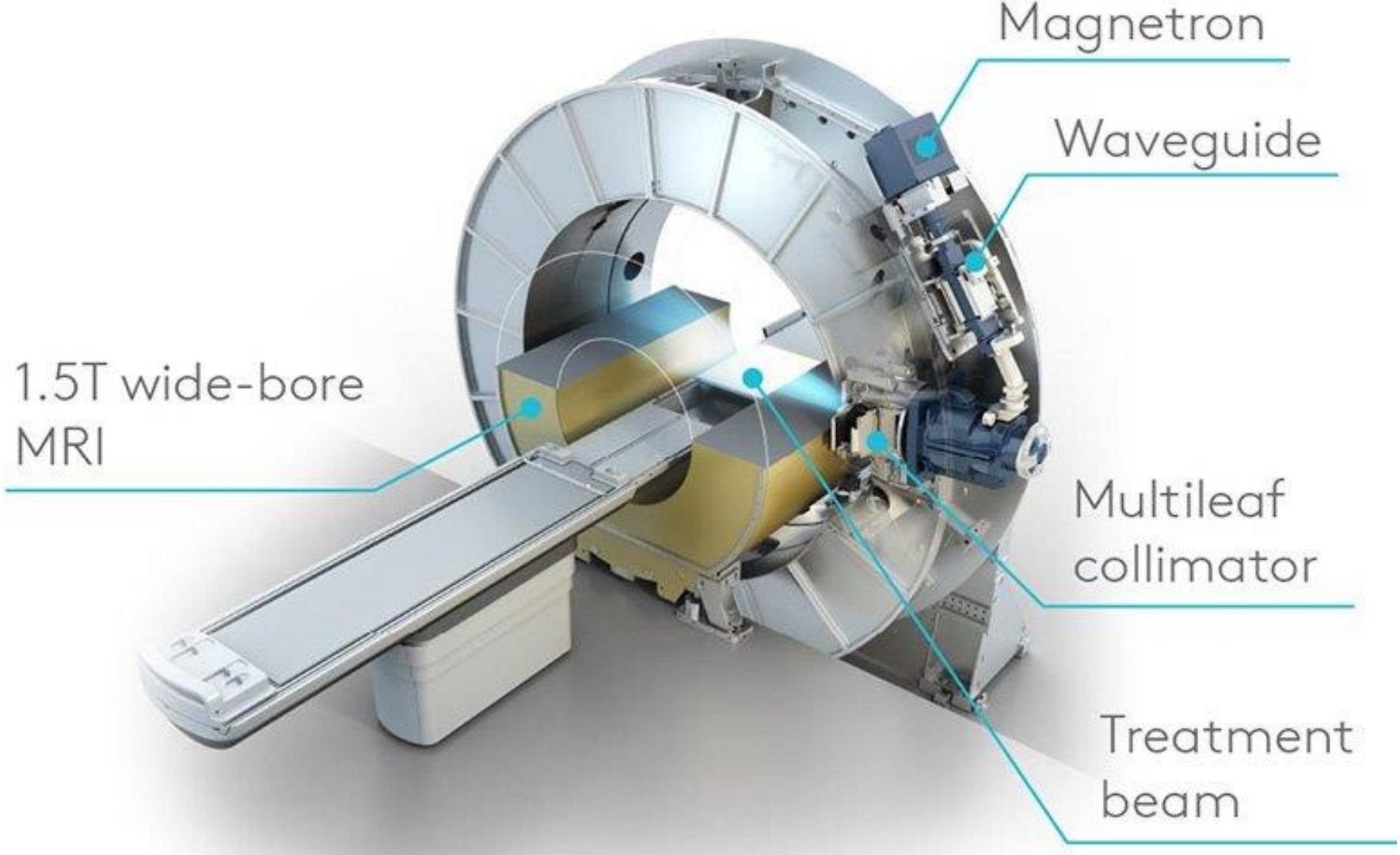
TrackRAD Project Overview

June 25, 2025

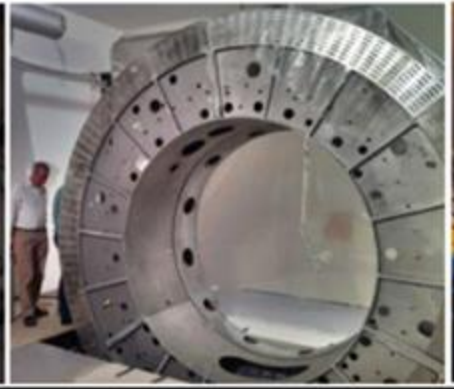
MRI-guided linear accelerator (MRI-linac)

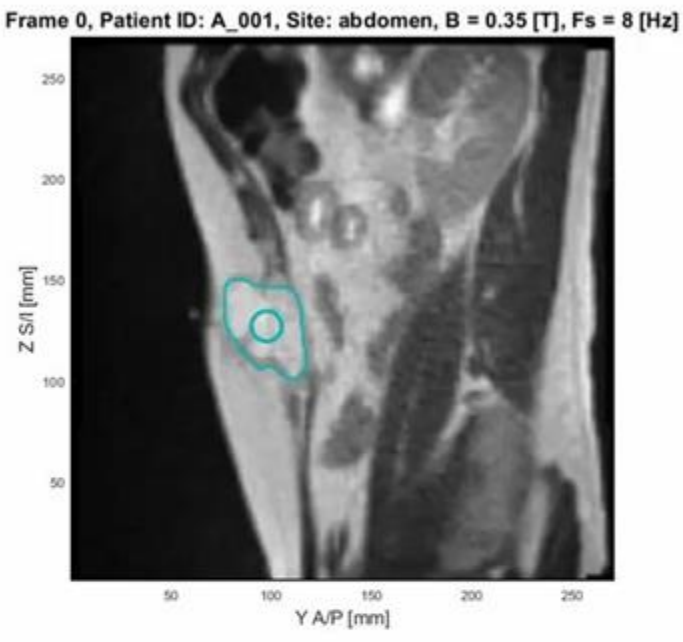
- Combine real-time MR imaging with a linear accelerator for radiation treatment
  - Use 0.35T or 1.5T field strength (less than diagnostic MR but good enough for soft tissue contrast)
  - Linac = high energy X ray beams
  - Beam is often perpendicular to magnetic field
- Uses
  - Real-time tumor tracking to account for motion during breathing
  - Adjust X ray beam in real-time based on MR imaging

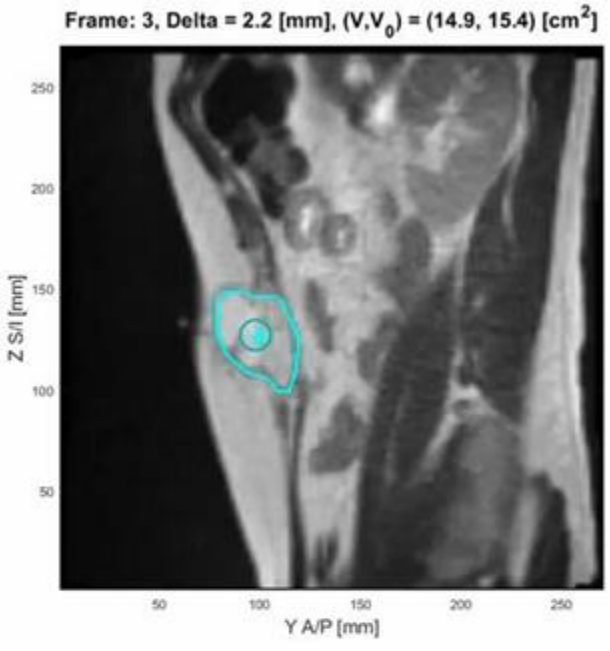

Magnetron
Waveguide
1.5T wide-bore MRI
Multileaf collimator
Treatment beam

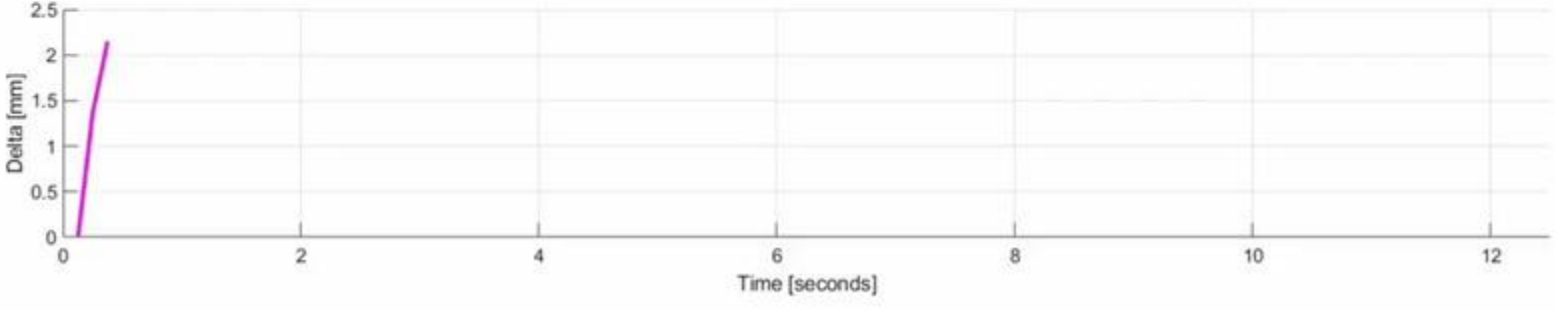
MR-LINAC TIMELINE (LAGENDIJK, RADONC 2022)

1999	2004	2005	2009	2012	2014	2018
MRI guided radiotherapy: a MRI based linear accelerator Lagendijk JJW and Bakker CJG 2000 Proc. ESTRO Istanbul 19th Annual Meeting Rad Onc vol 56, suppl. 1, pp S1–S255						
invention	design	Initial exp.	1st prototype	2nd prototype	3rd prototype	Clinical

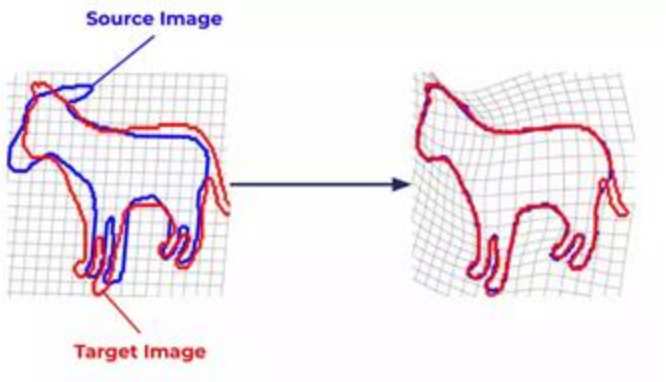


Frame 0, Patient ID: A_001, Site: abdomen, B = 0.35 [T], Fs = 8 [Hz]
Z S/I [mm]
Y A/P [mm]

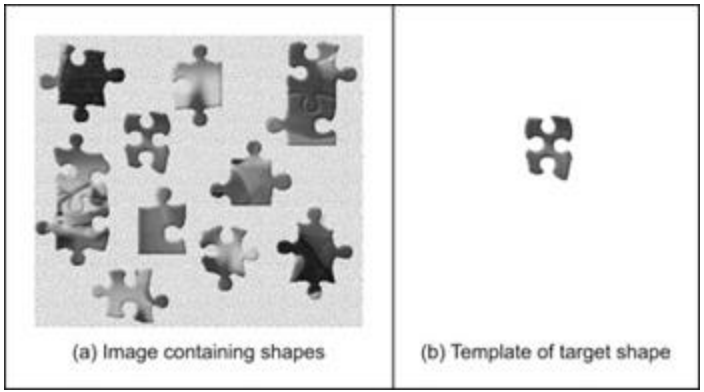

Frame: 3, Delta = 2.2 [mm], (V,V0) = (14.9, 15.4) [cm2]
Z S/I [mm]
Y A/P [mm]


Delta [mm]
Time [seconds]

MR object tracking status quo

- Current clinically available solutions use:
  - Deformable image registration – vector field maps points from frame to frame
  - Template matching – slide a reference image across the target, calculate similarity scores, optimize
- Both struggle with large, non-rigid motion
- Best practice = turn off the beam when “large motion” is detected (gating)
  - This wastes valuable scanner time
- MRI adds further complexity
  - Lack of standardization in field strengths, imaging protocols
  - Variance in image quality and artifacts
  - Pixel intensities are relative and vary b/w scans (vs. CT where 0 intensity == water)



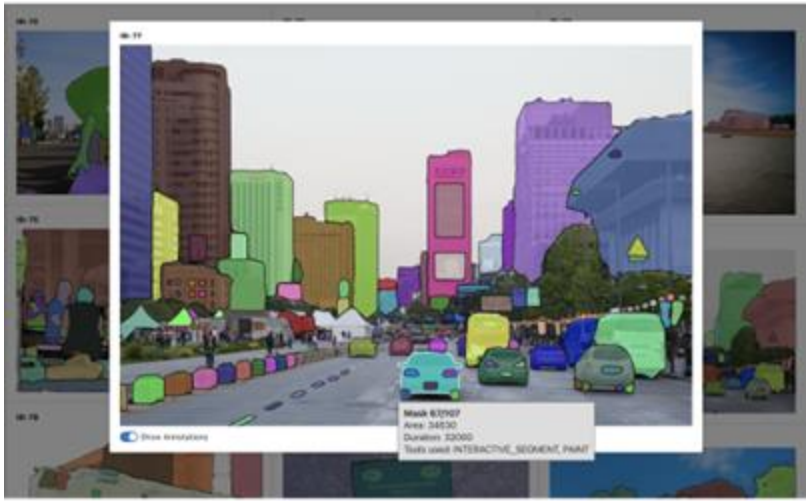


TrackRAD2025

- Goal = improve real-time tumor tracking algorithm for MRI-linac
- Task
  - Input
    - **Sequence of 2D MRI frames (AKA a video), 100-200 frames (~10-30s at 8 fps)**
    - **First frame labeled (all pixels pertaining to the tumor)**
    - Metadata (field strength, frame rate)
  - Output = labels for all subsequent frames
- Metrics = image similarity (Dice coefficient), runtime (<1s/frame)
- Dataset
  - From 6 institutions (3 Dutch, 1 German, 1 Australian, 1 Chinese)
  - 0.35T and 1.5T MRI
  - Labeled = 108 patient cases = 10_000 frames
    - 50 for training, 8 + 50 held out for validation and testing
  - **Unlabeled = 477 patient cases = 2_800_000 frames**

Segment Anything Model (SAM) (2023)

- Foundation vision model by Meta
- Transformer architecture
- Trained on 11M images + 1.1B segmentation masks
- “Promptable” = use positive/negative points, bounding box
- Good “zero-shot” performance = applied to a task without fine-tuning
- Can be fine-tuned (trained with domain-specific data)
- https://segment-anything.com/demo



SAM2 (2024)

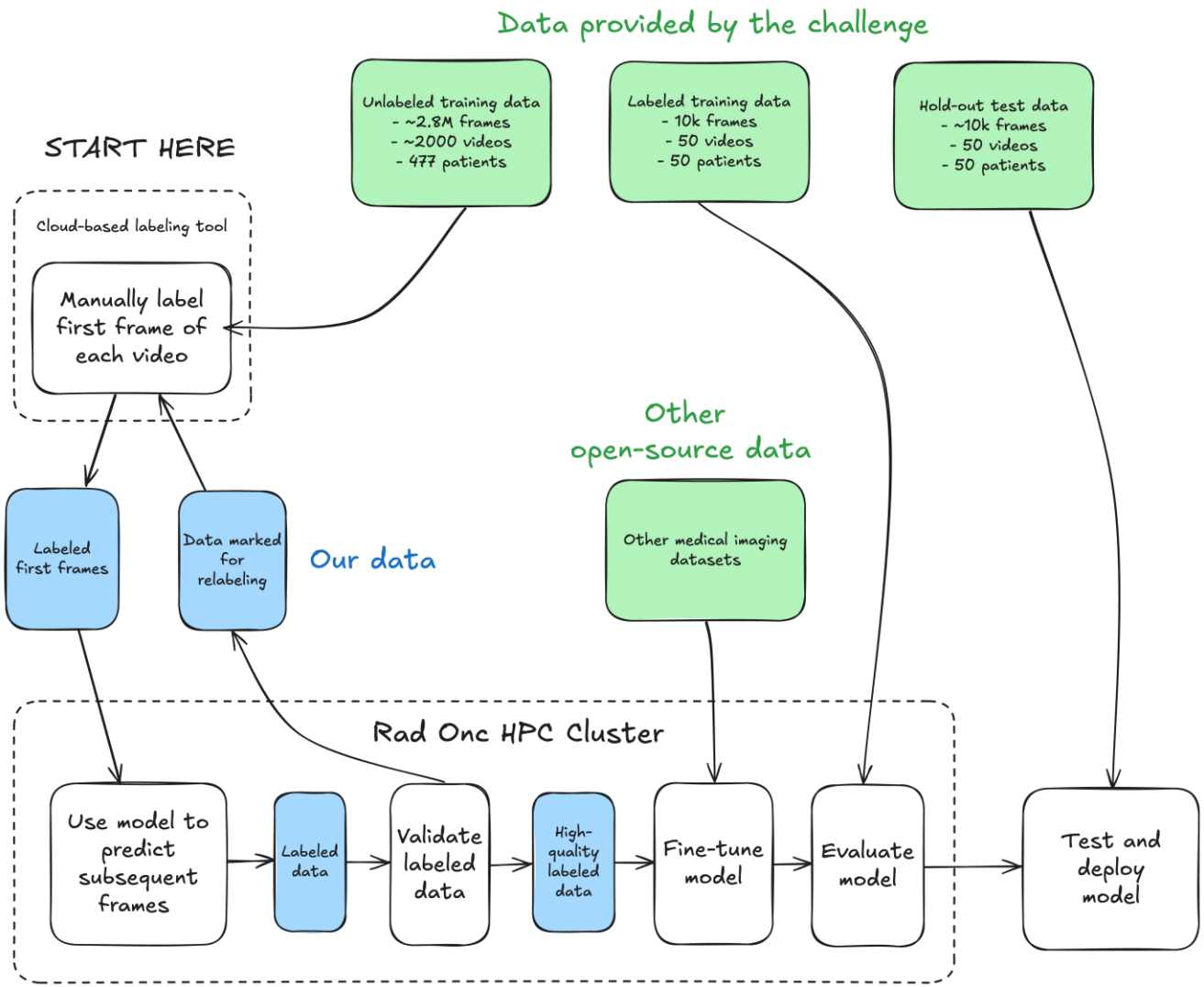
- Adds video object segmentation (VOS) functionality
  - **Given one frame + label, infer labels in all subsequent frames**
  - Robust to deformation, object occlusion, entering and leaving the field of view
- Trained on video
  - Training data includes the previous + following 8 frames
- https://ai.meta.com/sam2/

SAM2 evaluation – it's pretty good!

```
# Zero-shot with 50 labeled test cases
  "aggregates": {
      "dice_similarity_coefficient": 0.9071, # 2 * image overlap / total pixels, [0,1]
      "hausdorff_distance_95": 3.546, # for each point, min dist to any other point, 95th pct
      "surface_distance_average": 1.303, # for each point, min dist to any point, mean
      "center_distance": 1.595, # dist b/w centers of mass
      "relative_d98_dose": 0.9434, # est accuracy of X-ray dose, scatter and stuff
      "time_per_frame": 0.1900, # must be <1
  }
```

Modeling plan

1. Use SAM2 for tumor tracking
    a. Frame sequence + first frame label → Infer subsequent frames
2. Use SAM2 to expedite annotation of unlabeled frame sequences
    a. Assist in annotating first frames w/bounding box → generate mask
    b. Infer masks for subsequent frames
    c. Validate mask quality using traditional rule-based methods
    d. Flag bad frames for relabeling
3. Fine-tune SAM2 using the labeled data
4. (Maybe) Find-tune with other open-access datasets
5. Deploy fine-tuned model


Data provided by the challenge
Unlabeled training data
- ~2.8M frames
- ~2000 videos
- 477 patients
Labeled training data
- 10k frames
- 50 videos
- 50 patients
Hold-out test data
- ~10k frames
- 50 videos
- 50 patients
START HERE
Cloud-based labeling tool
Manually label first frame of each video
Other open-source data
Other medical imaging datasets
Labeled first frames
Data marked for relabeling
Our data
Rad Onc HPC Cluster
Use model to predict subsequent frames
Labeled data
Validate labeled data
High-quality labeled data
Fine-tune model
Evaluate model
Test and deploy model

Link

Labeling tool

https://huggingface.co/spaces/mayo-radonc/bouncing-target

- HuggingFace = tech company that provides software + services for ML apps
  - Hosts datasets (including TrackRAD)
  - Hosts applications, managing all the hardware
- Unlabeled data = 477 patients, ~2000 sequences, 2.8M frames
- What do I label?
  - Any and all objects
  - Goal = not “find the tumor” but “get better at black and white, fuzzy, medical images”

The importance of labeled data

- Labeling is a critical part of this project
- Anyone can download a foundation model from the internet and run it
- **Our competitive edge = rapidly labeling a large dataset for finetuning**
- 2.8M frames is a lot of data, but our plan is to can automate 99% of it, manually labeling a few key frames per video
- We will also look into other open-access imaging datasets

Up next

1. Complete labeling tool (ETA early July)
2. **Start labeling!**
3. Develop annotation pipeline
    a. Download labeled frames from HuggingFace
    b. Use SAM2 to infer labels for subsequent frames
    c. **Validate frames, flag bad frames**
4. Fine-tune SAM2 with labeled frames in batches
5. **(Maybe) Fine-tune SAM2 with other open source datasets**
6. **Optimize SAM2 parameters (confidence threshold)**
7. **Submit model (August 15)**

What’s in it for me?

- First-hand experience with AI tools in medical imaging + computer vision
- The top 5 teams (team limit = 5 people) get
  - A cash prize ($1000 to $200, will distribute amongst everyone who participates)
  - Invited to present at MICCAI 2025, 9/23-27 in S. Korea
  - Included in a “challenge” paper submitted to Medical Image Analysis or a similar journal
  - https://trackrad2025.grand-challenge.org/teams/
- We're currently in 1st place among preliminary submissions (as of June 25), so it's definitely doable!
  - https://trackrad2025.grand-challenge.org/evaluation/preliminary-testing/leaderboard/
- Additional paper(s) as appropriate