Colorectal Cancer diagnosis: A comparative study between Neural Networks vs. Boosted Forest Deep learning for business use cases

June 20, 2022

Abstract

The advancement of AI & ML in the medical field, particularly cancer diagnosis, has long been held back on the basis of accuracy limitations and a lack of trust from the practitioners. Since the diagnosis changes the course of action taken on a patient, any error, whether a false positive or negative, can lead to loss of life, adverse action, or potentially an unnecessary treatment. As a result, ML has not had the role of a primary predictor, acting as an optional aid instead.

This particular innovation fosters the 'time is money' mantra, accelerating the diagnosis of a patient in emergency ward, empowering gastroenterologists to recommend the right procedure reviewing the outcome of the diagnosis with the patient simply swallowing a smart pill with a camera.

The scope of my research is to get to a consistently accurate diagnosis as is possible by highlighting the areas of interest to the Physician (whether it's a polyp, ulcer, etc), allowing faster conclusion several hours in advance compared to a traditional procedure with no invasive procedure to the patient. The accuracy is realized by comparing the two dominant ML models; Neural Networks and Boosted Forest and will concludes that the best training and inference approaches for the pragmatic business use cases.

Both models are fed images of various colorectal cancer biomarkers, including but not limited to Polyps, Ulcers, and Colon Erosions, and photos of a regular, functional colon pathway. The training and test dataset represents a wide variety of use cases in upper and lower GI tract. The research utilized up to 110,000 images, and other 890,000 images for rest purposes. For better accuracy, the research utilized the images from traditional scope and Smart Pill as well. Their performance is measured using accuracy, precision-recall, F-1, log-loss, and confusion matrix.

I am incredibly to offer the conclusion of my research to the Healthcare community and push the envelope on the accuracy of ML models in the medical diagnosis and plethora of future imagery use cases. I am grateful to have had the opportunity to work with the industry superior HPCC Platform and to have been able to utilize its incredibly supercomputing and expansive libraries.