



# A Novel Attribute-based Symmetric Multiple Instance Learning for Histopathological Image Analysis

*Trung Vu<sup>1</sup>, Phung Lai<sup>1</sup>, Raviv Raich<sup>1</sup>, Anh Pham<sup>1</sup>, Xiaoli Z. Fern<sup>1</sup>, UK Arvind Rao<sup>2</sup>*

<sup>1</sup> - School of Electrical Engineering & Computer Science, Oregon State University, Corvallis, OR 97331-5501

<sup>2</sup> - Department of Computational Medicine and Bioinformatics, and Department of Radiation Oncology,  
The University of Michigan, Ann Arbor, MI 48109

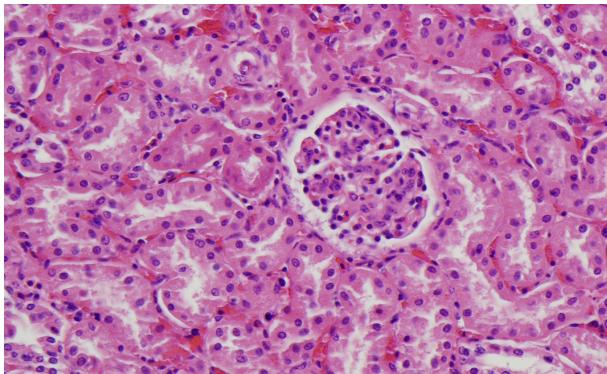
# Outline

- Introduction
- Problem Formulation
- Proposed Model
- Experiments

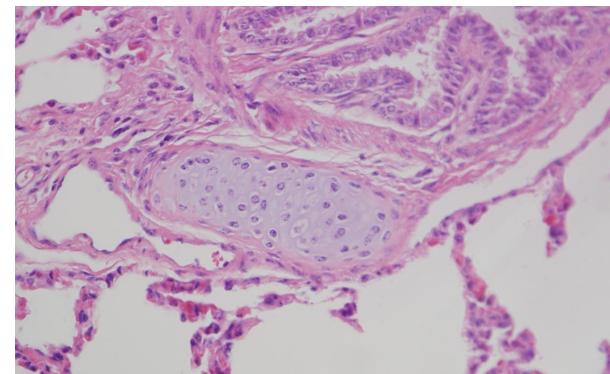
# Histopathological Image Classification (HIC)

Normal

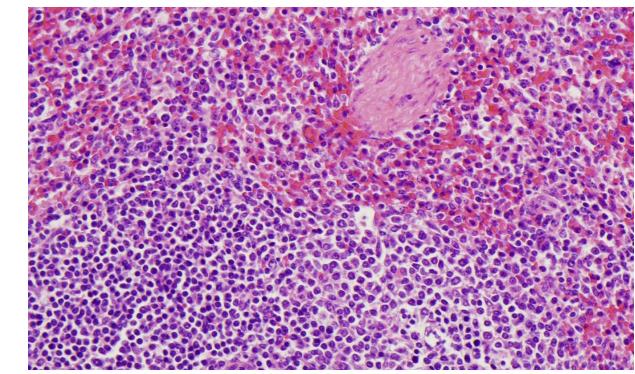
Kidney



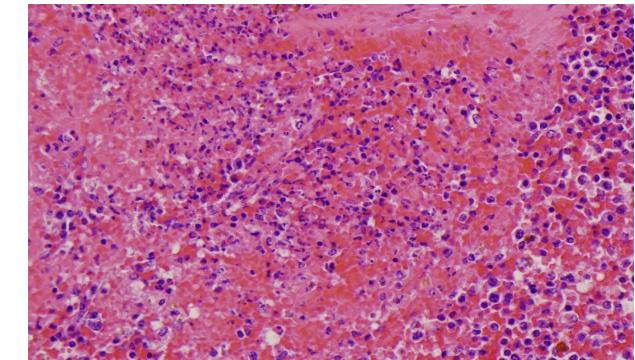
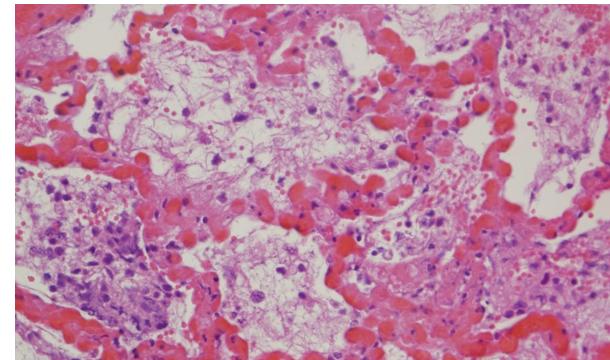
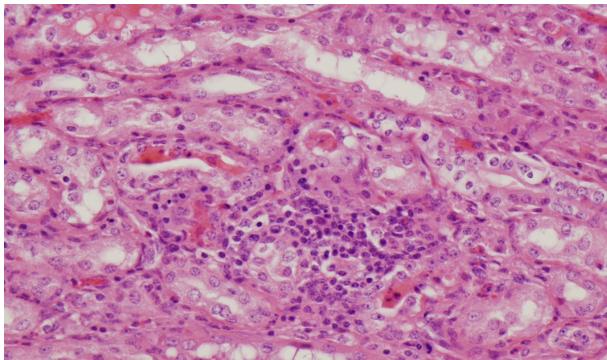
Lung



Spleen



Inflammation

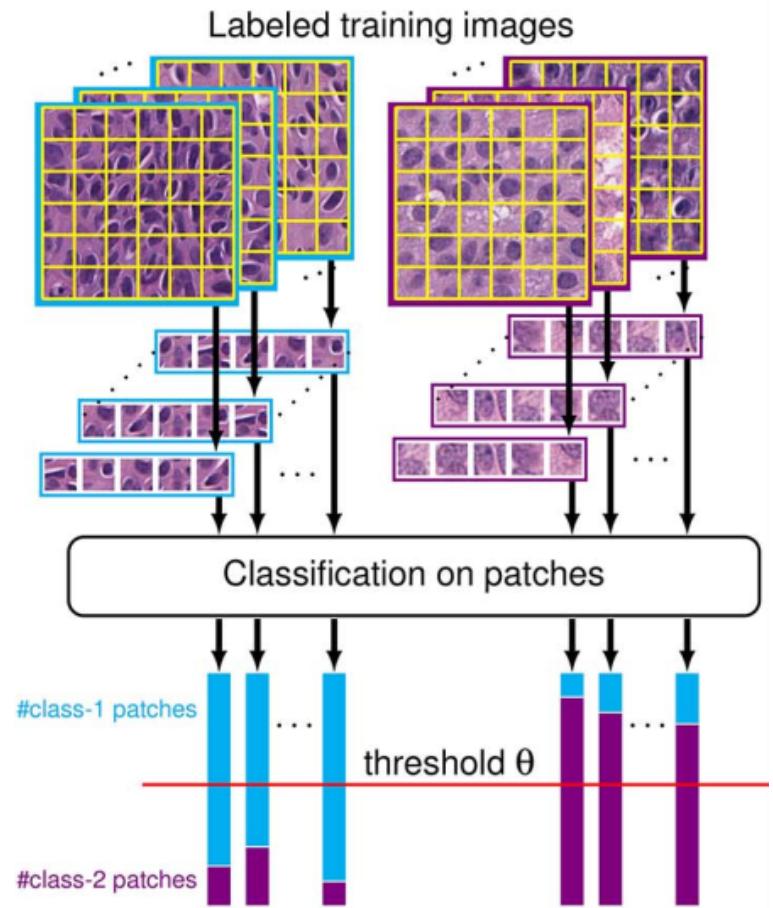


# Automatic HIC

- *Traditional approach* relies on pathologists who are capable of identifying problem-specific cues in a whole slide image (WSI)
  - shape, color, size of cells
  - spatial structure of cells
  - presence of isolated clusters
- *Computer-assisted diagnosis* has become widely-used in recent years
  - reduce variability in human interpretations
  - eliminate significant effort from experts on trivial cases
  - provide quantitative analysis of a specific disease

# Related Work

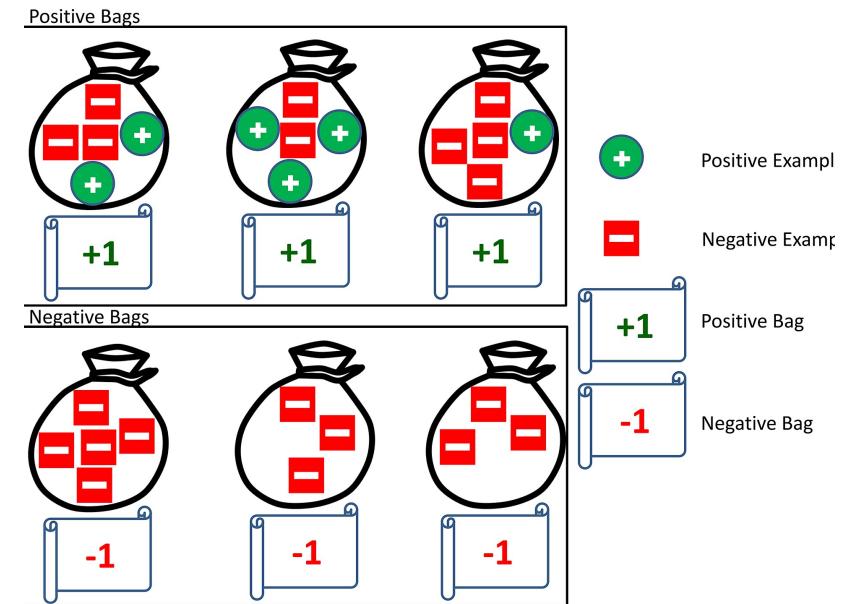
- Fully supervised learning
    - training labels are required for not just the WSI but all of its patches
    - common classification methods can be applied in HIC
      - feature extraction techniques (texture, spatial, graph-based, morphological, ...)
      - supervised training algorithms (SVMs, CNNs, ...)
  - DFDL [1] takes steps forward
    - uses dictionary learning to automatically extract features
- labelling patches is challenging
  - each WSI comprises hundreds to thousands patches
  - need expert knowledge



\* Figure from T. H. Vu, H. S. Mousavi, V. Monga, G. Rao, and U. A. Rao "Histopathological image classification using discriminative feature-oriented dictionary learning," IEEE Trans. Med. Imag., vol. 35, no. 3, pp. 738–751, 2016.

- Weakly supervised learning
  - coarse-grain annotations
    - only the bag-level labels are required
  - common methods in multiple instance learning (MIL) can be applied in HIC
    - WSI → bag, patch → instance
    - EM-DD, MI Kernel, mi-SVM, miGraph
  - **MCIL [2]** is the state-of-the-art method designed for HIC
    - adopts the clustering concept into MIL
- all are based on asymmetric assumption
  - not suited for cancer type classification

## Asymmetric assumption



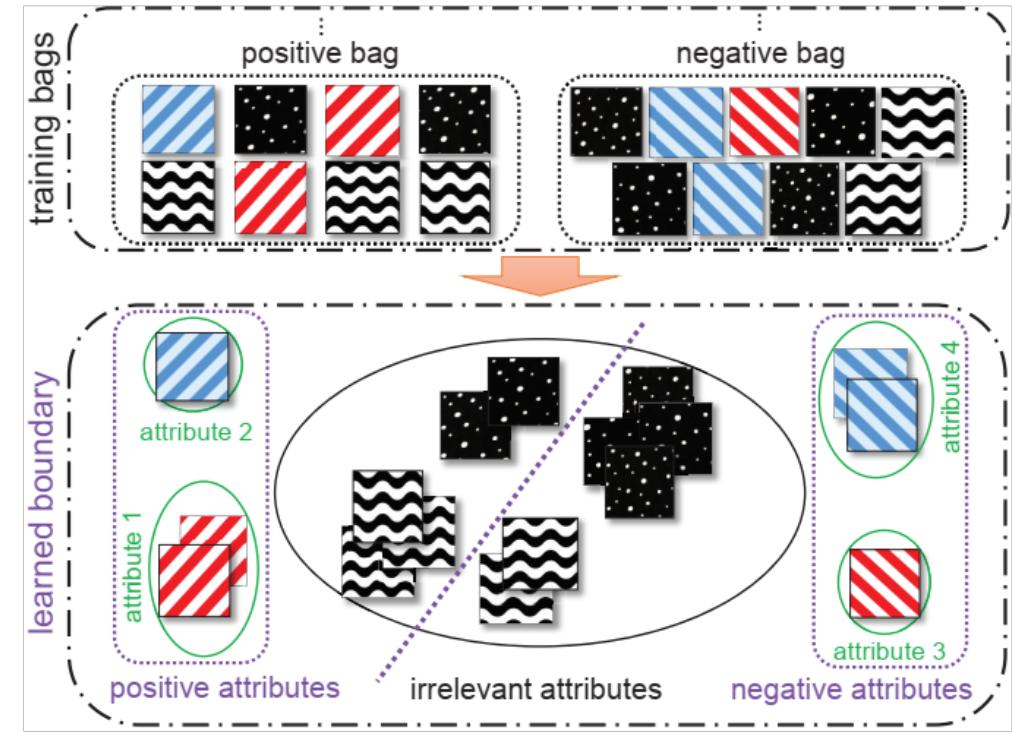
\* Figure from E.D. Ross and A. Ben-Hur. "Amino acid composition predicts prion activity". PLoS computational biology, 13(4), p.e1005465, 2017.

# Our contribution

- A novel symmetric MIL framework for classifying cancer types based on histopathological images.
- A probabilistic graphical model that incorporates the proposed paradigm and an efficient inference for learning model parameters.
- Extensive experiment results:
  - runtime evaluation,
  - real-world datasets.

# Problem Formulation

- Given bags and their labels  $\{X_b, Y_b\}_{b=1}^B$ 
  - $X_b = \{x_{bi}\}_{i=1}^{n_b} \subset \mathbb{R}^d$
  - $Y_b \in \{0,1\}$
- Symmetric MIL assumptions:
  - Symmetric attributes:** each instance in a bag can be associated with an attribute that is either *relevant* or *irrelevant* to the bag label.
  - No mixed-class attributes:** positive bags do not contain negative attributes and vice versa.
  - Cardinality constraint:** each bag has a limited number of relevant instances.



# Attribute-Based Symmetric MIL

- Attribute  $z_{bi}$  of  $i$ th instance in  $b$ th bag follows a multinomial LR

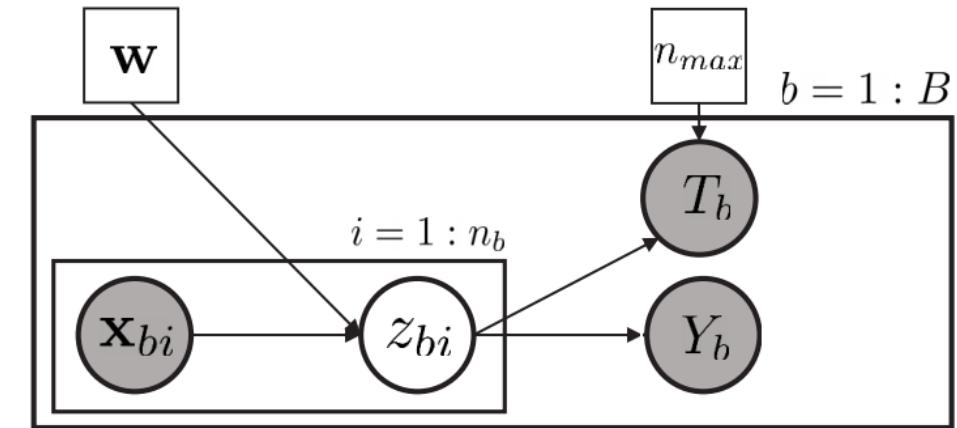
$$P_{bic}(\mathbf{w}) = P(z_{bi} = c \mid \mathbf{x}_{bi}, \mathbf{w}) = \frac{e^{\mathbf{w}_c^T \mathbf{x}_{bi}}}{\sum_{c=0}^{\mathbb{C}} e^{\mathbf{w}_c^T \mathbf{x}_{bi}}}$$

- Cardinality constraint is enforced on every bag

$$T_b = \left( I_{\sum_{i=1}^{n_b} I_{z_{bi} \neq 0} \geq 1} \right) \left( I_{\sum_{i=1}^{n_b} I_{z_{bi} \neq 0} \leq n_{\max}} \right) = 1$$

- Bag label is computed based on the presence of relevant attributes

$$Y_b = \begin{cases} 0 & \text{if } \bigcup_{i=1}^{n_b} \{z_{bi}\} \subseteq \{0, 1, \dots, \frac{\mathbb{C}}{2}\} \\ 1 & \text{if } \bigcup_{i=1}^{n_b} \{z_{bi}\} \subseteq \{0, \frac{\mathbb{C}}{2} + 1, \dots, \mathbb{C}\} \\ 2 & \text{otherwise} \end{cases}$$



Positive attributes

$$\mathcal{C}^+ = \left\{ 1, \dots, \frac{\mathbb{C}}{2} \right\}$$

Negative attributes

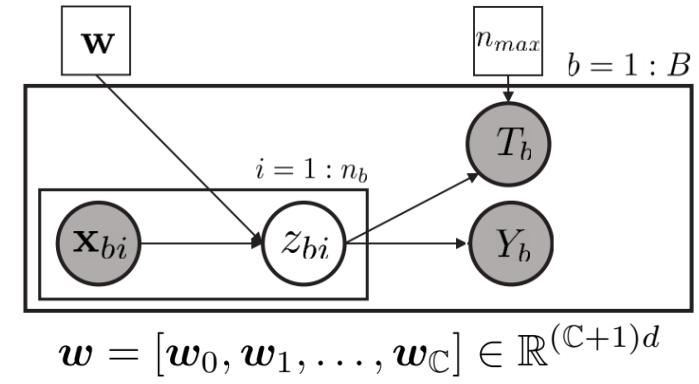
$$\mathcal{C}^- = \left\{ \frac{\mathbb{C}}{2} + 1, \dots, \mathbb{C} \right\}$$

Irrelevant attributes

$$\{0\}$$

# Inference

- Regularized maximum likelihood



$$\mathbb{L}(\mathbf{w}) = \frac{1}{B} \sum_{b=1}^B \left( \mathbb{L}_b(\mathbf{w}) + \lambda_e \mathbb{H}_b(\mathbf{w}) \right) + \frac{\lambda_q \|\mathbf{w}\|^2}{2}$$

normalized negative incomplete log-likelihood

$$\mathbb{L}_{incm}(\mathbf{w}) = -\frac{1}{B} \sum_{b=1}^B \left( \log P(Y_b, T_b \mid \mathbf{X}_b, \mathbf{w}) + \log P(\mathbf{X}_b) \right)$$

entropy regularization

$$\mathbb{H}_b(\mathbf{w}) = \sum_{i=1}^{n_b} \left( - \sum_{c=0}^{\mathbb{C}} P_{bic}(\mathbf{w}) \log P_{bic}(\mathbf{w}) \right)$$

- Challenge: minimizing  $\mathbb{L}$  w.r.t.  $\mathbf{w}$  requires computing  $P(Y_b, T_b \mid \mathbf{X}_b, \mathbf{w})$

$$P(Y_b = 1, T_b \mid \mathbf{X}_b, \mathbf{w}) = \sum_{\mathbf{z}_b \in \{0,1,\dots,\frac{\mathbb{C}}{2}\}^{n_b}} I_{\sum_{i=1}^{n_b} I_{z_{bi} \neq 0} \geq 1} I_{\sum_{i=1}^{n_b} I_{z_{bi} \neq 0} \leq n_{\max}} \prod_{i=1}^{n_b} P(z_{bi} \mid \mathbf{x}_{bi}, \mathbf{w})$$

# Expectation Maximization (EM) Approach

- Negative complete log-likelihood

$$\mathbb{L}_{cm}(\boldsymbol{w}) = -\frac{1}{B} \sum_{b=1}^B \log P(Y_b, T_b, \boldsymbol{z}_b \mid \boldsymbol{X}_b, \boldsymbol{w}) + const$$

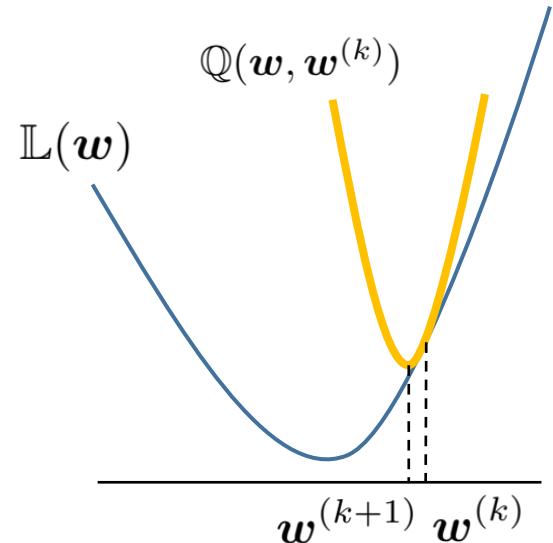
- E-step:

$$\mathbb{Q}(\boldsymbol{w}, \boldsymbol{w}^{(k)}) = \frac{1}{B} \sum_{b=1}^B \left( \mathbb{J}_b(\boldsymbol{w}, \boldsymbol{w}^{(k)}) + \lambda_e \mathbb{H}_b(\boldsymbol{w}) + \frac{\lambda_q \|\boldsymbol{w}\|^2}{2} \right)$$

$$\boxed{\mathbb{J}_b(\boldsymbol{w}, \boldsymbol{w}^{(k)}) = E_{\boldsymbol{z}_b | Y_b, T_b, \boldsymbol{X}_b, \boldsymbol{w}^{(k)}} \left[ -\log P(Y_b, T_b, \boldsymbol{z}_b \mid \boldsymbol{X}_b, \boldsymbol{w}) \right]}$$

- M-step:

$$\boldsymbol{w}^{(k+1)} = \operatorname{argmin}_{\boldsymbol{w}} \mathbb{Q}(\boldsymbol{w}, \boldsymbol{w}^{(k)}) \xrightarrow{\text{Generalized EM}} \text{find } \boldsymbol{w}^{(k+1)} \text{ s.t. } \mathbb{Q}(\boldsymbol{w}^{(k+1)}, \boldsymbol{w}^{(k)}) \leq \mathbb{Q}(\boldsymbol{w}^{(k)}, \boldsymbol{w}^{(k)})$$



## E-step

$$-\sum_{c=0}^{\mathbb{C}} I_{z_{bi}=c} \mathbf{w}_c^T \mathbf{x}_{bi} + \log \left( \sum_{c=0}^{\mathbb{C}} e^{\mathbf{w}_c^T \mathbf{x}_{bi}} \right)$$

$$\mathbb{J}_b(\mathbf{w}, \mathbf{w}^{(k)}) = E_{\mathbf{z}_b | Y_b, T_b, \mathbf{X}_b, \mathbf{w}^{(k)}} \left[ -\log P(Y_b, T_b, \mathbf{z}_b | \mathbf{X}_b, \mathbf{w}) \right]$$

$$= \sum_{i=1}^{n_b} \left( - \sum_{c=0}^{\mathbb{C}} P_{bic}^{post}(\mathbf{w}^{(k)}) \cdot \mathbf{w}_c^T \mathbf{x}_{bi} + \log \left( \sum_{c=0}^{\mathbb{C}} e^{\mathbf{w}_c^T \mathbf{x}_{bi}} \right) \right)$$

$$\begin{aligned} P_{bic}^{post}(\mathbf{w}) &= P(z_{bi} = c | Y_b, T_b, \mathbf{x}_{bi}, \mathbf{w}) \\ &= \frac{P(z_{bi} = c, Y_b, T_b | \mathbf{x}_{bi}, \mathbf{w})}{\sum_{t=0}^{\mathbb{C}} P(z_{bi} = t, Y_b, T_b | \mathbf{x}_{bi}, \mathbf{w})} \end{aligned}$$

Requires marginalization over all  
 $z_{bj}$  for  $j = 1, \dots, n_b$  and  $j \neq i$   
 $\rightarrow O(\mathbb{C}^{n_b-1})!$

# Dynamic Programming for E-step

- Forward message passing

$$\alpha_{bi}(l) \triangleq P(N_{bi} = l \mid \mathbf{X}_b, \mathbf{w})$$

- Backward message passing

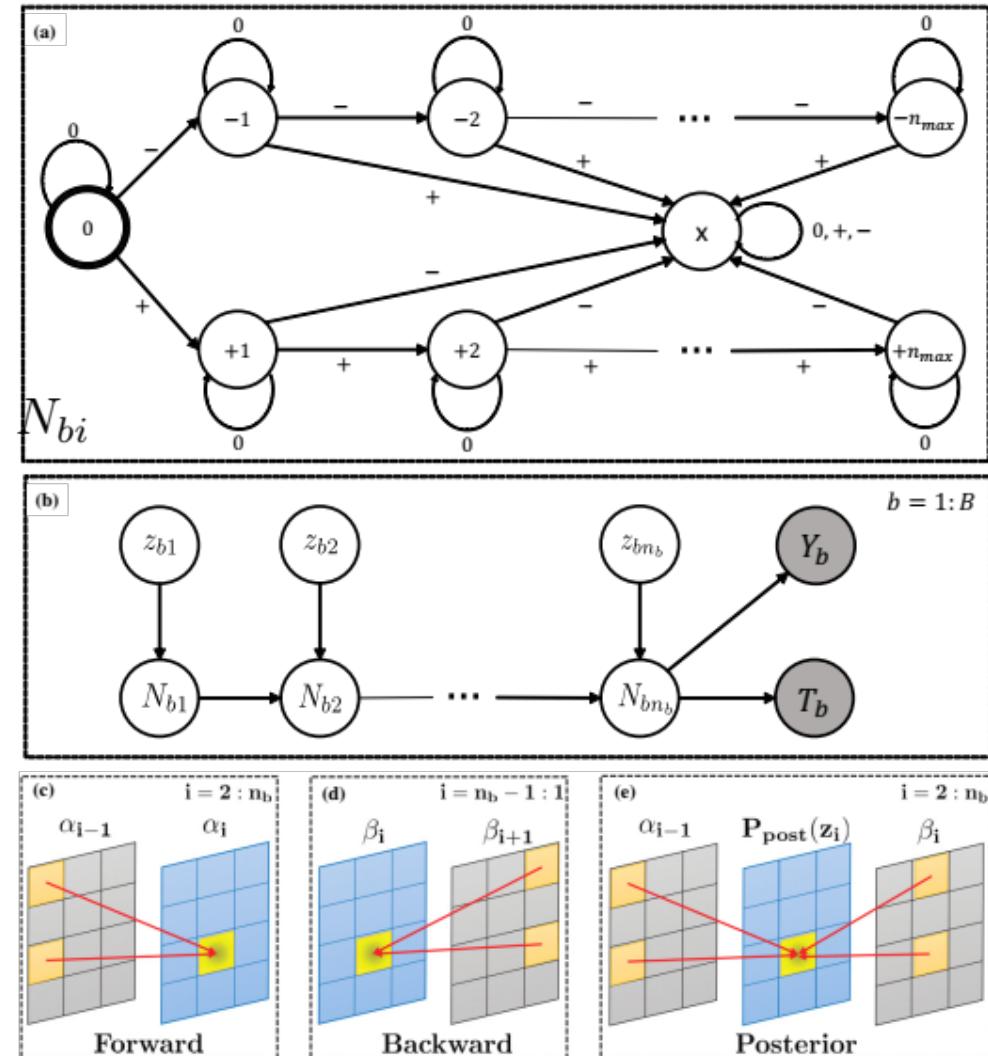
$$\beta_{bi}(l) \triangleq P(Y_b, T_b \mid N_{bi} = l, \mathbf{X}_b, \mathbf{w})$$

- Joint probability calculation

$$P_{bic}^{joint}(\mathbf{w}) \triangleq P(z_{bi} = c, Y_b, T_b \mid \mathbf{x}_{bi}, \mathbf{w})$$

- Posterior probability calculation

$$\begin{aligned} P_{bic}^{post}(\mathbf{w}) &= P(z_{bi} = c \mid Y_b, T_b, \mathbf{x}_{bi}, \mathbf{w}) \\ &= \frac{P(z_{bi} = c, Y_b, T_b \mid \mathbf{x}_{bi}, \mathbf{w})}{\sum_{t=0}^{\mathbb{C}} P(z_{bi} = t, Y_b, T_b \mid \mathbf{x}_{bi}, \mathbf{w})} \end{aligned}$$



# AbSMIL Algorithm

- M-step with stochastic GD

$$\begin{aligned} \frac{\partial \mathbb{Q}_b(\mathbf{w}, \mathbf{w}')}{\partial \mathbf{w}_c} = & \sum_{i=1}^{n_b} \left( P_{bic}(\mathbf{w}) - P_{bic}^{post}(\mathbf{w}') \right) \mathbf{x}_{bi} + \\ & \lambda_e \sum_{i=1}^{n_b} P_{bic}(\mathbf{w}) \left( \sum_{t=0}^{\mathbb{C}} P_{bit}(\mathbf{w}) (\mathbf{w}_t - \mathbf{w}_c)^T \mathbf{x}_{bi} \right) \mathbf{x}_{bi} + \lambda_q \mathbf{w}_c \end{aligned}$$

- Prediction

$$\hat{z}_{bi} = \operatorname{argmax}_{0 \leq c \leq \mathbb{C}} P(z_{bi} = c \mid \mathbf{x}_{bi}, \mathbf{w})$$

$$\hat{Y}_b = \operatorname{argmax}_{m \in \{0,1\}} P(Y_b = m, T_b = 1 \mid \mathbf{X}_b, \mathbf{w})$$

---

**Algorithm 1** Attribute-based Symmetric Multiple Instance Learning (AbSMIL)

---

```

1: Input: Training data  $\{\mathbf{X}_b, Y_b\}_{b=1}^B$ , cardinality constraint  

    $n_{\max}$ , positive constants  $\lambda_q$  and  $\lambda_e$ , initial weight  $\mathbf{w}^{(0)}$ 
2: Output:  $\{\mathbf{w}^{(k)}\}$ 
3:  $k = 0$ 
4: repeat
5:   Select a random bag  $b$ 
6:   // E-step:
7:   Compute prior probability  $P_{bic}(\mathbf{w}^{(k)})$  using (1)
8:   Compute prior probability  $P_{bi}^0, P_{bi}^+$  and  $P_{bi}^-$  using (9)
9:   Compute forward message  $\alpha_{bi}(l)$  for  $i = 1, \dots, n_b$  and  

    $l = 0, \pm 1, \dots, \pm n_{\max}$  using (10) and (11)
10:  Compute backward message  $\beta_{bi}(l)$  for  $i = n_b, \dots, 1$   

    and  $l = 0, \pm 1, \dots, \pm n_{\max}$  using (12) and (13)
11:  Compute joint probability  $P_{bic}^{joint}(\mathbf{w}^{(k)})$  for  $i =$   

     $1, \dots, n_b$  and  $c = 0, 1, \dots, \mathbb{C}$  using (14) and (15)
12:  Compute posterior probability  $P_{bic}^{post}(\mathbf{w}^{(k)})$  for  $i =$   

     $1, \dots, n_b$  and  $c = 0, 1, \dots, \mathbb{C}$  using (7)
13: // M-step:
14:  $\tau = \sqrt{\frac{2(\lambda_e+1)n_b}{\lambda_q} \log(\mathbb{C}+1)}$ 
15: Compute  $\frac{\partial \mathbb{Q}_b(\mathbf{w}, \mathbf{w}^{(k)})}{\partial \mathbf{w}_c}$  for  $c = 0, 1, \dots, \mathbb{C}$  using (8)
16:  $\mathbf{w}^{(k+1)} = \Pi_\tau \left( \mathbf{w}^{(k)} - \frac{1}{k\lambda_q} \frac{\partial \mathbb{Q}_b(\mathbf{w}, \mathbf{w}^{(k)})}{\partial \mathbf{w}^{(k)}} \right)$ 
17:  $k = k + 1$ 
18: until stopping criteria is met

```

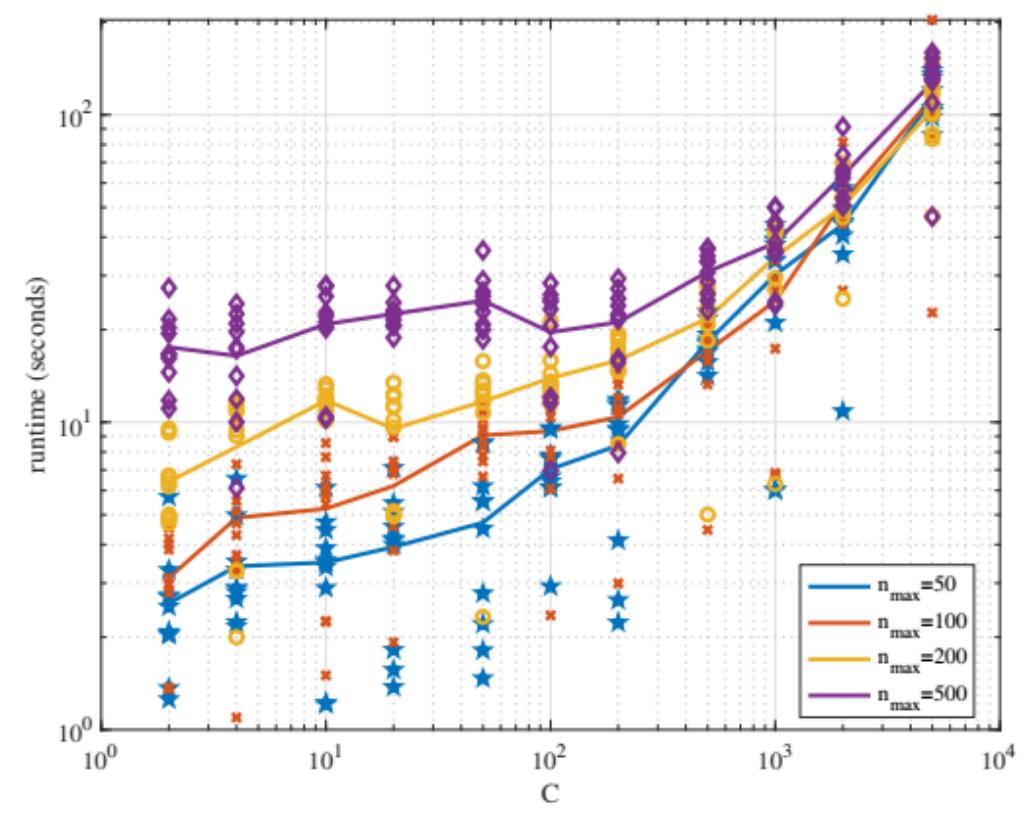
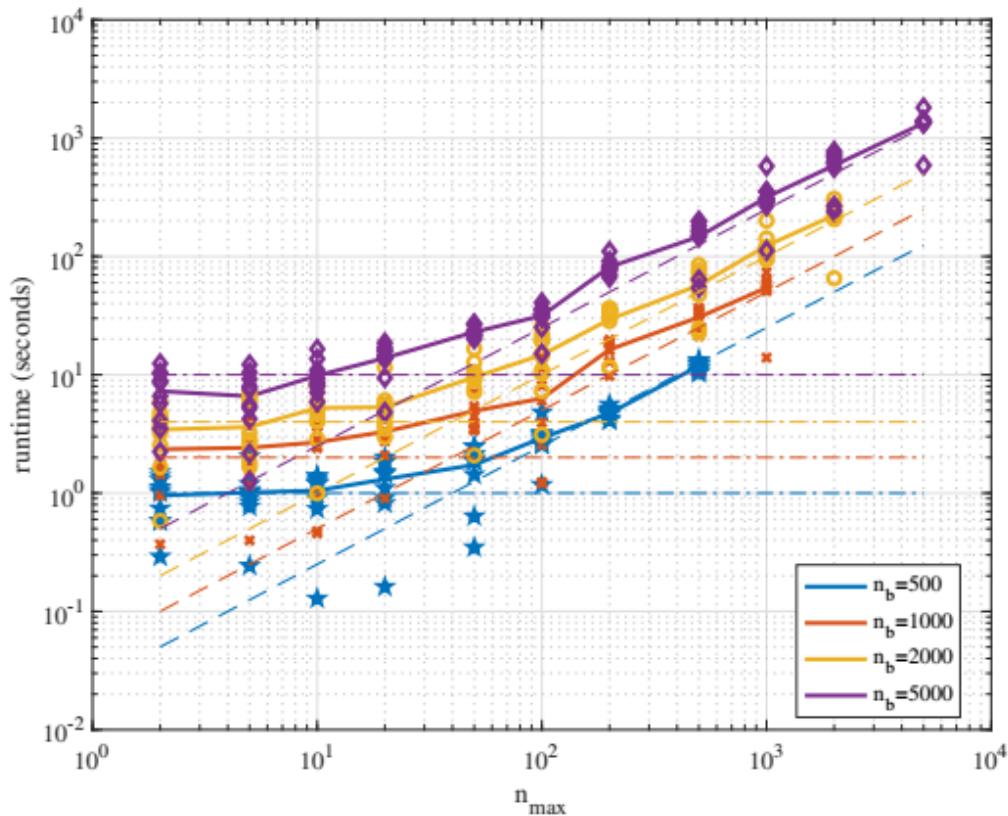
---

$$\Pi_\tau(\mathbf{v}) = \min \left\{ 1, \frac{\tau}{\|\mathbf{v}\|} \right\} \mathbf{v}$$

# Experiments

# Runtime Evaluation

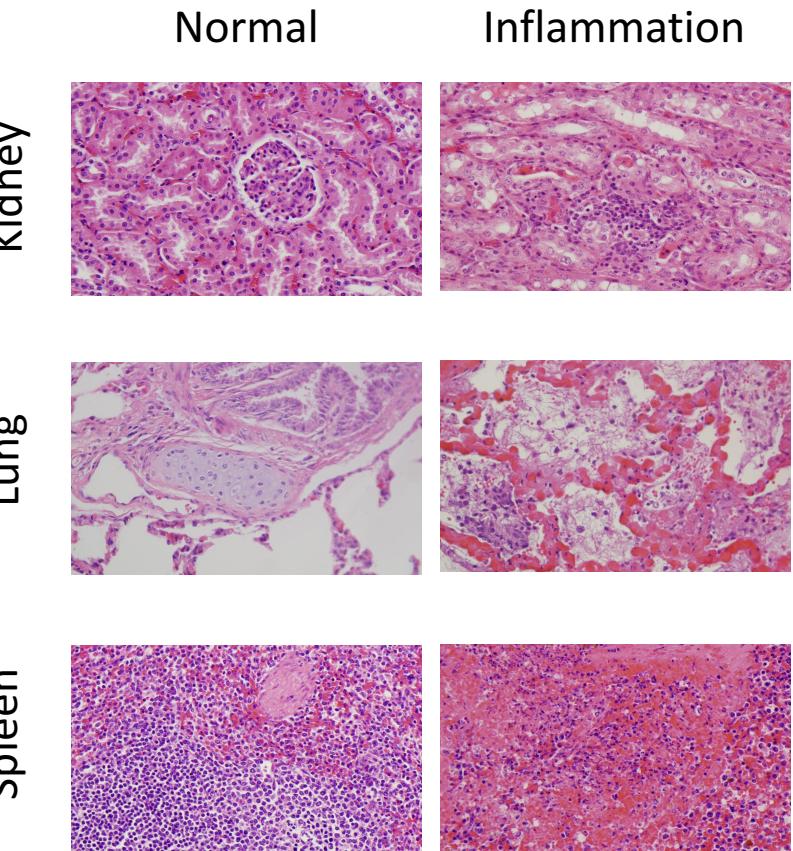
$d = 30$



$$O(n_b(\mathbb{C}d + n_{\max}))$$

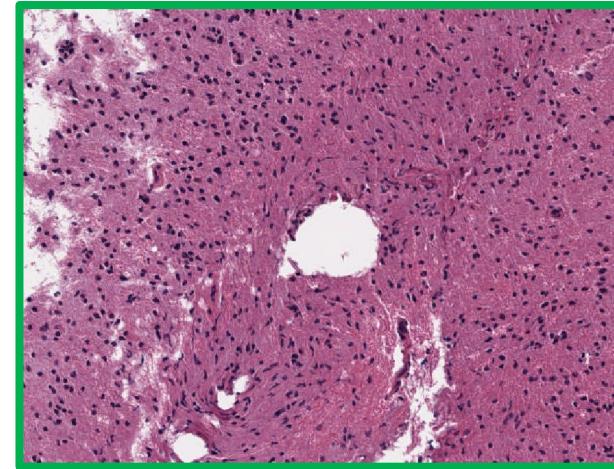
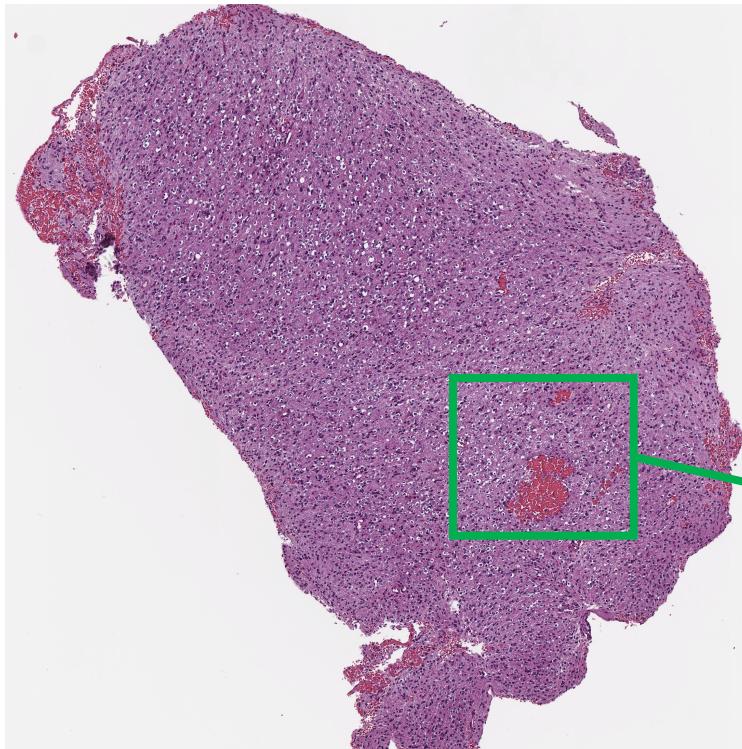
# Real-world Datasets

- Datasets
  - Tiger, Fox, and Elephant datasets
    - popularly used in studies of MIL
  - Animal Diagnostics Lab datasets:
    - Kidney, Lung, and Spleen
    - histopathology images of mammalian organs
    - 300 WSIs of size 4000x3000 in each dataset
  - The Cancer Genome Atlas (TCGA) dataset
    - 48 samples for astrocytoma
    - 48 samples for oligodendrogloma
    - Various sizes and shapes
- Baselines
  - Fully-supervised: [DFDL](#)
  - MIL-based: mi-SVM, MIL-Boost, miGraph, [MCIL](#)
  - MIML-based: ORLR, MIML-NC

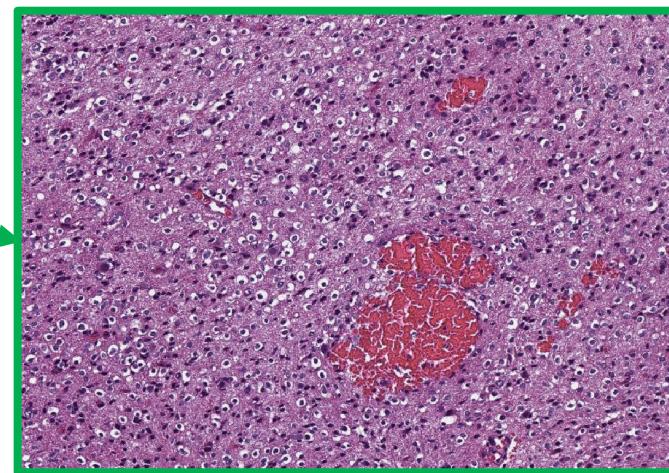
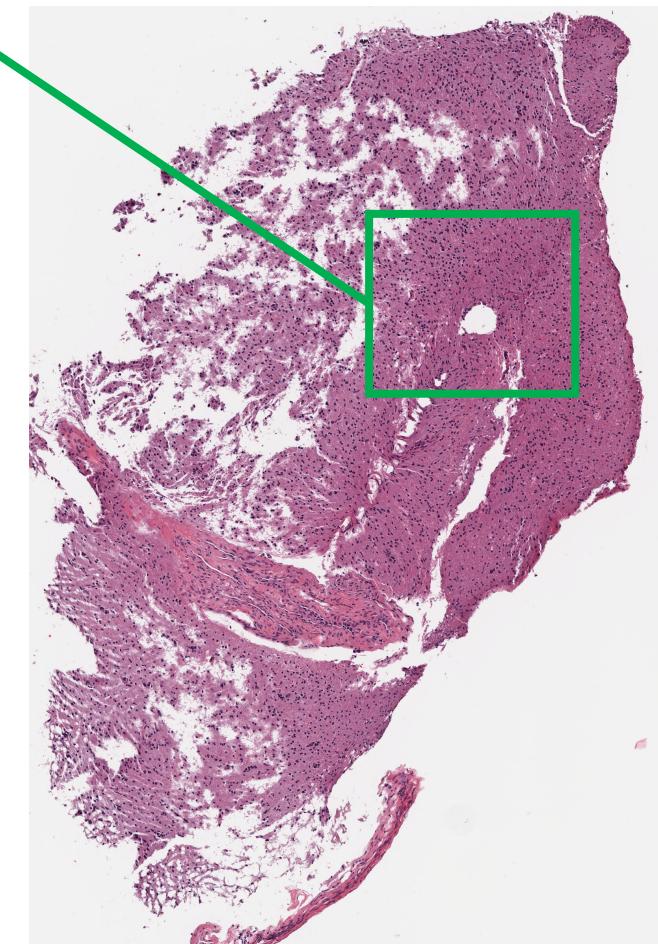


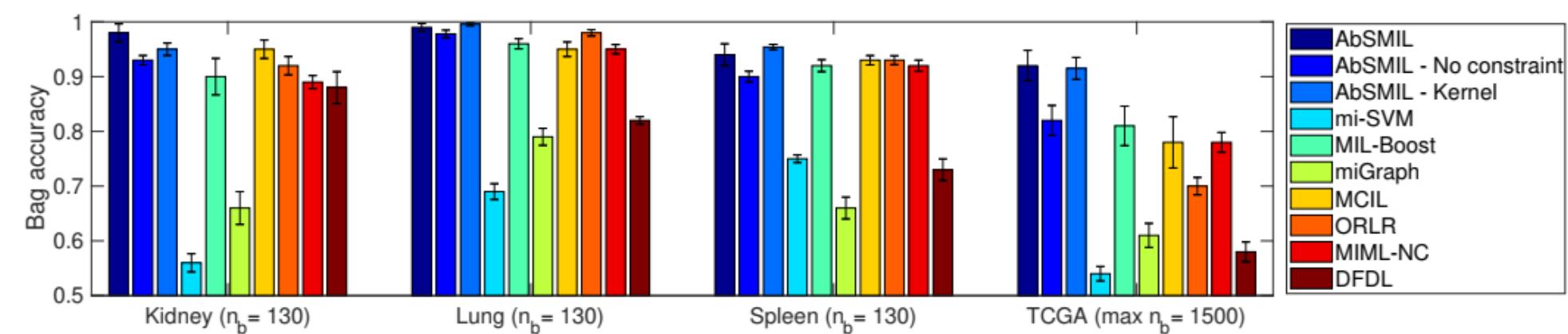
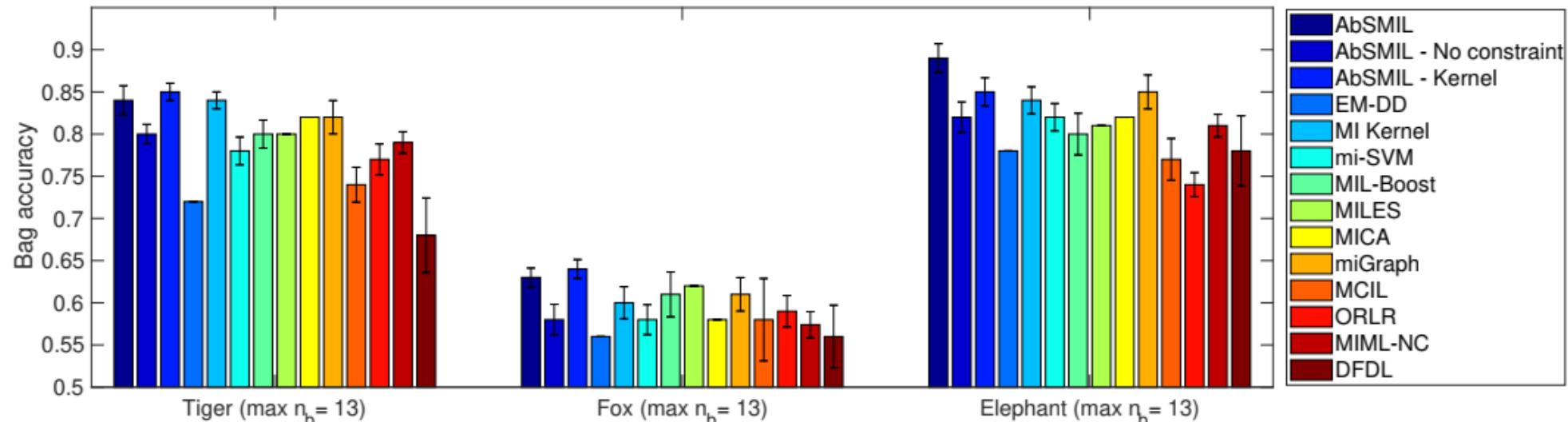
# TCGA dataset

Astrocytoma

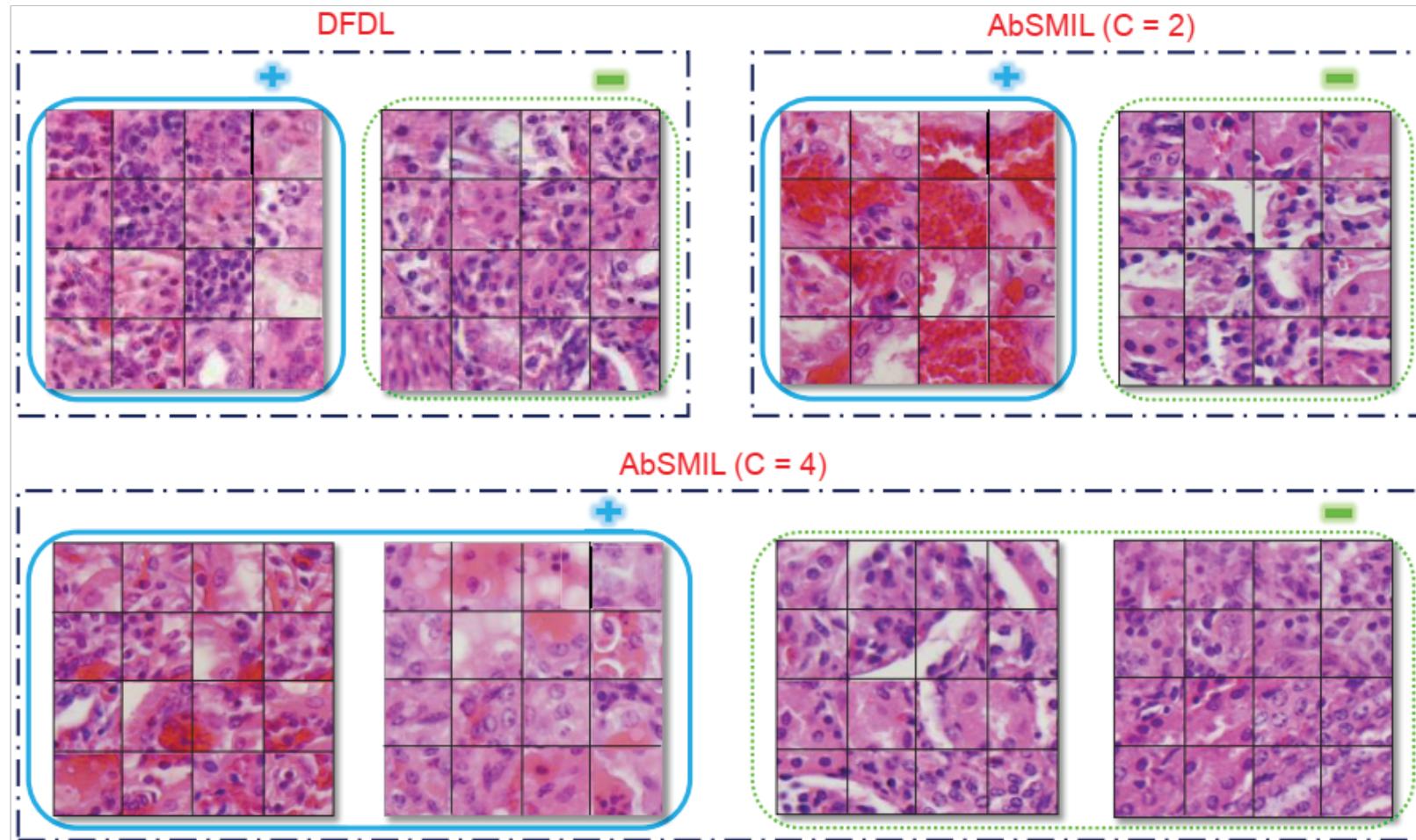


Oligodendrogioma





# Visualization from Kidney dataset



# Conclusion

- Introduced a Symmetric MIL setting where both positive and negative bags contain relevant class-specific instances as well as irrelevant instances.
- Proposed a probabilistic model that takes into account prior information about the sparsity of the relevant instances.
- Developed an efficient inference that is linear in the number of instances and is suitable for online learning scenarios.
- Evaluated our framework on the real-world datasets and obtained competitive results on all datasets and in particular for TCGA where bags contain mainly irrelevant instances

# References

1. T. H. Vu, H. S. Mousavi, V. Monga, G. Rao, and U. A. Rao  
*“Histopathological image classification using discriminative feature oriented dictionary learning,”* IEEE Trans. Med. Imag., vol. 35, no. 3, pp. 738–751, 2016.
2. Y. Xu, J.-Y. Zhu, E. Chang, and Z. Tu, “Multiple clustered instance learning for histopathology cancer image classification, segmentation and clustering,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. IEEE, 2012, pp. 964–971.

Thank you!