



## **AI Improves Agreement and Reduces Time for Quantifying Metabolic Tumour Burden in Hodgkin Lymphoma †**

Downloaded from: <https://research.chalmers.se>, 2026-03-12 13:30 UTC




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

Sadik, M., Barrington, S., Ulén, J. et al (2025). AI Improves Agreement and Reduces Time for Quantifying Metabolic Tumour Burden in Hodgkin Lymphoma †. *Hematology Reports*, 17(6). <http://dx.doi.org/10.3390/hematolrep17060060>

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

## Article

# AI Improves Agreement and Reduces Time for Quantifying Metabolic Tumour Burden in Hodgkin Lymphoma <sup>†</sup>

May Sadik <sup>1,\*</sup>, Sally F. Barrington <sup>2</sup>, Johannes Ulén <sup>3</sup>, Olof Enqvist <sup>3,4</sup>, Elin Trägårdh <sup>5</sup> , Babak Saboury <sup>6</sup> , Anne Lerberg Nielsen <sup>7</sup>, Annika Loft <sup>8</sup>, Jose Luis Loaiza Gongora <sup>9</sup>, Jesus Lopez Urdaneta <sup>1</sup>, Rajender Kumar <sup>10</sup>, Martijn van Essen <sup>1</sup> and Lars Edenbrandt <sup>1</sup> 

- <sup>1</sup> Department of Molecular and Clinical Medicine, Clinical Physiology, Sahlgrenska University Hospital, Sahlgrenska Academy at the University of Gothenburg, 413 90 Gothenburg, Sweden; [jesus.lopez.urdaneta@vregion.se](mailto:jesus.lopez.urdaneta@vregion.se) (J.L.U.); [martijn.van.essen@vregion.se](mailto:martijn.van.essen@vregion.se) (M.v.E.); [lars.edenbrandt@gu.se](mailto:lars.edenbrandt@gu.se) (L.E.)
- <sup>2</sup> Clinical PET Centre, School of Biomedical Engineering and Imaging Sciences Kings College, London SE5 9RS, UK; [sally.barrington@kcl.ac.uk](mailto:sally.barrington@kcl.ac.uk)
- <sup>3</sup> Eigenvision AB, 211 30 Malmö, Sweden; [johannes@eigenvision.se](mailto:johannes@eigenvision.se) (J.U.); [olof.enqvist@gmail.com](mailto:olof.enqvist@gmail.com) (O.E.)
- <sup>4</sup> Department of Electrical Engineering, Chalmers University of Technology, 412 96 Gothenburg, Sweden
- <sup>5</sup> Clinical Physiology and Nuclear Medicine, Lund University and Skåne University Hospital, 205 02 Malmö, Sweden; [elin.tragardh@med.lu.se](mailto:elin.tragardh@med.lu.se)
- <sup>6</sup> Department of Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda, MD 20892, USA; [babak.saboury@nih.gov](mailto:babak.saboury@nih.gov)
- <sup>7</sup> Department of Nuclear Medicine, Odense University Hospital, 5000 Odense, Denmark; [anne.l.nielsen@rsyd.dk](mailto:anne.l.nielsen@rsyd.dk)
- <sup>8</sup> Department of Clinical Physiology, Nuclear Medicine and PET, Centre of Diagnostic Investigations, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark; [annika.loft.jakobsen@regionh.dk](mailto:annika.loft.jakobsen@regionh.dk)
- <sup>9</sup> Department of Diagnostic Imaging, Akershus University Hospital, 1478 Oslo, Norway; [loaiza\\_joseluis@hotmail.com](mailto:loaiza_joseluis@hotmail.com)
- <sup>10</sup> Department of Nuclear Medicine, Post Graduate Institute of Medical Education and Research, Chandigarh 160012, India; [drrajender2010@gmail.com](mailto:drrajender2010@gmail.com)
- \* Correspondence: [may.sadik@gu.se](mailto:may.sadik@gu.se); Tel.: +46-31-3428028; Fax: +46-31-411735
- <sup>†</sup> This paper is an extended version of our paper published in M. Sadik, S. Barrington, J. Ulén, O. Enqvist, E. Trägårdh, B. Saboury, A. Lerberg Nielsen, A. Loft Jakobsen, J. Loaiza Gongora, J. López-Urdaneta, R. Kumar, L. Edenbrandt. Artificial intelligence increases agreement among physicians quantifying metabolic tumour burden in Hodgkin's lymphoma patients staged with [18F]FDG PET/CT. In Proceedings of 35th Annual Congress of the European Association of Nuclear Medicine (EANM'22), Barcelona, Spain, 15–19 October 2022; Abstract EP-226.



check for updates

Academic Editor: Claudio Cerchione

Received: 4 September 2025

Revised: 27 October 2025

Accepted: 29 October 2025

Published: 7 November 2025

**Citation:** Sadik, M.; Barrington, S.F.; Ulén, J.; Enqvist, O.; Trägårdh, E.; Saboury, B.; Lerberg Nielsen, A.; Loft, A.; Loaiza Gongora, J.L.; Lopez Urdaneta, J.; et al. AI Improves Agreement and Reduces Time for Quantifying Metabolic Tumour Burden in Hodgkin Lymphoma. *Hematol. Rep.* **2025**, *17*, 60. <https://doi.org/10.3390/hematolrep17060060>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## Abstract

**Background:** The aim was to evaluate whether an artificial intelligence (AI)-based tool for the automated quantification of the total metabolic tumour volume (tMTV) in patients with Hodgkin lymphoma (HL) could support nuclear medicine specialists in lesion segmentation and thereby enhance inter-observer agreement. **Methods:** Forty-eight consecutive patients who underwent staging with [18F]FDG PET/CT were included. Eight invited specialists from different hospitals were asked to manually segment lesions for tMTV calculations in 12 cases without AI advice, and to use automated AI segmentation in a further 12 cases, with editing as required, i.e., segmenting/adjusting 24 cases each. Each case was segmented by two specialists manually and by two different specialists using the AI tool, allowing for the pairwise comparison of inter-observer variability. **Results:** The median difference between two specialists performing manual tMTV segmentations was 26 cm<sup>3</sup> (IQR 10–86 cm<sup>3</sup>) corresponding to 23% (IQR 7–50%) of the median tMTV in the dataset, while the median difference between two specialists tMTV adjustments using AI segmentations was 12 cm<sup>3</sup> (IQR 4–39 cm<sup>3</sup>) corresponding to 9% (IQR 2–21%) ( $p = 0.023$ ). The median difference in tMTV between measurements with and without AI was 3.3 cm<sup>3</sup>, corresponding to 2.3% of the median tMTV. **Conclusions:** An automated AI-based tool can

significantly increase agreement among specialists quantifying tMTV in HL patients staged with [18F]FDG PET/CT, without markedly changing the measurements.

**Keywords:** total metabolic tumour volume; artificial intelligence; Hodgkin disease; observer variation; Fluorodeoxyglucose F18

## 1. Introduction

Total metabolic tumour volume (tMTV) and/or total tumour lesion glycolysis (tTLG) has/have been reported to be associated with progression-free and, sometimes, with overall survival in Hodgkin lymphoma (HL) patients staged with [18F]FDG PET/CT [1]. The manual segmentation of every abnormal area of uptake in the [18F]FDG PET/CT is, however, a subjective and time-consuming process, causing inter-observer disagreement. Semi-automated methods based on absolute or relative standardised uptake value (SUV) thresholds have been shown to significantly under- or overestimate visible tumours, thereby limiting the utility of tMTV and tTLG measurements in both clinical practice and clinical trials [2,3]. The calculation of tMTV and tTLG are currently not widely used in clinical practice. We believe an artificial intelligence (AI)-based tool, which has been trained to mimic human readers, has the potential of assisting image readers analysing [18F]FDG PET/CT from HL patients.

A convolutional neural network (CNN) has been developed for non-Hodgkin lymphoma [4], while we developed a tool (RECOMIA) specifically trained on HL patients [5]. As a first step, we compared our tool with a system by PARS (Siemens Medical Solutions USA, Inc.) [6], trained with lesions from lung cancer and lymphoma patients. We found that the RECOMIA and PARS AI tools could be applied without major manual adjustments in 69% (33/48) and 58% (28/48) of patients with Hodgkin lymphoma (HL), respectively.

In this study, we aim to investigate whether the RECOMIA AI-based tool for the automated quantification of the metabolic tumour burden in HL patients staged with [18F]FDG PET/CT could assist nuclear medicine specialists in segmenting focal lesions and thereby improve inter-observer agreement in the quantitative results, and reducing segmentation time.

## 2. Methods

### 2.1. Patients

A total of 49 newly diagnosed, untreated patients with biopsy-proven Hodgkin lymphoma (HL) who underwent staging with [18F]FDG PET/CT at Sahlgrenska University Hospital between 2017 and 2018 were initially included. One patient was excluded due to a failure in recording uptake time. The final cohort comprised 48 patients, with a median age of 35 years (range 7–75), of whom 46% were female. This is the same patient group as reported in previous publications [5,7].

The training of the AI tool is described in [5].

### 2.2. Image Acquisitions

[18F]FDG PET/CT scans were acquired using a Siemens Biograph 64 TruePoint integrated PET/CT system. Patients fasted for at least 6 hours before FDG administration. Adult patients received an injection of 4 MBq/kg [18F]FDG (maximum 400 MBq), while paediatric doses were administered according to the EANM Dosage Card (Version 5.7.2016). The standard uptake time was 60 min. Image acquisition was performed with 3 min per bed position, covering the base of the skull to the mid-thigh. PET images were recon-

structured using an iterative OSEM 3D algorithm (4 iterations, 8 subsets) with a matrix size of  $168 \times 168$ , slice thickness of 5 mm, and slice spacing of 3 mm. CT-based attenuation and scatter corrections were applied. A low-dose CT scan (64-slice helical, 120 kV, 30 mAs,  $512 \times 512$  matrix) was acquired over the same field of view as the PET scan. CT reconstruction was performed using a filtered back projection algorithm with slice thickness and spacing matched to the PET images [7,8].

### 2.3. Image Interpretation

Eight nuclear medicine specialists (S.F.B., E.T., B.S., A.L.N., A.L., J.L.L., J.L.U., and R.K.) from 8 different hospitals, each with more than five years of experience in interpreting PET/CT studies, were invited to participate. They were asked to segment FDG uptake in tumour sites listed below that should be included in the tMTV and the tTLG ( $tMTV \times SUV_{mean}$ ) calculations with and without the AI tool. The specialists were informed that the cohort consisted of untreated HL patients, but no other clinical information such as stage, histology, gender, or age were provided. The following FDG uptake should be segmented as “tumour” based on the recommendations by [1]:

- Viable regions within lymph nodes showing increased FDG uptake;
- Focal FDG uptake in bone marrow or other extranodal sites;
- Focal FDG uptake in the spleen, regardless of splenic size;
- Diffuse splenic uptake exceeding liver uptake (spleen/liver ratio  $> 1.5$  and bone marrow/liver ratio  $< 1.0$ ), in the absence of reactive bone marrow changes.

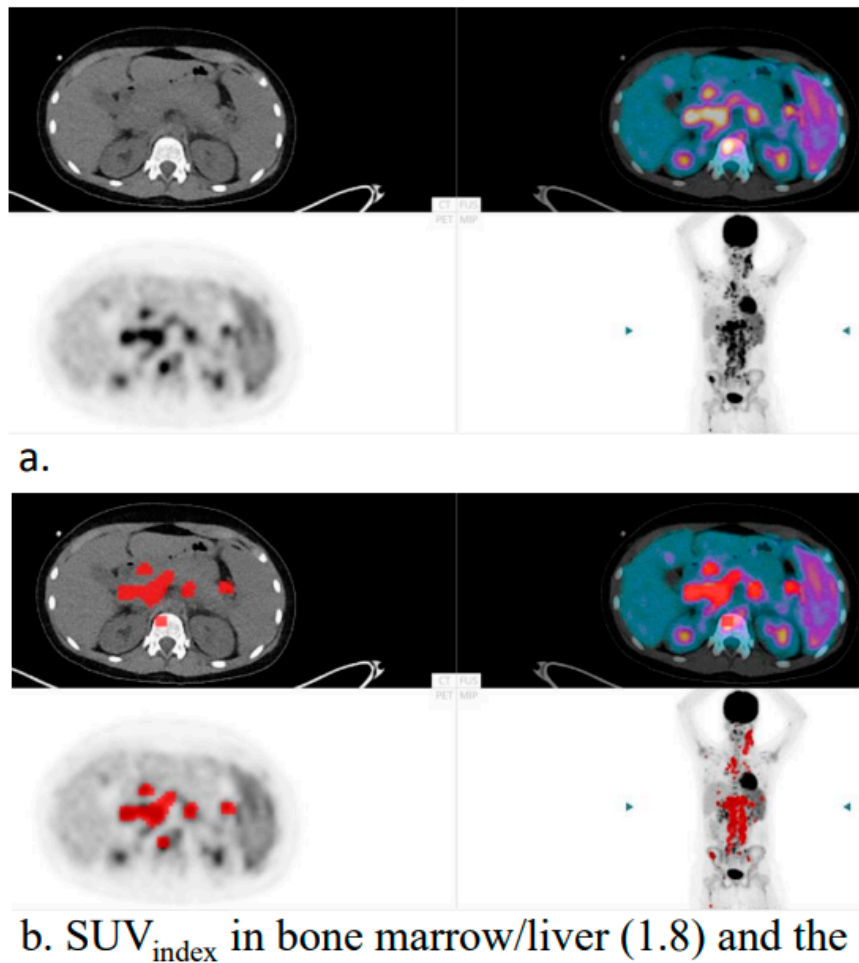
The RECOMIA cloud-based software (The AI “Lymphoma FDG-PET/CT segmentation:1.0” can be found at the RECOMIA platform.), was used, providing each examination with PET, CT, fused  $[^{18}F]FDG$  PET/CT, and maximum intensity projection images [9]. Interpreters could navigate coronal, sagittal, and transverse planes, and PET images were available in multiple colour scales. By default, images were scaled to an upper SUV threshold of 5, but both the SUV threshold and colour scale could be adjusted based on the reader’s preference. The CT images could be viewed using standard settings, e.g., bone, soft tissue, and lung. The segmentation brush could be adjusted in size and to mark only activity above a defined SUV threshold specified by the user. Before beginning, each specialist received two help videos showing how to perform the analysis and an instruction document explaining the purpose of the study.

Each of the eight specialists analysed 24 cases: 12 cases without the AI tool and 12 other cases with the AI tool (see below). Each case was analysed by four different specialists, two without the AI tool and the other two with the AI tool. The cases were randomly distributed to the specialists. In order to limit learning effects, four of the specialists started with the 12 cases without the AI tool and the other four started with the 12 cases with AI tool. The specialists were asked to record the segmentation time for each case.

Without AI tool: The specialist was asked to manually segment lesions that should be included in the tMTV and tTLG calculations.

With AI tool: The specialist was asked to adjust, as required, the AI lesion segmentations that should be included in the tMTV and tTLG calculations. Focal nodal and extra-nodal lesions detected by the AI tool were highlighted in the PET images (Figure 1). The  $SUV_{index}$  in bone marrow ( $SUV_{median}$  bone marrow/ $SUV_{median}$  liver) and the spleen ( $SUV_{median}$  spleen/ $SUV_{median}$  liver) were calculated by the AI tool and displayed together with the images (Figure 1). The tool calculated the median values for bone marrow, spleen, and liver. Diffused increased uptake in the spleen that was  $>1.5$  times the liver in the absence of diffusely increased bone marrow uptake was highlighted for the reader in the

images (Figure 1). The median SUVs in bone marrow, liver, and spleen were chosen in order to use the most common value and avoid extreme values due to focal lesions [5,8].



**Figure 1.** Example of a patient without (a) and with (b) the automated artificial intelligence (AI) pre-segmentation of tumour burden presented to the specialists (in red). The  $SUV_{index}$  in bone marrow/liver and the spleen/liver, calculated by the AI tool, were presented to the specialists (b).

#### 2.4. AI Tool

The tool, described in detail in [5], is composed of two convolutional neural networks (CNNs), one using only the CT image as input, used to segment tumour in spleen, bone, and liver, and one that uses CT, PET, and an auxiliary mask derived from the CT image as input, designed to directly segment lymph node tumours.

Focal spleen and liver uptake were defined as pixels with SUV above  $SUV_{mean} + 2SD$  for that organ. The lymph node CNN uses U-net 3D architecture with two 25% dropout layers. A training set of 101 PET/CT studies from lymphoma patients. Two nuclear medicine specialists performed the segmentation in the training group. A more detailed description of the AI methods can be found in [5].

#### 2.5. Statistical Analysis

To evaluate whether the agreement in tMTV measurements had changed using the AI tool, the absolute difference between each pair of specialists who segmented the same cases was calculated. For each image, this gives one difference for values obtained with the AI tool and one without. Kolmogorov–Smirnov and Shapiro–Wilk tests were performed for all analysed parameters showing significant difference ( $p < 0.001$ ); i.e., the data were not normally distributed. Wilcoxon signed-rank test (two-sided) was used to test whether

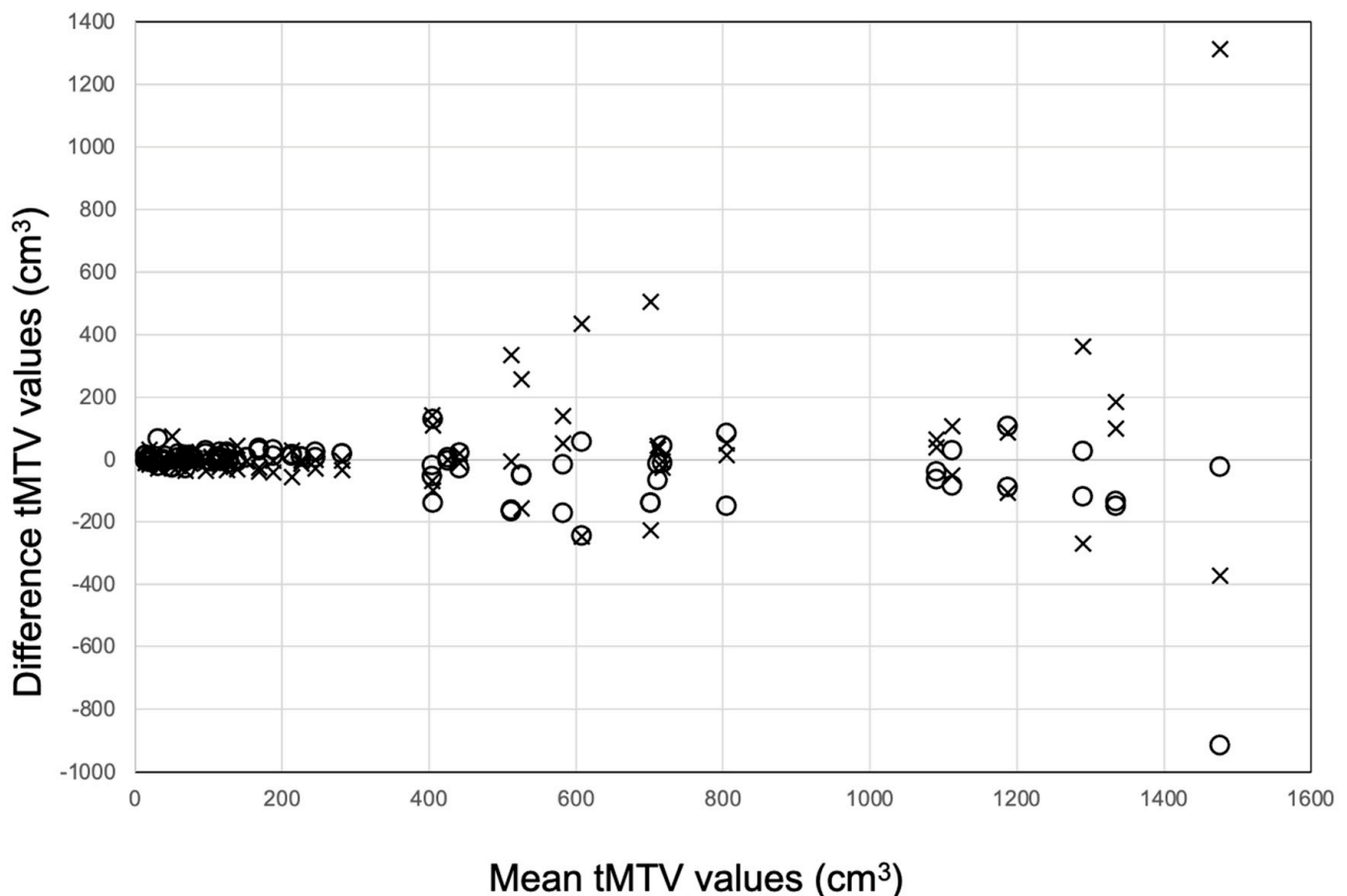
there was a significant difference between these paired measurements. The exact same setup was used to examine the tTLG measurements.

Some readers forgot to record the time for their examinations, but there were 32 cases with timings for both segmentations with and without AI tool. There are several ties since the timings are measured in whole minutes. Therefore, a two-sided sign test was used.

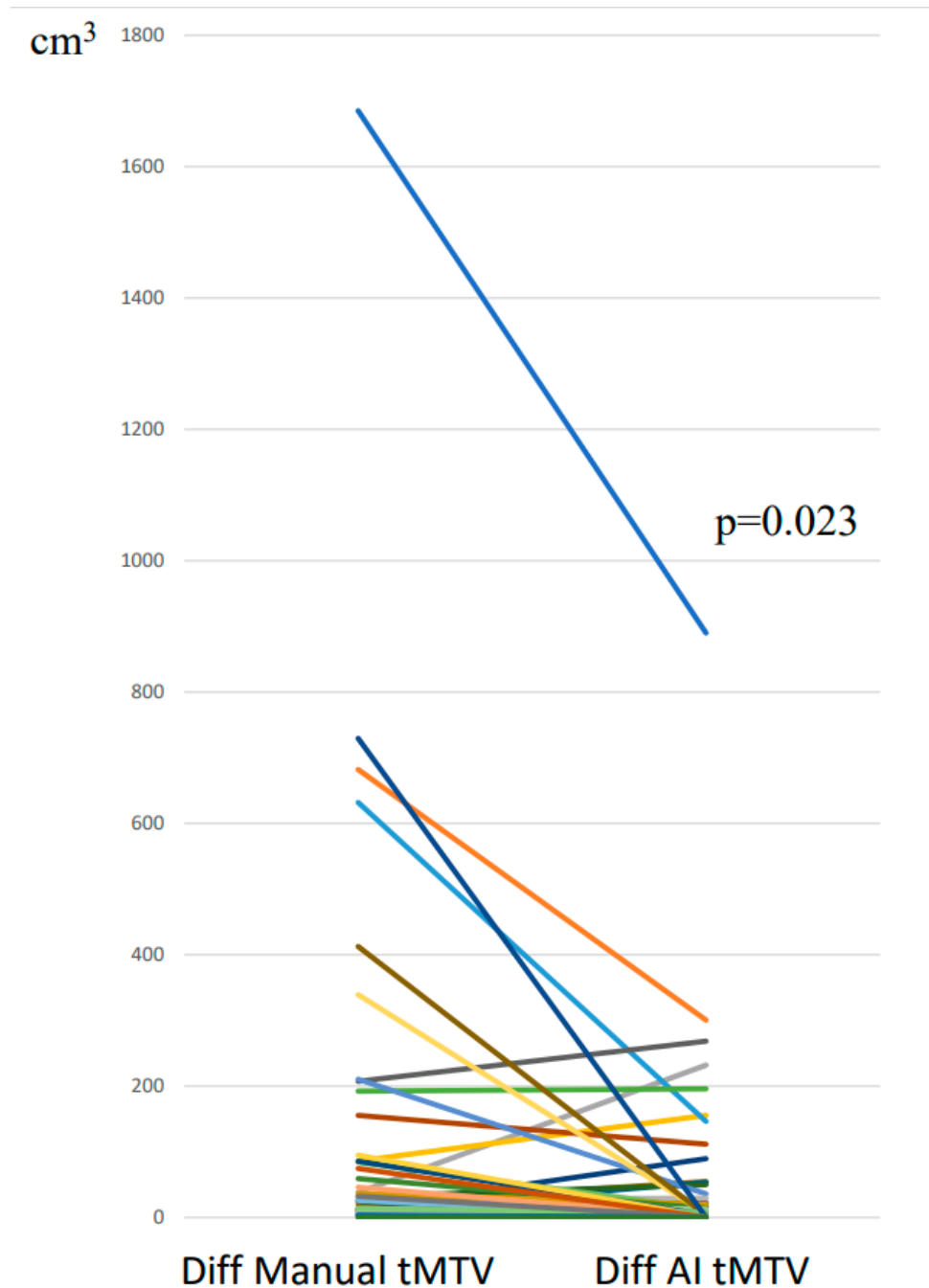
### 3. Results

#### 3.1. tMTV: Segmentations with and Without AI Tool

The manual tMTV segmentation ranged between 2–2789 cm<sup>3</sup>, while the specialists' tMTV segmentation using the AI tool ranged between 10–1451 cm<sup>3</sup>. The median difference between two specialists' manual tMTV segmentations was 26 cm<sup>3</sup> (interquartile range (IQR) 10–86 cm<sup>3</sup>) corresponding to 23% (IQR 7–50%) of the median tMTV in the study cohort, while the median difference between two specialists' tMTV adjustments of AI segmentations was 12 cm<sup>3</sup> (IQR 4–39 cm<sup>3</sup>) corresponding to 9% (IQR 2–21%). The Bland–Altman plot shows the manual and AI-supported tMTV values (Figure 2). The line graph in Figure 3 indicates significantly less variability in tMTV between two specialists' segmentations when using the AI tool compared with manual segmentations ( $p = 0.023$ ). The median difference in tMTV between measurements with and without AI was 3.3 cm<sup>3</sup> corresponding to 2.3% of the median tMTV. In 31 (65%) of the 48 cases, the absolute difference in tMTV between two specialists were lower using the AI tool: in one case, no difference was found with and without the AI tool, and, in 16 (33%) cases, the difference was higher with the AI tool (Figure 3).



**Figure 2.** Bland–Altman plot showing the difference between individual specialists' tMTV values without (X) and with (O) AI support, and the mean of all 4 tMTV values.



**Figure 3.** Line graph showing the difference in total metabolic tumour volume (tMTV) between two specialists' manual segmentations versus adjustments of artificial intelligence (AI) segmentations. Each color line represent the change in each case.

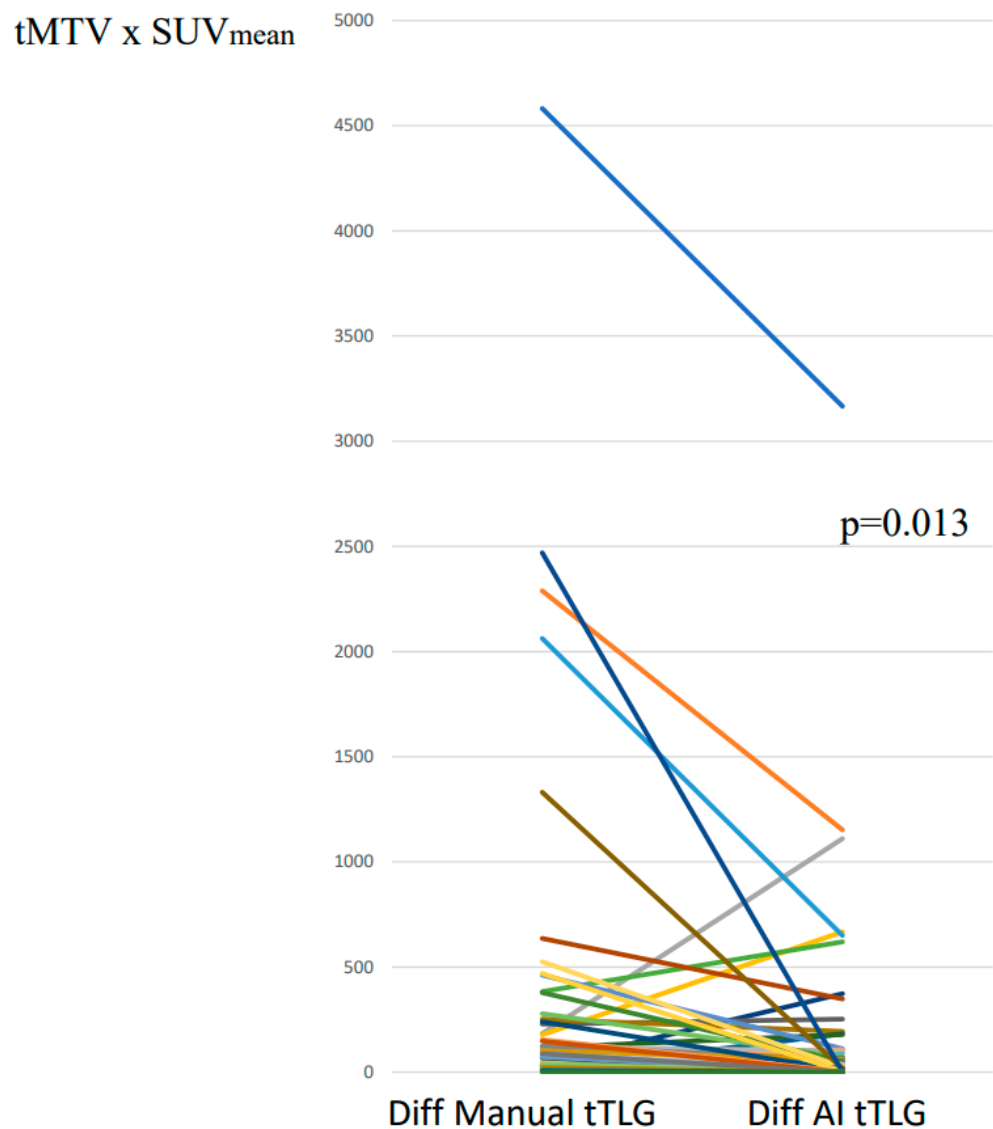
### 3.2. tTLG: Segmentations with and Without AI-Tool

The median difference between two specialists' manual tTLG segmentations was 89 (IQR 29–259) corresponding to 14% (IQR 4–32%), while the median difference between two specialists' tTLG adjustments of the AI segmentations was 45 (IQR 8–128) corresponding to 6% (IQR 1–13%). The line graph in Figure 4 indicates significantly less variability in tTLG between two specialists' segmentations when using the AI tool compared with manual segmentations ( $p = 0.013$ ). The median difference in tTLG between measurements with and without AI was 0.1 corresponding to 0.0% of the median tTLG.

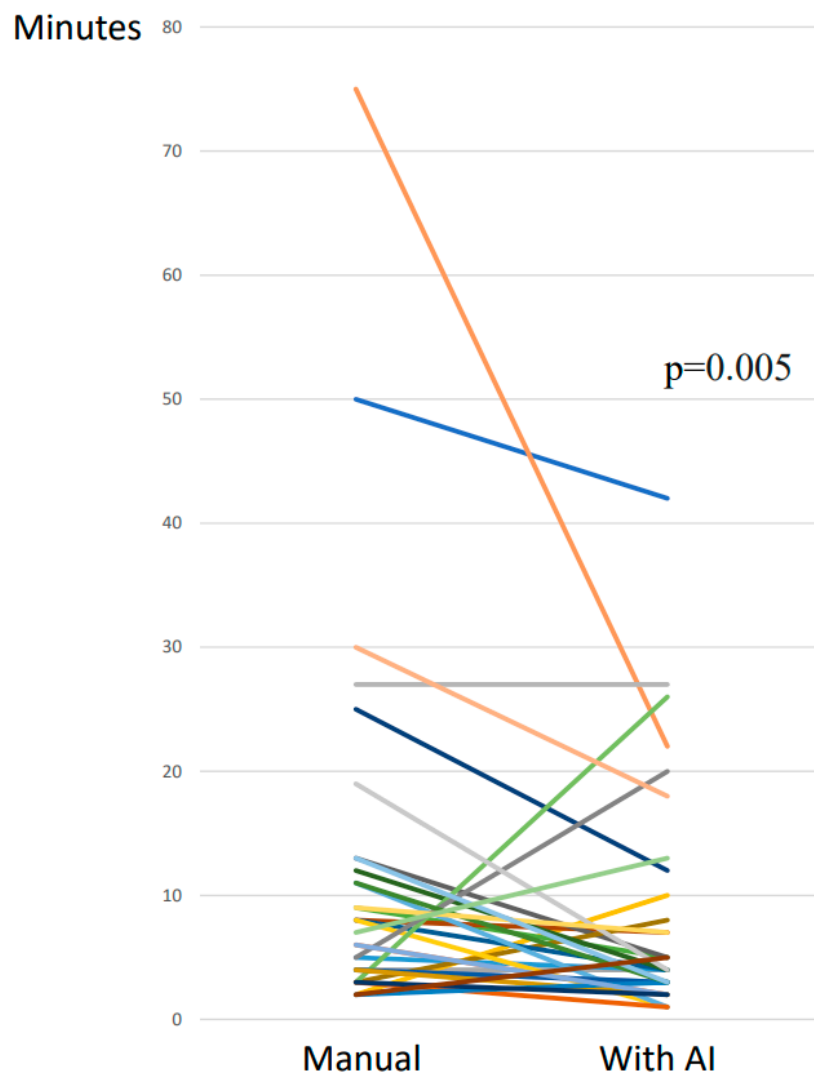
### 3.3. Time Registration

Only four specialists reported the time for each case, as intended. In 32 of the 48 cases, at least one time report with and one time report without the AI tool were available. In cases with two time reports for the same method, one of the time reports was randomly selected.

The median time taken for the manual segmentations was 7.5 min (IQR 4–12), while the median time taken for the adjustments of the AI segmentations was 4 min (IQR 3–10.5). The median difference between the time taken for the adjustment of the AI segmentations versus manual segmentations decreased significantly ( $p = 0.005$ ) by 2 min (IQR 0–8) (Figure 5). In 23 (72%) of the 32 cases, the time decreased using AI compared with manual segmentations; in 2 (6%) cases, the time required was the same, while, in 7 (22%) cases, the time increased using AI.



**Figure 4.** Line graph showing the difference in total tumour lesion glycolysis (tTLG) between two specialists' manual segmentations versus adjustments of artificial intelligence (AI) segmentations. Each color line represent the change in each case.



**Figure 5.** Line graph showing time taken for manual versus artificial intelligence (AI) segmentations. Each color line represent the change in each case.

#### 4. Discussion

The present study demonstrated that an AI-based-tool for the automated quantification of the total metabolic tumour burden (tMTV and tTLG) in untreated HL patients staged with [ $^{18}\text{F}$ ]FDG PET/CT could assist nuclear medicine specialists in segmenting lesions and significantly improve inter-observer agreement ( $p = 0.023$  (for tMTV) and  $p = 0.013$  (for tTLG)).

The problem with large deviations in the tMTV between different readers was highlighted in a recent study by Boellaard et. al. in which 10 readers from different hospitals analysed five PET/CT studies from patients with B-cell lymphoma [10]. Despite giving instructions on how to segment the lesions, the authors found large deviations in a first round, because some readers did not follow the instructions but rather acted as they normally do in everyday clinical work. The deviations reported in the first round could indicate how large the tMTV variations are in the clinical routine currently. The authors concluded that there is an urgent need to improve tMTV segmentation workflows in clinical practice. The results in the present study indicate that AI could play an important role in increasing the agreement in tMTV values in clinical practice and research trials.

The median time taken for the manual segmentations was 7.5 min and it decreased significantly when using the AI tool (Figure 5). The analysis time for tMTV segmentations

without AI support was reported to be similar in the study of Boellaard et. al. [10], despite the fact that the readers in their study used the software they were used to from their clinical work, while, in this study, they used the RECOMIA platform.

It was not within the scope of the present study to assess the accuracy of the tMTV values. However, the change in tMTV values with and without AI was only 3.3 cm<sup>3</sup> corresponding to 2.3% of the median tMTV, indicating that this change is not clinically important. As no gold standard exists to validate specialist segmentations due to the unknown ground truth, we compared the tMTV and tTLG measurements obtained by eight experts with each other. A similar approach was used in a study by Boellaard et al. who presented a benchmark method as an alternative to a gold standard. tMTV was measured by 12 PET/CT lymphoma experts in order to establish a reference value for each of the 60 PET/CT studies from lymphoma patients [11]. A strength of our study is that we included eight specialists from eight different hospitals around the world which makes the results generalizable. Both their and our study aim to contribute in different ways to allow tMTV to be a widely used reproducible biomarker.

Several reports have demonstrated that the baseline metabolic tumour burden on [18F]FDG PET/CT has prognostic potential in lymphoma [12–14]; however, no consensus exists on how to measure tMTV and tTLG [15–17]. Semi-automated methods based on absolute or relative SUV thresholds have been reported to frequently under- or overestimate visible tumours [1,3]. Therefore, we developed a novel approach using AI, trained to mimic human readers, avoiding a specific threshold method.

Despite attempts to validate former threshold methods by phantom studies, no cut-off boundaries have yet been agreed upon to discriminate a good from bad prognosis, nor a cut-off value that justifies a more aggressive treatment regimen [13]. To answer these questions, large cohorts of patients with long follow-up period are needed. We instructed the readers to segment the FDG uptake based on the recommendations by Barrington et. al. [1]. No SUV threshold boundaries were recommended, nor minimum lesion size. This approach mimics the everyday clinical work, which is the intention of this study, rather than to present results from a more artificially arranged study setting.

We aimed to develop an AI tool specifically designed for HL patients and trained the tool using lesions from patients with HL, outlined by two nuclear medicine specialists. To our knowledge, no such AI tool has specifically been developed for adult HL patients, meaning a comparison with other published reports is difficult. Weisman et. al. has developed an AI tool for paediatric HL patients and showed excellent agreement with manual segmentations carried out by specialists both for tMTV and tTLG [18].

Our test set comprised 48 untreated HL patients and eight physicians from different hospitals analysing 24 cases each. These sample sizes were not based on formal power calculations, but experience of what is feasible. The design is similar to the studies by Boellaard et al., who included 10 readers analysing 5 cases each and 12 readers analysing 20 cases each, respectively [10,11].

The segmentation of the total tumour burden is time-consuming and currently not performed in everyday clinical work. We have showed that using AI could significantly reduce the segmentation time and hope that this could increase the feasibility and willingness of reporters to quantify tMTV and tTLG in the daily routine. Our results should be interpreted with caution since only half of the physicians remembered to register the time taken for each case.

Patients were included from a single institution, who had undergone staging for [18F]FDG PET/CT using the same PET/CT system. Interestingly, this small training group could achieve significant improvements, which is promising and suggests that the model has learned robust features even from a limited set. Future work will focus on increasing

the training cohort with patients from different hospitals, examined using different cameras to show more lesion varieties to the AI tool. This improved AI tool will also be validated using the standardised tMTV benchmark dataset created by Boellaard and co-workers [11].

In a previous work, we have described the training and testing of the RECOMIA AI tool used here [5]. The current work investigated whether the RECOMIA tool can increase the agreement between specialists quantifying the baseline metabolic tumour burden and decrease the time needed for that. In both papers, the same patient group are used.

The limitations include the fact that the specialists were not familiar with the software used in this study; however, the cases were randomly distributed to the readers, with some starting with manual segmentations while others started with the adjustment of AI segmentations where required. The pairwise comparison of the segmentations, both without and with AI, was made without the knowledge of which segmentation procedure was performed first. Another limitation might be that the comparator with manual segmentation may not be entirely fair, since most commercial software includes tools that facilitate segmentation, such as region-growing algorithms that propagate contours across slices or automatic contouring based on SUV thresholds. However, the manual tool could be adjusted only to segment voxels with SUV values above a specific threshold. Furthermore, no statical adjustments were applied for the three variables investigated (tMTV, tTLG, and time).

The scope of this study was not to test the impact of an AI tool ready for clinical use—to develop such a tool requires a step-by-step huge effort to be taken, including pre-analytical, analytical, and clinical validation of the final product [19–21]—nor was the intention to investigate clinical significance of the median reduction of the inter-observer variability of tMTV and tTLG. Large prospective trials are needed for the latter. Furthermore, the test set of 48 patients is too small for a sub-analysis. At present, we are in the early phases of tool development for PET/CT in HL patients. The aim of this project was to investigate if the current tool could improve inter-observer agreement in the quantitative results and reduce the segmentation time and we showed that this could be accomplished despite that the AI tool being based on a relatively small training set, manually annotated by two nuclear medicine specialists. Moreover, the patient examinations used in this study cannot be publicly shared due to ethical considerations. However, the AI tool can freely be accessed on [recomia.org](http://recomia.org).

## 5. Conclusions

An automated AI-based-tool can significantly reduce the segmentation time and improve agreement among specialists quantifying the metabolic tumour burden in HL patients staged with [18F]FDG PET/CT, without markedly altering the absolute tMTV and tTLG measurements. Future work will aim at increasing the training group in order to present more variations to the network and test the tool with an external patient cohort.

**Author Contributions:** Conceptualization, M.S., S.F.B., J.U., O.E., E.T., J.L.U. and L.E.; Methodology, M.S., S.F.B., J.U., O.E., E.T., J.L.U. and L.E.; Software, J.U. and O.E.; Validation, J.U., O.E., E.T., A.L., J.L.L.G., R.K., M.v.E. and L.E.; Formal analysis, M.S., S.F.B., J.U., O.E., B.S., A.L.N., A.L., J.L.L.G., J.L.U., R.K., M.v.E. and L.E.; Investigation, M.S., E.T., B.S., A.L.N., A.L., J.L.L.G., J.L.U., R.K., M.v.E. and L.E.; Resources, M.S., E.T., B.S., A.L.N., A.L., J.L.L.G., J.L.U., R.K., M.v.E. and L.E.; Data curation, S.F.B., J.U., O.E., E.T., B.S., A.L.N., A.L., J.L.L.G., J.L.U., R.K., M.v.E. and L.E.; Writing – original draft, M.S., J.U. and L.E.; Writing—review & editing, M.S., S.F.B., J.U., O.E., E.T., B.S., A.L.N., A.L., J.L.L.G., J.U., R.K., M.v.E. and L.E.; Visualization, A.L., J.L.L.G., J.L.U., R.K. and M.v.E.; Supervision, S.F.B., O.E., J.L.U. and L.E.; Project administration, L.E.; Funding acquisition, S.F.B. and L.E. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded by the Swedish State through the ALF agreement (70380), established between the Swedish Government and the County Councils. The funding source had no involvement in the study's conceptualisation, design, data collection, analysis, decision to publish, or manuscript preparation. S.F.B. acknowledges the support from the National Institute for Health and Care Research (NIHR) [RP-2-16-07-001]. The King's College London and UCL Comprehensive Cancer Imaging Centre is funded by Cancer Research UK (CRUK) and the Engineering and Physical Sciences Research Council (EPSRC), in collaboration with the Medical Research Council (MRC) and the Department of Health and Social Care (England). Additional support was provided by core funding from the Wellcome/EPSCRC Centre for Medical Engineering at King's College London [WT203148/Z/16/Z], the NIHR Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London, and the NIHR Clinical Research Facility. The views expressed are those of the author(s) and do not necessarily reflect those of the NHS, the NIHR, or the Department of Health and Social Care.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Gothenburg University, Etikprövningsmyndigheten, which also waived the requirement for written informed consent (#2019-01274) on 28 March 2019.

**Informed Consent Statement:** The requirement for written informed consent was waived due to the ethic committee decision.

**Data Availability Statement:** The datasets generated and/or analysed during this study are not publicly available due to ethical restrictions.

**Acknowledgments:** During the preparation of this manuscript, the authors used the AI tool "Lymphoma FDG-PET/CT segmentation:1.0" available at the RECOMIA research platform (recomia.org) for the purpose of conducting the corresponding study. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

**Conflicts of Interest:** J.U. and O.E. are board members and stockholders of Eigenvision AB, which is a company working with research and development in automated image analysis, computer vision, and machine learning. The other authors declare that they have no conflict of interest.

## References

1. Barrington, S.F.; Meignan, M. Time to Prepare for Risk Adaptation in Lymphoma by Standardizing Measurement of Metabolic Tumor Burden. *J. Nucl. Med.* **2019**, *60*, 1096–1102. [[CrossRef](#)]
2. Weisman, A.J.; Kieler, M.W.; Perlman, S.; Hutchings, M.; Jeraj, R.; Kostakoglu, L.; Bradshaw, T.J. Comparison of 11 automated PET segmentation methods in lymphoma. *Phys. Med. Biol.* **2020**, *65*, 235019. [[CrossRef](#)]
3. Barrington, S.F.; Zwezerijnen, B.G.J.C.; de Vet, H.C.W.; Heymans, M.W.; Mikhaeel, N.G.; Burggraaff, C.N.; Eertink, J.J.; Pike, L.C.; Hoekstra, O.S.; Zijlstra, J.M.; 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. [[CrossRef](#)]
4. Pomykala, K.L.; Fendler, W.P.; Vermesh, O.; Umutlu, L.; Herrmann, K.; Seifert, R. Molecular Imaging of Lymphoma: Future Directions and Perspectives. *Semin. Nucl. Med.* **2023**, *53*, 449–456. [[CrossRef](#)]
5. Sadik, M.; Barrington, S.F.; Trägårdh, E.; Saboury, B.; Nielsen, A.L.; Jakobsen, A.L.; Gongora, J.L.L.; Urdaneta, J.L.; Kumar, R.; Edenbrandt, L. Metabolic tumour volume in Hodgkin lymphoma-A comparison between manual and AI-based analysis. *Clin. Physiol. Funct. Imaging* **2024**, *44*, 220–227. [[CrossRef](#)]
6. Sibille, L.; Seifert, R.; Avramovic, N.; Vehren, T.; Spottiswoode, B.; Zuehlsdorff, S.; Schäfers, M. <sup>18</sup>F-FDG PET/CT Uptake Classification in Lymphoma and Lung Cancer by Using Deep Convolutional Neural Networks. *Radiology* **2020**, *294*, 445–452. [[CrossRef](#)]
7. Sadik, M.; López-Urdaneta, J.; Ulén, J.; Enqvist, O.; Krupic, A.; Kumar, R.; Andersson, P.O.; Trägårdh, E. Artificial intelligence could alert for focal skeleton/bone marrow uptake in Hodgkin's lymphoma patients staged with FDG-PET/CT. *Sci. Rep.* **2021**, *17*, 10382.
8. Sadik, M.; Lind, E.; Polymeri, E.; Enqvist, O.; Ulén, J.; Trägårdh, E. 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* **2019**, *39*, 78–84. [[CrossRef](#)]

9. Trägårdh, E.; Borrelli, P.; Kaboteh, R.; Gillberg, T.; Ulén, J.; Enqvist, O.; Edenbrandt, L. RECOMIA—a cloud-based platform for artificial intelligence research in nuclear medicine and radiology. *EJNMMI Phys.* **2020**, *7*, 51. [[CrossRef](#)]
10. Boellaard, R.; Zwezerijnen, G.J.C.; Buvat, I.; Champion, L.; Hovhannisyani-Baghdasarian, N.; Orhac, F.; Arens, A.I.J.; Lobeek, D.; Celik, F.; Mitea, C.; et al. Measuring Total Metabolic Tumor Volume from 18F-FDG PET: A Reality Check. *J. Nucl. Med.* **2025**, *66*, 802–805. [[CrossRef](#)]
11. Boellaard, R.; Buvat, I.; Nioche, C.; Ceriani, L.; Cottreau, A.S.; Guerra, L.; Hicks, R.J.; Kanoun, S.; Kobe, C.; Loft, A.; 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. [[CrossRef](#)]
12. Prieto Prieto, J.C.; Casas, J.A.V.; Hatzimichael, E.; Fotopoulos, A.; Kiortsis, D.N.; Sioka, C. The contribution of metabolic parameters of FDG PET/CT prior and during therapy of adult patients with lymphomas. *Ann. Nucl. Med.* **2020**, *34*, 707–717. [[CrossRef](#)]
13. Meignan, M.; Cottreau, A.S.; Specht, L.; Mikhaeel, N.G. Total tumor burden in lymphoma—An evolving strong prognostic parameter. *Br. J. Radiol.* **2021**, *94*, 20210448. [[CrossRef](#)]
14. El-Galaly, T.C.; Villa, D.; Cheah, C.Y.; Gormsen, L.C. Pre-treatment total metabolic tumour volumes in lymphoma: Does quantity matter? *Br. J. Haematol.* **2022**, *197*, 139–155. [[CrossRef](#)]
15. Meignan, M.; Cottreau, A.S.; Versari, A.; Chartier, L.; Dupuis, J.; Boussetta, S.; Grassi, I.; Casasnovas, R.O.; Haioun, C.; Tilly, H.; 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. [[CrossRef](#)]
16. Cottreau, A.S.; Versari, A.; Loft, A.; Casasnovas, O.; Bellei, M.; Ricci, R.; Bardet, S.; Castagnoli, A.; Brice, P.; Raemaekers, J.; 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. [[CrossRef](#)] [[PubMed](#)]
17. Mikhaeel, N.G.; Heymans, M.W.; Eertink, J.J.; de Vet, H.C.W.; Boellaard, R.; Dührsen, U.; Ceriani, L.; Schmitz, C.; Wiegers, S.E.; Hüttmann, A.; et al. Proposed New Dynamic Prognostic Index for Diffuse Large B-Cell Lymphoma: International Metabolic Prognostic Index. *J. Clin. Oncol.* **2022**, *40*, 2352–2360. [[CrossRef](#)]
18. Weisman, A.J.; Kim, J.; Lee, I.; McCarten, K.M.; Kessel, S.; Schwartz, C.L.; Kelly, K.M.; Jeraj, R.; Cho, S.Y.; Bradshaw, T.J. Automated quantification of baseline imaging PET metrics on FDG PET/CT images of pediatric Hodgkin lymphoma patients. *EJNMMI Phys.* **2020**, *7*, 76. [[CrossRef](#)]
19. Anand, A.; Morris, M.J.; Kaboteh, R.; Reza, M.; Trägårdh, E.; Matsunaga, N.; Edenbrandt, L.; Bjartell, A.; Larson, S.M.; Minarik, D. A Preanalytic Validation Study of Automated Bone Scan Index: Effect on Accuracy and Reproducibility Due to the Procedural Variabilities in Bone Scan Image Acquisition. *J. Nucl. Med.* **2016**, *57*, 1865–1871. [[CrossRef](#)]
20. Anand, A.; Morris, M.J.; Kaboteh, R.; Báth, L.; Sadik, M.; Gjertsson, P.; Lomsky, M.; Edenbrandt, L.; Minarik, D.; Bjartell, A. Analytic Validation of the Automated Bone Scan Index as an Imaging Biomarker to Standardize Quantitative Changes in Bone Scans of Patients with Metastatic Prostate Cancer. *J. Nucl. Med.* **2016**, *57*, 41–45. [[CrossRef](#)]
21. Armstrong, A.J.; Anand, A.; Edenbrandt, L.; Bondesson, E.; Bjartell, A.; Widmark, A.; Sternberg, C.N.; Pili, R.; Tuveson, H.; Nordle, Ö.; et al. Phase 3 Assessment of the Automated Bone Scan Index as a Prognostic Imaging Biomarker of Overall Survival in Men With Metastatic Castration-Resistant Prostate Cancer: A Secondary Analysis of a Randomized Clinical Trial. *JAMA Oncol.* **2018**, *4*, 944–951. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.