HISTOPATHOLOGIC CANCER DETECTION BASED ON DEEP MIL

Ph.D. student: Xiankun Yan

Outline

① Introduction of project

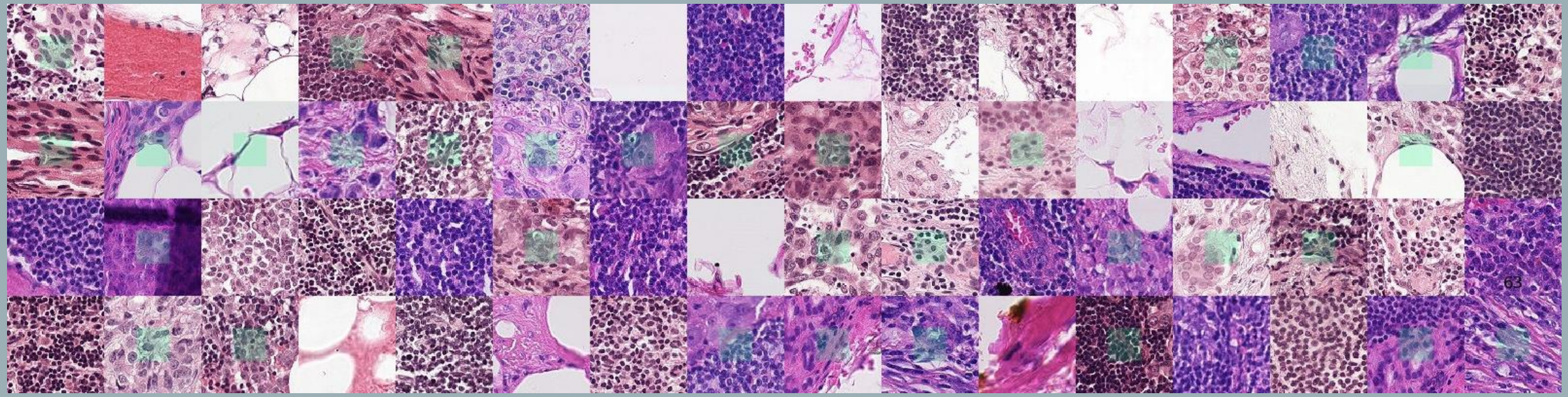
② My major work

Deep Multiple instance learning approach

③ Experiment and Results

④ Next Work and Conclusion

The dataset



Histopathologic image:

Extracted from histopathologic scans of lymph node sections

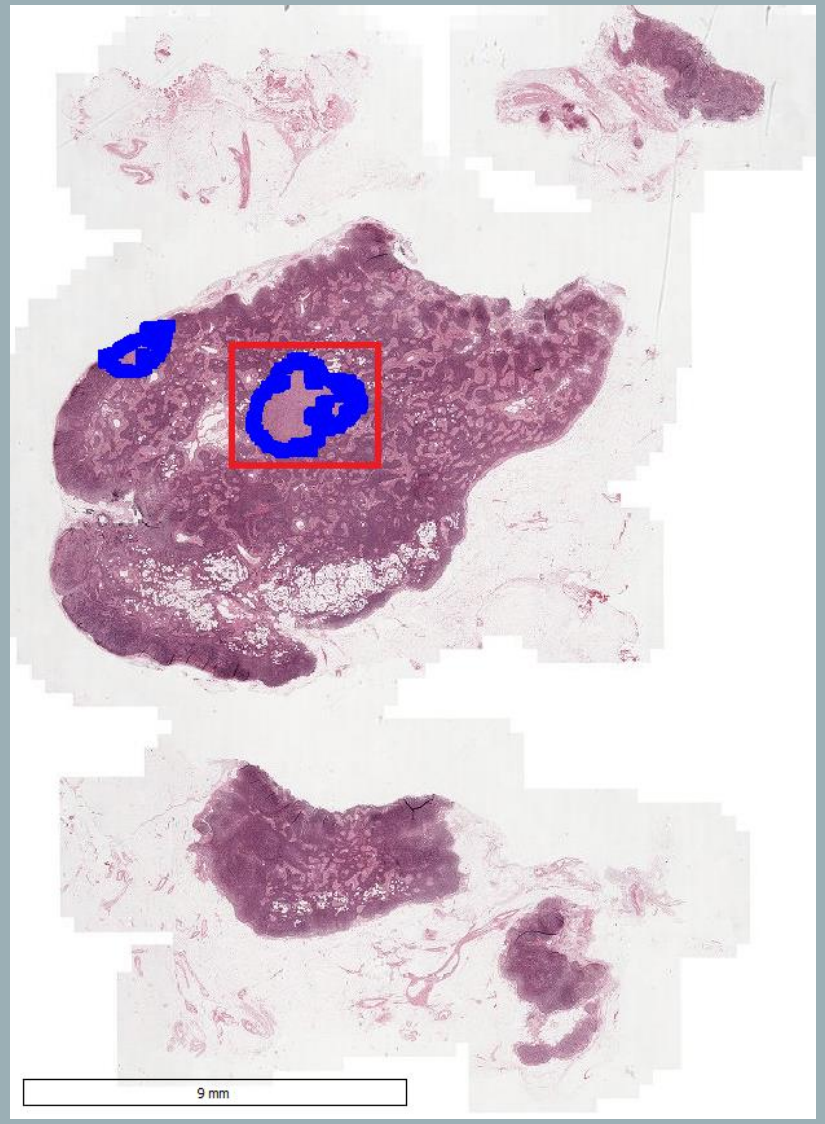
Reference: https://github.com/basveeling/pcam

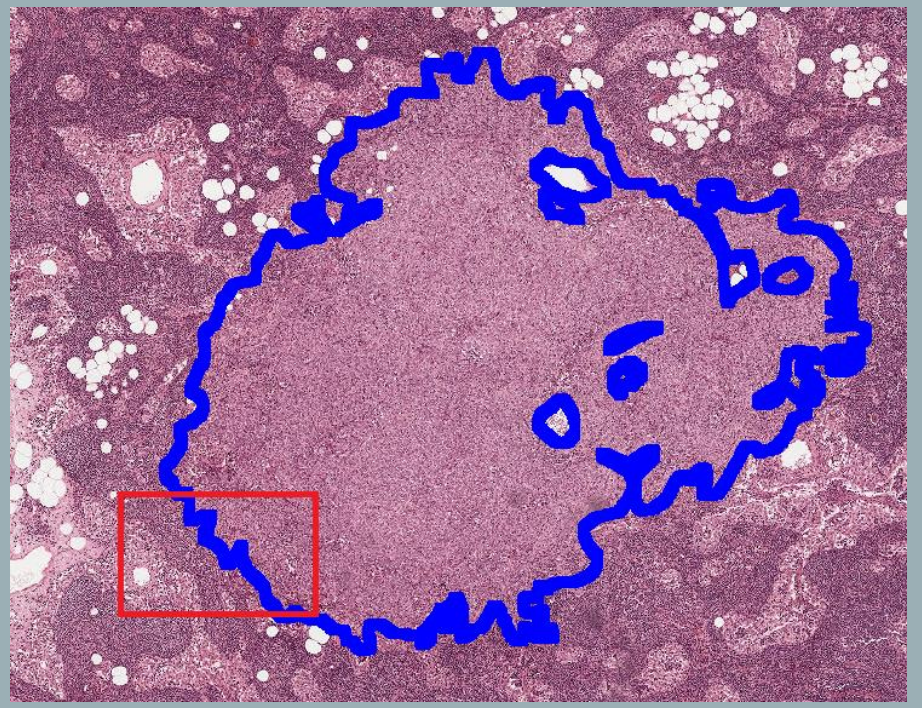
The dataset

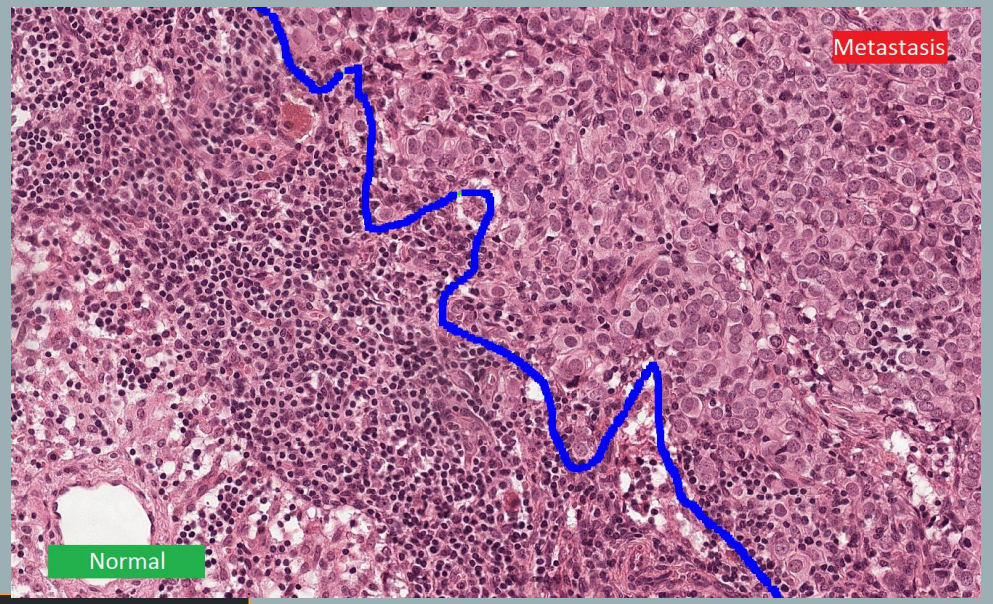
The histopathological images are glass slide microscope images of lymph nodes that are stained with hematoxylin and eosin (H&E). This staining method is one of the most widely used in medical diagnosis and it produces blue, violet and red colors. Dark blue hematoxylin binds to negatively charged substances such as nucleic acids and pink eosin to positively charged substances like amino-acid side chains (most proteins). Typically nuclei are stained blue, whereas cytoplasm and extracellular parts in various shades of pink.

组织病理学图像是用苏木精和曙红（H＆E）染色的淋巴结的玻璃载玻片显微镜图像。 这种染色方法是医学诊断中最广泛使用的方法之一，它产生蓝色，紫色和红色。 深蓝苏木精与带负电荷的物质（如核酸和粉红素）结合到带正电荷的物质如氨基酸侧链（大多数蛋白质）。 通常，细胞核染成蓝色，而细胞质和细胞外部分染成各种粉红色。

Reference: https://camelyon17.grand-challenge.org/Background/







Reference:https://camelyon17.grand-challenge.org/Background/

The dataset

The data-set is the PatchCamelyon (PCam). The size of image is 96 x 96px.

Training set: 262.144 (2^18) examples

Validation and test set: 32.768 (2^15) examples

No overlap.

All splits have a 50/50 balance between positive and negative examples.

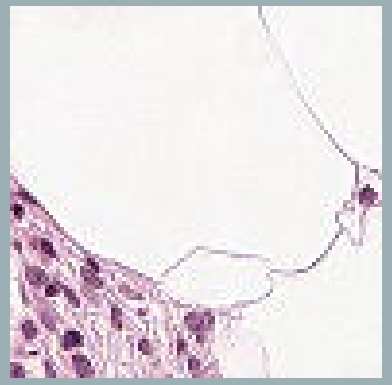
Reference: https://github.com/basveeling/pcam

The aim:

Identify the metastatic cancer in histopathologic scans of lymph nodes sections.

The main problems:

The image is too complex and the cancer is not easy to identify due to shape and size.



1) The normal image patch

2) The image patch with cancer

Multiple instance learning approach

A weakly supervisor learning approach

The label of bag is binary: -1 or 1.

At positive bag, at least one instance's label is positive.

The label of instance is unknown.

The goal of MIL is to learn an instance classifier.

The bag classifier is built by those instance classifiers.

Multiple instance learning approach

Bags: $x^m = \{x_1, x_2, x_3, \dots, x_n\}$

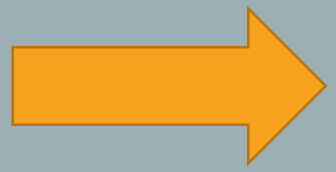
Instances in each bag: $x_i = \{x_{i1}, x_{i2}, \dots, x_{ij}\}$

Bag-level label: $y_i \in y = \{-1, 1\}$

Instance-level label: $y_{ij} \in y = \{-1, 1\}$

The operator $y_i = \max_j (y_{ij})$

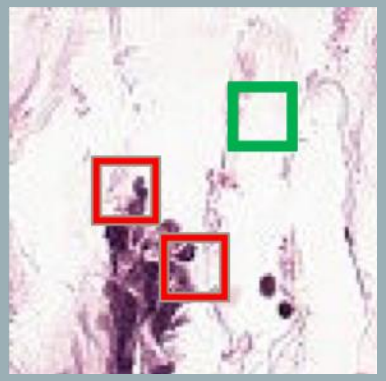
The instance classifier $h(x_{ij}) = x \rightarrow y$



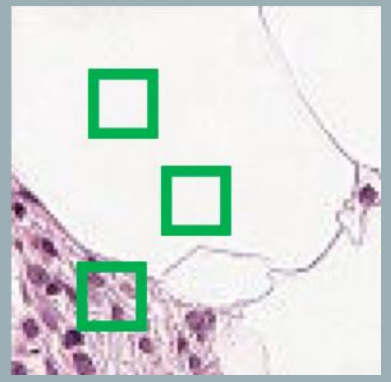
$H(x_i) = \max_j h(x_{ij})$

The bag classifier $H(x_i) = x^m \rightarrow y$

Histopathologic cancer detection based on MIL



Cancerous Image



Normal Image

Advantage:

- Reduce the pressure of labelling work.
- Keep the high accuracy.

References: Xu, Yan, et al. "Weakly supervised histopathology cancer image segmentation and classification." *Medical image analysis* 18.3 (2014): 591-604.

Histopathologic cancer detection based on MIL

Question:

1) how many instance we need in each bag?

2) how do we generate the instance from the bag?

3) For the Multi-classification problem, how do MIL work?

Deep Multiple instance learning

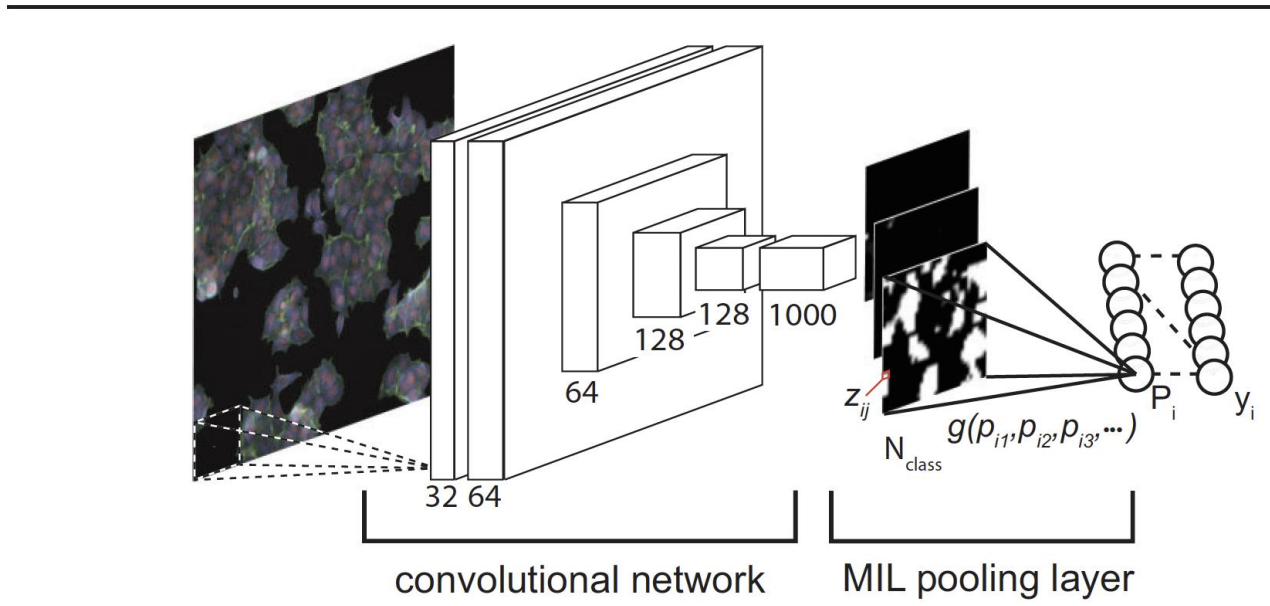


Fig. 1. Convolutional MIL model. $g(\cdot)$ is global pooling function that aggregates instance probabilities p_{ij}

$$x_{ij} \equiv z_{ij}$$

Feature level probability of an instance belonging to the class of bag:

$$p_{ij} = \sigma(z_{ij})$$

Bag level probability for class of bag built by instance probability:

$$p_i = g(p_{i1}, p_{i2}, \cdots, p_{ij})$$

$$P_i = g_i(\{p_{ij}\}) = \frac{\sigma(a(p_{i\bar{j}} - b_i)) - \sigma(-ab_i)}{\sigma(a(1 - b_i)) - \sigma(-ab_i)},$$

Where $\quad p_{i\bar{j}} = \frac{1}{|V|}\sum_j p_{ij}$

Reference: Kraus, Oren Z., Jimmy Lei Ba, and Brendan J. Frey. "Classifying and segmenting microscopy images with deep multiple instance learning." *Bioinformatics* 32.12 (2016): i52-i59.

Deep Multiple instance learning

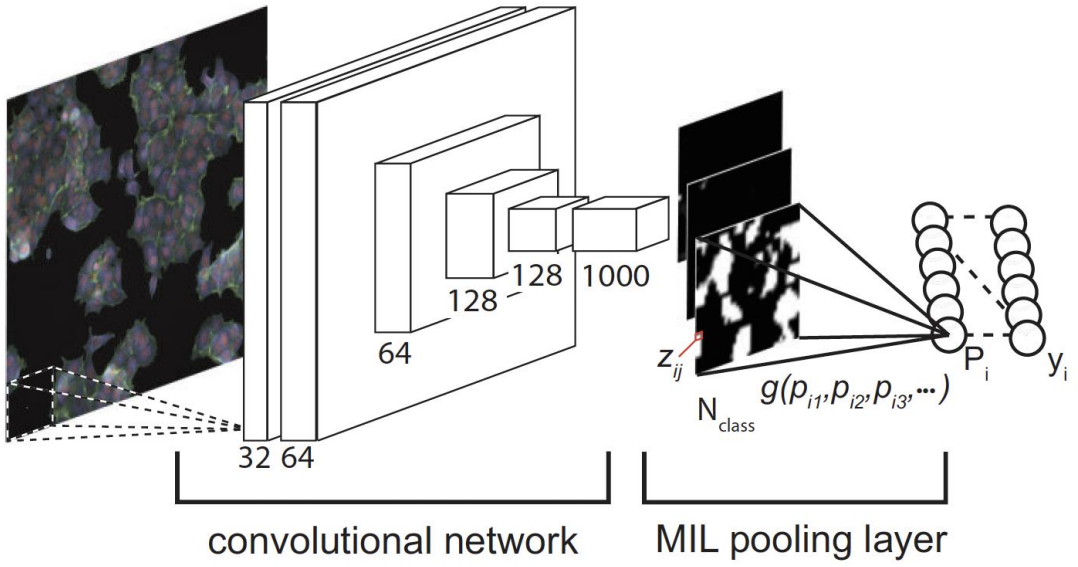
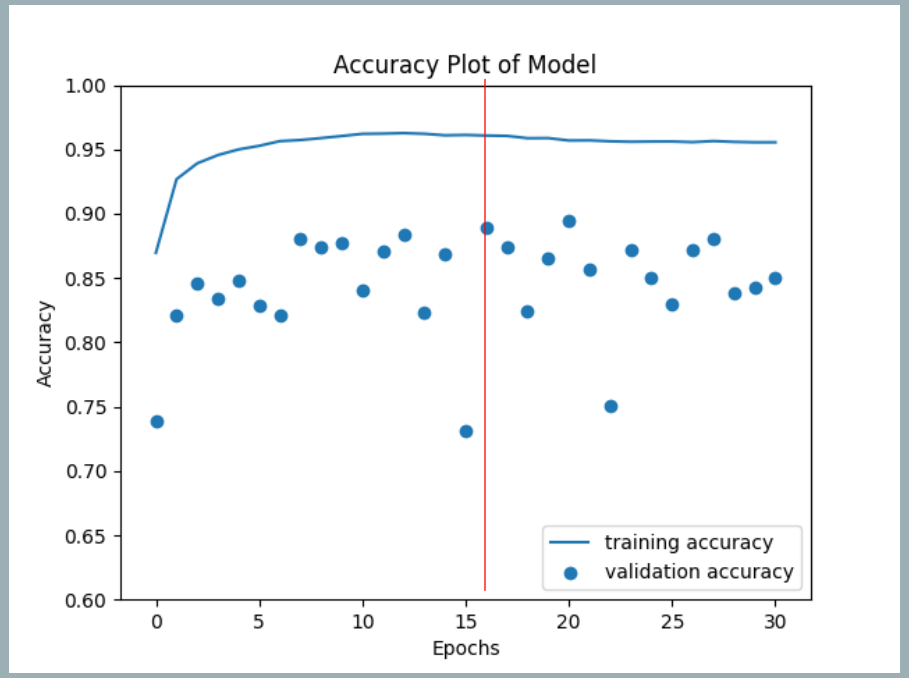
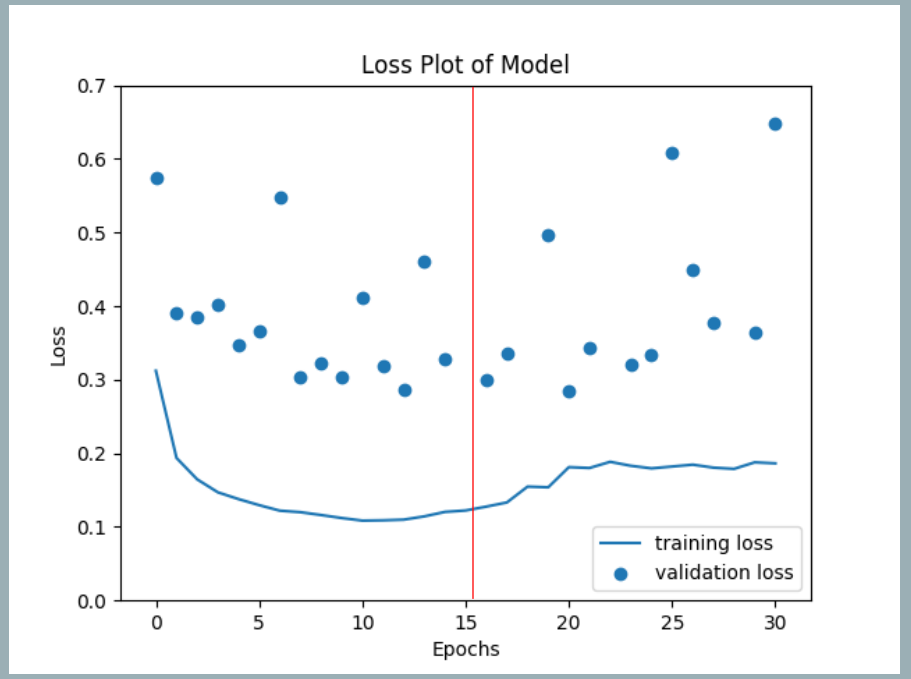


Fig. 1. Convolutional MIL model. $g(\cdot)$ is global pooling function that aggregates instance probabilities p_{ij}

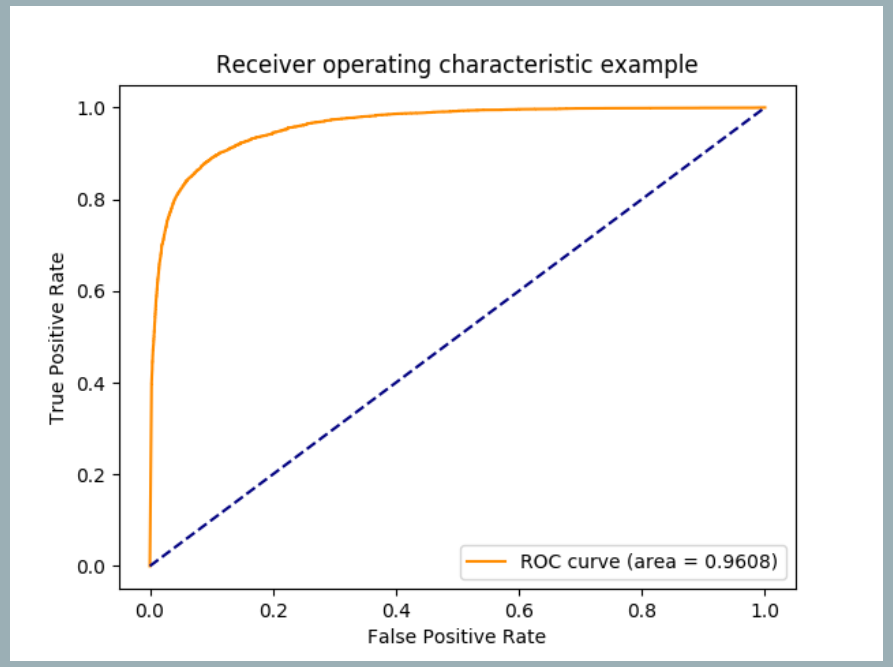
Additional fully connected layer can learn relationships between classes.

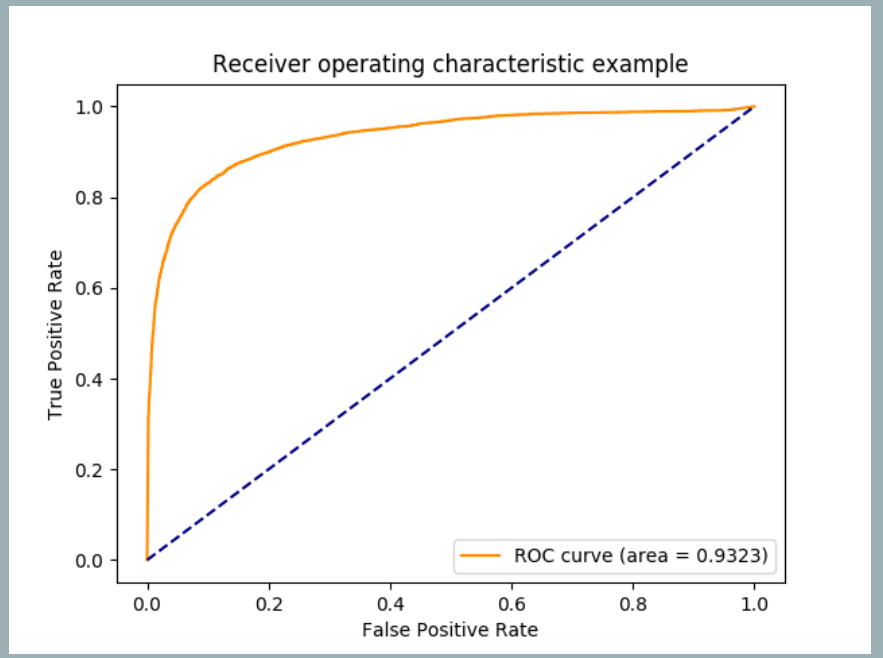
Reference: Kraus, Oren Z., Jimmy Lei Ba, and Brendan J. Frey. "Classifying and segmenting microscopy images with deep multiple instance learning." *Bioinformatics* 32.12 (2016): i52-i59.





Validation ACC:
0.8903





Test ACC:
0.8642

The Future work:

1. DenseNet and MIL.
2. Evolutionary Neural Networks
3. Fuzzy/ Neutrosophic Logic

Conclusion

No Free Lunch Theorem: Without a prior assumption about the specific problem, no strategy can be expected to perform better than any other.

Thank you for your listening.