UNIVERSITY OF TARTU
Institute of Computer Science

BCV Lab

Interactive AI for Efficient Annotation in Medical Imaging

Author: Venkata Narayana Bommanaboina, Supervisor: Dmytro Fishman, PhD

Abstract

Manual delineation of tumors in 3D CT volumes is time-consuming and requires significant radiology expertise, limiting the scalability of training data for deep learning segmentation models. Yet most public CT datasets contain only weak labels or none at all. We present a semi-automated human-in-the-loop annotation pipeline that converts weakly annotated and unlabeled CT data into expert-validated volumetric tumor annotations. Using nnInteractive and nnU-Net models for pseudo-label generation, radiology experts review and validate cases. Evaluated on DeepLesion and PanTS, the pipeline demonstrates iterative segmentation improvement, showing that interactive AI can efficiently produce curated 3D tumor annotations at scale.

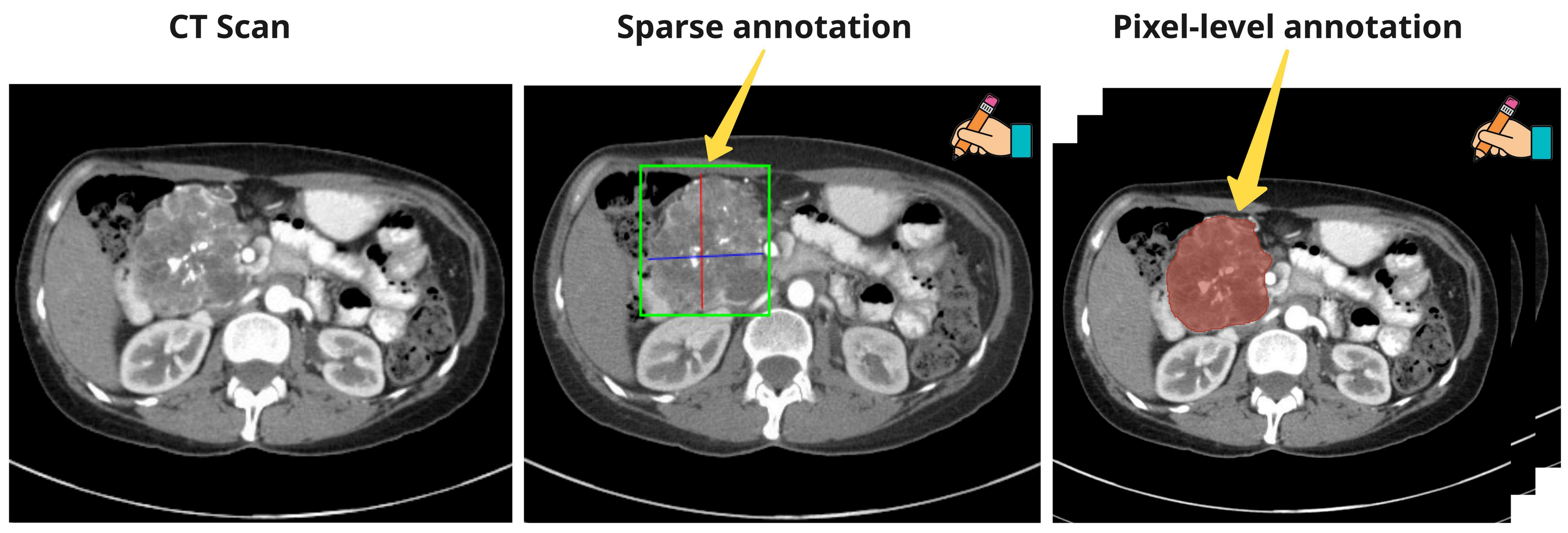

Figure 1. Manual 3D tumor annotation is time-consuming, expertise-intensive and costly.

Proposed Annotation Pipeline

For input CT scans, the initial 3D pseudo-labels are generated either by

- the nnInteractive model using available sparse 2D slice-level lesion labels as interactive scribble prompts as input, or
- an organ-specific trained nnU-Net when no slice-level labels are available,

A radiology expert validates each prediction. Accurate cases are added to the training cohort and inaccurate cases are reprocessed by a retrained nnU-Net model. This loop repeats iteratively, progressively improving annotation quality.

Annotation Efficiency

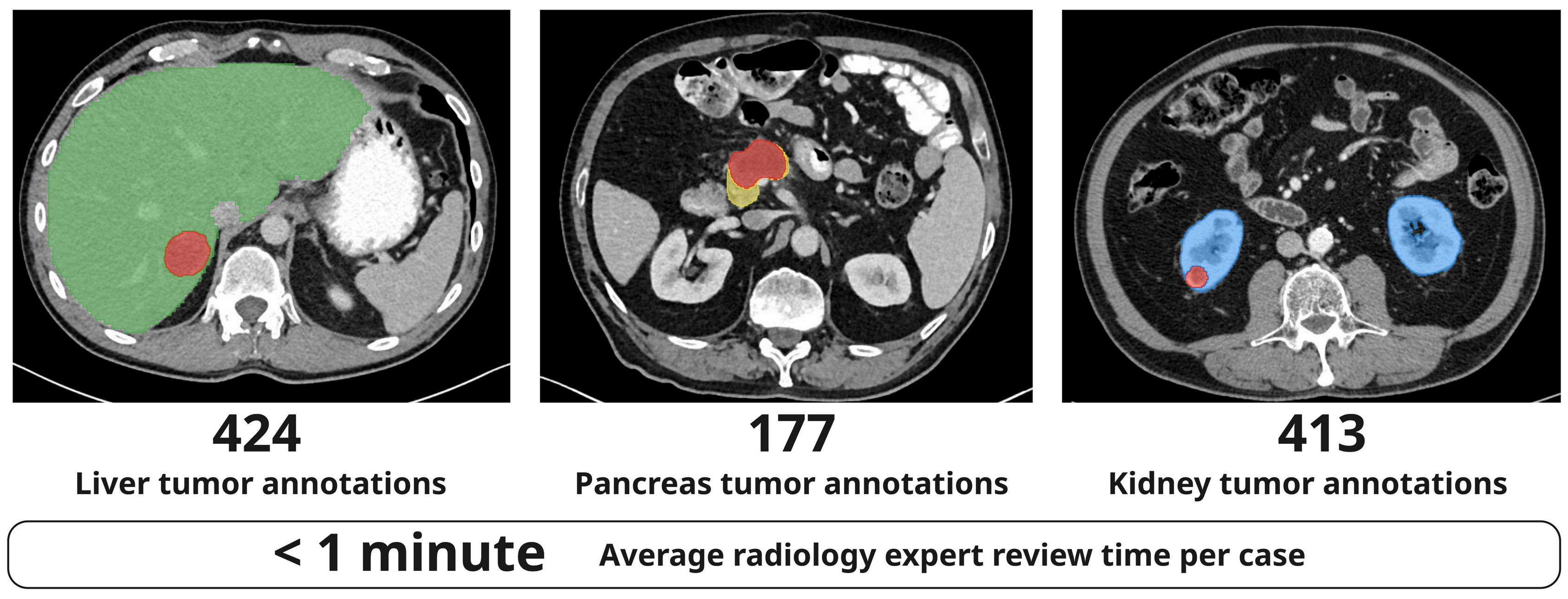

Figure 3. Expert-validated annotations generated by the proposed pipeline.

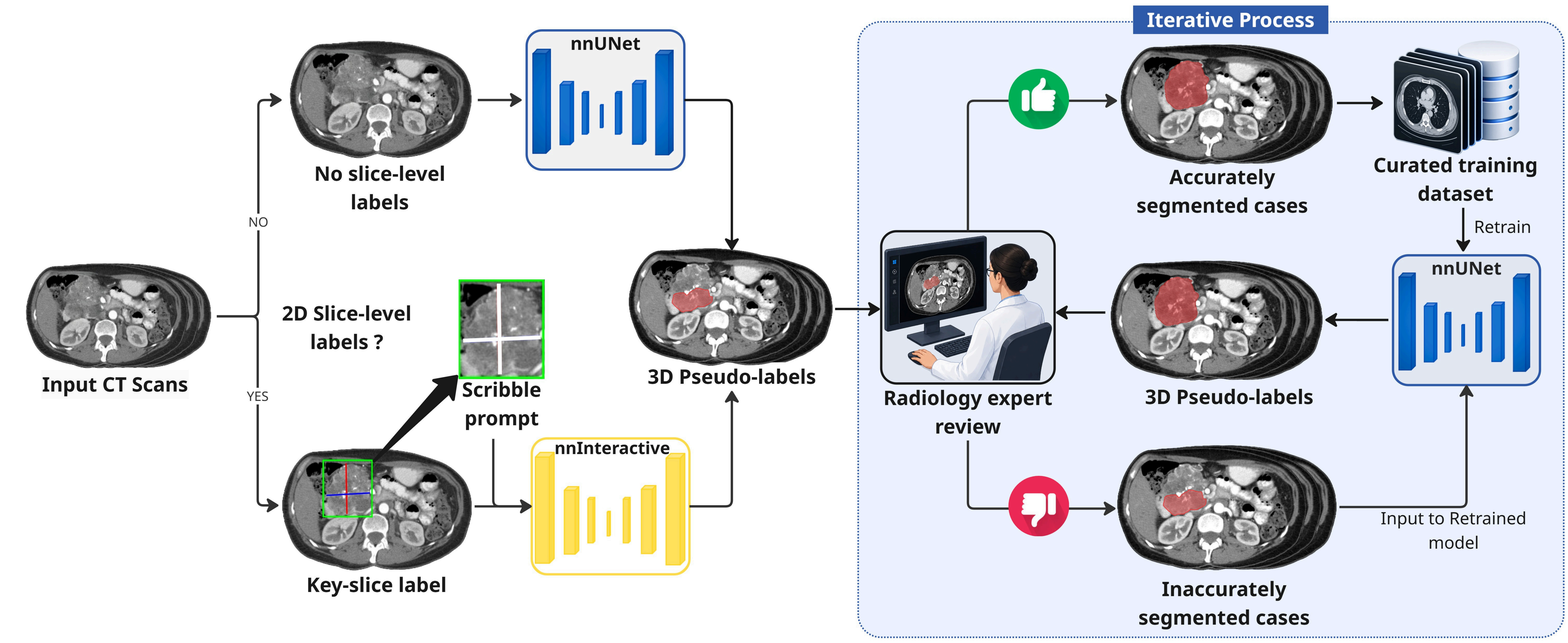

Figure 2. Overview of the proposed **semi-automated human-in-the-loop annotation pipeline**.

Iterative Improvement Results

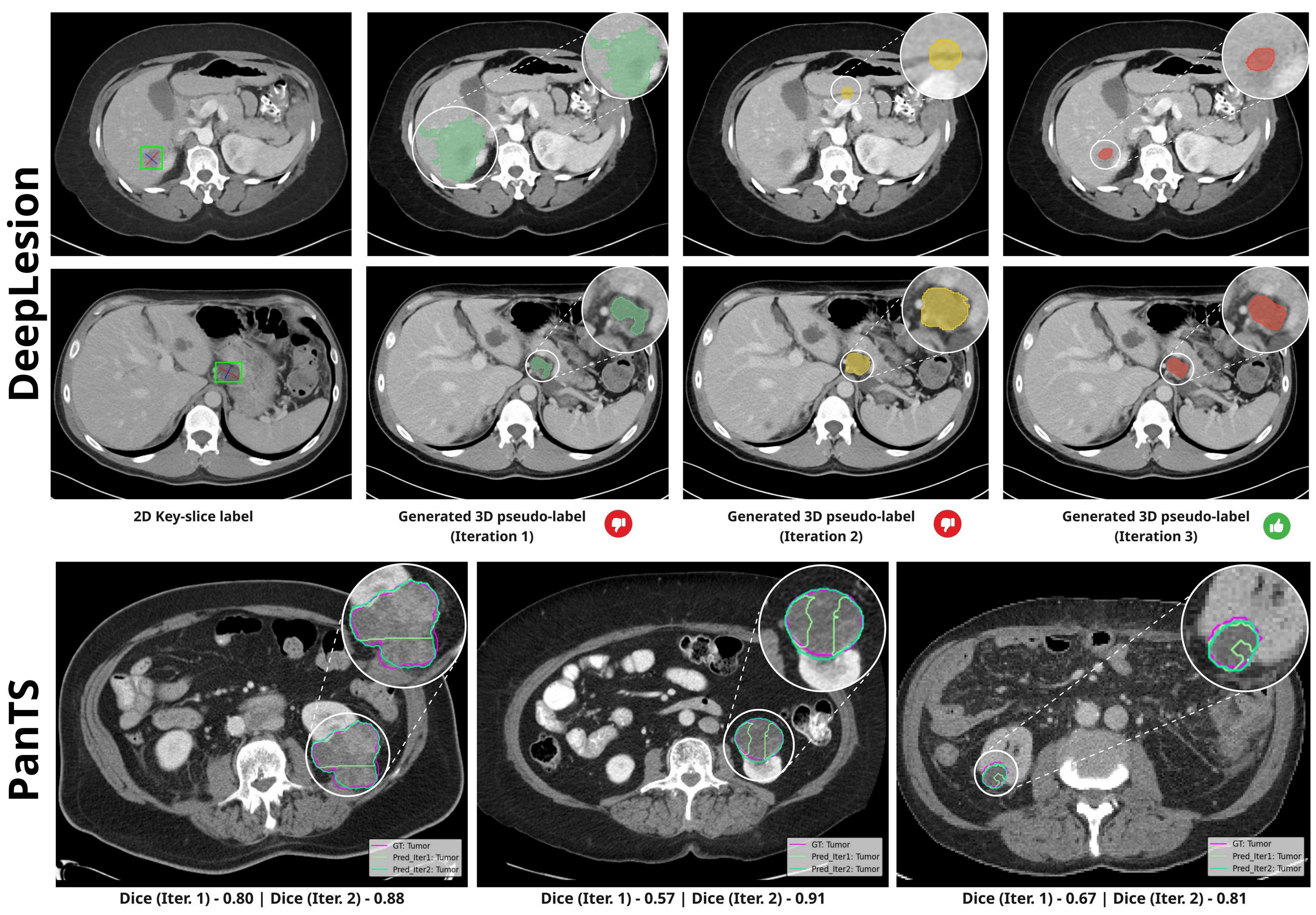

Figure 4. Qualitative results on DeepLesion and Quantitative results on PanTS

Acknowledgments

Special thanks to my supervisor, Dmytro Fishman, for invaluable discussions, and to the Biomedical Computer Vision Lab for their continuous support and collaboration. We thank the High-Performance Computing Cluster at the University of Tartu for providing the computational resources.

Downstream Segmentation Evaluation

We assessed whether expert-validated pseudo-labels generated by the pipeline improve supervised nnU-Net training. For DeepLesion, liver and pancreas pseudo-labels were added to public training cohorts and evaluated across external test sets. For PanTS, expert-validated kidney tumor pseudo-labels were added to KiTS training and evaluated on KiTS and AbdomenAtlas 2.0.

Model ↓	MSD Liver	LiTS	HCC	CRLM
MSD Liver	**0.812**	**0.809**	0.615	0.774
MSD Liver w/ DL	0.518	0.494	0.585	0.679
All data w/o DL	0.648	0.701	0.657	**0.858**
All data w/ DL	0.686	0.618	**0.666**	0.794

Table 1. Comparison of liver tumor segmentation across different trained nnU-Net models

Model ↓	KiTS		AbdomenAtlas 2.0	
	Kidney	Tumor	Kidney	Tumor
KiTS w/o PanTS	0.956	0.768	0.857	0.366
KiTS w/ PanTS	**0.960**	**0.802**	**0.860**	**0.470**

Table 2. Comparison of kidney tumor segmentation across different trained nnU-Net models

Conclusions

This work shows that interactive AI can efficiently transform weakly annotated and unlabeled CT scans into expert-validated 3D tumor annotations. The pipeline reduces manual annotation effort through radiologist-guided validation and iterative retraining, while downstream results highlight that the value of pseudo-labels depends on dataset compatibility.