PeerJ Computer Science

MuGNet: a graph-based framework for multi-tissue integration in computational pathology

Francesca Miccolis[1], Martina Demarco[1], Elisa Ficarra[1] and Marta Lovino[2]

[1] Department of Engineering Enzo Ferrari, University of Modena, Modena, Italy
[2] Life Science Department, University of Modena, Modena, Italy

ABSTRACT

Computational pathology has seen significant advancements through the application of deep learning to Whole-Slide Images (WSIs). However, most existing models operate under a mono-tissue assumption, using only a single WSI to inform the prediction. This limitation is primarily driven by the structure of publicly available datasets, which rarely provide more than one slide per patient. Consequently, multi-tissue data integration remains underexplored in computational pathology, not due to a lack of clinical relevance but rather to a shortage of both public datasets and modeling strategies capable of handling multi-sample input. To address this gap, this study introduces MuGNet, a novel graph-based deep learning framework that enables patient-level prediction from a variable number of WSIs. The MuGNet framework represents each patient as a network of tissue samples, with each WSI encoded as a node in a patient-specific graph. Edges are established to capture structural or biological relationships between tissues. A Graph Neural Network (GNN) with an attention-based message passing mechanism is then employed to process the patient graphs and generate patient-level predictions. MuGNet was evaluated on a private High Grade Serous Carcinoma (HGSC) dataset across three distinct prediction tasks to assess its robustness: one classification task (Homologous Recombination Deficiency status prediction, HRD) and two regression tasks (Overall Survival and Platinum Free Interval prediction, OS and PFI). MuGNet's performance was compared against three state-of-the-art Multiple Instance Learning (MIL) models: Attention-Based MIL (ABMIL), Dual-Stream MIL (DSMIL), and Transformer-based MIL (TransMIL). For the HRD status prediction task, MuGNet achieved an AUC of 0.7220 ± 0.0739, approximately 6% higher than the most performant MIL model. For the OS and PFI tasks, the model demonstrated an improvement in the C-index of approximately 10% over ABMIL and DSMIL and up to 2% over TransMIL, highlighting its superior performance and robustness in handling multi-sample input. Beyond its predictive accuracy, MuGNet provides interpretability. The attention weights learned across the graph nodes allow for the identification of tissue sites that contribute most to the prediction, revealing that the model consistently prioritizes metastasis-rich tissue regions such as the lymph nodes.

Subjects Bioinformatics, Computational Biology, Artificial Intelligence, Computer Vision, Data Mining and Machine Learning

Keywords Computational pathology, Whole-slide image analysis, Graph-neural networks, Deep learning, Multi-tissue integration, Attention mechanisms, Multi instance learning

Submitted 27 October 2025
Accepted 13 March 2026
Published 15 June 2026

Corresponding author
Marta Lovino,
marta.lovino@unimore.it

Academic editor
Paulo Jorge Coelho

Additional Information and Declarations can be found on page 13

DOI **10.7717/peerj-cs.3820**

Copyright
2026 Miccolis et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

How to cite this article Miccolis F, Demarco M, Ficarra E, Lovino M. 2026. MuGNet: a graph-based framework for multi-tissue integration in computational pathology. **PeerJ Comput. Sci. 12:e3820 DOI 10.7717/peerj-cs.3820**

INTRODUCTION

Computational pathology has emerged as a transformative field, leveraging deep learning to analyze gigapixel whole-slide images (WSIs) of histopathology specimens. The application of Convolutional Neural Networks (CNNs) and Transformer-based architectures has led to significant progress in tasks such as tumor detection, grading, and patient-level prognostication (*Shao et al., 2021*; *Li, Li & Eliceiri, 2021*; *Ilse, Tomczak & Welling, 2018*; *Lu et al., 2021*). These advancements have demonstrated the potential of artificial intelligence to assist pathologists and clinicians by providing quantitative and reproducible insights from complex morphological data.

However, a fundamental limitation in many existing computational pathology models is the reliance on a "mono-tissue assumption," where predictions are derived from a single WSI per patient. This constraint is primarily driven by the structure of publicly available datasets, which rarely provide more than one slide per patient. Consequently, multi-tissue data integration remains an underexplored area in computational pathology, not due to a lack of clinical relevance but rather to a shortage of both public datasets and modeling strategies capable of handling multi-sample input. This gap is particularly critical for cancers with pronounced spatial and morphological heterogeneity, such as High-Grade Serous Carcinoma (HGSC), where the metastatic process involves multiple tissue sites. In the context of HGSC, accurate prognostication is crucial for guiding treatment decisions and requires the prediction of key clinical outcomes, such as chemotherapy response and survival endpoints (*Liu et al., 2025*). These include the Homologous Recombination Deficiency (HRD) status (*Konstantinopoulos et al., 2015*), a critical biomarker that indicates a patient's likelihood of responding to platinum-based chemotherapy and PARP inhibitor therapies (*Mirza et al., 2016*); the Overall Survival (OS), defined as the time from diagnosis to death from any cause; and the Progression-Free Interval (PFI), which measures the time from diagnosis until the disease progresses or recurs (*Rustin et al., 2011*). The ability to accurately predict these outcomes from histopathology data is therefore of immense clinical importance.

To date, the most common approach for patient-level prediction from WSIs is Multiple Instance Learning (MIL), which is currently considered the state-of-the-art for this task (*Wen, Chen & Xu, 2025*; *Xu et al., 2025*; *Wang, Li & Zhang, 2024*). In the traditional MIL paradigm, a single WSI is treated as a "bag" of instances (*e.g.*, image patches extracted from the slide). The model learns to aggregate the features of these instances to predict the outcome for the entire bag. Several MIL models have been proposed, each with a distinct aggregation mechanism. Attention-Based MIL (AB-MIL) (*Ilse, Tomczak & Welling, 2018*) uses a learned attention mechanism to assign a weight to each instance (patch), allowing the model to focus on the most discriminative regions of the slide. Dual-Stream MIL (DSMIL) (*Li, Li & Eliceiri, 2021*) operates on a different principle, projecting the instances into a lower-dimensional subspace and then using a form of max-pooling to select the most relevant features. Transformer-based MIL (TransMIL) (*Shao et al., 2021*), a more recent and powerful model, employs a transformer-based encoder to learn the relationships between instances within the bag, using a self-attention mechanism to

capture complex dependencies and produce a robust bag-level representation. These models are designed for single-slide analysis but can be adapted to a multi-tissue setting by treating each whole-slide embedding as an instance in a single patient-level bag, which serves as a strong baseline for comparison.

To address this challenge, this study introduces MuGNet, a novel graph-based deep learning framework that enables patient-level prediction from a variable number of tissue slides. In the MuGNet framework, each WSI is represented as a node in a patient-specific graph, with edges established to capture structural or biological relationships between tissues. A Graph Neural Network (GNN) with an attention-based message passing mechanism is then employed to process these graphs and generate patient-level predictions. This approach moves beyond the conventional "one slide per patient" paradigm, providing a more holistic and clinically relevant view of the patient's disease.

The main contributions of this work are as follows:

- A novel GNN-based framework, MuGNet, is proposed for patient-level prognosis by integrating an arbitrary number of WSIs.
- The framework is validated on a unique multi-tissue dataset of HGSC, demonstrating its robustness on three distinct prediction tasks: one classification (HRD status) and two regression (OS and PFI) tasks.
- MuGNet is compared against three state-of-the-art MIL models (ABMIL, DSMIL, and TransMIL), consistently outperforming them and highlighting the superiority of the graph-based approach for multi-sample integration.
- The model's interpretability is demonstrated through the analysis of attention weights. This reveals that MuGNet learns to prioritize clinically relevant tissue sites, such as the lymph nodes, suggesting it captures meaningful biological patterns.

MATERIALS

Dataset

This study utilized histopathological and clinical data from the European DECIDER project (*European Union, 2024*) (DECIDER observational clinical trial—Multi-layer Data to Improve Diagnosis, Predict Therapy Resistance and Suggest Targeted Therapies in HGSC; ClinicalTrials.gov identifier: NCT04846933), focused on overcoming chemotherapy resistance in ovarian high-grade serous cancer. The dataset includes patient cohorts with two distinct clinical trajectories: a Neoadjuvant Chemotherapy (NACT) group of patients who received neoadjuvant chemotherapy followed by interval debulking surgery, and a Primary Debulking Surgery (PDS) group of patients who underwent primary debulking surgery prior to any systemic treatment. Slide embeddings from these WSIs are available at https://github.com/AI-BioInformatics/MuGNet/tree/main/data; the original WSIs are protected under NDA and cannot be shared publicly. Label files are available at https://github.com/AI-BioInformatics/MuGNet/tree/main/docs: the *XLSX* file contains labels for the PFI and OS prediction tasks, while the *CSV* file provides the labels

used for the HR prediction task. WSIs were collected from 243 patients, with samples from various anatomical sites including the ovary, adnexa, omentum, peritoneum, lymph nodes, and other tissues.

Prediction tasks

Three distinct and clinically relevant prediction tasks were defined:

- **Homologous Recombination (HR) status prediction (HRD/HRP):** HRD (Homologous Recombination Deficient) tumors have impaired double-strand DNA repair mechanisms, typically due to BRCA1/2 mutations, promoter methylation, or alterations in HR-related genes (*e.g.*, RAD51C/D, PALB2). Conversely, HRP (Homologous Recombination Proficient) tumors have intact DNA repair pathways and are generally less responsive to DNA-damaging agents. HR status is a key biomarker that influences post-surgical treatment, particularly eligibility for PARP inhibitors and response to platinum-based chemotherapy. Labels were derived based on a full consensus across three orthogonal genomic indicators: HRDscarScore, SBS3, and ID6. Ambiguous cases were excluded to ensure label robustness. This binary classification task was performed on the PDS group to assess the intrinsic DNA repair capacity of untreated tumors, including 107 patients and 2,328 WSIs, with 53 HRD patients and 54 HRP patients.
- **Overall Survival (OS) prediction:** The OS prediction task was applied to the NACT interval group, which includes patients who received neoadjuvant chemotherapy followed by interval debulking surgery. Since these patients had already undergone systemic treatment at the time of tissue collection, they represent a clinically meaningful population for estimating long-term outcomes based on post-treatment tissue characteristics. To enable supervised learning, the continuous OS variable was discretized into ordinal categories. In the absence of universally accepted clinical thresholds for OS stratification in this context, we developed a balanced binning function based on quantile discretization. We initially tested fixed clinical thresholds (0–180, 181–365, >365) obtaining sparsely populated bins. Hence, we conducted comparative experiments with three and four bins that showed that the 4-bin configuration consistently provided the most stable and robust performance across both MuGNet and competitor baselines, avoiding sparsely populated outcome groups. This pragmatic strategy, balancing clinical relevance with statistical robustness, has been adopted in previous WSI-based survival studies. This task included 136 patients and 1,595 WSIs.
- **Progression-Free Interval (PFI) prediction:** The PFI prediction task was also performed on the NACT interval group. PFI is defined as the time between the completion of initial treatment and disease recurrence, and serves as a proxy for sensitivity to platinum-based chemotherapy. Because NACT patients had already undergone systemic treatment before tissue sampling, PFI prediction from their tissue morphology can offer insights into post-treatment response. To reflect clinically meaningful categories, PFI was discretized using thresholds based on established

guidelines for response stratification in ovarian cancer. This task included 133 patients and 1,586 WSIs.

Both the OS and PFI tasks were treated as ordinal classification problems and evaluated using survival ranking metrics, as described in the dedicated subsection on evaluation metrics.

WSI preprocessing and feature extraction

To prepare the WSIs and obtain a meaningful vector representation for each slide, a two-stage feature extraction pipeline was employed. The goal of this pipeline was to generate a single embedding for each WSI that would serve as a node in the GNN structure. Initially, each WSI was partitioned into non-overlapping 256 × 256 pixel patches at 20× resolution using the CLAM framework (*Lu et al., 2021*). Subsequently, two state-of-the-art models were used, which were trained on extensive and varied histopathological image contexts, making them ideal for generating robust WSI representations. CONCH1.5 (*Lu et al., 2024*), a contrastive foundation model, was used to generate patch-level embeddings. These patch-level embeddings were then aggregated by TITAN (*Ding et al., 2024*), a multimodal transformer model trained to align image and text modalities from pathology reports, to produce a final 768-dimensional vector representation for the entire WSI. Neither CONCH nor TITAN were retrained; they were used directly for inference to extract features, with the resulting 768-dimensional vector constituting the input node for the GNN framework.

METHODS

Proposed architecture

This work introduces a novel method for structuring histopathological data in predictive modeling. Instead of analyzing individual slides or isolated tissue patches, the framework builds a comprehensive graph representation for each patient, explicitly capturing multiple tissue samples simultaneously.

Each node in the patient graph represents a WSI embedding extracted from a distinct anatomical site. Node features are derived from TITAN embeddings, which encode a compact representation of the tissue morphology. Edges are established using a similarity-driven strategy: pairwise distances between tissue-level embeddings are computed in a lower-dimensional space, and a representative point is selected for each tissue type *via* k-Nearest Neighbors (k-NN) density estimation. These distances are transformed into edge weights using a Gaussian kernel, yielding a weighted adjacency matrix that reflects morphological similarity. An edge is retained only if its similarity exceeds a predefined threshold.

For readability, tissue names are reported in their abbreviated form throughout the text; a table with the corresponding full names and brief descriptions is provided in Table 1.

Before constructing the patient-level graphs, we established a site matrix to model similarity between anatomical sites. Using the entire training cohort, we computed a representative centroid $\mathbf{r}_i$ for each of the 12 tissue types listed in Table 1. The centroid $\mathbf{r}_i$

Table 1 Tissue abbreviations, full names, and brief descriptions.

Abbreviation	Full name	Description
Ome	Omentum	Fatty layer in the abdomen
Adn	Adnexa	Supporting structures of the uterus
Per	Peritoneum	Lining of the abdominal cavity
Ova	Ovary	Female reproductive organ producing eggs
Ute	Uterus	Female reproductive organ where embryo develops
Mes	Mesentery	Tissue attaching intestines to the abdominal wall
Bow	Bowel	Intestinal tissue
Tub	Fallopian Tube	Duct connecting ovary to uterus
LN	Lymph Node	Small immune system organ
LNR	Right Lymph Node	Lymph node on the right side
LNL	Left Lymph Node	Lymph node on the left side
Vag	Vagina	Female reproductive canal

was defined *via* k-Nearest Neighbors (k-NN) density estimation as the point with the minimum average distance to its $k = 5$ nearest neighbors within that tissue class. We then computed the site matrix $\mathbf{S}^{site} \in \mathbb{R}^{12\times12}$, where the entry s_{ij} represents the affinity between tissue type i and tissue type j:

$$s_{ij} = exp(-\gamma \cdot \|\mathbf{r}_i - \mathbf{r}_j\|^2).$$

Here $\|\mathbf{r}_i - \mathbf{r}_j\|$ represents the Euclidean distance between the centroids of tissue types i and j in the PCA-reduced embedding space, and γ is a scaling parameter set to 0.5 (corresponding to a Gaussian $\sigma = 1$, given $\gamma = \frac{1}{2\sigma^2}$). This fixed matrix serves to integrate prior knowledge into the model, encoding learned priors regarding tissue relationships (*e.g.*, the proximity between Omentum and Peritoneum). These similarity values s_{ij} define the edge weights in the graph. A threshold is applied, keeping connections if similarity is greater than or equal to 0.1, a common choice for stable and interpretable graph structures (*Garrison et al., 2015*).

For each patient, we constructed a patient-level graph $G = (V, E)$ where nodes V represent the patient's specific WSIs. If a patient has N_s slides, the graph has N_s nodes and an adjacency matrix $A \in \mathbb{R}^{N_s\times N_s}$. The edges of this patient-level graph are populated using the values from the site matrix S^{site}. Formally, for a graph with n nodes, the adjacency matrix $A \in \mathbb{R}^{n\times n}$ is symmetric and weighted, with each entry A_{ij} encoding the strength of the connection between nodes i and j as:

$$A_{ij} = \begin{cases} S^{site}_{T_i,T_j} & \text{if } S^{site}_{T_i,T_j} \geq 0.1 \\ 0 & \text{otherwise} \end{cases}.$$

This approach ensures that the graph structure ($N_s \times N_s$) is patient-specific, while the edge weights reflect robust, population-level biological similarities between the anatomical sites present.

The site matrix derived from the entire cohort is shown in Fig. 1. This matrix encapsulates the prior knowledge regarding inter-tissue similarities, which drives the weight assignment in individual patient graphs.

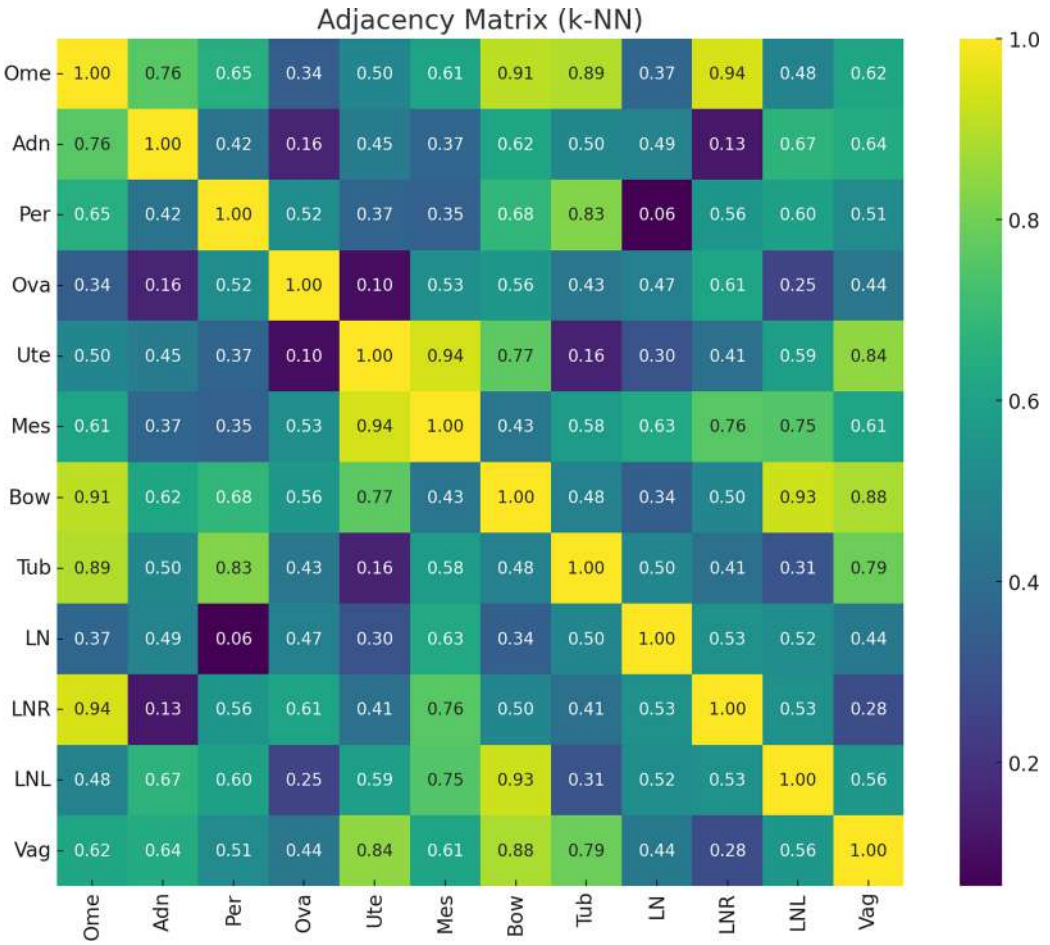


Figure 1 Global k-NN adjacency matrix (S^{site}) derived from the entire cohort. This matrix represents the morphological similarity between the 12 tissue types and serves as the source of edge weights (prior knowledge) for constructing patient-specific graphs. Full-size DOI: 10.7717/peerj-cs.3820/fig-1

This graph-based representation allows the model to capture inter-tissue dependencies and contextual cues that are not accessible when analyzing WSIs independently. By integrating both local (tissue-specific) and global (inter-tissue) patterns, the approach overcomes the core limitations of mono-tissue models.

Graph Attention Networks (GATs) (*Veličković et al., 2018*) are employed to model relationships among tissue regions. Given a graph $G = (V, E)$ with nodes $v \in V$ and edges $(v, u) \in E$, and node features $h_v^{(l)}$ at layer l, a GAT layer updates each node's embedding by aggregating information from its neighbors, assigning learnable, instance-specific weights during message passing:

$$h_v^{(l)} = \sigma\left(\sum_{u \in \mathcal{N}(v)} f\left(h_v^{(l-1)}, h_u^{(l-1)}\right)\right) \tag{1}$$

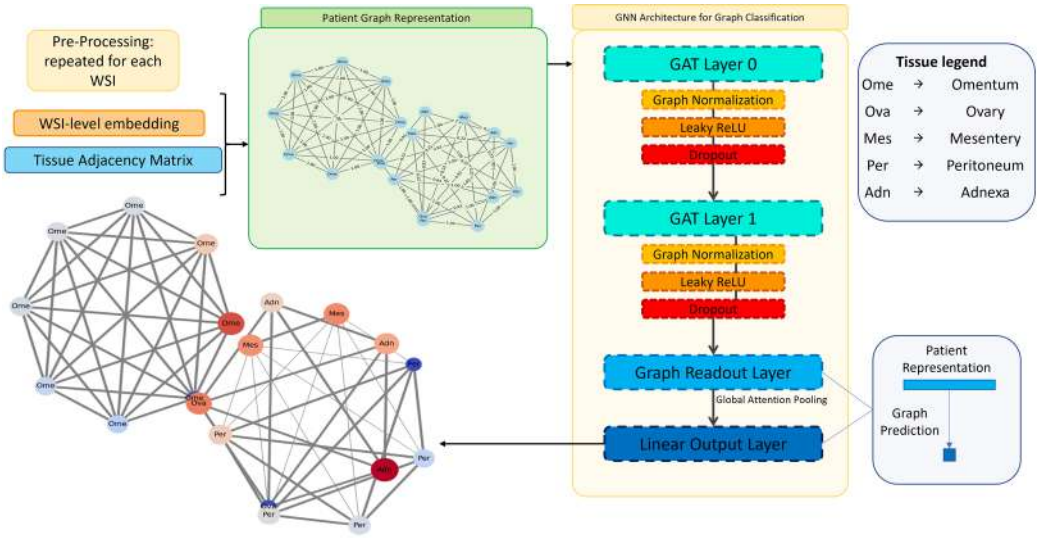


Figure 2 Overview of the proposed multi-tissue GNN framework. Input WSIs are first embedded *via* a two-stage pipeline, then structured as a patient-level graph. Node features are processed through two GATConv layers, and the resulting representations are aggregated *via* a global attention pooling module to produce the final patient-level prediction. Full-size DOI: 10.7717/peerj-cs.3820/fig-2

where $\mathcal{N}(v)$ denotes the neighborhood of node v, f is a learnable function, and σ is a non-linear activation function. The attention coefficient α_{vu} quantifies the importance of node u in updating node v and is computed as:

$$\alpha_{vu} = \frac{\exp(\text{LeakyReLU}(a^\top [Wh_v \| Wh_u]))}{\sum_{k \in \mathcal{N}(v)} \exp(\text{LeakyReLU}(a^\top [Wh_v \| Wh_k]))} \tag{2}$$

where W is a shared linear transformation, a is a learnable attention vector, and $\|$ denotes concatenation. The updated node embeddings are then computed as a weighted sum of the neighbors' representations:

$$h_v^{(l)} = \sigma\left(\sum_{u \in \mathcal{N}(v)} \alpha_{vu} W h_u^{(l-1)} \right). \tag{3}$$

To enhance model robustness, multi-head attention was used, computing multiple independent attention coefficients in parallel and combining them *via* concatenation or averaging.

The graph is processed through a 2-layer GAT architecture represented in Fig. 2. Each layer applies multi-head self-attention (four heads, 64 features per head), allowing the model to dynamically weigh neighbor contributions. This is followed by GraphNorm, a normalization strategy applied per graph that improves convergence and generalization in graph-level tasks (*Cai et al., 2020*). A LeakyReLU activation function introduces non-linearity, and dropout with a rate of 0.6 is used for regularization.

After message passing, node representations are aggregated using a *Global Attention Pooling* mechanism, which learns a weighted summary of node features based on learned

attention scores. The resulting pooled graph-level embedding is passed through a fully-connected layer to generate predictions. This design is compatible with both binary classification tasks (*e.g.*, HR status) and regression tasks (*e.g.*, OS or PFI), with the output dimensionality and the activation function of the final layer adjusted according to the task. This architecture captures both local and global tissue interactions, allowing dynamic weighting of each WSI *via* attention. Compared to mono-tissue models, the multi-tissue GNN provides improved predictive performance and interpretability in clinical outcome prediction.

Reproducibility

Our code is publicly available at https://github.com/AI-BioInformatics/MuGNet.git. All experiments were performed on a Linux-based environment running Ubuntu 22.04.5 LTS (Jammy Jellyfish) and a range of NVIDIA GPUs, including GeForce GTX 1080 (8 GB), RTX 2080 Ti (11 GB), RTX 5000 (16 GB), RTX 6000 (24 GB), and RTX A5000 (24 GB). The framework can be executed even on mid-range GPUs with 8 GB of memory, completing a full analysis in approximately 2 h using 32 GB of system RAM, ensuring reproducibility and accessibility for typical research environments.

Experimental setup and evaluation

Three clinically motivated prediction tasks were formulated and applied to the DECIDER dataset. All tasks were evaluated using 5-fold stratified cross-validation, ensuring consistent distribution of outcome classes across folds. The results presented in the 'Results' correspond to the mean performance across the five folds, accompanied by the corresponding standard deviation, to reflect both central tendency and variability across splits.

MuGNet has been compared against three state-of-the-art (SOTA) MIL models (ABMIL, DSMIL, and TransMIL), which were adapted to operate on the WSI-level embeddings, treating each WSI as an instance in a single patient-level bag. We acknowledge that applying MIL approaches to WSI-level aggregated embeddings, rather than strictly operating on patch-level bags, represents a limitation of our baseline implementation. However, this adaptation was necessary to ensure a fair, direct comparison with MuGNet within the same representation space, preventing discrepancies arising from differing input data structures. Model performance was assessed using task-specific metrics. For the prediction of HR status, standard classification metrics were used: accuracy, precision, recall, F1-score, and AUC (the area under the ROC curve, summarizing sensitivity *vs* specificity across thresholds).

For ordinal classification tasks involving OS and PFI, performance was evaluated using the concordance index (*c-index*), which measures agreement between predicted risk scores and actual survival times.

RESULTS

This section presents the results of the experiments conducted to evaluate the performance of the MuGNet framework against baseline models. Quantitative results for the three

Table 2 Classification performance comparison on HR status prediction task for the ABMIL, DSMIL and TransMIL models. Best results are reported in bold.

Model	Accuracy	Precision	Recall	F1-score	AUC
ABMIL	0.6117 ± 0.0912	0.6458 ± 0.1518	0.6255 ± 0.0775	0.6203 ± 0.0363	0.6583 ± 0.0870
DSMIL	0.6494 ± 0.0917	0.6436 ± 0.0813	0.7182 ± 0.2414	0.6598 ± 0.1160	0.6379 ± 0.1098
TransMIL	0.5000 ± 0.0168	0.2000 ± 0.2739	0.4000 ± 0.5477	0.2667 ± 0.3651	0.5750 ± 0.0520
Ours	**0.7000 ± 0.1093**	**0.6897 ± 0.1166**	**0.7564 ± 0.1154**	**0.7181 ± 0.1026**	**0.7220 ± 0.0739**

Table 3 Comparison of C-index scores for the ABMIL, DSMIL and TransMIL for the OS and PFI tasks. Best results are reported in bold.

Model	C-index (OS)	C-index (PFI)
ABMIL	0.4269 ± 0.1887	0.5778 ± 0.2134
DSMIL	0.5870 ± 0.1033	0.4444 ± 0.1628
TransMIL	0.6626 ± 0.1016	0.6674 ± 0.0919
Ours	**0.6814 ± 0.0819**	**0.6714 ± 0.1335**

prediction tasks (HR status, OS, and PFI) are reported in the following tables, which will be followed by an analysis of attention weights to provide insights into the model's interpretability.

Overall performance

The performance of the proposed MuGNet framework was evaluated against three state-of-the-art MIL models (ABMIL, DSMIL, and TransMIL) on three distinct prediction tasks. The results, summarized in the tables below, demonstrate that MuGNet consistently outperformed all baseline models across all tasks, thereby highlighting the efficacy of the graph-based approach for multi-tissue data integration.

For the HR task, the proposed method achieved the best classification performance across all standard metrics (Table 2), including accuracy (0.7000 ± 0.1093) and F1-score (0.7181 ± 0.1026). In contrast, TransMIL failed to generalize, with an accuracy close to chance level. Both ABMIL and DSMIL offered moderate performance, but fell short in recall and precision. Notably, the approach also yielded the highest AUC (0.7220 ± 0.0739), highlighting its ability to distinguish HRD from HRP cases.

MuGNet achieved the highest performance in the 4-bin ordinal survival classification setting, with a C-index score of 0.6814 ± 0.0819 (Table 3). Notably, it outperformed all MIL baselines, including TransMIL, which previously reported strong results on similar tasks. ABMIL exhibited limited generalization in this configuration, while DSMIL underperformed.

On the PFI prediction task (Table 3), the method again outperformed all baselines, reaching a C-index of 0.6714 ± 0.1335. While TransMIL achieved a comparable score (0.6674), the presented approach delivered more stable performance across folds. The gains over ABMIL and DSMIL were more pronounced, confirming the benefit of explicitly modeling multi-tissue relationships *via* a graph representation.

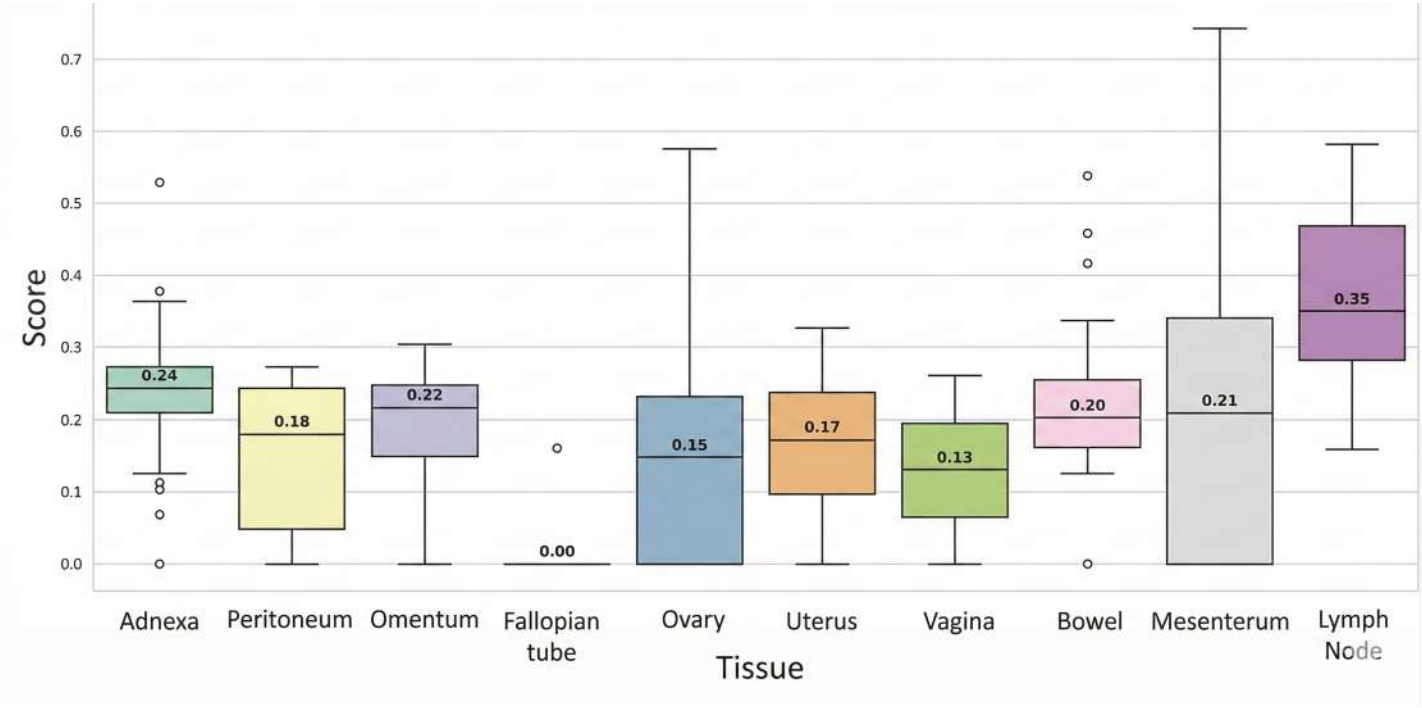


Figure 3 **Boxplot of node-level layer 1 attention scores across tissue types in the OS task, for both GAT layers.**
Full-size DOI: 10.7717/peerj-cs.3820/fig-3

Alternative adjacency schemes, including anatomical proximity, PCA correlation, and other k-NN thresholds, were also explored and are reported in the Supplemental Material (File S1), along with visualizations of both global and fold-specific adjacency matrices.

DISCUSSION

The results demonstrate that the presented graph-based framework consistently outperforms standard MIL baselines across all evaluated tasks. The gains are particularly notable in HR status prediction, where attention-based integration of multiple tissues appears to be critical. This likely stems from the inability of single-site models to capture morphologic patterns that are distributed across different anatomical contexts.

By leveraging a flexible, patient-level graph structure, the method accommodates heterogeneous sampling, both in terms of tissue composition and number of samples, without requiring a fixed input size. This design aligns more closely with real-world clinical scenarios than traditional MIL pipelines.

Importantly, the attention scores learned by the GAT layers provide tissue-level interpretability. To investigate which tissues contribute most to prediction, node-level attention scores were analyzed in the OS prediction task (Fig. 3). Although attention is computed at the edge level, scores were aggregated at the node level by summing the incoming and outgoing edge weights and normalizing by node degree. For interpretability, the focus was placed on source-side aggregation, highlighting tissues that act as

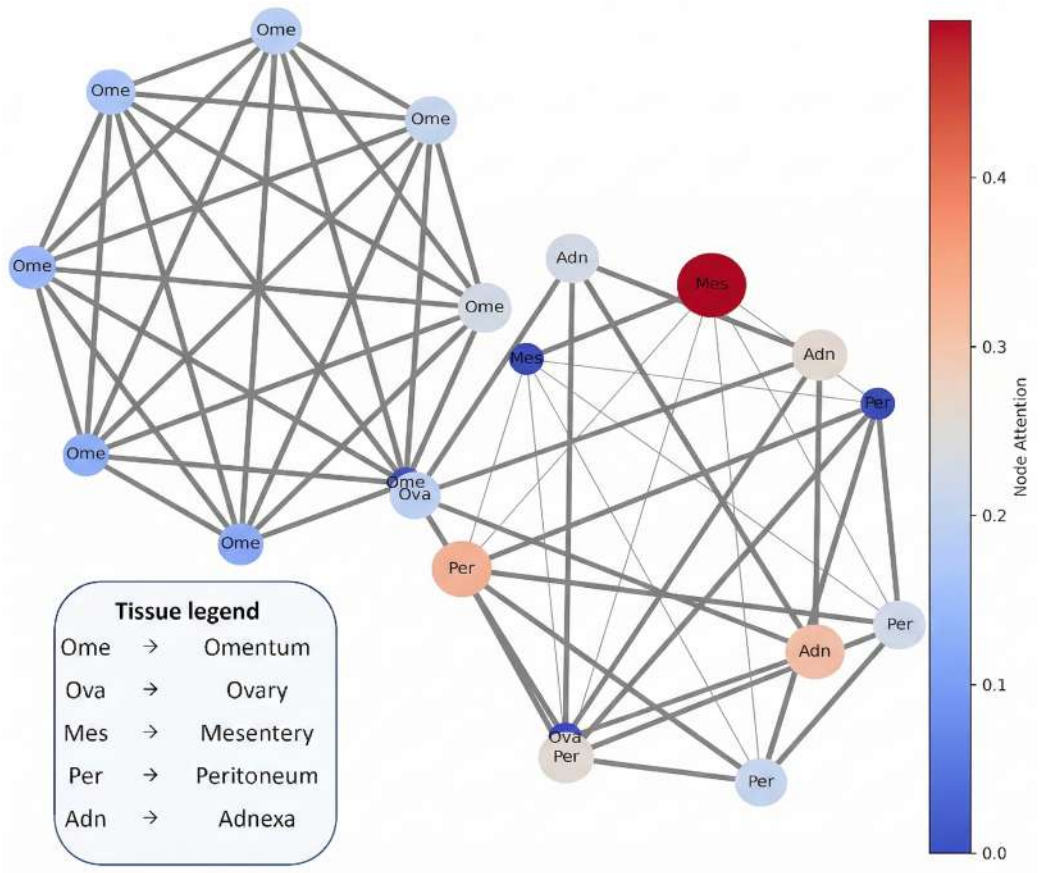


Figure 4 Visualization of a patient-specific graph: node size and color indicate attention score; edge thickness indicates connection strength. Full-size DOI: 10.7717/peerj-cs.3820/fig-4

informative providers during message passing[1]. This biologically motivated perspective emphasizes tissues whose features are referenced by others during prediction, offering a more mechanistic understanding of the model's decision process.

Notably, lymph node samples received consistently elevated attention across both GAT layers. These tissues are known to host metastatic lesions and are frequently associated with poor prognosis in ovarian HGSC, suggesting that the model captures clinically meaningful patterns. In contrast, adnexal and peritoneal tissues showed lower median attention scores, potentially reflecting reduced discriminative power in this context.

As illustrated in Fig. 4, the model tends to prioritize metastasis-rich regions (*e.g.*, lymph nodes). These insights not only enhance model trustworthiness but may also inform future research into site-specific pathophysiological mechanisms.

Furthermore, future investigations could benefit from implementing hierarchical modeling workflows. Such approaches would prioritize the analysis of intra-sample heterogeneity before modeling the interactions between multiple tissue sites, potentially capturing finer-grained pathological dynamics that global aggregation might overlook.

While the evaluation is limited to a single dataset, this choice reflects the scarcity of publicly available multi-tissue cohorts. To mitigate this limitation, the framework was

[1] Destination nodes represent the recipients of information and may receive attention from both relevant and irrelevant sources. Focusing on source nodes highlights those that actively transmit predictive signals.

extensively validated across three distinct clinical prediction tasks, demonstrating both flexibility and robustness.

CONCLUSIONS

Computational pathology has made significant strides through the application of deep learning to WSIs. However, a key challenge remains: the majority of existing models operate under a mono-tissue assumption, relying on a single WSI to inform a prediction. This limitation has largely been driven by the scarcity of publicly available datasets that provide more than one slide per patient. To bridge this gap, this study introduced MuGNet, a novel graph-based deep learning framework designed to enable patient-level prediction by integrating a variable number of tissue slides. This approach moves beyond the single-sample paradigm by modeling each patient as a structured graph, thereby facilitating a more holistic understanding of the disease from a multi-tissue perspective.

This approach consistently outperformed state-of-the-art MIL baselines in both survival prediction (OS and PFI) and biomarker classification (HR status) tasks, demonstrating robust generalization across various clinical endpoints. Beyond predictive performance, the use of Graph Attention Networks introduced a valuable layer of interpretability. The source-side attention scores allowed for the identification of the tissues that contribute most significantly to model decisions, with a consistent emphasis on biologically relevant sites such as the lymph nodes.

Although the evaluation was conducted on a single ovarian HGSC cohort, the model's inherent flexibility regarding the number and type of tissue samples per patient highlights its potential for future application in varied clinical settings. In future work, the framework can be readily extended to integrate multimodal or multi-center datasets, further enhancing its clinical applicability.

Overall, this study provides a scalable and interpretable modeling strategy for personalized pathology, representing a significant step toward the integration of artificial intelligence into precision oncology workflows and supporting informed, patient-specific clinical decision-making.

ACKNOWLEDGEMENTS

The authors used Gemini 3 for the text editing and grammatical refinement, improving the readability and language flow of the manuscript. After using the tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This study was funded by the European Union's Horizon 2020 research and innovation programme DECIDER under Grant Agreement 965193 and supported by funding from the Enzo Ferrari Engineering Department of the University of Modena and Reggio Emilia,

under the FARD-2024 grant program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
European Union's Horizon 2020 Research and Innovation Programme DECIDER: 965193.
Enzo Ferrari Engineering Department of the University of Modena and Reggio Emilia: FARD-2024.

Competing Interests

Marta Lovino is an Academic Editor for PeerJ.

Author Contributions

- Francesca Miccolis conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Martina Demarco conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Elisa Ficarra conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Marta Lovino conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The code is available at GitHub and Zenodo:

- https://github.com/AI-BioInformatics/MuGNet.git.

- Francesca Miccolis. (2026). AI-BioInformatics/MuGNet: v1.0 (v1.0). Zenodo. https://doi.org/10.5281/zenodo.18136144.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.3820#supplemental-information.

REFERENCES

Cai T, Luo S, Xu K, He D, Liu T-Y, Wang L. 2020. GraphNorm: a principled approach to accelerating graph neural network training. ArXiv DOI 10.48550/arXiv.2009.03294.

Ding T, Wagner SJ, Wagner SJ, Song AH, Chen RJ, Lu MY, Zhang A, Vaidya AJ, Jaume G, Shaban M, Kim A, Williamson DFK, Robertson H, Chen B, Almagro-Pérez C, Doucet P, Sahai S, Chen C, Chen CS, Komura D, Kawabe A, Ochi M, Sato S, Yokose T, Miyagi Y, Ishikawa S, Gerber G, Peng T, Le LP, Mahmood F. 2024. Multimodal whole slide foundation model for pathology. ArXiv DOI 10.48550/arXiv.2411.19666.

European Union. 2024. DECIDER project. *Available at https://www.deciderproject.eu/* (accessed 15 July 2025).

Garrison KA, Scheinost D, Finn ES, Shen X, Constable RT. 2015. The (in)stability of functional brain network measures across thresholds. *NeuroImage* **118**:651–661 DOI 10.1016/j.neuroimage.2015.05.046.

Ilse M, Tomczak JM, Welling M. 2018. Attention-based deep multiple instance learning. In: *Proceedings of the 35th International Conference on Machine Learning*, 2127–2136.

Konstantinopoulos PA, Ceccaldi R, Shapiro GI, D'Andrea AD. 2015. Homologous recombination deficiency: exploiting the fundamental vulnerability of ovarian cancer. *Cancer Discovery* **5(11)**:1137–1154 DOI 10.1158/2159-8290.cd-15-0714.

Li B, Li Y, Eliceiri KW. 2021. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 14318–14326.

Liu J, Song Y, Liu Q, Li L, Pan J, Luo L, Zhu S, Wu D, Hu D, Xu Q. 2025. Prognostic significance of chemotherapy response score in predicting outcomes for high-grade serous ovarian carcinoma patients undergoing neoadjuvant chemotherapy. *Translational Cancer Research* **14(4)**:2319–2330 DOI 10.21037/tcr-24-1654.

Lu MY, Chen B, Williamson DFK, Chen RJ, Liang I, Ding T, Jaume G, Odintsov I, Zhang A, Le LP, Gerber GK, Parwani AV, Mahmood F. 2024. A visual-language foundation model for computational pathology. *Nature Medicine* **30(3)**:863–874 DOI 10.1038/s41591-024-02856-4.

Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. 2021. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature Biomedical Engineering* **5**:1305–1315 DOI 10.1038/s41551-020-00682-w.

Mirza MR, Monk BJ, Herrstedt J, Oza AM, Mahner S, Redondo A, Fabbro M, Ledermann JA, Lorusso D, Vergote I, Ben-Baruch NE, Marth C, Madry R, Christensen RD, Berek JS, Döring H, Gorbunova V, Scambia G, Bartunkova J, Bover I, Tinker AV, du Bois A, González-Martín A, Sehouli J, Glasspool RM, O'Malley DM, Avall-Lundqvist E, Rustin GJS, Follana P, Wickens M, Paul J, McNeish IA. 2016. Niraparib maintenance therapy in platinum-sensitive, recurrent ovarian cancer. *New England Journal of Medicine* **375(22)**:2154–2164 DOI 10.1056/nejmoa1611310.

Rustin GJS, Vergote I, Eisenhauer E, Pujade-Lauraine E, Quinn M, Thigpen T, du Bois A, Kristensen G, Jakobsen A, Sagae S, Greven K, Parmar M, Friedlander M. 2011. Definitions for response and progression in ovarian cancer clinical trials incorporating RECIST 1.1 and CA 125 agreed by the gynecological cancer intergroup (GCIG). *International Journal of Gynecological Cancer* **21(2)**:419–423 DOI 10.1097/igc.0b013e3182070f17.

Shao Z, Bian H, Chen Y, Wang Y, Zhang J, Ji X, Zhang Y. 2021. TransMIL: transformer based correlated multiple instance learning for whole slide image classification. In: *NIPS'21: Proceedings of the 35th International Conference on Neural Information Processing Systems.*

Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. 2018. Graph attention networks. In: *International Conference on Learning Representations (ICLR).*

Wang J, Li X, Zhang M. 2024. Advances in multiple instance learning for whole slide image analysis: techniques, challenges, and future directions. ArXiv DOI 10.48550/arXiv.2408.09476.

Wen J, Chen L, Xu F. 2025. MsaMIL-Net: an end-to-end multi-scale aware multiple instance learning network for efficient whole slide image classification. ArXiv DOI 10.48550/arXiv.2503.08581.

Xu H, Wang M, Shi D, Qin H, Zhang Y, Liu Z, Madabhushi A, Gao P, Cong F, Lu C. 2025. When multiple instance learning meets foundation models. *Medical Image Analysis* **92**:102233 DOI 10.1016/j.media.2025.103456.