Colorectal Cancer Diagnosis: Achieving consistent accuracy in medical imagery analysis from a swallowable smart pill; a comparative study between Deep Learning and Random Forest models

Sarvesh Prabhu, Senior, Lambert High School, Suwanee, GA

Advisors: Roger Dev, Sr. Architect & Bob Foreman, LexisNexis Risk Solutions, GA

Abstract

The advancement of ML in the medical field, particularly cancer diagnosis, has long been held back based on accuracy limitations and a lack of trust from practitioners. Since the diagnosis changes the course of action taken on a patient, any false +ve or -ve leads to potentially unnecessary treatment or even a loss of life.

As a result, ML has not played the role of a primary predictor, acting as an optional aid.

My research aims to perform a Gastrointestinal (GI) diagnosis by **accurately highlighting the biomarkers** (polyp, ulcer, etc.) from the pictures captured by a **swallowable smart pill** as it travels through the GI tract. The research empowers Physicians to reach a faster conclusion for patient treatment by several hours in advance compared to a traditional colonoscopy, all while being non-invasive.

The goal is to achieve the highest consistent accuracy in diagnosing colorectal cancer using medical imagery analysis utilizing HPCC's Generalized Neural Network (GNN) & HPCC ECL-ML libraries. The model is trained and tested to identify biomarkers (features) of 1m images of the upper & lower GI tract using the GNN bundle. GI diagnosis was performed in two ways: Deep Learning, using the convolutional neural network and TensorFlow, and Random Forest model using the learning trees bundle.

The research concludes that **ensembling the GNN and multi-layered Random Forest yields the highest accuracy** and consistent results with 99.8% precision and 100% recall, enabling the practitioner to make an informed treatment decision for the patient in need of urgent care.