell DA, et al: Generating colloquial radiology reports with large language models. J Am Med Informatics Assoc 31:2660-2667, 2024. https://doi.org/10.1093/jamia/ocae223
60. Rogasch JMM, Metzger G, Preisler M, et al: ChatGPT: can you prepare my patients for [18F]FDG PET/CT and explain my reports? J Nucl Med 64:1876-1879, 2023. https://doi.org/10.2967/jnumed.123.266114
61. Shen Y, Xu Y, Ma J, et al: Multi-modal large language models in radiology: Principles, applications, and potential. Abdom Radiol 2024. https://doi.org/10.1007/s00261-024-04708-8
62. Javan R, Kim T, Mostaghni N: GPT-4 vision: Multi-modal evolution of ChatGPT and potential role in radiology. Cureus 16:e68298, 2024. https://doi.org/10.7759/cureus.68298
63. Dogra S, Silva E, Rajpurkar P: Reimbursement in the age of generalist radiology artificial intelligence. NPJ Digit Med 7:350, 2024. https://doi.org/10.1038/s41746-024-01352-w