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**
  - Clinically applicable and referral in retina
  - Clinical-grade computational pathology using weakly supervised deep learning on whole slide images

- **FDA/NMPA Approved AI-based Medical Products**
  - Mammogram
  - Ultrasound
  - Fundus
  - Pathology
  - CT & MRI
Diverse Medical Data

### Mono-Modal

- Medical Records [1]
- Genomic Data [2]
- Dermoscopy Images [3]
- ECG Signal [4]

### Multi-Modal

- X-ray Imaging and Report [5]

**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

Multimodal Medical Tasks

Medical Vision Tasks

- Classification
- Detection & Segmentation
- Registration

Medical Multimodal Tasks

Finding: 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 postoperative course is now complicated by decompensated CHF.

Plan Subsection: Respiratory failure with MRSA pneumonia: continue seven days of vancomycin.

Relation: Direct
Bigger Data, Larger Model

~200,000 CT Volumes
~100,000 with Reports

~1M Chest X-Ray
~300,000 with Reports

~300,000 MR Volumes
~100,000 with Reports

~3 Billion Sequence Genomics

We are not short of data, but high-quality labels!

~300,000 WSIs
Partially with Reports
How to deploy general models to clients with different computing resources?

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

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

2019 MICCAI Young Scientist Impact Award

2023 Asian Young Scientist Fellow

**Winners of 15+ Grand Medical Challenges**

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

Computational Pathology

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

Ophthalmology

Ophthalmology Disease Screening including Glaucoma, DME, etc.

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.

Multimodal Foundation Model for Healthcare

Uni-modal Foundation Model
- Applications: Disease diagnosis, Quantitative evaluation, Disease prognosis, Augmented procedures, Grounded radiology reports, Bedside support

Multi-modal Foundation Model
- Applications: Disease diagnosis, Quantitative evaluation, Disease prognosis, Augmented procedures, Grounded radiology reports, Bedside support

Federated Foundation Model

MAIF

[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.
Pathology Foundation Model

Enhance Long Sequence Modeling in Pathology

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

Prompt-Guided Model Adaptation for WSI Diagnosis

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

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.

  - Cross-modal Feature Enhancement
  - Self-Adaptive Disease-Balanced Learning
  - Diagnosis

- **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.

  - Ours
  - Thorax is symmetrical. The lung window shows that bilateral lung markings are increased and disordered, and the light transmittance of the lung field is increased. There are multiple calcifications in the right lung. Multiple nodules can be seen in both lungs. The largest nodule is located in the anterior segment of the left upper lobe (S3, S1, and S1, and the solid nodule shadow is 6 mm x 4 mm in size. Bilateral pulmonary hilum are not large. The shape of heart shadow and heart big vessels is normal, and no obvious mass or enlarged lymph node is found in mediastinal. Left pleural effusion and pleural thickening.

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.

---


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.

---


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

3D 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
Application: Computer-assisted Intervention

Preoperative Planning and Intraoperative Navigation for Assisting Surgery

**Preoperative**
- **3D reconstruction**

**Intraoperative**
- **Computer-assisted Intervention**

**Dynamic registration**

**AR & VR visualization**

**Real-time navigation**

**System integration**

**System assessment**
Explainable AI (XAI)

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

Explainable AI
- Explainer
- Transparent
- Blackbox
- Human makes decision

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

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

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.

Multimodal Intrinsic XAI for Survival Prediction

How pairwise instances from histology and genomics are matched?


LLM-enhanced XAI in Healthcare

- **LLM-based Prompt Learning**
  
  - Explainable prompt learning for computer-aided diagnosis via concept-guided context optimization.

  ![LLM-based Clinical Prompt Design](image)

  ![Multi-Granularity Soft-Hard Prompt Alignment & Global-Local Image-Prompt Alignment](image)

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

  ![Concept Activation Vectors Learning](image)

  ![Explainable Disease Diagnosis](image)

---


Scalable and Sustainable AI (SAI)

Multimodal self-supervised training

Applications

Multimodal foundation model

Medical domain knowledge

Text
Video
EHRs
Omics
Images
Audio
Signals
Graphs

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

Efficient MLP-permutation for Segmentation

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

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

Sustainable Deployment and Continual Learning


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?

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**

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

https://cse.hkust.edu.hk/~jhc/

Smart Lab

https://hkustsmartlab.netlify.app/