Insights from GNN and Cellular Explanations: A Multistage Framework to Interpret Automatic Diagnosis in Histopathology

Catarina Palmeirão[1] CATARINA.PALMEIRAO@TECNICO.ULISBOA.PT

Rita Pereira[1] RITA.MARTINS.PEREIRA@TECNICO.ULISBOA.PT

Maria Rita Verdelho[1] RITAVERDELHO@HOTMAIL.COM

Catarina Barata[1] ANA.C.FIDALGO.BARATA@TECNICO.ULISBOA.PT

Carlos Santiago[1] CARLOS.SANTIAGO@TECNICO.ULISBOA.PT

[1] *Institute for Systems and Robotics, LARSyS, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal*

Editors: Under Review for MIDL 2026

Abstract

Whole Slide Imaging has transformed digital pathology by capturing tissue architecture at cellular resolution, yet its gigapixel scale and complex spatial organization challenge automated analysis. Traditional deep learning methods often overlook these spatial dependencies, limiting their diagnostic reliability. Graph-based learning addresses this limitation by representing tissue as interconnected cellular and structural entities, preserving spatial and morphological context essential for accurate cancer diagnosis. This paper investigates how Graph Neural Networks and Explainable Artificial Intelligence can jointly enhance the performance and interpretability of histopathological diagnosis. By modeling both cellular and tissue level relationships, GNN capture biologically meaningful patterns, while methods such as GNNExplainer reveal the rationale behind predictions. The integration with HoVer-Net further enables multiscale interpretability, reflecting the hierarchical reasoning process of pathologists. Extensive experiments show that attention-based GNN architectures outperform standard graph convolutional models while remaining efficient and interpretable. Beyond accuracy, the results demonstrates that graph-based learning combined with XAI provides a robust, biologically grounded foundation for reliable diagnostic systems in computational pathology.

Keywords: Whole Slide Imaging, Digital Pathology, Graph Neural Networks, Explainable Artificial Intelligence, GNNExplainer, HoVer-Net, Multiscale Interpretability

1. Introduction

Cancer is a large and complex group of diseases characterized by the abnormal and uncontrolled proliferation of cells. These cancerous cells are morphologically characterized by their irregular size and shape, enlarged nuclei, and limited cytoplasmic content, often resulting in intense staining (Baba and Câtoi, 2007). When cancer spreads beyond its origin boundaries, known as metastasis, it becomes the leading cause of cancer-related mortality, with the World Health Organization (WHO) projecting approximately 16.3 million deaths by 2040 (for Research on Cancer, 2024).

Traditional diagnosis relies on histopathology, where tissue samples are examined under the microscope (Cotta et al., 2021). Although effective, this approach is labor-intensive and subject to inter-pathologist variability. The advent of digital pathology has revolutionized this workflow by digitizing slides into high-resolution images, the Whole Slide Images (WSIs), enabling computational analysis and remote collaboration (Kanwal et al., 2022). However, its gigapixel size and heterogeneity turns manual assessment impractical, motivating the adoption of Machine Learning (ML) and Deep Learning (DL) for automated detection and classification (Oliveira et al., 2021).

Despite their success, DL models are frequently perceived as "black boxes," hindering their application in clinical practice where interpretability and reliability are critical (Teng et al., 2022). Explainable Artificial Intelligence (XAI) has emerged to address this limitation by providing transparent and interpretable model outputs. In pathology, XAI techniques highlight image regions or features that contribute to predictions, enabling clinicians to validate AI findings and identify potential biases (Plass et al., 2023; Abhishek and Kamath, 2022).

A major challenge in WSI analysis lies in capturing spatial context. Conventional patch-based methods and Multiple Instance Learning (MIL) frameworks typically process image tiles independently, overlooking spatial correlations that pathologists instinctively consider. Graph Neural Networks (GNNs) have recently been introduced to address this limitation by modeling patches as graph nodes and their spatial or morphological relationships as edges, thereby capturing both local and global structural dependencies (Mirabadi et al., 2024; Pereira et al., 2025) . However, their decision-making process remains largely opaque.

Integrating XAI with GNNs enables visualization of influential regions and relationships that guide model predictions, enhancing interpretability. The concept of causality, which measures the clarity and clinical relevance of explanations, ensures alignment between AI reasoning and medical decision-making (Plass et al., 2023). This mirrors pathologists' diagnostic reasoning, which combines a global tissue overview with detailed cellular assessment.

The objective of this paper is to investigate how XAI can improve the interpretability and diagnostic performance of GNN-based models in digital pathology. By identifying the most informative nodes and regions, XAI can provide insights into the model's decision-making process that contribute to diagnostic predictions. Specifically, this work aims to:

- Explore graph-based representations for modeling spatial and morphological structures essential for diagnosis
- Evaluate XAI's contribution to enhancing GNN diagnostic accuracy and transparency;

- Examine whether explainability replicates the hierarchical reasoning applied by pathologists, from tissue level patterns to cell based features.

2. Related Work

MIL is a weakly supervised strategy widely used in computational pathology when pixel-level annotations are unavailable. WSIs are divided into small portions of the original WSI, patches, forming a bag that receives a slide-level label, assuming positivity if at least one instance is positive (Brand et al., 2024; Waqas et al., 2024). Traditional MIL methods use pooling operations, which either treat all patches equally or prioritize only the most discriminative ones (Liu et al., 2024). Attention-based MIL (ABMIL) (Abhishek and Kamath, 2022) improves both interpretability and performance by learning instance-specific relevance weights. Extensions to MIL incorporate hierarchical or multi-scale information (Shi et al., 2024), enabling joint modeling of local and global features. However, MIL inherently treats instances independently, ignoring spatial context, an essential component of histopathological reasoning. To address this limitation, recent research has shifted towards graph-based learning, which explicitly models spatial and morphological relationships between tissue components (Mirabadi et al., 2024; Lu et al., 2020).

Graph-based approaches represent WSIs as graphs, where nodes correspond to biological entities (e.g., cells, patches, or tissue regions) and edges capture spatial or morphological interactions (Jaume et al., 2021). Multiple paradigms exist: cell-graphs focus on fine-grained cellular relationships, tissue-graphs capture broader structural organization, and patch-graphs provide a balance between detail and computational efficiency (Brussee et al., 2025). GNNs, such as Graph Convolutional Networks (GCNs) (Kipf, 2016), Graph Attention Networks (GATs) (Veličković et al., 2017), and GraphSAGE (Hamilton et al., 2017), enable relational learning through message passing. While GCNs aggregate neighborhood information uniformly, GATs assign attention weights to emphasize more influential neighbors, offering improved expressiveness and interpretability. Attention-based GNNs are particularly effective in modeling biologically meaningful relationships in histopathology (Brussee et al., 2025).

Despite their success, GNNs remain largely opaque, which difficult into clinical adoption. XAI methods aim to increase transparency by highlighting influential substructures or features guiding predictions. GNNExplainer (Ying et al., 2019) identifies relevant nodes and edges via soft masking, while CF-GNNExplainer (Lucic et al., 2022) extends this with counterfactual reasoning. More recently, ZORRO (Funke et al., 2022) improves explanation stability and sparsity. These frameworks align with the hierarchical reasoning process used by pathologists, who integrate global tissue patterns with detailed cellular examination. Therefore, combining GNNs with XAI represents a crucial step toward interpretable, trustworthy, and clinically applicable computational pathology systems.

3. Proposed Approach

This paper aims to explore graph representation, evaluate the impact of XAI on GNN transparency and assess whether explainability mirrors hierarchical reasoning from tissue to cell level. To achieve this, we propose a hierarchical pipeline, in Figure 1 that combines

graph learning, explainability and traditional ML. Together, they form a hierarchical and interpretable framework for Whole Slide Image (WSI) analysis, enabling both accurate classification and meaningful explanation of diagnostic decisions.

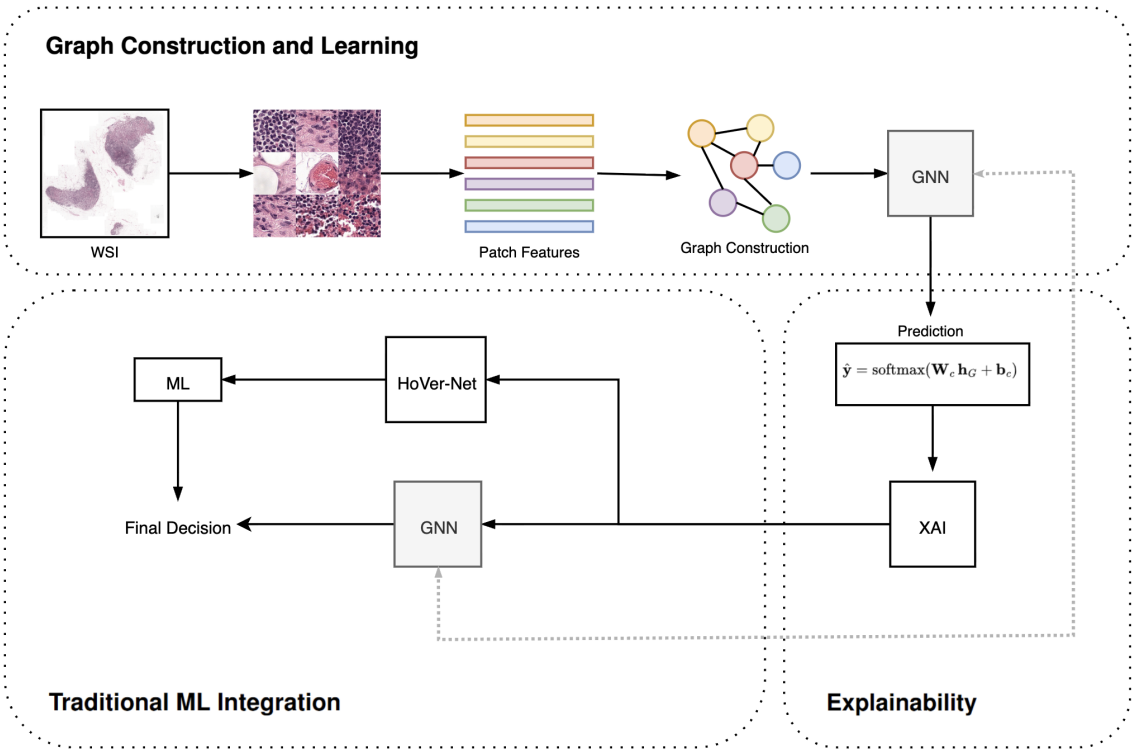

Figure 1: The pipeline begins with WSI processing, patch extraction, and feature embedding via a pretrained CNN. These patches are modeled as nodes in a graph, with spatial adjacency defining edges. GNN models perform classification by integrating local cellular features with global tissue context. Post-hoc explainability via GNNExplainer identifies key subgraphs and features influencing predictions. Finally, nuclei features from the highlighted regions, extracted using HoVer-Net, are evaluated with traditional ML classifiers to validate the biological and clinical relevance of the explanations.

Each WSI is first segmented using the CLAM toolbox (Lu et al., 2021)to isolate tissue regions and then divided into non-overlapping 256×256 patches. Patch-level feature extraction is performed using a pretrained CNN, ResNet-50, generating 1024-dimensional embeddings for each patch. These patches are represented as nodes in a graph, an approach described in (Chen et al., 2021), with edges connecting spatially adjacent patches (distance ≤ 256), capturing the local tissue architecture.

The resulting patch-graph is processed by a GNN, which performs neighborhood aggregation to produce node embeddings. A global pooling operation summarizes these embeddings for graph-level classification. Both GCN and GAT are evaluated: GCN aggregates

normalized neighbor features, while GAT learns attention-based weights to model the importance of each neighbor adaptively.

Post-hoc explainability is implemented using GNNExplainer, which identifies the subgraphs and node features that most strongly influence model predictions. The regions highlighted by the explanations are further analyzed by extracting nuclei-level features using HoVer-Net (Graham et al., 2019). These features are aggregated into histogram-based signatures and used as input to classical ML models, such as Support Vector Machines, Random Forest, and Gradient Boosting classifiers, to assess the biological relevance of the explanations.

By integrating patch-level morphology, spatial context, model explainability, and cellular composition, this multistage pipeline provides a robust and interpretable framework for histopathology image classification, supporting both accurate diagnosis and insight into the underlying tissue characteristics.

4. Experimental Setup

The experimental evaluation was performed using two publicly breast cancer datasets: Camelyon16 (Bandi, 2016), focused on lymph node metastasis detection, and BRACS (Brancati et al., 2022), which includes seven lesion types collected from different patients. Both datasets contain annotated regions of interest and slides were scanned at high resolution.

To address the class imbalance and clinical relevance, model performance was assessed using Balanced Accuracy, AUC and F1 Score ensuring sensitivity to both majority and minority classes.

The WSIs were processed using the CLAM framework, which segments tissue regions, removes background areas, and extracts non-overlapping patches. Each patch was converted into a compact feature embedding using a pretrained feature extractor and stored alongside its spatial coordinates. A 5 fold cross-validation setup was employed and feature normalization was applied using statistics (mean and standard deviation) computed from the training data in each fold.

For classification, GCN and GAT models were trained for graph-level prediction, using up to three layers with 256 hidden units, ReLU activation, and 0.25 dropout. Models were optimized with Adam (learning rate 1×10^{-3}, weight decay 1×10^{-4}), trained for 100 epochs, with batch size 9, and evaluated using 5 fold cross-validation. Graph pooling strategies (max, mean, add) were compared.

Post-hoc interpretability was implemented using GNNExplainer, configured with 100 optimization epochs, attribute-based node masking, and object-based edge masking. Explanations were generated per graph using the predicted class as target, and ranked nodes were compared to expert-annotated regions for qualitative and quantitative validation. All experiments were implemented in Python using PyTorch, PyTorch Geometric, and scikit-learn, and executed on an NVIDIA GeForce RTX 4080 Super GPU with 16 GB memory.

5. Experimental Results

This section presents the experimental results and corresponding analyses for the proposed graph-based diagnostic framework. Two histopathology datasets were used to evaluate

model performance and interpretability at different diagnostic levels: (1) a binary classification dataset, and (2) the multiclass dataset, which includes seven lesion categories ranging from benign to malignant. For both datasets, two GNN architectures were implemented, GCN and GAT. Each model was evaluated in terms of classification accuracy, AUC, F1-score, and explainability through GNNExplainer.

5.1. Binary Classification Dataset

5.1.1. Graph Construction and Learning

Table 1: Performance comparison of GCN and GAT models with max pooling for graph classification.

Model	Layers	Pooling	Balanced Acc.	AUC	F1-score
GCN	3	Max	0.903 ± 0.048	0.933 ± 0.052	0.885 ± 0.060
GAT	3	Max	0.950 ± 0.034	0.958 ± 0.043	0.942 ± 0.040

The results in Table 1 demonstrates the advantages of attention mechanisms in GNNs. While GCNs rely on fixed weights for aggregating information from neighboring nodes, GATs dynamically learn the importance of each neighbor through an attention mechanism. This allows GATs to prioritize the most relevant information, which is particularly beneficial in the context of histopathological analysis, where certain regions or features of the tissue may carry more diagnostic significance than others. The choice of analyzing only 2 combinations of GNN model, architecture and pooling strategy was chosen based on the recent literature, (Pereira et al., 2025), where the best results on balanced accuracy were achieved on a 3 layer GAT model with max pooling.

5.1.2. Explainability Analysis

To understand he decision-making processes of the proposed GNN architectures, GNNExplainer was applied to identify the substructures and node features most influential to model predictions. Both GCN and GAT models were trained with three layers and max pooling, and the explainer was evaluated using three different explanation targets: (i) explanations for the model's prediction, (ii) explanations assuming the graph belongs to the Tumor class, and (iii) explanations assuming the graph belongs to the Normal class. By comparing explanations across these targets, it was possible to observe how each model distinguishes between tumor and normal tissue patterns and which regions most strongly influence these decisions.

Table 2 presents the mean precision and recall scores that quantify the quality of the regions highlighted by GNNExplainer, obtained by comparing the identified nodes with the ground-truth annotated regions of interest. Although the precision values were relatively low, indicating that some highlighted nodes may not be essential for predictions, the recall values were higher, particularly for GAT, suggesting that the attention mechanism captures a broader set of potentially relevant nodes. These findings are consistent with previous

studies showing that attention-based models, such as GAT, generate more comprehensive explanations that incorporate contextual information, while GCNs tend to produce more localized explanations focused on compact neighborhoods.

Despite modest precision, the results provide meaningful insights into model behavior. The lower precision likely reflects the inherent complexity of histopathological data, where relevant features are diffuse and overlapping rather than sharply defined. Additionally, discrepancies between explainer outputs and expert annotations may arise because annotations typically mark only the most prominent diagnostic regions, while other informative yet subtle structures remain unannotated.

Overall, the higher recall observed for the GAT model indicates that attention mechanisms are more effective at capturing a wider range of subgraph regions relevant to classification. Applying GNNExplainer, in this context, provides valuable insights into how graph-based models identify diagnostically meaningful tissue patterns in histopathological images.

Table 2: Mean precision, recall for GCN and GAT models with max pooling under different explanation targets.

Model	**Layers**	**Pooling**	**Target**	**Precision**	**Recall**
GCN	3	Max	Prediction	0.155 ± 0.025	0.314 ± 0.063
			1-Tumor	0.154 ± 0.026	0.305 ± 0.069
			0-Normal	0.156 ± 0.024	0.306 ± 0.073
GAT	3	Max	Prediction	0.139 ± 0.041	0.371 ± 0.052
			1-Tumor	0.138 ± 0.039	0.375 ± 0.051
			0-Normal	0.140 ± 0.043	0.389 ± 0.044

5.1.3. Traditional ML Integration

To further analyze and interpret the graph-based models, two complementary experiments were conducted: Subgraph Masking and Cell Nuclei based Classification. The Subgraph Masking approach assessed how the GNNs' predictions depended on the most informative substructures identified by GNNExplainer. By selectively retaining or removing these subgraphs, it was possible to evaluate their contribution to overall classification accuracy. Results showed that both GCN and GAT maintained high performance when using only explainer-identified nodes, confirming that these subgraphs captured the most discriminative regions. When medical annotations were integrated, the GCN experienced a drop in performance, suggesting reliance on broader contextual information, whereas the GAT remained stable, demonstrating greater robustness and focus on essential discriminative features. Expanding the annotation regions through morphological dilation partially recovered GCN performance, emphasizing the importance of surrounding tissue context.

In the Cell Nuclei based Classification analysis, nodes highlighted by GNNExplainer were used as input features for traditional ML classifiers, including Random Forest, SVM,

and Gradient Boosting. These experiments evaluated whether the explanatory nodes retained sufficient information for accurate cell based classification, using cell distribution histograms as features. Overall, moderate accuracies (0.53–0.61) were achieved, with GAT nodes slightly outperforming GCN ones, particularly when using Random Forest and Gradient Boosting. These results indicate a biological correlation between the cell distribution histograms from the explainability nodes and the original WSI label.

5.2. BRACS Multiclass dataset

5.2.1. Graph Construction and Learning

To evaluate the effectiveness of different GNN configurations for multiclass classification on the BRACS dataset, we compared GCN and GAT models across 1–3 layers and three pooling strategies (max, mean, add). Results in Table 3 shows that overall, the GAT with two layers and max pooling achieved the best performance, reaching a balanced accuracy of 0.487 ± 0.023, an AUC of 0.793 ± 0.021, an F1-macro of 0.473 ± 0.027, and an F1-weighted of 0.536 ± 0.032. Performance across classes varied, with lesion types such as Invasive Carcinoma (IC), Ductal Carcinoma in Situ (DCIS), and Pathological benign (PB) showing the highest F1 scores due to their distinctive morphological features. Increasing model depth generally led to oversmoothing, reducing classification performance, as observed in both 3 layer GCN and GAT models. These results, confirm that attention mechanisms enhance feature discrimination and robustness, making them the most effective architecture for multiclass tissue classification in the BRACS dataset.

Table 3: Performance comparison of GCN and GAT models with different pooling strategies grouped by layer.

Model	Layers	Pooling	Balanced Acc.	AUC	F1 Macro	F1 Weighted
GCN	1	Max	0.476 ± 0.044	0.809 ± 0.018	0.447 ± 0.039	0.515 ± 0.027
	2	Max	0.476 ± 0.042	0.796 ± 0.023	0.440 ± 0.030	0.498 ± 0.040
	3	Max	0.448 ± 0.031	0.768 ± 0.018	0.414 ± 0.031	0.479 ± 0.035
GAT	1	Max	0.487 ± 0.042	0.806 ± 0.019	0.461 ± 0.048	0.521 ± 0.053
	2	Max	$\mathbf{0.487 \pm 0.023}$	$\mathbf{0.793 \pm 0.021}$	$\mathbf{0.473 \pm 0.027}$	$\mathbf{0.536 \pm 0.032}$
	3	Max	0.477 ± 0.036	0.799 ± 0.016	0.431 ± 0.037	0.494 ± 0.032

5.2.2. Explainability

The explainability analysis of the multiclass GNN models focused on qualitatively and quantitatively evaluating how the networks identify histologically meaningful regions in the BRACS dataset. Figure 2 shows a qualitative analysis of the explanation in BRACS.

5.2.3. Traditional ML Integration

Through Subgraph Masking, it was shown that the explanatory subgraphs alone preserve most of the discriminative information needed for accurate predictions, as the 2 layer GAT

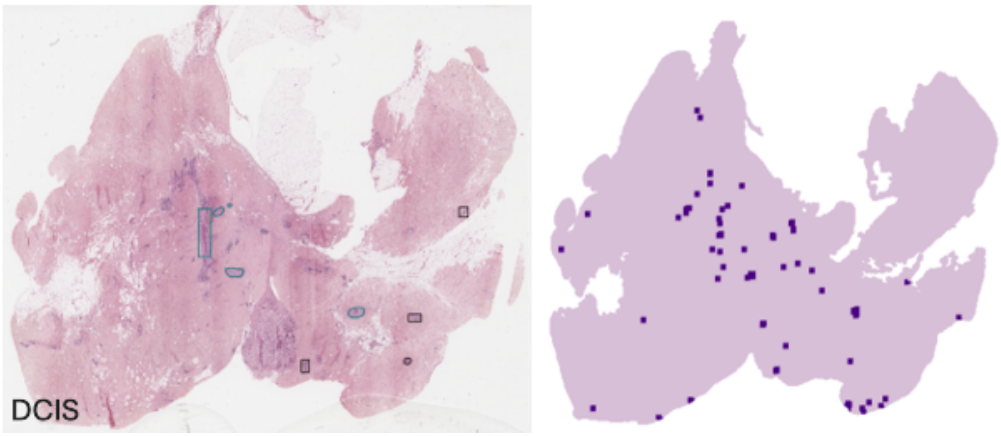

Figure 2: Qualitative explainability analysis using GNNExplainer. The left column shows QuPath annotations for DCIS lesion type, while the right column displays the corresponding graph representations. Dark purple nodes denote those identified as most influential by GNNExplainer.

maintained strong performance in Table 4. This indicates that GNNExplainer effectively isolates the most influential regions in the graph structure.

Finally, Cell Nuclei based Classification linked the GNN subgraphs with HoVer-Net cellular segmentation, exemplified Figure 3, to assess biological interpretability. While the nuclei compositions in explainer identified regions mirrored pathologist annotations, traditional classifiers trained only on nuclei classification described in Table 5 achieved low performance, confirming that single cell features lack sufficient discriminative power.

Table 4: Performance of the best GNN model (2 layer GAT) under the Subgraph Masking experiment using GNNExplainer-derived subgraphs.

Model	**Layers**	**Pooling**	**Target**	**Balanced Acc.**	**AUC**	**F1-score**
GAT	2	Max	Prediction	0.479 ± 0.034	0.793 ± 0.024	0.531 ± 0.042

Table 5: Classification results of cell nuclei using histogram features extracted from HoVer-Net. Only HoVer-Net derived histogram features were used as input.

GNN Model	**XAI Method**	**Target**	**ML Method**	**Accuracy**	**Balanced Acc.**
GAT	GNNExplainer	Prediction	Random Forest	0.289 ± 0.049	0.267 ± 0.043
			SVM	0.310 ± 0.047	0.275 ± 0.032
			Gradient Boosting	0.371 ± 0.014	0.250 ± 0.013

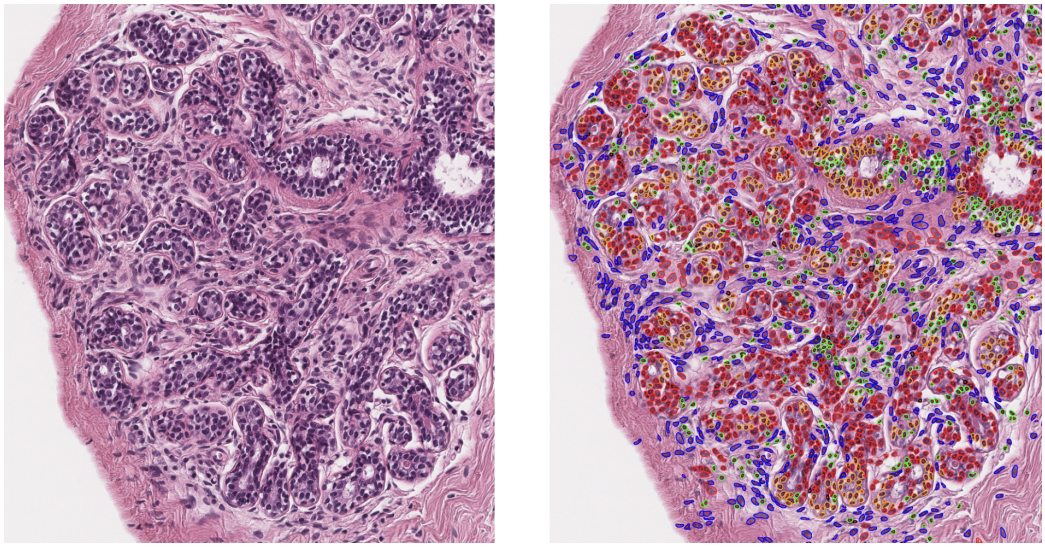

Figure 3: Comparison of ROI for patient 745 with PB. Left: original ROI image. Right: corresponding processed ROI with HoVer-Net overlay segmentation

6. Conclusion

This research work aimed to investigate how cellular and tissue-level insights provided by XAI can enhance both the performance and interpretability of GNN-based diagnostic models. The study addressed four primary objectives: evaluating the impact of XAI insights on GNN performance, assessing whether explainability methods can replicate the hierarchical reasoning of pathologists, exploring how XAI can bridge automated predictions with clinical decision-making, and examining the potential of graph-based representations for capturing critical spatial and morphological relationships in histopathological data. The results demonstrated the effectiveness of the proposed framework across binary and multiclass classification tasks, highlighting the ability of GNNs to model complex tissue structures while providing interpretable insights. Notably, GAT consistently outperformed GCN by leveraging attention mechanisms to focus on the most relevant graph features, which is particularly valuable in histopathology where spatial context and cell-tissue relationships are crucial for accurate diagnosis. The integration of GNNExplainer and HoVer-Net enabled a multiscale analysis, linking tissue-level patterns to cellular-level features, with the identified subgraphs aligning closely with expert annotations and established histopathological knowledge. While GNNs effectively captured higher-order tissue context, traditional ML approaches based on individual nuclear morphology proved insufficient, emphasizing the importance of graph-based modeling for leveraging spatial and contextual information. Overall, this work demonstrates that GNN-based models, when combined with XAI, offer a robust, interpretable, and clinically relevant framework for histopathological diagnosis, providing both competitive performance and biologically meaningful explanations.

References

Kumar Abhishek and Deeksha Kamath. Attribution-based xai methods in computer vision: A review. *arXiv preprint arXiv:2211.14736*, 2022.

Alina I. Baba and Cornel Câtoi. The tumor cell. In *Comparative Oncology*, chapter 3.1. The Publishing House of the Romanian Academy, Bucharest, Romania, 2007.

Peter Bandi. Camelyon16 dataset. *Grand Challenge Dataset*, 2016. URL https://camelyon16.grand-challenge.org/data.

Nadia Brancati, Anna Maria Anniciello, Pushpak Pati, Daniel Riccio, Giosuè Scognamiglio, Guillaume Jaume, Giuseppe De Pietro, Maurizio Di Bonito, Antonio Foncubierta, Gerardo Botti, et al. Bracs: A dataset for breast carcinoma subtyping in h&e histology images. *Database*, 2022:baac093, 2022.

Lodewijk Brand, Hoon Seo, Lauren Zoe Baker, Carla Ellefsen, Jackson Sargent, and Hua Wang. A linear primal–dual multi-instance svm for big data classifications. *Knowledge and Information Systems*, 66(1):307–338, 2024.

Siemen Brussee, Giorgio Buzzanca, Anne MR Schrader, and Jesper Kers. Graph neural networks in histopathology: Emerging trends and future directions. *Medical Image Analysis*, page 103444, 2025.

Richard J Chen, Ming Y Lu, Muhammad Shaban, Chengkuan Chen, Tiffany Y Chen, Drew FK Williamson, and Faisal Mahmood. Whole slide images are 2d point clouds: Context-aware survival prediction using patch-based graph convolutional networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 339–349. Springer, 2021.

Ana Cotta, Elmano Carvalho, Antonio Lopes da Cunha-Júnior, Jaquelin Valicek, Monica M Navarro, Sidney Baptista Junior, Eni Braga da Silveira, Maria Isabel Lima, Bruno Arrivabene Cordeiro, Alexandre Faleiros Cauhi, et al. Muscle biopsy essential diagnostic advice for pathologists. *Surgical and Experimental Pathology*, 4(1):3, 2021.

International Agency for Research on Cancer. Putting an end to cancer before it can begin. Brochure, 2024. 12-page brochure.

Thorben Funke, Megha Khosla, Mandeep Rathee, and Avishek Anand. Zorro: Valid, sparse, and stable explanations in graph neural networks. *IEEE Transactions on Knowledge and Data Engineering*, 35(8):8687–8698, 2022.

Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical image analysis*, 58:101563, 2019.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

Guillaume Jaume, Pushpak Pati, Behzad Bozorgtabar, Antonio Foncubierta, Anna Maria Anniciello, Florinda Feroce, Tilman Rau, Jean-Philippe Thiran, Maria Gabrani, and Orcun Goksel. Quantifying explainers of graph neural networks in computational pathology. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8106–8116, 2021.

Neel Kanwal, Fernando Pérez Bueno, Arne Schmidt, Rafael Molina Soriano, et al. The devil is in the details: Whole slide image acquisition and processing for artifacts detection, color variation, and data augmentation: A review. 2022.

TN Kipf. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Xin Liu, Weijia Zhang, and Min-Ling Zhang. Position: From correlation to causation: Max-pooling-based multi-instance learning leads to more robust whole slide image classification. *arXiv preprint arXiv:2408.09449*, 2024.

Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Richard J Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, 5(6):555–570, 2021.

Wenqi Lu, Simon Graham, Mohsin Bilal, Nasir Rajpoot, and Fayyaz Minhas. Capturing cellular topology in multi-gigapixel pathology images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 260–261, 2020.

Ana Lucic, Maartje A Ter Hoeve, Gabriele Tolomei, Maarten De Rijke, and Fabrizio Silvestri. Cf-gnnexplainer: Counterfactual explanations for graph neural networks. In *International Conference on Artificial Intelligence and Statistics*, pages 4499–4511. PMLR, 2022.

Ali Khajegili Mirabadi, Graham Archibald, Amirali Darbandsari, Alberto Contreras-Sanz, Ramin Ebrahim Nakhli, Maryam Asadi, Allen Zhang, C Blake Gilks, Peter Black, Gang Wang, et al. Grasp: graph-structured pyramidal whole slide image representation. *arXiv preprint arXiv:2402.03592*, 2024.

Sara P Oliveira, Pedro C Neto, Joao Fraga, Diana Montezuma, Ana Monteiro, João Monteiro, Liliana Ribeiro, Sofia Gonçalves, Isabel M Pinto, and Jaime S Cardoso. Cad systems for colorectal cancer from wsi are still not ready for clinical acceptance. *Scientific Reports*, 11(1):14358, 2021.

Rita Pereira, M. Rita Verdelho, Catarina Barata, and Carlos Santiago. The role of graph-based mil and interventional training in the generalization of wsi classifiers, 2025. URL https://arxiv.org/abs/2501.19048.

Markus Plass, Michaela Kargl, Tim-Rasmus Kiehl, Peter Regitnig, Christian Geißler, Theodore Evans, Norman Zerbe, Rita Carvalho, Andreas Holzinger, and Heimo Müller. Explainability and causability in digital pathology. *The Journal of Pathology: Clinical Research*, 9(4):251–260, 2023.

Zhan Shi, Jingwei Zhang, Jun Kong, and Fusheng Wang. Integrative graph-transformer framework for histopathology whole slide image representation and classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 341–350. Springer, 2024.

Qiaoying Teng, Zhe Liu, Yuqing Song, Kai Han, and Yang Lu. A survey on the interpretability of deep learning in medical diagnosis. *Multimedia Systems*, 28(6):2335–2355, 2022.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Muhammad Waqas, Syed Umaid Ahmed, Muhammad Atif Tahir, Jia Wu, and Rizwan Qureshi. Exploring multiple instance learning (mil): A brief survey. *Expert Systems with Applications*, 250:123893, 2024.

Zhitao Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. Gnnexplainer: Generating explanations for graph neural networks. *Advances in neural information processing systems*, 32, 2019.