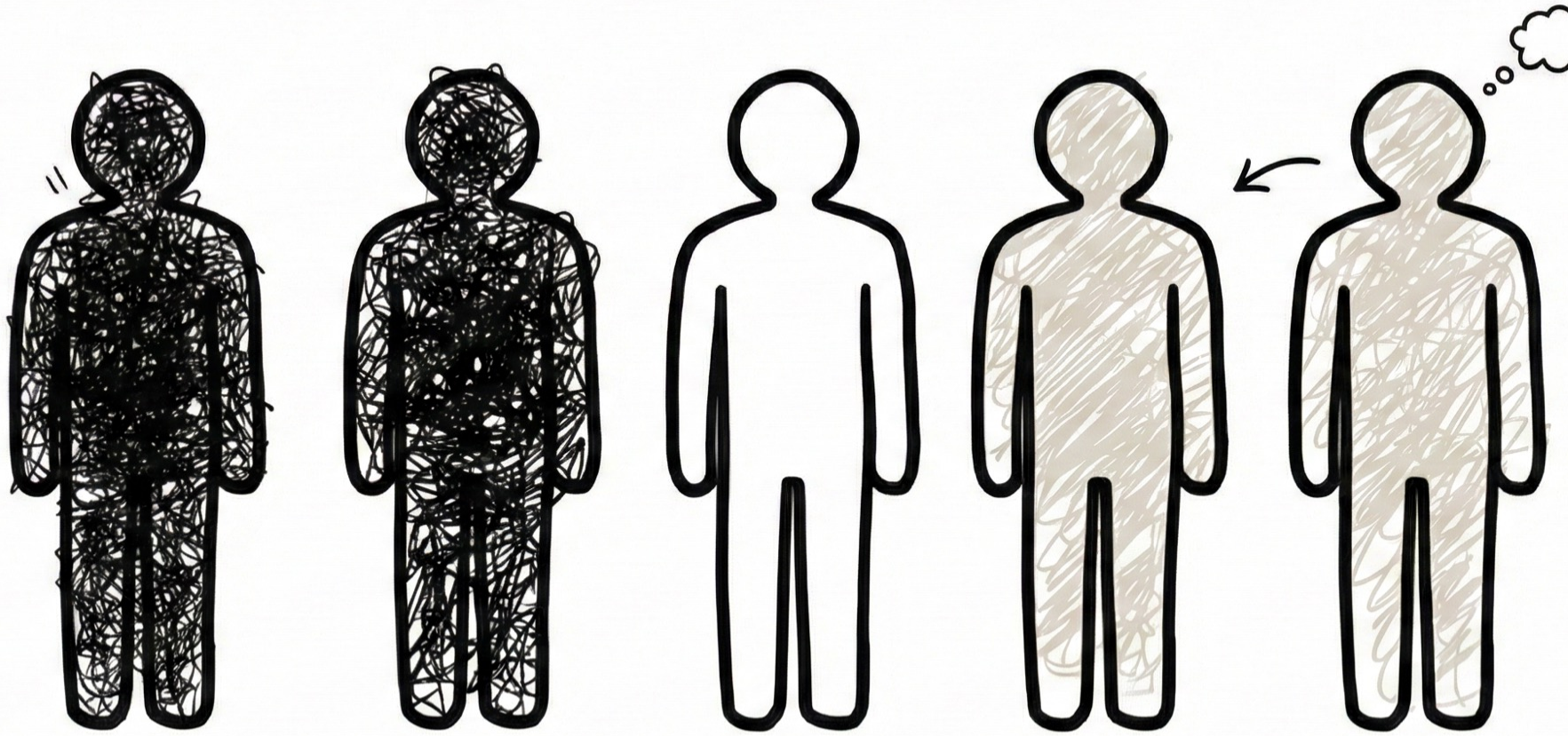
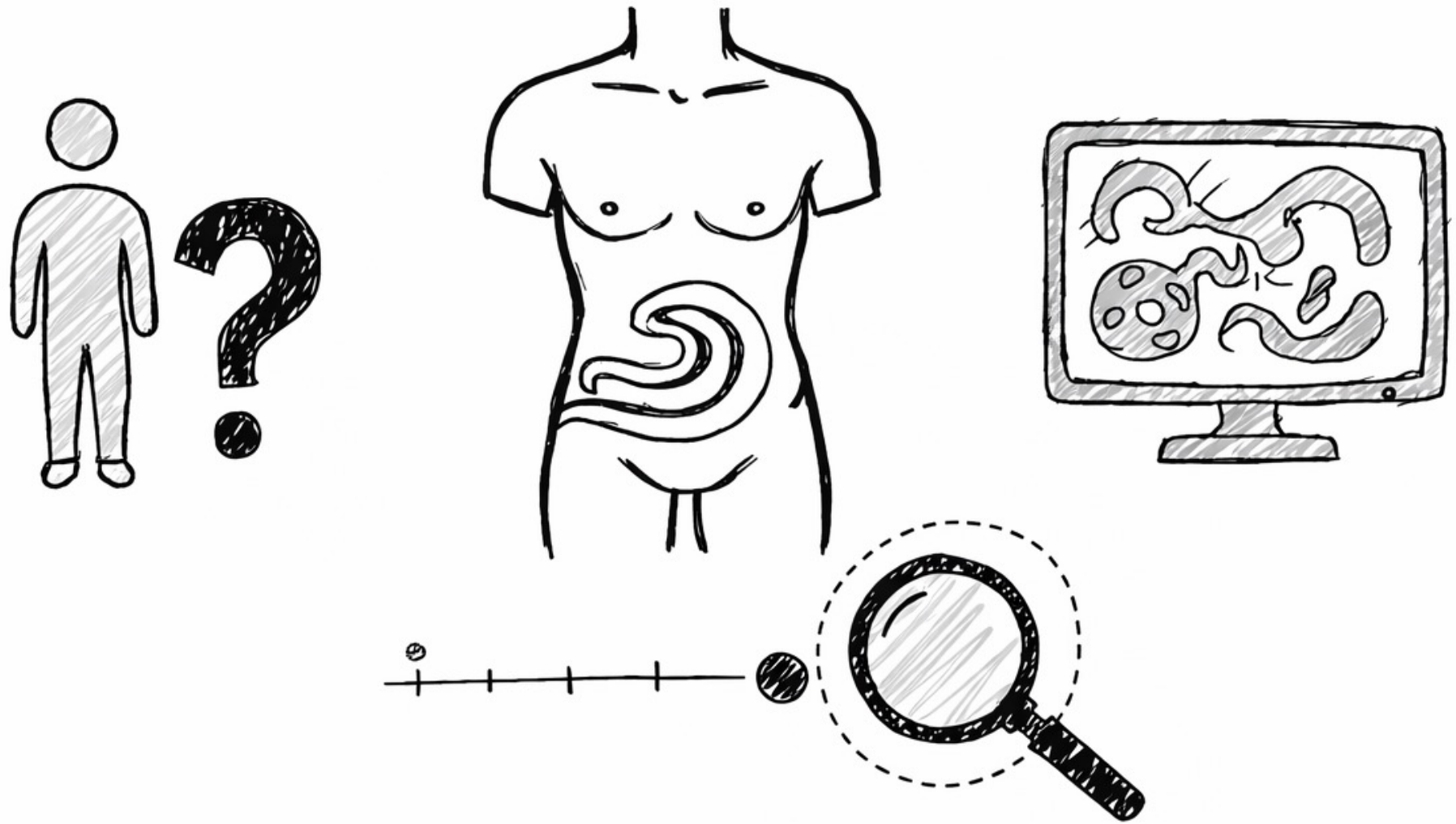
PanTS: The Pancreatic Tumor Segmentation Dataset

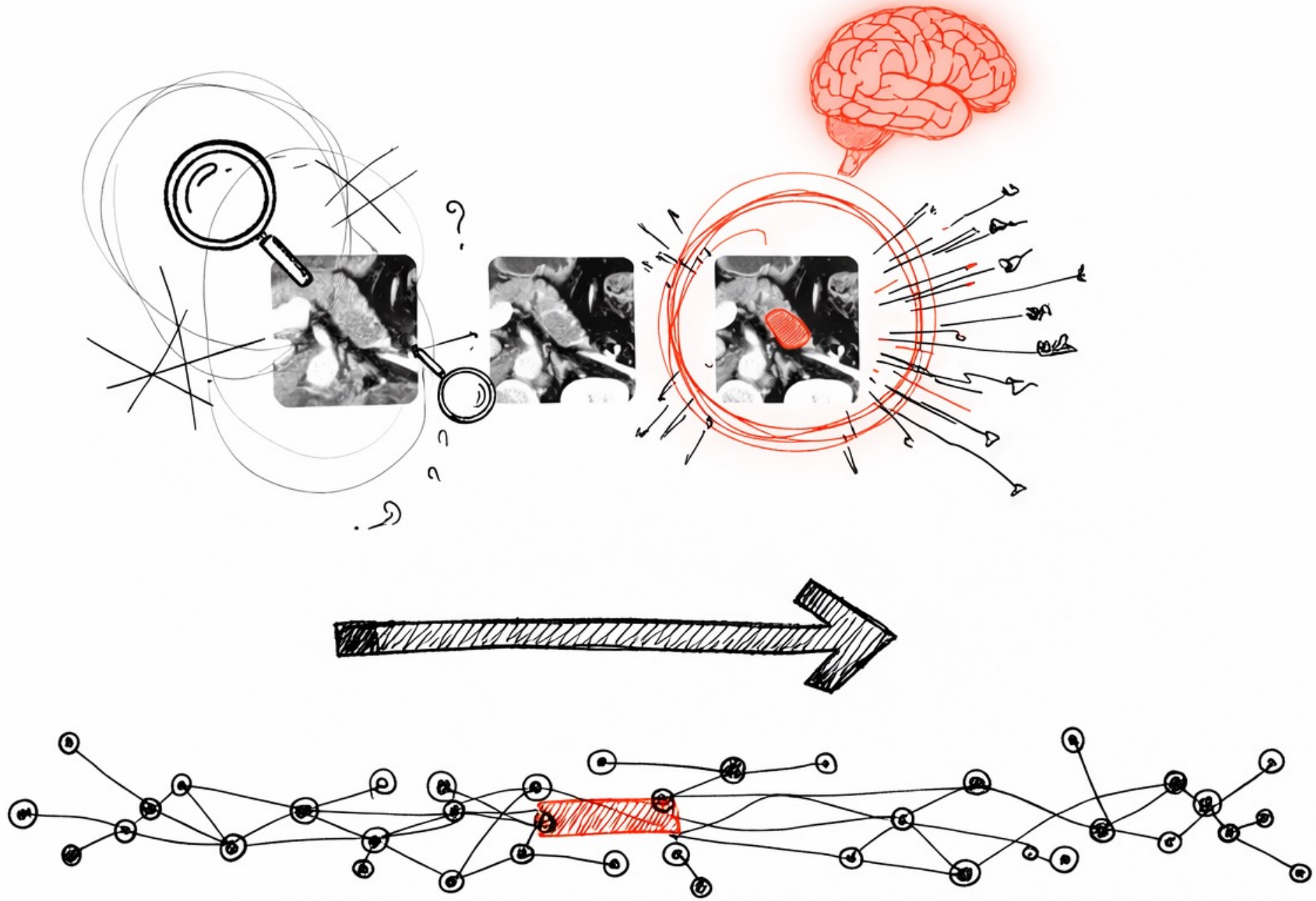
Wenxuan Li · Xinze Zhou · Qi Chen · Tianyu Lin · Pedro R. A. S. Bassi · Xiaoxi Chen · Chen Ye · Zheren Zhu · Kai Ding · Heng Li · Kang Wang · Yang Yang · Yucheng Tang · Daguang Xu · Alan L. Yuille · Zongwei Zhou

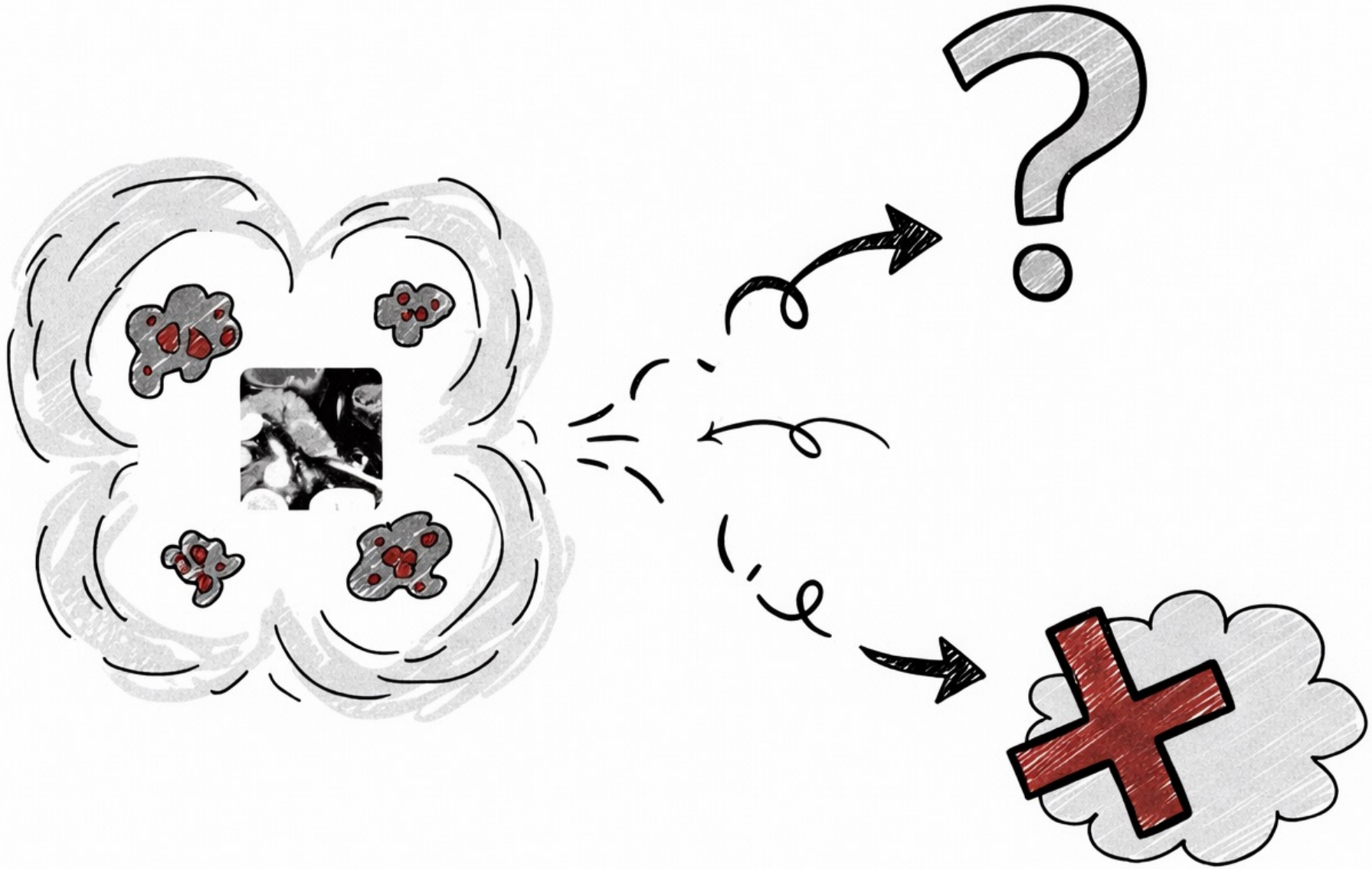


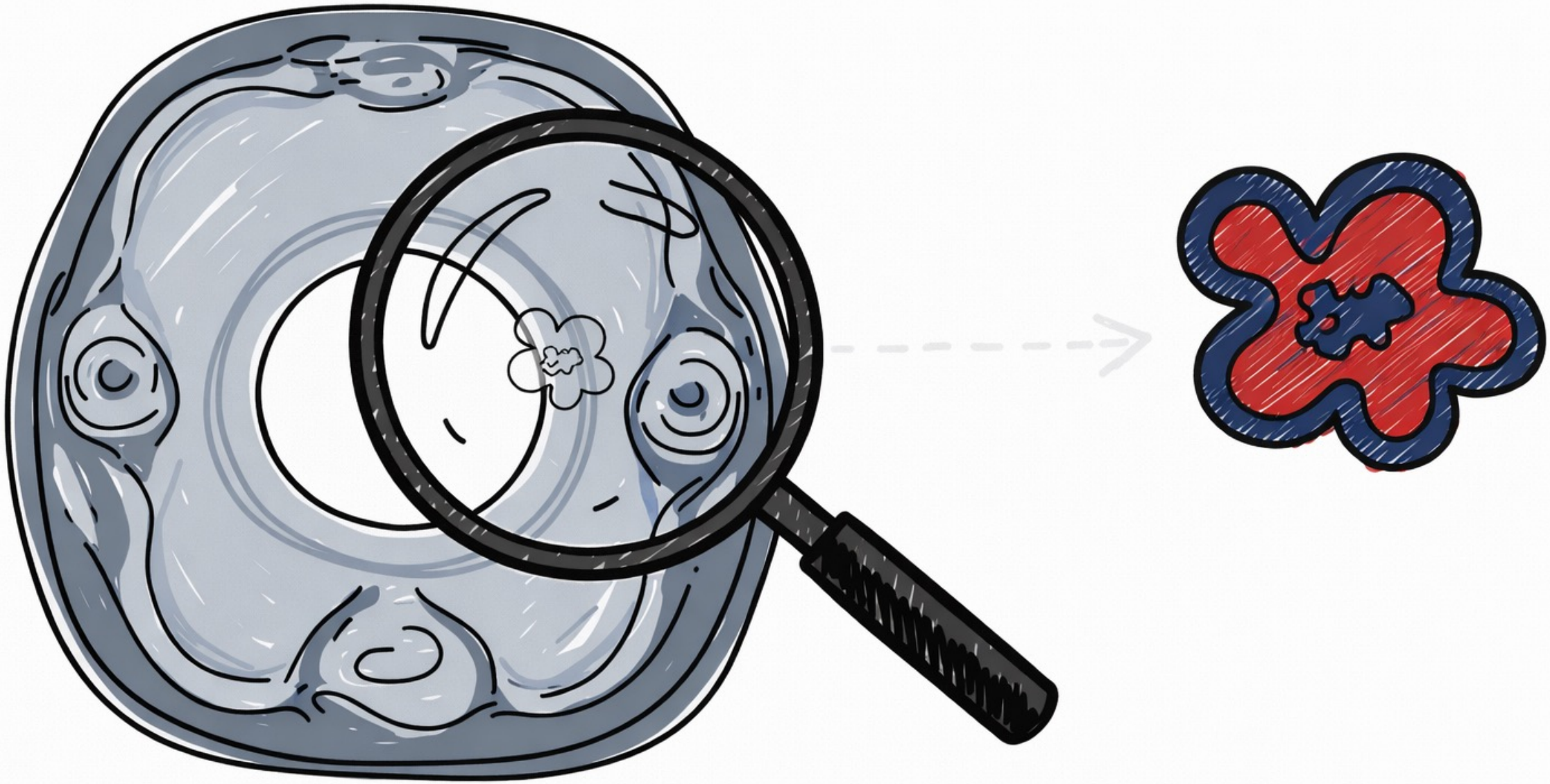
Why is it so hard to spot before it's too late?

The Detection Challenge

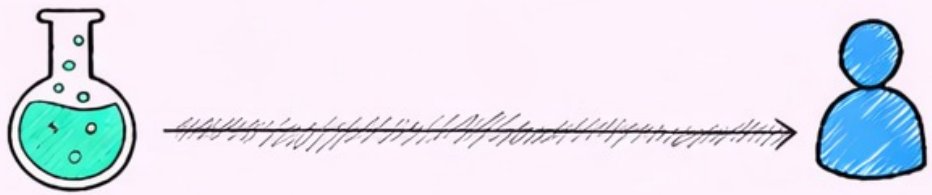


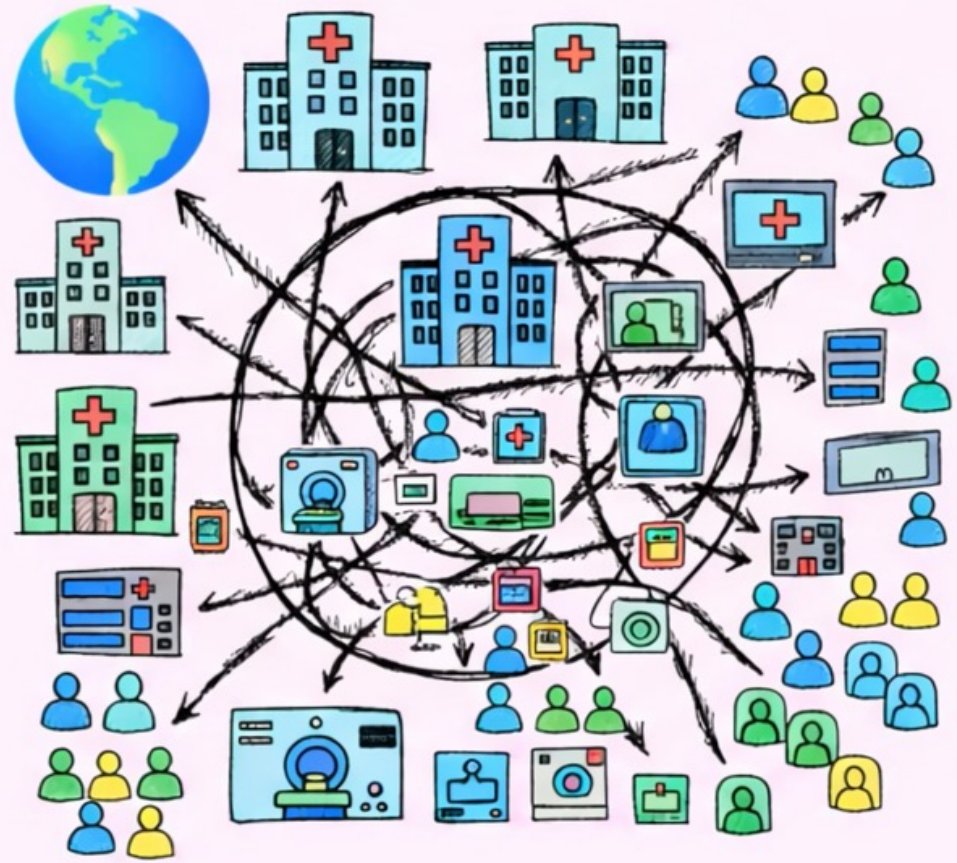


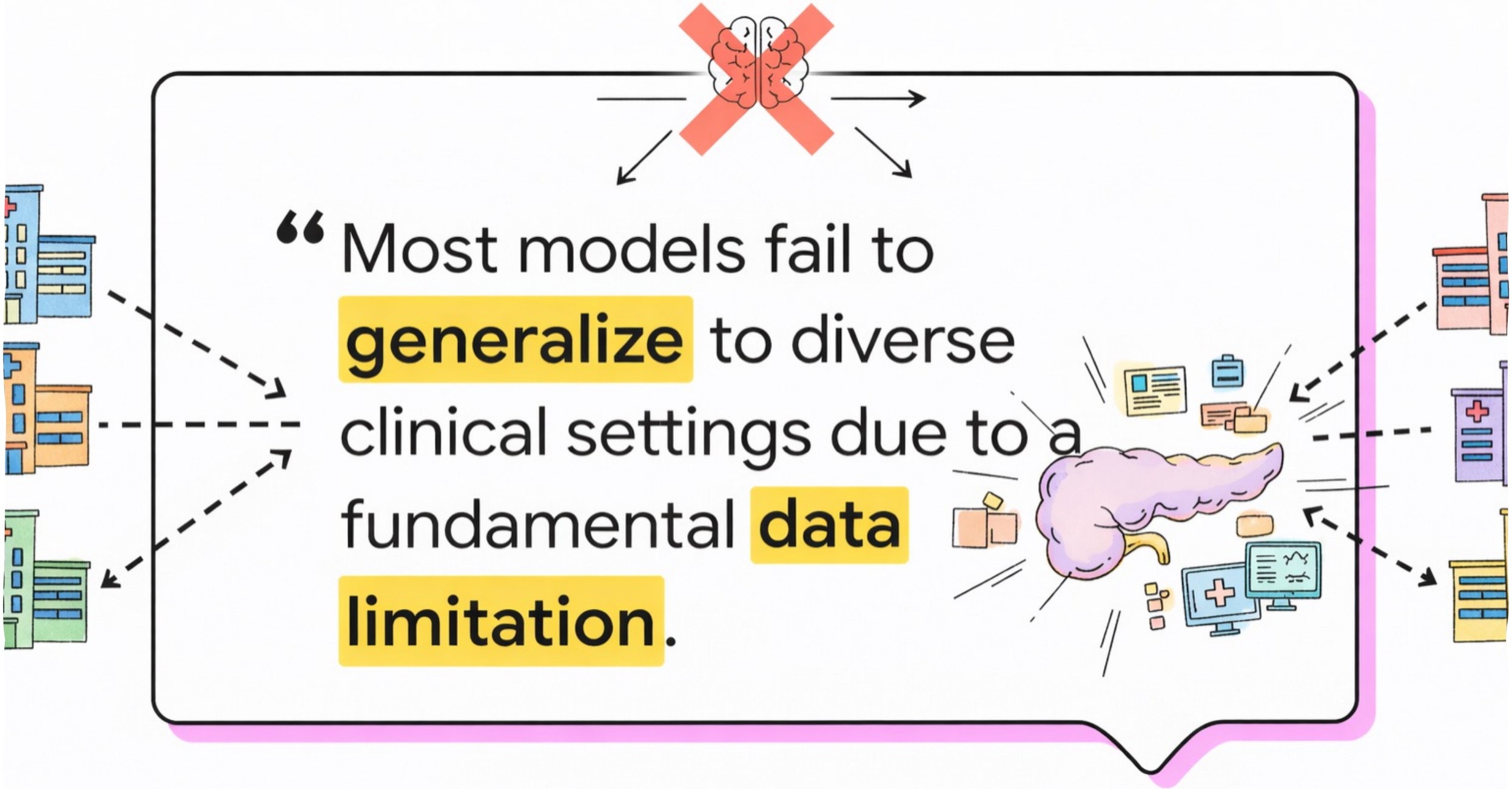





E

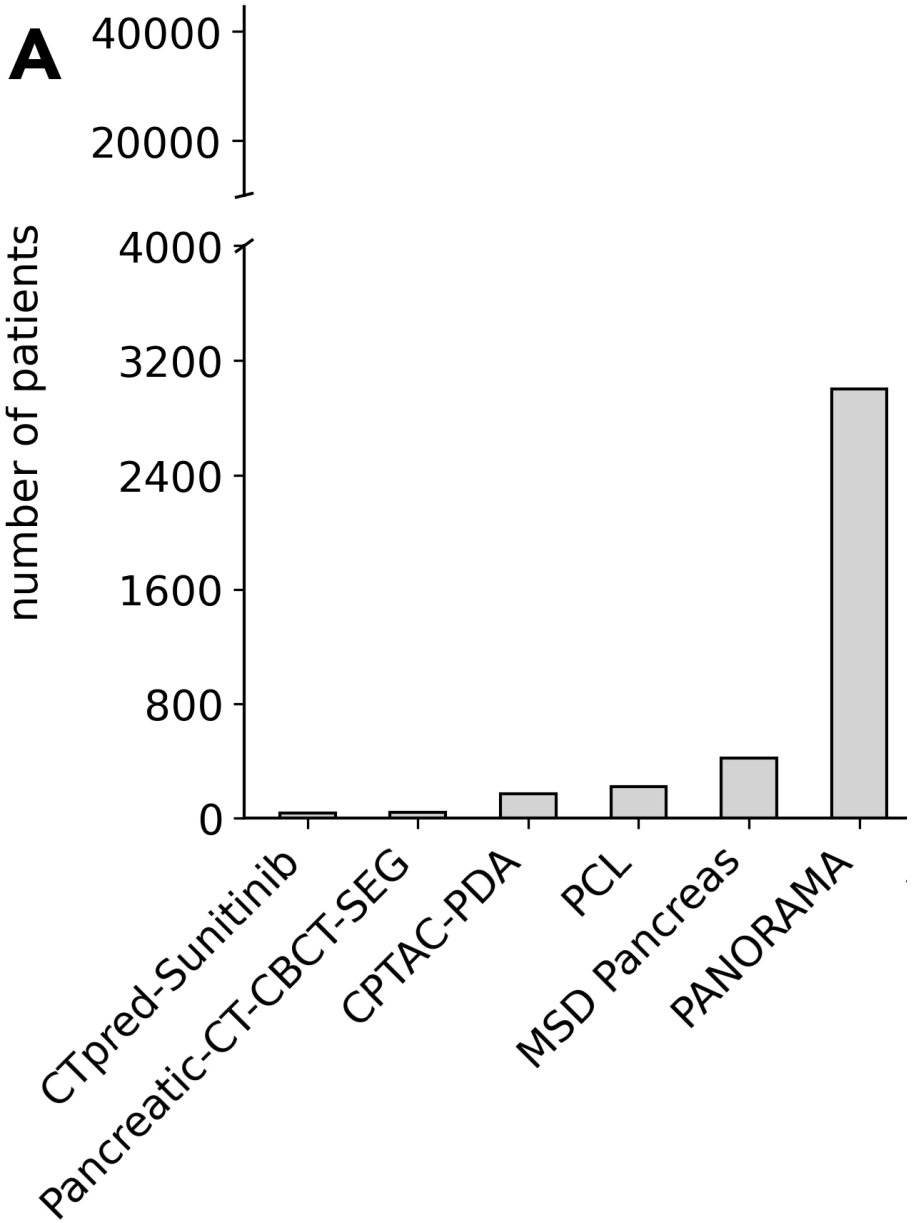


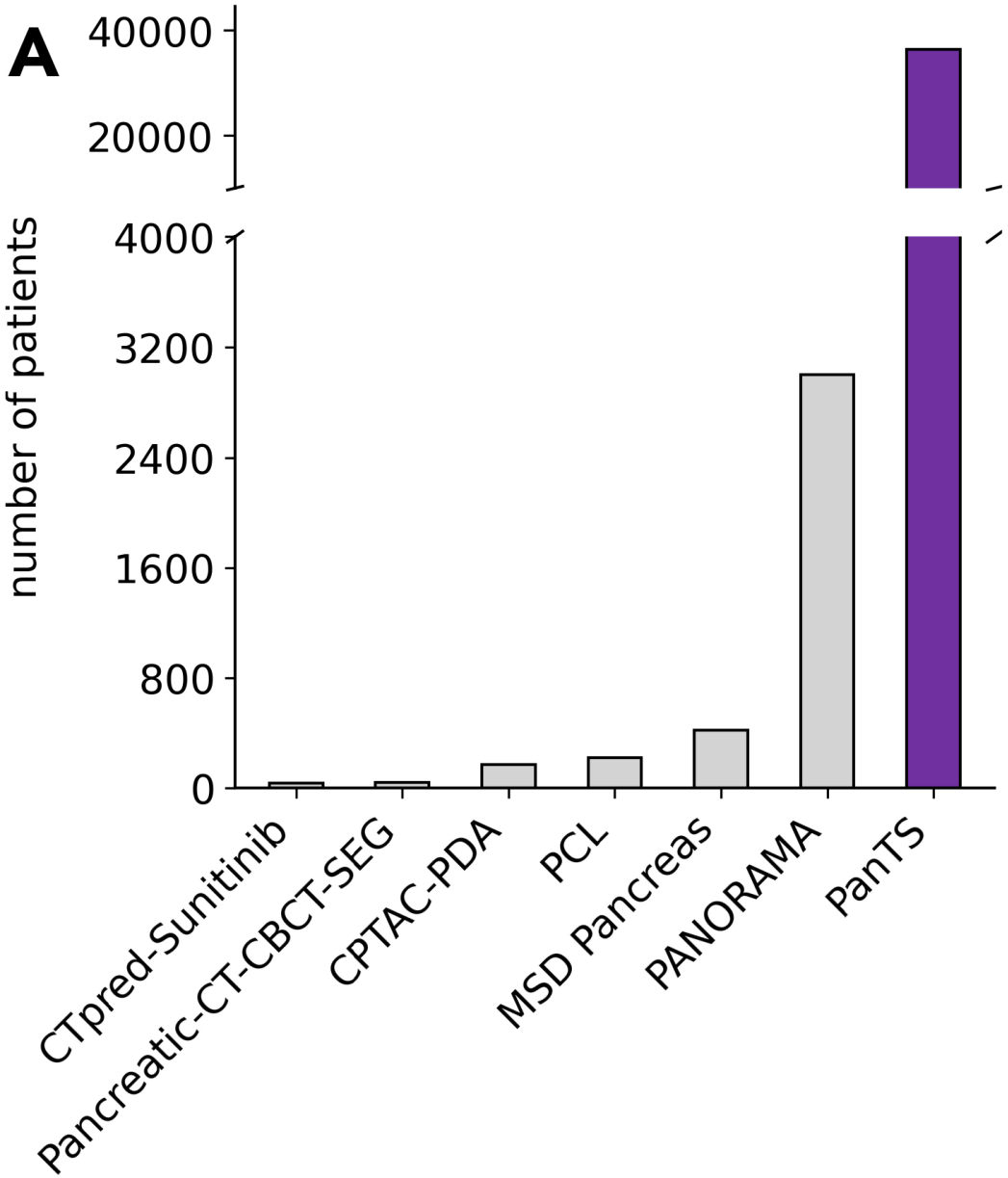



“Most models fail to generalize to diverse clinical settings due to a fundamental data limitation.

Challenges (1/3)

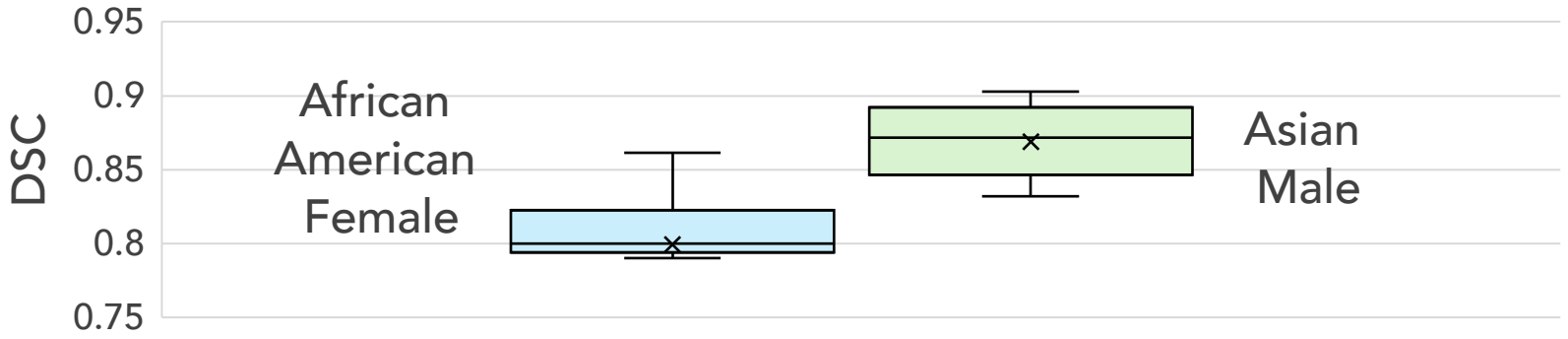
- **Data:** What data we need to collect?
- **Annotations:** How to annotate the data?
- **Justifications:** How to validate the usefulness of the dataset?


A
40000
20000
4000
3200
2400
1600
800
0
number of patients
CTpred-Sunitinib
Pancreatic-CT-CBCT-SEG
CPTAC-PDA
PCL
MSD Pancreas
PANORAMA

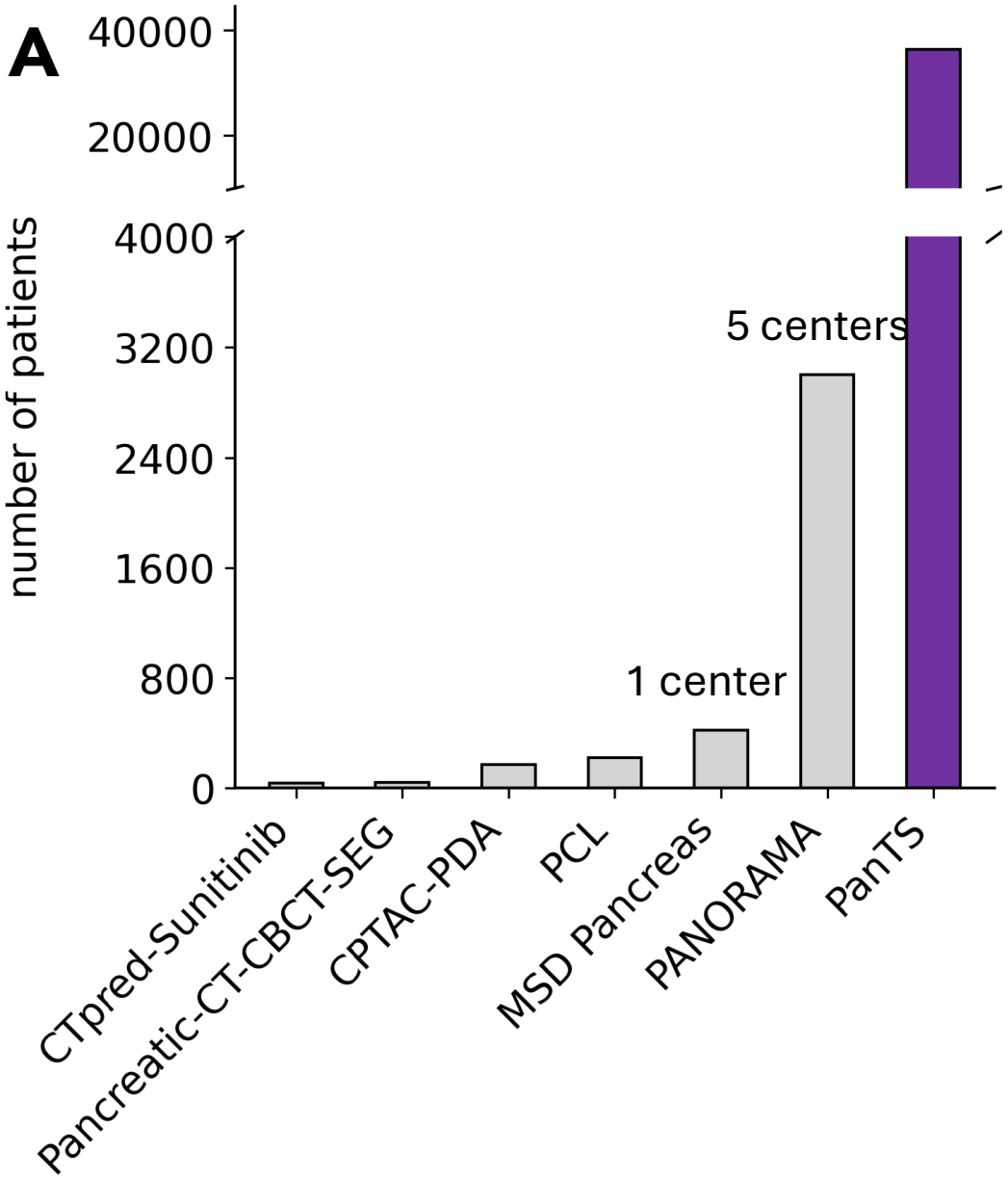

A
40000
20000
4000
3200
2400
1600
800
0
number of patients
CTpred-Sunitinib
Pancreatic-CT-CBCT-SEG
CPTAC-PDA
PCL
MSD Pancreas
PANORAMA
PanTS

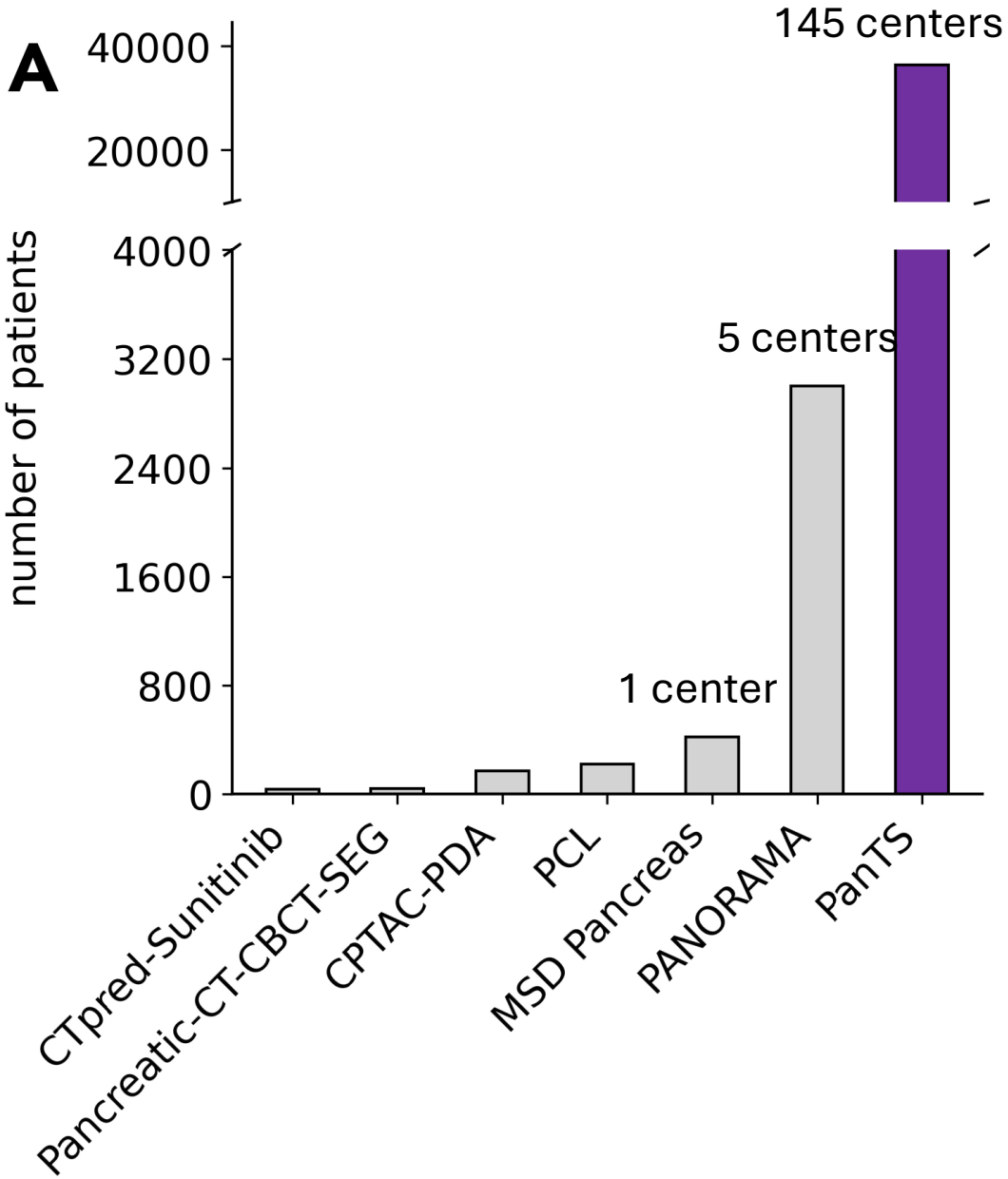
Evaluating Segmentation Architectures

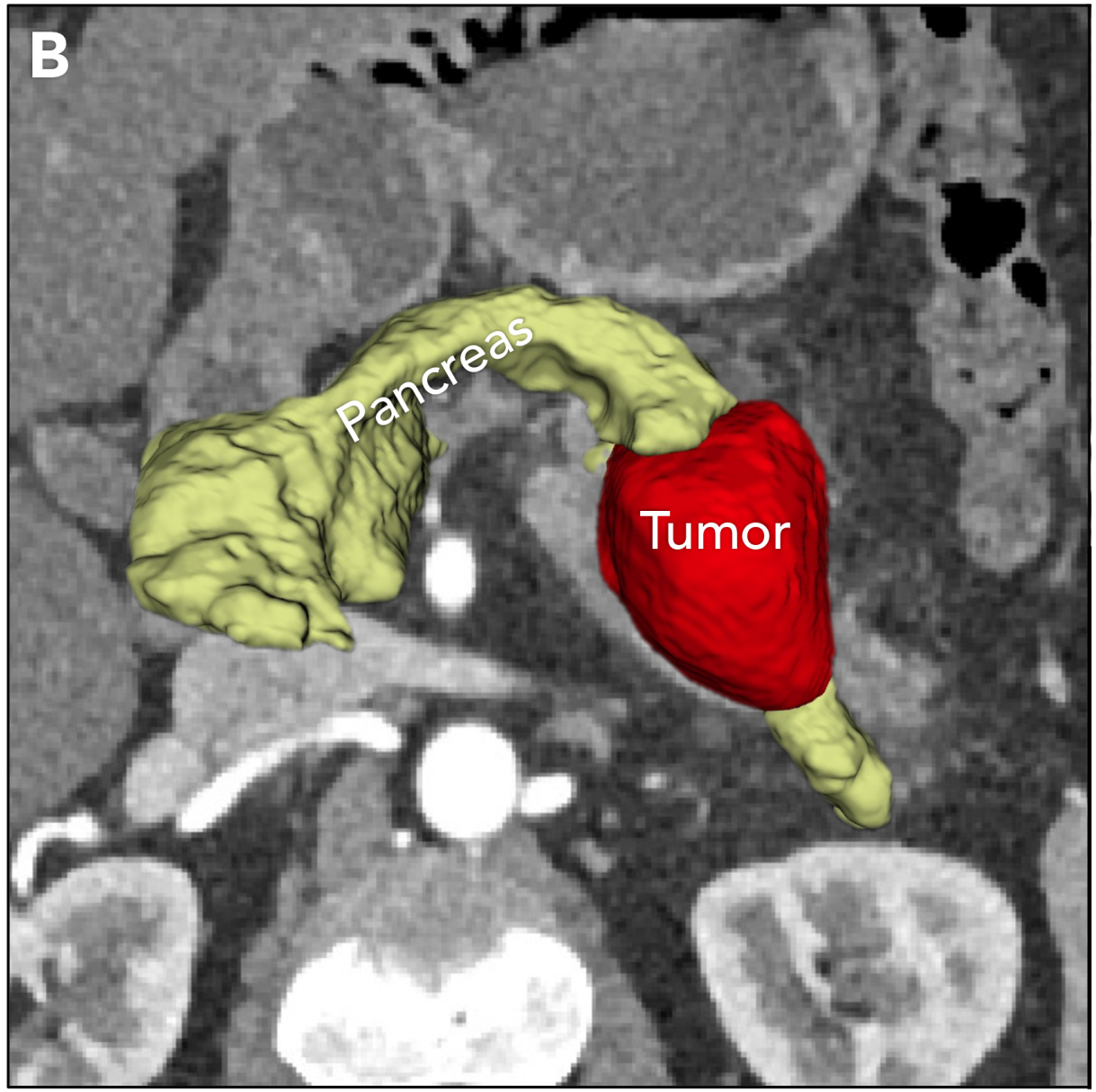
- It is critical to test CT scans from other hospitals, as they may use different scanners and imaging protocols, and patient demographics (e.g., race, gender, age) can vary even within the same hospital (Bassi et al., NeurIPS 2024).




GitHub.com/MrGiovanni/Touchstone


A
40000
20000
4000
3200
2400
1600
800
0
number of patients
5 centers
1 center
CTpred-Sunitinib
Pancreatic-CT-CBCT-SEG
CPTAC-PDA
PCL
MSD Pancreas
PANORAMA
PanTS


A
40000
20000
4000
3200
2400
1600
800
0
number of patients
145 centers
5 centers
1 center
CTpred-Sunitinib
Pancreatic-CT-CBCT-SEG
CPTAC-PDA
PCL
MSD Pancreas
PANORAMA
PanTS


B
Pancreas
Tumor

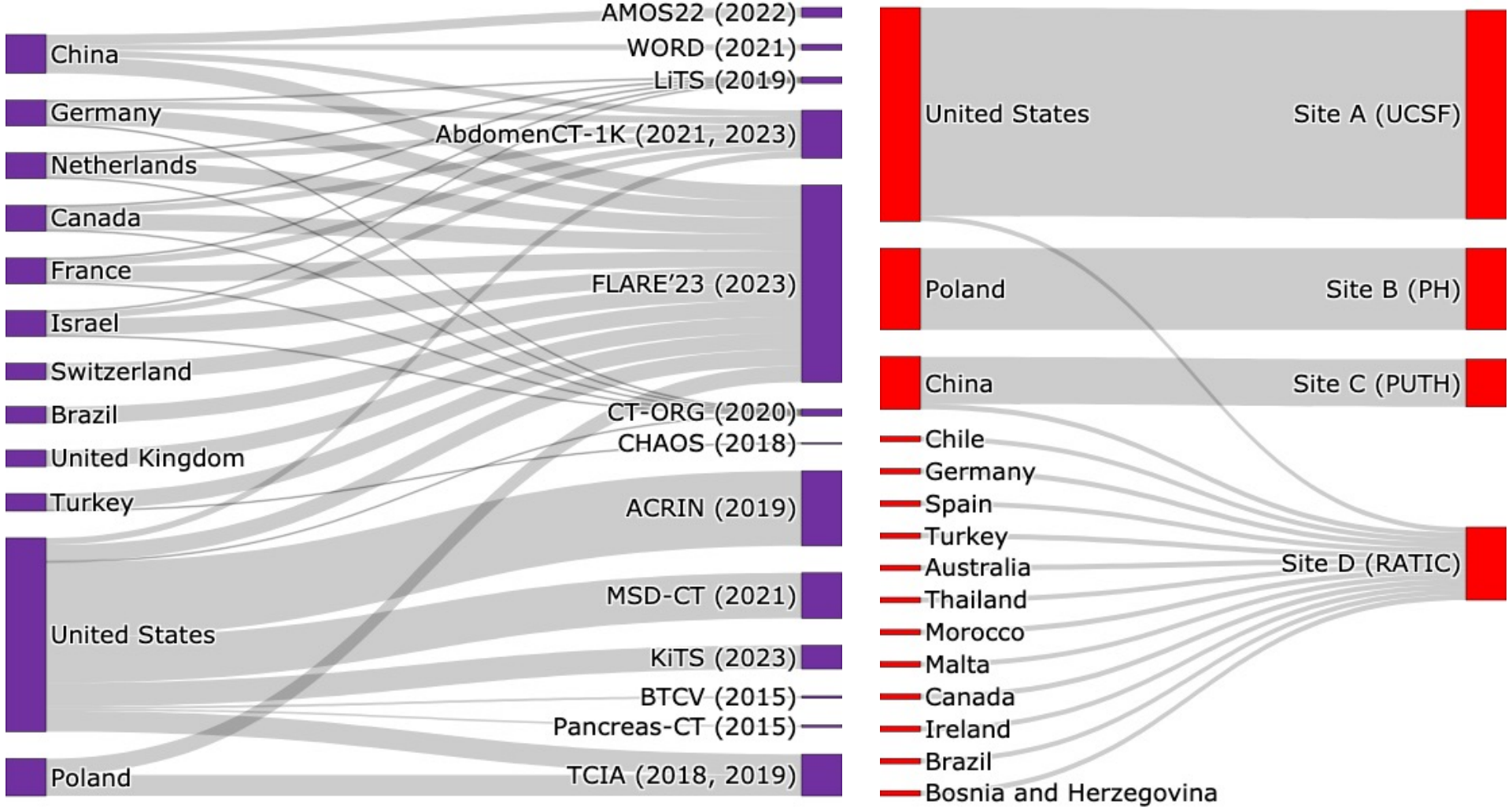
Training set 9,901 CT scans + 28 classes masks + metadata

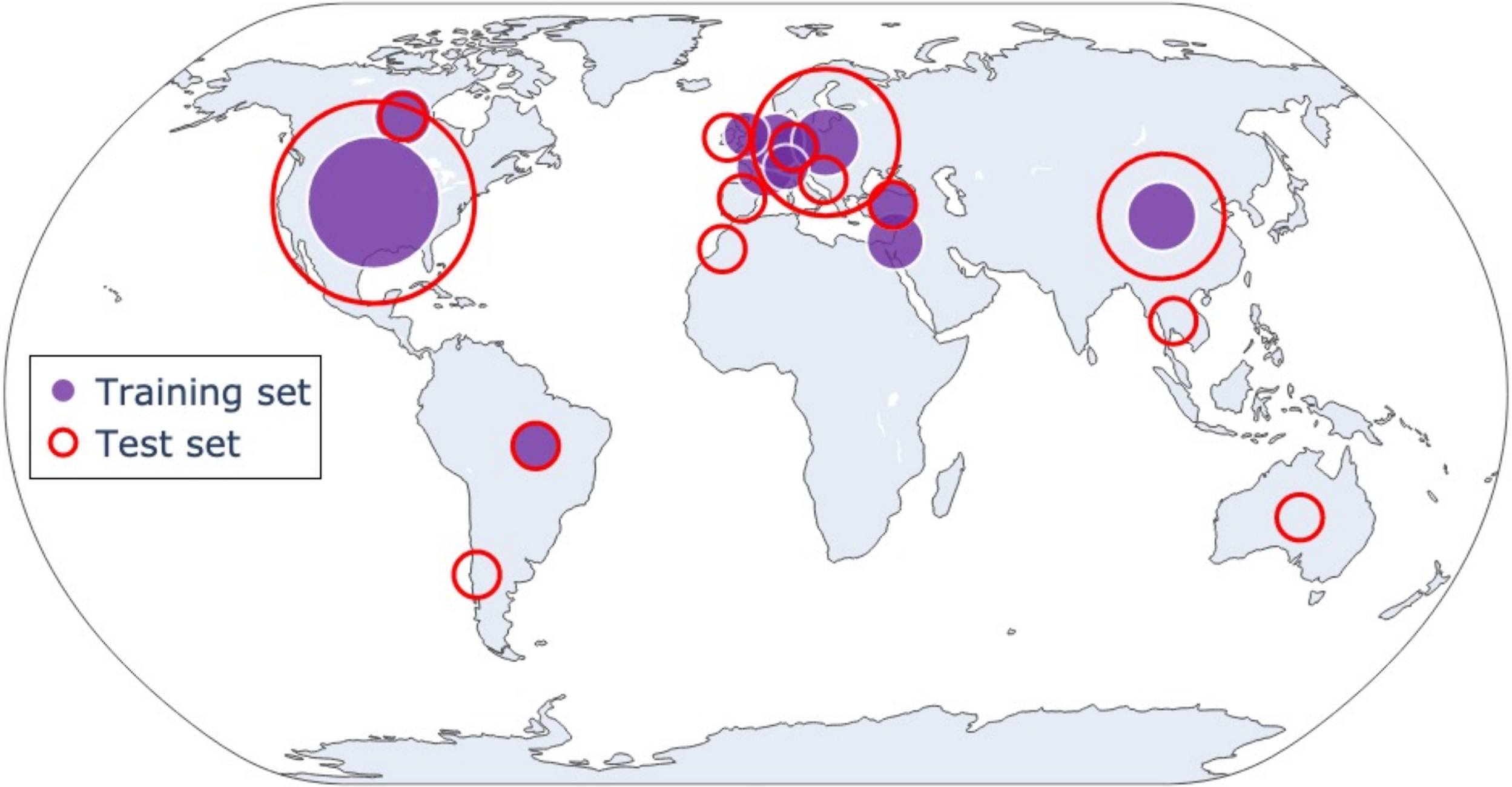
New structured reports are now publicly available!

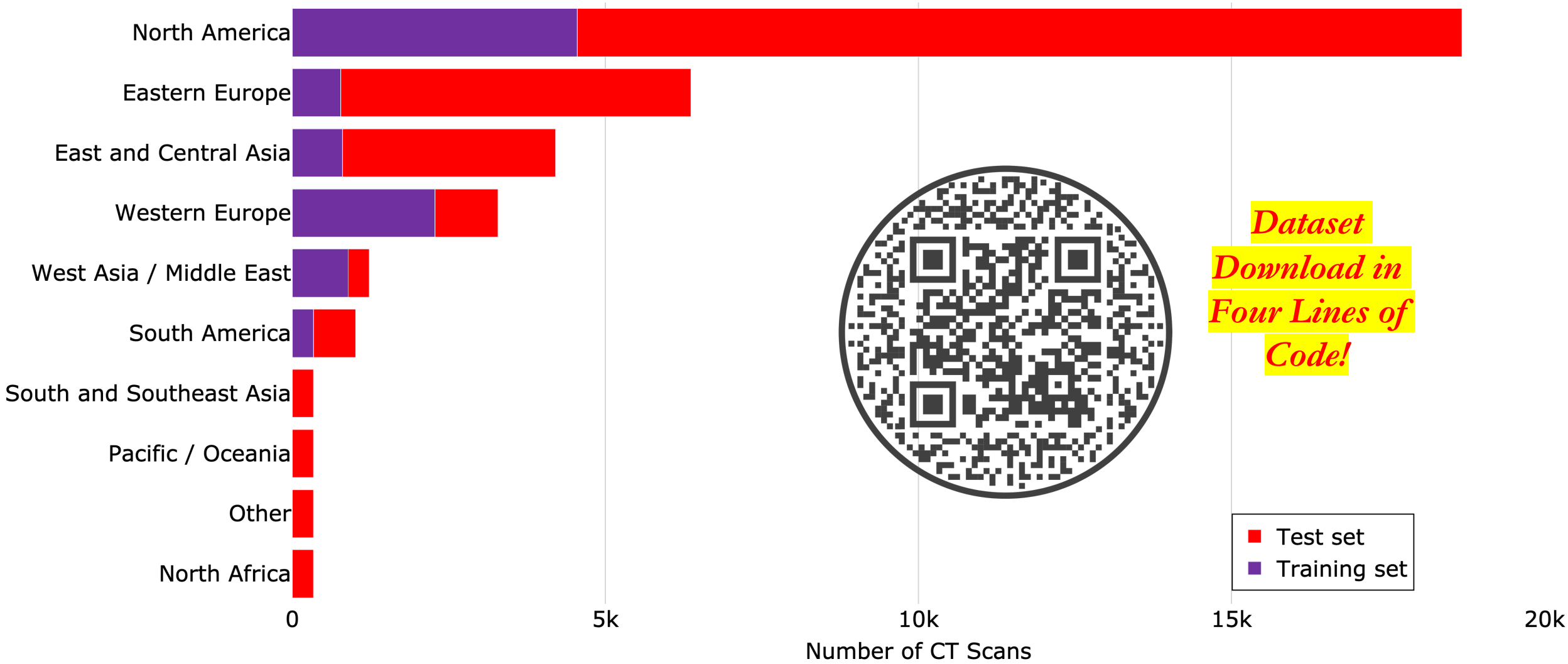
Variable	Training set (n = 9,901)	Test set (n = 26,489)	p-value
Age, mean (SD)	60.6 (13.0)	58.5 (17.0)	1.78×10^{-7}
Sex			7.87×10^{-27}
Female, no. (%)	2,358 (23.8)	13,090 (49.4)	
Male, no. (%)	2,923 (29.5)	11,714 (44.2)	
Unknown, no. (%)	4,620 (46.7)	1,685 (6.4)	
In-plane spacing, mm (IQR)	0.81 (0.74, 0.98)	0.75 (0.70, 0.83)	0.00
Slice thickness, mm (IQR)	1.25 (0.80, 2.50)	1.25 (1.25, 2.50)	5.13×10^{-169}
Contrast phase			0.00
Non-contrast, no. (%)	4,488 (45.3)	3,920 (14.8)	
Portal venous, no. (%)	2,895 (29.2)	20,296 (76.6)	
Arterial, no. (%)	2,450 (24.7)	2,273 (8.6)	
Delayed, no. (%)	68 (0.8)	0 (0.0)	
Pancreatic tumor			
Yes, no. (%)	1,077 (10.9)	2,829 (10.7)	
No, no. (%)	8,824 (89.1)	23,660 (89.3)	
Dilated duct			
Yes, no. (%)	3,387 (34.2)	11,180 (42.2)	
No, no. (%)	6,514 (65.8)	15,309 (57.8)	
Tumors per positive CT, no. (IQR)	1.00 (1.00, 1.00)	1.00 (1.00, 2.00)	1.48×10^{-65}
Tumor volume, mm^3 (IQR)	4,749 (1,658, 11,479)	12,667 (3,347, 32,238)	4.07×10^{-53}
Tumor HU value, mean (SD)	57.3 (30.7)	78.2 (59.0)	1.54×10^{-10}
Pancreas volume, mm^3 (IQR)	74,669 (52,806, 95,892)	74,480 (56,676, 92,892)	8.75×10^{-2}
Pancreas HU value, mean (SD)	56.8 (36.4)	85.6 (54.8)	0.00

*Having **normal** scans helps reduce false positives from oversensitive AI models*

Test set UCSF (13,458 CT scans) & PH (5,259 CT scans) & PUTH (3,066 CT scans) & RATIC (4,706 CT scans)




Training set
Test set


North America
Eastern Europe
East and Central Asia
Western Europe
West Asia / Middle East
South America
South and Southeast Asia
Pacific / Oceania
Other
North Africa
0
5k
10k
15k
20k
Number of CT Scans
Test set
Training set
Dataset Download in Four Lines of Code!

Challenges (2/3)

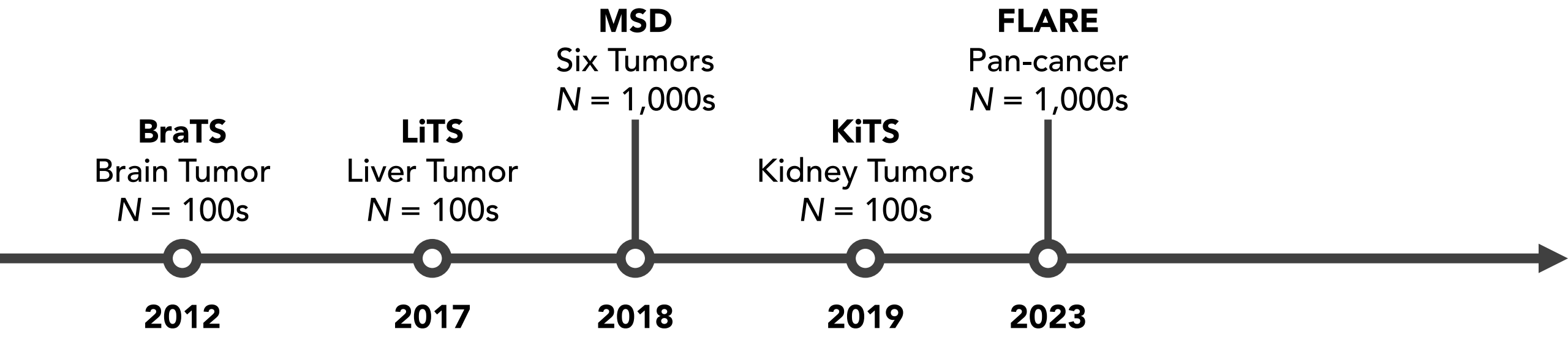
- **Data:** What data we need to collect?
- **Annotations:** How to annotate the data?
- **Justifications:** How to validate the usefulness of the dataset?

The JHU Dataset (Private ☹)

- **>5,000** voxel-wise annotated CT scans, requiring **25** person years.
- It trains high-performance AI algorithms (Xia et al., medRxiv, 2022)
- Sensitivity = 97%, Specificity = 99%
- Exceed radiologist performance.
- Generalizable to multiple centers.
- *But it is private!*
- *$6M, five-year annotations*

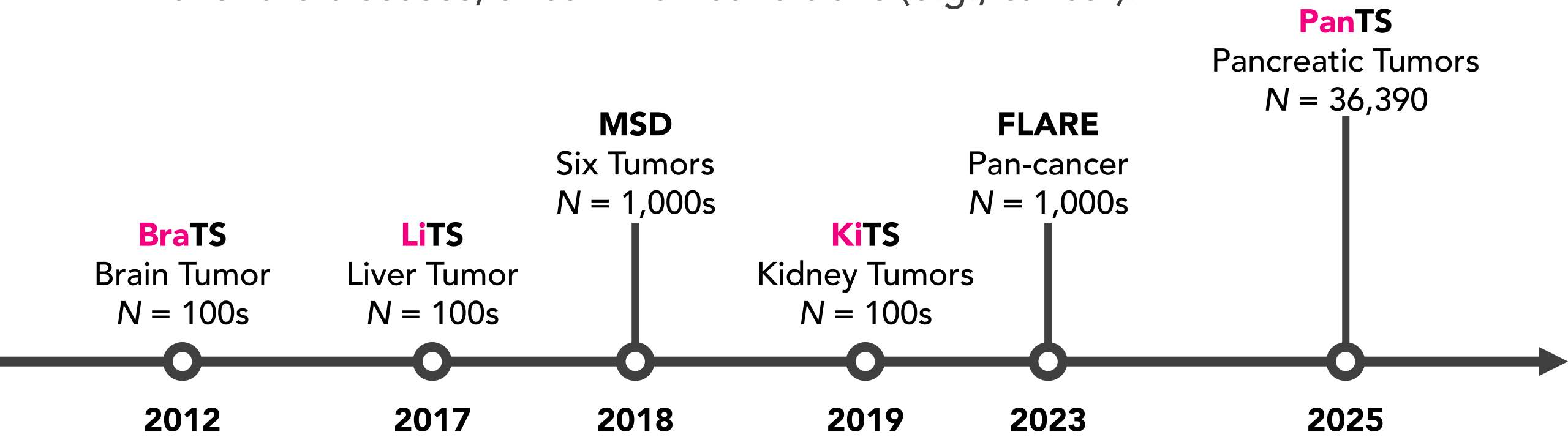
Annotated Tumor Datasets Should Be Open to More Researchers

- There's a huge data gap in medical AI right now, particularly when you have rare diseases, uncommon conditions (e.g., cancer).

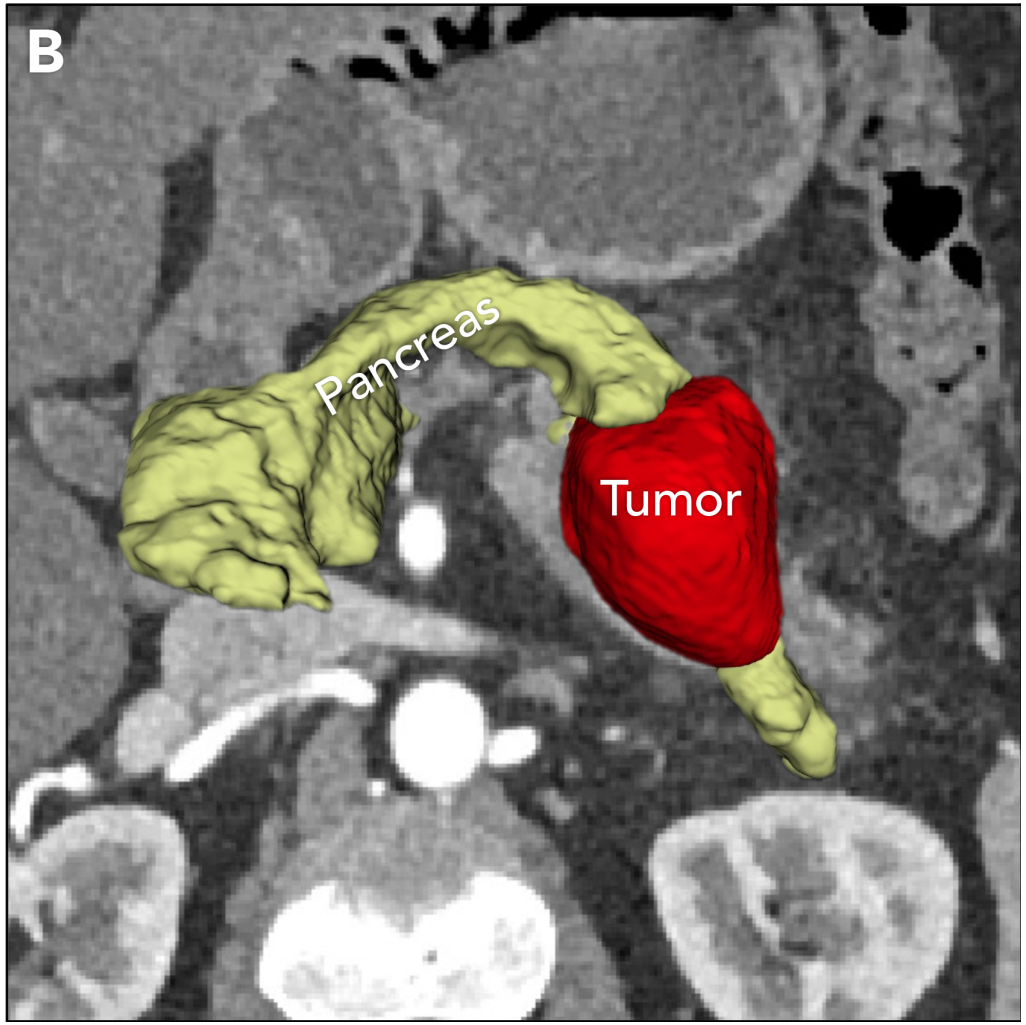


Annotated Tumor Datasets Should Be Open to More Researchers

- There's a huge data gap in medical AI right now, particularly when you have rare diseases, uncommon conditions (e.g., cancer).



Pancreatic Tumor Annotations

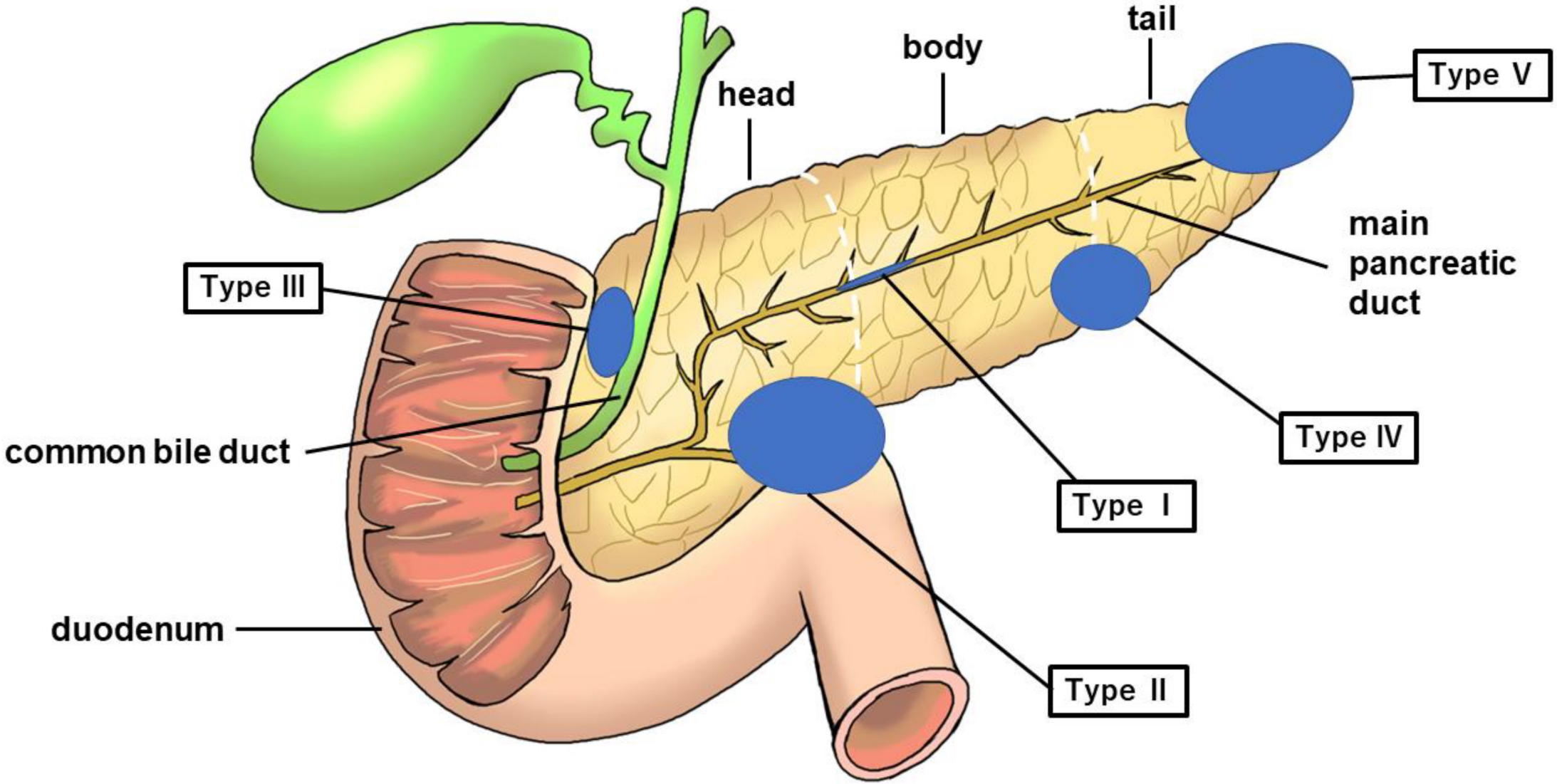



23

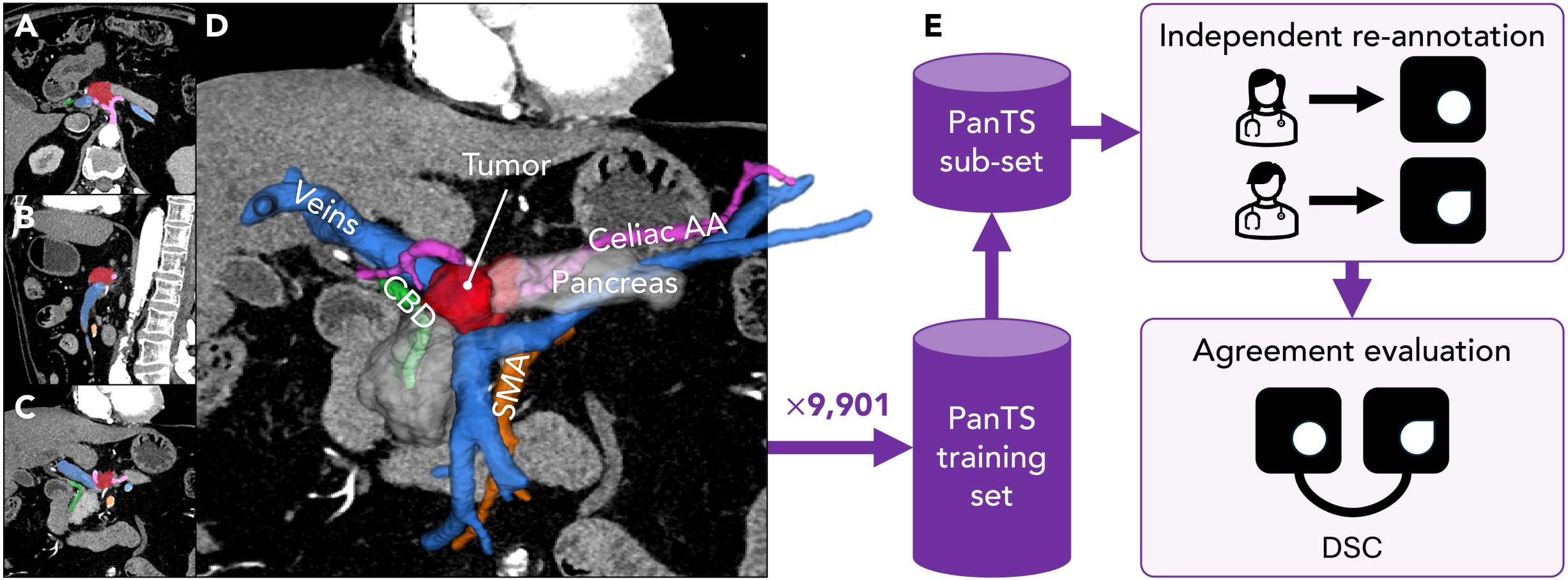
Pancreatic Tumor Annotations

No.	Annotator ID	Experience (yr)	CT read / year	No.	Annotator ID	Experience (yr)	CT read / year
1	Specialist 1 (S1)	24	12,000	2	Specialist 2 (S2)	22	12,000
3	Specialist 3 (S3)	35	8,000	4	Specialist 4 (S4)	30	8,000
5	Specialist 5 (S5)	28	9,000	6	Specialist 6 (S6)	19	13,000
7	Specialist 7 (S7)	23	11,000	8	General 1 (G1)	12	18,000
9	General 2 (G2)	8	18,000	10	General 3 (G3)	9	18,000
11	General 4 (G4)	10	18,000	12	General 5 (G5)	8	18,000
13	General 6 (G6)	13	18,000	14	General 7 (G7)	11	18,000
15	General 8 (G8)	10	18,000	16	General 9 (G9)	10	18,000
17	General 10 (G10)	13	18,000	18	General 11 (G11)	10	18,000
19	Resident 1 (R1)	5	16,000	20	Resident 2 (R2)	3	16,000
21	Resident 3 (R3)	4	16,000	22	Resident 4 (R4)	5	16,000
23	Resident 5 (R5)	5	16,000				

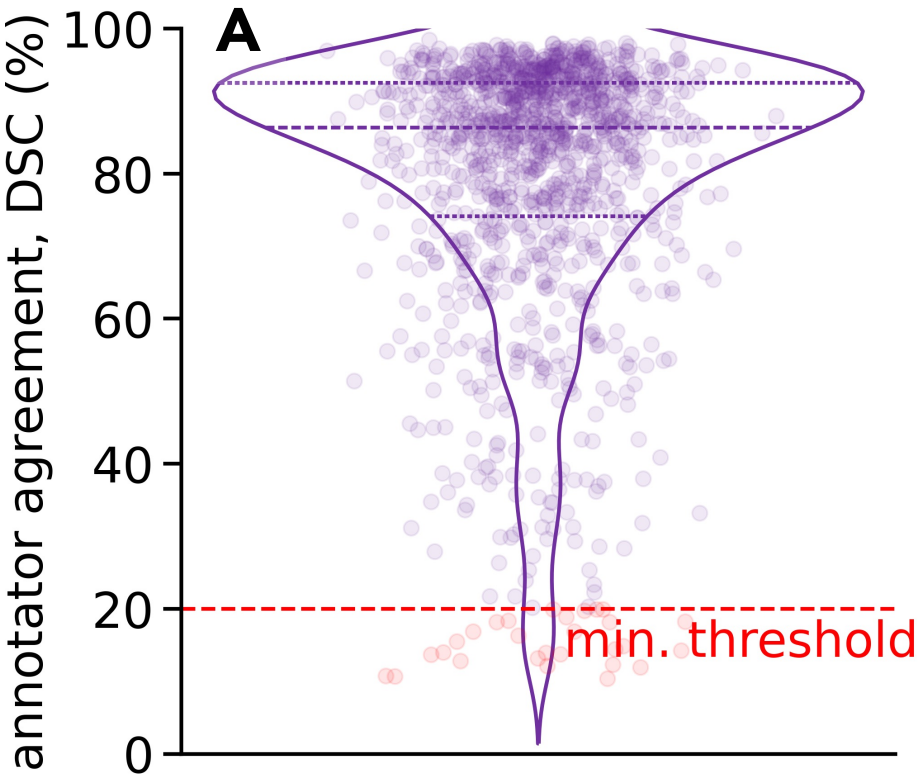
Pancreatic Tumor Annotations

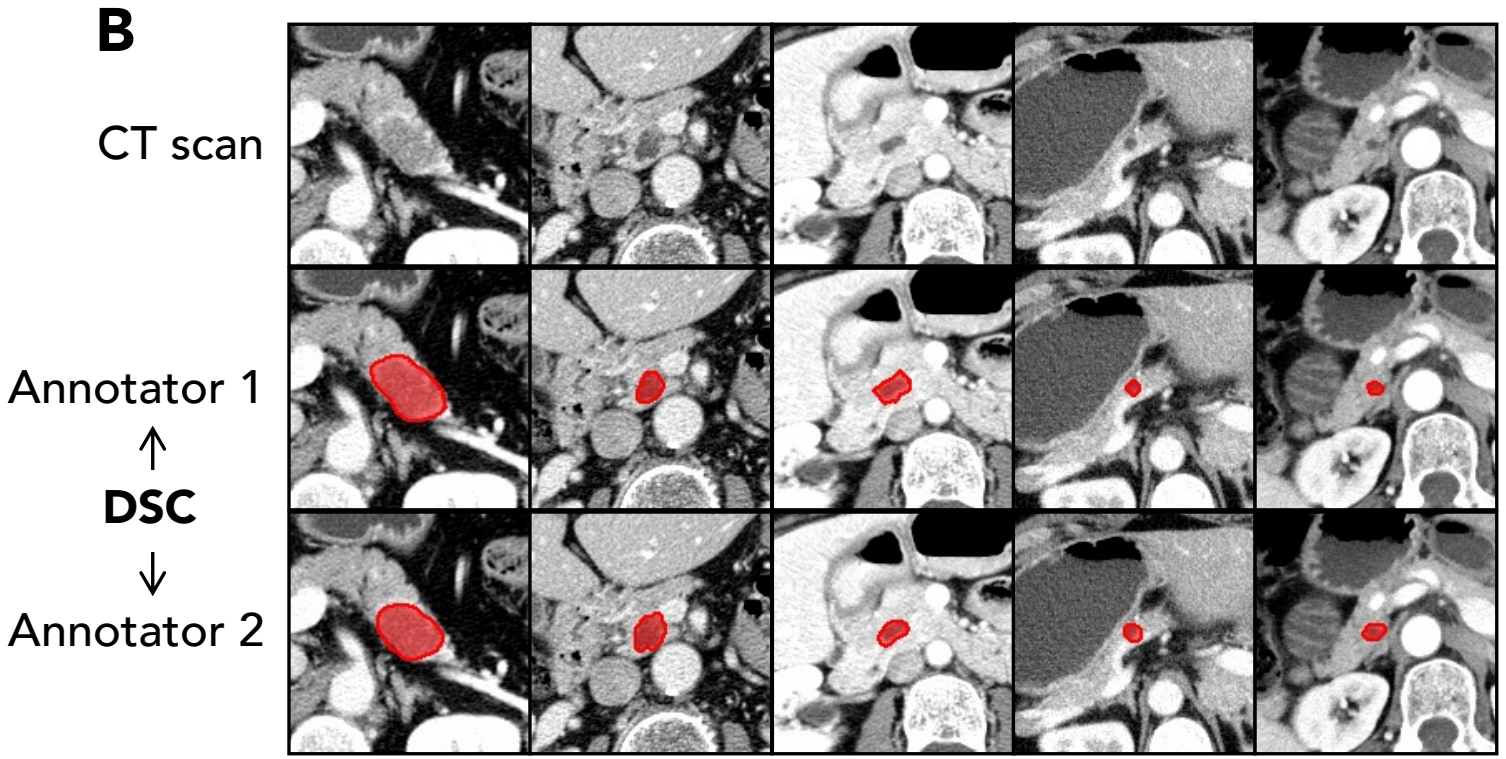


Pancreatic Tumor Annotations

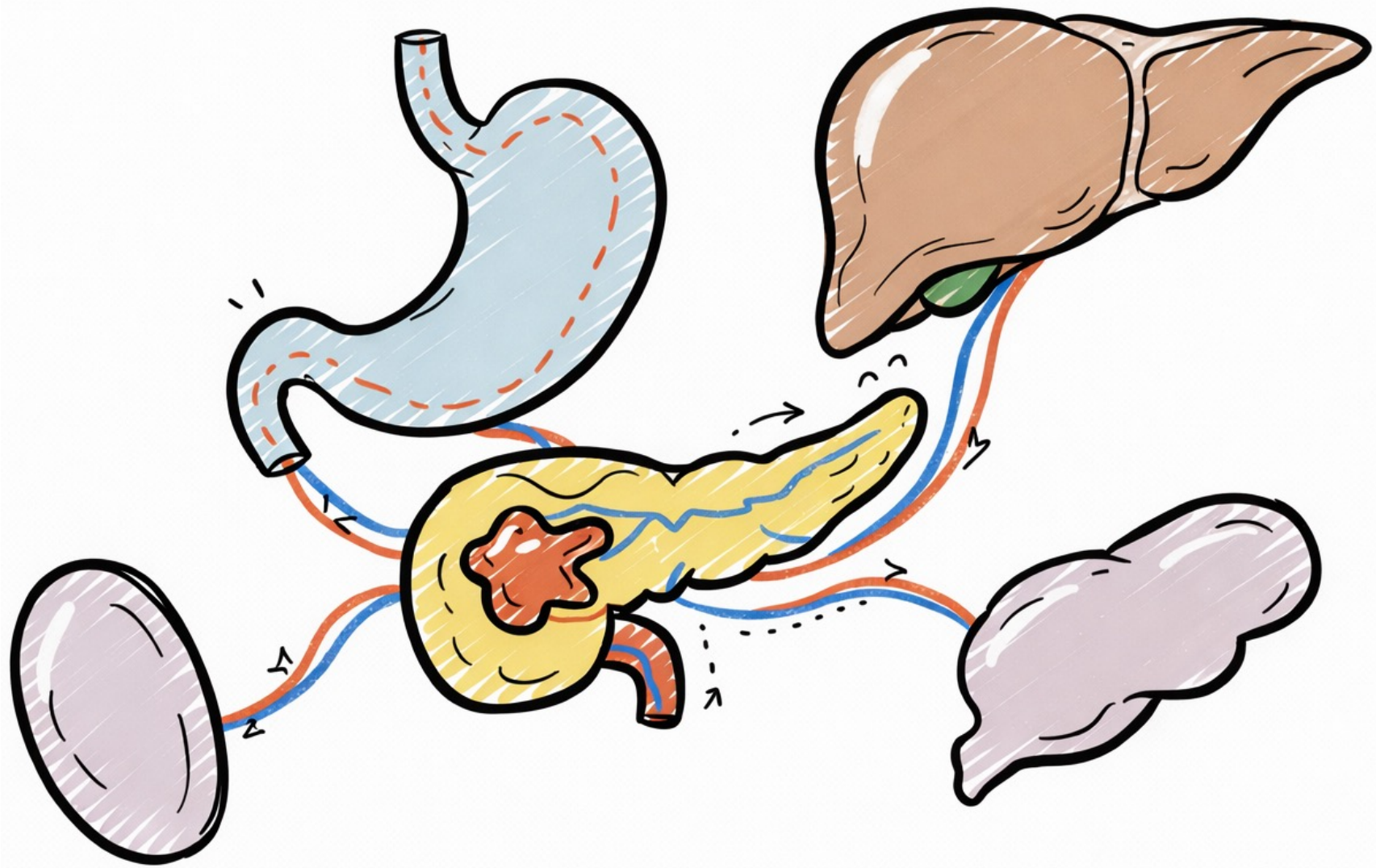


Pancreatic Tumor Annotations

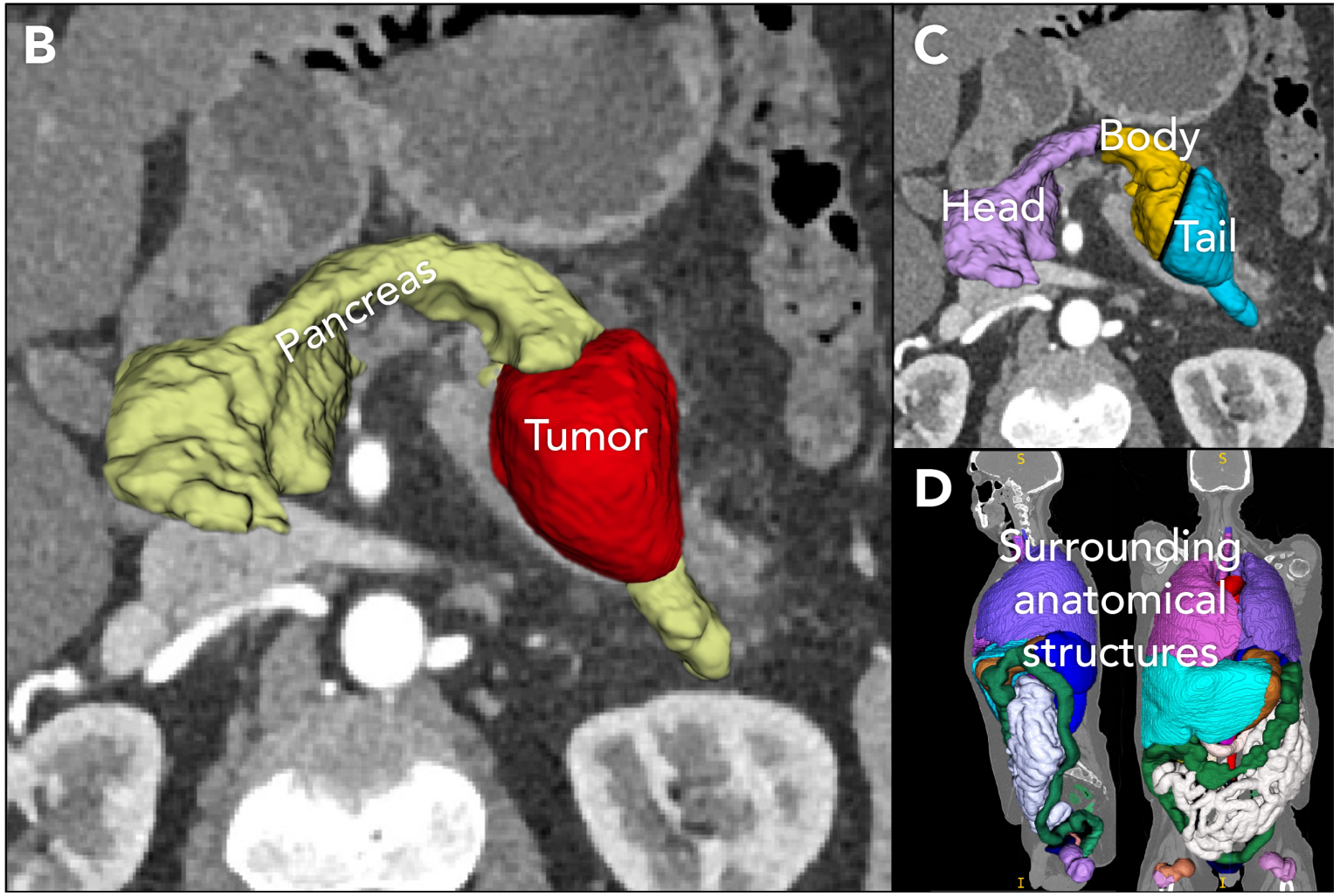




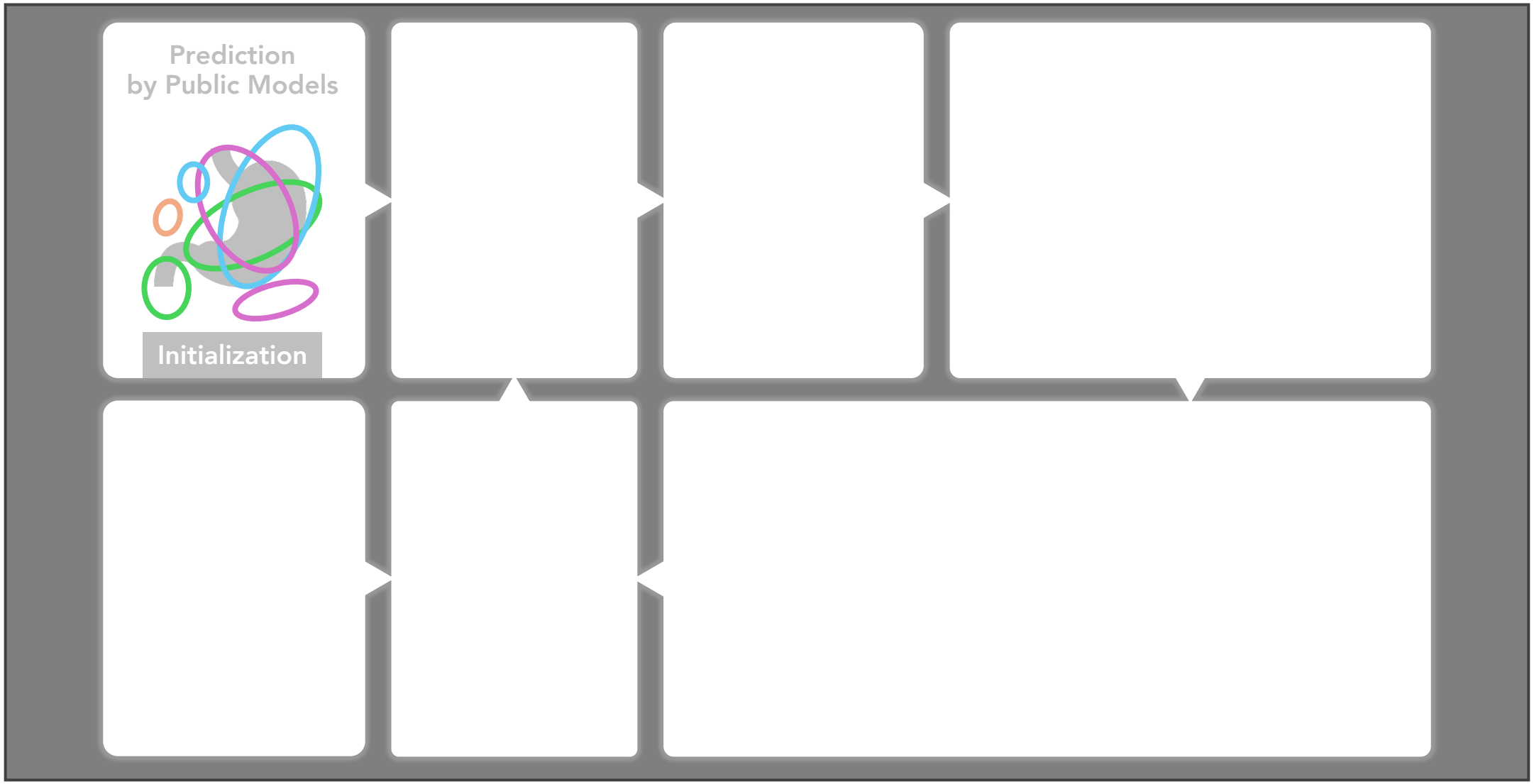
Surrounding Structures Annotations



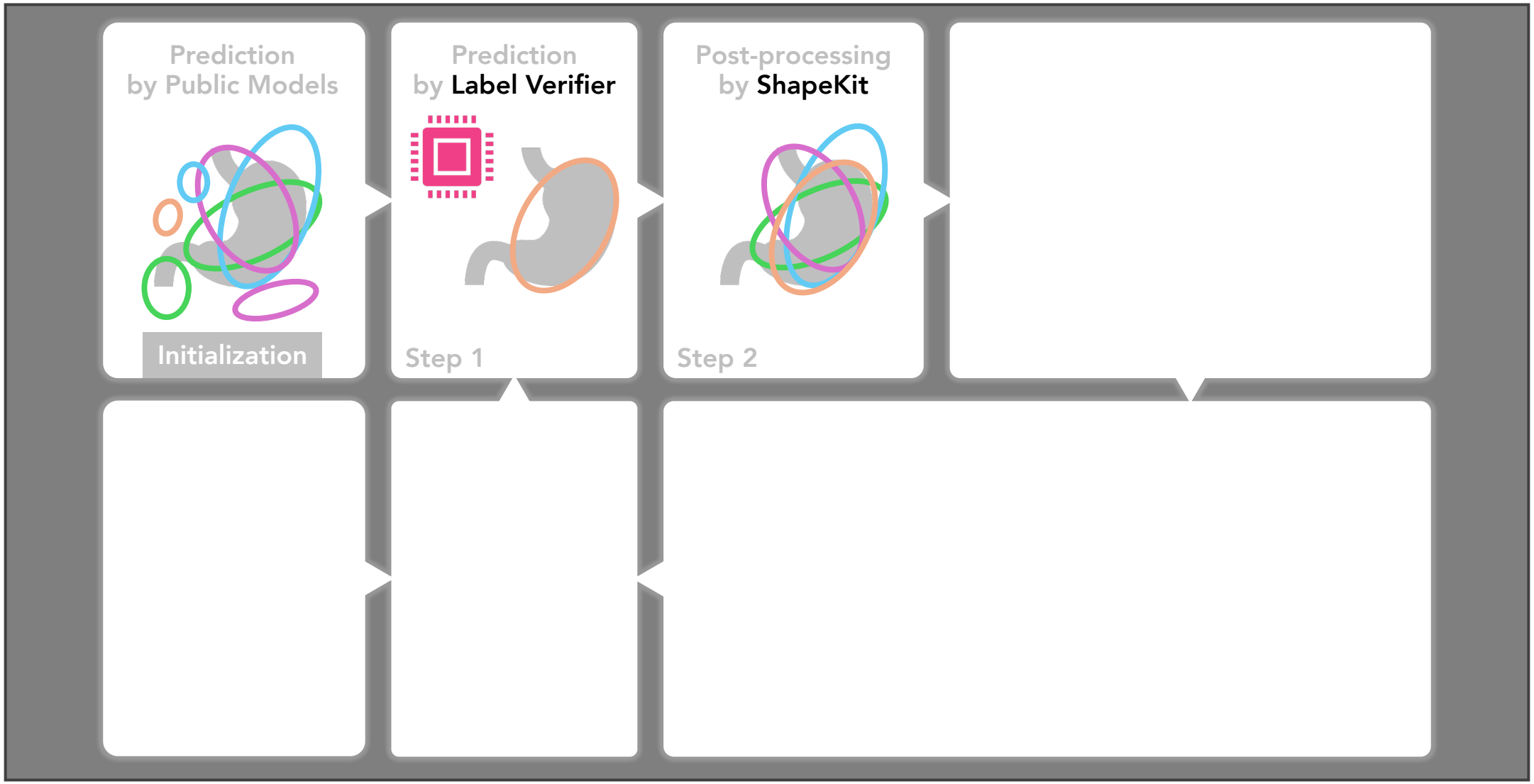
Surrounding Structures Annotations



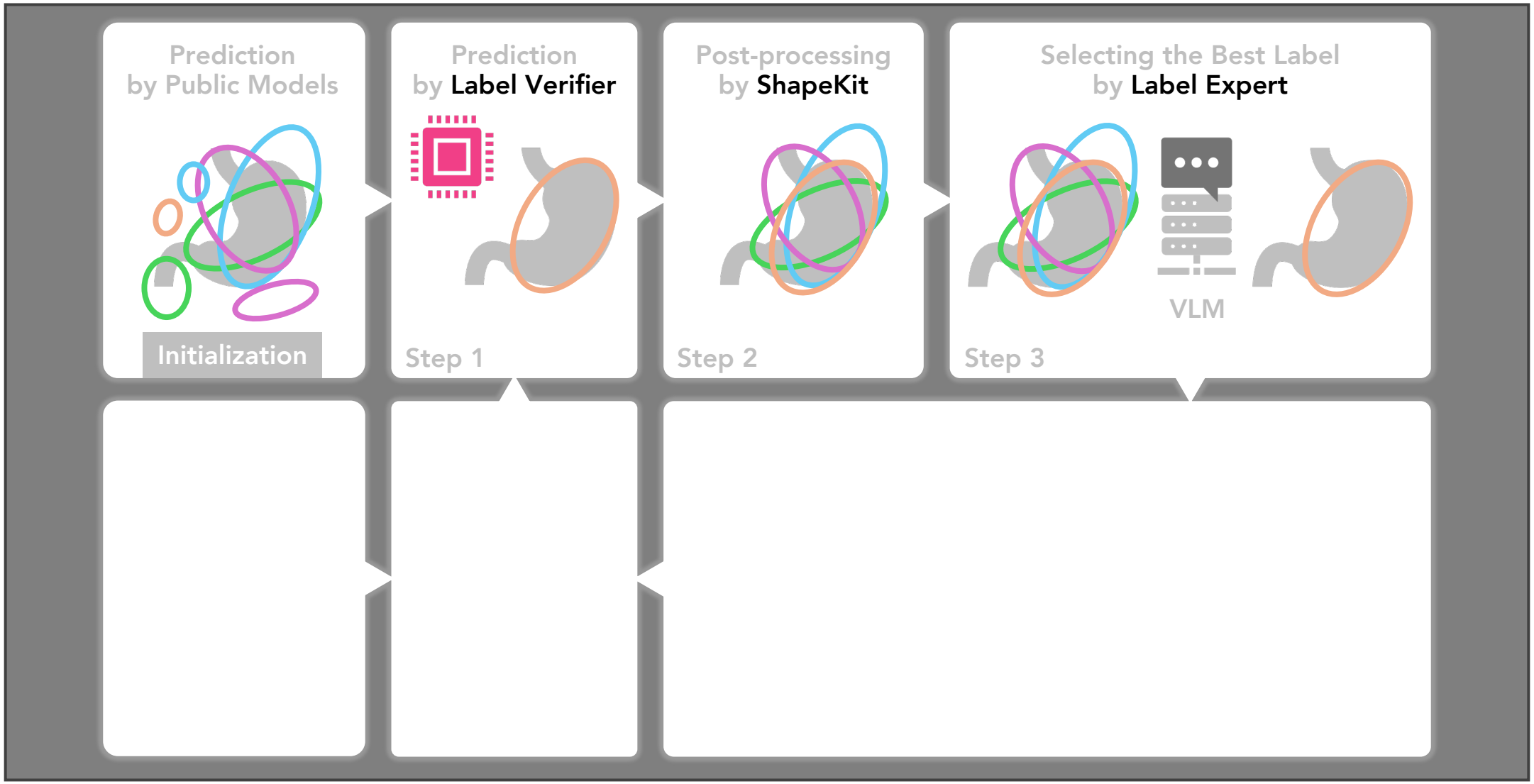
Surrounding Structures Annotations



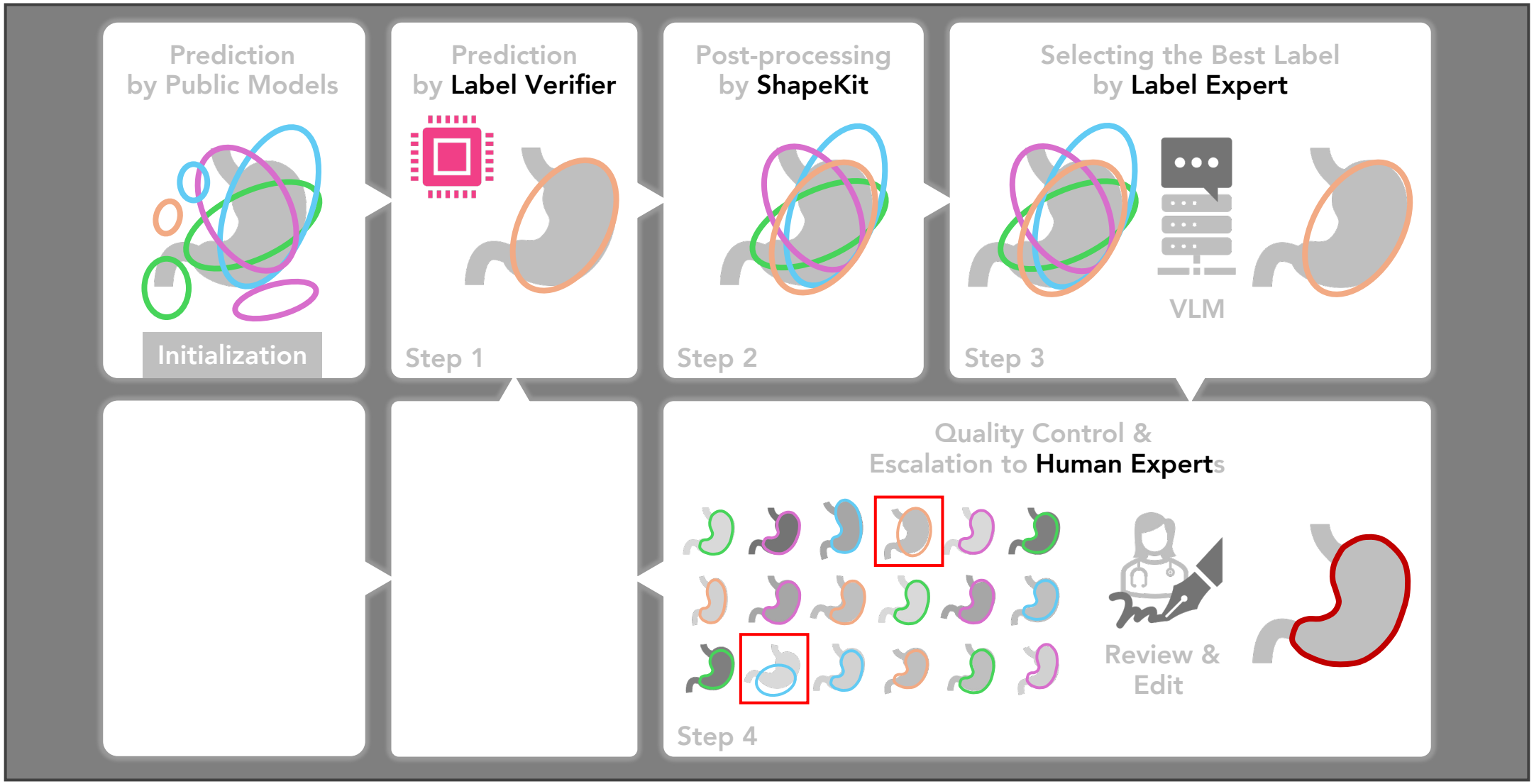
Surrounding Structures Annotations



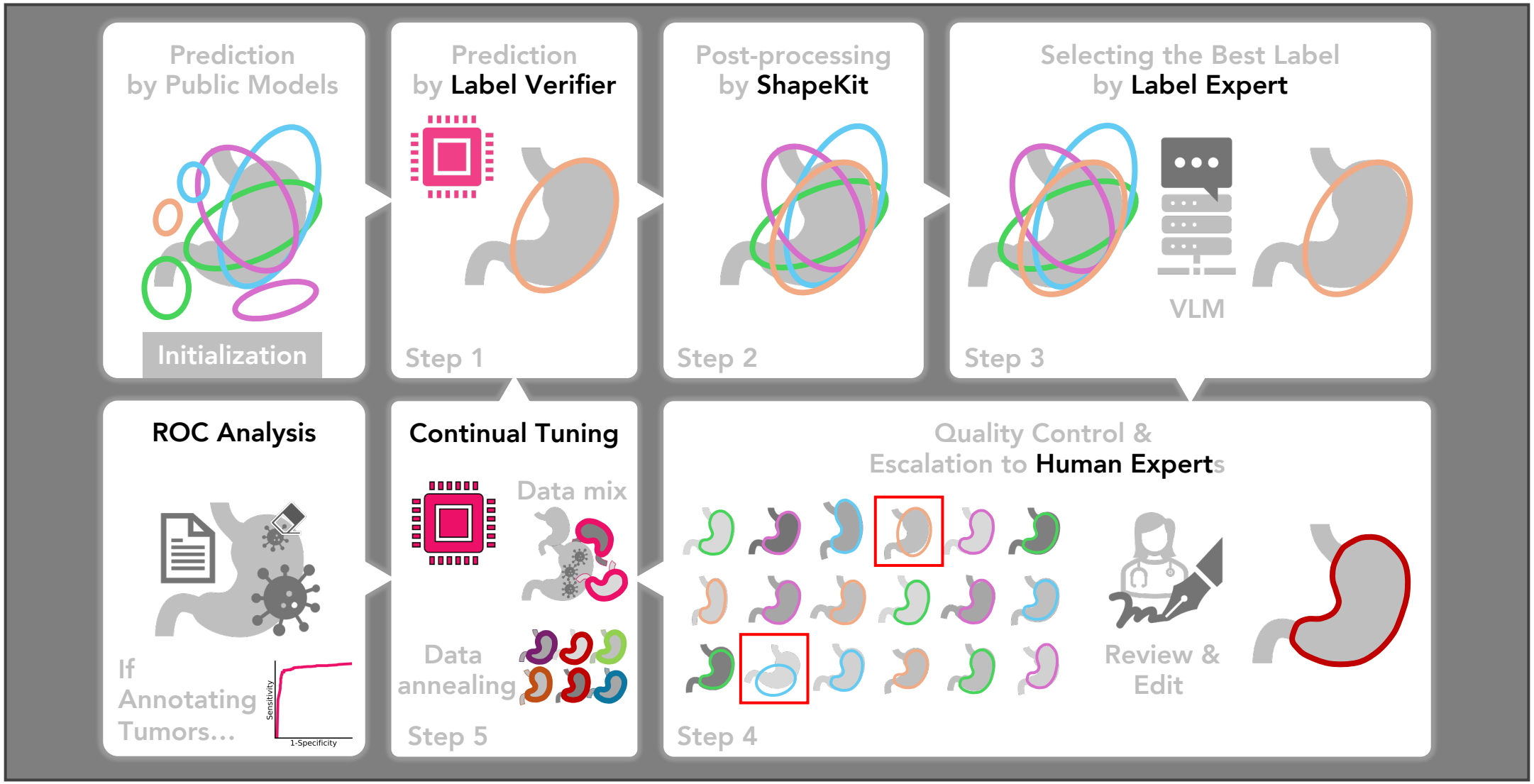
Surrounding Structures Annotations



Surrounding Structures Annotations



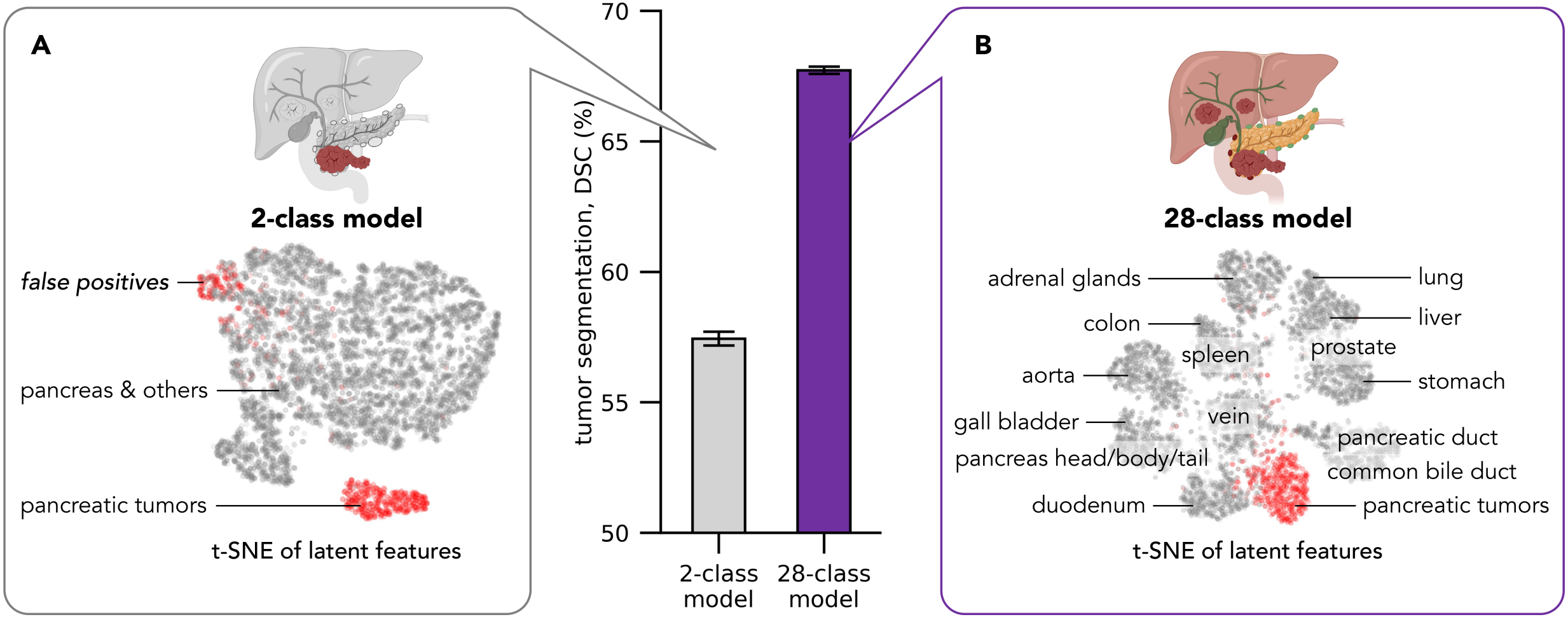
Surrounding Structures Annotations



Challenges (3/3)

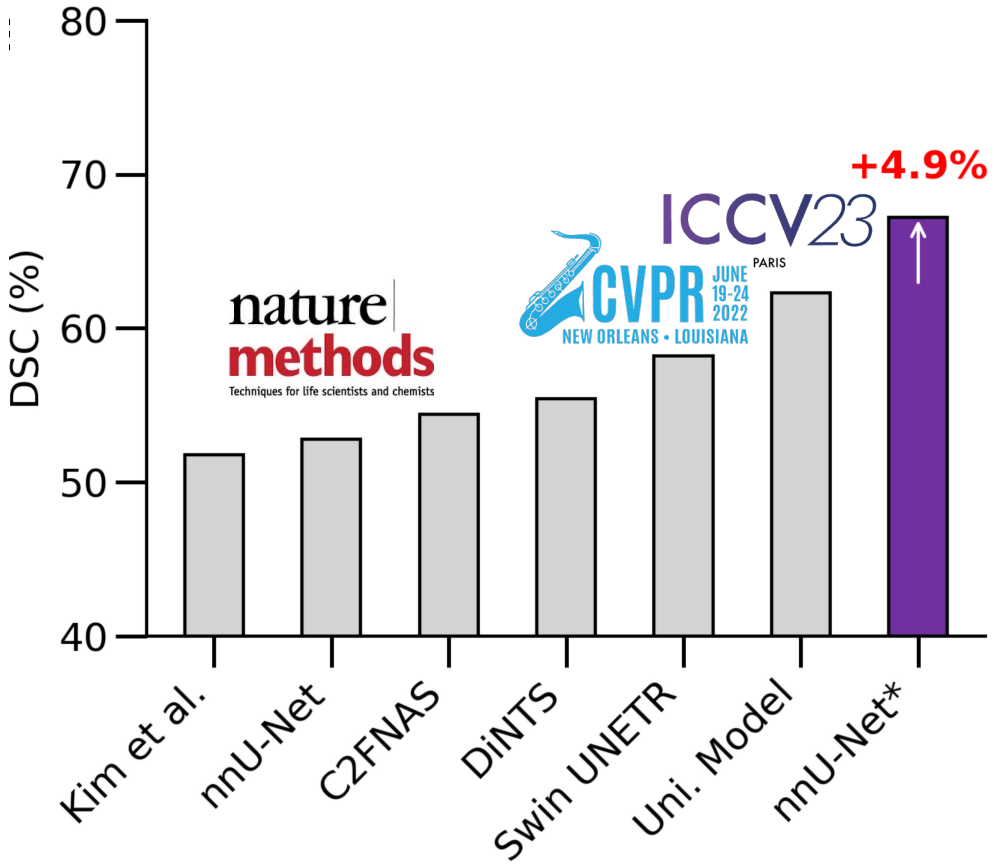
- **Data:** What data we need to collect?
- **Annotations:** How to annotate the data?
- **Justifications:** How to validate the usefulness of the dataset?

Justification: Anatomical Structures



Justification: Large-Scale Tumor Datasets

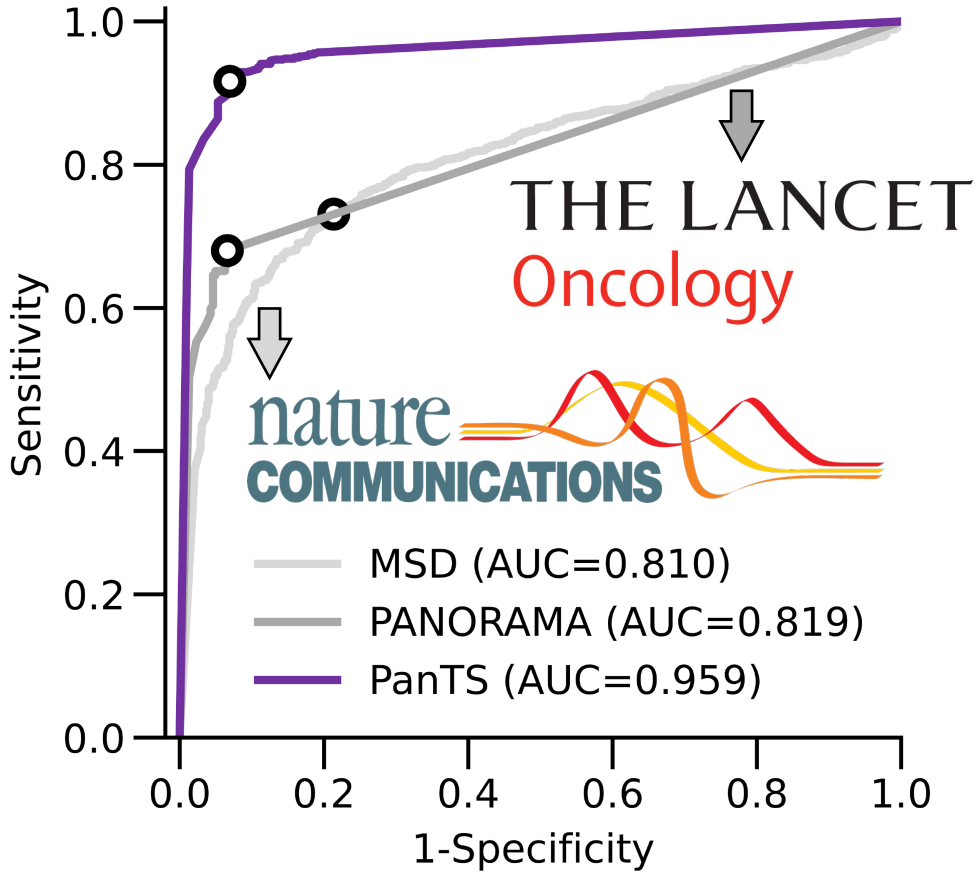
R1. Benchmark on open leaderboard, Medical Segmentation Decathlon (MSD)



R1. AI trained on our PanTS vs. AI trained on publicly available datasets. The performance is tested on the official MSD-Pancreas test set (third-party evaluation).

Justification: Large-Scale Tumor Datasets

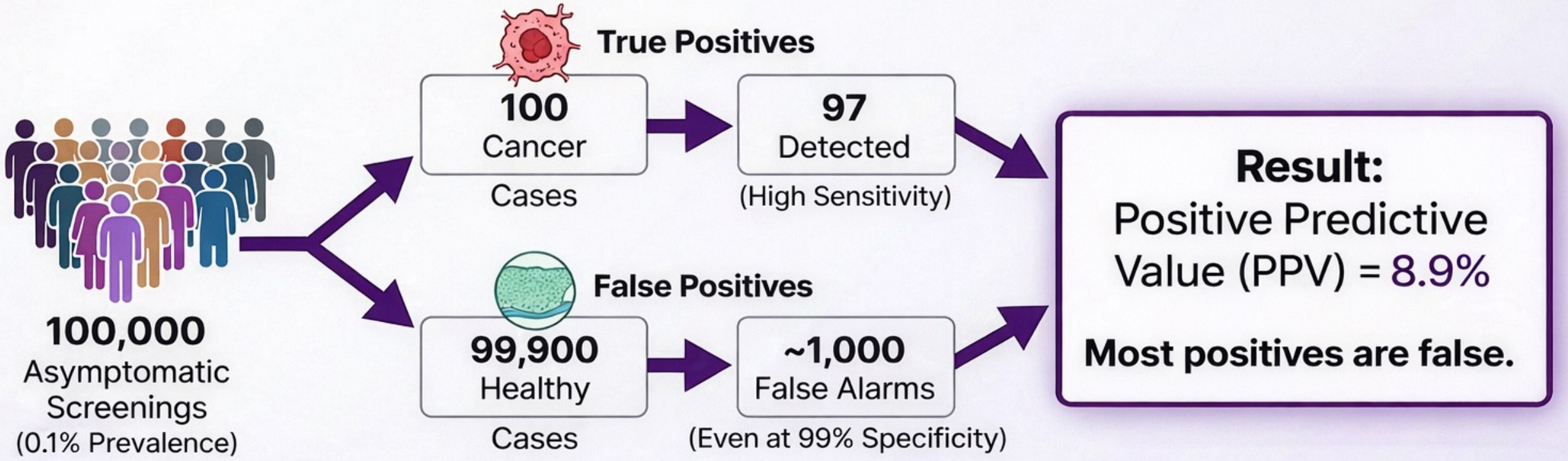
R2. Benchmark on the PanTS-test leaderboard



R1. AI trained on our PanTS vs. AI trained on publicly available datasets. The performance is tested on the official MSD-Pancreas test set (third-party evaluation).
R2. The ROC curve of AI trained on different scale of datasets, i.e., MSD-Pancreas (n=281), PANORAMA (n=2,238), and our PanTS dataset (n=9,901).
Observation: the larger training set, the higher voxel-wise annotation quality, the better pancreatic tumor detection performance on out-of-distribution test sets.
PanTS enables tasks of semantic segmentation, vision-language models, metadata prediction, and many more.

Solving the Screening Paradox

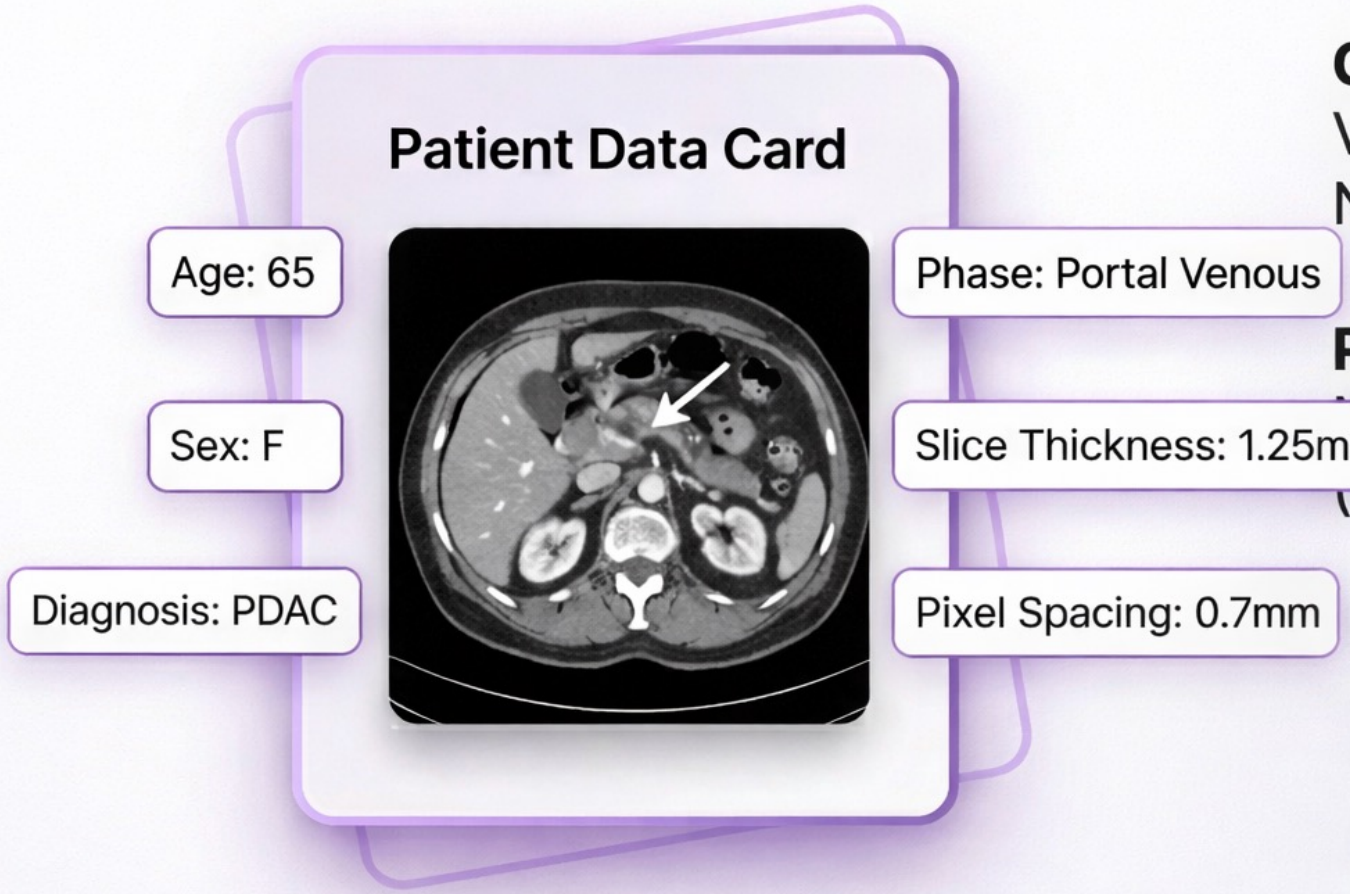
Why ‘Normal’ scans are critical for clinical utility.



The PanTS Solution

Previous datasets (MSD, KiTS) have **0** normal scans. PanTS includes **89%** normal scans, enabling true specificity estimation and reduction of false positives.

Beyond Pixels: Rich Metadata for Risk Stratification



Clinical Nuance: Includes Venous, Arterial, Delayed, Non-contrast phases.

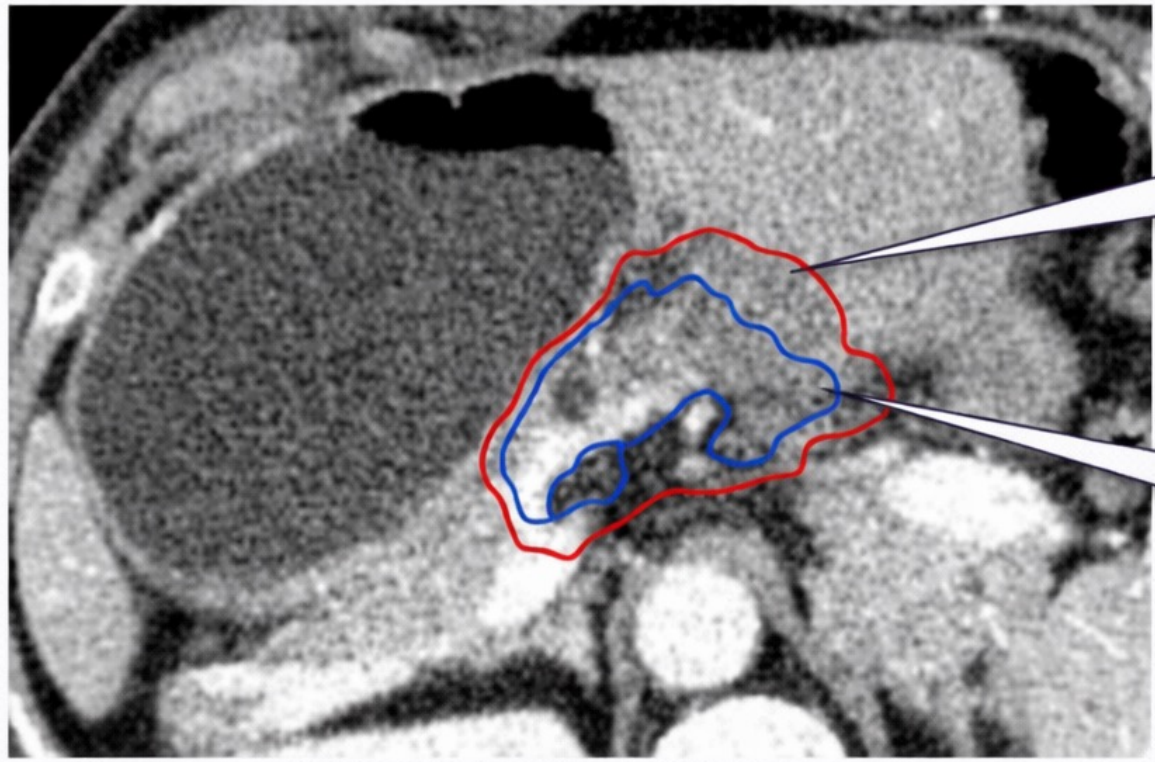
Pathology Coverage: PDAC, Neuroendocrine tumors (PNETs), Cystic neoplasms.

Impact: Enables opportunist screening algorithms that incorporate patient demographics.

The Persistent Challenge of Ambiguity

Even with 993k annotations, edge cases remain.

Low Agreement Case Study



False Positives:
Texture irregularities without ductal dilation.

False Negatives:
Exophytic growths or diffuse thinning.

High-quality data reduces uncertainty, but multimodal learning…

```
git clone https://github.com/MrGiovanni/PanTS.git; cd PanTS
bash download_PanTS_data.sh
bash download_PanTS_label.sh
http://www.cs.jhu.edu/~zongwei/dataset/PanTSMini_Label.tar.gz
```

PanTS is a large-scale, multi-institutional dataset, containing **36,390** three-dimensional CT volumes from **145** medical centers, with expert-validated, voxel-wise annotations of over **993,000** anatomical structures, including

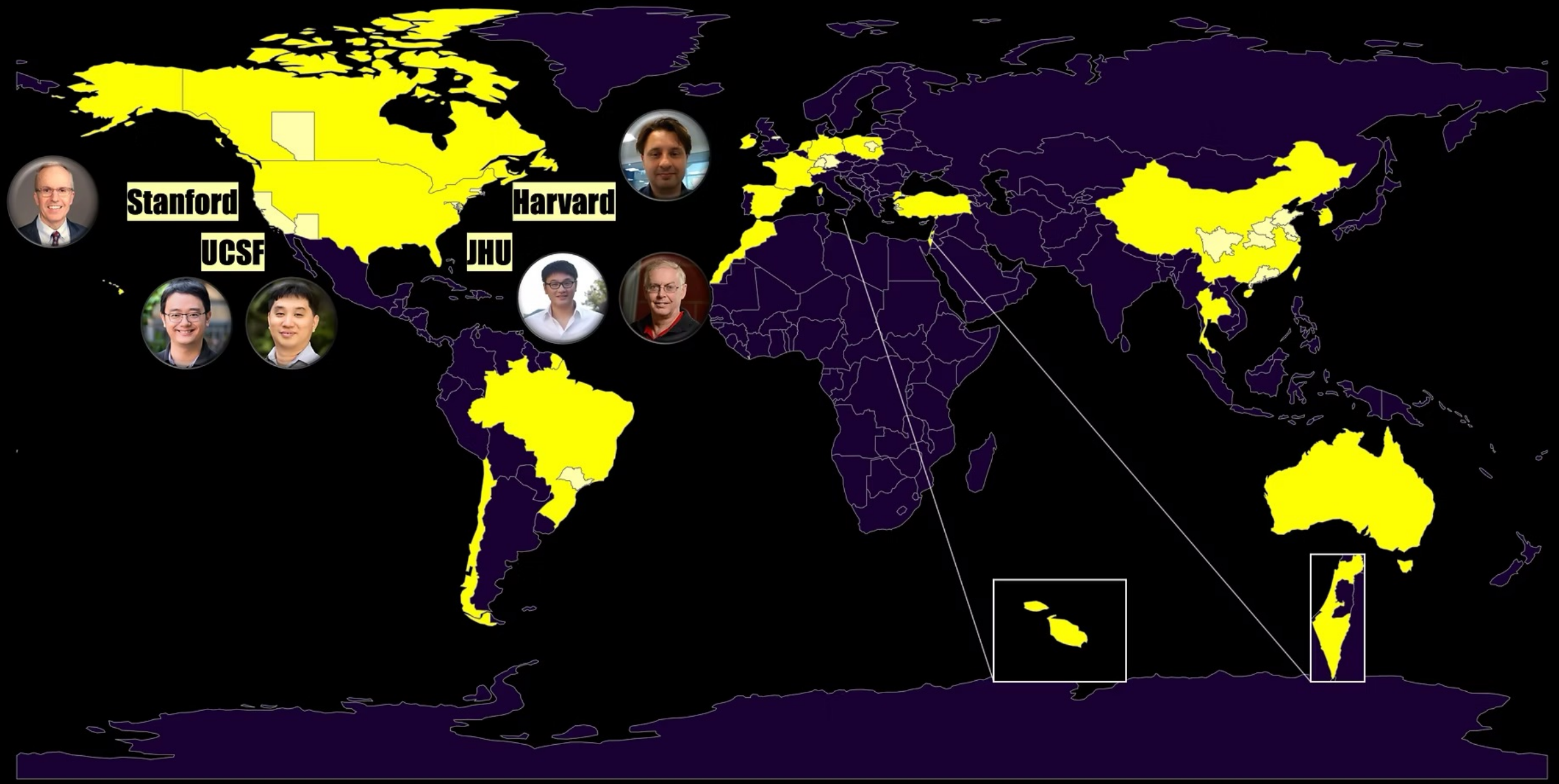
pancreatic tumors, pancreas head, body, and tail, and 24 surrounding anatomical structures such as vascular/skeletal structures and abdominal/thoracic organs.

(Li et al., NeurIPS 2025)

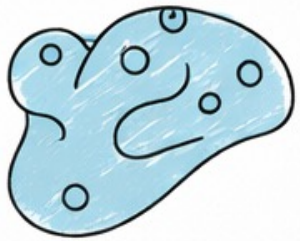


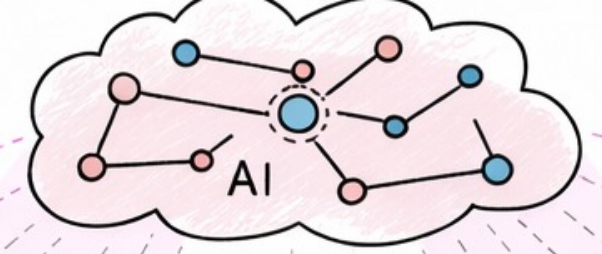
GitHub.com/MrGiovanni/PanTS

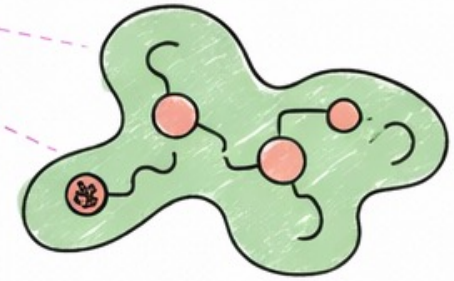
We have access to CT scans from **145 hospitals** worldwide,
and our **collaboration** is expanding



We have access to CT scans from **145 hospitals** worldwide,
and our **collaboration** is expanding







How many lives could be saved by an earlier diagnosis?