Graph Laplacian Transformer with Progressive Sampling for Prostate Cancer Grading

Masum Shah Junayed[1], John Derek Van Vessem[2], Qian Wan[2], Gahie Nam[2] and Sheida Nabavi[1]

School of Computing, University of Connecticut (UConn), Storrs, CT 06269, USA[1]
Pathology and Laboratory Medicine, UConn Health, Farmington, CT 06030, USA[2]

Abstract. Prostate cancer grading from whole-slide images (WSIs) remains a challenging task due to the large-scale nature of WSIs, the presence of heterogeneous tissue structures, and difficulty of selecting diagnostically relevant regions. Existing approaches often rely on random or static patch selection, leading to the inclusion of redundant or non-informative regions that degrade performance. To address this, we propose a Graph Laplacian Attention-Based Transformer (GLAT) integrated with an Iterative Refinement Module (IRM) to enhance both feature learning and spatial consistency. The IRM iteratively refines patch selection by leveraging a pretrained ResNet50 for local feature extraction and a foundation model in no-gradient mode for importance scoring, ensuring only the most relevant tissue regions are preserved. The GLAT models tissue-level connectivity by constructing a graph where patches serve as nodes, ensuring spatial consistency through graph Laplacian constraints and refining feature representations via a learnable filtering mechanism that enhances discriminative histological structures. Additionally, a convex aggregation mechanism dynamically adjusts patch importance to generate a robust WSI-level representation. Extensive experiments on five public and one private dataset demonstrate that our model outperforms state-of-the-art methods, achieving higher performance and spatial consistency while maintaining computational efficiency.

Keywords: Progressive Sampling · High Informative Patch · Graph Laplacian Attention · Transformer · Histopathology.

1 Introduction

Prostate cancer remains a leading cause of cancer-related mortality worldwide, with whole-slide image (WSI) analysis being essential for grading and risk assessment [3]. The Gleason grading system, which evaluates glandular structures, is the standard for prognosis but is challenging due to differences in expert interpretation, high computational costs, and tissue artifacts like folding and staining inconsistencies [17], [26]. Many computational pathology models treat all tissue patches equally, failing to focus on the most relevant areas [4]. This leads to lower grading accuracy and increased computational burden, highlighting the need for a more efficient and attention based sampling approach [10].

Deep learning-based prostate cancer grading predominantly relies on multiple instance learning (MIL) frameworks, where WSI-level labels are inferred from patch-level features. While attention-based MIL models [20], [16], [19] attempt to highlight relevant regions, they struggle with non-informative patches due to their reliance on static attention mechanisms, leading to performance degradation. Correlation-based MIL methods [24], [4], [21] improve inter-patch dependencies but lack explicit spatial constraints, resulting in inconsistent Gleason grading predictions. Graph-based approaches, such as GNN-based models [1], [2], [23], build local neighborhood graphs from high-attended patches to capture tissue-level connectivity. However, they require extensive computations and high memory usage, limiting their practicality for real-world applications. Furthermore, transformer-based models [24], [5], [14], [13] use self-attention mechanisms to model long-range dependencies, yet they struggle with random patch selection, often discarding critical regions while retaining less informative ones. These challenges highlight the need for a method that adaptively refines patch selection while enforcing spatial constraints to preserve histological consistency.

To overcome these challenges, this work introduces the iterative refinement module (IRM) for adaptive patch selection and graph laplacian attntion-based transfomrer (GLAT) for spatially coherent feature learning. The IRM leverages a pretrained ResNet50 for local feature extraction and a foundation model (FM) [7] operating in no-gradient mode to iteratively refine patch importance scores. This ensures that only the most relevant tissue regions contribute while eliminating redundant or non-informative areas. However, IRM does not explicitly model spatial dependencies, which are essential for preserving glandular structures and histological patterns. To address this, the GLAT incorporates graph Laplacian constraints to maintain spatial consistency by modeling histologically similar patches as graph nodes and enforcing smooth feature transitions between them. Additionally, a learnable filtering mechanism refines feature representation by dynamically adjusts the influence of neighboring patches through graph-based feature propagation, ensuring spatial coherence while preserving glandular boundaries and tissue morphology. Finally, a convex aggregation mechanism consolidates refined patch features into a robust WSI-level representation, ensuring proportional contribution from the most informative patches for accurate classification. The key contributions of this work are as follows:

- This work presents a novel transformer-based model that dynamically selects and processes high-relevance regions to improve prostate cancer grading.
- The Iterative Refinement Module (IRM) introduces an efficient patch selection strategy by refining patch importance scores, eliminating irrelevant regions while reducing computational overhead.
- The Graph Laplacian Attention-Based Transformer (GLAT) enforces spatial consistency through graph Laplacian constraints and enhances feature representation via learnable filtering.
- Extensive experiments on five public and one private dataset demonstrate the superiority of the proposed framework over state-of-the-art methods.

2   Proposed Method

Figure 1 depicts the overview of proposed model. As shown in Figure 1, the model first extracts patch embeddings using a pretrained ResNet50, followed by an IRM to refine high-informative patch selection. Then, Graph Laplacian Transformer to model spatial relationships followed by convex aggregation and classification head for Gleason grading.
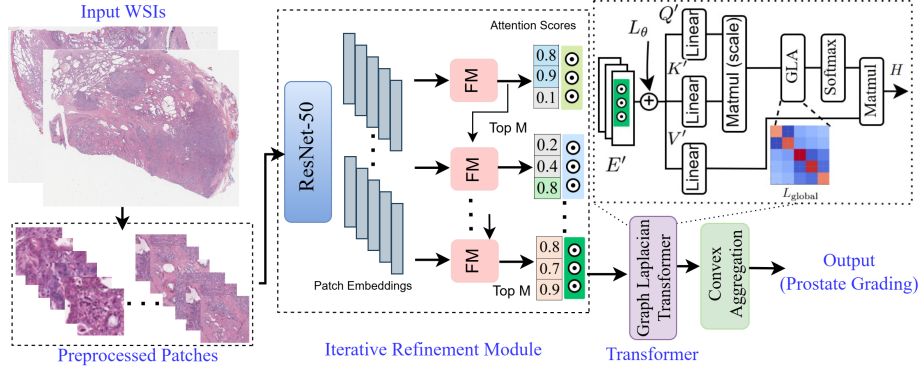

Fig. 1. Overview of the proposed prostate cancer grading model. The model extracts patches from WSIs, scores their relevance using an IRM with ResNet-50 and a no-gradient FM. These selected patches are then processed through a Graph Laplacian Transformer to capture spatial relationships, followed by convex aggregation for WSI-level feature representation, and a classification head for final Gleason grading.

2.1   Iterative Refinement Module

The IRM operates in two stages: (1) local feature extraction using ResNet50 [11] and (2) context-aware scoring using a frozen FM, specifically the UNI model [7]. ResNet50 is used to extract local feature embeddings for each patch, while the FM model, operating in no-gradient mode, assigns attention-based importance scores that capture inter-patch relationships, enabling efficient global reasoning without additional training overhead. The IRM leverages these scores to progressively refine the patch subset across multiple iterations. At each step, patches are rescored based on their contextual relevance, and the least informative ones are discarded. This iterative filtering mechanism allows the model to concentrate on the most diagnostically relevant tissue regions while significantly reducing computational cost. Each patch $P_i \in \mathcal{P}$ is first passed through a pretrained ResNet50 model to extract feature embeddings: $E_i = f_{\text{ResNet}}(P_i)$, $E_i \in \mathbb{R}^d$, where f_{ResNet} represents the ResNet50 feature extractor, and $d = 512$ is the dimensionality of the output embeddings. The embeddings $E = \{E_1, E_2, \ldots, E_N\}$ encapsulate local characteristics of the tissue patches. Next, the embeddings are

passed through the FM, which employs a self-attention mechanism to capture pairwise relationships between patches while remaining frozen (i.e., weights are not updated during training). The attention weights A_{ij} between patches i and j are computed as:

$$A_{ij} = \text{softmax}\left(\frac{Q_i K_j^\top}{\sqrt{d_k}}\right), \quad Q_i = W_Q E_i, \ K_j = W_K E_j, \tag{1}$$

where $W_Q, W_K \in \mathbb{R}^{d \times d_k}$ are fixed projection matrices from the FM that define the query and key representations, ensuring that the learned attention mechanism is based on pretrained knowledge. The softmax function normalizes the attention scores, ensuring that the influence of each patch is effectively distributed across all other patches.

Using the computed attention weights, the FM refines the embeddings (E_i') of each patch by aggregating information from its neighbors: $E_i' = \sum_{j=1}^{N} A_{ij} V_j$, $V_j = W_V E_j$, where $W_V \in \mathbb{R}^{d \times d_v}$ projects the value vectors. The refined embeddings $E' = \{E_1', E_2', \ldots, E_N'\}$ encode both local features and global contextual dependencies, making them suitable for scoring the patches. These embeddings are then used to compute patch importance scores: $S_i = \frac{\sum_{j=1}^{N} A_{ij}}{N}$, where S_i represents the average attention weight of patch P_i across all other patches, reflecting its overall contribution to the contextual structure of the WSI.

Patch embeddings are divided in several (T) non-overlapping subsets. At the first iteration $(t = 0)$, the first subset of patch embeddings are passed through the FM, which assigns initial importance scores $S_i^{(0)}$. Based on these scores, the top M patches with the highest scores are selected: $\mathcal{P}_{\text{selected}}^{(0)} = \{P_i^{(0)} : S_i^{(0)} \text{ is among the top } M\}$. The subset of selected patches are denoted as $\mathcal{P}_{\text{selected}}^{(0)} = \{P_{i_1}^{(0)}, P_{i_2}^{(0)}, \ldots, P_{i_M}^{(0)}\}$. In each subsequent iteration t, the embeddings of the selected patches from the previous iteration are combined with the next subset of patches and reprocessed using the FM, which recalculates their contextual relationships with the other patches in the subset. The refined embeddings $\{E_i'^{(t)}\}$, and scores $\{S_i^{(t)}\}$ are updated as :

$$E_i'^{(t)} = \sum_{j \in N/T} A_{ij}^{(t)} V_j^{(t)}, \quad S_i^{(t)} = \frac{\sum_{j=1}^{N/T} A_{ij}^{(t)}}{N/T}. \tag{2}$$

At each iteration, the FM updates the patch importance scores, progressively refining the selection process by keeping only the top M patches with the highest scores: $\mathcal{P}_{\text{selected}}^{(t)} = \{P_i^{(t)} : S_i^{(t)} \text{ is among the top } M\}$. At the end of T iterations, the IRM process produces the final set of selected patches $\mathcal{P}_{\text{selected}}^{(T)}$ is passed to the next stage of the framework for downstream analysis.

2.2   Graph Laplacian Transformer

To capture both spatial coherence and long-range contextual dependencies among high-informative patches, the GLAT is introduced. The GLAT addresses a criti-

cal limitation of standard multihead self attention (MSA) by explicitly enforcing spatial consistency. Unlike MSA, which lacks spatial regularization, GLAT models spatial and morphological relationships by connecting histologically similar patches in a graph structure. To explicitly model the spatial relationships and tissue-level connectivity among selected patches, we represent them as a node in a graph $G = (Vt, X)$. Here, Vt corresponds to the high-informative patches and X defines the edges that capture histological feature similarity. To construct the graph, an edge X_{ij} is established between patches i and j based on their feature similarity. The adjacency matrix W is computed using a Gaussian kernel function to quantify the pairwise similarity between patches:

$$W_{ij} = \exp\left(-\frac{\|E_i' - E_j'\|^2}{2\sigma^2}\right), \tag{3}$$

where W_{ij} measures the feature similarity between patches i and j, with E_i' and E_j' denoting their respective feature embeddings. The parameter σ controls the sensitivity of similarity weighting. The degree matrix D, defined as : $D_{ii} = \sum_j W_{ij}$, Using these matrices, the global graph Laplacian is computed as: $L_{\text{global}} = D - W$.

To refine feature representations before self-attention, GLAT employs a learnable filtering mechanism that dynamically adjusts feature propagation across patches. This is formulated as: $Q' = L_\theta Q, \quad K' = L_\theta K, \quad V' = L_\theta V$, where L_θ is a trainable filter optimized during training to control feature propagation. L_θ learns adaptive transformations that enhance discriminative features while preserving local structural details. Using the learnable-filtered queries, keys, and values, the modified graph laplacian attention (GLA) mechanism is computed as:

$$A' = \text{softmax}\left(\frac{Q'K'^\top + \lambda L_{\text{global}}}{\sqrt{d_k}}\right), \tag{4}$$

where λ is a tunable hyperparameter that regulates the influence of the spatial constraints, ensuring an optimal balance between feature-driven attention and structured spatial coherence within the GLA mechanism.. The resulting attention scores refine feature embeddings as: $H = A'V', \quad H \in \mathbb{R}^{M \times d}$, where H contains individual refined embeddings for each selected patch. To generate a global WSI representation, we apply convex aggregation [12], which ensures that the most relevant refined patches contribute proportionally:

$$H_{\text{WSI}} = \sum_{i=1}^{M} w_i H_i', \quad w_i = \frac{\exp(\theta_i)}{\sum_{j=1}^{M} \exp(\theta_j)}, \tag{5}$$

where θ_i are trainable parameters that determine the relative importance of each patch. The softmax function is applied to θ_i to obtain normalized weights w_i, ensuring that they are non-negative $w_i \geq 0,$ and $\sum_{i=1}^{M} w_i = 1$. This normalization allow the model to dynamically adjust patch importance during training while maintaining a balanced feature aggregation. Finally, the WSI representation H_{WSI} is passed through a classification head: $y = \text{Softmax}(\text{Linear}(H_{\text{WSI}}))$.

The model is trained using categorical cross-entropy loss for Gleason grading. To further encourage spatial consistency, a Graph-based feature smoothness constraint is incorporated into the loss function:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \alpha \sum_{i,j} W_{ij} \|H_i - H_j\|^2, \tag{6}$$

where $\mathcal{L}_{\text{CE}}$ represents the loss of standard classification, and the second term encourages the consistency of characteristics between spatially similar patches. The hyperparameter α balances classification performance with spatial coherence.

3   Experimental Results

3.1   Datasets and Preprocessing

This study utilizes five diverse publicly availabe datasets for evaluating the proposed framework: TCGA-PRAD [9], SICAPv2 [25], GLEASON19 [22], PANDA, DiagSet [15], and a Private dataset. The private (UConn Health) dataset includes 79 WSIs, enhancing the evaluation of the model on a smaller yet clinically relevant dataset. For grading, the ISUP classification system was employed, categorizing samples into four classes: Grade 1 and 2 representing normal tissue, and Grades 3, 4, and 5 indicating varying levels of malignancy. For preprocessing, the CLAM [19] method was employed to generate high-quality patches from WSIs and TMAs. The preprocessing pipeline included stain normalization, tissue segmentation, and patch extraction with a fixed size, ensuring consistency across datasets. Patches with minimal tissue content were excluded to enhance data quality, and each patch was normalized to reduce staining variability.

3.2   Experimental Setup

All models, including the proposed method and existing baseline methods, were trained on high-performance GPUs (NVIDIA RTX A6000) to handle large-scale histopathological datasets. The input patches were extracted using the CLAM [19] standard preprocessing pipeline with a patch size of 224, followed by standard data augmentation techniques such as random flipping and rotation to enhance model robustness. A batch size of 16 was used with an initial learning rate of 1×10^{-4}, optimized using the Adam optimizer with a weight decay set to 1×10^{-5}. The early stopping strategy was applied to prevent overfitting based on the validation performance, and all models were trained for up to 100 epochs. The performance of the model was assessed primarily using AUC and Cohen's Kappa (CK) reported as mean over five-fold cross-validation for statistical reliability. The proposed method was evaluated against state-of-the-art baseline models. To ensure fair comparison, publicly available codebases were used, and all hyperparameters were aligned with the original implementations.

Table 1. Quantitative comparison of the proposed method against state-of-the-art approaches on six prostate cancer grading datasets. Performance is measured using AUC and CK. The best results are in bold, and the second-best results are underlined.

Dataset/ Methods	SICAPv2 AUC	SICAPv2 CK	TCGA-PRAD AUC	TCGA-PRAD CK	GLESON19 AUC	GLESON19 CK	PANDA AUC	PANDA CK	Diagset AUC	Diagset CK	Private AUC	Private CK
ABMIL [20]	0.658	0.598	0.616	0.591	0.648	0.601	0.631	0.605	0.598	0.546	0.534	0.496
TranMIL [24]	0.593	0.567	0.587	0.538	0.612	0.582	0.605	0.580	0.535	0.486	0.481	0.437
MST [4]	0.918	0.796	0.863	0.792	0.839	0.810	0.895	0.851	0.720	0.679	0.679	0.609
DASMIL [5]	0.915	0.819	0.867	0.799	0.846	0.808	0.897	0.846	0.736	0.685	0.683	0.628
WSDMPC [2]	0.819	0.785	0.835	0.805	0.826	0.808	0.860	0.806	0.678	0.618	0.650	0.605
MaskHIT [13]	0.938	0.868	0.909	0.856	0.887	0.853	0.922	0.901	0.785	0.724	0.746	0.701
SMAHM [14]	0.941	0.881	0.918	0.876	**0.914**	0.876	<u>0.958</u>	<u>0.913</u>	<u>0.819</u>	<u>0.768</u>	<u>0.755</u>	<u>0.719</u>
HEAT [6]	<u>0.943</u>	0.875	<u>0.921</u>	**0.886**	0.905	<u>0.889</u>	0.946	0.909	0.805	0.759	0.748	0.706
Proposed	**0.951**	**0.892**	**0.923**	<u>0.885</u>	<u>0.913</u>	**0.892**	**0.963**	**0.936**	**0.836**	**0.791**	**0.781**	**0.731**

3.3 Results and Discussions

Table 1 presents the performance comparison of the proposed method with existing state-of-the-art approaches on six benchmark datasets, including SICAPv2, TCGA-PRAD, GLEASON19, PANDA, DiagSet, and Private. The evaluation metrics used are AUC and Cohen's Kappa, both of which assess model reliability and agreement with ground truth annotations. The proposed method consistently achieves the highest AUC and Kappa scores across all datasets, demonstrating superior grading performance. Notably, the model outperforms the closest competitor, HEAT, on SICAPv2, TCGA-PRAD, and PANDA, while maintaining competitive performance on the remaining datasets. The significant improvement in CK highlights the model's robustness in handling class imbalance and variability in histopathological features.

Table 2. Ablation study evaluating the impact of key components in the proposed framework for prostate cancer grading.

Exp.	IRM ResNet50	FM	IP	Transformer GLA	MSA	SWA	CA	Performance AUC	Performance CK	Comp. Cost #P(M)	Comp. Cost Flops
1	✓	✓	✓	✓	×	×	✓	**0.781**	**0.731**	**83.3**	**32.53**
2	×	✓	✓	✓	×	×	✓	0.779	0.725	89.9	43.56
3	✓	×	✓	✓	×	×	✓	0.737	0.696	98.6	49.46
4	✓	✓	×	✓	×	×	✓	0.763	0.708	130.5	91.63
5	✓	✓	✓	×	✓	×	✓	0.754	0.709	86.7	41.58
6	✓	✓	✓	×	×	✓	✓	0.751	0.710	84.2	33.9
7	✓	✓	✓	✓	×	×	×	0.769	0.662	84.1	35.94

Table 2 presents an ablation study assessing the impact of individual components within the proposed framework on Private dataset. The full model (Exp. 1), incorporating IRM with GLA and CA, achieved the highest performance (AUC: 0.781, CK: 0.731) with an optimal computational balance (83.3M pa-

rameters, 32.53 FLOPs). Replacing ResNet50 with ViT [8] (Exp. 2) slightly reduced performance while increasing computational cost. Excluding FM (Exp. 3) significantly lowered AUC (0.737) and CK (0.696), demonstrating its importance in refining patch selection. Removing iterative process (IP) (Exp. 4) and relying on random patch selection led to degraded performance and increased overhead (130.5M parameters, 91.6 FLOPs), emphasizing the necessity of adaptive sampling. Substituting GLA with MSA (Exp. 5) or SWA (Exp. 6) resulted in lower AUC, confirming the advantage of GLA in modeling spatial relationships. Lastly, replacing CA with mean pooling (Exp. 7) significantly reduced AUC (0.969 to 0.662), highlighting CA's role in forming a robust WSI representation. Figure 2 presents a comparative analysis of different attention mechanisms—
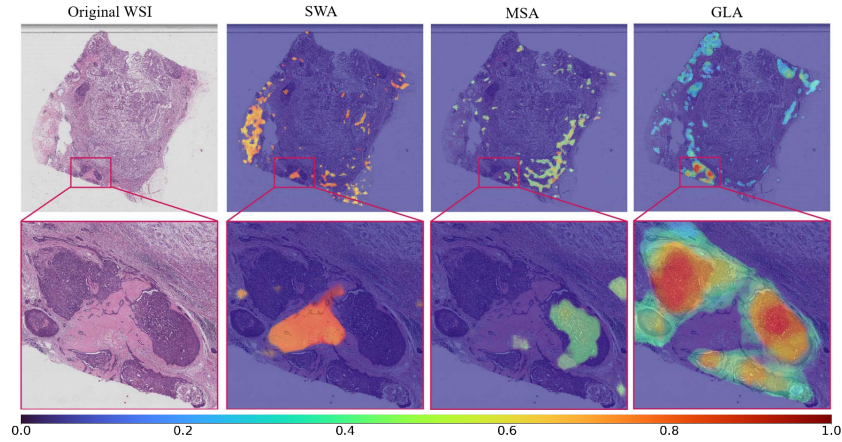

Fig. 2. Visualization of attention score maps for different mechanisms in prostate cancer grading.

shifted window attention (SWA) [18], MSA [8], and GLA—for prostate cancer grading. The first column displays the original WSI. The second, third, and fourth columns illustrate the attention heatmaps generated by SWA, MSA, and GLA, respectively. The magnified views in the second row further highlight the differences in feature localization. SWA captures localized features but lacks a global perspective, resulting in fragmented attention. MSA expands attention to broader regions but does not enforce spatial consistency, sometimes misaligning critical areas. In contrast, GLA integrates graph Laplacian constraints to maintain spatial coherence across histologically similar regions while refining feature embeddings. Expert pathologists confirmed that the highlighted areas in MSA and GLA correspond to cancerous regions, validating the effectiveness of our approach. Unlike SWA and MSA, GLA preserves high-frequency histological structures , such as glandular boundaries and morphological variations, by applying learnable filtering at the feature level. The heatmaps demonstrate that

GLA provides a more structured and precise attention distribution, highlighting the importance of spatial constraints for accurate prostate cancer grading.

4   Conclusions and Future Work

This work presents a novel framework combining an Iterative Refinement Module (IRM) for adaptive patch selection and a Graph Laplacian Attention-Based Transformer (GLAT) for spatially coherent feature learning, achieving state-of-the-art performance in prostate cancer grading. By refining patch selection and enforcing spatial constraints, the model effectively captures histopathological structures while reducing computational overhead. Extensive experiments on public and private datasets validate its effectiveness and generalizability. However, the framework relies on fixed patch sizes, which may not fully capture multi-scale tissue variations. Future work will explore adaptive patch selection and cross-slide attention mechanisms to enhance contextual awareness across distant tumor regions.

References

1. Anklin, V., Pati, P., Jaume, G., Bozorgtabar, B., Foncubierta-Rodriguez, A., Thiran, J.P., Sibony, M., Gabrani, M., Goksel, O.: Learning whole-slide segmentation from inexact and incomplete labels using tissue graphs. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24. pp. 636–646. Springer (2021)
2. Behzadi, M.M., Madani, M., Wang, H., Bai, J., Bhardwaj, A., Tarakanova, A., Yamase, H., Nam, G.H., Nabavi, S.: Weakly-supervised deep learning model for prostate cancer diagnosis and gleason grading of histopathology images. Biomedical Signal Processing and Control **95**, 106351 (2024)
3. Bhattacharyya, R., Das, S.P., Mitra, S.: Efficient self-supervised grading of prostate cancer pathology. arXiv preprint arXiv:2501.15520 (2025)
4. Bian, H., Shao, Z., Chen, Y., Wang, Y., Wang, H., Zhang, J., Zhang, Y.: Multiple instance learning with mixed supervision in gleason grading. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 204–213. Springer (2022)
5. Bontempo, G., Porrello, A., Bolelli, F., Calderara, S., Ficarra, E.: Das-mil: Distilling across scales for mil classification of histological wsis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 248–258. Springer (2023)
6. Chan, T.H., Cendra, F.J., Ma, L., Yin, G., Yu, L.: Histopathology whole slide image analysis with heterogeneous graph representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15661–15670 (2023)

7. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F., Jaume, G., Song, A.H., Chen, B., Zhang, A., Shao, D., Shaban, M., et al.: Towards a general-purpose foundation model for computational pathology. Nature Medicine **30**(3), 850–862 (2024)

8. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)

9. Goldman, M.J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., et al.: Visualizing and interpreting cancer genomics data via the xena platform. Nature biotechnology **38**(6), 675–678 (2020)

10. Gustafsson, F.K., Rantalainen, M.: Evaluating computational pathology foundation models for prostate cancer grading under distribution shifts. arXiv preprint arXiv:2410.06723 (2024)

11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)

12. Iso, H., Wang, X., Suhara, Y., Angelidis, S., Tan, W.C.: Convex aggregation for opinion summarization. arXiv preprint arXiv:2104.01371 (2021)

13. Jiang, S., Hondelink, L., Suriawinata, A.A., Hassanpour, S.: Masked pre-training of transformers for histology image analysis. Journal of Pathology Informatics p. 100386 (2024)

14. Junayed, M.S., Nabavi, S.: A scaled mask attention-based hybrid model for survival prediction. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 2100–2107. IEEE (2024)

15. Koziarski, M., Cyganek, B., Niedziela, P., Olborski, B., Antosz, Z., Żydak, M., Kwolek, B., Wąsowicz, P., Bukała, A., Swadźba, J., et al.: Diagset: a dataset for prostate cancer histopathological image classification. Scientific Reports **14**(1), 6780 (2024)

16. Lin, T., Yu, Z., Hu, H., Xu, Y., Chen, C.W.: Interventional bag multi-instance learning on whole-slide pathological images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19830–19839 (2023)

17. Linkon, A.H.M., Labib, M.M., Hasan, T., Hossain, M., et al.: Deep learning in prostate cancer diagnosis and gleason grading in histopathology images: An extensive study. Informatics in Medicine Unlocked **24**, 100582 (2021)

18. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10012–10022 (2021)

19. 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)

20. Moranguinho, J., Pereira, T., Ramos, B., Morgado, J., Costa, J.L., Oliveira, H.P.: Attention based deep multiple instance learning approach for lung cancer prediction using histopathological images. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 2852–2855. IEEE (2021)

21. Myronenko, A., Xu, Z., Yang, D., Roth, H.R., Xu, D.: Accounting for dependencies in deep learning based multiple instance learning for whole slide imaging. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 329–338. Springer (2021)

22. Nir, G., Hor, S., Karimi, D., Fazli, L., Skinnider, B.F., Tavassoli, P., Turbin, D., Villamil, C.F., Wang, G., Wilson, R.S., et al.: Automatic grading of prostate cancer in digitized histopathology images: Learning from multiple experts. Medical image analysis **50**, 167–180 (2018)
23. Pati, P., Jaume, G., Ayadi, Z., Thandiackal, K., Bozorgtabar, B., Gabrani, M., Goksel, O.: Weakly supervised joint whole-slide segmentation and classification in prostate cancer. Medical Image Analysis **89**, 102915 (2023)
24. 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**, 2136–2147 (2021)
25. Silva-Rodríguez, J., Colomer, A., Sales, M.A., Molina, R., Naranjo, V.: Going deeper through the gleason scoring scale: An automatic end-to-end system for histology prostate grading and cribriform pattern detection. Computer methods and programs in biomedicine **195**, 105637 (2020)
26. Talaat, F.M., El-Sappagh, S., Alnowaiser, K., Hassan, E.: Improved prostate cancer diagnosis using a modified resnet50-based deep learning architecture. BMC Medical Informatics and Decision Making **24**(1), 23 (2024)