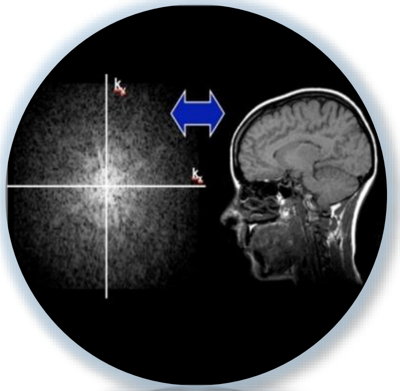
A Multimodal AI Foundation Platform for Trustable and Scalable Biomedical Data Analytics

Hao CHEN
Dept. of Computer Science and Engineering
Hong Kong University of Science and Technology
jhc@cse.ust.hk

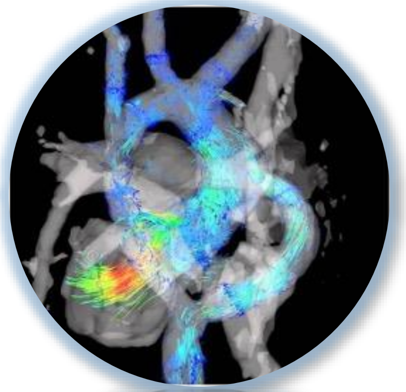
AI Shapes Medical Imaging

From imaging to prognosis

Acquisition



Reconstruction



Visualization (XR)

Analysis & Diagnosis

Treatment & Prognosis

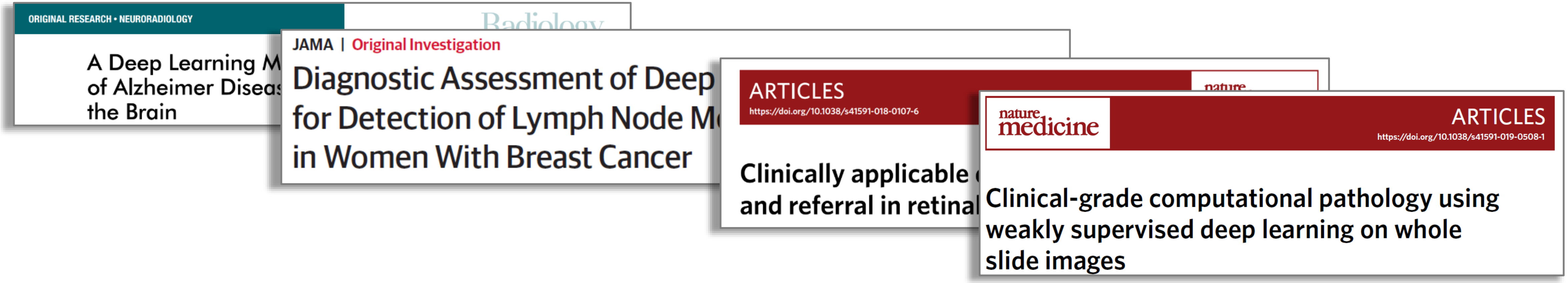
Safer, Faster, Better

See the Invisible, Accurate, Quantitative

Decision Support, Minimize Risk

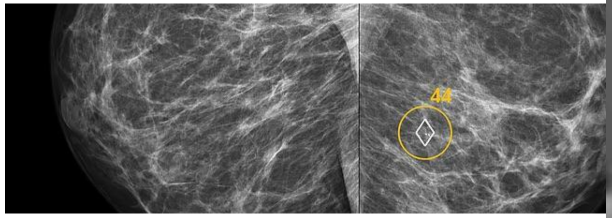
Background and Impact

❑ DL in Medical Image Analysis



❑ FDA/NMPA Approved AI-based Medical Products

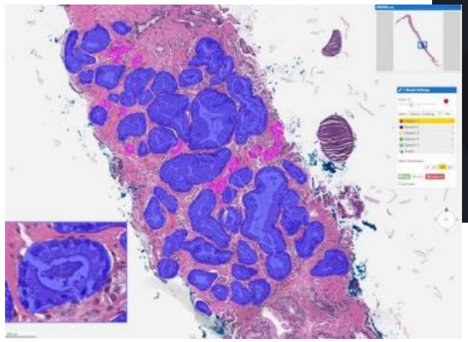
ScreenPoint Medical



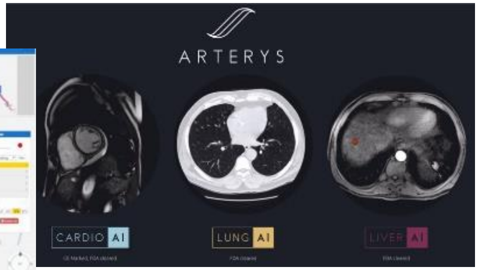
Mammogram

Ultrasound

Fundus



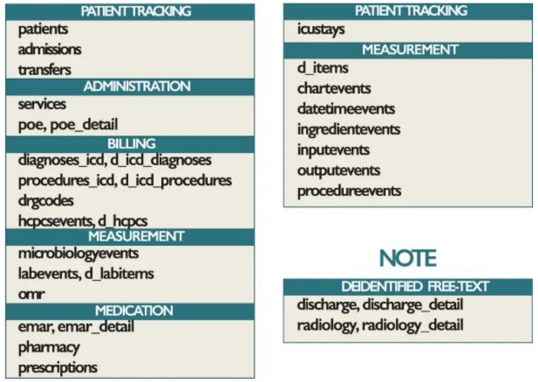
Pathology



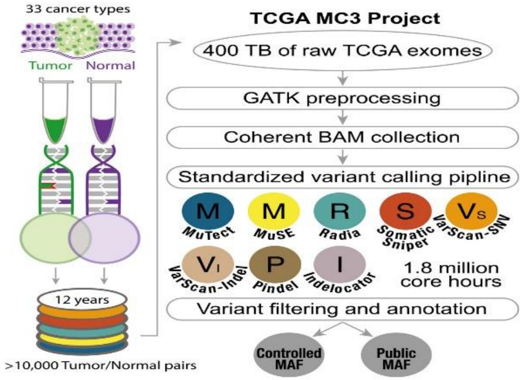
CT & MRI

Diverse Medical Data

❑ Mono-Modal



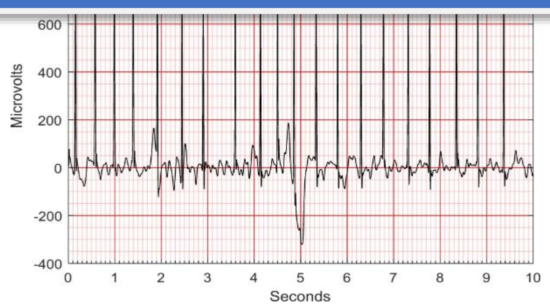
Medical Records [1]



Genomic Data [2]

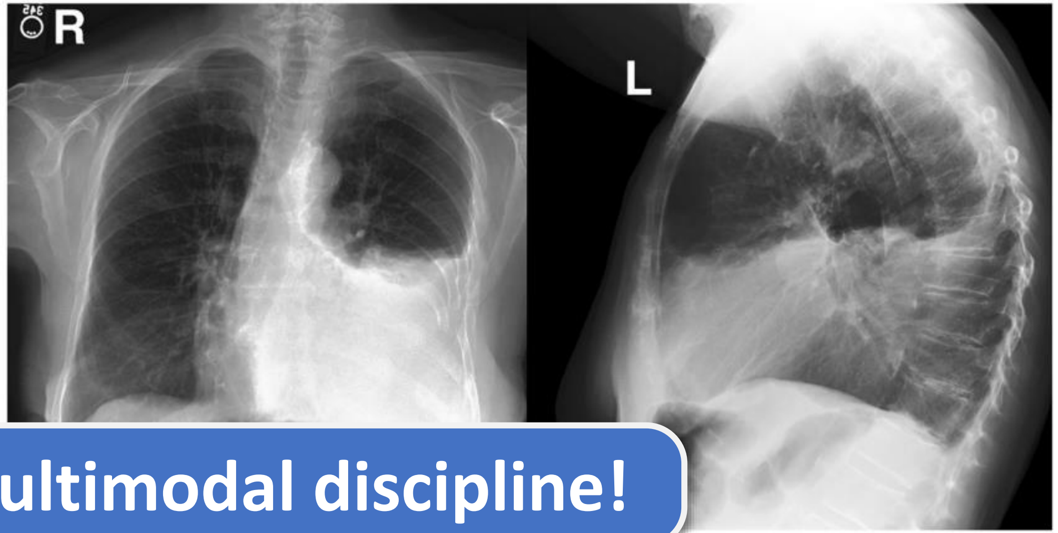
Medicine is an inherently multimodal discipline!

Dermoscopy Images [3]



ECG Signal [4]

❑ Multi-Modal



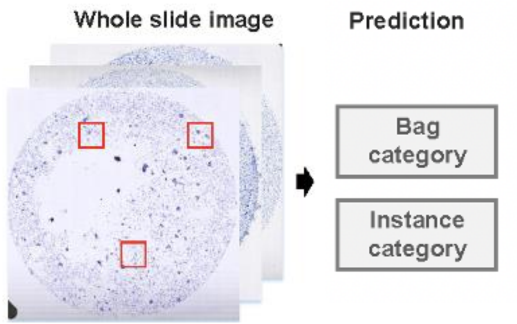
FINDINGS: Cardiac size cannot be evaluated. Large left pleural effusion is new. Small right effusion is new, The upper lungs are clear, Right lower lobe opacities are better seen in prior CT. There is no pneumothorax. There are mild degenerative changes in the thoracic spine
IMPRESSION: Large left pleural effusion

X-ray Imaging and Report [5]

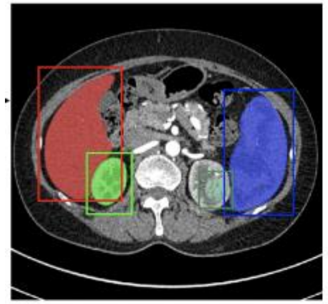
[1] Johnson et al. MIMIC-IV, a freely accessible electronic health record dataset. Scientific data 2023.
[2] Jha et al. Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. Cell Systems 2018.
[3] Tschandl et al. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data 2018.
[4] Zheng et al. A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. Scientific data 2020.
[5] Johnson et al. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Scientific data 2019.

Multimodal Medical Tasks

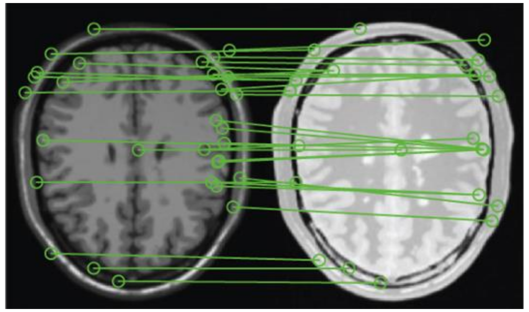
❑ Medical Vision Tasks



Classification



Detection & Segmentation



Registration

❑ Medical Multimodal Tasks

Input Prompt

<image-1> <image-2> Patient Data Age: 40 years Gender: Male Perianal purulent discharge and tenderness. Please caption this scan with finding and impression. <image-3> <image-4>

RadFM Prediction

Finding: Injection of contrast via the orifice of purulent discharge showed opacification of the perianal abscess. No evidence of obvious contrast

...rast

❑ Medi

How to integrate the information from different modalities?

Question:
Do preoperative statins reduce atrial fibrillation after coronary artery bypass grafting?
Context:
(Objective) Recent studies have demonstrated that statins have pleiotropic effects, including anti-inflammatory effects and atrial fibrillation (AF) preventive effects [...]
(Methods) 221 patients underwent CABG in our hospital from 2004 to 2007. 14 patients with preoperative AF and 4 patients with concomitant valve surgery [...]
(Results) The overall incidence of postoperative AF was 26%. *Postoperative AF was significantly lower in the Statin group compared with the Non-statin group (16% versus 33%, p=0.005).* Multivariate analysis demonstrated that independent predictors of AF [...]
Long Answer:
(Conclusion) Our study indicated that preoperative statin therapy seems to reduce AF development after CABG.
Answer: yes

Medical Question Answering

Findings: there is an intraparenchymal hemorrhage in the right cerebellar hemisphere measuring 1.7 cm with vasogenic edema …

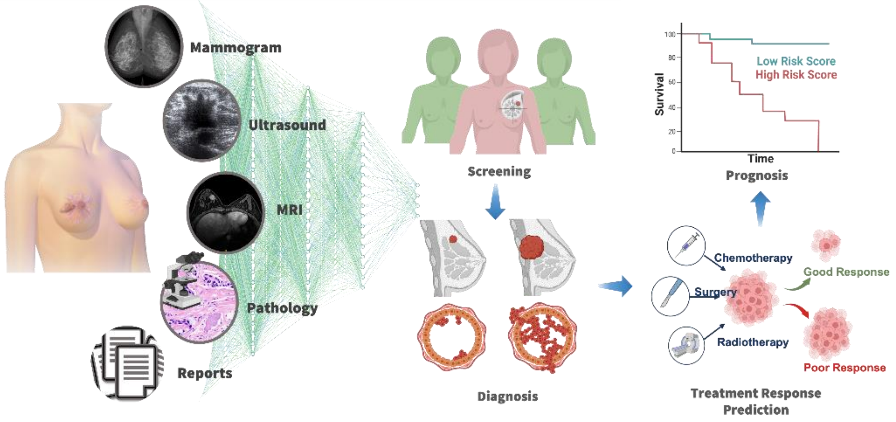
Summary: Target 1. 1.7-cm right cerebellar parenchymal hemorrhage with surrounding vasogenic edema …

Medical Report Summarization

Assessment: Ms. *** is an 87-year-old woman now s/p left craniotomy for a traumatic subdural hematoma whose post-operative course is now complicated by decompensated CHF,

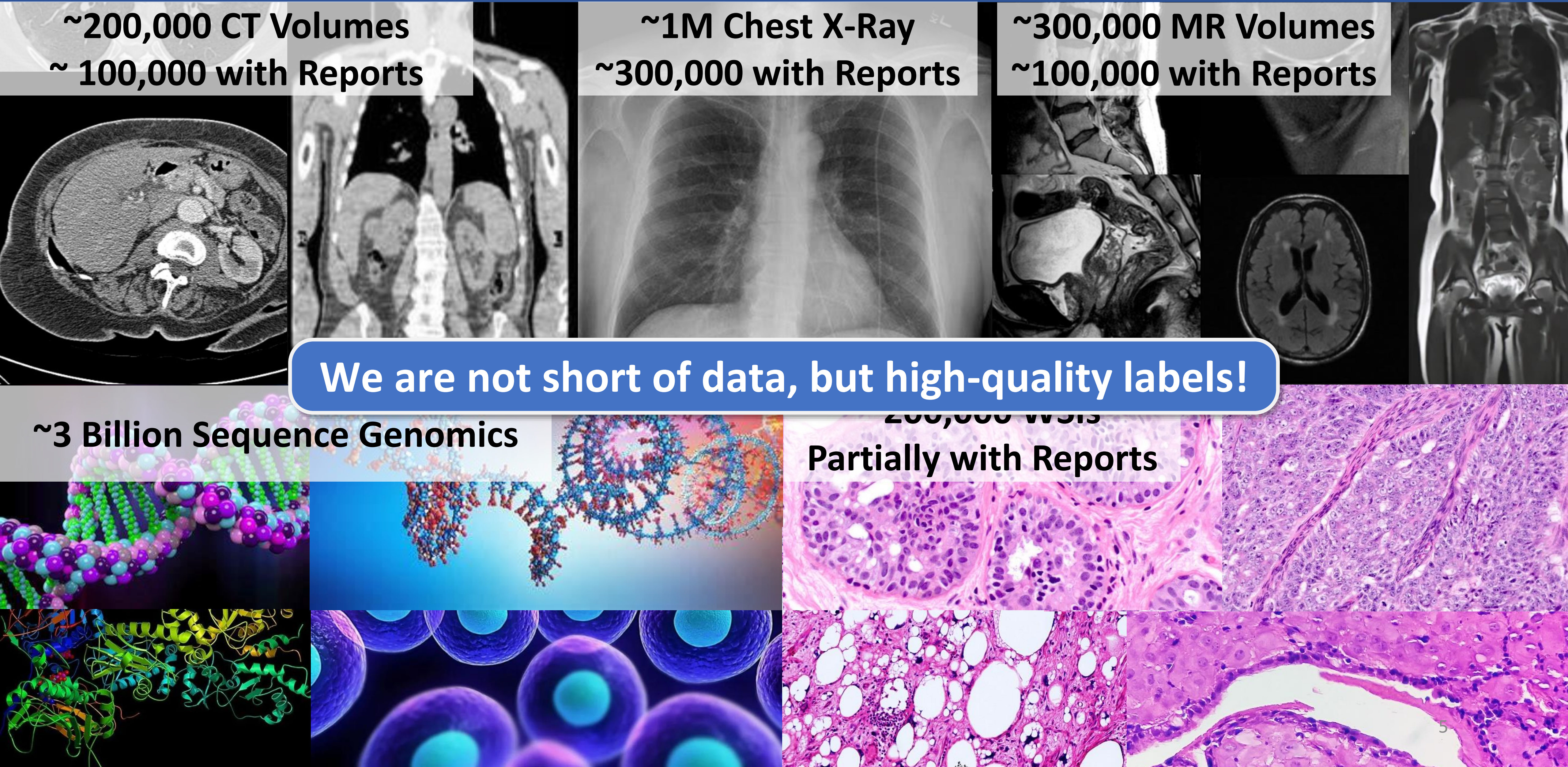
Plan Subsection: Respiratory failure with MRSA pneumonia: continue seven days of vancomycin, **Relation:** Direct

Assessment and Plan Reasoning

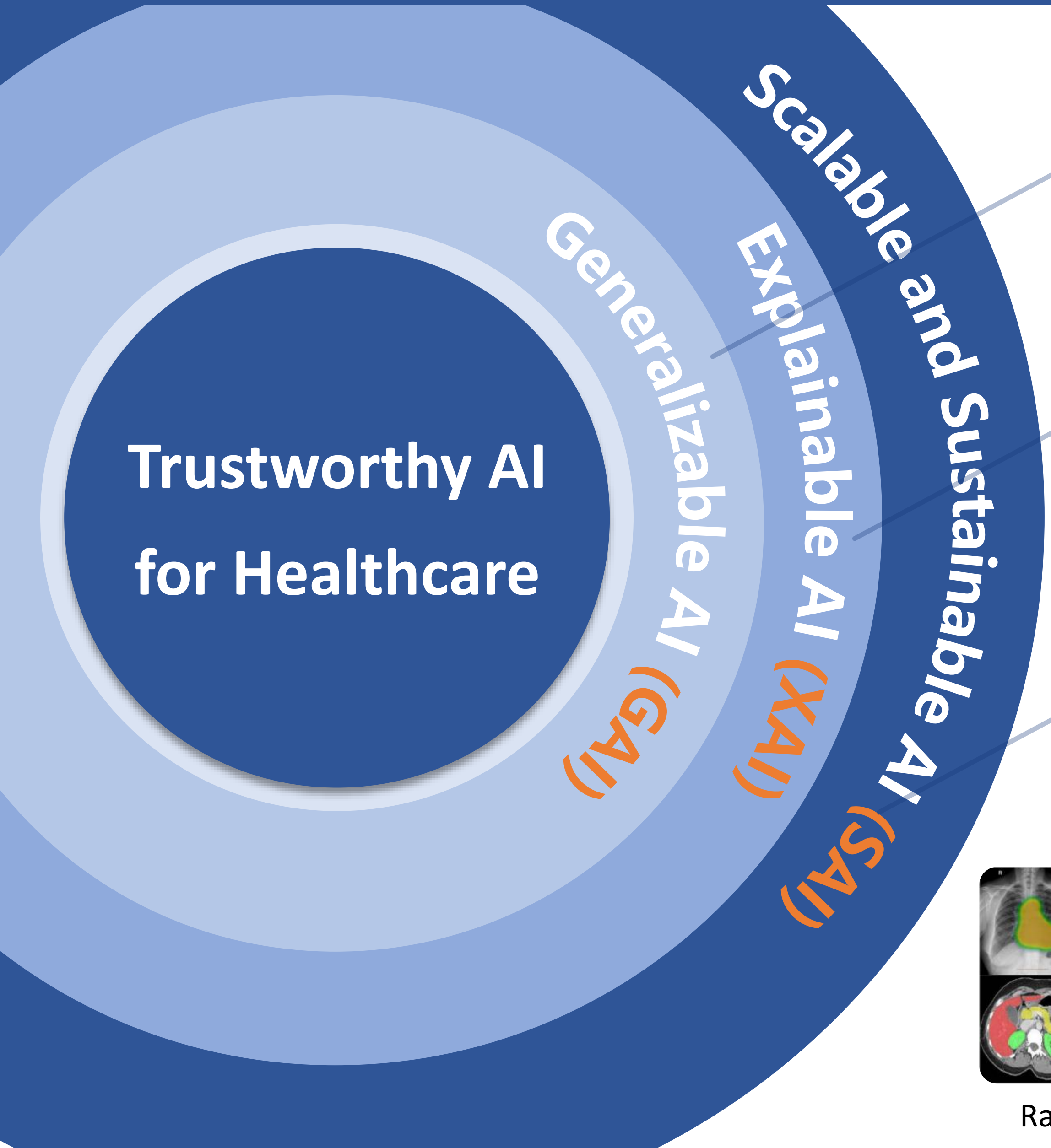


Cancer Diagnosis and Prognosis

Bigger Data, Larger Model



Smart Lab: Trustworthy AI for Healthcare



Multimodal foundation model

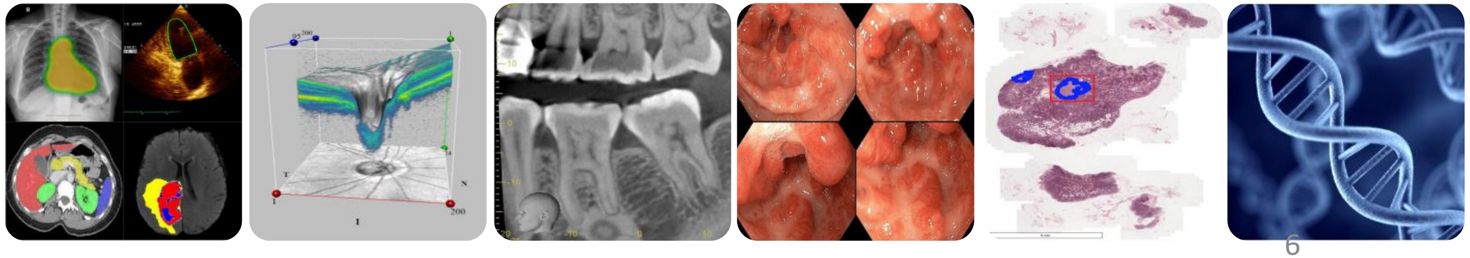
Can we use one general model for different modalities and diverse tasks?

Interpretable model

How to provide explanations for a decision-making process, thus enhancing the trust and confidence of doctors and patients?

Scalable and sustainable model

How to deploy general models to clients with different computing resources?

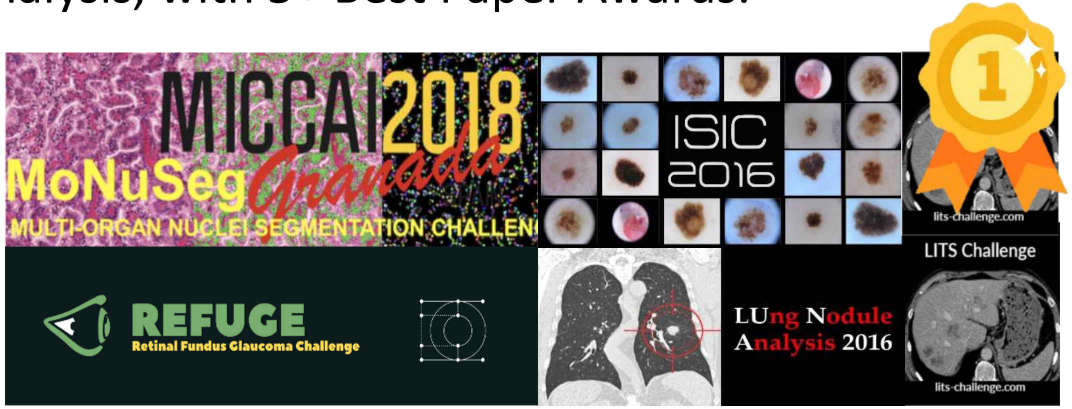


Radiology Ophthalmology Dentistry Surgery & Endoscopy Pathology Genomics

Artificial Intelligence for Healthcare

- We have achieved **state-of-the-art results** on **15+** international grand medical challenges.
- **100+** top-tier publications (e.g., IEEE TMI, MedIA, CVPR, MICCAI, ICCV, JAMA, Lancet Digital Health; Google Scholar Citations **24K+, h-index 63**) in AI for multimodal analysis, with **5+** Best Paper Awards.

Best Paper Awards and Championships



Winners of 15+ Grand Medical Challenges

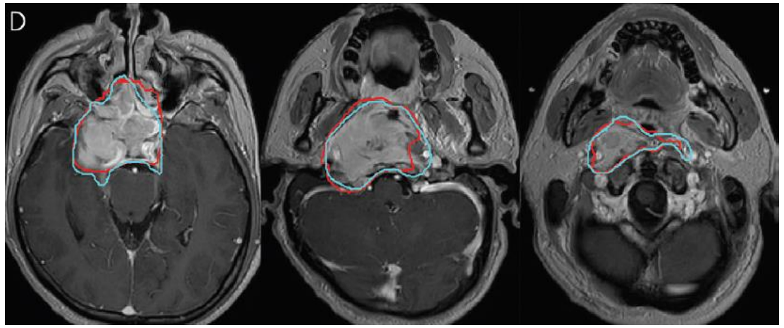
2019 MICCAI Young Scientist Impact Award

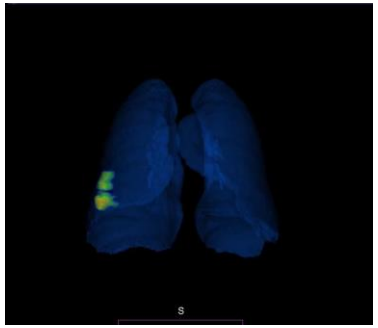
2023 Asian Young Scientist Fellow

2022&2023 Top 2% of the World's Top Scientists
2023 OMIA-X Prestigious Achievement Award
Ministry of Education Higher Education Outstanding Scientific Research Output Awards
国家教育部高等学校科学研究优秀成果奖
The First Prize of Beijing Science&Technology Award
北京市科技进步一等奖
World AI Conference Superior AI Leader (SAIL) Award
Forbes China 30 under 30

Artificial Intelligence for Healthcare

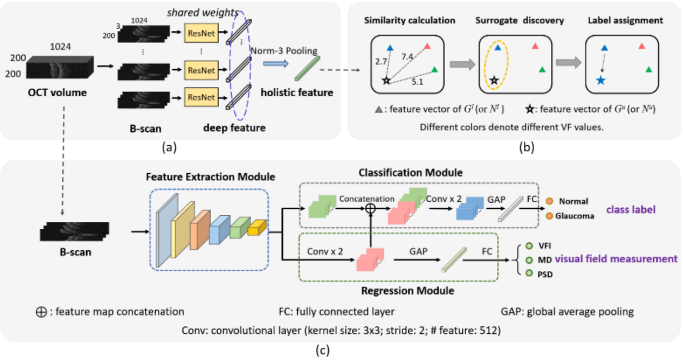
Radiology & Radiotherapy

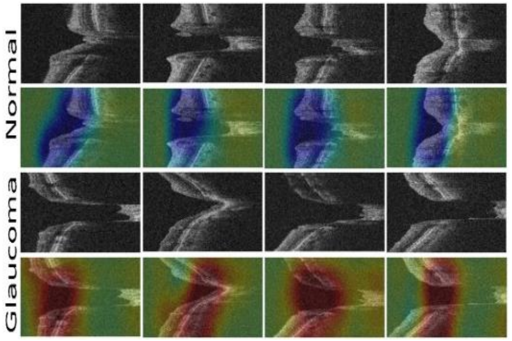




Disease Diagnosis, Quantitative Evaluation, and Radiotherapy
[Radiology 2019, TMI 2020, MICCAI 2021&2022, Rad AI 2021&2023]

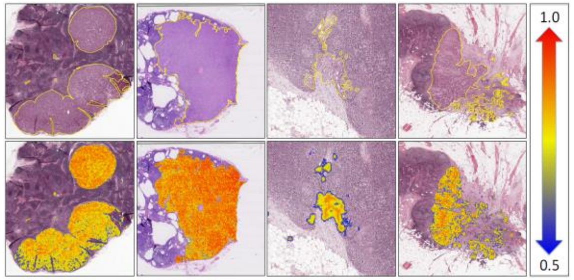
Ophthalmology

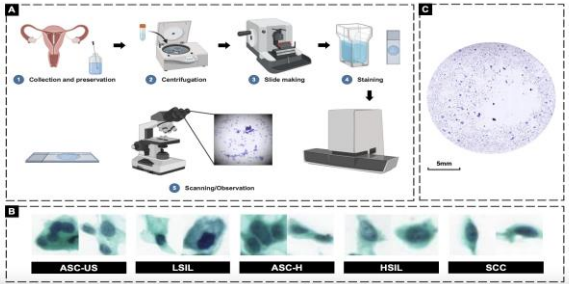




Ophthalmology Diease Screening including Glaucoma, DME, etc.
[Lancet Digital Health 2019 Cover Page, MIA 2023, CVPR 2023]

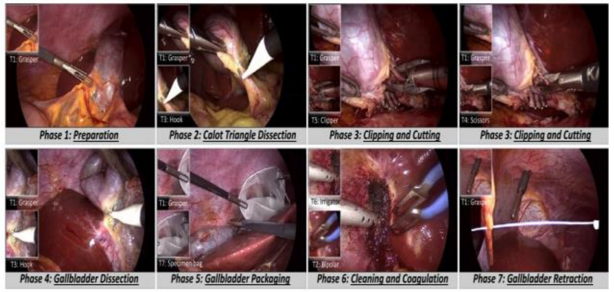
Computational Pathology

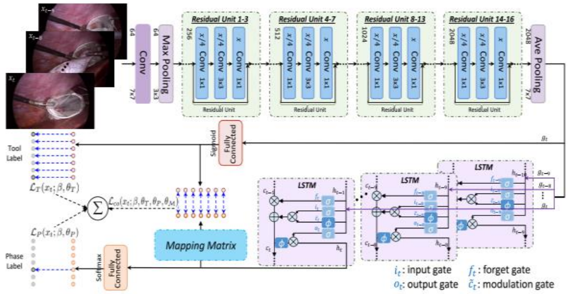




Cancer Screening and Analysis from WSI
[JAMA 2017, TMI 2019, CVPR 2023, IJCAI 2023, MIA 2023]

Surgery & Endoscopy





Surgical Data Science
[JBHI 2016, TMI 2017, MIA 2020, Two winners of MICCAI challenges]

Multimodal Foundation Model for Healthcare

Multimodal self-supervised training

Text | Video | EHRs | Omics
Images | Audio | Signals | Graphs

Medical domain knowledge

Literature | Publications | Clinical notes | Knowledge graphs

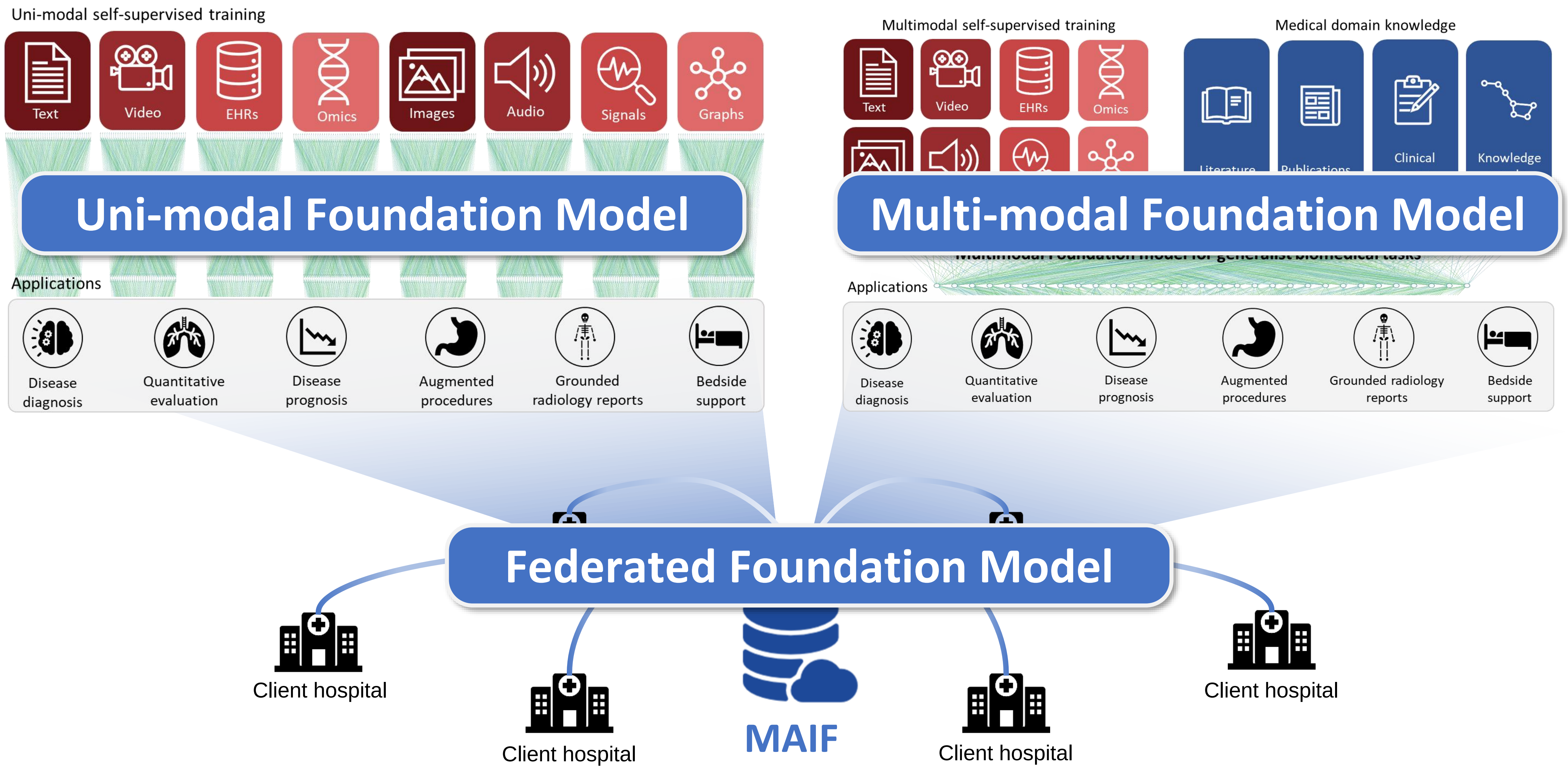
Applications

Disease diagnosis | Quantitative evaluation | Disease prognosis | Augmented procedures | Grounded radiology reports | Bedside decision support

- **Can we have one generalist model for different modalities and diverse tasks?**
- **The generalist model is then finetuned on the target modality and specific task to obtain a specialist model.**
- **Medical domain knowledge should be incorporated to enhance the specialist model.**

[1] Moor et al. Foundation models for generalist medical artificial intelligence. Nature 2023.
[2] Tu et al. Towards Generalist Biomedical AI. arXiv 2023.

Multimodal Foundation Model for Healthcare

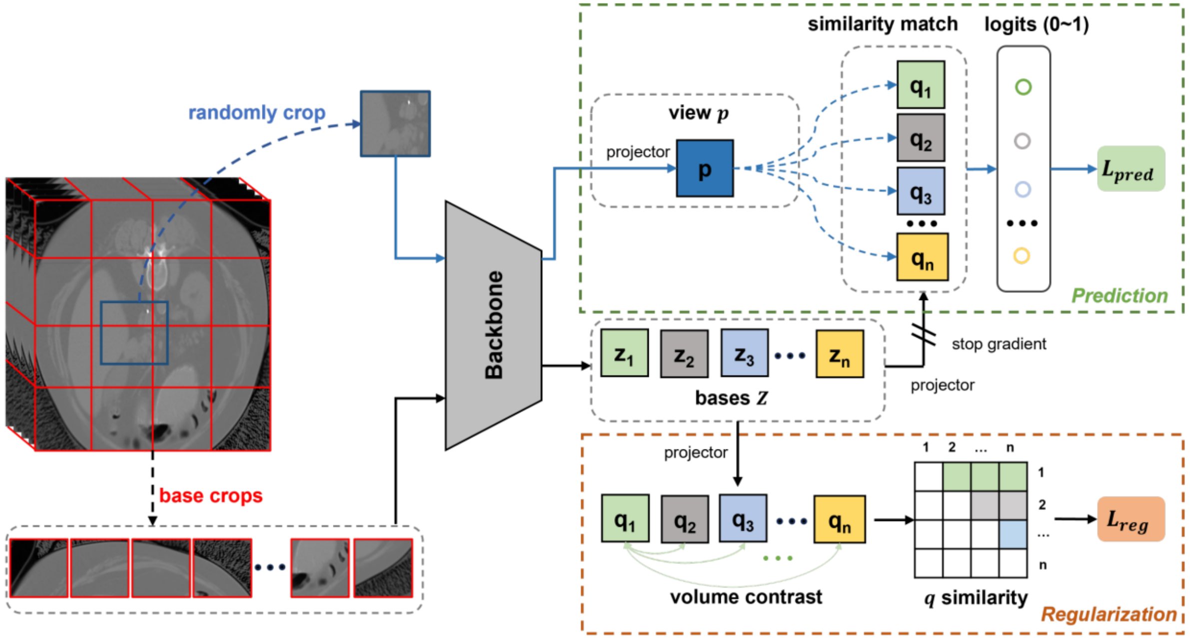


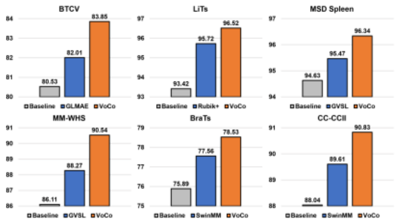
[1] Moor et al. Foundation models for generalist medical artificial intelligence. Nature 2023.
[2] Deng et al. Scale Federated Learning for Label Set Mismatch in Medical Image Classification. MICCAI 2023

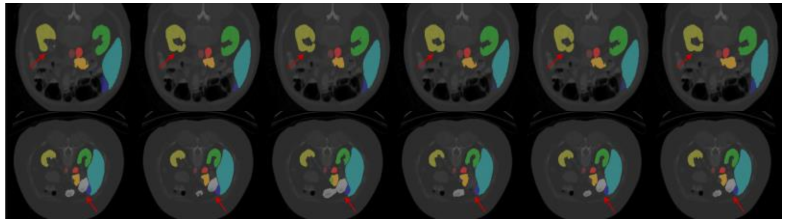
Radiology Foundation Model

❑ Volume Contrastive Learning for 3D Medical Image Analysis

- A self-supervised learning framework leverages the contextual position priors for pre-training.

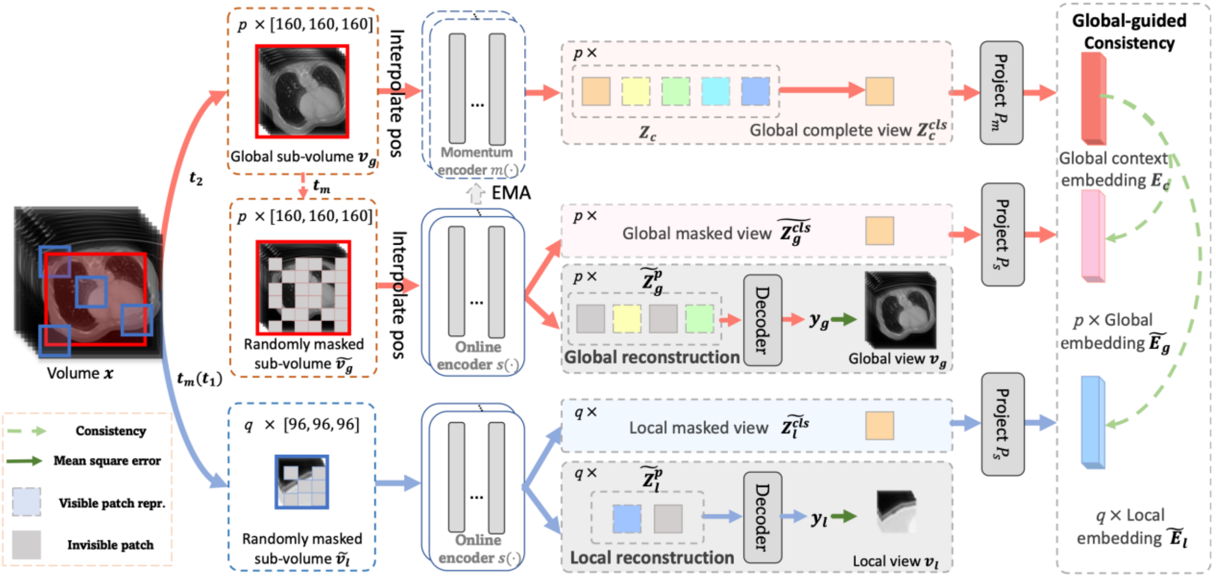


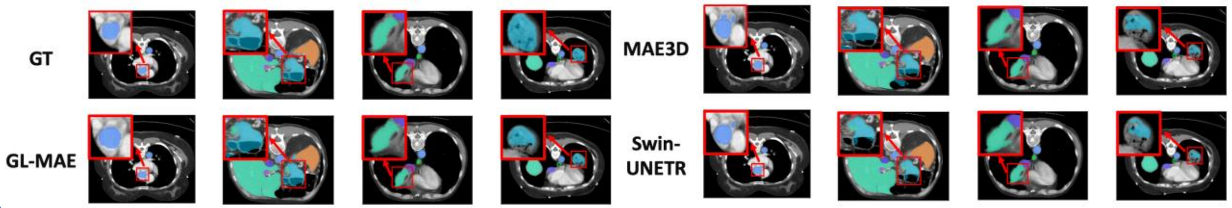




❑ Global-Local MAE for Volumetric Medical Image Analysis

- A 3D Mask Autoencoder with global and local reconstruction and global-guided consistency learning.



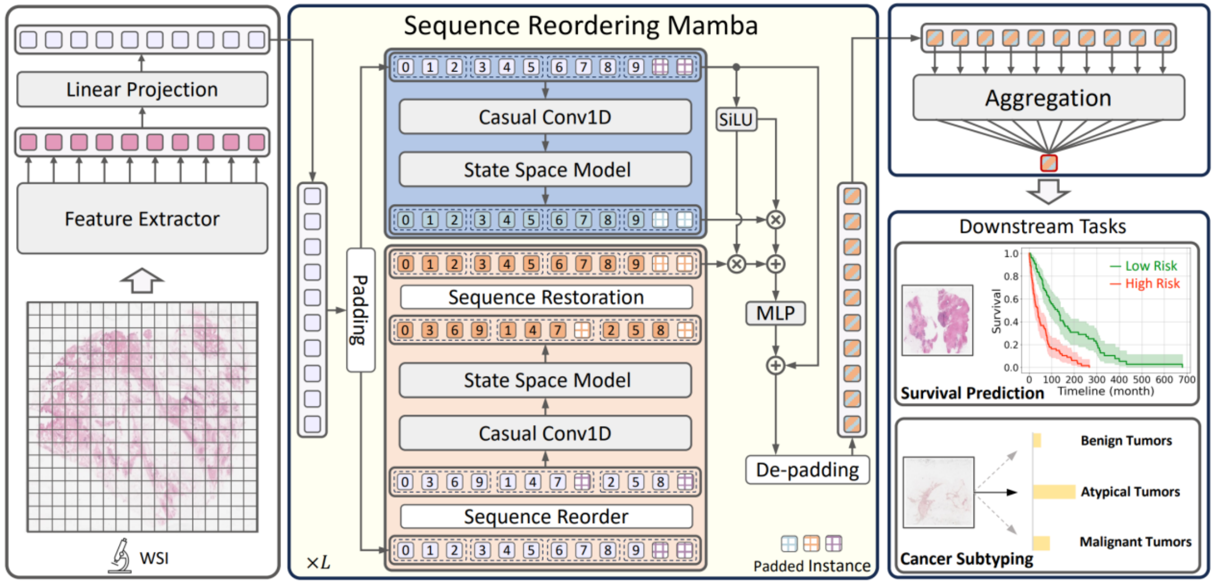


[1] Wu et al. VoCo: A Simple-yet-effective Volume Contrastive Learning Framework for 3D Medical Image Analysis. CVPR 2024.
[2] Zhuang et al. Advancing Volumetric Medical Image Segmentation via Global-Local Masked Autoencoder. arXiv 2023.

Pathology Foundation Model

❑ Enhance Long Sequence Modeling in Pathology

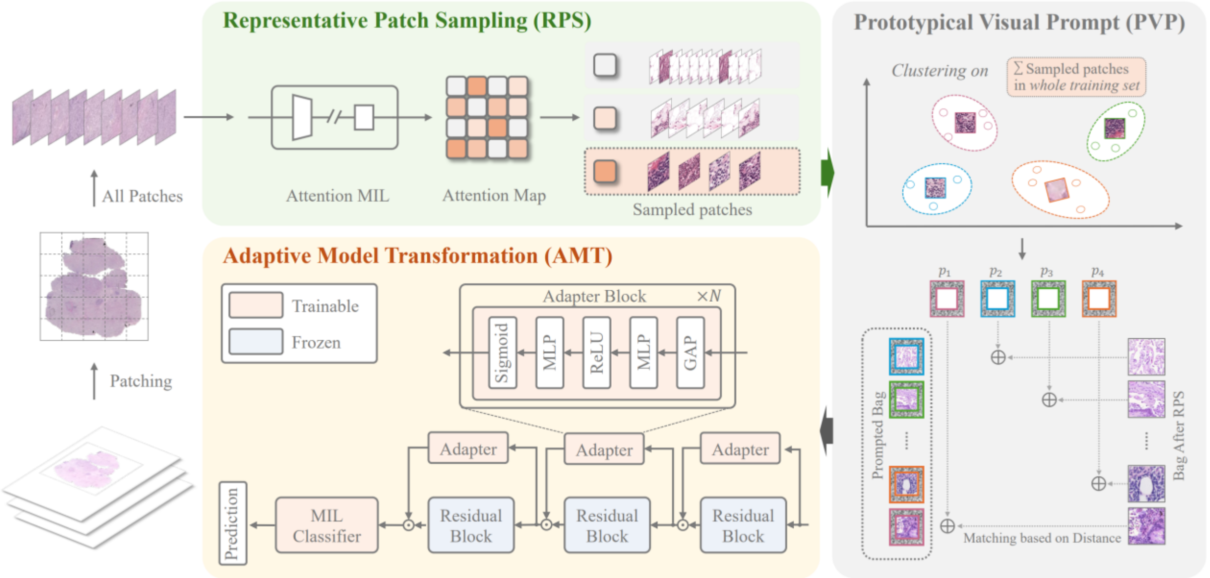
- Mamba-based framework in MIL via long sequence modeling to capture long-range dependencies.

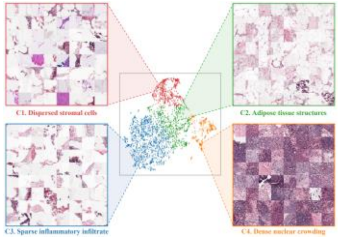


Method \ Dataset	BLCA	BRCA	COADREAD	KIRC	KIRP	LUAD	STAD	MEAN
ResNet-50								
Mamba	0.622±0.053	0.664±0.034	0.650±0.066	0.700±0.058	0.734±0.062	0.643±0.027	0.621±0.056	0.662
Bi-Mamba	0.647±0.024	**0.675±0.065**	0.662±0.058	0.690±0.048	0.737±0.052	0.628±0.059	0.622±0.068	0.665
SR-Mamba	**0.652±0.028**	0.673±0.063	**0.671±0.066**	**0.721±0.064**	**0.748±0.094**	**0.653±0.059**	**0.639±0.076**	**0.680**

❑ Prompt-Guided Model Adaptation for WSI Diagnosis

- A novel prompt-guided adaptive model transformation framework that enhances MIL classification on WSI.





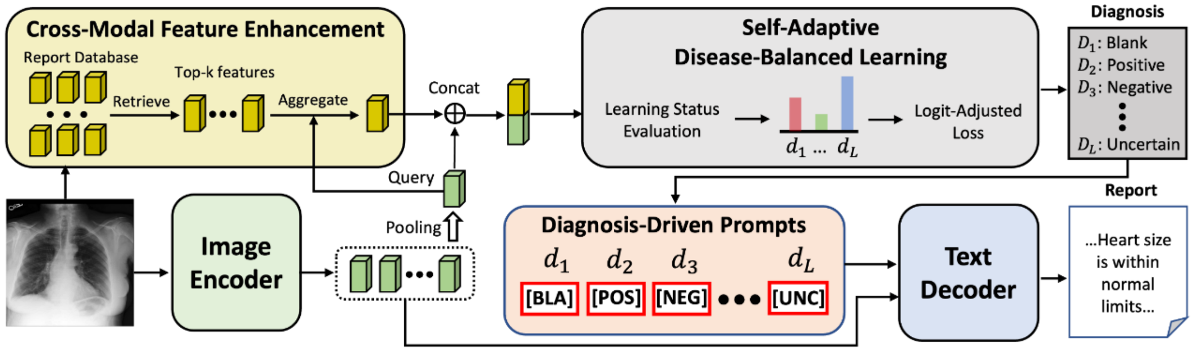
Method		Camelyon16 AUC	Camelyon16 F1	Camelyon16 Acc	TCGA-NSCLC AUC	TCGA-NSCLC F1	TCGA-NSCLC Acc
PAMT	Mean-pooling	$0.810_{0.010}$	$0.714_{0.013}$	$0.803_{0.017}$	$0.925_{0.011}$	$0.874_{0.014}$	$0.871_{0.010}$
	Max-pooling	$0.890_{0.015}$	$0.777_{0.016}$	$0.824_{0.025}$	$0.926_{0.013}$	$0.873_{0.011}$	$0.870_{0.013}$
	AB-MIL [9]	$0.921_{0.011}$	$0.828_{0.018}$	$0.866_{0.008}$	$0.931_{0.009}$	$0.884_{0.007}$	$0.879_{0.011}$
	CLAM-SB [15]	$0.924_{0.009}$	$0.832_{0.007}$	$0.881_{0.006}$	$0.933_{0.011}$	$0.891_{0.006}$	$0.888_{0.005}$
	DTFD-MIL [22]	$\mathbf{0.929}_{0.010}$	$\mathbf{0.859}_{0.007}$	$\mathbf{0.897}_{0.009}$	$\mathbf{0.942}_{0.008}$	$\mathbf{0.895}_{0.013}$	$\mathbf{0.893}_{0.014}$

[1] Yang et al. MambaMIL: Enhancing Long Sequence Modeling with Sequence Reordering in Computational Pathology. arXiv 2024.
[2] Lin et al. Prompt-Guided Adaptive Model Transformation for Whole Slide Image Classification. arXiv 2024.

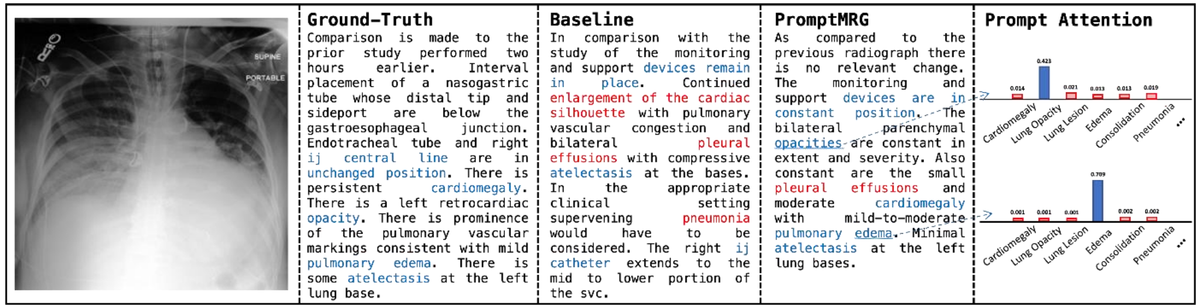
Vision-Language Model for Report Generation

Diagnosis-driven Prompts for Report Generation

- Diagnosis-driven prompts for medical report generation with cross-modal feature enhancement and self-adaptive disease-balanced learning.

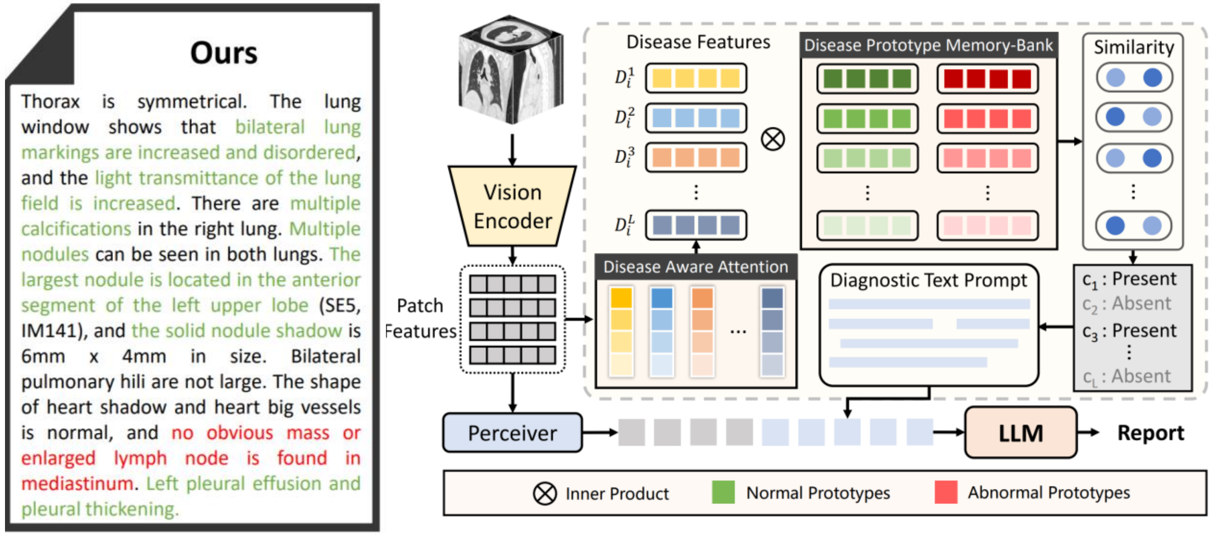


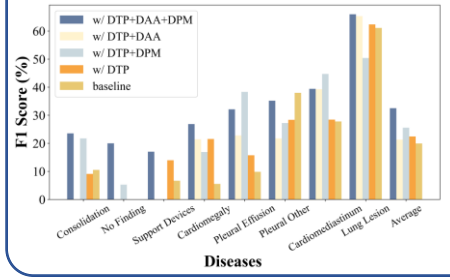
The proposed PromptMRG covers most key descriptions



Large Language Model-driven CT Report Generation

- Adapt LLaMA2-7B for CT report generation via the disease prototype memory-bank and incorporation of diagnostic information.





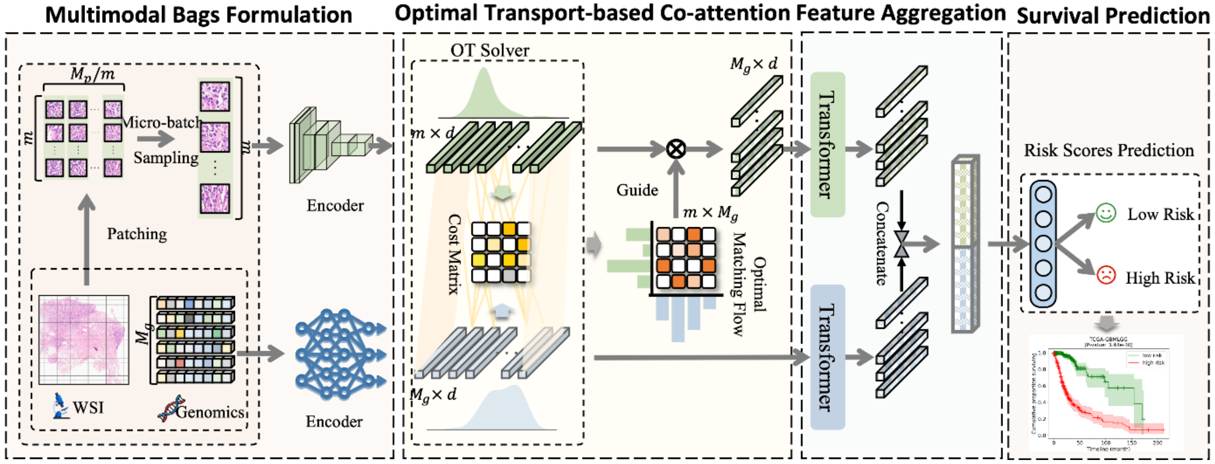
METHOD	YEAR	CE Metrics			NLG Metrics			
		Pre.	Rec.	F1	BL-1	BL-4	MTR	RG-L
R2Gen [3]	2020	0.207	0.121	0.144	34.11	23.39	21.40	**47.75**
R2GenCMN [2]	2022	0.158	0.100	0.114	35.88	23.37	21.43	45.94
M2KT [23]	2023	0.220	0.119	0.145	46.09	21.93	25.20	36.47
PromptMRG [8]	2023	0.290	0.330	0.290	47.73	23.02	22.87	37.35
SL-DG* [19]	2024	-	-	-	-	23.70	21.90	43.80
RadFM [22]	2023	0.403	0.361	0.345	46.70	24.70	24.01	38.98
Ours	-	**0.421**	**0.387**	**0.372**	**51.16**	**29.64**	**26.28**	42.15

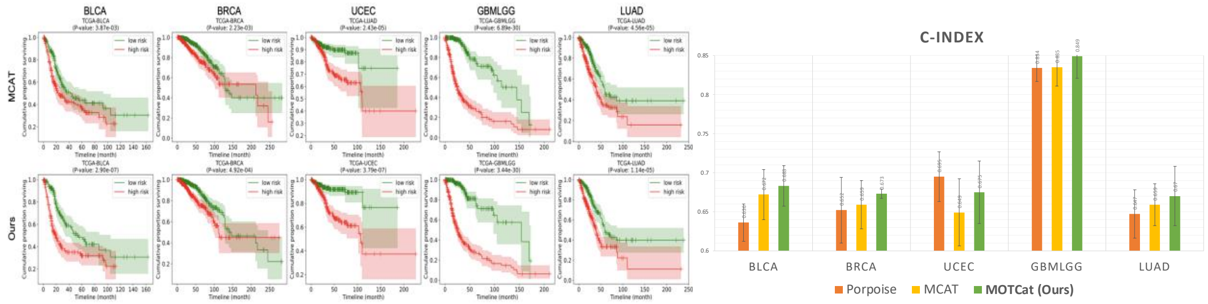
[1] Jin et al. PromptMRG: Diagnosis-Driven Prompts for Medical Report Generation. AAAI 2023.
[2] Chen et al. Dia-LLaMA: Towards Large Language Model-driven CT Report Generation. arXiv 2024.

Multimodal Fusion for Precision Oncology

❑ Optimal Transport-based Co-Attention Transformer

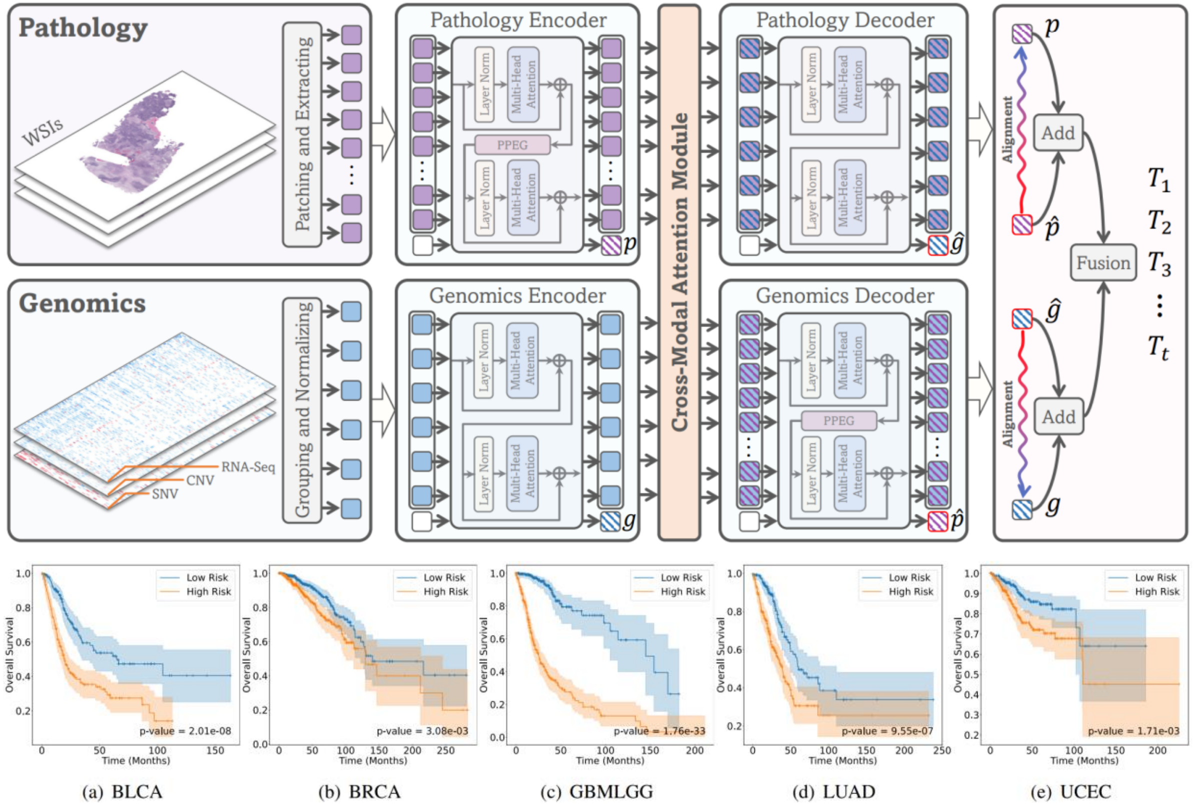
- Construct an optimal matching solution with the overall minimum matching cost for histology and genomics alignment.





❑ Cross-Modal Translation and Alignment

- Integrate intra-modal information and generate cross-modal representations on pathological images and genomic profiles.

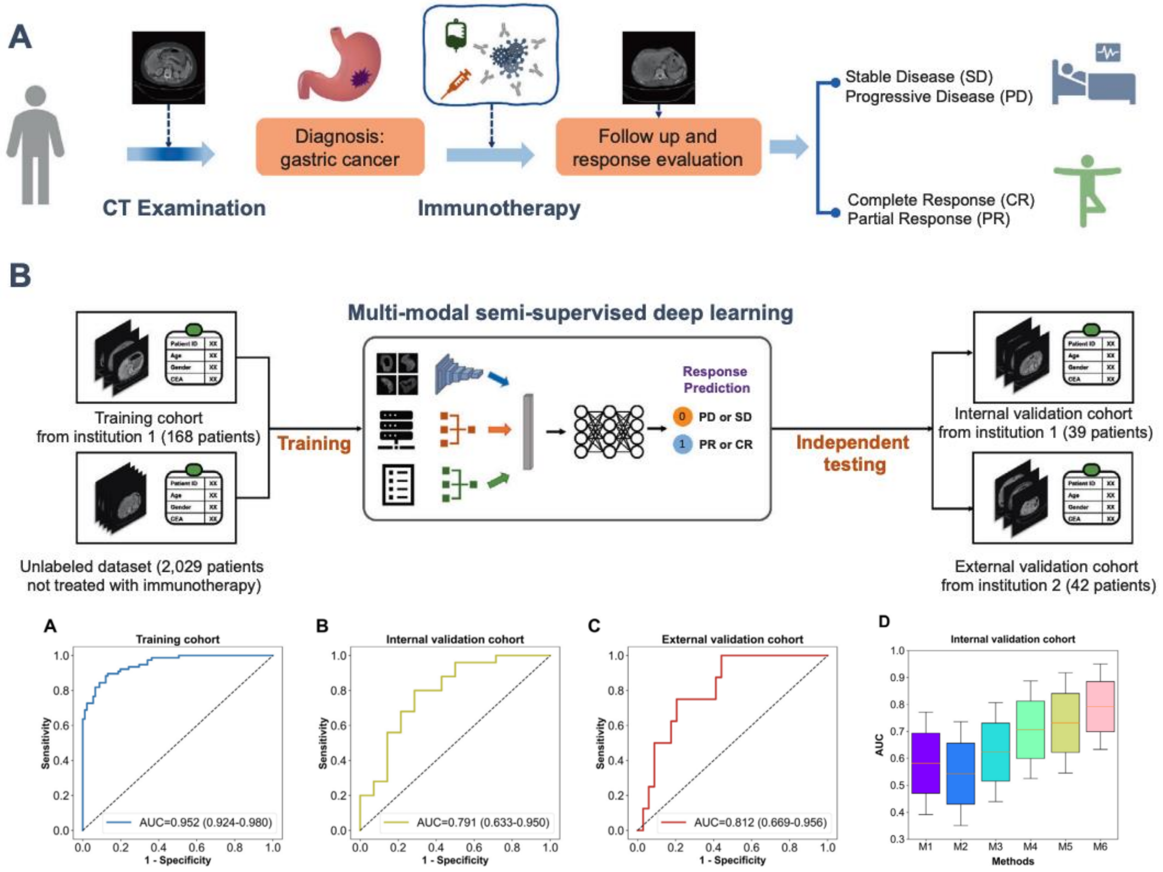


[1] Xu et al. Multimodal Optimal Transport-based Co-Attention Transformer with Global Structure Consistency for Survival Prediction. ICCV 2023.
[2] Zhou et al. Cross-Modal Translation and Alignment for Survival Analysis. ICCV 2023.
[3] Zhang et al. Prototypical Information Bottlenecking and Disentangling for Multimodal Cancer Survival Prediction. ICLR 2024.

Multimodal Fusion for Precision Oncology

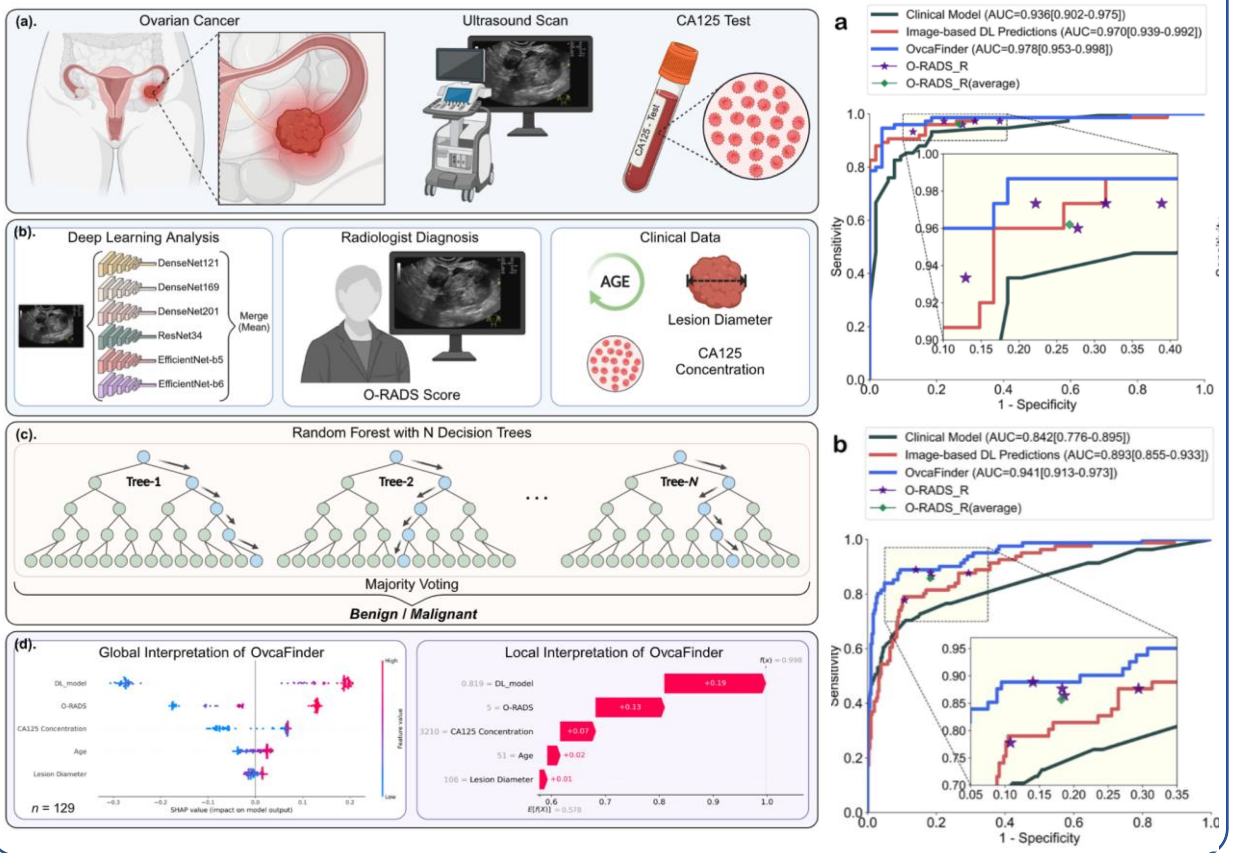
❑ Multimodal Cancer Immunotherapy Response Prediction

- Using multi-modal clinical and image data for predicting immunotherapy response.



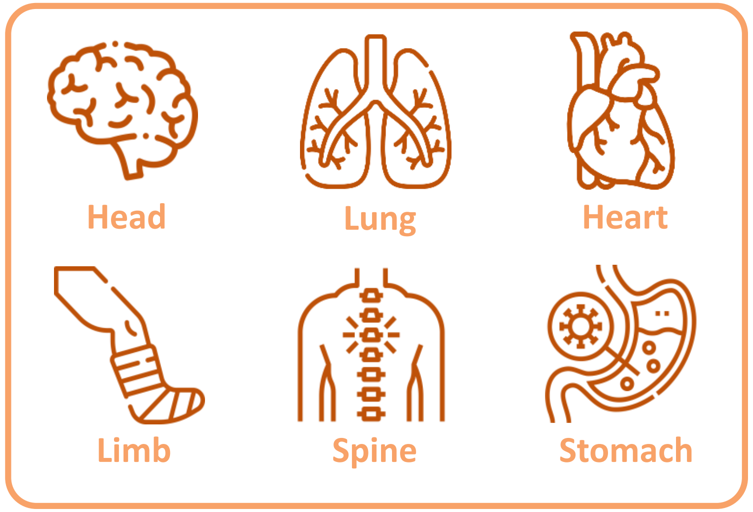
❑ Multimodal Information for Ovarian Cancer Diagnosis

- Develop and validate the OvcaFinder to discriminate benign from ovarian cancer via a multimodal AI model.

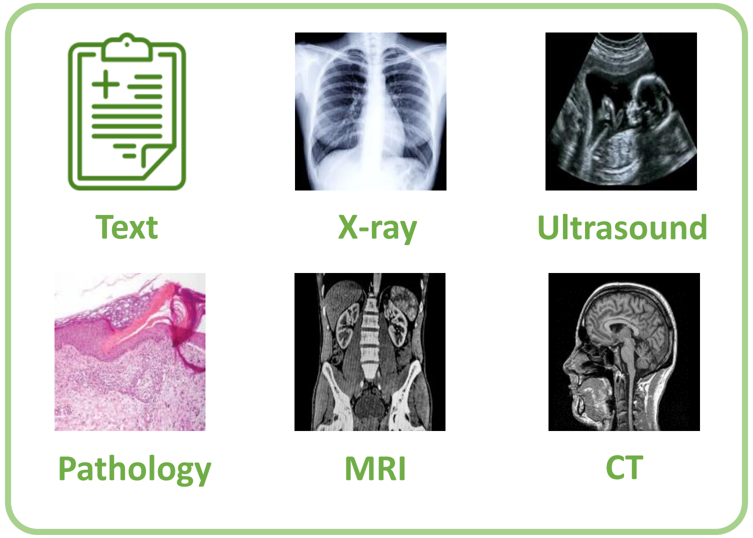


[1] Wang et al. Cancer immunotherapy response prediction from multi-modal clinical and image data using semi-supervised deep learning. Radiotherapy and Oncology 2023.
[2] Xiang . et al. Development and validation of an interpretable model integrating multimodal information for improving ovarian cancer diagnosis. Nature Communications 2024.

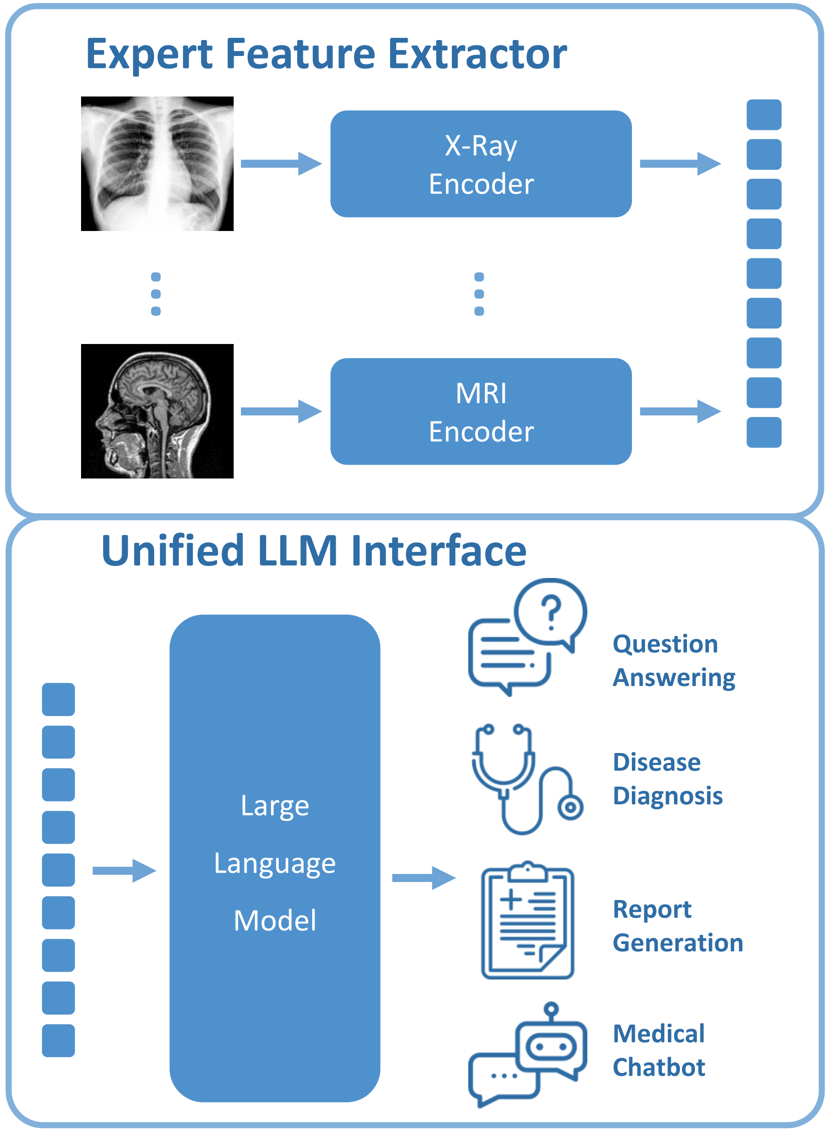
Generalist Foundation Model



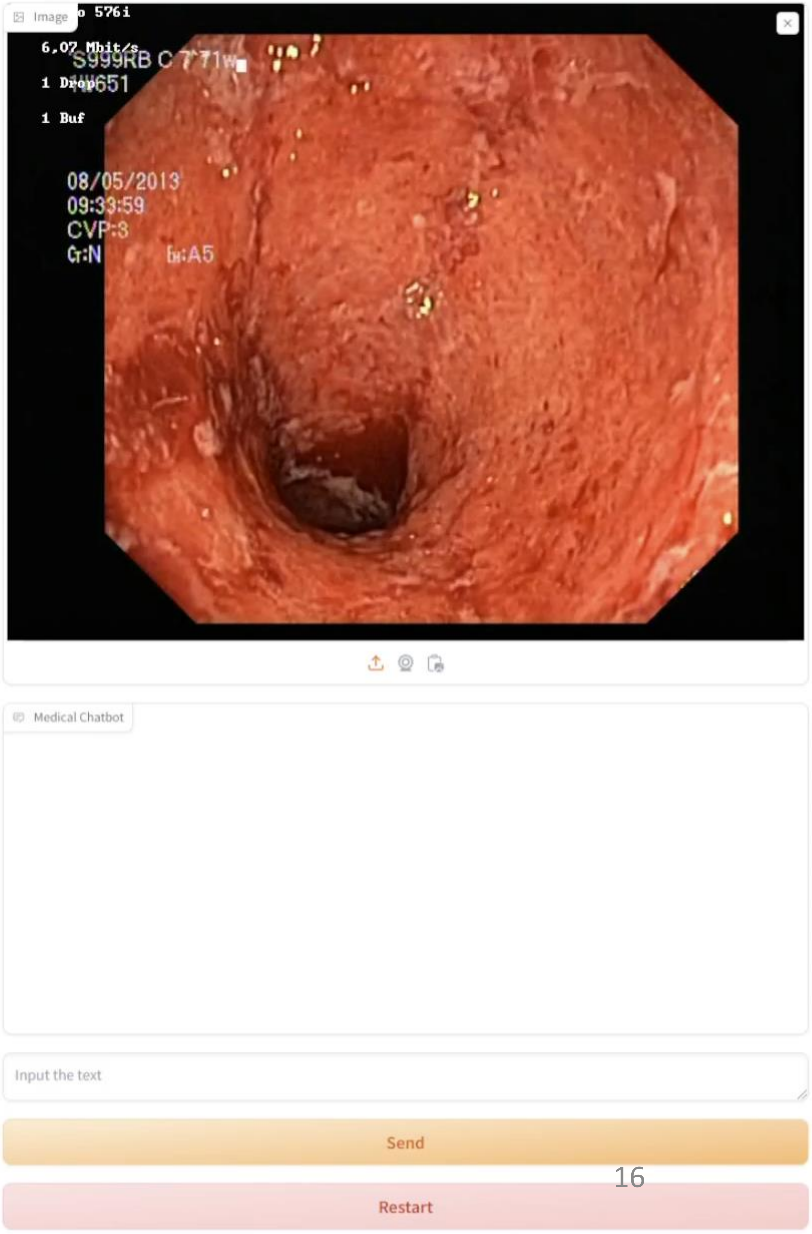
(a) Anatomical Regions.



(b) Multimodal Medical Data.



(c) Multimodal AI Foundation Model



Application: Smart Dental Design and Treatment

AI-Enabling Direct Surgical Execution

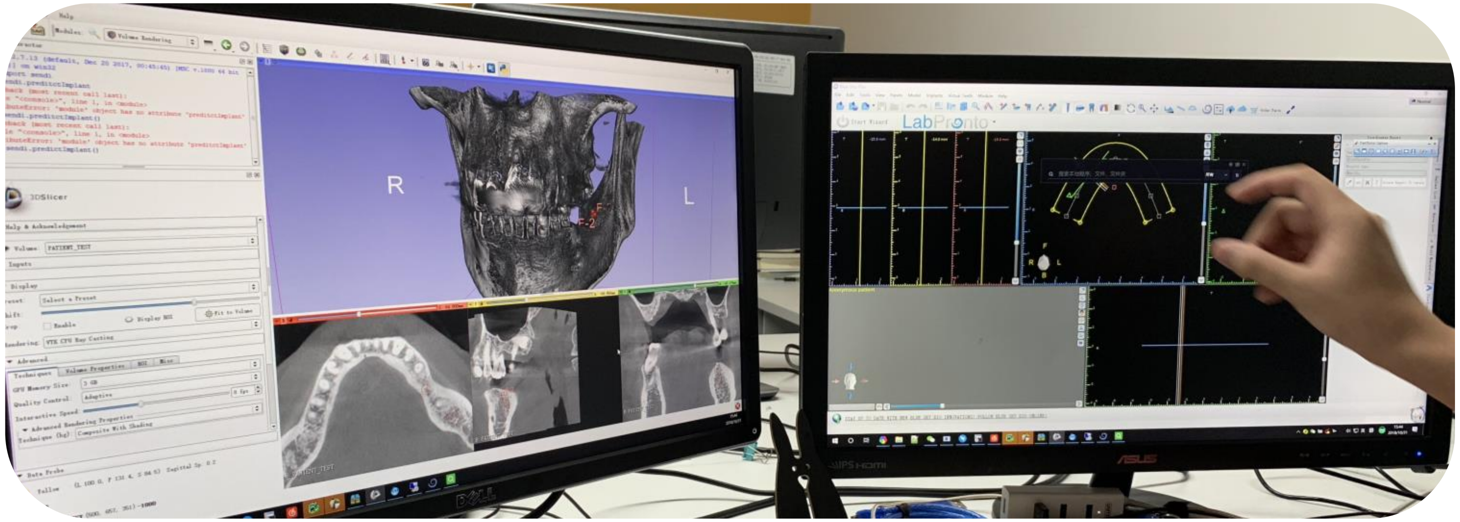
Cloud-based powerful AI dental treatment design system

Printing technology

AI-designed **Crown**

AI-designed **Surgical Guide** for **Implant** surgery execution

AI-designed **Clear Aligners** for **Orthodontic**



AI reduces manpower	Solution to human bottleneck	>50,000 sets of patient data training	Alleviate dental service supply shortage	Minimally invasive surgery
Generative AI generates instant dental treatments	Reduce cost of dental implants to <$2,000	Much more efficient and safer surgery	Instant restoration of teeth function	2 patents

HKU Dentistry
Institute for Advanced Dentistry
Global Dental Campus
先進牙醫學研究所
環球牙科校園

NYU DENTISTRY

DBT Technology

HK AI LAB

HKSTP

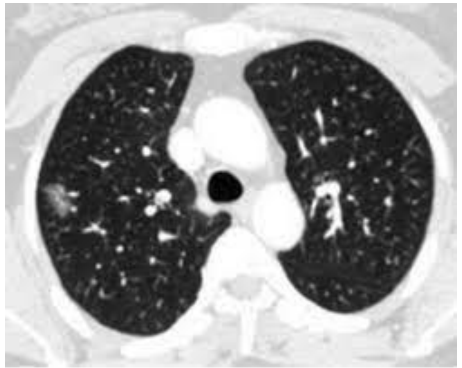
香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

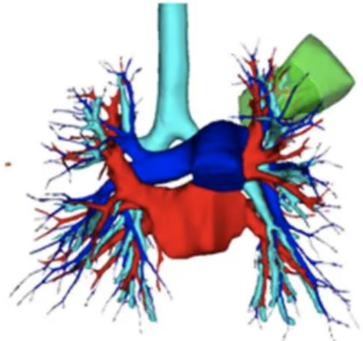
Application: Computer-assisted Intervention

Preoperative Planning and Intraoperative Navigation for Assisting Surgery

Preoperative

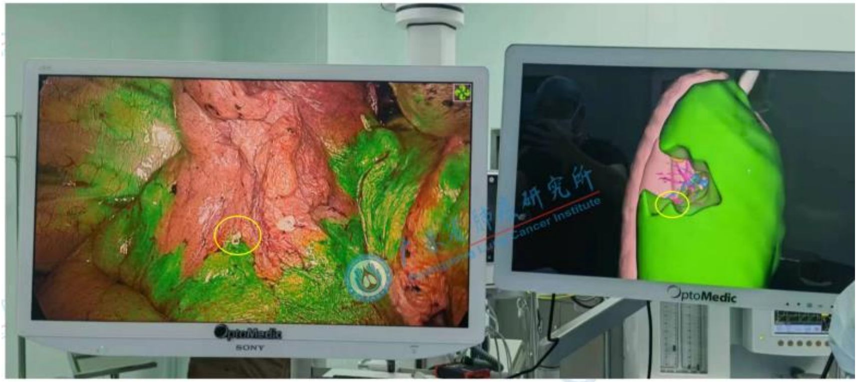
3D reconstruction





Dynamic registration

Intraoperative

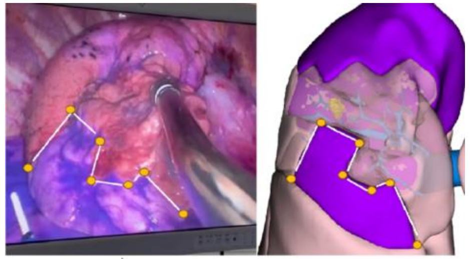


Computer-assisted Intervention

AR & VR visualization

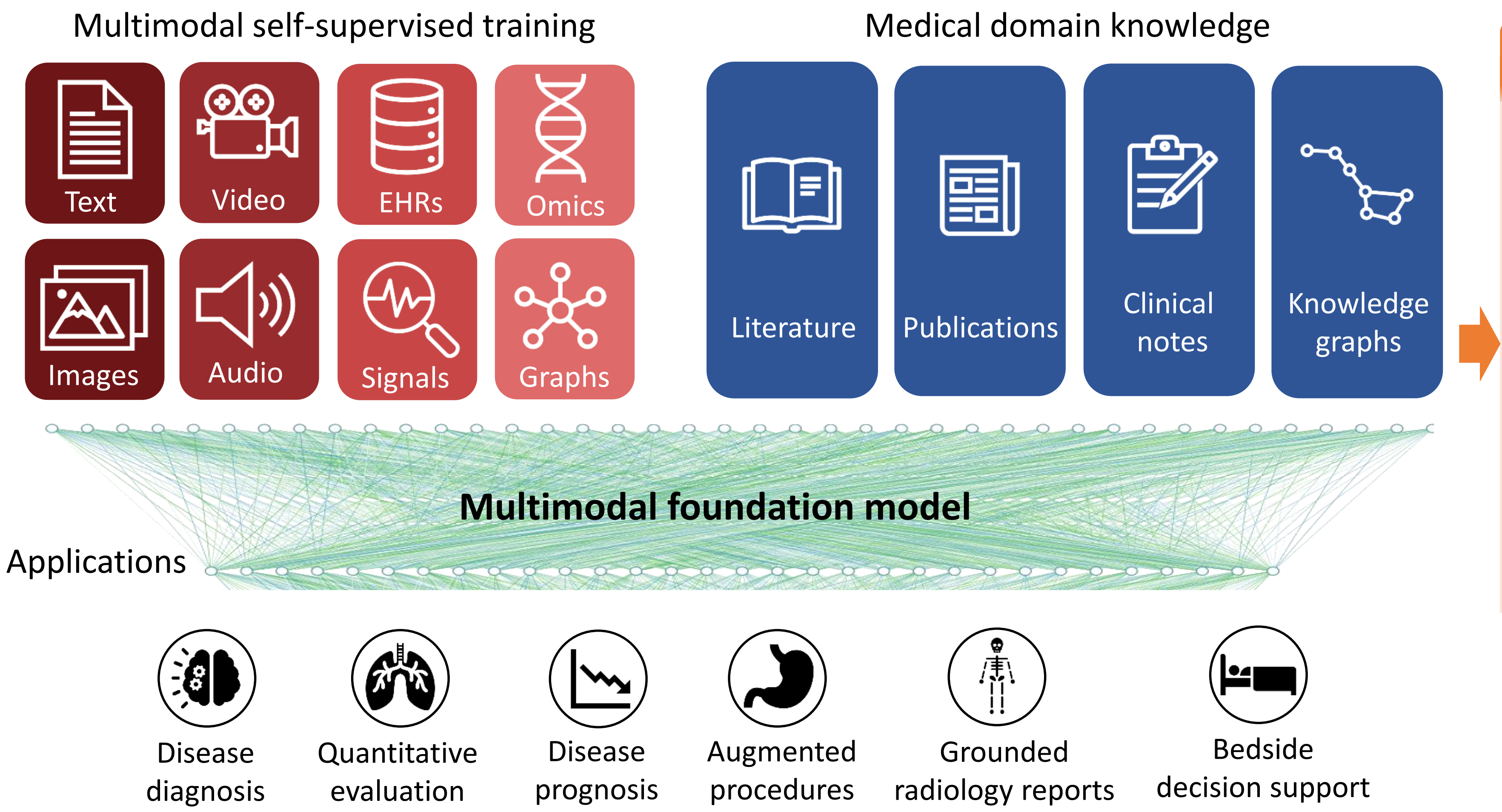
Real-time navigation

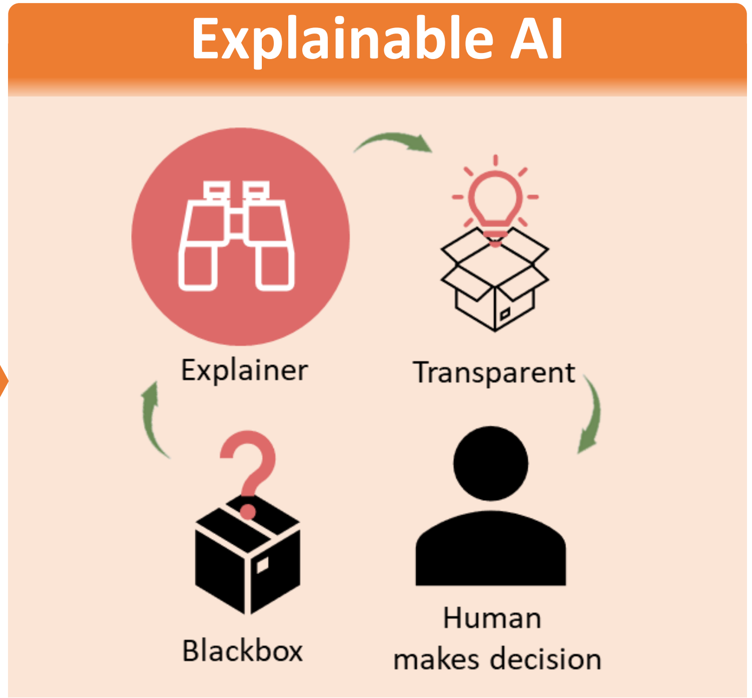
System integration



System assessment

Explainable AI (XAI)





Is the prediction trustable?
Which feature is important?
How to estimate uncertainty?

[1] Moor, et al. Foundation models for generalist medical artificial intelligence. Nature 2023.
[2] Tu et al. Towards Generalist Biomedical AI. arXiv 2023.
[3] He et al. Foundation Model for Advancing Healthcare: Challenges, Opportunities and Future Directions. arXiv 2024.

XAI Evaluation

Clinical Explainable AI Guidelines

Explainable AI algorithms

Guideline 1
Understandability
No technical knowledge is required to understand the explanation

Guideline 2
Clinical relevance
Explanation is relevant to clinical decision-making

Guideline 3
Truthfulness
Explanation should truthfully reflect model decision process

Guideline 4
Informative plausibility
Human judgment on explanation plausibility can reveal decision quality

Guideline 5
Computational efficiency
Computational speed is within clinical users' tolerable waiting time

FAST

Suitable for clinical use

Evaluation results on 16 heatmap methods

G1	G2	G3	G4	G5
Passed	Partially passed	Not passed	Not passed	Mostly passed

STOP The evaluated heatmap methods did not meet G3 and G4, thus cannot be recommended for clinical use.

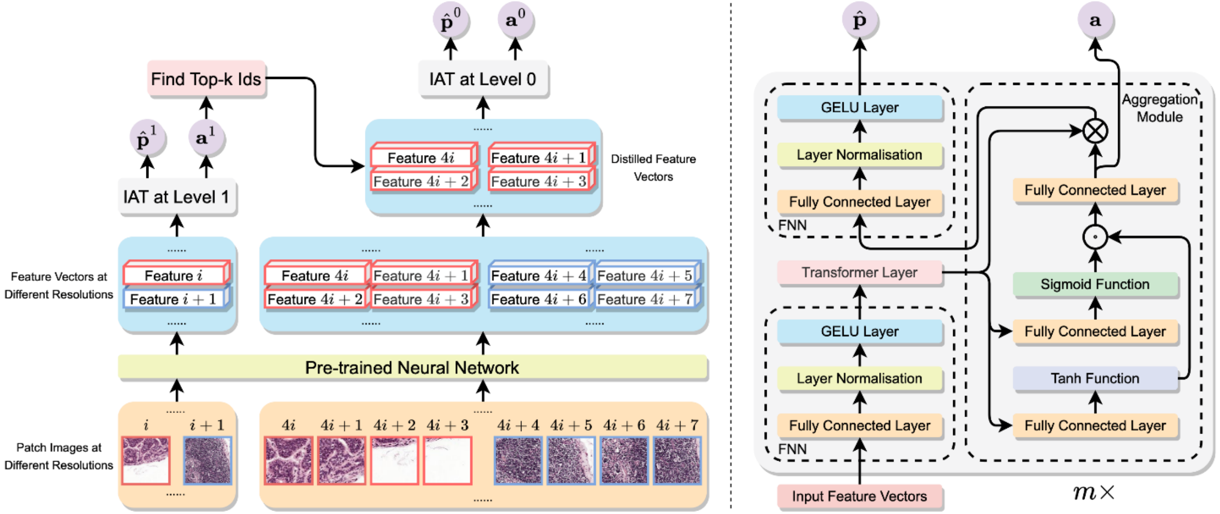
Most post-hoc heatmap-based methods fail in the clinical practice!

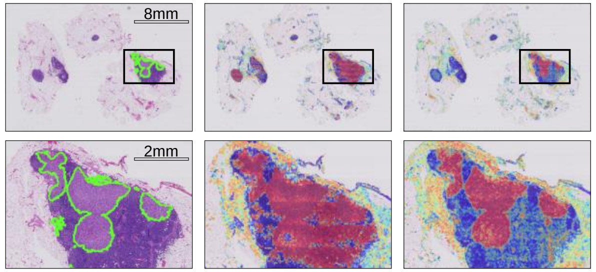
[1] Jin, Weina, et al. Guidelines and evaluation of clinical explainable AI in medical image analysis. MIA 2023.

XAI for Healthcare

❑ Visual Attention Interpretation for WSI Classification

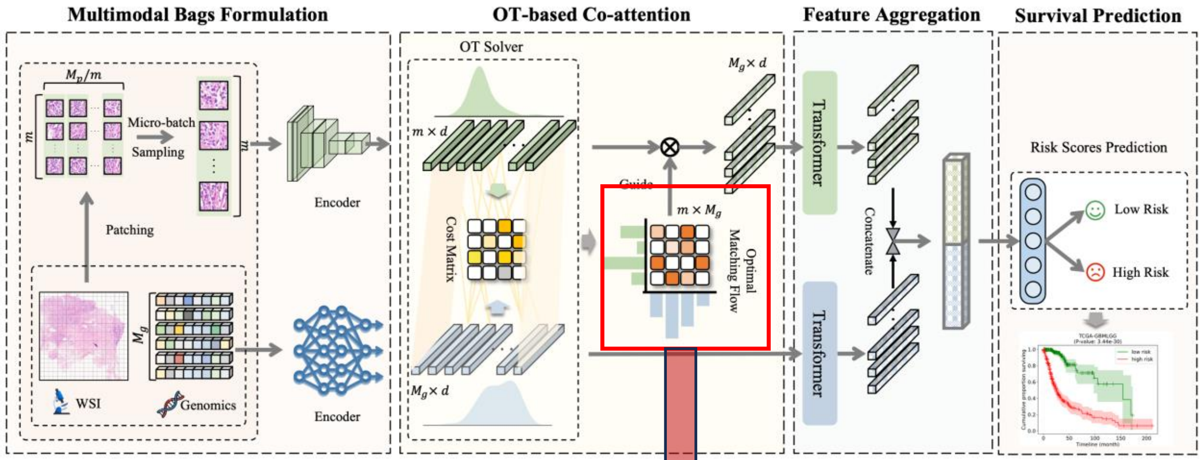
- Attention scores are used as a guide in higher resolutions, like a pathologist zooming in the regions of interests.



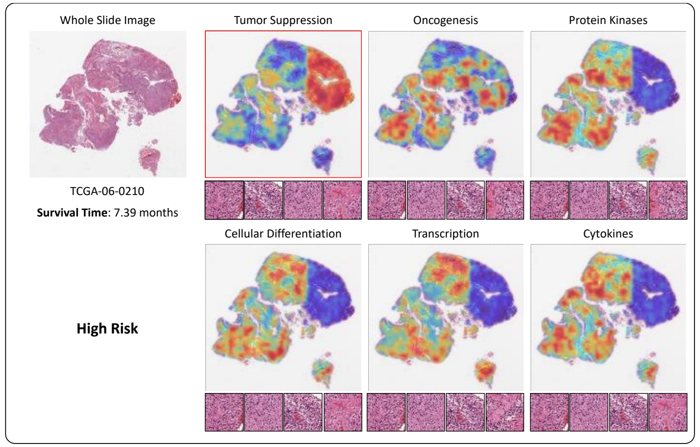


(a) The original image (b) CLAM heatmap (c) IAT heatmap

❑ Multimodal Intrinsic XAI for Survival Prediction



How pairwise instances from histology and genomics are matched?

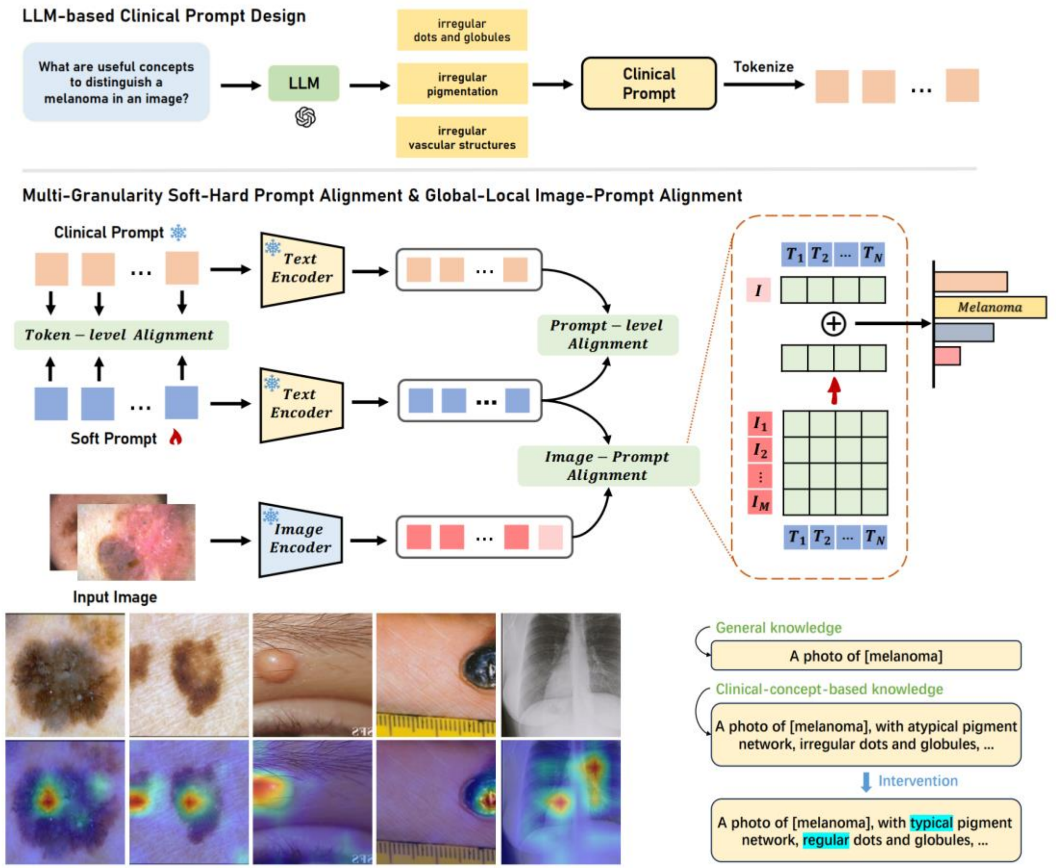


[1] Xiong, et al. Diagnose Like a Pathologist: Transformer-Enabled Hierarchical Attention-Guided Multiple Instance Learning for Whole Slide Image Classification. IJCAI 2023.
[2] Xu, et al. Multimodal optimal transport-based co-attention transformer with global structure consistency for survival prediction. ICCV 2023.

LLM-enhanced XAI in Healthcare

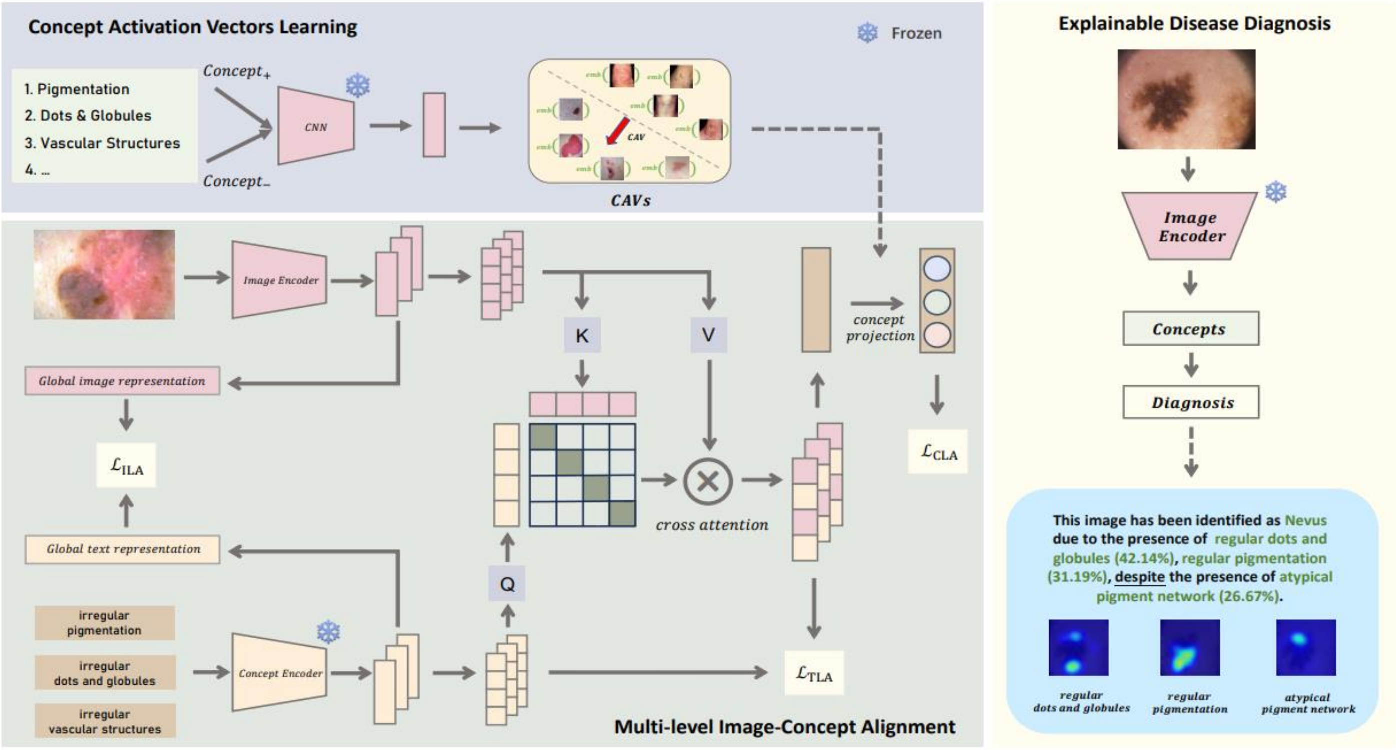
❑ LLM-based Prompt Learning

- Explainable prompt learning for computer-aided diagnosis via concept-guided context optimization.



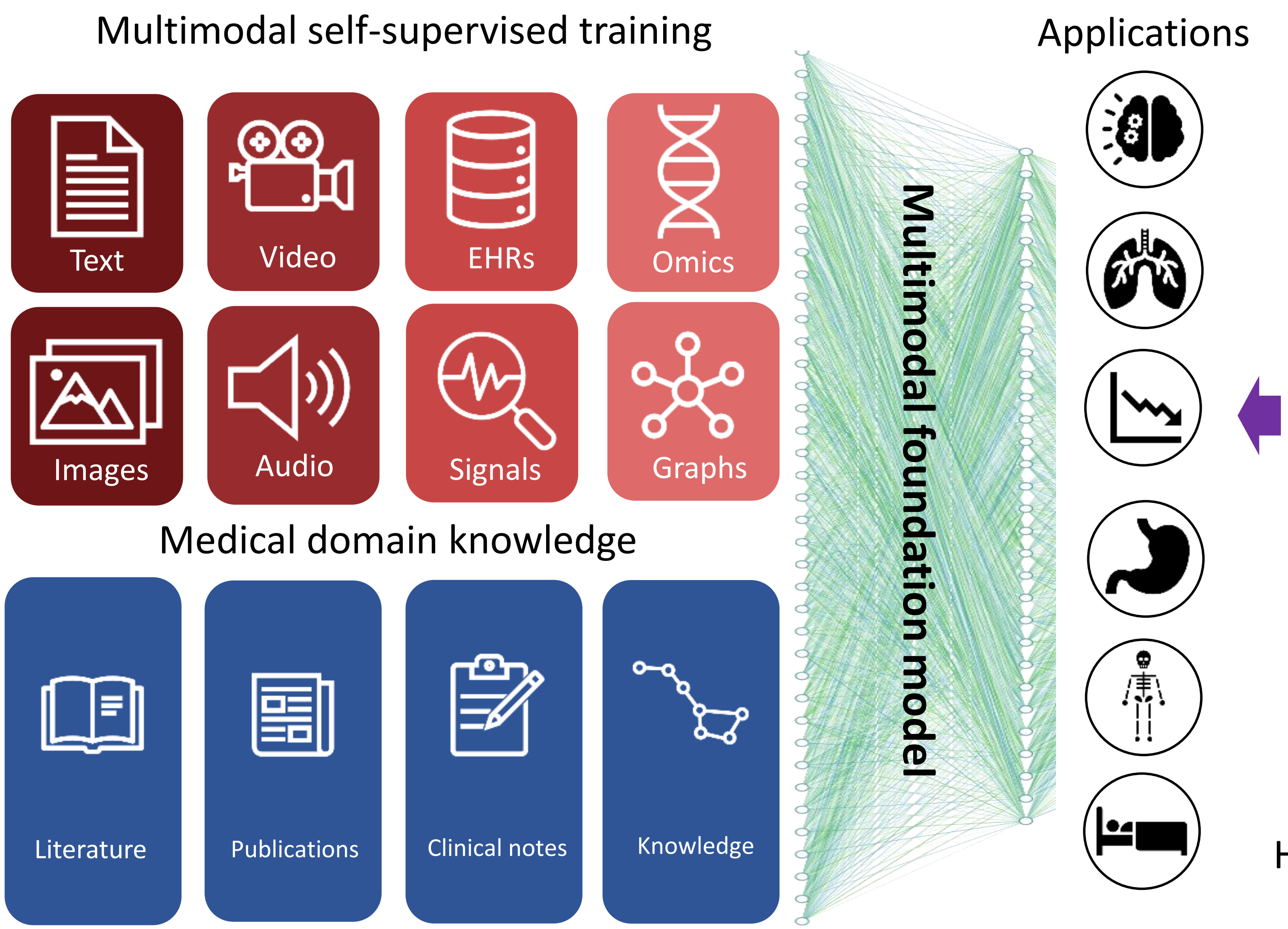
❑ Concept-based Learning

- A multimodal explainable disease diagnosis framework that meticulously aligns medical images and clinical-related concepts semantically at multiple levels.



[1] Bie, et al. XCoOp: Explainable Prompt Learning for Computer-Aided Diagnosis via Concept-guided Context Optimization. arXiv 2024.
[2] Bie, et al. MICA: Toward Explanable Skin Lesion Diagnosis via Multi-level Image-Concept Alignment. AAAI 2023.

Scalable and Sustainable AI (SAI)



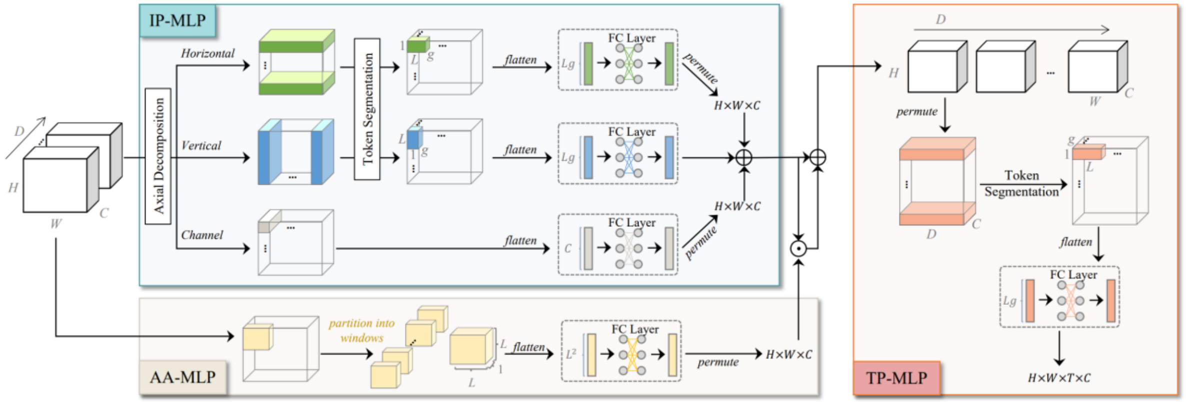
How to deploy systems in source-limited scenarios?
How to dynamically adapt to target domain?
How to sustainably learn the new knowledge?

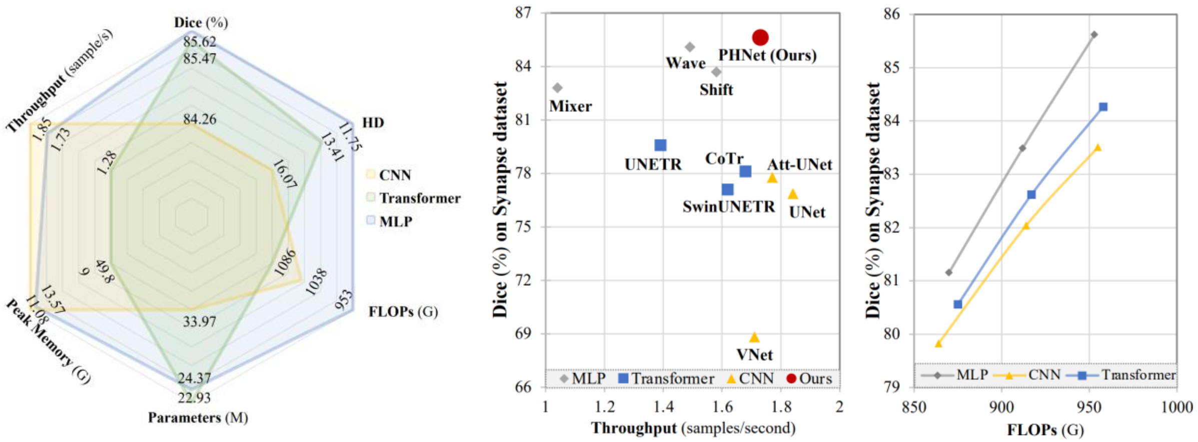
[1] Moor, et al. Foundation models for generalist medical artificial intelligence. Nature 2023.
[2] Tu et al. Towards Generalist Biomedical AI. arXiv 2023.
[3] He et al. Foundation Model for Advancing Healthcare: Challenges, Opportunities and Future Directions. arXiv 2024.

Scalable and Sustainable AI (SAI)

❑Efficient MLP-permutation for Segmentation

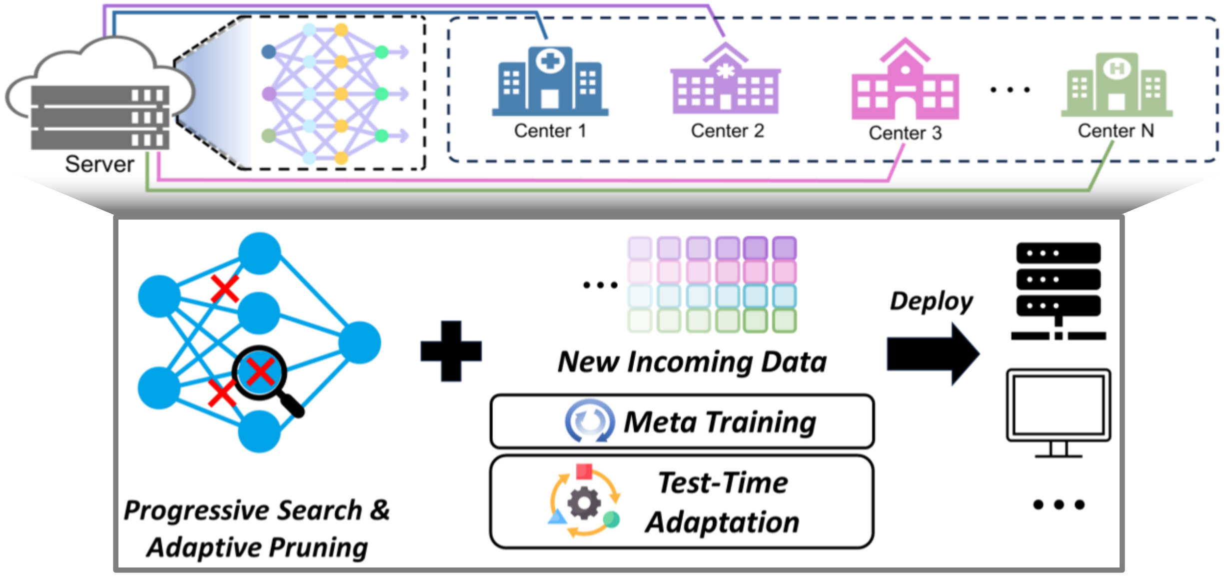
- **Efficient multi-layer permute** perceptron module captures long-range dependence with positional information.

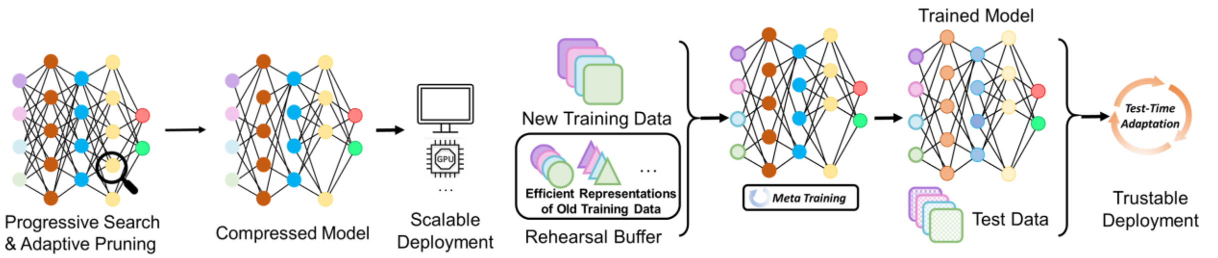




❑Sustainable Deployment and Continual Learning

- Deploy large-scale model to devices via **test-time adaptation, meta learning, continue update,** etc.





[1] Pang et al. Slim UNETR: Scale Hybrid Transformers to Efficient 3D Medical Image Segmentation Under Limited Computational Resources. IEEE TMI, 2023.
[2] Lin et al. Boosting Convolution with Efficient MLP-Permutation for Volumetric Medical Image Segmentation, arXiv, 2023.

Challenges

Data

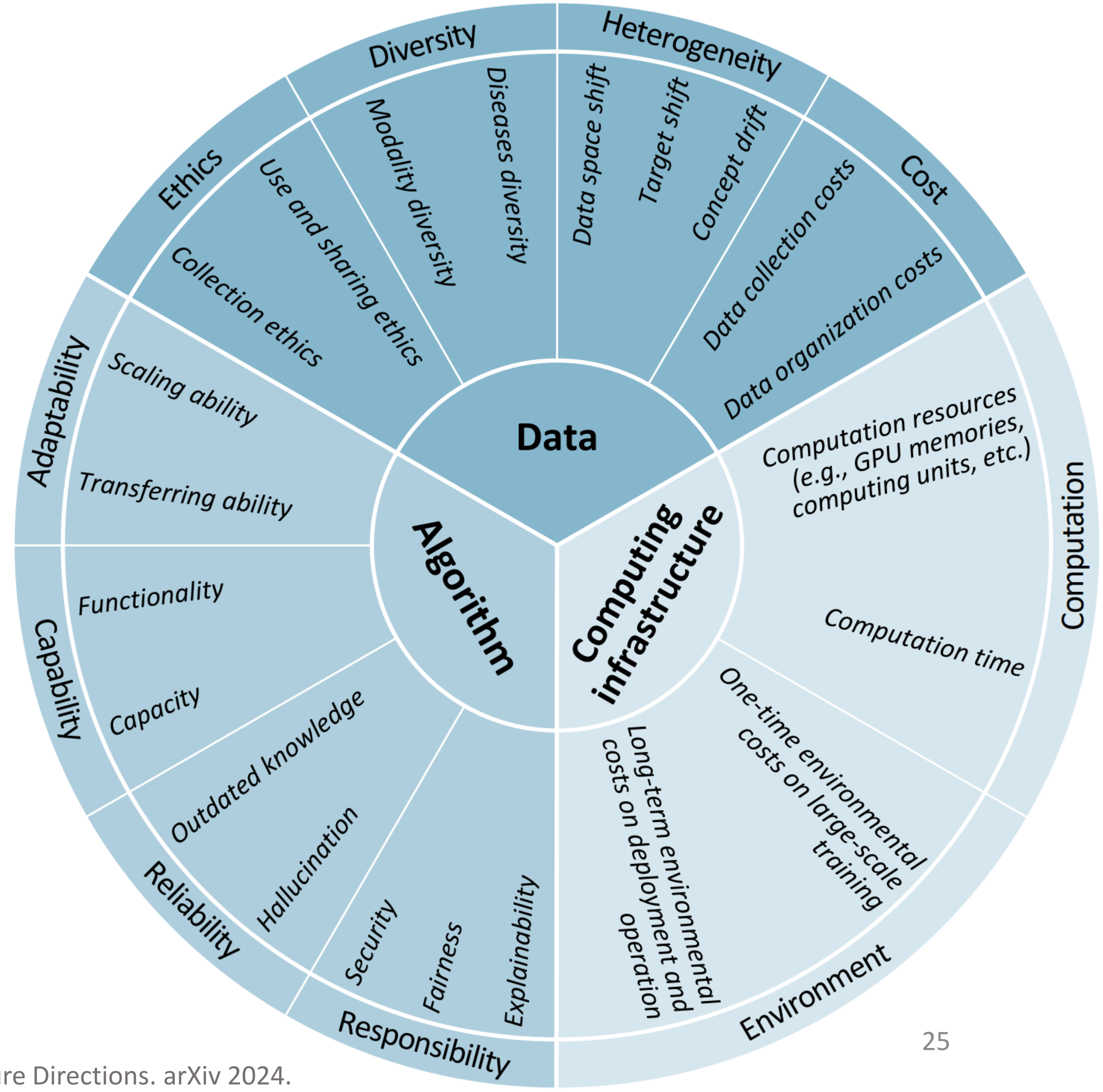
- *How to get **large-scale high-quality medical data** for foundation model training?*
- *It is still facing the ethical issue, heterogeneity, cost, etc., challenges.*

Algorithms

- *How to construct **powerful enough AI algorithms** for medical knowledge learning?*
- *It is still facing the challenges including adaptability, capability, reliability, responsibility, etc.*

Computing infrastructures

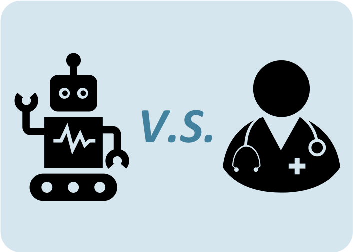
- *How to **widespread deploy** AI models?*
- *How to **sustainably learn** the large AI models?*



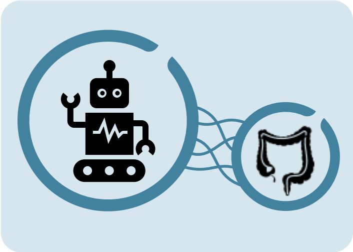
[1] He et al. Foundation Model for Advancing Healthcare: Challenges, Opportunities and Future Directions. arXiv 2024.

Future Directions

Existing paradigms



- AI ***versus*** humans to automatically perform ***repetitive*** healthcare tasks

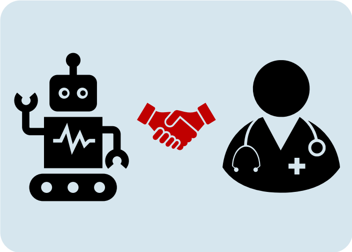


- On ***ideal*** condition for ***single*** issue and ***certain*** situation

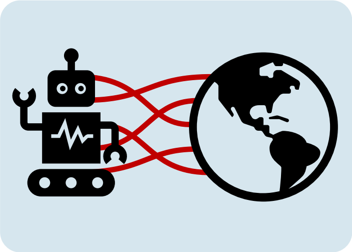
- ***Static*** AI model is fixed to ***specific*** healthcare tasks

- Explore AI methods for ***capability***

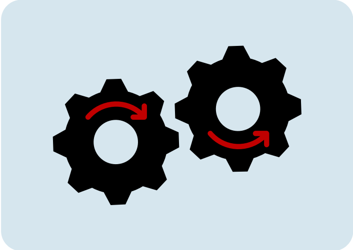
Future directions



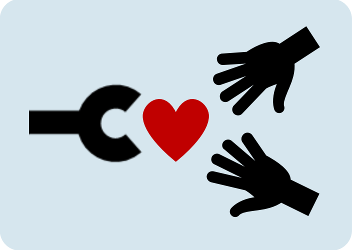
- AI ***cooperates*** with humans to jointly energize ***challenging*** healthcare tasks



- In ***real world*** for ***complex*** issues and ***uncertain*** situation



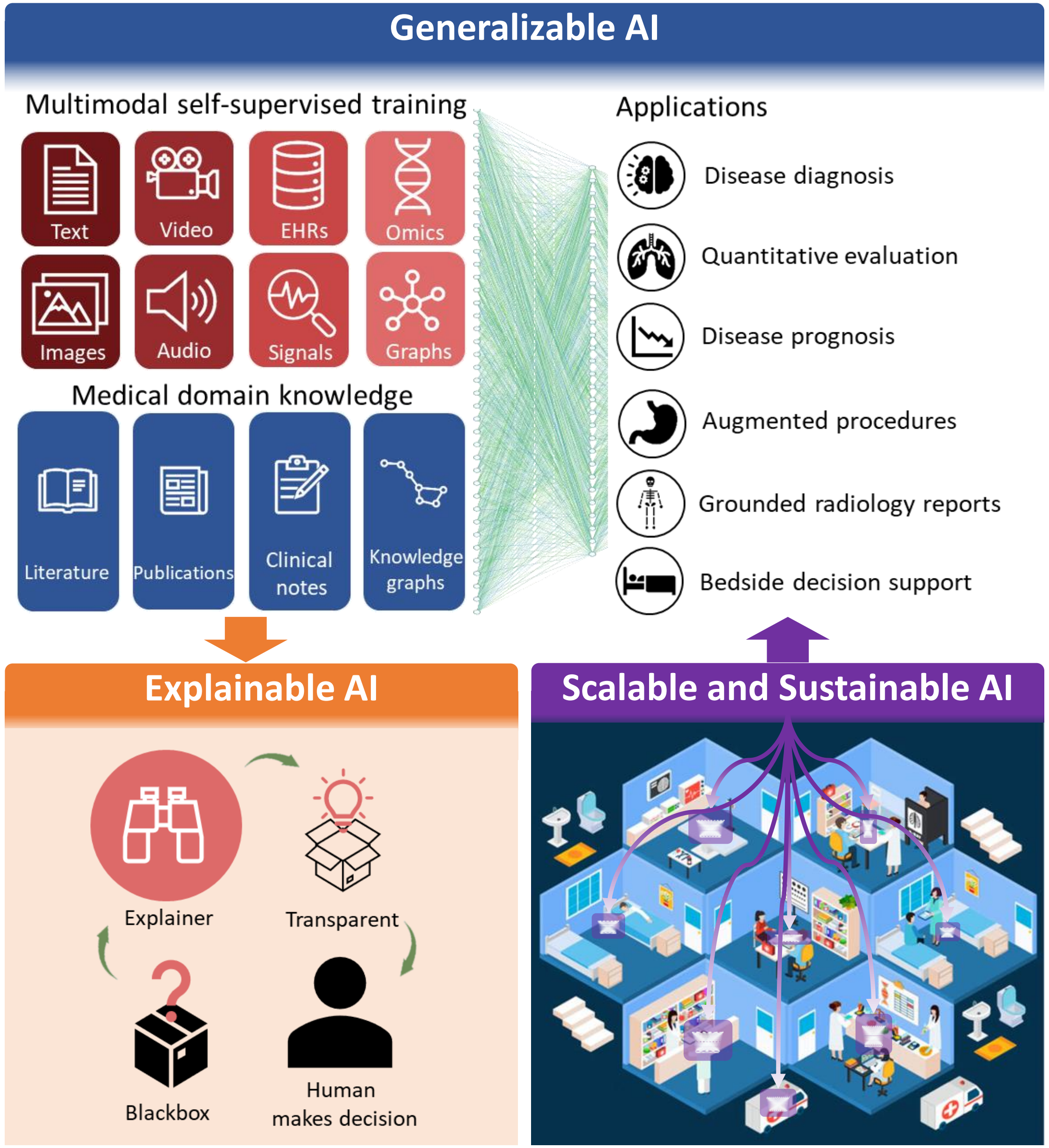
- ***Dynamic*** AI model adapts to ***general*** healthcare tasks



- Trust AI behaviors for ***responsibility***

[1] He et al. Foundation Model for Advancing Healthcare: Challenges, Opportunities and Future Directions. arXiv 2024.

Conclusions



- Medical data is **multimodal** in nature. Clinicians leverage multi-modality data (e.g., images, genomics) for precise diagnosis and treatment.

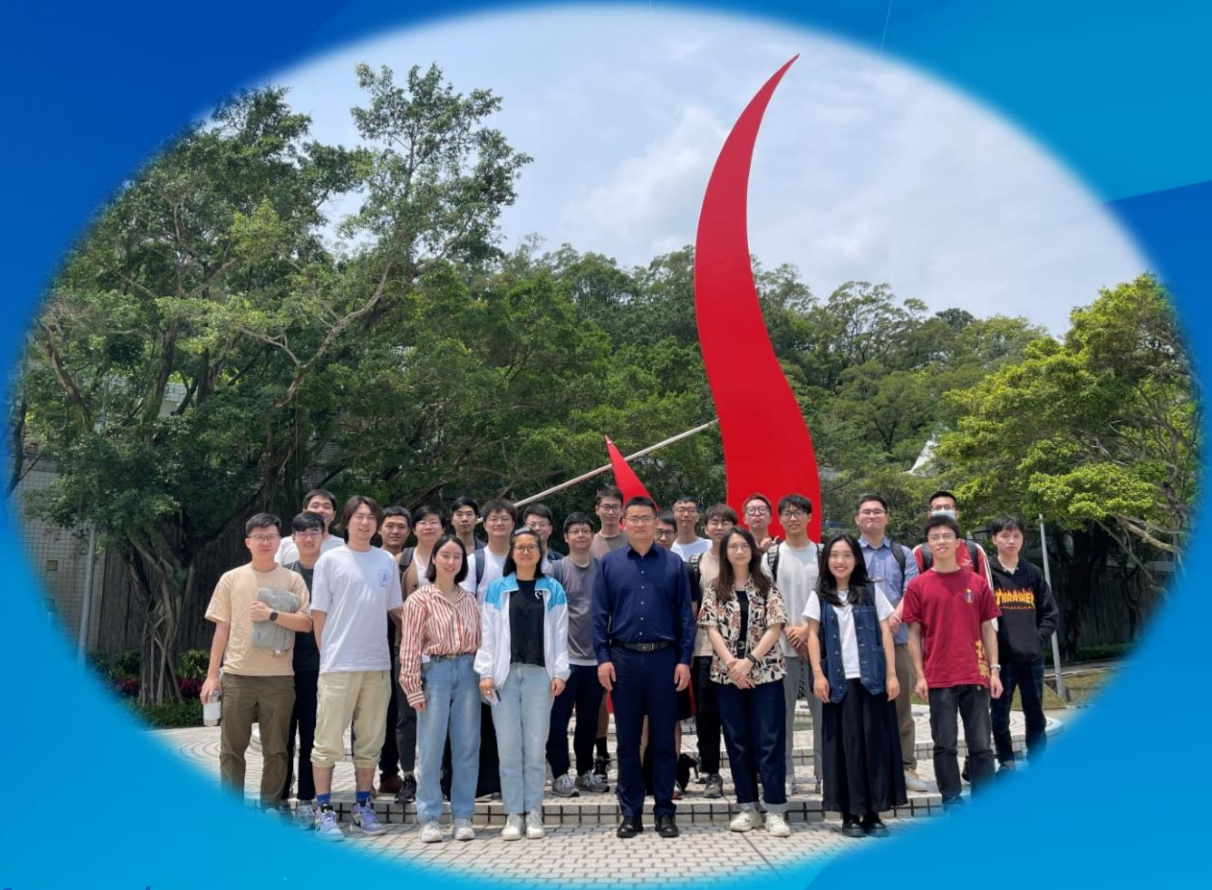
- Foundation model advances healthcare, but it **is still far from** humans' expectations. The way to healthcare foundation model faces many open questions to be explored.

- Trustworthy AI (including generalizability, scalability, explainability, sustainable deployment, benchmark construction, etc.) are key aspects in the **real-word applications** of healthcare foundation models.

Thank You!

Smart Lab: Trustworthy AI for Healthcare

Email: jhc@ust.hk



About Me

Smart Lab

https://cse.hkust.edu.hk/~jhc/ https://hkustsmartlab.netlify.app/