Image Classifier for colon disease Dataset

A dataset was obtained from kaggle containing images of colons broken down by test, train, and validation sets. Each set contains 4 classes: normal, ulcerative colitis, polyps, and esophagitis. To a trained eye characterizing these diseases visually may be no problem, especially for the physician performing a colonoscopy. An article by the CDC states that an estimated 15 million colonoscopies were performed in 2012, but only half of people aged 50-75 had been up to date on their screening. New technology has become available to make this process less invasive, and an initiative has been launched by the CDC to increase screenings of this population to 80 percent. The new technology available to patients is a camera that they can swallow that takes images as it passes through the digestive track.

With an increase in colonoscopies performed and this new tech, the workload of those reviewing these images is bound to increase. A deep learning model which performs image classification for the colon diseases stated in this data set could be applied to patient images to flag them as diseased for review. This project is constrained to the diseases contained within the dataset; however, future projects should focus on diversifying diseased images to widen the application of this model.

https://www.kaggle.com/datasets/francismon/curated-colon-dataset-for-deep-learning

https://www.mayoclinic.org/tests-procedures/capsule-endoscopy/about/pac-20393366

https://www.cdc.gov/media/releases/2016/p0622-colorectal-cancer-screening.html#print