Enhancing PFI Prediction with GDS-MIL: a Graph-based Dual Stream MIL Approach

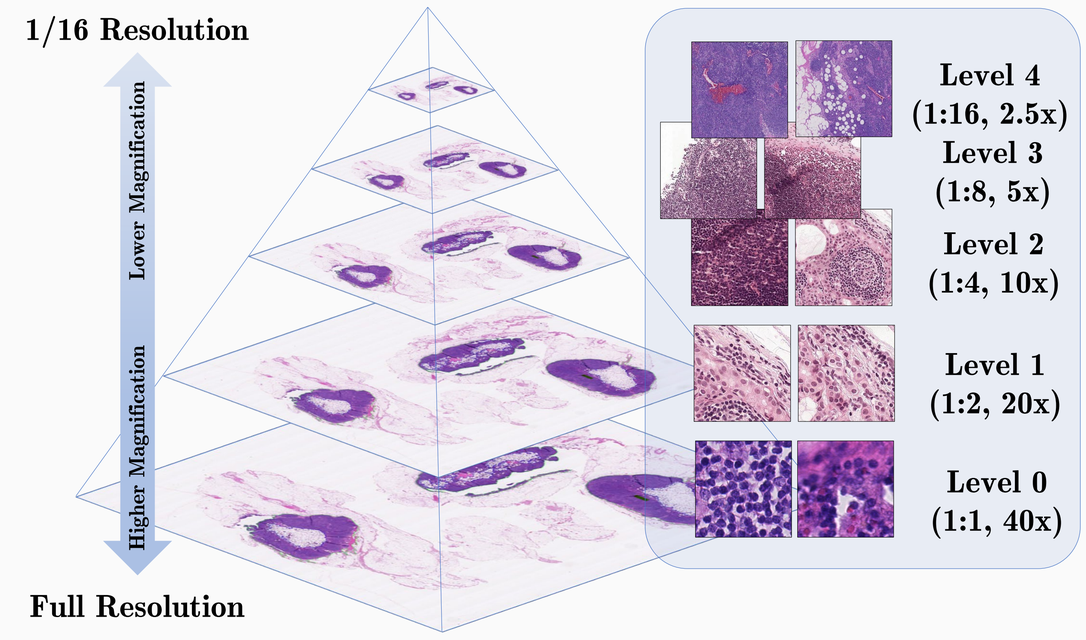
Gianpaolo Bontempo[1,2], Nicola Bartolini[1], Marta Lovino[1], Federico Bolelli[1], Anni Virtanen[3], and Elisa Ficarra[1]

[1]University of Modena and Reggio Emilia, Italy
{name.surname}@unimore.it

[2]University of Pisa, Italy

[3]University of Helsinki, Finland

Introduction



Whole-Slide Images (WSIs) present challenges for deep learning frameworks due to their large size and lack of pixel-level annotations.

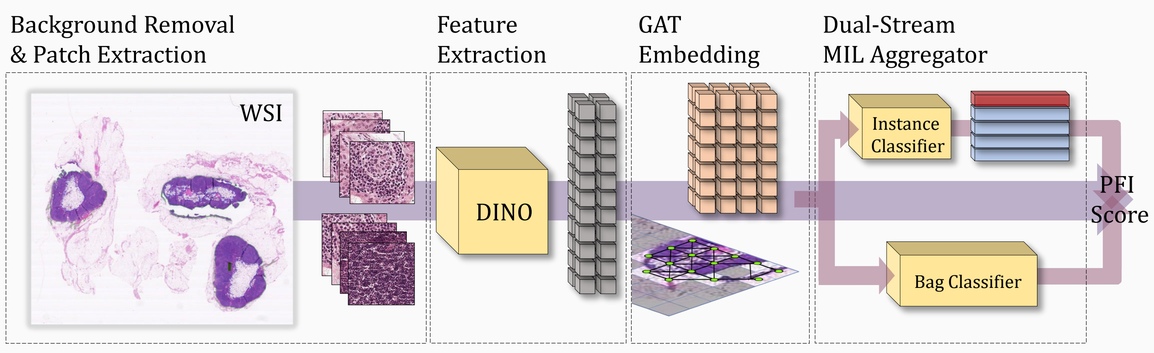
Multi-Instance Learning (MIL) approaches consider the image slide as a bag composed of many patches, called instances; afterward, they weight the instances through attention mechanisms and aggregate them into a single representation to provide a classification score for the entire bag.

Problem Statement

The prediction of **Platinum-Free Interval (PFI)**, defined as the time interval between the end of chemotherapy and disease recurrence, is a determinant for treatment planning and is performed by analyzing WSIs.

While some tasks can rely on morphology (e.g., tumor detection), others would benefit from a **more contextualized tissue analysis**. An example is the aforementioned prediction of PFI on chemotherapy-treated HGSOC tissue.

Architecture



The embeddings are fed to a GAT module to capture context and generate a more contextualized representation.

Graph Context

GNNs are a class of deep learning models designed specifically for processing data structured as graphs, where nodes represent entities, and edges represent relationships between them.

Message Passing: GNNs perform message passing or information aggregation across neighboring nodes in multiple iterations (typically called layers or message-passing steps). In each iteration, nodes aggregate information from their neighbors and update their own representations.

Results

Scale	Approach	Best Epoch Accuracy	Best Epoch AUC	Last Epoch Accuracy	Last Epoch AUC
5×	MaxPooling	0.579 ± 0.067	0.432 ± 0.165	0.579 ± 0.055	0.419 ± 0.161
	MeanPooling	0.596 ± 0.072	0.427 ± 0.166	$\mathbf{0.594 \pm 0.069}$	0.413 ± 0.148
	AB-MIL [1]	0.606 ± 0.076	0.467 ± 0.171	0.577 ± 0.076	0.413 ± 0.166
	DS-MIL [2]	0.582 ± 0.090	0.478 ± 0.145	0.574 ± 0.075	$\mathbf{0.458 \pm 0.136}$
	GDS-MIL (our)	$\mathbf{0.620 \pm 0.045}$	$\mathbf{0.512 \pm 0.096}$	0.566 ± 0.045	0.402 ± 0.139
20×	MaxPooling	0.676 ± 0.055	0.637 ± 0.083	0.661 ± 0.072	0.598 ± 0.107
	MeanPooling	0.610 ± 0.089	0.446 ± 0.196	0.605 ± 0.086	0.443 ± 0.193
	AB-MIL [1]	0.594 ± 0.095	0.510 ± 0.141	0.576 ± 0.090	0.438 ± 0.156
	DS-MIL [2]	0.681 ± 0.033	0.650 ± 0.049	0.656 ± 0.028	0.572 ± 0.065
	GDS-MIL (our)	$\mathbf{0.704 \pm 0.070}$	$\mathbf{0.661 \pm 0.099}$	$\mathbf{0.663 \pm 0.064}$	$\mathbf{0.611 \pm 0.092}$

# Layer	# Heads	AUC	Acc.
1	1	0.664	0.704
1	2	0.667	0.732
1	3	**0.726**	**0.764**
2	1	0.634	0.722
2	2	0.607	0.648
2	3	0.619	0.694
3	1	0.641	0.694
3	2	0.639	0.704
3	3	0.697	0.732

Increasing the **number of layers**, an over-smoothing effect could occur.
If the smoothing is too strong, it becomes challenging for the MIL module to distinguish what is actually important.

Using a **multi-head** approach enhances the ability to capture the most important information from the neighborhood and build a more contextualized representation of each instance.

References

[1] Li, B., Li, Y., & Eliceiri, K. W. (2021). Dual-stream Multiple Instance Learning Network for Whole Slide Image Classification with Self-Supervised Contrastive Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 14318-14328).

[2] Ilse, M., Tomczak, J., & Welling, M. (2018). Attention-based Deep Multiple Instane Learning. In International Conference on Machine Learning (pp. 2127-2136). PMLR.