

# Validation of Convolutional Neural Networks for Fast Determination of Whole-Body Metabolic Tumor Burden in Pediatric Lymphoma

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

<sup>1</sup>University of Campinas, Campinas, Brazil; <sup>2</sup>Medicina Nuclear de Campinas, Campinas, Brazil; <sup>3</sup>University of Cape Town, Cape Town, South Africa; <sup>4</sup>QUANTA Diagnóstico e Terapia, Curitiba, Brazil; <sup>5</sup>University of British Columbia, Vancouver, British Columbia, Canada; <sup>6</sup>All India Institute of Medical Sciences, New Delhi, India; <sup>7</sup>Tata Memorial Centre, Mumbai, India; <sup>8</sup>Ludwig-Maximilian University of Munich, Munich, Germany; <sup>9</sup>Rambam Health Care Campus, Haifa, Israel; <sup>10</sup>Centro Uruguayo de Imagenología Molecular, Montevideo, Uruguay; <sup>11</sup>National Institute of Nuclear Medicine and Allied Sciences, Dhaka, Bangladesh; <sup>12</sup>Nuclear Medicine, Oncology and Radiotherapy Institute, Islamabad, Pakistan; <sup>13</sup>University of the Philippines, Manila, Philippines; <sup>14</sup>Institute of Nuclear Medicine, London, United Kingdom; and <sup>15</sup>Nuclear Medicine and Diagnostic Imaging Section, Division of Human Health, International Atomic Energy Agency, Vienna, Austria

<sup>18</sup>F-FDG PET/CT quantification of whole-body tumor burden in lymphoma is not routinely performed because of the lack of fast methods. Although the semiautomatic method is fast, it is not fast enough to quantify tumor burden in daily clinical practice. Our purpose was to evaluate the performance of convolutional neural network (CNN) software in localizing neoplastic lesions in whole-body <sup>18</sup>F-FDG PET/CT images of pediatric lymphoma patients. **Methods:** The retrospective image dataset, derived from the data pool of the International Atomic Energy Agency (coordinated research project E12017), included 102 baseline staging <sup>18</sup>F-FDG PET/CT studies of pediatric lymphoma patients (mean age, 11 y). The images were quantified to determine the whole-body tumor burden (whole-body metabolic tumor volume [wbMTV] and whole-body total lesion glycolysis [wbTLG]) using semiautomatic software and CNN-based software. Both were displayed as semiautomatic wbMTV and wbTLG and as CNN wbMTV and wbTLG. The intra-class correlation coefficient (ICC) was applied to evaluate concordance between the CNN-based software and the semiautomatic software. **Results:** Twenty-six patients were excluded from the analysis because the software was unable to perform calculations for them. In the remaining 76 patients, CNN and semiautomatic wbMTV tumor burden metrics correlated strongly (ICC, 0.993; 95% CI, 0.989–0.996;  $P < 0.0001$ ), as did CNN and semiautomatic wbTLG (ICC, 0.999; 95% CI, 0.998–0.999;  $P < 0.0001$ ). However, the time spent calculating these metrics was significantly ( $<0.0001$ ) less by CNN (mean, 19 s; range, 11–50 s) than by the semiautomatic method (mean, 21.6 min; range, 3.2–62.1 min), 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.

**Key Words:** <sup>18</sup>F-FDG PET/CT; whole-body tumor burden; pediatric; lymphoma

J Nucl Med Technol 2022; 50:256–262

DOI: 10.2967/jnmt.121.262900

For pediatric staging and treatment response evaluation of Hodgkin and non-Hodgkin lymphoma, <sup>18</sup>F-FDG PET/CT is an invaluable tool and an established modality (1–7). Visual interpretation of <sup>18</sup>F-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). <sup>18</sup>F-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 SUVs (10–13). However, the prognostic determination, although easily measured in primary solid tumors (14–17), has not been applied in daily clinical practice to 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 radiologic 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.

Received Jul. 14, 2021; revision accepted Feb. 10, 2022.  
For correspondence or reprints, contact Elba Etchebehere (elba@hc.unicamp.br).

\*Contributed equally to this work.

Published online Apr. 19, 2022.

COPYRIGHT © 2022 by the Society of Nuclear Medicine and Molecular Imaging.

Computer deep learning and functioning as a neural network have evolved substantially, achieving remarkable success in tumor segmentation and diagnosis and ultimately transforming and optimizing clinical practice (18,20–23), providing objective and accurate diagnoses in medicine by building diagnostic models. For example, software for multimodality imaging using deep convolutional neural networks (CNNs) automatically localizes and delineates metastases in whole-body  $^{18}\text{F}$ -FDG PET/CT scans. Deep CNN seems capable of correctly localizing and classifying uptake patterns in  $^{18}\text{F}$ -FDG PET/CT images into foci suggestive and nonsuggestive of cancer. These extracted features help the semantic interpretation and may simplify the PET workflow with a 1-click calculation of whole-body tumor burden (24–26). 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.

## MATERIALS AND METHODS

This dataset, retrospectively studied, is derived from a subset of 102 baseline staging  $^{18}\text{F}$ -FDG PET/CT studies of pediatric lymphoma patient images from the data pool of the prospective multicenter research project coordinated by the International Atomic Energy Agency (coordinated research project E12017).

### Research Regulation and Data Protection

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

### Patients

The eligibility criteria consisted of pediatric patients (age < 18 y) with newly diagnosed Hodgkin lymphoma or non-Hodgkin lymphoma who underwent a staging  $^{18}\text{F}$ -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 stage were evaluated, such as the age at diagnosis, the final clinical stage, spleen disease, additional nodal sites, disease volume, B symptoms, lactate dehydrogenase level, leukocytosis, erythrocyte sedimentation rate, anemia,

albumin level, bone marrow  $^{18}\text{F}$ -FDG uptake, Deauville score, MTV, and TLG.

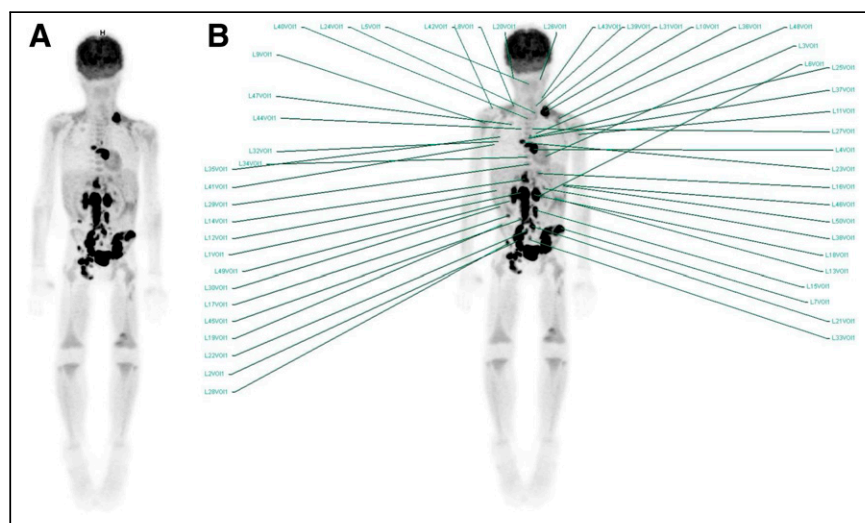
### $^{18}\text{F}$ -FDG PET/CT Imaging and Quantification

All patients underwent staging whole-body  $^{18}\text{F}$ -FDG PET/CT, from the top of the skull to the toes. All scans were obtained according to standard Society of Nuclear Medicine and Molecular Imaging or European Association of Nuclear Medicine procedure guidelines (30).

The whole-body MTV (wbMTV) and whole-body TLG (wbTLG) metrics were calculated using semiautomatic and CNN software. All images on both types of software were processed by 2 observers. Differences in the wbMTV and wbTLG metrics (if any) were recalculated to reach consensus. The semiautomatic software was used as the reference standard to evaluate the CNN software's performance.

*Semiautomatic Quantification of Whole-Body Tumor Burden.* The wbMTV and wbTLG metrics were calculated using semiautomatic multifocal segmentation software (Syngovia VB20; Siemens Medical Solutions), previously validated for clinical use (19) using a fixed threshold.

With this software, the whole-body tumor burden metrics (semiautomatic wbMTV and wbTLG) were obtained. The semiautomatic whole-body tumor burden was obtained 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 in 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  $\text{SUV}_{\text{mean}}$  of the liver. A VOI threshold of 41% of the  $\text{SUV}_{\text{max}}$  using isocontour drawings was applied for all automatically delineated lesions. The image and VOIs were then reviewed to exclude physiologic areas incorrectly



**FIGURE 1.** Whole-body tumor burden quantification on baseline staging  $^{18}\text{F}$ -FDG PET/CT using semiautomatic software on patient with non-Hodgkin lymphoma. (A) Maximum-intensity projection shows hypermetabolic lymphoma infiltration in left supraclavicular and cervical lymph nodes, mediastinal lymph nodes, and extensively in abdominopelvic lymph nodes; lung nodules; and bone infiltration. (B) For calculation, liver is set as background reference, and VOIs automatically surround each lymphoma lesion with uptake higher than  $\text{SUV}_{\text{mean}}$  of liver. VOIs also include physiologic areas incorrectly selected as cancer to include metastatic foci with relatively low uptake, such as lung nodule metastasis with mild  $^{18}\text{F}$ -FDG uptake in right upper lobe.

selected as cancer (such as brain, kidneys, bladder, and ureters) and include metastatic foci with relatively low uptake that were missed by the software (e.g., small lymph nodes). Afterward, whole-body MTV and TLG calculations were readily available and displayed as semiautomatic wbMTV and wbTLG (Fig. 1).

**CNN Quantification of Whole-Body Tumor Burden.** The wbMTV and wbTLG metrics were calculated using software based on deep CNN (Syngovia VB50; Siemens Healthineers). With this software, the whole-body tumor burden metrics (CNN wbMTV and wbTLG) were obtained.

Computation of the whole-body tumor burden on the CNN software was automatically performed by the deep CNN method as described by Sibille et al. (24). Unlike the semiautomatic software, the CNN software does not require an initial positioning of a VOI surrounding the body. The CNN automatically computes the maximum-intensity-projection  $^{18}\text{F}$ -FDG PET image and integrates the anatomic CT image using an intuitive interface. Afterward, the software automatically detects  $^{18}\text{F}$ -FDG-avid anatomic landmarks and discriminates hypermetabolic areas related to the physiologic activity that will be automatically excluded (Fig. 2) from cancer. Briefly, the PET VOIs are segmented using a fixed threshold algorithm and evaluated by the deep CNN. Whole-body CT examinations are aligned to an anatomic atlas. Finally, a maximum-intensity projection of the whole-body  $^{18}\text{F}$ -FDG PET/CT is reconstructed, and the lesions are classified. The deep CNN uses a combination of multiplanar

reconstructions of PET and CT,  $^{18}\text{F}$ -FDG PET maximum-intensity projections, and anatomic atlases to predict the anatomic localization of  $^{18}\text{F}$ -FDG foci and determine whether a focus was suggestive (or not) for malignancy. The advantage of the CNN algorithm is that it does not require the initial positioning of a VOI. This specific CNN software is not yet validated for pediatric patients.

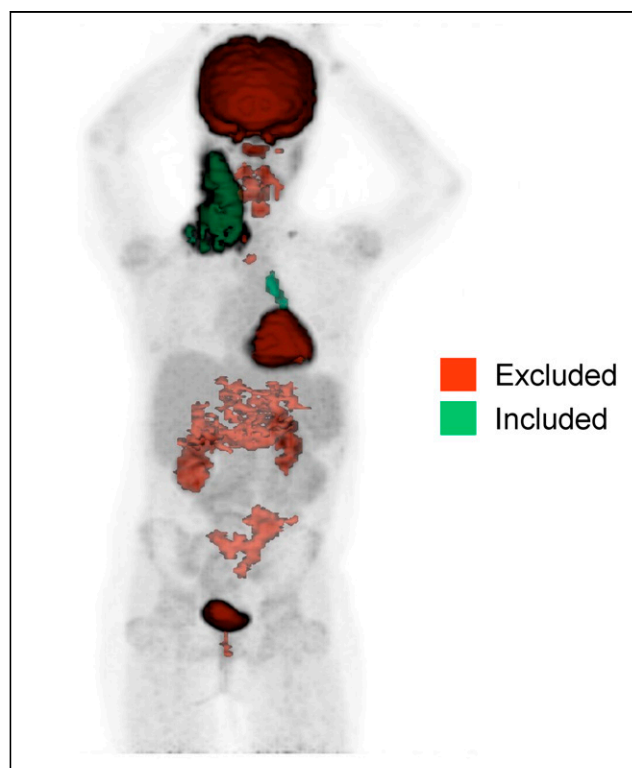
Two forms of analyses were undertaken on the CNN software: the observer method, in which all VOIs automatically generated by the multifocal segmentation tool were reviewed (in a masked manner) by both observers to determine whether the VOIs were wrongly included or excluded from the results (afterward, values were calculated and displayed as CNN + observer wbMTV and wbTLG), and the no-observer method, in which the VOIs automatically obtained were accepted and did not undergo a masked review by each of the observers. The calculations were readily available and displayed as CNN wbMTV and wbTLG.

### 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, SD, median, minimum and maximum).

**TABLE 1**  
Clinical Characteristics of Patients ( $n = 102$ )

Parameter	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	Yes	29	28.4%
Disease	No	73	71.6%
Extranodal sites	0	67	65.7%
	1	15	14.7%
	$\geq 2$	20	19.6%
Disease bulk	Bulky	63	61.8%
	Nonbulky	39	38.2%
B symptoms	Yes	43	43.0%
	No	57	57.0%
LDH	High	47	52.8%
	Normal	42	47.2%
Leukocytosis	Yes	32	31.7%
	No	69	68.3%
Erythrocyte sedimentation 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	Diffuse	12	11.9%
$^{18}\text{F}$ -FDG	Focal	16	15.8%
Uptake	Negative	73	72.3%
Event	Yes	10	9.8%
	No	92	90.2%
Status	Alive	101	99.0%
	Dead	1	1.0%



**FIGURE 2.** Whole-body tumor burden quantification on staging  $^{18}\text{F}$ -FDG PET/CT using CNN. Displayed in red are regions that software excluded from analysis (regions related to physiologic uptake: brain, head and neck, heart, intestines, kidneys, and bladder), and displayed in green are regions that software included in calculation of whole-body tumor burden. In this patient, extensive cervical lymph node bulky mass and mediastinal lymph nodes were included.

The  $\chi^2$  test or Fisher exact test was used to check associations or compare proportions, and the Mann–Whitney test was used to compare continuous or orderable measurements between the 2 groups. Risk factors associated with the event were identified with univariate and multiple Cox regression analyses. The variable selection process used was stepwise.

To verify the relationship between continuous measurements, the Spearman correlation coefficient was used ranging from  $-1$  to  $1$ .

To assess agreement between the semiautomatic and CNN software, the intraclass correlation coefficient (ICC) was used (values above  $0.7$  were considered to indicate substantial agreement). The Friedman test and the Wilcoxon test for related samples were used to compare the times. The time was defined as the moment that the physician began focusing on the task until the moment that the whole-body tumor burden calculation was completed. The level of significance was  $0.05$ .

## RESULTS

The whole-body tumor burden was quantified using both types of software in 102  $^{18}\text{F}$ -FDG PET/CT baseline scans of pediatric lymphoma patients. There were 32 (31.4%) girls and 70 (68.6%) boys. The mean age at lymphoma diagnosis was  $11.1 \pm 4.3$  y (range, 4.0–18.0 y). Among these, 80 (78.4%) patients had Hodgkin lymphoma, and 22 (21.6%) had

non-Hodgkin lymphoma. Table 1 displays the clinical characteristics.

### Semiautomatic Calculation of Whole-Body Tumor Burden

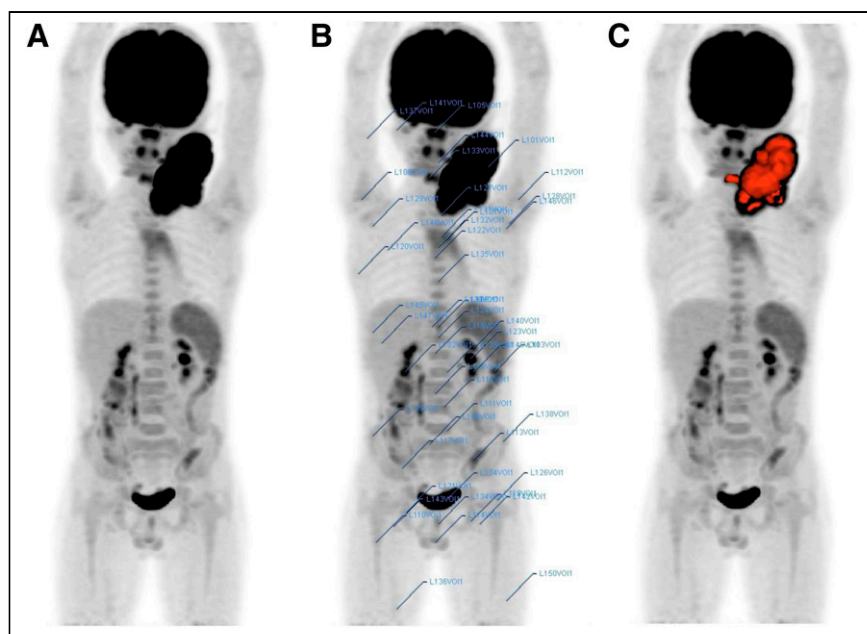
The semiautomatic wbMTV and wbTLG were calculated in all 102 patients. The average time spent on this calculation was 21.6 min, ranging from 3.2 to 62.1 min. Notably, in patients with widespread lesions in multiple organs or confluent with areas of physiologic excretion, the software took longer to identify and delineate abnormal areas.

### CNN-Based Calculation of Tumor Burden

The CNN + observer wbMTV and wbTLG were also calculated in all 102 patients. The average time spent on this calculation, with the CNN software having the observers evaluate the images before calculation, was 3.8 min, ranging from 0.5 to 19.6 min.

On the other hand, CNN wbMTV and wbTLG (i.e., without any observer evaluating the CNN software's performance before calculation) were calculated in 76 of the 102 patients. Twenty-six patients were excluded from the analyses because the software could not perform calculations because of patient movement or misregistration ( $n = 6$ ), because the software could not recognize small lymph nodes as diseased ( $n = 8$ ), or because there was widespread brown fat ( $n = 3$ ), diffuse bone infiltration ( $n = 5$ ), diffuse homogeneous mild infiltration of the spleen ( $n = 2$ ), or subcutaneous infiltration of  $^{18}\text{F}$ -FDG at the injection site ( $n = 2$ ) (Fig. 3).

Impressively, the average total time spent calculating CNN wbMTV and wbTLG was 19 s, ranging from 11 to 50 s. This total time begins when the physician begins focusing on the task and ends at completion of the whole-body tumor burden calculation. Thus, the times spent calculating CNN, CNN + observer, and semiautomatic wbMTV metrics in 76 paired patients were significantly different ( $P < 0.0001$ ). The CNN software alone was much faster and more precise than either the semiautomatic or the CNN + observer method (Table 2).



**FIGURE 3.** Baseline staging  $^{18}\text{F}$ -FDG PET/CT of patient with Hodgkin lymphoma. (A) Maximum-intensity projection reveals cervical hypermetabolic bulky mass. (B) Image displayed with different whole-body tumor burden quantification methods shows that using semiautomatic method, VOIs are delineated in cancer lesions and also in physiologic regions not related to cancer; these regions must be deleted before quantification. Whole-body tumor burden calculation showed semiautomatic wbMTV of 104 and TLG of 1,663; time spent calculating these metrics was 5 min. (C) CNN whole-body tumor burden quantification does not delineate regions nonrelated to cancer and demonstrates similar metrics: CNN + observer wbMTV of 105 and CNN + observer wbTLG of 1,671. Time spent calculating was significantly less (13 s) even though CNN software failed to delineate spleen, which had to be performed manually.

### Comparison of Semiautomatic and CNN Tumor Burden Measurements

The CNN + observer and semiautomatic wbMTV metrics calculated on the 102 patients correlated strongly (ICC, 0.993; 95% CI, 0.989–0.996;  $P < 0.0001$ ), as did the CNN + observer and semiautomatic wbTLG metrics (ICC, 0.999; 95% CI, 0.998–0.999;  $P < 0.0001$ ). Among the 76  $^{18}\text{F}$ -



**TABLE 2**  
Time Spent Quantifying Whole-Body Tumor Burden Metrics on Semiautomatic Software and CNN Software With and Without Observer Input

Variable	<i>n</i>	Time (s)					<i>P</i>
		Mean	SD	Minimum	Median	Maximum	
Semiautomatic	76	1,301.3	863.5	198.0	1,107.0	3,724.0	<0.0001
CNN + observer	76	221.1	204.4	31.0	155.0	1,176.0	
CNN	76	19.6	8.0	11.0	17.0	50.0	

FDG PET/CT studies in which the fully automatic CNN was performed, the CNN + observer, CNN, and semiautomatic wbMTV metrics also correlated strongly, as did the CNN + observer, CNN, and semiautomatic wbTLG metrics (Table 3).

Impressively, the correlation between CNN and semiautomatic wbMTV was significantly high (ICC, 0.950; 95% CI, 0.922–0.968;  $P < 0.0001$ ), as was CNN and semiautomatic wbTLG (ICC, 0.947; 95% CI, 0.917–0.966;  $P < 0.0001$ ). Therefore, the CNN software performed equally well, similar to the semiautomatic tool in which an experienced observer evaluated the images.

More impressive, however, was the fact that the correlation between CNN + observer and CNN wbMTV was significantly high (ICC, 0.946; 95% CI, 0.912–0.966;  $P < 0.0001$ ), as was CNN + observer and CNN wbTLG (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 was the first study to quantify the whole-body tumor burden of pediatric lymphoma patients using CNN and deep learning. Despite the difference in  $^{18}\text{F}$ -FDG biodistribution between children and 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 semiquantitative software in calculating whole-body tumor burden.

The CNN-based software allows a review of the VOIs provided automatically (i.e., VOIs can be added manually or incorrect ones deleted). Ultimately 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 semiautomatic software was longer, because it depends primarily on the extent of the disease. The semiautomatic quantification does not allow preselection of VOIs by the operator before creating the definitive findings and thus does not distinguish diseased areas from physiologic areas, creating many VOIs that overload the program.

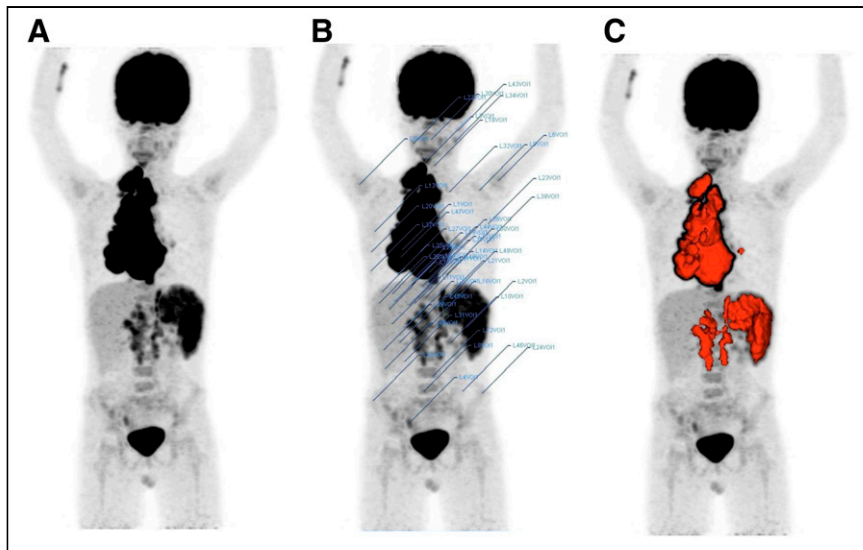
On the other hand, quantifying the whole-body tumor burden through CNN-based software was significantly faster, with and without the observer reviewing the VOIs. Impressively, when we compared quantification of the whole-body tumor burden on the CNN-based software (without observer interference) with the semiautomatic 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 s to calculate the patient's entire tumor burden, without the need to review the VOIs (Figs. 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 patient cohort to predict prognosis and response evaluation. Most (78.4%) of the patients had Hodgkin lymphoma, and there were only 2 deaths; therefore, it was not possible to determine overall survival. A

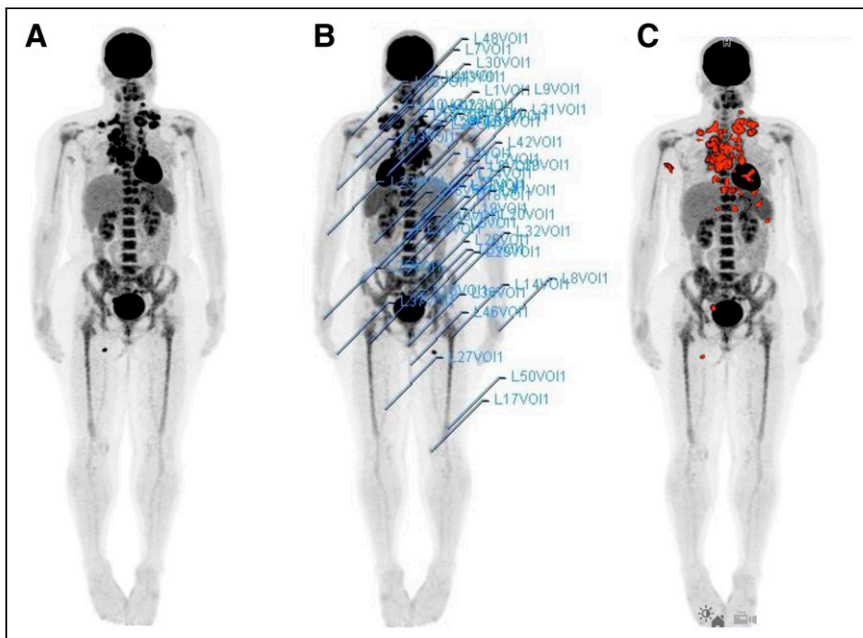
**TABLE 3**  
Correlation of Whole-Body Tumor Burden Metrics on Semiautomatic Software and CNN-Based Software With and Without Observer Input in 76 Patients

Variable	Mean	SD	Min	Median	Max	ICC	95% CI	<i>P</i>
<b>MTV</b>						0.960	0.942–0.974	<0.0001
Semiautomatic	242.8	205.9	4.6	149.0	772.6			
CNN + observer	254.8	212.8	4.1	178.3	778.3			
CNN	234.8	206.9	11.7	147.6	784.4			
<b>TLG</b>						0.963	0.947–0.975	<0.0001
Semiautomatic	1,626.4	1,674.6	50.0	894.7	6,963.1			
CNN + observer	1,647.3	1,685.8	50.1	902.1	5,963.4			
CNN	1,647.7	1,811.2	31.0	871.3	8,218.6			

Min = minimum; Max = maximum; ICC = intraclass correlation coefficient.



**FIGURE 4.** Baseline staging  $^{18}\text{F}$ -FDG PET/CT of Hodgkin lymphoma. (A) Maximum-intensity projection reveals mediastinal hypermetabolic bulky mass and extensive infiltration of cervical lymph nodes, abdominal lymph nodes, and spleen. (B) Semiautomatic quantification reveals semiautomatic wbMTV of 548 and semiautomatic wbTLG of 5,238; time spent calculating was 15 min. (C) CNN whole-body tumor burden quantification demonstrates similar metrics: CNN wbMTV of 570 and CNN wbTLG of 5,213, but time spent calculating was significantly less (14 s). CNN software excludes focal areas of physiologic uptake such as right ureter and includes areas of mild uptake such as left hilar lymph node.



**FIGURE 5.**  $^{18}\text{F}$ -FDG PET/CT of patient with Hodgkin lymphoma. (A) Maximum-intensity projection reveals mediastinal hypermetabolic bulky mass and cervical, axillary, and inguinal nodes. (B) Semiautomatic VB20 whole-body tumor burden quantification reveals MTV of 194 and TLG of 1,007; time spent calculating these metrics was 30 min because of extent of lesions and need to exclude multiple areas of physiologic uptake. (C) CNN whole-body tumor burden quantification demonstrates similar metrics: MTV of 200 and TLG of 968. However, time spent calculating was significantly less (36 s). CNN software excludes physiologic areas with high uptake such as heart and includes lymph nodes with less uptake adjacent to heart.

larger number of patients with events are required to determine whether the measurements predicted by CNN-based software can predict prognosis. Another limitation is 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 semiautomatic quantification. The CNN software we tested was not initially designed or 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. For example, small lymph nodes with mild  $^{18}\text{F}$ -FDG uptake were excluded; extensive brown fat was erroneously included as lymphomatous infiltration; extensive diffuse bone marrow infiltration (5/12 patients) was missed; and radiopharmaceutical extravasation sites and bladder catheter were erroneously included. Most likely, with further CNN and deep-learning development and specific training in pediatric patients regarding differentiation of normal biodistribution from cancer tissue, failure rates will decrease.

CNN-based software with CNN and deep learning still requires the input of the observer (26–28). In 25% of the patients, CNN could not depict the correct neoplastic tissue or added non-neoplastic tissue; thus, quantification had to be excluded because the software was not performing the calculations. Therefore, errors and failure to detect proper tissue will occur even in CNN and DL software, arguing in favor of the observer input. Most likely, the largest errors may be associated with unsupervised quantification.

## CONCLUSION

CNN-based quantification of whole-body tumor burden in pediatric lymphoma patients is an emerging field. Determination of whole-body tumor

burden using CNN-based software is extremely fast and feasible in clinical practice in pediatric lymphoma patients. CNN-based software requires CNN and deep-learning development and specific training in pediatric patients, as well as the input of the observer to minimize failure rates. Tumor burden should be evaluated in most if not all tumors and age groups for therapy purposes.

## DISCLOSURE

The whole-body metrics were calculated using a loaned Siemens device equipped with a software based on deep CNN (Syngovia VB50).

## ACKNOWLEDGMENT

We thank 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 reliable quantification data regarding whole-body metabolic tumor burden in  $^{18}\text{F}$ -FDG PET/CT pediatric lymphoma patients?

**PERTINENT FINDINGS:** Quantification of whole-body metabolic tumor burden using CNN correlates strongly with semiautomatic quantification (ICC, 0.993; 95% CI, 0.989 – 0.996;  $P < 0.0001$ ).

**IMPLICATIONS FOR PATIENT CARE:** In addition to reliable data, implementation of CNN quantification tools in 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 Clin Proc.* 2012;87:571–580.
2. Cheson BD. PET/CT in lymphoma: current overview and future directions. *Semin Nucl Med.* 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. *Semin Nucl Med.* 2018;48:28–36.
4. Gillman J, States LJ, Servaes S. PET in pediatric lymphoma. *PET Clin.* 2020;15:299–307.
5. D'souza MM, Jaimini A, Bansal A, et al. FDG-PET/CT in lymphoma. *Indian J Radiol Imaging.* 2013;23:354–365.
6. Cronin CG, Swords R, Truong MT, et al. Clinical utility of PET/CT in lymphoma. *AJR.* 2010;194:W91–W103.
7. London K, Cross S, Onikul E, Dalla-Pozza L, Howman-Giles R.  $^{18}\text{F}$ -FDG PET/CT in paediatric lymphoma: comparison with conventional imaging. *Eur J Nucl Med Mol Imaging.* 2011;38:274–284.
8. Lin C, Itti E, Haioun C, et al. Early  $^{18}\text{F}$ -FDG PET for prediction of prognosis in patients with diffuse large b-cell lymphoma: SUV-based assessment versus visual analysis. *J Nucl Med.* 2007;48:1626–1632.
9. Jhanwar YS, Straus DJ. The role of PET in lymphoma. *J Nucl Med.* 2006;47:1326–1334.
10. Yang J, Zhu S, Pang F, et al. Functional parameters of  $^{18}\text{F}$ -FDG PET/CT in patients with primary testicular diffuse large B-cell lymphoma. *Contrast Media Mol 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. *BMC Cancer.* 2015;15:198.
12. Albano D, Camoni L, Giubbini R, Bertagna F. Prognostic value of  $^{18}\text{F}$ -FDG PET/CT metabolic parameters in splenic marginal zone lymphoma. *Clin Lymphoma Myeloma Leuk.* 2020;20:e897–e904.
13. Wang X-Y, Zhao Y-F, Liu Y, Yang Y-K, Wu N. Prognostic value of metabolic variables of  $^{18}\text{F}$ -FDG PET/CT in surgically resected stage I lung adenocarcinoma. *Medicine (Baltimore).* 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. *Eur J Nucl Med Mol Imaging.* 2017;44:1969–1983.
15. Wang L, Bai J, Duan P. Prognostic value of  $^{18}\text{F}$ -FDG PET/CT functional parameters in patients with head and neck cancer: a meta-analysis. *Nucl Med Commun.* 2019;40:361–369.
16. Brito AE, Mourato F, Santos A, Mosci C, Ramos C, Etchebehere E. Validation of the semiautomatic quantification of  $^{18}\text{F}$ -fluoride PET/CT whole-body skeletal tumor burden. *J Nucl Med Technol.* 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? *Nucl Med Mol 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. *Inf Process Med Imaging.* 2011;22:245–256.
19. Camacho MR, Etchebehere E, Tardelli N, et al. Validation of a multifocal segmentation method for measuring metabolic tumor volume in Hodgkin lymphoma. *J Nucl Med Technol.* 2020;48:30–35.
20. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J.* 2019;6:94–98.
21. Deo RC. Machine learning in medicine. *Circulation.* 2015;132:1920–1930.
22. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. *J Med Internet Res.* 2019;21:e13659.
23. Recht MP, Dewey M, Dreyer K, et al. Integrating artificial intelligence into the clinical practice of radiology: challenges and recommendations. *Eur Radiol.* 2020;30:3576–3584.
24. Sibille L, Seifert R, Avramovic N, et al.  $^{18}\text{F}$ -FDG PET/CT uptake classification in lymphoma and lung cancer by using deep convolutional neural networks. *Radiology.* 2020;294:445–452.
25. Froelich JW, Salavati A. Artificial intelligence in PET/CT is about to make whole-body tumor burden measurements a clinical reality. *Radiology.* 2020;294:453–454.
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. *J Nucl Med Technol.* 2019;47:273–281.
28. Currie G, Rohren E. Intelligent imaging in nuclear medicine: the principles of artificial intelligence, machine learning and deep learning. *Semin Nucl Med.* 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. *Am Soc Clin Oncol Educ Book.* 2017;37:535–546.
30. Boellaard R, Delgado-Bolton R, Oyen WJG, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging—version 2.0. *Eur J Nucl Med Mol Imaging.* 2015;42:328–354.