XAICancer: Explainable Artificial Intelligence for Cancer Imaging
Neo Christopher Chung

Routine screening based on medical imaging has been a major contributor to reduction in mortality and morbidity for several types of cancer. Currently, cancer detection and diagnosis based on computed tomography (CT), positron emission tomography (PET), and other imaging technology require labor-intensive manual examination by clinicians. With the global population both growing and aging, cancer imaging and diagnosis will become more critical than ever. How can artificial intelligence (AI) driven by deep neural networks (DNNs) help clinicians and patients to improve diagnosis and prognosis based on cancer images?

Recent developments of DNNs have demonstrated unprecedented performance in a wide range of computer vision tasks. Therefore, AI could help improve and automate diagnosis and prognosis of cancer based on medical images. To achieve this translational potential, we propose a comprehensive research project towards explainable AI (XAI) for cancer images. The prevailing paradigm is that AI is trained in an end-to-end fashion, without considering how the trained models reach their decisions. These types of black-box AI are largely unacceptable for healthcare. Being able to understand AI's decision making process will have two major consequences: First, it will increase our trust and willingness to accept AI as part of a clinical process. Second, explainability helps discover biases and weaknesses in AI models that can be reasoned and improved.

The XAICancer project plans to achieve this goal by developing a series of DNN architectures, interpretability methods, and open-source models that advance explainable AI for oncology. In particular, we target how state of the art AI models can provide accurate and explainable diagnosis for lung and head-and-neck cancers. We plan three interconnected objectives:

Objective 1. Generative AI for Synthetic Medical Images: How do we overcome a lack of massive quantities of CT/PET images needed to train AI? We will develop a controllable generative AI model, called *CancerDiffusion*. Trained on heterogenous CT/PET images from multiple cancer types, *CancerDiffusion* will be able to generate realistic CT/PET images that can be used to train downstream AI models.

Objective 2. Cancer Foundation Models using Self-Supervised Learning: Most cancer images do not include clinical annotations that are generally essential for traditional machine learning classification tasks. How do we leverage unannotated CT/PET images to improve classification and diagnostic performances? Self-supervised learning (SSL) is a state of the art approach to learn inherent characteristics of data without labels. We will develop specialized SSL architectures and training schemes for both real and synthetic cancer images. We plan to train and release open source *Cancer Foundation Models (CFM)*.

Objective 3. Diagnosis and Interpretability using Vision Transformers: When AI is used for cancer diagnosis in practice, clinicians need to understand the underlying decision-making process. Specifically, when AI makes a classification that a CT/PET image does have a malignant (or benign) tumor, which parts of the image did it *look* at? To this end, we develop interpretability methods and interpretable classifiers based on vision transformers (ViT). All of our proposed models and methods will be developed into a proof-of-concept software *XAICancer*, which provides interactivity and visualizations.