Abstract

With the introduction of whole slide imaging, the field of pathology has experienced a rapid growth in discoveries. However, many of these discoveries have resulted from manual, labor-intensive methods that require the supervision of trained professionals whose time is scarce. This has led to the exploration of automated, machine learning based approaches which are intended to alleviate the stress put on trained professionals who are qualified to annotate such resources. Even with the introduction of artificial intelligence based methods, much of the data used to develop these technologies exists in sizeable formats, with few labels, and additional nuances such as localized targets and subjective annotations.

A computationally lightweight supervised model is explored with a variety of different training and implementation methods for the classification of targets that are present in localized regions of large resolution imagery using few labeled data points. A simplified structure featuring a pretrained convolutional neural network feature extractor and a support vector machine classifier, with a dimensionality reduction in between is implemented with a trivial one step training process. This straightforward approach is then built upon with a full pipeline, multistage training process that attempts to circumvent the issue of localization for a more specialized, accurate model. The functionality of both techniques are evaluated on an intercalated disc classification heart muscle imaging dataset provided by The Human Protein Atlas [7] and Johns Hopkins University [8] and cross compared to gauge the requirements of future needs. This research lays the groundwork for semi-supervised machine learning on partially labeled pathological imaging datasets.