# Abstract

Developments in whole slide imaging and in computational analysis methods have opened the door to new discoveries in the field of medicine. In particular, machine learning has gained attention as a tool for analyzing large amounts of medical image data. However, while the ease of using machine learning systems makes them appealing, building the labeled datasets required to train these systems remains a difficult problem due to the need for trained professionals to annotate each data point. This work presents a machine learning-based classification pipeline that incorporates both labeled and unlabeled images, reducing the burden of building a fully-labeled dataset. The pipeline consists of a pretrained convolutional neural network feature extractor and a support vector machine classifier, with the feature extractor finetuned with MixMatch, a semi-supervised learning algorithm. Different training methods are explored based on the pretrained weights of the feature extractor and the ratio of labeled to unlabeled images used during training. The final classification pipelines are evaluated on a heart tissue dataset from The Human Protein Atlas that was annotated for the presence of intercalated discs by doctors at Johns Hopkins University. This approach achieves higher accuracy in classifying intercalated disc presence than a supervised learning-based pipeline.