

Recent Advances in Medical AI for Cancer and Chronic Disease Screening and Precision Diagnosis and Treatment

Le Lu

le.lu@alibaba-inc.com

Medical AI Lab, DAMO Academy

2024.12

Elysium is a mythical place that means a state of bliss or a place of perfect happiness



0-Current Status of Cancer Screening

Currently, **WHO** and internationally recognized clinical standards have standardized recommended screening methods for only four types of cancer/premalignant lesions, which are also eligible for insurance reimbursement.

- 1. Lung Cancer/Pulmonary Nodules:** Low-Dose Chest CT (LDCT) for high-risk populations such as smokers and individuals with a family history, recommended annually or biennially.
- 2. Breast Cancer:** Mammography for women aged 40 and older or those with a family history, recommended annually or biennially.
- 3. Cervical Cancer:** Pap Smear, a medical procedure for screening cervical cancer in women, requires the collection of cervical cells.
- 4. Colorectal Cancer/Colonic Polyps:** Colonoscopy, starting at age 40-45 for men and 45 for women, recommended every two years.

Other cancers: Although numerous clinical studies exist, the information is inconsistent, and there is currently no universal standard of care, hence no screenings are recommended at this time.

1-MCED Liquid Biopsy?

MCED (Multi-Cancer Early Detection) liquid biopsy encountered an increasing number of difficulties in early multi-cancer screening over nearly 20 years of development, many of which currently seem insurmountable. Why? (Note: This is based on my understanding from the literature, and there are ongoing debates)

- **Blood tests measuring ctDNA and protein sequencing have a low signal-to-noise ratio, or even a very low ratio.** “Because the tissue being sampled is blood, hematologic cancer signals are disproportionately detected. Developers have had to work hard to more stringently exclude cancer like hematologic signals.” [2]
- **For the patient groups that may benefit the most, such as those in the surgically resectable stages (T1, T2), the sensitivity of MCED screening is low, with literature suggesting it is only about 13% to 40%, depending on the type of cancer.** “the most commonly detected solid tumors are late-stage or recurrent cancers.” [2]
- **Patients who test positive cannot be confirmed or diagnosed effectively, leading to challenges in follow-up; there is a high rate of false positives, which imposes a societal burden and causes anxiety in the general population.** “Most patients with a positive test result do not have a cancer identified. This is typical of screening: The low prevalence of cancer means positive tests are less likely to represent clinically relevant disease and more likely to represent false-positive results or overdiagnosis.” [2]
- **Due to intrinsic issues with the computational and biological mechanisms, MCED is fundamentally a problematic issue that AI cannot resolve (GRAIL's Galleri product is based on AI computation [6]). Additionally, the generalizability across multiple centers and locations is poor; cancer targets can change over time and cannot be captured in a timely manner, which adds another layer of complexity to the problem.**

[1] Questions Swirl Around Screening for Multiple Cancers With a Single Blood Test, <https://jamanetwork.com/journals/jama/fullarticle/2816615> JAMA

[2] Assessing the Clinical Utility of Liquid Biopsies Across 5 Potential Indications From Therapy Selection to Population Screening, <https://jamanetwork.com/journals/jamainternalmedicine/article-abstract/2808647> JAMA

[3] Deploying blood-based cancer screening, https://www.science.org/doi/10.1126/science.adk1213?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%20pubmed Science

[4] Why are so many young people getting cancer? What the data say <https://www.nature.com/articles/d41586-024-00720-6#:~:text=The%20prominence%20of%20gastrointestinal%20cancers,contributing%20to%20rising%20case%20rates.>
Nature

[5] Blood Test for Multicancer Detection in Symptomatic Individuals, <https://ascopubs.org/doi/10.1200/PO.23.00305> JCO, most authoritative journal in oncology

[6] Performance of a Cell-Free DNA-Based Multi-cancer Detection Test in Individuals Presenting With Symptoms Suspicious for Cancers, <https://ascopubs.org/doi/10.1200/PO.22.00679> JCO Precision Oncology JCO

2-Using Imaging Means for (Early) Cancer Screening

(Early) cancer screening can be effective if the following three conditions are fully met:

1. Enhanced CT, Enhanced MRI, preferably PET-CT, or PET-MRI.
2. Tumor markers (at least CA199; if financial conditions allow, it's best to screen for major digestive system tumor markers: CA199, CEA, CA125, CA19-9, AFP, CA724) & more...
3. Specialized radiology and surgical oncologists with rich experience in interpreting images and scientifically managing patient follow-ups.

[1] <https://www.q.bio/> "We collect blood, vitals and urine, and perform a whole-body MRI from which extensive quantitative data is derived by our Anatomical Foundation Model. This is all seamlessly combined with your medical history, lifestyle and other data to create your Gemini Digital Twin Dashboard — the most comprehensive view of your health and risks available today."

[2] <https://humanlongevity.com/> Human Longevity Inc. healthnucleus100plus.com <https://www.youtube.com/watch?v=QwS-b-stG7o&t=223s>

[3] <https://rossdawson.com/futurist/companies-creating-future/top-11-longevity-companies-life-extension-immortality/>

[4] CancerUniT: Towards a Single Unified Model for Effective Detection, Segmentation, and Diagnosis of Eight Major Cancers Using a Large Collection of CT Scans. [ICCV 2023](#): 21270-21281

[5] Devil is in the Queries: Advancing Mask Transformers for Real-world Medical Image Segmentation and Out-of-Distribution Localization. [CVPR 2023](#): 23879-23889

3. Do we have a (new) way out?

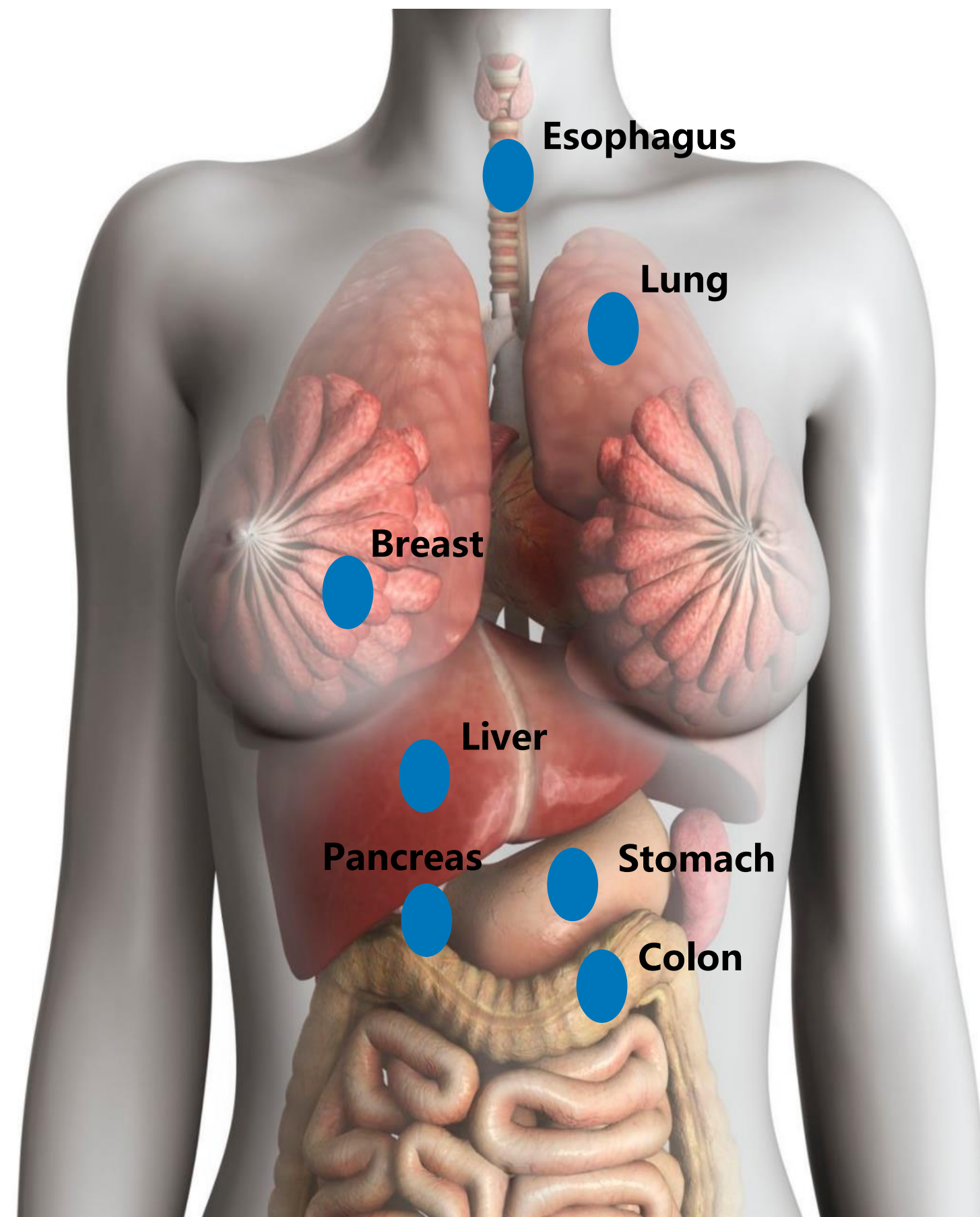
Do we have an early cancer screening solution that is affordable for ordinary people, very effective, and easily accessible?

- ❖ “8+5” multi-cancer and chronic disease screening
using non-contrast CT
- ❖ via “AI + Cloud Computing”

Attempt to solve the "impossible trinity" problem in healthcare



Multi-cancer Early Screening with Non-contrast CT and Deep Learning



AI + Non-contrast CT Technology helps to Achieve Detection of Multiple Cancers in the Early Stages

- Technological Breakthroughs: For the first time in the world that a single CT scan could be used to screen for multiple cancers and diseases.
- Over 10 million times of usage
- 30 Papers & 60 Patents

Cancer Screening and Precision Diagnosis

Pancreatic Cancer

Liver Cancer

Breast Cancer

Esophageal Cancer

Gastric Cancer

Kidney Cancer

Lung Cancer

Colorectal Cancer

Precision Radiotherapy and Prognosis for Cancers

Complete Solution for Identification of benign and malignant **lymph nodes** + Segmentation of lymph nodes

Auto-delineation of target volumes and **organs** covering more than 200 target volumes and organs at risk

Multi-cancer Early Screening with Non-contrast CT and Deep Learning

胸部CT平扫

布局 滚动 调窗 缩放 移动 反片 旋转 左右翻转 上下翻转 播放 放大镜 标注 隐藏/显示 MPR 重置 更多 帮助 AI结果

检查日期: 序列: 1 series 数量: 76 Images

姓名: N2D_PATIENT 性别/年龄: **/** 出生日期: ** PID: id

机构名称: ** studyDate: ** ** Acc No: yes

5.00 mm 76 PCS

疑似(100%)胰腺癌 隐藏AI标注

分类	体积	概率
胰腺癌	6.37	100%
非胰腺癌病变	2.92	0%
正常胰腺	82.86	0%

检查所见 复制

肺结节

- 1) 右肺中叶可见一实性结节影(IM:31)最大直径约6.43mm;
- 2) 右肺下叶可见一实性结节影(IM:40)最大直径约8.12mm;
- 3) 右肺下叶可见一实性结节影(IM:43)最大直径约4.72mm;
- 4) 右肺下叶可见一实性结节影(IM:43)最大直径约6.26mm;
- 5) 左肺下叶可见一实性结节影(IM:45)最大直径约5.63mm;
- 6) 左肺上叶可见一实性结节影(IM:50)最大直径约8.88mm;
- 7) 右肺上叶可见一实性结节影(IM:51)最大直径约5.35mm;
- 8) 右肺上叶可见一实性结节影(IM:52)最大直径约7.17mm;

Ser: 1
Img: 11 / 76
512 x 512
Loc: ** Thick: 5.00 mm

Zoom: 147%
W: 350 L: 50
Lossless / Uncompressed

本产品分析结果仅供医生参考, 非最终临床诊断意见。

(1) Pancreatic cancer screening using non-contrast CT with AI

Article <https://doi.org/10.1038/s41591-023-02640-w>

Large-scale pancreatic cancer detection via non-contrast CT and deep learning

Received: 9 February 2023
Accepted: 12 October 2023
Published online: 20 November 2023

Kai Cao^{1,19}, Yingda Xia^{2,19}, Jiawen Yao^{3,4,19}, Xu Han^{5,19}, Lukas Lambert^{6,19}, Tingting Zhang^{7,19}, Wei Tang^{8,19}, Gang Jin⁹, Hui Jiang¹⁰, Xu Fang¹, Isabella Nogues¹¹, Xuezhou Li¹, Wenchao Guo^{3,4}, Yu Wang^{3,4}, Wei Fang^{3,4}, Mingyan Qiu^{3,4}, Yang Hou¹², Tomas Kovarnik¹³, Michal Vocka¹⁴, Yimei Lu⁸, Yingli Chen⁹, Xin Chen¹⁵, Zaiyi Liu¹⁵, Jian Zhou¹⁶, Chuanmiao Xie¹⁶, Rong Zhang¹⁶, Hong Lu¹⁷, Gregory D. Hager¹⁸, Alan L. Yuille¹⁸, Le Lu², Chengwei Shao¹, Yu Shi¹², Qi Zhang⁵, Tingbo Liang⁵, Ling Zhang² & Jianping Lu¹

nature medicine

[Check for updates](#)

Large-scale pancreatic cancer detection via non-contrast CT and deep learning, *Nature Medicine*, volume 29, pages: 3033 – 3043 (2023)

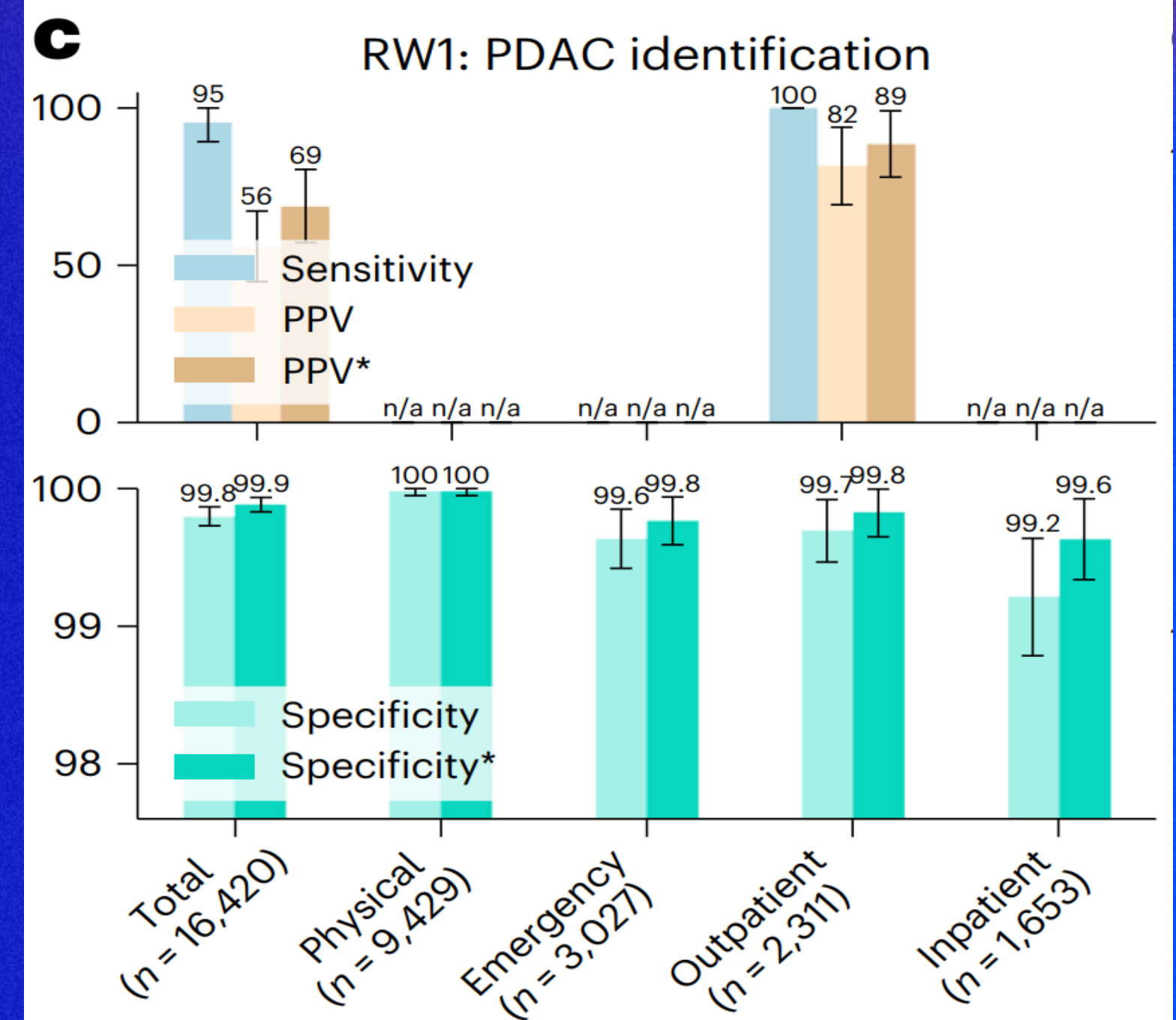
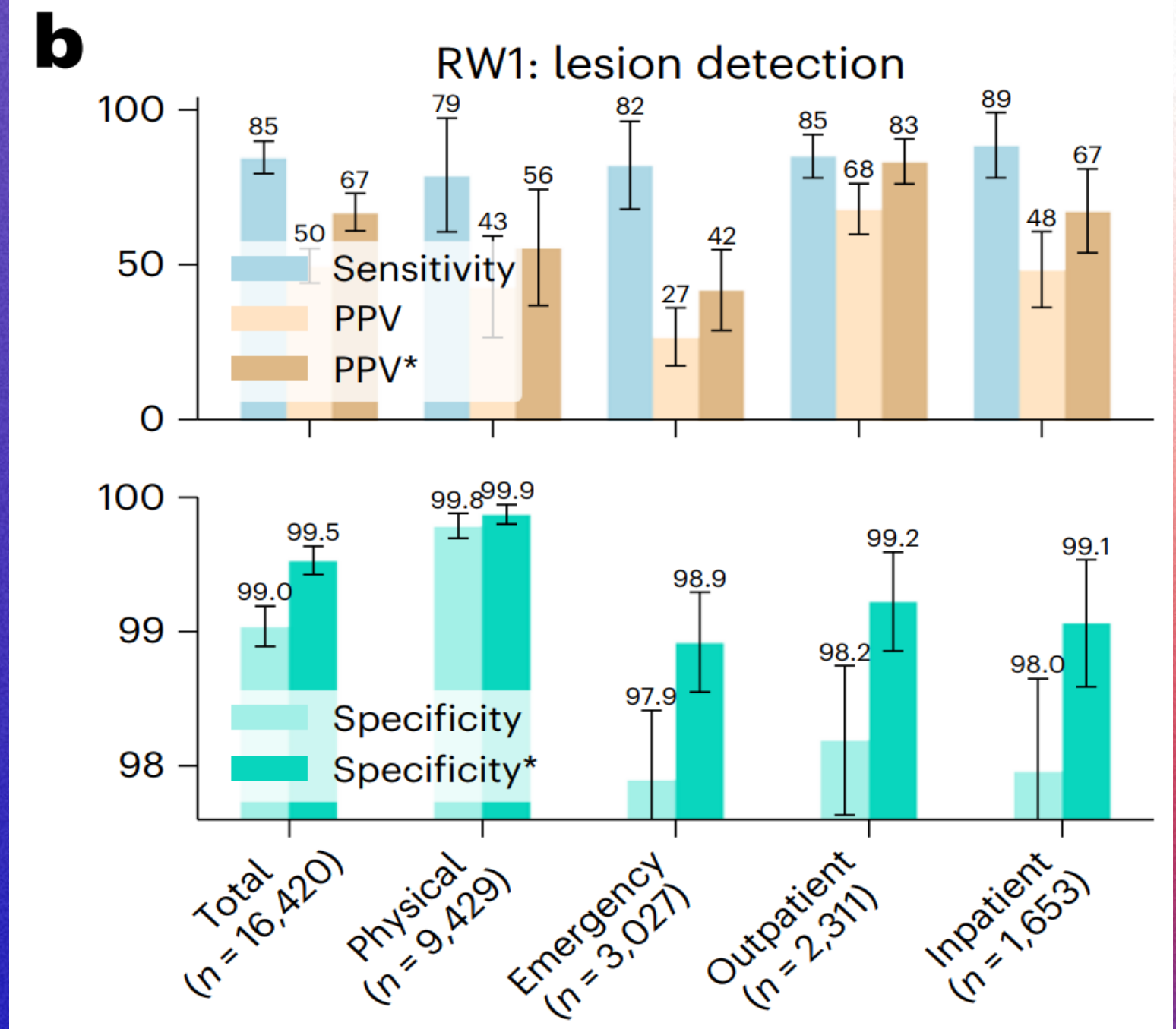
Real-world validation in a health check population revealed a sensitivity of 84.6%, specificity of 99.0%, and a positive predictive value of 56%.

Among hospitalized patients, the sensitivity was highest at 88.6%,

The positive predictive value (PPV) for opportunistic screening in hospitalized and emergency patients was between 0.75 and 0.85.

A multidisciplinary team (MDT) found that 51% (80 out of 156) of the false positives were due to peripancreatic diseases.

The adjusted specificity for tumor detection was 99.5%, and the adjusted specificity for identifying PDAC was 99.9%.

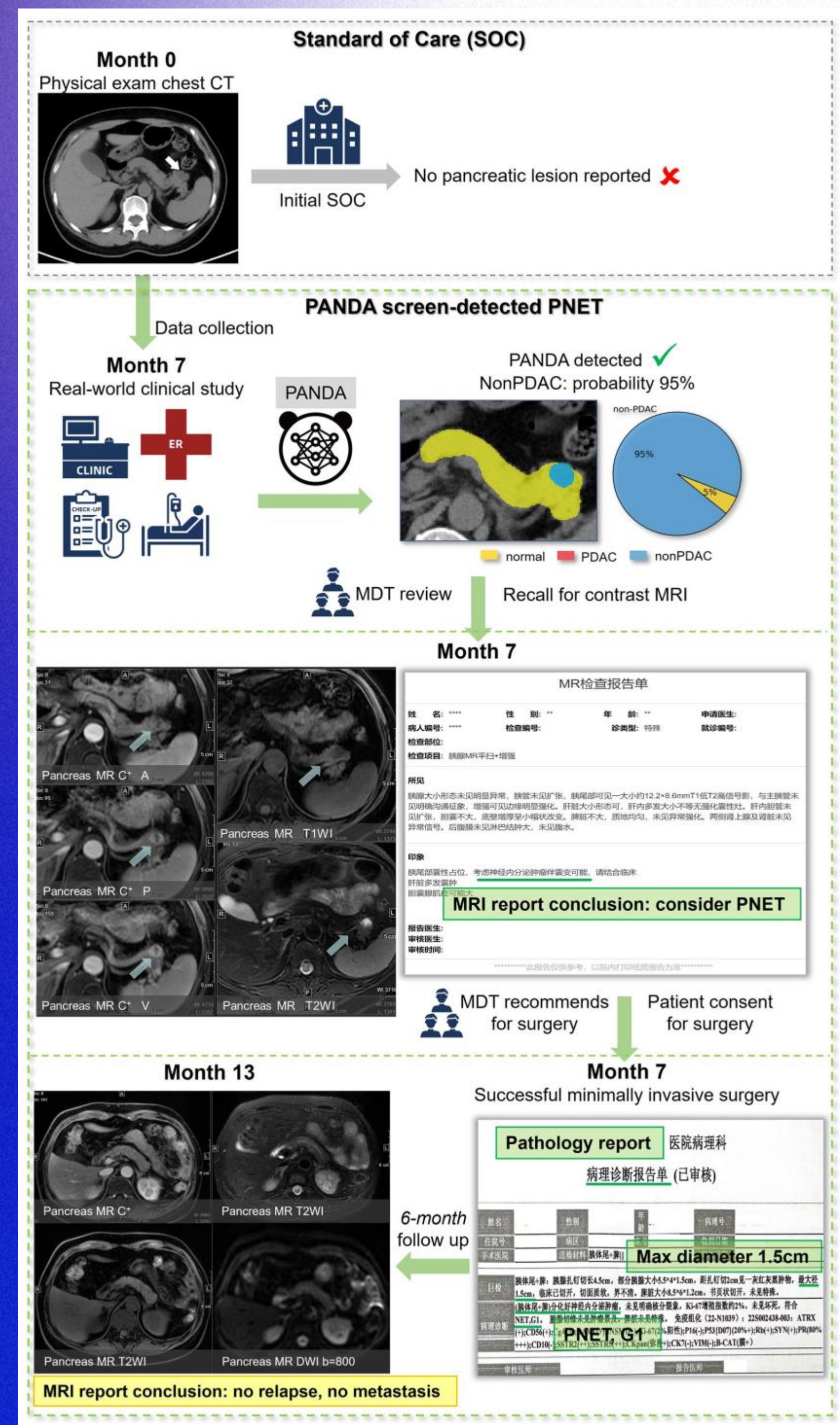


(1) Validation – Changing Clinical Pancreatic cancer screening

Non-contrast CT + AI Real-World Practice

- ✓ The PANDA model detected **26 clinically missed pancreatic lesions [1]**
 - 1 case of PDAC, 1 case of PNET, 3 cases of IPMNs, 1 case of metastatic cancer, 6 cases of pancreatitis, 1 case of peripancreatic tumor, and 13 cases of SCN/cysts (ranging from 10 to 33 mm).
 - One patient was followed up and underwent contrast-enhanced MRI, followed by minimally invasive surgery with the aim of achieving a cure for a **G1 PNET** sized at 1.5 cm [1].
- With opportunistic screening using the PANDA++ model, one of our collaboration hospitals in Shanghai identified **five PDAC patients** measuring 2 cm or less among over 6200 consecutive patients (all hospitalized for other reasons) within 12 days;
- **16 PDAC patients** detected from 24000+ patients with chest CT scans by AI, 11 of them could be surgically resected, 3 out them already done with surgery removal.

[1] Cao, Xia, Yao, et al., Large-scale pancreatic cancer detection via non-contrast CT and deep learning, *Nature Medicine*, volume 29, pages:3033 – 3043 (2023)
 [2] Kleeff & Ronellenfitsch, AI and imaging-based cancer screening: getting ready for prime time, *Nature Medicine*, volume 29, pages:3002 – 3003 (2023)



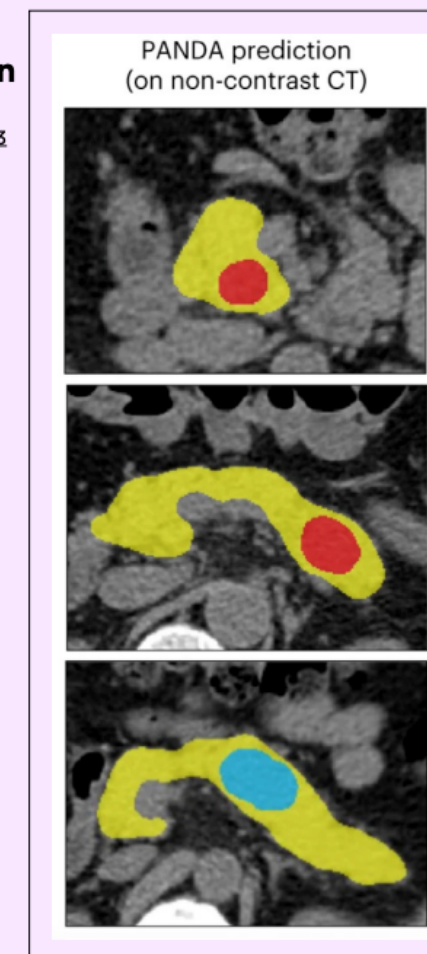
5.2 AI in Medicine 12

Notable Medical Systems	12
SynthSR	12
Coupled Plasmonic Infrared Sensors	14
EVEscape	15
AlphaMissence	17
Human Pangenome Reference	18
Clinical Knowledge	19
MedQA	19
Highlighted Research: GPT-4 Medprompt	20
Highlighted Research: MediTron-70B	22
Diagnosis	23
Highlighted Research: CoDoC	23
Highlighted Research: CT Panda	24
Other Diagnostic Uses	25
FDA-Approved AI-Related Medical Devices	26
Administration and Care	28
Highlighted Research: MedAlign	28

Highlighted Research:
CT Panda

Pancreatic ductal adenocarcinoma (PDAC) is a particularly lethal cancer, often detected too late for surgical intervention. Screening for PDAC in asymptomatic individuals is challenging due to its low prevalence and the risk of false positives. This year, a Chinese research team developed **PANDA** (pancreatic cancer detection with artificial intelligence), an AI model capable of efficiently detecting and classifying pancreatic lesions in X-rays (Figure 5.2.16). In validation tests, PANDA surpassed the average radiologist in sensitivity by 34.1% and in specificity by 6.3% (Figure 5.2.17). In a large-scale, real-world test involving approximately 20,000 patients, PANDA achieved a sensitivity of 92.9% and a specificity of 99.9% (Figure 5.2.18). AI medical tools like PANDA represent significant advancements in diagnosing challenging conditions, offering cost-effective and accurate detection previously considered difficult or prohibitive.

PANDA detection
Source: Cao et al., 2023
Figure 5.2.16



PANDA vs. mean radiologist on multicenter validation (6,239 patients)
Source: Cao et al., 2023 | Chart: 2024 AI Index report

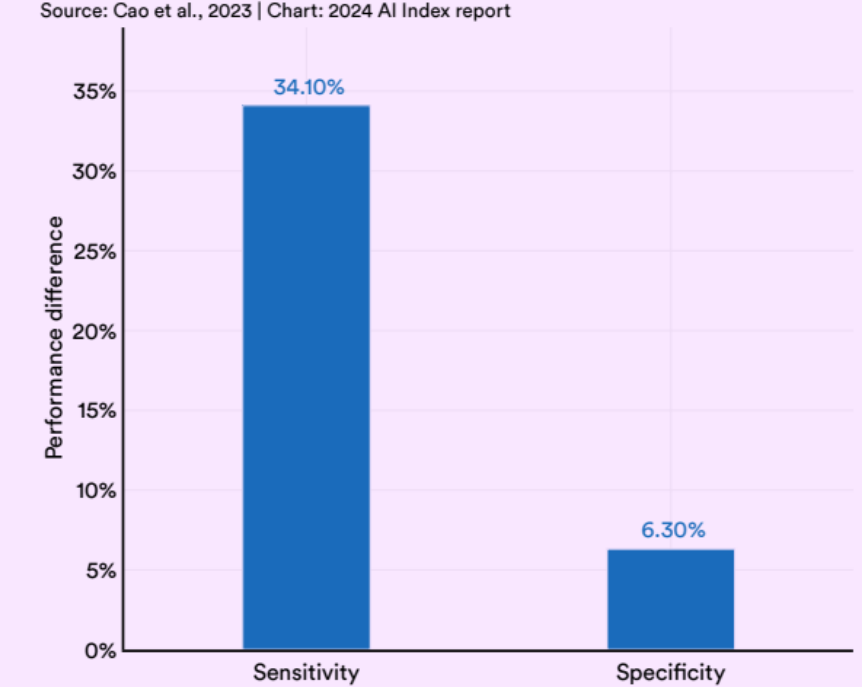


Figure 5.2.17

PANDA performance on real-world multi-scenario validation (20,530 patients)
Source: Cao et al., 2023 | Chart: 2024 AI Index report

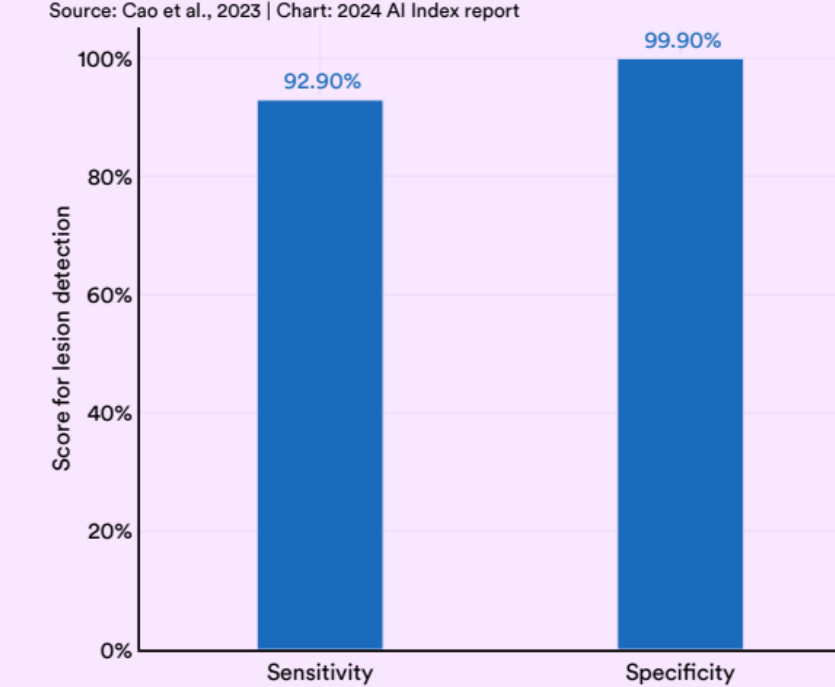
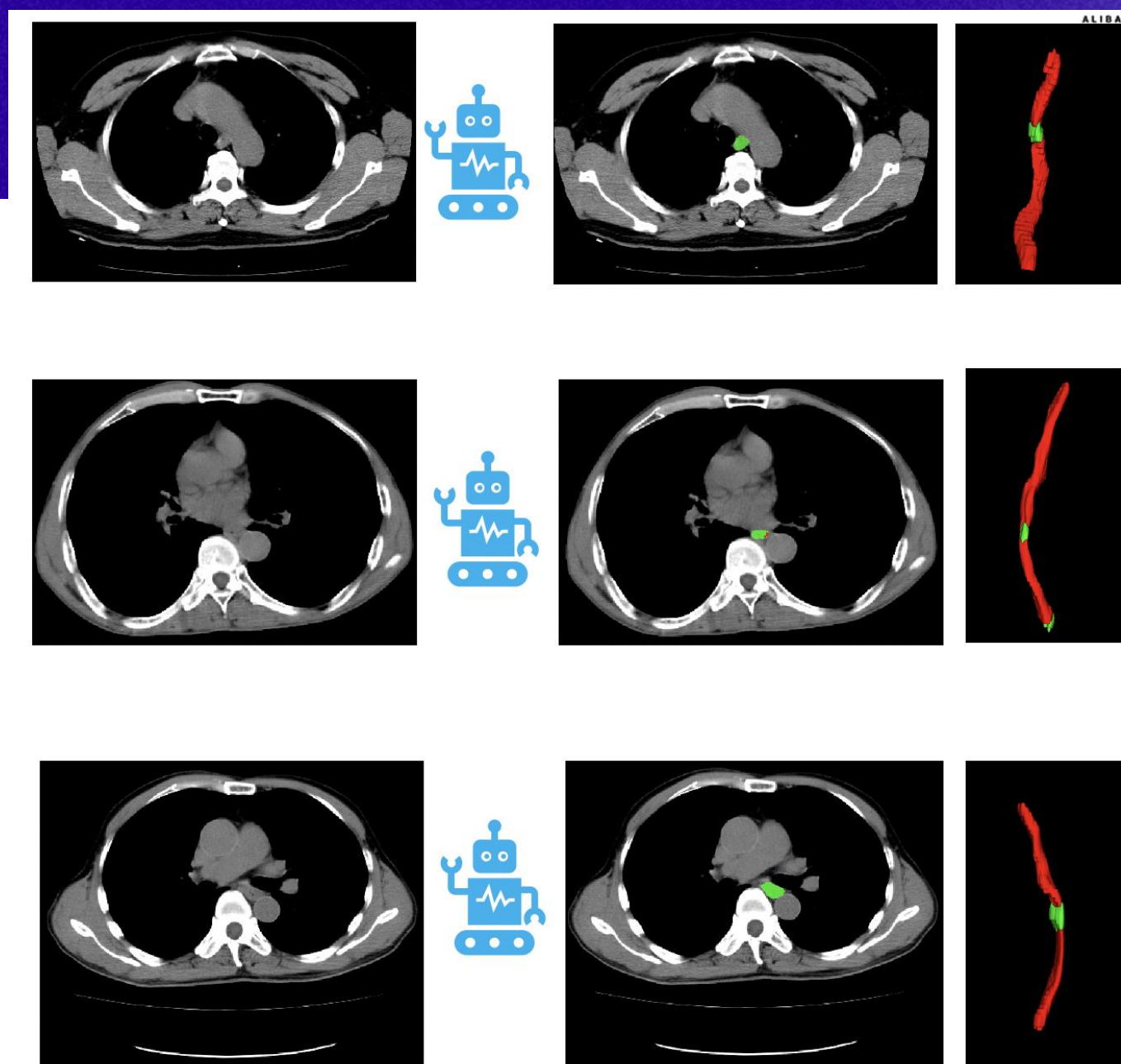
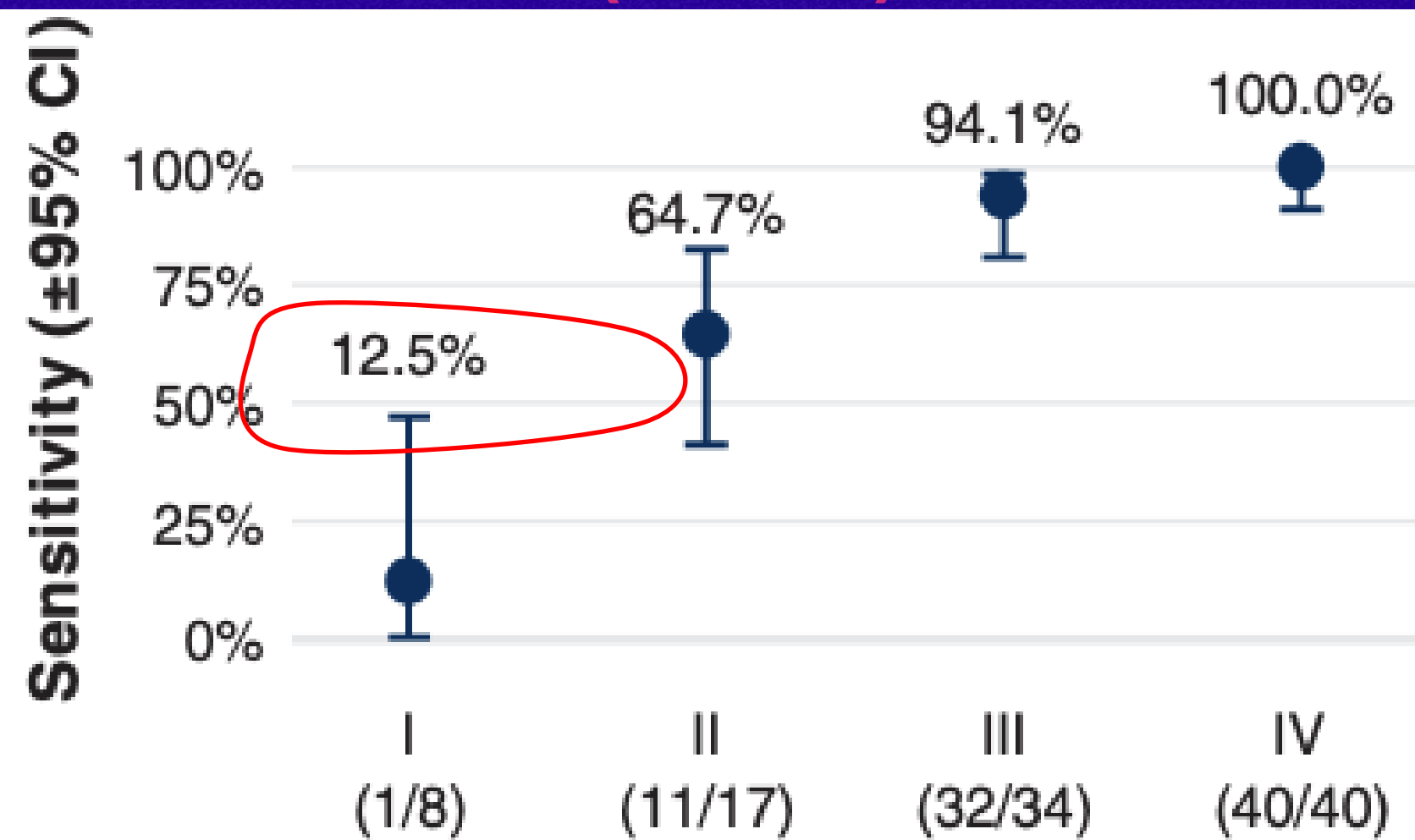


Figure 5.2.18

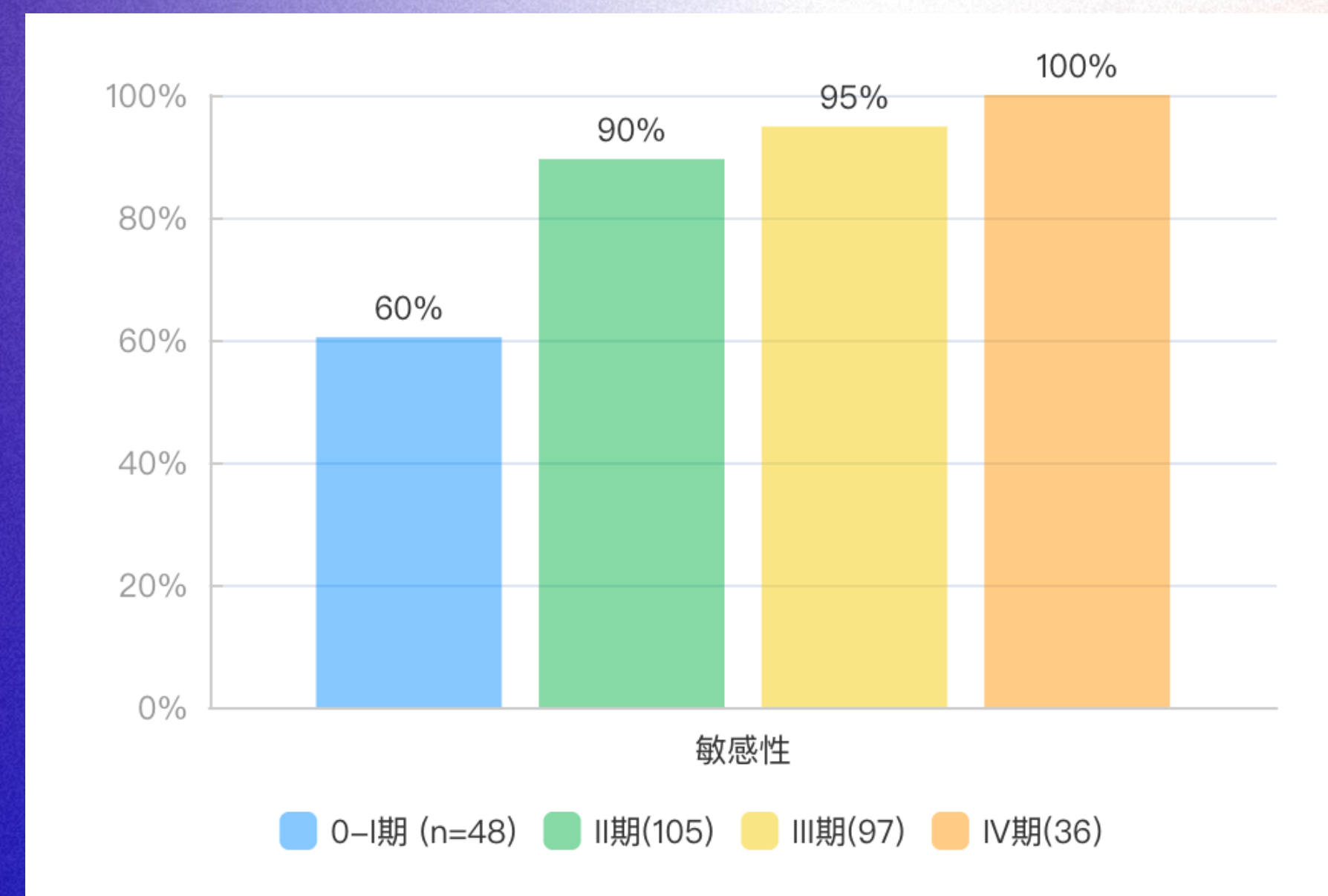
(2) CT screening for Esophageal cancer



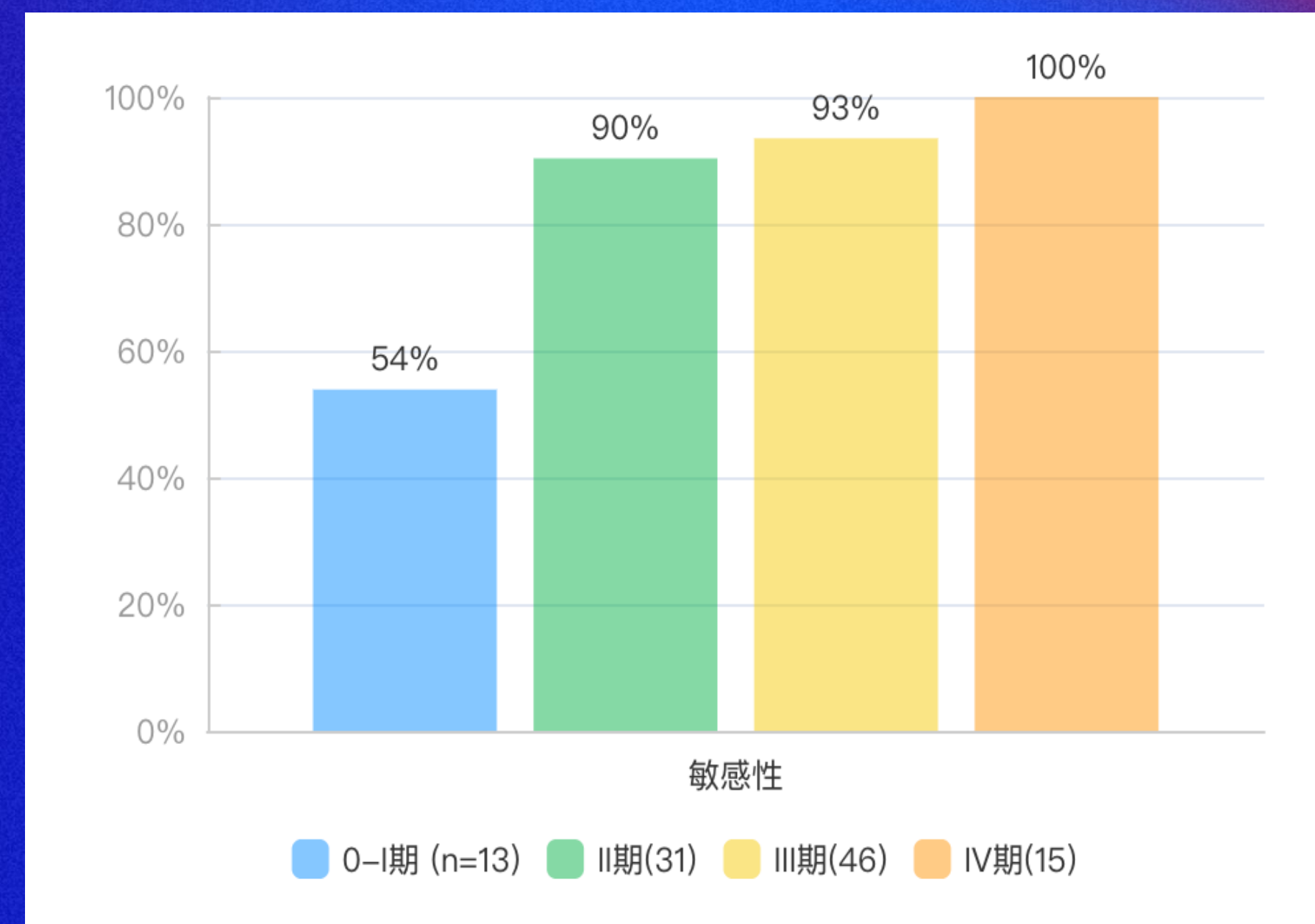
The doctors were unable to detect esophageal cancer on 3 non-contrast CTs, but the model correctly identified it (confirmed by enhanced CT and endoscopy).

- ✓ The sensitivity for stage I esophageal cancer reported by MCED in AO is 12.5%, and for stage II is 64.7%. In our results, the sensitivity is 59% for stage I and 90% for stage II (with 99.1% specificity).
- ✓ We are even able to effectively detect T0 (in collaboration with a top hospital in Sichuan, China).
- ✓ From May 8 to July 23, 25,756 patients underwent AI screening for esophageal cancer, with 7 cases confirmed by pathology; 97 patients are being followed up.

Hospital A

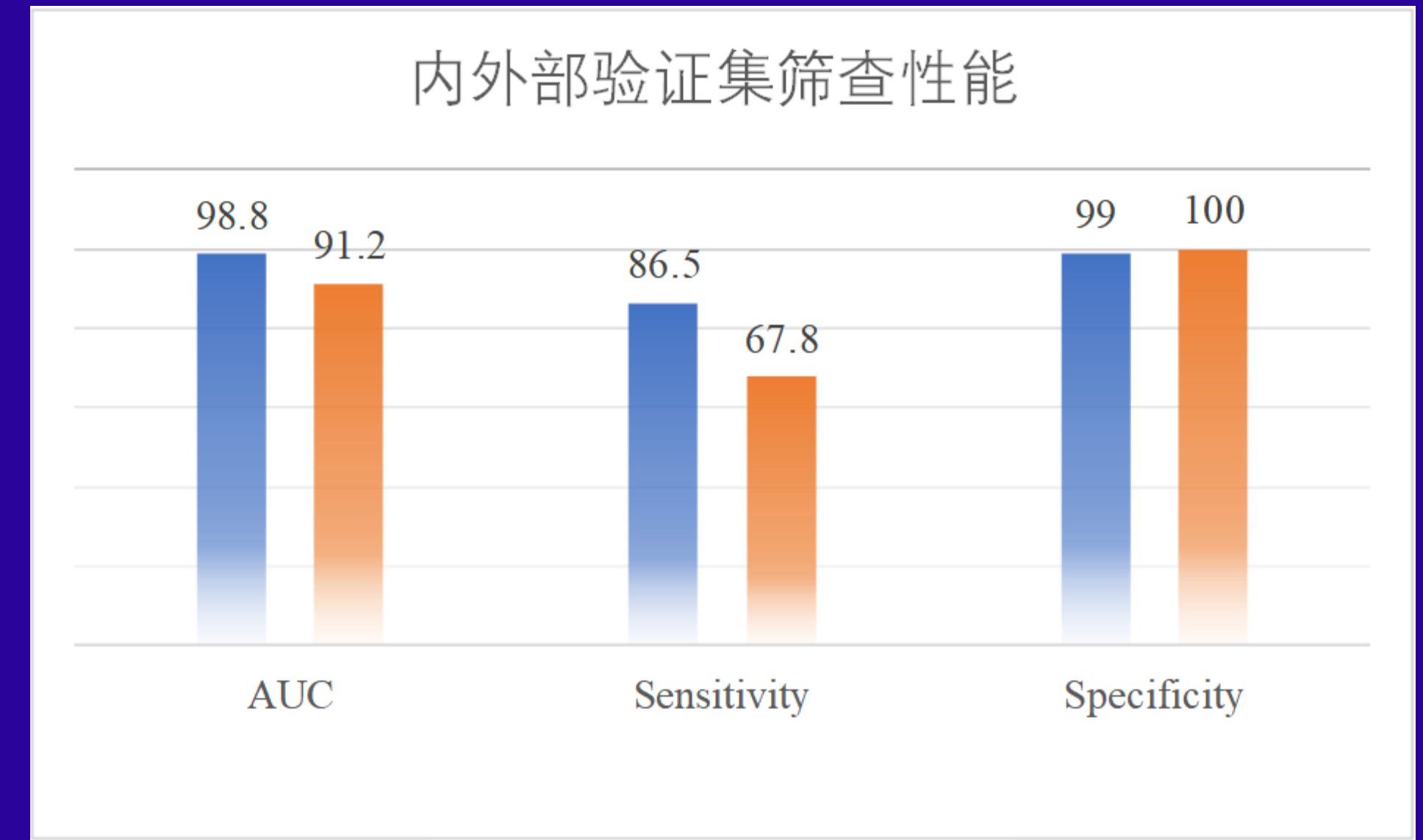


Hospital B



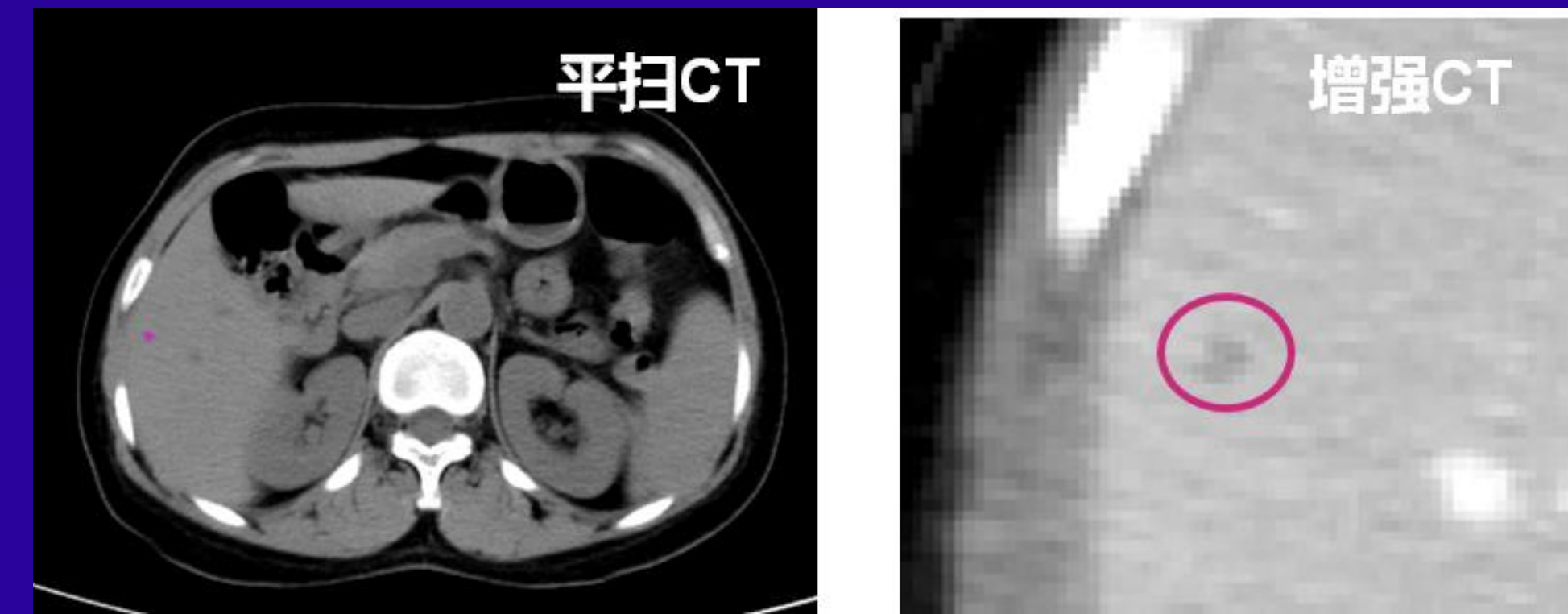
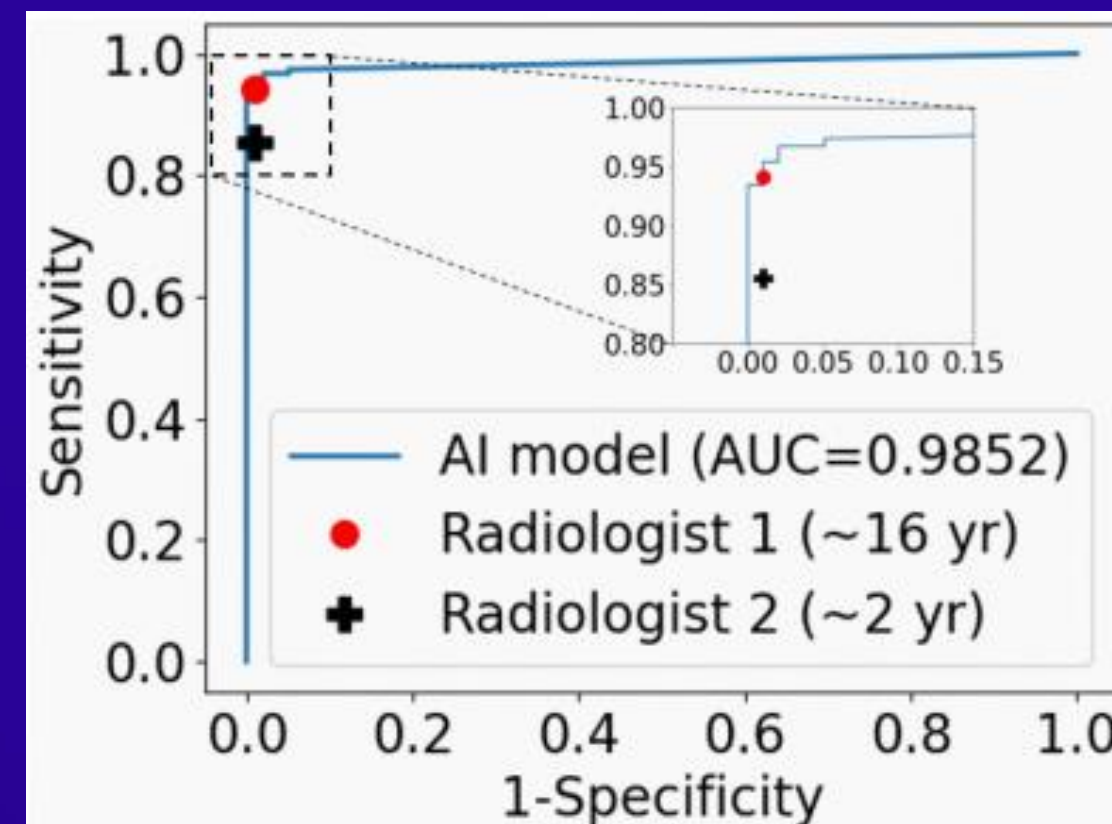
(3) Liver cancer screening using non-contrast CT

- A screening AUC of 0.99 on the internal testing set with a specificity of 99% and sensitivity exceeding 85%.
- Two real-world validation studies, under a specificity of 99%, reporting sensitivities of 87.6%, 78.3%, respectively.
- From June 14 to July 9, AI screening found 161 positive cases, where **155 cases later confirmed by imaging or pathology**; six positive patients are followed up with 5 malignant tumors.



Reader study

	Sens.	Spec.	3-class Acc.
Radiologist 1	94.1	99.0	90.8
Radiologist 2	85.5	99.0	72.0
PLAN	96.7	98.0	91.3



A case of missed diagnosis of a metastatic tumor based on **non-contrast CT**, 4 months faster than clinical discovery.

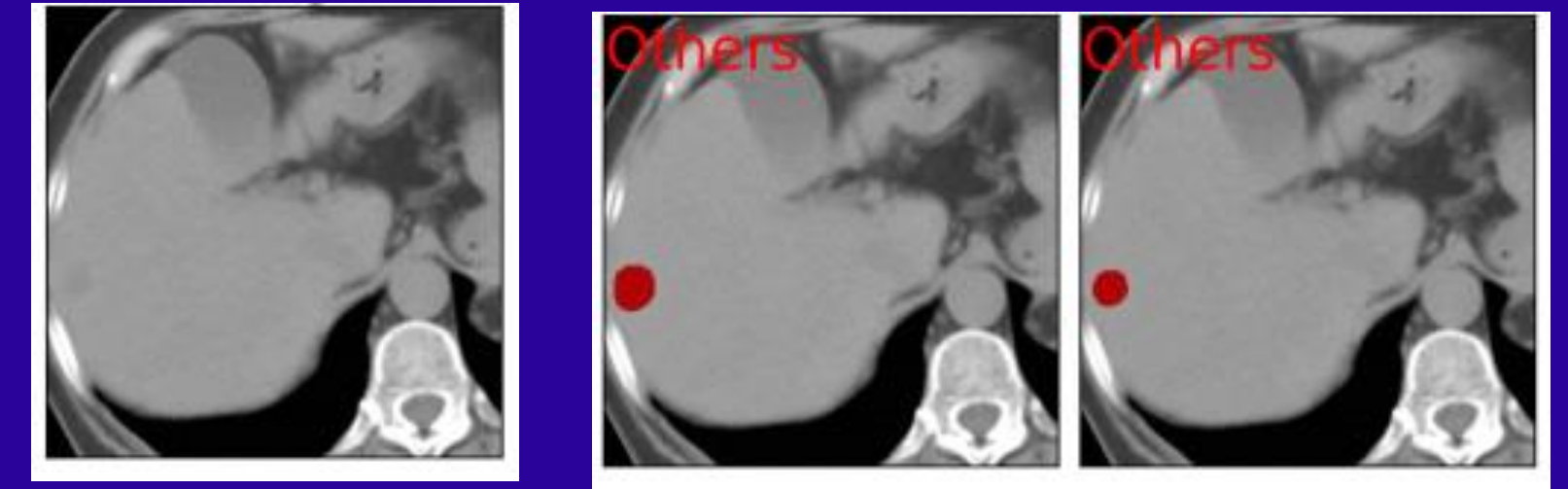
[1] "Liver Tumor Screening and Diagnosis in CT with Pixel-Lesion-Patient Network", MICCAI 2023
 [2] "A Two-streamed Network Approach for Effective Liver Tumor Screening on Non-Contrast CT using Label Distillation", 2024
 [3] "Improving Liver Tumor Identification in Non-contrast CT Scans using Label Distillation", RSNA 2024
 [4] "Automatic Liver Tumor Diagnosis in Contrast-Enhanced CT Scans with Variable Input Phases", RSNA 2024
 [5] "Automatic Liver Tumor Screening and Differential Diagnosis in CT Using Pixel-Lesion-Patient Network with Reader Study and External Validation", Oral, RSNA 2023
 [6] "Accurate Liver Tumor Detection on Non-contrast CT Scans via Annotation-Efficient Semi-Supervised Learning (#14687)", RSNA 2022

(3) Liver tumor diagnosis using contrast-enhanced CT (8 sub-types)

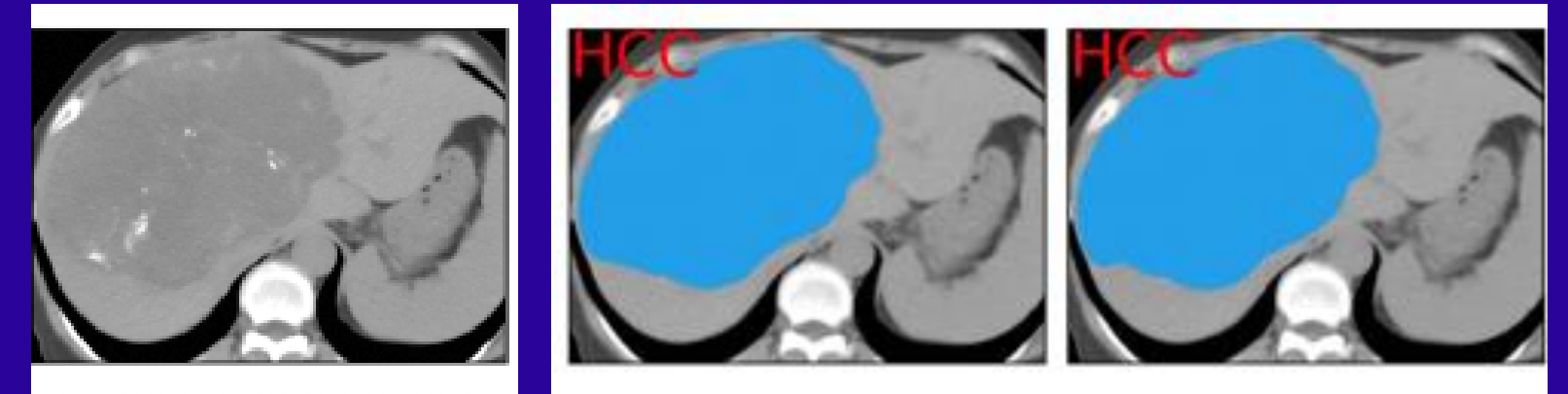
- Task: On multi-phase enhanced CT, screen for liver tumors (such as those with cirrhosis and tumors in other organs)
- If present, differentiate the categories of liver tumors

AUC	HCC	ICC	Meta	Heman	FNH	Cyst	Calc	others	Average
Internal	0.951	0.850	0.951	0.979	0.970	0.948	0.963	0.874	0.936
External	0.934	0.857	0.960	0.963	0.946	0.944	0.919	0.619	0.893

Case 1



Case 2



- Challenges: Differential diagnosis of malignant categories (HCC, ICC, metastasis) and detection of subtle lesions.
- In a reader study with 150 samples, accuracy comparable to that of experienced physicians.
- Two difficult cases were sent by physicians, with algorithm diagnoses similar to those by experienced doctors using MRI.

Reader study	8-subtype accuracy
Junior radiologist (2 yrs exp)	40.5%
Senior radiologist (16 yrs exp)	75.6%
Our algorithm	75.6%

	Algorithm	Senior doc.	Pathology
Hard case1	“others” or meta	Possibly meta of neuroendocrine tumor	N/A
Hard case2	Meta or ICC	ICC or meta	ICC

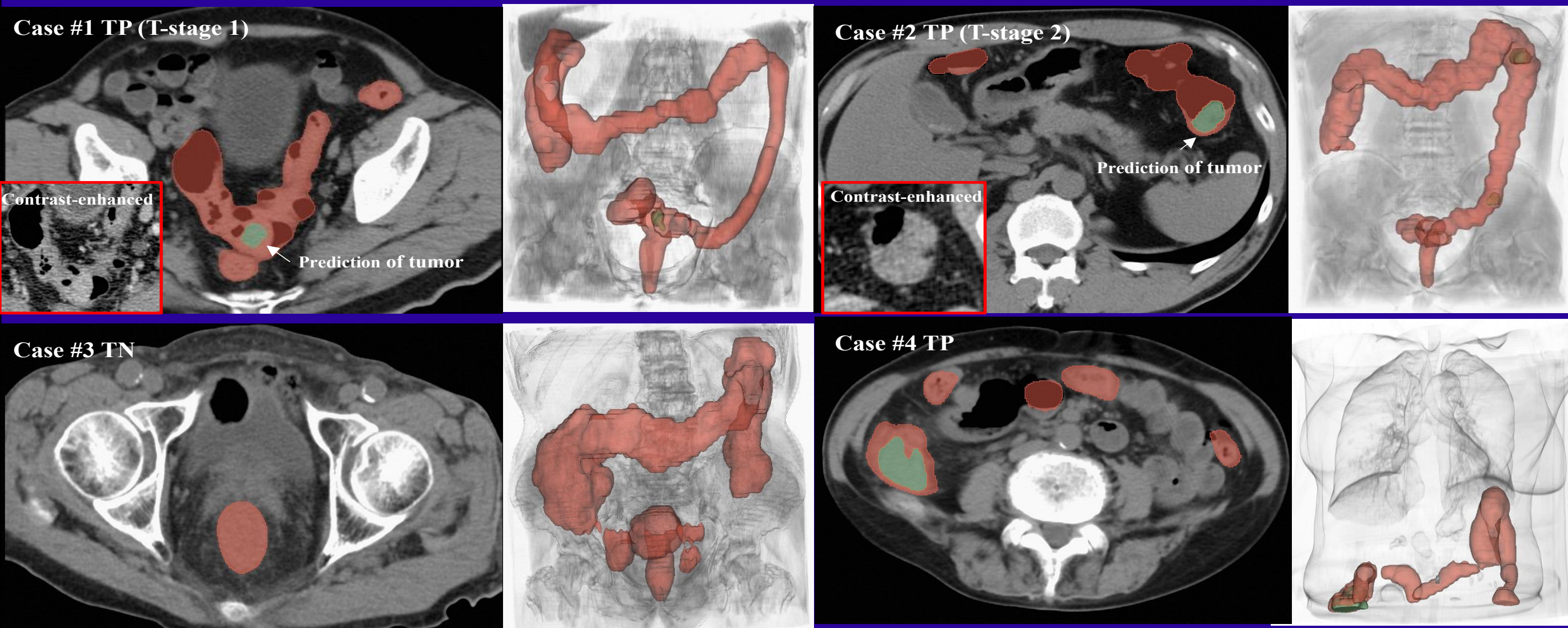
True Labels	HCC	ICC	Meta	Heman	FNH	Cyst	Calc	Others
HCC	123	0	3	1	0	0	0	1
ICC	9	9	5	0	0	0	0	1
Meta	12	6	274	0	0	6	0	2
Heman	3	0	1	100	0	1	0	1
FNH	3	0	1	0	10	0	0	1
Cyst	0	0	4	1	0	179	0	1
Calc	0	0	0	1	0	2	7	4
Others	10	1	10	0	2	6	0	32
	HCC	ICC	Meta	Heman	FNH	Cyst	Calc	Others

[1] “LIDIA: Precise Liver Tumor Diagnosis on Multi-Phase Contrast-Enhanced CT via Iterative Fusion and Asymmetric Contrastive Learning”, MICCAI 2024

[2] "Automatic Liver Tumor Diagnosis in Contrast-Enhanced CT Scans with Variable Input Phases", RSNA 2024

[3] "Automatic Liver Tumor Screening and Differential Diagnosis in CT Using Pixel-Lesion-Patient Network with Reader Study and External Validation", Oral, RSNA 2023

(4) Colorectal Cancer Screening using Non-Contrast CT



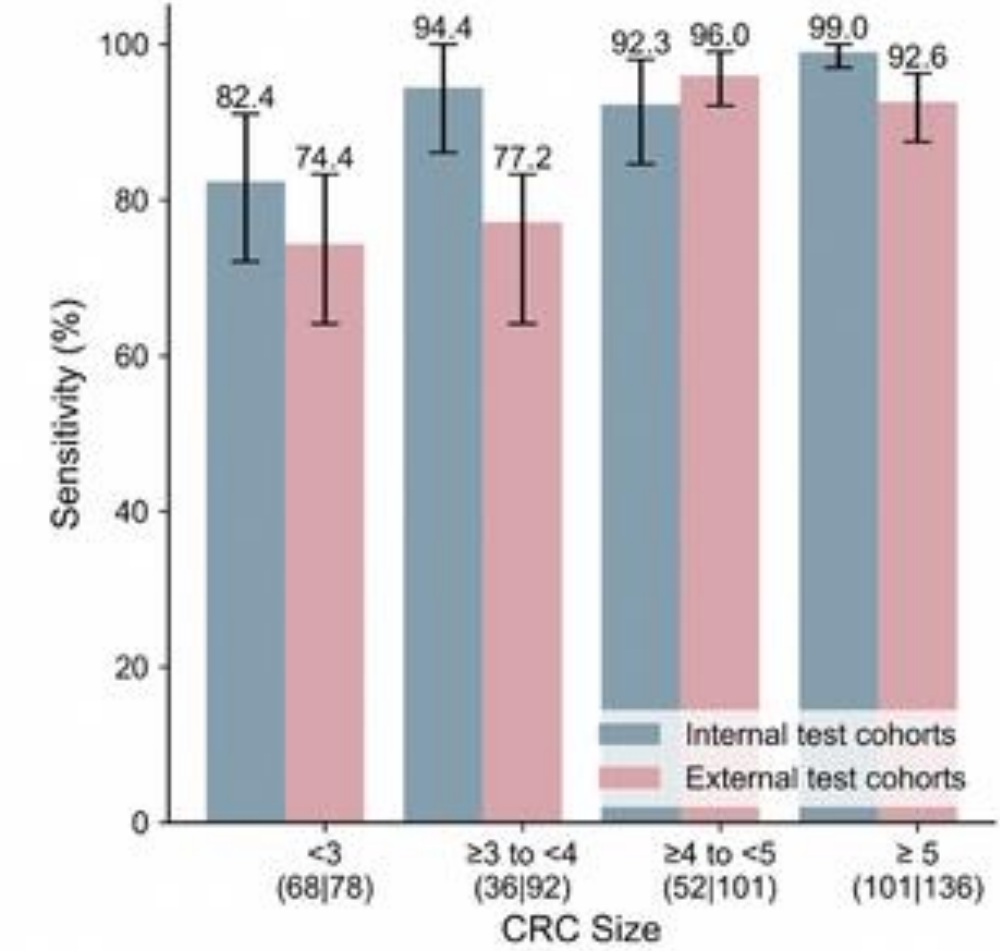
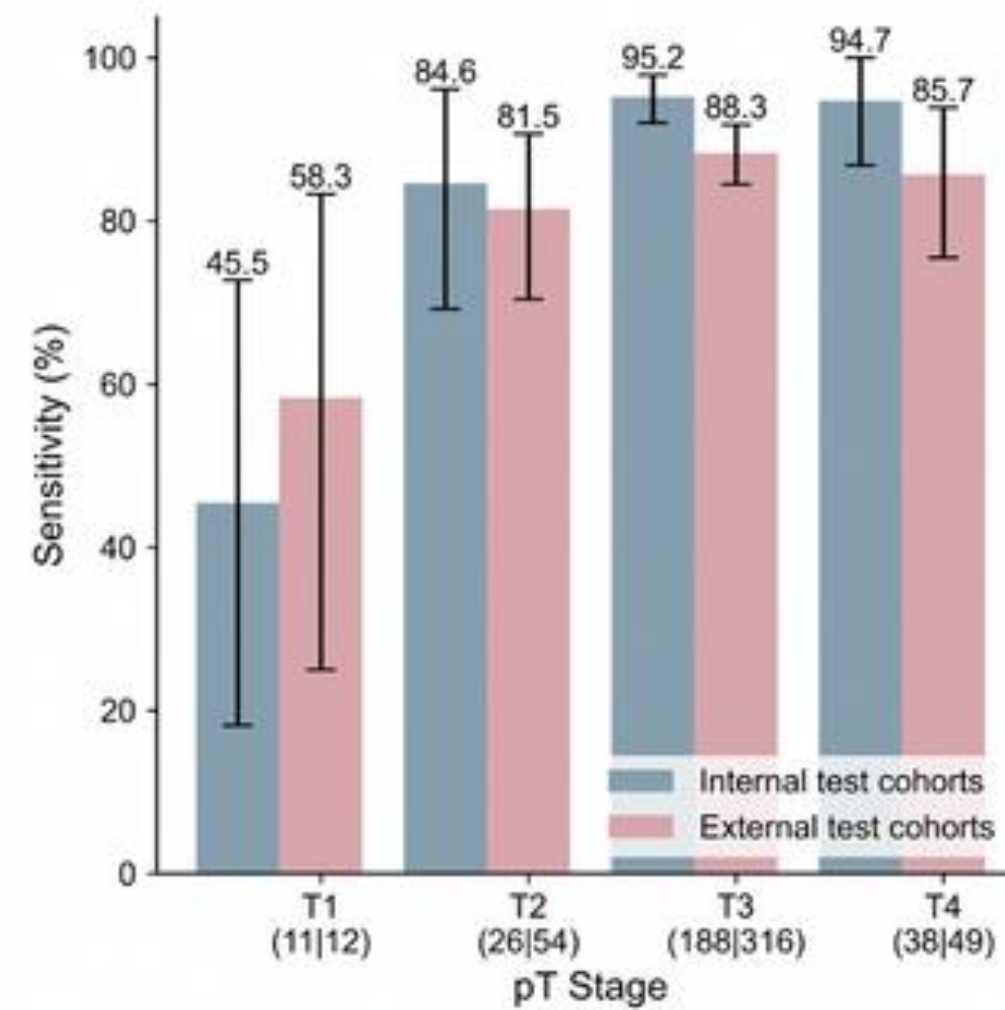
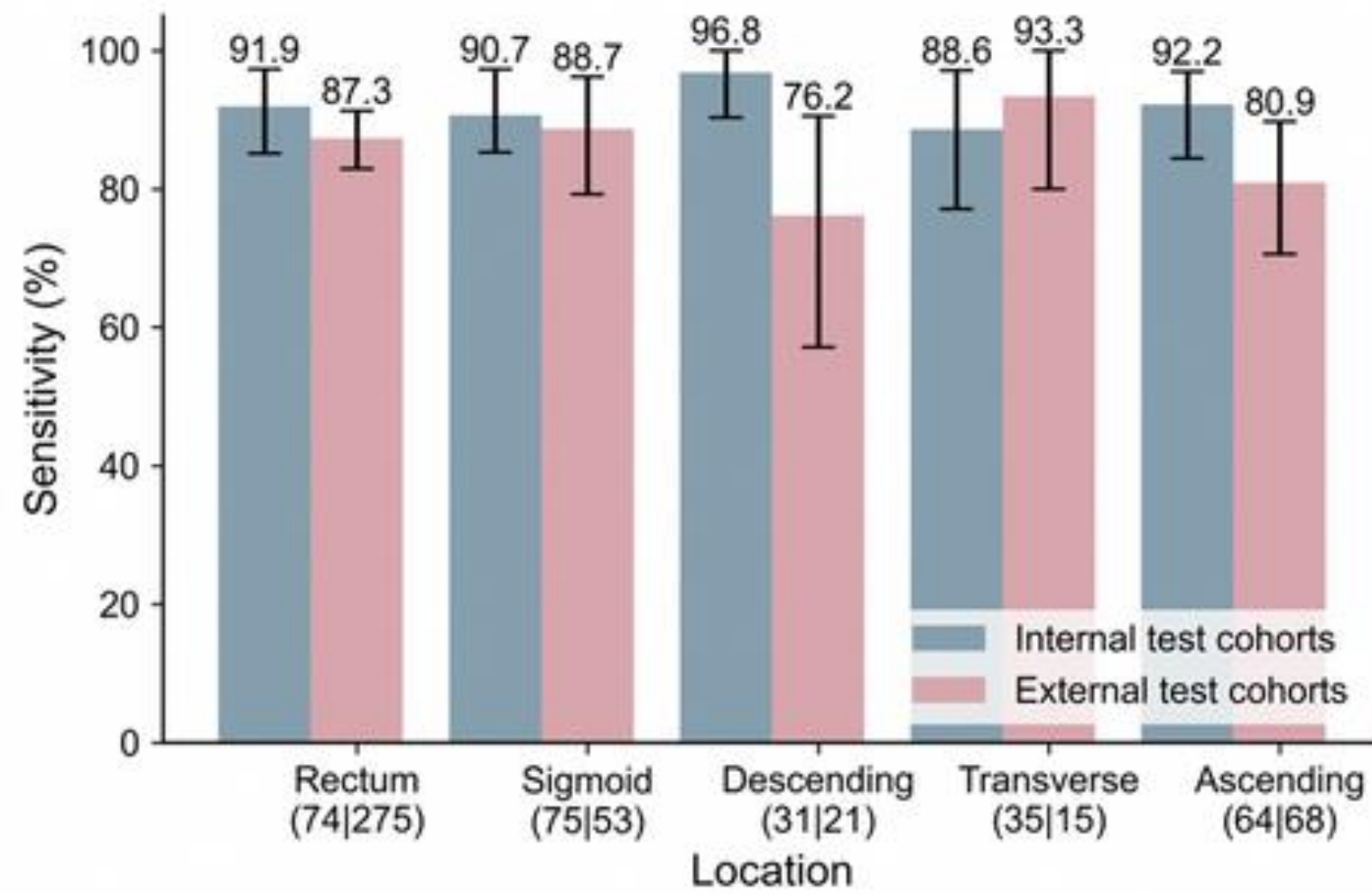
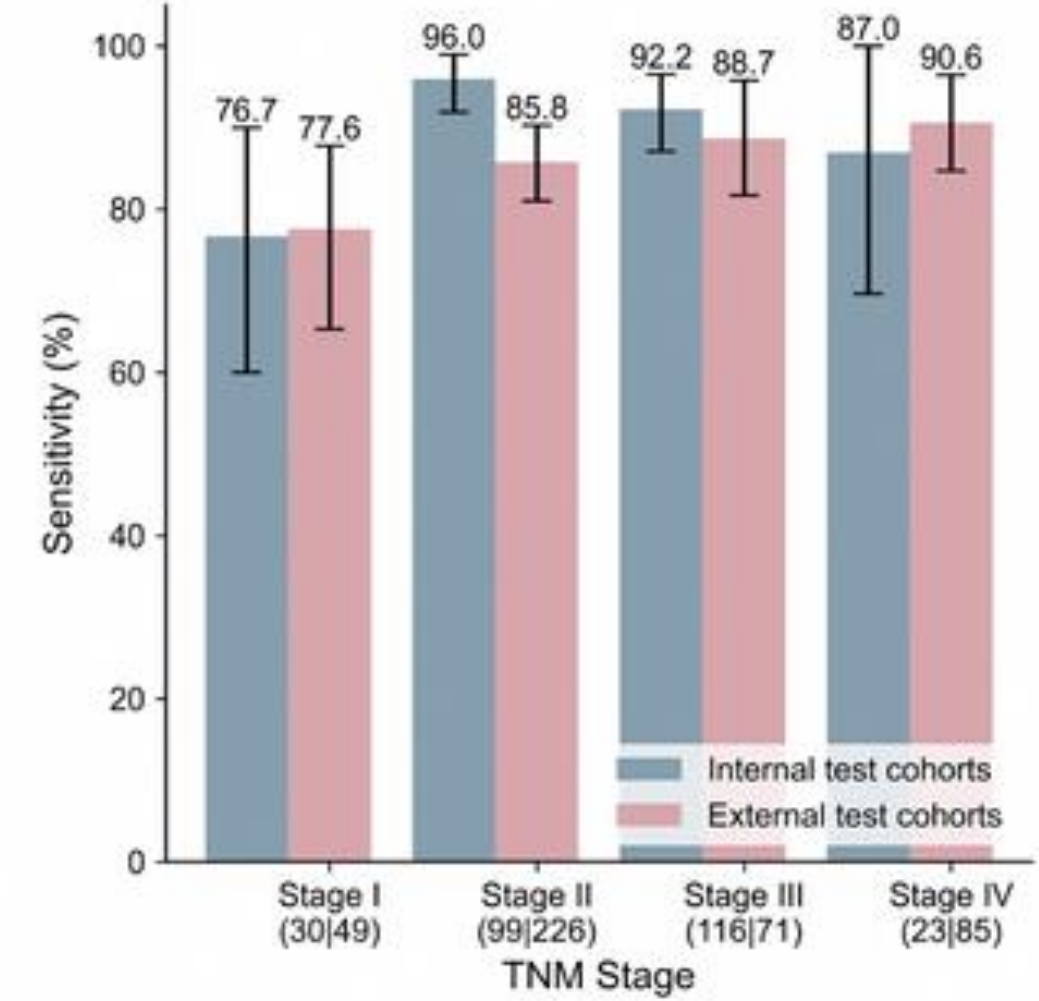
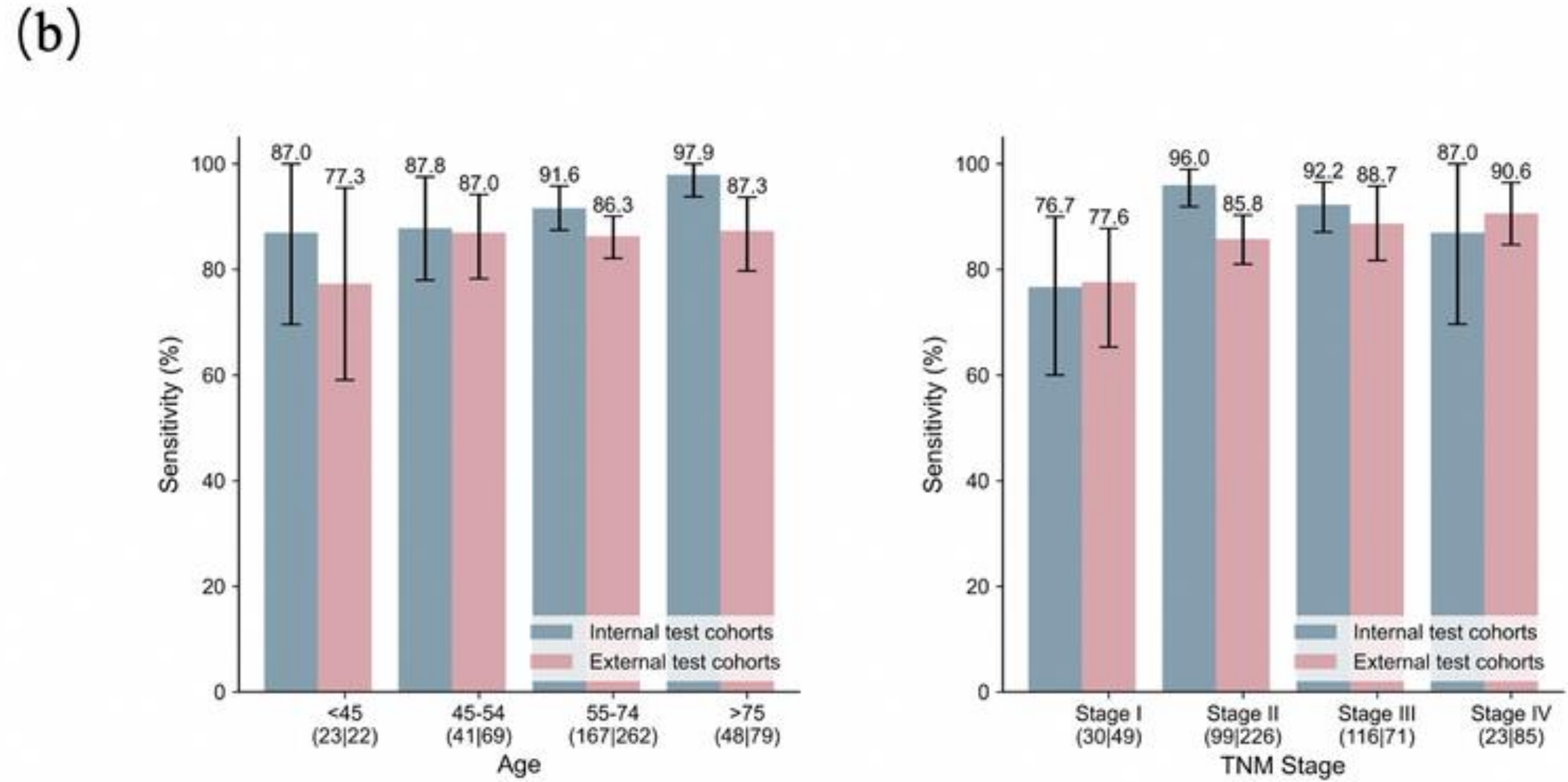
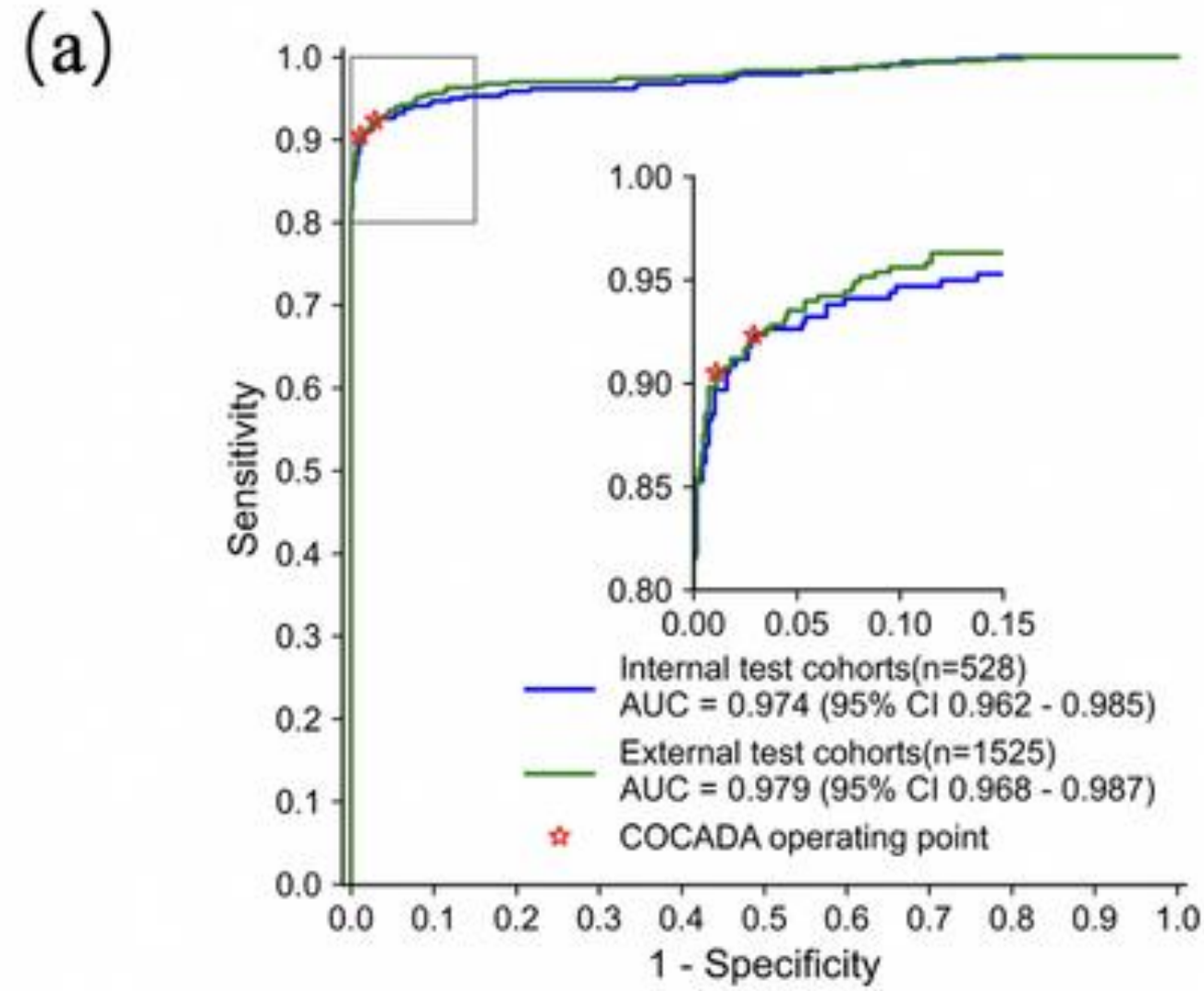
AUC, sensitivity, and Specificity of CRC detection on internal, external center cohorts and real-world

	AUC	Sensitivity (%)	Specificity (%)
Internal Test Cohorts	0.992 (0.987 - 0.996)	91.8 (88.5-94.9)	98.8 (97.3-100.0)
External Test Cohorts (abdominal CT)	0.978 (0.969 - 0.987)	86.3 (83.3-89.6)	99.5 (99.1-99.9)
min	0.967 (0.942-0.987)	88.8 76.5 (83.6-93.4)	99.4 (98.5-100.0)
max	0.996 (0.992-0.999)	89.5 (82.1-96.0)	99.6 (98.2-100.0)
External Test Cohorts (chest CT)	-	75.0 (45.5-100.0)	99.9 (99.8-100.0)
Real-World clinical evaluation	-	88.2 (79.1-96.3)	99.5 (99.3-99.6)
Chest	4 missed diagnoses,	-	99.8 (99.7-99.9)
Abdomen	11 valuable discoveries	88.2 (79.1-96.3)	98.0 (97.3-98.6)

	Sensitivity	Specificity
Chest CT	72%	99.96%
Abdominal CT	90%	99%

Real world evaluation: From May 8 to July 23, AI screening for colon cancer was conducted on 32,739 patients. Among the positive cases, there were 21 cases confirmed by imaging and 39 cases confirmed by pathology; there were 56 positive patients still being followed up.

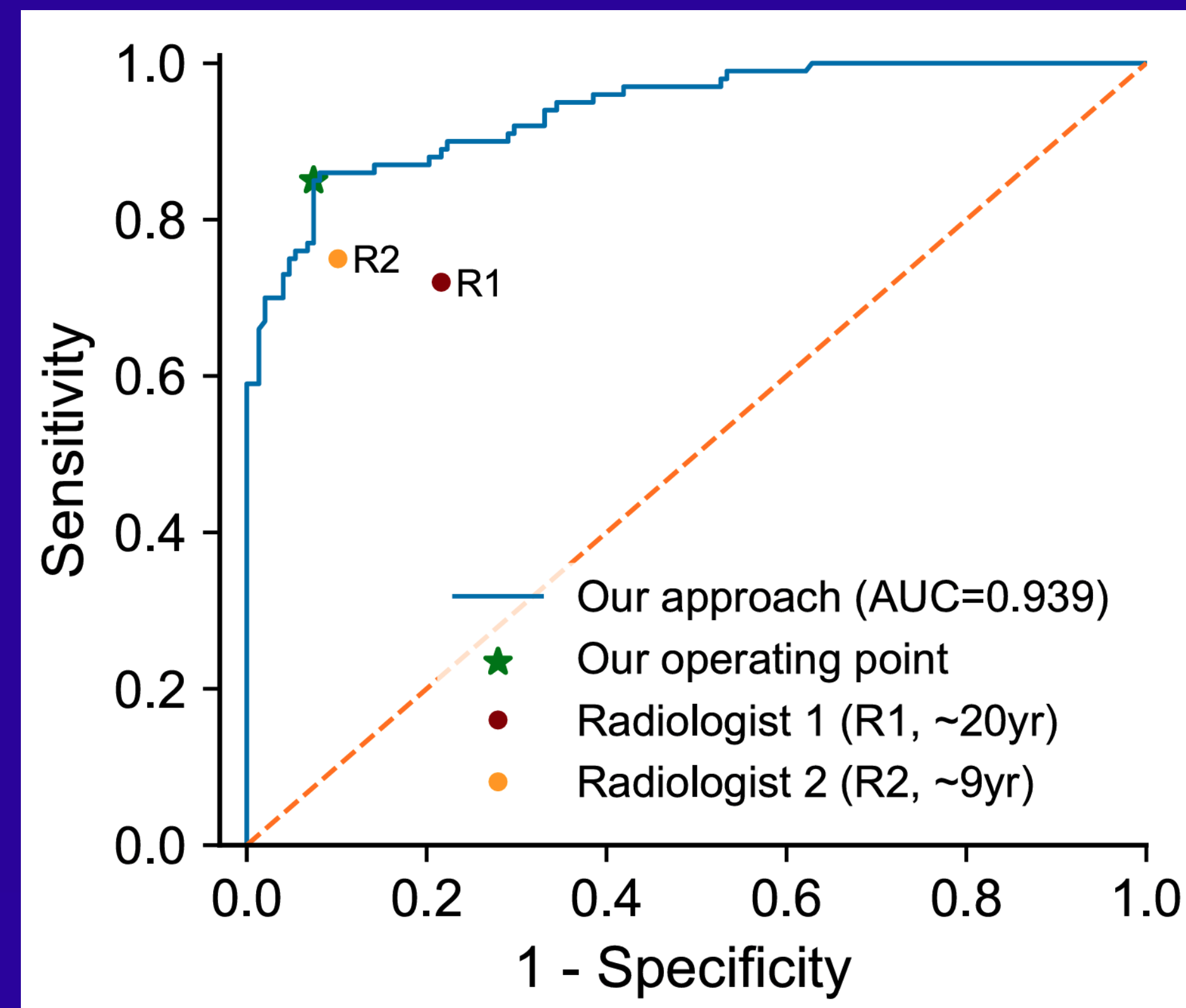
(4) Colorectal Cancer Screening using Non-Contrast CT



(5) Gastric Cancer Screening using Non-Contrast CT



Real world evaluation: from May 8 to July 23, a total of 34,937 patients were screened for gastric cancer via AI. Among the positive cases, there were 16 cases confirmed by imaging and 43 cases confirmed by pathology; there were 131 patients still being followed up.



• Reader Study结果

Reader	Patient (n)	Sensitivity	Specificity
Reader #1 (20 yrs)	300	72.0	78.4
Reader #2 (9 yrs)	300	75.0	89.9

Sensitivity	Specificity
80%	99.7%

(6) Lung nodule screening and diagnosis using non-contrast CT

Differentiating benign and malignant lung nodules: best quantitative performance

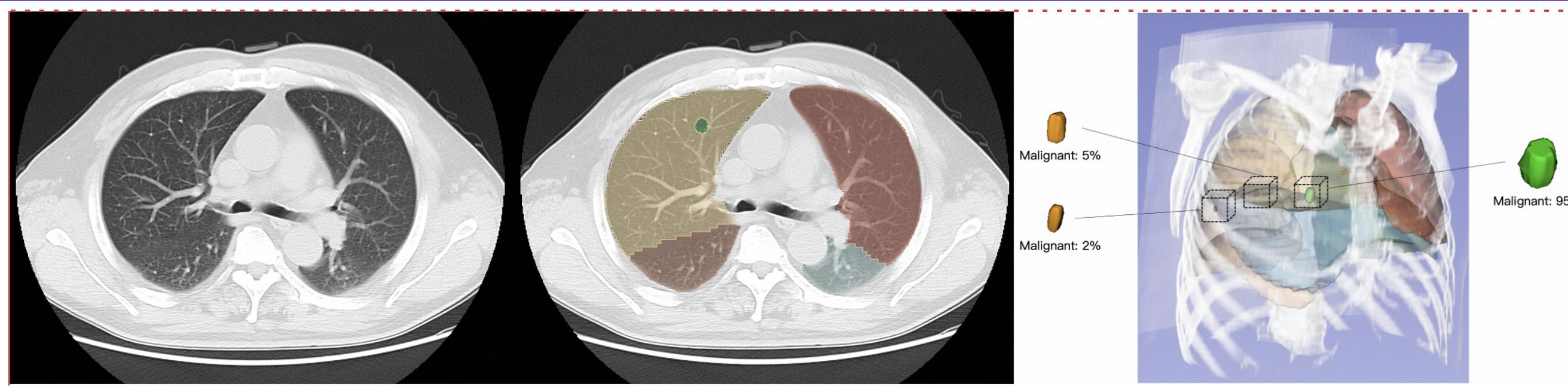


Table 3. Comparison of different methods on both NLST and in-house test sets. †: pure classification; ‡: pure segmentation; ◊: multi-task learning; *: ensemble of deep supervision heads. Note that we add the segmentation task in CA-Net.

Method	NLST test set				In-house test set			
	<10mm	10~20mm	>20mm	All	<10mm	10~20mm	>20mm	All
CNN [†]	0.742	0.706	0.780	0.894	0.851	0.797	0.744	0.901
ASPP [5] [†]	0.798	0.716	0.801	0.902	0.854	0.788	0.743	0.901
MiT [24] [†]	0.821	0.755	0.810	0.908	0.858	0.784	0.751	0.904
nnUnet [8] [‡]	0.815	0.736	0.815	0.910	0.863	0.804	0.750	0.911
CA-Net [12] [◊]	0.833	0.759	0.807	0.916	0.878	0.786	0.779	0.918
PARE [◊]	0.882	0.770	0.826	0.928	0.892	0.817	0.783	0.927
PARE ^{◊*}	0.890	0.781	0.827	0.931	0.899	0.821	0.780	0.931

Lung nodule screening

肺结节检测性能在多个数据来源组成的测试集上进行了评估，结果如表1所示。

	所有已标注结节			5mm以上结节		
	1	2	4	1	2	4
每数据平均假阳数量	1	2	4	1	2	4
检出率 (%)	60.1	69.9	79.3	72.7	81.9	86.5

表1. 在3278例多中心数据上的肺结节检测性能。

Lung mass screening

- New capabilities for detecting large lung masses, with a detection sensitivity of over 98%.
- Can distinguish between large lung masses and pneumonia.

Currently undergoing clinical multicenter validation and multiple scenarios in Shanghai for the differentiation of benign and malignant lung nodules.

Other company	Internal AUC	External AUC
XX-1	0.910	0.795
XX-2	0.969	0.847
Optellum (FDA)	0.921	0.864

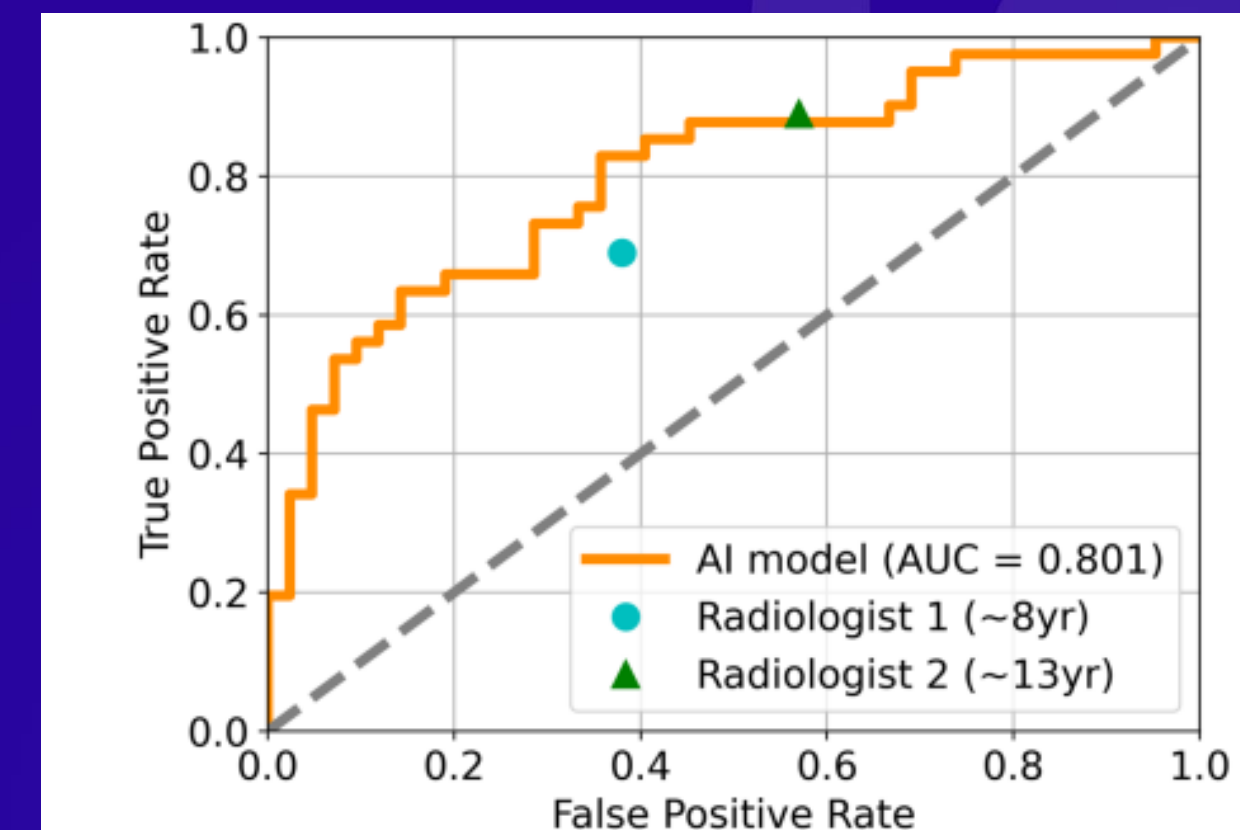
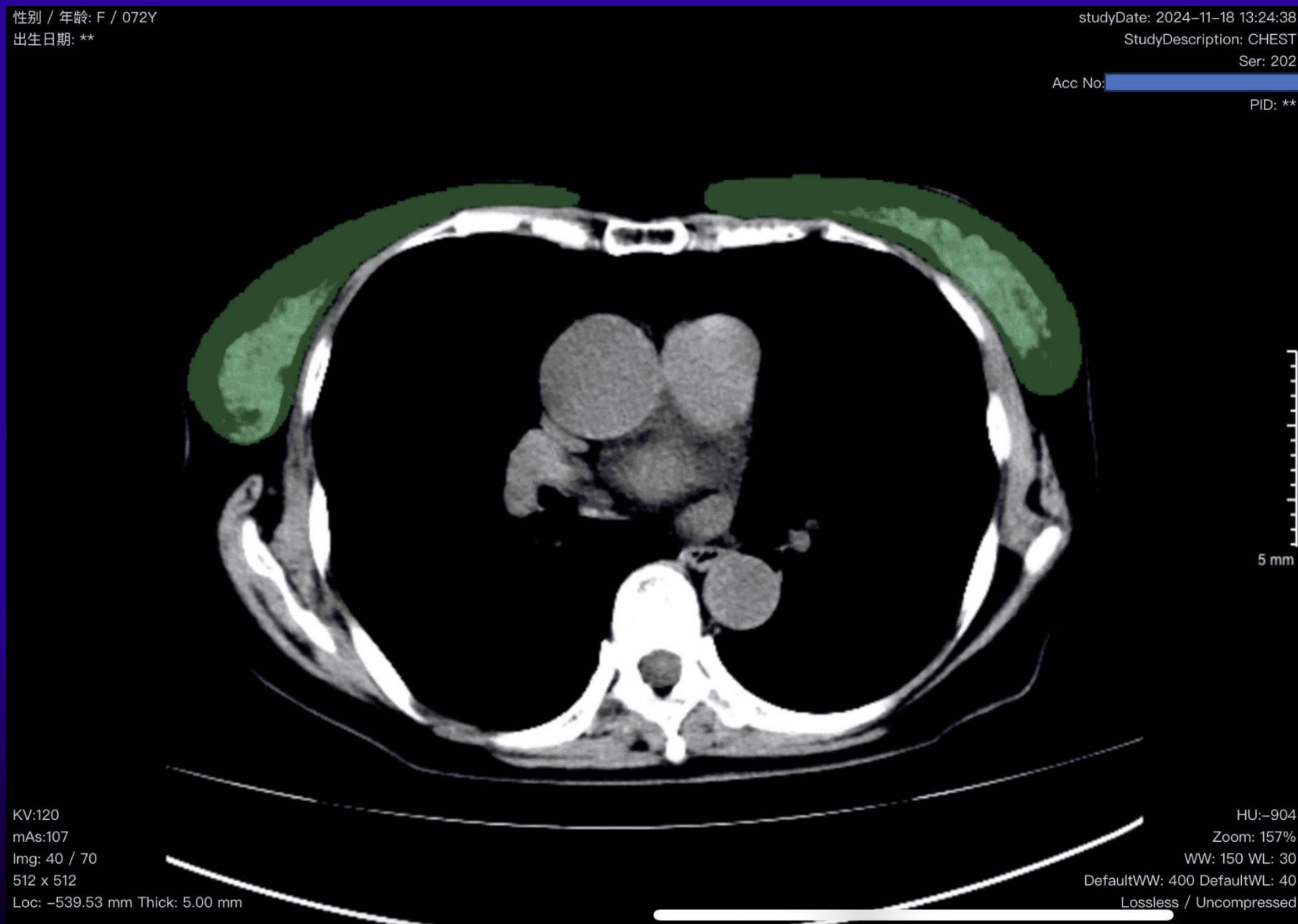


Fig. 3. Reader study compared with AI.

(7) Breast cancer screening using Non-Contrast CT

Non-standard breast cancer screening model (non-SOC) → Maximizing the clinical value of multi-cancer screening for patients.

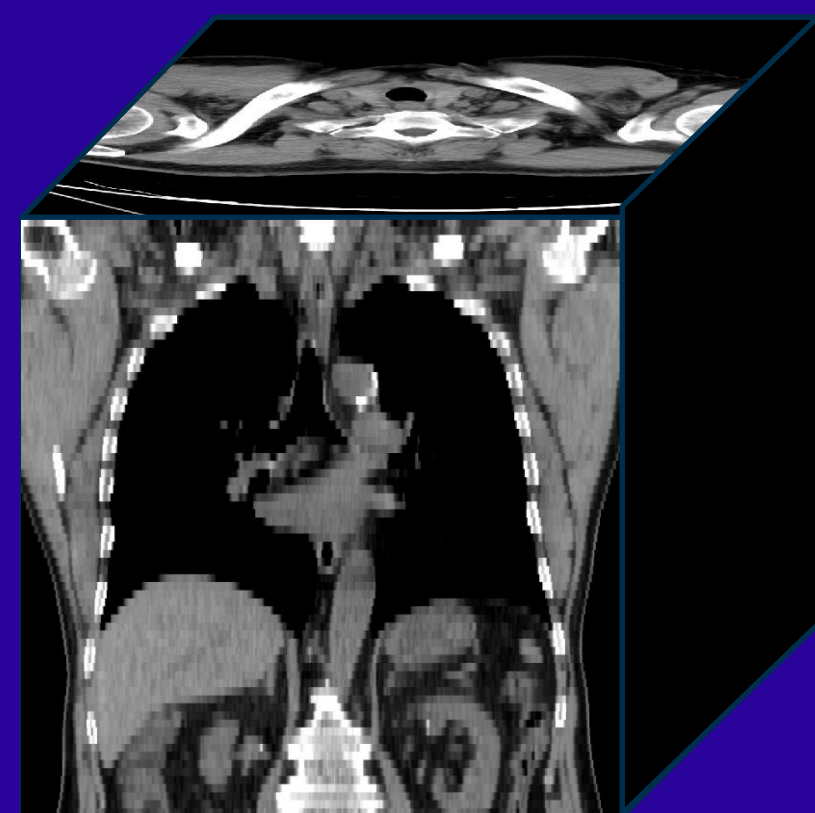


- Recall of 80% @ 0.1FPs per Breast under NCT
- In a preliminary reader study conducted on 100 patients with a total of 200 breasts, a comparison of the performance between doctors and the model was performed. Doctors had a low accuracy rate for non-contrast CT breast cancer detection (76%), while the model's performance significantly outperformed that of the doctors (87%).
- When both the model and Doctor 1 exhibited the same sensitivity, the specificity of the model was 20 percentage points higher than that of Doctor 1; conversely, when both exhibited the same specificity, the model's sensitivity was 15 percentage points higher. Both doctors had an accuracy rate of 76% with high specificity (99%); the model achieved an accuracy rate of 87%, which was 11 percentage points higher than that of the doctors.

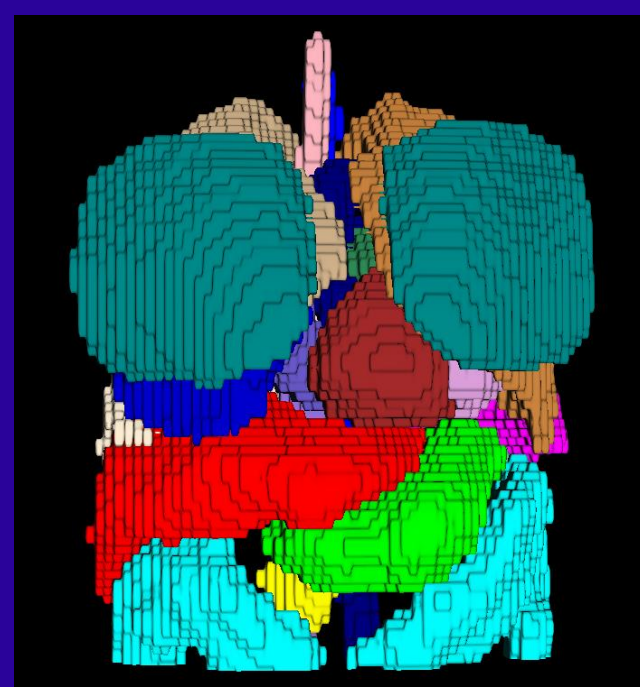
	AUC	Sensitivity (%)	Specificity (%)
Internal test	0.9686	74.45	99
External test 1	0.9158	52.84	99.78
External test 2	0.9416	63.37	95.8

Screening of seven cancers - Cloud API integration

- Initial version, 2 min/case



3D chest / abdominal non-contrast CT

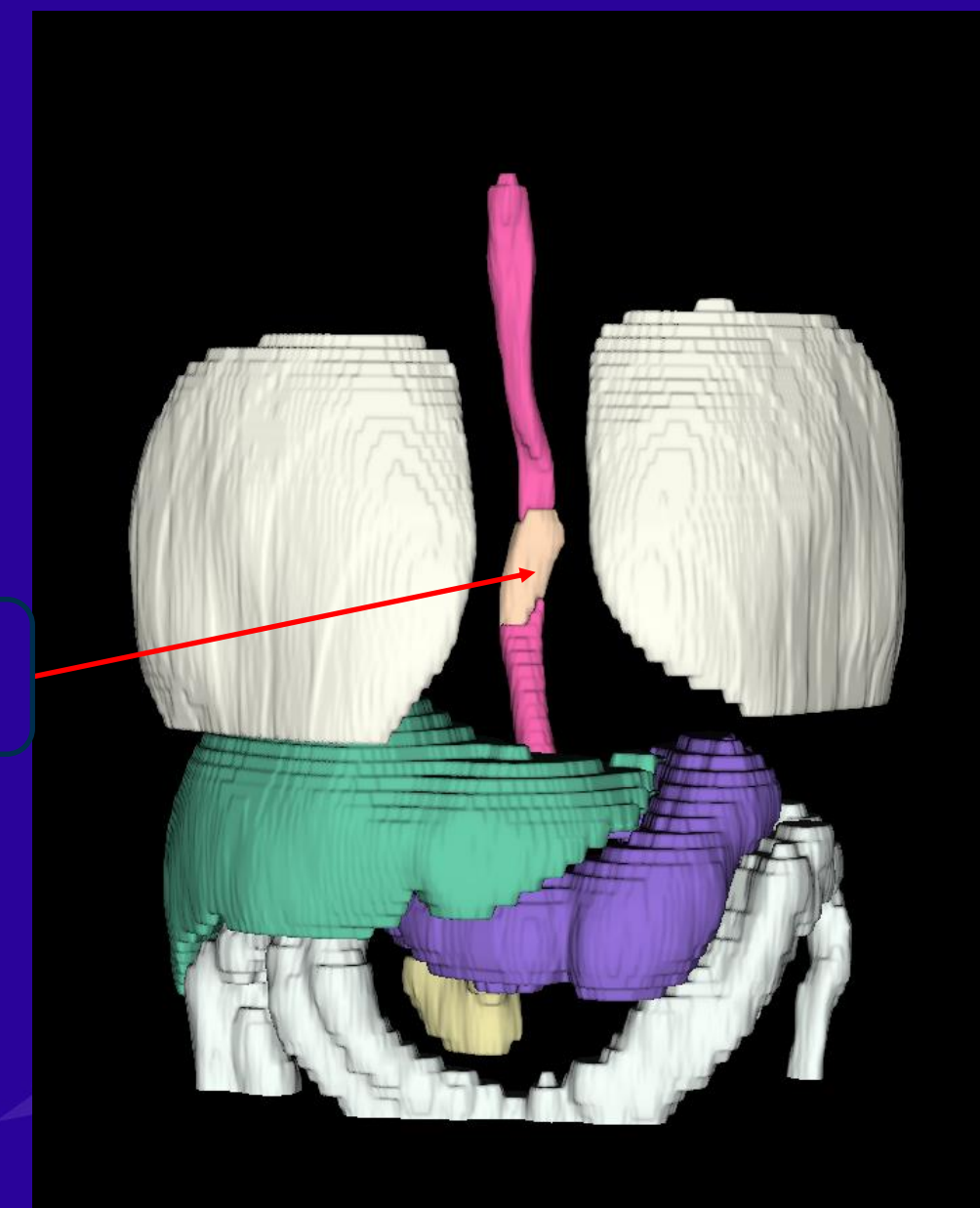


24 organs and vessels (<10s)

Fast organ segmentation and localization

```
"filename": "eso_esoCancer_ZSEC0500_NC_z5.00",  
"Tumor probabilities": {  
  "lung": {  
    "Cancer": 0.0  
  },  
  "breast": {  
    "BC": 0.114  
  },  
  "liver": {  
    "HCC": 0.0,  
    "other_malignant": 0.0,  
    "other_benign": 0.0,  
    "cyst": 0.0  
  },  
  "eso": {  
    "EC": 0.9994,  
    "nonEC": 0.0002  
  },  
  "stomach": {  
    "GC": 0.1642  
  },  
  "crc": {  
    "CRC": 0.2092  
  },  
  "pancreas": {  
    "PDAC": 0.0392,  
    "nonPDAC": 0.0351  
  }  
},
```

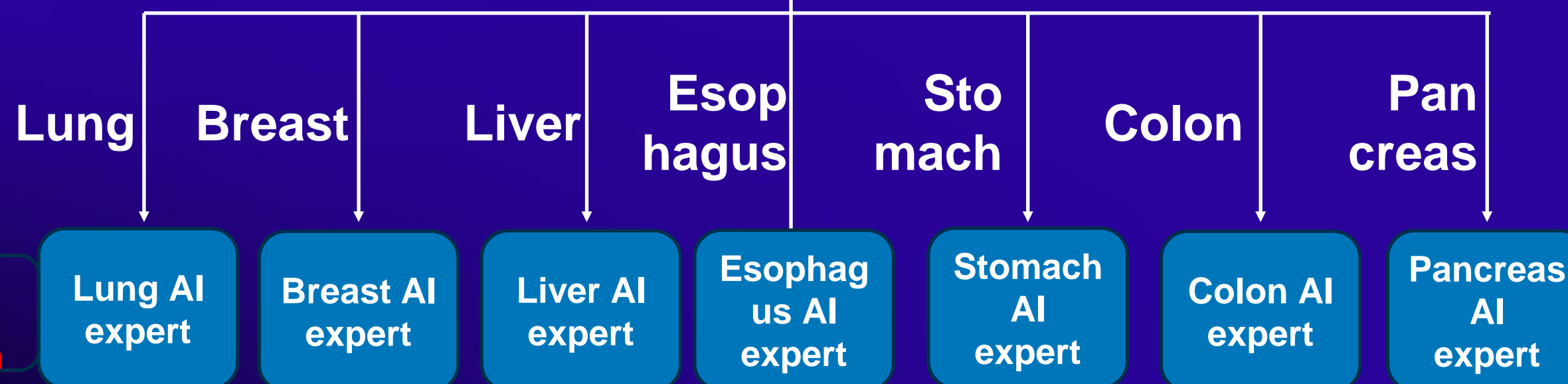
Cancer and other tumor probabilities



Esophageal cancer prob. 99%

Tumor segmentation

Organ ROIs

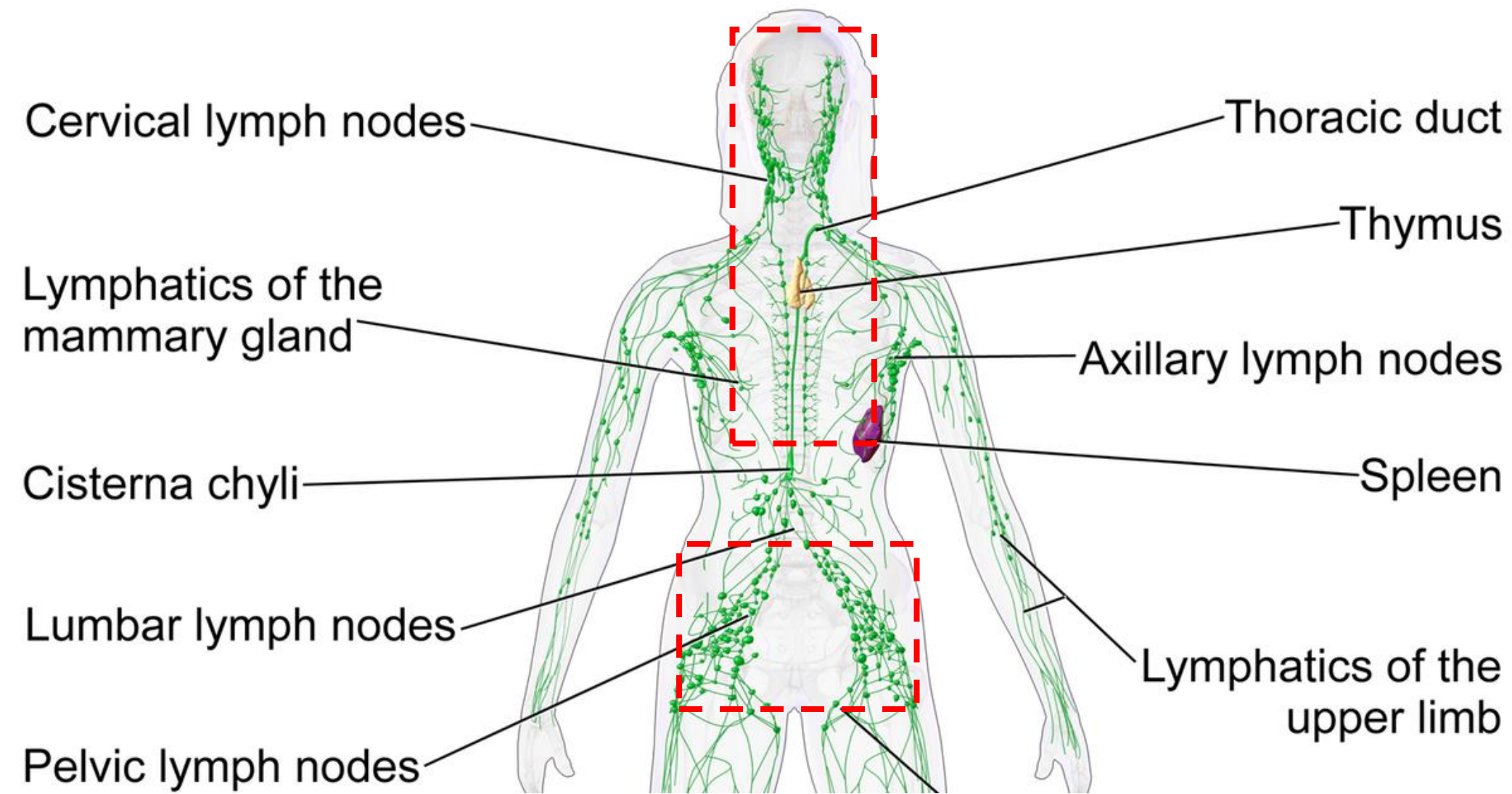


Expert models
15s/organ

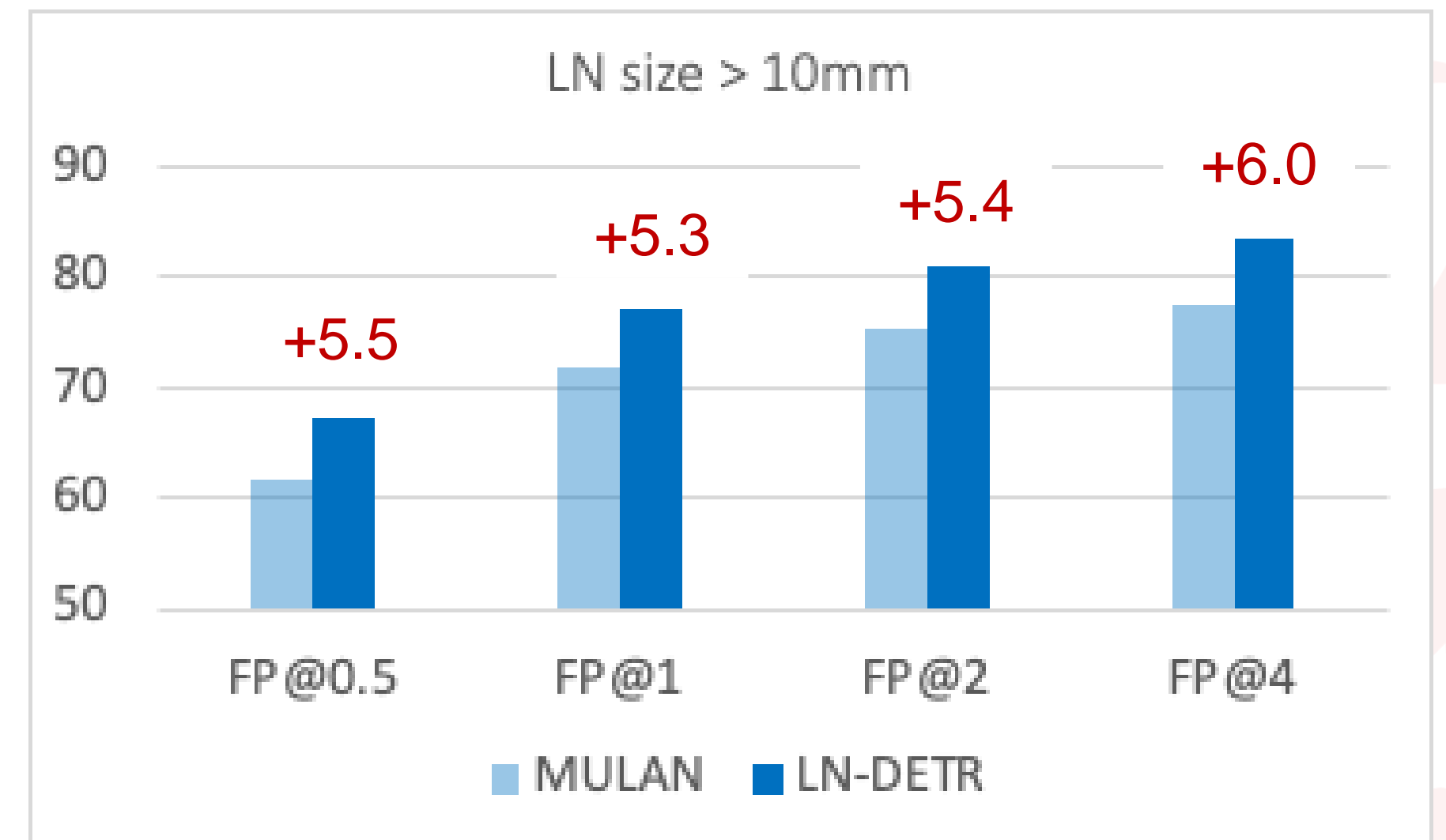
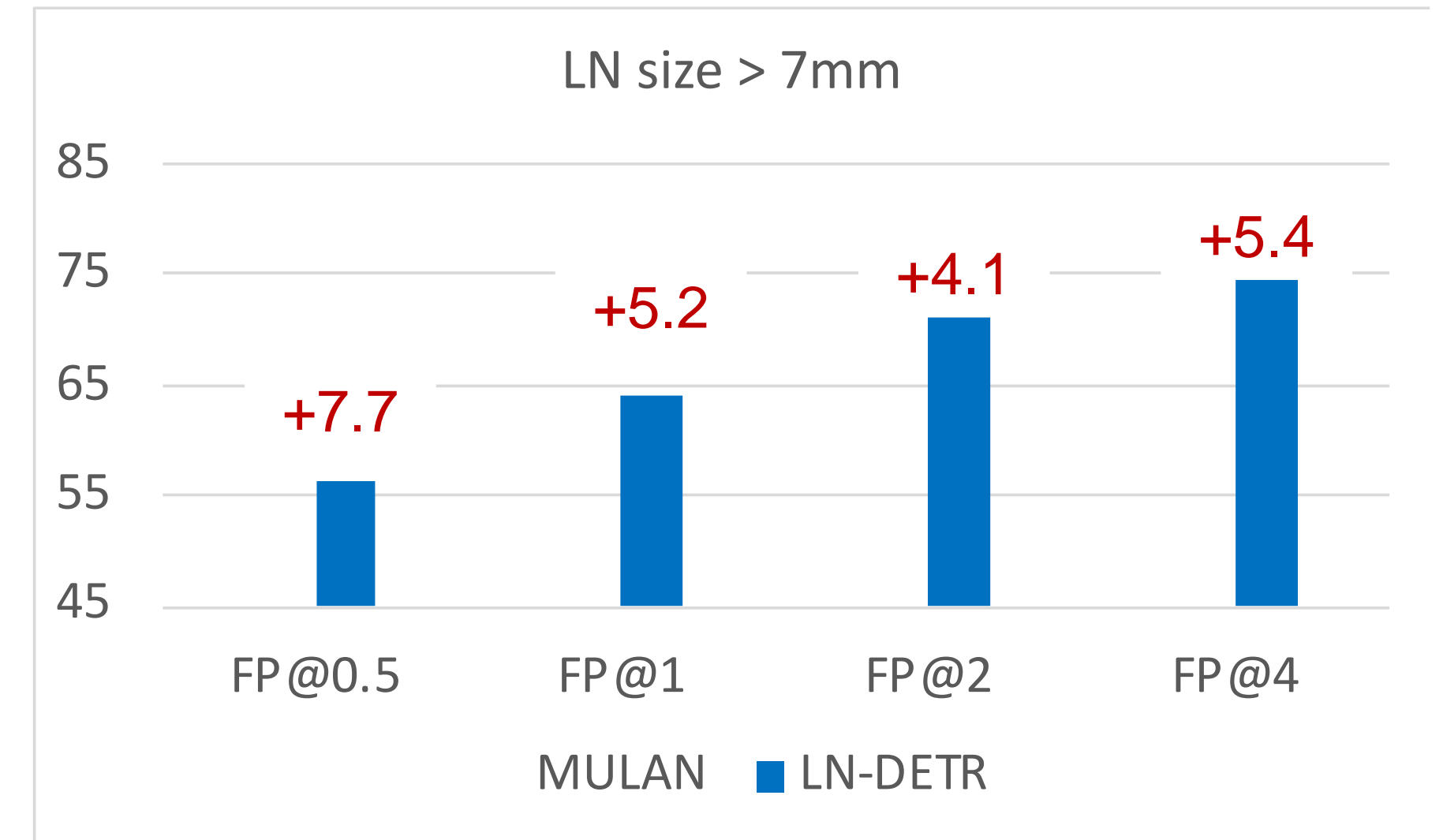
Segmentation + probability fusion

120s per case on a single GPU

(8) Whole-body abnormal lymph node screening on non-contrast CT



- Lymph nodes are distributed throughout the body. Currently, there are no methods for comprehensive screening of abnormal lymph nodes. Existing CT-based detection methods for abnormal lymph nodes focus on specific areas, resulting in a high false positive rate and can only screen lymph nodes with a short diameter greater than 1 cm (lymph nodes <1 cm have at least a 20-30% chance of being malignant [1]).
- We cover the head and neck, mediastinum, upper abdomen, and pelvis, which are the most common sites for cancer metastasis. Previously, the universal lesion detection model [2] showed an average of 0.5 to 4 false positives when detecting lymph nodes >1 cm, with a detection sensitivity of 71.6%. Our method achieves a sensitivity of 77.2% (+5.6%). For lymph nodes larger than 7 mm, the lesion detection method has a sensitivity of 61.7%, while our method has a sensitivity of 67.3% (+5.6%).



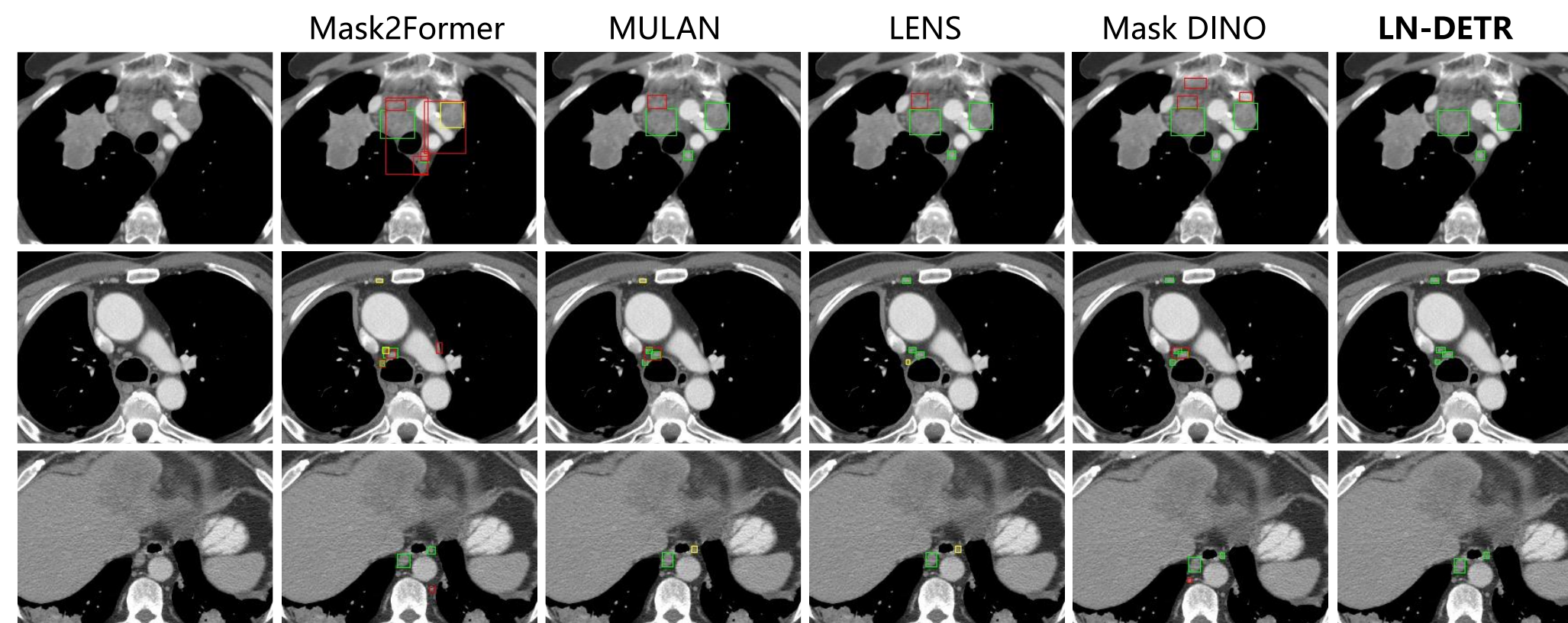
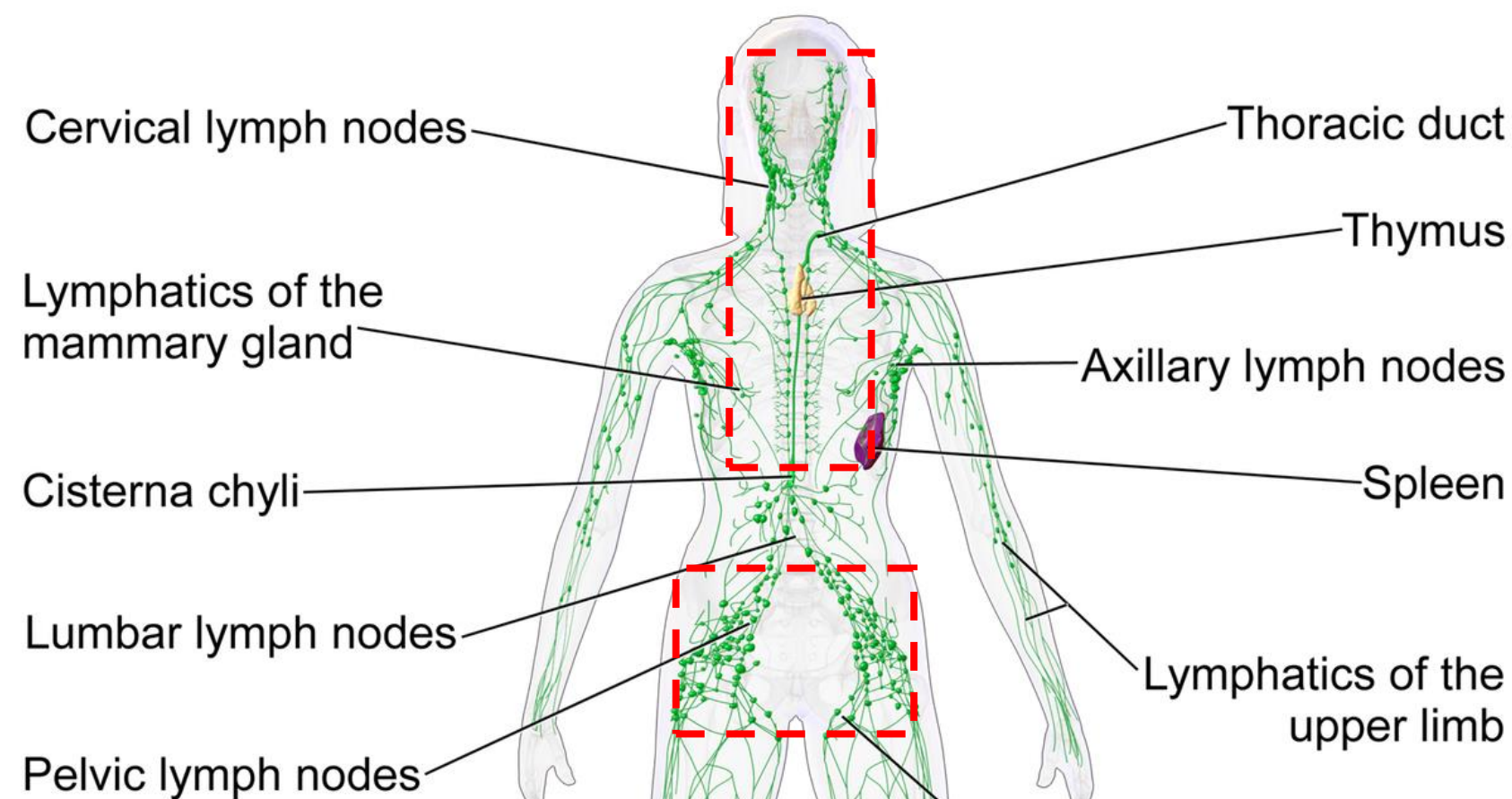
[1] “Bronchogenic carcinoma: analysis of staging in the mediastinum with CT by correlative lymph node mapping and sampling”, Radiology, 1992

[2] “Learning from multiple datasets with heterogeneous and partial labels for universal lesion detection in CT”, IEEE TMI, 2020

[3] “Effective Lymph Nodes Detection in CT Scans Using Location Debiased Query Selection and Contrastive Query Representation in Transformer”, ECCV 2024

[4] “Slice-Consistent Lymph Nodes Detection Transformer in CT Scans via Cross-slice Query Contrastive Learning”, MICCAI 2024

(8) Universal Whole-body Abnormal Lymph Node Detection



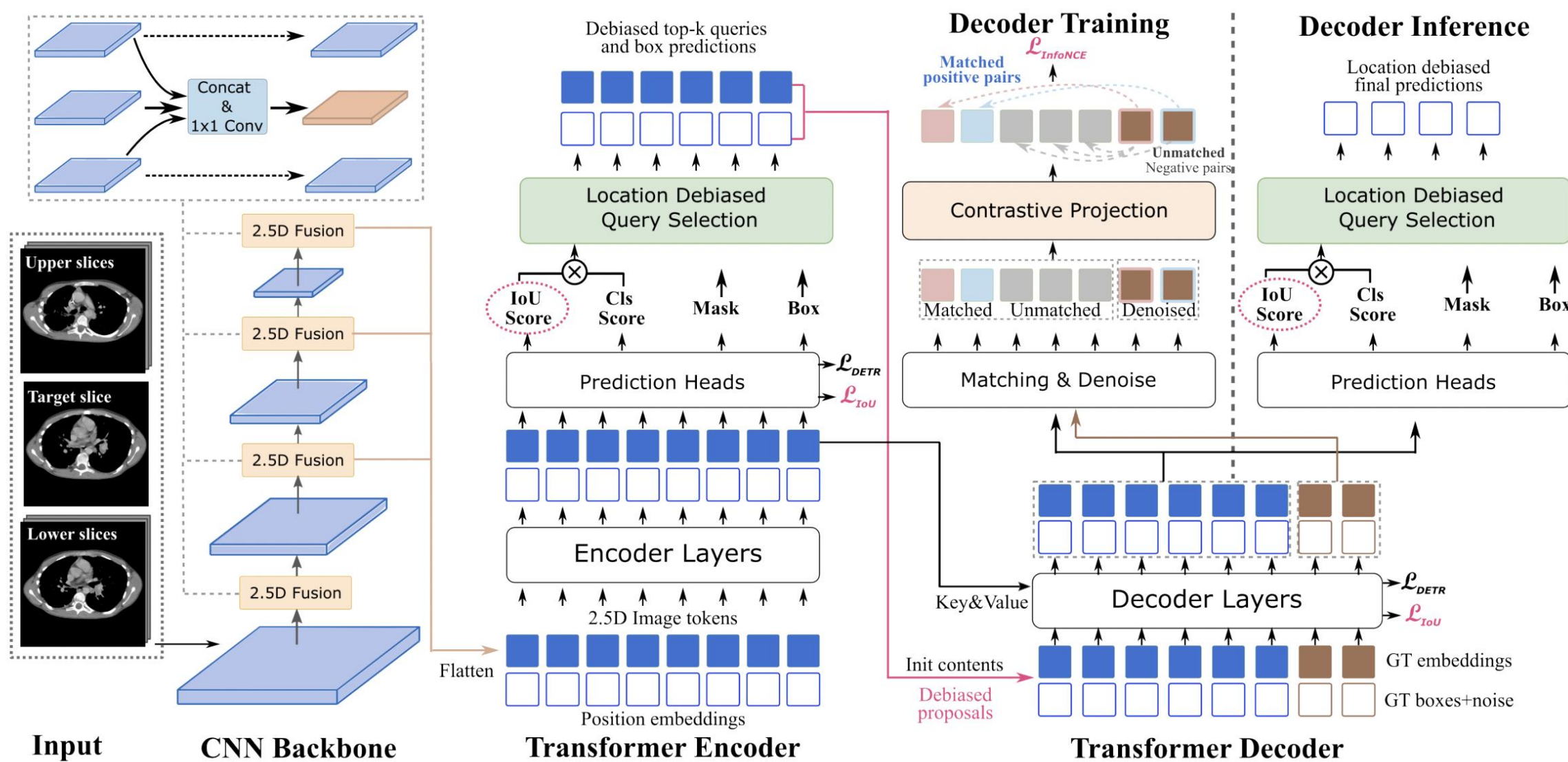
LN-DETR achieves **top performance** on **DeepLesion benchmark** results

Methods	Recall(%)@FPs				
	@0.5	@1	@2	@4	Avg.
3DCE [57]	62.48	73.37	80.70	85.65	75.55
RetinaNet [72]	72.18	80.07	86.40	90.77	82.36
MVP [31]	73.83	81.82	87.60	91.30	83.64
MULAN [60]	76.12	83.69	88.76	92.30	85.22
A3D [62]	79.24	85.04	89.15	92.71	86.54
LENS [58]	78.60	85.50	89.60	92.50	86.60
DKMA [51]	78.48	85.95	90.48	93.48	87.16
A3D [62]+SATr [27]	<u>81.03</u>	86.64	90.70	93.30	87.92
P3D* [68]	82.22	87.42	90.91	93.65	88.55
LN-DETR	79.89	<u>87.05</u>	92.00	94.89	<u>88.46</u>

- Propose the first transformer-based universal lymph node detection model – **LN-DETR**

- Datasets: 7 datasets covering head & neck, mediastinal, abdomen regions with **1000+ patients with 10,000+ lymph nodes**

- LN-DETR substantially outperforms leading detectors in both **internal** and **external** testing.



Model	Internal test						External test					
	Recall(%)@FPs					AP ₁₀ ^{box}	Recall(%)@FPs					AP ₁₀ ^{box}
	@0.5	@1	@2	@4	Avg.		@0.5	@1	@2	@4	Avg.	
Mask2Former [10]	32.63	39.76	49.77	58.28	45.11	49.91	28.69	37.49	46.14	54.57	41.72	43.66
MP-Former [68]	34.07	44.33	51.44	60.93	47.69	49.86	29.34	38.05	46.02	54.69	42.03	43.10
nnDetection [3]	26.54	33.34	41.34	50.89	38.03	50.35	25.57	31.98	39.43	49.36	36.59	39.21
Mask-RCNN [19]	31.24	37.05	49.21	60.33	44.45	46.63	27.00	34.52	47.87	53.46	40.71	42.18
LENS [60]	33.55	43.84	54.17	65.87	49.36	55.39	26.83	36.52	44.42	53.62	40.35	45.55
MULAN [62]	34.79	46.29	57.75	64.58	50.85	54.34	<u>34.99</u>	<u>44.36</u>	<u>53.00</u>	<u>59.80</u>	<u>48.03</u>	<u>47.43</u>
DINO [67]	37.55	44.80	57.43	65.16	51.23	54.42	32.32	42.04	49.85	58.13	45.60	45.20
Mask DINO [27]	38.48	47.09	55.27	64.43	51.32	54.72	34.80	43.05	51.49	58.20	46.89	46.90
LN-DETR	42.48	51.02	60.60	70.96	56.27	58.47	35.26	48.18	57.92	66.78	52.04	50.24
	+4.00%	+3.93%	+2.85%	+5.09%	+4.95%	+3.08%	+0.27%	+3.82%	+4.92%	+6.98%	+4.01%	+2.81%

[3] Baumgartner, et al. "ndetection: a self-configuring method for medical object detection", *MICCAI*, 2021.

[27] Li, et al. "Mask dino: Towards a unified transformer-based framework for object detection and segmentation", *CVPR*, 2023

[60] Yan, et al. "Learning from multiple datasets with heterogeneous and partial labels for universal lesion detection in CT", *IEEE TMI*, 2020

[62] Yan et al. "Mulan: multitask universal lesion analysis network for joint lesion detection, tagging, and segmentation", *MICCAI*, 2019

(1) Opportunistic screening for major CVD risk using non-contrast CT

Overall Goal: To achieve an algorithm for opportunistic screening of CVD risk based on CT imaging and its productization for major cardiovascular events.

Objectives:

Technical and clinical validation: **interpretability, AUC reaching SOTA (>0.85).**

Multi-center retrospective validation: retrospective validation

Productization for health check-up scenarios

Progress:

1. Interpretable algorithm, achieving optimal performance across different test cohorts.
2. Testing on a data set of 10,000 cases composed of various diseases, with **an average sensitivity close to 80%.**

Testing completed and clinical paper under review

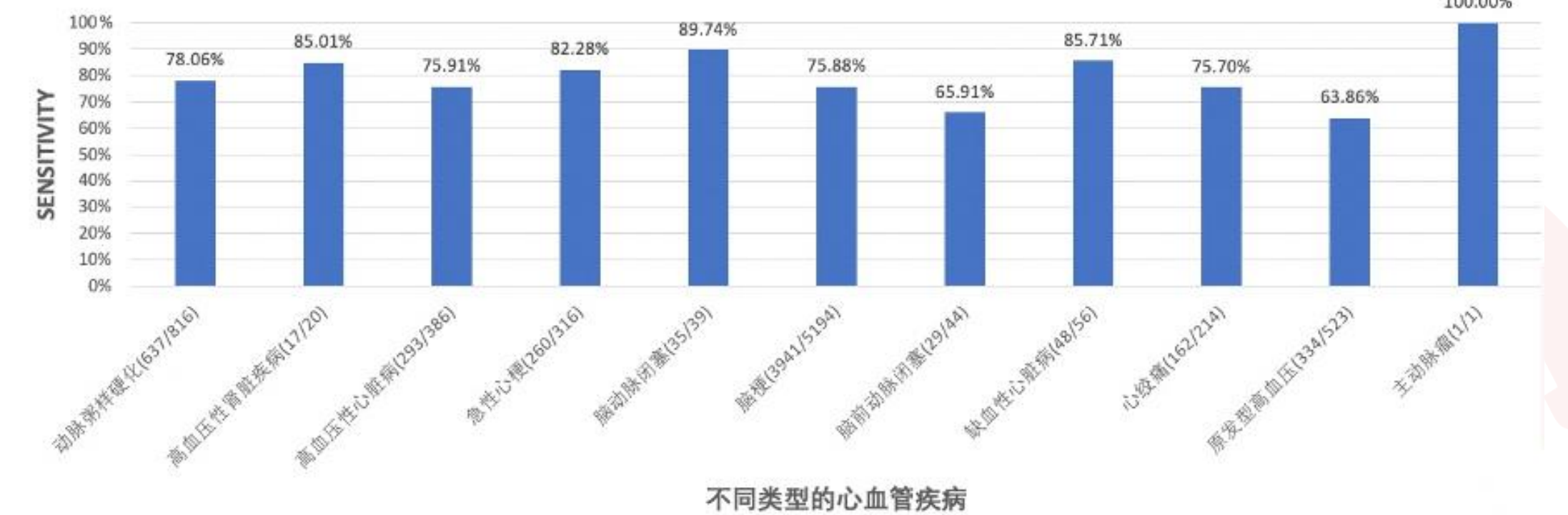
Mastered the performance boundaries of the algorithm and completed the design for the algorithm's product interface and output reports.

Xu, et al., “**A Joint Representation using Continuous (Deep) and Discrete Features for Cardiovascular Disease Risk Prediction on Chest CT Scans**”, under review, 2024

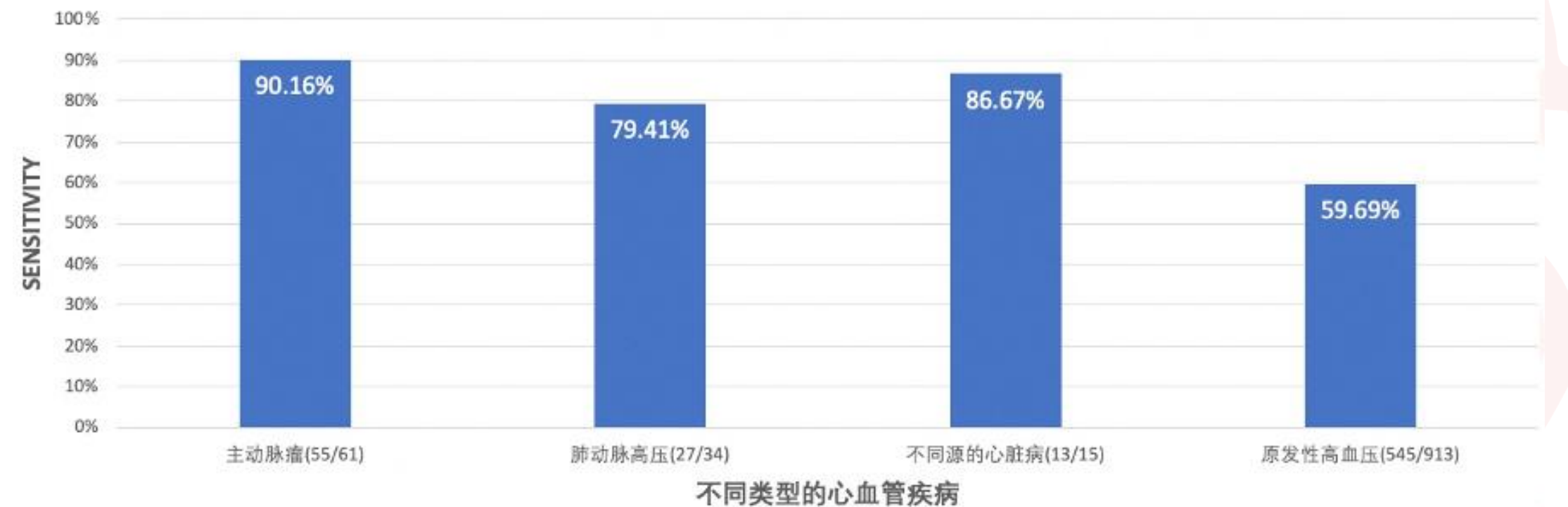
Chao, et al., “**Deep learning predicts cardiovascular disease risks from lung cancer screening low dose computed tomography**”, Nature Communications 12 (1), 2963, 2021

Feature Representation Manner	Method	Accuracy	Sensitivity	Specificity	F1-Score	AUC	
Discrete	Xgboost [44]*	0.771	0.264	0.971	0.394	0.835	
Internal NLST cohort	ResNet34 [45]*	0.796	0.543	0.896	0.601	0.844	
	nnUNet [33]*	0.825	0.587	0.919	0.655	0.874	
	Continuous	ViT-B [46]*	0.651	0.560	0.686	0.475	0.676
	nnFormer [47]*	0.788	0.692	0.825	0.648	0.837	
	Tri2D-Net [18]	0.819	0.485	0.952	0.603	0.869	
Joint	Ours	0.835	0.613	0.922	0.677	0.875	

External cohort 1



External cohort 2



Interpretability and Usability of Medical AI Diagnosis & Collaboration with Physicians

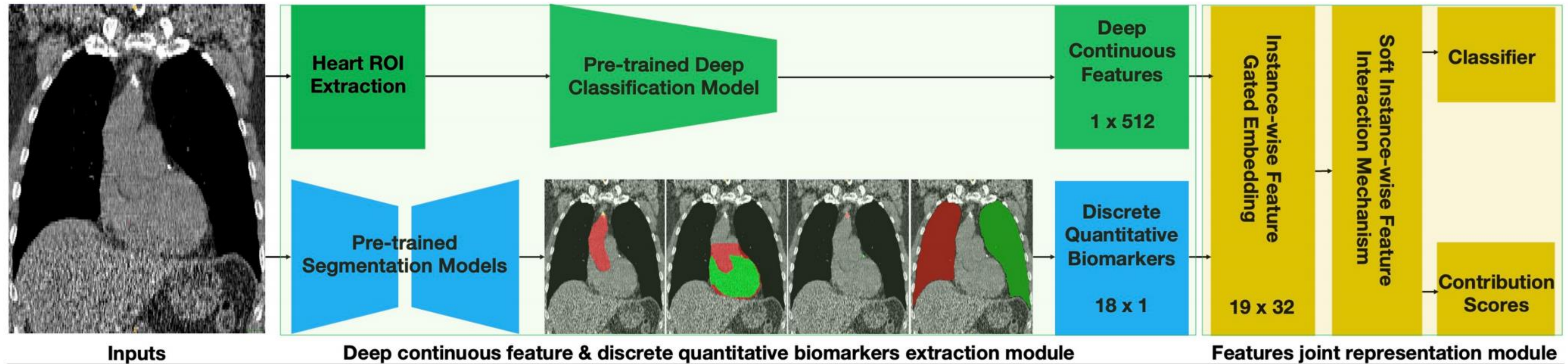


Fig. 1. The overview of our proposed continuous and discrete features joint representation framework for CVD risk prediction. It consists of two main modules: one is the deep continuous features and discrete CT quantitative biomarkers extraction module based on the pre-trained models, and another is the features joint representation module. The input of this approach is Chest CT solely, the output is the prediction of CVD risk and contribution score of each feature.

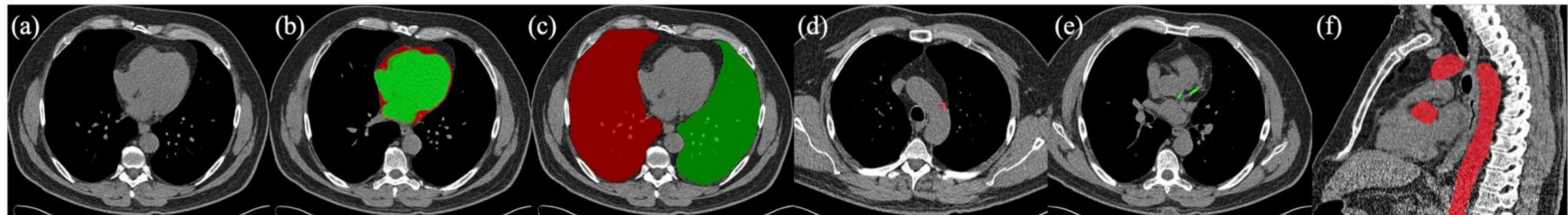


Fig. 2. The fully automated body part segmentation models used in this study. (a) Input Chest CT scan. (b) Heart chambers and pericardium segmentation. (c) Lungs segmentation. (d) Aortic calcium segmentation. (e) Coronary calcium segmentation. (f) Aorta segmentation.

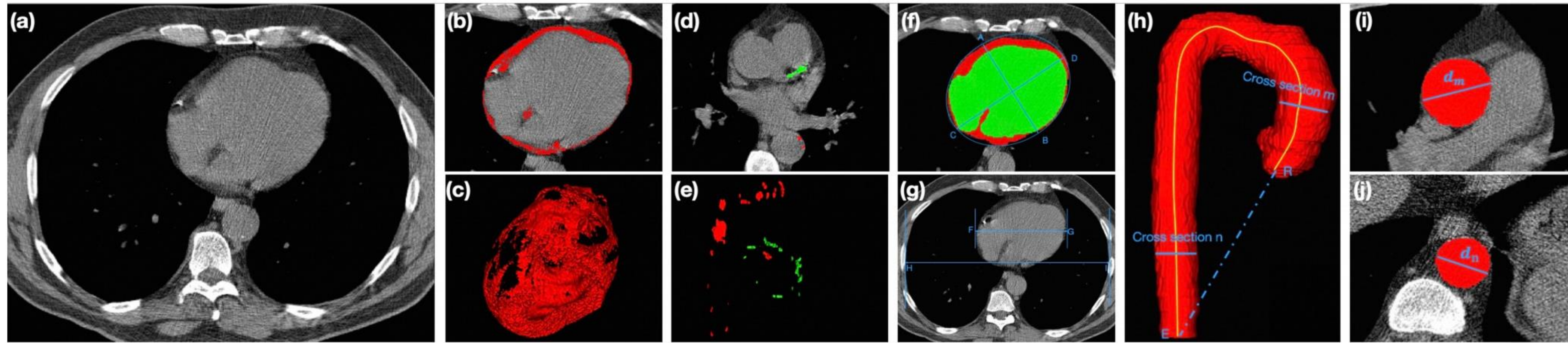


Fig. 6. It visually illustrates discrete CT quantitative biomarkers related to the heart and vasculature. (a) The input CT scan is in the axial view. (b) Segmentation results of pericardial fat in the axial view. (c) 3D visualization of the segmented pericardial fat. (d) Segmentation results for coronary artery calcification and aortic calcification in the axial view. (e) 3D visualization of the segmented coronary artery and aortic calcifications. (f) Axial view showing the segmentation results for cardiac chambers and pericardium, as well as a schematic representation of the calculation of cardiac long and short axes. (g) Schematic representation of calculating the cardiothoracic ratio. (h) 3D visualization of the thoracic aorta. (i) Cross-sectional view illustrating the segmentation at the ascending aorta. (j) Cross-sectional view depicting the segmentation at the descending thoracic aorta

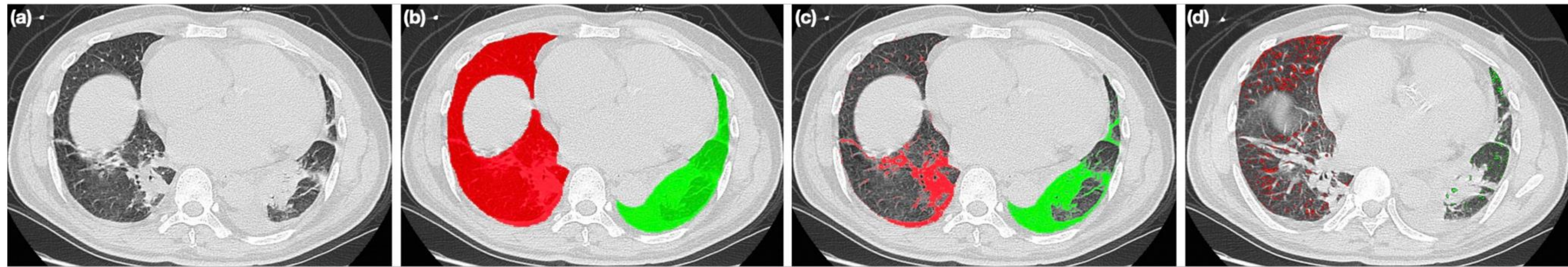


Fig. 7. It visually represents discrete CT quantitative biomarkers related to the lungs. (a) The input CT scan is in the axial view. (b) Lung segmentation results are displayed in the axial view. (c) Segmentation outcome of high attenuation regions within the lungs in the axial view. (d) Segmentation results of low attenuation regions within the lungs in the axial view.

(2.A) Diagnosis of Acute Aortic Syndrome (AAS) via non-contrast CT

Overall Goal: To develop a product that completes the clinical diagnostic pathway for AAS, saving lives and achieving equitable healthcare.

Stage Goals:

1. Technical and Clinical Validation: Retrospective validation across 8 centers (NPV > 0.98, SEN > 0.95, SPE > 0.97).
2. Real-World Value Validation: Multi-center verification of time savings in diagnosis, identification of missed diagnoses, and enhancement of trainee competency.
3. Prospective validation in tiered Hospital System: Complete prospective multi-center validation covering community – county – municipal - provincial hospital levels.

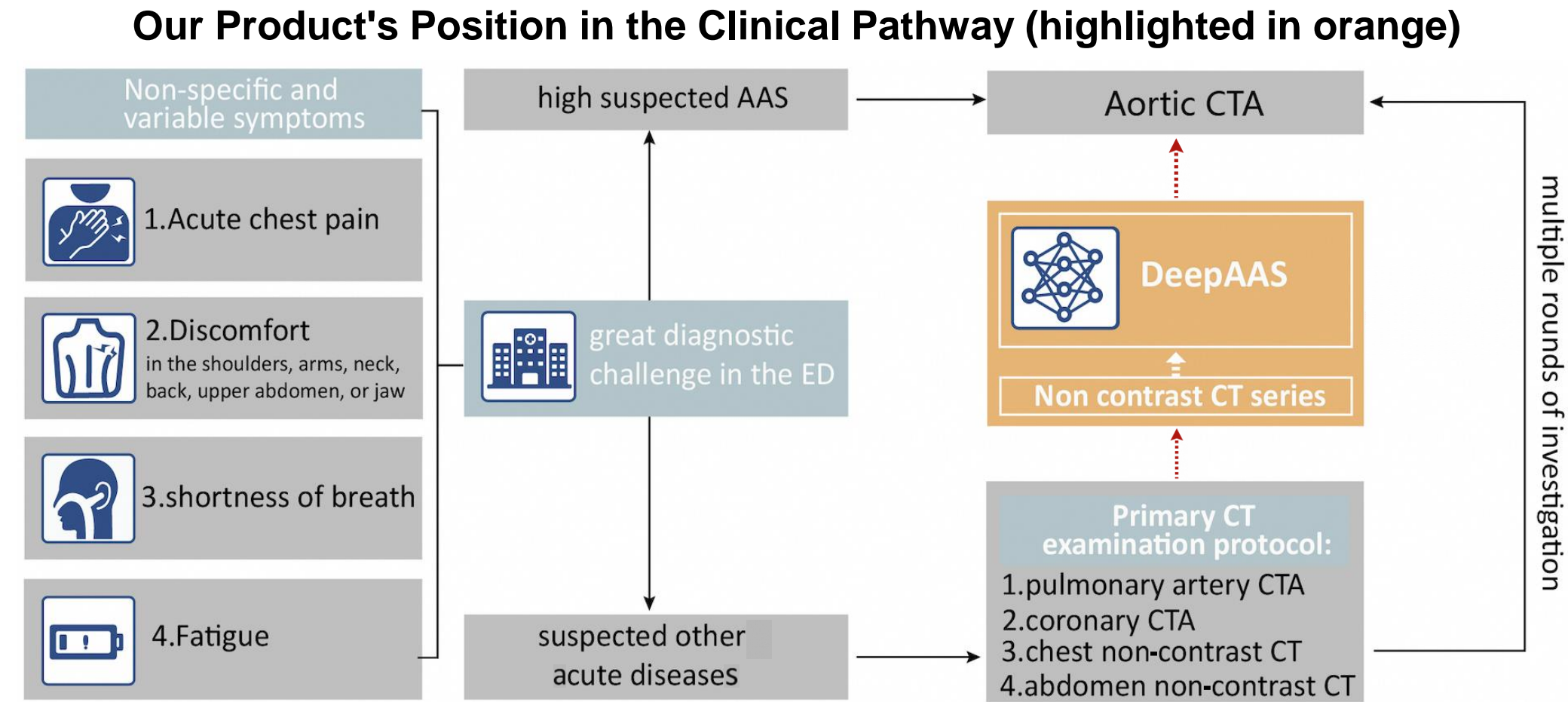
Current Progress:

1. Completed retrospective validation
2. Conducted a user study (positive patients diagnosed an average of 11 hours earlier, ongoing statistics for missed diagnoses show an average SEN of 0.928 and SPE of 0.992).

Next Steps:

1. Complete the full productization chain including OCR, H5, and algorithm API, followed by deployment in prospective hospitals.
2. Conduct a multi-center prospective validation in Zhejiang Province across community – county – municipal - provincial hospitals over a duration of 1 year.

Hu, Xiang, Zhou, et al., **Rapid and Accurate Diagnosis of Acute Aortic Syndrome using Non-contrast CT: A Large-scale, Retrospective, Multi-center and AI-based Study.** CoRR abs/[2406.15222](https://arxiv.org/abs/2406.15222) in revision, (2024)



User study validates our value in equitable access

	DeepAAS alone	Radiologist w/o DeepAAS			Radiologist w/ DeepAAS		
		Medical trainee	Board-certified	Special expert	Medical trainee	Board-certified	Special expert
Sensitivity	0.984 (0.972-0.990)	0.488 (0.478-0.501)	0.649 (0.637-0.661)	0.786 (0.772-0.802)	0.843 (0.829-0.856)	0.856 (0.843-0.867)	0.924 (0.912-0.936)
Specificity	0.947 (0.935-0.957)	0.804 (0.796-0.812)	0.856 (0.849-0.864)	0.922 (0.912-0.934)	0.914 (0.905-0.921)	0.913 (0.906-0.921)	0.948 (0.942-0.956)
Accuracy	0.960 (0.951-0.967)	0.689 (0.672-0.703)	0.781 (0.767-0.796)	0.873 (0.862-0.882)	0.888 (0.877-0.900)	0.892 (0.878-0.905)	0.939 (0.927-0.953)
PPV	0.908 (0.887-0.926)	0.589 (0.573-0.602)	0.722 (0.706-0.736)	0.855 (0.842-0.869)	0.852 (0.840-0.864)	0.850 (0.838-0.861)	0.911 (0.901-0.921)
NPV	0.991 (0.984-0.995)	0.733 (0.724-0.745)	0.811 (0.796-0.822)	0.883 (0.869-0.899)	0.912 (0.897-0.929)	0.917 (0.907-0.926)	0.956 (0.948-0.966)

Data are presented as the mean number (95% CI). PPV = Positive predictive value. NPV = negative predictive value.

Real-world verification involving 110,000 individuals across 3 centers

	Cohort 1 Zheyi	Cohort 2 Shaoxing	Cohort 3 Quzhou
Sensitivity (DeepAAS)	0.947	0.923	0.913
Specificity (DeepAAS)	0.990	0.991	0.993

(2.B) Pulmonary Embolism (PE) Diagnosis using Non-Contrast CT

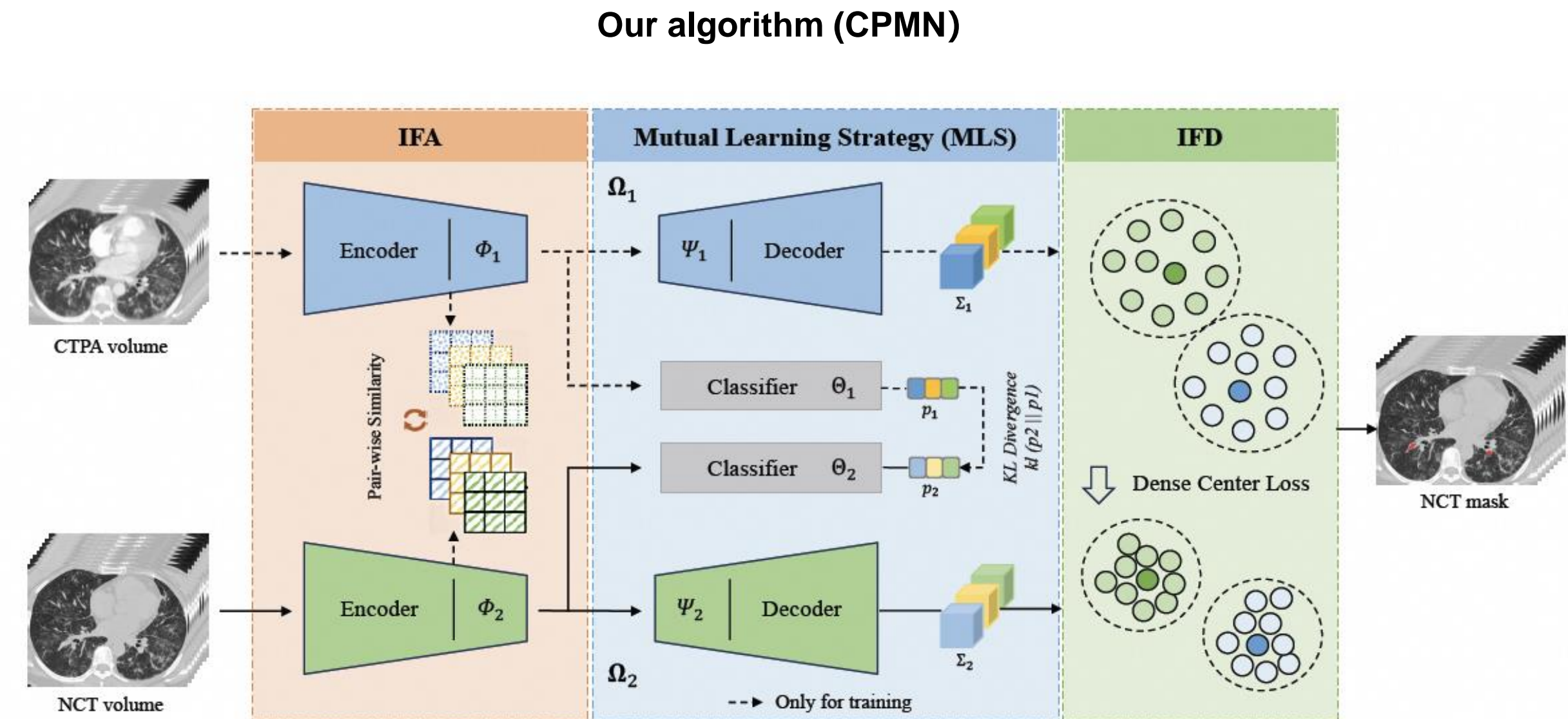
Overall Goal: To achieve precise PE diagnosis based on CT, complementing the solution for chest pain triad with AAS.

Stage Goals:

1. Technical and Clinical Validation: Approach the sensitivity (SEN) and specificity (SPE) of CTA, which are 0.969 and 0.996 respectively, using non-contrast CT scans.
2. Multi-Center Retrospective Validation: Conduct validation, completing a major clinical paper.
3. Prospective validation in tiered Hospital System: Complete prospective multi-center validation covering community – county – municipal - provincial hospital levels.

Current Progress:

1. Completed technical and clinical validation on a secondary vessel dataset, with the manuscript accepted by MICCAI 2024.
2. Ongoing construction of a large cohort with big data from hospital for level three and lower vessels, alongside iterative improvements to the model's segmentation capabilities.



Validation of CPMN on Zheyi dataset

Phase	Method	Classification			Segmentation
		Sens. (%)	Spec. (%)	AUC	Dice (%)
<i>Single</i>	CTPA model	96.9	99.6	0.996	79.9
	NCT model	84.6	97.8	0.973	68.8
<i>Dual</i> ^Δ	+ MLS ⁽¹⁾	92.3	99.1	0.989	70.1
	+ MLS + IFA ⁽²⁾	92.3	99.1	0.988	75.7
	+ MLS + IFA + IFD ⁽³⁾	95.4[†]	99.6[†]	0.990[‡]	78.5

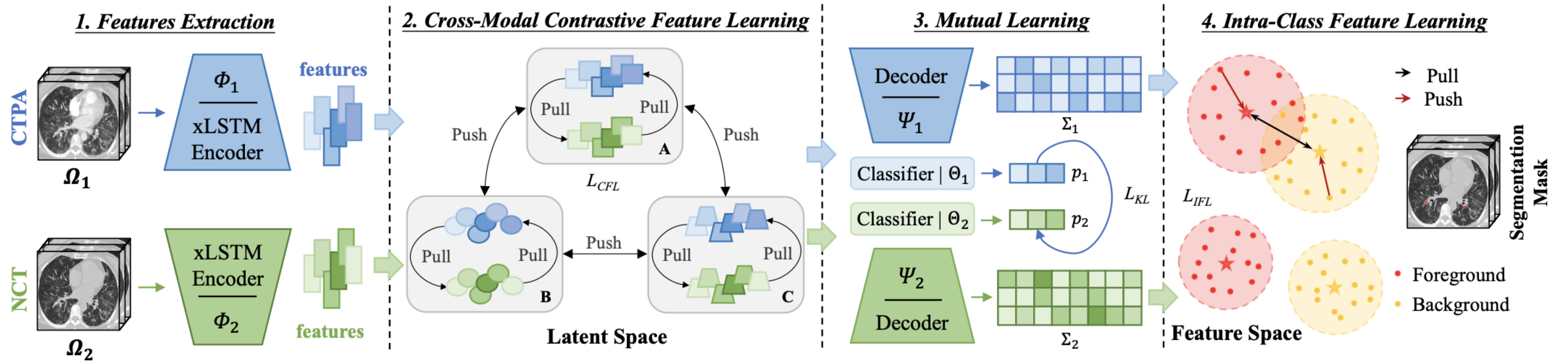
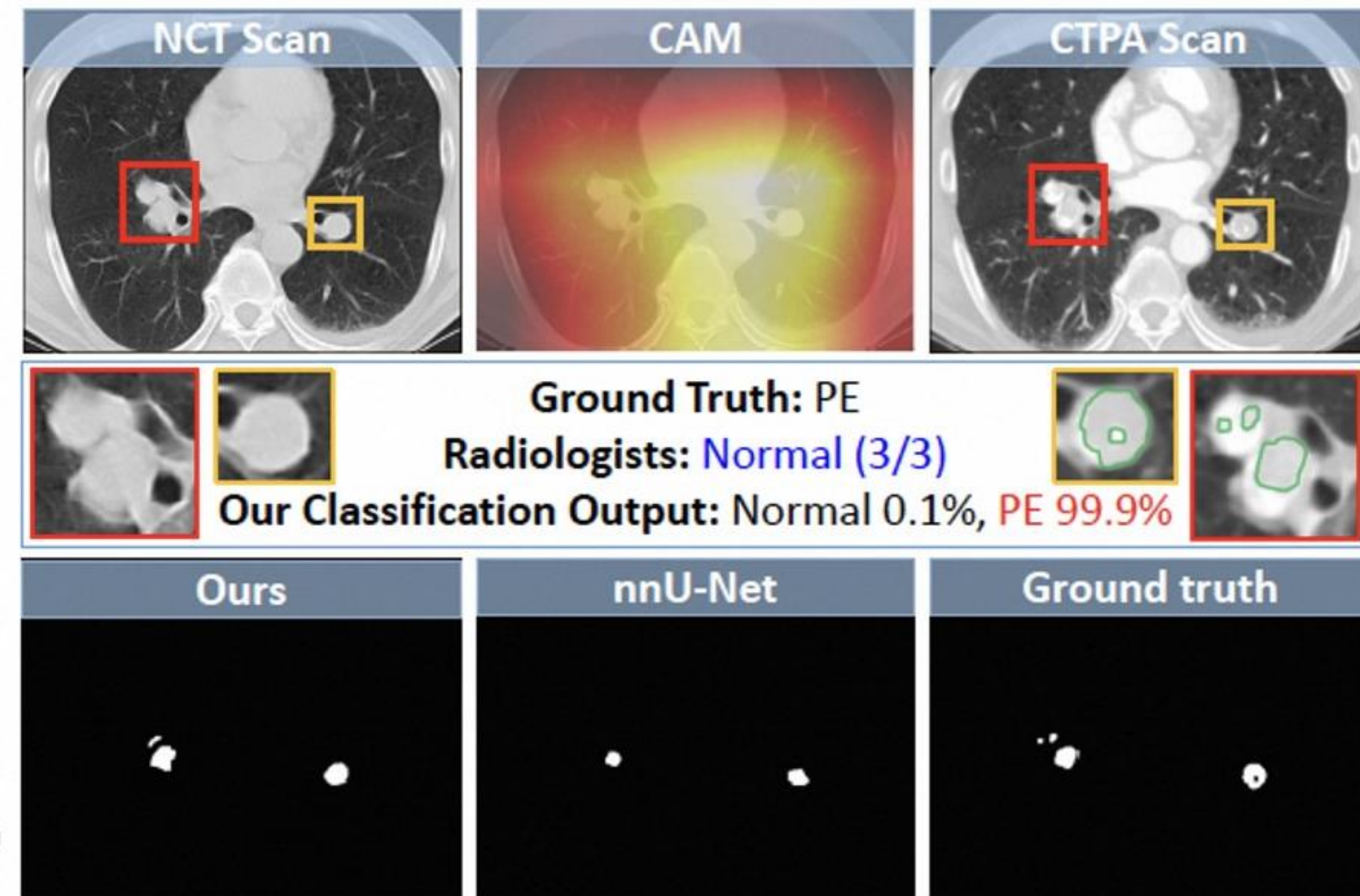
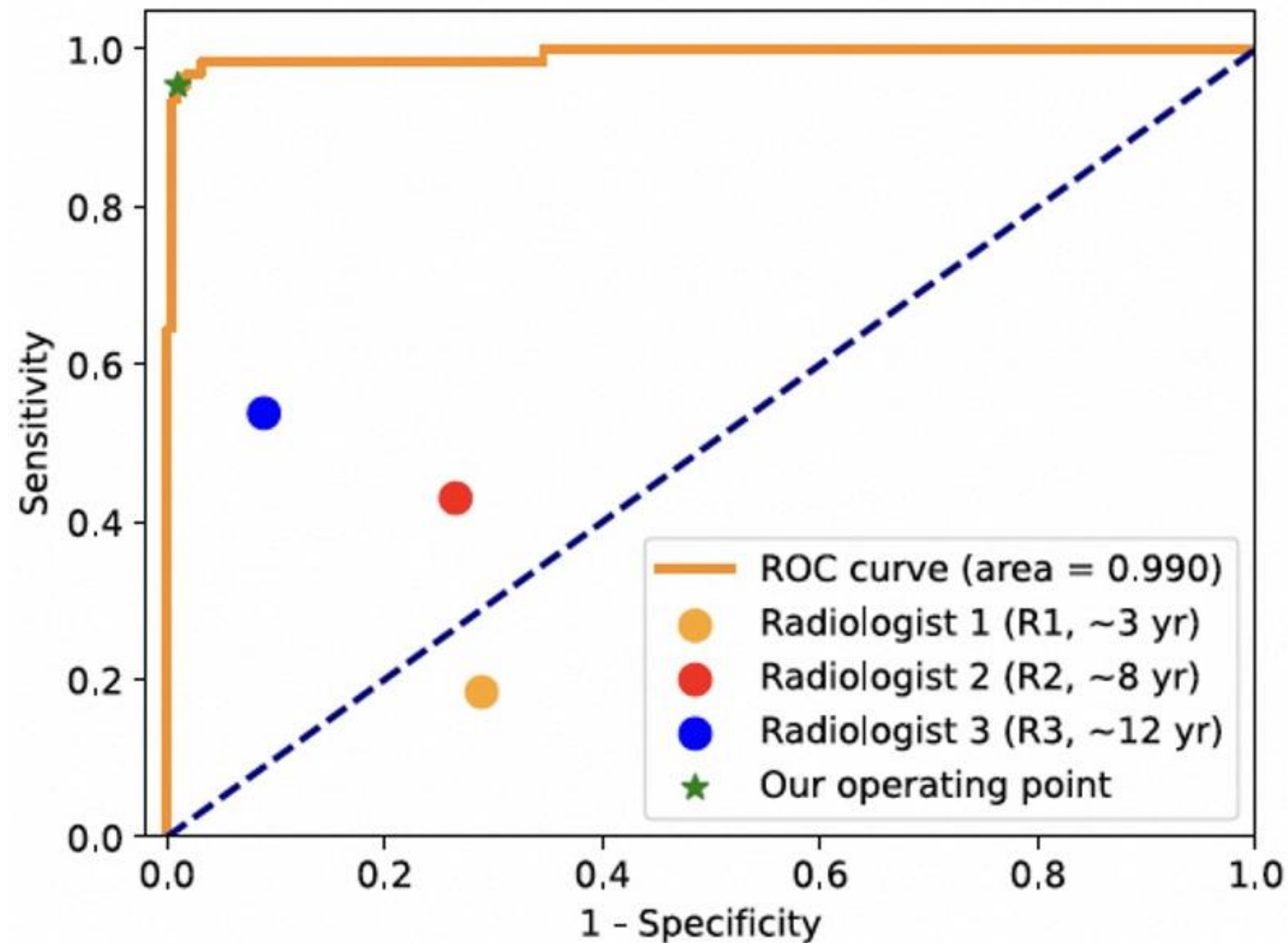


Fig. 1. Overview of our proposed contrastive mutual learning framework that contains the CTPA-pathway network (Ω_1) and the NCT-pathway network (Ω_2). Each pathway network comprises an encoder-decoder pair (Φ_1/Ψ_1 , Φ_2/Ψ_2) that extracts features from the corresponding volume. The presented Inter-Feature Alignment (IFA) strategy through an affinity graph captures pair-wise spatial feature similarities in the encoder. The predicted PE probabilities (p_1 , p_2) are harmonized using KL divergence to align feature distributions without altering the CTPA-pathway network. The dense center loss is designed to refine the segmentation feature space (Σ_1 , Σ_2).

User study results and visualization



(3) Chronic liver diseases: steatosis (fatty liver), cirrhosis, and esophageal varices

- Products for screening steatosis based on non-contrast CT serve a broad patient population
- Products for cirrhosis screening based on non-contrast CT, as well as esophageal varices screening for the cirrhotic population

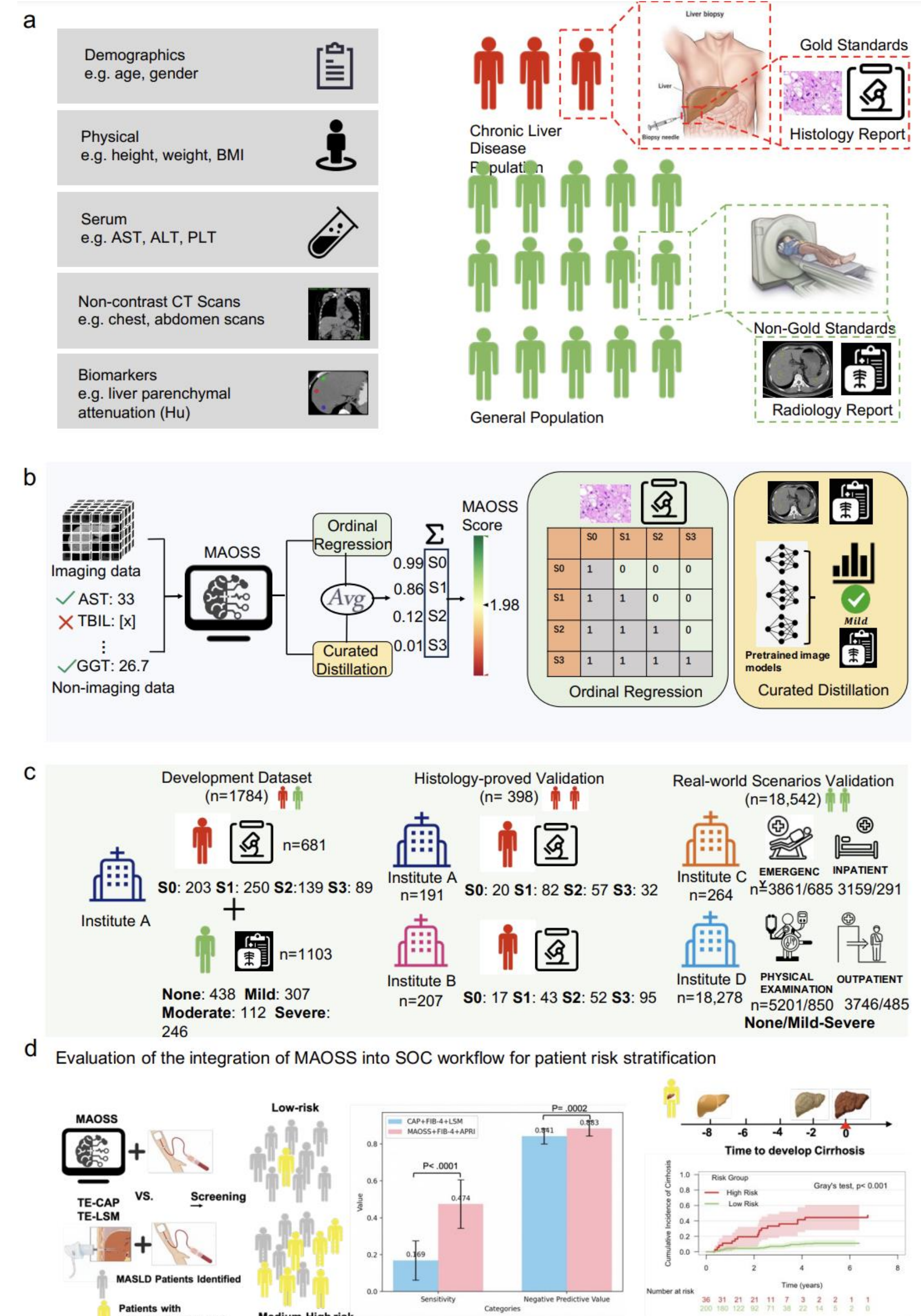
[1] “Effective Opportunistic Hepatic Steatosis Screening Using Imaging and Non-Imaging Cues via Robust Multi-modal Learning”, 2024

[2] “Improved Esophageal Varices Assessment on Non-Contrast CT Scans”, MICCAI 2024

Bowen Li, et al., **Accurate and Generalizable Quantitative Scoring of Liver Steatosis from Ultrasound Images via Scalable Deep Learning**, YIA Silver Medal, **AFSUMB 2021** Conventional grayscale ultrasound diagnosis surpasses elastographic ultrasound (Software Medical Device - Hardware Medical Device). A Computer Science Ph.D. candidate received the Young Scientist Award/Silver Award at the 2021 Asian Ultrasound Congress.

Bowen Li, et al., **Accurate and Reliable Liver Steatosis Assessment From Conventional Ultrasound Images Trained With Subjective Ratings**, Scientific Oral, **RSNA 2021**

Bowen Li, et al., **Learning from Subjective Ratings Using Auto-Decoded Deep Latent Embeddings**. (early accept, MICCAI Student Travel Award, ORAL) MICCAI 2021, CoRR abs/2104.05570 (2021)



(3.a) Steatosis: Real-World Opportunistic Screening via Non-Contrast CT

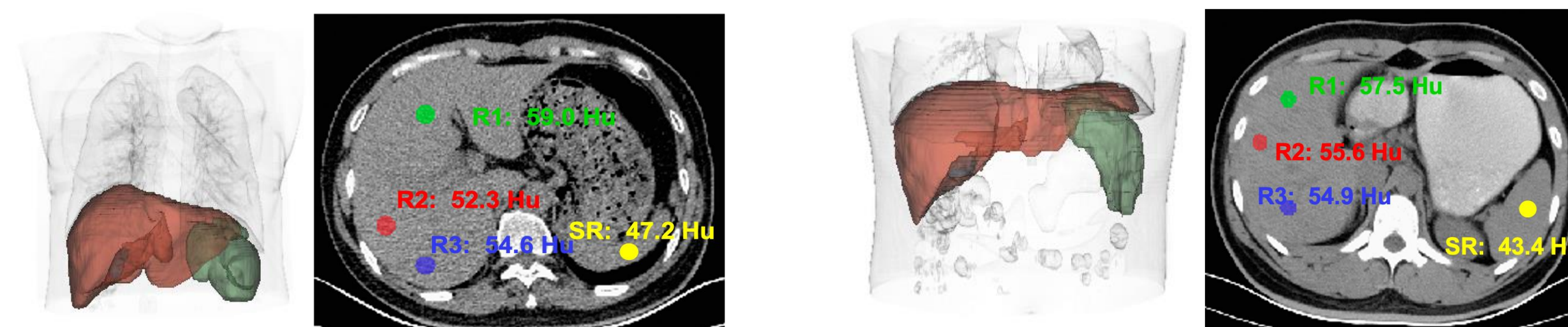
- ◆ A clinical product solution based on non-contrast CT for steatosis screening
- ◆ Steatosis Screening Product: Completed the v2 version upgrade and real-world testing and gather feedback to optimize product capabilities.

- ✓ Greatest challenge: early diagnosis of mild steatosis, which can hardly be detected by doctors on CT – yet, with appropriate intervention, mild steatosis is reversible without medication.
- ✓ Another issue is the scarcity of gold standards, which makes deep models prone to overfitting and poor generalization. To address this, we propose a novel learning framework that leverages a limited number of pathologically confirmed gold standard steatosis cases. It also employs distillation to curate labels from a vast amount of real-world data with imaging reports, incorporating reliable annotations into an iterative learning and refinement process!
- ✓ High-quality algorithms remain crucial!
- ✓ Currently, this solution shows good generalizability across four internal and external centers (in China and Brazil), and validation results on a pathology gold standard patient test set outperform elastography ultrasound / SOC significantly, approaching the diagnostic accuracy of MRI-PDFF!

Deep-bio screens steatosis achieving SOTA in real-world validation

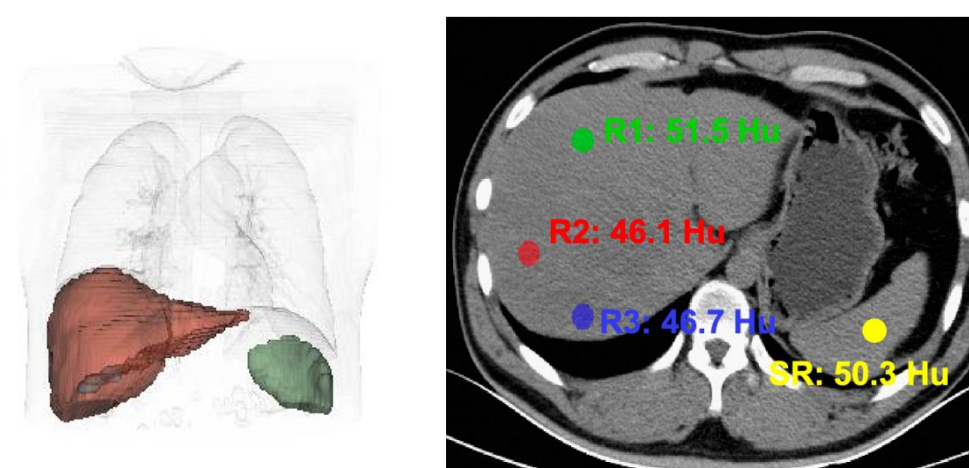
Deep-bio	Inpatient (3450)	Outpatient (4231)	Emergency (4546)	Health check (4177)
Sensitivity	0.835	0.905	0.88	0.943
Specificity	0.926	0.898	0.834	0.908
AUC	0.947	0.963	0.926	0.970

More accurate in edge cases than the method of Radiology [1]

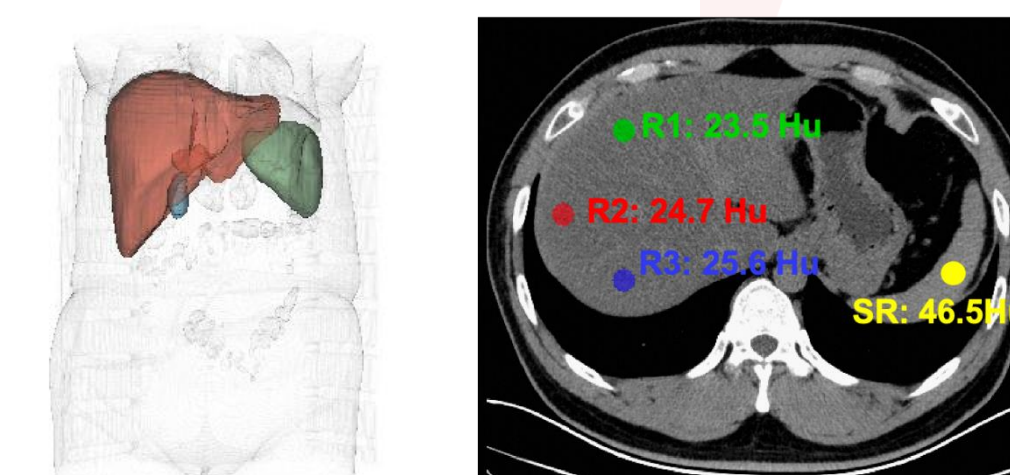


Pathology: S0
 QuanLF [1]: Mild ❌
 Deep-bio [2]: None ✓

Pathology: S1
 QuanLF [1]: None ❌
 Deep-bio [2]: Mild ✓



Pathology: S2
 QuanLF [1]: Mild ❌
 Deep-bio[2]: Moderate ✓



Pathology: S3
 QuanLF [1]: Moderate ❌
 Deep-bio [2]: Severe ✓

Pathological staging of steatosis (S0: None; S1: Mild; S2: Moderate; S3: Severe)

[1] Quantification of liver fat content with ct and mri: state of the art. Radiology, 301(2):250–262, 2021.

[2] Gao et al., “Effective Opportunistic Hepatic Steatosis Screening Using Imaging and Non-Imaging Cues via Robust Multi-modal Learning”, 2024

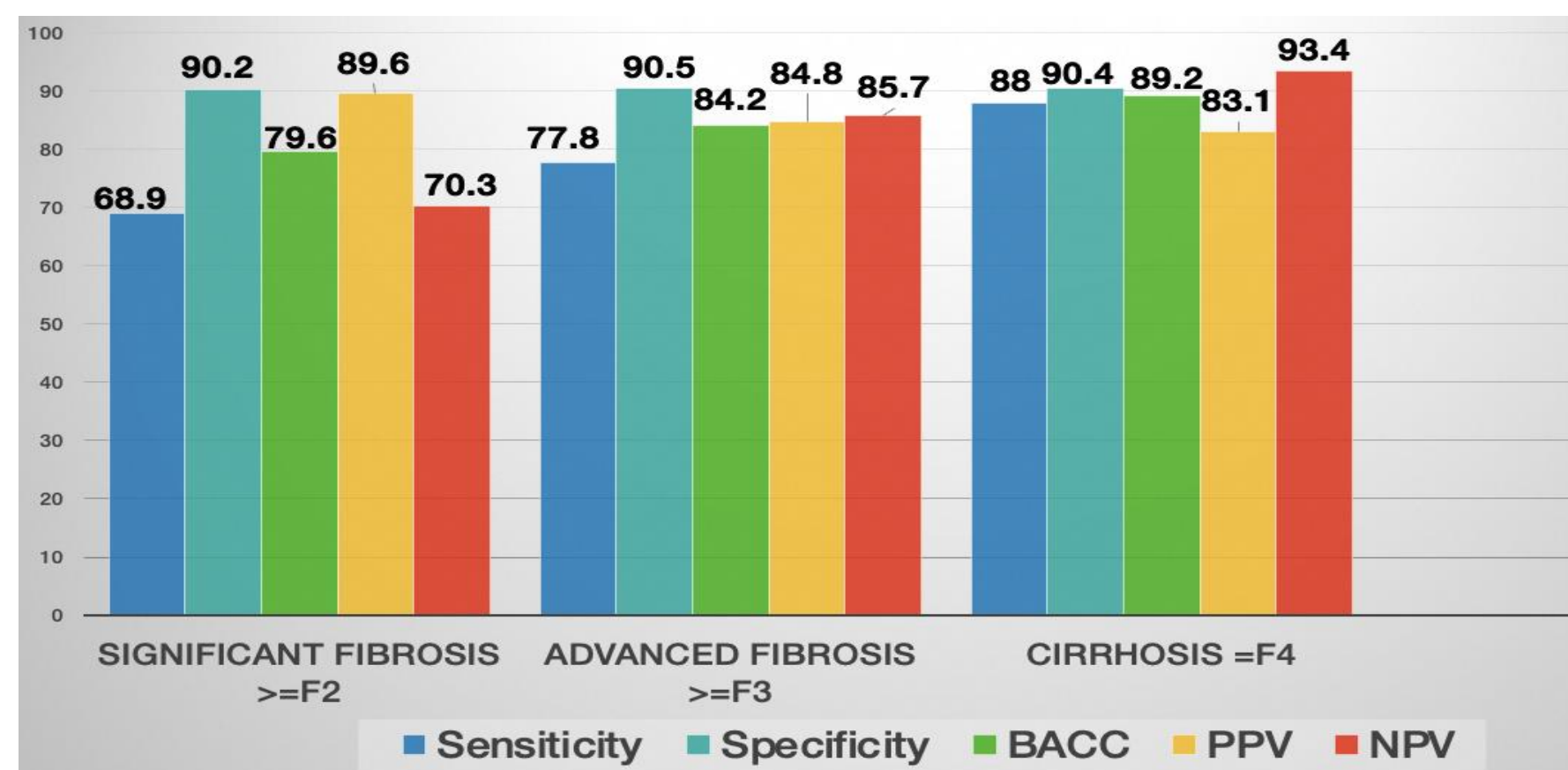
(3.b) Screening for Liver Cirrhosis and Esophageal Varices using Non-Contrast CT

- ◆ Screening for liver cirrhosis based on non-contrast CT
- ◆ Screening for esophageal varices in liver cirrhosis population using non-contrast CT

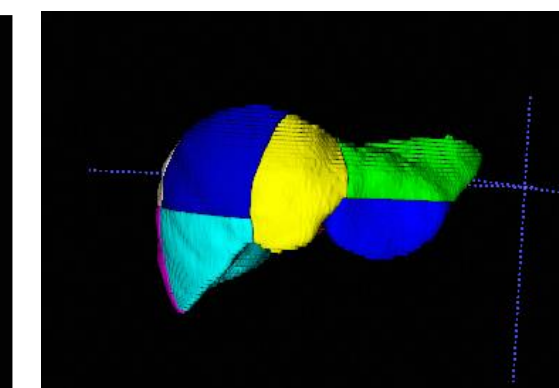
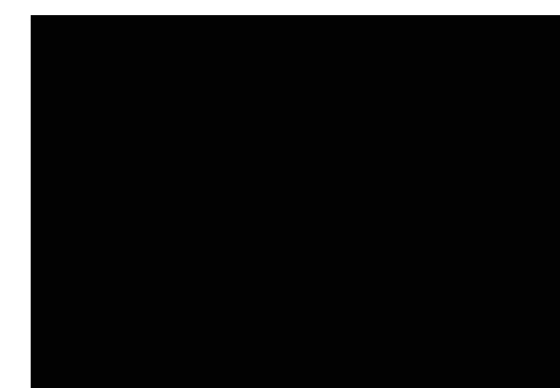
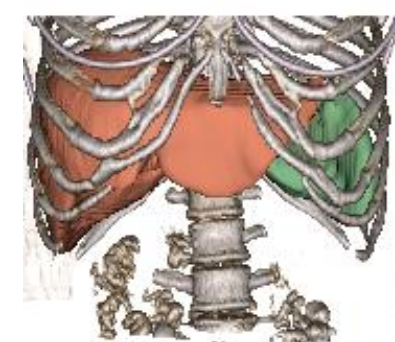
Current Project Progress

- ◆ Version 1 of the algorithm has been developed and internally independently tested
- ◆ Further validation of algorithm accuracy and generalization capability is required using data from external centers

Grading of Liver Fibrosis: Internal independent testing



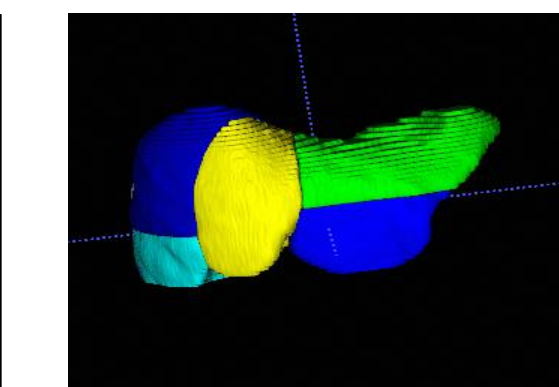
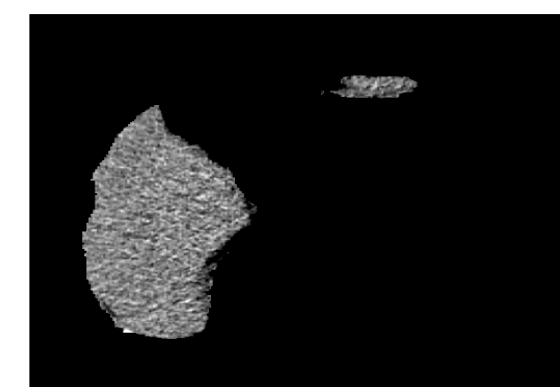
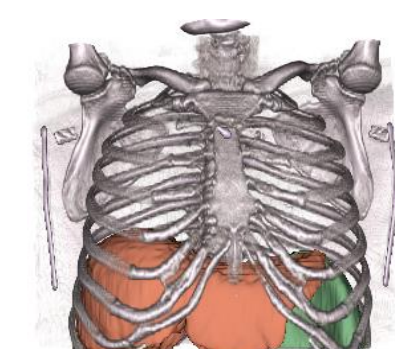
Patient X baseline: chronic hepatitis, mild, G2/F0, steatosis; scattered small cysts within the liver. 2012/11/24



AI prediction
F0, no fibrosis

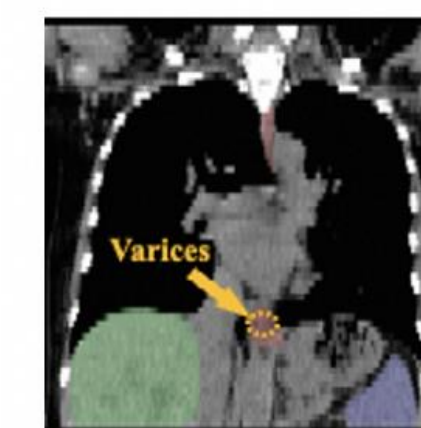
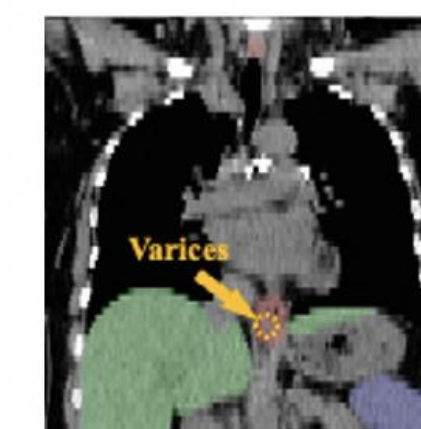
Precise patient longitudinal management

Patient X Outcome: liver cirrhosis and splenomegaly persisting as before. 2019/11/11

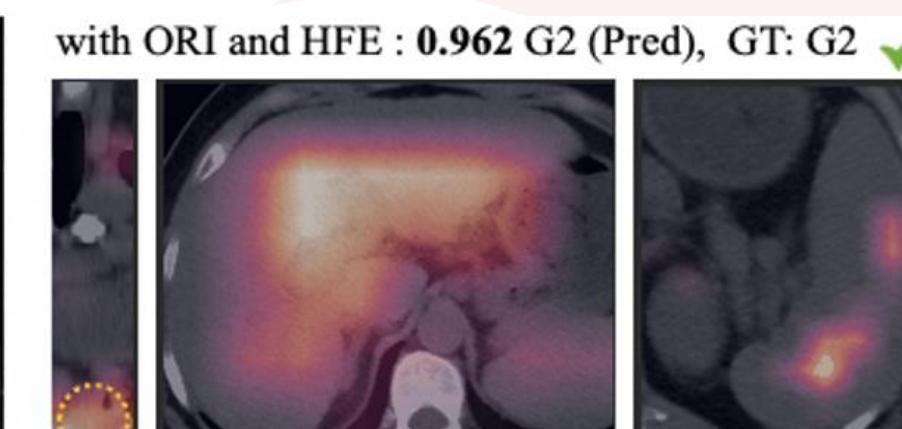


AI prediction
F3, advanced fibrosis

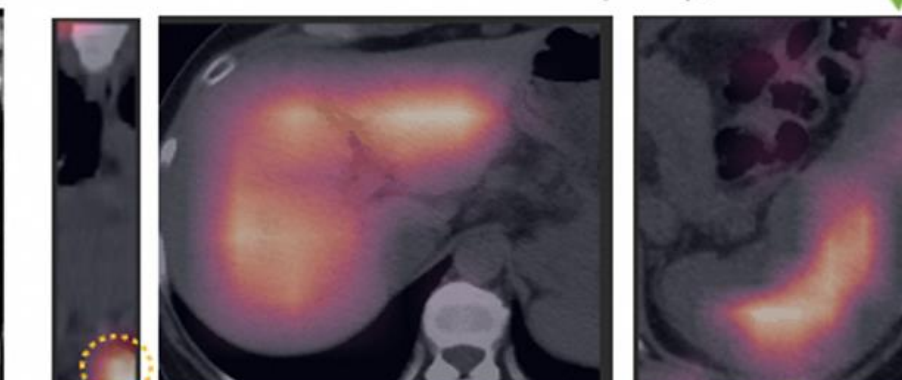
Esophageal varices screening in cirrhosis patients
AUC = 0.832, 0.864



Non Contrast CT



with ORI and HFE : 0.962 G2 (Pred), GT: G2 ✓



with ORI and HFE : 0.718 G3 (Pred), GT: G3 ✓

Esophagus Liver Spleen

Affected regions in the esophagus, liver, and spleen is highlighted

(4) Osteoporosis screening and diagnosis via non-contrast CT

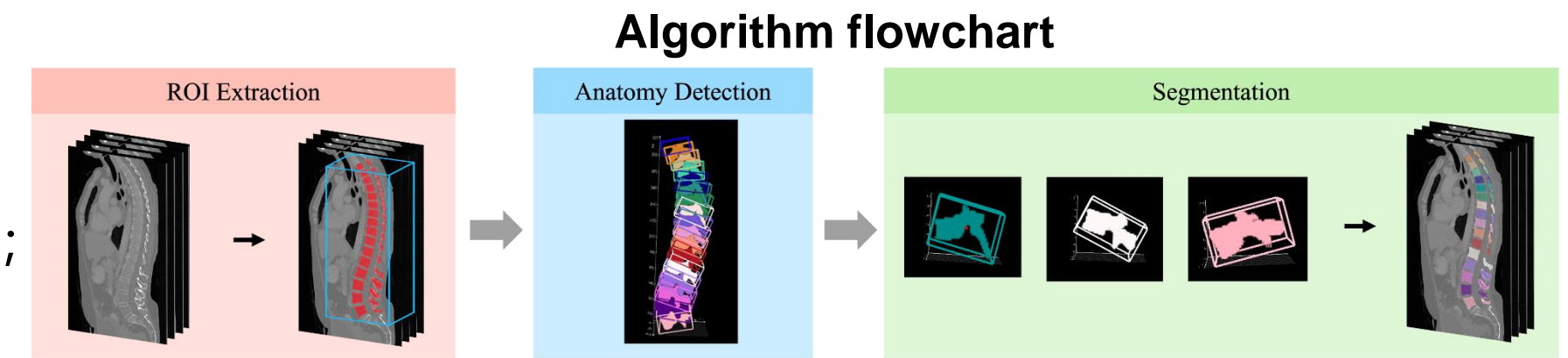
Overall Objective: To achieve intelligent estimation/tri-classification of bone mineral density (BMD) in any vertebral body of the human spine through non-contrast CT : Normal Bone Density/0; Bone Loss/Deficiency/1; Osteoporosis/2.

1. Optimization of Vertebral Body Instance Segmentation Model: Achieve a failure rate of less than 1% in vertebral identification and segmentation across multi-center datasets.
2. Prediction Model for Vertebral Body BMD Values: Demonstrate universal consistency with readings from at least three common DXA devices (Correlation Coefficient $R > 0.85$).
3. Product Implementation and Validation: Validate algorithms and refine models in 4-6 collaborating hospitals, leading to productization.

Progress Update:

1. Finished Algorithm V1.0 and conducted tests in Zhuji and Lishui with a correlation coefficient over $R > 0.85$.
2. Collected pairs of BMD data from three different DXA machines and corresponding CT scans, initiating the upgrade to V1.0.
3. A clinical paper with multi-center validation is being authored.

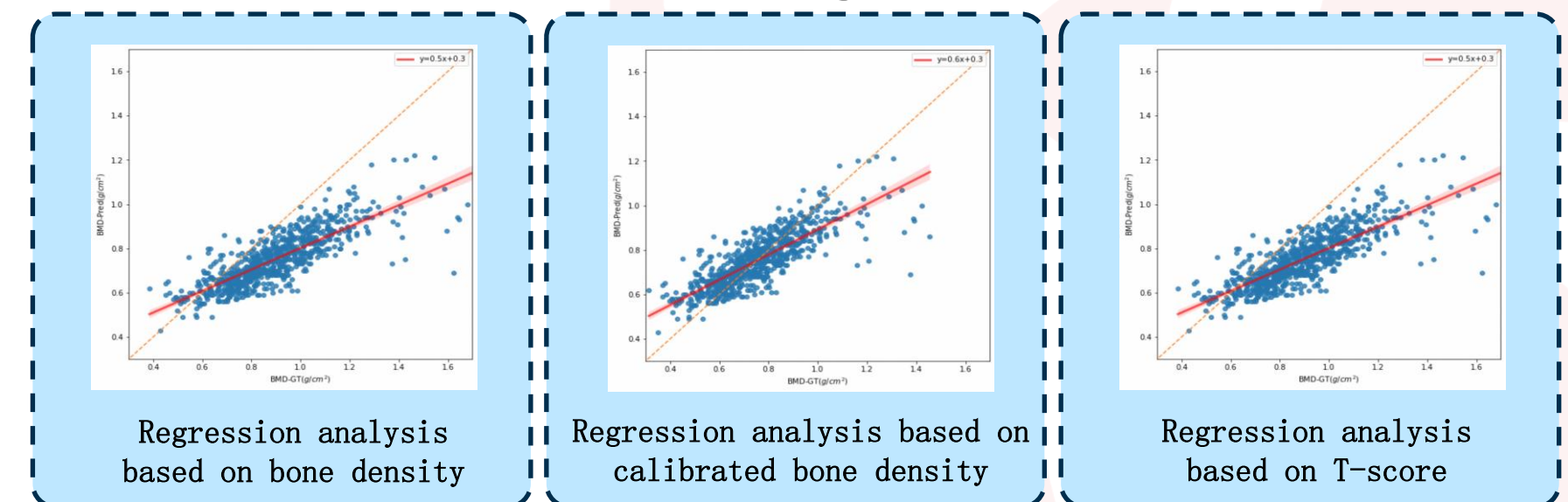
[1] Rib-Query: Steerable 9-DoF Rib Instance Segmentation and Labeling with Query, (JBHI, 2024) segmentation tool
 [2] Towards a Comprehensive, Efficient and Promptable Anatomic Structure Segmentation Model using 3D Whole-body CT Scans, CoRR abs/2403.15063 (2024)
 DAMO Academy's universal multi-organ interactive segmentation tool is open-sourced in the Alibaba Cloud ModelScope Community <https://modelscope.cn/models/xiuan123/CT-SAM3D>



DXA-CT Paired Data from Multiple Centers and Devices

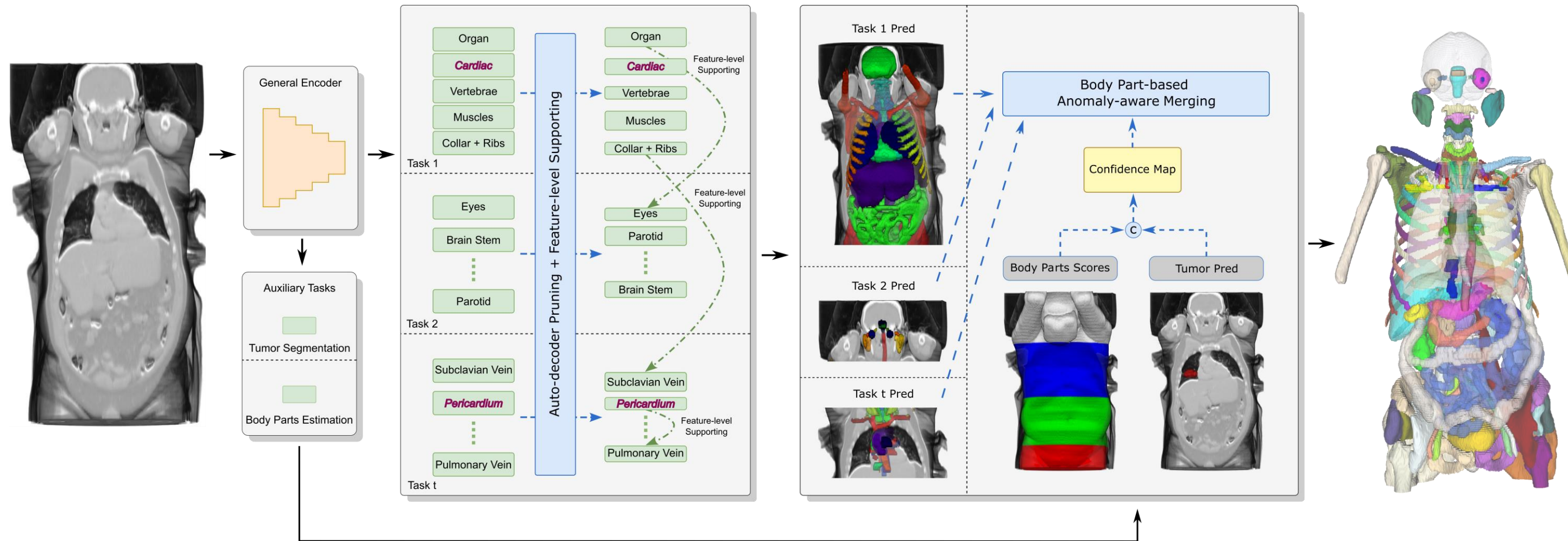
医院	数据来源	CT数据量	DXA数据量	DXA设备	备注
	湘雅骨病专科	2633	2633	Hologic	
	诸暨人民医院	372	372	Hologic	
	武汉协和本部	475	475	GE Lunar	
	武汉协和金银湖院区	880	TBD	TBD	
	武汉协和武科大院区	527	TBD	MEDIX 90	法国设备
	丽水中心医院	800	800	GE Lunar	仅有L1结果

Quantitative Test Results of Algorithm V1.0 on Lishui Dataset



(5) Whole-body organ segmentation on non-contrast CT scans / 236 organs

- By integrating multiple partial label datasets and employing a framework of universal encoder coupled with scalable and pruned decoders, we establish a **unified segmentation model for all 236 organs and lesions** throughout the body. This acts as a large-scale medical segmentation model empowering downstream segmentation, chronic disease management, muscle mass assessment, fat analysis, radiation therapy planning, and diagnosis.
- When multiple datasets are **not simultaneously accessible**, our **Continual Segmentation System (CSS)** dynamically updates the segmentation capability for select organs while preserving the segmentation accuracy for others (no forgetting) [1].
- When multiple datasets are **concurrently available**, our **Universal Segmentation Model** facilitates easier optimization, enabling the model to gain broader knowledge across these datasets, thereby enhancing segmentation precision [2,3].



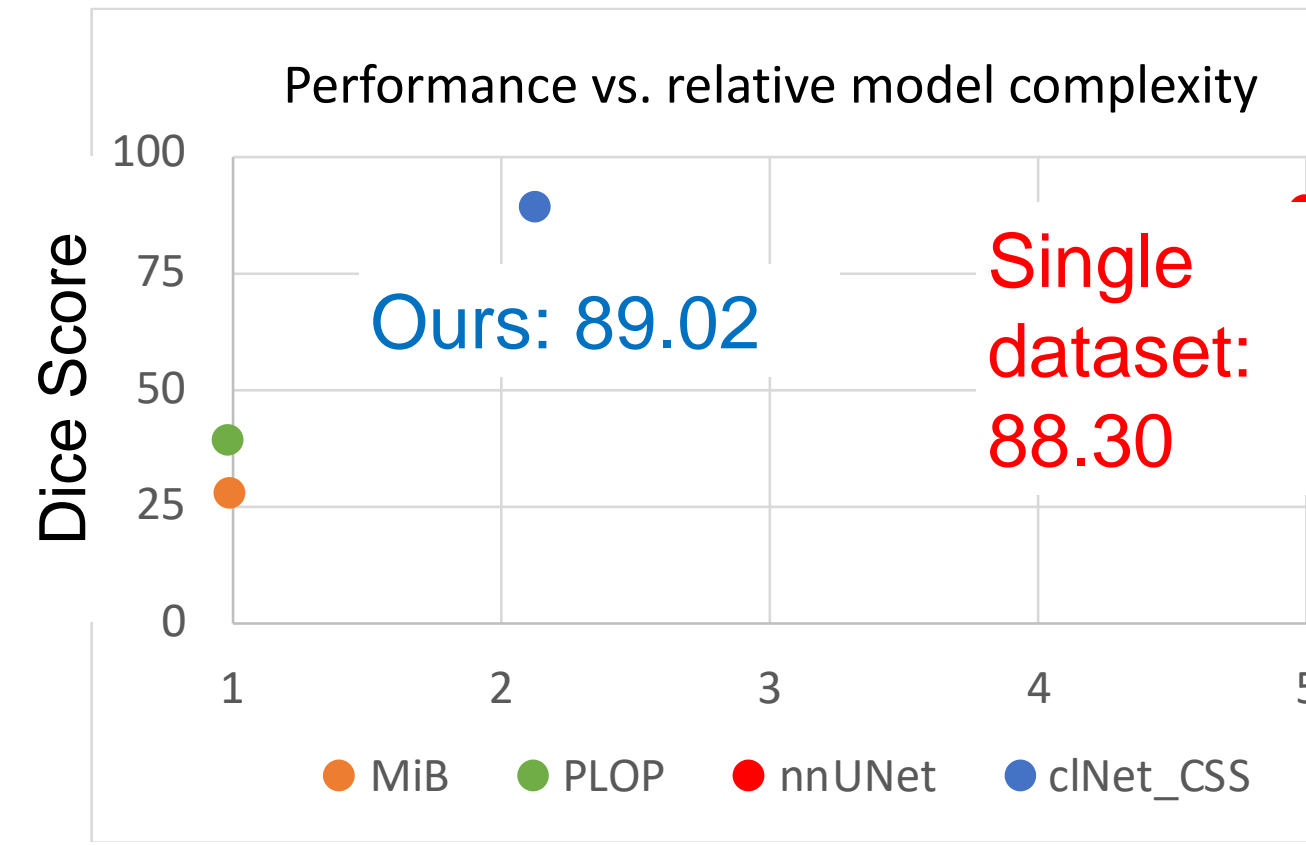
Currently, we are compiling additional results and optimizing inference for parallel outputs. Open-sourced within the Alibaba Cloud ModelScope Community. Model as a Service!

<https://github.com/alibaba-damo-academy/ct-sam3d>
<https://modelscope.cn/models/xiuan123/CT-SAM3D>



[1] Ji, Z. et al., "Continual segment: Towards a single, unified and non-forgetting continual segmentation model of 143 whole-body organs in CT scans", IEEE ICCV, 2023
 [2] Zhu, V. et al., "Low-Rank Continual Pyramid Vision Transformer: Incrementally Segment Whole-Body Organs in 3D CT Scans with Light-Weighted Adaptation", MICCAI 2024
 [3] Guo, H. et al., "Towards a Comprehensive, Efficient and Promptable Anatomic Structure Segmentation Model using 3D Whole-body CT Scans", CoRR abs/2403.15063 (2024, in review)
 [4] Ulrich et al., "Multitalent: A multi-dataset approach to medical image segmentation", MICCAI, 2023

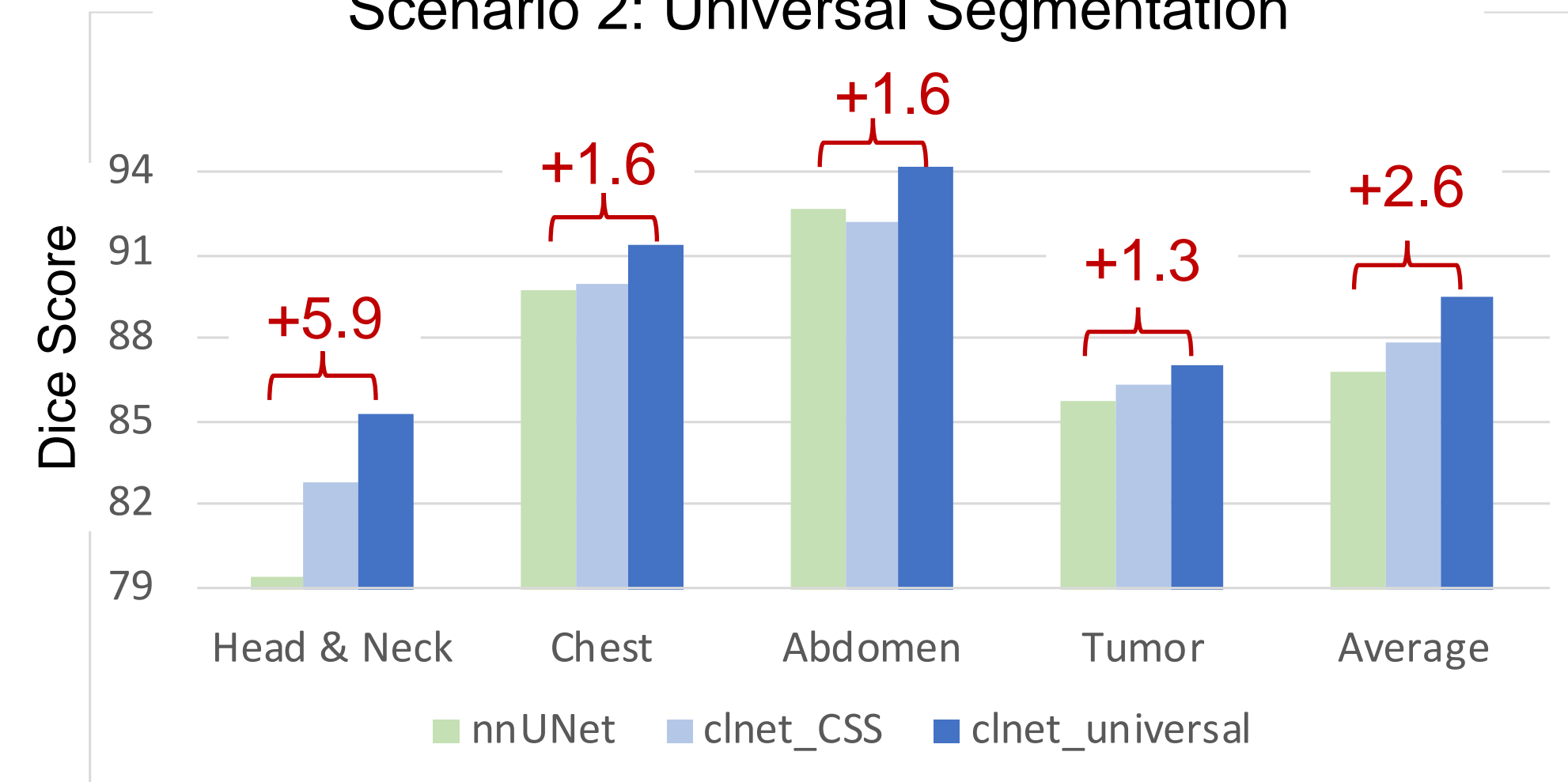
Scenario 1: Continuous Segmentation



5 testing datasets:

- ✓ Surpasses nnUnet trained on single datasets.
- ✓ Parameter count is only about 40%

Scenario 2: Universal Segmentation



- ✓ Our Universal Model surpasses CSS model, achieving 2.6% in Dice Similarity Coefficient (DSC) per organ compared to nnUNet trained on single datasets!

1 **Unified, Accurate, Generalizable and Non-forgetting Con-**
2 **tinual Segmentation Models of Fine-grained 236 Whole-**
3 **body Organs, Lymph Node Stations and Lesions in 3D CT**
4 **Scans: Is Organ Segmentation in CT a Solved Problem?**

5 Dazhou Guo^{1†}, Zhanghexuan Ji^{1†}, Yanzhou Su^{1,3†}, Dandan Zheng^{2†}, Heng Guo^{1,3}, Puyang Wang⁴,
6 Ke Yan^{1,3}, Yirui Wang¹, Zi Li^{1,3}, Minfeng Xu^{1,3}, Qifeng Wang⁵, Na Shen⁶, Tsung-Ying Ho⁷, Jia
7 Ge², Yun Bian⁸, Hua Zhang⁹, Alan L. Yuille¹⁰, Ronald M. Summers¹¹, Zhiyong Lu¹², Perry J.
8 Pickhardt¹³, Cher Heng Tan¹⁴, Chunyan Miao^{15,16}, Senxiang Yan^{2*}, Le Lu^{1*}, Dakai Jin^{1*}, Xianghua
9 Ye^{2*}

10 ¹*DAMO Academy, Alibaba Group*

11 ²*Department of Radiation Oncology, The first Affiliated Hospital, Zhejiang University, Hangzhou,*
12 *China*

13 ³*Hupan Laboratory, Hangzhou, China*

14 ⁴*School of Medicine, Johns Hopkins University, Baltimore, United States of America*

15 ⁵*Department of Radiation Oncology, Sichuan Cancer Hospital, Sichuan, China*

16 ⁶*Department of Otolaryngology-Head & Neck Surgery, Zhongshan Hospital, Fudan University,*
17 *Shanghai, China*

18 ⁷*Department of Nuclear Medicine, Chang Gung Memorial Hospital, LinKou, ROC*

19 ⁸*Department of Radiology, Shanghai Institution of Pancreatic Disease, Shanghai, China*

20 ⁹*Linking Med Inc., Beijing, China*

21 ¹⁰*Department of Computer Science, Johns Hopkins University, Baltimore, United States of America*

22 ¹¹*Radiology and Imaging Sciences, National Institutes of Health Clinical Center, Bethesda, United*
23 *States of America*

24 ¹²*National Library of Medicine, National Institutes of Health, Bethesda, United States of America*

25 ¹³*Radiology & Medical Physics, School of Medicine & Public Health, University of Wisconsin,*
26 *Madison, United States of America*

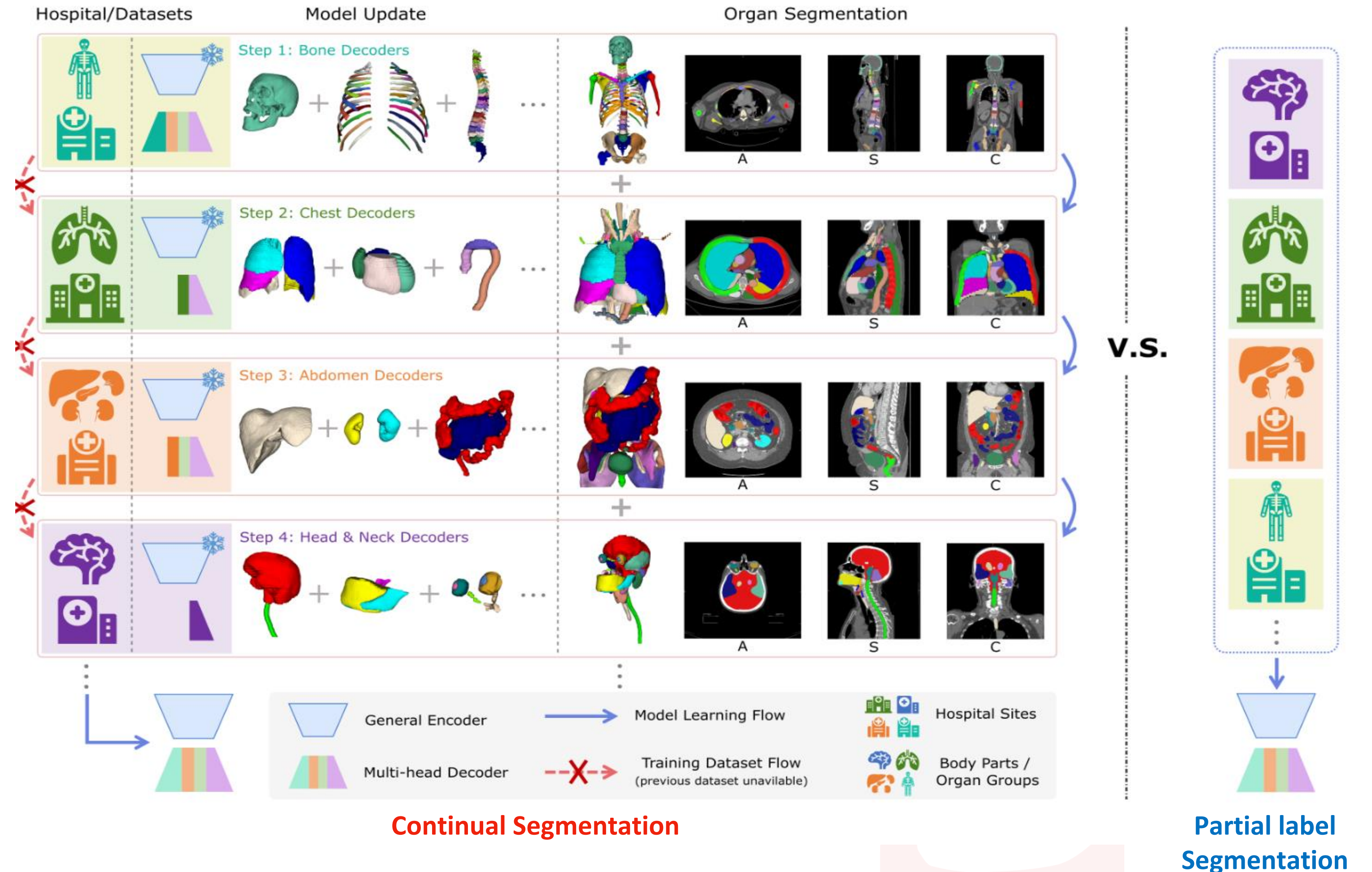
27 ¹⁴*Department of Diagnostic Radiology, Tan Tock Seng Hospital, Singapore*

28 ¹⁵*Centre of AI in Medicine, Nanyang Technological University, Singapore*

29 ¹⁶*Alibaba-NTU Global e-Sustainability CorpLab (ANGEL), Nanyang Technological University,*
30 *Singapore*



- Despite claimed as foundation models, none of recent medical SAM-type models^[1,2] have truly surpassed the performance of dataset-specific trained nnUNet^[3] under rigorous and fair evaluation.
- We tackle universal segmentation from a new perspective of **continual learning** and propose a **unified, scalable and non-forgetting** Continual Learning model, **CL-Net^[4,5]**
 - ❑ Harness the **synergies** of a large set of partially labeled datasets to accurately segment all anatomies present across all datasets
 - ❑ **Dynamically extend** to segment new anatomies without compromising previously learned knowledge, even without access to prior training datasets
 - ❑ **20 public and 16 private datasets** (collection of our previous peer-reviewed work) from different institutions with various vendors, phases and pathologies, resulting in a substantial total of **13,952 CT scans** covering **236 whole-body anatomies: 193 organs/sub-organs, 33 lymph node stations, and 10 tumor GTVs.**



[1] Zhao, Z. et al. One model to rule them all: Towards universal segmentation for medical images with text prompts. *Nature Communications under review* (2024)

[2] Zhao, T. et al. Biomedparse: a biomedical foundation model for image parsing of everything everywhere all at once. *Nature Methods* (2024)

[3] Isensee, F., et al. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods* (2021)

[4] Zhang et al. "Continual segment: Towards a single, unified and non-forgetting continual segmentation model of 143 whole-body organs in ct scans", *IEEE ICCV* (2023)

[5] Guo, D. et al., "Unified, Accurate, Generalizable and Non-forgetting Continual Segmentation Model of Fine-grained 236 Whole body Organs, Lymph Node Stations and Lesions in CT scans", *in submission* (2024)

Universal Segmentation for Fine-grained 236 Whole-body anatomies

Table 1 | Evaluation results of nnUNet^{E36}, CL-Net^{U36_{unpruned}}, CL-Net^{C36} and CL-Net^{U36} on different whole body organ groups. All: average over all 236 classes. There metrics are reported to evaluate

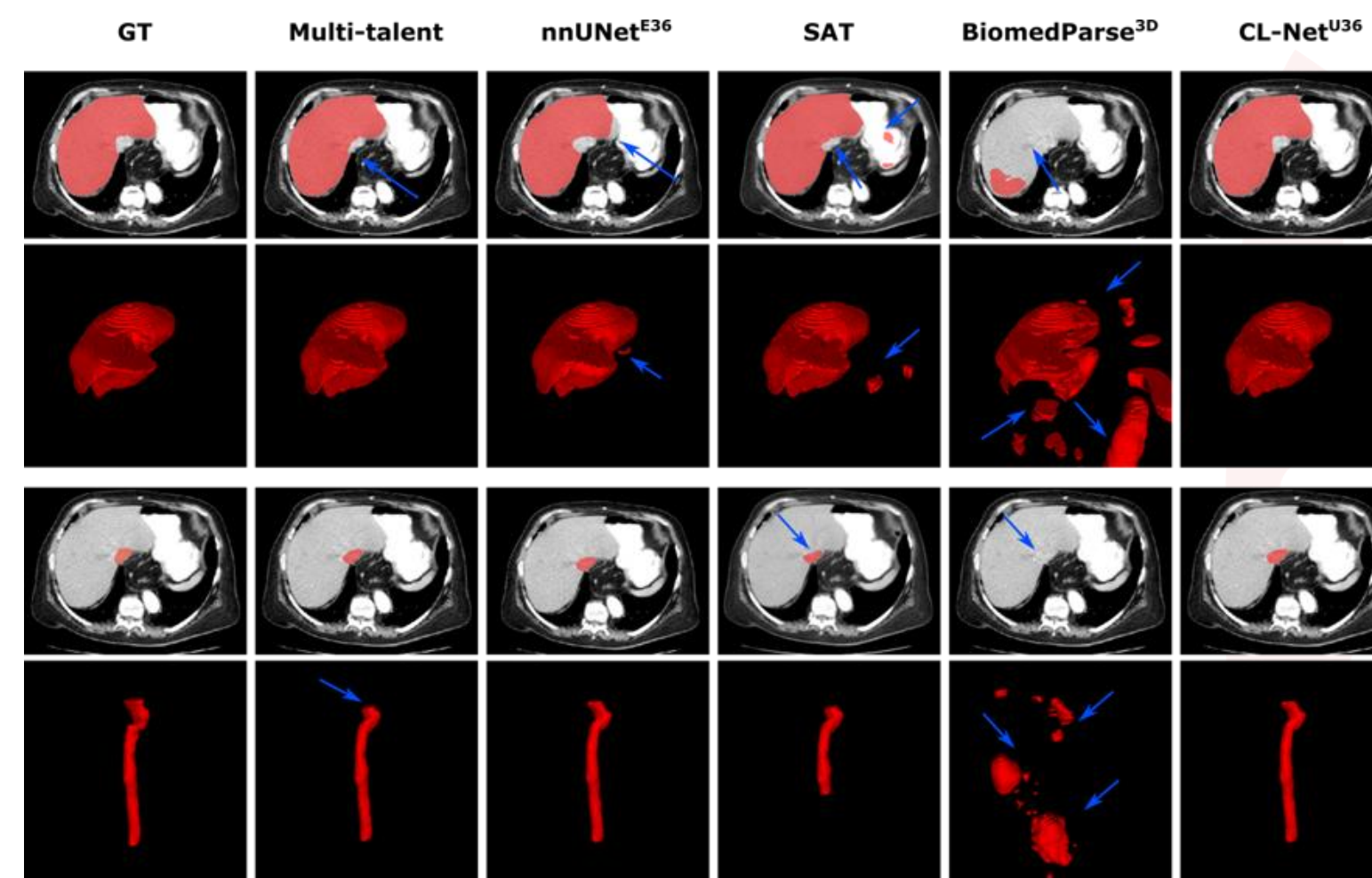
Metric	Method	Head & Neck (49)	Chest (52)	Abdomen (28)	Bone (63)	LNS (33)	Lesion (10)	Body (1)	All (236)
DSC↑	nnUNet ^{E36}	74.26	85.45	90.81	91.92	72.11	70.31	96.97	83.03
	CL-Net ^{U36_{unpruned}}	80.63	88.15	91.85	92.26	72.27	70.81	97.17	85.21
	CL-Net ^{C36}	79.60	88.03	91.51	92.14	72.23	70.67	97.09	84.89
	CL-Net ^{U36}	80.64	88.15	91.84	92.28	72.27	70.89	97.11	85.22
ASD↓	nnUNet ^{E36}	1.18	1.37	1.14	0.86	1.35	3.59	2.2	1.26
	CL-Net ^{U36_{unpruned}}	0.96	1.18	0.98	0.83	1.29	3.64	2.29	1.14
	CL-Net ^{C36}	1.01	1.19	1.02	0.83	1.30	3.67	2.33	1.16
	CL-Net ^{U36}	0.95	1.17	0.97	0.82	1.29	3.48	2.5	1.13
Dec. Param#	nnUNet ^{E36}								1126.8*
	CL-Net ^{U36_{unpruned}}	438.2	485.2	297.4	172.2	31.3	156.5	15.7	1596.5
	CL-Net ^{C36}	12.2	6.9	3.5	2.0	3.8	11.6	0.1	40.0
	CL-Net ^{U36}	12.1	6.9	3.5	2.0	3.8	11.6	0.1	39.9

Table 2 | Performance comparison of PLS models on five representative public datasets. The mean DSC (%) and ASD (mm) of CL-Net^{U5}, nnUNet^{E5}, MultiTalent^[2], and SAT-pro^[3] are evaluated on five representative public datasets. CL-Net^{U5} achieves the top performance in both DSC and ASD. Note that the officially released SAT-pro model removes image metadata headers during inference, preventing the calculation of ASD.

Metric	Method	TotalSegmentator	StructSeg19	FLARE22	SegThor	KiTS21	Mean	
DSC↑	nnUNet ^{E5}	94.06	86.39	89.87	92.58	86.70	92.53	
	MultiTalent ^[2]	92.55	78.96	89.29	90.66	79.57	90.21	
	SAT-pro ^[3]	86.38	62.87	85.13	90.34	79.98	83.00	
	CL-Net ^{U5}	94.28	86.95	90.73	93.71	87.16	92.88	
ASD↓	nnUNet ^{E5}	0.97	0.29	1.14	0.34	1.18	0.87	
	MultiTalent ^[2]	0.81	1.95	1.83	1.44	9.94	1.13	
	SAT-pro ^[3]							
	CL-Net ^{U5}	0.85	0.27	1.04	0.32	1.03	0.77	

Table 3 | Performance comparison of CL-Net^{U5}, nnUNet^{E5}^[1], and BiomedParse^[24] on 13 organs of FLARE22. Note that DSC* (2D slice-level) are 2D segmentation results directly copied from BiomedParse^[24].

Metric	Method	Liver	Kidney_R	Kidney_L	Spleen	Pancreas	A.Aorta	V.VenaCava.I
DSC* (2D slice-level)	BiomedParse ^[6]	96.6	95.7	96.3	96.1	86.7	94.8	89.6
DSC↑	BiomedParse ^[6]	24.1±19.8	53.9±7.5	64.9±3.6	62.3±6.1	2.4±0.4	27.2±0.5	11.6±1.5
	nnUNet ^{E5}	98.1±14.5	92.4±16.9	92.8±16.9	97.9±15.6	91.6±17.2	96.7±16.5	88.8±10.1
	CL-Net ^{U5}	98.3±5.8	94.3±7.1	92.9±6.7	98.0±6.6	92.4±7.0	97.1±16.5	88.6±9.9
ASD↓	BiomedParse ^[6]	48.34±26.44	57.48±14.48	69.38±10.18	51.41±8.09	60.92±0.61	64.15±3.13	70.89±8.90
	nnUNet ^{E5}	0.83±0.98	1.81±0.74	92.8±16.9	0.30±0.37	0.80±0.69	0.21±0.95	1.08±1.03
	CL-Net ^{U5}	0.81±1.00	1.44±0.76	1.18±0.69	0.19±0.37	0.72±0.38	0.18±0.83	1.12±0.98
Metric	Method	GlnD_Adrenal.R	GlnD_Adrenal.L	Gallbladder	Eso	Stomach	Duodenum	Mean
DSC* (2D slice-level)	BiomedParse ^[6]	76.3	79.0	86.8	85.0	92.8	81.1	89.0
DSC↑	BiomedParse ^[6]	0.0±0.0	0.0±0.0	1.0±1.0	59.4±13.7	0.8±0.0	1.0±1.0	23.7
	nnUNet ^{E5}	85.9±12.3	88.0±13.3	80.2±24.7	84.4±3.1	91.7±16.4	79.8±25.9	89.9
	CL-Net ^{U5}	87.9±13.2	87.7±12.5	83.1±8.7	86.0±3.0	92.4±6.5	81.0±16.1	90.7
ASD↓	BiomedParse ^[6]	79.70±2.33	102.87±16.02	134.81±19.78	34.80±28.08	59.16±2.60	58.08±7.51	68.62
	nnUNet ^{E5}	0.39±0.37	0.26±1.16	2.16±0.83	1.70±0.15	1.24±0.67	2.71±1.29	1.14
	CL-Net ^{U5}	0.35±0.39	0.21±0.38	2.12±0.83	1.61±0.14	1.17±0.69	2.39±1.18	1.04



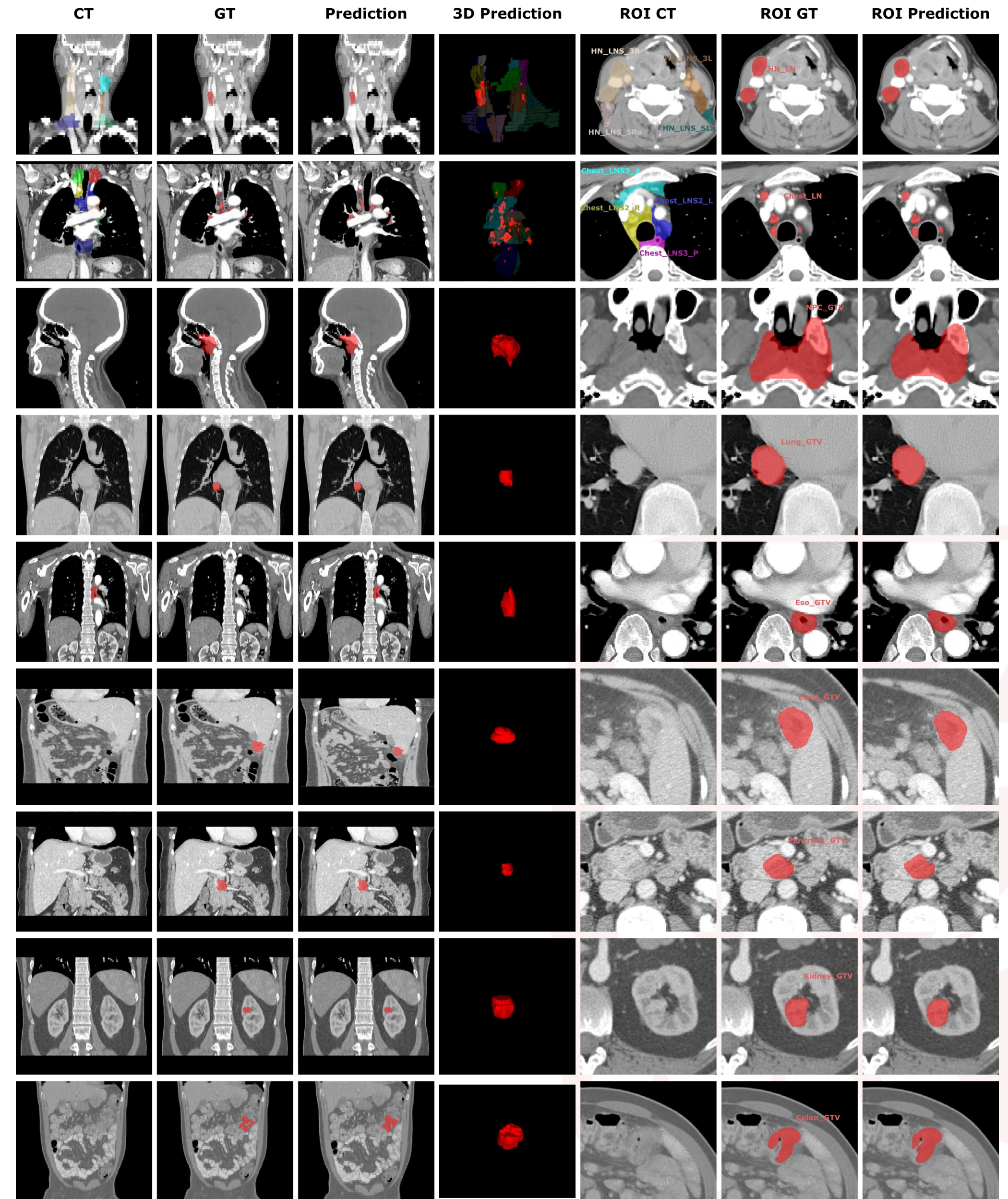
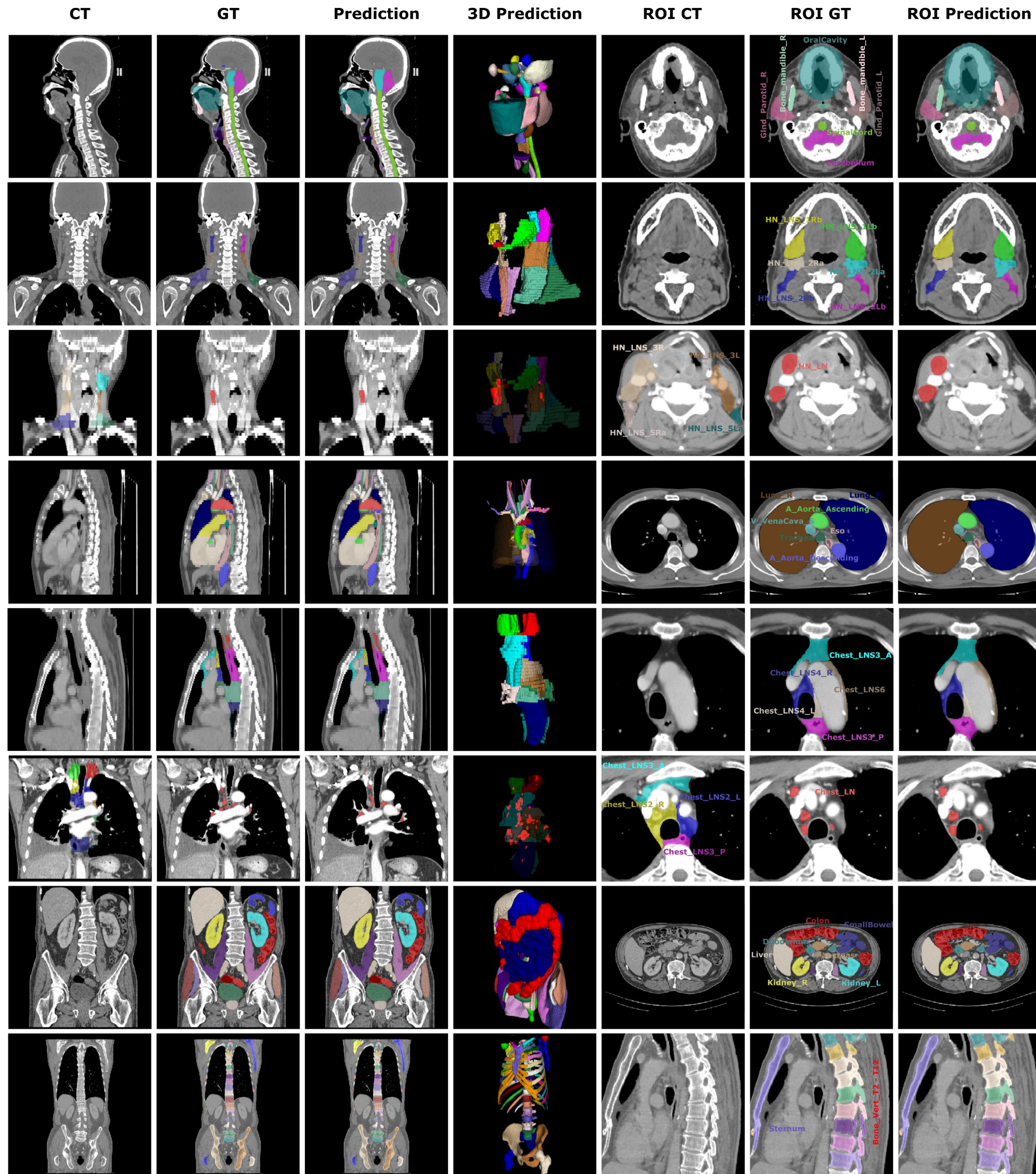
- CL-Net achieves consistent superior or on par accuracy with dataset-specific trained nnUNet
- CL-Net outperforms nnUNet-extended universal model MultiTalent^[2] by +2.64% DS
- Significantly outperforms SAM-type medical imaging foundation models by large margins, e.g., SAT^[3] (+9.9% DS), BiomedParse^[6] (+67.0% DS).

[2] Ulrich et al. "Multitalent: A multi-dataset approach to medical image segmentation", *MICCAI*, 2023

[3] Zhao, Z. et al. One model to rule them all: Towards universal segmentation for medical images with text prompts. *Nature Communications under review* (2024)

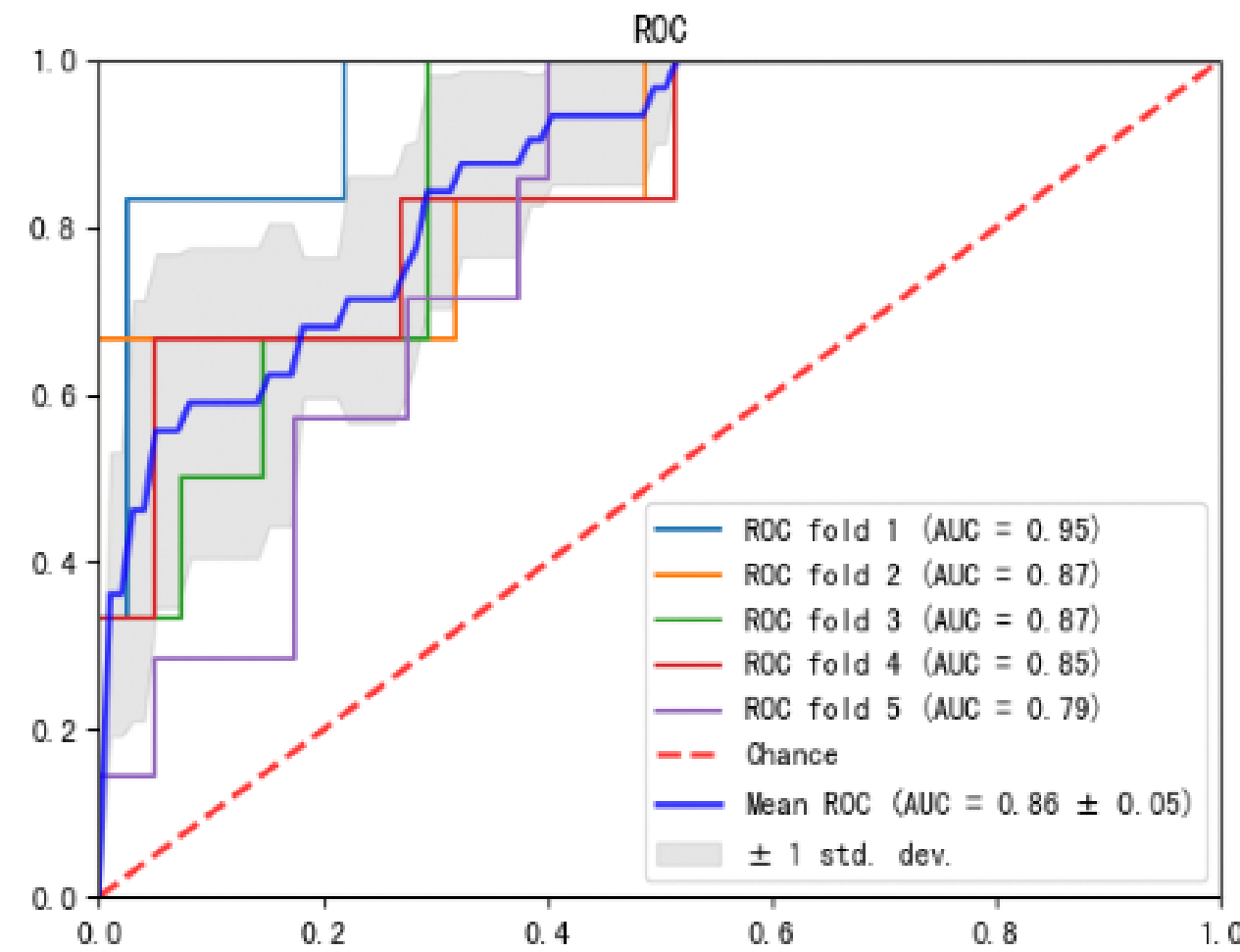
[6] Zhao, T. et al. Biomedparse: a biomedical foundation model for image parsing of everything everywhere all at once. *Nature Methods* (2024)

Universal Segmentation for Fine-grained 236 Whole-body anatomies

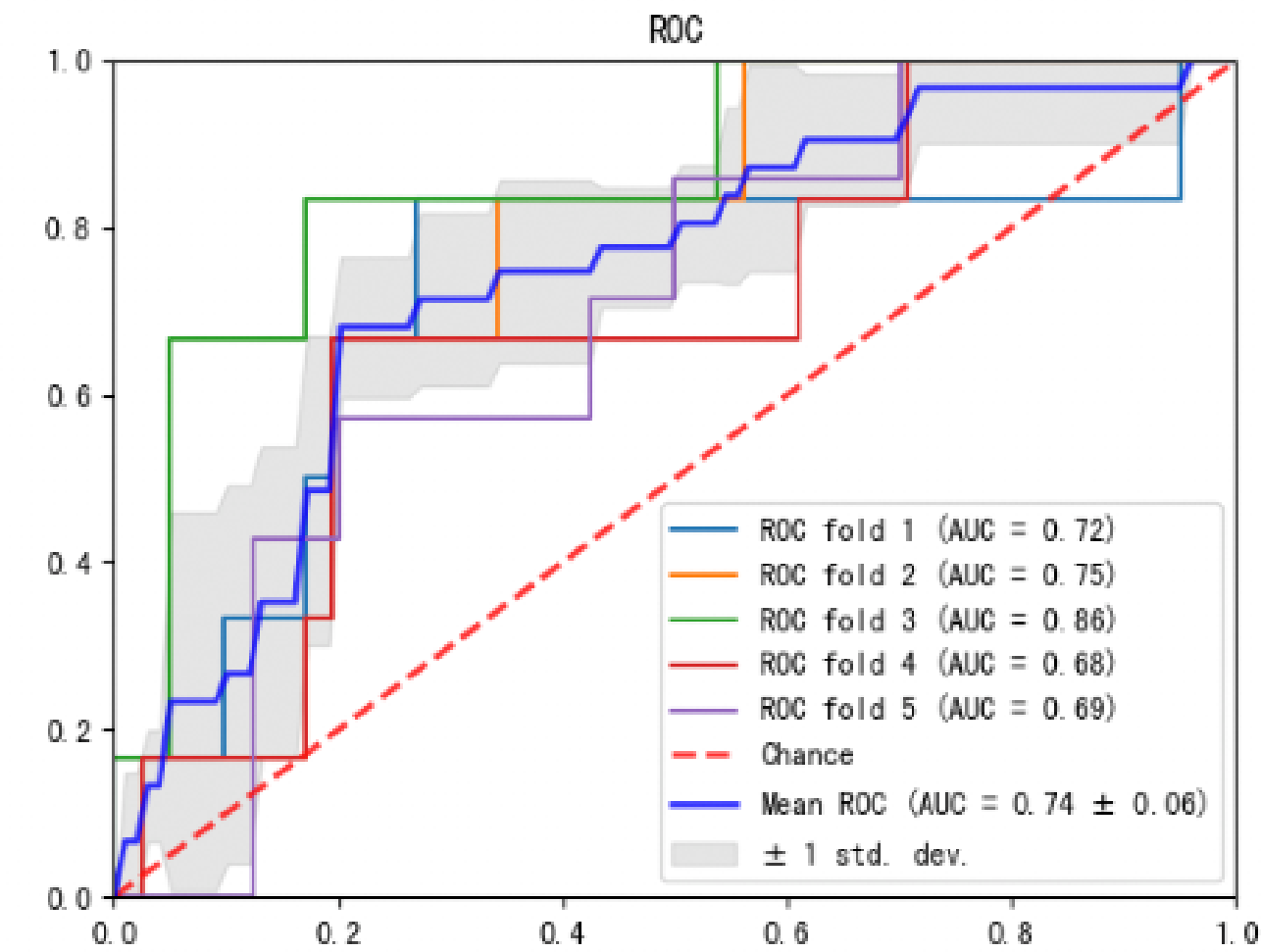


Whole-body/chest/abdominal muscle and fat segmentation on non-contrast CT scans - management of chronic disease treatment response (Severe Fever with Thrombocytopenia Syndrome, SFTS)

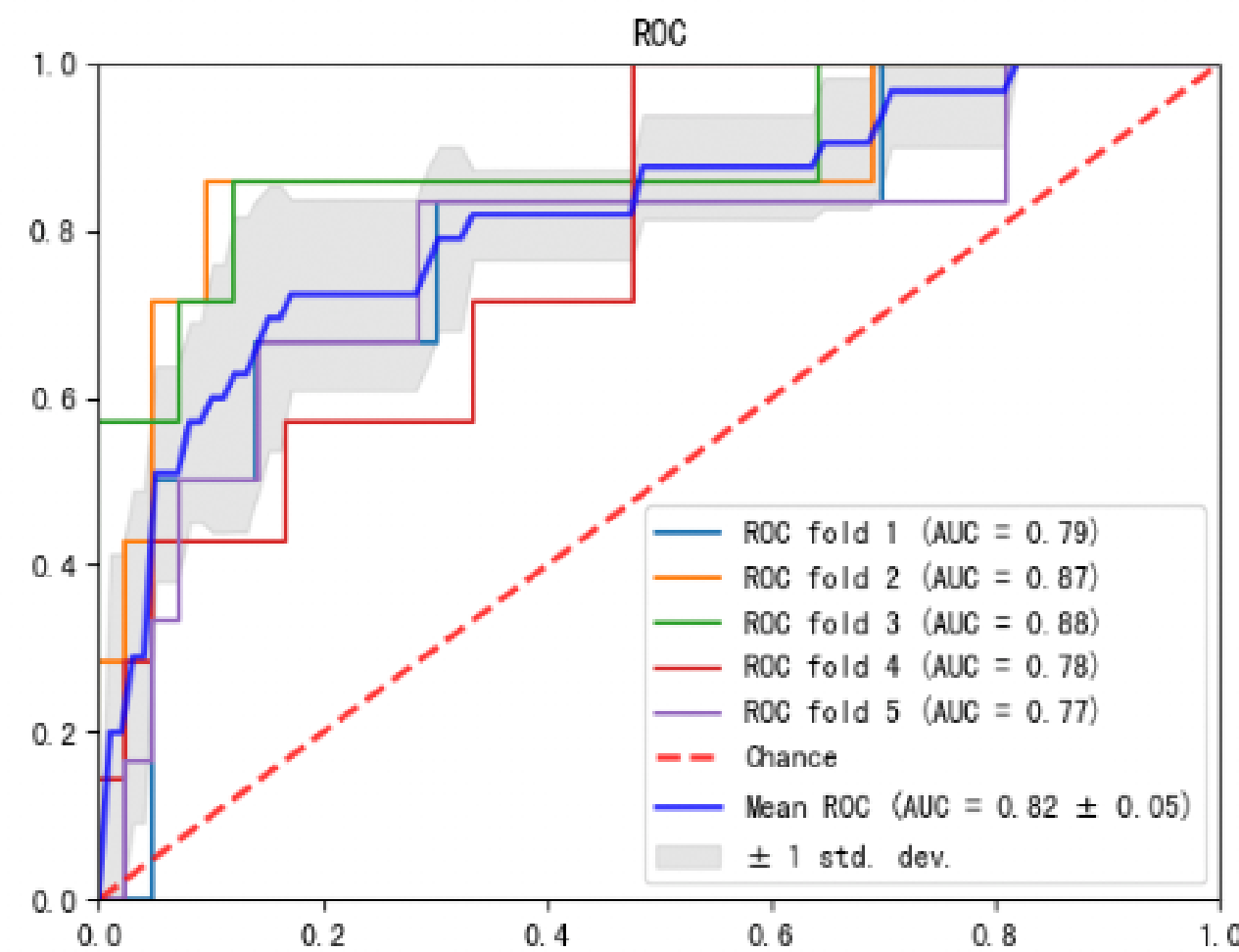
仅使用 radiomics 特征:



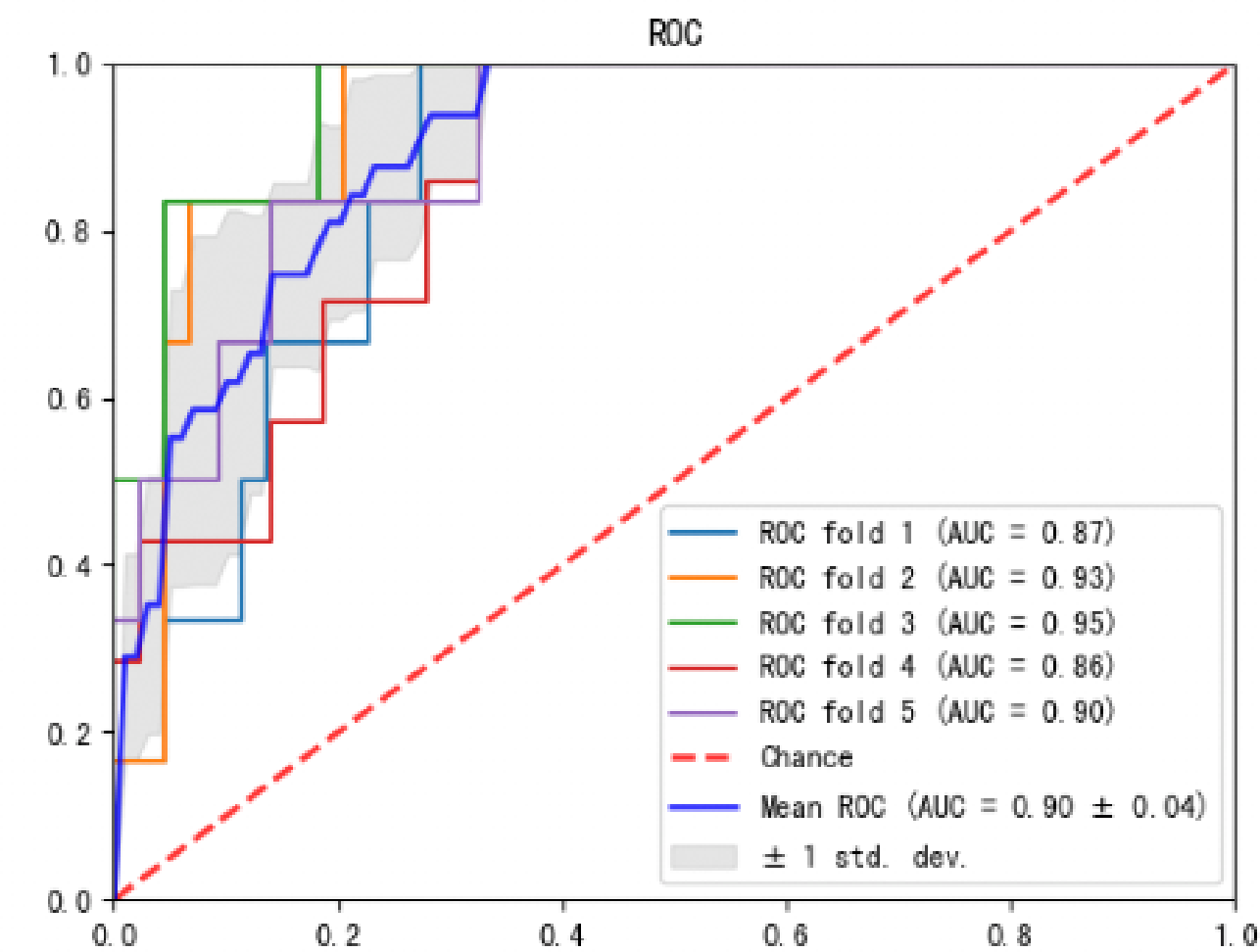
仅使用肌肉脂肪特征:



仅使用临床特征:



使用组合特征:



(6) Full brain area segmentation based on MRI / 106 brain regions

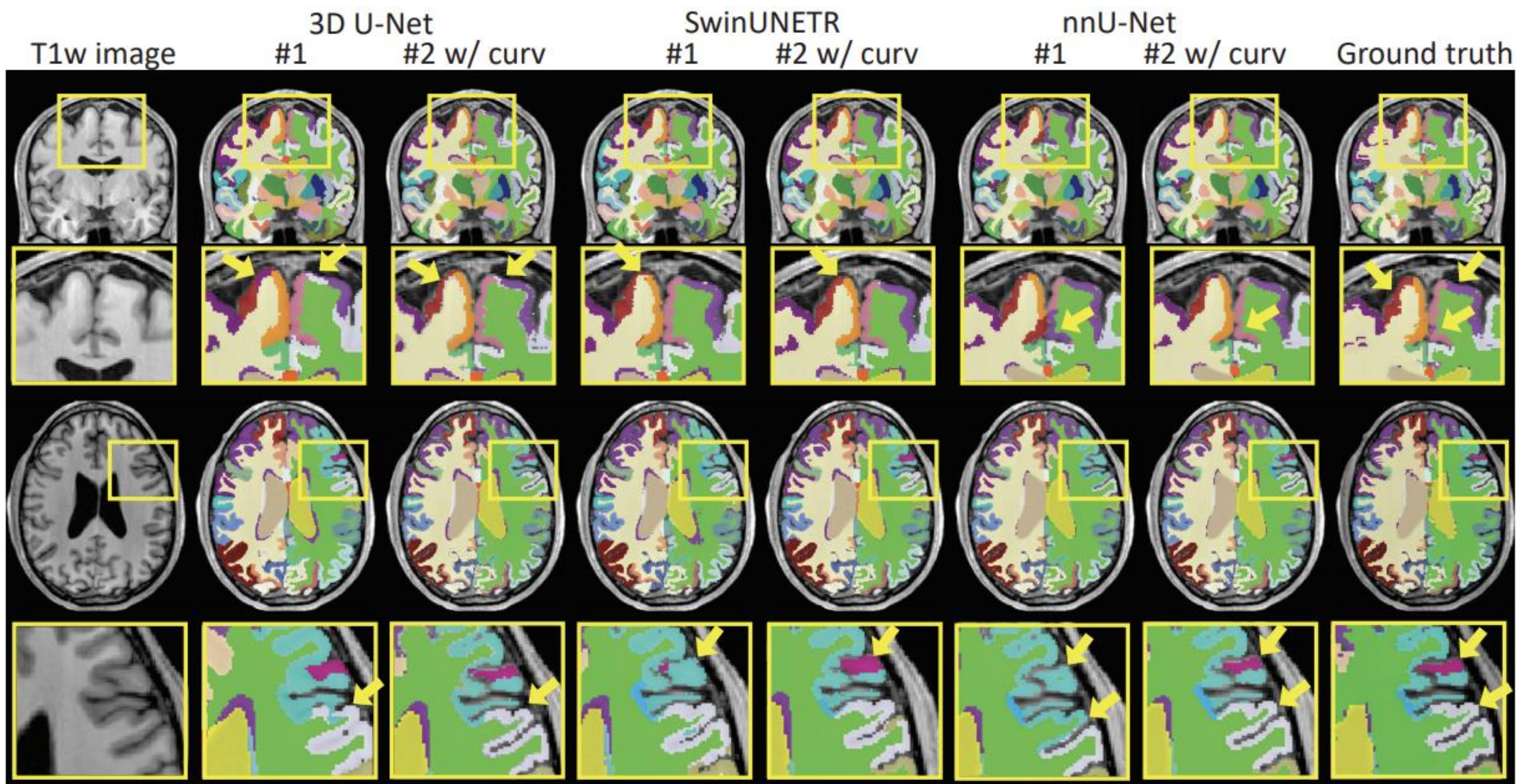
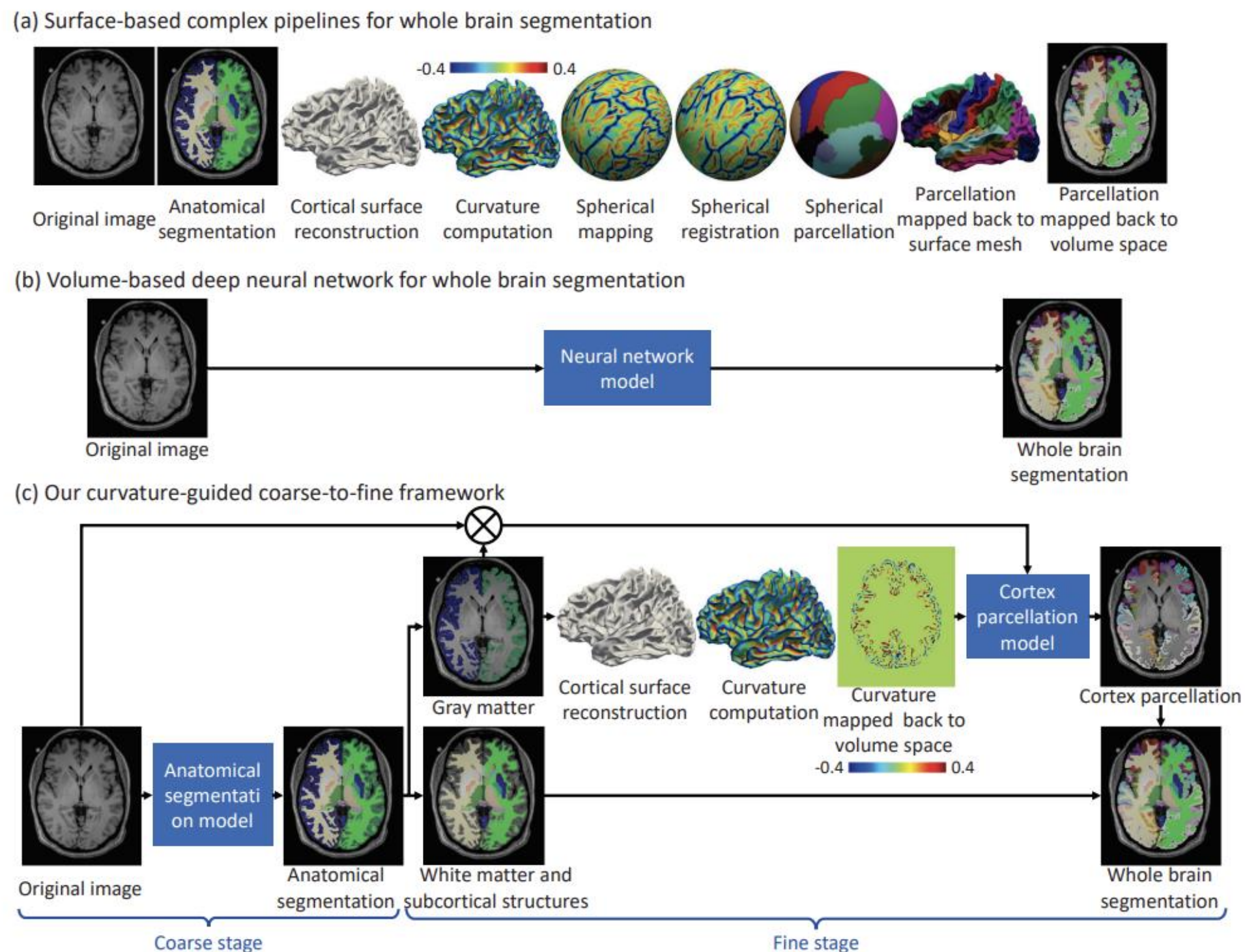


Table 2. Comparison with state-of-the-art whole brain segmentation methods.

	SS Dice (%)	WM Dice (%)	GM Dice (%)	All Dice (%)	Time
FreeSurfer [6]	76.28	91.22	86.45	82.84	~5h
UNesT [24]	74.59	88.32	73.49	74.20	~10s
Our method (nnU-Net #2 w/ curv)	77.42	94.46	83.13	81.36	~10s

106 manually labeled ROIs include 31 cortex labels in each hemisphere, 1 WM label in each hemisphere, 41 SS labels, and 1 background label.

[https://modelscope.cn/models/...](https://modelscope.cn/models/) Open-sourced in the Alibaba Cloud ModelScope community: Model as a Service!

4. How to adequately and rapidly address the very next step clinical needs for our screened-positive patient population ...

Cancer Precision Diagnosis & Treatment Recommendation (AI-MDT)

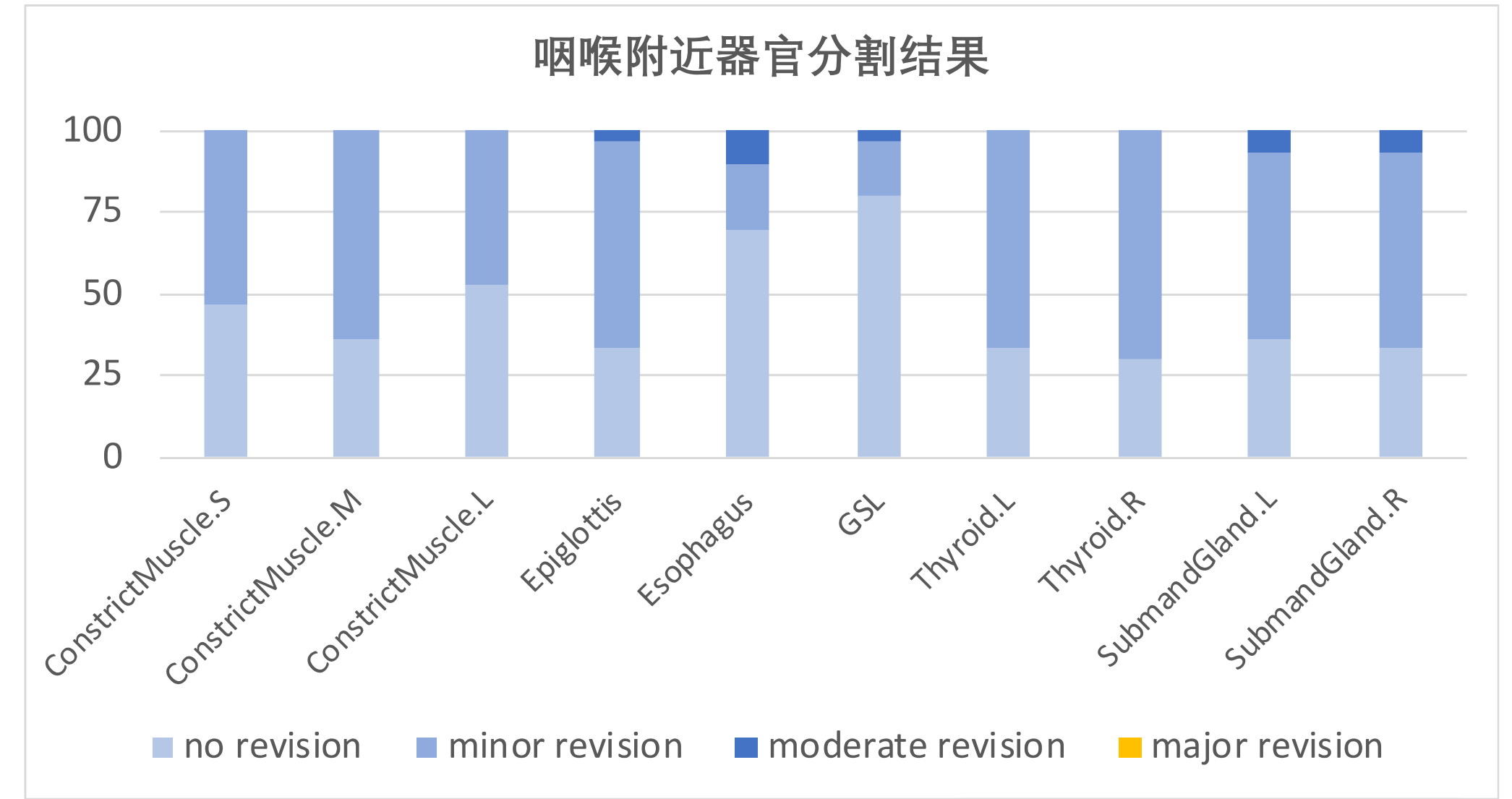
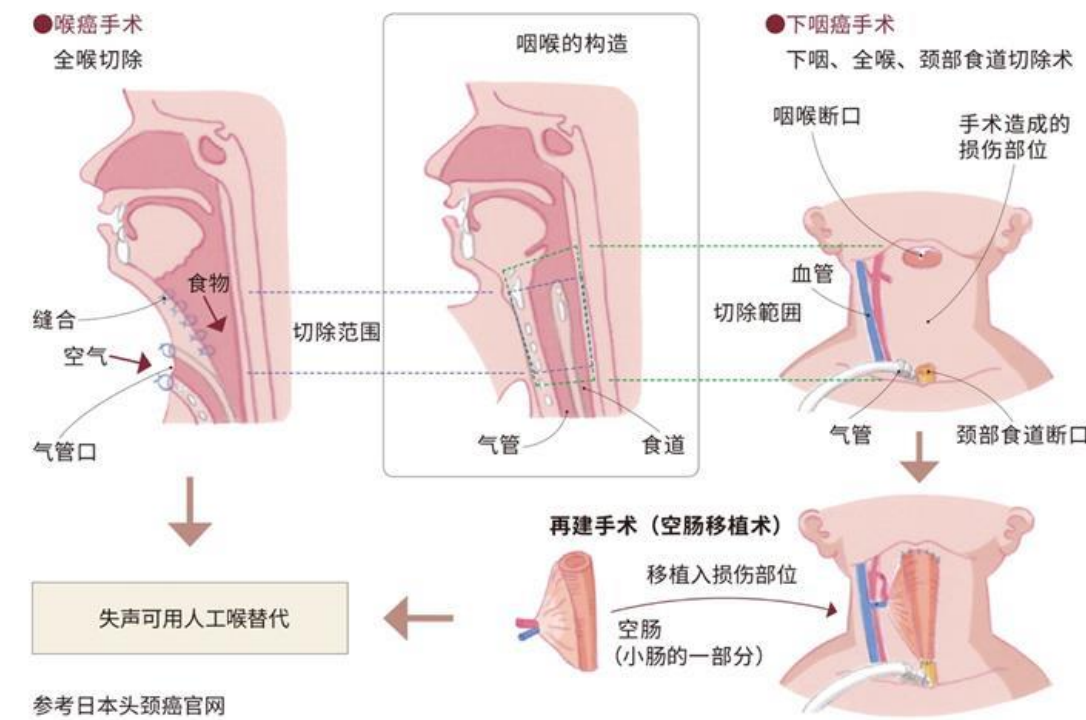
- ❖ Contrast CT+ multi-modal **head and neck cancer** TNM
 - ❖ Contrast CT+ multi-modal **lung cancer** TNM
- ❖ Contrast CT+ multi-modal **esophageal cancer** TNM
- ❖ Contrast CT+ multi-modal **pancreatic cancer** TNM
- ❖ Contrast CT unified detection and diagnosis model of at least **eight major cancers** (M)

(1) T-staging for laryngeal cancer based on CT

- ❑ The precise clinical T staging before treatment for laryngeal cancer patients determines the therapeutic approach, such as larynx preserving surgery (for T1-T3 stages) or total laryngectomy (for T4 stage), which in turn leads to varying
- ❑ Currently, clinical diagnosis of T staging relies on PET-CT; however, even PET often fails to clearly demonstrate invasion of the thyroid cartilage and extralaryngeal tissues [2].
- ❑ Objective: To establish preoperative T staging for laryngeal cancer based on CT, reducing the need for additional PET-CT radiation exposure and patient expenses, concurrently devising the most accurate surgical approach.



Total laryngectomy surgery



- Over 300 pathologically confirmed patients for early N stage, with plans for more cases and PET-CT data.
- 54-organ segmentation model for head and neck.
- Tumor segmentation accuracy of 78% for nasopharyngeal cancer based on non-contrast CT scans (approaching the 80% consistency of physician delineation on multi-modal MRI).
- Our current system scores at 3.1.

[1] Lydiatt et al. "Head and neck cancers—major changes in the American Joint Committee on cancer eighth edition cancer staging manual", CA: A Cancer Journal for Clinicians, 2017

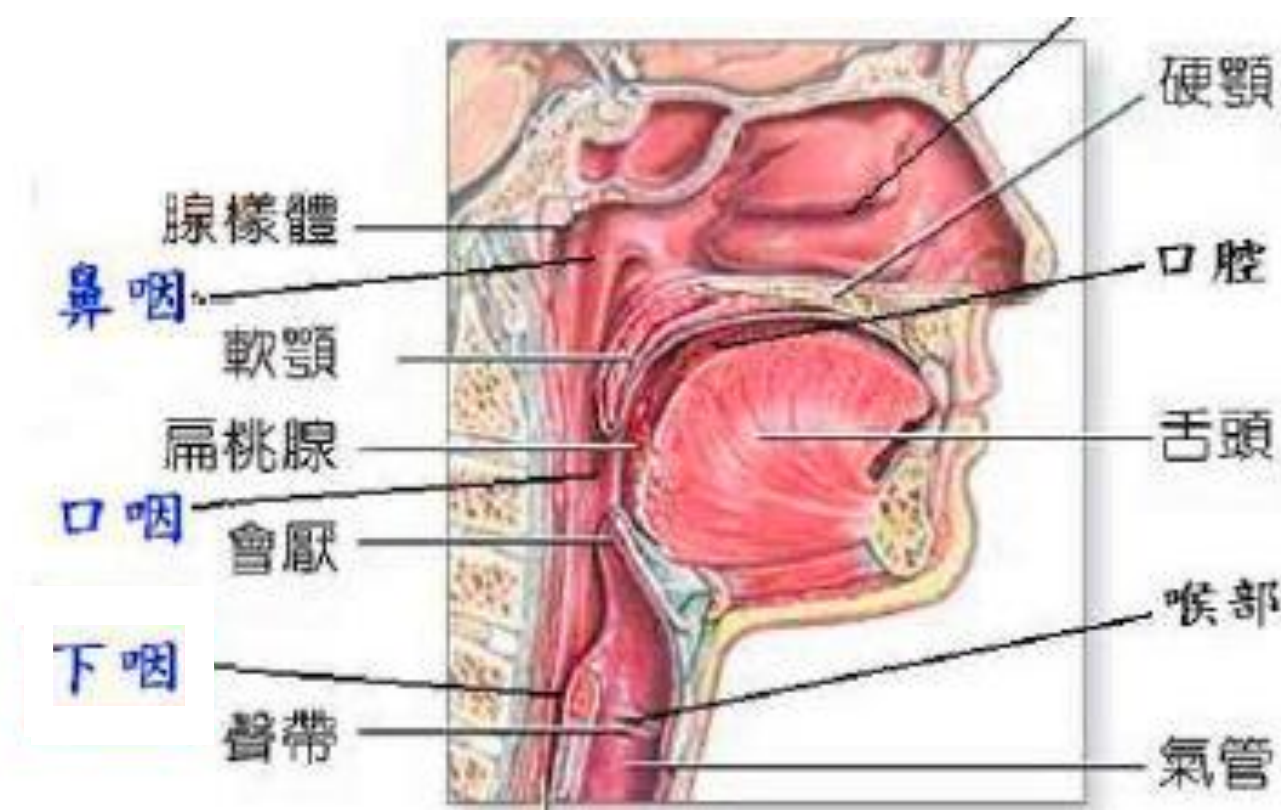
[2] Forastiere et al. "Use of Larynx-Preservation Strategies in the Treatment of Laryngeal Cancer: American Society of Clinical Oncology Clinical Practice Guideline Update", Journal of Clinical Oncology, 2018

[3] "Comprehensive and Clinically Accurate Head and Neck Cancer Organs-at-risk Delineation on a Multi-institutional Study", Nature Communications, 2022

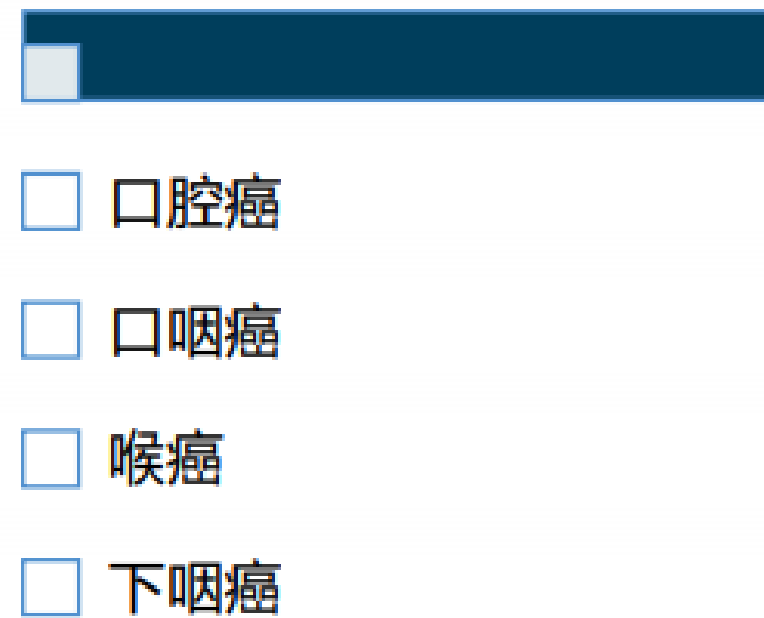
[4] "Deep learning for fully automated prediction of overall survival in patients with oropharyngeal cancer using FDGPET imaging: an international retrospective study", Clinical Cancer Research, 2021

(1) N-staging in head and neck squamous cell carcinoma using CT

- Head and neck squamous cell carcinoma refers to malignant epithelial tumors originating from the oral cavity, pharynx, and other sites, ranking 6th in overall cancer incidence and accounting for over 90% of all head and neck cancers.
- These cancers often affect vital functions such as chewing, speaking, swallowing, and aesthetics due to their involvement of multiple sites, highlighting the importance of precise N staging for HNSCC [1,2], which enables reasonable treatment plans that balance patients' quality of life, aesthetic appearance, and prognosis.
- Since the 8th edition of the UICC/AJCC TNM staging system, extranodal extension (ENE), defined as the spread of cancer beyond the lymph node capsule, has been incorporated into the N staging, recommending a combined modality of treatment.

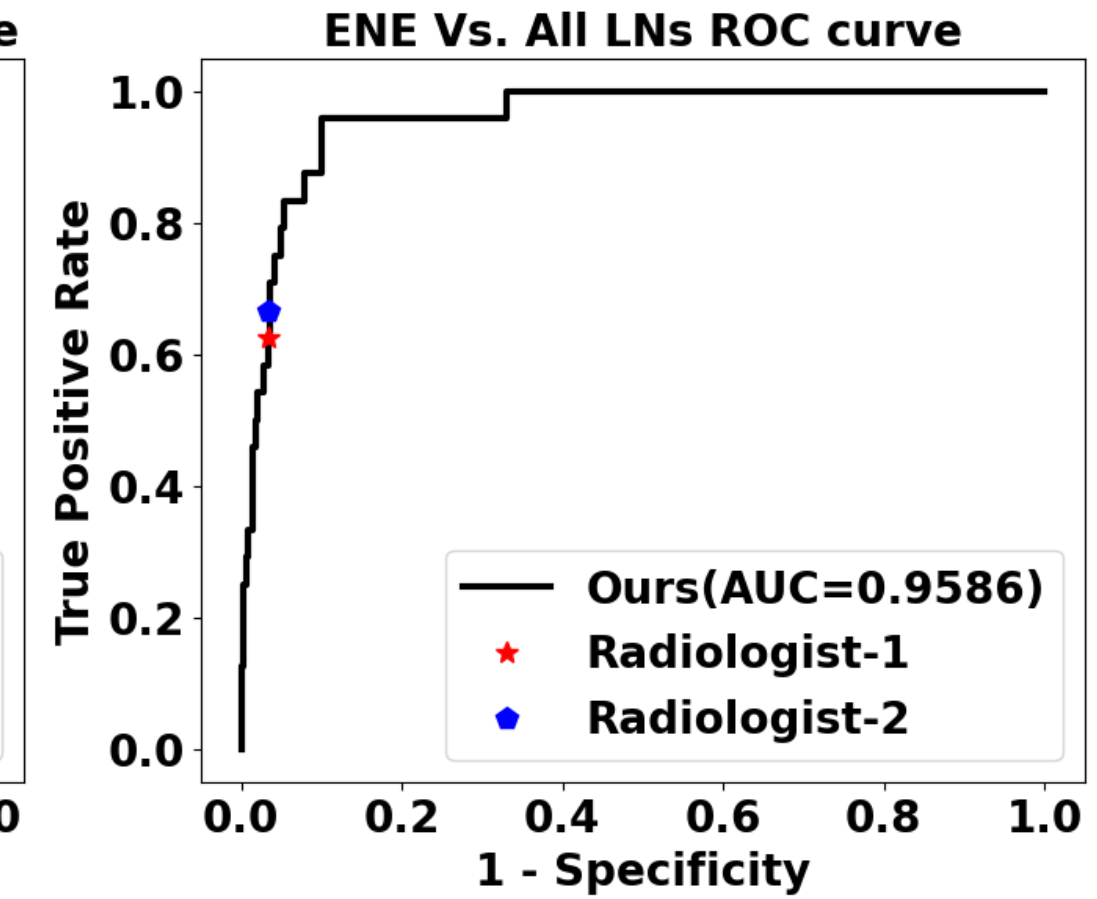
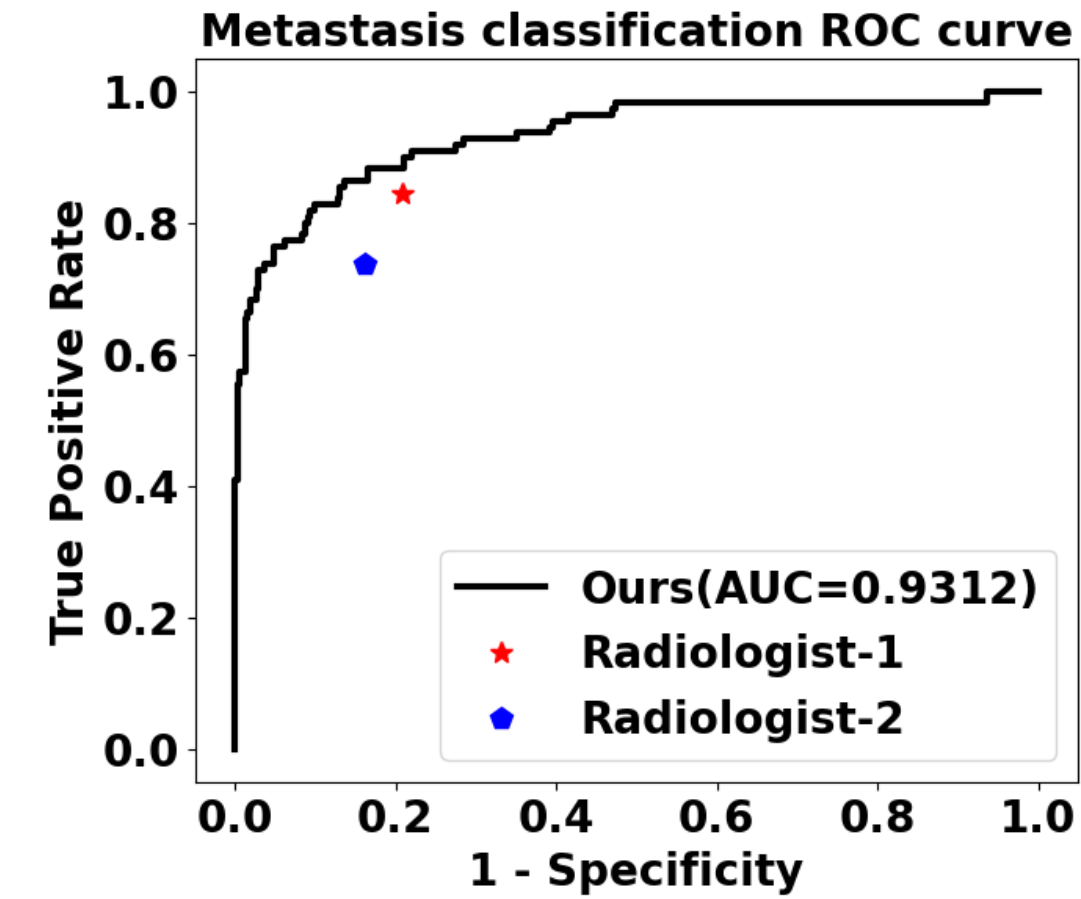


淋巴结包膜外侵犯 ENE

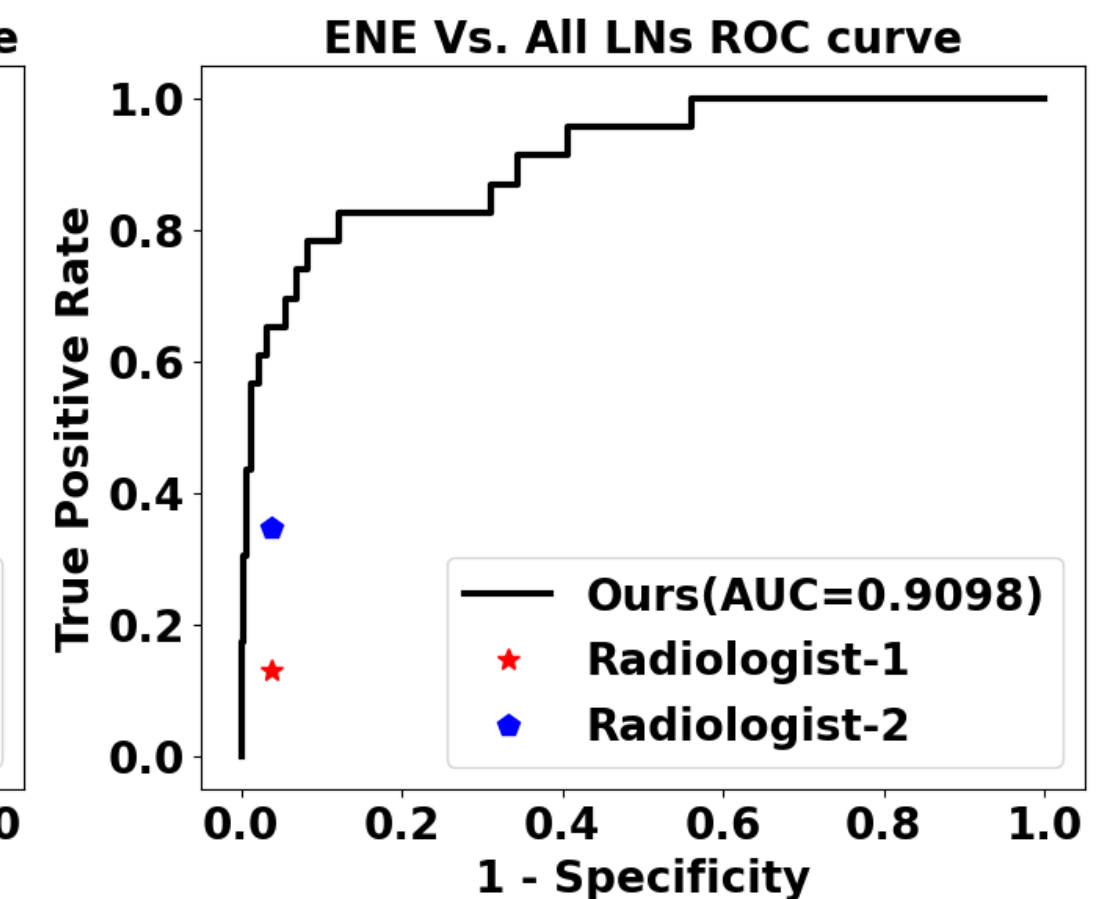
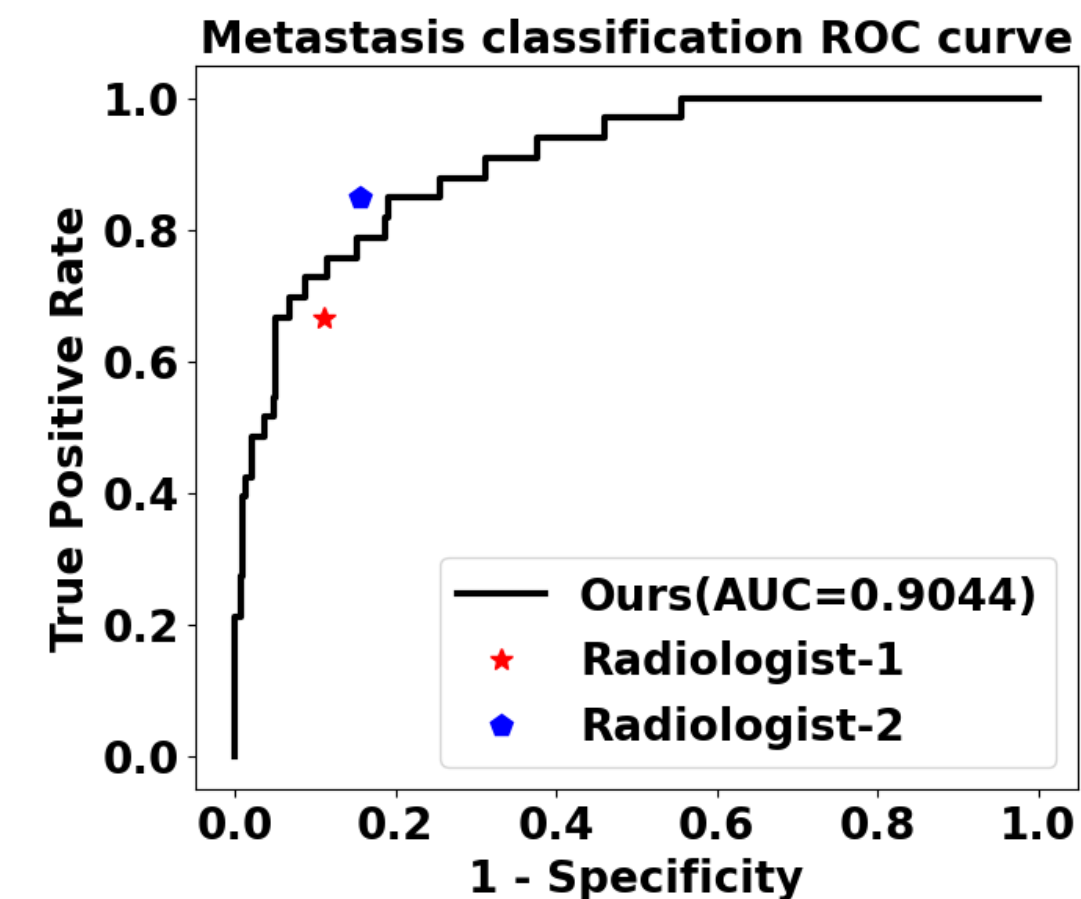


- JCO ENE diagnosis AUC 0.87; Lancet DH AUC 0.86; Our **AUC for metastatic lymph node diagnosis is 0.90-0.93**, **ENEdiagnosis AUC 0.91-0.96**; ENE diagnosis for two radiologists in internal and external data is only **64% and 28%** in average.

Internal test



External test



[1] Lydiatt et al. "Head and neck cancers—major changes in the American Joint Committee on cancer eighth edition cancer staging manual", CA: A Cancer Journal for Clinicians, 2017

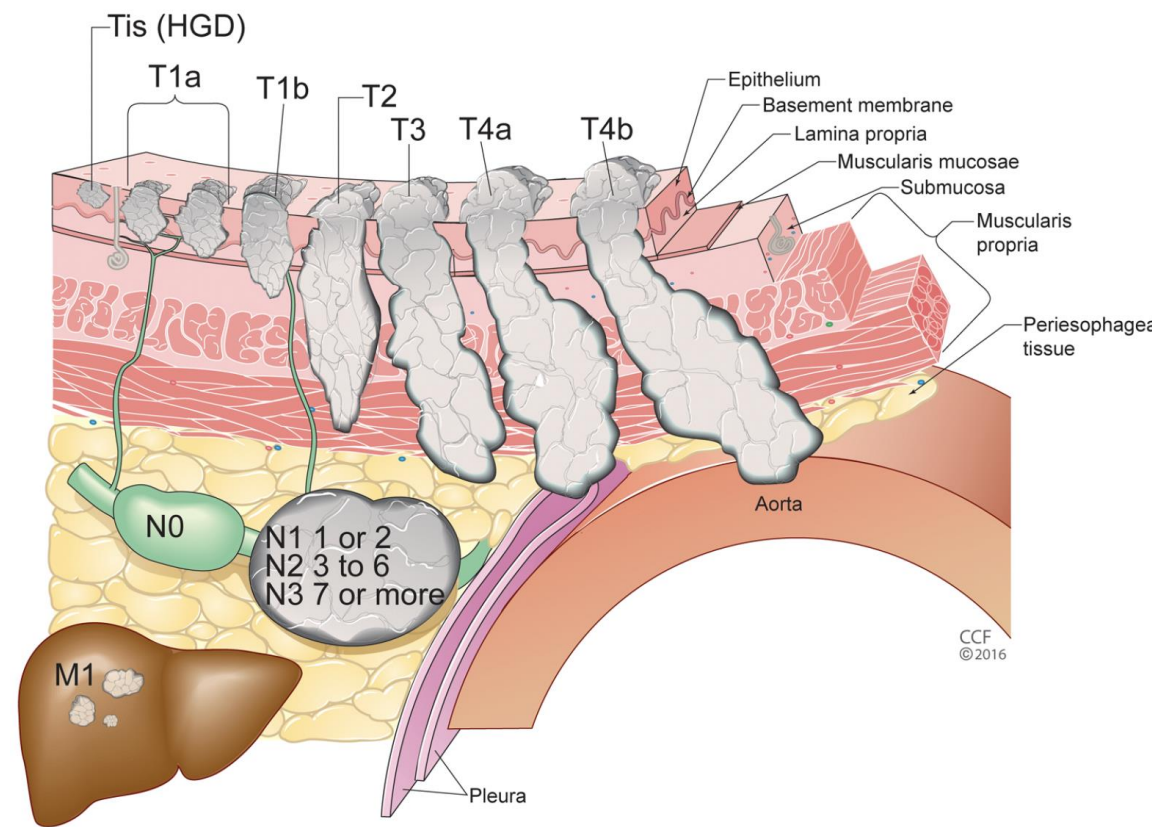
[2] Ho et al. "Association of Quantitative Metastatic Lymph Node Burden With Survival in Hypopharyngeal and Laryngeal Cancer", JAMA Oncology, 2018

[3] Kann et al. "Multi-Institutional Validation of Deep Learning for Pretreatment Identification of Extranodal Extension in Head and Neck Squamous Cell Carcinoma", Journal of Clinical Oncology, 2020

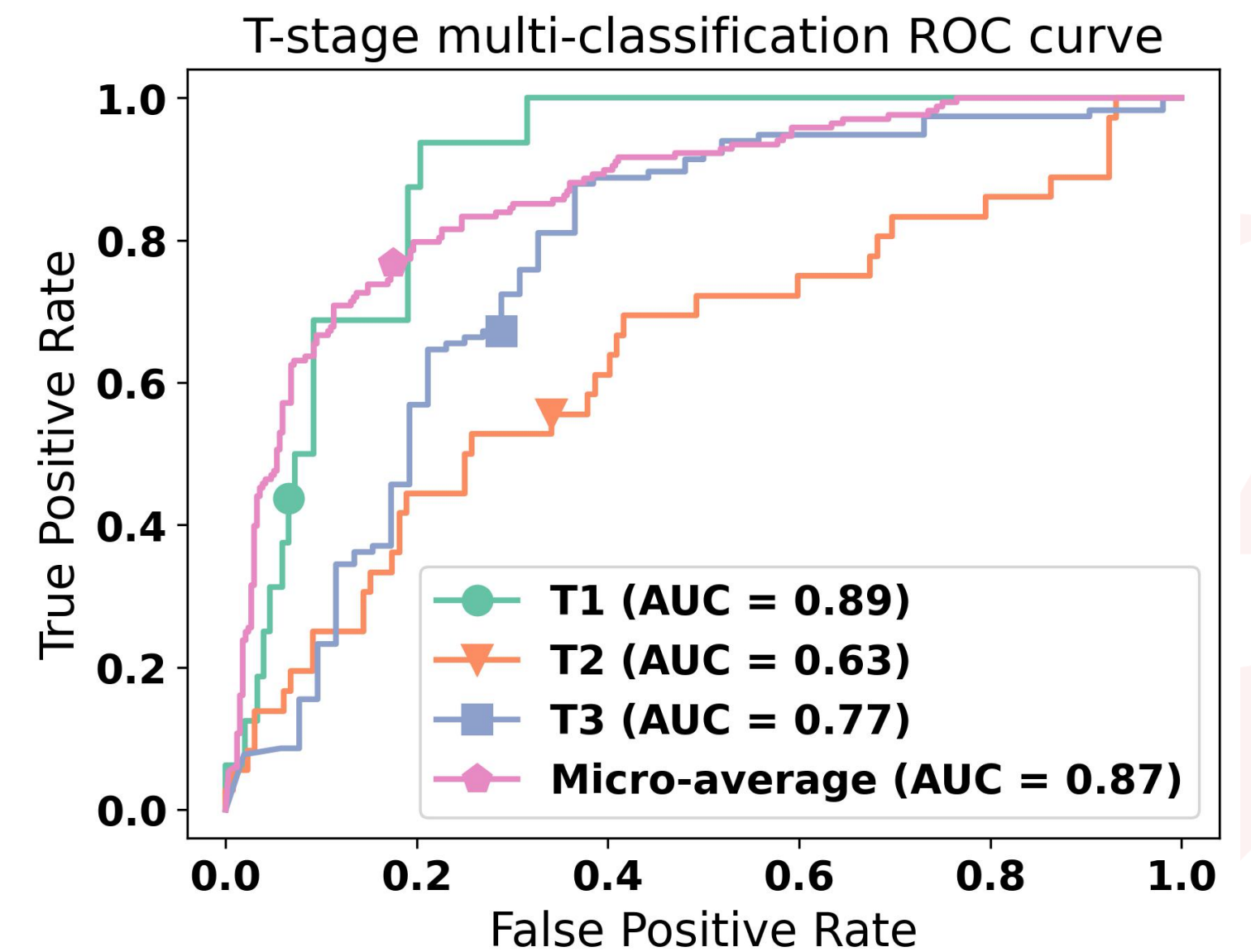
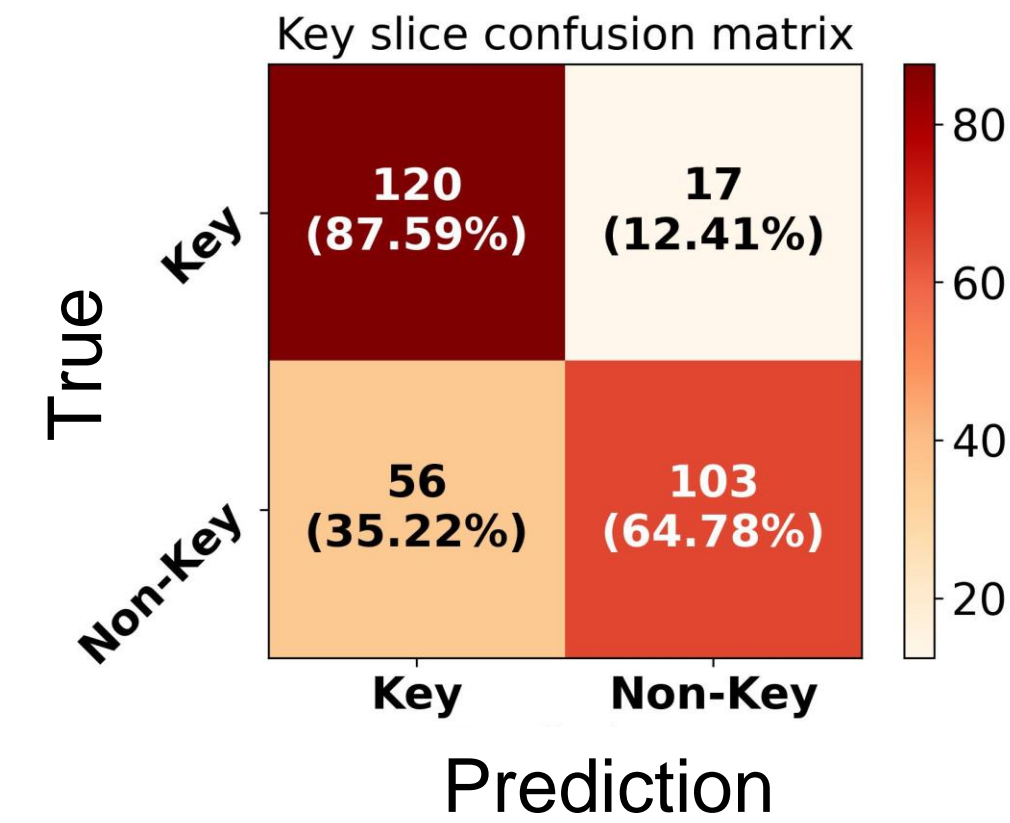
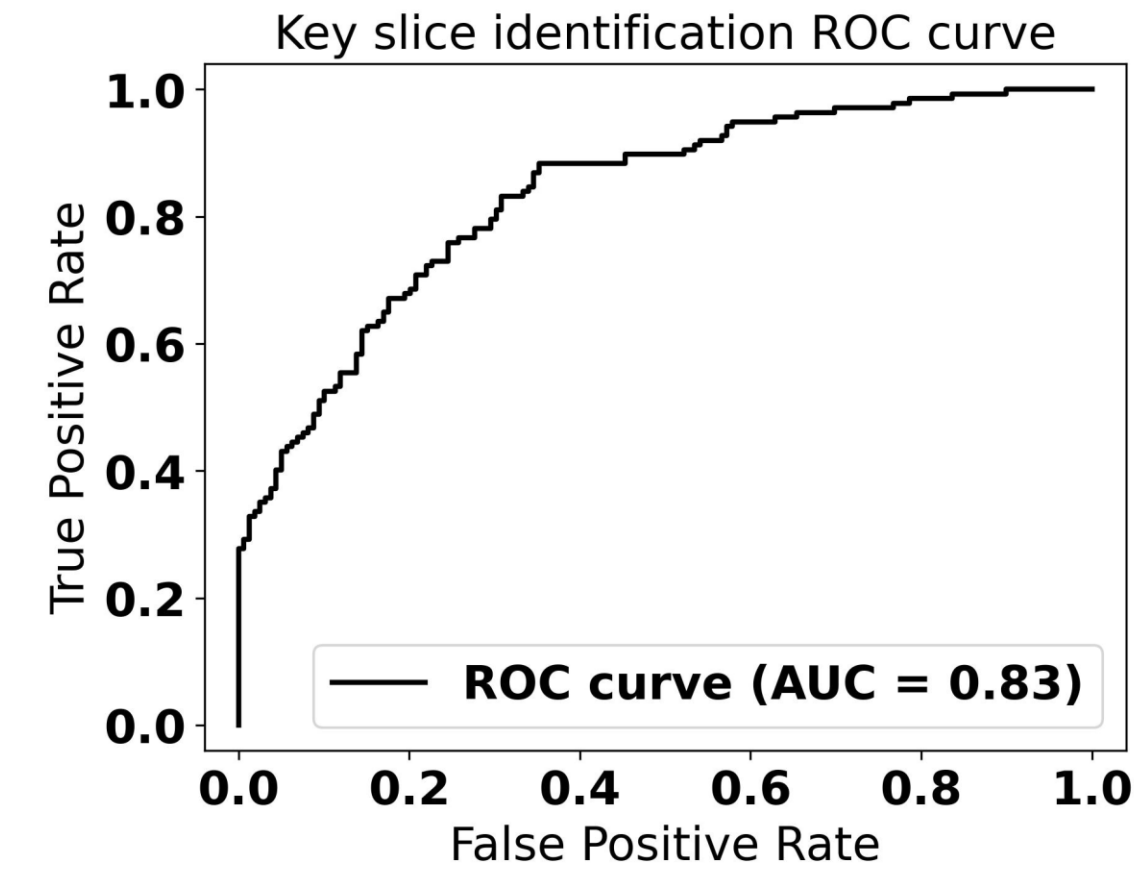
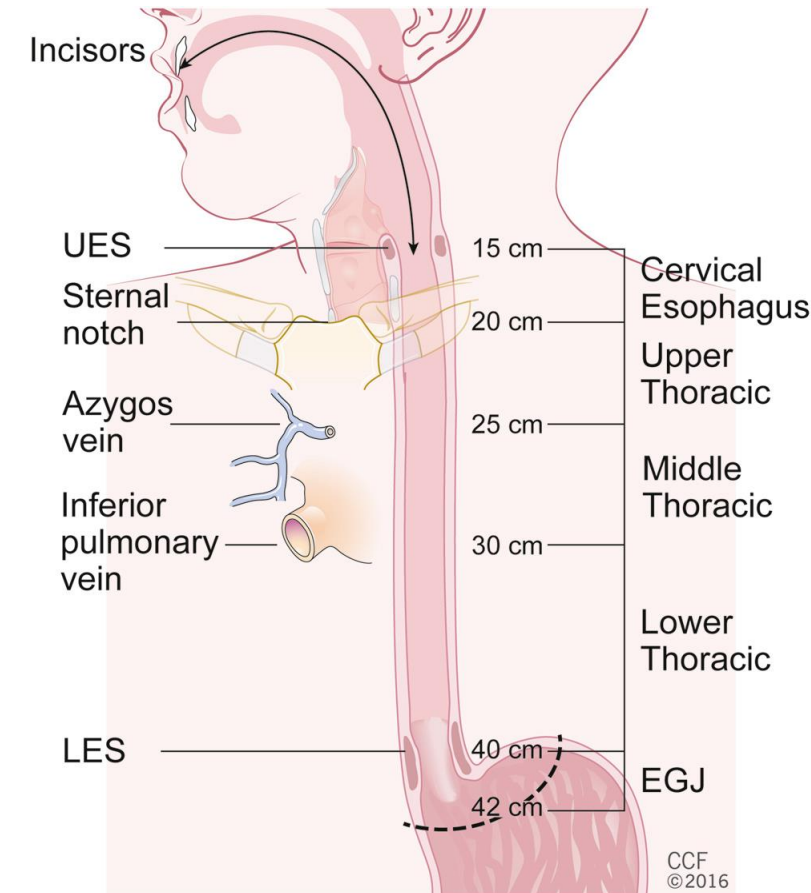
[4] Kann et al. "Screening for extranodal extension in HPV-associated oropharyngeal carcinoma: evaluation of a CT-based deep learning algorithm in patient data from a multicentre, randomised de-escalation trial", Lancet Digital Health, 2023

(2) T-staging for Esophageal Cancer using CT

- The T stage refers to the status of the primary tumor, T1 to T4 showing the tumor volume enlarges and involvement of adjacent tissues expands.
- The precise clinical T staging of esophageal cancer patients before treatment determines the therapeutic approach taken (e.g. surgery versus neoadjuvant chemoradiotherapy), impacting patients' prognoses and survival outcomes [1].
- The accuracy of clinical T staging based on CT/MRI/PET by physicians is low, from 30% to 70% [2,3], while endoscopic ultrasound achieves an accuracy rate of 85%; however, it is invasive, not universally offered and relies on the operator's skill [4].



Invasion:
 T1: Mucosal layer
 T2: muscularis propria
 T3: esophageal fibrous membrane
 T4: adjacent organs



Currently, our precision in identifying the key slices determining T staging in esophageal tumors stands at 87.6%, with an area under the curve (AUC) of 83%. For T1, T2, and combined T3+T4 staging, the respective AUCs are 89%, 63%, and 77%, with a micro AUC of 87%.

[1] Rice et al. "Cancer of the Esophagus and Esophagogastric Junction: An Eighth Edition Staging Primer", Journal of Thoracic Oncology, 2017

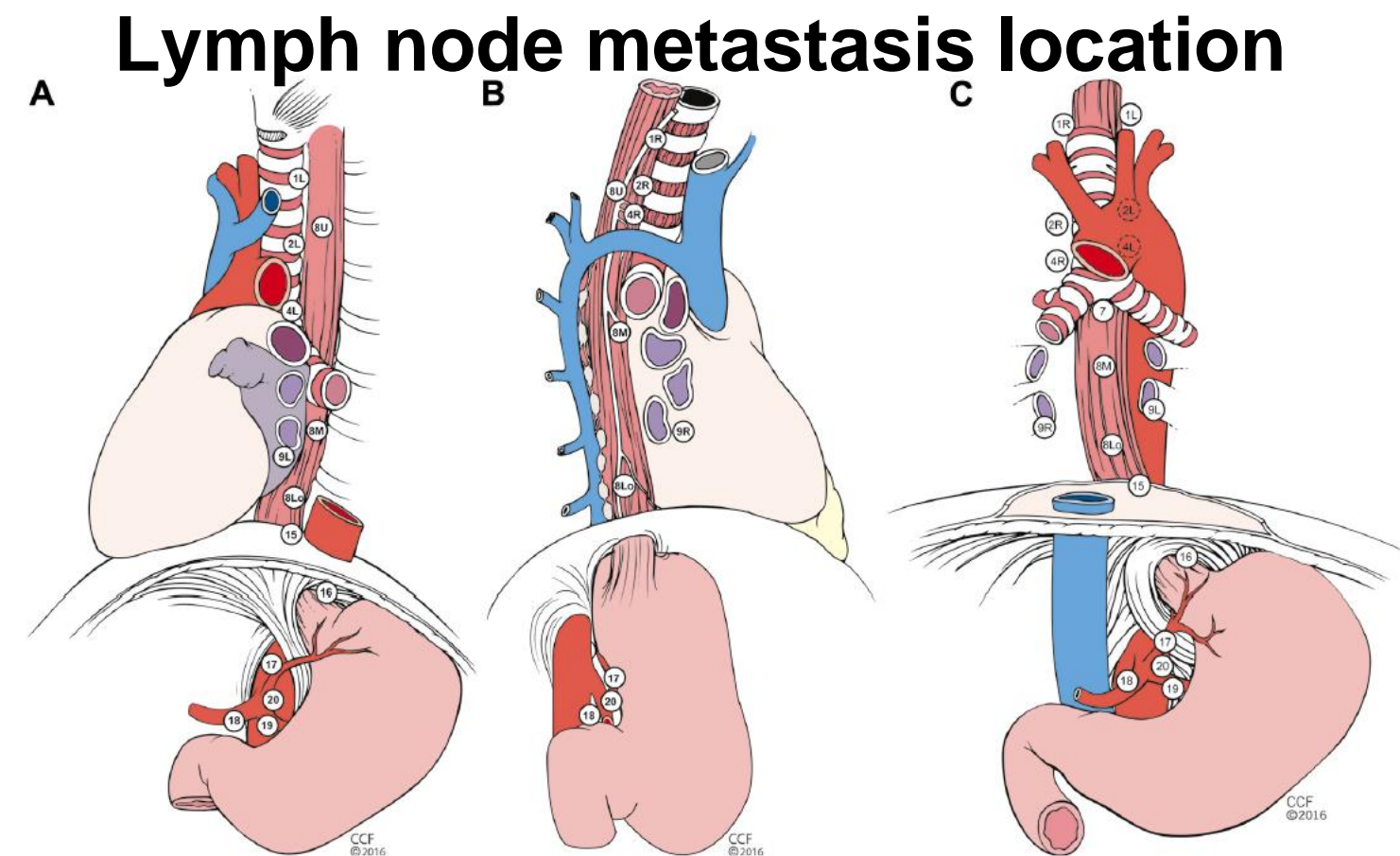
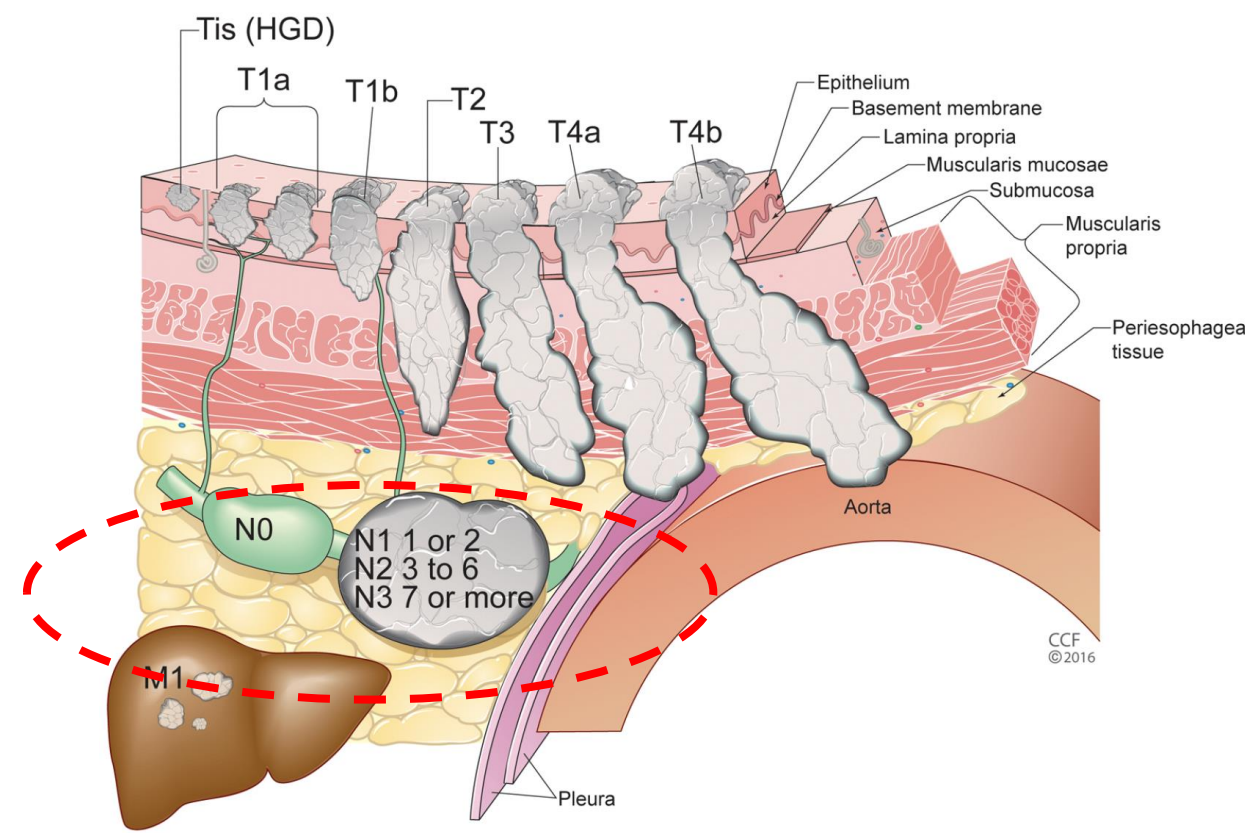
[2] Lee et al. "Clinical implication of PET/MR imaging in preoperative esophageal cancer staging: comparison with PET/CT, EUS, and CT", Journal of Nuclear Medicine, 2014

[3] Lee et al. "Diagnostic Performance of MRI for Esophageal Carcinoma: A Systematic Review and Meta-Analysis", Radiology, 2021

[4] Puli et al. "Staging accuracy of esophageal cancer by endoscopic ultrasound: A meta-analysis and systematic review", World Journal of Gastroenterology, 2008

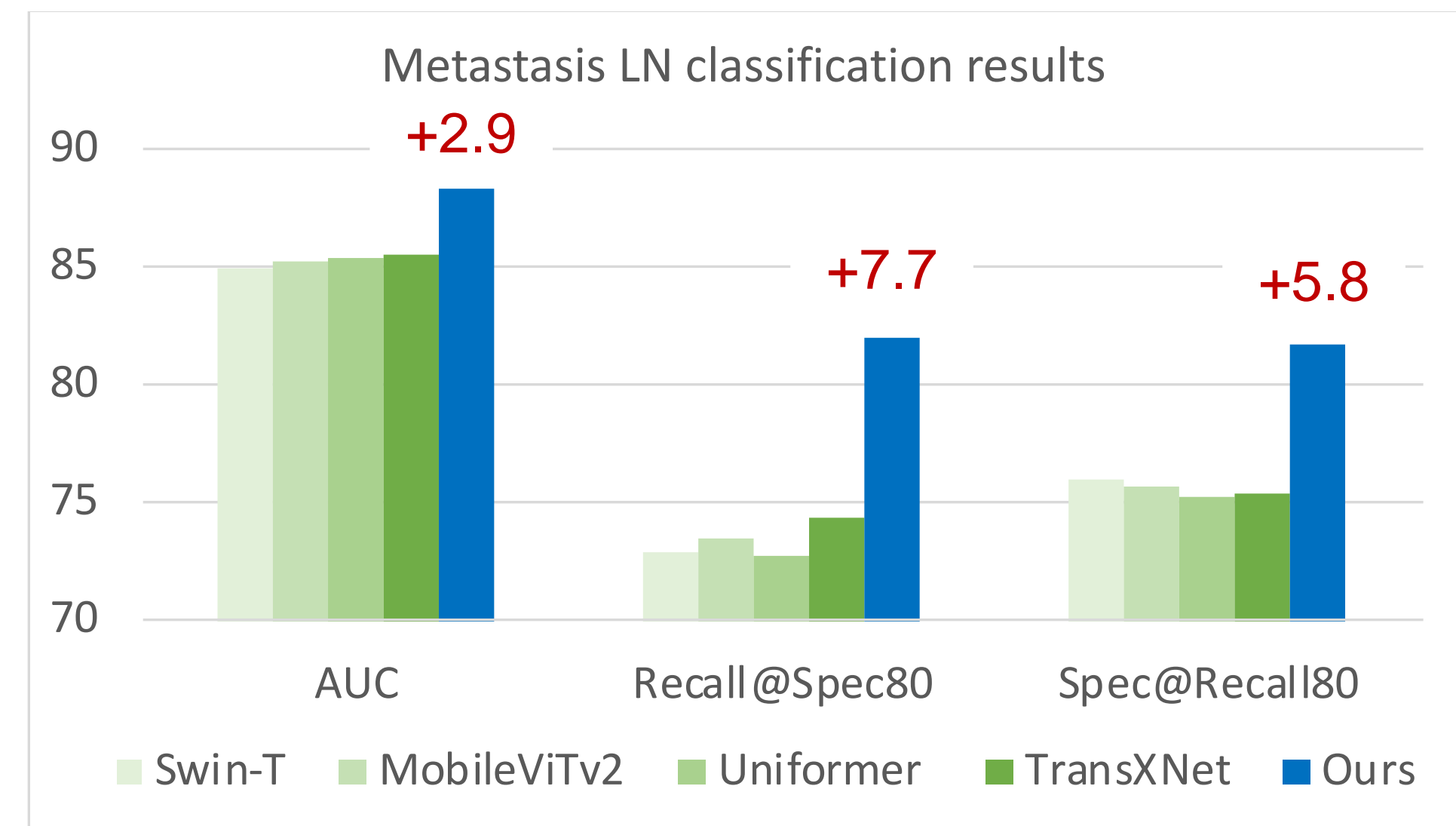
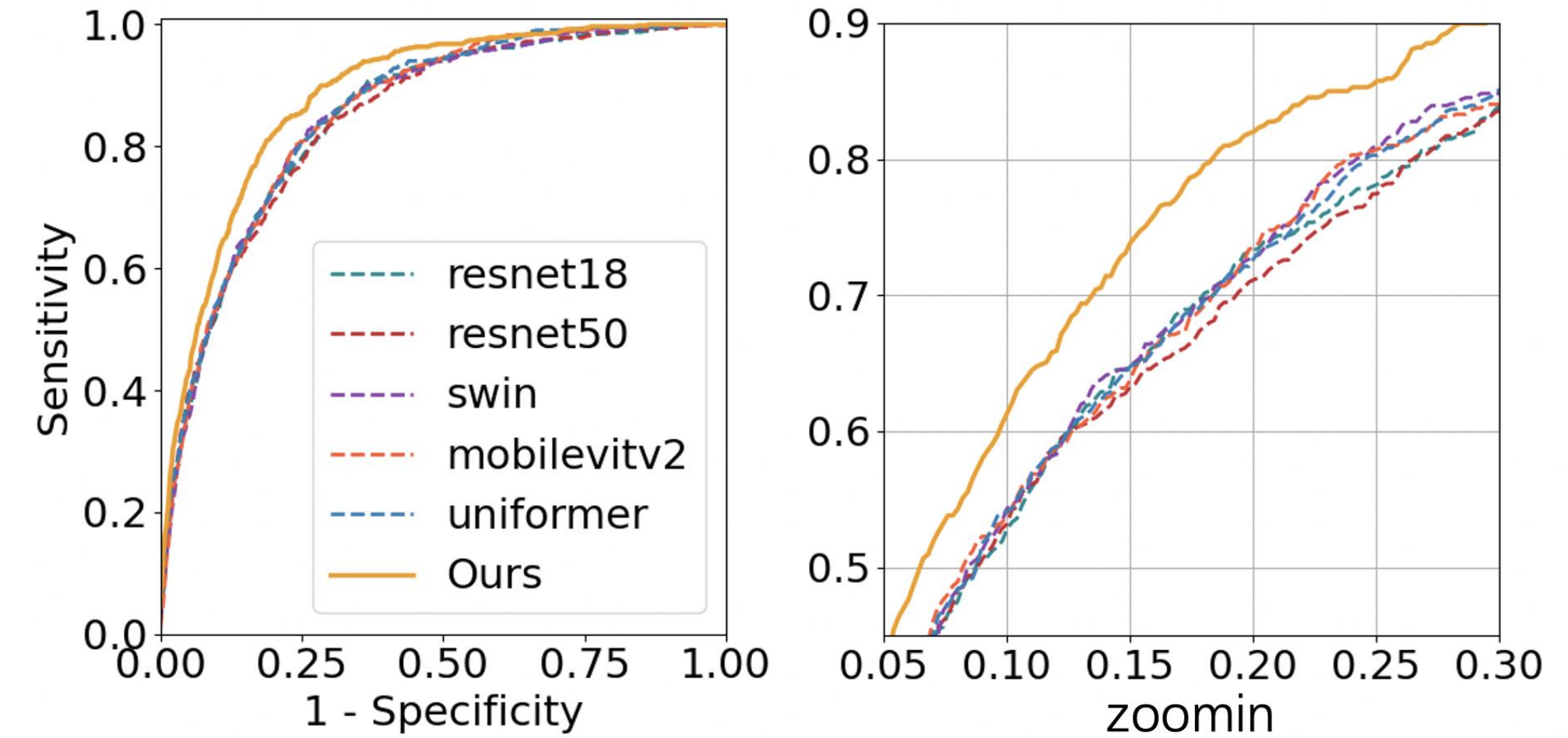
(2) N-staging for Esophageal Cancer using CT

- The accurate clinical N staging [1] before treatment for esophageal cancer patients determines the treatment (such as surgery versus neoadjuvant chemoradiotherapy) and the precision (the scope of lymph node dissection, the range of radiotherapy clinical target volume [CTV]), which directly impacts patients' survival outcomes [2].
- The accuracy of differentiating between benign and malignant esophageal lymph nodes based on EUS/CT/PET by clinicians is low: **sensitivity ranges from 50% to 80% [2,3], and specificity ranges from 70% to 85% [3,4].**



Our AUC to identify malignant lymph nodes in esophageal cancer is 88.4%, with a sensitivity of 82.02% at 80% specificity, and a specificity of 81.76% at 80% sensitivity.

Malignancy classification ROC



[1] Rice et al. "Cancer of the Esophagus and Esophagogastric Junction: An Eighth Edition Staging Primer", Journal of Thoracic Oncology, 2017

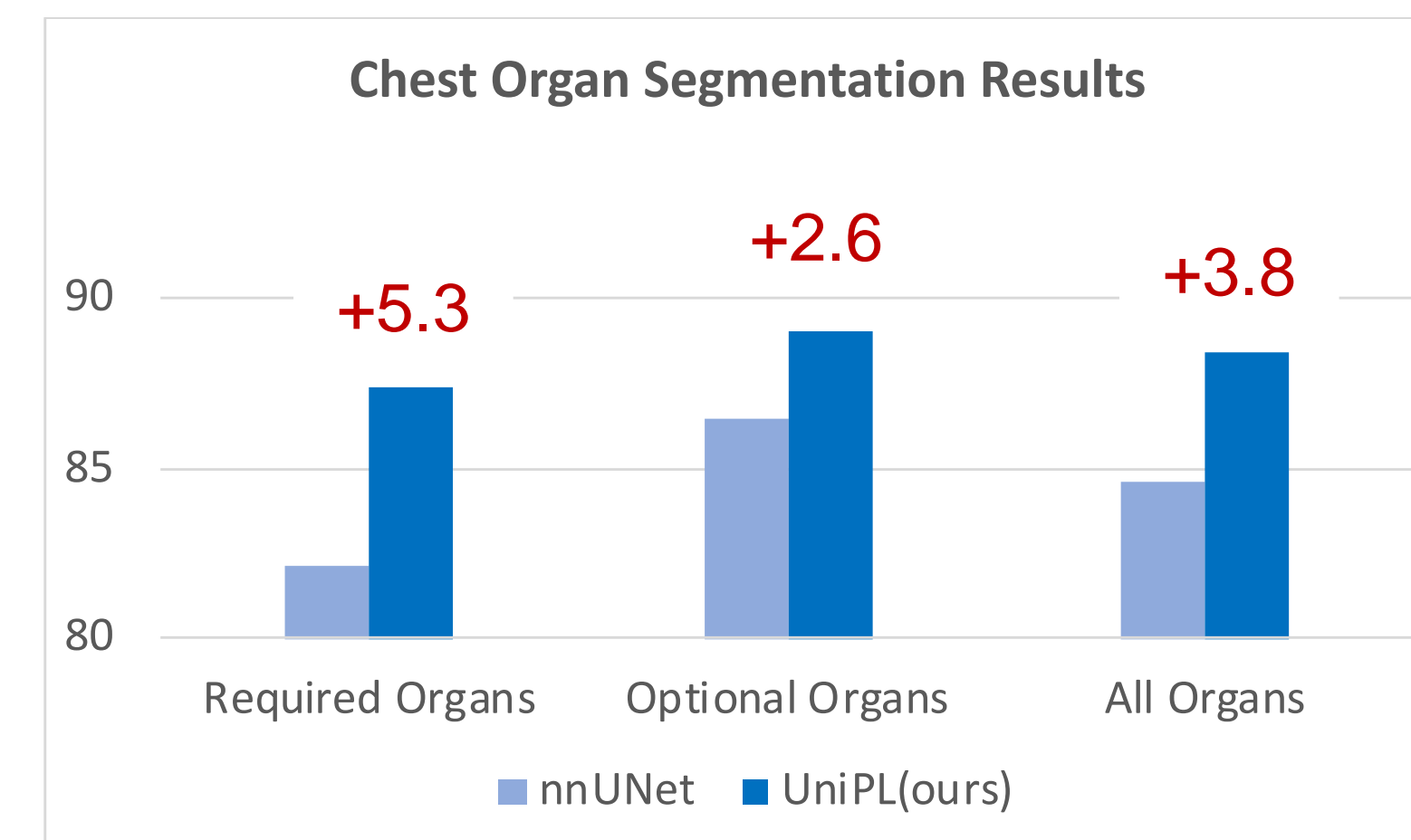
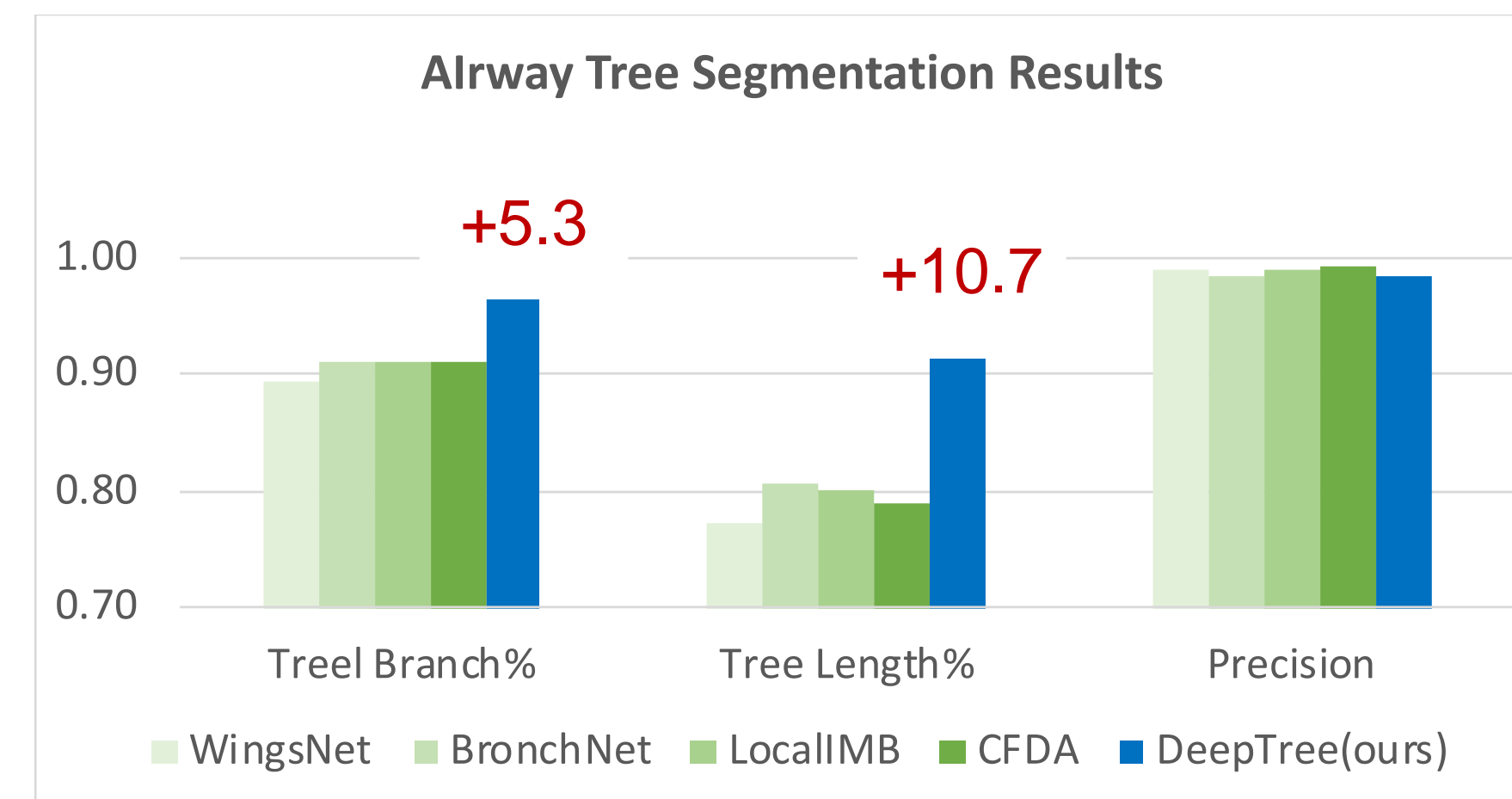
[4] Shah et al. "Improving outcomes in patients with oesophageal cancer", Nature Review Clinical Oncology, 2023

[2] van Vliet et al. "Staging investigations for oesophageal cancer: a meta-analysis", British Journal of Cancer, 2008

[3] Obermannová et al. "Oesophageal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up", Annals of Oncology, 2022

(3) T-staging for lung cancer using CT

- The precise clinical T staging prior to treatment in lung cancer patients determines the therapeutic approach adopted (such as varying surgical procedures, combined treatment modalities, etc.), which in turn leads to differing prognosis and quality of life outcomes for the patients [1].
- Lung cancer T staging relies on accurate tumor segmentation, coupled with the spatial relationship to the bronchial tree, as well as the spatial relationship with surrounding tissues (including the pericardium, chest wall, nerves, esophagus, diaphragm, vertebral bodies, etc.) [2].



Our accuracy for lung tumor segmentation is 81% (Inter-observer consistency ~ 80%).

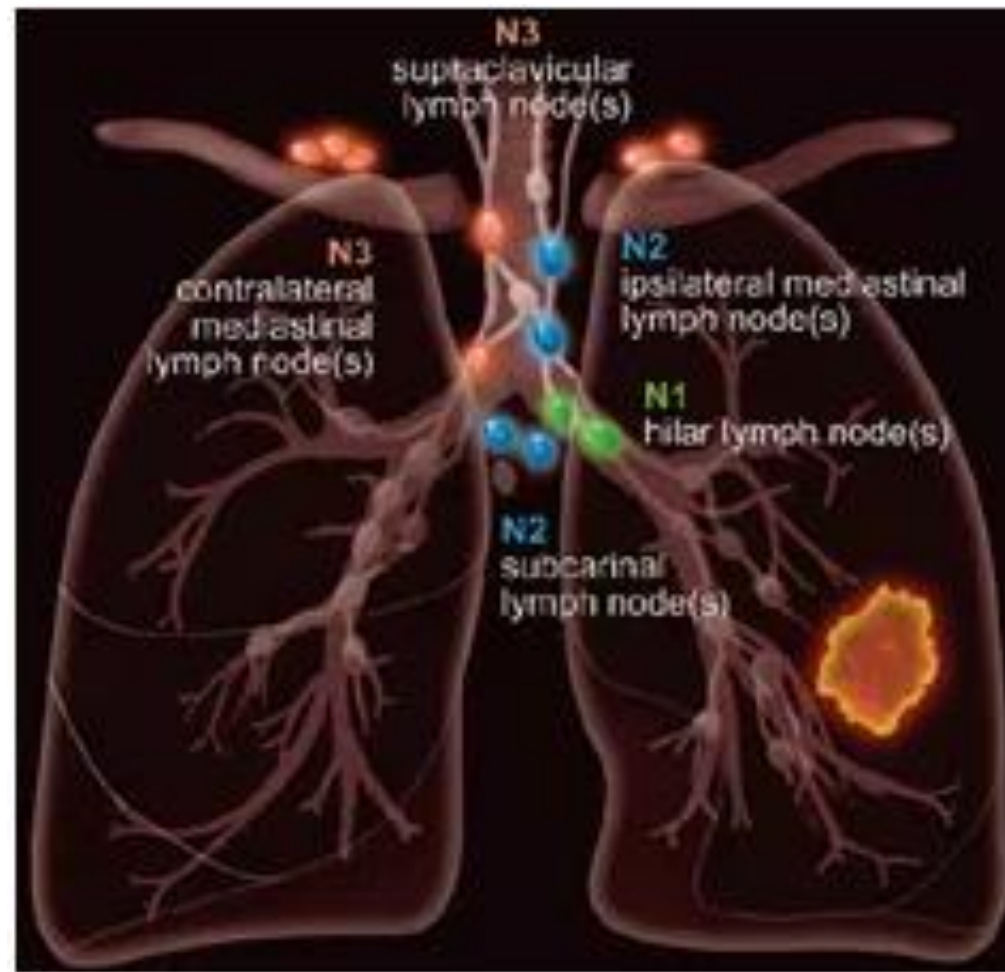
[1] “The International Association for the Study of Lung Cancer Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groups in the Forthcoming (Ninth) Edition of the TNM Classification for Lung Cancer”, Journal of Thoracic Oncology, 2024

[2] “The International Association for the Study of Lung Cancer Lung Cancer Staging Project: Proposals for the Revisions of the T-Descriptors in the Forthcoming Ninth Edition of the TNM Classification for Lung Cancer”, Journal of Thoracic Oncology, 2024

(3) N-staging for lung cancer using PET/CT

○ Accurate clinical N staging prior to treatment for lung cancer determines the therapeutic approach (such as various surgical methods, combined treatment strategies, etc.), which in turn results in differing prognoses and quality of life for the patients [1].

□ Radiology only detected N2 patients from enhanced CT images of tumors [2]; whereas Nature Communications identifies both N1 and N2 patients using tumors on PET/CT scans [3].

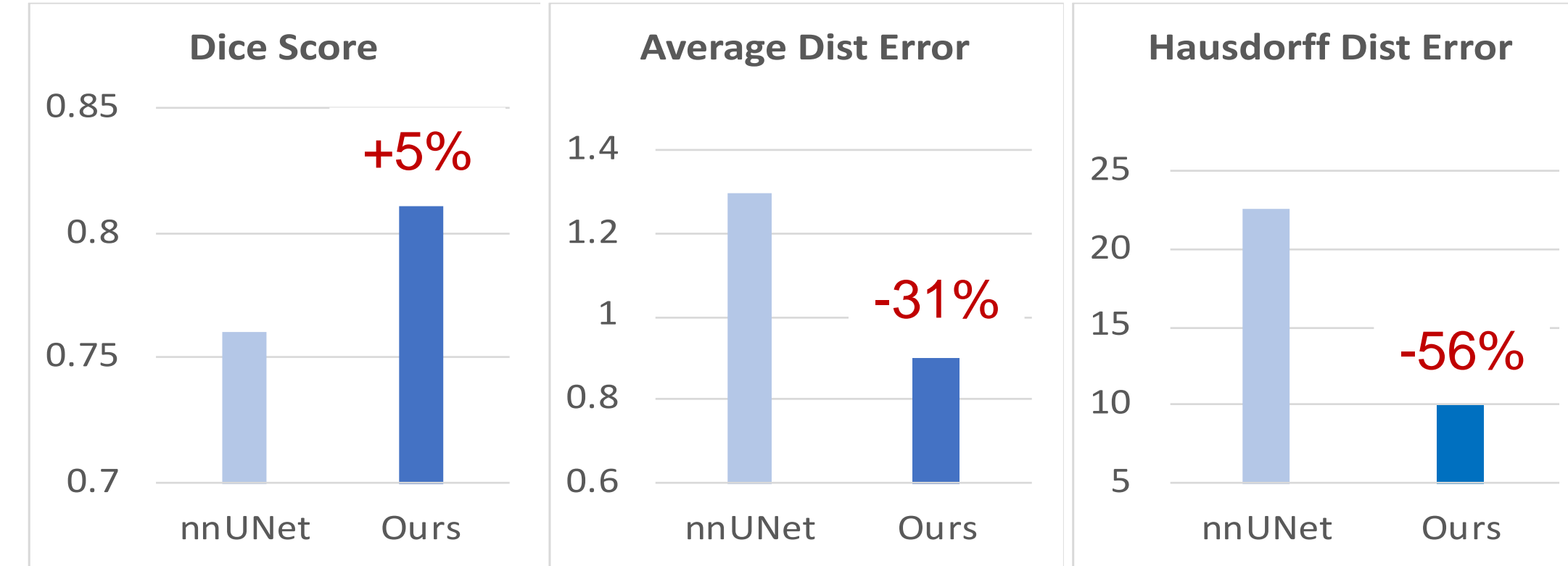


Our Objectives and Strategies:

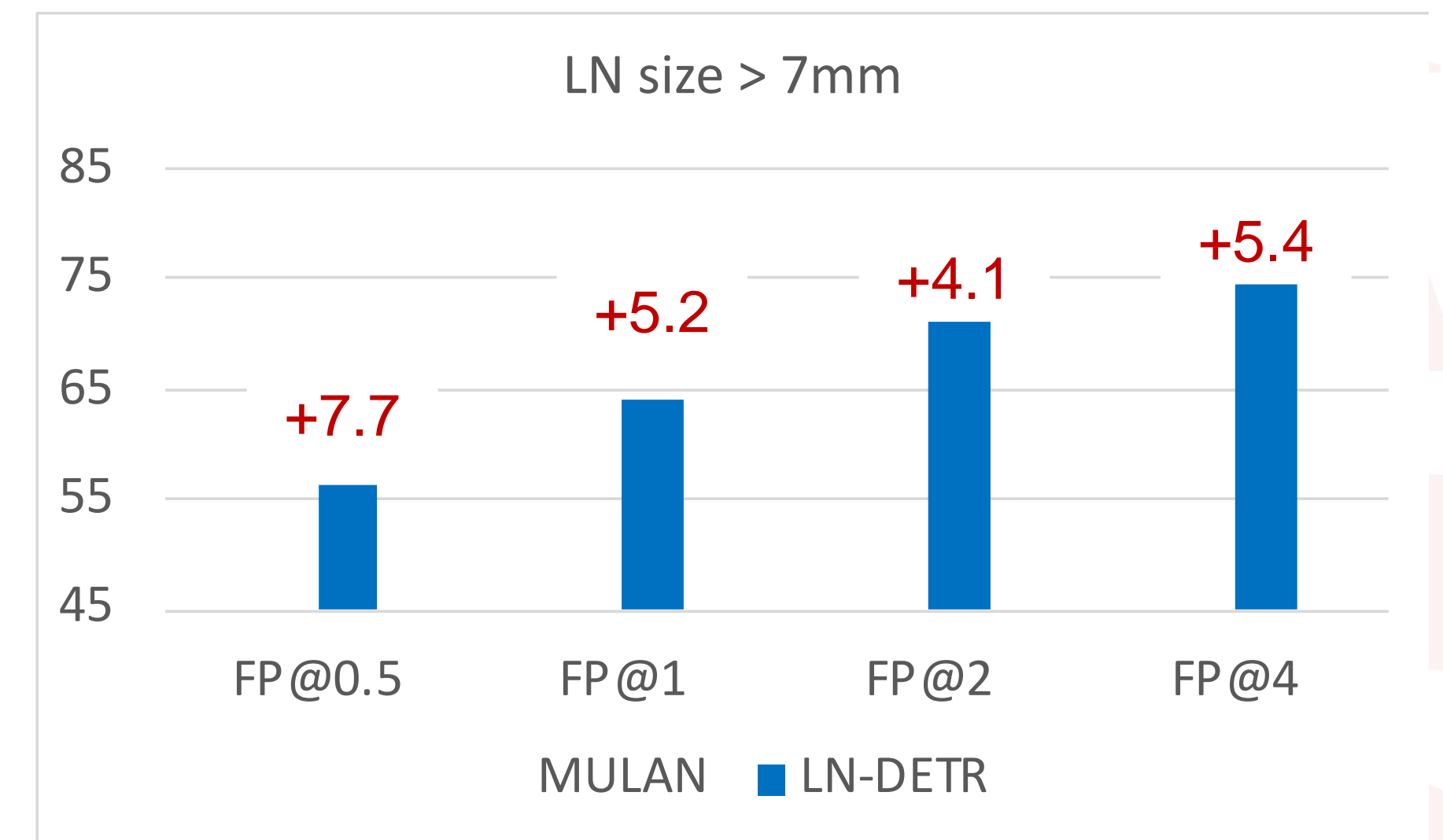
- Full Automation
- Enhanced multi-modal imaging combining Thin-Layer CT with PET/CT
- Lymph node instance + Multi-dimensional Tumor inputs
- Reduce false positives (cN1/2 vs pN0)
- Reduce false negatives (cN0 vs pN1/2)

The preliminary techniques has been established; data collection for lung cancer lymph node differentiation is still ongoing.

Lung cancer lymph node station segmentation results



Lymph node detection results

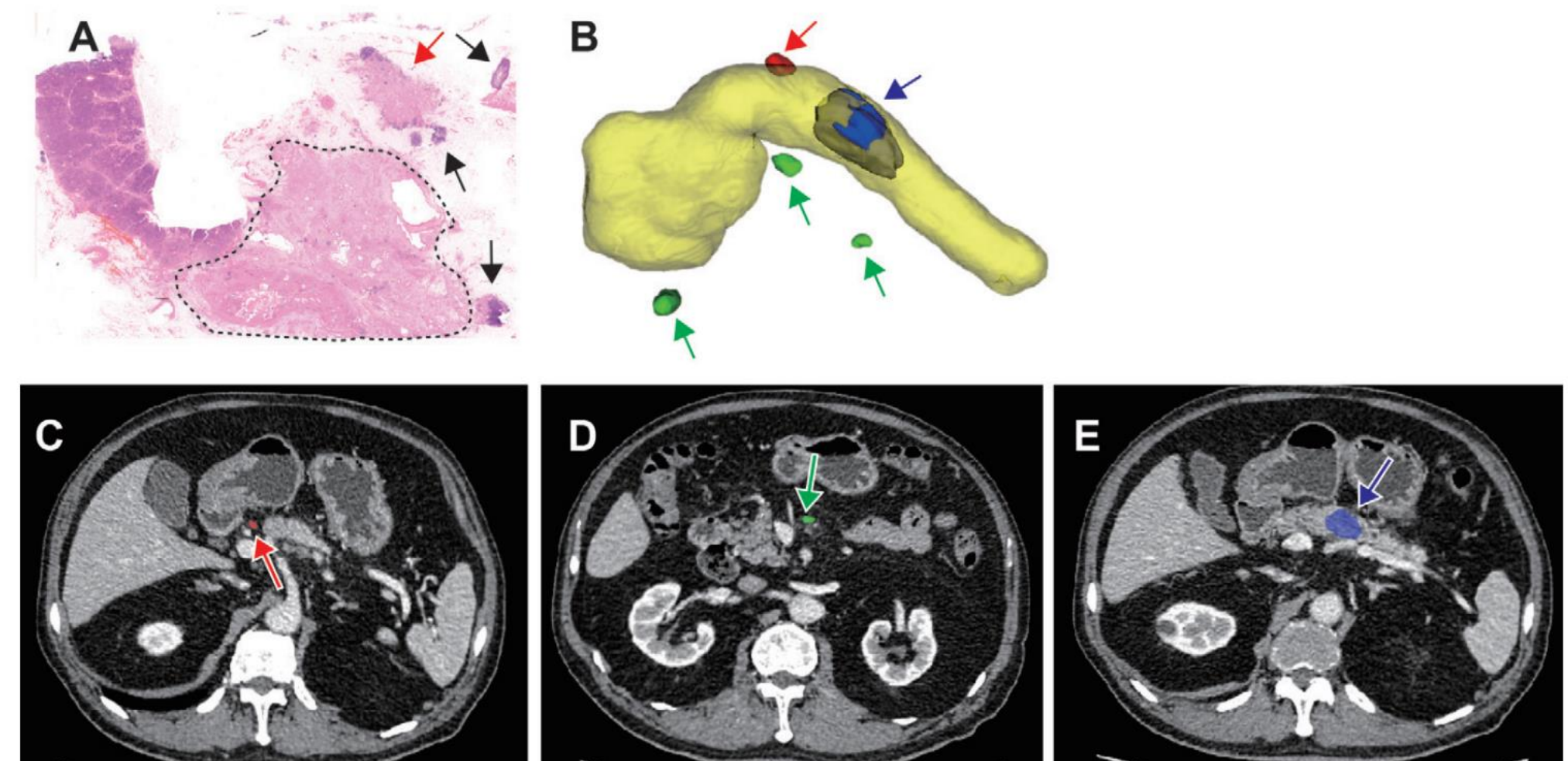
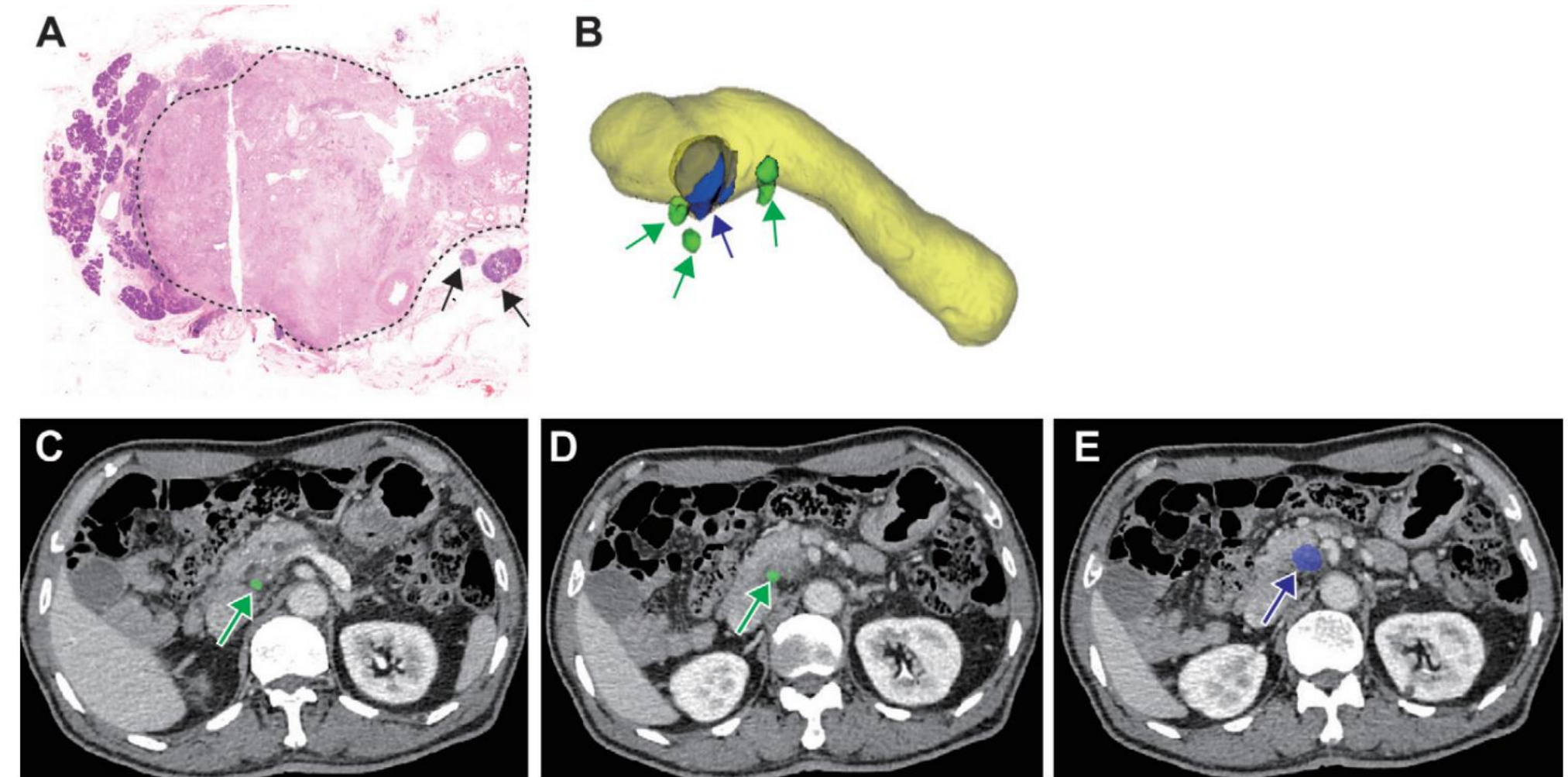
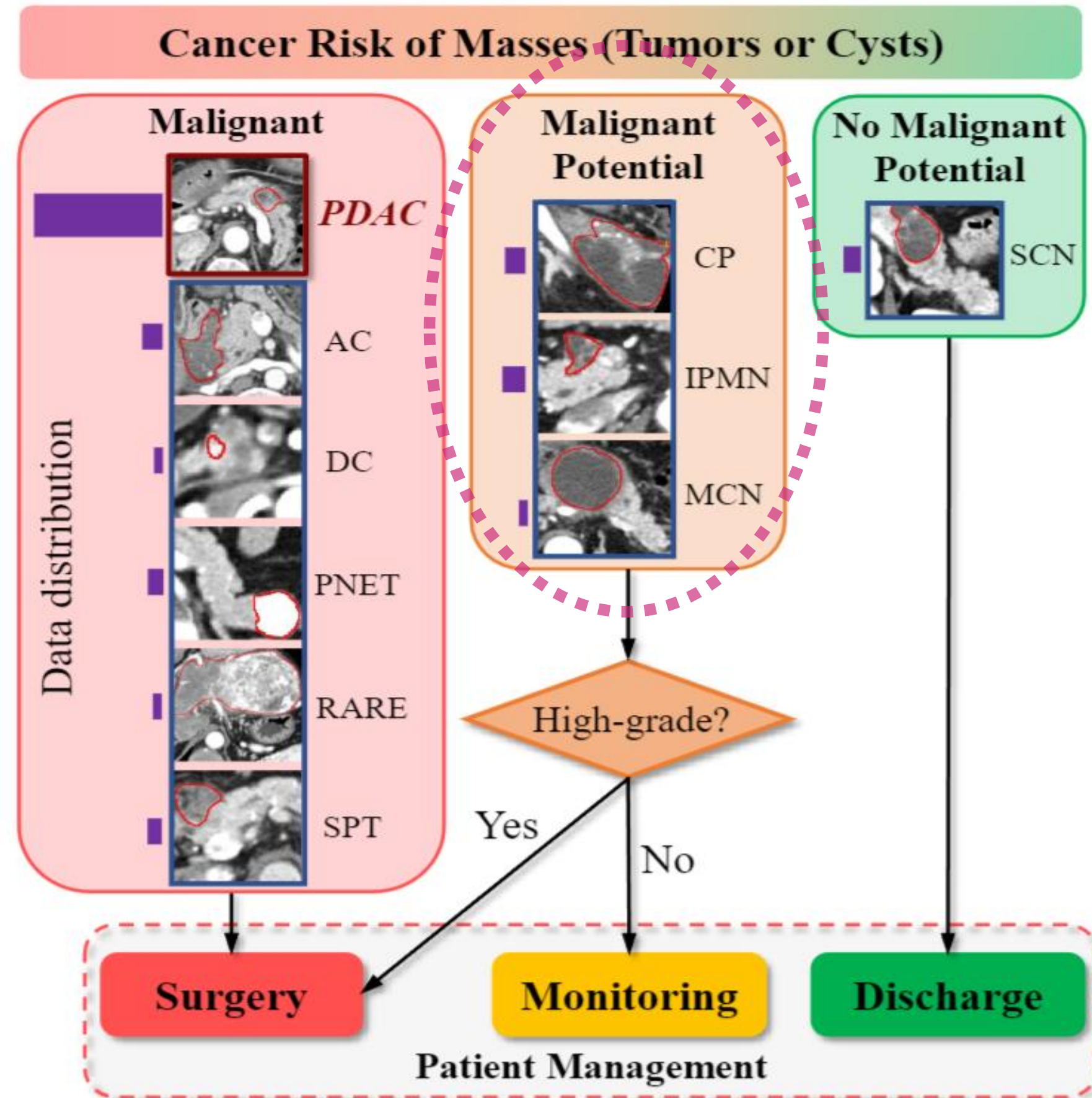


[1] “The International Association for the Study of Lung Cancer Lung Cancer Staging Project: Proposals for the Revisions of the N-Descriptors in the Forthcoming Ninth Edition of the TNM Classification for Lung Cancer”, Journal of Thoracic Oncology, 2024

[2] “Deep learning for prediction of N2 metastasis and survival for clinical stage I non–small cell lung cancer”, Radiology, 2022

[3] “PET/CT based cross-modal deep learning signature to predict occult nodal metastasis in lung cancer”, Nature Communications, 2023

(4) Precision Diagnosis of Pancreatic Cancer using multi-phase CT Management of Precancerous Pancreatic Lesions; Resectability, N-Stage Assessment



[1] "Artificial Intelligence to Predict Lymph Node Metastasis at CT in Pancreatic Ductal Adenocarcinoma", Radiology, 2023

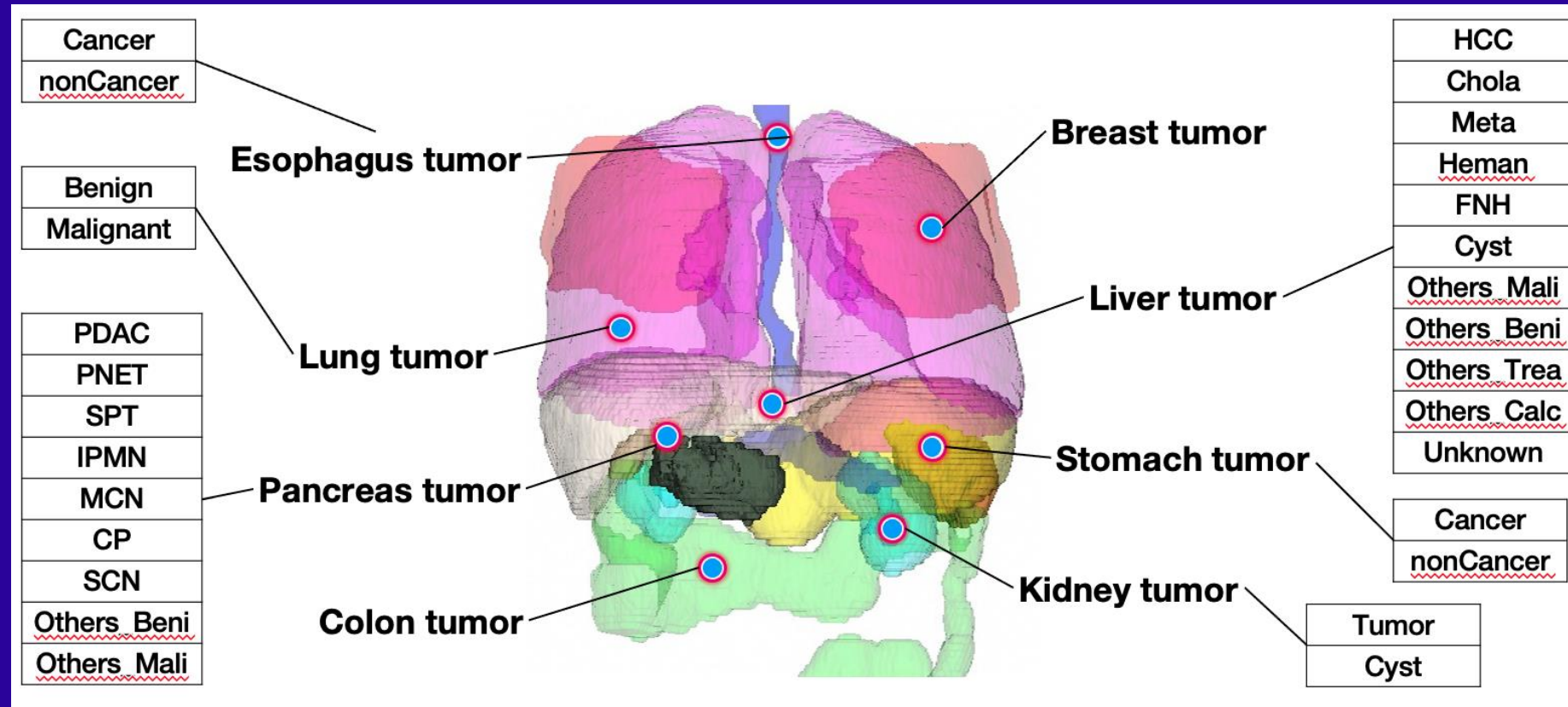
[2] "Deep Attention Learning for Pre-operative Lymph Node Metastasis Prediction in Pancreatic Cancer via Multi-object Relationship Modeling", International Journal of Computer Vision, 2024

[3] "Deep Learning for Fully Automated Prediction of Overall Survival in Patients Undergoing Resection for Pancreatic Cancer: A Retrospective Multicenter Study". Annals of Surgery, 2023

[4] "Graph Anatomy Geometry-Integrated Network for Pancreatic Mass Segmentation, Diagnosis, and Quantitative Patient Management", IEEE CVPR 2021

[5] S Springer, DL Masica, M Dal Molin, C Douville..., "A multimodality test to guide the management of patients with a pancreatic cyst" - Science translational Medicine, 2019

(5) Eight Major Cancers Image Reading AI on Enhanced CT: Detection, Measurement, Diagnosis & Beyond → can we go to M (detecting cancer metastasis)?



#Task	#Dataset	Training Set	Testing Set
Breast	Shenyi	332	100
colon	Shengyi	847	100
eso	esoBenign	1710	100
eso	esoCancer		
Eso	esoSiChuan		
eso	esoZJU		
eso	esoZJU-EarlyEC24		
eso	esoZJU-ECA300		
kidney	KITS23	360	100
Liver	LivShengjing16-22	1736	200
lung	GDLC	1783	100
lung	ZDY670		
lung	ZDY2000	2296	759
pancreas	shangyi		
pancreas	shangyi2		
stomach	shengyi	2186	100
stomach	shenyiAddData		
stomach	shiyiNormal		
CT Scans		11250	1559

Sens. / Spec. = 96% / 90%

1. Techniques:

- (1) Data & Annotation: A training set of 10,000 labeled cases
- (2) Algorithm: Hierarchical optimization with background modeling (HOBM), Partial label learning

2. Product Readiness:

Applicable to chest and abdominal enhanced CT scans. Requires enhancement of specificity and clinical validation.

3. Next Steps:

Deepening collaboration for clinical iteration.

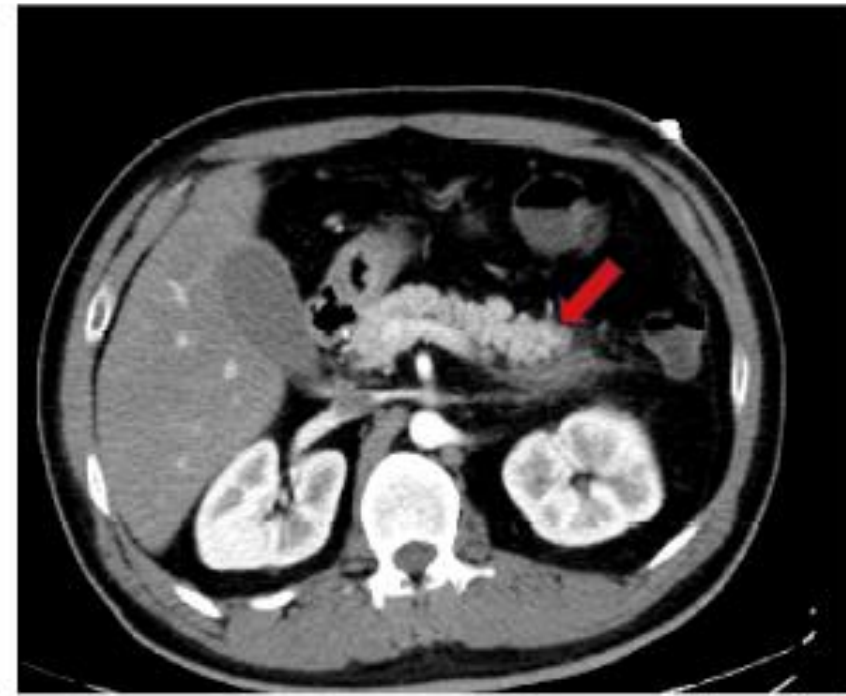
Advancing towards universal lesion detection capabilities: Establishing an ultra-large-scale real-world dataset, integrating with DeepLesion, and incorporating Vision-Language Pretraining.



8-cancer Summary									
	breast	colon	eso	sto	kid	lung	liver	panc	avg.
Detection	100.0	97.0	83.0	95.0	100.0	99.0	97.5	96.4	96.0
Detection_hit	99.0	94.0	80.0	92.0	99.0	98.0	97.0	94.3	94.2
Segmentation	57.7	33.7	40.4	19.0	46.8	30.4	48.7	64.1	42.6
Segmentation_TP	71.0	61.9	73.0	61.5	83.8	76.9	82.2	69.9	72.5
Specificity	98.2	93.3	96.1	83.6	89.4	86.5	78.1	96.9	90.3
Diagnosis	/	/	61.7	95.0	97.0	86.4	60.3	71.3	78.6

[1] CancerUniT: Towards a Single Unified Model for Effective Detection, Segmentation, and Diagnosis of Eight Major Cancers Using a Large Collection of CT Scans, IEEE ICCV, 2023

(6) Towards bigger models ... in Radiology & Pathology



(a) CT image

Finding: The pancreas demonstrates normal morphology, with slight fullness in the body and tail regions. The peripancreatic fat planes surrounding the body and tail are indistinct, with evidence of minor exudative changes.

Impression: Acute pancreatitis with minimal peripancreatic exudative changes.

(b) Diagnostic report



(c) Visual activation map of CLIP



(d) Visual activation map of our fVLM

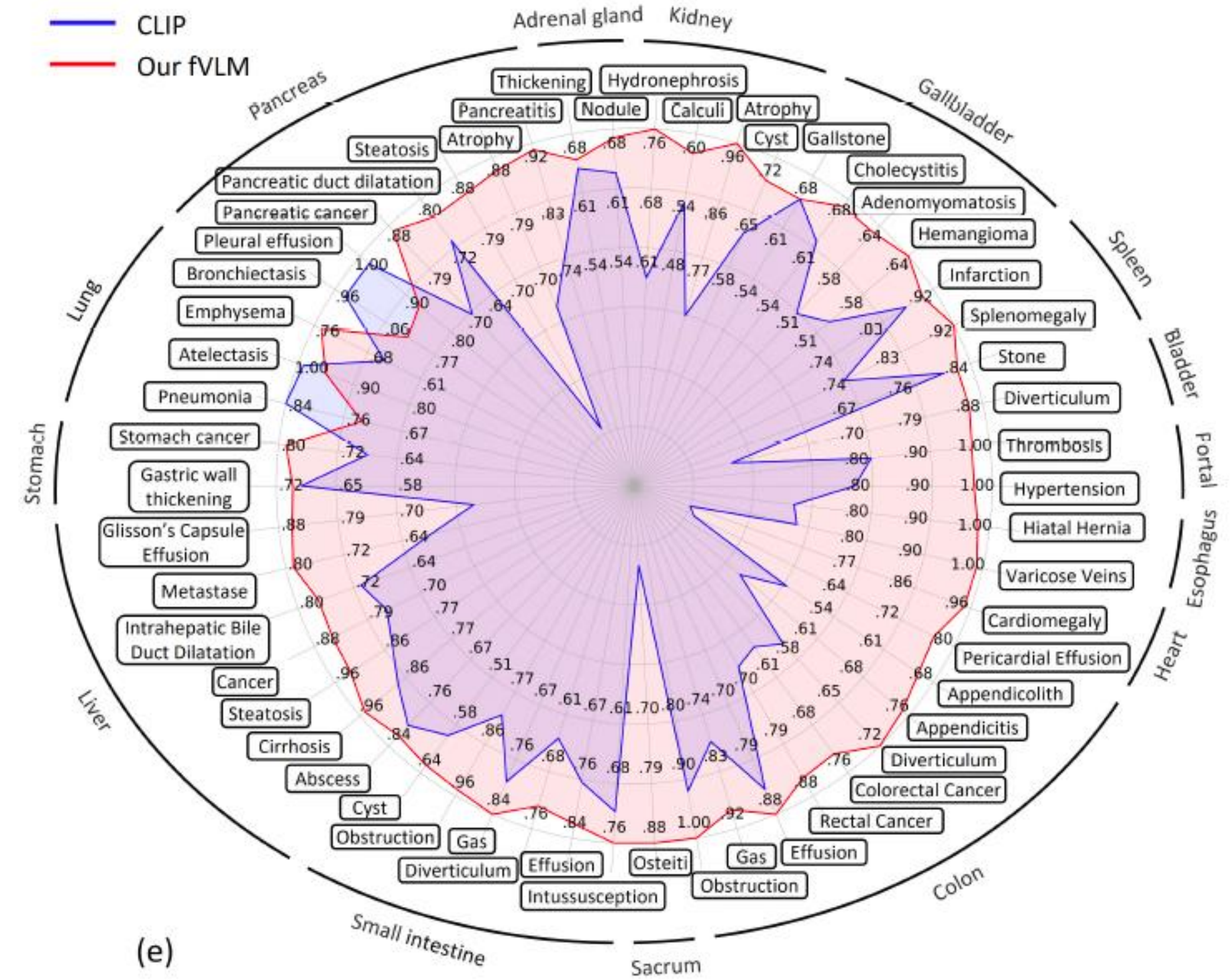
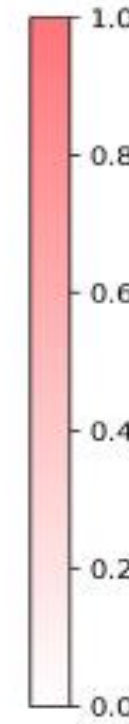
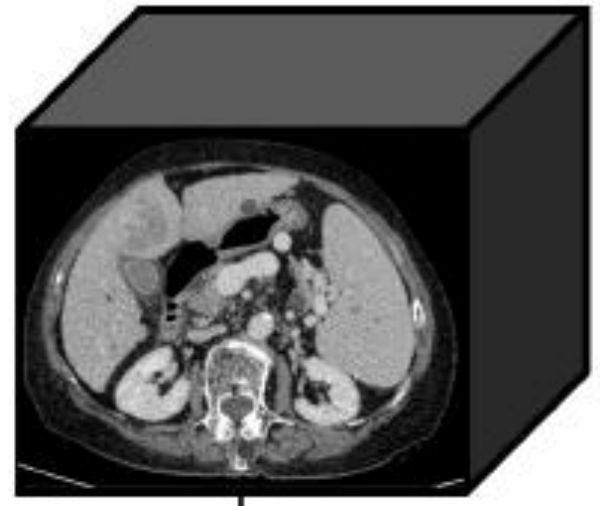


Figure 1: Comparative analysis of vanilla VLM (CLIP) and our fine-grained VLM (fVLM). (a,b) A representative CT slice and its corresponding radiological report. (c,d) Visual activation maps generated by CLIP and fVLM respectively, illustrating regions of interest for pancreatitis diagnosis. (e) Quantitative comparison of AUC scores across 54 disease diagnosis tasks in 15 anatomies.



Anatomy Segmentation




Original reports


Findings. The **liver** demonstrates normal position, size, and morphology. Diffuse decreased density is observed in the liver parenchyma. A small, round, non-enhancing low-density lesion with well-defined margins is visible in the right lobe of the **liver**. No dilation of intrahepatic or extrahepatic bile ducts is noted. The **kidneys** are normal in position and morphology; the right kidney shows punctate high-density shadows, with no enhancing small cystic lesions in both **kidneys**. The **pancreas** is normal in size and density, with no abnormal enhancement foci. There is no dilation of the pancreatic duct. A slightly hypo-dense nodule is observed on the greater curvature side of the gastric wall, with indistinct margins from the **stomach** wall, measuring approximately 14mm in diameter.


Impression. Fatty **liver**. Small cyst in the right lobe of the **liver**. There are small calculi in the right **kidney** and small cysts in both **kidneys**. Nodule on the greater curvature side of the **stomach**; consideration of a gastrointestinal stromal tumor is suggested.



Anatomy-level decomposed reports

 **(Findings)** The liver demonstrates normal position, size, and morphology. Diffuse decreased density is observed in the liver parenchyma. A small, round, non-enhancing low-density lesion with well-defined margins is visible in the right lobe of the liver. No dilation of intrahepatic or extrahepatic bile ducts is noted. **(Impression)** Fatty liver. Small cyst in the right lobe of the liver.

 **(Findings)** Punctate high-density shadows in the right kidney, with no enhancing small cystic lesions in both kidneys. **(Impression)** Small calculi in the right kidney and small cysts in both kidneys.

 **(Findings)** Normal in size and density, with no abnormal enhancement foci. No dilation of the pancreatic duct.


 **(Findings)** A slightly hypo-dense nodule is observed on the greater curvature side of the gastric wall, with indistinct margins from the stomach wall, measuring approximately 14mm in diameter. **(Impression)** Nodule on the greater curvature side of the stomach; consideration of a gastrointestinal stromal tumor is suggested.

Figure 2: Illustration of CT anatomy parsing (left) and diagnostic report decomposition (right).

(6) Towards bigger models ... in Radiology & Pathology

Table 2. The results of the tumor staging and survival analysis on pathology images. V-FM: Vision Foundation Model. VL-FM: Vision-Language Foundation Model. The best results are highlighted in **bold**, and the second-best results are in underlined.

Tasks			Tumor Staging (5 Fold - Macro F1 Score)			Survival Analysis (5 Fold - C-index)		
Methods	Backbone	Params.	BLCA	BRCA	LUAD	BLCA	BRCA	LUAD
<i>Two-stage MIL; Patch features: ResNet-50 [14]; Pre-training: ImageNet</i>								
MeanMIL	ResNet-50	25M + 4.1K	0.3760 ± 0.0689	0.1860 ± 0.0049	0.1800 ± 0.0089	0.5256 ± 0.0564	0.5303 ± 0.0535	0.5883 ± 0.0803
MaxMIL		25M + 4.1K	0.3640 ± 0.0745	0.1840 ± 0.0049	0.1880 ± 0.0223	0.5250 ± 0.0548	0.5036 ± 0.0707	0.5009 ± 0.0703
ABMIL [19]		25M + 0.9M	0.4100 ± 0.0498	0.2300 ± 0.0245	0.2540 ± 0.0242	0.5813 ± 0.0349	0.6118 ± 0.0331	0.6130 ± 0.0270
TransMIL [33]		25M + 2.7M	0.4200 ± 0.0268	0.2900 ± 0.0126	0.3020 ± 0.0417	0.5610 ± 0.0223	0.5689 ± 0.0273	0.5973 ± 0.0338
ILRA-MIL [41]		25M + 3.7M	0.4460 ± 0.0441	0.2400 ± 0.0261	0.2540 ± 0.0432	0.5570 ± 0.0219	0.5998 ± 0.0333	0.5725 ± 0.0519
<i>Two-stage MIL; Patch features: V-FM (GigaPath) [45]; pre-training: 171K WSIs</i>								
MeanMIL	ViT-G	1.1G + 6.1K	0.4880 ± 0.0426	0.3160 ± 0.0287	0.3140 ± 0.0233	0.6052 ± 0.0575	0.6255 ± 0.0391	0.6004 ± 0.0506
MaxMIL		1.1G + 6.1K	0.4240 ± 0.0939	0.1940 ± 0.0233	0.2360 ± 0.0233	0.5798 ± 0.0290	0.5803 ± 0.0357	0.5263 ± 0.0552
ABMIL [19]		1.1G + 1.2M	0.5220 ± 0.0662	0.3520 ± 0.0286	0.3820 ± 0.0331	0.5990 ± 0.0698	0.6594 ± 0.0437	0.6083 ± 0.0461
TransMIL [33]		1.1G + 2.9M	0.4720 ± 0.0458	0.2980 ± 0.0133	0.3200 ± 0.0420	0.6141 ± 0.0549	0.6291 ± 0.0580	0.5770 ± 0.0741
ILRA-MIL [41]		1.1G + 4.2M	<u>0.5320 ± 0.0487</u>	0.3640 ± 0.0294	0.3880 ± 0.0643	0.6153 ± 0.0397	<u>0.6528 ± 0.0430</u>	0.6006 ± 0.0714
<i>Two-stage MIL; Patch features: VL-FM (CONCH) [27]; pre-training: 1.17 million pathology image-caption pairs</i>								
MeanMIL	ViT-B	86M + 2.1K	0.5160 ± 0.0427	0.3040 ± 0.0326	0.3260 ± 0.0459	0.5977 ± 0.0305	0.6451 ± 0.0637	0.6174 ± 0.0704
MaxMIL		86M + 2.1K	0.4660 ± 0.0408	0.1860 ± 0.0080	0.2460 ± 0.0388	0.5701 ± 0.0588	0.5483 ± 0.0742	0.5731 ± 0.0626
ABMIL [19]		86M + 0.7M	0.5260 ± 0.0753	0.3340 ± 0.0320	0.3840 ± 0.0700	0.6057 ± 0.0344	0.6444 ± 0.0770	<u>0.6399 ± 0.0578</u>
TransMIL [33]		86M + 2.4M	0.5180 ± 0.0312	<u>0.3700 ± 0.0940</u>	0.3500 ± 0.0253	<u>0.6404 ± 0.0253</u>	0.6380 ± 0.0379	0.5879 ± 0.0389
ILRA-MIL [41]		86M + 3.2M	0.5160 ± 0.0561	0.3380 ± 0.0264	<u>0.3880 ± 0.0271</u>	0.6030 ± 0.0363	0.6532 ± 0.0409	0.6218 ± 0.0881
<i>Two-Stage Hierarchical Representation; pre-training: 104M pathology patches + 400K WSI regions</i>								
HIPT [4]	HIPT-ViT-6	24M + 2.2M	0.4660 ± 0.0185	0.3240 ± 0.0287	0.3480 ± 0.0299	0.5731 ± 0.0331	0.6139 ± 0.0446	0.5895 ± 0.0478
Prov-GigaPath (CLS token) [45]	ViT-LongNet	1.2G + 6.1K	0.4820 ± 0.0466	0.3060 ± 0.0388	0.2940 ± 0.0224	0.5435 ± 0.0635	0.5697 ± 0.0964	0.6044 ± 0.0294
Prov-GigaPath (feature) [45]	ViT-LongNet	1.2G + 1.2M	0.5200 ± 0.0486	0.3300 ± 0.0219	0.3500 ± 0.0452	0.5954 ± 0.0311	0.6193 ± 0.0313	0.6210 ± 0.0724
<i>End-to-End; pre-training: ImageNet + 10K WSIs</i>								
LongViT [38] (0.6x)	ViT-S	22M	0.2310 ± 0.0731	0.3049 ± 0.0137	0.2757 ± 0.0240	0.5885 ± 0.0439	0.6453 ± 0.0699	0.5890 ± 0.0236
LongViT [38] (2.5x)	ViT-S	22M	0.4963 ± 0.0908	0.3068 ± 0.0276	0.3155 ± 0.0358	0.5789 ± 0.0506	0.6403 ± 0.0588	0.6085 ± 0.0140
LongViT [38] (5.0x)	ViT-S	22M	0.4757 ± 0.0847	0.2979 ± 0.0265	0.2809 ± 0.0196	0.5708 ± 0.0377	0.6316 ± 0.0832	0.6030 ± 0.0267
<i>End-to-End; pre-training: ImageNet</i>								
Pixel-Mamba-Stage/Surv (2.5x)	Pixel-Mamba	6.2M	0.5334 ± 0.0608	0.3744 ± 0.0163	0.3917 ± 0.0125	0.6507 ± 0.0485	0.6707 ± 0.0728	0.6468 ± 0.0331

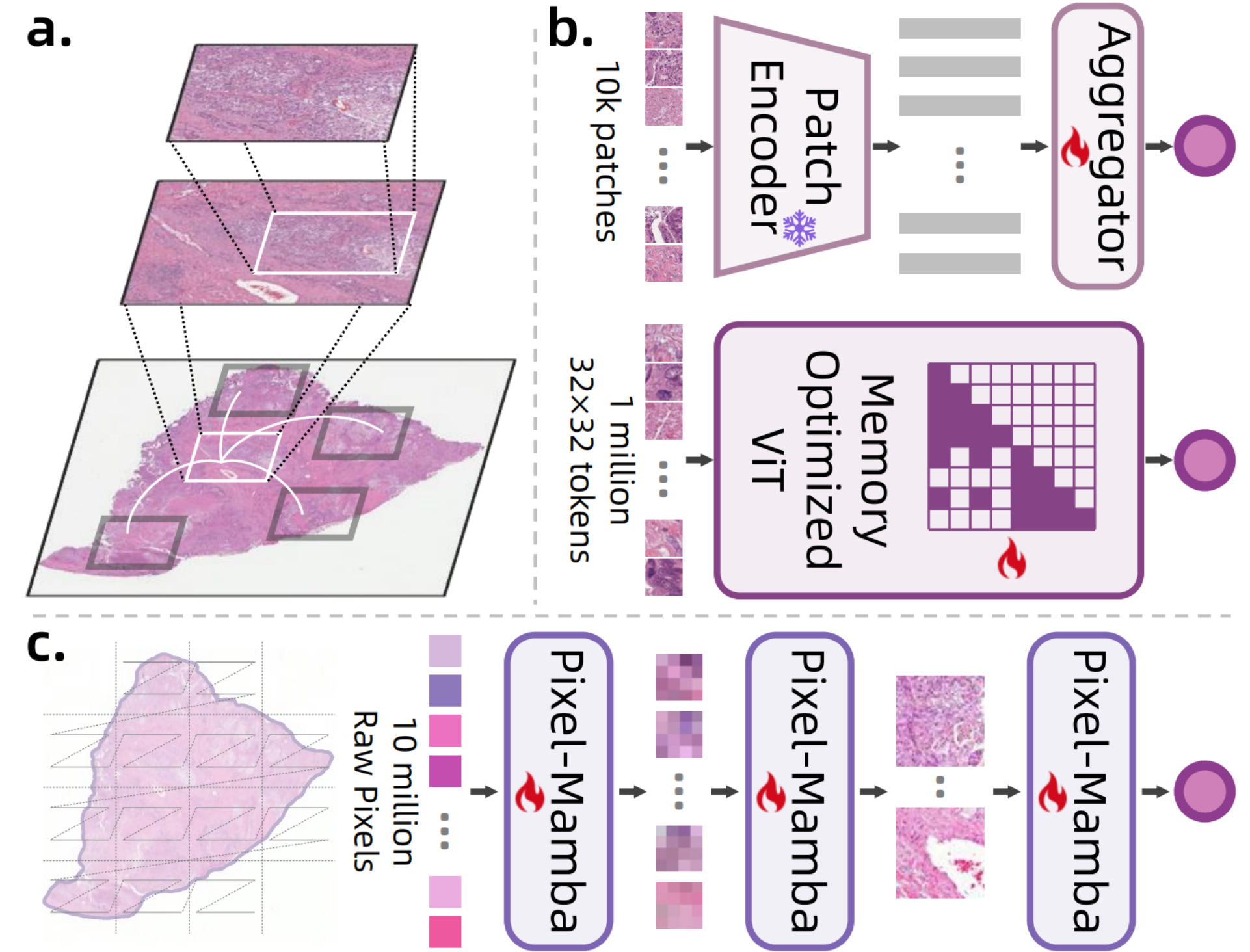
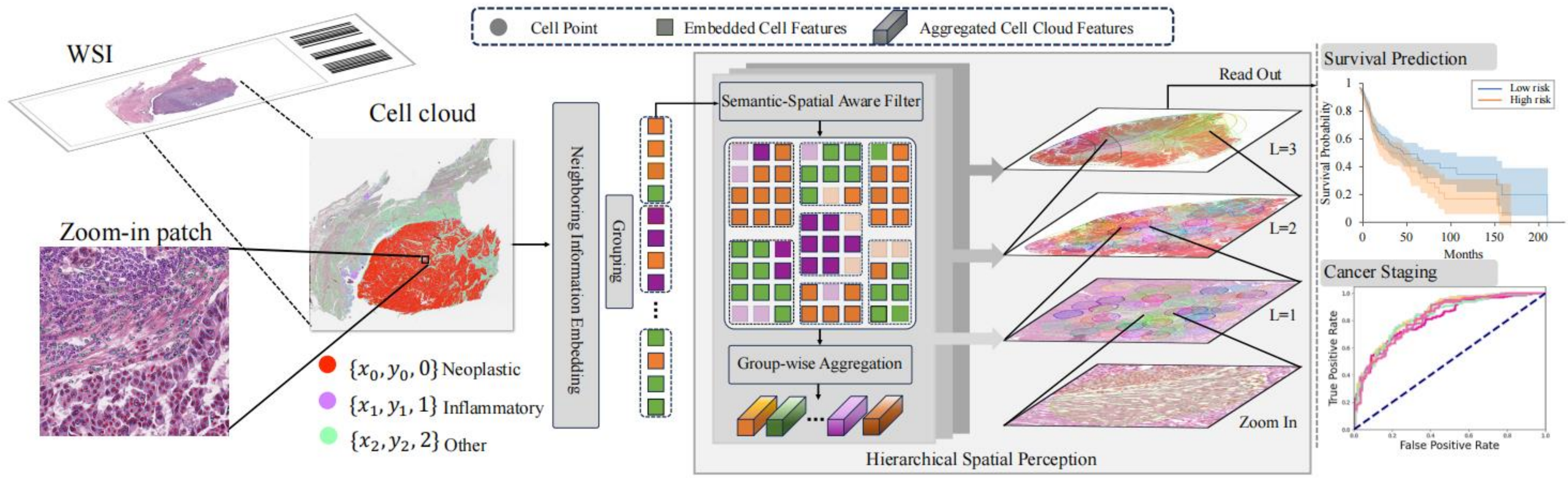


Figure 1. (a) Pathologists integrate observations from multiple regions across different scales to make a comprehensive assessment. (b) Frameworks of mainstream WSI analysis methods: a two-stage pipeline (top) and memory-optimized ViT (bottom, often with heavily pruned attention). (c) The proposed Pixel-Mamba, an end-to-end framework that combines progressive token expansion and the Mamba module to effectively integrate local inductive biases with long-range dependencies in a hierarchical manner.



		Method	Params.	BLCA	BRCA	COAD READ	HNSC	KIRC	LUAD	LUSC	PAAD	STAD	UCEC
Patch feature MIL	MeanPool (Patch)	4.1K	0.610 ± 0.024	0.643 ± 0.031	0.629 ± 0.186	0.572 ± 0.037	0.676 ± 0.051	0.571 ± 0.057	0.526 ± 0.051	0.672 ± 0.097	0.579 ± 0.085	0.727 ± 0.041	
	MaxPool (Patch)	4.1K	0.510 ± 0.038	0.589 ± 0.063	0.585 ± 0.078	0.561 ± 0.068	0.590 ± 0.076	0.480 ± 0.037	0.493 ± 0.075	0.404 ± 0.122	0.474 ± 0.102	0.650 ± 0.036	
	ABMIL [17]	0.9M	0.609 ± 0.028	0.656 ± 0.055	0.668 ± 0.167	0.606 ± 0.044	0.712 ± 0.057	0.614 ± 0.066	0.595 ± 0.073	0.696 ± 0.080	0.650 ± 0.098	<u>0.735</u> ± 0.039	
	TransMIL [36]	2.7M	0.600 ± 0.061	0.663 ± 0.058	0.634 ± 0.105	0.587 ± 0.041	0.634 ± 0.051	0.587 ± 0.076	0.589 ± 0.052	0.636 ± 0.121	0.581 ± 0.081	0.704 ± 0.052	
Graph	Patch-GCN [4]	1.4M	0.597 ± 0.022	0.628 ± 0.036	0.634 ± 0.121	0.566 ± 0.030	0.648 ± 0.074	0.617 ± 0.043	0.590 ± 0.063	0.668 ± 0.115	0.563 ± 0.048	0.678 ± 0.037	
	WiKG [20]	2.0M	0.638 ± 0.030	0.649 ± 0.036	0.722 ± 0.069	0.635 ± 0.033	0.657 ± 0.067	0.632 ± 0.038	0.635 ± 0.044	0.661 ± 0.112	0.672 ± 0.089	0.723 ± 0.036	
Point Cloud	MeanPool (Cell)	1.5K	0.535 ± 0.045	0.573 ± 0.070	0.639 ± 0.063	0.584 ± 0.027	0.536 ± 0.084	0.552 ± 0.068	0.563 ± 0.036	0.647 ± 0.045	0.569 ± 0.074	0.594 ± 0.034	
	MaxPool (Cell)	1.5K	0.476 ± 0.024	0.571 ± 0.055	0.514 ± 0.156	0.548 ± 0.054	0.497 ± 0.030	0.535 ± 0.035	0.521 ± 0.038	0.544 ± 0.065	0.527 ± 0.090	0.563 ± 0.095	
	PointNet [31]	3.5M	0.633 ± 0.025	0.665 ± 0.021	0.732 ± 0.044	0.650 ± 0.032	0.649 ± 0.035	0.638 ± 0.012	0.612 ± 0.021	0.715 ± 0.047	0.682 ± 0.075	0.706 ± 0.048	
	PointNet++ [32]	1.5M	0.613 ± 0.038	0.656 ± 0.036	0.743 ± 0.028	0.637 ± 0.030	0.627 ± 0.045	0.645 ± 0.020	<u>0.634</u> ± 0.036	0.702 ± 0.043	0.633 ± 0.055	0.673 ± 0.045	
	PTv3 [46]	38.8M	0.553 ± 0.039	0.536 ± 0.054	0.616 ± 0.069	0.571 ± 0.034	0.498 ± 0.037	0.591 ± 0.050	0.579 ± 0.041	0.631 ± 0.114	0.560 ± 0.023	0.600 ± 0.087	
CCFormer (ours)		2.0M	<u>0.645</u> ± 0.031	<u>0.688</u> ± 0.040	<u>0.753</u> ± 0.069	0.649 ± 0.032	0.658 ± 0.052	<u>0.657</u> ± 0.012	0.633 ± 0.019	<u>0.739</u> ± 0.044	0.687 ± 0.062	0.693 ± 0.049	
CCFormer (ours) + MeanPool (Patch)		2.7M	0.664 ± 0.019	0.704 ± 0.065	0.756 ± 0.076	0.652 ± 0.032	<u>0.696</u> ± 0.038	0.660 ± 0.018	0.626 ± 0.017	0.741 ± 0.066	0.704 ± 0.062	0.738 ± 0.052	

Table 1. Comparison of survival prediction with SOTA methods in C-Index (\uparrow).



Thanks for your interests!

In this rapid-changing era, **PATIENT FIRST** should always be our top priority of principles!

May Peace and Prosperity Prevail!