Kevin Lin
Jason Wang

Confidential Computing in Computer Vision

Description: Eosinophilic Esophagitis (EoE) is a disease characterized by white blood cell (eosinophils) accumulation in the lining of the esophagus. Patients with EoE often complain of chest pain and have trouble swallowing as a result of the excess white blood cells (Adorno, 2021). Symptoms are exacerbated by allergic reactions or acid reflux that can damage the esophageal tissue. In order to diagnose EoE, epithelial samples are collected from patients; the presence of 15 or more eosinophils in a single sample is the generally accepted criteria to confirm the diagnosis (Furuta, 2007).

Complicating the issue is the fact that doctors are not able to assess EoE risk, predict patient outcomes, or determine treatment plans based on the biopsy data. The identification of these eosinophils is typically performed manually, which is tedious and time consuming, and makes conducting research exceedingly difficult. As with all medical image machine learning systems, patient confidentiality must be maintained throughout the entire process.

Dataset:
514 Images/Masks from 30 UVA hospital patients (consent obtained from all subjects under conditions of **academic use only**, data contain no PHI).
Image Size: 512x512x3 pixels (RGB).

Scaling and Efficiency: Medical imaging machine learning algorithms struggle to generalize and our dataset is no exception. Each medical center captures patient data differently and each patient has a different set of medical conditions. For most of our initial analysis, we assumed that the data was well prepared with maximum effort given to minimize the impact of other medical conditions besides EoE. To fully prove that our approach can scale to larger datasets and incorporate other datasets, we can use transfer learning to see if models trained on our dataset work on other datasets. This would also address the efficiency concern since we do not need to retrain a full model on a new given dataset. Other approaches for efficiency could be few shot learning since it does not require the model to train on the full dataset while still providing surprisingly strong results.

Security and Privacy: Patient data should always be protected. Although hacks in other domains are more common, hacks on medical data do occur and are extremely damaging to the public. We would need to create a Trusted Execution Environment (TEE) for patients to access systems with their health data and the entire system would need to be secure. To accomplish this, we encrypt all patient data when it is transferred from system to system and when unauthorized parties attempt to access it. Currently, encryption is one of the most useful data protection methods for healthcare organizations.