Sarvesh Prabhu

A comparative study between Neural Networks vs. Random Tree Deep learning in Colorectal Cancer diagnosis use case using HPCC supercomputing and libraries

June 20, 2022

Abstract

The advancement of Artificial Intelligence and Machine Learning in the medical field, particularly cancer diagnosis, has long been held back on the basis of accuracy limitations and a lack of trust from practitioners and hospitals. Since the diagnosis changes the course of action taken on a patient, any error, whether this be a false positive or a false negative, can lead to loss of life, or loss in the form of unnecessary treatment. As a result, ML has not had the role of a primary predictor, acting as an aid instead (if even that).

To identify or create a model that hospitals and practitioners alike can trust for primary diagnosis in colorectal cancer, I compared the performance metrics of two popular machine learning models which heavily utilize two of the expansive HPCC platform libraries. For this use case, I have developed a Neural Network model using the HPCC GNN library and a Random Forest model using the HPCC Learning Trees library.

Both models will be fed images of various colorectal cancer biomarkers, including but not limited to Polyps, Ulcers, and Colon Erosions, as well as photos of a regular, functional colon pathway. The images will come from the kvasir and hyper-kvasir dataset, containing 48,000 and 110,000 respectively. Their performance will be measured using accuracy, precision-recall, and F-1.

By completing this research, I will be able to give this vital information back to the Healthcare community and push the AI-medical-integration envelope further while also giving back to the HPCC community with beneficial insight on imagery analysis, which fellow ML engineers can use in a plethora of image use cases.

Colorectal Cancer diagnosis: A comparative study between Neural Networks vs. Random 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, leads to loss of life or potentially unnecessary treatment; As a result, ML has not played the role of a primary predictor, acting as an optional aid.

This innovation fosters the ‘time is money mantra, accelerating the diagnosis of a patient admitted into the emergency ward, empowering gastroenterologists to recommend the correct procedure by 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, possible by highlighting the areas of interest to the Physician (whether it’s a polyp, ulcer, etc.), allowing faster conclusion several hours faster than a traditional procedure while being non-invasive. The accuracy is realized by comparing the two dominant ML models: Neural Networks and Random Forest. They will conclude 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 standard, functional colon pathway. The training and test dataset represents various upper and lower GI tract use cases. The research utilized up to 110,000 images and 890,000 images for rest purposes. For better accuracy, the research utilized the images from both traditional scope and Smart Pill. Their performance is measured using accuracy using precision-recall, F-1, log-loss, and confusion matrix.

I am ecstatic to offer the conclusion of my research to the Healthcare community and push the envelope on the accuracy of ML models in medical diagnosis and a plethora of future imagery use cases. I am grateful to have had the opportunity to work with the industry's superior HPCC Platform, my awesome mentor Mr. Bob Foreman, and to have been able to utilize its incredible supercomputing and expansive libraries.