UNIVERSIDAD DE GRANADA

torchmil

Multiple Instance Learning for Histopathology

An Introduction to torchmil

RISE-MICCAI Tutorial

Francisco Miguel Castro-Macías

Outline

1. Motivation
2. Multiple Instance Learning
3. torchmil

Motivation

Machine Learning for Histopathology

Machine Learning for Histopathology

$$\begin{Bmatrix}\text{Computational}\\ \text{Pathology}\end{Bmatrix} = \begin{Bmatrix}\text{Computer}\\ \text{Science}\end{Bmatrix} \bigcap \begin{Bmatrix}\text{Pathology}\end{Bmatrix}$$

For what?

- Automatic analysis of pathology slides.
- Enhanced diagnostic accuracy.
- Biomarker discovery.
- Many more!

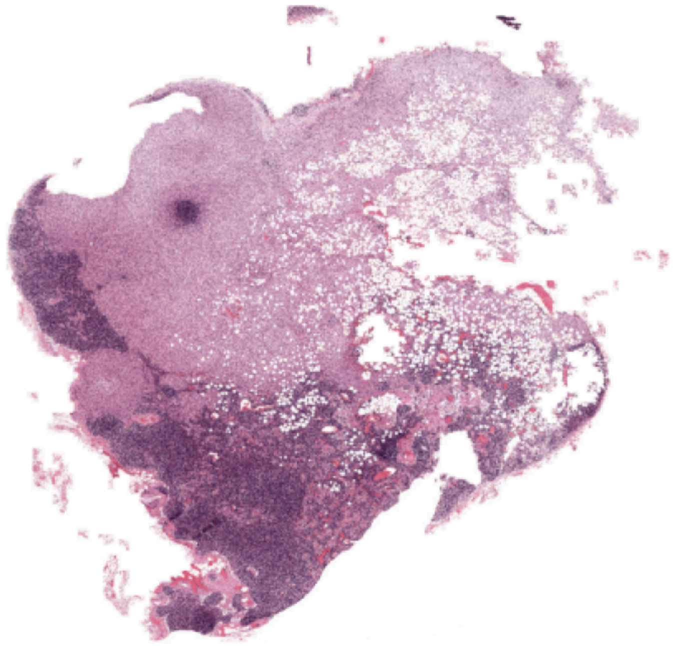


Figure: Whole Slide Image (WSI).

Our End Goal

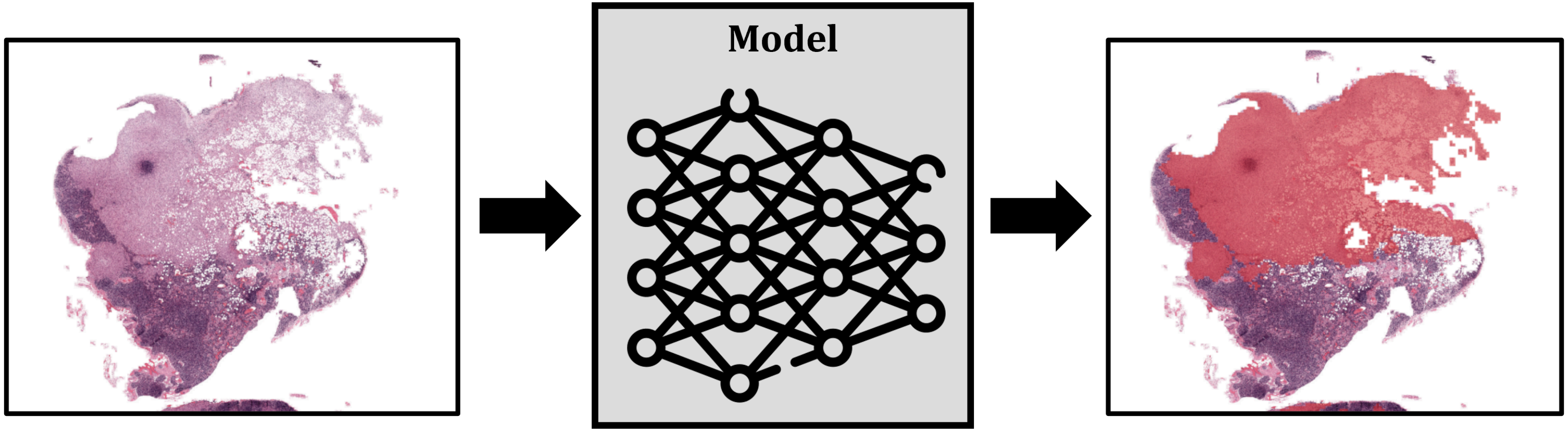


Figure: A trained model predicts the presence and location of tumor tissue in a WSI.

Our End Goal

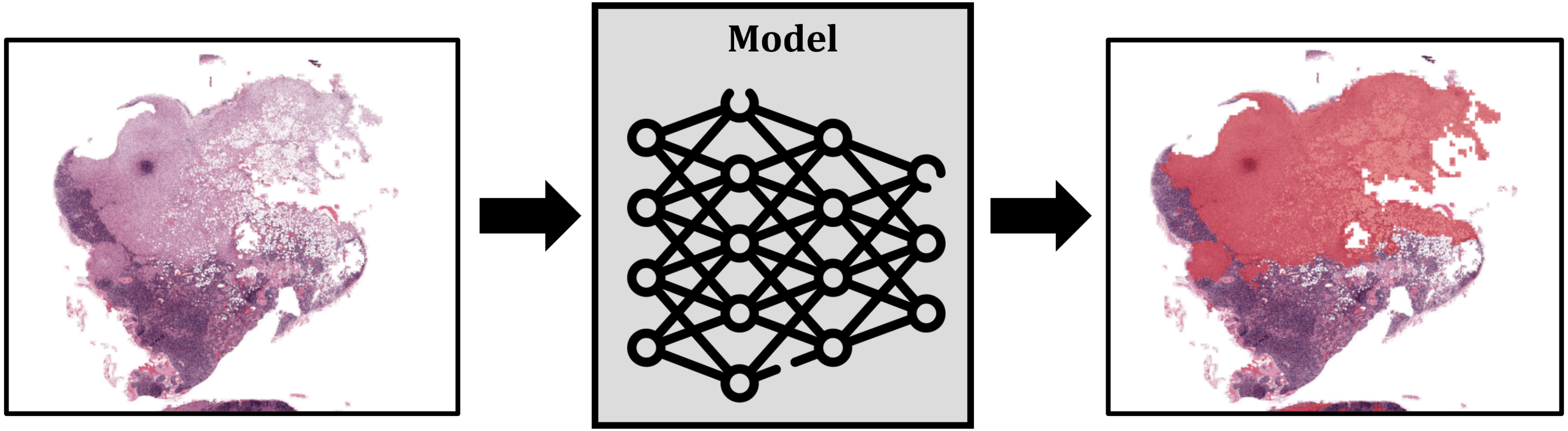


Figure: A trained model predicts the presence and location of tumor tissue in a WSI.

Problem. Training models on WSIs is not easy 😢

Working with WSIs

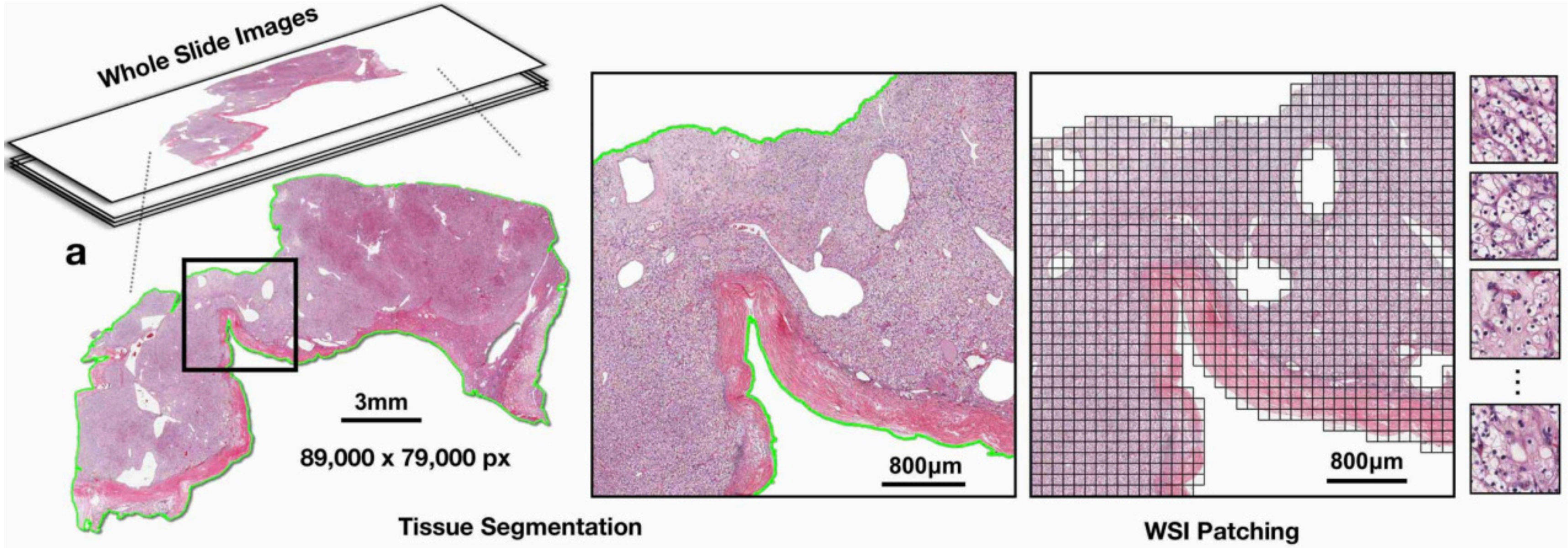


Figure: WSIs must be segmented and divided into patches (Lu et al., 2021).

Good news. This has been standardized by TRIDENT (Zhang et al., 2025).

Training a WSI classifier: first attempt

Idea. (1) Ask one or more pathologists to label all the patches in a set of WSIs. Then, (2) train a model to predict the label of each path.

Training a WSI classifier: first attempt

Idea. (1) Ask one or more pathologists to label all the patches in a set of WSIs. Then, (2) train a model to predict the label of each path.

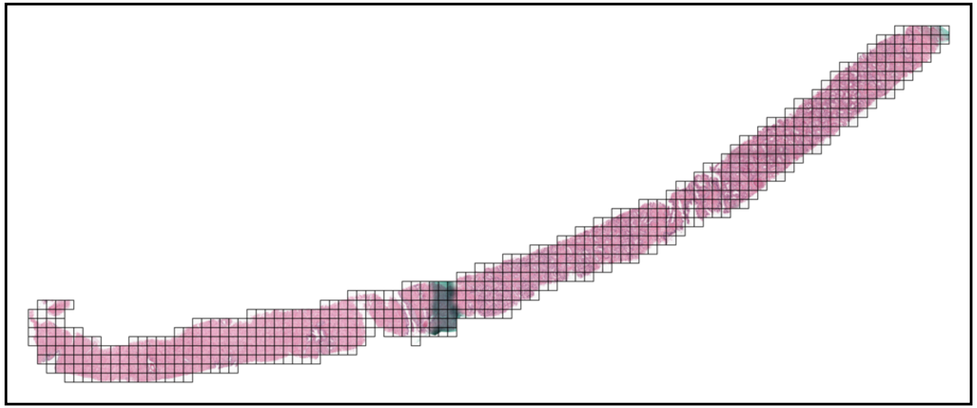


Figure: Patches without labels.

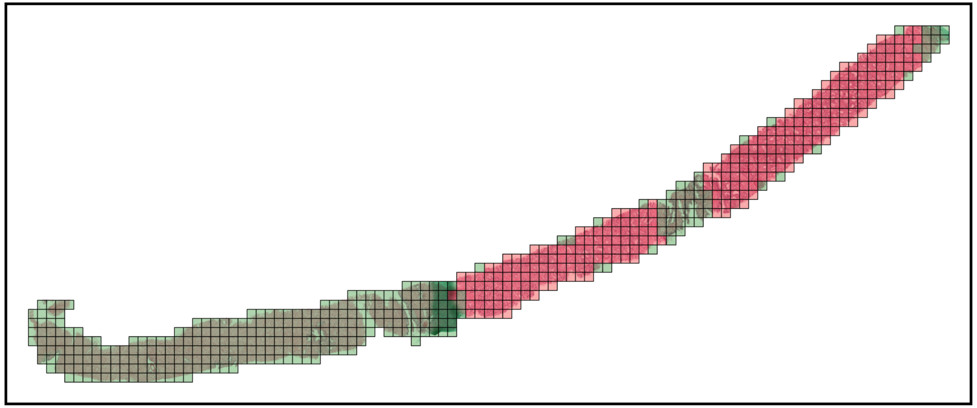


Figure: Labelled patches.

Training a WSI classifier: first attempt

Idea. (1) Ask one or more pathologists to label all the patches in a set of WSIs. Then, (2) train a model to predict the label of each path.

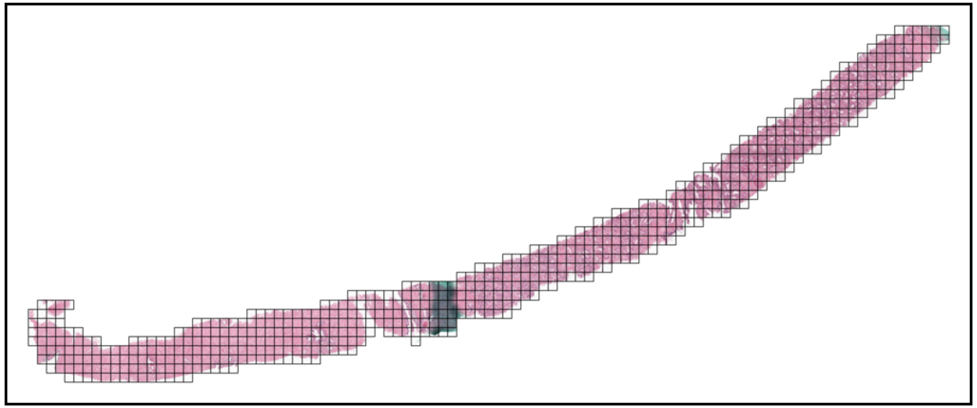


Figure: Patches without labels.

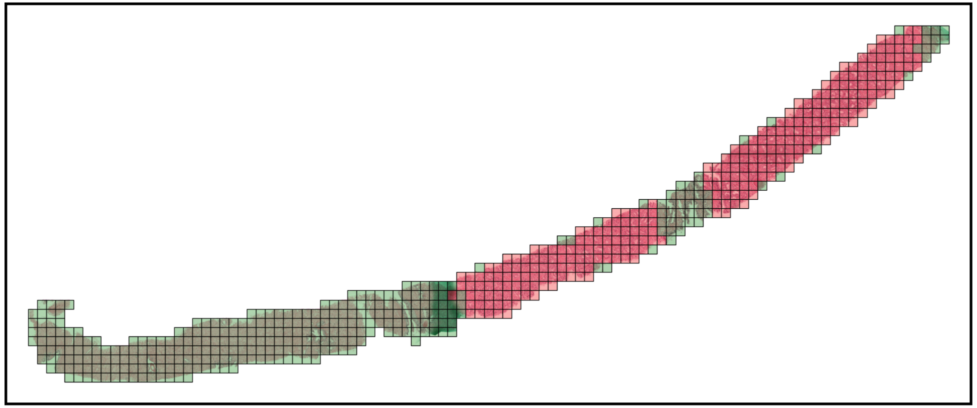


Figure: Labelled patches.

Problems.

- The labelling process is extremely time-consuming.
- Pathologists don't always agree on the labelling.
- The label of one patch might depend on the patches surrounding it.
- And more…

Multiple Instance Learning (MIL)

Learning (almost) without labels

An observation about WSIs

A WSI can be seen as a set of patches.

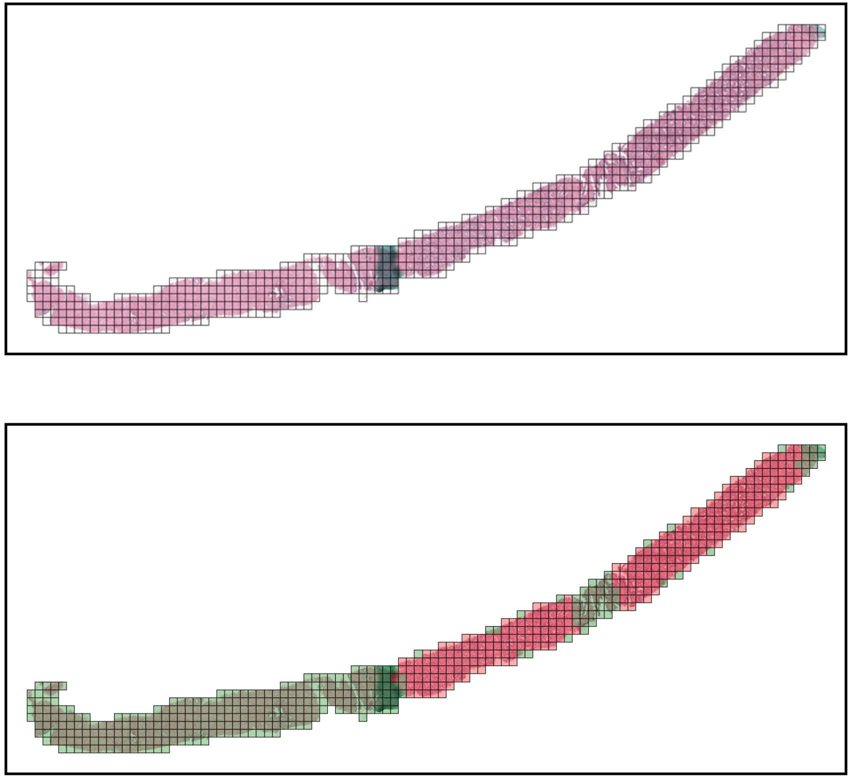
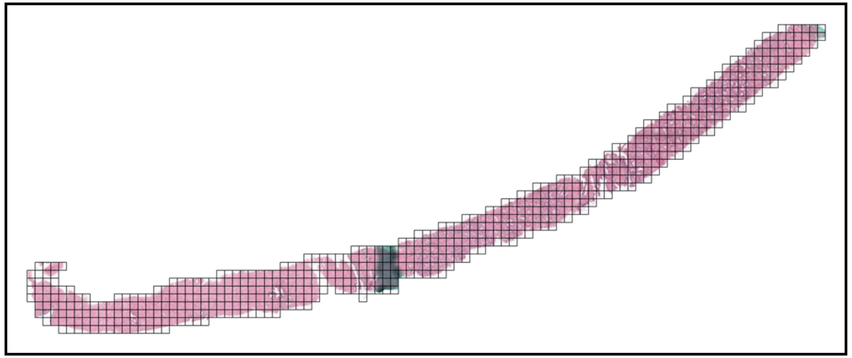


Figure: WSI (top) with patch labels (bottom).

An observation about WSIs

A WSI can be seen as a set of patches.

The labels of the patches are not available.



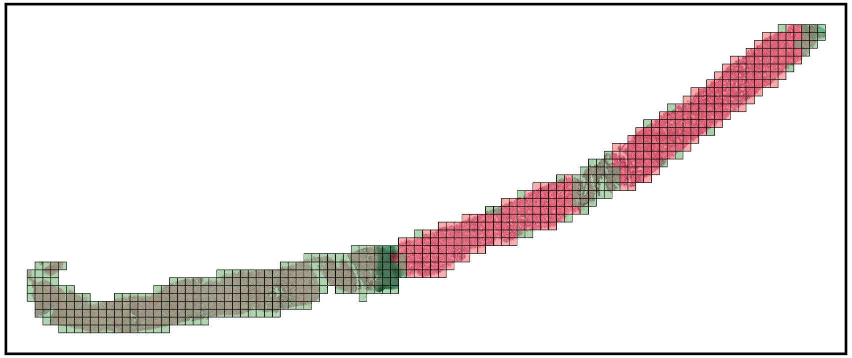


Figure: WSI (top) with patch labels (bottom).

An observation about WSIs

A WSI can be seen as a set of patches.

The labels of the patches are not available.

But the WSI has a **global label**:
- It is “healthy” if all the patches are “healthy”.
- It is “cancerous” if at least one patch is “cancerous”.

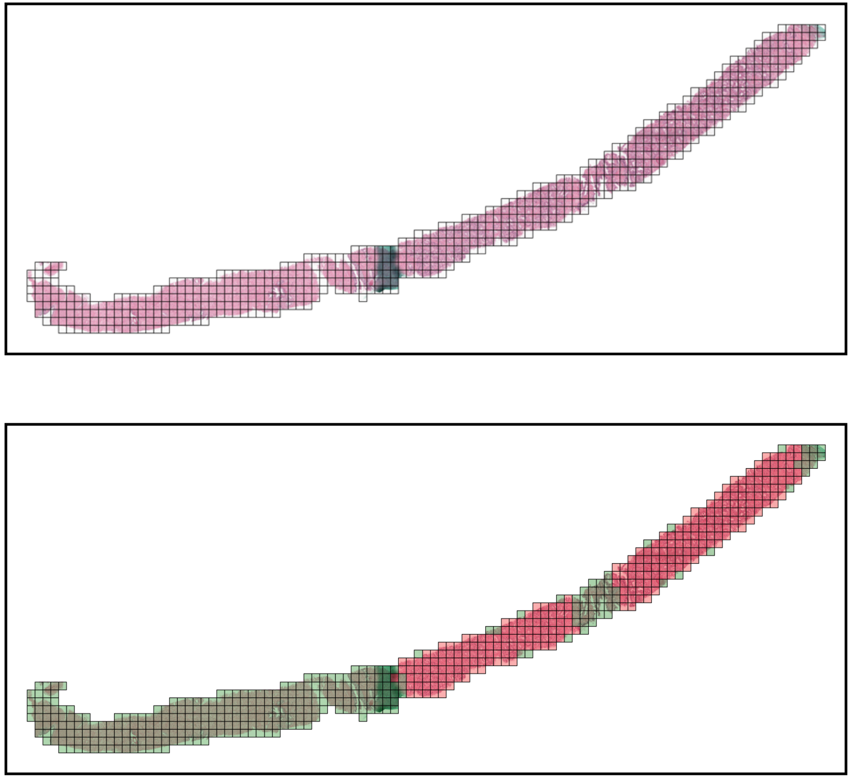


Figure: WSI (top) with patch labels (bottom).

An observation about WSIs

A WSI can be seen as a set of patches.

The labels of the patches are not available.

But the WSI has a **global label**:
- It is "healthy" if all the patches are "healthy".
- It is "cancerous" if at least one patch is "cancerous".

This **global label** is easy to obtain: medical records.

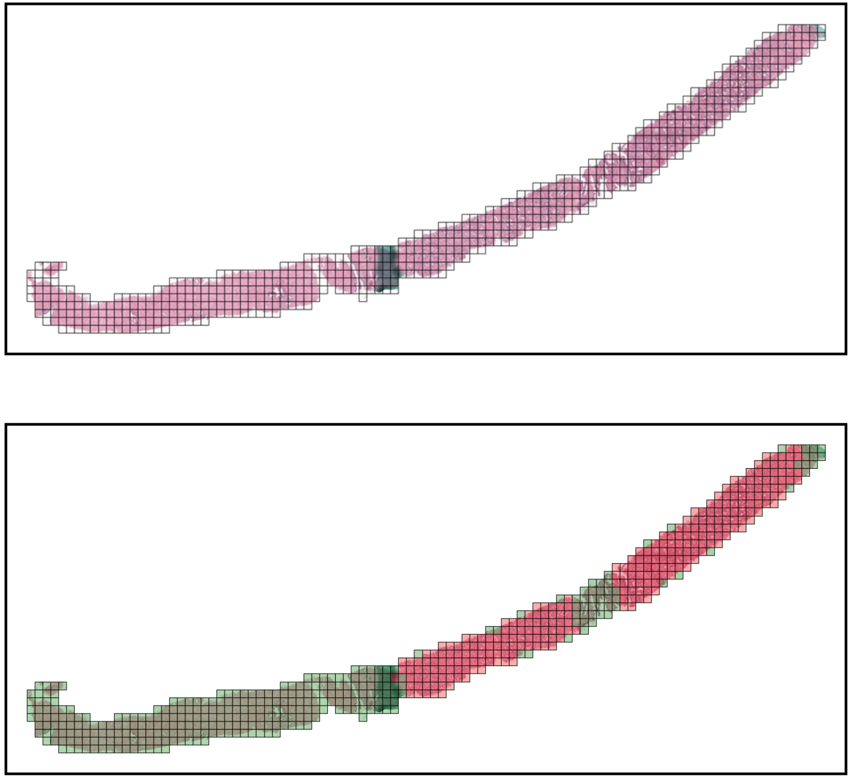


Figure: WSI (top) with patch labels (bottom).

An observation about WSIs

A WSI can be seen as a set of patches.

The labels of the patches are not available.

But the WSI has a **global label**:
- It is "healthy" if all the patches are "healthy".
- It is "cancerous" if at least one patch is "cancerous".

This **global label** is easy to obtain: medical records.

Can we learn using only the **global label**?

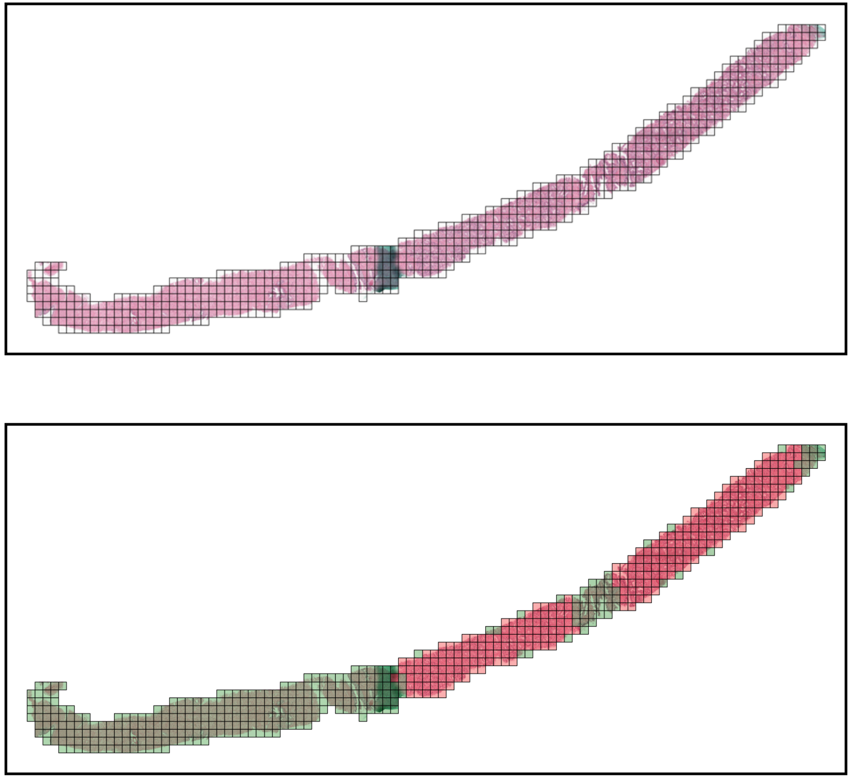


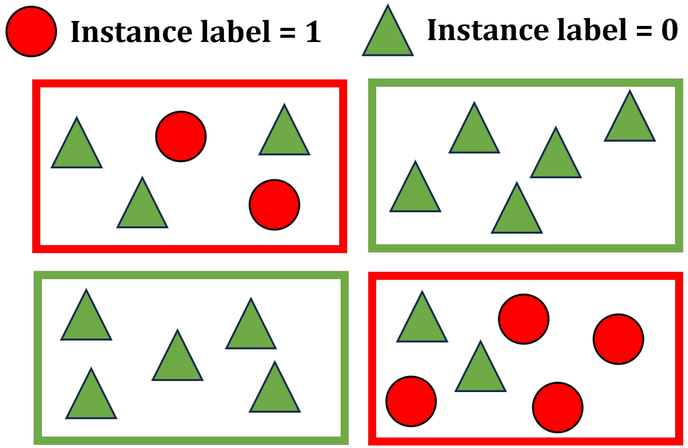
Figure: WSI (top) with patch labels (bottom).

Multiple Instance Learning (MIL)

Data. Pairs of the form $(\mathbf{X}, Y)$:

$$\underbrace{\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}}_{\text{bag = set of instances}} \subset \mathbb{R}^D, \quad \underbrace{\mathbf{x}_n}_{\text{instance}} \in \mathbb{R}^D,$$

$$\underbrace{y_n}_{\text{instance label}} \in \{0, 1\}, \quad \underbrace{Y}_{\text{bag label}} \in \{0, 1\}$$



Multiple Instance Learning (MIL)

Data. Pairs of the form $(\mathbf{X}, Y)$:

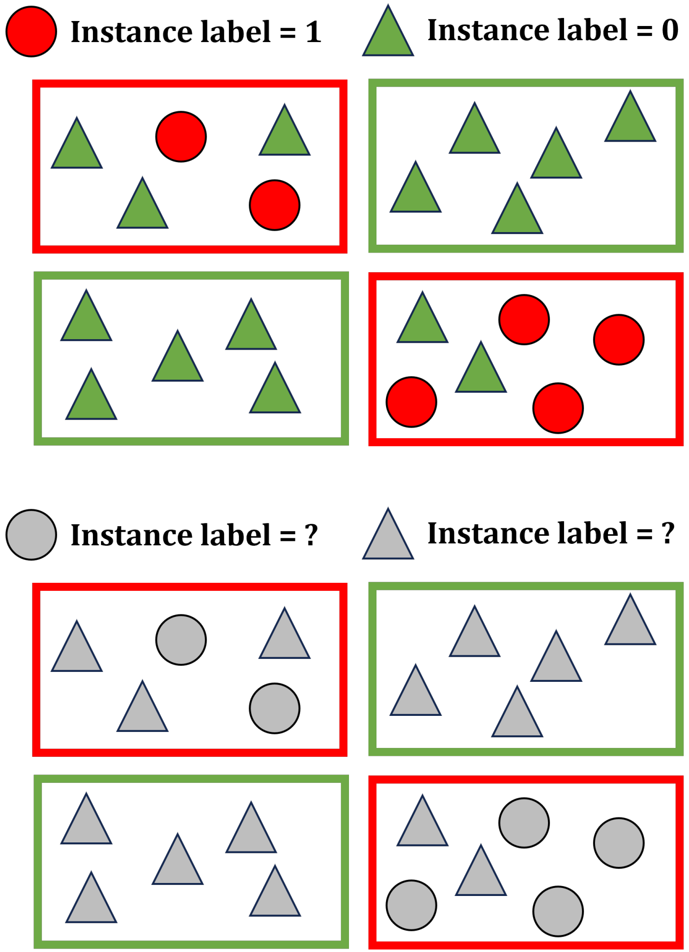
$$\underbrace{\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}}_{\text{bag = set of instances}} \subset \mathbb{R}^D, \quad \underbrace{\mathbf{x}_n}_{\text{instance}} \in \mathbb{R}^D,$$

$$\underbrace{y_n}_{\text{instance label}} \in \{0,1\}, \quad \underbrace{Y}_{\text{bag label}} \in \{0,1\}$$

The instance labels y_n are **not observed during training**.

It holds:

$$Y = \max\{y_1, \ldots, y_N\} = \begin{cases} 1 & \text{exists } n \text{ such that } y_n = 1 \\ 0 & y_n = 0 \text{ for every } n. \end{cases}$$



Multiple Instance Learning (MIL)

Data. Pairs of the form $(\mathbf{X}, Y)$:

$$\underbrace{\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}}_{\text{bag = set of instances}} \subset \mathbb{R}^D, \quad \underbrace{\mathbf{x}_n}_{\text{instance}} \in \mathbb{R}^D,$$

$$\underbrace{y_n}_{\text{instance label}} \in \{0, 1\}, \quad \underbrace{Y}_{\text{bag label}} \in \{0, 1\}$$

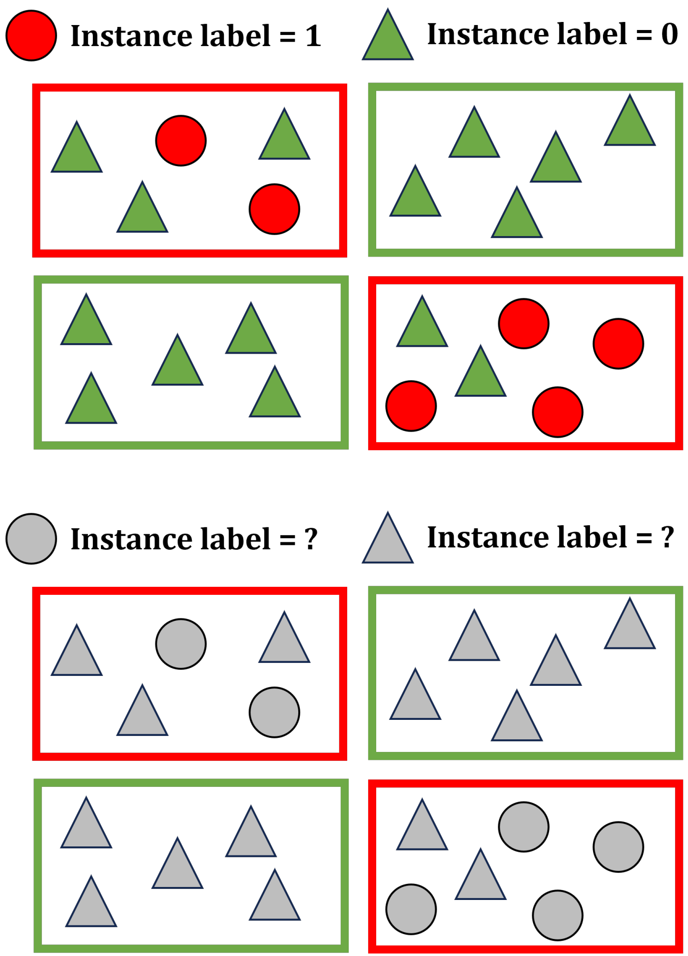
The instance labels y_n are **not observed during training**.

It holds:

$$Y = \max\{y_1, \ldots, y_N\} = \begin{cases} 1 & \text{exists } n \text{ such that } y_n = 1 \\ 0 & y_n = 0 \text{ for every } n. \end{cases}$$

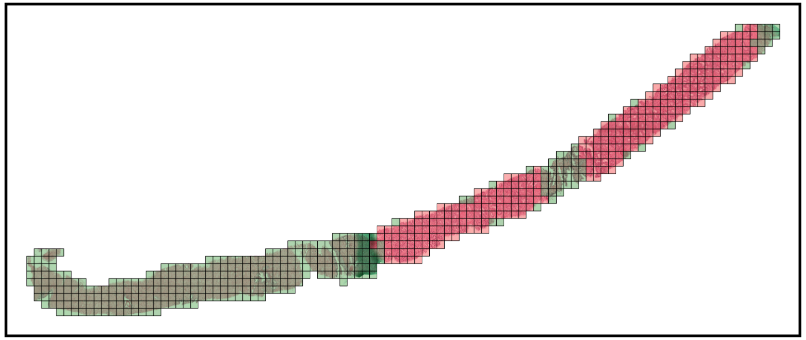
At test time. Given a new bag, we want to predict:

- The bag label (classification task).
- The instance labels (localization task).



Examples

Histopathology (Tumor detection):



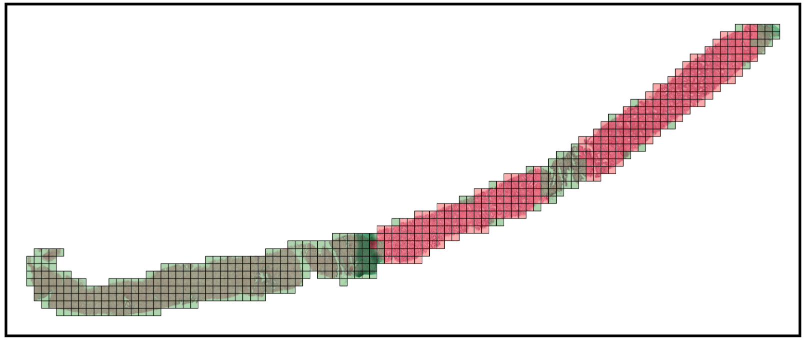
$$\text{Bag} \equiv \text{WSI}, \quad \text{Instance} \equiv \text{Patch}$$

- The WSI "is cancerous" ($Y = 1$) if and only if some patch "is cancerous" ($y_n = 1$ for some n).
- The WSI "is healthy" ($Y = 0$) if and only if all patches "are healthy" ($y_n = 0$ for every n).

Examples

Histopathology (Tumor detection):

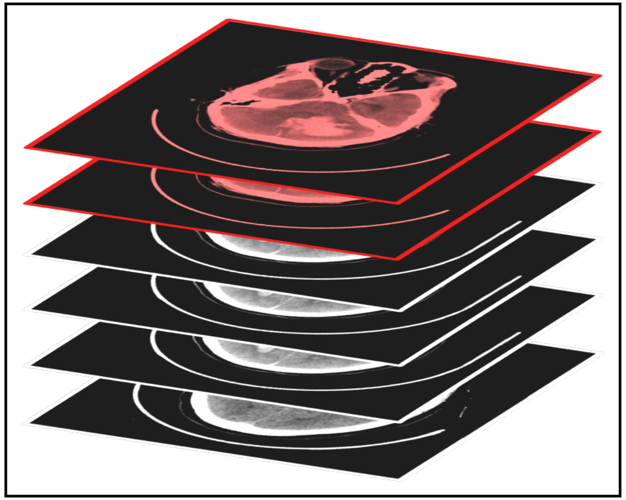
$$\text{Bag} \equiv \text{WSI}, \quad \text{Instance} \equiv \text{Patch}$$



- The WSI "is cancerous" ($Y = 1$) if and only if some patch "is cancerous" ($y_n = 1$ for some n).
- The WSI "is healthy" ($Y = 0$) if and only if all patches "are healthy" ($y_n = 0$ for every n).

CT scans (Intracranial hemorrhage detection):

$$\text{Bag} \equiv \text{CT scan}, \quad \text{Instance} \equiv \text{Slice}$$



- The CT scan "has hemorrhage" ($Y = 1$) if and only if some slice "has hemorrhage" ($y_n = 1$ for some n).
- The CT scan "is healthy" ($Y = 0$) if and only if all slices "are healthy" ($y_n = 0$ for every n).

Attention-based MIL (ABMIL)

Originally proposed by (Ilse et al., 2018).

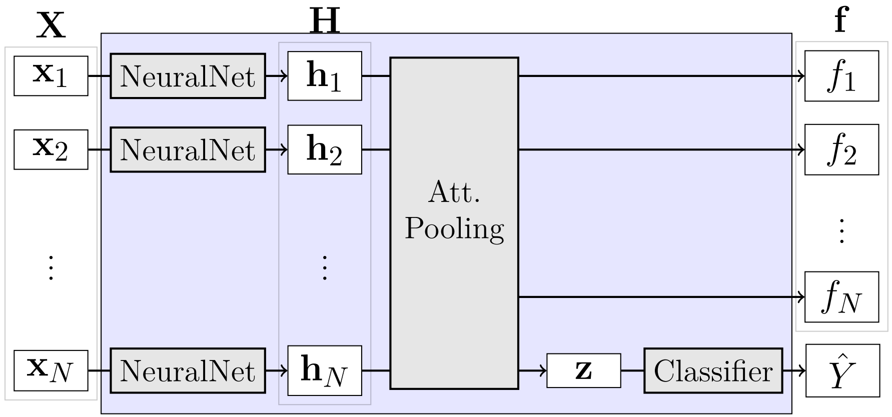


Figure: The ABMIL model.

Attention-based MIL (ABMIL)

Originally proposed by (Ilse et al., 2018).

It introduced the **Attention Pooling** mechanism, which assigns an attention value $f_i \in \mathbb{R}$ to each instance $\mathbf{x}_i$.

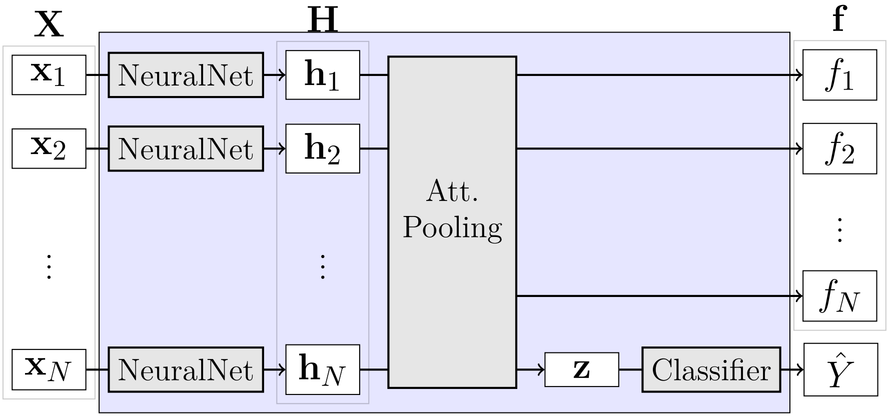


Figure: The ABMIL model.

Attention-based MIL (ABMIL)

Originally proposed by (Ilse et al., 2018).

It introduced the **Attention Pooling** mechanism, which assigns an attention value $f_i \in \mathbb{R}$ to each instance $\mathbf{x}_i$.

The attention values are used to **aggregate** all the instances into a bag representation $\mathbf{z}$.

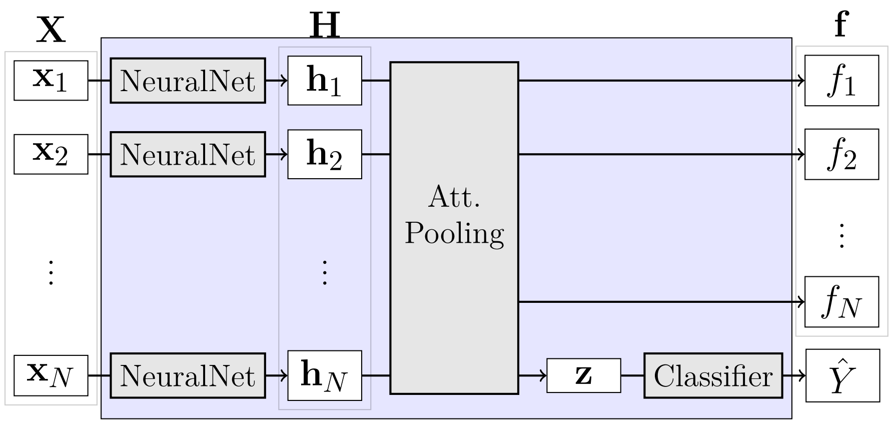


Figure: The ABMIL model.

Attention-based MIL (ABMIL)

Originally proposed by (Ilse et al., 2018).

It introduced the **Attention Pooling** mechanism, which assigns an attention value $f_i \in \mathbb{R}$ to each instance $\mathbf{x}_i$.

The attention values are used to **aggregate** all the instances into a bag representation $\mathbf{z}$.

It is defined by:

$$\mathbf{h}_i = \mathrm{NeuralNet}(\mathbf{x}_i) \in \mathbb{R}^P,$$

$$\boxed{\begin{aligned} f_i &= \mathbf{w}^\top \tanh(\mathbf{h}_i \mathbf{V}) \in \mathbb{R}, \\ s_i &= \frac{\exp(f_i)}{\sum_j \exp(f_j)} \in (0,1), \quad \text{Attention Pooling} \\ \mathbf{z} &= \sum_i s_i \mathbf{h}_i \in \mathbb{R}^P, \end{aligned}}$$

$$\hat{Y} = \mathrm{Classifier}(\mathbf{z})$$

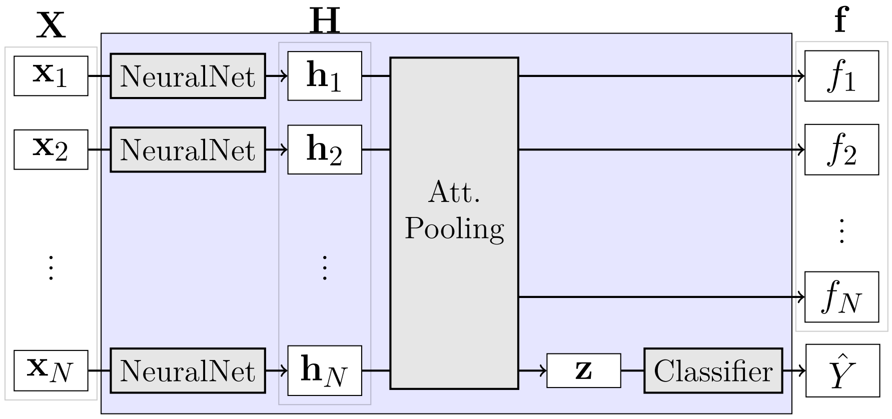


Figure: The ABMIL model.

ABMIL: important of the attention values

Important: the attention values f_i represent the importance of the corresponding instance within the bag.

Observation: the attention values f_i can be used to identify positive instances. For cancer prediction, this means detecting the cancerous tissue.

ABMIL: important of the attention values

Important: the attention values f_i represent the importance of the corresponding instance within the bag.

Observation: the attention values f_i can be used to identify positive instances. For cancer prediction, this means detecting the cancerous tissue.

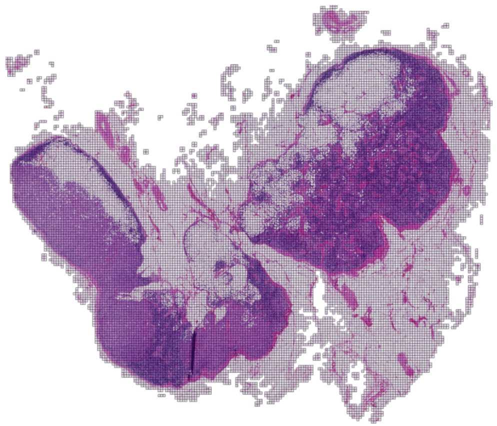


Figure: WSI.

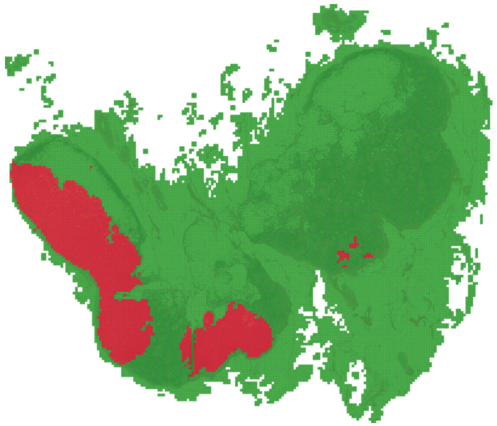


Figure: Patches labelled by a pathologist.

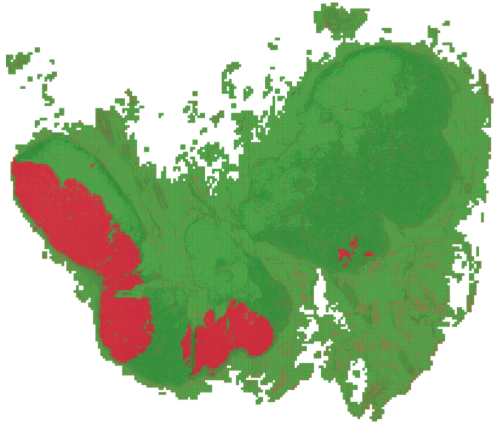


Figure: ABMIL attention values.

torchmil

A PyTorch library for MIL

torchmil: A PyTorch library for MIL

torchmil is an open-source Python library built on top of PyTorch. It provides a flexible and extensible framework for building, training, and evaluating deep MIL models

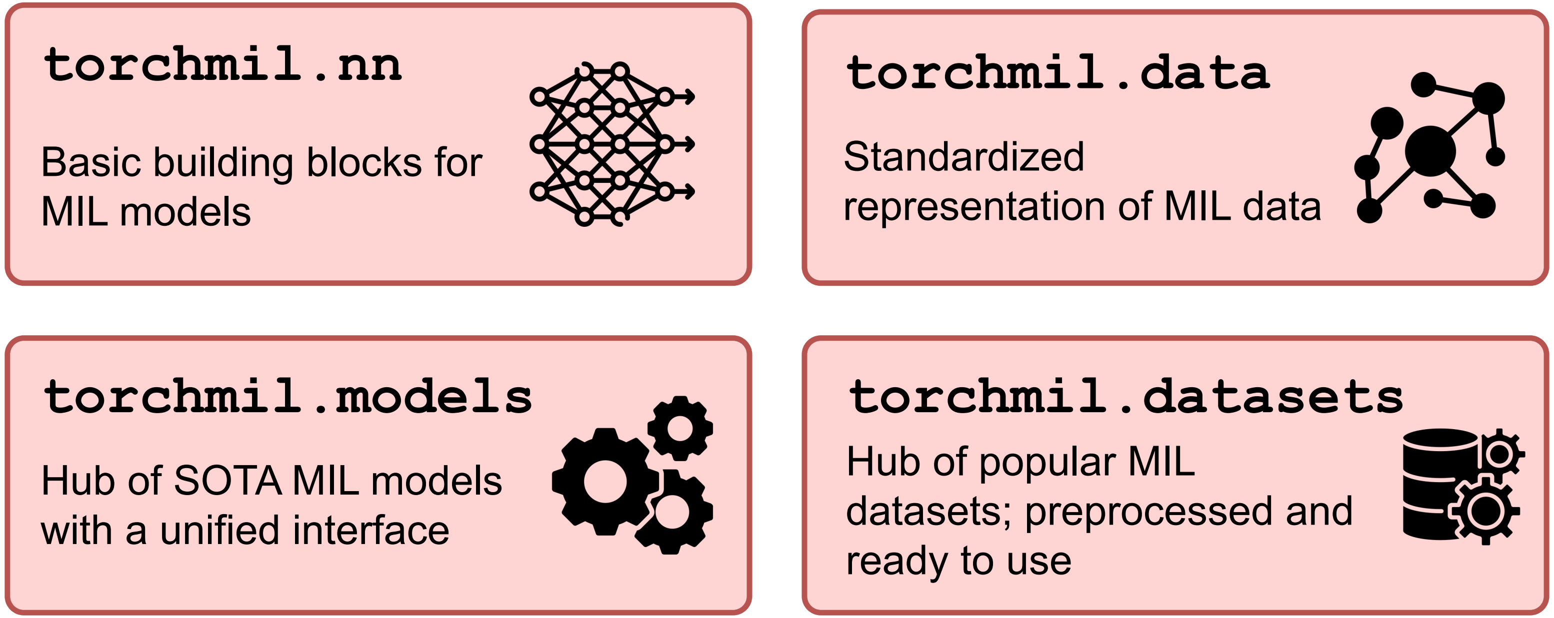


Figure: An overview of torchmil.

torchmil in action

Some attention maps generated by the models available in torchmil:

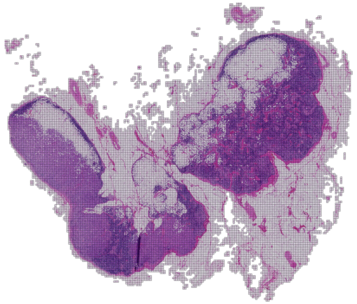


Figure: WSI.

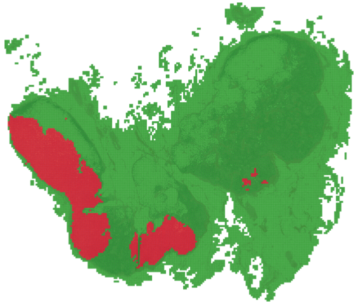


Figure: Patches labelled by a pathologist.

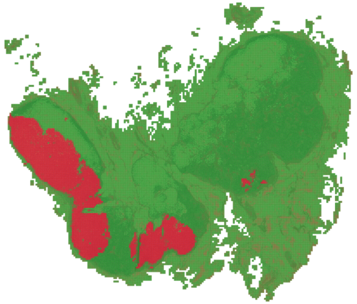


Figure: ABMIL (Ilse et al., 2018).

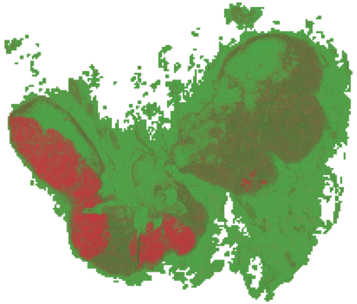


Figure: DTFDMIL (Zhang et al., 2022).

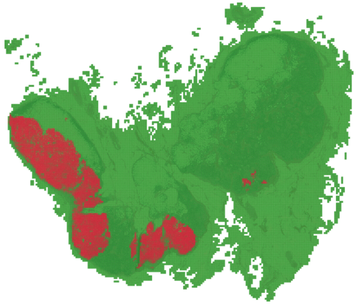


Figure: CAMIL (Fourkioti et al., 2023).

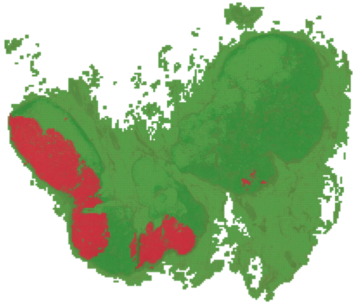


Figure: SmTABMIL (Castro-Macías et al., 2024).

Questions?

torchmil.readthedocs.io

github.com/Franblueee/torchmil

Next up: Jupyter Notebook

References

Castro-Macías, F. M., Morales-Álvarez, P., Wu, Y., Molina, R., & Katsaggelos, A. K. (2024). Sm: enhanced localization in multiple instance learning for medical imaging classification. *Advances in Neural Information Processing Systems*, *37*, 77494–77524.

Fourkioti, O., De Vries, M., Jin, C., Alexander, D. C., & Bakal, C. (2023). CAMIL: Context-aware multiple instance learning for cancer detection and subtyping in whole slide images. *Arxiv Preprint Arxiv:2305.05314*.

Ilse, M., Tomczak, J., & Welling, M. (2018). Attention-based deep multiple instance learning. *International Conference on Machine Learning*, 2127–2136.

Lu, M. Y., Williamson, D. F., Chen, T. Y., Chen, R. J., Barbieri, M., & Mahmood, F. (2021). Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature Biomedical Engineering*.

Zhang, A., Jaume, G., Vaidya, A., Ding, T., & Mahmood, F. (2025). Accelerating data processing and benchmarking of AI models for pathology. *Arxiv Preprint Arxiv:2502.06750*.

Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S. E., & Zheng, Y. (2022). Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18802–18812.