Learning Without Labels

Unsupervised and Weakly Supervised Learning of Deep Models

Presented by Dr. Shazia Akbar
shazia@altislabs.com

AISC Premium Workshop 2019

Attention-based Deep Multiple Instance Learning

Maximilian Ilse [* 1] **Jakub M. Tomczak** [* 1] **Max Welling** [1]

Abstract

Multiple instance learning (MIL) is a variation of supervised learning where a single class label is assigned to a bag of instances. In this paper, we state the MIL problem as learning the Bernoulli distribution of the bag label where the bag label probability is fully parameterized by neural networks. Furthermore, we propose a neural network-based permutation-invariant aggregation operator that corresponds to the attention

model that predicts a bag label, *e.g.*, a medical diagnosis. An additional challenge is to discover *key instances* (Liu et al., 2012), *i.e.*, the instances that trigger the bag label. In the medical domain the latter task is of great interest because of legal issues[1] and its usefulness in clinical practice. In order to solve the primary task of a bag classification different methods are proposed, such as utilizing similarities among bags (Cheplygina et al., 2015b), embedding instances to a compact low-dimensional representation that is further fed to a bag-level classifier (Andrews et al., 2003; Chen et al., 2006), and combining responses of an instance-level

Task

Purpose here is to learn one label for a very high resolution image

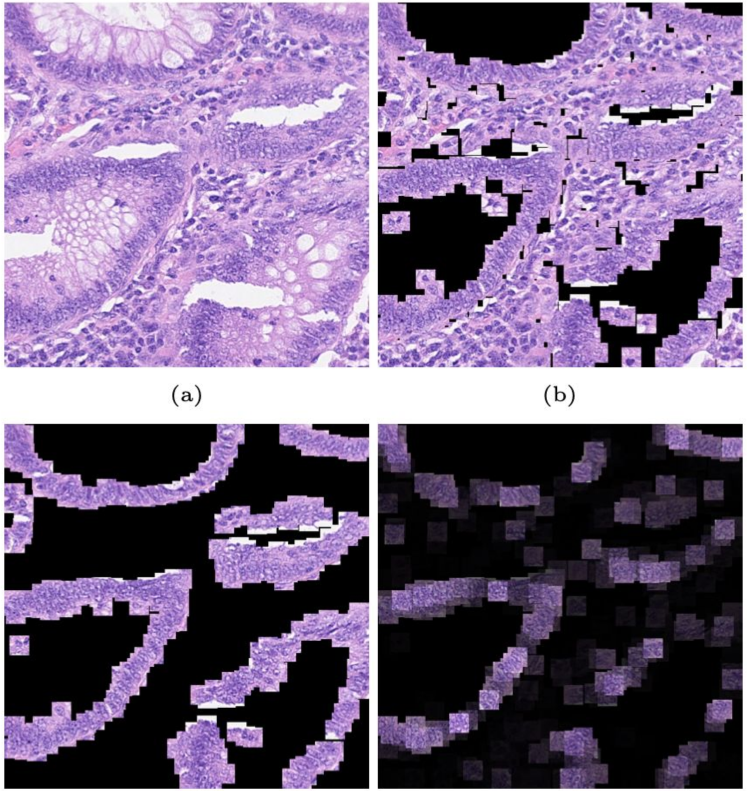
Patches are provided instead but **labels of each patch are not provided**!

Example

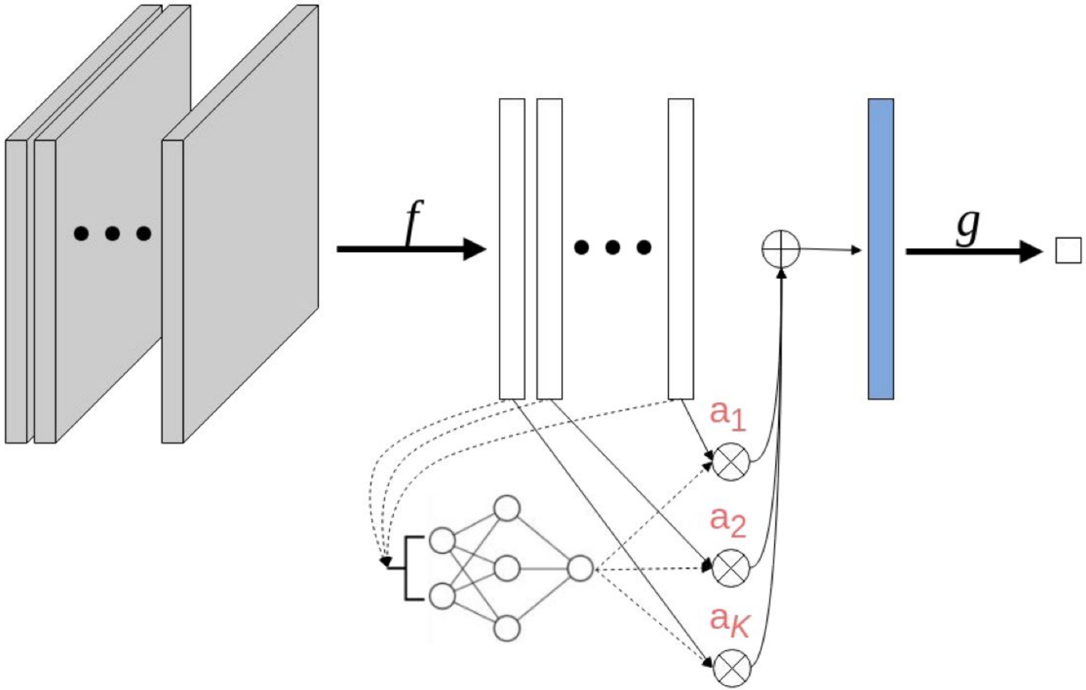
We know that this image contains cancer, therefore y = [1]

However each patch does not contain cancer

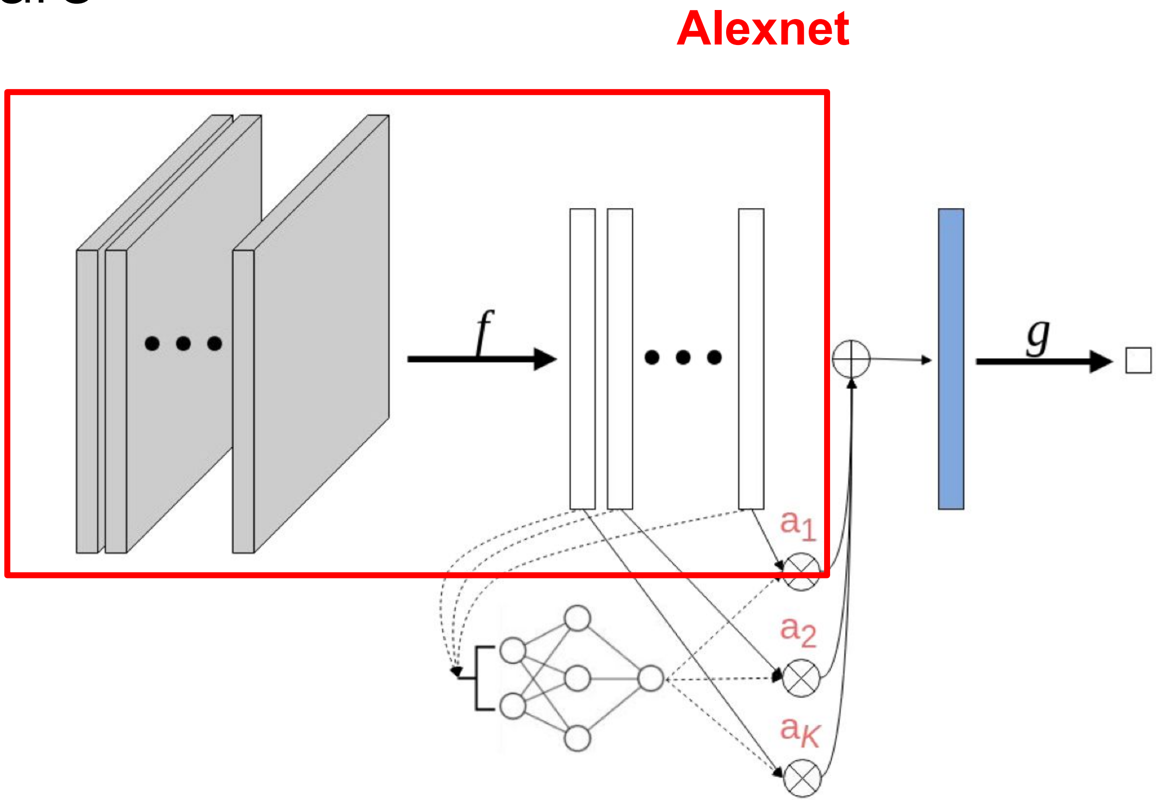
- Only those you can see in bottom left do
- Task: can the model learn which patches do and which patches do not share label y



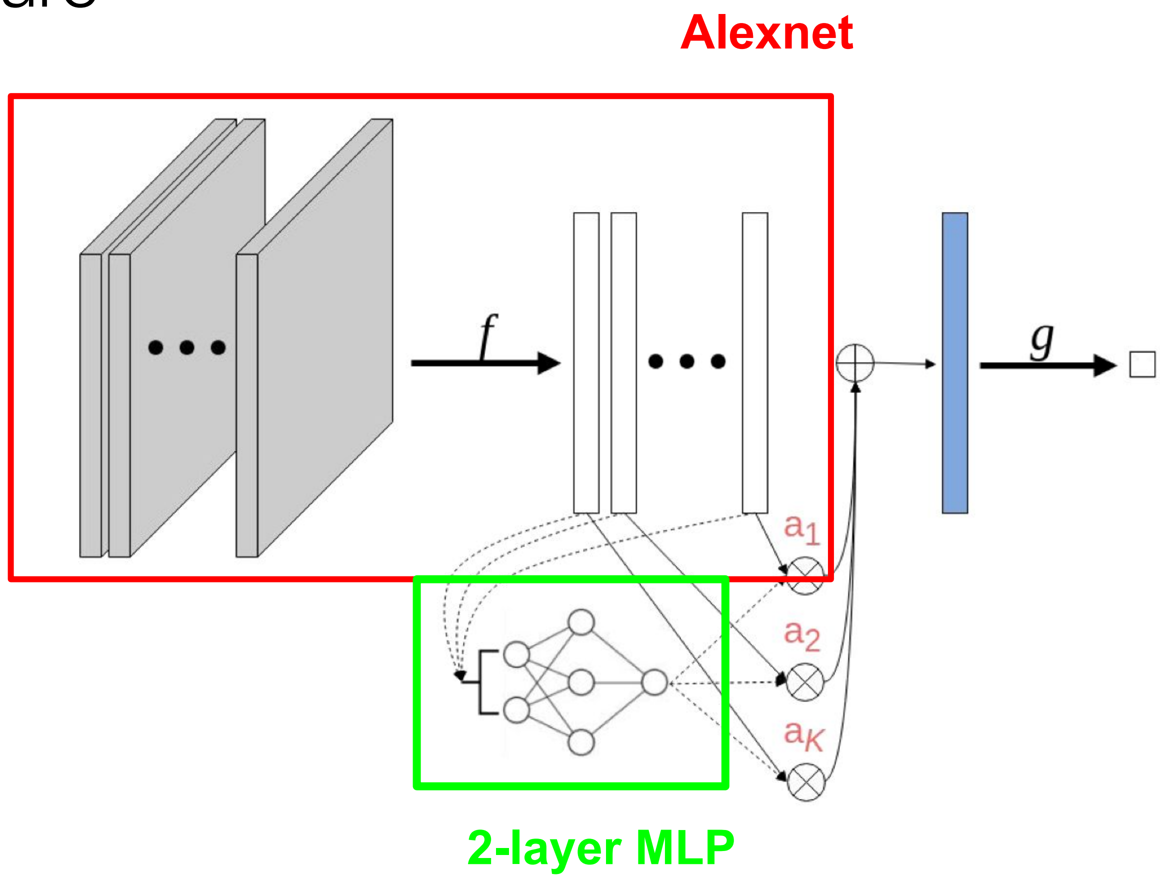
Architecture



Architecture



Architecture



Gated Attention

Take advantage of RNN connections but without temporal constraint

Borrowing techniques from LSTM

- Because we don't have temporal information, no need for a forget gate
- Input gates can be formulated as:

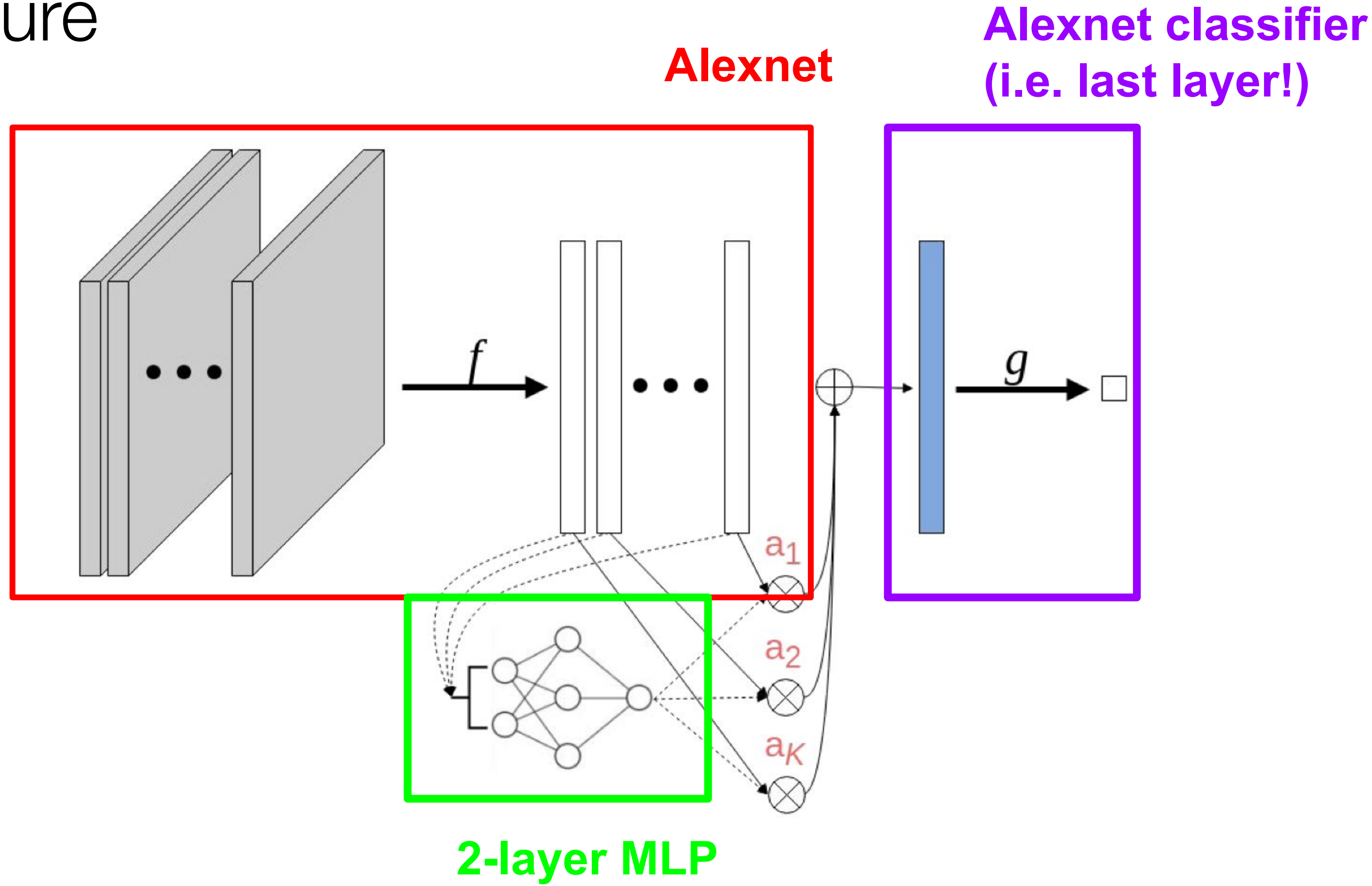
Gates controlling how much information is passed through

$$h_l(\mathbf{X}) = (\mathbf{X} * \mathbf{W} + \mathbf{b}) \otimes \boxed{\sigma(\mathbf{X} * \mathbf{V} + \mathbf{c})}$$

More commonly known as "GLU"

Language Modeling with Gated Convolutional Networks, Dauphin et al (2017)

Architecture



Attention class

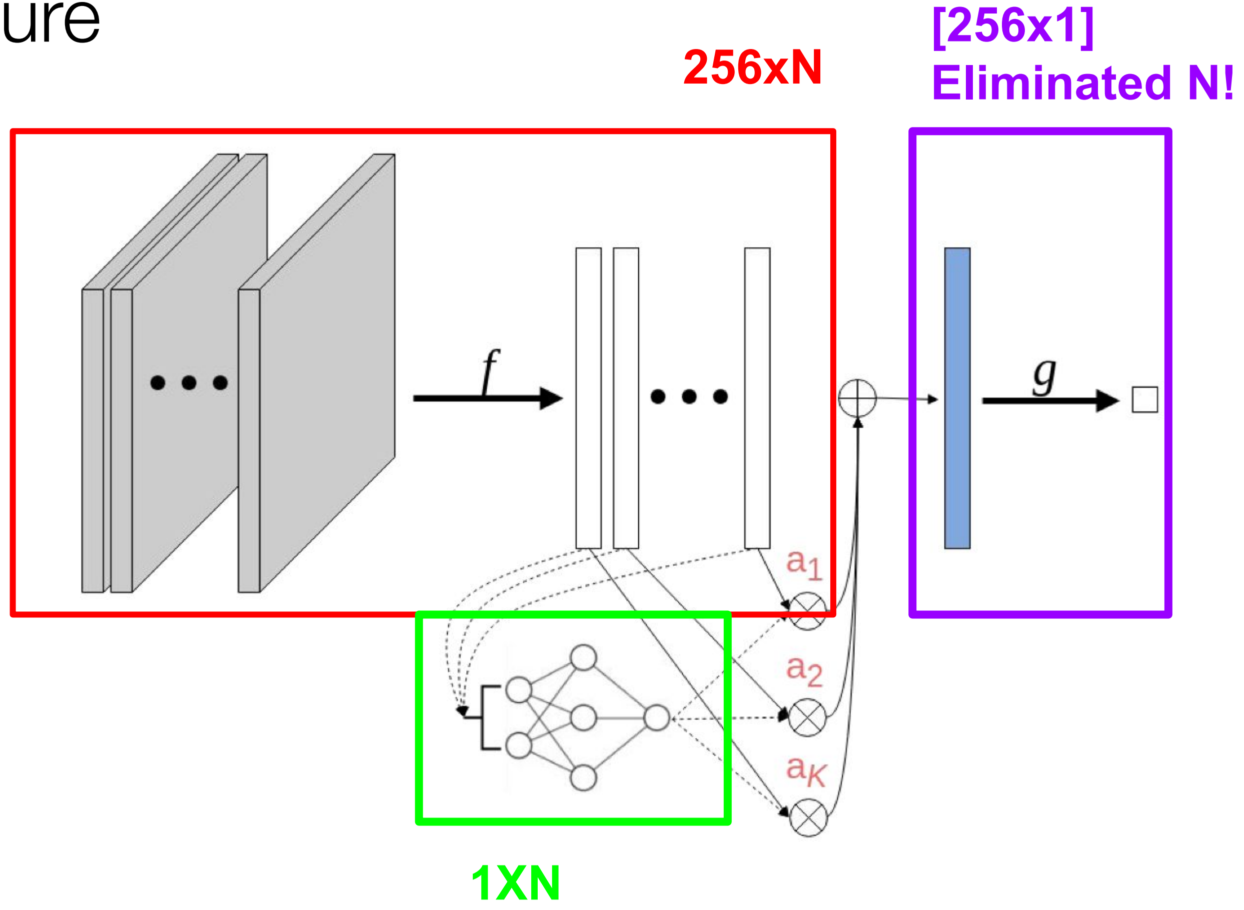
Three models:

1. AlexNet - up to the second last layer (i.e. the last fully connected layer before output)
2. AlexNet - last layer (as per normal)
3. MIL Attention: two layer MLP which will take as input Alexnet features and output a weight for each feature component

```
nn.Linear([alexnet_feature_size], [mlp_layer1_size])
nn.Tanh()
nn.Linear([mlp_layer1_size], 1)
```

Remember to apply softmax after as well!

Architecture



Dataset

You have been provided with code in the colab notebook which creates MNIST_BAG as described in the paper

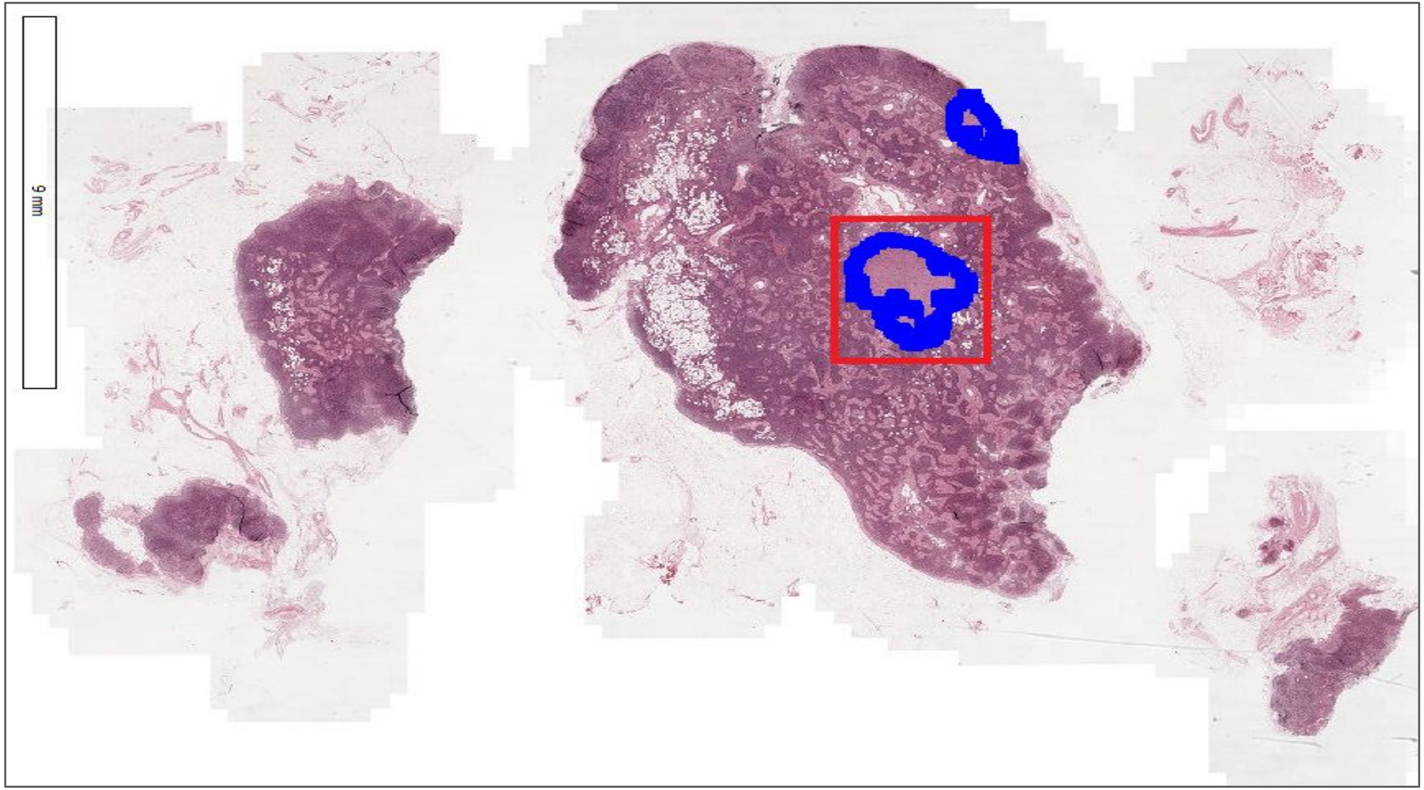
Dataset

Also put together a real-world dataset similar to the cancer classification dataset

A subset of PCam: https://github.com/basveeling/pcam

100 patches randomly taken from each WSI in training set

Camelyon 2017



Dataset

Also put together a real-world dataset similar to the cancer classification dataset

A subset of PCam: https://github.com/basveeling/pcam

100 patches randomly taken from each WSI in training set

- Cancer (1): may contain some healthy structures
- Healthy (0): all healthy

Increase throughput in pathology lab

Predict cancer subtypes