

Sarvesh Prabhu

A comparative study between Neural Networks vs. Boosted 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 Boosted 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.