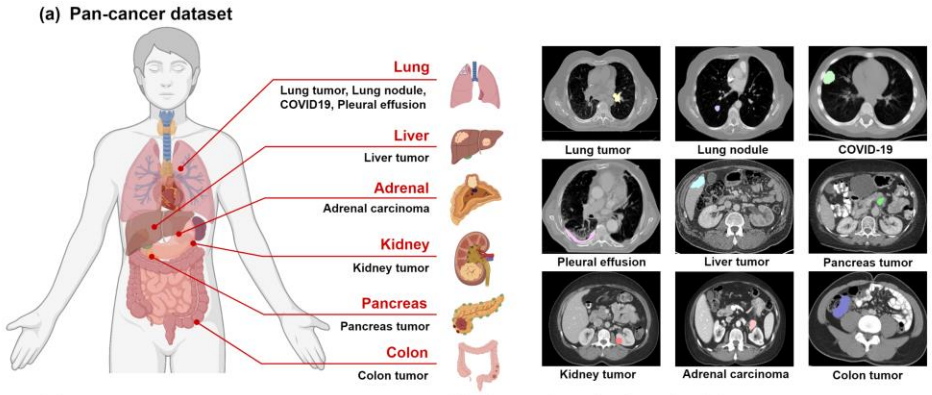
Glance and Focus Reinforcement for Pan-cancer Screening

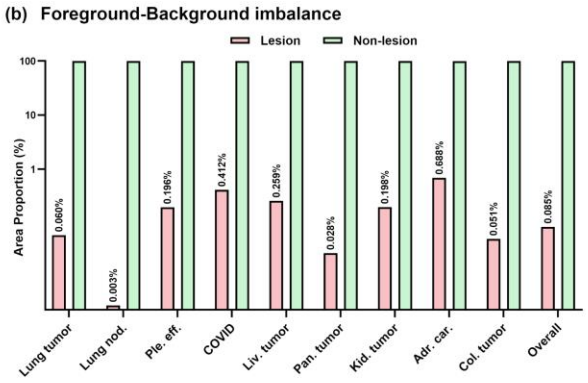
Linshan Wu

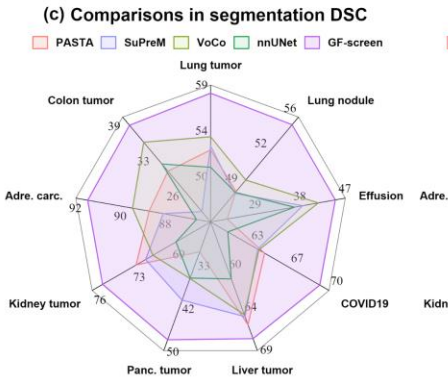
Overview

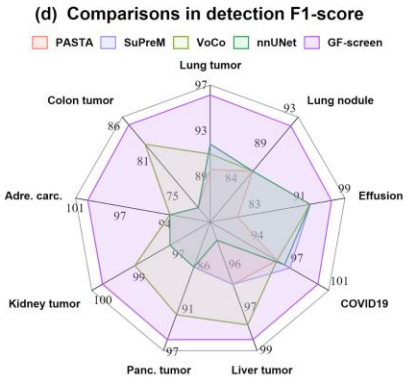
➢ Towards Precise and Efficient Cancer Screening

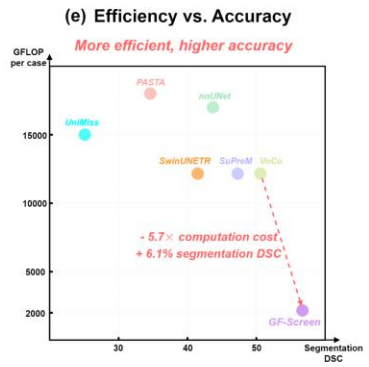
- ❑ **Motivation:** Severe foreground-background imbalance poses a significant challenge for cancer screening. Redundant regions increase false positives and decrease inference efficiency.
- ❑ **Method:** We introduce GF-Screen, a Glance and Focus reinforcement learning framework for pan-cancer screening, enabling precise and efficient pan-cancer screening.
- ❑ **Dataset:** 16 internal and 7 external datasets (5,117 CT scans) across 9 diverse lesion types.
- ❑ **Result:** Leading performance on FLARE25 leaderboard, surpassing the second-best by 25.6% DSC and 28.2% NSD.

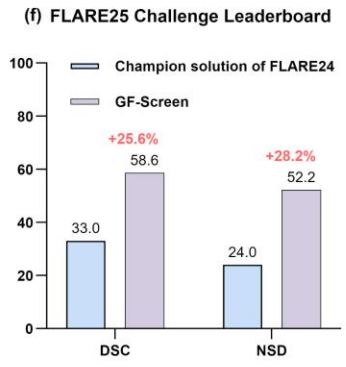












MICCAI2025 Daejeon REPUBLIC OF KOREA

FLARE 2025

Pan-cancer Segmentation

Task:	Results				FLARE25 Task1 Eval
#	Participant	Date	ID	Lesion DSC	Lesion NSD
1	linshanwu	2025-07-23 22:27	336484	**0.59**	0.52
2	ziyanhuang	2025-05-26 09:51	298216	**0.33**	0.24

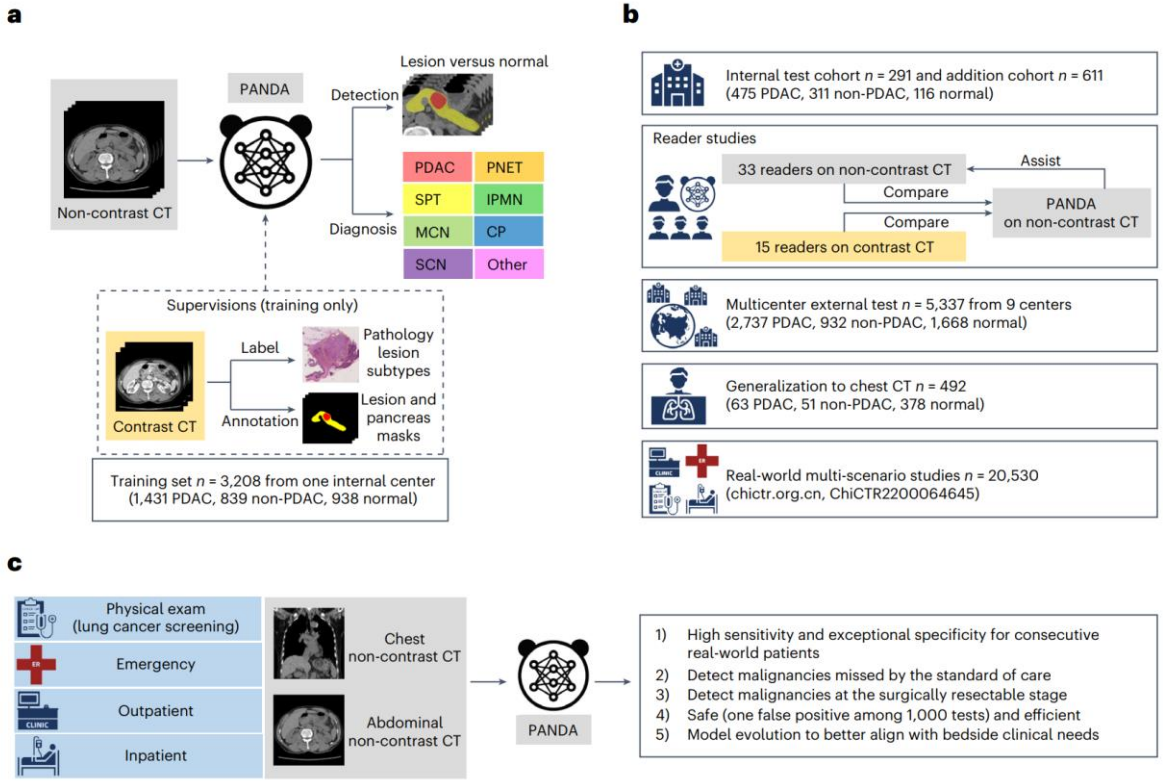
Outperform FLARE 22'23'24 champion solution (Shanghai AI lab) by a large margin

Linshan Wu, Jiaxin Zhuang, Hao Chen*. Glance and Focus Reinforcement for Pan-cancer Screening. ICLR, 2026

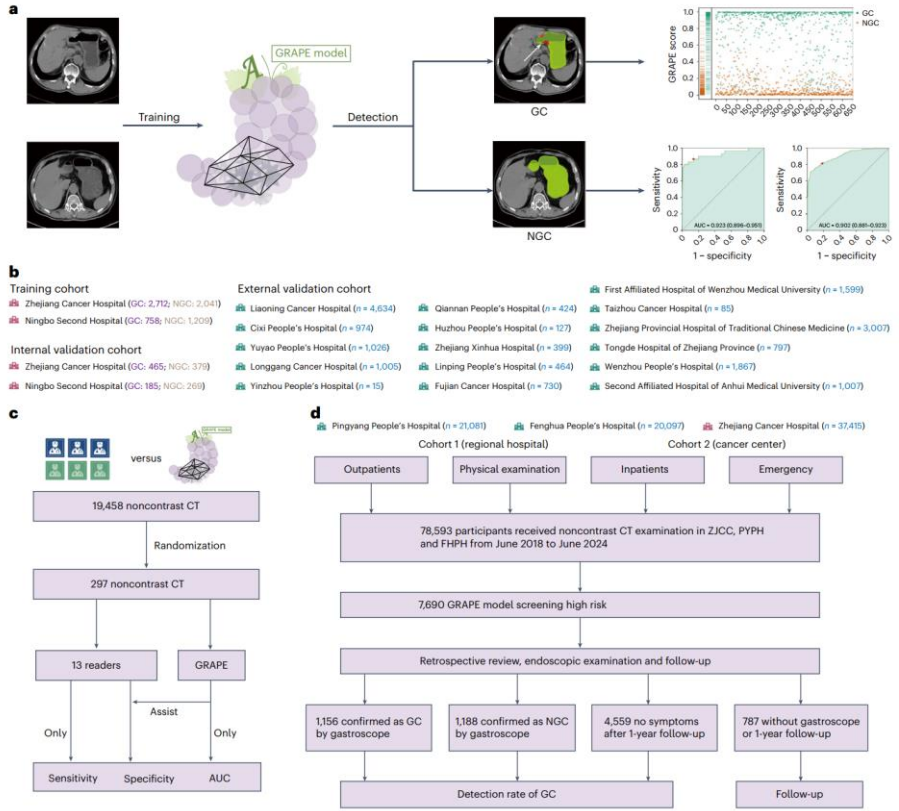
Related Works

➢ AI-driven Cancer Screening

- ❑ Cancer is a leading cause of death worldwide, accounting for nearly 10 million deaths annually according to the WHO. Effective cancer screening is crucial to reducing the mortality rate.
- ❑ AI-driven cancer screening in Computed Tomography (CT) scans has received increasing attention in clinical applications, since **CT is a low-cost and commonly used imaging protocol in large-scale routine physical examination.**



Pancreatic cancer screening



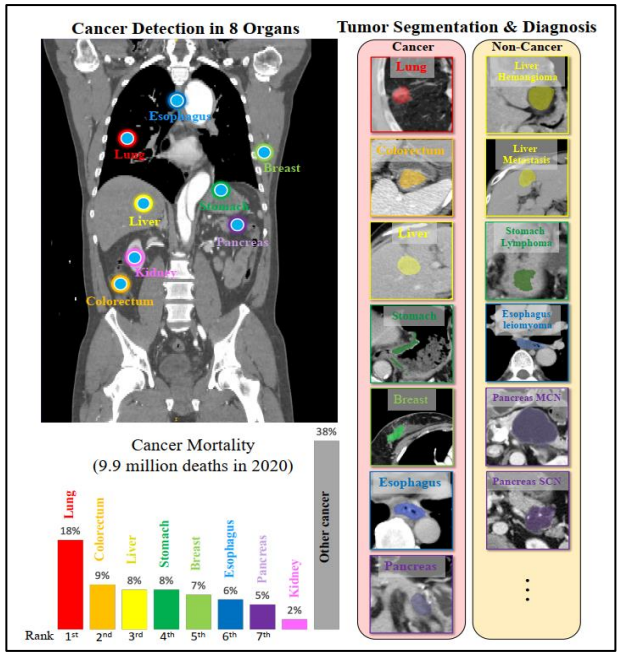
Gastric cancer screening

Kai Cao, et al. Large-scale pancreatic cancer detection via non-contrast CT and deep learning. Nature Medicine, 2023 (Alibaba DAMO)
Can Hu, et al. AI-based large-scale screening of gastric cancer from non-contrast CT imaging. Nature Medicine, 2025 (Alibaba DAMO)

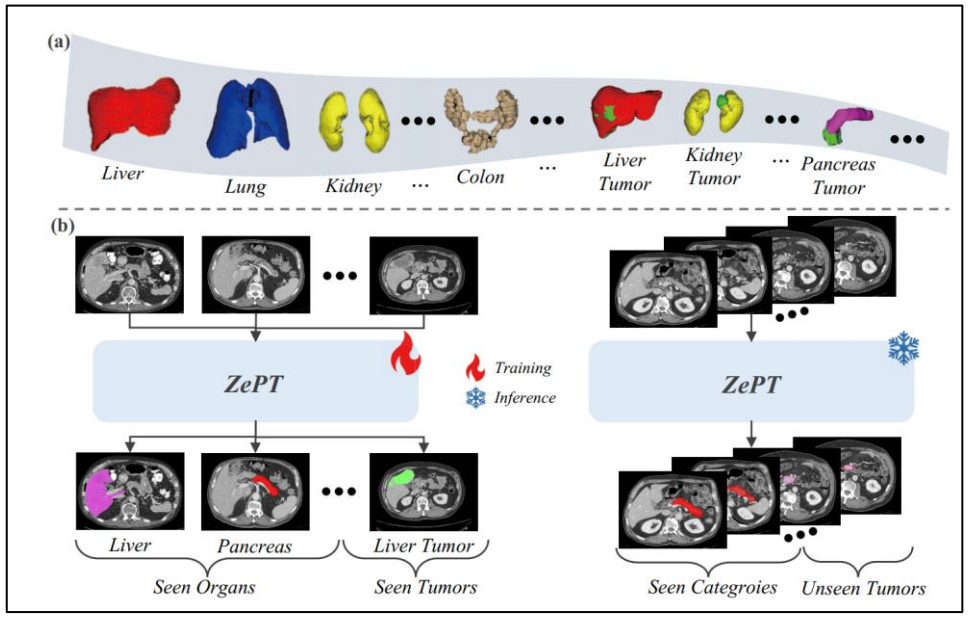
Related Works

➢ Pan-cancer Screening

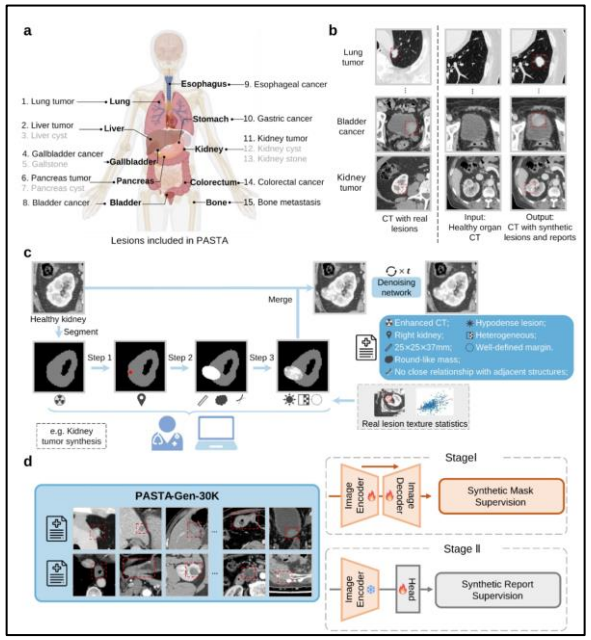
- ❑ Pan-cancer screening aims to develop a universal model for different lesion types, with profound significance in clinical practice.
- ❑ Pan-cancer screening is more challenging, which poses a significant challenge for AI models to localize diverse lesions across different organs.



Cancerunit (ICCV 23): Mask-Transformer Segmentation



ZEPT (CVPR 24): Zero-shot Tumor Segmentation



PASTA (arxiv 25): Tumor synthesis and segmentation

Jun Ma, et al. Unleashing the strengths of unlabelled data in deep learning-assisted pan-cancer abdominal organ quantification. The Lancet Digital Health, 2024 (University of Toronto)
Jieneng Chen, et al. 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 (JHU Alan Y.)
Yankai Jiang, et al. Zept: Zero-shot pan-tumor segmentation via query-disentangling and self-prompting. CVPR 2024. (Alibaba Damo)
Wenhui Lei, et al. A data-efficient pan-tumor foundation model for oncology CT interpretation. arxiv 2025. (SJTU & Harvard, Shaoting Z. & Pranav)
Maximilian Rokuss, et al. Lesionlocator: Zero-shot universal tumor segmentation and tracking in 3d whole-body imaging. CVPR 2025 (DKFZ group, German)

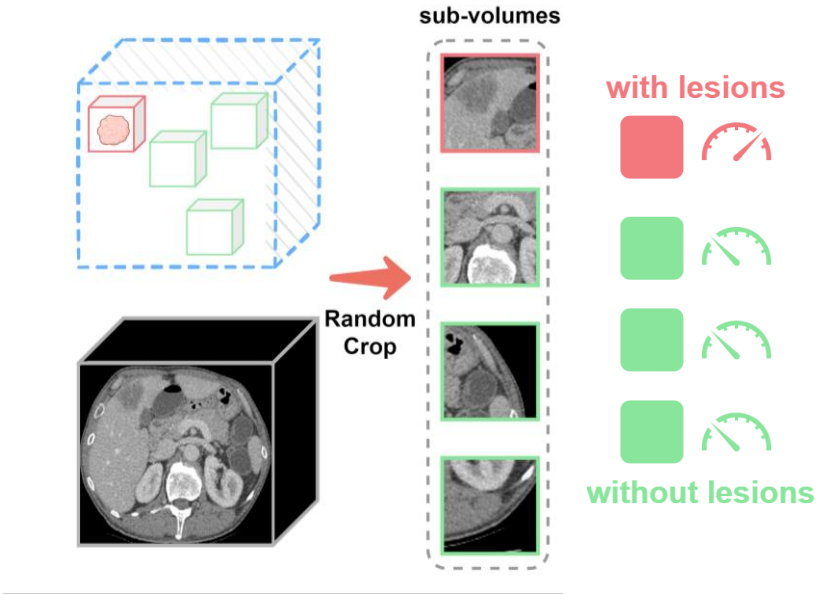
Related Works

➢ Pan-cancer Screening

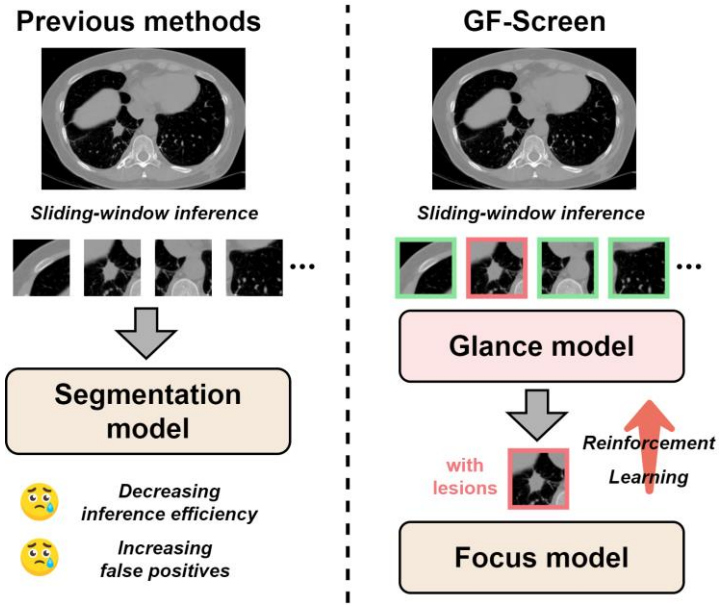
❑ Due to the limitation of computation costs, existing methods generally crop sub-volumes for segmentation. However,

- Lesions are tiny in CT and typically occupy small proportions. Thus, most of the cropped sub-volumes are without lesions.
- The redundant sub-volumes not only decrease inference efficiency but also increase false positives.

❑ **Motivation:** In contrast to AI models, experienced radiologists can rapidly disregard irrelevant regions and concentrate on potentially diseased regions. Typically, radiologists will glance at the entire CT, then focus on specific regions for precise diagnosis [1], inspiring us to explore a similar **glance and focus strategy** in AI-based cancer screening.



Lesions are tiny in CT



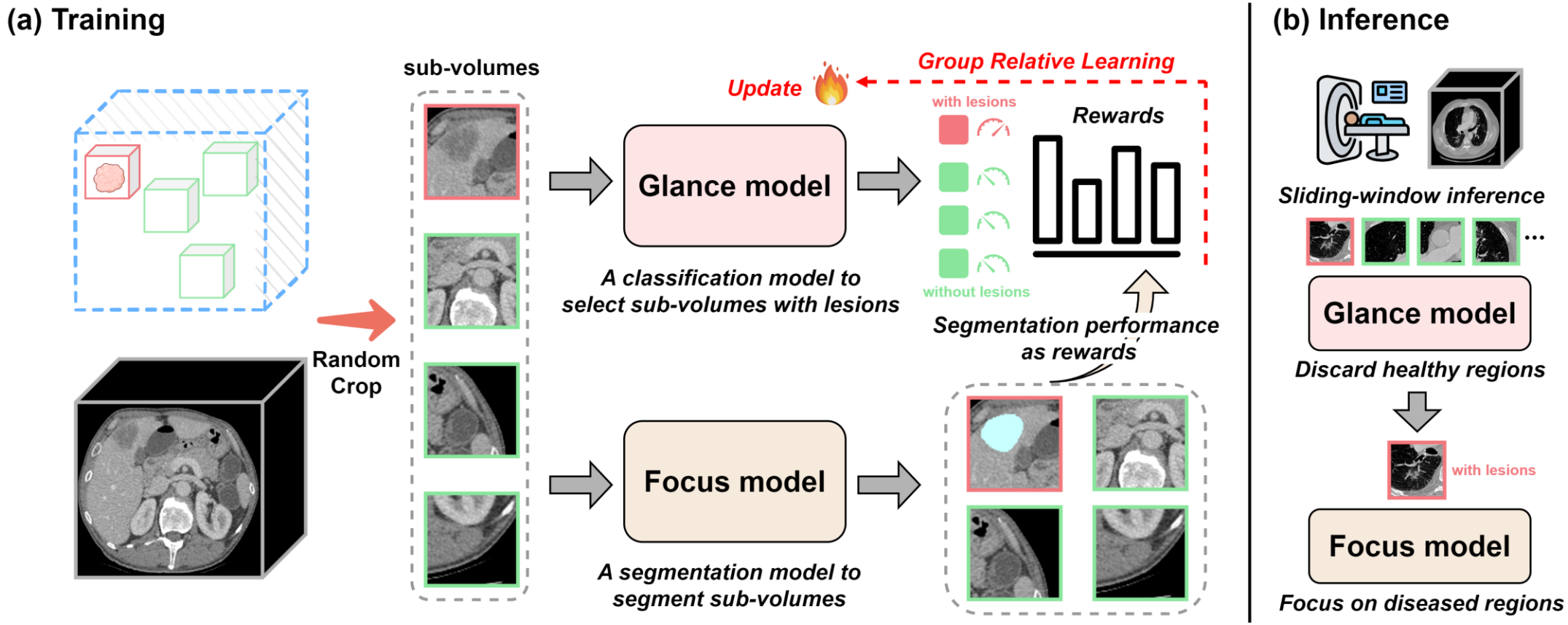
Paradigm Comparisons

[1] Kai Cao, et al.: Combining the strengths of radiologists and AI for breast cancer screening. The Lancet Digital Health, 2022

Method

➢ Towards Precise and Efficient Pan-cancer Screening

❑ We develop GF-Screen, a simple-yet-effective framework enabling both precise and efficient pan-cancer screening.



(a) In the training stage, we conduct segmentation on all sub-volumes and leverage the segmentation performance to reward the Glance model via a group relative learning paradigm. (b) In the inference stage, only the sub-volumes classified as "with lesions" will be input to the Focus model for segmentation, where the redundant regions will be discarded.

How to train the Glance model?

Method

➢ Towards Precise and Efficient Pan-cancer Screening

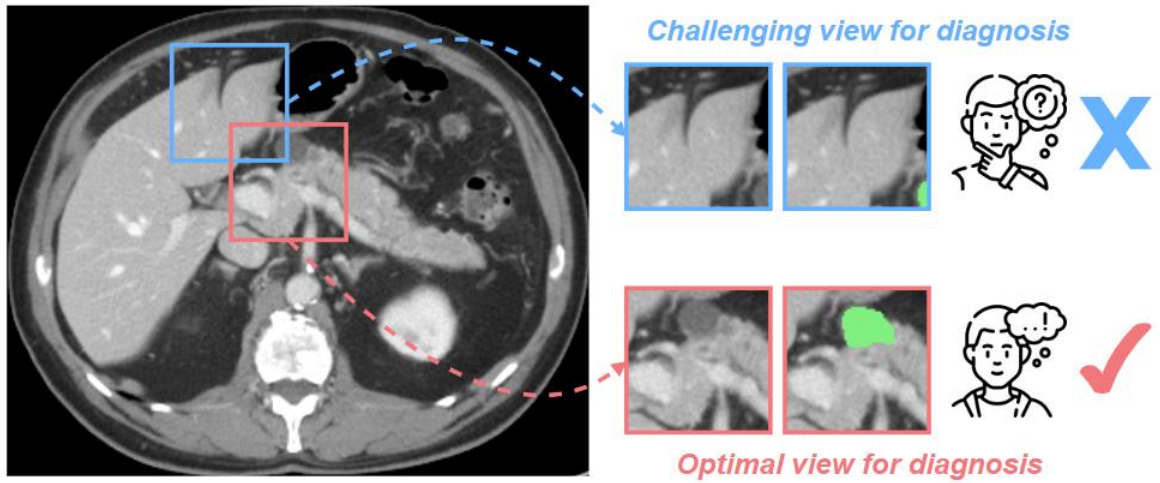
❑ ***How to effectively train the Glance model for discarding redundant healthy regions without compromising lesion recognition performance?***

❑ A straightforward way is to degrade the lesion masks m as binary categories y (“with lesions” or “without lesions”) for each sub-volume v_i, then employing a cross-entropy loss to train the Glance model G:

$$\mathbb{CE}(o_i, y_i) = -y_i * log(o_i) - (1 - y_i) * (1 - log(o_i)), \quad o_i = G(v_i),$$

❑ **Two fundamental shortcomings**:

- Most sub-volumes are without lesions, the **imbalance** will cause the model to overfit to negative cases and fail to pass diseased regions to the Focus model.
- Many lesion-containing sub-volumes present challenges due to either **partial inclusion** or **suboptimal viewing angles,** which significantly impede the classification training.



Visualization of sub-volume variation: blue regions represent challenging cases with partial lesions and poor viewing angles, while red regions indicate optimal diagnostic views containing complete lesion information.

✓ Compared to AI models, radiologists employ an intelligent strategy: they only **focus on the most diagnostically informative views** for accurate cancer diagnosis.

✓ GF-Screen aims to emulate this clinical expertise, which is developed to **identify the optimal views for segmentation**.

Key challenge: introduce extra information to identify the optimal views

Method

➢ Glance and Focus Reinforcement Learning

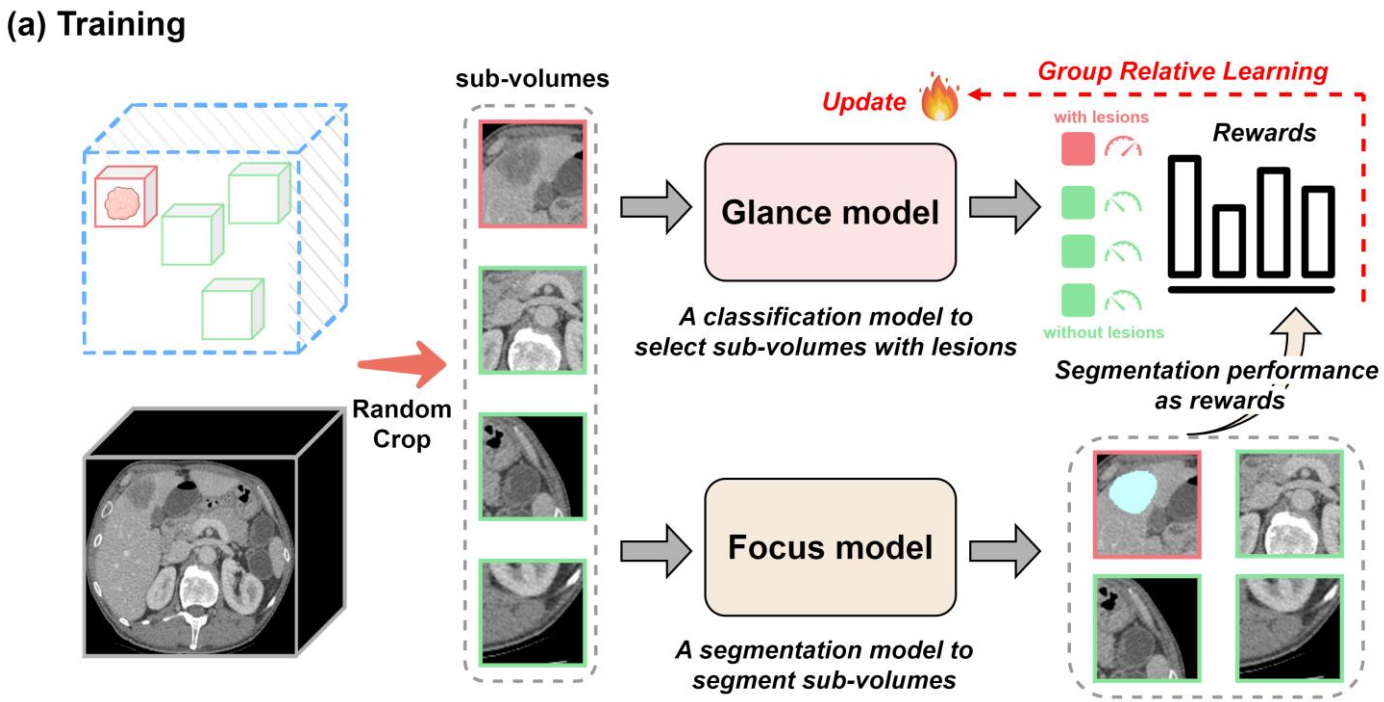
❑ We propose to **leverage the segmentation results of the Focus model to guide the Glance model,** encouraging the Glance model to prioritize the optimal diagnostic views that can yield better segmentation results.

❑ Since the selection operation of the Glance model is **non-differentiable** in segmentation training, we employ reinforcement learning for training.

❑ The Glance model acts as a policy model to deliver actions (select this sub-volume or not), and the Focus model serves as a reward model to reward the Glance model.

❑ Binary reward function: once the Focus model segments the lesions, reward $r = 1$; otherwise, $r = 0$.

$$r_i = \mathbb{1}(s_i \cap m_i \neq \emptyset).$$



How to optimize the policy model by reinforcement learning?

Method

➢ Glance and Focus Reinforcement Learning

- ❑ Traditional RL algorithms like Proximal Policy Optimization (PPO) require training an additional value model, which is unstable and difficult to converge.
- ❑ Group Relative Policy Optimization (GRPO) has received increasing attention in recent advanced LLMs, eliminating the requirement of value models.

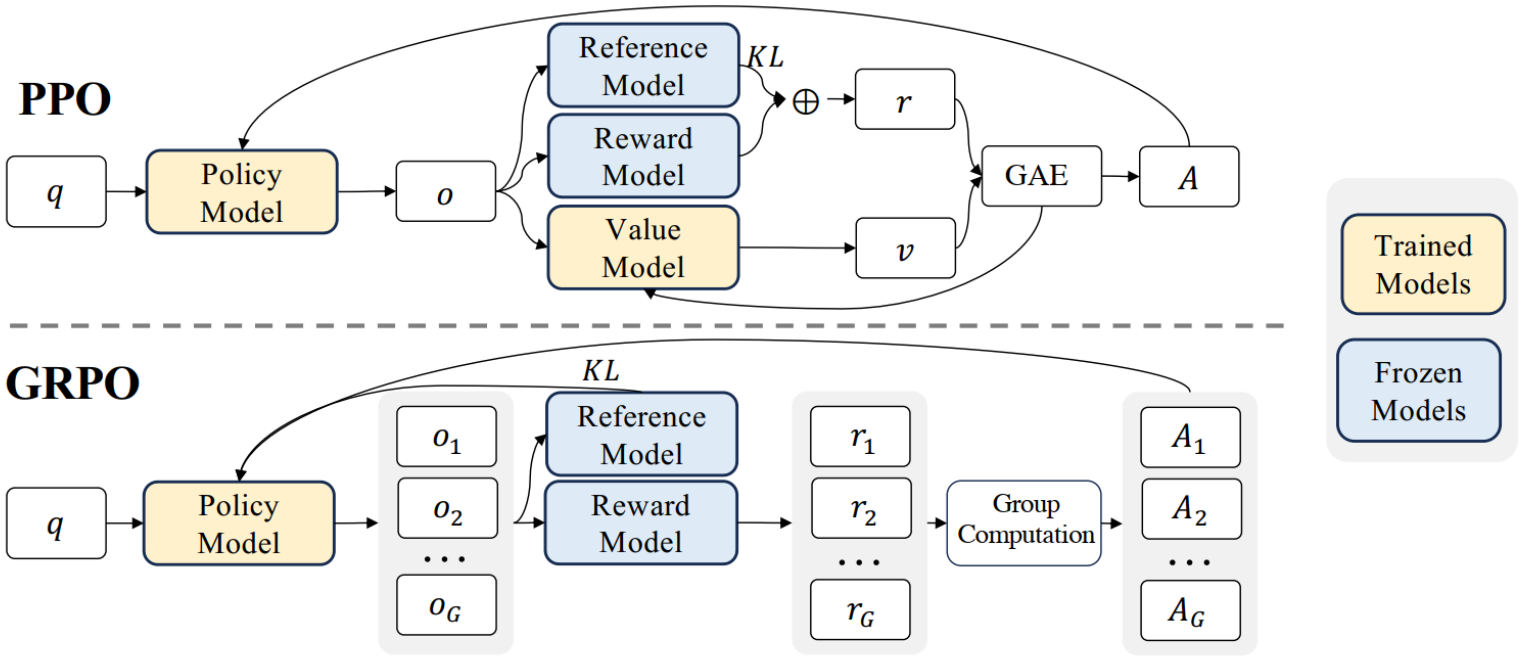


Figure 4 | Demonstration of PPO and our GRPO. GRPO foregoes the value model, instead estimating the baseline from group scores, significantly reducing training resources.

How to shift GRPO from LLMs to vision models?

John Schulma, et al. Proximal policy optimization algorithms. arxiv, 2017
Zhihong Shao, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. arxiv, 2024

Method

➢ Group Relative Learning for Sub-volumes Selection

- ❑ In GF-Screen, the group of cropped sub-volumes readily provides comparable candidates for GRPO.
- ❑ We introduce a group relative learning paradigm, which employs group relative comparison to prioritize high-advantage predictions and discard low-advantage predictions within the sub-volume group.
- ❑ In this way, we encourage the Glance model to select sub-volumes with higher advantages in the group, **improving inference efficiency and reducing false positives by eliminating low-advantage predictions.**

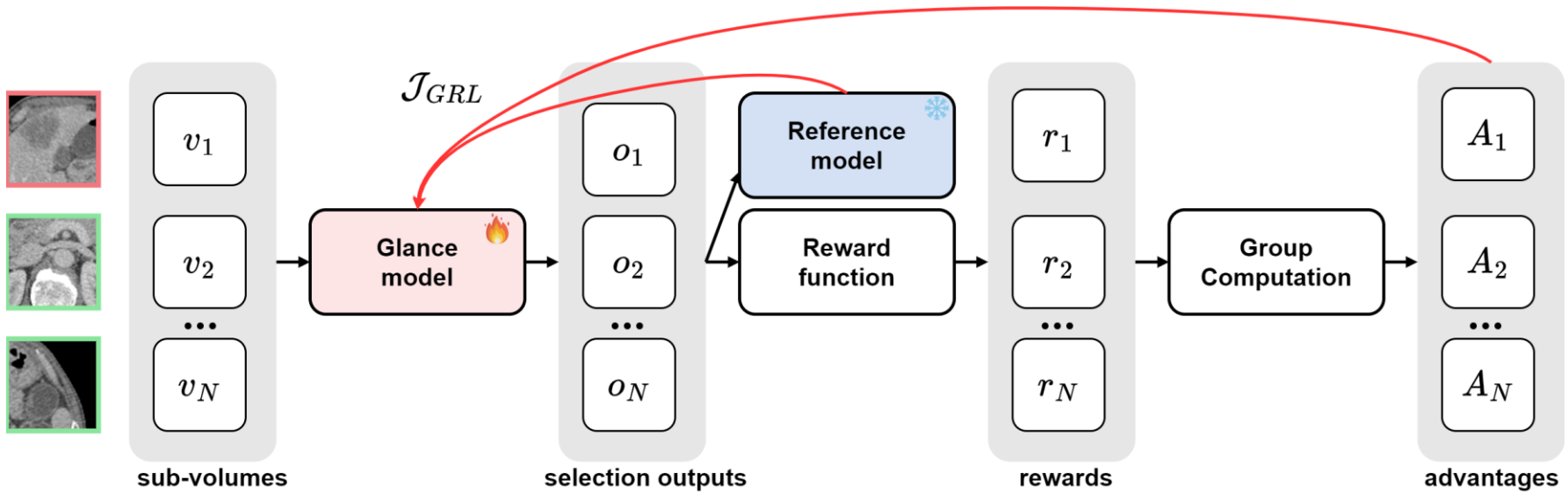
$$A_i = \frac{r_i - mean\{r_1, r_2, ..., r_N\}}{std\{r_1, r_2, ..., r_N\}},$$

$$\mathcal{J}_{GRPO}(\theta) = \mathbb{E}[\{o_i\}_{i=1}^{N} \sim G(v)]$$

$$= \frac{1}{N}\sum_{i=1}^{N}\{\min[\hat{s}_1 \cdot A_i,\ \hat{s}_2 \cdot A_i] - \beta\mathbb{D}_{KL}[G||G_{ref}] - \alpha\mathbb{CE}(o_i, y_i)\},$$

$$\mathbb{D}_{KL}(G||G_{ref}) = \frac{G_{ref}(o_i|v_i)}{G(o_i|v_i)} - \log\frac{G_{ref}(o_i|v_i)}{G(o_i|v_i)} - 1,$$

$$\hat{s}_1 = \frac{G(v_i)}{G_{ref}(v_i)}, \quad \hat{s}_2 = \text{clip}\left(\frac{G(v_i)}{G_{ref}(v_i)}, 1-\epsilon, 1+\epsilon\right),$$



The Glance model is trainable while the reference model is frozen. The referenced model is initialized as the Glance model and updated by fixed epochs. We generate selection outputs o from the input sub-volumes v, then use the reward function to calculate the rewards r. Finally, we compute the relative advantages A via Generalized Advantage Estimation (GAE).

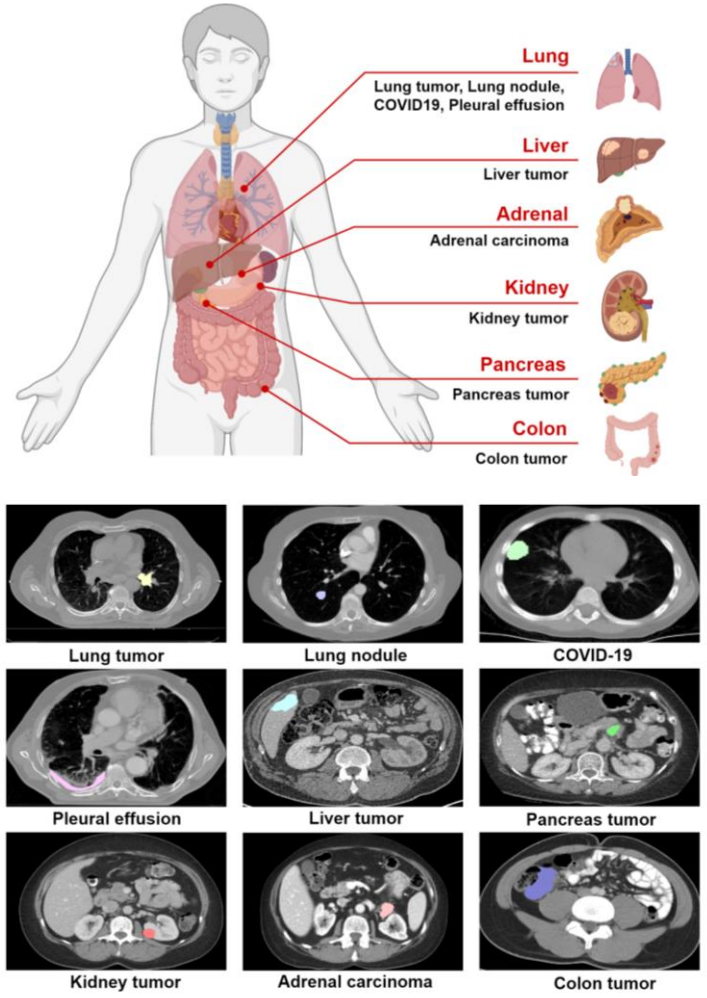
Results

➢ Extensive experiments on 16 internal and 7 external datasets

- ❑ We curate 5,117 CT scans from 16 internal and 7 external datasets for training and evaluation, across 9 different types of lesions.

Table A1: The datasets used for training and validation in pan-cancer screening. For the Atlas dataset (Bassi et al., 2025), we select the healthy cases based on the provided reports and ensure they do not overlap with other datasets. Overall, there are a total of 5,117 CT scans from 16 internal and 7 external datasets covering 9 different types of lesions.

Dataset	Tumor/Lesion Types	Scans (Train/Val)
Internal datasets		
MSD06-Lung (Antonelli et al., 2022)	Lung tumor	46/17
NSCLC-Radiogenomics (Bakr et al., 2018)	Lung tumor	68/20
NSCLC-Radiomics (Zhang et al., 2017)	Lung tumor	323/92
LIDC (Armato III et al., 2011)	Lung nodule	826/184
NSCLC-PleuralEffusion (Arrieta et al., 2019)	Pleural effusion	63/15
COVID-19 (Roth et al., 2022)	COVID-19 infection	158/41
MSD03-Liver (Antonelli et al., 2022)	Liver tumor	95/23
MSD08-Hepatic (Antonelli et al., 2022)	Liver tumor	248/55
WAWTACE (Bartnik et al., 2024)	Liver tumor	172/47
HCC (Morshid et al., 2019)	Liver tumor	60/14
MSD07-Pancreas (Antonelli et al., 2022)	Pancreas tumor	226/55
PANORAMA (Alves et al., 2024)	Pancreas tumor	376/106
KiTS23 (Heller et al., 2023)	Kidney tumor	389/99
Adrenal (Ahmed et al., 2020)	Adrenocortical carcinoma	44/8
MSD10-Colon (Antonelli et al., 2022)	Colon tumor	103/23
Atlas (Bassi et al., 2025)	Without lesions	628/157
External datasets		
Rider (Aerts et al., 2014)	Lung tumor	0/56
Corona (Paiva, 2020)	COVID-19 infection	0/10
IRCADb (Soler et al., 2010)	Liver tumor	0/20
CHAOS (Kavur et al., 2021)	Without lesions	0/20
TCIA-Pancreas (Roth et al., 2016)	Without lesions	0/80
FLARE23 (Ma et al., 2024)	Pan-cancer	0/50
FLARE25 (val&test) (Ma et al., 2024)	Pan-cancer	0/100
Total		**3,825/1,292 (5,117)**



Results

➢ Superior performance on public leaderboard

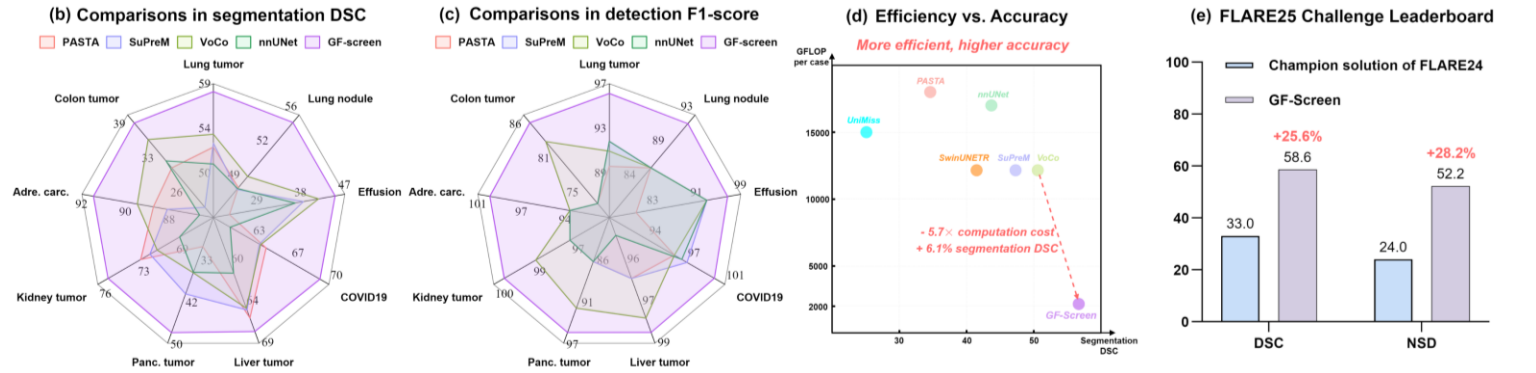
❑ FLARE25 challenge requires models with high accuracy and efficiency (less than 60s per CT scan). GF-Screen achieves high accuracy and efficiency, demonstrating superior performance on the public leaderboard.

Table 2: Pan-cancer segmentation performance (DSC %) on internal datasets across 9 types of lesions. Note that for pan-cancer screening, we train one universal model for different types of lesions. Interactive segmentation methods rely on manual prompts, which gain higher performance but cannot be applied to automatic screening. '-' denotes that this model cannot apply to this lesion type. The dataset-wise results are shown in Appendix Table A5.

Method	Lung tumor	Lung nodule	Pleural effusion	COVID-19 infection	Liver tumor	Pancreas tumor	Kidney tumor	Adreno. carcinoma	Colon tumor	Average (per type)
Interactive segmentation (require manual prompts)										
SegVol (point+text)	68.9	-	-	-	71.8	-	68.3	90.5	69.7	-
SAT (text)	51.0	28.0	-	67.2	60.1	34.2	67.9	-	35.3	-
ULS (point)	76.5	-	-	-	65.1	-	57.9	82.9	68.3	-
LesionLocator (point)	77.1	-	-	-	76.8	-	67.9	89.1	75.3	-
Pan-cancer segmentation (automatic screening)										
nnUNet	50.4	48.1	36.8	61.5	61.2	35.9	68.3	86.6	30.6	53.3
SwinUNETR	47.8	43.6	26.3	58.5	57.0	27.6	70.8	83.0	21.0	48.6
3D UX-Net	42.4	30.1	10.2	56.1	54.4	16.3	62.3	82.5	6.9	40.1
CLIP-driven	46.7	36.8	22.8	63.1	61.2	23.7	67.0	87.5	20.2	47.6
TransUNet	51.2	36.7	17.9	60.8	59.2	29.9	64.7	87.7	14.5	46.9
UniMiSS+	49.1	40.3	23.6	62.0	55.5	22.1	64.1	83.6	12.1	45.9
VoCo	53.4	49.4	41.6	64.2	65.1	36.0	70.5	89.3	34.6	56.1
SuPreM	52.5	48.0	38.5	64.1	65.3	40.2	71.2	88.0	21.9	54.4
PASTA	52.1	48.2	23.3	64.7	66.2	30.8	72.1	88.6	29.3	52.8
GF-Screen	**57.7**	**55.2**	**45.0**	**69.5**	**67.7**	**47.9**	**75.3**	**91.2**	**37.7**	**60.8**

Table 3: Pan-cancer detection performance (F1-Score %) on internal datasets across 9 types of lesions. We did not compare the interactive-based methods since the prompt will already provide whether this scan contains lesions or not. We report the average results for each CT scan.

Method	Lung tumor	Lung nodule	Pleural effusion	COVID-19 infection	Liver tumor	Pancreas tumor	Kidney tumor	Adreno. carcinoma	Colon tumor	Average (per type)
nnUNet	92.0	86.3	92.8	96.9	95.5	85.8	97.5	93.3	72.2	90.2
SwinUNETR	92.5	90.3	88.9	96.2	97.8	91.9	99.5	93.3	72.0	92.5
3D UX-Net	89.7	87.5	84.6	97.4	95.5	88.9	97.5	65.8	35.7	86.3
CLIP-driven	90.2	88.2	88.9	97.4	96.7	85.8	98.8	75.5	46.7	88.1
TransUNet	90.2	87.5	79.9	96.2	95.5	87.8	99.9	93.3	64.7	89.1
UniMiSS+	91.1	86.0	79.9	98.9	95.5	85.8	97.5	93.3	51.6	86.7
VoCo	91.1	86.3	92.8	96.2	97.8	91.9	98.8	93.3	82.0	92.2
SuPreM	92.0	86.3	92.8	97.4	96.7	85.8	97.5	93.3	72.2	90.4
PASTA	89.7	86.3	79.9	96.2	96.7	85.8	97.5	93.3	72.2	88.6
GF-Screen	**96.4**	**91.9**	**96.6**	**100.0**	**98.2**	**95.1**	**100.0**	**100.0**	**85.0**	**95.9**



MICCAI2025 Daejeon REPUBLIC OF KOREA

FLARE 2025

Pan-cancer Segmentation

Task:	Results				FLARE25 Task1 Eval
#	Participant	Date	ID	Lesion DSC	Lesion NSD
1	linshanwu	2025-07-23 22:27	336484	**0.59**	0.52
2	ziyanhuang	2025-05-26 09:51	298216	**0.33**	0.24

Outperform FLARE 22'23'24 champion solution (Shanghai AI lab) by a large margin

Jun Ma, et al. Unleashing the strengths of unlabelled data in deep learning-assisted pan-cancer abdominal organ quantification. The Lancet Digital Health, 2024 (University of Toronto)