Implementation of Delineation Error Detection Systems in Clinical Practice: Do AI-Supported Optimization and Human Preferences Meet?

Chaves-de-Plaza, Nicolas F.

Publication date
2023

Document Version
Final published version

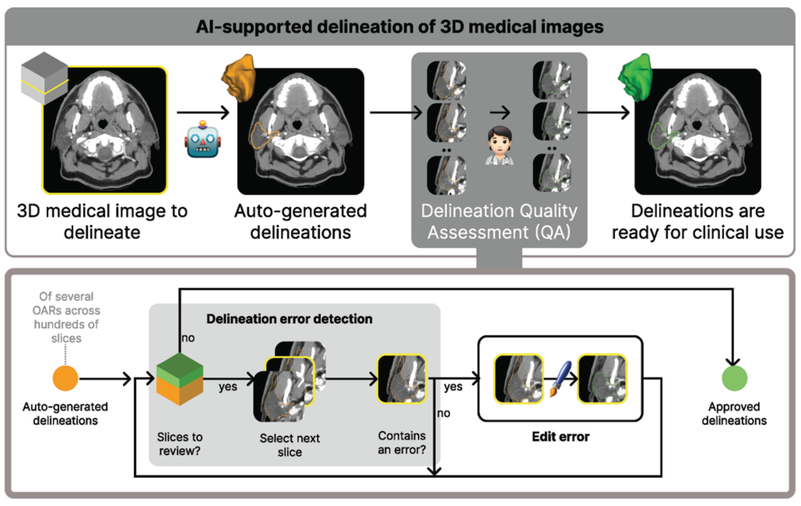
Citation (APA)
Chaves-de-Plaza, N. F. (2023). *Implementation of Delineation Error Detection Systems in Clinical Practice: Do AI-Supported Optimization and Human Preferences Meet?*. Poster session presented at International Scientific Symposium of the HollandPTC R&D Consortium.

Implementation of Delineation Error Detection Systems in Clinical Practice: Do AI-Supported Optimization and Human Preferences Meet?

Nicolas F. Chaves-de-Plaza[1,2], Prerak Mody[2,3], Klaus Hildebrandt[1], Marius Staring[3], Eleftheria Astreinidou[3], Mischa de Ridder[5], Huib de Ridder[1], Anna Vilanova[4] and René van Egmond[1]

Introduction and Motivation

- Delineation is a bottleneck in the planning process [1]
- AI advances have dramatically accelerated delineation [2]
- In the AI-supported delineation workflow, the clinicians perform quality assessment (QA)



The QA process is fatiguing and time-consuming due to the **number** of errors and their **extent** [3]

Delineation Error Detection Systems (**DEDS**) aim at reducing time but lack validation with clinicians:

Would clinicians adopt DEDS in practice?
Can DEDS effectively speed up the QA process?

We investigate these questions in user and simulation studies

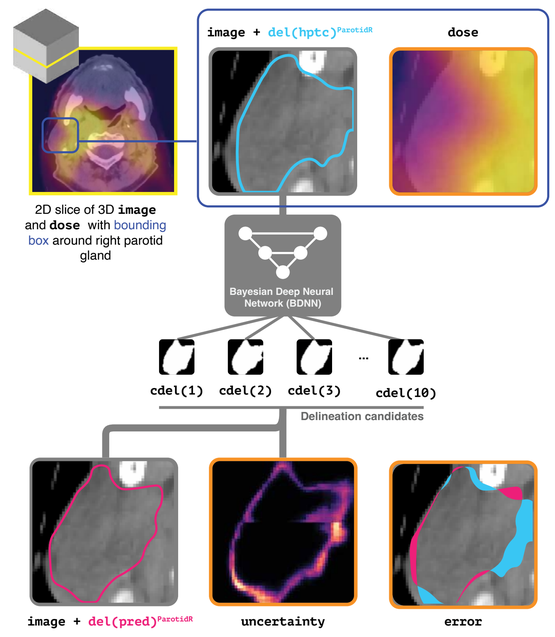
Anatomy of Delineation Error Detection Systems (DEDS)

We designed a DEDS which permitted clinicians address most relevant slices first by sorting them by their priority score:

Clinicians define slices's priority as a weighted combination of AI uncertainty [4], error and dose:

$$priority = w1.agg(unc)+w2.agg(error)+w3.agg(dose)$$

Top row presents available information sources and bottom row the once we derived



We derive the scores by aggregating volumes per slice using functions aggregation functions (agg) like sum, max and mean

User Study

Would clinicians adopt DEDS in practice? Which information sources and workflows do users prefer?

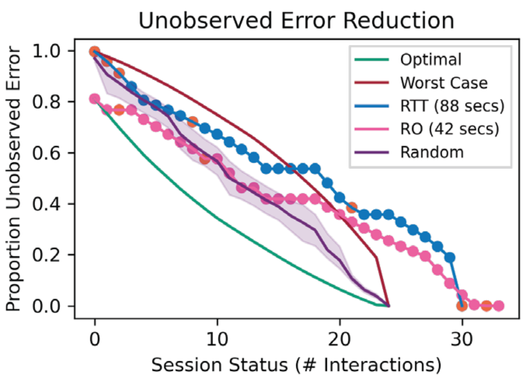
Method

- Patient data: CT, dose, del(hptc), del(pred) [4], and uncertainty volumes of two head and neck patients from HollandPTC
- Participants: experienced head and neck RTT and RO from HollandPTC
- Task: detecting clinically significant delineation errors within a 5-minute time window
- Conditions: with and without DEDS assistance

Results

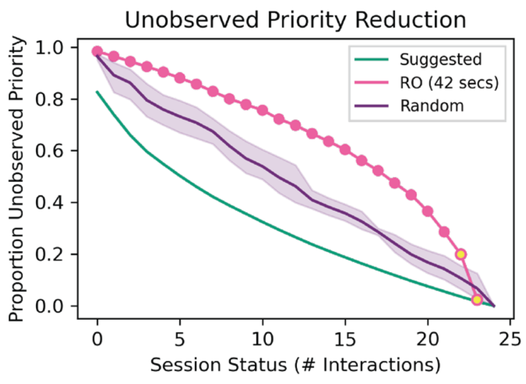
The figures below present the reduction in unattended error or user-defined priority for the RO and RTT workflows without and with DEDS assistance for the brainstem. We also include the optimal (suggested by DEDS), worst case (the opposite of the DEDS suggestions), and random workflows.

Without DEDS assistance



RO and RTT check the same slice several times

With DEDS assistance



RO refused to follow DEDS suggestion. Loos like the worst case!

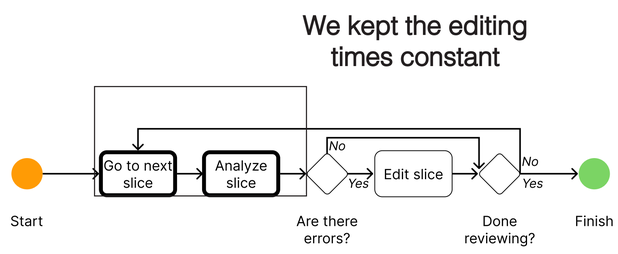
Simulation Study

Can DEDS effectively speed up the QA process?

Method

- Patient data: CT, dose, del(hptc), del(pred) [4] and uncertainty volumes of 41 head and neck patients treated at HollandPTC
- Approach: simulated QA process using model
- Conditions:
  - The information source used for sorting: mean error and max dose
  - The time it takes to analyze a slice

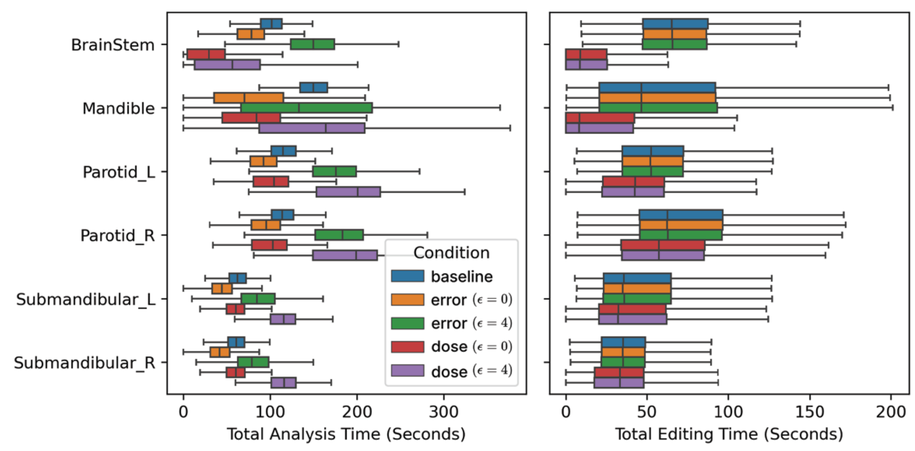
We kept the editing times constant



Times for analyzing and editing from [5]

Results

Summary statistics of the simulated analysis (left) and editing (right) times across 41 patients and 100 runs of the simulation. Epsilon=0 represents the currrent workflow and Epsilon=4 the workflow with DEDS suggestions, which require extra analysis time.



Conclusions

User study:
- Sequential error finding is suboptimal
- The proposed DEDS workflow fatigues and frustrates clinicians
- Clinicians prioritize based on dose distribution

Simulation study:
- Priority (based on dose) DEDS have the potential to reduce error detection time
- AI error modes and how the DEDS presents the errors (slice vs 3d regions) influence analysis time and QA duration

References

[1] Sonke, J. J., Aznar, M., & Rasch, C. (2019, July). Adaptive radiotherapy for anatomical changes. In Seminars in radiation oncology (Vol. 29, No. 3, pp. 245-257). WB Saunders.
[2] Cardenas, C. E., Yang, J., Anderson, B. M., Court, L. E., & Brock, K. B. (2019, July). Advances in auto-segmentation. In Seminars in radiation oncology (Vol. 29, No. 3, pp. 185-197). WB Saunders.
[3] Chaves-de-Plaza, N. F., Mody, P., Hildebrandt, K., Staring, M., Astreinidou, E., de Ridder, M., ... & van Egmond, R. (2022). Towards fast human-centred contouring workflows for adaptive external beam radiotherapy. In Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2022 Annual Conference.
[4] Mody, P. P., Chaves-de-Plaza, N., Hildebrandt, K., van Egmond, R., de Ridder, H., & Staring, M. (2022, April). Comparing Bayesian models for organ contouring in head and neck radiotherapy. In Medical Imaging 2022: Image Processing (Vol. 12032, pp. 100-109). SPIE.
[5] Aselmaa, A., van Herk, M., Laprie, A., Nestle, U., Götz, I., Wiedenmann, N., ... & Goossens, R. H. (2017). Using a contextualized sensemaking model for interaction design: a case study of tumor contouring. Journal of Biomedical Informatics, 65, 145-158.

Institutions

1 TU Delft
2 HollandPTC Protonen Therapie Centrum
3 LUMC Leids Universitair Medisch Centrum
4 TU/e Eindhoven University of Technology
5 UMC Utrecht

Acknowledgements

The research for this work was funded by Varian, a Siemens Healthineers Company, through the HollandPTC-Varian Consortium (grant id 2019022), and partly financed by the Surcharge for Top Consortia for Knowledge and Innovation (TKIs) from the Ministry of Economic Affairs and Climate.