



**HAL**  
open science

## Deep-learning-based 3D lesion segmentation on whole-body [18F]-FDG PET images including automated quality control: method and external validation

Kibrom Girum, Louis Rebaud, Anne-Ségolène Cottereau, Thibault Escobar, Jérôme Clerc, Laetitia Vercellino, Olivier Casasnovas, Franck Morschhauser, Catherine Thieblemont, Irene Buvat

### ► To cite this version:

Kibrom Girum, Louis Rebaud, Anne-Ségolène Cottereau, Thibault Escobar, Jérôme Clerc, et al.. Deep-learning-based 3D lesion segmentation on whole-body [18F]-FDG PET images including automated quality control: method and external validation. Society of Nuclear Medicine and Molecular Imaging meeting, Jun 2023, Chicago, United States. hal-04248828

**HAL Id: hal-04248828**

**<https://hal.science/hal-04248828>**

Submitted on 18 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Deep-learning-based 3D lesion segmentation on whole-body [18F]-FDG PET images including automated quality control: method and external validation

Kibrom Girum, Louis Rebaud, Anne-Ségolène Cottreau, Thibault Escobar, Jérôme Clerc, Laetitia Vercellino, Olivier Casasnovas, Franck Morschhauser, Catherine Thieblemont and Irene Buvat

**Introduction:** Reliable and automatic lesion segmentation might facilitate the investigation of prognostic image-based biomarkers by reducing the delineation time and inter/intra-observer variability. This work proposes an artificial intelligence (AI) pipeline to automatically detect and segment metabolically active lesions on [18F]-FDG PET images in 3D. Importantly, an associated quality control (QC) method is developed to detect possibly unreliable segmentations. Image-based biomarkers extracted by nuclear medicine experts and the proposed AI are compared.

**Methods:** This was a retrospective analysis of multi-center multi-trial Diffuse Large B-Cell Lymphoma patients. We used the [18F]-FDG PET/CT of 287 patients from the REMARC ([NCT01122472](#)) cohort to train and cross-validate the AI method and evaluated it on the PET/CT of 174 LNH073B ([NCT00498043](#)) patients. All lesions on the PET/CT scans were delineated by experts using a semi-automated procedure, providing the lesions' &ldquo;mask&rdquo;. The 3D PET image was projected into coronal and sagittal maximum intensity projection (MIP) images. The coronal and sagittal MIP of the lesion mask obtained from the experts were also calculated. The AI method was trained to segment the lesion areas on the sagittal and coronal PET images (Girum et al., J Nucl Med, 2022). A robust post-processing algorithm was developed to reconstruct the 3D segmentations by exploiting the lesion areas identified by the AI on the sagittal and coronal MIP and the SUV values from the 3D PET images. The segmented masks on the coronal and sagittal MIP images were inserted at the location of the coronal and sagittal planes corresponding to the MIP values. The regions where the coronal and sagittal insertions coincided were considered seed voxels for the 3D segmentation. The minimum SUV values of each cluster of seed voxels were obtained and used as a local SUV threshold to refine the 3D segmentation.

For QC, we retrieved a sagittal MIP segmentation from the coronal MIP segmentation by projecting the segmented SUV values in the coronal PET MIP into sagittal PET MIP. The ratio between the number of pixels in the retrieved sagittal MIP segmentation and the number of pixels in the AI sagittal MIP segmentation was calculated and called the confidence score. Total metabolic tumor volume (TMTV), tumor dissemination (Dmax), and total lesion glycolysis (TLG) extracted by the AI method in 3D were measured and compared with the corresponding values obtained by nuclear medicine experts.

**Results:** The concordance index (C-index) between the AI and expert-based TLG, Dmax, and TMTV were 0.905, 0.800, and 0.865 respectively on the REMARC cohort. The time-dependent areas under the ROC curves for progression-free survival (PFS) from the expert segmentation were TMTV: 0.67 (Confidence interval (CI): 0.60-0.73), Dmax: 0.65 (CI: 0.58-0.72), TLG: 0.65 (CI: 0.56-0.73), and from the AI segmentation were rTMTV: 0.65 (CI: 0.58-0.72), rDmax: 0.65 (CI: 0.58-0.72), rTLG 0.63 (CI: 0.56-0.70). The hazard ratio between the expert and AI-driven features were similar. The C-index on the LNH073B between the AI and expert-based TLG, Dmax, and TMTV were respectively 0.915, 0.814, and 0.874. Patients' classifications into three-risk groups using the tumor burden (TMTV and TLG) and dissemination (Dmax) biomarkers were predictive, and the expert and AI-driven features agreed. Interestingly, the cut-off values obtained on the expert-based biomarkers can be directly applied on the AI-based biomarkers (Figure 1). The mean QC of the detected low-quality segmentations (QC < 0.6) was 0.35 ( $\pm 0.196$  standard deviation), against a mean QC of 0.933 ( $\pm 0.072$ ) for all other cases.

**Conclusions:** The proposed fully automatic [18F]-FDG lesion detection pipeline enabled calculating prognostic image-based biomarkers (TMTV, Dmax, and TLG) that had similar prognostic values compared to when those biomarkers were calculated by the experts. The AI results were associated with a confidence score that could make them reliable and suitable for clinical practice.

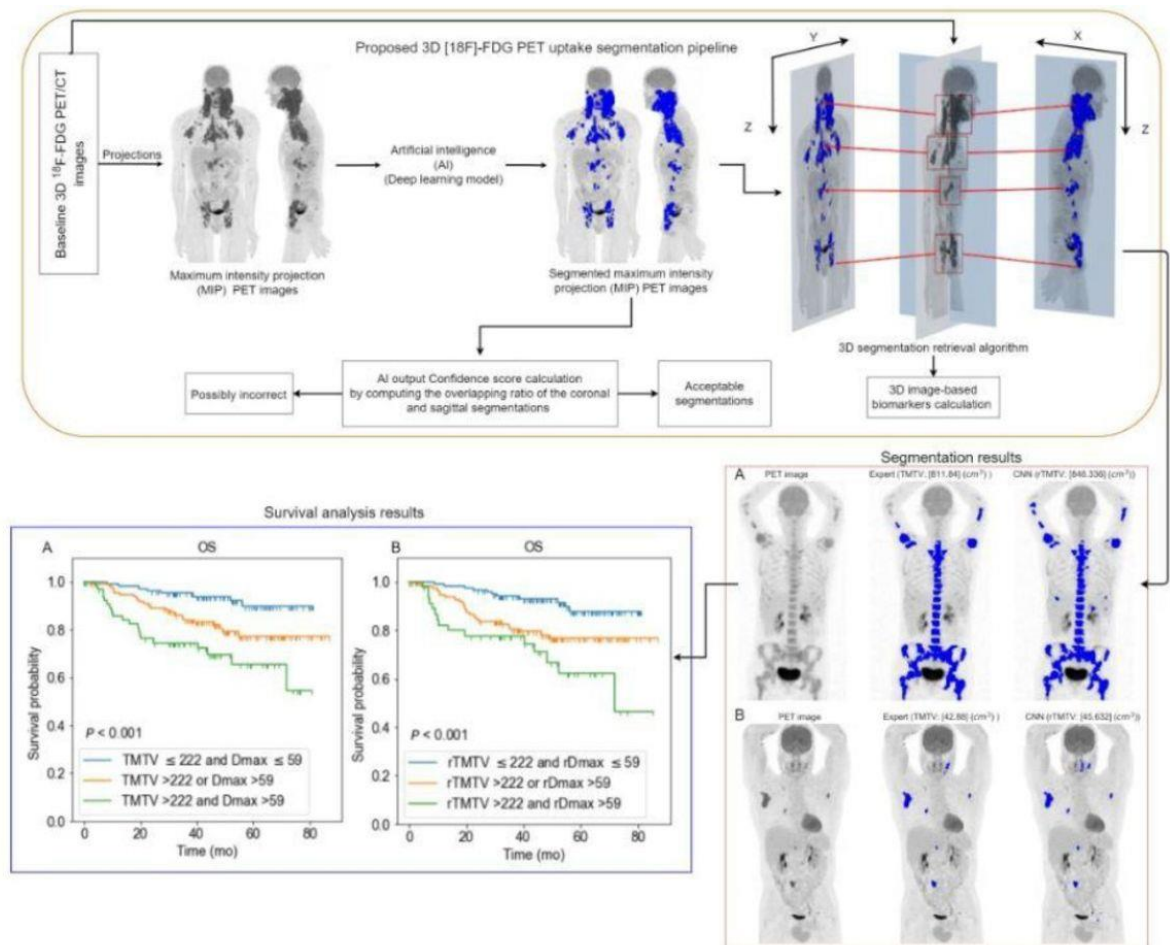


Figure 1: The proposed 3D [18F]-FDG PET uptake segmentation pipeline. The quality control (QC) values for the two segmentation examples shown in coronal views were  $> 0.85$ . The C-index of the QC and the Dice score on the REMARC data was 0.741 and 0.691 on the LNH073B data. Three-risk group stratification of patients using the expert-based biomarkers (TMTV, and Dmax) agrees with when using the A-based biomarkers (rTMTV, and rDmax). AI-driven biomarkers (*survival analysis results*) using the expert-based cut-off values were predictive, and when the cut-off values were obtained from the AI.

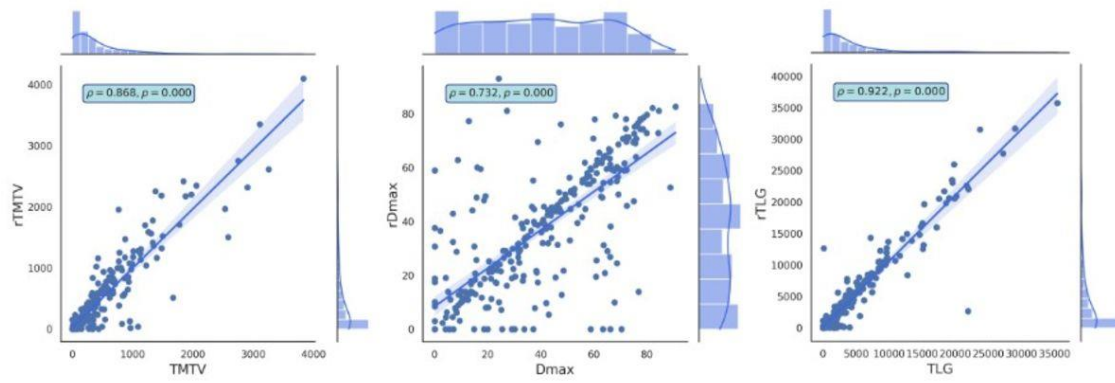


Figure 3: Spearman correlation between the 3D image segmentation-based biomarkers extracted from the expert including total metabolic tumor volume (TMTV), tumor dissemination (Dmax), and total lesion glycolysis (TLG) and the AI model (rTMTV, rDmax, and rTLG).