Early and Prediagnostic Detection of Pancreatic Cancer from Computed Tomography

Zongwei Zhou, PhD
Assistant Research Professor
Dept. Computer Science, Johns Hopkins University

JOHNS HOPKINS
UNIVERSITY

2018 **FELIX** Lustgarten

2021 Joined JHU

2023 **FELIX-Civitas** Lustgarten & McGovern

2025 **FELIX-Civitas** Lustgarten & NIH

This talk summarizes a lot of research over the last four years

ZONGWEI ZHOU AWARDED $2.8 MILLION NIH GRANT

HOME / NEWS / ZONGWEI Z

Early Detection of Cancer (#2 Killer)

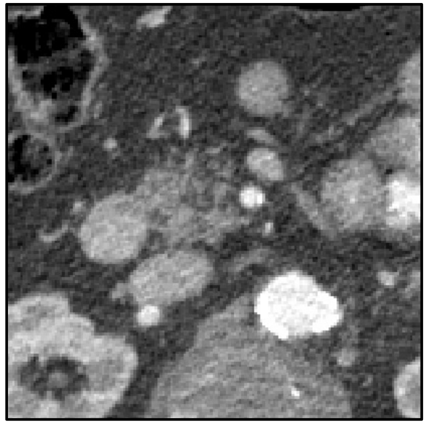
- Early detection of cancer enables surgery and will save many lives.
- For pancreatic cancer, the five-year survival rate increases from about **7-10%** to **40-45%** if detected at an early stage.
- **80,000,000** Computer Tomography (CT) scans taken each year in the US, enabling to screen many people.
- Radiologists can detect pancreatic cancer from CT scans, but the sensitivity of early pancreatic cancer is only **33-44%**.
- This motivates the development of AI algorithms for detecting and localizing early cancer from CT scans, less than 2 cm, and even before tumors are visible.

A Successful Story

- We formulate this problem as *Semantic Segmentation*.
- We developed an AI algorithm and train it to classify voxels as Healthy Pancreas, Tumor, or Background.
- For the pancreas, our AI has achieved very high performance.

	Sensitivity early tumors ≤2 cm	Sensitivity all-size tumors	Specificity
Radiologists	33–44%	76–92%	82–96%
Our AI	94%	97%	99%

A Successful Story



Prediagnostic
CT scans

Diagnostic
CT scans

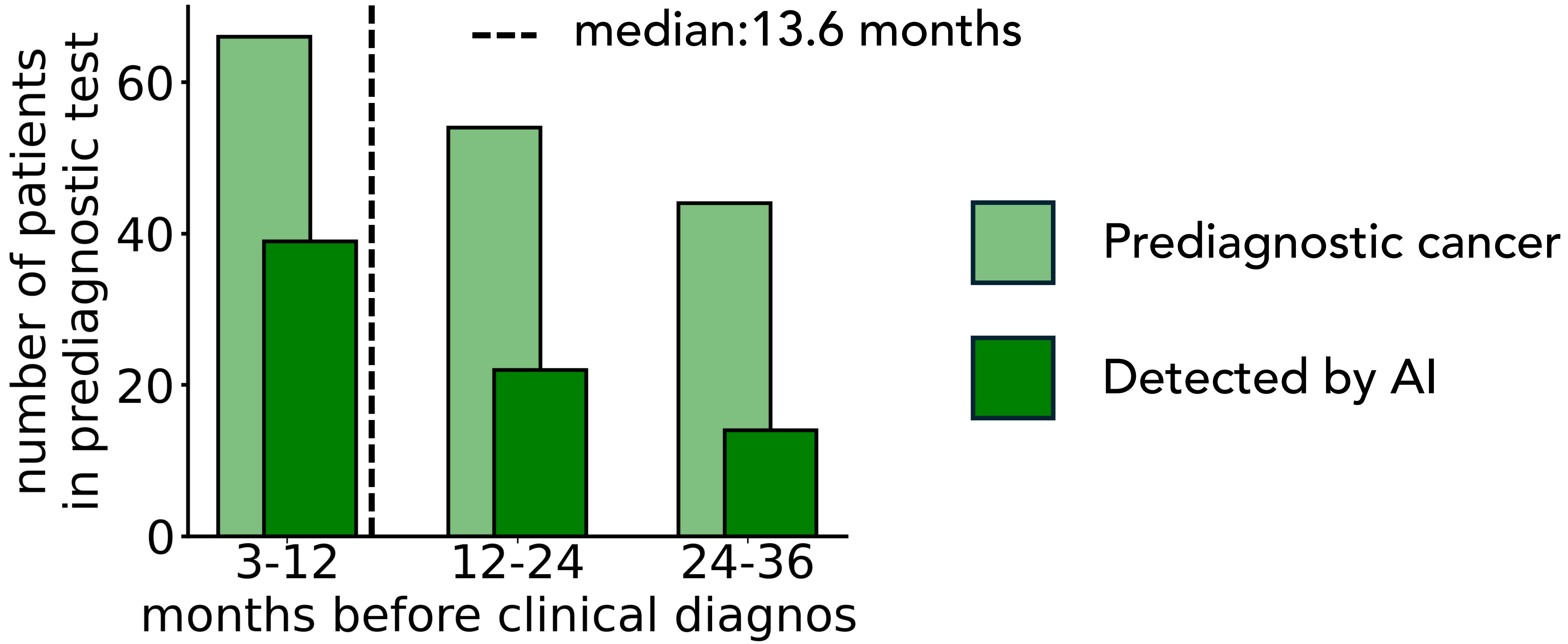
2021

3-36 months
before diagnosis

2024

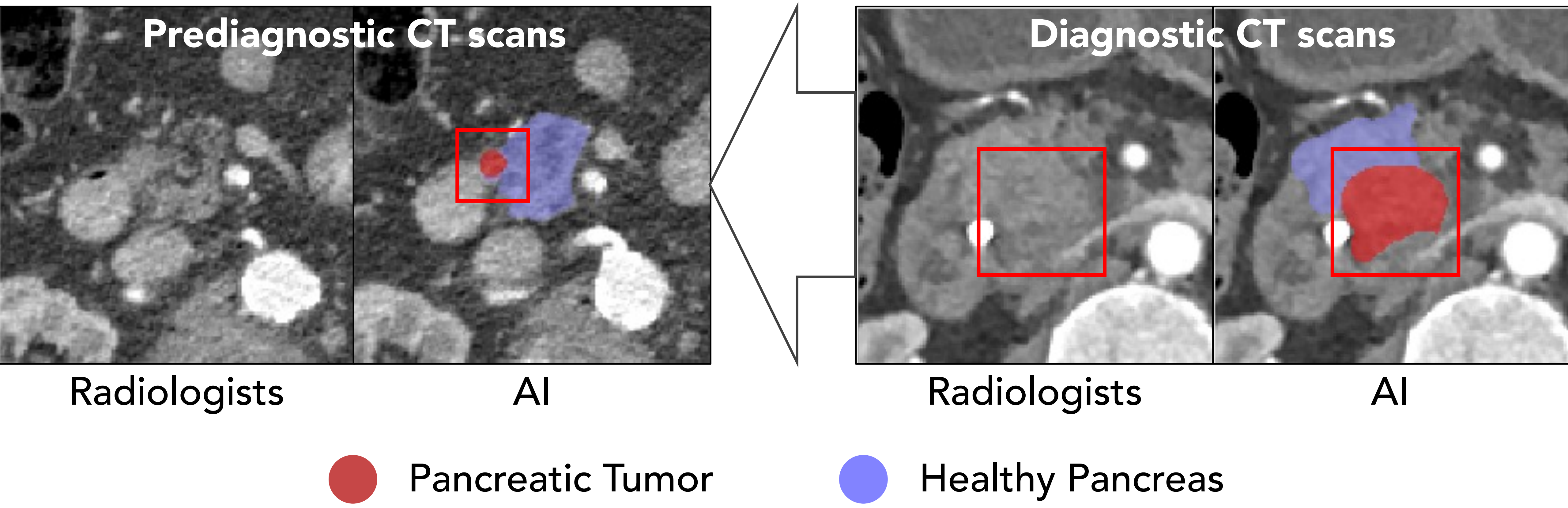
A Successful Story

- Our AI can detect cancer **13.6 months** earlier than radiologists.



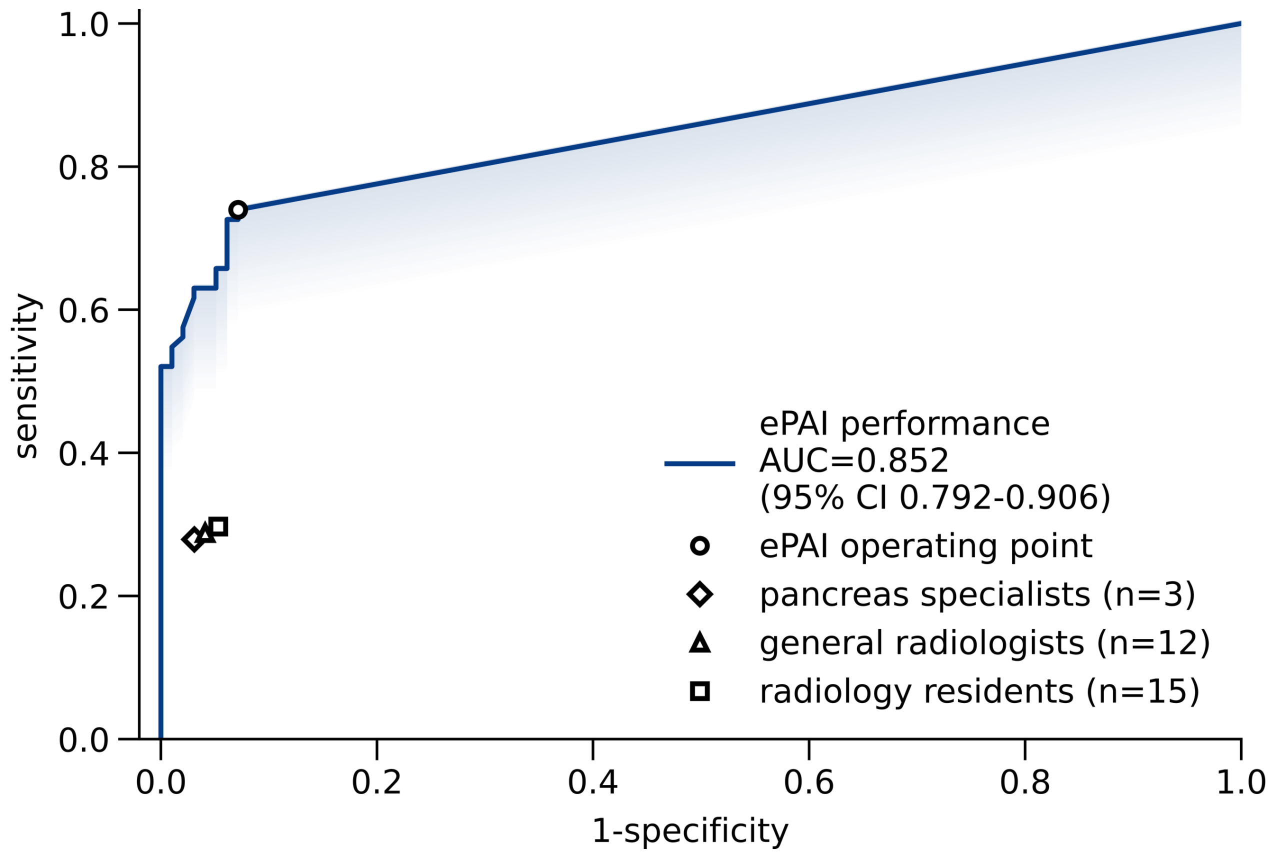
A Successful Story

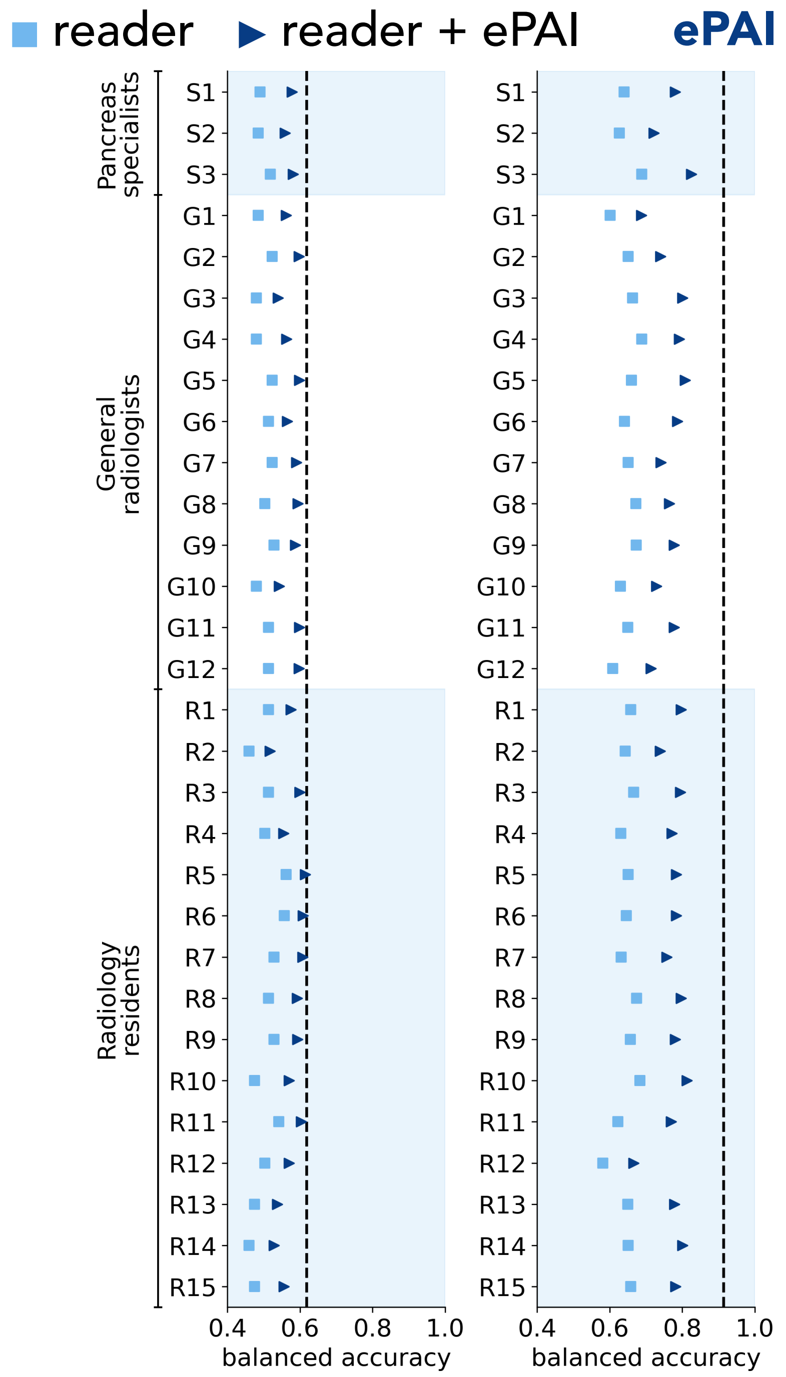
- Our AI can also **localize** these cancer in prediagnostic scans.



A Successful Story

- Our AI can be an assistive tool for radiologists.

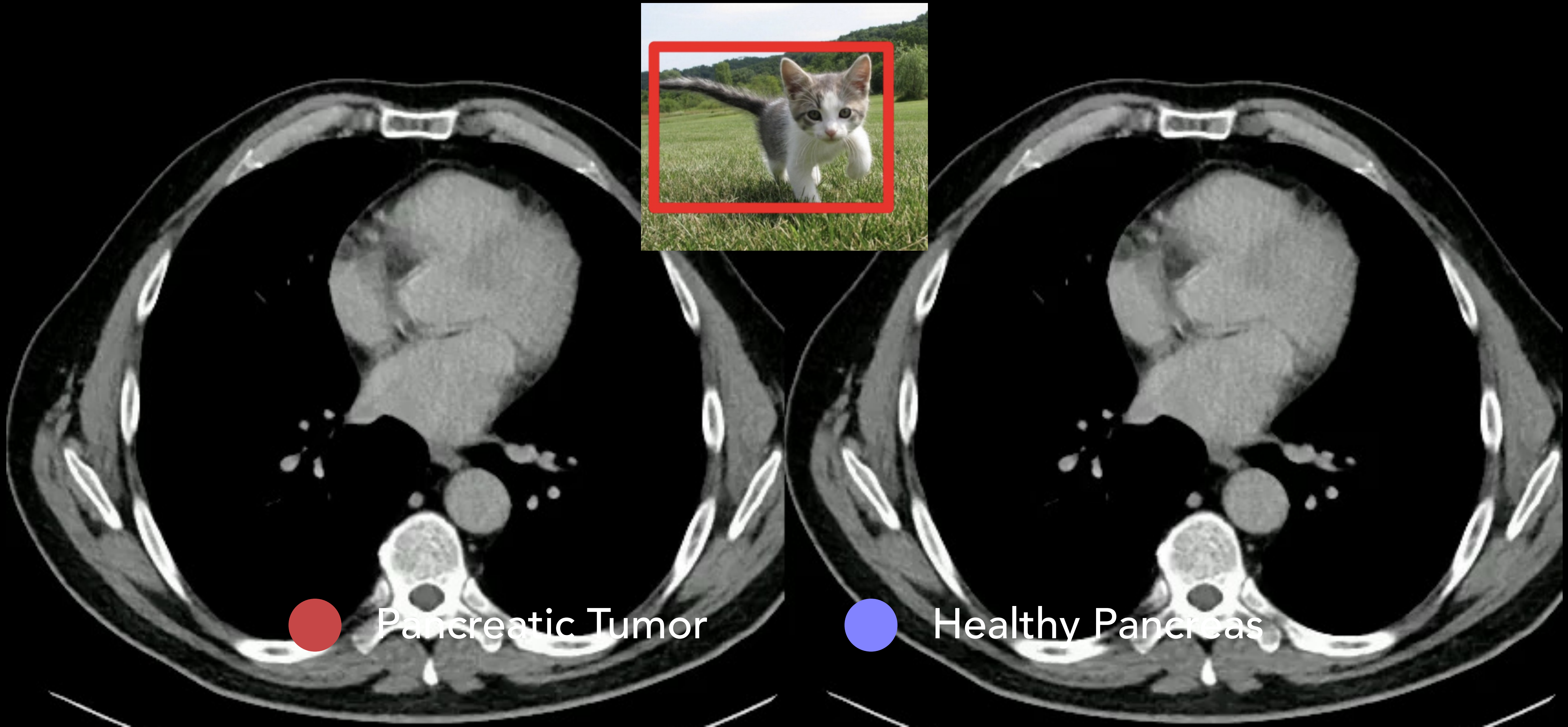




Challenges (1/3)

- **Algorithms:** How to detect small cancer?
- **Data:** Where to get data from different hospitals?
- **Annotations:** How to annotate the data?

Tumors (0.0001%) in 3D medical images vs. objects (5-50%) in 2D natural images



Chapter I. Segmentation Architectures

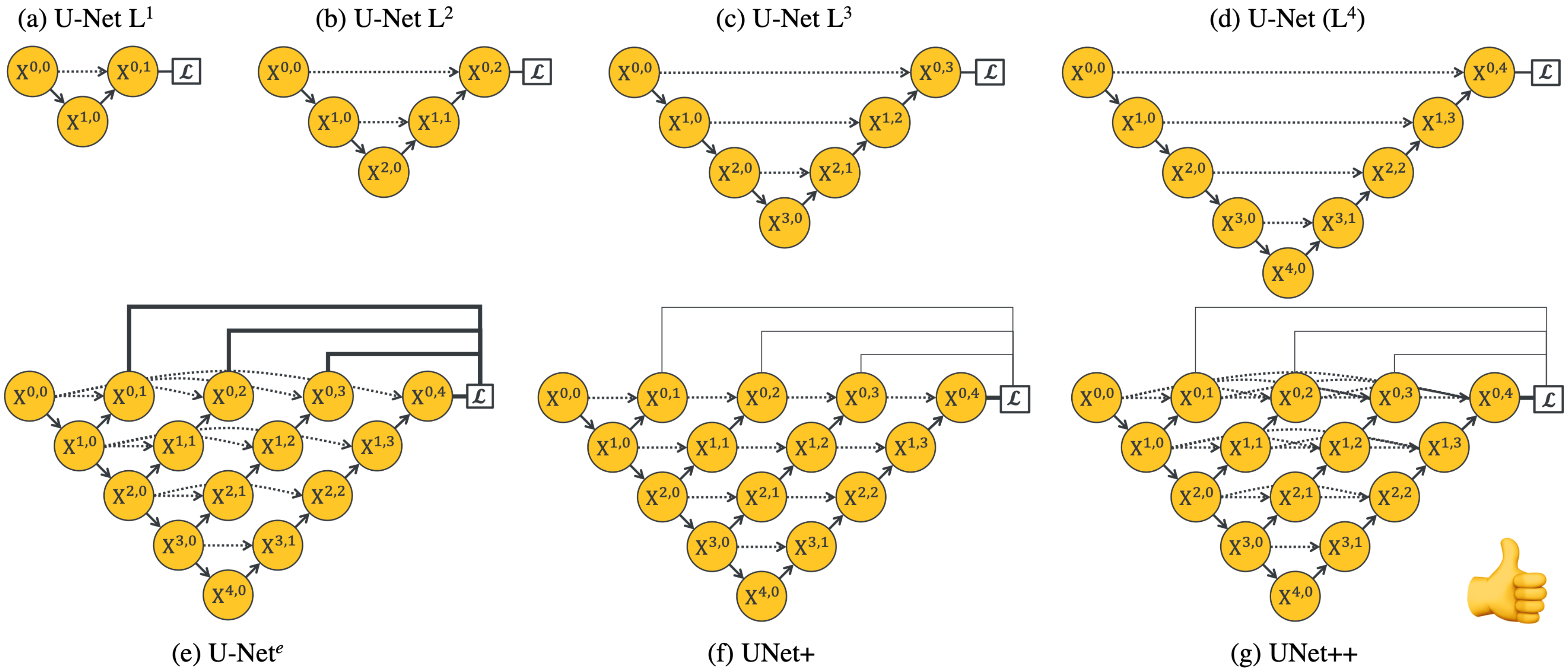
- AI is an extremely dynamic research field. Novel AI algorithms are continually being created and improved.

Medical Image Segmentation

1.	U-Net	(O. Ronneberger et al., 2015)	125,000 cites
2.	UNet++	(Z. Zhou et al., 2019)	16,000 cites
3.	TransU-Net	(J. Chen et al., 2021)	8,000 cites
4.	nnU-Net	(F. Isensee et al., 2020)	8,000 cites
5.	…		

Developing Segmentation Architectures

- UNet++ is suitable for segmenting tumors of a wide range of sizes (Z. Zhou et al., TMI 2019; AMIA Edward H. Shortliffe Doctoral Dissertation Award).

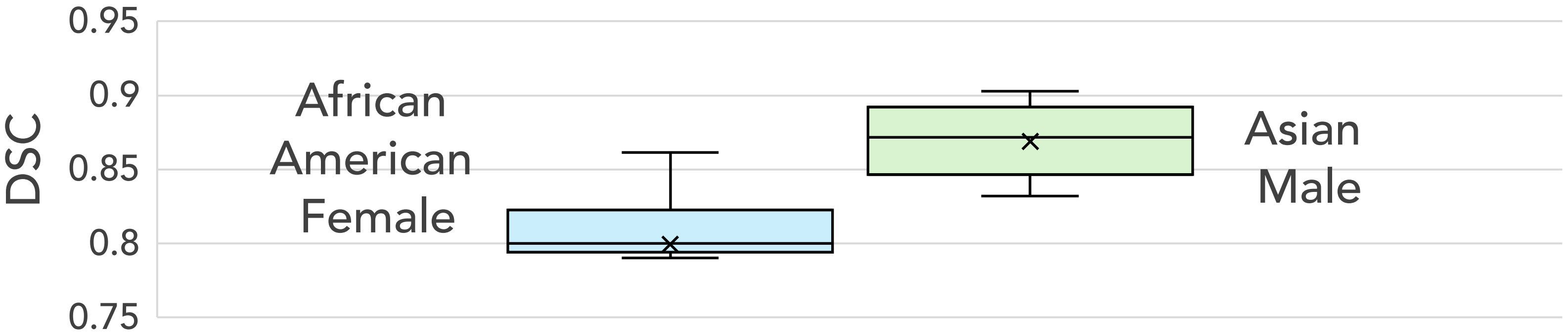


Challenges (2/3)

- **Algorithms:** How to detect small cancer?
- **Data:** Where to get data from different hospitals?
- **Annotations:** How to annotate the data?

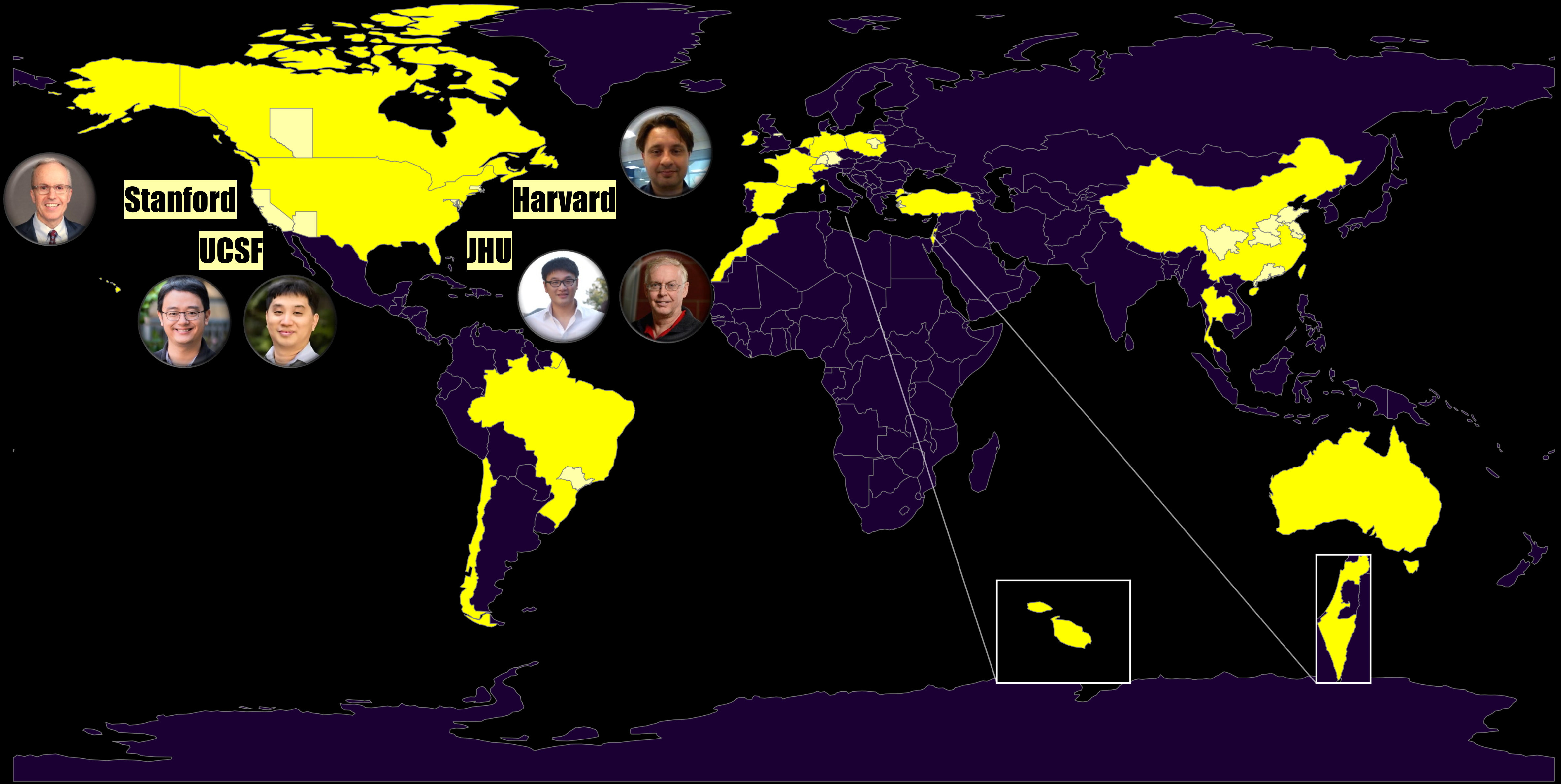
Rigorously Evaluating Medical AI

- 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.
- This is called the *Domain Transfer (DT)* problem (A. Lubonja et al., MICCAIW 2025).

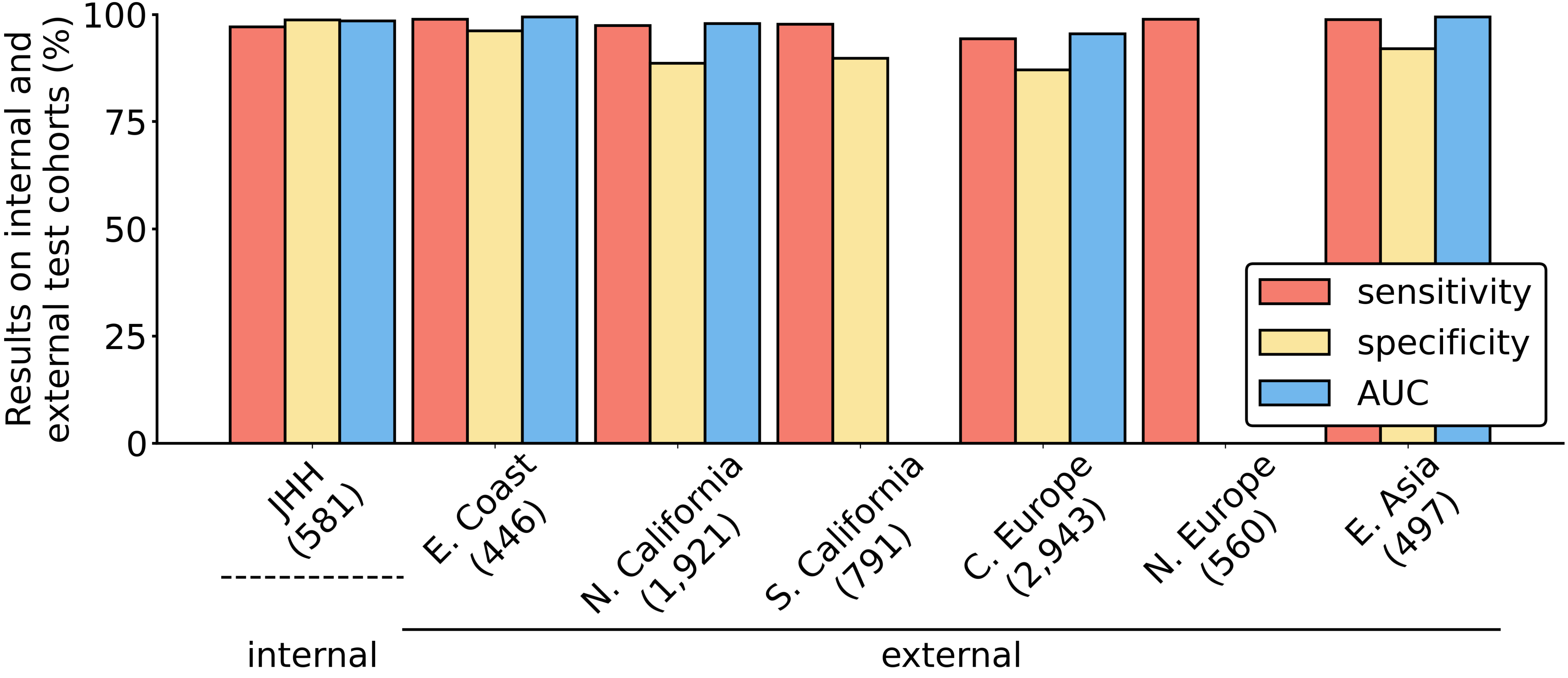


GitHub.com/ariellubonja/RankInsight

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



Rigorously Evaluating Medical AI



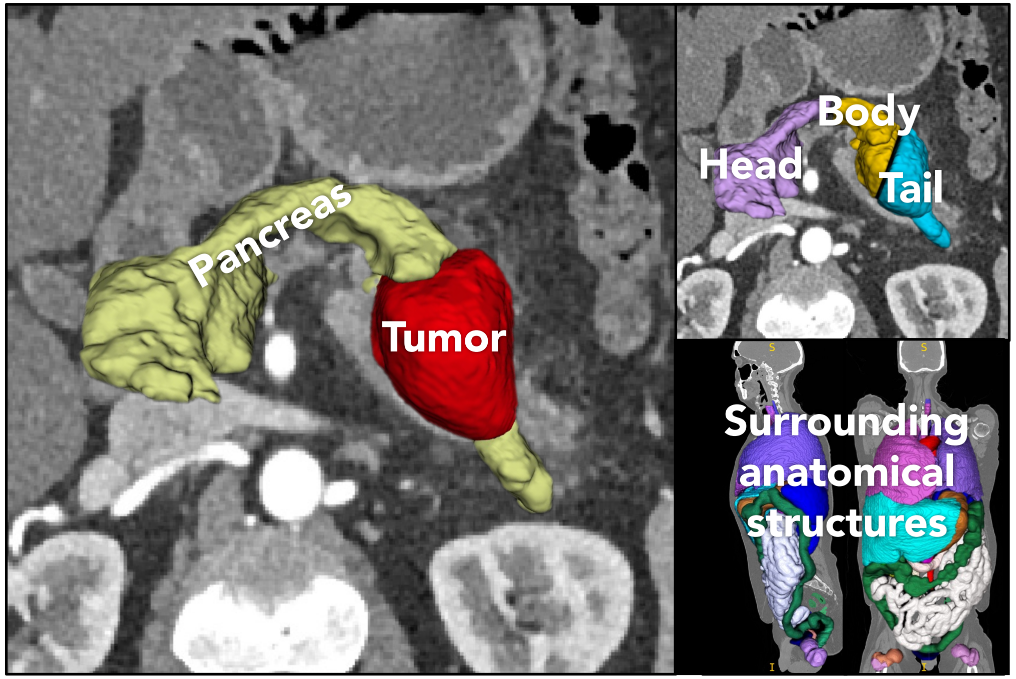
Challenges (3/3)

- **Algorithms:** How to detect small cancer?
- **Data:** Where to get data from different hospitals?
- **Annotations:** How to annotate the data?

Behind the Success: The JHU Dataset

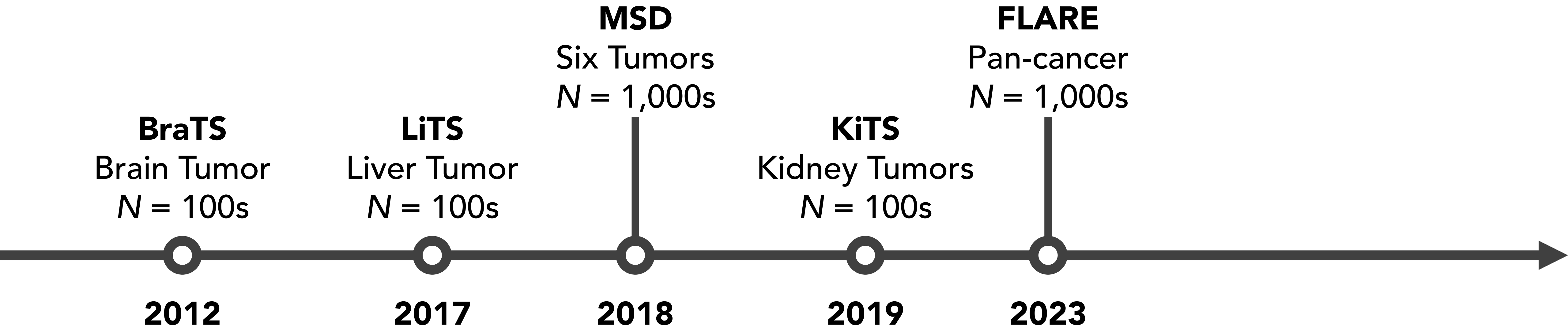
- **>5,000** voxel-wise annotated CT scans, requiring **25** person years.
- It trains high-performance AI algorithms (Xia et al., medRxiv, 2022)
  - Sensitivity = 97%, Specificity = 99%
  - Exceed radiologist performance.
  - Generalizable to multiple centers.

- *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).

2012 — **BraTS**
Brain Tumor
N = 100s

2017 — **LiTS**
Liver Tumor
N = 100s

2018 — **MSD**
Six Tumors
N = 1,000s

2019 — **KiTS**
Kidney Tumors
N = 100s

2023 — **FLARE**
Pan-cancer
N = 1,000s

2025 — **PanTS**
Pancreatic Tumors
N = 36,390

JOHNS HOPKINS
UNIVERSITY

Our AI Strategy (Two Chapters)

I. Active learning, quickly creating voxel-wise annotations
II. Novel strategies, reducing the need of voxel-wise annotations
    A. Radiology reports as weak supervision
    B. Synthetic tumors as additional training data

Chapter I. Annotations with Active Learning

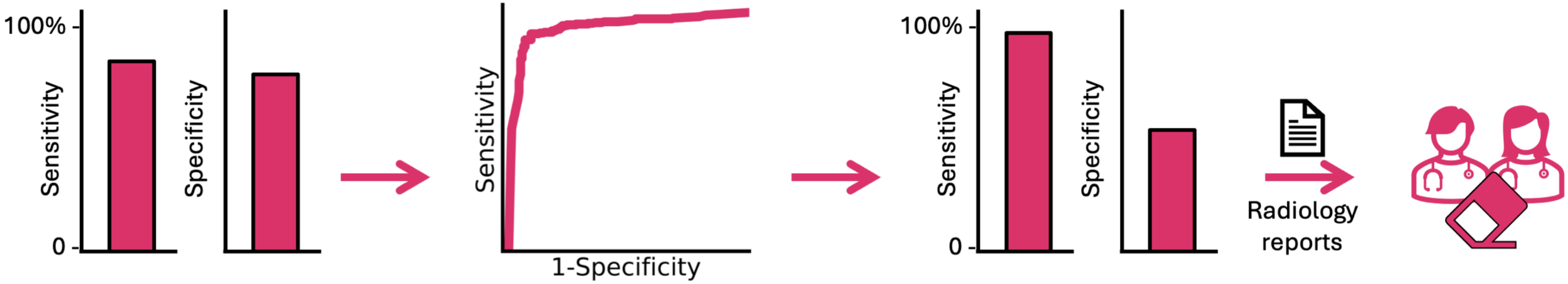
- We created the largest, annotated, public CT dataset.
- It provided **9,901** patients' CT scans of human subjects with and without cancer collected from **145** hospitals worldwide.
- We created voxel-wise annotations in this dataset by active learning.
- We speed up
  - organ annotations by **533x** (W. Li et al., MIA 2024)
  - tumor annotations by **80x**

https://www.zongweiz.com/dataset

Efficient Tumor Annotations

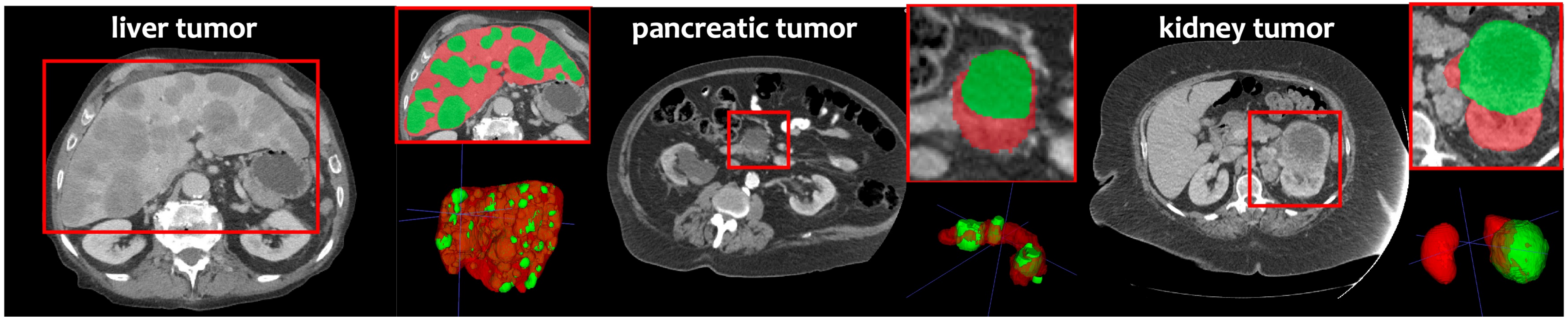
- Make the AI highly sensitive, offering a strong starting point for radiologist review and edit at least **80×** faster (Zhou et al., ISBI 2024).



Adjust AI prediction threshold to make the AI overly sensitive,
i.e., false negatives -- false positives ++

Efficient Tumor Annotations

- Make the AI highly sensitive, offering a strong starting point for radiologist review and edit at least **80×** faster (Zhou et al., ISBI 2024).
- (I) Editing an AI-generated tumor takes **~1 minute**. (rarely needed)
- (II) Removing a false positive takes **<5 seconds**.
- In contrast, manual annotation from scratch takes **4–5 minutes**.



PanTS

Pancreatic Tumor Segmentation

GitHub.com/MrGiovanni/PanTS

```
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

A Huge Annotated Internal Dataset

BodyMaps

Internal use only; open for collaboration

Funded by

NIH R01 (PI: Zongwei Zhou, Yang Yang, Kang Wang),

Lustgarten Foundation (PI: Alan Yuille), and

McGovern Foundation (PI: Alan Yuille)

81.7 million	**241,336**	**377**	**145**
2D CT images	3D CT volumes	anatomical structures	hospitals

A team of 23 board-certified radiologists

A Huge Annotated Internal Dataset

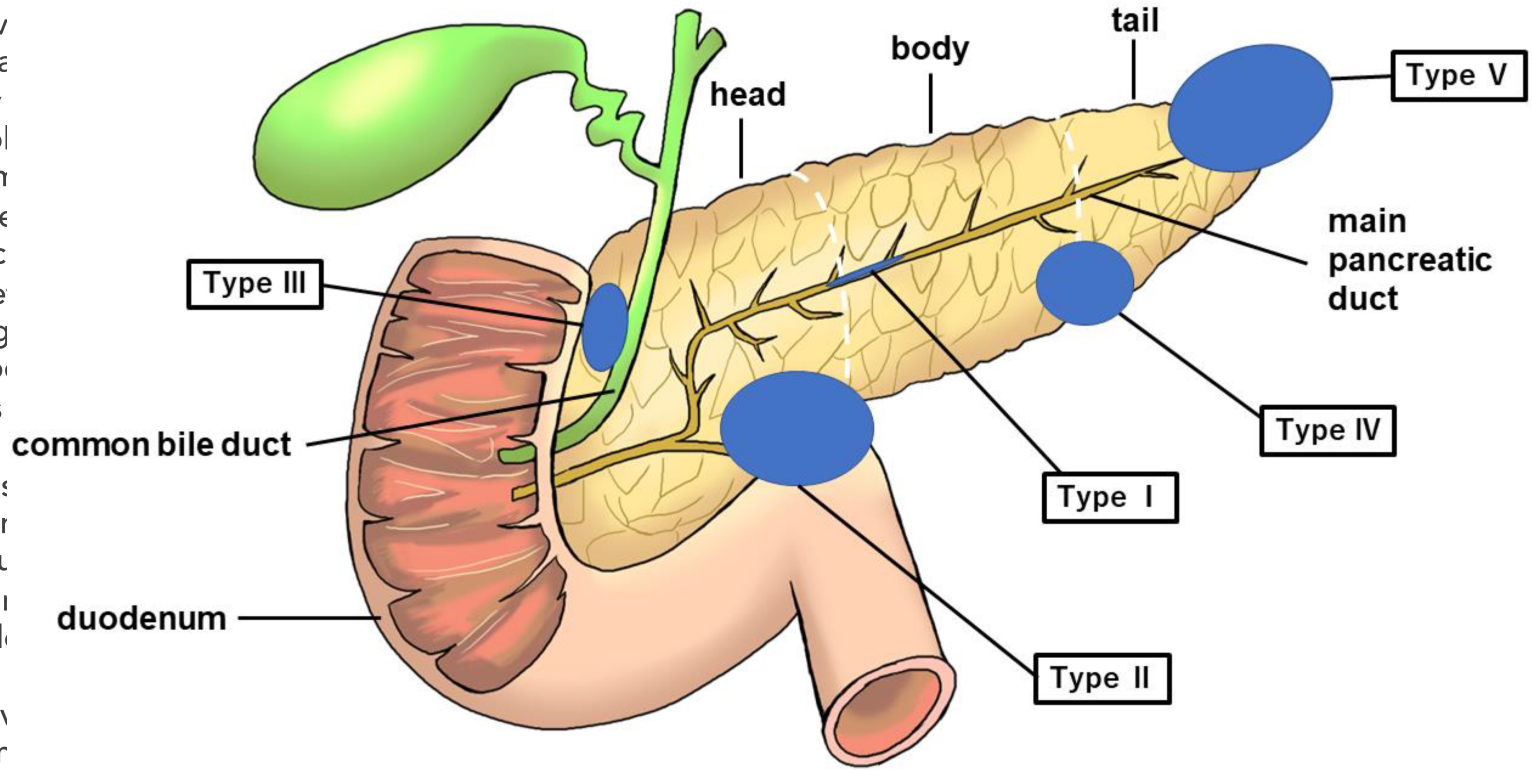
Voxel-wise annotated 16 tumor types: adrenal · bladder · bone · breast · colon · duodenum · esophagus · gallbladder · kidney · liver · lung · pancreas · prostate · spleen · stomach · uterus

81.7 million	**241,336**	**377**	**145**
2D CT images	3D CT volumes	anatomical structures	hospitals

A team of 23 board-certified radiologists

Voxel-wise annotated 377 anatomical structures: abdominal cavity · adrenal gland left · adrenal gland right · airway · anterior eyeball segment left · anterior eyeball segment right · anterior scalene left · anterior scalene right · aorta · arm left · arm right · arms · artery brachiocephalic · artery common carotid left · artery common carotid right · artery internal carotid left · artery internal carotid right · artery subclavian left · artery subclavian right · atrial appendage left · arytenoid cartilage · auditory canal left · auditory canal right · autochthon left · autochthon right · bladder · blood · body · body extremities · body trunc · bones · both lips · brachiocephalic trunk · brachiocephalic vein left · brachiocephalic vein right · brain · brain ventricle · brainstem · breast left · breast right · bronchus · buccal mucosa · carotid artery left · carotid artery right · carpal · carpal left · carpal right · caudate nucleus · cbd stent · celiac artery · celiac trunk · central sulcus · cerebellum · cerebrospinal fluid · cervical esophagus · cheek left · cheek right · clavicula left · clavicula right · cochlear left · cochlear right · colon · colostomy bag · common bile duct · common carotid artery left · common carotid artery right · common iliac artery left · common iliac artery right · common iliac vein left · common iliac vein right · compact bone · coronary artery · costal cartilages · cricoid cartilage · cricopharyngeus · digastric left · digastric right · duodenum · esophagus · eye lens left · eye lens right · eyeball · eyeball left · eyeball right · face · fat · femur left · femur right · fibula · fibula left · fibula right · fingers left · fingers right · frontal lobe · gall bladder · gland structure · glottis · gluteus maximus left · gluteus maximus right · gluteus medius left · gluteus medius right · gluteus minimus left · gluteus minimus right · gonads · gray matter · hard palate · head · heart · heart atrium left · heart atrium right · heart myocardium · heart tissue · heart ventricle left · heart ventricle right · hepatic vein · hepatic vessel · hip left · hip right · humerus left · humerus right · hyoid · hypopharynx · iliac artery left · iliac artery right · iliac vena left · iliac vena right · iliopsoas left · iliopsoas right · inferior oblique muscle left · inferior oblique muscle right · inferior pharyngeal constrictor · inferior rectus muscle left · inferior rectus muscle right · inferior vena cava · insular cortex · intermuscular adipose tissue · internal capsule · internal carotid artery left · internal carotid artery right · internal jugular vein left · internal jugular vein right · intestine · kidney · kidney left · kidney right · lacrimal gland left · lacrimal gland right · larynx · lateral pterygoid left · lateral pterygoid right · lateral rectus muscle left · lateral rectus muscle right · lentiform nucleus · leg left · leg right · legs · levator palpebrae superioris left · levator palpebrae superioris right · levator scapulae left · levator scapulae right · liver · liver segment 1 · liver segment 2 · liver segment 3 · liver segment 4 · liver segment 5 · liver segment 6 · liver segment 7 · liver segment 8 · lung · lung left · lung lower left lobe · lung lower right lobe · lung middle right lobe · lung right · lung trachea bronchia · lung upper left lobe · lung upper right lobe · lung vessels · mammary gland left · mammary gland right · mandible · masseter left · masseter right · medial pterygoid left · medial pterygoid right · medial rectus muscle left · medial rectus muscle right · mediastinal tissue · mediastinum · metacarpal · metacarpal left · metacarpal right · metatarsal · metatarsal left · metatarsal right · middle pharyngeal constrictor · middle scalene left · middle scalene right · muscle · muscle fat · muscle of head · nasal cavity left · nasal cavity right · nasopharynx · occipital lobe · optic chiasm · optic nerve left · optic nerve right · oral cavity · oropharynx · pancreas · pancreas body · pancreas head · pancreas tail · pancreatic duct · parietal lobe · parotid gland left · parotid gland right · parotid glands · patella · patella left · patella right · pericardium · phalanges feet · phalanges hand · pituitary gland · platysma left · platysma right · portal vein · portal vein and splenic vein · postcava · posterior eyeball segment left · posterior eyeball segment right · posterior scalene left · posterior scalene right · prevertebral left · prevertebral right · prostate · prosthetic breast implant · psoas major muscle left · psoas major muscle right · pulmonary artery · pulmonary vein · radius · radius left · radius right · rectum · rectus abdominis muscle left · rectus abdominis muscle right · renal vein left · renal vein right · rib cartilage · rib left 1 · rib left 2 · rib left 3 · rib left 4 · rib left 5 · rib left 6 · rib left 7 · rib left 8 · rib left 9 · rib left 10 · rib left 11 · rib left 12 · rib right 1 · rib right 2 · rib right 3 · rib right 4 · rib right 5 · rib right 6 · rib right 7 · rib right 8 · rib right 9 · rib right 10 · rib right 11 · rib right 12 · sacrum · scalp · scapula left · scapula right · seminal vesicle · septum pellucidum · sigmoid colon · skeletal muscle · skin · skull · small bowel · soft palate · spinal canal · spinal cord · spleen · spongy bone · sterno thyroid left · sterno thyroid right · sternocleidomastoid left · sternocleidomastoid right · sternum · sternum corpus · sternum manubrium · stomach · styloid process left · styloid process right · subarachnoid space · subclavian artery left · subclavian artery right · subcutaneous adipose tissue · submandibular gland left · submandibular gland right · submandibular glands · superior mesenteric artery · superior oblique muscle left · superior oblique muscle right · superior rectus muscle left · superior rectus muscle right · superior vena cava · supraglottis · tarsal · tarsal left · tarsal right · temporal lobe · temporalis left · temporalis right · thalamus · thoracic cavity · thymus · thyrohyoid left · thyrohyoid right · thyroid cartilage · thyroid gland · thyroid left · thyroid right · tibia · tibia left · tibia right · toes left · toes right · tongue · trachea · trapezius left · trapezius right · ulna · ulna left · ulna right · uterocervix · uterus · veins · venous sinuses · vertebrae C1 · vertebrae C2 · vertebrae C3 · vertebrae C4 · vertebrae C5 · vertebrae C6 · vertebrae C7 · vertebrae L1 · vertebrae L2 · vertebrae L3 · vertebrae L4 · vertebrae L5 · vertebrae S1 · vertebrae T1 · vertebrae T2 · vertebrae T3 · vertebrae T4 · vertebrae T5 · vertebrae T6 · vertebrae T7 · vertebrae T8 · vertebrae T9 · vertebrae T10 · vertebrae T11 · vertebrae T12 · visceral adipose tissue · white matter · zygomatic arch left ·

Voxel-wise annotated 377 anatomical structures: abdominal cav
segment right · anterior scalene left · anterior scalene right · aorta · a
right · artery internal carotid left · artery internal carotid right · artery
auditory canal right · autochthon left · autochthon right · bladder · bl
vein left · brachiocephalic vein right · brain · brain ventricle · brainsten
carpal left · carpal right · caudate nucleus · cbd stent · celiac artery · ce
· clavicula left · clavicula right · cochlear left · cochlear right · colon · c
iliac artery left · common iliac artery right · common iliac vein le
cricopharyngeus · digastric left · digastric right · duodenum · esophag
· fibula · fibula left · fibula right · fingers left · fingers right · frontal lobe
left · gluteus medius right · gluteus minimus left · gluteus minimus
myocardium · heart tissue · heart ventricle left · heart ventricle right
iliac artery left · iliac artery right · iliac vena left · iliac vena right · iliops
constrictor · inferior rectus muscle left · inferior rectus muscle right · ir
left · internal carotid artery right · internal jugular vein left · internal ju
larynx · lateral pterygoid left · lateral pterygoid right · lateral rectus r
superioris left · levator palpebrae superioris right · levator scapulae l
liver segment 5 · liver segment 6 · liver segment 7 · liver segment 8
trachea bronchia · lung upper left lobe · lung upper right lobe · lung v
pterygoid left · medial pterygoid right · medial rectus muscle left · m



right · metatarsal · metatarsal left · metatarsal right · middle pharyngeal constrictor · middle scalene left · middle scalene right · muscle · muscle fat · muscle of head · nasal cavity left · nasal cavity right · nasopharynx · occipital lobe · optic chiasm · optic nerve left · optic nerve right · oral cavity · oropharynx · pancreas · pancreas body · pancreas head · pancreas tail · pancreatic duct · parietal lobe · parotid gland left · parotid gland right · parotid glands · patella · patella left · patella right · pericardium · phalanges feet · phalanges hand · pituitary gland · platysma left · platysma right · portal vein · portal vein and splenic vein · postcava · posterior eyeball segment left · posterior eyeball segment right · posterior scalene left · posterior scalene right · prevertebral left · prevertebral right · prostate · prosthetic breast implant · psoas major muscle left · psoas major muscle right · pulmonary artery · pulmonary vein · radius · radius left · radius right · rectum · rectus abdominis muscle left · rectus abdominis muscle right · renal vein

Focusing on specific type of cancer will require finer annotations.

right · seminal vesicle · septum pellucidum · sigmoid colon · skeletal muscle · skin · skull · small bowel · soft palate · spinal canal · spinal cord · spleen · spongy bone · sterno thyroid left · sterno thyroid right · sternocleidomastoid left · sternocleidomastoid right · sternum · sternum corpus · sternum manubrium · stomach · styloid process left · styloid process right · subarachnoid space · subclavian artery left · subclavian artery right · subcutaneous adipose tissue · submandibular gland left · submandibular gland right · submandibular glands · superior mesenteric artery · superior oblique muscle left · superior oblique muscle right · superior rectus muscle left · superior rectus muscle right · superior vena cava · supraglottis · tarsal · tarsal left · tarsal right · temporal lobe · temporalis left · temporalis right · thalamus · thoracic cavity · thymus · thyrohyoid left · thyrohyoid right · thyroid cartilage · thyroid gland · thyroid left · thyroid right · tibia · tibia left · tibia right · toes left · toes right · tongue · trachea · trapezius left · trapezius right · ulna · ulna left · ulna right · uterocervix · uterus · veins · venous sinuses · vertebrae C1 · vertebrae C2 · vertebrae C3 · vertebrae C4 · vertebrae C5 · vertebrae C6 · vertebrae C7 · vertebrae L1 · vertebrae L2 · vertebrae L3 · vertebrae L4 · vertebrae L5 · vertebrae S1 · vertebrae T1 · vertebrae T2 · vertebrae T3 · vertebrae T4 · vertebrae T5 · vertebrae T6 · vertebrae T7 · vertebrae T8 · vertebrae T9 · vertebrae T10 · vertebrae T11 · vertebrae T12 · visceral adipose tissue · white matter · zygomatic arch left ·

A Series of AI-Ready Datasets

"Annotating **240,000 CT scans** with **72 million anatomical shapes** would require an expert radiologist to have started working around 420 BCE—the era of Hippocrates—to complete the task by 2025. *We did it in two years*." says lead author Zongwei Zhou

Chapter II. Strategies to Further Reduce Need of Voxel-Wise Annotations

A. **Radiology reports** as weak supervision for multi-cancer detection
B. **Synthetic tumors** as additional training data for small tumors

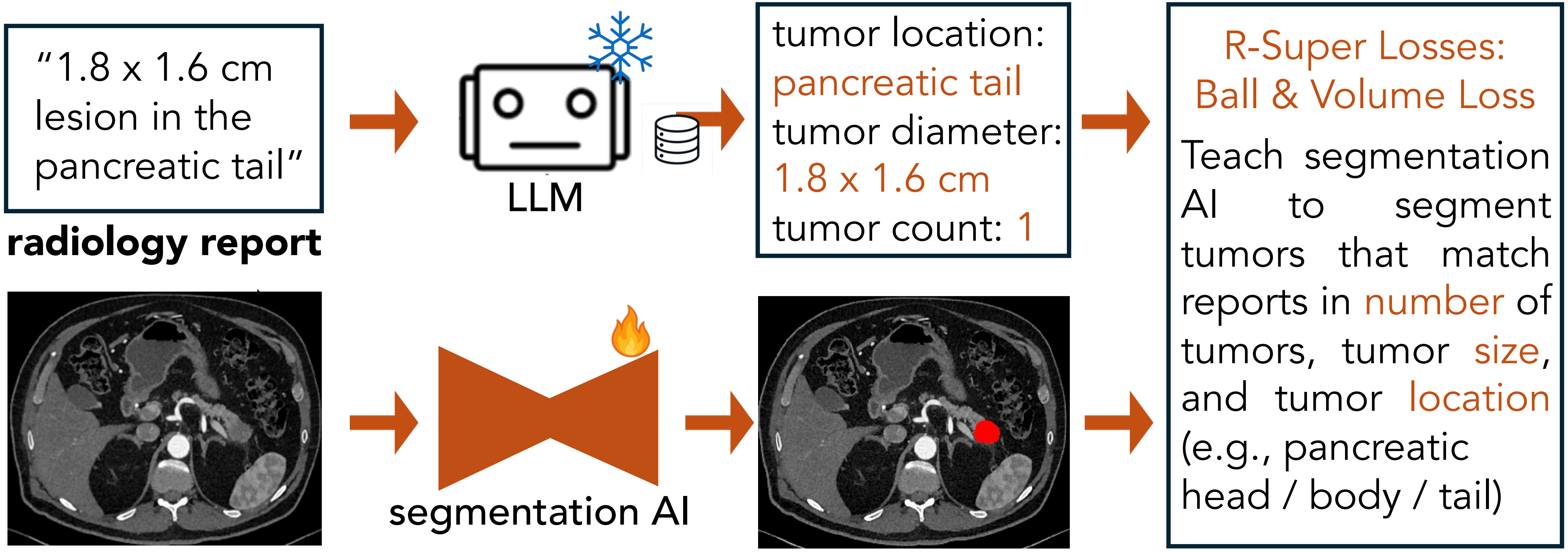
Chapter II.A. Reports as Weak Supervision

- Public datasets have few tumor Image-Mask pairs, only **10s to 100s**.
- By contrast, reports are written every day by radiologists—Public datasets already provide around **25,000** CT-Report pairs, and a single hospital can easily accumulate over **500,000** CT-Report pairs.
- We enable AI to learn tumor segmentation directly from these reports (P. Bassi et al., MICCAI 2025 Best Paper Award, Runner-up; RSNA 2025 Science Poster Award).

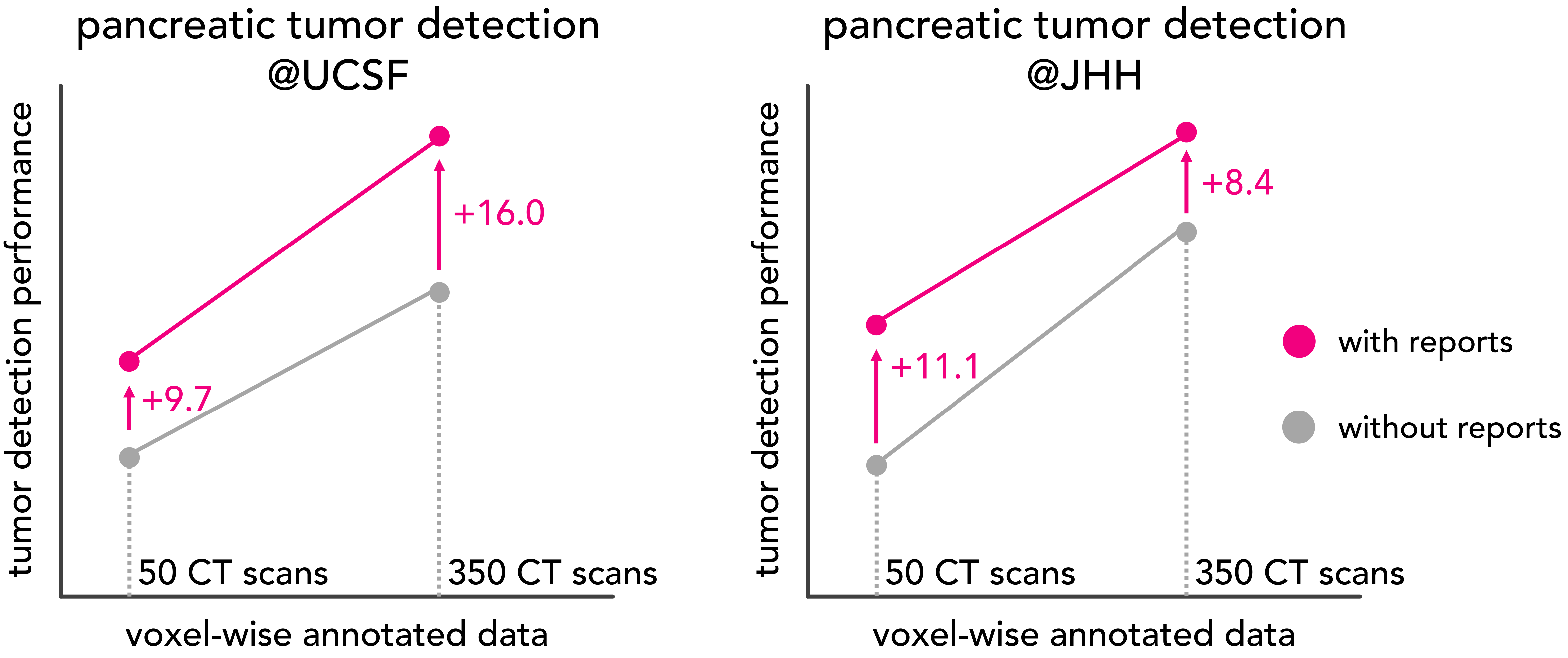
GitHub.com/MrGiovanni/R-Super

Chapter II.A. Reports as Weak Supervision

- R-Super, a novel AI training method that enforce the consistency between AI segmented tumors and report descriptions such as tumor **number, size, and location**.

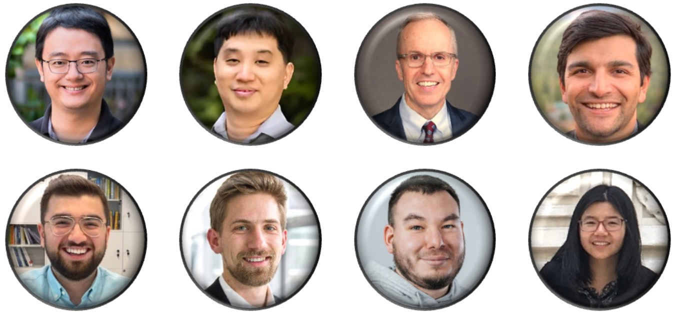


Scaling Laws in Reports Supervision



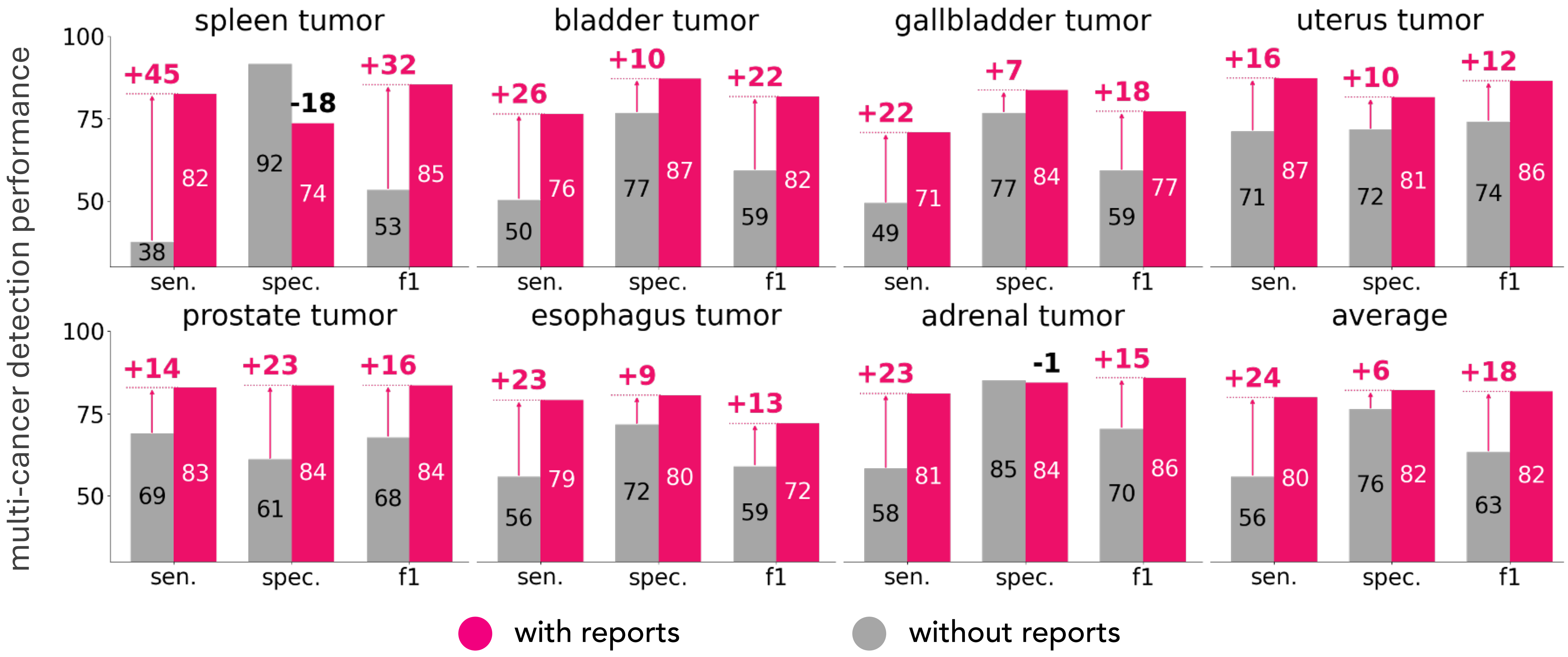
Reports Supervision for Multi-Cancer

- We have curated a dataset of **117,000** Image–Report and **270** CT-Mask pairs for tumors in the adrenal, bladder, esophagus, gallbladder, prostate, spleen, and uterus.
- No publicly available Image–Mask pairs exist for these tumor types.
- We will release the **first** AI model that can segment these tumor types.
- *Many thanks to the Image & Image-Report data providers!*



GitHub.com/MrGiovanni/R-Super

Reports Supervision for Multi-Cancer



Reports for Different Hospitals

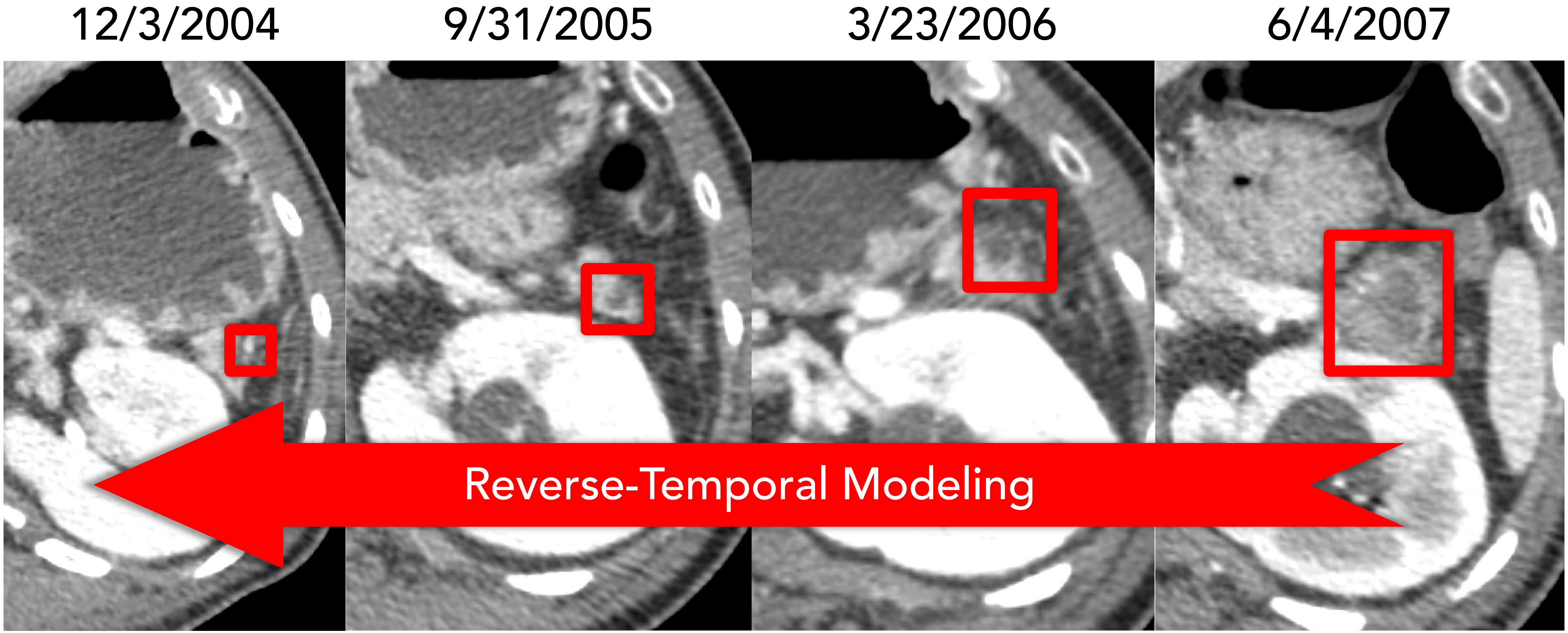
Chapter II.C. Synthetic Data

- There's a huge data gap in medical AI right now, particularly when you have rare diseases, uncommon conditions (e.g., cancer).
- Early-stage tumor scans are 10–20 times less common than late-stage scans in clinical datasets.
- We don't have enough early-stage tumor scans to train these models; unfortunate these are the tumors we must detect to improve survival.

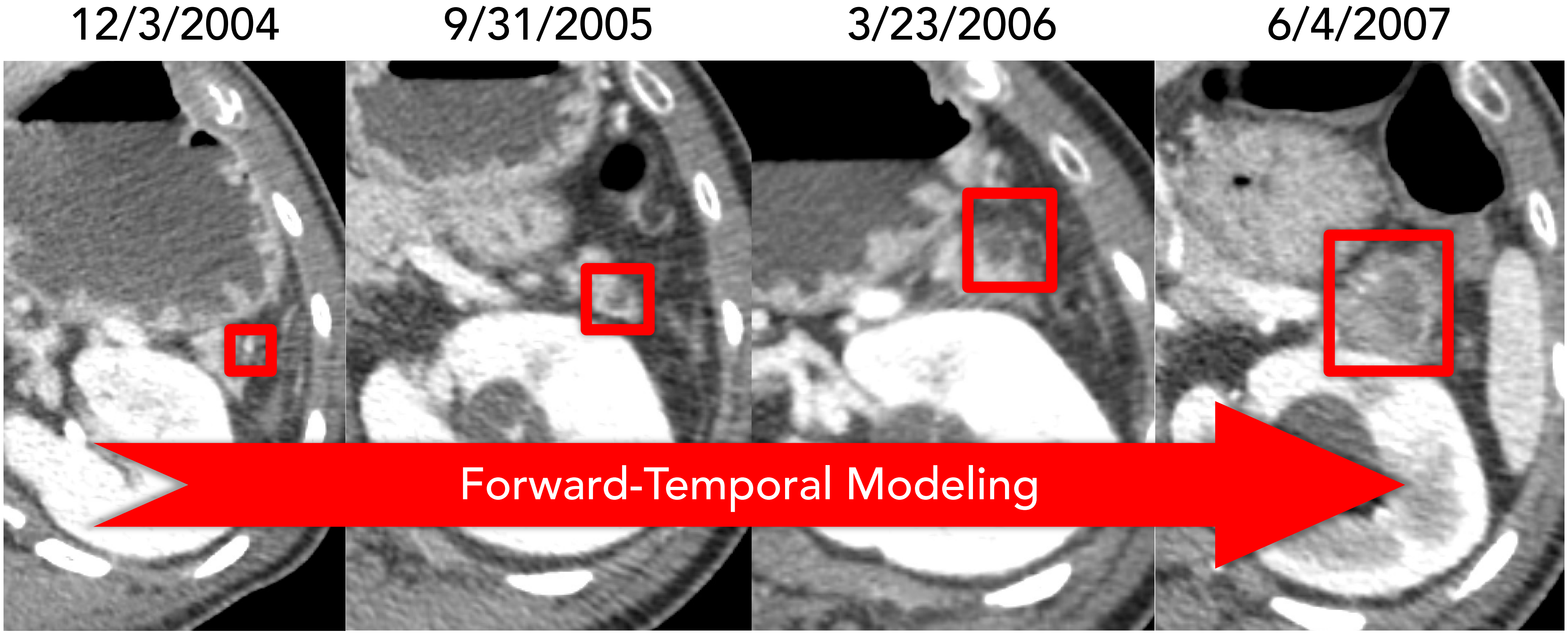
- Synthetic data can be a big piece of that puzzle (Lai et al., MICCA 2024).

GitHub.com/MrGiovanni/Pixel2Cancer

Synthetic Tumors as Time Machine

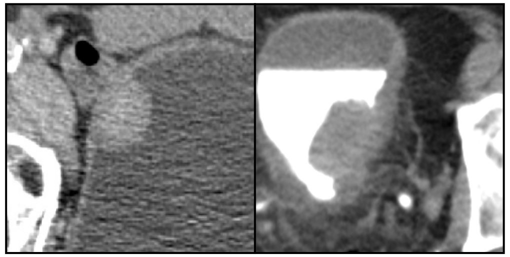


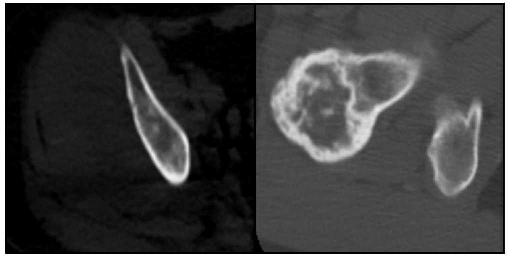
Synthetic Tumors as Time Machine

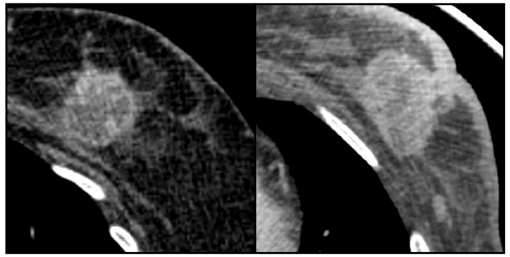


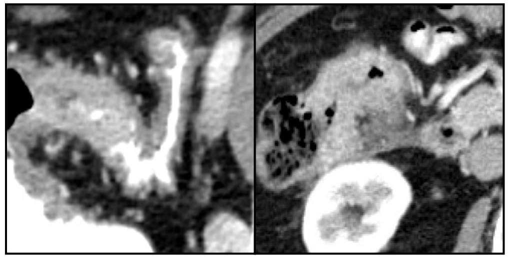
Visual Turing Test for Radiologists

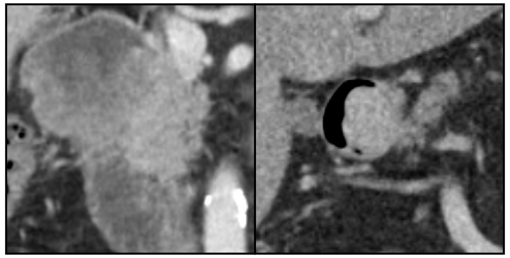
(a) real or fake test

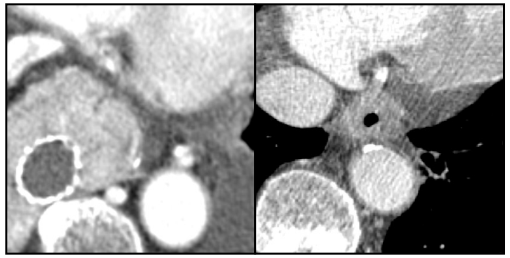

(b) bladder tumor

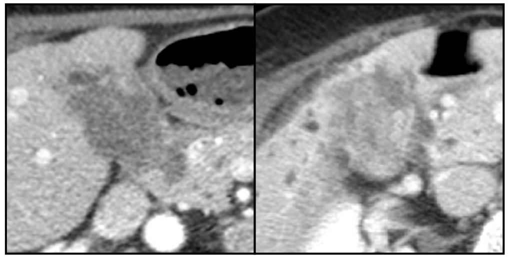

(c) bone tumor

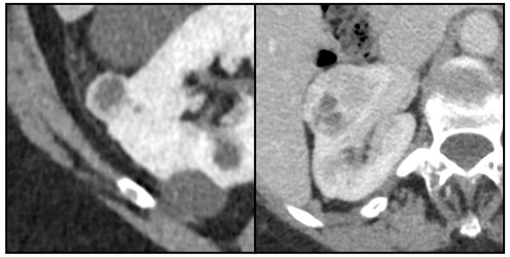

(d) breast tumor

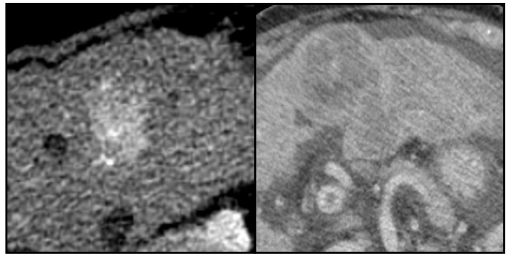

(e) colon tumor

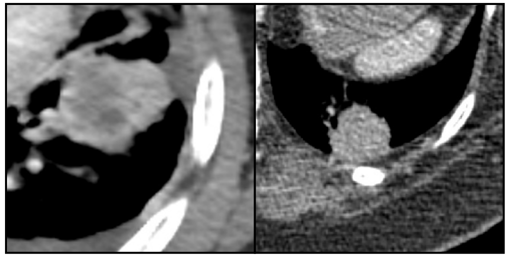

(f) duodenum tumor

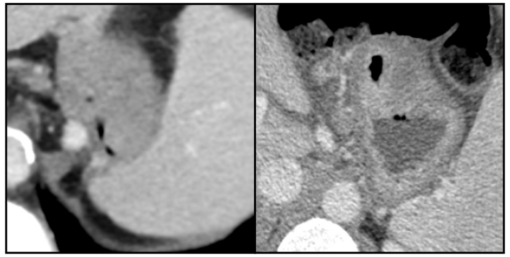

(g) esophagus tumor


(h) gallbladder tumor


(i) kidney tumor

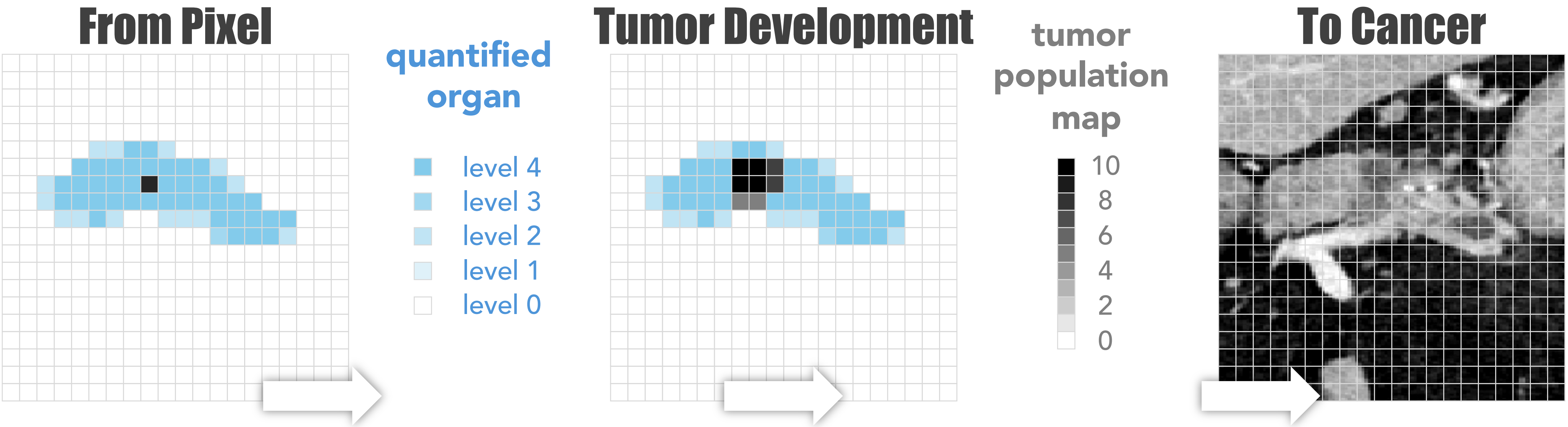

(j) liver tumor


(k) lung tumor


(l) stomach tumor

Tumor/Vessel/Duct/Organ Synthesis

- We developed "game of life" to simulate tumor development (Lai et al., MICCAI 2024) and applied diffusion models to create synthetic tumors.



applying three generic rules

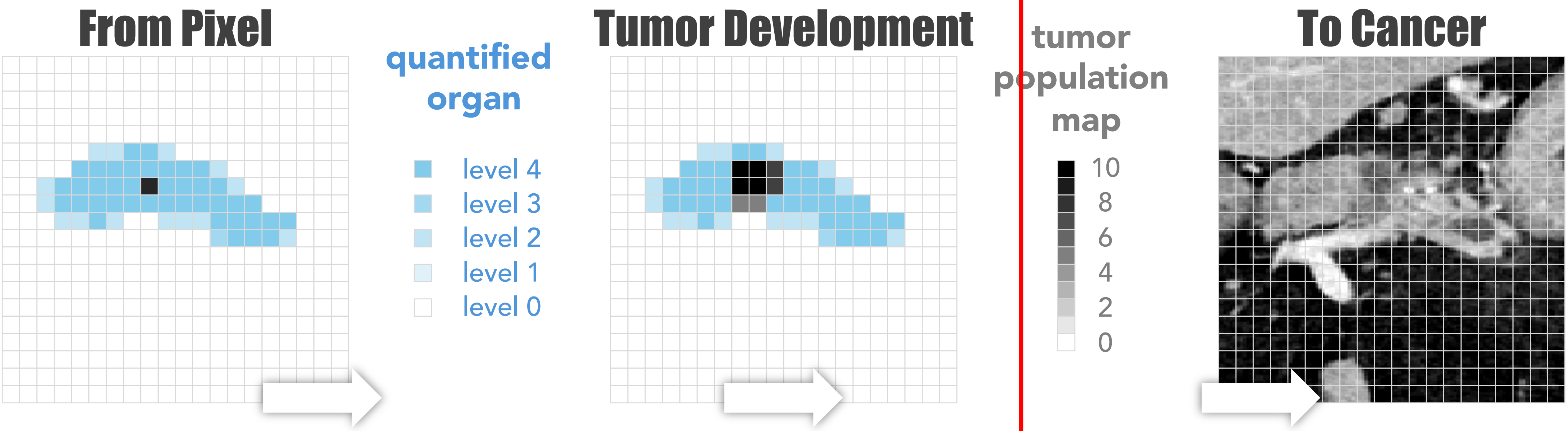
simulating tumor & organ interaction

mapping tumors to CT images

Tumor/Vessel/Duct/Organ Synthesis

Cellular Automata

a mathematical model that uses simple rules to simulate complex systems



applying three generic rules

simulating tumor & organ interaction

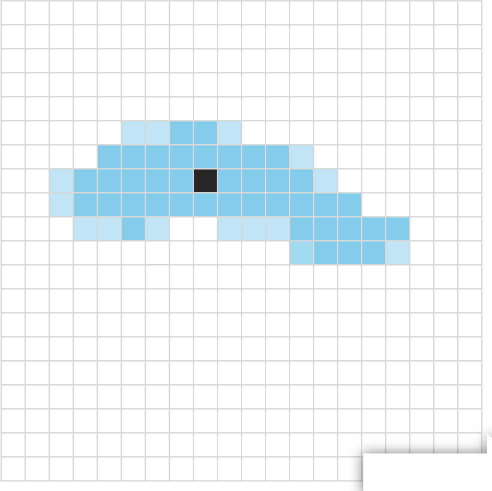
mapping tumors to CT images

Tumor/Vessel/Duct/Organ Synthesis

Diffusion Models

conditioned on tumor/vessel/duct/organ shapes simulated by cellular automata

From Pixel

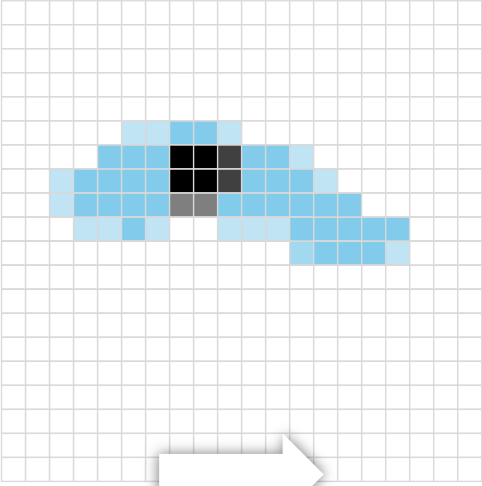


quantified organ

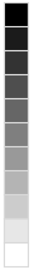
level 4
level 3
level 2
level 1
level 0

applying three generic rules

Tumor Development

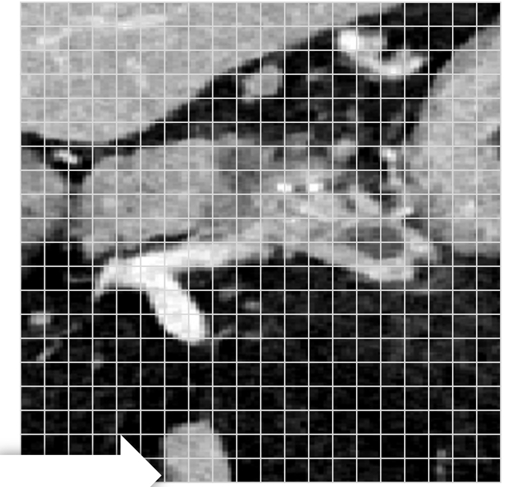


tumor population map



simulating tumor & organ interaction

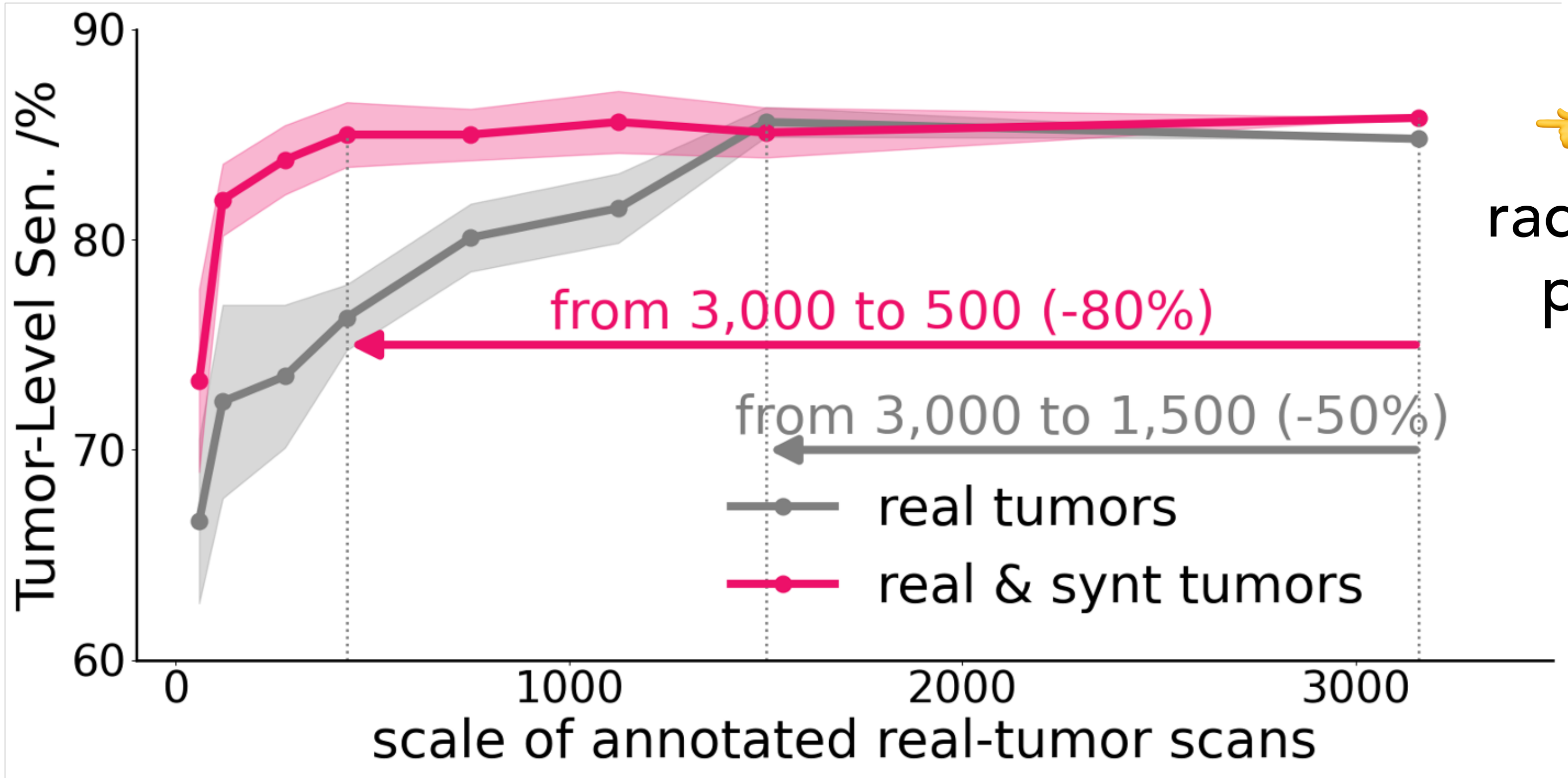
To Cancer



mapping tumors to CT images

Scaling Laws in Synthetic Tumors

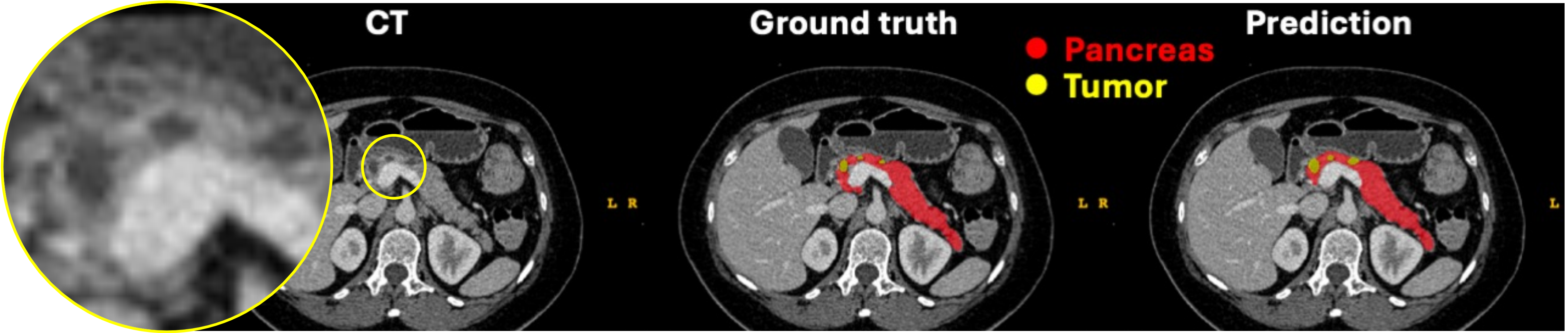
- Synthetic data reduces the need for voxel-wise annotated real data from 1,500 down to 500. (Q. Chen et al., ICCV 2025).



👈 exceeding radiologist-level performance

Synthetic Data for Small/Tiny Tumors

- Synthetic data improves sensitivity of detecting small tumors (≤2 cm) by **5% (89% → 94%)** (Q. Chen et al., CVPR 2024; Q. Hu et al., CVPR 2023)
- The smallest lesion we detected was 2 mm.



GitHub.com/MrGiovanni/SyntheticTumors

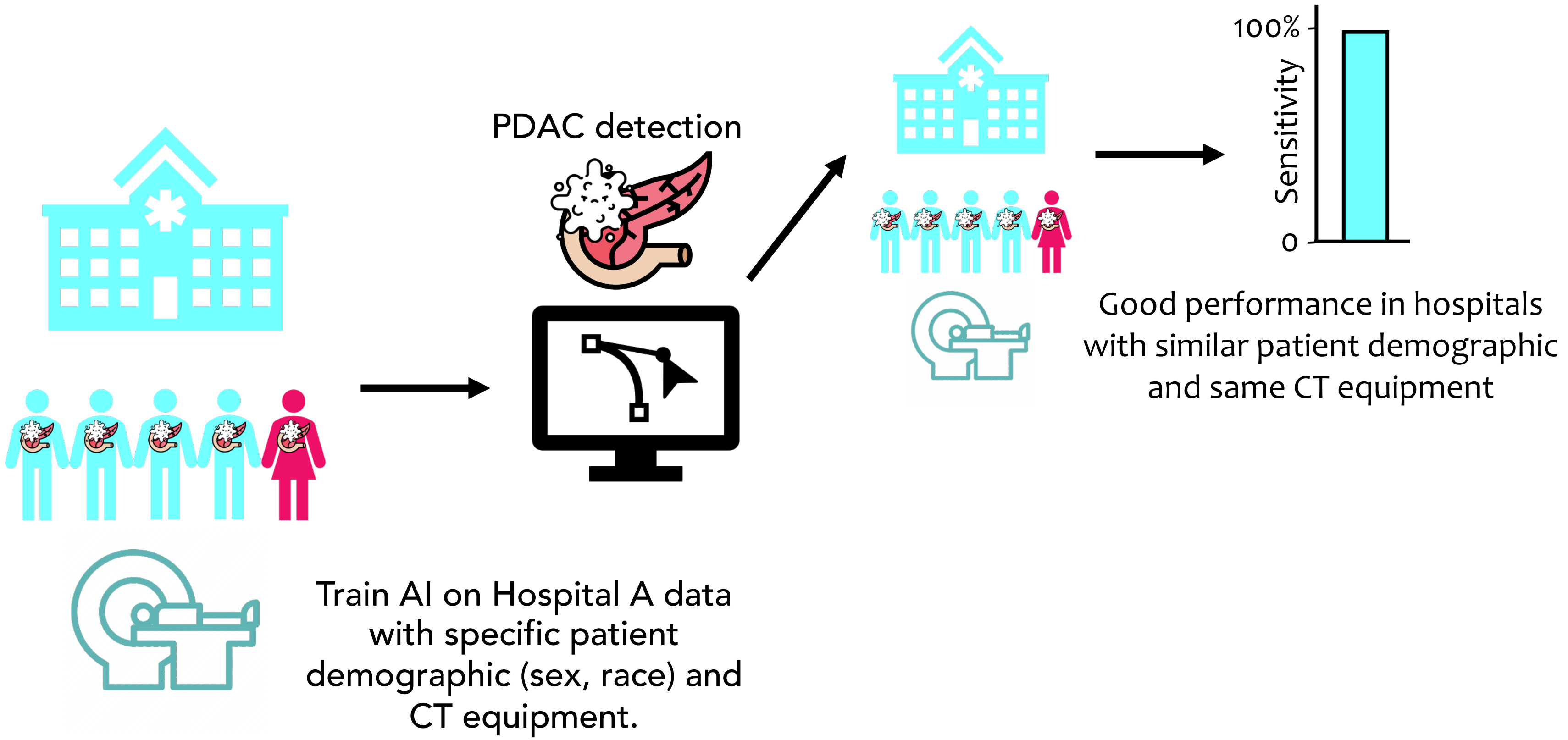
Synthetic Data for Small/Tiny Tumors

- Real-world tumor datasets are often biased.
  - E.g., Tumor location: about 65% of pancreatic tumors arise in the head of the pancreas with the remaining roughly one-third in the body or tail.

- Targeted data augmentation enabled by error analysis and synthetic data (X. Li et al., TMI 2025)
  - The AI often misses small tumors in the body or tail of the pancreas.
  - We can add more synthetic small tumors in these regions during training.

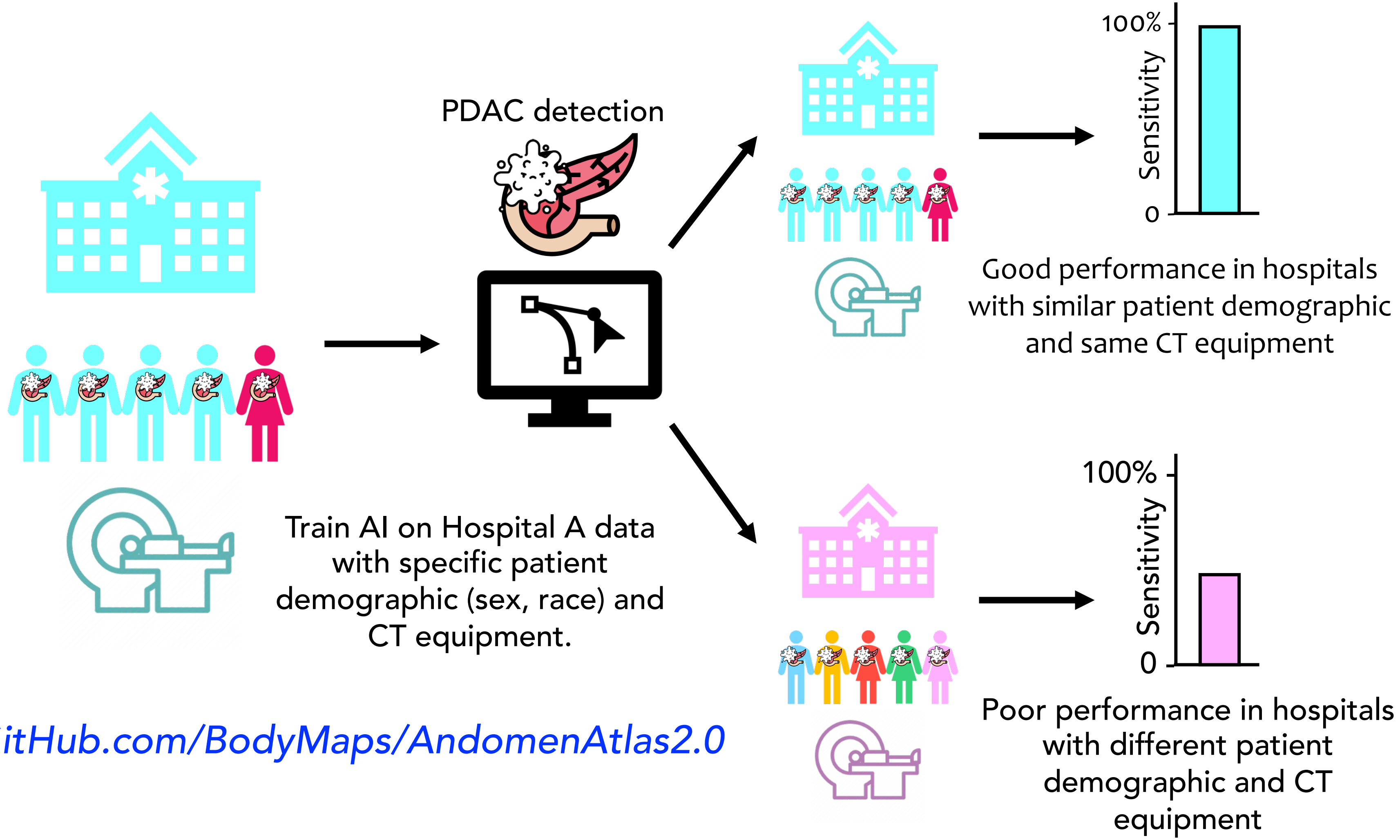
GitHub.com/MrGiovanni/TextoMorph

Synthetic Data for Different Hospitals



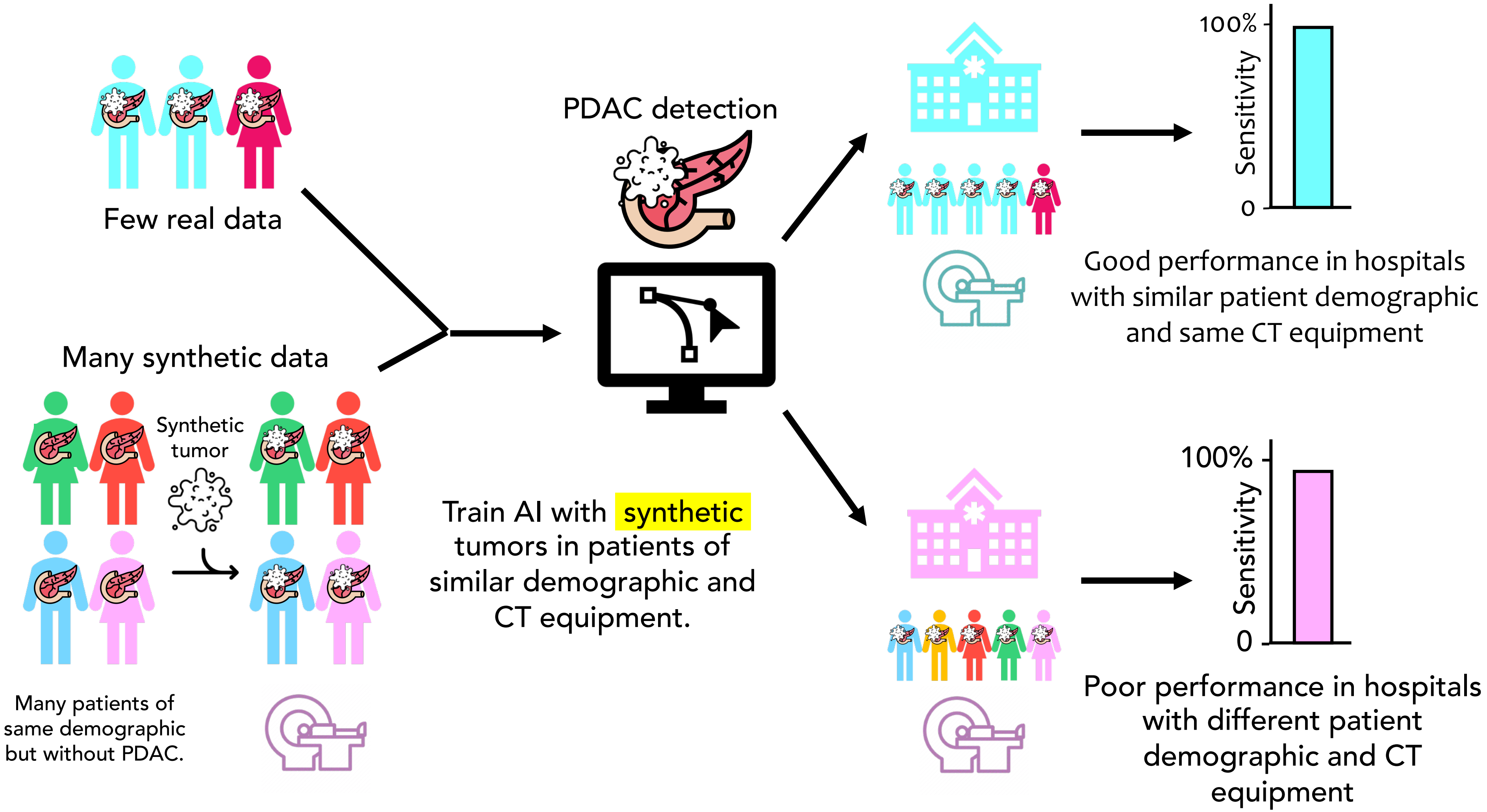
GitHub.com/BodyMaps/AndomenAtlas2.0

Synthetic Data for Different Hospitals



GitHub.com/BodyMaps/AndomenAtlas2.0

Synthetic Data for Different Hospitals



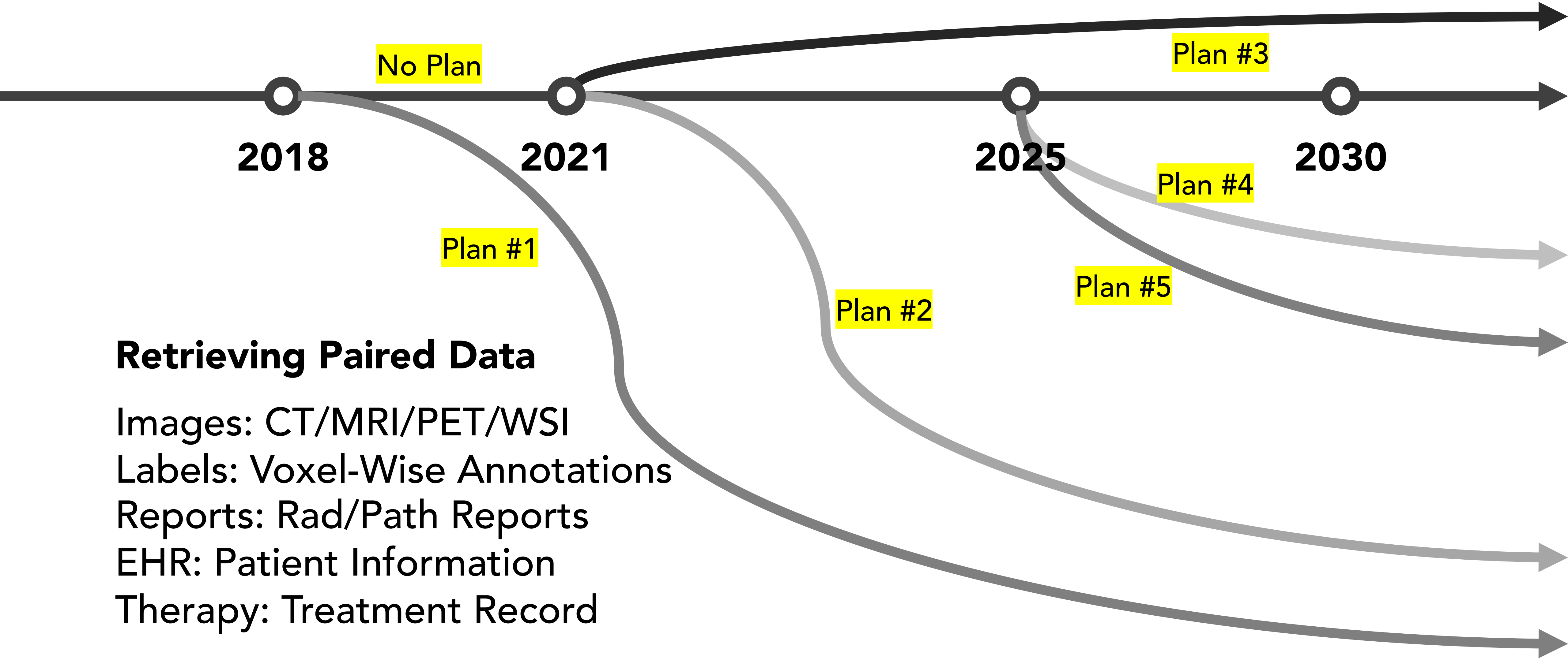
Other Projects & Future Work

- Even bigger datasets to ensure algorithms perform well at most institutions (1-10 Million).
- Pre-diagnostic detection – years before conventional detection in all abdominal organs. (NIH R01 Under Review)
- Understanding cancer development prior to diagnosis through causal models that incorporate causality and treatment. (Yang et al., ICCV 2025)
- Increasing resolution of CT scans using AI techniques. 2MM is almost of the limit of what can be detected. (Lin et al., NeurIPS 2025)

If you had unlimited resources to build a "dream" dataset for AI in cancer research, especially cancer prevention, detection, and treatment, what would it look like?

What modalities, metadata, and labels would you prioritize?

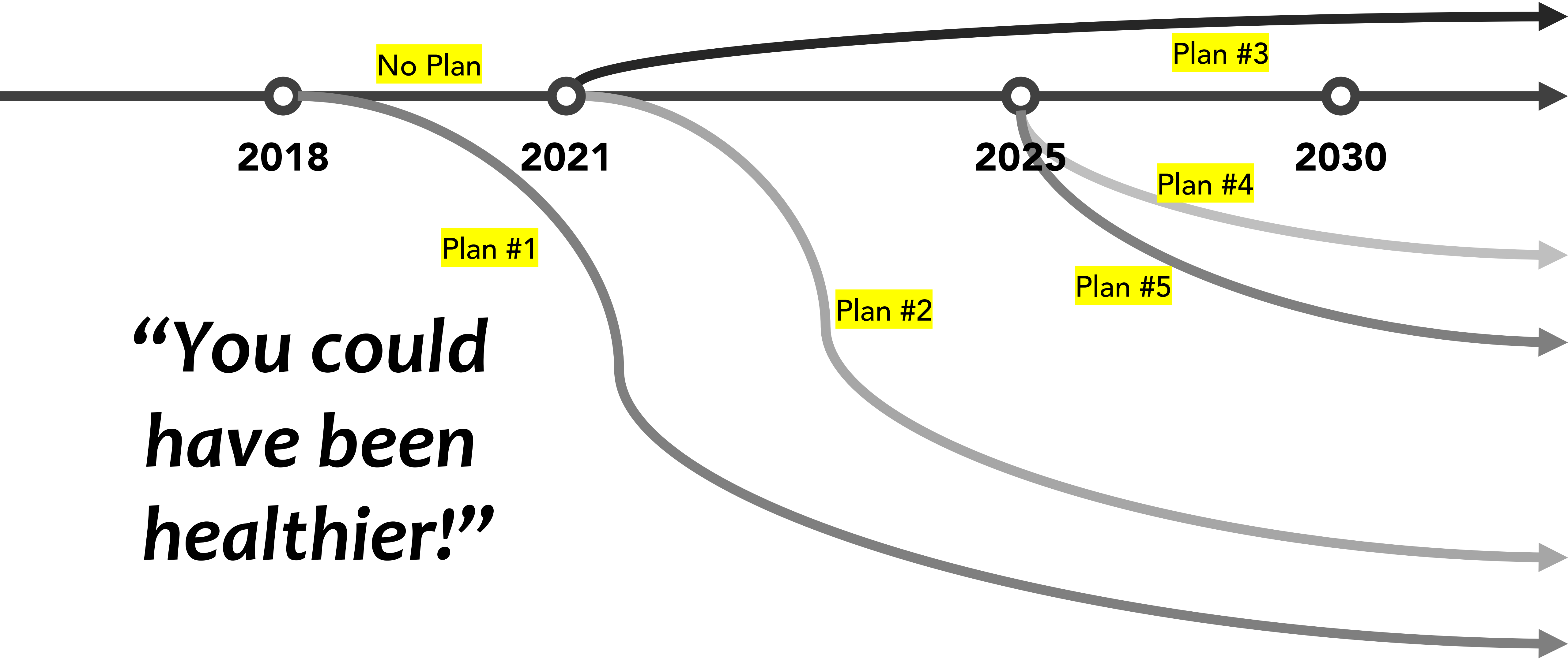
NIH R01: Multimodal+Longitudinal Analysis



Retrieving Paired Data

Images: CT/MRI/PET/WSI
Labels: Voxel-Wise Annotations
Reports: Rad/Path Reports
EHR: Patient Information
Therapy: Treatment Record

NIH **R01**: Multimodal+Longitudinal Analysis



"You could have been healthier!"

The One Possibility

Over 14 Million Futures

Report Generation | Early Cancer Detection | Organ Segmentation | Patient Retrieval

Applications

Algorithms

Research Topics	Featured Achievements	Peer-Reviewed Publications
I. Segmentation Architectures	UNet++, 15,000 Citations	TMI, MIA, ICCV, NeurIPS, RSNA
II. Annotations with Active Learning	AbdomenAtlas, N = 240,000	MIAx2, CVPRx2, MICCAI, NeurIPS, RSNAx12
III.A. Medical Foundation Models	Models Genesis, MICCAI Best Paper Award & MIA Best Paper Award	TPAMI, TMI, MIA, CVPRx2, ICLRx2, ICCV, MICCAIx4, RSNAx12
III.B. Vision & Language	Finalist, MICCAI Best Paper Award	ICCV, MICCAI, ISBI, RSNAx2
III.C. Tumor Synthesis & Generation	Segmentation of 16 Cancer Types	CVPRx2, ICCVx3, MICCAI, RSNAx10

Wishlist for my "dream" dataset

1. Multi-modal data of tumors in multiple organs, including CT of different contrast phases, MRI of different sequences, PET, and their reports
2. Long-term follow-up with clear timelines
3. Large screening + non-cancer cohorts (including healthy and high-risk people)
4. Imaging + pathology linked for the same lesions
5. High-quality outcome labels (recurrence, survival, toxicity)
6. Standardized imaging protocols + scanner metadata
7. Treatment details (drug/radiation/surgery with dates and doses)
8. Multi-omics where feasible, plus repeat blood or DNA samples over time
9. Diverse populations and many hospitals (to reduce bias)
10. Multi-expert annotations with disagreement recorded

Thank you!