Towards more efficient tumor follow-up assessment using AI assistance

Alessa Hering[1,2] ALESSA.HERING@MEVIS.FRAUNHOFER.DE
[1] *Fraunhofer Institute for Digital Medicine MEVIS, Bremen, Germany*
[2] *Diagnostic Image Analysis Group, Radboud University Medical Center, Nijmegen, Netherlands*
Felix Peisen[3] FELIX.PEISEN@MED.UNI-TUEBINGEN.DE
[3] *Department of Diagnostic and Interventional Radiology, University Hospital Tübingen, Germany*
Jan Moltz[1] JAN.MOLTZ@MEVIS.FRAUNHOFER.DE

Editors: Under Review for MIDL 2022

Abstract

Measurement of metastatic tumors on longitudinal computer tomography (CT) scans is essential to evaluate the efficacy of cancer treatment. Manual measurements for the diameter-based RECIST (Response Evaluation Criteria In Solid Tumors) criteria are often time-consuming and error-prone. However, those criteria and the execution of the measurements undergo continuous changes. Lesion segmentation assistance based on artificial intelligence (AI) might significantly speed up response evaluation and help to handle the ever-growing mass of image-based staging and follow-up evaluations. Various technical papers investigate the segmentation accuracy of AI algorithms. While these technical measures give a first impression of the performance, they do not yet tell us whether we can add value to the assessment of cancer patients. As a first step to quantify this, the goal of the presented reader study was to compare the workflow of reading follow-up examinations with and without AI assistance to evaluate the impact of the proposed AI-assisted workflow. Our findings support our research hypothesis of an assisted workflow which is superior with respect to processing time and non-inferior with respect to accuracy compared to the manual workflow.

1. Introduction

In our last year's MIDL paper (Hering et al., 2021), we presented a pipeline that automates the segmentation and measurement of matching lesions, given a point annotation in the baseline lesion. We have validated our pipeline on the challenging task of whole-body soft- tissue lesion tracking and segmentation. We showed that the registration accurately propagated the center of gravity of the lesions from the baseline to the follow-up scan with a mean Euclidean distance of 7.66 mm and, more importantly, it was in almost all cases close enough to find the corresponding lesion in the follow-up scan. In the baseline and follow-up scans our segmentation approach achieved an average Dice Score of approximately 0.80.

While these technical measures give a first impression of the performance, they do not indicate whether the presented method will also add value in the clinical routine by speeding up the reading or improving patient outcomes. Therefore, we conducted a reader study to investigate the following three hypotheses: 1.) Assessment time for follow-up lesion segmentation is reduced using an AI-assisted workflow 2.) The inter-reader variability of the resulting segmentation is reduced with AI assistance. 3.) The quality of the AI-assisted segmentation is non-inferior to a fully manual segmentation. The complete reader study has been submitted to a clinical journal.

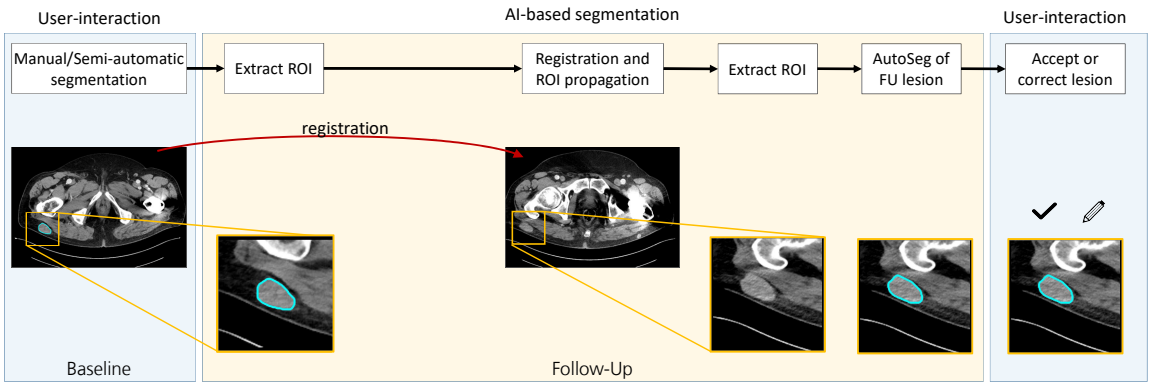

Figure 1: Schema of the proposed pipeline for AI-assisted segmentation of metastases in follow-up CT scans.

2. Experiments and results

Dataset: The testing dataset consists of 262 lesions (126 soft tissue metastases and 137 lymph node metastases) of 55 patients with metastatic melanoma (Stage IV, AJCC) treated at the Center for Dermato-Oncology at the University Hospital Tuebingen, Germany.

Experiments For each examination, baseline and follow-up CTs were imported into SATORI – a custom-made image post-processing software system. Two experienced radiologists performed a manual and AI-assisted segmentation for all lesions. The radiologists were able to see the baseline CT with the already segmented metastases and the follow-up CT in one shared window. For the manual workflow, masks were created by manually segmenting the lesions on the follow-up CT images using a cursor to contour the lesions with optional interpolation. For AI-assisted segmentations, follow-up CT examinations with lesion masks created by the proposed pipeline were imported into SATORI, and each mask was manually edited by radiologists. They had the choice to (a) accept the automated segmentation as perfect and move on to the next metastasis, (b) accept the automated segmentation as passable and make manual corrections on various slides using a cursor or (c) dismiss the automated segmentation and perform a manual segmentation instead. User interaction time was recorded for manual segmentations and for manual corrections of automated segmentations per patient.

Results: AI-assisted segmentation achieves a significant reduction of user interaction time compared to manual segmentation (1.6 min vs 3.6 min) and similar Dice scores (0.84 vs 0.82), sensitivity (0.94 vs 0.92) and PPV (0.95 vs 0.96) compared to a manual segmentation of the reference reader. Results show a higher AI-assisted inter-reader agreement (median Dice 1.0) compared to the manual inter-reader agreement (median Dice 0.82). The inter-method agreement of the automatic segmentation to the respective AI-assisted segmentations (automatic to A1 and automatic to A2) also achieved a median Dice score of 1.0 This means that in more than 50% of the lesions, the reader accepted the segmentation without any further corrections.

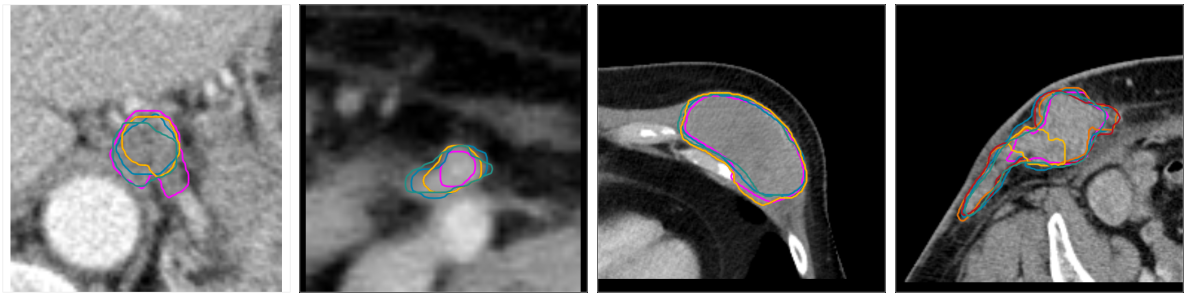

Figure 2: Exemplary segmentation results for the reference reader (pink), the manual segmentation M1 (green) and M2 (blue), the assisted segmentation A1 (orange) and A2 (red) as well as the automatic segmentation (yellow). If no assisted segmentation is shown, it is the same as the automatic segmentation.

3. Limitations and future work

The study focused on lymph nodes and soft tissue lesions in CT only. However, other lesion types and imaging modalities could likely be added by providing a sufficient number of reference segmentations. Furthermore, we have only performed an evaluation on the segmentation masks. However, the current guideline of metastatic tumor evaluation on CT scans RECIST is based on the diameter of lesions. Therefore, the next logical step is to further evaluate the accuracy of the calculated diameters for several lesion types.

An independent evaluation with additional readers is required to support the generalizability of our results to other readers. The analysis is conservative in the sense that further training with the assisted workflow might lead to an additional improvement in one or both outcomes over time.

4. Conclusion

Our findings support our research hypothesis of an assisted workflow for tumor follow-up assessment which is superior with respect to processing time and non-inferior with respect to accuracy compared to the manual workflow.

Acknowledgments

The research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 428216905 / SPP 2177.

References

Alessa Hering, Felix Peisen, Teresa Amaral, Sergios Gatidis, Thomas Eigentler, Ahmed Othman, and Jan Hendrik Moltz. Whole-body soft-tissue lesion tracking and segmentation in longitudinal CT imaging studies. In *Medical Imaging with Deep Learning*, 2021.