

Title: A Data-Efficient Pan-Tumor Foundation Model for Oncology CT Interpretation

Wenhui Lei^{1,2*}, Hanyu Chen^{3*}, Zitian Zhang^{4*}, Luyang Luo^{5*}, Qiong Xiao³, Yannian Gu¹, Peng Gao³, Yankai Jiang², Ci Wang⁴, Guangtao Wu³, Tongjia Xu³, Yingjie Zhang³, Xiaofan Zhang^{1,2#}, Pranav Rajpurkar^{5#}, Shaoting Zhang^{1#}, Zhenning Wang^{3#}

¹Shanghai Jiao Tong University, Shanghai, P.R. China

²Shanghai AI Lab, Shanghai, P.R. China

³Department of Surgical Oncology and General Surgery, Key Laboratory of Precision Diagnosis and Treatment of Gastrointestinal Tumors, Ministry of Education, The First Hospital of China Medical University, Liaoning, P.R. China

⁴Department of Radiology, The First Hospital of China Medical University, Liaoning, P.R. China

⁵Harvard Medical School, Boston, MA, USA

*Contributed equally as first authors

#Correspondence to Xiaofan Zhang, Pranav Rajpurkar, Shaoting Zhang or Zhenning Wang

Keywords: Pan-Tumor Foundation Model, Synthetic Data, Computed Tomography, Oncology AI

Abstract

Artificial intelligence-assisted imaging analysis has made substantial strides in tumor diagnosis and management. Here we present PASTA, a pan-tumor CT foundation model that achieves state-of-the-art performance on 45 of 46 representative oncology tasks—including lesion segmentation, tumor detection in plain CT, tumor staging, survival prediction, structured report generation, and cross-modality transfer learning, significantly outperforming the second-best models on 35 tasks. This remarkable advancement is driven by our development of PASTA-Gen, an innovative synthetic tumor generation framework that produces a comprehensive dataset of 30,000 CT scans with pixel-level annotated lesions and paired structured reports, encompassing malignancies across ten organs and five benign lesion types. By leveraging this rich, high-quality synthetic data, we overcome a longstanding bottleneck in the development of CT foundation models—specifically, the scarcity of publicly available, high-quality annotated datasets due to privacy constraints and the substantial labor required for scaling precise data annotation. Encouragingly, PASTA demonstrates exceptional data efficiency with promising practical value, markedly improving performance on various tasks with only a small amount of real-world data. The open release of both the synthetic dataset and PASTA foundation model effectively addresses the challenge of data scarcity, thereby advancing oncological research and clinical translation.

Main Text

Malignant tumors continue to rank among the leading causes of death worldwide¹, with imaging techniques—particularly computed tomography (CT)—serving as a cornerstone for tumor diagnosis and evaluation. While specialized AI-driven CT imaging models for tumor detection, staging, and prognostic prediction have substantially advanced conventional medical imaging, they often face limitations including narrow disease coverage, the need for large amounts of specialized data, and suboptimal generalizability²⁻⁷. Recently, foundation models pre-trained on large, heterogeneous datasets have demonstrated strong label efficiency and robustness in downstream tasks⁸⁻¹⁶. However, the development of a pan-tumor CT foundation model presents several challenges.

Firstly, tumor regions occupy only a small fraction of the entire CT volume, and existing CT foundation models, typically based on self-supervised learning algorithms, learn broad CT features from unlabeled datasets¹⁷⁻²⁰. While these models are effective for non-disease-related tasks, such as organ segmentation, their pre-training objectives do not focus on lesions, limiting their ability to extract tumor-specific features. An exception is Foundation Model for Cancer Imaging Biomarkers (FMCIB)²¹, which pre-trains models to distinguish between lesions and non-lesions at the patch level. However, the absence of fine-grained tumor annotations limits its ability to capture detailed feature representations. Secondly, high-quality publicly available datasets that include both imaging reports and pixel-level lesion segmentation annotations are scarce, and the available data often cover a limited range of cancer types⁹. This hinders the development of a generalized cross-tumor foundation model through supervised learning. A notable attempt in this direction is the Supervised Pre-trained Model (SuPreM)²², which pre-trains on 9,262 3D CT scans, including 25 anatomical structures and five organs with tumor pseudo-labels. While promising, it still suffers from limited tumor type coverage and lacks precise tumor annotations, limiting generalizability. Lastly, the creation and sharing of large, multi-tumor lesion datasets is constrained by low tumor incidence, high annotation costs, and privacy concerns.

To address these challenges, we introduce Pan-tumor Analysis with Synthetic Training Augmentation

(PASTA), a 3D-CT foundation model designed to overcome data scarcity and enable unified cross-tumor analysis. We first developed PASTA-Gen, a generative model capable of synthesizing lesions across ten organs and 15 lesion types, including ten common malignancies and five benign lesions. Using PASTA-Gen, we created the PASTA-Gen-30K dataset, consisting of 30,000 synthetic 3D-CT scans with lesion masks, and structured textual descriptions. We then pre-trained PASTA in two stages: one focused on lesion semantic segmentation and the other on vision-language alignment. To evaluate the performance of the PASTA foundation model, we conducted systematic comparisons with leading medical image analysis frameworks, including nnUNet²³, Universal²⁴, and outstanding 3D CT foundation models such as ModelsGenesis²⁰, SupreM²², and FMCIB²¹. PASTA was further evaluated across 46 downstream tasks, achieving the best performance in 45 tasks and significantly outperforming the second-best model in 35 of them. These tasks included lesion segmentation, tumor identification in plain CT, tumor staging, survival prediction, structured report generation, and cross-modality transfer learning. Notably, PASTA excels in few-shot learning, outperforming the next best model by 31.2% in average Dice Similarity Coefficient (DSC) for lesion segmentation with just two labeled examples, and achieving comparable performance to full-data training across ten lesion types. In tumor detection from plain CT scans, PASTA reaches AUCs of 0.96 or higher for gallbladder, rectal, and bladder cancers. Overall, by leveraging the PASTA-Gen-30K dataset with precise lesion masks and reports, PASTA demonstrated its potential as a highly effective foundation model. We have made both PASTA-Gen-30K and PASTA publicly available, advancing AI in the domain of pan-tumor analysis in CT.

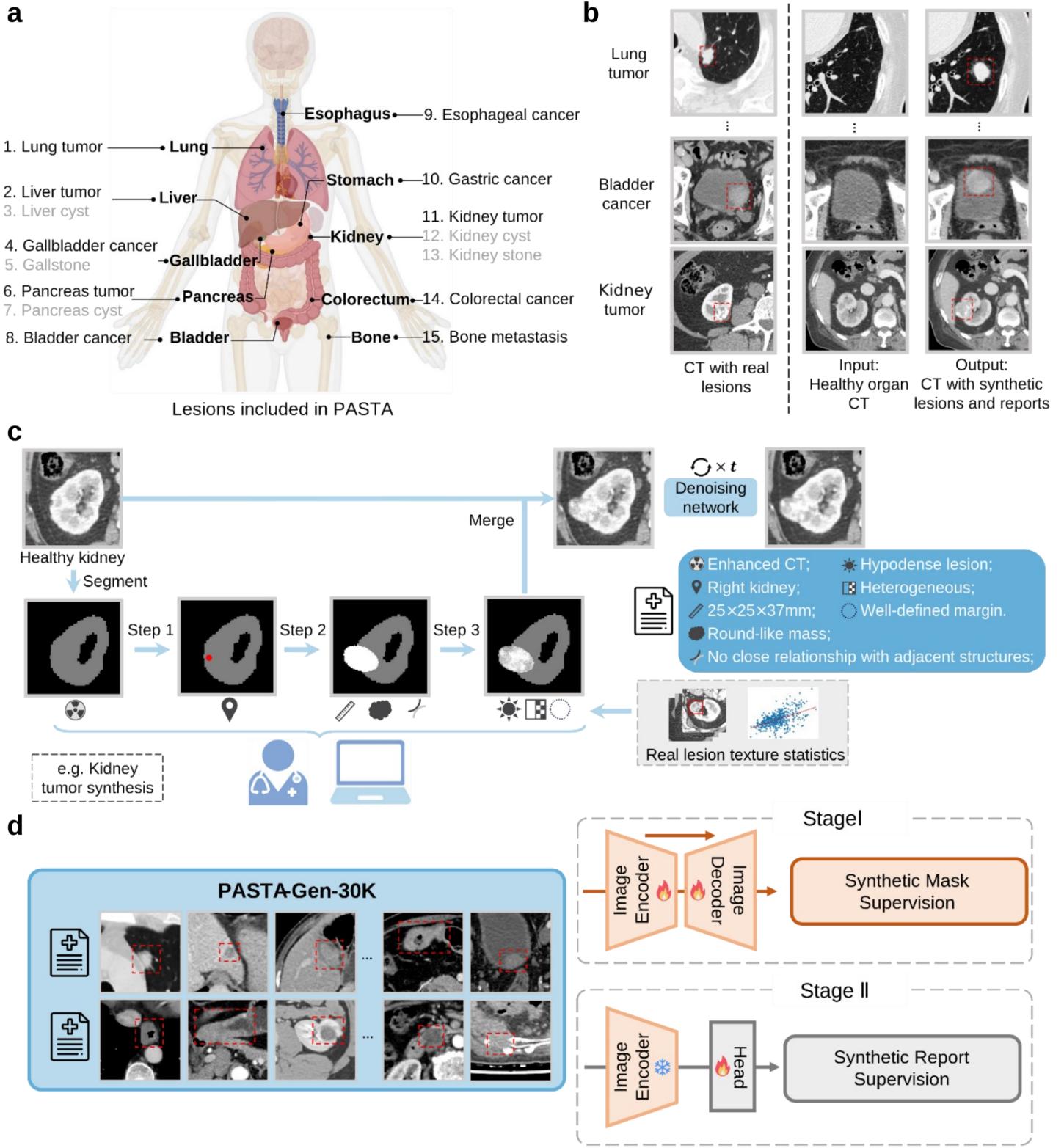


Fig. 1 | Workflow of PASTA Model Development and Training Pipeline. **a**, Overview of organs and lesion types involved in PASTA training. **b**, Examples of lesions generated by PASTA-Gen from healthy organs. **c**, Lesion generation process pipeline of PASTA-Gen. **d**, Two-stage training of PASTA using the PASTA-Gen-30K dataset.

Results

1. Dataset Generation and Model Training

Through multiple iterative sessions with both radiologists and computer scientists, we developed a CT lesion generative model named PASTA-Gen, which has the capability to simulate various lesions on healthy organs in CT scans. We then assembled the PASTA-Gen-30K dataset and employed a supervised, two-stage pretraining strategy to obtain the PASTA foundation model (Fig. 1).

1.1 Construction of PASTA-Gen and PASTA-Gen-30K

To ensure that the synthesized lesions closely match clinical realities, we collected a large set of in-house CT scans paired with radiology reports. Two senior radiologists thoroughly analyzed the lesion descriptions and identified eight key attributes (e.g., shape, density, density variation) that define solid space-occupying lesions across multiple organs. Building on these findings, we systematically categorized the otherwise varied textual attributes into several structured subcategories, enabling a unified, cross-organ description (Methods and Extended Data Table 1). These structured attributes served as the basis for a universal framework that guided lesion simulation in PASTA-Gen.

During synthesis, we used organ segmentation masks for precise lesion localization. We then applied specific graphical operations to capture the full spectrum of lesion attributes and used a denoising network to enhance image realism (Fig. 1c and Extended Data Fig. 1). Since PASTA-Gen simulates lesions only in corresponding healthy organs, we first gathered 10,767 in-house CT scans along with their radiology reports. From these, we selected subsets containing specific healthy organs for the synthesis task (Extended Data Tables 2,3 and Methods). Based on the PASTA-Gen generative model, we created the PASTA-Gen-30K dataset comprising 30,000 image-mask-text pairs. These include 10 malignancies and 5 benign lesion types, each with 2,000 cases. The dataset provides both lesion and organ masks for each sample and is publicly available. Since it does not contain any real patient data, there are no privacy concerns. Representative samples are shown in Extended Data Fig. 2.

1.2 Construction of PASTA

Using the PASTA-Gen-30K dataset, we trained the PASTA foundation model via a two-stage process focused on both semantic segmentation and text-image alignment. PASTA learns to segment the synthetic lesions from target organs in the first stage and classify their attributes in the second stage. In contrast to previous imaging foundation models, PASTA leverages a large volume of high-quality synthetic data, overcoming the limitations imposed by data scarcity. Moreover, because tumor lesions constitute only a small fraction of the total CT volume and given the availability of pixel-level lesion masks, PASTA learns the nuanced relationships and subtle differences among lesions, their surrounding tissues, and various tumor types during pre-training. On top of this, PASTA undergoes further training to align lesion-related visual and textual information, facilitating a wide array of downstream tasks (Fig. 1d).

2. Model Performance and Downstream Task Evaluation

We first collected various publicly available tumor imaging datasets with comprehensive lesion annotations. For diseases lacking public datasets necessary for subsequent evaluations of PASTA-Gen and PASTA, we conducted data collection and annotation at the First Hospital of China Medical University (Extended Data Table 4). The authenticity of images generated by PASTA-Gen and the accuracy of the corresponding textual reports were assessed through expert blind evaluations. Subsequently, the PASTA foundation model was systematically evaluated across multiple downstream oncological tasks.

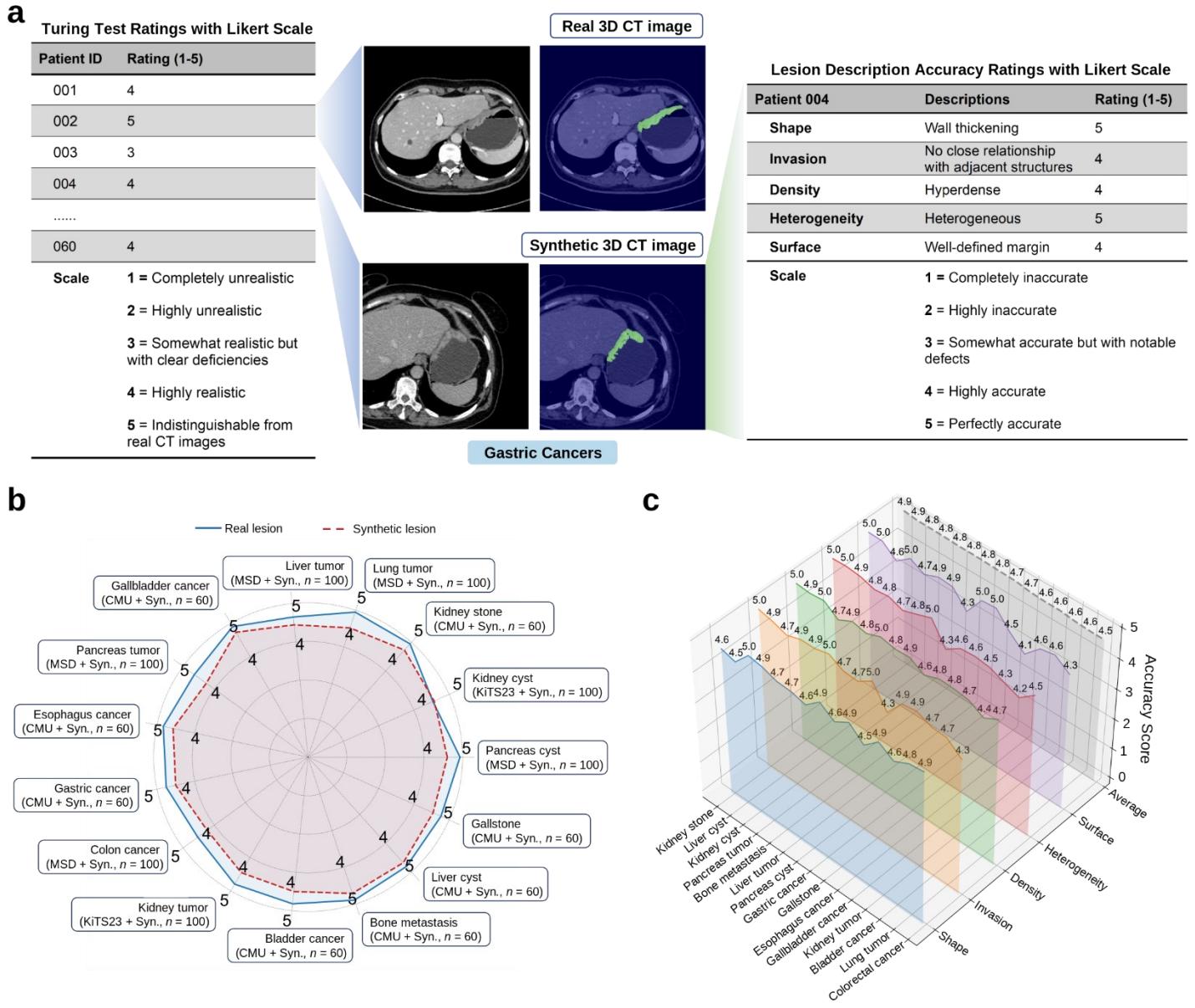


Fig. 2 | Generated Image and Description Accuracy Evaluation for PASTA-Gen. **a**, Illustration of Turing test ratings with Likert scale and lesion description accuracy ratings with Likert scale. **b**, Average Turing test scores of real and synthetic lesions. **c**, Average accuracy scores of structured report descriptions for each lesion type.

2.1 Generated Image and Description Accuracy Evaluation for PASTA-Gen

A systematic evaluation was conducted by 4 radiologists to assess both the realism of images generated by PASTA-Gen and the accuracy of the corresponding structured text descriptions (Fig. 2a).

For image realism, real and generated data for each lesion type were mixed in a 1:1 ratio (Methods and Extended Data Table 5), resulting in a combined dataset of 1,180 cases. All four radiologists performed a blind Turing test, rating each sample on a scale of 1 (completely unrealistic) to 5 (indistinguishable from real CT images). The

results showed that radiologists assigned PASTA-Gen's simulated data an average score ranging from 4.23 to 4.73 (Fig. 2b), with most scores exceeding 4. This indicates that the majority of data simulated by PASTA-Gen consistently met the "highly realistic" rating standard. Notably, for kidney cysts, the simulated data achieved an average score of 4.57, slightly surpassing the real data score of 4.54.

For the lesion descriptions, PASTA-Gen generated 50 images with corresponding reports for each of the 15 lesion types, resulting in a total of 750 image-mask-text pairs. The four radiologists evaluated the alignment between descriptions and five key attributes—shape, density, density variations, surface characteristics, and adjacency to surrounding organs—using a scale from 1 (completely inaccurate) to 5 (perfectly accurate). Across all lesion types, the average consistency scores ranged from 4.54 to 4.93. Cysts and calcifications exhibited exceptionally high text-image consistency, while tumor-like lesions also demonstrated a match level exceeding the "highly accurate" standard (Fig. 2c).

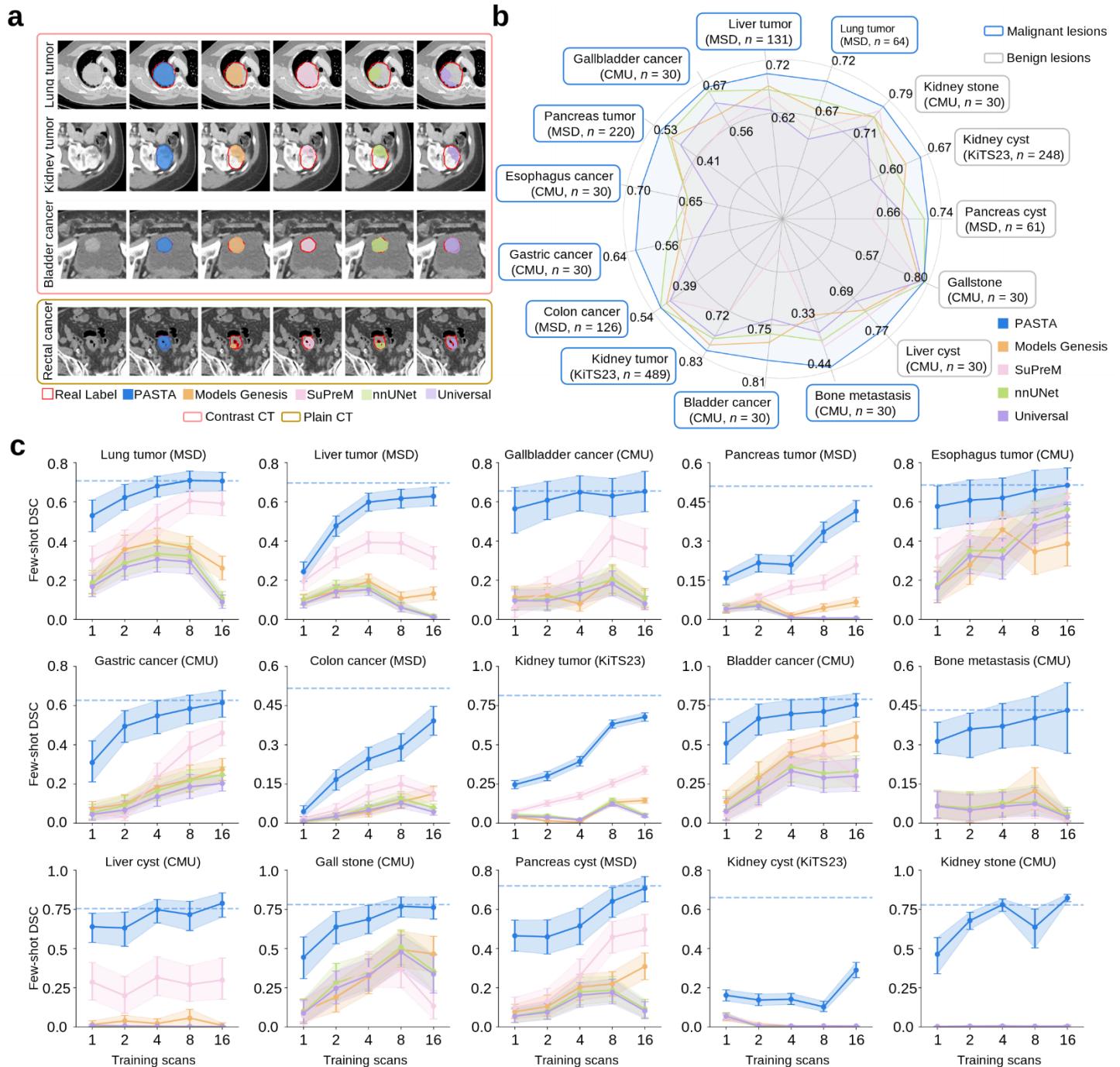


Fig. 3 | Comparison on Lesion Segmentation. **a**, Example images of lesion segmentation results from various models. **b**, Comparison of model performance in lesion segmentation with sufficient data, measured by Dice Similarity Coefficients (DSC). **c**, Lesion segmentation performance of models under few-shot settings, with blue dashed lines indicating PASTA’s full-data training results. Error bands denote 95% confidence intervals.

2.2 Lesion Segmentation with Sufficient Data

We first evaluated PASTA’s lesion segmentation performance under conditions of abundant annotated data and sufficient training. Publicly available datasets with lesion mask annotations were prioritized, including the MSD²⁵

dataset (lung tumor, liver tumor, pancreas tumor and cyst, colon cancer) and the KiTS23²⁶ dataset (kidney tumor and cyst). For lesion types lacking publicly available datasets with annotated masks, we collected and annotated additional data from the First Hospital of CMU, covering gallbladder cancer, esophageal tumor, gastric cancer, bladder cancer, bone metastasis, liver cysts, gallstones, and kidney stones, resulting in a total of 1,535 scans.

To assess PASTA’s performance, we compared it with other pretrained models (Models Genesis²⁰ and SuPerM²²) and established segmentation frameworks (nnUNet²³ and Universal²⁴). Details regarding the segmentation tasks, experimental setup, and performance are provided in the Methods section, Extended Table 4, and Supplementary Tables 1.

Across all 15 lesion segmentation tasks, PASTA consistently outperformed other pretrained models, achieving Dice Similarity Coefficients (DSC) ranging from 0.433 to 0.814 (Fig. 3a, b, and Supplementary Table 1), except for gallstone segmentation, where it performed slightly below nnUNet. In seven tumor segmentation tasks, PASTA significantly surpassed the next best-performing model, with performance improvements as follows: lung tumor (+1.9%, $P = 0.018$), liver tumor (+2.3%, $P < 0.001$), pancreatic tumor (+1.9%, $P = 0.010$), esophageal cancer (+3.7%, $P = 0.004$), gastric cancer (+4.6%, $P = 0.004$), kidney tumor (+1.4%, $P < 0.001$), and bone metastasis (+4.4%, $P = 0.020$). Notably, for tumors that typically exhibit lower segmentation performance, such as gastric cancer and bone metastasis, PASTA demonstrated more pronounced improvements in segmentation accuracy.

2.3 Few-Shot Lesion Segmentation

An essential capability of foundation models is label-efficient learning and effective transfer learning, which are particularly valuable for analyzing rare diseases with limited data availability. For label-efficiency-related tests, we used the same datasets as in the fully supervised lesion segmentation task. In scenarios with limited training data ($K \in \{1, 2, 4, 8, 16\}$) and a small number of training iterations (2,000 iterations), PASTA significantly outperformed all baseline models, achieving segmentation performance improvements ranging from 0.025 to

0.463 (Fig. 3c).

Notably, under extremely limited data conditions ($n \leq 2$), PASTA demonstrated remarkable performance for certain cancers. For gallbladder cancer ($n=2$), PASTA achieved a Dice Similarity Coefficient (DSC) of 0.608, significantly outperforming SuPerM (46.3%, $P < 0.001$), and approaching the performance of models trained on the full dataset (DSC of 0.654). Similarly, for bladder cancer ($n=2$), PASTA achieved a DSC of 0.667, outperforming Models Genesis by +37.7% ($P < 0.001$), and approaching the full dataset performance (DSC of 0.790). These results underscore PASTA's superior ability to generalize from minimal training data, particularly for rare lesion types, where its segmentation accuracy far exceeded the second-best models (Fig. 3c and Supplementary Table 1). These findings highlight PASTA's ability to achieve high segmentation accuracy with minimal training data, making it especially effective for rare lesion types with limited annotated data. By outperforming competing models under constrained conditions, PASTA shows strong potential to enhance radiology workflows, particularly for screening and diagnosing rare diseases.

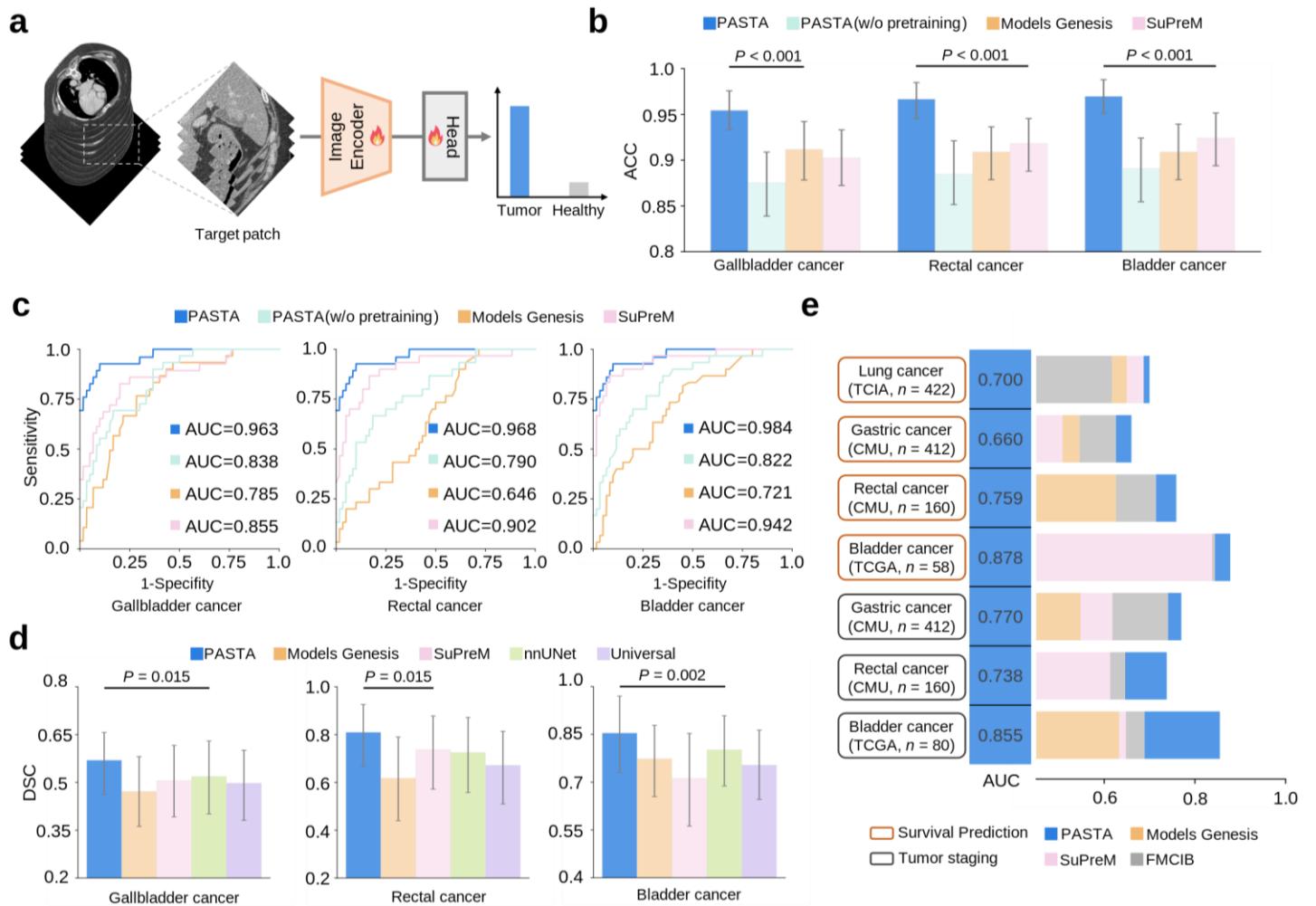


Fig. 4 | Comparison on Downstream Tasks. **a**, Workflow of the classification tasks: target patches are cropped and passed through the encoder, followed by an MLP head to predict class probabilities. For tumor detection in plain CT scans, the target patch corresponds to the organ of interest, while for survival prediction and tumor staging tasks, the target patch is centered around the tumor region. **b**, **c**, Tumor identification performance of accuracy (**b**) and AUC (**c**) of models on Plain-CT data. **d**, Tumor segmentation performance on plain-CT data. **e**, Performance of models in tumor staging and survival prediction across various tumor types, with the top model highlighted by the highest AUC. Bars in **b** and **d** plot displaying 95% confidence intervals as error bands. The listed P-values indicate the significance of PASTA outperforming the best comparison model, determined by a one-sided paired permutation test, utilizing 10,000 permutations.

2.4 Plain-CT Tumor Identification

Tumor assessment typically relies on contrast-enhanced CT scans, but routine screenings in hospitals and health check facilities often use plain CT scans. By pre-training on both contrast-enhanced and plain CT scans across various cancer types, the PASTA foundation model demonstrated its ability to detect tumors that are difficult to identify in plain CT scans. We evaluated PASTA's performance in classification and segmentation tasks on plain

CT data for three cancer types: gallbladder, rectal, and bladder (Fig. 4a, b, c, d, and Supplementary Tables 1, 2, 3).

For the classification task, we collected 30 manually annotated tumor cases for each cancer type, along with 300 healthy CT scans matched at a 1:10 ratio to the tumor cases from the First Hospital of CMU. Multi-layer perceptrons (MLPs) were added after the encoder for classification (Fig. 4a). For segmentation, two board-certified radiologists provided annotations to generate 30 tumor masks for each cancer type. Additional details about the plain CT tumor identification experiments and performance results are provided in the Methods and Supplementary Tables 1, 2, and 3.

In the classification task, PASTA achieved accuracy ranging from 0.954 to 0.970 and AUC values ranging from 0.963 to 0.984 across the three cancer types, outperforming both the PASTA model without pre-training and other pre-trained models, including Models Genesis and SuPerM. PASTA's accuracy and AUC scores surpassed the second-best models by +5.1% ($P<0.001$) and +0.108 ($P=0.031$) for gallbladder cancer, +4.9% ($P<0.001$) and +0.066 ($P=0.031$) for rectal cancer, and +4.4% ($P<0.001$) and +0.042 ($P=0.031$) for bladder cancer. These results highlight PASTA's potential for early cancer detection in routine plain CT screenings, advancing cancer screening and diagnostic workflows.

In the segmentation task, PASTA achieved Dice Similarity Coefficients (DSC) ranging from 0.570 to 0.705, significantly outperforming the next best model. Notable improvements included +5.1% for gallbladder cancer ($P = 0.015$), +3.6% for rectal cancer ($P = 0.015$), and +3.5% for bladder cancer ($P = 0.002$). These results suggest that PASTA's pre-training allowed it to distinguish subtle differences between tumors and normal tissues, enhancing its ability to delineate tumor boundaries in plain CT scans.

2.5 Tumor Staging and Survival Predictions

For frontline clinicians, the ability to perform tumor staging and survival predictions is crucial in clinical applications. We tested PASTA's performance in tumor staging prediction for three types of tumors (gastric

cancer, rectal cancer, and bladder cancer) and survival prediction for four types of tumors (lung cancer, gastric cancer, rectal cancer, and bladder cancer). For the tumor staging prediction task, we compared the models' ability to differentiate between relatively early and late stages of tumors (gastric cancer: stage I-II vs. stage III-IV; rectal cancer: stage I-III vs. stage IV; bladder cancer: stage I-II vs. stage III-IV). For the staging task, we utilized the TCGA-BLCA²⁷ public dataset (bladder cancer) and the CMU dataset (gastric cancer and rectal cancer). In survival prediction, considering the practical aspects of survival data collection, we compared the models' ability to distinguish between short and long tumor-related overall survival (OS) durations (lung tumor: OS < 2 yrs vs. OS \geq 2 yrs; gastric cancer: OS < 2 yrs vs. OS \geq 2 yrs; rectal cancer: OS < 3 yrs vs. OS \geq 3 yrs; bladder cancer: OS < 3 yrs vs. OS \geq 3 yrs). For the survival task, we used the TCIA²⁸ public dataset (lung cancer), TCGA-BLCA²⁷ public dataset (bladder cancer), and the CMU dataset (gastric cancer and rectal cancer). In evaluating model performance, we specifically included the FMCIB²¹ model, a CT pre-trained model that places greater emphasis on clinical phenotype-related biomarkers, for comparison (Methods). The results revealed that PASTA achieved outstanding performance in these tasks (Fig.4e and Supplementary Table 4). In the tumor staging tasks, PASTA's predictions achieved an AUC of 0.770–0.855, with significant improvements over the next best-performing model: +0.029 in gastric cancer ($P = 0.031$), +0.092 in rectal cancer ($P = 0.031$), and +0.166 in bladder cancer ($P = 0.031$). Meanwhile, for the survival prediction tasks, PASTA reached an AUC of 0.660–0.878, outperforming the next best-performing model by +0.034 in gastric cancer ($P = 0.031$) and +0.045 in rectal cancer ($P = 0.094$).

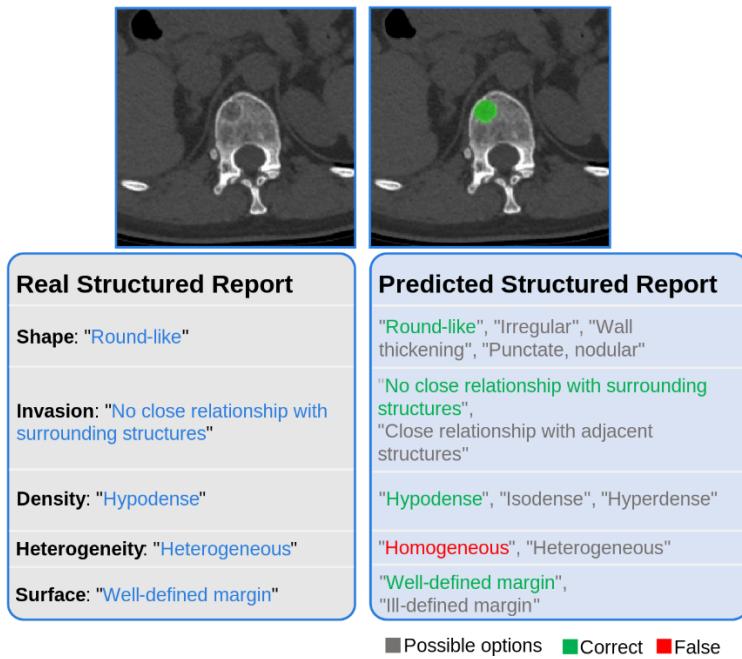
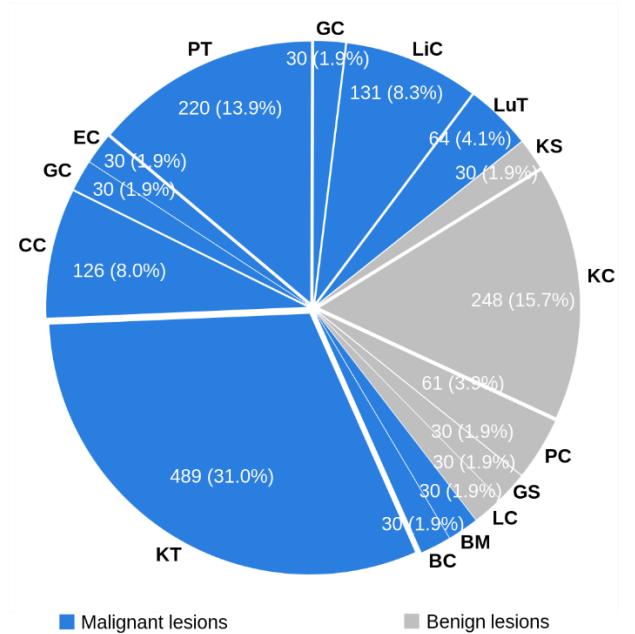
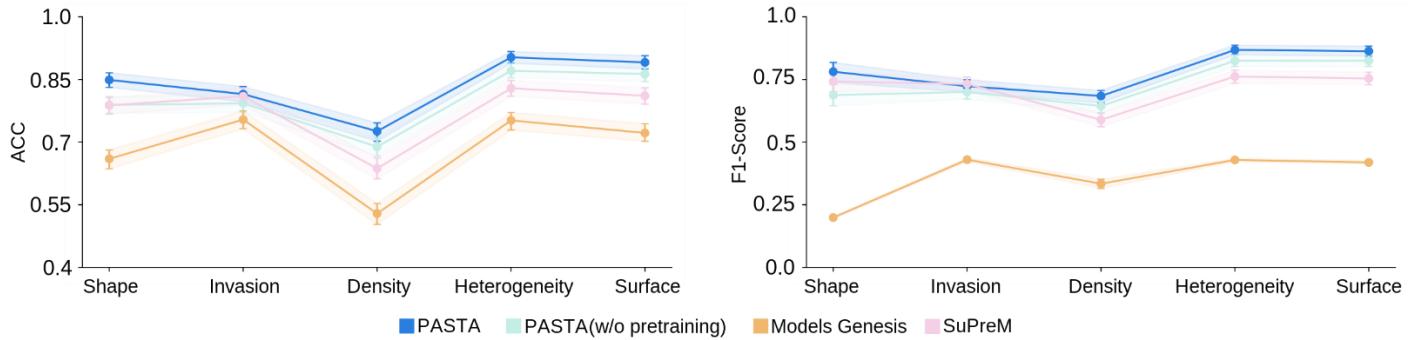
a**b****c**

Fig. 5 | Comparison on Structured Lesion Report Generation. **a**, Example of real and predicted lesion structure reports for bone metastasis generated by PASTA. **b**, Composition of the structured lesion report dataset, including 10 malignant lesion types (LuT: lung tumor, LiC: liver cancer, GC: gallbladder cancer, PT: pancreas tumor, EC: esophagus cancer, GC: gastric cancer, CC: colon cancer, KT: kidney tumor, BC: bladder cancer, and BM: bone metastasis) and 5 benign lesion types (LC: liver cyst, GS: gallstone, PC: pancreas cyst, KC: kidney cyst, and kidney stone). **c**, Comparison of Accuracy (ACC) and F1-scores for five structured report attributes across different models. Error bands denote 95% confidence intervals.

2.6 Structured Lesion Report Generation

Linguistic descriptions of lesion imaging characteristics are vital for understanding and evaluating lesion properties. To assess PASTA's capability in generating structured reports, we followed the structured report format of PASTA-Gen and annotated 1,535 scans corresponding to 15 lesion types (Fig. 5a, Methods, Extended Data Table 1, 4). The structured report generation task was formulated as a multi-class classification problem for

each attribute. For PASTA, the encoder and MLP components were fine-tuned for this task, while for Models Genesis and SuPreM, their encoders were retained and supplemented with an MLP classifier.

Systematic evaluation revealed that PASTA outperformed competing models in generating structured reports across all lesion types. Both F1 scores and accuracy metrics surpassed those of the next best-performing models (Fig. 5c and Supplementary Table 5), except for the invasion attribute, where PASTA's performance was on par with SupreM. Compared to the next best-performing model, PASTA demonstrated significant improvements in both accuracy and F1-score, including +6.0% ($P < 0.001$) and +3.8% ($P = 0.017$) for shape, +3.7% ($P < 0.001$) and 4.0% ($P < 0.001$) for density, +3.2% ($P < 0.001$) and +4.3% for heterogeneity ($P < 0.001$), and +2.8% ($P < 0.001$) and +3.8% for invasion ($P < 0.001$).

2.7 Efficient Cross-Modality Oncology Transfer Learning

Foundation models should excel in rapid transfer learning, allowing efficient adaptation to new imaging modalities. While PASTA was pre-trained on CT data, MRI is a widely used 3D imaging modality with both similarities and complementarities to CT. To evaluate PASTA's label-efficient transfer performance, we tested it on MRI brain tumor and liver tumor datasets (Methods and Extended Table 4). Under limited training data conditions ($K \in \{1, 2, 4, 8, 16\}$) with only 2,000 training iterations, PASTA consistently outperformed other models, demonstrating superior generalization. In the cross-domain MRI brain tumor dataset, which contained head images not included in PASTA's training data, the model achieved a Dice Similarity Coefficient (DSC) of 0.504 at $n=16$, significantly exceeding the second-best model, SuPreM, which scored 0.468 ($P < 0.001$). Similarly, in the cross-domain MRI liver tumor dataset—comprising chest and abdominal images that were present in PASTA's training data—the model attained a DSC of 0.603 at $n=16$, outperforming SuPreM by +13.4% ($P < 0.001$) (Supplementary Table 1).

Overall, these findings highlight PASTA's exceptional capacity for efficient cross-modality transfer learning compared to other CT-pretrained models. We hypothesize that this may reflect PASTA's ability to effectively

differentiate between normal and abnormal tissues across diverse imaging modalities.

Discussion

This study introduces PASTA (Pan-Tumor Analysis with Synthetic Training Data), a novel 3D-CT foundation model designed to advance pan-tumor analysis across multiple cancer types. A significant achievement of this work is the development of PASTA-Gen, a generative model capable of synthesizing realistic lesions across ten organs and fifteen lesion types. Utilizing PASTA-Gen, we assembled the PASTA-Gen-30K dataset, comprising 30,000 synthetic 3D-CT scans with corresponding lesion masks and structured reports. The two-stage pretraining paradigm, focusing on semantic segmentation and vision-language alignment, enabled PASTA to effectively leverage this extensive synthetic dataset, overcoming traditional data scarcity and privacy constraints. Comprehensive evaluations demonstrated that PASTA not only outperforms existing models in lesion segmentation and few-shot learning scenarios but also exhibits robust generalizability across diverse datasets and imaging modalities, including MRI. Additionally, PASTA excels in downstream tasks such as tumor detection in plain CT scans, tumor staging, survival prediction, and structured report generation. These findings underscore PASTA's potential to enhance clinical diagnostics and prognostics, facilitate personalized treatment planning, and support a wide range of oncology research applications.

PASTA distinguishes itself from existing AI-driven imaging models through its comprehensive pan-tumor approach and the utilization of synthetic training data. Traditional models like Models Genesis²⁰, FMCIB²¹, and SuPreM²² focus on specific cancer types or require large annotated datasets, limiting their scalability. While FMCIB distinguishes lesions, it lacks fine-grained annotations. SuPreM suffers from narrow tumor type coverage and imprecise annotations, restricting its generalizability. In contrast, PASTA leverages the PASTA-Gen-30K synthetic dataset to encompass a broader spectrum of tumor types and organ systems, effectively addressing the data scarcity issue inherent in rare cancers. First and foremost, by generating a large volume of publicly available,

precisely controlled image-text pairs, PASTA-Gen itself provides a groundbreaking solution to the chronic shortage of real-world datasets with pixel-level lesion masks and comprehensive imaging reports. Manual annotation is notoriously time-consuming and resource-intensive²⁹, and privacy regulations further restrict the sharing of patient data³⁰. High-fidelity synthetic data not only circumvent these barriers but also open new avenues for researchers and clinicians to explore advanced computational methods and translational applications without compromising patient confidentiality.

Second, large-scale imaging foundation models—particularly those spanning multiple organs and tumor types—are exceedingly rare, in large part due to the scarcity of annotated data. Yet biological commonalities among different malignancies across organ systems suggest that a unified model can yield powerful insights, as evidenced by breakthroughs in pathology^{31,32} and molecular biology research³³. By systematically simulating tumors across ten organs, PASTA is the first to establish a truly pan-tumor 3D imaging foundation model, thus overcoming limitations posed by fragmented real-world datasets. Moreover, while models like nnUNet²³ and Universal²⁴ have set benchmarks in lesion segmentation, PASTA consistently achieves higher DSC across all evaluated tumor types, demonstrating superior segmentation accuracy and reliability.

Lastly, the pan-tumor nature of PASTA translates into tangible benefits for a variety of downstream tasks, including few-shot learning—one of the most pressing challenges in medical imaging³⁴. Many institutions, especially those focusing on specific or rare tumor subtypes, can only gather minimal datasets, limiting their ability to train robust models. PASTA addresses this gap by maintaining high segmentation accuracy with as few as one or two labeled cases, thereby substantially reducing the burden of large-scale data collection and annotation. This advantage is particularly impactful for research centers or clinical settings with constrained resources, where even a handful of reliable training samples can yield meaningful results. Beyond few-shot learning, PASTA excels in structured report generation, staging, survival prediction, and other tasks, consistently outperforming established baselines. Taken together, these capabilities position PASTA as a unified, highly adaptable solution

that not only matches but often exceeds the performance of specialized models—paving the way for broader, more equitable adoption of AI-driven oncological imaging.

Despite these promising outcomes, several limitations merit attention. First, although the synthetic data generation pipeline was carefully validated through radiologist assessments, artificial lesions may still exhibit subtle deviations from real-world complexity. Second, the model’s performance in rare tumor subtypes or atypical lesion presentations requires further validation with clinically acquired data. Lastly, while we integrated multi-phase CT scans from different institutions, additional data from broader patient populations could further enhance the model’s robustness and mitigate potential biases introduced by specific scanners or patient demographics.

In conclusion, PASTA represents a significant advancement in the development of pan-tumor foundation models, effectively addressing key challenges related to data scarcity, privacy, and model generalizability. By leveraging synthetic training data, PASTA not only enhances lesion segmentation and classification performance but also extends its utility to a broad range of downstream clinical tasks. Moving forward, future research should focus on integrating additional imaging modalities, refining synthetic data generation to capture more complex clinical variations, and conducting extensive real-world validations across diverse populations. Moreover, exploring methods to incorporate unstructured clinical data and enhancing the model’s interpretability could further augment its clinical utility. As AI-driven medical imaging continues to evolve, PASTA provides a foundation step toward developing more generalized, data-efficient, and clinically impactful models that can be seamlessly deployed in diverse healthcare settings.

Reference

1. Bray, F., *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* **74**, 229–263 (2024).
2. Kleeff, J. & Ronellenfitsch, U. AI and imaging-based cancer screening: getting ready for prime time. *Nature medicine* **29**, 3002–3003 (2023).
3. Cao, K., *et al.* Large-scale pancreatic cancer detection via non-contrast CT and deep learning. *Nature medicine* **29**, 3033–3043 (2023).
4. Ardila, D., *et al.* End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature medicine* **25**, 954–961 (2019).
5. McKinney, S.M., *et al.* International evaluation of an AI system for breast cancer screening. *Nature* **577**, 89–94 (2020).
6. Preetha, C.J., *et al.* Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study. *The Lancet Digital Health* **3**, e784–e794 (2021).
7. Lipkova, J., *et al.* Artificial intelligence for multimodal data integration in oncology. *Cancer cell* **40**, 1095–1110 (2022).
8. Moor, M., *et al.* Foundation models for generalist medical artificial intelligence. *Nature* **616**, 259–265 (2023).
9. Zhang, S. & Metaxas, D. On the challenges and perspectives of foundation models for medical image analysis. *Medical image analysis* **91**, 102996 (2024).
10. Brown, T., *et al.* Language models are few-shot learners. *Advances in neural information processing systems* **33**, 1877–1901 (2020).
11. Acosta, J.N., Falcone, G.J., Rajpurkar, P. & Topol, E.J. Multimodal biomedical AI. *Nature Medicine* **28**, 1773–1784 (2022).
12. Rajpurkar, P. & Topol, E.J. A clinical certification pathway for generalist medical AI systems. *The Lancet* **405**, 20 (2025).
13. Chen, R.J., *et al.* Towards a general-purpose foundation model for computational pathology. *Nature Medicine* **30**, 850–862 (2024).
14. Wang, X., *et al.* A pathology foundation model for cancer diagnosis and prognosis prediction. *Nature* **634**, 970–978 (2024).
15. Hua, S., Yan, F., Shen, T., Ma, L. & Zhang, X. Pathoduet: Foundation models for pathological slide analysis of H&E and IHC stains. *Medical Image Analysis* **97**, 103289 (2024).
16. Lei, W., Xu, W., Li, K., Zhang, X. & Zhang, S. MedLSAM: Localize and segment anything model for 3D CT images. *Medical Image Analysis* **99**, 103370 (2025).
17. Wang, G., *et al.* Mis-fm: 3d medical image segmentation using foundation models pretrained on a large-scale unannotated dataset. *arXiv preprint arXiv:2306.16925* (2023).
18. Wald, T., *et al.* Revisiting MAE pre-training for 3D medical image segmentation. *arXiv preprint arXiv:2410.23132* (2024).
19. Jiang, Y., *et al.* Anatomical invariance modeling and semantic alignment for self-supervised learning in 3d medical image analysis. in *Proceedings of the IEEE/CVF International Conference on Computer Vision* 15859–15869 (2023).
20. Zhou, Z., Sodha, V., Pang, J., Gotway, M.B. & Liang, J. Models genesis. *Medical image analysis* **67**, 101840 (2021).
21. Pai, S., *et al.* Foundation model for cancer imaging biomarkers. *Nature machine intelligence* **6**, 354–367 (2024).
22. Li, W., Yuille, A. & Zhou, Z. How Well Do Supervised 3D Models Transfer to Medical Imaging Tasks? in *The Twelfth International Conference on Learning Representations*.
23. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J. & Maier-Hein, K.H. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods* **18**, 203–211 (2021).
24. Liu, J., *et al.* Clip-driven universal model for organ segmentation and tumor detection. in *Proceedings of the IEEE/CVF International Conference on Computer Vision* 21152–21164 (2023).
25. Antonelli, M., *et al.* The medical segmentation decathlon. *Nature communications* **13**, 4128 (2022).
26. Heller, N., *et al.* The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging:

- Results of the KiTS19 challenge. *Medical image analysis* **67**, 101821 (2021).
- 27. Kirk, S., *et al.* The Cancer Genome Atlas Urothelial Bladder Carcinoma Collection (TCGA-BLCA)(Version 8)[Data set]. *The Cancer Imaging Archive* **10**, K9 (2016).
 - 28. Aerts, H., *et al.* Data From NSCLC-Radiomics (version 4)[Data set]. *The Cancer Imaging Archive* (2014).
 - 29. Aljabri, M., AlAmir, M., AlGhamdi, M., Abdel-Mottaleb, M. & Collado-Mesa, F. Towards a better understanding of annotation tools for medical imaging: a survey. *Multimedia tools and applications* **81**, 25877-25911 (2022).
 - 30. Thapa, C. & Camtepe, S. Precision health data: Requirements, challenges and existing techniques for data security and privacy. *Computers in biology and medicine* **129**, 104130 (2021).
 - 31. Arslan, S., *et al.* A systematic pan-cancer study on deep learning-based prediction of multi-omic biomarkers from routine pathology images. *Communications Medicine* **4**, 48 (2024).
 - 32. Fu, Y., *et al.* Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nature cancer* **1**, 800-810 (2020).
 - 33. Chen, R.J., *et al.* Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer Cell* **40**, 865-878. e866 (2022).
 - 34. Pachetti, E. & Colantonio, S. A systematic review of few-shot learning in medical imaging. *Artificial intelligence in medicine*, 102949 (2024).

Method

1 Development of PASTA-Gen

PASTA-Gen is a lesion editing model designed to simulate lesions on scans of target organs with healthy anatomy. To develop PASTA-Gen, we first constructed a universal structured lesion report template, followed by the collection and annotation of real lesion data as a reference set. Using this reference data and the structured template, we iteratively refined the model through close collaboration with radiologists, ensuring its clinical relevance and accuracy. Finally, a denoising network was incorporated to enhance the realism of the generated images (Figure 1. C).

1.1 Structured Lesion Report Template

Building on previous studies on structured radiology reporting³⁵, we synthesized insights from existing research and analyzed a large collection of in-house radiology reports across various disease types. While no universal standard currently exists, radiologists consolidated lesion descriptions for solid tumors into eight key attributes, including enhancement status, location, size, shape, density, heterogeneity, surface, and invasion. These attributes form a comprehensive and objective template for characterizing lesions, ensuring consistency and enabling generalization across diverse lesion types (Extended Data Table 1).

One of these attributes, "density," typically relies on assessments from multiple CT modalities. For instance, in contrast-enhanced CT scans, radiologists often integrate information from different imaging phases to provide a nuanced evaluation of a lesion's density. However, as our study focuses on developing a cross-phase CT analysis model, the concept of "density" in PASTA-Gen refers to a relative assessment within a single-phase CT image. Specifically, it captures the density of the lesion area relative to the surrounding normal tissues in the given phase. This approach enhances the model's adaptability in representing lesions with complex or variable densities.

1.2 Real Lesion Reference Data

To ensure comprehensive modeling and evaluation of synthetic lesion data, we developed a diverse reference dataset encompassing 15 target lesion types across 10 organs. This dataset includes lesion segmentations and structured reports derived from both public sources and in-house data from the First Hospital of CMU (Extended Data Table 4). For instances with incomplete labels, two senior radiologists meticulously supplemented the missing information, ensuring the dataset's accuracy and clinical relevance. We reformat all CT scans so that the first axis points from right to left, the second from anterior to posterior, and the third from inferior to superior. We then resample the spacing in all directions to 1 mm using linear spline interpolation.

1.3 Modeling Process for PASTA-Gen

The development of PASTA-Gen involved a multidimensional statistical analysis of collected real reference lesions, combined with iterative feedback from experienced radiologists. This collaborative approach enabled us to systematically model the eight structured attributes outlined in our framework.

Lesion size distributions were statistically represented using a log-normal distribution across various axes, with random sampling used to generate realistic dimensions. Lesion shapes were modeled based on the specific morphological characteristics of each type, such as the round-like shape of liver cysts or the wall thickening observed in gastric cancer. These shape parameters ensure the model captures the clinical nuances of different lesions. Density attributes were linearly modeled with controlled variability, determined by calculating the intensity of normal tissues surrounding the lesion on target organ scans. The lesion density was then derived by sampling from these calculated differences, ensuring alignment with clinical observations. Heterogeneity attributes, including brightness deviation and texture features, were modeled based on radiologist input. For instance, heterogeneous density patterns were simulated to reflect clinical variability. For solid tumors in

parenchymal organs like liver and kidney tumors, we emphasized the boundary between the tumor and surrounding tissues, controlling this characteristic by adjusting the degree of Gaussian blur applied to the tumor edges. For hollow organs such as those in the gastrointestinal or urinary systems, surface features were modeled to highlight serosal roughness and outer surface irregularities. Invasion was modeled based on whether the lesion extended beyond the target organ into adjacent tissues. For benign lesions, invasion into surrounding organs was explicitly avoided to maintain consistency with clinical presentations. Each simulated lesion is accompanied by a corresponding structured report containing eight attributes, with each attribute reflecting one of its possible values (Extended Data Table 1). This comprehensive approach ensures that PASTA-Gen generates synthetic lesions with high fidelity and clinical relevance, capable of capturing the diverse characteristics of real-world lesions.

1.4 Collection and Preprocessing of Template CT Scan and Radiology Report

To ensure the diversity and scalability of PASTA-Gen, we collected an additional 10,767 in-house CT scans to serve as the template scan set, which include both contrast-enhanced and plain phases, covering various parts of the body, along with their corresponding radiology reports (Extended Data Table 2). We reformat all CT scans so that the first axis points from right to left, the second from anterior to posterior, and the third from inferior to superior. We then resample the spacing in all directions to 1 mm using linear spline interpolation. We use TotalSegmentator³⁶ to segment the standardized scans. This tool segments 104 classes of organs, encompassing all the organs for which PASTA-Gen simulates lesions.

For the template CT scans. Based on the presence of specific organs, we only retain the following three study types of CT scans: (1) Chorax CT: must include the T1 vertebra, left upper lobe of the lung, and right upper lobe of the lung, but not include the L5 vertebra. (2) Abdomen-pelvis CT: must include the T8 vertebra and the bladder but not include the T1 vertebra. (3) Thorax-abdomen-pelvis CT: must include the T1 vertebra, left upper lobe of

the lung, and right upper lobe of the lung and the bladder (Extended Data Table 3). Next, we determine whether each selected CT scan is contrast-enhanced by evaluating the Hounsfield unit (HU) values of the aorta and inferior vena cava. If the average HU of both the aorta and inferior vena cava is less than 80, the scan is classified as non-contrast. Otherwise, it is classified as contrast-enhanced. Based on the diagnostic modalities typically used by radiologists for different types of lesions, we selected the appropriate scanning regions and modalities for simulating each type of lesion (Supplementary Table 6).

To ensure that PASTA-Gen simulates target lesions on healthy organs, we implemented a filtering process based on the radiology report associated with each CT scan. For the target organ in each CT scan, the criteria for being classified as healthy are as follows: (1) the volume of the organ is larger than 4000 mm^3 , and (2) the corresponding radiology report impression does not mention the organ, as determined by keyword search. By applying these criteria, we ensure that only healthy organs are used for simulating lesions (Extended Data Table 4).

1.5 Details of Denoising Network

The denoising network in PASTA-Gen aims to optimize the initially synthesized lesions. This network is based on a 3D medical denoising diffusion probabilistic model (DDPM)^{37,38}, which involves adding random Gaussian noise to the original image and learning how to remove it. We leverage this property to eliminate unnatural parts of the simulated lesions. During the training phase, the input image size is $128 \times 128 \times 128$. We concatenate the corresponding organ segmentation mask as the input condition for the model (Figure 1, Extended Figure 1). We set the total timesteps T to 1000 and utilize the L_1 loss to measure the difference between predicted and added noise. The model channel number is set to 16. We use an AdamW optimizer³⁹ with an initial learning rate of 1×10^{-4} and weight decay of 1×10^{-5} . We employ a total batch size of 8 across 4 NVIDIA GTX4090 GPUs. During the lesion generation stage, the initially simulated lesions are input into the denoising network for $t = 5$ optimization iterations. A sliding window approach is applied to ensure comprehensive and seamless optimization

across the entire image volume.

2 Dataset Curation for PASTA-Gen-30K

Using PASTA-Gen, we simulated 2,000 instances for each of the 15 target lesion types, resulting in a dataset of 30,000 lesion image-mask-text pairs. Since the template scan set includes organ segmentation results generated by TotalSegmentator, the labels for the simulated images retain all 10 target organ labels alongside the lesion labels, totaling 25 categories (Supplementary Table 7).

3 Details of PASTA Pretraining

The PASTA model leverages a 3D UNet⁴⁰ architecture for its encoder and decoder, with a three-layer MLP serving as the classification head. The encoder and decoder of the 3D UNet architecture each consist of five convolutional blocks. Each convolutional block is composed of two sub-blocks, and within each sub-block, there is a 3D convolution operation with a kernel size of $3 \times 3 \times 3$ and a stride of 1 along each dimension. Each convolution is followed by a 3D Instance Normalization^{41,42} activation function. In the encoder, each convolutional block is followed by a max pooling operation with a kernel size of $2 \times 2 \times 2$, while in the decoder, each block is followed by a transposed convolution operation with a stride of 2, which up-samples the spatial resolution. The MLP classification head consists of an input layer with 1024 channels, followed by two hidden layers with channel sizes of 512 and 256, and a final output layer mapping to the number of classes. Each intermediate layer is followed by a LeakyReLU activation function.

During the initial segmentation pretraining stage, we employed the nnUNet framework²³, which has demonstrated state-of-the-art performance across various biomedical image segmentation tasks^{43,44}. The training target is to segment all 25 categories in PASTA-Gen-30K. The training process used an initial learning rate of 1×10^{-2} and was conducted over 2,000 epochs, with each epoch comprising 250 iterations. Random cropping was applied to

extract 3D patches of size $128 \times 128 \times 128$ voxels for training. The loss function combined Cross-Entropy Loss and Dice Loss⁴⁵, optimizing for both pixel-wise classification accuracy and volumetric overlap. Training was conducted with a total batch size of 16, distributed across 8 NVIDIA GTX 4090 GPUs.

In the second stage of lesion attribute classification pretraining, the 3D UNet component is frozen, and only the MLP connected to the encoder is trained. The training objective is to predict lesion attributes from the structured reports in the PASTA-Gen-30K dataset, including Shape, Density, Heterogeneity, Surface, and Invasion attributes. Each attribute comprises fixed categories: Shape (4 classes), Density (3 classes), and Heterogeneity, Surface, and Invasion (2 classes each). The task is formulated as a multi-class classification problem. The initial learning rate is set to 1×10^{-3} , and the training runs for a total of 100,000 iterations. The loss function used is Cross-Entropy Loss. During training, patches of size $96 \times 96 \times 96$ voxels, centered on the lesion, are cropped and fed into the network to focus on the lesion region. Training was conducted with a total batch size of 64, distributed across 4 NVIDIA GTX 4090 GPUs.

4 Competing Methods and Baselines

We compare PASTA to 5 comparison approaches. Models Genesis²⁰ was a released model pretrained on 623 Chest CT scans in LUNA 2016⁴⁶. It utilized a 3D UNet architecture and employed a self-supervised learning approach by recovering the original sub-volumes of images from their transformed versions. SuPreM²² was pretrained in a supervised segmentation setting on the AbdomenAtlas 1.1 dataset⁴⁷ which comprises 9,262 CT volumes with detailed annotations of 25 anatomical structures and pseudo annotations for seven tumor types. SuPreM incorporates multiple backbone architectures, and we selected the 3D UNet model for evaluation as it demonstrated the best performance in SuPreM experimental results. FMCIB²¹ used a Resnet⁴⁸ as the backbone and was pretrained with a comprehensive dataset of 11,467 radiographic lesions in DeepLesion⁴⁹. It is tailored for cancer imaging biomarker discovery by contrasting volumes with and without lesions. nnUNet²³ and

Universal²⁴ are both state-of-the-art frameworks for biomedical image segmentation.

5 Lesion Segmentation with sufficient data

From the lesion reference dataset, we constructed 15 lesion segmentation tasks (Extended Data Table 4), including MSD²⁵ dataset (lung tumor, liver tumor, pancreas tumor and cyst, colon cancer) and the KiTS23²⁶ dataset (kidney tumor and cyst), as well as private scans from the First Hospital of CMU, covering liver cysts, gallbladder cancer, gallstones, esophageal cancer, gastric cancer, kidney stones, bladder cancer, and bone metastases. For the MSD Pancreas dataset, the original annotations for pancreatic cancer and pancreatic cysts shared the same label value. To address this, a senior radiologist manually differentiated and re-annotated them. For PASTA, Models Genesis, and SuPreM, we fine-tuned their encoder-decoder part using the nnUNet training framework with a base learning rate of 1×10^{-3} for 500 epochs, with each epoch comprising 250 iterations. Following the default nnUNet settings, the baseline nnUNet model was trained with a learning rate of 1×10^{-3} for 1,000 epochs. For the Universal model, we adhered to its standard configuration, training it with a learning rate of 1×10^{-4} for 2,000 epochs. Training random-crop patch-size is $128 \times 128 \times 128$ voxel. All experiments were conducted on a single NVIDIA GTX 4090 GPU with a batch size of 2. The performances were evaluated in terms of the DSC using the 5-fold cross-validation.

6 Few-shot Lesion Segmentation

Few-shot learning is a common approach for evaluating the quality of features extracted by a pre-trained model and its ability to quickly adapt to new tasks with limited annotated data and transferring iterations. In this study, the few-shot lesion segmentation experiment was designed using the same 15 segmentation tasks and datasets as described in the fully supervised lesion segmentation section (Extended Data Table 4). To simulate few-shot settings, the number of labeled training scans per task was limited to $K \in \{1, 2, 4, 8, 16\}$, and the total training

iterations for all models were capped at 2,000. For PASTA, Models Genesis, and SuPreM, we fine-tuned their UNet part using the nnUNet training framework with a base learning rate of 1×10^{-4} . The baseline nnUNet model was trained with a learning rate of 1×10^{-3} . For the Universal model, the learning rate is set to 1×10^{-4} . Training random-crop patch-size is $128 \times 128 \times 128$ voxel. All experiments were conducted on a single NVIDIA GTX 4090 GPU with a batch size of 2. The same 5-fold cross-validation splits used in the fully supervised setting were adopted, with K training samples randomly selected for each fold, repeated over five runs to ensure robustness. For the KiTS23 dataset, training samples were specifically chosen to ensure that each selected scan included both kidney tumors and kidney cysts.

7 Plain-CT Tumor Detection

Detecting tumors in plain CT scans demonstrates the ability of models to identify subtle lesions, showcasing their potential for clinical applications. This study includes 3 tumor classification tasks and 3 tumor segmentation tasks, focusing on gallbladder cancer, rectal cancer, and bladder cancer. A total of 30 annotated non-contrast CT scans for each cancer type were collected from the First Hospital of CMU, with tumor regions delineated by an experienced radiologist. Additionally, 300 healthy scans corresponding to each organ were extracted from the template CT dataset to serve as controls. All CT scans are reoriented such that the first axis runs from right to left, the second from anterior to posterior, and the third from inferior to superior. The voxel spacing is then uniformly resampled to 1 mm in all directions using linear spline interpolation. For the tumor classification tasks, a conventional two-stage detection paradigm was adopted. This involved pre-extraction of target organs followed by classification of the extracted ROI volumes. TotalSegmentator was applied to segment the target organs in each scan, and the segmented regions were cropped with a 24-voxel margin in all directions. The cropped volumes were resized and padded to a uniform size of $128 \times 128 \times 128$ voxels. The PASTA model retained its encoder and MLP components, with the MLP’s final layer adjusted to output two channels (non-cancer and cancer). Models

Genesis and SuPreM retained their encoders, with a three-layer MLP appended, comprising two hidden layers (channel sizes of 512 and 256) and a final output layer with two channels; each intermediate layer was followed by a LeakyReLU activation function. The training process utilized the Adam optimizer with a total batch size of 32. A weighted cross-entropy loss function addressed class imbalance by assigning a 10:1 weight ratio to non-cancer and cancer scans. Performance evaluation was conducted using the Area Under the Curve (AUC) metric with 5-fold cross-validation. For the tumor segmentation tasks, the UNet components of PASTA, Models Genesis, and SuPreM were fine-tuned using the nnUNet training framework. The training was conducted with a base learning rate of 1×10^{-4} over 500 epochs, with each epoch consisting of 250 iterations. The baseline nnUNet model was trained with a learning rate of 1×10^{-4} for 1,000 epochs, while the Universal model followed its default configuration, using a learning rate of 1×10^{-4} for 2,000 epochs. Training utilized random-crop patches of size $128 \times 128 \times 128$ voxels. All experiments were performed on a single NVIDIA GTX 4090 GPU with a batch size of 2. Performance was assessed using the Dice Similarity Coefficient (DSC) metric with 5-fold cross-validation to ensure robustness.

8 Tumor Staging and Survival Predictions

This study addresses seven tasks, including four cancer survival prediction tasks and three cancer staging tasks. The Lung1 dataset⁵¹ consists of 422 pretreatment CT scans from non-small cell lung cancer patients, annotated with primary gross tumor volumes and overall survival information. Patients were classified into two categories: OS less than two years and OS of two years or more. The CMU gastric cancer dataset comprises 412 contrast-enhanced pretreatment abdominal CT scans from gastric cancer patients, annotated with tumor center points, survival data, and overall staging. The survival prediction task divides patients into OS less than two years and OS of two years or more, while staging distinguishes between stages I-II and III-IV. Similarly, the CMU rectal cancer dataset includes 160 contrast-enhanced pretreatment abdominal CT scans annotated with tumor center

points, survival information, and overall staging, with tasks involving OS less than three years and OS of three years or more, as well as staging differentiation between stages I-III and stage IV. The TCGA-BLCA²⁷ dataset features 120 bladder cancer patients with imaging data and survival information, complemented by radiologist-annotated tumor center points. Survival prediction categorizes patients into OS less than three years and OS of three years or more, while staging separates stages I-II from III-IV. All CT scans are reformatted with the first axis oriented from right to left, the second from anterior to posterior, and the third from inferior to superior. The spacing in all directions is then resampled to 1 mm using linear spline interpolation. For all tasks, the FMCIB pipeline was adopted, which involves feature extraction from ROI regions centered on annotated tumor points for classification. Fine-tuning was performed on PASTA, Models Genesis, SuPreM, and FMCIB models, each using a base learning rate of 1×10^{-4} . For Models Genesis, SuPreM, and FMCIB models, encoders were retained, with a three-layer MLP appended for attribute prediction. The training process involved Cross-Entropy Loss and the Adam optimizer with a total batch size of 32. Performance was assessed using the Area Under the Curve (AUC) metric and 5-fold cross-validation to ensure robust evaluation.

9 Tumor Structure Report Generation

Evaluation of structured lesion reports utilized a reference dataset covering five attributes: shape, density, heterogeneity, surface, and invasion. The prediction tasks were framed as multi-label classification problems, requiring the model to predict an option for each attribute. CT scans are standardized by orienting the first axis from right to left, the second from anterior to posterior, and the third from inferior to superior. The voxel spacing is subsequently resampled to 1 mm in all directions using linear spline interpolation. For all tasks, regions of size $96 \times 96 \times 96$ voxels, centered on the lesion, are extracted and input into the network for classification. Fine-tuning was performed on PASTA, Models Genesis, and SuPreM models, each using a base learning rate of 1×10^{-4} . For Models Genesis and SuPreM, encoders were retained, with a three-layer MLP appended for attribute prediction.

The training process involved Cross-Entropy Loss and the Adam optimizer with a total batch size of 32. Performance was assessed using the Area Under the Curve (AUC) metric and 5-fold cross-validation to ensure robust evaluation.

10 Efficient Cross-Modality Oncology Transfer Learning

A cross-modality evaluation was performed on two cancer segmentation tasks: (1) the MSD-Brain Tumors segmentation task²⁵, comprising 484 patients with FLAIR, T1w, T1gd, and T2w MRI scans. The segmentation targets included necrotic/active tumor regions and edema in gliomas. (2) The ATLAS⁵⁰ dataset, which includes 60 contrast-enhanced T1-weighted MRI scans of the liver from 60 patients with unresectable hepatocellular carcinoma (HCC), along with segmentation masks for the liver and liver tumors. In a few-shot learning setup, the number of labeled training scans was limited to $K \in \{1, 2, 4, 8, 16\}$, and total training iterations were capped at 2,000. PASTA, Models Genesis, and SuPreM models were fine-tuned using their UNet components within the nnUNet training framework, employing a base learning rate of 1×10^{-4} . The nnUNet baseline model was trained with a learning rate of 1×10^{-3} , while the Universal model used a learning rate of 1×10^{-4} . The training employed random-cropped patches of $128 \times 128 \times 128$ voxels. All experiments were performed on a single NVIDIA GTX 4090 GPU with a batch size of 2. The 5-fold cross-validation splits from the fully supervised setting were applied. The 5-fold cross-validation splits from the fully supervised setting were applied. For each fold, K training samples were randomly selected from the training set, and this process was repeated across five runs.

Data Availability

To advance oncology research, we release the PASTA-Gen-30K dataset, which consists of 30,000 synthetic 3D-CT scans with corresponding lesion masks and structured report descriptions at <https://huggingface.co/datasets/LWHYC/PASTA-Gen-30K>. The publicly available datasets used in this study can

be accessed through the following sources: MSD (<http://medicaldecathlon.com/>); KiTS23 (<https://kits-challenge.org/kits23/>); Lung1 (<https://www.cancerimagingarchive.net/collection/nsclc-radiomics/>); TCGA-BLCA (<https://www.cancerimagingarchive.net/collection/tcga-blca/>); ATLAS (<https://atlas-challenge.u-bourgogne.fr/>). Structured report annotations for publicly available datasets, as well as lesion center point annotations for Lung1 and TCGA-BLCA, can be accessed at <https://github.com/LWHYC/PASTA>. Additional imaging data and associated clinical records were obtained from the First Hospital of China Medical University. Due to privacy regulations, these data cannot be publicly shared.

Code Availability

The PASTA model, along with the full training and evaluation pipeline, is publicly available to facilitate further research and replication of our findings. The source code, model weights, and implementation details can be accessed at <https://github.com/LWHYC/PASTA>, ensuring accessibility for the broader clinical and scientific community.

Methods References

35. Nobel, J.M., van Geel, K. & Robben, S.G. Structured reporting in radiology: a systematic review to explore its potential. *European radiology*, 1-18 (2022).
36. Wasserthal, J., et al. TotalSegmentator: robust segmentation of 104 anatomic structures in CT images. *Radiology: Artificial Intelligence* **5**(2023).
37. Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. *Advances in neural information processing systems* **33**, 6840-6851 (2020).
38. Yu, Y., et al. CT Synthesis with Conditional Diffusion Models for Abdominal Lymph Node Segmentation. *arXiv preprint arXiv:2403.17770* (2024).
39. Loshchilov, I. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017).
40. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T. & Ronneberger, O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II* 19 424-432 (Springer, 2016).
41. Ulyanov, D. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022* (2016).
42. Xu, B. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853* (2015).
43. Tu, T., et al. Towards generalist biomedical AI. *NEJM A/1*, Aloa2300138 (2024).
44. Ma, J., et al. Unleashing the strengths of unlabelled data in deep learning-assisted pan-cancer abdominal organ quantification: the FLARE22 challenge. *The Lancet Digital Health* **6**, e815-e826 (2024).
45. Milletari, F., Navab, N. & Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. in *2016 fourth international conference on 3D vision (3DV)* 565-571 (Ieee, 2016).
46. Setio, A.A.A., et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *Medical image analysis* **42**, 1-13 (2017).
47. Li, W., et al. Abdomenatlas: A large-scale, detailed-annotated, & multi-center dataset for efficient transfer learning and open algorithmic benchmarking. *Medical Image Analysis* **97**, 103285 (2024).
48. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. in *Proceedings of the IEEE conference on computer vision and pattern recognition* 770-778 (2016).
49. Yan, K., Wang, X., Lu, L. & Summers, R.M. DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *Journal of medical imaging* **5**, 036501-036501 (2018).
50. Quinton, F., et al. A tumour and liver automatic segmentation (atlas) dataset on contrast-enhanced magnetic resonance imaging for hepatocellular carcinoma. *Data* **8**, 79 (2023).
51. Aerts, H.J.W.L., Wee, L., Rios Velazquez, E., Leijenaar, R. T. H., Parmar, C., Grossmann, P., Carvalho, S., Bussink, J., Monshouwer, R., Haibe-Kains, B., Rietveld, D., Hoebers, F., Rietbergen, M. M., Leemans, C. R., Dekker, A., Quackenbush, J., Gillies, R. J., & Lambin, P. Data From NSCLC-Radiomics (Version 4) [Data set]. *The Cancer Imaging Archive*. (2019).

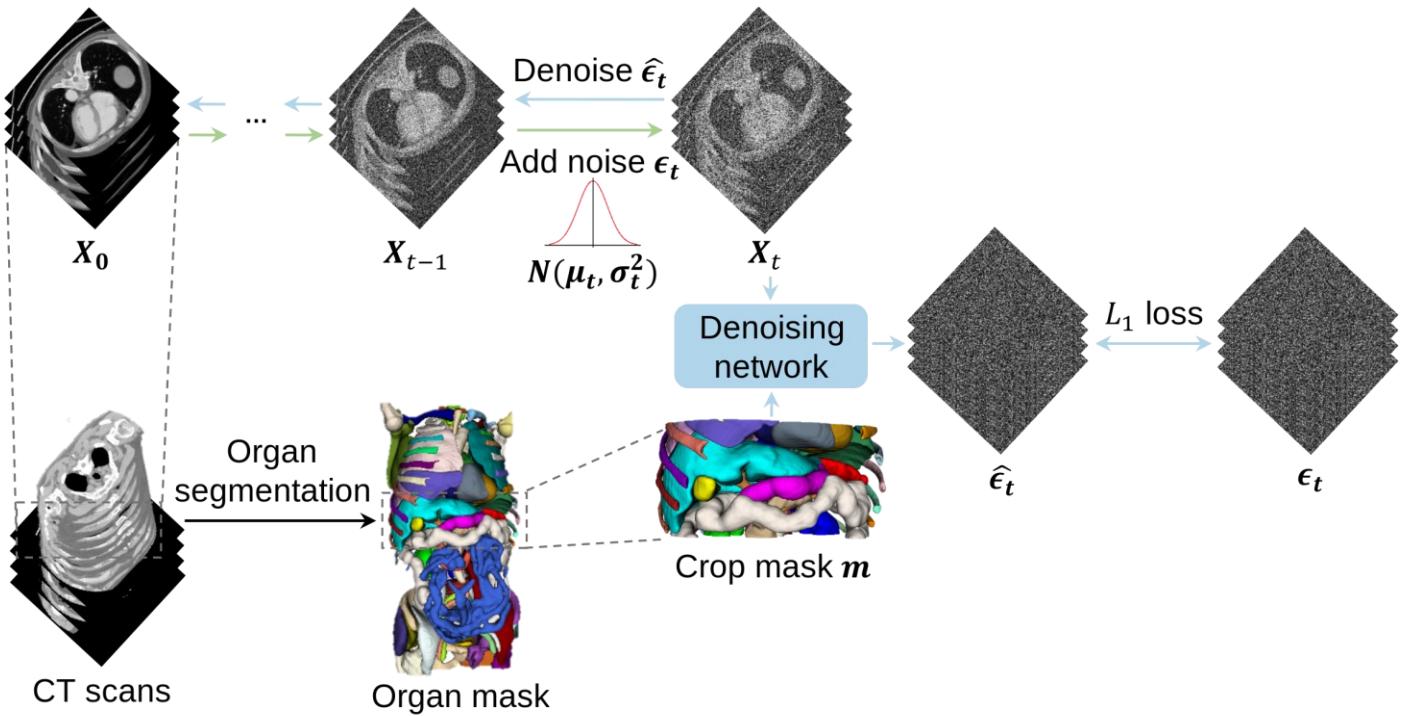
Acknowledgements This work was supported by the National Natural Science Foundation of China (62301311) (X.Z.), Shanghai Municipal Commission of Economy and Informatization (204694) (S. Z.), the Noncommunicable Chronic Diseases – National Science and Technology Major Project (2023ZD0501500) (Z.W.), the National Natural Science Foundation of China (U123A20457) (Z.W.), and the Traditional Chinese Medicine Multidisciplinary Innovation Team Program of Liaoning Province (LNZYYCXTD-JCCX-002) (Z.W.), National Natural Science Foundation of China (82203199) (H.C.).

Author contributions Conceptualization: W.L., H.C., X.Z., P.R., S.Z., and Z.W. Methodology: W.L., H.C., Z.Z., and L.L. Investigation: W.L., H.C., Z.Z., L.L., Q.X., Y.G., P.G., Y.J., C.W., G.W., T.X., and Y.Z. Visualization: W.L. and H.C. Funding acquisition: H.C., X.Z., P.R., S.Z., and Z.W. Supervision: X.Z., P.R., S.Z., and Z.W. Writing: W.L., H.C., and L.L.

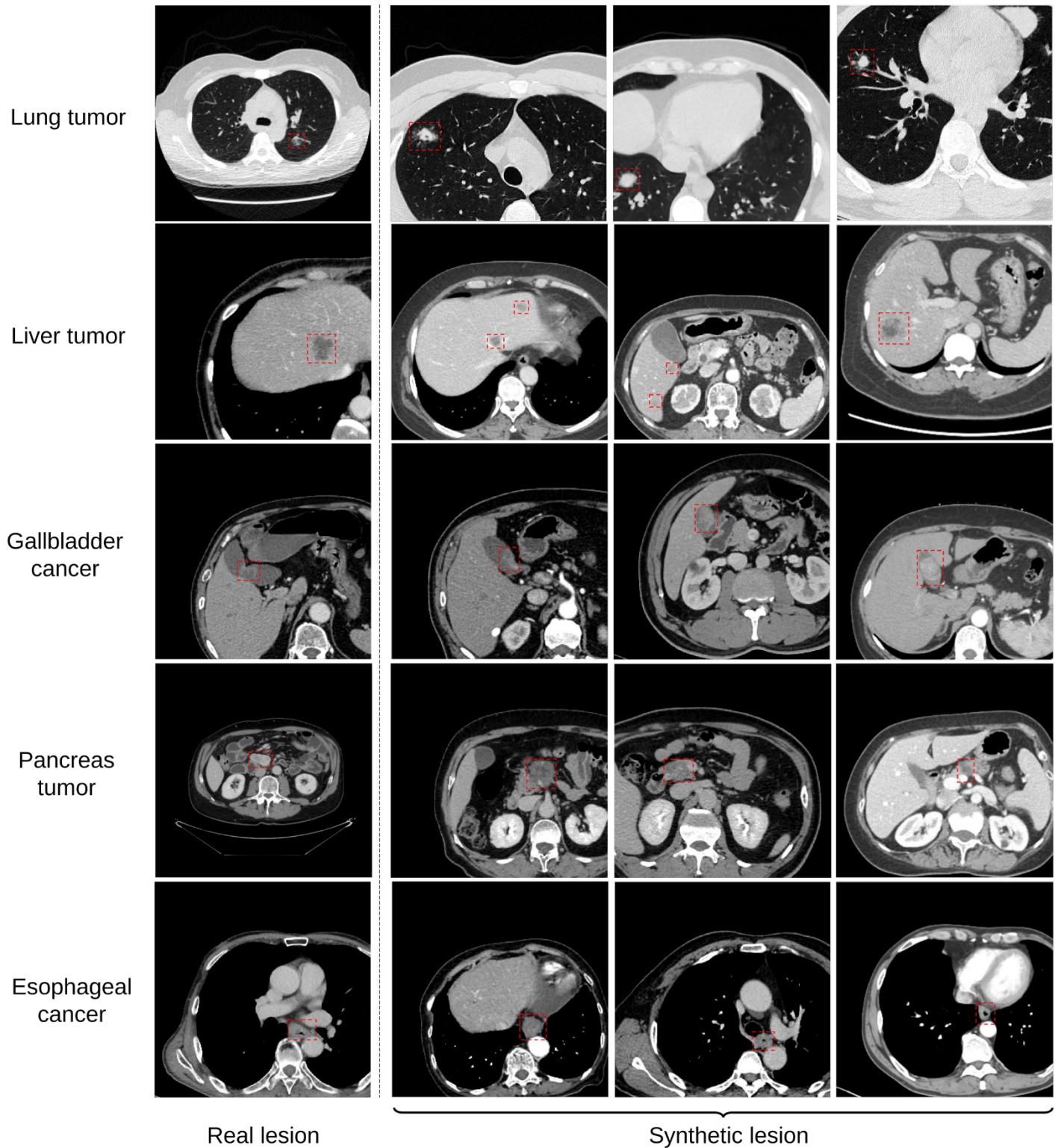
Competing interest The authors declare no competing interests.

Additional information

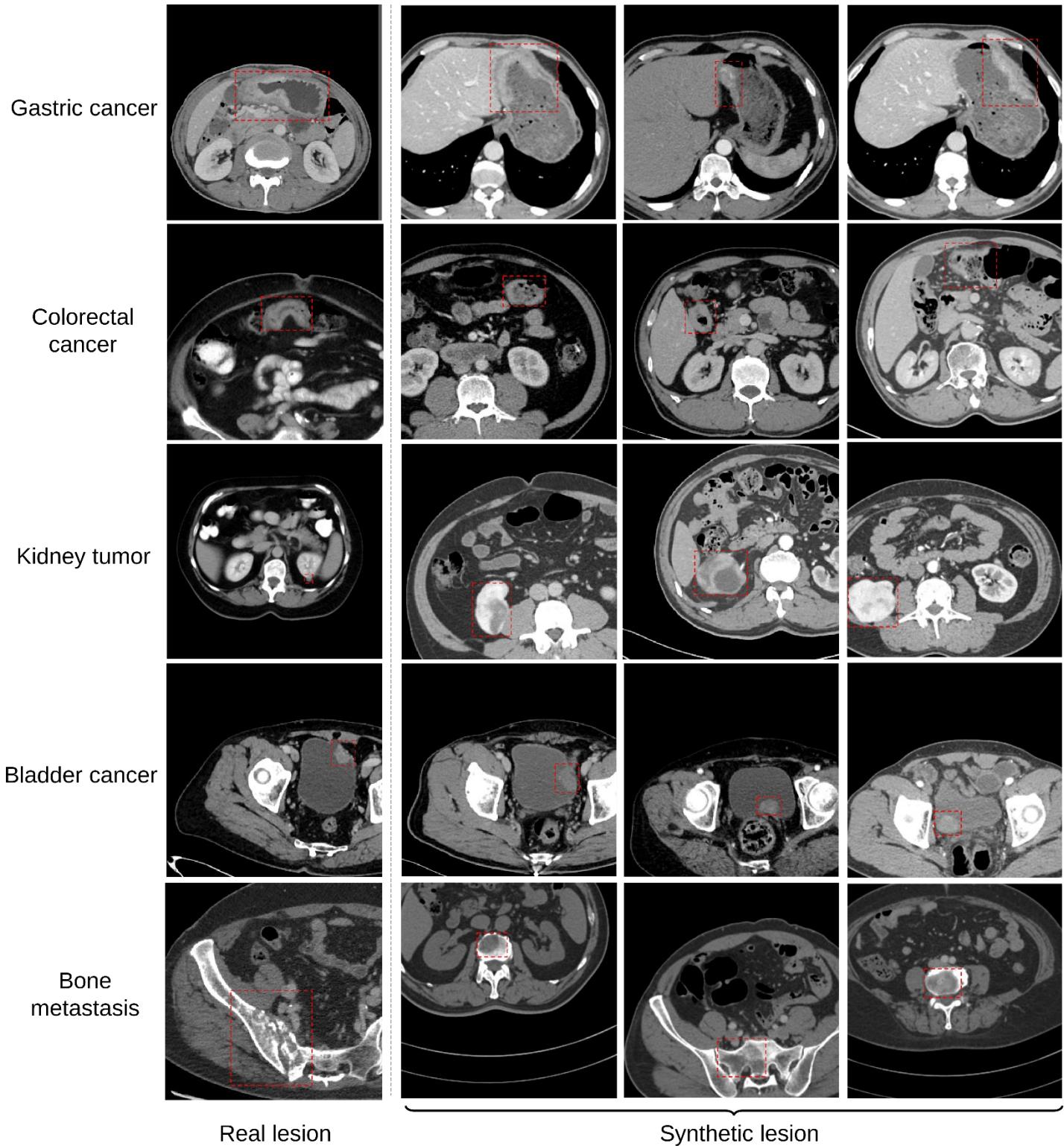
Correspondence and requests for materials should be addressed to Xiaofan Zhang, Pranav Rajpurkar, Shaoting Zhang or Zhenning Wang



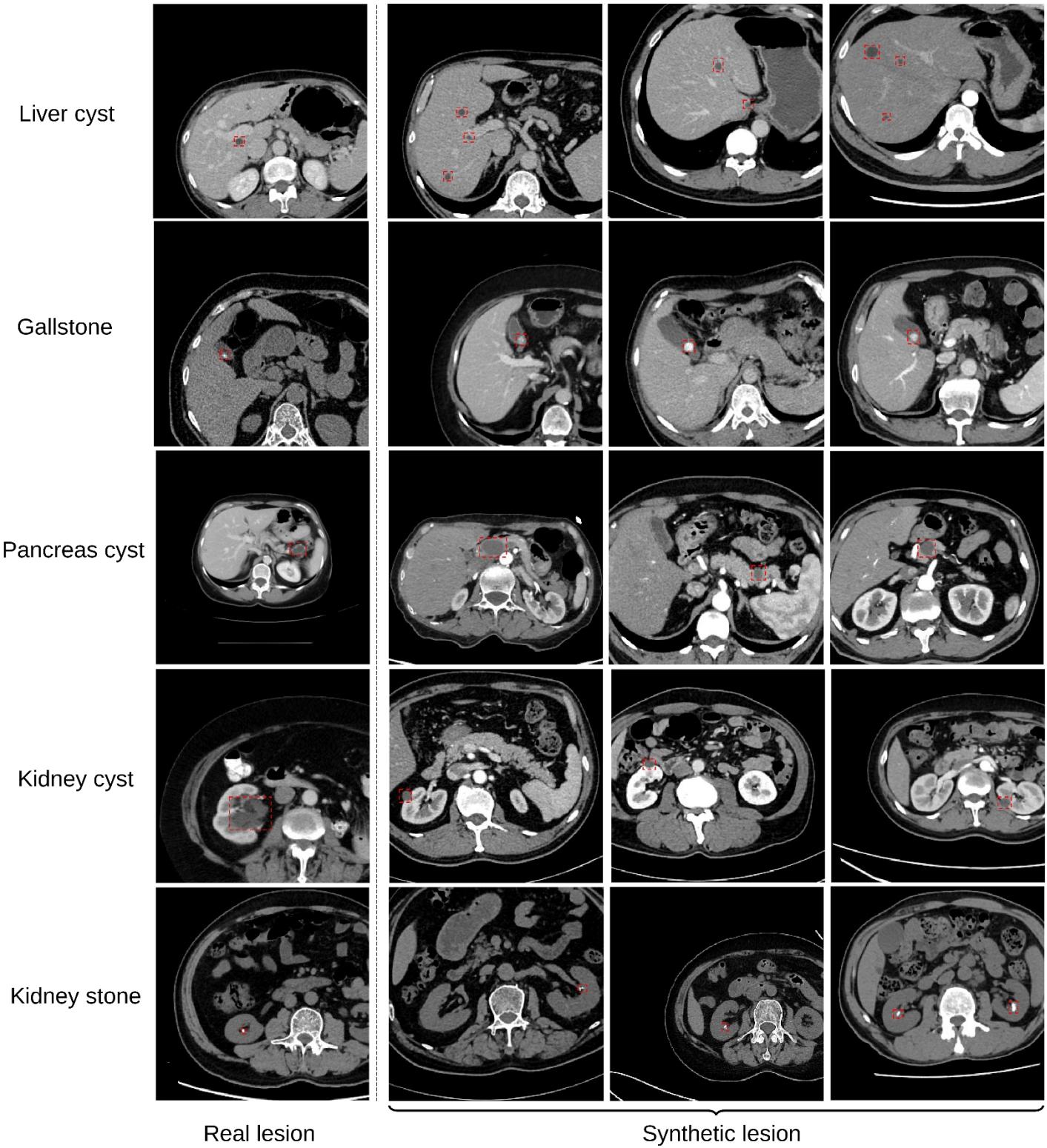
Extended Data Figure 1 | Overview of denoising network training process. During training, random Gaussian noise is added to the cropped original CT images (X_0), and the network learns to predict and remove this noise. The corresponding organ segmentation mask is concatenated as an input condition to guide the denoising process.



Extended Data Fig.2 | Examples from PASTA-Gen-30K (Part 1)



Extended Data Fig.2 | Examples from PASTA-Gen-30K (Part 2)



Extended Data Fig.2 | Examples from PASTA-Gen-30K (Part 3)

Extended Data Table 1 | Structured Attributes for Lesion Simulation in PASTA-Gen

Attribute	Definition	Structured descriptions/options
Enhancement status	Intravenous contrast agent usage	"Enhanced CT", "Plain CT"
Location	Organ-specific anatomical regions	"Lung", "Liver", "Gallbladder", "Pancreas", "Esophagus", "Stomach", "Colorectal", "Kidney", "Bladder", "Bone"
Size	Dimensions across axial, sagittal, and coronal planes	Z×X×Y mm
Shape	Morphological characteristics of the lesion	"Round-like", "Irregular", "Wall thickening", "Punctate, nodular"
Density	Radiographic attenuation properties of the lesion	"Hypodense", "Isodense", "Hyperdense"
Heterogeneity	Uniformity of attenuation within the lesion	"Homogeneous", "Heterogeneous"
Surface	Features of the lesion's border and surface texture	"Well-defined margin", "Ill-defined margin"
Invasion	Invasion of or proximity to adjacent structures	"No close relationship with surrounding structures", "Close relationship with adjacent structures"

Extended Data Table 2 | Distribution of scan ranges in the in-house dataset.

Scan Range	No. of scans (%)
Thorax CT	5,642 (47.5%)
Abdomen-pelvis CT	1,601 (16.4%)
Thorax-abdomen-pelvis CT	763 (7.8%)
Others	2,761 (28.3%)
Total	10,767 (100.0%)

Extended Data Table 3 | Number of healthy organ scans in the template CT dataset

Organ	No. of scans
Healthy Lung	1,720
Contrast CT	0
Plain CT	1,720
Healthy Liver	816
Contrast CT	600
Plain CT	216
Healthy Gallbladder	1,901
Contrast CT	1,389
Plain CT	512
Healthy Pancrease	2,222
Contrast CT	1,620
Plain CT	602
Healthy Esophagus	752
Contrast CT	545
Plain CT	207
Healthy Stomach	2,281
Contrast CT	1,664
Plain CT	617
Healthy Colorectum	1,668
Contrast CT	1,292
Plain CT	376
Healthy Kidney	975
Contrast CT	716
Plain CT	259
Healthy Bladder	2,216
Contrast CT	1,616
Plain CT	600
Healthy Trunk and Extremity	6,811
Bones	
Contrast CT	4,972
Plain CT	1,839

Extended Data Table 4 | PASTA-Gen and PASTA Validation Datasets

Dataset name	Lesion target	No. of scans
Public datasets		
KiTS23	Kidney tumor, kidney cyst	489
MSD-Colon Tumor	Colon tumor	126
MSD-Liver Tumor	Liver tumor	303
MSD-Lung Tumor	Lung tumor	96
MSD-Pancreas*	Pancreas tumor	216
MSD-Pancreas*	Pancreas cyst	65
CMU datasets		
	Liver cyst	30
	Gallbladder cancer	30
	Gallstones	30
	Esophageal cancer	30
	Gastric cancer	30
	Kidney stone	30
	Bladder cancer	30
	Bone metastasis	30
Total		1,535

* MSD-Pancreas is reannotated to pancreas tumor and pancreas cyst.

Extended Data Table 5 | Composition of the Image Realism Evaluation Dataset

Lesion targets	No. of real CT scans	No. of synthetic CT scans
Lung tumor	50	50
Liver tumor	50	50
Gallbladder cancer	30	30
Pancreas tumor	50	50
Esophageal Cancer	30	30
Gastric cancer	30	30
Colorectal cancer	50	50
Kidney tumor	50	50
Bladder cancer	30	30
Bone metastasis	30	30
Liver cyst	30	30
Gallstone	30	30
Pancreas cyst	50	50
Kidney cyst	50	50
Kidney stone	30	30

Supplementary Table 1 | Sufficient Data and Few-Shot Segmentation DSC across Models.

	PASTA	Models Genesis	SuPerM	nnUNet	Universal	p-value*
Lung tumor						
No. of Few-shot Samples†						
1	0.529 (0.458-0.599)	0.179 (0.127-0.230)	<u>0.302</u> <u>(0.230-0.383)</u>	0.188 (0.133-0.254)	0.167 (0.118-0.219)	<0.001
2	0.622 (0.559-0.683)	0.358 (0.291-0.425)	<u>0.383</u> <u>(0.310-0.457)</u>	0.288 (0.220-0.360)	0.266 (0.203-0.330)	<0.001
4	0.680 (0.621-0.728)	0.396 (0.321-0.464)	<u>0.513</u> <u>(0.444-0.590)</u>	0.335 (0.265-0.405)	0.307 (0.247-0.378)	<0.001
8	0.709 (0.656-0.757)	0.365 (0.309-0.423)	<u>0.605</u> <u>(0.546-0.660)</u>	0.323 (0.257-0.389)	0.293 (0.233-0.361)	<0.001
16	0.706 (0.651-0.751)	0.261 (0.206-0.322)	<u>0.591</u> <u>(0.531-0.654)</u>	0.106 (0.073-0.142)	0.088 (0.059-0.121)	<0.001
Full-scale data training (n=64)	0.708 (0.639-0.767)	0.680 (0.616-0.735)	0.661 (0.595-0.720)	<u>0.689</u> <u>(0.630-0.744)</u>	0.661 (0.602-0.708)	0.018
Liver tumor						
No. of Few-shot Samples†						
1	0.243 (0.199-0.290)	0.101 (0.078-0.129)	<u>0.194</u> <u>(0.148-0.247)</u>	0.097 (0.070-0.125)	0.080 (0.057-0.106)	0.004
2	0.477 (0.430-0.527)	0.145 (0.112-0.176)	<u>0.313</u> <u>(0.258-0.369)</u>	0.164 (0.133-0.202)	0.141 (0.110-0.176)	<0.001
4	0.598 (0.548-0.643)	0.193 (0.157-0.233)	<u>0.392</u> <u>(0.341-0.451)</u>	0.170 (0.134-0.206)	0.152 (0.121-0.186)	<0.001
8	0.617 (0.568-0.658)	0.108 (0.081-0.138)	<u>0.390</u> <u>(0.334-0.447)</u>	0.064 (0.041-0.090)	0.060 (0.040-0.084)	<0.001
16	0.629 (0.580-0.677)	0.131 (0.100-0.164)	<u>0.316</u> <u>(0.261-0.376)</u>	0.018 (0.011-0.025)	0.012 (0.008-0.016)	<0.001
Full-scale data training (n=131)	0.696 (0.652-0.738)	<u>0.673</u> <u>(0.627-0.716)</u>	0.654 (0.607-0.701)	0.666 (0.619-0.713)	0.640 (0.591-0.687)	<0.001
Gallbladder cancer						
No. of Few-shot Samples†						
1	0.565 (0.454-0.673)	<u>0.112</u> <u>(0.062-0.171)</u>	0.065 (0.009-0.131)	0.101 (0.051-0.160)	0.096 (0.042-0.152)	<0.001
2	0.608 (0.509-0.696)	0.122 (0.067-0.179)	<u>0.145</u> <u>(0.074-0.226)</u>	0.104 (0.055-0.162)	0.095 (0.046-0.153)	<0.001
4	0.649 (0.554-0.736)	0.079 (0.039-0.125)	<u>0.217</u> <u>(0.137-0.301)</u>	0.152 (0.097-0.216)	0.130 (0.080-0.190)	<0.001
8	0.630 (0.528-0.721)	0.206 (0.128-0.286)	<u>0.419</u> <u>(0.317-0.518)</u>	0.203 (0.137-0.269)	0.181 (0.116-0.242)	<0.001
16	0.653 (0.544-0.748)	0.107 (0.059-0.158)	<u>0.365</u> <u>(0.268-0.465)</u>	0.105 (0.065-0.148)	0.081 (0.047-0.116)	<0.001
Full-scale data training (n=30)	0.654 (0.556-0.744)	0.598 (0.484-0.700)	0.575 (0.470-0.671)	<u>0.644</u> <u>(0.541-0.742)</u>	0.613 (0.517-0.708)	0.464

Pancreas tumor

No. of Few-shot Samples†						
1	0.158 (0.133-0.186)	0.037 (0.024-0.052)	0.041 (0.027-0.058)	<u>0.046</u> <u>(0.032-0.062)</u>	0.040 (0.026-0.054)	<0.001
2	0.216 (0.187-0.252)	0.073 (0.054-0.092)	<u>0.079</u> <u>(0.060-0.101)</u>	0.055 (0.038-0.075)	0.051 (0.035-0.065)	<0.001
4	0.209 (0.175-0.244)	0.015 (0.007-0.026)	<u>0.122</u> <u>(0.096-0.150)</u>	0.005 (0.002-0.010)	0.007 (0.004-0.010)	<0.001
8	0.335 (0.294-0.375)	0.043 (0.029-0.057)	<u>0.141</u> <u>(0.113-0.169)</u>	0.000 (0.000-0.000)	0.005 (0.004-0.006)	<0.001
16	0.414 (0.374-0.454)	0.066 (0.049-0.085)	<u>0.207</u> <u>(0.176-0.241)</u>	0.000 (0.000-0.000)	0.005 (0.004-0.007)	<0.001
Full-scale data training (n=220)	0.510 (0.468-0.551)	0.482 (0.442-0.520)	0.408 (0.365-0.449)	<u>0.491</u> <u>(0.452-0.531)</u>	0.468 (0.428-0.508)	0.010

Esophageal Cancer

No. of Few-shot Samples†						
1	0.577 (0.466-0.683)	0.166 (0.089-0.253)	<u>0.319</u> <u>(0.218-0.419)</u>	0.175 (0.089-0.268)	0.163 (0.089-0.248)	<0.001
2	0.608 (0.507-0.698)	0.279 (0.184-0.377)	<u>0.422</u> <u>(0.324-0.518)</u>	0.351 (0.254-0.458)	0.323 (0.232-0.417)	<0.001
4	0.620 (0.520-0.719)	<u>0.458</u> <u>(0.357-0.552)</u>	0.384 (0.255-0.512)	0.350 (0.242-0.464)	0.312 (0.214-0.415)	<0.001
8	0.659 (0.555-0.758)	0.345 (0.234-0.460)	0.491 (0.376-0.604)	<u>0.506</u> <u>(0.405-0.605)</u>	0.476 (0.389-0.566)	<0.001
16	0.684 (0.582-0.775)	0.385 (0.275-0.500)	<u>0.624</u> <u>(0.527-0.709)</u>	0.562 (0.481-0.637)	0.526 (0.447-0.604)	0.007
Full-scale data training (n=30)	0.683 (0.576-0.783)	0.635 (0.508-0.741)	<u>0.646</u> <u>(0.535-0.746)</u>	0.634 (0.514-0.739)	0.602 (0.479-0.703)	0.004

Gastric Cancer

No. of Few-shot Samples†						
1	0.309 (0.209-0.418)	<u>0.074</u> <u>(0.039-0.110)</u>	0.045 (0.015-0.083)	0.054 (0.021-0.091)	0.045 (0.017-0.080)	<0.001
2	0.494 (0.413-0.574)	<u>0.098</u> <u>(0.056-0.142)</u>	0.033 (0.013-0.056)	0.090 (0.049-0.132)	0.067 (0.031-0.112)	<0.001
4	0.548 (0.469-0.623)	0.183 (0.129-0.241)	<u>0.235</u> <u>(0.165-0.302)</u>	0.165 (0.110-0.220)	0.135 (0.081-0.190)	<0.001
8	0.584 (0.505-0.655)	0.222 (0.159-0.293)	<u>0.383</u> <u>(0.295-0.455)</u>	0.217 (0.161-0.276)	0.185 (0.129-0.245)	<0.001
16	0.615 (0.544-0.677)	0.275 (0.219-0.330)	<u>0.461</u> <u>(0.398-0.524)</u>	0.247 (0.204-0.288)	0.203 (0.162-0.241)	<0.001
Full-scale data training (n=30)	0.627 (0.542-0.697)	0.561 (0.496-0.631)	0.544 (0.460-0.617)	<u>0.581</u> <u>(0.505-0.646)</u>	0.554 (0.486-0.624)	0.004

Colon Cancer

No. of Few-shot Samples†						
1	0.044 (0.027-0.066)	0.006 (0.000-0.015)	<u>0.019</u> <u>(0.008-0.034)</u>	0.005 (0.002-0.008)	0.008 (0.006-0.010)	0.001
2	0.167 (0.126-0.208)	0.024 (0.012-0.041)	<u>0.050</u> <u>(0.029-0.071)</u>	0.023 (0.011-0.038)	0.026 (0.014-0.040)	<0.001
4	0.245 (0.204-0.292)	0.042 (0.024-0.064)	<u>0.115</u> <u>(0.083-0.147)</u>	0.061 (0.042-0.085)	0.050 (0.031-0.069)	<0.001

8	0.290 (0.240-0.343)	0.082 (0.059-0.107)	<u>0.148</u> <u>(0.117-0.182)</u>	0.093 (0.069-0.119)	0.077 (0.054-0.101)	<0.001
16	0.391 (0.338-0.445)	<u>0.112</u> <u>(0.086-0.142)</u>	0.110 (0.083-0.139)	0.059 (0.046-0.074)	0.043 (0.031-0.056)	<0.001
Full-scale data training (n=126)	0.517 (0.458-0.583)	0.496 (0.436-0.556)	0.481 (0.425-0.542)	<u>0.517</u> <u>(0.454-0.576)</u>	0.491 (0.438-0.547)	0.901

Kidney tumor

No. of Few-shot Samples†						
1	0.246 (0.220-0.271)	0.039 (0.030-0.049)	<u>0.072</u> <u>(0.059-0.086)</u>	0.052 (0.042-0.061)	0.043 (0.034-0.052)	<0.001
2	0.301 (0.269-0.330)	0.015 (0.010-0.019)	<u>0.127</u> <u>(0.110-0.146)</u>	0.046 (0.037-0.054)	0.037 (0.030-0.046)	<0.001
4	0.394 (0.361-0.428)	0.006 (0.003-0.010)	<u>0.174</u> <u>(0.152-0.198)</u>	0.021 (0.014-0.029)	0.021 (0.014-0.027)	<0.001
8	0.632 (0.606-0.658)	0.133 (0.116-0.150)	<u>0.254</u> <u>(0.229-0.279)</u>	0.143 (0.128-0.159)	0.124 (0.107-0.139)	<0.001
16	0.678 (0.654-0.706)	0.144 (0.129-0.161)	<u>0.336</u> <u>(0.309-0.361)</u>	0.051 (0.042-0.062)	0.046 (0.036-0.055)	<0.001
Full-scale data training (n=489)	0.814 (0.793-0.834)	<u>0.800</u> <u>(0.780-0.819)</u>	0.722 (0.695-0.748)	0.784 (0.763-0.805)	0.756 (0.736-0.776)	<0.001

Bladder cancer

No. of Few-shot Samples†						
1	0.509 (0.381-0.631)	<u>0.136</u> <u>(0.064-0.213)</u>	0.052 (0.000-0.122)	0.081 (0.018-0.158)	0.075 (0.014-0.152)	<0.001
2	0.667 (0.566-0.764)	<u>0.290</u> <u>(0.196-0.396)</u>	0.194 (0.105-0.294)	0.219 (0.140-0.311)	0.201 (0.115-0.297)	<0.001
4	0.697 (0.600-0.781)	<u>0.445</u> <u>(0.355-0.535)</u>	0.396 (0.289-0.500)	0.360 (0.264-0.461)	0.332 (0.242-0.426)	<0.001
8	0.712 (0.611-0.797)	<u>0.500</u> <u>(0.400-0.588)</u>	0.441 (0.325-0.553)	0.318 (0.215-0.421)	0.289 (0.194-0.385)	<0.001
16	0.757 (0.674-0.828)	<u>0.550</u> <u>(0.466-0.637)</u>	0.309 (0.226-0.394)	0.331 (0.230-0.432)	0.301 (0.205-0.404)	<0.001
Full-scale data training (n=30)	0.790 (0.717-0.844)	<u>0.769</u> <u>(0.694-0.830)</u>	0.660 (0.550-0.756)	0.759 (0.680-0.829)	0.738 (0.658-0.806)	0.127

Bone metastasis

No. of Few-shot Samples†						
1	0.313 (0.266-0.386)	0.064 (0.015-0.120)	0.063 (0.019-0.122)	<u>0.067</u> <u>(0.021-0.124)</u>	0.065 (0.020-0.123)	<0.001
2	0.360 (0.251-0.421)	0.049 (0.001-0.118)	0.050 (0.009-0.106)	<u>0.058</u> <u>(0.011-0.128)</u>	0.050 (0.007-0.111)	<0.001
4	0.371 (0.287-0.458)	0.068 (0.023-0.123)	0.053 (0.013-0.108)	<u>0.076</u> <u>(0.026-0.135)</u>	0.064 (0.022-0.113)	<0.001
8	0.402 (0.298-0.485)	<u>0.124</u> <u>(0.059-0.205)</u>	0.060 (0.020-0.115)	0.081 (0.033-0.138)	0.074 (0.025-0.145)	<0.001
16	0.432 (0.267-0.538)	<u>0.023</u> <u>(0.010-0.040)</u>	0.002 (0.000-0.007)	0.031 (0.011-0.058)	0.023 (0.006-0.051)	<0.001
Full-scale data training (n=30)	0.433 (0.317-0.574)	0.321 (0.210-0.431)	<u>0.389</u> <u>(0.284-0.489)</u>	0.376 (0.252-0.488)	0.346 (0.245-0.458)	0.020

Liver cyst

No. of Few-shot Samples†						
1	0.639 (0.545-0.727)	0.015 (0.000-0.042)	<u>0.286</u> <u>(0.172-0.407)</u>	0.000 (0.000-0.000)	0.009 (0.004-0.015)	<0.001
2	0.631 (0.533-0.740)	0.038 (0.008-0.075)	<u>0.199</u> <u>(0.087-0.328)</u>	0.000 (0.000-0.000)	0.006 (0.003-0.010)	<0.001
4	0.748 (0.671-0.814)	0.022 (0.001-0.050)	<u>0.315</u> <u>(0.204-0.436)</u>	0.000 (0.000-0.000)	0.004 (0.001-0.008)	<0.001
8	0.716 (0.619-0.802)	0.056 (0.009-0.114)	<u>0.271</u> <u>(0.161-0.391)</u>	0.000 (0.000-0.000)	0.002 (0.000-0.006)	<0.001
16	0.789 (0.694-0.854)	0.010 (0.001-0.025)	<u>0.299</u> <u>(0.165-0.454)</u>	0.000 (0.000-0.000)	0.002 (0.000-0.005)	<0.001
Full-scale data training (n=30)	0.755 (0.649-0.846)	0.713 (0.611-0.799)	<u>0.724</u> <u>(0.628-0.802)</u>	0.716 (0.612-0.794)	0.686 (0.588-0.773)	<0.001

Gallstone

No. of Few-shot Samples†						
1	0.444 (0.312-0.581)	<u>0.100</u> <u>(0.030-0.182)</u>	0.072 (0.000-0.153)	0.092 (0.022-0.185)	0.087 (0.024-0.172)	<0.001
2	0.637 (0.531-0.732)	0.190 (0.097-0.286)	0.254 (0.159-0.354)	<u>0.280</u> <u>(0.164-0.396)</u>	0.247 (0.145-0.354)	<0.001
4	0.687 (0.589-0.776)	0.321 (0.208-0.427)	<u>0.364</u> <u>(0.254-0.486)</u>	0.348 (0.244-0.454)	0.329 (0.226-0.437)	<0.001
8	0.769 (0.692-0.829)	0.493 (0.368-0.611)	0.366 (0.255-0.492)	<u>0.510</u> <u>(0.397-0.619)</u>	0.477 (0.360-0.580)	<0.001
16	0.762 (0.684-0.833)	<u>0.465</u> <u>(0.352-0.580)</u>	0.133 (0.048-0.234)	0.356 (0.244-0.471)	0.337 (0.226-0.457)	<0.001
Full-scale data training (n=30)	<u>0.781</u> <u>(0.714-0.836)</u>	0.760 (0.681-0.827)	0.754 (0.696-0.808)	0.785 (0.726-0.838)	0.756 (0.694-0.812)	0.358

Pancreas cyst

No. of Few-shot Samples†						
1	0.465 (0.382-0.541)	0.077 (0.043-0.118)	<u>0.099</u> <u>(0.054-0.153)</u>	0.056 (0.022-0.100)	0.055 (0.023-0.092)	<0.001
2	0.460 (0.375-0.546)	0.104 (0.054-0.156)	<u>0.130</u> <u>(0.074-0.197)</u>	0.081 (0.040-0.132)	0.075 (0.040-0.118)	<0.001
4	0.515 (0.428-0.609)	0.203 (0.138-0.277)	<u>0.261</u> <u>(0.178-0.352)</u>	0.179 (0.109-0.252)	0.162 (0.099-0.229)	<0.001
8	0.641 (0.552-0.710)	0.220 (0.158-0.285)	<u>0.458</u> <u>(0.389-0.534)</u>	0.187 (0.119-0.253)	0.175 (0.115-0.241)	<0.001
16	0.707 (0.636-0.768)	0.308 (0.239-0.378)	<u>0.497</u> <u>(0.412-0.579)</u>	0.092 (0.048-0.140)	0.082 (0.044-0.125)	<0.001
Full-scale data training (n=61)	0.722 (0.653-0.788)	0.680 (0.599-0.751)	0.637 (0.555-0.712)	<u>0.715</u> <u>(0.641-0.779)</u>	0.682 (0.606-0.750)	0.020

Kidney cyst

No. of Few-shot Samples†						
1	0.161 (0.132-0.192)	0.045 (0.030-0.061)	0.048 (0.033-0.066)	<u>0.056</u> <u>(0.039-0.076)</u>	0.054 (0.038-0.071)	<0.001
2	0.137 (0.108-0.167)	<u>0.013</u> <u>(0.006-0.021)</u>	0.010 (0.004-0.019)	0.000 (0.000-0.000)	0.005 (0.004-0.007)	<0.001
4	0.141 (0.114-0.171)	0.004 (0.001-0.009)	0.000 (0.000-0.000)	0.000 (0.000-0.000)	<u>0.005</u> <u>(0.004-0.006)</u>	<0.001

8	0.102 (0.077-0.127)	0.000 (0.000-0.000)	0.000 (0.000-0.000)	0.000 (0.000-0.000)	<u>0.005</u> <u>(0.004-0.007)</u>	<0.001
16	0.289 (0.252-0.329)	0.000 (0.000-0.000)	0.000 (0.000-0.000)	0.000 (0.000-0.000)	<u>0.005</u> <u>(0.004-0.006)</u>	<0.001
Full-scale data training (n=489)	0.660 (0.622-0.696)	<u>0.641</u> <u>(0.607-0.677)</u>	0.616 (0.582-0.649)	0.624 (0.589-0.655)	0.598 (0.562-0.630)	<0.001

Kidney stone

No. of Few-shot Samples†						
1	0.464 (0.350-0.573)	0.000 (0.000-0.000)	0.000 (0.000-0.000)	0.000 (0.000-0.000)	<u>0.002</u> <u>(0.000-0.005)</u>	<0.001
2	0.678 (0.622-0.730)	0.000 (0.000-0.000)	0.000 (0.000-0.000)	0.000 (0.000-0.000)	<u>0.005</u> <u>(0.002-0.009)</u>	<0.001
4	0.780 (0.733-0.819)	0.000 (0.000-0.000)	0.000 (0.000-0.000)	0.000 (0.000-0.000)	<u>0.006</u> <u>(0.002-0.011)</u>	<0.001
8	0.636 (0.502-0.752)	0.000 (0.000-0.000)	0.000 (0.000-0.000)	0.000 (0.000-0.000)	<u>0.005</u> <u>(0.001-0.009)</u>	<0.001
16	0.822 (0.799-0.844)	0.004 (0.000-0.011)	0.000 (0.000-0.000)	0.000 (0.000-0.000)	<u>0.005</u> <u>(0.001-0.009)</u>	<0.001
Full-scale data training (n=30)	0.777 (0.713-0.820)	0.755 (0.682-0.809)	<u>0.768</u> <u>(0.707-0.814)</u>	0.755 (0.675-0.823)	0.724 (0.638-0.789)	0.261

Gallbladder cancer (plain-CT)

Full-scale data training (n=30)	0.570 (0.462-0.657)	0.472 (0.362-0.581)	0.506 (0.392-0.616)	0.519 (0.401-0.630)	0.498 (0.381-0.601)	0.015
---------------------------------	--------------------------------------	------------------------	------------------------	------------------------	------------------------	-------

Rectal cancer (plain-CT)

Full-scale data training (n=30)	0.705 (0.633-0.763)	0.609 (0.520-0.695)	0.669 (0.586-0.739)	0.663 (0.579-0.736)	0.636 (0.555-0.707)	0.015
---------------------------------	--------------------------------------	------------------------	------------------------	------------------------	------------------------	-------

Bladder cancer (plain-CT)

Full-scale data training (n=30)	0.703 (0.620-0.780)	0.649 (0.570-0.719)	0.608 (0.508-0.702)	0.668 (0.592-0.739)	0.636 (0.564-0.709)	0.002
---------------------------------	--------------------------------------	------------------------	------------------------	------------------------	------------------------	-------

Brain tumor (MRI)

No. of Few-shot Samples†						
1	0.310 (0.288-0.333)	0.303 (0.281-0.325)	<u>0.312</u> <u>(0.287-0.337)</u>	0.288 (0.265-0.309)	0.253 (0.231-0.273)	0.208
2	0.344 (0.318-0.372)	0.290 (0.269-0.311)	<u>0.310</u> <u>(0.287-0.333)</u>	0.270 (0.247-0.292)	0.240 (0.219-0.264)	<0.001
4	0.438 (0.415-0.462)	0.376 (0.353-0.399)	<u>0.403</u> <u>(0.376-0.428)</u>	0.348 (0.324-0.370)	0.315 (0.293-0.336)	<0.001
8	0.465 (0.441-0.488)	0.420 (0.397-0.443)	<u>0.450</u> <u>(0.425-0.475)</u>	0.423 (0.399-0.447)	0.386 (0.362-0.409)	0.011
16	0.504 (0.479-0.528)	0.447 (0.425-0.471)	<u>0.468</u> <u>(0.443-0.490)</u>	0.457 (0.434-0.483)	0.418 (0.393-0.440)	<0.001

Liver tumor (MRI)

No. of Few-shot Samples†						
1	0.267 (0.219-0.315)	0.211 (0.173-0.255)	<u>0.237</u> <u>(0.185-0.287)</u>	0.213 (0.171-0.259)	0.181 (0.141-0.227)	0.030
2	0.339 (0.282-0.397)	0.225 (0.185-0.269)	<u>0.232</u> <u>(0.185-0.291)</u>	0.193 (0.152-0.236)	0.166 (0.127-0.209)	<0.001

4	0.492 (0.422-0.555)	0.284 (0.233-0.341)	<u>0.341</u> <u>(0.279-0.405)</u>	0.252 (0.204-0.307)	0.232 (0.183-0.280)	<0.001
8	0.562 (0.495-0.628)	0.289 (0.245-0.337)	<u>0.430</u> <u>(0.368-0.501)</u>	0.296 (0.244-0.343)	0.264 (0.212-0.319)	<0.001
16	0.603 (0.532-0.672)	0.292 (0.238-0.339)	<u>0.469</u> <u>(0.400-0.540)</u>	0.283 (0.235-0.326)	0.244 (0.197-0.292)	<0.001

Non-parametric bootstrapping with 1,000 bootstrap replicates is employed for statistical analysis. The 95% CI is included in parentheses.

* *p*-values were calculated using a one-sided Wilcoxon test comparing the **top-performing model** and the second-best model for each task

† 2000 iterations.

Supplementary Table 2 | Comparison of Tumor Detection on Plain-CT Data Based on accuracy.

	PASTA	PASTA (w/o pretraining)	Models Genesis	SuPerM	p-value*
Gallbladder cancer	0.954 (0.933-0.976)	0.875 (0.842-0.909)	0.912 (0.881-0.942)	<u>0.903</u> <u>(0.869-0.933)</u>	<0.001
Rectal cancer	0.967 (0.945-0.985)	0.885 (0.848-0.918)	0.909 (0.879-0.939)	<u>0.918</u> <u>(0.888-0.945)</u>	<0.001
Bladder cancer	0.970 (0.948-0.988)	0.891 (0.858-0.924)	0.909 (0.879-0.939)	<u>0.924</u> <u>(0.891-0.949)</u>	<0.001

Non-parametric bootstrapping with 1,000 bootstrap replicates is employed for statistical analysis. The 95% CI is included in parentheses.

* *p*-values were calculated using a one-sided permutation test with 10,000 permutations comparing the **top-performing model** and the second-best model for each task

Supplementary Table 3 | Comparison of Tumor Detection on Plain-CT Data Based on AUC.

	PASTA	PASTA (w/o pretraining)	Models Genesis	SuPerM	p-value*
Gallbladder cancer	0.963 (0.919-1.000)	0.838 (0.750-0.911)	0.785 (0.678-0.887)	<u>0.855</u> <u>(0.788-0.936)</u>	0.031
Rectal cancer	0.968 (0.897-1.000)	0.790 (0.722-0.844)	0.646 (0.590-0.682)	<u>0.903</u> <u>(0.822-0.983)</u>	0.031
Bladder cancer	0.984 (0.958-1.000)	0.822 (0.742-0.869)	0.721 (0.613-0.875)	<u>0.942</u> <u>(0.872-0.997)</u>	0.031

The values in parentheses represent the minimum and maximum AUC values obtained across the 5-fold cross-validation.

* p-values were calculated using a one-sided Wilcoxon test comparing the **top-performing model** and the second-best model for each task

Supplementary Table 4 | Comparison of Tumor Staging and Survival Prediction AUC Values.

	PASTA	Models Genesis	SuPerM	FMCIB	p-value*
Tumor staging					
Gastric cancer	0.770 (0.636-0.858)	0.548 (0.456-0.656)	0.618 (0.520-0.751)	<u>0.741</u> <u>(0.593-0.851)</u>	0.031
Rectal cancer	0.738 (0.643-0.846)	<u>0.646</u> <u>(0.617-0.683)</u>	0.613 (0.421-0.804)	0.645 (0.461-0.779)	0.031
Bladder cancer	0.855 (0.750-0.950)	0.633 (0.489-0.800)	0.648 (0.500-0.889)	<u>0.689</u> <u>(0.646-0.756)</u>	0.031
Survival prediction					
Lung cancer	0.700 (0.595-0.779)	0.650 (0.562-0.701)	<u>0.687</u> <u>(0.597-0.730)</u>	0.617 (0.591-0.660)	0.156
Gastric cancer	0.660 (0.525-0.784)	0.546 (0.480-0.652)	0.508 (0.344-0.647)	<u>0.626</u> <u>(0.481-0.763)</u>	0.031
Rectal cancer	0.759 (0.718-0.858)	0.626 (0.468-0.750)	<u>0.714</u> <u>(0.656-0.790)</u>	0.713 (0.613-0.805)	0.094
Bladder cancer	0.878 (0.750-1.000)	<u>0.844</u> <u>(0.722-1.000)</u>	0.838 (0.700-1.000)	0.843 (0.750-1.000)	0.500

The values in parentheses represent the minimum and maximum AUC values obtained across the 5-fold cross-validation.

* p-values were calculated using a one-sided Wilcoxon test comparing the **top-performing model** and the second-best model for each task

Supplementary Table 5 | Accuracy and F1-Scores of Various Models in Structured Report Generation

	PASTA	UNet	Models Genesis	SuPerM	p-value*
Shape					
ACC	0.849 <u>(0.831-0.866)</u>	<u>0.789</u> <u>(0.769-0.809)</u>	0.660 (0.636-0.681)	0.788 (0.767-0.807)	<0.001
F1-score	0.781 <u>(0.738-0.818)</u>	0.688 (0.645-0.730)	0.199 (0.194-0.203)	<u>0.742</u> <u>(0.701-0.775)</u>	0.017
Invasion					
ACC	0.815 <u>(0.797-0.833)</u>	0.793 (0.772-0.815)	0.754 (0.732-0.775)	<u>0.810</u> <u>(0.790-0.830)</u>	0.319
F1-score	<u>0.723</u> <u>(0.699-0.749)</u>	0.700 (0.672-0.728)	0.430 (0.422-0.437)	0.732 (0.706-0.759)	0.730
Density					
ACC	0.726 <u>(0.702-0.746)</u>	<u>0.689</u> <u>(0.667-0.711)</u>	0.529 (0.503-0.553)	0.637 (0.612-0.662)	<0.001
F1-score	0.684 <u>(0.657-0.706)</u>	<u>0.644</u> <u>(0.617-0.668)</u>	0.334 (0.315-0.352)	0.589 (0.561-0.615)	<0.001
Heterogeneity					
ACC	0.903 <u>(0.889-0.917)</u>	<u>0.871</u> <u>(0.854-0.888)</u>	0.752 (0.729-0.771)	0.829 (0.810-0.848)	<0.001
F1-score	0.868 <u>(0.849-0.887)</u>	<u>0.825</u> <u>(0.802-0.847)</u>	0.429 (0.422-0.435)	0.762 (0.735-0.787)	<0.001
Surface					
ACC	0.891 <u>(0.875-0.907)</u>	<u>0.863</u> <u>(0.845-0.880)</u>	0.722 (0.702-0.744)	0.811 (0.791-0.830)	<0.001
F1-score	0.863 <u>(0.843-0.883)</u>	<u>0.825</u> <u>(0.802-0.847)</u>	0.419 (0.413-0.427)	0.754 (0.729-0.779)	<0.001

Non-parametric bootstrapping with 1,000 bootstrap replicates is employed for statistical analysis. The 95% CI is included in parentheses.

* p-values were calculated using a one-sided permutation test with 10,000 permutations comparing the **top-performing model** and the second-best model for each task

Supplementary Table 6 | CT modalities selected as templates for simulating each lesion in PASTA-Gen

Lesion	Modality of template CT
Lung tumor	Non-contrast CT
Liver tumor	Enhanced CT
Gallbladder cancer	Enhanced CT
Pancreas tumor	Enhanced CT
Esophageal Cancer	Enhanced CT
Gastric cancer	Enhanced CT
Colorectal cancer	Enhanced CT
Kidney tumor	Enhanced CT
Bladder cancer	Enhanced CT
Bone metastasis	Enhanced CT
Liver cyst	Enhanced CT & Non-contrast CT
Gallstone	Enhanced CT & Non-contrast CT
Pancreas cyst	Enhanced CT
Kidney cyst	Enhanced CT
Kidney stone	Non-contrast CT

Supplementary Table 7 | Class name and value in PASTA-Gen-30K

Lesion	Value
Lung	1
Liver	2
Gallbladder	3
Pancreas	4
Esophagus	5
Stomach	6
Colon & rectal	7
Kidney	8
Bladder	9
Bone	10
Lung tumor	11
Liver tumor	12
Gallbladder cancer	13
Pancreas tumor	14
Esophageal Cancer	15
Gastric cancer	16
Colorectal cancer	17
Kidney tumor	18
Bladder cancer	19
Bone metastasis	20
Liver cyst	21
Gallstone	22
Pancreas cyst	23
Kidney cyst	24
Kidney stone	25