Utilizing Machine Learning to Identify Metastatic Tissue in Histopathologic Scans of Lymph Node Sections

Kini, Siddharth (School: Crooms Academy of Information Technology)

Multiple studies have found that Metastasis is responsible for between 66 to 90 percent of cancer deaths. Metastasis involves cancerous cells detaching from their initial tumor and entering the circulatory or lymphatic systems and proliferating in a separate part of the body. Currently, doctors must review multiple scans of lymph node sections, to ensure that a cancer patient does not have metastatic cancer. With digital pathology becoming more widespread, computers can utilize convolutional neural networks to identify metastatic tissue in histopathologic scans with a high accuracy. Neural networks can be trained from scratch, or by adapting an existing one for a new application, which is called transfer learning. How does a machine learning model trained from scratch compare to a model trained using transfer learning? A dataset of 120,000 images of breast cancer metastasis was used to train, validate, and test the two different types of models. Nine classifiers with different configurations of sequential model "sections" and neural layers per "section" were trained from scratch for 15 epochs. A "section" was defined as a set of convolutional layers with a dropout and max_pooling layer. A VGG19 model using transfer learning was also trained for 15 epochs. The best performing model trained from scratch had 4 sections with 2 convolutional layers per section. When this was compared against the VGG19 model, it was found that the model utilizing transfer learning had an F1 score of 92.47 percent, 1.76 percent higher than the best performing model trained from scratch.