Using Machine Learning to Predict Colon Cancer Tumors

Scott Chen and Arav Srivastava

Abstract:

Today, Colon cancer is one of the leading causes of cancer-related death, with approximately 1.9 million cases in the year 2022, 930,000 being deaths. Early detection and proper diagnosis play very important roles in improving patient outcome and survival rate. However, the diagnostic procedure falls heavily on the manual interpretation of Whole Slide Images (WSI) by pathologists,which can be very time-consuming. The main goal of this project is oriented toward developing a machine learning model which integrates Convolutional Neural Networks (CNN) with Graph Neural Networks (GNN) for the prediction of colon cancer tumor outcomes from Whole Slide Images. The proposed models use the application of The CNN arctitecture, ResNet50, to extract features from the Whole Slide Images, the derived features are submitted to a GNN workflow to further refine predictions based on the structual relationships of detected features. The combination of the two techniques helps improve accuracy and stability in the detection of tumors, which ultimately aims to provide a reliable colon cancer diagnosis tool to assist pathologists. This, in turn, settles more emphasis on researching how to streamline the analysis of WSIs to reduce the workload on pathologists, scale up to accommodate large data sets, and increase diagnostic consistency. The application of this project could significantly contribute to the early detection and treatment of colon cancer, ultimately improving patient outcomes.

Goals:

1. Achieve high accuracy and precision in tumor predictions
2. Create a pipeline for processing, analyzing, and predicting WSIs
3. Evaluate and train the model
4. Develop a Machine Learning model to predict Colon Cancer Tumors from Whole Slide Images
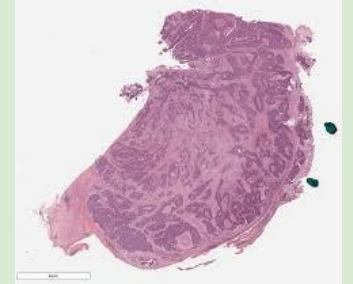
Methods:

Data: Whole Slide Images (WSI) will be used from the Dartmouth Hitchcock Medical Center.

Model: A workflow integrating a Graph Neural Network (GNN) with a supervised Convolutional Neural Network (CNN) will be utilized.

- **Model Build:** A Convolutional Neural Network which utilizes either ResNet or DenseNet architecture.
- **Model Training:** The model will then be trained multiple times with the Whole Slide Images provided of the Colon Cancer.
- **Test Model and Improve:** The model will finally be tested on Whole Slide Images it hasn't seen before and an Accuracy will be determined.After this the model is trained until it reaches the certain criteria
- **Analysis:** The study model is analyzed and considered final when the success criteria is met.