

Bachelor's Final Project Report

Classification of Histopathological Images Using Multi-Instance Learning with Attention

Internship conducted at:

LITIS – Laboratoire d’informatique, de traitement de l’information et des systèmes

BENKHEDDA Mohamed Serradj Eddine

Supervised by: Caroline Petitjean (Professor), Zoé Lambert (Associate Professor)

University of Rouen Normandy - Bachelor in Computer Science, Data Science

Acknowledgments

I would like to express my deepest gratitude to my supervisors for their guidance throughout this project. Their expertise, patience, and advice were essential to the success of this enriching experience.

Introduction

The analysis of histopathological images plays a crucial role in medical diagnosis, particularly in detecting cancerous metastases in lymphatic tissues. These images are extremely large and cannot be processed directly by conventional CNN architectures. To overcome this limitation, we adopted an approach based on **Multi-Instance Learning (MIL)**, combined with **attention** and **gated attention** mechanisms, enabling the model to dynamically weight patches within an image for global classification.

Below is the architecture used in our work, taken from the original paper (Ilse et al., 2018):

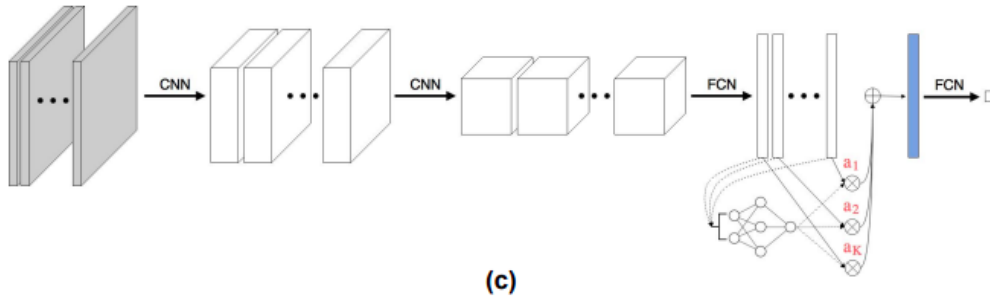


Figure 1: Architecture of the attention-based multi-instance learning model. Source: Ilse et al., 2018.

The code developed during this project is available at: <https://github.com/benk27Serradj/Attention-MIL-CAMELYON16>

We worked on a subset of the **CAMELYON16** dataset and implemented the models described in the paper "*Attention-based Deep Multiple Instance Learning*" (Ilse et al., 2018), available on GitHub: <https://github.com/AMLab-Amsterdam/AttentionDeepMIL>

Previous Work

Multi-Instance Learning (MIL)

In MIL, each image (called a *bag*) contains multiple patches (*instances*). The bag’s label depends on the presence or absence of a positive patch. This allows image classification without fine-grained annotations (pixel-level), which is crucial in medical applications.

Why does the model accept variable-sized bags?

The model is designed to process histopathological images divided into patches, and these images can vary in size or tissue density. This results in a variable number of patches per image (i.e., per *bag*). Fortunately, thanks to the **attention mechanism**, the model can easily handle this variability:

- Each patch is independently encoded into a feature vector $h_k \in R^{500}$.
- Attention weights are computed for each patch and normalized via a **softmax** function, ensuring their sum always equals 1, regardless of the number of patches.
- The final vector representing the entire image is obtained by a **weighted sum** of the feature vectors, weighted by the attention scores.

Thus, whether an image yields 20 or 60 patches, the model always produces a fixed-dimension vector, enabling it to handle **variable-sized bags without architectural modifications**.

Methods Used

Attention-Based Model

The model proposed by Ilse et al. (2018) uses an **attention** layer that assigns each patch a weight reflecting its importance for the final decision. This mechanism improves interpretability while maintaining a flexible architecture. The attention function computes, for each patch $h_k \in R^{500}$, an attention score A_k via a linear transformation followed by a tanh function:

$$A_k = \text{softmax}(w^\top \cdot \tanh(Vh_k))$$

A weighted sum is then performed to obtain a global vector Z :

$$Z = \sum_{k=1}^K A_k \cdot h_k$$

This vector Z is projected into a scalar via a linear layer and passed through a sigmoid function to produce the probability of the image being tumorous.

Gated Attention

A variant proposed in the same paper is **gated attention**, introduced to address some limitations of the standard tanh: > "the $\tanh(\cdot)$ non-linearity could be inefficient to learn complex relations. Our concern follows from the fact that $\tanh(x)$ is approximately linear for $x \in [-1, 1]$, which could limit the final expressiveness of learned relations among instances." This version replaces the tanh function with a product between $\tanh(Vh_k)$ and $\sigma(Uh_k)$, where σ is the sigmoid function. This allows greater

flexibility in learning complex relationships between patches.

$$A_k = \text{softmax} \left(w^\top \cdot [\tanh(Vh_k) \odot \sigma(Uh_k)] \right)$$

Dataset Description

We worked on a subset of the **CAMELYON16** dataset, comprising 48 images in total:

- **32 images for training** (16 normal, 16 tumorous)
- **16 images for testing** (8 normal, 8 tumorous)

Each image has an average size of **2 GB** and is stored in a pyramidal format with **11 resolution levels**. We chose to work with **level 5**, corresponding to an image size of (7168, 3072, 3) pixels (RGB). The images were divided into **non-overlapping patches** of size (700, 700, 3) pixels. Remaining pixels after division were ignored.

Dataset Visualization

Below are examples of Whole Slide Images (WSI) from the CAMELYON16 dataset:



Figure 2: Example of a normal WSI from the CAMELYON16 dataset.

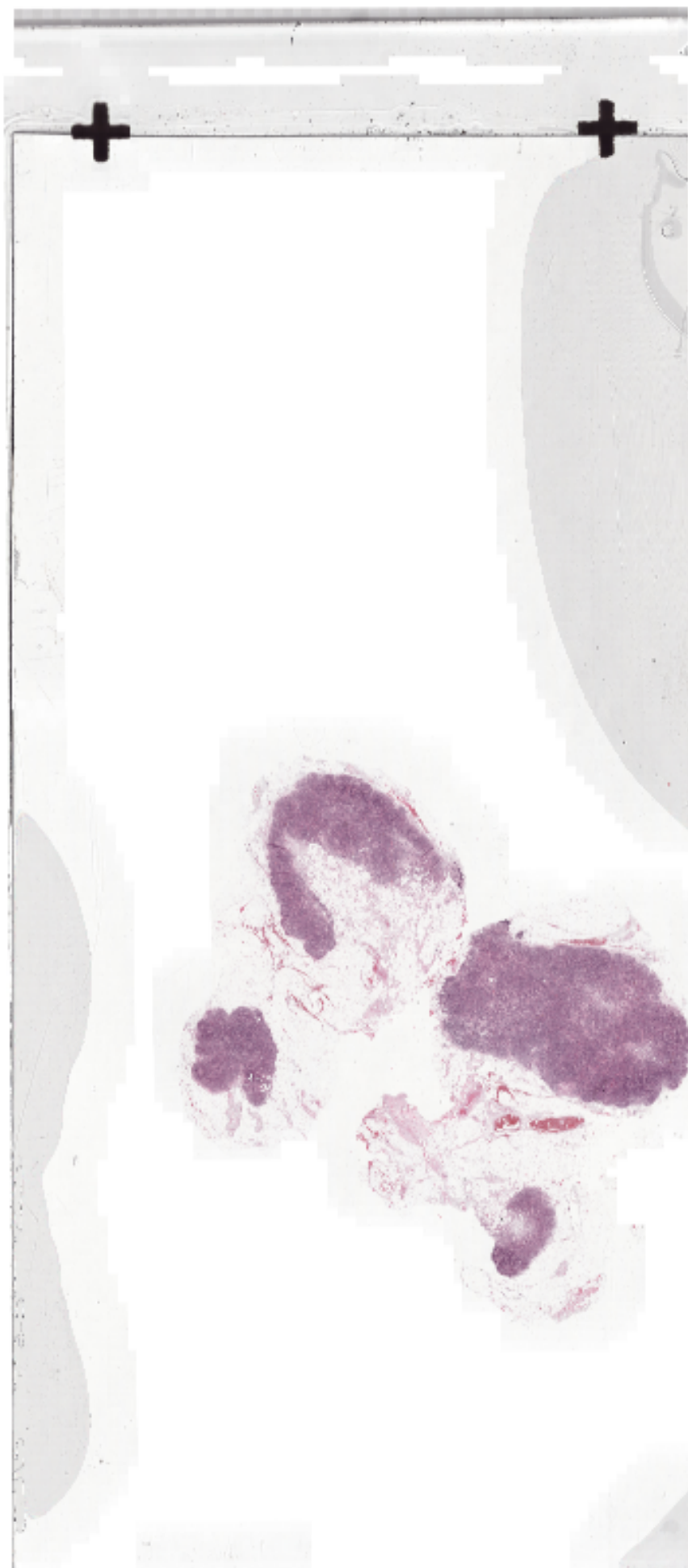


Figure 3: Example of a tumorous WSI from the CAMELYON16 dataset.

Patch Preprocessing

To reduce the number of irrelevant patches, two filtering functions were added.

`_is_tissue` Function

This function identifies patches containing biological tissue by applying three simple criteria:

- **Brightness threshold:** if the average luminance is below 40, the patch is rejected.
- **Color variation:** if the difference between the maximum and minimum RGB channel values is less than 2, the patch is likely homogeneous.
- **Color mask:** a heuristic rule $(R > 120) \& (B > 80) \& (G < 150)$ recognizes typical tissue stains (Hematoxylin-Eosin).

This ensures only relevant patches are retained for analysis, reducing computational load and improving prediction quality.

`is_not_blurry` Function

This function rejects blurry or out-of-focus patches that could impair feature extraction. It relies on visual sharpness analysis: a blurry image will have low gradient variance, while a sharp image will show high spatial variation. A threshold is applied to this variance to decide whether the patch is kept. Its goal is to retain only high-quality patches, increasing the reliability of the learning process.

Patch Extraction and Filtering

Below are examples of patches before and after filtering:

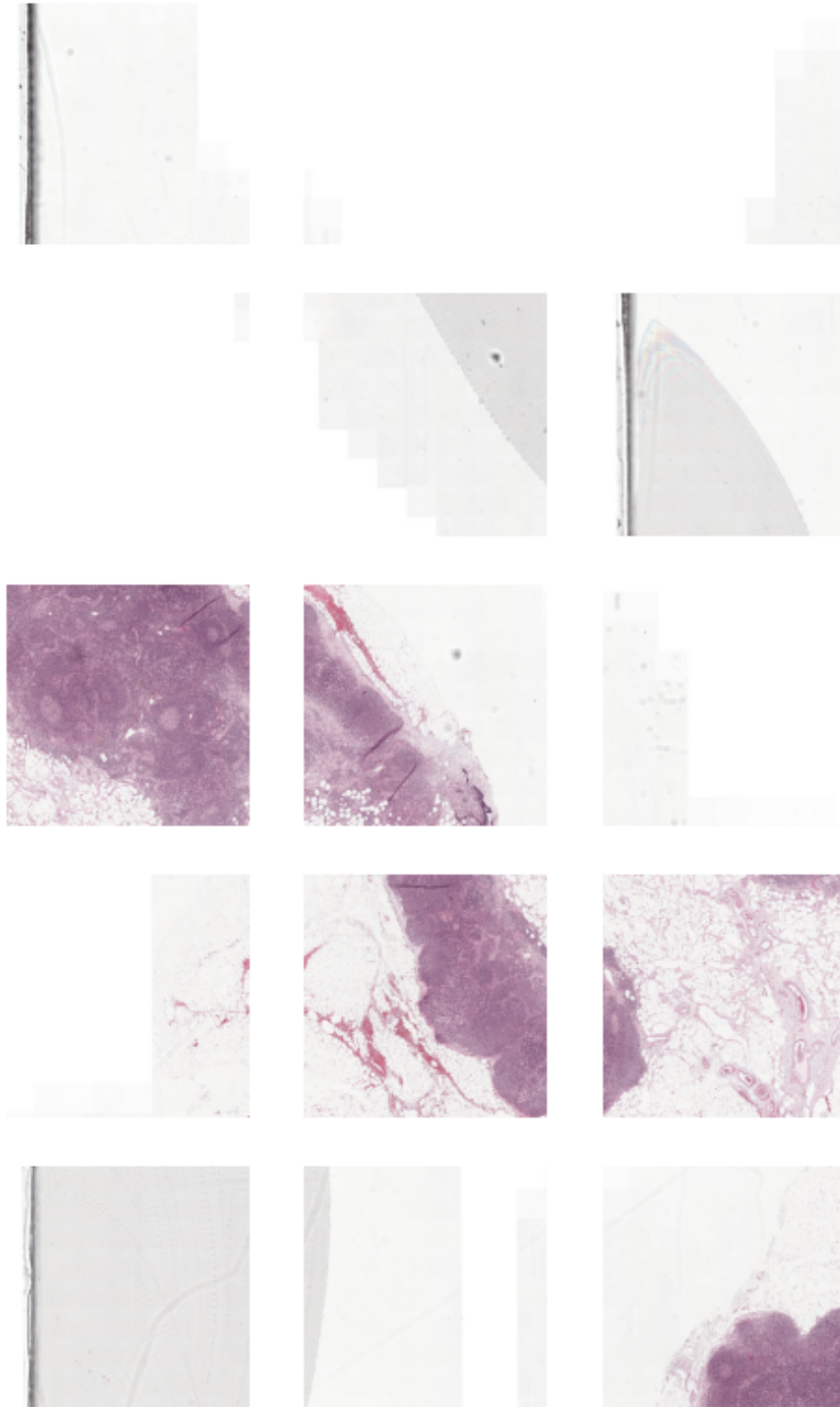


Figure 4: Set of patches extracted from an image without filtering (152 patches).

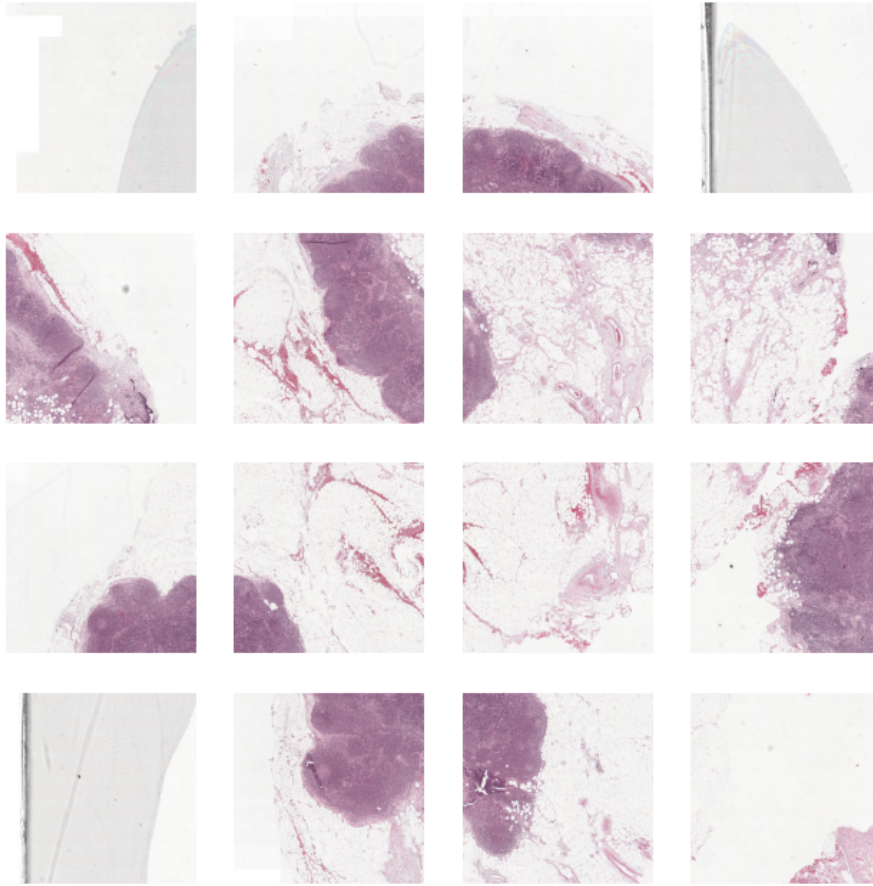


Figure 5: Patches remaining after applying `_is_tissue` (39 patches).

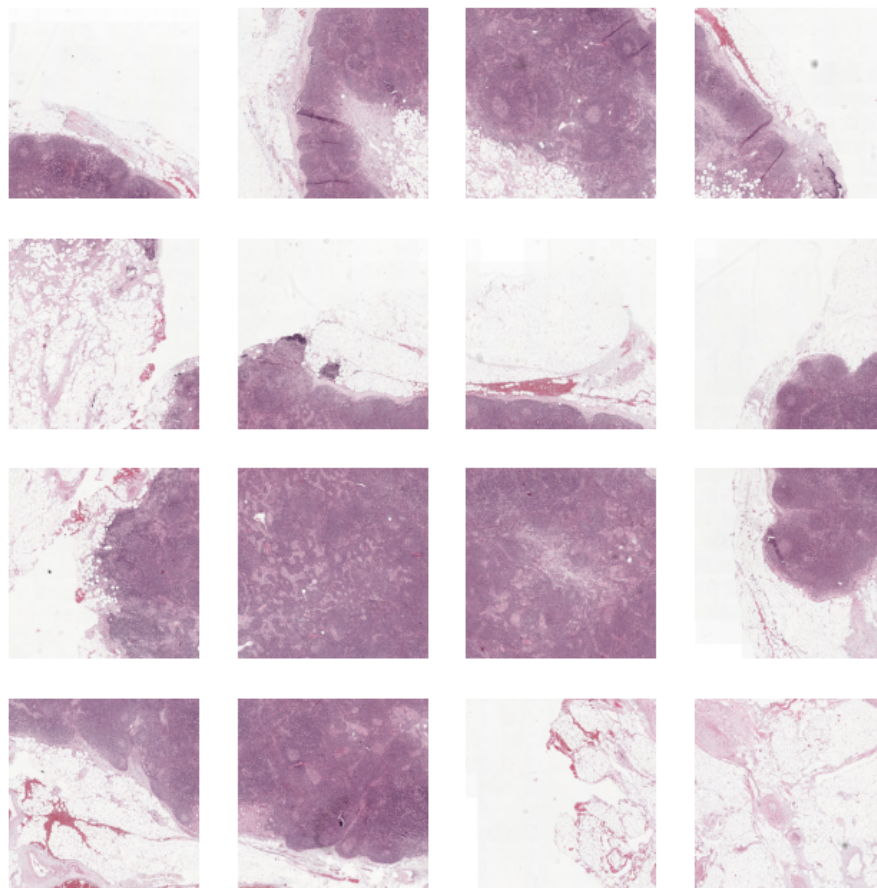


Figure 6: Patches remaining after applying `is_not_blurry` (28 patches).

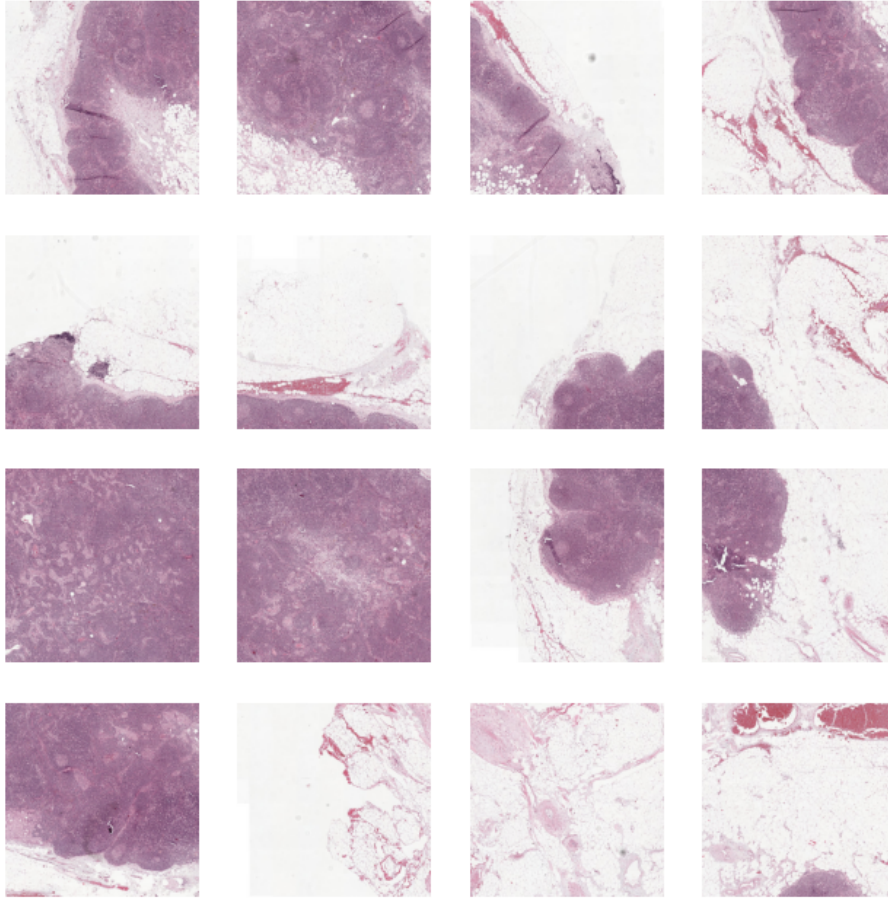


Figure 7: Patches remaining after applying both `_is_tissue` and `is_not_blurry` (28 patches).

Model Configuration

The following parameters were used during model training:

Parameter	Value
Pyramid level	Level 5
Patch size	(700, 700, 3)
Batch size	1
Number of epochs	50
Bag size (no filter)	40
Bag size (with filter)	Variable
Dimension of Hk	500
Loss function	neg_log_likelihood
Optimizer	Adam
Learning rate	0.0005
Weight decay	1e-5

Table 1: Model settings

Implementation

The code was adapted from the official implementation provided by the authors on GitHub: <https://github.com/AMLab-Amsterdam/AttentionDeepMIL> Major modifications were made: [leftmargin=2cm]

- Rewriting `main.py` to integrate custom parameters and CAMELYON16 dataset handling.
- Adapting `model.py` to support the dimensions used in our case.
- Complete reimplementaion of the `dataloader` to load WSIs, extract patches, and apply the two new filtering functions (`_is_tissue`, `is_not_blurry`).

Results

We evaluated **8 different configurations**, each tested over **5 independent runs**. Below are the average performances:

Table 2: Performance of the 8 tested configurations

Configuration	Accuracy	Recall	Precision	F1-score	AUC
Gated Attention	0.6750 \pm 0.1000	0.6750 \pm 0.1000	0.6771 \pm 0.2154	0.6253 \pm 0.1618	0.6750 \pm 0.1000
Gated Attention + <code>is_tissue</code>	0.8125 \pm 0.0685	0.8125 \pm 0.0685	0.8293 \pm 0.0744	0.8103 \pm 0.0686	0.8125 \pm 0.0685
Gated Attention + <code>is_not_blurry</code>	0.8750 \pm 0.0685	0.8750 \pm 0.0685	0.8912 \pm 0.0606	0.8731 \pm 0.0699	0.8750 \pm 0.0685
Gated Attention + <code>is_tissue</code> + <code>is_not_blurry</code>	0.8500 \pm 0.0306	0.8500 \pm 0.0306	0.8670 \pm 0.0404	0.8485 \pm 0.0300	0.8500 \pm 0.0306
Attention	0.7625 \pm 0.0468	0.7625 \pm 0.0468	0.8237 \pm 0.0424	0.7497 \pm 0.0561	0.7625 \pm 0.0468
Attention + <code>is_tissue</code>	0.8375 \pm 0.0637	0.8375 \pm 0.0637	0.8492 \pm 0.0640	0.8360 \pm 0.0646	0.8375 \pm 0.0637
Attention + <code>is_not_blurry</code>	0.7875 \pm 0.0848	0.7875 \pm 0.0848	0.7997 \pm 0.0941	0.7861 \pm 0.0845	0.7875 \pm 0.0848
Attention + <code>is_tissue</code> + <code>is_not_blurry</code>	0.8125 \pm 0.0968	0.8125 \pm 0.0968	0.8556 \pm 0.0462	0.7993 \pm 0.1202	0.8125 \pm 0.0968

Discussion

The experimental results show that **attention mechanism** and **gated attention** models yield promising results in histopathological image classification but require **careful preprocessing** to reach their full potential, especially with limited datasets.

Comparative Analysis of Configurations

Gated Attention + `is_not_blurry` stands out as the best-performing configuration, with an **accuracy of 87.50%** and **precision of 89.12%**, surpassing all others, including those with `is_tissue`.

This suggests that **patch visual quality** is critical for this model, which benefits more from **sharpness-based filtering** than simple tissue detection.

The `is_tissue` function consistently improves both models’ performance by eliminating biologically irrelevant regions. It is particularly effective for the **attention** model, where accuracy increases from **76.25% to 83.75%** and precision increases from **82.37% to 84.92%** when applied.

Combining Both Filters

For the **gated attention** model, the best performance is achieved with the *is_not_blurry* filter alone (F1-score = 0.8731), but combining both filters still outperforms the baseline and the *is_tissue* configuration, indicating a strong benefit from quality-based filtering. On the other hand, the **attention** model performs best with the *is_tissue* filter alone (F1-score = 0.8360), though the combination of both filters (F1-score = 0.7993) still provides a notable improvement over the baseline and *is_not_blurry*

configurations. These results highlight the importance of tailoring filtering strategies to the model architecture to maximize performance.

Behavioral Differences Between Models

- The **standard attention model** shows **good stability** and **better tolerance to small datasets**.
 - The **gated attention** model, while theoretically more powerful, requires **rigorous preprocessing** to achieve good performance. Without filtering, it is significantly less effective than standard attention.

These observations confirm that more complex models, while theoretically appealing, may suffer from **noise exploitation** or **optimization difficulties** when trained on small datasets.

Conclusion

This project evaluated the effectiveness of two **multi-instance learning models with attention**, applied to the classification of histopathological images from the **CAMELYON16** dataset. Key conclusions include:

- **Attention mechanism** and **gated attention** models are well-suited for histopathological image classification, as shown in the original study (Ilse et al., 2018).
- On a **small dataset**, their performance can be **significantly improved** using **automatic filtering functions**.
 - I developed two filtering functions:
 - `_is_tissue`: to isolate biologically relevant patches.
 - `_is_not_blurry`: to reject blurry or low-quality patches.
 - These tools have a measurable impact on model accuracy and stability, particularly for **gated attention**, which achieved its best performance with `is_not_blurry` (**87.50% accuracy**).

This work highlights the value of **MIL with attention** for efficiently processing large medical images while demonstrating that, under practical constraints (small dataset, high processing time), **simple yet relevant technical additions** can have a significant impact.

References

1. Ilse, M., Tomczak, J. M., & Welling, M. (2018). *Attention-based Deep Multiple Instance Learning*. arXiv preprint arXiv:1802.04712.
2. Official CAMELYON16 challenge website: <https://camelyon16.grand-challenge.org/Data/>
3. Model source code: <https://github.com/AMLab-Amsterdam/AttentionDeepMIL>