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
{name.surname}@phd.unipi.it
[3] University of Helsinki, Finland
{name.surname}@hus.fi

Abstract. Whole-Slide Images (WSI) are emerging as a promising resource for studying biological tissues, demonstrating a great potential in aiding cancer diagnosis and improving patient treatment. However, the manual pixel-level annotation of WSIs is extremely time-consuming and practically unfeasible in real-world scenarios. Multi-Instance Learning (MIL) have gained attention as a weakly supervised approach able to address lack of annotation tasks. MIL models aggregate patches (*e.g.,* cropping of a WSI) into bag-level representations (*e.g.,* WSI label), but neglect spatial information of the WSIs, crucial for histological analysis. In the High-Grade Serous Ovarian Cancer (HGSOC) context, spatial information is essential to predict a prognosis indicator (the Platinum-Free Interval, PFI) from WSIs. Such a prediction would bring highly valuable insights both for patient treatment and prognosis of chemotherapy resistance. Indeed, NeoAdjuvant ChemoTherapy (NACT) induces changes in tumor tissue morphology and composition, making the prediction of PFI from WSIs extremely challenging. In this paper, we propose GDS-MIL, a method that integrates a state-of-the-art MIL model with a Graph ATtention layer (GAT in short) to inject a local context into each instance before MIL aggregation. Our approach achieves a significant improvement in accuracy on the "Ome18" PFI dataset.
In summary, this paper presents a novel solution for enhancing PFI prediction in HGSOC, with the potential of significantly improving treatment decisions and patient outcomes.

1 Introduction

High-Grade Serous Ovarian Cancer (HGSOC) is a form of ovarian cancer characterized by multiple treatment recurrences with variable response to platinum-based chemothereapy. The prediction of Platinum-Free Interval (PFI), defined as the time interval between the end of chemotherapy and disease recurrence [27], is determinant for treatment planning and is usually performed by analyzing the histological tissue digitalized in Whole-Slide Images (WSIs). Unfortunately,

NeoAdjuvant ChemoTherapy (NACT), recommended for HGSOC patients who are ineligible for Primary Debulking Surgery (PDS) [10,22], causes strong variable changes and heterogeneity in tumor morphology and composition, making the prediction of PFI from WSI extremely challenging.

Whole-slide imaging has emerged in recent years as a promising technology to enable the digitalization and the analysis of tissue sections [12]. The creation of multi-resolution gigapixel WSIs provides the opportunity of developing novel diagnostic tools for treatment and monitoring [6,25]. However, the manual pixel-level annotation of WSIs is a time-consuming and labor-intensive task. As an alternative, WSIs are often labelled with metadata (*e.g.,* genetic or other molecular features) characterizing the disease. In addition, because of their gigapixel size, WSIs are usually clipped into patches before being fed into a deep learning model. Given all these conditions, Convolutional Neural Networks (CNNs), which have provided amazing results for a multitude of tasks [13,16,21,31], cannot be directly applied to such data.

Consequently, Multi-Instance Learning (MIL) methods have gained considerable attention in WSI analysis [15,32], avoiding the need for pixel-level annotations. MIL is a weakly supervised learning approach used to assign a label to a set (or bag) composed of unlabelled instances. The label of the bag (*e.g.,* a WSI) is determined by the presence or absence of at least one positive instance (*e.g.,* patch containing tumour), so it is generally assumed that negative bags only contain negative instances (*e.g.,* patch not containing tumour), while positive bags contain at least one positive instance. When dealing with histological images, such assumption cannot be enough and Attention-Based MIL (AB-MIL) [9,2] should be employed to improve patch aggregations [32,35]. However, AB-MIL approaches do not exploit any spatial dependency between instances, which may be crucial in some application [7]. While some tasks can rely solely on morphology analysis (*e.g.,* tumor detection), others would benefit from a more comprehensive tissue analysis. An example of such a task is the aforementioned prediction of PFI on chemotherapy treated HGSOC tissue.

This paper proposes GDS-MIL, which integrates a state-of-the-art MIL model with Graph Neural Networks (GNNs) to contextualize patch local interactions better. Specifically, we use Graph ATtention networks (GATs) [33] to capture the spatial relationships between instances before MIL aggregation, introducing a local context into each instance. This approach has shown promising results, achieving a significant improvement on the "Ome18" PFI dataset. Our study provides a novel solution to improve the accuracy of PFI prediction in HGSOC, which could ultimately lead to better treatment decisions and improved patient outcomes [27].

2 Related Works

In this section, we briefly review recent developments in MIL models, as well as relevant studies that employ MIL for WSI analysis, and existing strategies for PFI prediction.

2.1 Multi-Instance Learning for WSI Analysis

Consider a bag X^{bag} composed of a set of N feature vectors:

$$X^{bag} = \{x_1, x_2, ..., x_N\} \tag{1}$$

Each instance $x_i \in X^{bag}$, can be assigned to a class through a mapping process $f : X^{bag} \rightarrow \{0, 1\}$, where the negative and positive classes correspond to 0 and 1, respectively. While traditionally MIL approaches rely on simple aggregators like mean-pooling and max-pooling [8,24], recent studies have shown that there may be benefits in parameterizing the aggregation operator with neural networks [17,23]. The Attention-Based MIL (AB-MIL) [9] employs a side-branch network to calculate attention scores. Similarly, in [37], Zhang *et al.* apply an attention mechanism to support a double-tier feature distillation approach, where relevant features are distilled from pseudo-bags to the WSI using either "MaxMin" or Aggregated Feature Selection (AFS) [37]. Another approach, DS-MIL [15], applies non-local attention aggregation to measure the distance with the most relevant patch. In 2021, Lu *et al.* [18] propose an algorithm that applies a clustering loss to single or multiple branches (CLAM-SB and CLAM-MB), a variant of the classic AB-MIL. Shao *et al.* [28], instead, employ a transformer architecture named Trans-MIL.

2.2 PFI Prediction

A few algorithms for automatic PFI prediction have been proposed in the literature. Both Yu *et al.* [36] and Laury *et al.* [14] use pixel-level annotated WSI for their studies. Yu *et al.* propose a method based on a VGG [29], using portions of WSI for regression analysis finalized to PFI prediction, while Laury *et al.* develop a method based on multiple neural networks used in series, *i.e.,* the output of the first becomes the input of the following network, after human supervised rearrangements. The final aggregation is based on the ratio between digital biomarkers associated with a poor or good prognosis. Their approach employs WSI of treatment-naïve HGSOC. Only tumoral areas are analyzed for the PFI prediction, exploiting pixel-level annotations for the segmentation. Moreover, by focusing the method on treatment-naïve patients, the tumor tissue presents a higher homogeneity in its morphology and texture than tissues undergoing treatment.

Instead, our approach focuses on patients with HGSOC who underwent NACT therapy. Therefore, the WSIs analyzed in this paper are characterized by unique morphological characteristics resulting from the treatment effects. Furthermore, to better understand the effects of the treatment, our method analyzes different tissues and compartments in the WSI (*e.g.,* tumor, stroma, inflammatory cells, etc.), and not only tumoral areas, increasing data heterogeneity.

Finally, our method does not require pixel-level annotations to predict the PFI score, relaying only on the global label. To achieve this goal, a graph attention layer has been incorporated into the model to analyze tissue as a complex system composed of multiple interconnected parts.

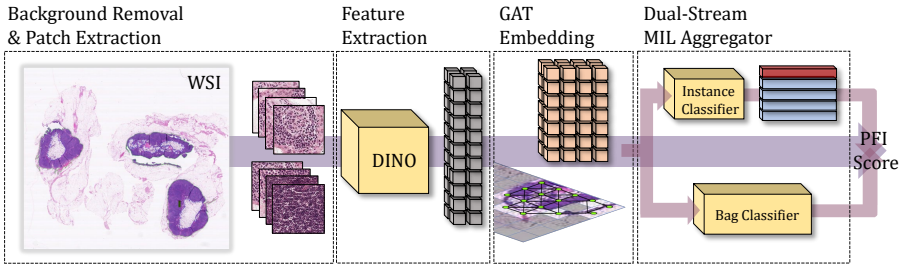

Fig. 1: DINO [4] features extractor is applied to patches tiled from the original WSI. The embeddings thus obtained are fed to a GAT module to capture patches' context and generate a more contextualized representation. A dual-stream MIL aggregation module is then employed to obtain the final prediction by averaging the scores of instance and bag classifiers.

3 Model

In this study, we propose the use of a GAT to contextualize instances (WSI patches) through local interaction before MIL aggregation. Fig. 1 summarizes the key elements of the proposed method, which are detailed in the following of this Section.

3.1 Graph Integration

Given the data as a set of instances $x_i^{ins} \in X^{bag}$ and a self-supervised feature extractor f, informative and discriminative embeddings are obtained as follows[4]:

$$E^{bag} = f(X^{bag}) = \{f(x_0^{ins}), ..., f(x_i^{ins})\} = \{E_0^{ins}, ..., E_i^{ins}\} \tag{2}$$

Each embedding contains important local information inside the patch (*e.g.,* representing the morphology). In order to also capture the micro and macro interaction between instances, we apply a GNN G [11,26,34], implemented with GATs. Given an adjacency matrix $\mathcal{A}$ considering the spatial coordinates of the instances (*e.g.,* each patch is connected to its at most 8 closest neighbors), a more contextualized instance representation is obtained as:

$$\widehat{E}^{bag} = G(E^{bag}, A) \tag{3}$$

3.2 Graph Attention Layer

The GAT applies a masked attention on each instance $E_i^{ins} \in E^{bag}$ and its neighborhood $E_j^{ins} \in \mathcal{N}_i$. The neighborhood of each instance can be found in the adjacency matrix $\mathcal{A}$. At the starting point, each instance is processed with

[4] x_i^{ins} represents a patch extracted from the X^{bag}, *i.e.,* the entire WSI.

a shared weight matrix W$\in \mathbb{R}$ as $H^{ins} = W(E^{ins})$. The instance interaction is measured by an α_{ij} computed as:

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(a(H_i \parallel H_j))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(a(H_i \parallel H_k))} \tag{4}$$

where $a \in \mathbb{R}^{2F}$ is a single-layer feedforward neural network and $\parallel$ is the concatenation operator. A multi-head attention produces a new instance representation as the average of the linear combinations of the neighborhood among each head $k \in K$:

$$\widehat{E}_i^{ins} = \sigma(\frac{1}{K} \sum_{k \in K} \sum_{j \in N_i} (\alpha_{ij}^k H_j^k)) \tag{5}$$

where σ is a softmax operation.

3.3 Bag-level Representation

Taking inspiration from DS-MIL [15], the bag representation is built through a dual stream approach. In particular, starting from the graph output $\widehat{E}_i^{ins} \in \widehat{E}^{bag}$ a first patch classifier f_{patch} is used to identify the most critical patch instance as:

$$\widehat{E}_{crit}^{ins} = \underset{\widehat{E}_i^{ins}}{\text{argmax}}\, f(\widehat{E}_i^{ins}) \tag{6}$$

Given the most relevant instance, $\widehat{E}_{crit}^{ins}$, and a linear-layer neural networks, U, it is possible to build the attention scores of the current instance, $\widehat{E}_i^{ins}$, considering its similarity with $\widehat{E}_{crit}^{ins}$:

$$A_i = \text{softmax}(< U(\widehat{E}_i^{ins}), U(\widehat{E}_{crit}^{ins}) >) \tag{7}$$

After that, the bag label is obtained applying a classifier $\mathbf{W}_{\text{CLS}}$ over the bag embedding built as:

$$y_{\text{BAG}} = \mathbf{W}_{\text{CLS}} \sum_{i}^{n} \underbrace{A_i}_{\substack{\text{Attention scores w.r.t.} \\ \text{the critical patch.}}} * \underbrace{\text{V}(E_i^{ins})}_{\text{Patch-level value.}} \tag{8}$$

where V is another linear-layer neural networks, and n is $|\widehat{E}^{bag}|$.

4 Experimental Setup

4.1 Dataset

The dataset is composed by 176 omentum-tissue-WSIs [20] belonging to 77 different HGSOC patients who underwent NACT therapy. The staining procedure used for the WSIs was Hematoxylin and Eosin (HE) [19]. Images have been

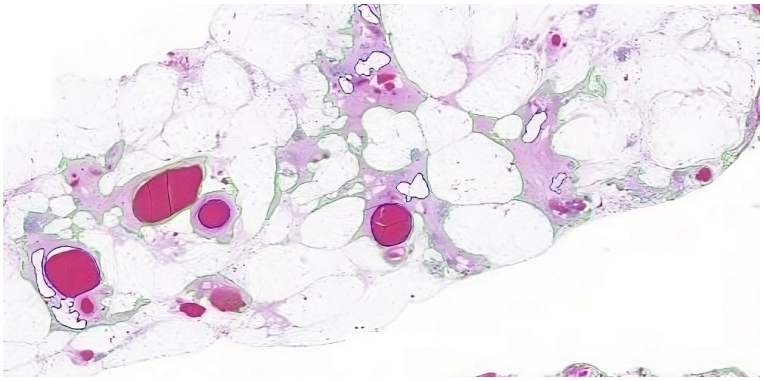

Fig. 2: Example of segmentation masks generated by the pre-processing algorithm. Green contours identify the considered tissue, blue ones are holes the algorithm will discard. The procedure allows for filtering out background, fat, and blood.

scanned by a Pannoramic SCAN 150 with a resolution of 0.22 μm/pixel at the 40× resolution. Each WSI is assigned a label based on the patients' PFI: those with a poor prognosis, *low-PFI* ($\leq$ 6 months), are 99 in total, while the other 77 scans have an *high-PFI* ($\geq$ 12 months). The dataset is split into 4-folds in order to perform cross-validation. For each split, a balance between low- and high-PFI was respected. We also ensured that WSIs from the same patient were not mixed between training and test sets.

4.2 Pre-processing

The state-of-the-art CLAM [18] framework has been employed to crop each WSI into multiple patches. This strategy involves selecting only relevant tissue by means of Otsu thresholding [38] and Connected Components Analysis [1].

Additionally, a red filter is used to remove blood[5]. An example of the resulting segmentation mask is shown in Fig. 2. The green contour delineates a portion of tissue that is preserved; the blue one indicates a removed area (holes). The preserved area is then cropped into non-overlapping 256×256 patches at different resolution scales. 20× and 5× resolutions were chosen to capture both micro and macro details in the dataset. On average, each WSI contains 5 960 patches at 20× resolution and 370 patches at 5× resolution.

DINO [4], a Vision Transformer (ViT) model [5], is then employed to produce high quality patch representations, while ensuring a fast processing with low computational resource requirements. This approach focuses on aligning exclusively the positive pairs by leveraging a teacher-student framework, which

[5] A fixed threshold is applied on the HSV (Hue Saturation Brightness) color space of the WSI thumbnail and later propagated to 5× and 20× resolutions.

Table 1: Performance comparison. Experiments were run 5 times, each with a 4-fold cross-validation. This table reports the average results and the corresponding standard deviation.

		Best Epoch		Last Epoch	
Scale	**Approach**	**Accuracy**	**AUC**	**Accuracy**	**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	0.606 ± 0.076	0.467 ± 0.171	0.577 ± 0.076	0.413 ± 0.166
	DS-MIL	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	0.594 ± 0.095	0.510 ± 0.141	0.576 ± 0.090	0.438 ± 0.156
	DS-MIL	0.681 ± 0.033	0.650 ± 0.049	0.656 ± 0.028	0.572 ± 0.065
	GDS-MIL (our)	$\mathbf{0.704 \pm 0.070}$	$\mathbf{0.661 \pm 0.099}$	$\mathbf{0.663 \pm 0.064}$	$\mathbf{0.611 \pm 0.092}$

Table 2: Performance comparison on an Out of Distribution (OOD) testset.

		Best Epoch		Last Epoch	
Scale	**Approach**	**Accuracy**	**AUC**	**Accuracy**	**AUC**
5×	MaxPooling	0.552 ± 0.052	0.422 ± 0.102	$\mathbf{0.573 \pm 0.021}$	0.414 ± 0.103
	MeanPooling	0.556 ± 0.001	0.385 ± 0.010	0.556 ± 0.001	0.384 ± 0.009
	AB-MIL	0.563 ± 0.029	0.392 ± 0.065	0.517 ± 0.042	0.335 ± 0.062
	DS-MIL	0.486 ± 0.015	0.411 ± 0.015	0.500 ± 0.001	0.405 ± 0.124
	GDS-MIL (our)	$\mathbf{0.618 \pm 0.036}$	$\mathbf{0.490 \pm 0.025}$	0.566 ± 0.044	$\mathbf{0.429 \pm 0.033}$
20×	MaxPooling	0.646 ± 0.041	0.611 ± 0.048	0.625 ± 0.065	0.530 ± 0.067
	MeanPooling	0.580 ± 0.010	0.388 ± 0.010	0.580 ± 0.010	0.388 ± 0.010
	AB-MIL	0.510 ± 0.039	0.430 ± 0.008	0.500 ± 0.001	0.386 ± 0.012
	DS-MIL	0.670 ± 0.023	0.632 ± 0.001	0.653 ± 0.015	0.540 ± 0.011
	GDS-MIL (our)	$\mathbf{0.764 \pm 0.039}$	$\mathbf{0.726 \pm 0.042}$	$\mathbf{0.712 \pm 0.063}$	$\mathbf{0.667 \pm 0.082}$

comprises two separate networks. We trained the model over the entire set of patches, separately for each resolution level.

4.3 Implementation Details

The optimization is performed using *Adam* with a learning rate of $2 * 10^{-4}$ and a weight decay of $5 * 10^{-3}$. The training is carried out for 200 epochs with the *CosineAnnealingLR* scheduler. We employ one single GAT layer with 3 heads used for multi-head attention. All the experiments are conducted using a unified codebase and under identical experimental conditions. Each bag is sub-sampled using a patch dropout probability of 0.5 to increase the number of bags and promote randomness during training. The Area Under the Curve (AUC) and the accuracy metrics are calculated as described in [30]. To ensure a fair comparison, all methods considered in our analysis are evaluated using the same metrics.

5 Results and Discussion

A comparison of the proposed solution with state-of-the-art MIL approaches is reported in Tab. 1 and Tab. 2. All the experiments have been performed on the

previously described dataset and repeated 5 times to stress the robustness of the algorithms. Tables report the average performance and the associated standard deviation at 5× and 20× resolutions.

We compared the proposed model GDS-MIL with MaxPooling and MeanPooling to understand the effectiveness of patch-level classifiers, and AB-MIL [9] and DS-MIL [15] as state-of-the-art attention based MIL solutions. The performance of each approach is measured with average accuracy and average AUC, both at the best and last epoch. The best epoch is the one where the model obtains the best performance considering the average between accuracy and AUC on the test set, while the last epoch is the end of the training phase.

Experimental results demonstrate that GDS-MIL outperforms all the other approaches on both scales, achieving the highest accuracy and AUC scores at the best epochs. DS-MIL also performs well, achieving good scores on both scales, while MeanPooling and AB-MIL show moderate performance. Overall, the results suggest that integrating a graph-based solution improves our baseline (DS-MIL) by 3.5% on accuracy and 2% on AUC.

Even when considering only the last epoch, GDS-MIL outperforms the baselines improving DS-MIL by 1.3%.

In Tab. 2 we investigated a specific dataset split characterized by significant tissue heterogeneity. In this case, the contextualization introduced with the graph plays an even more relevant role: our model outperforms DS-MIL by 9.4% on accuracy and 9.3% on AUC.

A further analysis is reported in Tab. 3, stressing the relevance of graph (main) hyper-parameters such as layer type, number of sequential layers, and number of heads within the same graph layer.

5.1 Model Analysis

Experimental results demonstrate that the 20× scale resolution is the most effective when tackling the PFI prediction task on omentum WSIs tacken from NACT patients. Specifically, a patch-level classifier such as MaxPooling can achieve surprisingly good performance at 20× resolution, with an accuracy of 0.676 and AUC of 0.637. This phenomenon implies the existence of morphology and patterns correlated to the PFI which can be exploited to solve the task. This conclusion is also supported by the effectiveness of DS-MIL which achieves an accuracy of 0.681 and an AUC of 0.649. The attention mechanism used by DS-MIL allows to identify the most relevant WSI regions, guiding the PFI classification.

However, adding a graph attention layer can significantly improve the performance at both considered resolutions. This finding suggests that incorporating spatial context into each instance, including both neighborhood morphology and interaction, allows to change the meaning of critical patch. In GDS-MIL, the relevance score of each instance is not limited to the instance itself, but also influenced by the area where it is located, allowing for a more fine-grained criticality assessment. These results suggest that the proposed model is highly effective and can offer significant improvements over existing state-of-the-art approaches. The

Table 3: Performance comparison changing the type of graph layer (type), the number of layers ($\mathcal{L}$) and heads ($\mathcal{H}$) used by the graph neural network.

Type	$\mathcal{L}$	$\mathcal{H}$	AUC	Acc.	Type	$\mathcal{L}$	$\mathcal{H}$	AUC	Acc.	Type	$\mathcal{L}$	$\mathcal{H}$	AUC	Acc.
GCN	1	1	0.625	0.667	GAT	1	1	0.664	0.704	GATv2	1	1	0.657	0.657
GCN	1	2	0.623	0.648	GAT	1	2	0.667	0.732	GATv2	1	2	**0.734**	**0.732**
GCN	1	3	**0.634**	0.648	GAT	1	3	**0.726**	**0.764**	GATv2	1	3	0.667	0.704
GCN	2	1	0.602	**0.676**	GAT	2	1	0.634	0.722	GATv2	2	1	0.679	0.722
GCN	2	2	0.608	0.648	GAT	2	2	0.607	0.648	GATv2	2	2	0.628	0.694
GCN	2	3	0.591	0.657	GAT	2	3	0.619	0.694	GATv2	2	3	0.655	0.713
GCN	3	1	0.564	0.648	GAT	3	1	0.641	0.694	GATv2	3	1	0.641	0.685
GCN	3	2	0.595	0.648	GAT	3	2	0.639	0.704	GATv2	3	2	0.642	0.713
GCN	3	3	0.572	0.639	GAT	3	3	0.697	0.732	GATv2	3	3	0.660	0.713

high standard deviation of all reported experiments is intrinsically connected to the small number of WSIs and to the high heterogeneity of the task.

5.2 Hyperparameter Analysis

To stress the contribution of different graph layers, Tab. 3 is reported. The results indicate that, in general, using layers of a Graph Convolutional Network [11] leads to worse performances compared to GAT [33] and GATv2 [3]. When relying on convolutional layers, the patch representation becomes similar to its neighborhood, resulting in a loss of important details. In contrast, leveraging an attention layer enables the patch to acquire context information, while preserving its own unique features. No significant difference can be observed between GAT and GATv2, with the latter performing slightly better than the former.

The experiments reported in Tab. 3 also reveal that a higher number of graph layers has a negative impact on the performance. This is mainly related to the smoothing operation performed by the graph on the patch representation. If the smoothing is too strong, it becomes challenging for the MIL module to distinguish what is actually important. Therefore, it is crucial to identify a trade-off between the number of layers and the overall performance.

Moreover, increasing the number of heads applied to the attention mechanism generally provide better performances. Indeed, 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.

In summary, our analysis highlights the importance of carefully selecting the graph hyper-parameters. Specifically, the adoption of attention layers usually provide better performance than convolutional graph layers. Limiting the number of graph layers, and considering an higher number of heads during the self-attention process can also improve the final results. This is the reason why we opted for a single GAT layer consisting of three heads.

6 Conclusions

This paper proposes GDS-MIL method which integrates a GAT into a MIL architecture for predicting the PFI of WSIs obtained from NACT patients. Our results demonstrate that introducing a spatial contextualization has beneficial effects on the MIL architecture. A future work will analyze what kind of biological patterns have major impact for the prediction in order to better explain the PFI task.

Acknowledgements This project has received funding from DECIDER, the European Union's Horizon 2020 research and innovation programme under GA No. 965193, and from the Department of Engineering "Enzo Ferrari" of the University of Modena through the FARD-2022 (Fondo di Ateneo per la Ricerca 2022).

References

1. Allegretti, S., Bolelli, F., Cancilla, M., Pollastri, F., Canalini, L., Grana, C.: How does Connected Components Labeling with Decision Trees perform on GPUs? In: Computer Analysis of Images and Patterns. vol. 11678, pp. 39–51. Springer (2019) 6
2. Bontempo, G., Porrello, A., Bolelli, F., Calderara, S., Ficarra, E.: DAS-MIL: Distilling Across Scales for MIL Classification of Histological WSIs. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023 (2023) 2
3. Brody, S., Alon, U., Yahav, E.: How Attentive are Graph Attention Networks? . In: International Conference on Learning Representations (2022) 9
4. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging Properties in Self-Supervised Vision Transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 9650–9660 (2021) 4, 6
5. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, Jakob anf Houlsby, N.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. International Conference on Learning Representations (2021) 6
6. Evans, A.J., Bauer, T.W., Bui, M.M., Cornish, T.C., Duncan, H., Glassy, E.F., Hipp, J., McGee, R.S., Murphy, D., Myers, C., et al.: US Food and Drug Administration Approval of Whole Slide Imaging for Primary Diagnosis: A Key Milestone Is Reached and New Questions Are Raised. Archives of Pathology & Laboratory Medicine **142**(11), 1383–1387 (2018) 2
7. Fatemi, M., Feng, E., Sharma, C., Azher, Z., Goel, T., Ramwala, O., Palisoul, S., Barney, R., Perreard, L., Kolling, F., et al.: Inferring spatial transcriptomics markers from whole slide images to characterize metastasis-related spatial heterogeneity of colorectal tumors: A pilot study. Journal of Pathology Informatics p. 100308 (2023) 2
8. Feng, J., Zhou, Z.H.: Deep MIML Network. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. p. 1884–1890. AAAI Press (2017) 3

9. Ilse, M., Tomczak, J., Welling, M.: Attention-based Deep Multiple Instance Learning. In: Proceedings of the 35th International Conference on Machine Learning. pp. 2127–2136. PMLR (2018) 2, 3, 8
10. Kehoe, S., Hook, J., Nankivell, M., Jayson, G.C., Kitchener, H., Lopes, T., Luesley, D., Perren, T., Bannoo, S., Mascarenhas, M., et al.: Primary chemotherapy versus primary surgery for newly diagnosed advanced ovarian cancer (CHORUS): an open-label, randomised, controlled, non-inferiority trial. The Lanchet **386**(9990), 249–257 (2015) 2
11. Kipf, T.N., Welling, M.: Semi-Supervised Classification with Graph Convolutional Networks. In: International Conference on Learning Representations. ICLR (2017) 4, 9
12. Kumar, N., Gupta, R., Gupta, S.: Whole Slide Imaging (WSI) in Pathology: Current Perspectives and Future Directions. Journal of Digital Imaging **33**(4), 1034–1040 (2020) 2
13. Landi, F., Baraldi, L., Corsini, M., Cucchiara, R.: Embodied Vision-and-Language Navigation with Dynamic Convolutional Filters. In: Proceedings of the British Machine Vision Conference (2019) 2
14. Laury, A.R., Blom, S., Ropponen, T., Virtanen, A., Carpén, O.M.: Artificial intelligence-based image analysis can predict outcome in high-grade serous carcinoma via histology alone. Scientific Reports **11**(1), 19165 (2021) 3
15. Li, B., Li, Y., Eliceiri, K.W.: Dual-stream Multiple Instance Learning Network for Whole Slide Image Classification with Self-supervised Contrastive Learning. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14318–14328 (2021) 2, 3, 5, 8
16. Lovino, M., Ciaburri, M.S., Urgese, G., Di Cataldo, S., Ficarra, E.: DEEPrior: a deep learning tool for the prioritization of gene fusions. Bioinformatics **36**(10) (2020) 2
17. Lu, M.Y., Chen, R.J., Wang, J., Dillon, D., Mahmood, F.: Semi-Supervised Histology Classification using Deep Multiple Instance Learning and Contrastive Predictive Coding. arXiv preprint arXiv:1910.10825 (2019) 3
18. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature Biomedical Engineering **5**(6), 555–570 (2021) 3, 6
19. Martina, J.D., Simmons, C., Jukic, D.M.: High-definition hematoxylin and eosin staining in a transition to digital pathology. Journal of Pathology Informatics **2**(1), 45 (2011) 5
20. Meza-Perez, S., Randall, T.D.: Immunological Functions of the Omentum. Trends in Immunology **38**(7), 526–536 (2017) 5
21. Morelli, D., Fincato, M., Cornia, M., Landi, F., Cesari, F., Cucchiara, R.: Dress Code: High-Resolution Multi-Category Virtual Try-On. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2022) 2
22. Nikolaidi, A., Fountzilas, E., Fostira, F., Psyrri, A., Gogas, H., Papadimitriou, C.: Neoadjuvant treatment in ovarian cancer: New perspectives, new challenges. Frontiers in Oncology p. 3758 (2022) 2
23. Panariello, A., Porrello, A., Calderara, S., Cucchiara, R.: Consistency-Based Self-supervised Learning for Temporal Anomaly Localization. In: Computer Vision – ECCV 2022 Workshops (2022) 3
24. Pinheiro, P.O., Collobert, R.: From image-level to pixel-level labeling with Convolutional Networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1713–1721 (2015) 3

25. Ponzio, F., Urgese, G., Ficarra, E., Di Cataldo, S.: Dealing with Lack of Training Data for Convolutional Neural Networks: The Case of Digital Pathology. Electronics **8**(3), 256 (2019) 2
26. Porrello, A., Abati, D., Calderara, S., Cucchiara, R.: Classifying Signals on Irregular Domains via Convolutional Cluster Pooling. In: The 22nd International Conference on Artificial Intelligence and Statistics. pp. 1388–1397 (2019) 4
27. Pujade-Lauraine, E., Combe, P.: Recurrent ovarian cancer. Annals of Oncology **27**, i63–i65 (2016) 1, 2
28. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., et al.: TransMIL: Transformer based Correlated Multiple Instance Learning for Whole Slide Image Classification. Advances in Neural Information Processing Systems 34 (NeurIPS) **34**, 2136–2147 (2021) 3
29. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. In: 3rd International Conference on Learning Representations (ICLR 2015) (2015) 3
30. Sokolova, M., Japkowicz, N., Szpakowicz, S.: Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. In: AI 2006: Advances in Artificial Intelligence. pp. 1015–1021. Springer (2006) 7
31. Tomei, M., Baraldi, L., Calderara, S., Bronzin, S., Cucchiara, R.: RMS-Net: Regression and Masking for Soccer Event Spotting. In: 25th International Conference on Pattern Recognition (ICPR). pp. 7699–7706. IEEE (2021) 2
32. Tourniaire, P., Ilie, M., Hofman, P., Ayache, N., Delingette, H.: Attention-based Multiple Instance Learning with Mixed Supervision on the Camelyon16 Dataset. In: Proceedings of the MICCAI Workshop on Computational Pathology. pp. 216–226. PMLR (2021) 2
33. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., et al.: Graph Attention Networks. Stat **1050**(20), 10–48550 (2017) 2, 9
34. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Philip, S.Y.: A Comprehensive Survey on Graph Neural Networks. IEEE Transactions on Neural Networks and Learning Systems **32**(1), 4–24 (2020) 4
35. Yao, J., Zhu, X., Jonnagaddala, J., Hawkins, N., Huang, J.: Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. Medical Image Analysis **65**, 101789 (2020) 2
36. Yu, K.H., Hu, V., Wang, F., Matulonis, U.A., Mutter, G.L., Golden, J.A., Kohane, I.S.: Deciphering serous ovarian carcinoma histopathology and platinum response by convolutional neural networks. BMC medicine **18**(1), 1–14 (2020) 3
37. Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S.E., Zheng, Y.: DTFD-MIL: Double-Tier Feature Distillation Multiple Instance Learning for Histopathology Whole Slide Image Classification. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 18802–18812 (2022) 3
38. Zhang, J., Hu, J.: Image Segmentation Based on 2D Otsu Method with Histogram Analysis. In: International Conference on Computer Science and Software Engineering. vol. 6, pp. 105–108. IEEE (2008) 6