**EANM Abstract**

Authors

Daniel Abler*, Mario Jreige*, , Olivier Michielin, Niklaus Schaefer, John O. Prior, Adrien Depeursinge, Michel A. Cuendet, Paul Blanc-Durand

(*) Authors contributed equally

Topic:

https://eanm23.eanm.org/abstract-submission-information/

- D Technical Studies - D2 Data Analysis - D22 Other Data Analysis
- D Technical Studies - D2 Data Analysis - D25 Artificial Intelligence

Title:

Evaluation of PAIRE PET/CT segmentation software as fully-automated cancerous lesion contouring tool in image annotation workflows for research studies.

Aim/Introduction

Assistive automatic segmentation tools for clinical applications frequently prioritize detection of 'suspicious' regions of interest (ROIs), thereby accepting many false positives. While these can be discarded by expert review, fully-automated workflows are important for large-scale retrospective research [1]. This study assesses the PAIRE PET/CT segmentation software [2] and its potential role in such settings.

Materials and Methods

Tumour ROIs in 20 whole-body PET/CT images from 10 metastatic melanoma patients from the Lausanne University Hospital (CHUV) were contoured by an expert aware of the full clinical information and follow-up information to obtain a ground truth (GT). For these images, PAIRE (version 0.2.1) yielded: (a) contours of suspicious ROIs, and (b) maps indicating each voxel's likelihood of being cancerous.

Agreement between GT and PAIRE contours was investigated by establishing for each ROI the number of True Positive (TP), False Positive (FP) and False Negative (FN) voxels. PAIRE ROIs sharing at least one voxel with GT were considered TP lesions. To investigate alternative operating points, PAIRE's sensitivity-recall curve was computed by repeated assessment of contours resulting from varying decision thresholds.

Results

The GT dataset contained 98 malignant lesions across 19 PET/CT scans with an average metabolic tumour volume (MTV) of 33.9cm$^3$ per patient (1 scan not evaluable). PAIRE, using default settings, identified an average MTV of 125.1cm$^3$ and 399 ROIs. Of these, 92 TP and 307 FP (6 FN), corresponding to a (precision, recall) by lesion (PRL) of (0.23, 0.94) and by voxel (PRV) of (0.23, 0.84).. Decision thresholds between 0.01 and 0.99 yielded PRLs between (0.06, 1.00) and (0.69, 0.70), and PRVs between (0.08, 0.98) and (0.88, 0.22),

respectively. The highest average Dice index (0.75) for corresponding ROIs was achieved at a decision threshold of 0.6 with PRL of (0.53, 0.85) and PRV of (0.70, 0.67).

Conclusion

While PAIRE achieves almost perfect recall for cancerous lesion detection, any alternative operating point with acceptable recall (e.g., >0.8) still results in a large FP/TP ratio, thus requiring further FP removal steps for reliable fully automated workflows. Overall, PAIRE's voxel precision exceeds lesion detection precision, while the opposite applies to recall. Consequently, the optimal choice would depend on the specific research use case.

References

- [1] D. Abler, P. Courlet, M. Dietz, R. Gatta, P. Girard, A. Munafo, A. Wicky, M. Jreige, M. Guidi, S. Latifyan, R. De Micheli, C. Csajka, J. O. Prior, O. Michielin, N. Terranova, M. A. Cuendet. "Semiautomated Pipeline to Quantify Tumor Evolution From Real-World Positron Emission Tomography / Computed Tomography Imaging". In: JCO Clinical Cancer Informatics (2023). In press.
- [2] A. Van Der Gucht, M. Mantzarides, L. Lerman, E. Itti, P. Blanc Durand. Évaluation en routine clinique de PaIRe, outil d'intelligence artificielle pour la détection et la quantification automatique des lésions en TEP/TDM au 18F-FDG, Médecine Nucléaire, (2021) Volume 45, Issue 4, Page 192.