



CHALMERS
UNIVERSITY OF TECHNOLOGY

Validation of an AI Method for Automated Lymphoma Metabolic Tumor Volume Segmentation Using a Public Benchmark PET/CT Dataset

Downloaded from: <https://research.chalmers.se>, 2026-06-14 17:31 UTC

Citation for the original published paper (version of record):

Sadik, M., Larsson, M., Enqvist, O. et al (2026). Validation of an AI Method for Automated Lymphoma Metabolic Tumor Volume Segmentation Using a Public Benchmark PET/CT Dataset. Journal of nuclear medicine : official publication, Society of Nuclear Medicine, 67(6): 1001-1005. <http://dx.doi.org/10.2967/jnumed.125.271605>

N.B. When citing this work, cite the original published paper.

Validation of an AI Method for Automated Lymphoma Metabolic Tumor Volume Segmentation Using a Public Benchmark PET/CT Dataset

May Sadik¹, Måns Larsson², Olof Enqvist^{2,3}, Lars Edenbrandt¹, and Elin Trägårdh^{4,5}

¹Department of Molecular and Clinical Medicine, Sahlgrenska University Hospital, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden; ²Eigenvision AB, Lund, Sweden; ³Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden; ⁴Department of Clinical Physiology and Nuclear Medicine, Skåne University Hospital, Malmö, Sweden; and ⁵Department of Translational Medicine and Wallenberg Center for Molecular Medicine, Lund University, Malmö, Sweden

The aim of this study was to evaluate the performance of an artificial intelligence (AI)-based method for automated segmentation of total metabolic tumor volume (TMTV) in ¹⁸F-FDG PET/CT scans of patients with lymphoma, using an independent, publicly available benchmark dataset curated and segmented by expert readers in a previously published study. **Methods:** The AI model, based on a 3-dimensional U-Net architecture implemented in MONAI (the medical open-source network for AI framework), was trained on 1,500 ¹⁸F-FDG PET/CT scans of patients with lymphoma. It was tested on a benchmark dataset comprising 60 baseline scans (20 each of follicular lymphoma, Hodgkin lymphoma, and diffuse large B-cell lymphoma), each segmented by 3 or 4 nuclear medicine physicians using an SUV threshold of 4. Agreement between AI-derived and benchmark TMTVs was assessed using Bland-Altman analysis, with acceptable deviation defined as within 10% or 10 cm³, consistent with interreader variability reported in the benchmark study. **Results:** In 50 (83%) of the 60 benchmark cases, AI-derived TMTVs were within 10% or 10 cm³ of the benchmark reference. In 4 of the remaining 10 cases, AI-derived results were within the same margin of at least 1 of the expert readers, indicating partial concordance. **Conclusion:** The AI-based method achieved high concordance with expert-derived TMTVs in a standardized benchmark setting. The findings demonstrate that the AI model performs comparably to human experts in most cases, even in an externally curated dataset deliberately enriched with challenging cases by its original authors. The AI model's ability to produce accurate, reproducible segmentations without user interaction could significantly reduce manual workload and interreader variability in lymphoma imaging. However, human supervision is required to minimize errors.

Key Words: lymphoma; total metabolic tumor volume; artificial intelligence; deep learning; benchmark dataset

J Nucl Med 2026; 67:1001–1005
DOI: 10.2967/jnumed.125.271605

Accurate quantification of total metabolic tumor volume (TMTV) from baseline ¹⁸F-FDG PET/CT imaging has emerged as a powerful prognostic biomarker in lymphoma, offering superior risk stratification compared with traditional staging methods and clinical scores (1–7). TMTV has demonstrated predictive value across multiple lymphoma subtypes, including Hodgkin lymphoma, follicular lymphoma, and diffuse large B-cell lymphoma, and is increasingly being considered for use in guiding treatment decisions and monitoring therapeutic response (8–11).

Despite its clinical relevance, widespread adoption of TMTV in routine practice and clinical trials has been hindered by the lack of standardized segmentation methods. Variability in lesion delineation, driven by differences in software tools, thresholding strategies, and reader interpretation, can lead to significant discrepancies in measured TMTVs, complicating comparisons across studies and limiting the utility of TMTV as a reproducible biomarker (12–19). Recognizing this challenge, an international consortium recently developed and published a benchmark dataset comprising 60 baseline ¹⁸F-FDG PET/CT scans of patients with lymphoma, segmented by 12 nuclear medicine experts, with each case processed by 3 or 4 readers (20). This benchmark provides a valuable reference for the evaluation and validation of segmentation methods, with the goal of harmonizing TMTV measurements across institutions and software platforms.

In previous work, we developed an artificial intelligence (AI)-based method for automated segmentation of TMTV in ¹⁸F-FDG PET/CT images of patients with lymphoma (21–23). Our approach leverages deep learning to identify and delineate metabolically active tumor regions with no user interaction, aiming to improve reproducibility, reduce manual editing, and accelerate clinical workflows. In this study, we build on these previous studies and have expanded the training cohort from 156 to 1,500 patients and included images from more hospitals. We evaluated the performance of our AI-based segmentation method using the publicly available benchmark dataset described by Boellaard et al. (20)—a proposed international benchmark dataset intended to standardize TMTV measurements in lymphoma. We assessed the accuracy of our method by comparing our results to expert-derived reference TMTVs.

MATERIALS AND METHODS

Patients

The training set of 1,500 ¹⁸F-FDG PET/CT studies was obtained from 2 hospitals and 2 databases from the Cancer Imaging Archive

Received Nov. 4, 2025; revision accepted Feb. 12, 2026.
For correspondence, contact Elin Trägårdh (elin.tragardh@med.lu.se).
Published online Mar. 5, 2026.

Immediate Open Access: Creative Commons Attribution 4.0 International License (CC BY) allows users to share and adapt with attribution, excluding materials credited to previous publications. License: <https://creativecommons.org/licenses/by/4.0/>. Details: <https://jnm.snmjournals.org/page/permissions>.

COPYRIGHT © 2026 by the Society of Nuclear Medicine and Molecular Imaging.

(FDG PET/CT lesions (24) and CALGB50303 (25)). Patient characteristics are shown in Table 1.

The study was approved by the ethics committees at the University of Gothenburg and Lund University or by the Swedish Ethical Review Authority. The need for written informed consent was waived for patients from Gothenburg University (2019-01274 and 2024-08225-02), whereas all patients from Skåne University Hospital provided written informed consent before entering the following studies: #2016/417, #2018/117, #2018/753, and #2021-05734-02. This study was performed in accordance with the ethical standards established in the 1964 Declaration of Helsinki and its later amendments. The datasets from anonymized publicly available publications of the Cancer Imaging Archive data were approved by the local ethics committee and data protection officer.

Test Set

The test set comprising the international TMTV benchmark group has been previously described (20). In short, it consists of 60 baseline ¹⁸F-FDG PET/CT scans of 20 patients with follicular lymphoma, 20 with Hodgkin lymphoma, and 20 with diffuse large B-cell lymphoma. The cases were chosen to represent a broad range of ¹⁸F-FDG distribution patterns, including less common presentations (e.g., focal or diffuse high uptake in the liver and spleen) and were enriched with more challenging cases rather than representative cases of the true prevalence of disease distribution. TMTV was measured by 12 nuclear medicine physicians, each analyzing 20 cases, with each case processed by 3 or 4 readers. Lesions were segmented using an SUV threshold of 4. TMTV ranged from 8 to 2,290 cm³ across scans and readers. The reference value for TMTV for each patient was defined as the median TMTV provided by the 3 or 4 readers. An interreader difference in TMTV of less than 10 cm³ or 10% was found in 87% of cases in the benchmark dataset. In 13% of cases, discrepancies were caused by differences in lesion inclusion or high-uptake region removal for final TMTV.

AI Algorithm

Automated segmentation of metabolically active tumor regions was performed using a 3-dimensional convolutional neural network with a 4-level U-Net architecture (26) implemented in MONAI (the medical open network for AI framework) (27). The model inputs were coregistered ¹⁸F-FDG PET/CT images, resampled to a voxel size of 2.73 × 2.73 × 2.79 mm. Intensities were normalized to a 0–1 range

after limiting CT values to −1,024 to 3,072 Hounsfield units and SUVs to 0–100. The network produced a binary segmentation map distinguishing pathologic uptake from background.

The model was trained on 1,500 ¹⁸F-FDG PET/CT scans of patients with lymphoma, collected from 2 hospitals and 2 publicly available Cancer Imaging Archive datasets. Of these, 1,200 scans were used for training and 300 for validation. Training was performed for 200 epochs on 192 × 192 × 192 voxel patches, with 1 epoch defined as 10,000 samples. The Nesterov-accelerated Adaptive Moment Estimation optimizer (28) (initial learning rate, 5 × 10^{−5}; exponential decay factor, 0.985) was used together with a weighted categorical cross-entropy loss (foreground weight, 25.0; background, 1.0). Deep supervision with auxiliary losses (weights of 0.5, 0.25, and 0.125) was applied to improve gradient flow (29). Foreground and background patches were sampled in equal proportions, with adaptive resampling of challenging background regions every 20 epochs.

Inference was performed using a sliding-window approach with 94-voxel overlap to minimize edge artifacts. For voxels with multiple predictions, the value corresponding to the largest effective receptive field was retained. The resulting segmentation was postprocessed using an SUV-based thresholding step—voxels with an SUV of at least 4 and with their nearest local SUV maximum located inside the original AI segmentation were retained as foreground; all others were assigned to background. This produced full-volume, postprocessed segmentations of metabolically active tumor regions without user interaction, enabling direct computation of AI-based TMTVs.

Statistical Analysis

Agreement between the AI-based TMTVs and the benchmark reference was evaluated using Bland–Altman analysis. According to Boellaard et al. (20), an interreader difference in TMTV of less than 10 cm³ or 10% is acceptable. Thus, the proportion of studies fulfilling these criteria were noted.

RESULTS

The AI-derived TMTV ranged from 4 to 2,165 cm³, whereas the benchmark TMTV spanned from 8 to 2,290 cm³. Agreement between AI-derived and benchmark TMTVs is illustrated in the Bland–Altman plot (Fig. 1).

TABLE 1
Patient Characteristics

Characteristic	Dataset			
	Skåne UH	AutoPET	Sahlgrenska UH	CALGB
<i>n</i>	624	503	258	115
Sex				
Female	251	241	118	—
Male	373	262	140	—
Age (y)	52 (17–88)	55 (11–85)	43 (10–85)	—
Lymphoma type				
Hodgkin	281	—	177	—
DLBCL	322	—	60	115
Follicular	6	—	2	—
Other	15	138	19	—
Negative controls	—	365	—	—

UH = University Hospital; DLBCL = diffuse large B-cell lymphoma.

Data are presented as number; age is presented as mean with range in parentheses.

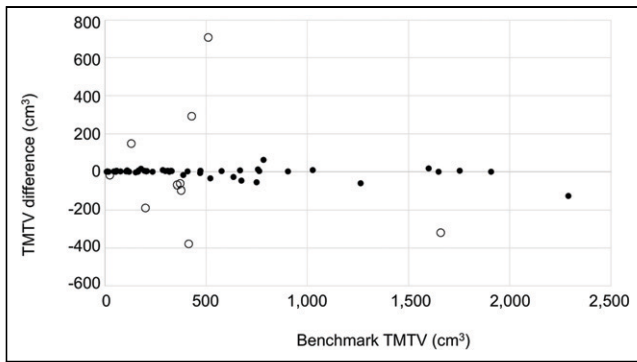


FIGURE 1. Bland-Altman plot showing difference between AI-derived and benchmark TMTV. Ten cases with largest discrepancies, described in Table 2, are represented by open circles; remaining 50 cases are shown as solid dots.

In 50 (83%) of the 60 cases, the AI-derived TMTVs were within 10% or 10 cm³ of the benchmark TMTV, indicating a high level of concordance. The remaining 10 cases, which showed larger discrepancies, are detailed in Table 2. Among these 10 outlier cases, 4 (H15, F02, B16, F05) correspond to patients for whom substantial interreader variability was observed during benchmark TMTV determination, as described by Boellaard (20). In these cases, the AI-derived TMTV either aligned with 1 or 2 of the individual expert readers or fell between the range of values provided by the readers.

The remaining 6 cases, shown in Figure 2, exhibited discrepancies attributable to specific segmentation issues, as annotated in Table 2. These included undersegmentation (e.g., missed bone metastases in B07 or partial tumor volumes in F13 and B15), oversegmentation (e.g., inclusion of the spleen in F04), and differences in tumor boundary delineation.

DISCUSSION

TMTV is a valuable prognostic and predictive biomarker in both clinical practice and research. However, realization of its full potential is dependent on accurate, reproducible, and efficient methods for tumor segmentation and TMTV calculation. In this study, we evaluated an automated AI-based approach. In 83% (50/60) of the benchmark cases, AI-derived TMTVs were in agreement with the benchmark TMTVs. This is close to the 87% agreement that Boellaard et al. reported between the final benchmark TMTVs and the TMTVs of the 3 or 4 readers involved in the benchmark process (20). In 4 additional cases, the AI-derived segmentations deviated from the benchmark reference but still aligned with at least 1 expert reader involved in the benchmark process.

It is important to note that the benchmark dataset used in this study was not designed for evaluating the clinical performance of segmentation models. Although the TMTVs derived by the AI model showed good agreement with the benchmark TMTVs, this does not necessarily imply that the same individual lesions were segmented by AI and the expert readers. As stated by Boellaard et al., one intended use of the dataset is to evaluate the performance of AI-based segmentations. Furthermore, the benchmark dataset was deliberately enriched with complex and challenging cases that are more difficult than those typically encountered in routine clinical settings (20). This context makes our results particularly encouraging, as they demonstrate robustness of the AI model under demanding conditions.

Generalizability is a crucial aspect of AI model development. Validation on external datasets is essential, as performance on test sets drawn from the same site or scanner as the training data may overestimate the model's effectiveness.

The strong performance of our AI tool can be partly attributed to the large and diverse training dataset, comprising 1,500 PET/CT studies from patients with various diagnoses across multiple hospitals. This dataset was made possible through publicly available

TABLE 2
Cases with Discrepancies Between AI-Derived and Benchmark TMTV Exceeding Both 10% and 10 cm³

Patient No.	Benchmark TMTV (cm ³)	AI-derived TMTV (cm ³)	Comment*
Hodgkin lymphoma			
H15	130	279	Individual TMTV estimates: 129 cm ³ , 275 cm ³ , 282 cm ³
Follicular lymphoma			
F02	376	279	Individual TMTV estimates: 252 cm ³ , 399 cm ³ , 426 cm ³ , 456 cm ³
F04	428	722	AI included spleen in segmentation
F05	1,658	1,338	Individual TMTV estimates: 567 cm ³ , 1,650 cm ³ , 1,680 cm ³ , 1,706 cm ³
F08	413	34	AI missed main areas of tumor burden
F13	356	287	AI missed minor parts of tumor bulk
F15	22	4	AI detected tumor but segmented it smaller
DLBCL			
B07	199	11	AI missed widespread bone metastases
B15	371	309	AI missed 1 liver lesion; others segmented accurately
B16	509	1,218	Individual TMTV estimates: 509 cm ³ , 546 cm ³ , 1,223 cm ³

*Individual TMTV estimates provided by expert readers during benchmark definition process (20).

DLBCL = diffuse large B-cell lymphoma.

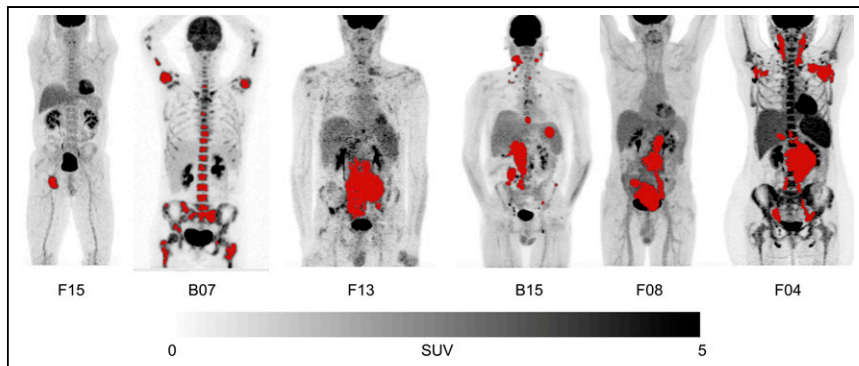


FIGURE 2. Maximum-intensity-projection images of 6 patients with discrepancies between AI-derived and benchmark TMTV exceeding both 10% and 10 cm³. Corresponding details and commentary are provided in Table 2.

resources such as the Cancer Imaging Archive, highlighting the importance of open-source initiatives in developing generalizable AI tools. However, a limitation of using such datasets is that key clinical metadata, such as diagnosis, age, sex, scan type (baseline or follow-up), were not consistently available because of privacy constraints.

Observed discrepancies between the AI and benchmark segmentations can often be explained by underrepresentation or inconsistencies in the training data. For instance, cases with diffuse splenic involvement or osseous lesions pose challenges not only for AI models but also for human experts, as noted in the benchmark study. These ambiguities underline the need for well-annotated and consistent training data. Furthermore, 6 of the 10 cases with discrepancies shown in Table 2 were patients with follicular lymphoma, a group only represented by 8 patients in the large training set, which is a limitation. Efforts should be made in future training of AI models to include more diverse patients. Further validation of the method is needed in different contexts, such as staging, early or late response assessment, and other diagnoses. The benchmark dataset only included baseline PET/CT scans.

Unlike some existing approaches that rely on fixed SUV thresholds (e.g., SUV of 2.5 or 4 or 41% of SUV_{max}), our AI model was trained on manually curated segmentations considered optimal for each case. This allowed for more nuanced learning. Furthermore, postprocessing steps can be added to apply SUV thresholds retrospectively if needed, offering flexibility for different clinical or research requirements.

Our findings demonstrate the feasibility of using an AI-based tool to generate TMTV and segmentations that are, in most cases, clinically usable without manual correction. The potential to reduce tumor segmentation time to near zero represents a significant advantage over current software solutions, which often require manual inclusion or exclusion of regions. As Boellaard et al. have noted, such manual steps not only increase processing time but also introduce interreader variability (20). However, it should be noted that human supervision and correction of the AI model, when needed, are still required.

For now, the AI model is freely available for research use on the RECOMIA platform, where over 75 research groups across 35 countries have already accessed our various AI tools for their own studies.

CONCLUSION

The AI-based method achieved high concordance with expert-derived TMTVs in a standardized benchmark setting, demonstrating that the AI model performed comparably to human experts in most

cases, even in an externally curated dataset deliberately enriched with challenging cases. The AI model's ability to produce accurate, reproducible segmentations without user interaction could significantly reduce manual workload and interreader variability in lymphoma imaging. However, human supervision is still required to minimize errors.

DISCLOSURE

This study was supported by the Swedish State under the agreement between the Swedish Government and the Country Councils, the ALF agreement (70380), and Malmö General Hospital Cancer Founda-

tion. Måns Larsson and Olof Enqvist are employees and stockholders of Eigenvision AB, a consultancy company working with research and development in automated image analysis, computer vision, and machine learning. Lars Edenbrandt serves on the board and is a stockholder of Slicevault AB, a company working with streamlining image management for biotech and pharmaceutical companies. No other potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENT

We acknowledge the team behind the benchmark dataset for their significant contribution in promoting standardized evaluation in this field.

KEY POINTS

QUESTION: Can an AI-based method accurately and reproducibly segment TMTV in lymphoma ¹⁸F-FDG PET/CT scans using a public benchmark dataset?

PERTINENT FINDINGS: In this validation study, a 3-dimensional U-Net AI model trained on 1,500 lymphoma PET/CT scans was tested on a benchmark dataset of 60 cases segmented by experts. The AI-derived TMTVs were within 10% or 10 cm³ of the benchmark in 83% of cases, demonstrating high concordance with expert readers.

IMPLICATIONS FOR PATIENT CARE: This AI tool could help physicians image lymphoma more quickly and consistently, which may lead to better treatment planning and less time spent on manual work.

REFERENCES

- Cottreau AS, Rebaud L, Trotman J, et al. Metabolic tumor volume predicts outcome in patients with advanced stage follicular lymphoma from the RELEVANCE trial. *Ann Oncol.* 2024;35:130–137.
- Eertink JJ, Zwezerijnen GJC, Heymans MW, et al. Baseline PET radiomics outperforms the IPI risk score for prediction of outcome in diffuse large B-cell lymphoma. *Blood.* 2023;141:3055–3064.
- 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.* 2022;40:2352–2360.
- Thieblemont C, Chartier L, Duhren U, et al. A tumor volume and performance status model to predict outcome before treatment in diffuse large B-cell lymphoma. *Blood Adv.* 2022;6:5995–6004.

5. Cottreau AS, Meignan M, Nioche C, et al. Risk stratification in diffuse large B-cell lymphoma using lesion dissemination and metabolic tumor burden calculated from baseline PET/CT. *Ann Oncol*. 2021;32:404–411.
6. Cottreau AS, 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*. 2018;131:1456–1463.
7. Meignan M, Cottreau AS, Versari A, et al. Baseline metabolic tumor volume predicts outcome in high-tumor-burden follicular lymphoma: a pooled analysis of three multicenter studies. *J Clin Oncol*. 2016;34:3618–3626.
8. Luminari S, Manni M, Galimberti S, et al.; Fondazione Italiana Linfomi. Response-adapted postinduction strategy in patients with advanced-stage follicular lymphoma: the FOLL12 study. *J Clin Oncol*. 2022;40:729–739.
9. El-Galaly TC, Villa D, Gormsen LC, Baech J, Lo A, Cheah CY. FDG-PET/CT in the management of lymphomas: current status and future directions. *J Intern Med*. 2018;284:358–376.
10. Mikhaeel NG, Smith D, Dunn JT, et al. Combination of baseline metabolic tumour volume and early response on PET/CT improves progression-free survival prediction in DLBCL. *Eur J Nucl Med Mol Imaging*. 2016;43:1209–1219.
11. Voltin CA, Mettler J, van Heek L, et al. Early response to first-line anti-PD-1 treatment in Hodgkin lymphoma: a PET-based analysis from the prospective, randomized phase II NIVAHL trial. *Clin Cancer Res*. 2021;27:402–407.
12. Boellaard R, Zwezerijnen GJC, Buvat I, et al. Measuring total metabolic tumor volume from ¹⁸F-FDG PET: a reality check. *J Nucl Med*. 2025;66:802–805.
13. Shiyam Sundar LK, Beyer T. Is automatic tumor segmentation on whole-body ¹⁸F-FDG PET images a clinical reality? *J Nucl Med*. 2024;65:995–997.
14. Barrington SF, Zwezerijnen B, de Vet HCW, et al. Automated segmentation of baseline metabolic total tumor burden in diffuse large B-cell lymphoma: which method is most successful? A study on behalf of the PETRA Consortium. *J Nucl Med*. 2021;62:332–337.
15. Barrington SF, Meignan M. Time to prepare for risk adaptation in lymphoma by standardizing measurement of metabolic tumor burden. *J Nucl Med*. 2019;60:1096–1102.
16. Ilyas H, Mikhaeel NG, Dunn JT, et al. Defining the optimal method for measuring baseline metabolic tumour volume in diffuse large B cell lymphoma. *Eur J Nucl Med Mol Imaging*. 2018;45:1142–1154.
17. Cottreau AS, Buvat I, Kanoun S, et al. Is there an optimal method for measuring baseline metabolic tumor volume in diffuse large B cell lymphoma? *Eur J Nucl Med Mol Imaging*. 2018;45:1463–1464.
18. Cottreau AS, Hapdey S, Chartier L, et al. Baseline total metabolic tumor volume measured with fixed or different adaptive thresholding methods equally predicts outcome in peripheral T cell lymphoma. *J Nucl Med*. 2017;58:276–281.
19. Kanoun S, Tal I, Berriolo-Riedinger A, et al. Influence of software tool and methodological aspects of total metabolic tumor volume calculation on baseline [¹⁸F]FDG PET to predict survival in Hodgkin lymphoma. *PLoS One*. 2015;10:e0140830.
20. Boellaard R, Buvat I, Nioche C, et al. International benchmark for total metabolic tumor volume measurement in baseline ¹⁸F-FDG PET/CT of lymphoma patients: a milestone toward clinical implementation. *J Nucl Med*. 2024;65:1343–1348.
21. Sadik M, Barrington SF, Trägårdh E, et al. Metabolic tumour volume in Hodgkin lymphoma—a comparison between manual and AI-based analysis. *Clin Physiol Funct Imaging*. 2024;44:220–227.
22. Sadik M, López-Urdaneta J, Ulén J, et al. Artificial intelligence increases the agreement among physicians classifying focal skeleton/bone marrow uptake in Hodgkin's lymphoma patients staged with [¹⁸F]FDG PET/CT—a retrospective study. *Nucl Med Mol Imaging*. 2023;57:110–116.
23. Sadik M, López-Urdaneta J, Ulén J, et al. Artificial intelligence could alert for focal skeleton/bone marrow uptake in Hodgkin's lymphoma patients staged with FDG-PET/CT. *Sci Rep*. 2021;11:10382.
24. Gatidis S, Kuestner T. A whole-body FDG-PET/CT dataset with manually annotated tumor lesions (FDG-PET-CT-Lesions). Cancer Imaging Archive website. <https://www.cancerimagingarchive.net/collection/fdg-pet-ct-lesions>. Published June 2, 2022. Accessed February 18, 2026.
25. Bartlett N, Wilson W, Jung S, et al. Rituximab and combination chemotherapy in treating patients with diffuse large B-cell non-Hodgkin's lymphoma (CALGB50303) (version 2). Cancer Imaging Archive website. <https://www.cancerimagingarchive.net/collection/calgb50303/>. Updated August 6, 2025. Accessed February 18, 2026.
26. Çiçek Ö, Abdulkadir A, Lienkamp SS, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin S, Joskowicz L, Sabuncu M, Unal G, Wells W, eds. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016*. Springer; 2016:424–432.
27. Cardoso JM, Li W, Brown R, et al. Monai: an open-source framework for deep learning in healthcare [preprint]. <https://arxiv.org/pdf/2211.02701>. Accessed February 18, 2026.
28. Dozat T. Incorporating Nesterov momentum into Adam. Presented at 4th International Conference on Learning Representations, Workshop Track; May 2016. San Juan, Puerto Rico.
29. Lee C-Y, Xie S, Gallagher P, Zhang Z, Tu Z. Deeply-supervised nets. *Proc Mach Learn Res*. 2015;38:562–70.