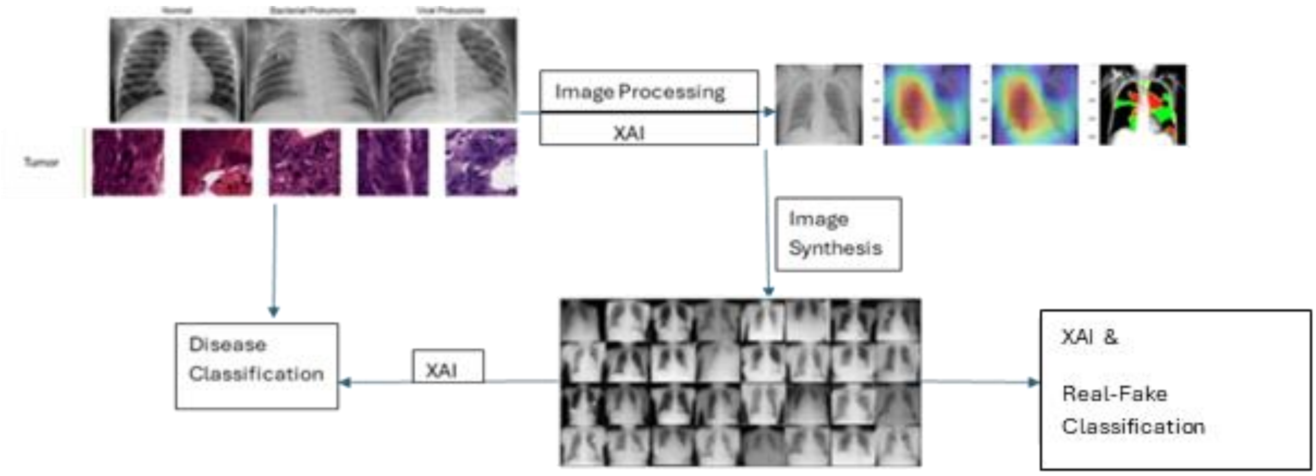
XAI Framework for Medical Image Synthesis, Classification, and Real-vs-Synthetic Differentiation

Aim: This project aims to develop an explainable artificial intelligence (XAI) framework that combines medical image synthesis, classification, and real-vs-synthetic differentiation. The goal is to analyze how synthetic medical images are generated, how classification models interpret both real and synthetic data, and how explainability insights can enhance the detection of synthetic images. By integrating image processing and explainable AI techniques, the study seeks to improve data efficiency, interpretability, and reliability in medical imaging applications.

Backgrounds: Medical image datasets are often limited and imbalanced, which restricts the performance of deep learning models. Synthetic image generation using generative models (e.g., GANs, diffusion) can alleviate this problem, but the interpretability and reliability of these synthetic samples remain uncertain. Explainable AI (XAI) helps visualize and understand how models generate and classify images, revealing key decision regions. By combining image synthesis, classification, and explainability, this study aims to improve model transparency and develop a reliable framework to distinguish real from synthetic medical images.



Objectives: This study aims to investigate how explainable AI (XAI) can provide insights into medical image synthesis and classification. It seeks to develop models capable of distinguishing real from synthetic images in both Chest X-ray and histopathology domains, while evaluating the impact of synthetic data on classification performance and interpretability. The project also focuses on visualizing the key features and latent patterns that guide both generative and classification models to enhance transparency.

Methods: The project will use the Chest X-Ray Image Dataset (5,216 training, 624 test, 16 validation images) and the Kather Texture 2016 dataset (5,000 histopathology images, 150×150 px, 8 tissue classes). Images will be preprocessed with resizing, normalization, and optional contrast enhancement. Synthetic images will be generated using conditional diffusion models or StyleGAN2-ADA, conditioned on class labels. Classification experiments will employ CNN-based models (e.g., ResNet, ViT) trained on real-only and real-plus-synthetic data. Explainable AI methods (Grad-CAM, Integrated Gradients, SHAP) will be applied to both generative and classification models to identify influential regions. Binary classifiers will also be trained to distinguish real from synthetic images, and evaluation will be performed using image quality metrics, classification performance measures, and explainability analyses.

Materials: Two open-access datasets will be used: the grayscale Chest X-Ray Image Dataset, consisting of 5,216 training, 624 test, and 16 validation images, and the Kather Texture 2016 dataset, containing 5,000 RGB histopathology images of 150×150 px each, covering 8 tissue categories. Experiments will be implemented in Python using PyTorch or TensorFlow, with Captum for XAI analysis, and GPU-enabled hardware to support training of generative and classification models. Preprocessing and visualization will use standard libraries such as OpenCV and scikit-image.