J of Nuclear Medicine Technology, first published online April 19, 2022 as doi:10.2967/jnmt.121.262900

VALIDATION OF CONVOLUTIONAL NEURAL NETWORKS FOR FAST DETERMINATION OF WHOLE-BODY METABOLIC TUMOR BURDEN IN PEDIATRIC LYMPHOMA

Elba Etchebehere[1,2], Rebeca Andrade[1], Mariana Camacho[2], Mariana Lima[1,2], Anita Brink[3], Juliano Cerci[4], Helen Nadel[5], Chandrasekhar Bal[6], Venkatesh Rangarajan[7],Thomas Pfluger[8], Olga Kagna[9], Omar Alonso[10], Fatima K. Begum[11], Kahkashan Bashir Mir[12], Vincent Peter Magboo[13], Leon J. Menezes[14], Diana Paez*[15], Thomas NB Pascual*[15].

1. University of Campinas, Campinas, BRAZIL;
2. Medicina Nuclear de Campinas, Campinas, BRAZIL;
3. University of Cape Town, Cape Town, SOUTH AFRICA;
4. QUANTA Diagnóstico e Terapia, Curitiba, BRAZIL;
5. University of British Columbia, Vancouver, CANADA;
6. All India Institute of Medical Sciences, New Delhi, INDIA;
7. Tata Memorial Centre, Mumbay, INDIA;
8. Ludwig-Maximillian University of Munich, Munich, GERMANY;
9. Rambam Health Care Campus, Haifa, ISRAEL;
10. Centro Uruguayo de Imagenología Molecular, Montevideo, URUGUAY;

11. National Institute of Nuclear Medicine and Allied Sciences, Dhaka, BANGLADESH;

12. Nuclear Medicine, Oncology and Radiotherapy Institute, Islamabad, PAKISTAN;

13. University of the Philippines, Manila, PHILIPPINES;

14. Institute of Nuclear Medicine, London, UNITED KINGDOM;

15. Nuclear Medicine and Diagnostic Imaging Section, Division of Human Health, International Atomic Energy Agency, Vienna, AUSTRIA

** The last two authors T.N.B. P and D.P. contributed equally*

Corresponding author:

Elba Etchebehere, MDPhD
Serviço de Medicina Nuclear do Hospital de Clínicas
R. Vital Brasil, 251 - Cidade Universitária, Campinas - SP, Brasil.
CEP: 13083-888
+55 (19) 3521 7772
E-mail: elba@hc.unicamp.br

Word Count: 3608 (excluding title page, tables and references)

Short Title: **Tumor Burden in Pediatric Lymphoma**

Conflict of Interest: **None to declare.**

ABSTRACT

INTRODUCTION: 18F-FDG PET/CT whole-body tumor burden in lymphoma is not routinely performed due to the lack of fast quantification methods. Although the semi-automatic method is fast, it still lacks the necessary speed required to quantify tumor burden in daily clinical practice.

PURPOSE: To evaluate the performance of the convolutional neural networks (CNN) software to localize neoplastic lesions in whole-body 18F-FDG PET/CT images of pediatric lymphoma patients.

METHODS: This retrospective image data set, derived from the data pool under the IAEA (CRP# E12017), included 102 baseline staging 18F-FDG PET/CTs of pediatric lymphoma patients (mean age 11 yrs). Images were quantified to determine the whole-body (wb) tumor burden (wbMTV and wbTLG) using a semi-automatic (**SEMI**) software and an **CNN-based** software. Both were displayed as $\text{wbMTV}_{\mathit{SEMI}}$ & $\text{wbTLG}_{\mathit{SEMI}}$ and $\text{wbMTV}_{\mathit{CNN}}$ & $\text{TLG}_{\mathit{CNN}}$*.* The intraclass correlation coefficient (ICC) was applied to evaluate concordance between the **CNN**-based software and the **SEMI** software.

RESULTS: Twenty-six patients were excluded from the analyses because the software was unable to perform calculation. In the remaining 76 patients, $\text{wbMTV}_{\mathit{CNN}}$ and $\text{wbMTV}_{\mathit{SEMI}}$ whole-body tumor burden metrics were highly correlated (ICC=0.993; 95%CI: 0.989 -0.996; p-value<0.0001) as were $\text{wbTLG}_{\mathit{CNN}}$ and $\text{wbTLG}_{\mathit{SEMI}}$ (ICC=0.999; 95%CI: 0.998-0.999; p-value<0.0001). However, the time spent calculating these metrics was significantly (<0.0001) faster by CNN (mean =

19 **seconds**; 11 - 50 seconds) compared to the semi-automatic method (mean = 21.6 **minutes**; 3.2 – 62.1 minutes), especially in patients with advanced disease.

CONCLUSION: Determining whole-body tumor burden in pediatric lymphoma patients using CNN is fast and feasible in clinical practice.

KEYWORDS: 18F-FDG PET/CT. Whole-body tumor burden. Pediatric. Lymphoma.

INTRODUCTION

Positron emission computed tomography with fluoro-deoxyglucose labeled with fluorine-18 (18F-FDG PET/CT) is an established modality for pediatric staging of Hodgkin's lymphoma and Non-Hodgkin's lymphoma as well as an invaluable tool for treatment response evaluation (1–7). Visual interpretation of 18F-FDG PET/CT studies to assess the extent of disease can be subjective; therefore qualitative interpretation is necessary to provide additional insight, reducing the subjectivity of visual interpretation (8,9). 18F-FDG PET/CT whole-body metabolic tumor burden parameters such as metabolic tumor volume (MTV) and total lesion glycolysis (TLG) bear a high prognostic value in lymphoma patients, much greater than SUV values (10–13). However, the prognostic determination, although easily measured in primary solid tumors (14–17), have not been applied in daily clinical practice in patients with widespread lymphoma disease because calculations are extremely time-consuming.

There is a wide variety of methods to quantify MTV and TLG, using threshold-based or algorithm-based methods. Specifically relating to the threshold-based methods, the most commonly applied is the volume of interest (VOI) isocontour method (15,17,18). Automatic multifocal segmentation quantification in patients with lymphoma uses VOI isocontour and has been validated before and proven to be quite fast [19]. Depending on patient tumor burden, the time spent calculating MTV and TLG could be impractical and still not feasible in daily clinical practice. The extraction and processing of imaging features from radiological data, also known as radiomics, may also link imaging features with patient outcome. However, radiomics also requires precise tumor ROI delineation, which is also time-consuming with delineation variabilities between observers.

Currently, computer deep learning and functioning as a neural network have evolved substantially, achieving remarkable success in tumor segmentation and diagnosis and ultimately transforming and optimized clinical practice (18,20, 21, 22, 23), providing objective and accurate diagnoses in medicine by building diagnostic models. For example, software for multi-modality imaging using deep convolutional neural networks (*d*CNN) method automatically localizes and delineates metastases in whole-body 18F-FDG PET/CT scans. *d*CNN seems capable of correctly localizing and classifying uptake patterns in 18F-FDG PET/CT images into foci suspicious and non-suspicious for cancer. These extracted features help the semantic interpretation, and may simplify the PET workflow with a one-click calculation of whole-body tumor burden (24,25, 26). However, the clinical applicability of this software has not yet been fully tested and unusual features may be identified if unsupervised by a physician (27,28).

The purpose of this study was to evaluate the performance of the recently developed CNN software in a clinical setting in pediatric lymphoma patients.

MATERIAL AND METHODS

This data set, retrospectively studied, is derived from a subset of 102 baseline staging 18F-FDG PET/CTs of pediatric lymphoma patient images from the data pool of the prospective multicenter research project coordinated by the International Atomic Energy Agency (IAEA) (CRP# E12017).

Research Regulation and Data Protection

The study protocol was approved by each center’s Institutional Review Board. A signed parental consent was an inclusion criterion for recruitment and all subjects signed a written informed consent. Cases and forms were anonymized to ensure confidentiality while sharing data internationally.

Patients

The eligibility criteria consisted of pediatric patients (age<18 years) with newly diagnosed Hodgkin’s lymphoma or non-Hodgkin’s lymphoma who underwent a staging 18F-FDG PET/CT scan. According to the World Health Organization classification criteria, the diagnosis was based on biopsy with immunohistochemistry (29). Exclusion criteria consisted of prior radiation therapy and chemotherapy and concurrent HIV infection.

The patient's clinical characteristics and tumor staging were evaluated, such as the age of diagnosis, the final clinical-stage, spleen disease, additional nodal sites, disease volume, B-symptoms, LDH, leukocytosis, elevated erythrocyte sedimentation rate, anemia, albumin, bone marrow 18F-FDG uptake, Deauville, MTV and TLG criteria.

18F-FDG PET/CT Imaging and quantification

All patients underwent a staging whole-body 18F-FDG PET/CT, from the top of the skull to the toes. All scans were performed according to standard SNMMI or EANM procedure guidelines (30).

The whole-body tumor burden (wbMTV and wbTLG) metrics were calculated using semi-automatic (SEMI) and convolutional neural networks (CNN) softwares. All images on both software were processed by two observers (M.R.C. & R.A.A). Differences in the wbMTV and wbTLG metrics (if any) were re-calculated to reach consensus. The *SEMI* software was used as the reference standard to evaluate the *CNN* software's performance.

Whole-body tumor burden semi-automatic quantification (SEMI)

The whole-body tumor burden (wbMTV and wbTLG) metrics were calculated using semi-automatic multifocal segmentation (*SEMI*) software (*Syngovia VB20* - Siemens Medical Solutions, Chicago, IL), previously validated for clinical use (19) using a fixed threshold.

With this software the whole-body tumor burden metrics (wbMTV$_{SEMI}$ and wbTLG$_{SEMI}$) were obtained. The SEMI whole-body tumor burden was performed by choosing the multifocal segmentation tool that automatically draws a rectangular VOI around the patient's entire body on the coronal axis. If necessary, the VOI is adjusted on the axial and sagittal planes. The liver is set as the background reference, and then volumes of interest are automatically determined surrounding each lymphoma lesion with uptake higher than the mean SUV of the liver. A VOI threshold of 41% of the SUVmax using isocontour drawings was applied for all automatically delineated lesions. The image and VOIs were then reviewed to exclude physiologic areas incorrectly selected as cancer (such as brain, kidneys, bladder, and ureters) and include metastatic foci with relatively low uptake were missed by the software (e.g., small lymph nodes). Afterward, whole-body MTV and TLG calculations were readily available and displayed as wbMTV$_{SEMI}$ and wbTLG$_{SEMI}$ (Figure 1).

Whole-body tumor burden Convolutional Neural Networks (CNN)

The whole-body tumor burden (wbMTV and wbTLG) metrics were calculated using *CNN* software based on *d*CNN *(Syngovia VB50* - Siemens Medical Solutions, Chicago, IL). The quantification was undertaken on a loaned Siemens equipment. With this software the whole-body tumor burden metrics (wbMTV$_{\mathrm{CNN}}$ and wbTLG$_{\mathrm{CNN}}$) were obtained.

The computation of the whole-body tumor burden on the CNN software was automatically performed by the *d*CNN method as described by Sibille L, et al [24]. Unlike the SEMI software, the CNN software does not require an initial positioning of a VOI surrounding the body. The CNN automatically computes the MIP 18F-FDG PET image and integrates the anatomical CT image using an intuitive interface. Afterward, the software automatically detects 18F-FDG-avid anatomical landmarks and discriminates hypermetabolic areas related to the physiologic activity that will be automatically excluded (Figure 2) from cancer. Briefly, the PET VOIs are segmented using a fixed threshold algorithm and evaluated by the *d*CNN. Whole-body CT examinations are aligned to an anatomic atlas. Finally, a MIP of the whole-body 18F-FDG PET/CT is reconstructed, and the lesions are classified. The *d*CNN uses a combination of multi-planar reconstructions of PET and CT, 18F-FDG PET MIPs, and anatomic atlases to predict the anatomic localization of 18F-FDG foci and determine whether a focus was suspicious (or not) for malignancy. The advantage of the CNN algorithm is that it does not require the initial positioning of a VOI. At the moment, this specific CNN software is not validated for pediatric patients.

Two forms of analyses were undertaken on the CNN software:

1. **Observer method:** All VOIs automatically generated by the multifocal segmentation tool were reviewed (blindly) by both observers to determine if the VOIs were wrongly included or excluded from the results. Afterward, values were calculated and displayed as $\text{wbMTV}_{\textit{CNN+observer}}$ and $\text{wbTLG}_{\textit{CNN+observer}}$.

2. **No-observer method**: The VOIs automatically obtained were accepted and not reviewed blindly by each of the observers. The calculations were readily available and displayed as $\text{wbMTV}_{\textit{CNN}}$ and $\text{wbTLG}_{\textit{CNN}}$.

Statistical Analysis

The sample was characterized by descriptive analysis, performed using frequency tables for categorical variables and measures of position and dispersion for continuous variables (mean values, standard deviation, median, minimum and maximum).

Chi-square test or Fisher's exact test were used to check associations or compare proportions, well the comparison of continuous or orderable measurements between the two groups was undertaken by the Mann-Whitney test. Identification of risk factors associated with the event was performed with univariate and multiple Cox regression analyses. The variable selection process employed was stepwise.

To verify the relationship between continuous measurements, Spearman's correlation coefficient was used wearing from -1 to 1.

To assess agreement between whole-body tumor burden quantification of the SEMI and CNN software, the intraclass correlation coefficient (Intraclass Correlation Coefficient - ICC) was used (values above 0.7 were considered as indicating substantial agreement). The Friedman's test was used to compare the times. The Wilcoxon test for related samples was used to compare the times. The time was defined when the physician began focusing on the task until completion of whole-body tumor burden calculation.

The level of significance adopted was 0.05.

RESULTS

The quantification of whole-body tumor burden was undertaken in a total of 102 18F-FDG PET/CT baseline scans of pediatric lymphoma patients using both software. There were 32 (31.4%) female patients and 70 (68.6%) males. The mean age of lymphoma diagnosis was 11.1 ± 4.3 years (ranging from 4.0 to 18.0 years). Among these, 80 (78.4%) patients had Hodgkin's lymphoma, and 22 (21.6%) patients had non-Hodgkin's lymphoma. Table 1 displays all patient's clinical characteristics.

Semi-automatic calculation of whole-body tumor burden (SEMI)

The $wbMTV_{SEMI}$ and $wbTLG_{SEMI}$ were undertaken in all 102 patients. The average time spent calculating $wbMTV_{SEMI}$ and $wbTLG_{SEMI}$ was 21.6 minutes, ranging from 3.2 – 62.1 minutes. Notably, in patients with widespread lesions in multiple organs or confluent with areas of physiological excretion, the software took longer to identify and delineate abnormal areas.

Convolutional neural network-based calculation of tumor burden (CNN)

The $wbMTV_{CNN+observer}$ and $wbTLG_{CNN+observer}$ were undertaken in all 102 patients. The average time spent calculating $wbMTV_{CNN+observer}$ and $wbTLG_{CNN+observer}$, with the CNN software having the observers evaluate the images before calculation, was 3.8 minutes, ranging from 0.5 – 19.6 minutes.

On the other hand, $wbMTV_{CNN}$ and $wbTLG_{CNN}$ (that is, without any observer evaluating the CNN software's performance before calculation) were undertaken in

76 of the 102 patients. Twenty-six patients were excluded from the analyses because the software was unable to perform calculation due to patient movement/misregistration (n=6), software non-recognition of small lymph nodes as a disease (n=8), widespread brown fat (n=3), diffuse bone infiltration (n=5), diffuse homogeneous mild infiltration of the spleen (n=2), and subcutaneous infiltration of 18F-FDG in the injection site (n=2) (Figure 3).

Impressively, the average total time spent calculating wbMTV$_{CNN}$ and wbTLG$_{CNN}$ was 19 seconds, ranging from 11 to 50 seconds. The total time relates to the time starting when the physician began focusing on the task until completion of whole-body tumor burden calculation. Thus, the time spent calculating wbMTV$_{CNN}$, wbMTV$_{CNN+observer,}$ and wbMTV$_{SEMI}$ metrics in 76 paired patients were significantly different (p<0.0001). The CNN software alone was much faster and more precise than both the SEMI and the CNN+observer methods (Table 2).

Comparison of SEMI and CNN tumor burden measurements

The wbMTV$_{CNN+\text{observer}}$ and wbMTV$_{SEMI}$ metrics calculated on 102 patients were highly correlated (ICC=0.993; 95%CI = 0.989 - 0.996; p<0.0001) as were the wbTLG $_{CNN+\text{observer}}$ and wbTLG$_{SEMI}$ (ICC=0.999; 95%CI = 0.998 - 0.999; p<0.0001). Among the 76 18F-FDG PET/CTs in which the fully automatic CNN was performed, the wbMTV$_{CNN+observer}$, wbMTV$_{CNN,}$ and wbMTV$_{SEMI}$ metrics were also highly correlated as were the wbTLG$_{CNN+observer}$, wbTLG$_{CNN,}$ and wbTLG$_{SEMI}$ metrics (Table 3).

Impressively, the correlation between $wbMTV_{CNN}$ and $wbMTV_{SEMI}$ was significantly high (ICC = 0.950; 95%CI: 0.922-0.968; $p < 0.0001$) as was $wbTLG_{CNN}$ and $wbTLG_{SEMI}$ (ICC = 0.947; 95%CI: 0.917-0.966; $p < 0.0001$). Therefore, the CNN software performed equally well, similar to the SEMI tool in which an experienced observer evaluated the images.

More impressive, however, was the fact that the correlation between $wbMTV_{CNN+observer}$ and $wbMTV_{CNN}$ was significantly high (ICC = 0.946; 95%CI: 0.912-0.966; $p<0.0001$) as was $wbTLG_{CNN+observer}$ and $wbTLG_{CNN}$ (ICC = 0.952; 95%CI: 0.925-0.969; $p < 0.0001$). Consequently, the CNN software performance did not require an observer to evaluate the images and validate all VOIs.

DISCUSSION

To our knowledge, this is the first study to quantify the whole-body tumor burden of pediatric lymphoma patients using convolution of neural networks (CNN) and deep learning (DL). Despite the different 18F-FDG biodistribution of pediatric patients compared to adults, the CNN-based software accurately delineated abnormal regions. The CNN-based software optimized the working time, was extremely fast, and performed better than the semi-quantitative software to calculate whole-body tumor burden.

The CNN-based software allows a review of the VOIs provided automatically (adding new VOIs manually or deleting incorrect VOIs). Ultimately the comparison of the CNN-based software with and without the observer's review of the VOIs rendered the same metrics. However, the time spent determining the whole-body tumor burden metrics by the semi-automatic software was longer because it depends primarily on the extent of the disease. The semi-automatic quantification does not allow a pre-selection of VOIs by the operator before creating the definitive findings and thus does not distinguish diseased areas from physiological areas, creating many VOIs that overload the program.

On the other hand, quantifying the whole-body tumor burden through CNN-based software was significantly faster, with and without the observer reviewing the VOIs. Impressively, when comparing quantification of the whole-body tumor burden on the CNN-based software (without observer interference) with the semi-automatic software and CNN-based software with observer interference, CNN-based software without the interference of the observer was significantly faster and just as precise. CNN-based software took as little as 20 seconds to calculate the patient’s entire tumor burden without the need to review the VOIs (Figures 4 and 5).

However, there were some limitations. It was not possible to show whether the measurements predicted by the CNN-based software could be applied to our patients’ cohort to predict prognosis and response evaluation. The majority (80%) of the patients had Hodgkin´s lymphoma and there were only two deaths; therefore, it was not possible to determine overall survival. A larger number of patients with events are required to determine whether the measurements predicted by CNN-based software could predict prognosis. Another limitation was that 25% of the patients were excluded from analyses because the CNN-based software could not recognize areas of metabolically active disease and could not perform calculation. In such situations, these patients had to be excluded because there was no ability to compare CNN quantification with manual or semi-automatic quantification. The CNN software we tested was not initially designed nor validated to quantify specifically pediatric patients, but even so, performed quite well. These exclusions were caused by either the wrong lesion being segmented or lesions being missed. These included small lymph nodes with mild 18F-FDG uptake; including extensive brown fat as lymphomatous infiltration; not including extensive diffuse bone marrow infiltration (5/12 patients); including radiopharmaceutical extravasation sites; and including bladder catheter. Most likely, with further CNN and DL development and specific training in pediatric patients regarding the differentiation of normal biodistribution versus cancer tissue, failure rates will possibly reduce.

CNN-based software with CNN and DL still requires the input of the observer (26-28). In 25% of the patients, CNN was not able to depict the correct neoplastic tissue or added non-neoplastic tissue, thus quantification had to be excluded as the software was not perform the calculations. Therefore, errors and failure to detect proper tissue will occur even in CNN and DL software, which argues in favor for the observer input. Most likely the largest errors may be associated with unsupervised quantification.

In conclusion, CNN-based quantification of whole-body tumor burden in pediatric lymphoma patients is an emerging field. Whole-body tumor burden determination using CNN-based software is extremely fast and feasible in clinical practice in pediatric lymphoma patients. CNN-based software requires CNN and DL development and specific training in pediatric patients and the input of the observer to minimize the failure rates. Tumor burden should be evaluated in most if not all tumors and age groups for therapy purposes.

ACKGNOWLEDGMENTS

The authors would like to thank Mrs. Cleide Silva from the Statistics Research Department of the University of Campinas for her invaluable help.

KEY POINTS

QUESTION: Will the use of CNN promote fast and also reliable quantification data regarding whole-body metabolic tumor burden in 18F-FDG PET/CT pediatric lymphoma patients?

PERTINENT FINDINGS: Whole-body metabolic tumor burden quantification using CNN is highly correlated to semi-automatic quantification (ICC=0.993; 95%CI: 0.989 -0.996; p-value<0.0001).

IMPLICATIONS FOR PATIENT CARE: In addition to reliable data, implementation of CNN quantification tools in the clinical practice may be able to quickly and accurately deliver prognostic information for better patient management.

REFERENCES

1. Ansell SM, Armitage JO. Positron emission tomographic scans in lymphoma: convention and controversy. *Mayo Clinic Proceedings.* 2012;87:571–80.

2. Cheson BD. PET/CT in lymphoma: current overview and future directions. *Seminars in Nuclear Medicine.* 2018;48:76–81.

3. Kobe C, Dietlein M, Hellwig D. PET/CT for lymphoma post-therapy response assessment in hodgkin lymphoma and diffuse large b-cell lymphoma. *Seminars in Nucearl Medicine*. 2018;48:28–36.

4. Gillman J, States LJ, Servaes S. PET in pediatric lymphoma. *PET Clinics*. 2020;15:299–307.

5. D'souza MM, Jaimini A, Bansal A, et al. FDG-PET/CT in lymphoma. *Indian Journal of Radiology and Imaging*. 2013;23:354–65.

6. Cronin CG, Swords R, Truong MT, et al. Clinical utility of PET/CT in lymphoma. *American Journal of Roenthenology*. 2010;194:W91–103.

7. London K, Cross S, Onikul E, Dalla-Pozza L, Howman-Giles R. 18F-FDG PET/CT in paediatric lymphoma: comparison with conventional imaging. *European Journal of Nuclear Medicine and Molecular Imaging*. 2011;38:274–84.

8. Lin C, Itti E, Haioun C, Petegnief Y, et al. Early 18F-FDG PET for prediction of prognosis in patients with diffuse large b-cell lymphoma: suv-based assessment versus visual analysis. *The Journal of Nuclear Medicine*. 2007;48:1626–32.

9. Jhanwar YS, Straus DJ. The role of PET in lymphoma. *The Journal of Nuclear Medicine*. 2006;47:1326–34.

10. Yang J, Zhu S, Pang F, et al. Functional parameters of 18F-FDG PET/CT in patients with primary testicular diffuse large b-cell lymphoma. *Contrast Media & Molecular Imaging*. 2018;2018:8659826.

11. Jung S-H, Ahn J-S, Kim Y-K, et al. Prognostic significance of interim PET/CT based on visual, SUV-based, and MTV-based assessment in the treatment of peripheral t-cell lymphoma. *BioMed Central Cancer*. 2015;15:198.

12. Albano D, Camoni L, Giubbini R, Bertagna F. Prognostic Value of 18F-FDG PET/CT metabolic parameters in splenic marginal zone lymphoma. *Clinical Lymphoma, Myeloma & Leukemia*. 2020; 20:e897-904.

13. Wang X-Y, Zhao Y-F, Liu Y, Yang Y-K, Wu N. Prognostic value of metabolic variables of 18F-FDG PET/CT in surgically resected stage I lung adenocarcinoma. *Medicine*. 2017;96:e7941.

14. Salavati A, Duan F, Snyder BS, et al. Optimal FDG PET/CT volumetric parameters for risk stratification in patients with locally advanced non-small cell lung cancer: results from the ACRIN 6668/RTOG 0235 trial. *European Journal of Nuclear Medicine and Molecular Imaging*. 2017;44:1969–83.

15. Wang L, Bai J, Duan P. Prognostic value of 18F-FDG PET/CT functional parameters in patients with head and neck cancer: a meta-analysis. *Nuclear Medicine Communications*. 2019;40:361–9.

16. Brito AE, Mourato F, Santos A, Mosci C, Ramos C, Etchebehere E. Validation of the semiautomatic quantification of 18F-fluoride PET/CT whole-body skeletal tumor burden. *Journal of Nuclear Medicine Technology*. 2018;46:378–383.

17. Im H-J, Bradshaw T, Solaiyappan M, Cho SY. Current methods to define metabolic tumor volume in positron emission tomography: which one is better?; *Nuclear Medicine and Molecular Imaging*. 2018;52:5–15.

18. Han D, Bayouth J, Song Q, et al. Globally optimal tumor segmentation in PET-CT images: a graph-based co-segmentation method. *Information Processing in Medical Imaging*. 2011;22:245–56.

19. Camacho MR, Etchebehere E, Tardelli N, et al. Validation of a multifocal segmentation method for measuring metabolic tumor volume in hodgkin lymphoma. *Journal of Nuclear Medicine Technology*. 2020;48:30–5.

20. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthcare Journal*. 2019; 6:94–8.

21. Deo RC. Machine learning in medicine. *Circulation*; 2015;132:1920–30.

22. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. *Journal of Medical Internet Research*. 2019;21:e13659.

23. Recht MP, Dewey M, Dreyer K, et al. Integrating artificial intelligence into the clinical practice of radiology: challenges and recommendations. *European Radiology*. 2020;30:3576–84.

24. 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*. 2020;294:445–52.

25. Froelich JW, Salavati A. Artificial intelligence in PET/CT is about to make whole-body tumor burden measurements a clinical reality. *Radiology*. 2020; 294: 453–4.

26. Currie G, Hawk EK, Rohren E, Vial A, Klein R. Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging. *J Med Imaging Radiat Sci*. 2019; 50:477-487.

27. Currie G. Intelligent imaging: anatomy of machine learning. *Journal of Nuclear medicine and technology.* 2019; 47:273-281.

28. Currie G and Rohren E. medicine: the principles of artificial intelligence, machine learning and deep learning. *Seminars in nuclear medicine*. 2021; 51:102-111.

29. Jaffe ES, Barr PM, Smith SM. Understanding the new WHO classification of lymphoid malignancies: why it’s important and how it will affect practice. *American Society of Clinical Oncology Educational book*. 2017;37:535–46.

30. Boellaard R, Delgado-Bolton R, Oyen WJG, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *European Journal of Nuclear Medicine and Molecular Imaging*. 2015;42:328–54.

TABLES

Table 1. Clinical characteristics of patients (N=102).

Variable		Number	Percentage
Sex	Female	32	31.4%
	Male	70	68.6%
Lymphoma type	Hodgkin	80	78.4%
	Non- Hodgkin	22	21.6%
Clinical Final Stage	1	8	7.8%
	2	34	33.3%
	3	34	33.3%
	4	26	25.5%
Spleen Disease	Yes	29	28.4%
	No	73	71.6%
Extra Nodal Sites	0	67	65.7%
	1	15	14.7%
	≥2	20	19.6%
Disease Bulk	Bulky	63	61.8%
	Non Bulky	39	38.2%
B Symptoms	Yes	43	43.0%
	No	57	57.0%
LDH	High	47	52.8%
	Normal	42	47.2%
Leucocytosis	Yes	32	31.7%

	No	69	68.3%
Erythrocyte Sedmentation Rate	Normal	34	52.3%
	Elevated	31	47.7%
Anemia	Yes	47	47.5%
	No	52	52.5%
Albumin	Yes	27	37.0%
	No	46	63.0%
Bone Marrow 18F-FDG Uptake	Diffuse	12	11.9%
	Focal	16	15.8%
	Negative	73	72.3%
Event	Yes	10	9.8%
	No	92	90.2%
Status	Alive	101	99.0%
	Dead	1	1.0%

Table 2. Time spent quantifying whole-body tumor burden metrics on the semi-automatic software (SEMI) and CNN software with observer input (CNN+observer) and without observer input (CNN).

Variable	N	Time in seconds					p-value
		Mean	Std Dev	Min	Med	Max	
SEMI	76	1301.3	863.5	198.0	1107.0	3724.0	<0.0001
CNN + OBSERVER	76	221.1	204.4	31.0	155.0	1176.0	
CNN	76	19.6	8.0	11.0	17.0	50.0	

* Friedman´s Test (all different); Std Dev = standard deviation; Min = Minimum; Med= median; Max = maximum. ICC = Intraclass Correlation Coefficient; CI = confidence interval.

Table 3. Correlation of whole-body tumor burden metrics on semi-automatic-based software and the CNN-based software with (CNN+observer) and without (CNN) observer input in 76 patients.

Variable	Mean	Std Dev	Min	Med	Max	ICC	95% CI	p-values
MTV_{SEMI}	242.8	205.9	4.6	149.0	772.6	0.960	0.942 - 0.974	<0.0001
$MTV_{CNN+observer}$	254.8	212.8	4.1	178.3	778.3			
MTV_{CNN}	234.8	206.9	11.7	147.6	784.4			
TLG_{SEMI}	1626.4	1674.6	50.0	894.7	6963.1	0.963	0.947 - 0.975	<0.0001
$TLG_{CNN+observer}$	1647.3	1685.8	50.1	902.1	5963.4			
TLG_{CNN}	1647.7	1811.2	31.0	871.3	8218.6			

Std Dev = standard deviation; Min = Minimum; Med= median; Max = maximum. ICC = Intraclass Correlation Coefficient; CI = confidence interval.

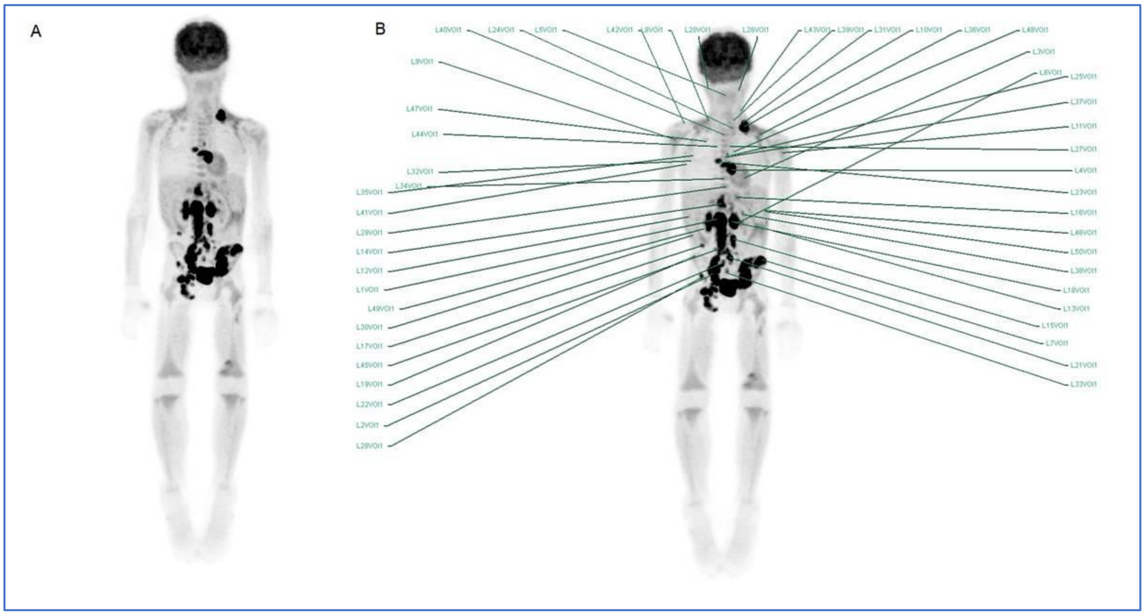


Figure 1. Whole-body tumor burden quantification of a baseline staging 18F-FDG PET/CT using the semi-automatic software of a patient with Non-Hodgkin's Lymphoma. A) MIP image shows hypermetabolic lymphoma infiltration in left supraclavicular/cervical lymph nodes, mediastinal lymph nodes and extensive lymph nodes in the abdomino-pelvic regions; lung nodules and bone infiltration. B) For calculation, the liver was set as the background reference and the VOIs automatically surrounded each lymphoma lesion with uptake higher than the mean SUV of the liver. Notice that the VOIs also include physiologic areas incorrectly selected as cancer in order to include metastatic foci with relatively low uptake, such as the right upper lobe lung nodule metastasis with mild 18F-FDG uptake.

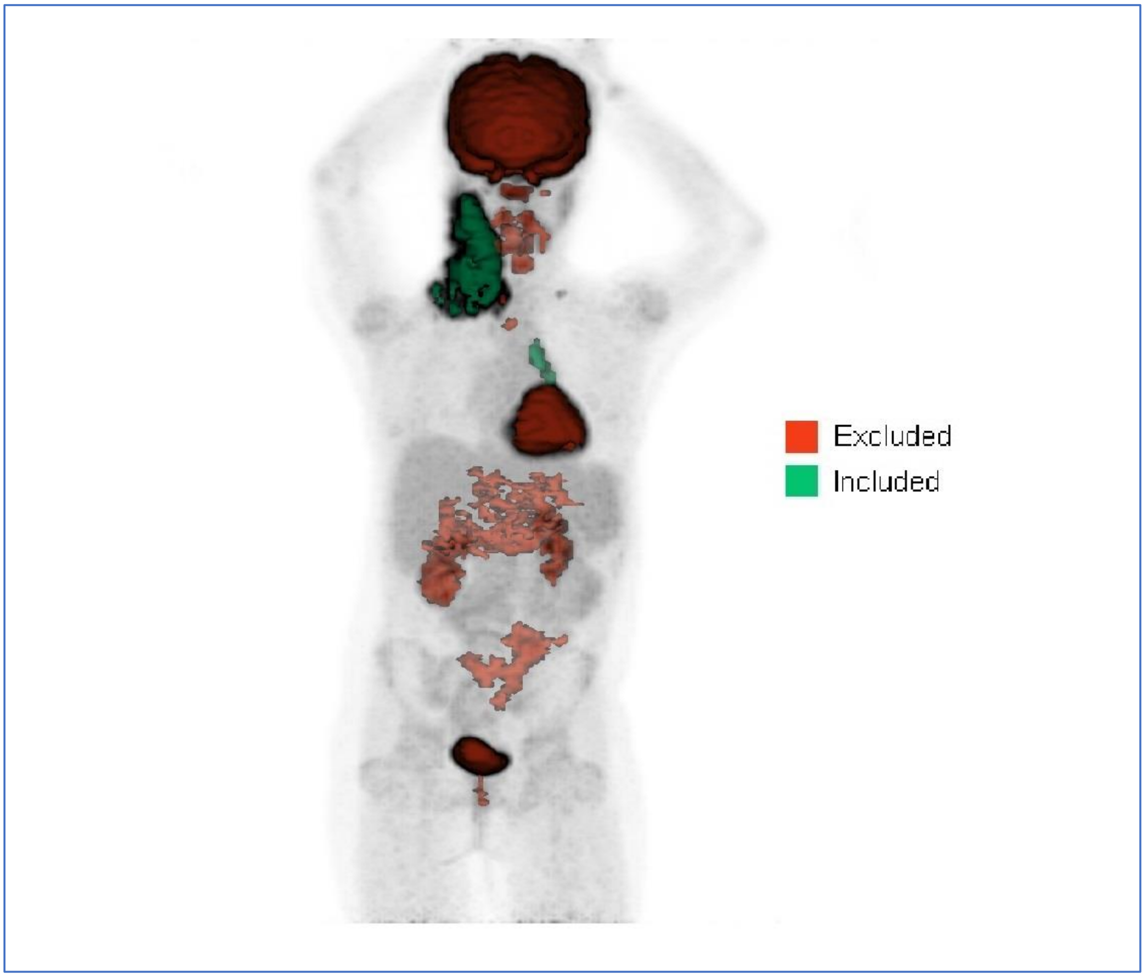


Figure 2. Staging 18F-FDG PET/CT whole-body tumor burden quantification using CNN. Displayed in RED are the regions that the software deems should be excluded from the analysis (regions related to physiological uptake: brain, head&neck, heart, intestines, kidneys and bladder) and in GREEN the regions that the software included in the calculation of whole-body tumor burden. In this patient, the extensive cervical lymph node bulky mass and mediastinal lymph nodes were included.

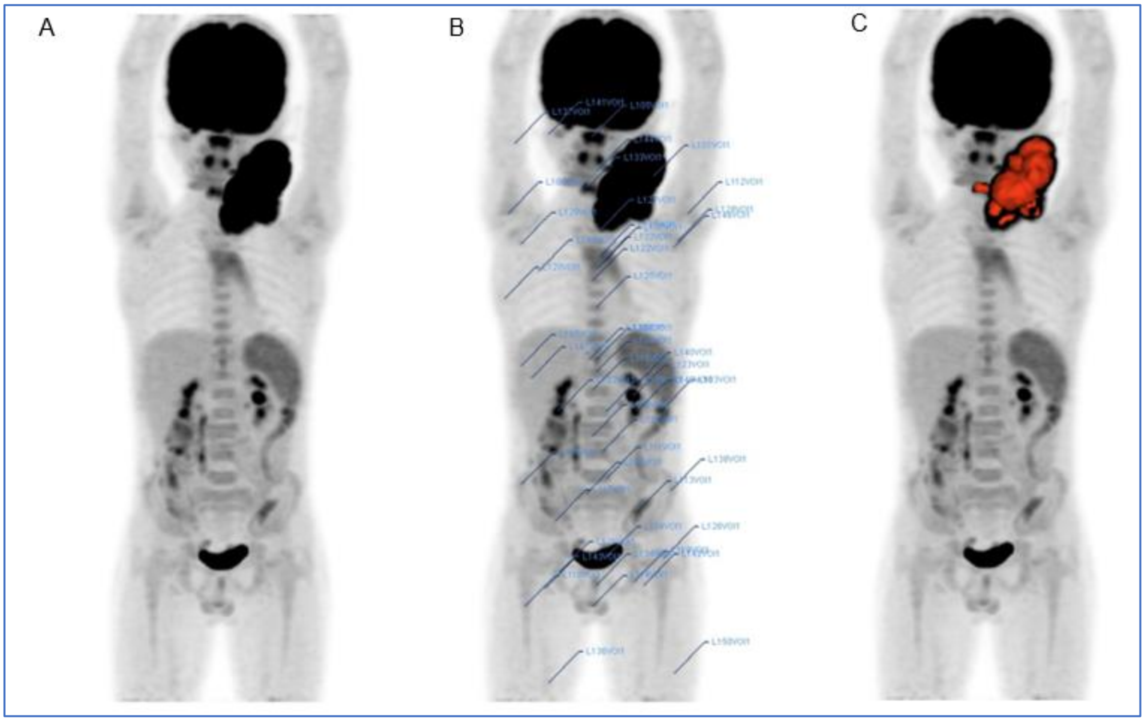


Figure 3. A baseline staging 18F-FDG PET/CT of a patient with Hodgkin's Lymphoma. (A) The MIP image reveals a cervical hypermetabolic bulky mass. The images displayed with different whole-body tumor burden quantification methods show that (B) using the semi-automatic method, VOIs are delineated in the cancer lesions and also in physiologic regions not related to cancer; these regions must be deleted prior to quantification. The whole-body tumor burden calculation showed SEMI-wbMTV = 104 and TLG $_{\mathrm{VB20}}$ =1663; the time spent calculating these metrics was 5 minutes. (C) On the other hand, the CNN whole-body tumor burden quantification did not delineate regions non-related to cancer and demonstrated similar metrics: CNN-wbMTV-$_{\mathrm{OBSERVER}}$ = 105 and CNN-wbTLG-$_{\mathrm{OBSERVER}}$ = 1671. Impressively, the time spent calculating was significantly faster (13 seconds) even though on CNN the software failed to delineate the spleen, which had to be performed manually.

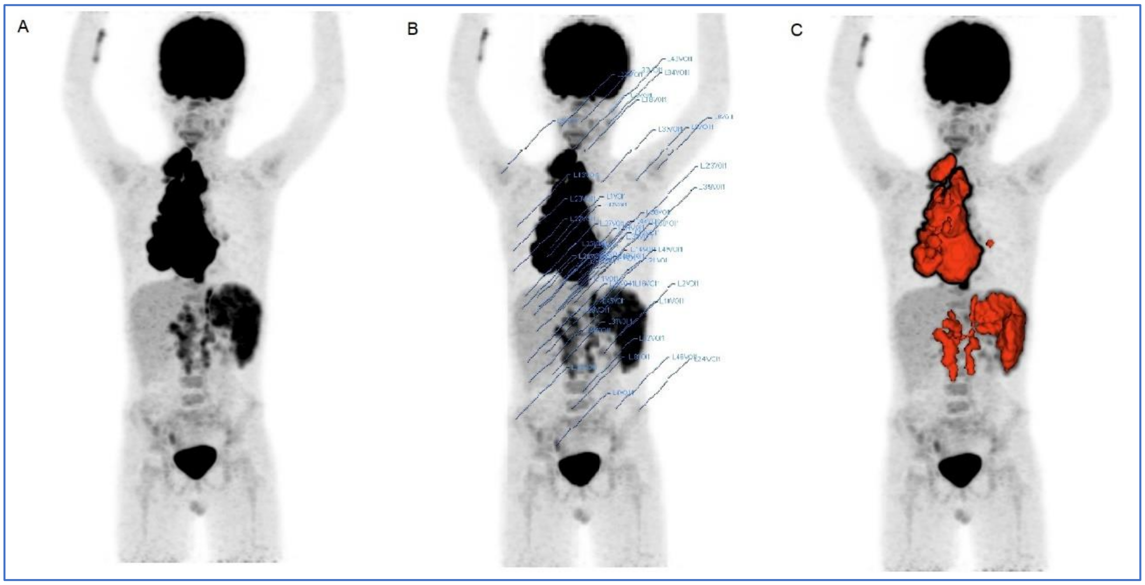


Figure 4. A baseline staging 18F-FDG PET/CT of a Hodgkin's Lymphoma. (A) MIP images reveals a mediastinal hypermetabolic bulky mass and extensive cervical, abdominal lymph nodes and spenic infiltration. (B) Semi-automatic quantification revealed SEMI-wbMTV = 548 and SEMI-wbTLG = 5238; the time spent calculating was 15 minutes. (C) On the other hand, the CNN whole-body tumor burden quantification demonstrated similar metrics: CNN-wbMTV = 570 and CNN-wbTLG = 5213 but the time spent calculating was significantly faster (14 seconds). Notice how the CNN software excludes focal areas of physiologic uptake such as the right ureter and includes areas of mild uptake such as the left hilar lymph node.

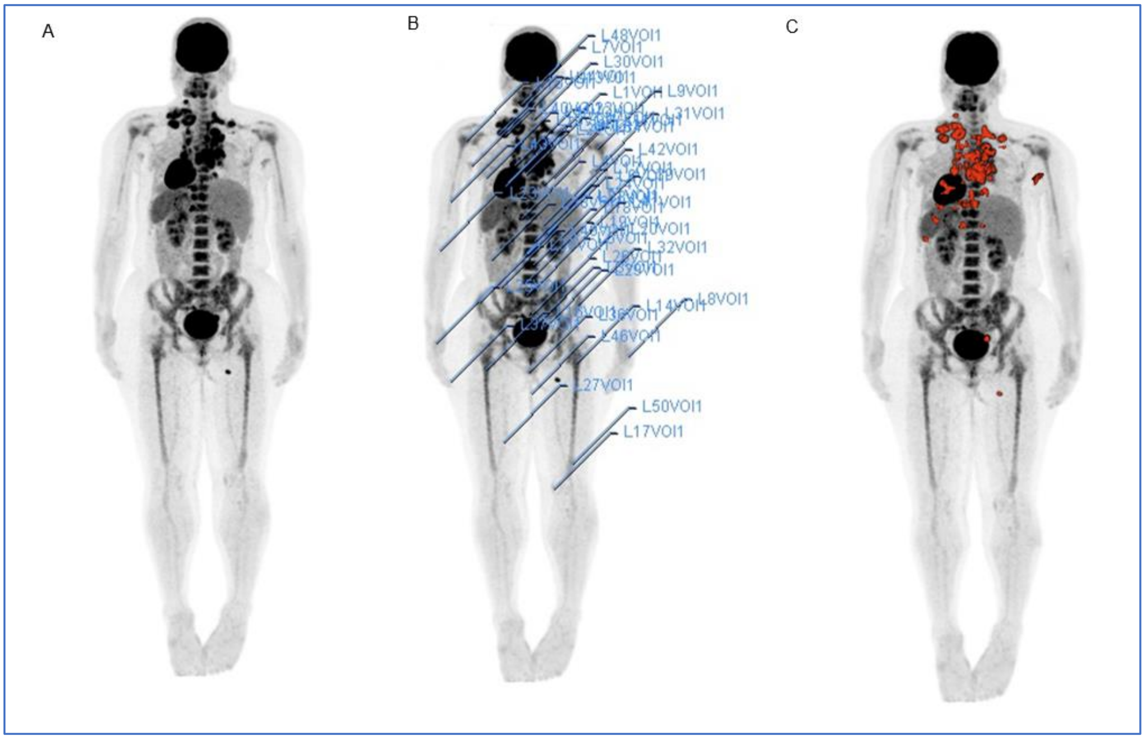
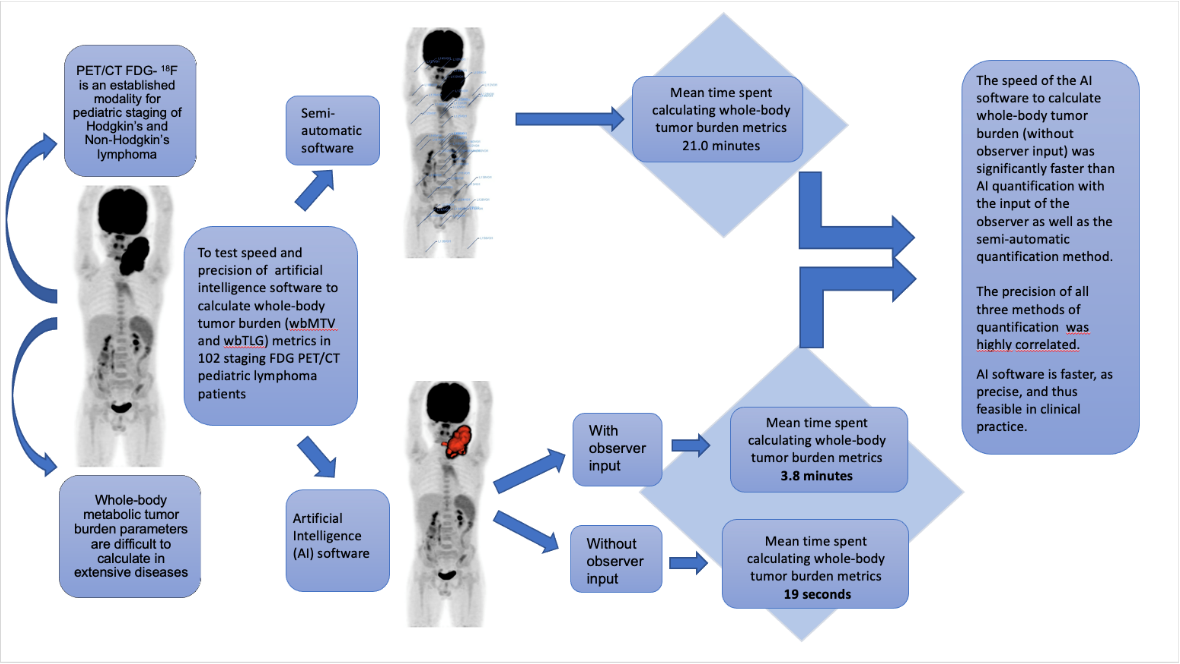


Figure 5. 18F-FDG PET/CT of a patient with Hodgkin´s Lymphoma. (A)The MIP images revealed a mediastinal hypermetabolic bulky mass, cervical, axillary and inguinal nodes. (B) The semi-automatic VB20 whole-body tumor burden quantification revealed MTV = 194 and TLG = 1007; the time spent calculating these metrics was 30 minutes because of the extent of lesions and the necessity to exclude multiple areas of physiological uptake. (C) On the other hand, the CNN whole-body tumor burden quantification demonstrated similar metrics: MTV = 200 and TLG = 968. However, the time spent calculating was significantly faster (36 seconds). Notice how the CNN software excludes physiologic areas with high uptake such as the heart and includes lymph nodes with less uptake adjacent to the heart.

Graphical Abstract



PET/CT FDG- ¹⁸F is an established modality for pediatric staging of Hodgkin's and Non-Hodgkin's lymphoma

Whole-body metabolic tumor burden parameters are difficult to calculate in extensive diseases

To test speed and precision of artificial intelligence software to calculate whole-body tumor burden (wbMTV and wbTLG) metrics in 102 staging FDG PET/CT pediatric lymphoma patients

Semi-automatic software

Mean time spent calculating whole-body tumor burden metrics 21.0 minutes

Artificial Intelligence (AI) software

With observer input

Mean time spent calculating whole-body tumor burden metrics **3.8 minutes**

Without observer input

Mean time spent calculating whole-body tumor burden metrics **19 seconds**

The speed of the AI software to calculate whole-body tumor burden (without observer input) was significantly faster than AI quantification with the input of the observer as well as the semi-automatic quantification method.

The precision of all three methods of quantification was highly correlated.

AI software is faster, as precise, and thus feasible in clinical practice.