Explainable Classifier for Malignant Lymphoma Subtyping via Cell Graph and Image Fusion

Daiki Nishiyama[1,2⋆], Hiroaki Miyoshi[3], Noriaki Hashimoto[2], Koichi Ohshima[3], Hidekata Hontani[4], Ichiro Takeuchi[5,2], and Jun Sakuma[1,2]

[1] Institute of Science Tokyo, Tokyo, Japan
nishiyama.d.2d7f@m.isct.ac.jp
[2] RIKEN AIP, Tokyo, Japan
[3] Kurume University, Fukuoka, Japan
[4] Nagoya Institute of Technology, Aichi, Japan
[5] Nagoya University, Aichi, Japan

Abstract. Lymphoma subtype classification has a direct impact on treatment and outcomes, necessitating models that are both accurate and explainable. This study proposes a novel explainable Multi-Instance Learning (MIL) framework that identifies subtype-specific Regions of Interest (ROIs) from Whole Slide Images (WSIs) while integrating features of cell distribution and image. Our framework simultaneously addresses three objectives: (1) indicating appropriate ROIs for each subtype, (2) explaining the frequency and spatial distribution of characteristic cell types, and (3) reaching accurate subtyping using both cell distribution and image modalities. Our method fuses cell graph and image features extracted for each patch in a WSI by a Mixture-of-Experts-based approach and classifies subtypes within an MIL framework. Experiments on a dataset of 1,233 WSIs demonstrate that our approach achieves state-of-the-art accuracy compared to ten other methods and provides region- and cell-level explanations that align with a pathologist's perspective.

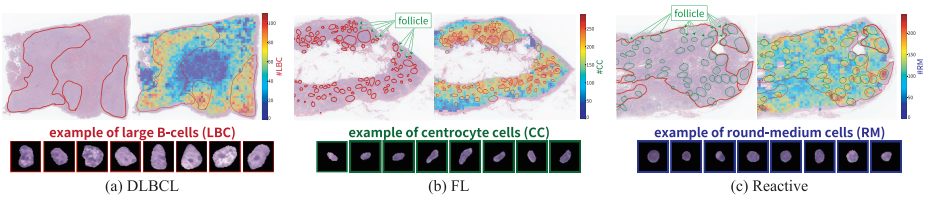
Keywords: Explainable AI · Multi-modality · Multiple Instance Learning · Malignant Lymphoma · Subtyping · Whole Slide Image · Cell Graph

1 Introduction

Subtyping of malignant lymphomas is important because it is essential for determining appropriate treatment strategies and directly affects patient prognosis. Recent advances in whole slide imaging (WSI) technology [10] have enabled the development of machine learning models capable of classifying malignant lymphoma subtypes from a WSI. However, to incorporate a model into actual clinical practice, it must not only achieve high diagnostic accuracy but also provide a reliable explanation for its decisions.

Approaches that mimic the diagnostic processes employed by pathologists are viewed as beneficial for attaining high classification accuracy and reliable explanations. Pathologists diagnose lymphoma subtypes by investigating H&E-stained

⋆ Corresponding Author



(a) DLBCL (b) FL (c) Reactive

Fig. 1. Left to right: DLBCL, FL, and Reactive case, and each subtype's characteristic cells, LBC, CC, and RM. Red lines indicate ROIs, and green and red lines in Reactive and FL, respectively, indicate follicles. Heat maps indicate the spatial distribution of each subtype's number of characteristic cells in a 512×512 pixels at 40x magnification.

slide specimens. They concentrate on local regions of interest (ROIs), which vary for each subtype, and analyze the characteristics, frequency, and spatial distribution of each cell type within these ROIs. During diagnosis, pathologists utilize not only cell-based features but also visual information from pathological images to understand the broader tissue architecture.

As examples, we explain using three clinically important subtypes that we address in this research: diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FL), and reactive lymphoid hyperplasia (Reactive). Figure 1 shows, for each subtype, a representative case along with its ROIs, cells corresponding to the subtype, and the distribution of these cells' counts across the WSI. ROIs of DLBCL show a high frequency of large B-cells (LBCs). ROIs of FL are inside nearly round structures called *follicles*, where specific cell types such as central cells (CC) overabundantly exist. ROIs of Reactive extend within and between follicles, and many round-medium cells (RMs) are observed, especially near the follicular boundaries. As Fig. 1 indicates, the distribution of cell types characteristic of each subtype is observed only when ROIs specific to that subtype can be properly determined.

Consequently, an explainable malignant lymphoma classification model must be able to (1) identify appropriate ROIs for each subtype, (2) explain the frequency and spatial distribution of characteristic cell types, and (3) achieve high-accuracy subtyping by leveraging both image and cell distribution modalities. Cell-level explanations (2) rely on accurately identifying subtype-specific ROIs (1), while the correct determination of ROIs is contingent on understanding cell-type distributions and sustaining overall classification accuracy (3). Therefore, our goal is to integrate all three capabilities simultaneously and accomplish a reliable and explainable subtype classification of malignant lymphomas.

1.1 Related Works

WSI is a gigapixel-scale image that cannot be directly processed by machine learning due to memory constraints. Multiple Instance Learning (MIL) [17] addresses this by partitioning WSIs into a set (called *bag*) of smaller patches

(called *instances*). To indicate ROIs in a WSI, attention-based MILs (ABMILs) [26,13,23] highlight important regions but provide only a single attention score. This is insufficient for lymphoma subtyping, where different regions are important for different subtypes. AdditiveMIL [15] can be a solution to this situation by providing attention scores of instances for each subtype, clarifying the contribution of each instance to the subtyping result. One limitation of these is that they rely solely on image features without explicitly utilizing cell information.

Cell graphs, where nodes denote cells and edges denote cell adjacencies, effectively capture spatial cell distribution in pathological images. They can be encoded into feature vectors by graph neural networks to classify subtypes and grades investigated in [1,2,16,14,19,22,24,28]. For instance, HACT-Net [22] introduced hierarchical cell graphs to deal with cell- and tissue-level graph structures. In [14], a post-hoc explainable method was proposed to show morphological attributes of cells important for prediction. However, these methods are designed for pre-defined and small-scale ROIs rather than WSI-scale, limiting their ability to (1) discover local ROIs specific to the assumed subtype in a WSI and (2) utilize cell graphs for subtyping and explanation across the discovered ROIs.

These limitations are interdependent in malignant lymphoma subtyping and cannot be solved by simply applying existing methods.

1.2 Contributions

To resolve these problems, a multi-modal MIL-based model utilizing both image and cell graphs is seen as effective. Our contributions are summarized as follows:

- We propose an explainable classification framework for malignant lymphoma WSIs that indicates localized ROIs through class-wise importance per patch and demonstrates the frequency and spatial distribution of characteristic cell types. Experiments confirm that our framework provides region- and cell-level explanations that are well-aligned with a pathologist's view.
- To achieve high accuracy, we propose a new multi-modal fusion method, Weak-Expert-based Gating Mixture-of-Experts (WEG-MoE), which fuses cell graph and image features for each patch. Our method achieves state-of-the-art accuracy compared to ten other methods in classifying three major lymphoma subtypes using our 1,233 WSIs dataset.
- According to our investigation, 16 computational pathology studies utilize cell graphs, yet none of them address malignant lymphoma. Our work is, therefore, the first to apply cell graphs specifically to malignant lymphoma.

2 Methodology

We treat each WSI of lymph node tissue as a *bag* comprising multiple *instances* (patches). Let $n \in \{1, \ldots, N\}$ be an index of bags (WSIs), $m \in \{1, \ldots, M\}$ an index of instances in a bag, $X_{n,m}$ be the m-th image instance in the n-th bag, and $\boldsymbol{y}_n \in \{0,1\}^K$ be the n-th one-hot label over the $K=3$ subtypes, i.e., DLBCL, FL, and Reactive. Here, we define $\mathbb{X}_n = \{X_{n,m}\}_{m=1}^{M} \in \mathcal{X}$ as the n-th image bag, where $\mathcal{X}$ is a space of image bags. Then the n-th bag is defined by $\mathcal{B}_n = \left(\mathbb{X}_n, \boldsymbol{y}_n\right)$.

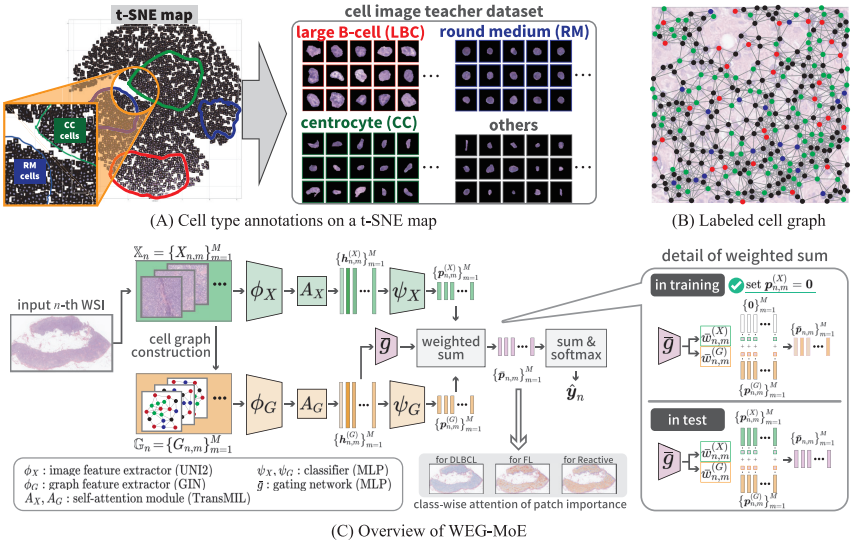


Fig. 2. (A) Annotations on a t-SNE map to label cells with LBC, CC, RM, or others. (B) Example of a cell graph to be constructed. The nodes' colors indicate the cell types. (C) An overview of our method, WEG-MoE, for classifying the n-th WSI.

2.1 Labeled Cell Graph Construction

We construct a labeled cell graph $G_{n,m}$ with cells in $X_{n,m}$, where each node label represents a corresponding cell's type. For cell labeling, we combine cell segmentation using HoVerNet [9] pre-trained on CoNSep [9] with a cell classifier. To train the cell classifier, as Fig. 2(A) shows, we construct the training dataset by having a pathologist label cell types (LBC, CC, RM, and others) on a 2D t-SNE map. The t-SNE map is created from segmented cells' feature vectors, encoded by ImageNet [12] pre-trained ResNet34 [7]. Identifying LBC, CC, and RM with high recall is crucial for lymphoma subtyping. Thus, we employed CAMRI loss [20] as the loss function, which enables the classifier to maintain relatively high recall for these three classes while preserving overall accuracy.

Finally, we get the labeled cell graph $G_{n,m} = (V_{n,m}, E_{n,m}, L_{n,m})$ corresponding to $X_{n,m}$ like Fig. 2(B), where $V_{n,m}$ is the set of cell indices, $E_{n,m}$ is the set of edges connecting cells within radius r to reflect the spatial density of cells ($r = 60$ in our setting), and $L_{n,m}$ is the set of corresponding cell labels. Let $\mathbb{G}_n = \{G_{n,m}\}_{m=1}^{M}$, then we reformulate the n-th WSI bag by $\mathcal{B}_n = (\mathbb{X}_n, \mathbb{G}_n, \boldsymbol{y}_n)$.

2.2 MoE-based Multimodal MIL

As we stated in Sec. 1, the frequency and spatial distribution of specific cell types are crucial for subtyping lymphoma. While labeled cell graphs provide this information, they lack visual elements like cell texture and tissue architecture in pathological images. Then, combining cell graphs with image fea-

tures is essential for accurate and explainable subtyping. Therefore, for an input $\mathcal{B}_n = \big(\mathbb{X}_n, \mathbb{G}_n, \boldsymbol{y}_n\big)$, we develop an explainable multi-modal classification model utilizing both graphs and images to achieve high classification performance.

Pretraining AdditiveMIL for Two Modalities. To capture ROIs with different features for each subtype, as we stated in Sec. 1, we adopt AdditiveMIL [15]. We first introduce AdditiveMIL, where only a single modality (cell graph or image patch) is available. Let f_G and f_X be MIL models for cell graphs and images, respectively, $\phi_G : \mathcal{G} \to \mathbb{R}^d$ be a feature extractor (e.g., GIN [27]) of f_G, and $\phi_X : \mathcal{X} \to \mathbb{R}^d$ be a feature extractor (e.g., UNI [4]) of f_X. Given an instance $(X_{n,m}, G_{n,m})$ in a bag $\mathcal{B}_n$, these models produce d-dimensional latent features:

$$\boldsymbol{h}^{(G)}_{n,m} = A_G\big(\phi_G(G_{n,1}), \ldots, \phi_G(G_{n,M})\big)_m, \boldsymbol{h}^{(X)}_{n,m} = A_X\big(\phi_X(X_{n,1}), \ldots, \phi_X(X_{n,M})\big)_m, \quad (1)$$

where $A_G, A_X : \mathbb{R}^{M\times d} \to \mathbb{R}^{M\times d}$ are self-attention modules (e.g., TransMIL [23]). After each $\boldsymbol{h}^{(\cdot)}_{n,m}$ is mapped to a K-dimensional logit via $\psi_G, \psi_X : \mathbb{R}^d \to \mathbb{R}^K$, which are learnable functions (e.g., multilayer perceptrons (MLPs)), finally, we obtain the bag-level class probability:

$$\boldsymbol{p}^{(G)}_{n,m} = \psi_G\big(\boldsymbol{h}^{(G)}_{n,m}\big), \quad \boldsymbol{p}^{(X)}_{n,m} = \psi_X\big(\boldsymbol{h}^{(X)}_{n,m}\big), \quad (2)$$

$$\hat{\boldsymbol{y}}^{(G)}_n = \mathrm{softmax}\big(\textstyle\sum_{m=1}^{M} \sigma\big(\boldsymbol{p}^{(G)}_{n,m}\big)\big), \quad \hat{\boldsymbol{y}}^{(X)}_n = \mathrm{softmax}\big(\textstyle\sum_{m=1}^{M} \sigma\big(\boldsymbol{p}^{(X)}_{n,m}\big)\big), \quad (3)$$

where σ is a sigmoid function to stabilize learning in our case (optional), and softmax is a softmax function. Here, the j-th element of $\boldsymbol{p}^{(\cdot)}_{n,m}$ represents the importance of the m-th instance, assuming the prediction result is the j-th class. Because $\hat{\boldsymbol{y}}^{(\cdot)}_n$ is given as an addition of the class-wise attention score $\boldsymbol{p}^{(\cdot)}_{n,m}$, the impact of each instance on the classification can be obtained for each class. f_G and f_X are pre-trained independently using cross-entropy loss $\ell(\hat{\boldsymbol{y}}^{(\cdot)}_n, \boldsymbol{y}_n)$.

Weak-Expert-based Gating MoE. To combine cell graph and image modalities, we employ an MoE-based approach. First, we introduce a naive MoE. Let $g : \mathbb{R}^{2d} \to [0,1]^2$ be a gating function, and the bag-level prediction by MoE is

$$\hat{\boldsymbol{y}}^{(\mathrm{moe})}_n = \mathrm{softmax}\big(\textstyle\sum_{m=1}^{M} w^{(G)}_{n,m}\, \boldsymbol{p}^{(G)}_{n,m} + w^{(X)}_{n,m}\, \boldsymbol{p}^{(X)}_{n,m}\big), \quad (4)$$

$$\big[w^{(G)}_{n,m},\, w^{(X)}_{n,m}\big] = \mathrm{softmax}\big(g(\mathrm{cat}[\boldsymbol{h}^{(G)}_{n,m}, \boldsymbol{h}^{(X)}_{n,m}])\big), \quad (5)$$

where $\mathrm{cat}[\cdot,\cdot]$ is a concatenation of two vectors. Each $w^{(\cdot)}_{n,m}$ in Eq. (4) and Eq. (5) can be interpreted as how much each modality contributes to the m-th instance in the n-th bag. g allows the model to combine cell graph and image modalities adaptively, considering the contributions of both modalities.

Each modality, i.e., image (ResNet50, UNI2) and graph (GIN), shows decent classification accuracy as shown in Tab. 1, and the naive MoE simply fuses them without considering modality differences. However, the naive MoE actually tends

to ignore the graph modality, possibly because the image modality contributes to achieving loss minimization more rapidly. To better utilize both modalities while preserving explainability, we propose Weak-Expert-based Gating MoE (WEG-MoE), shown in Fig. 2(C). WEG-MoE gates each modality using only the weaker modality (i.e., graph), which typically shows lower accuracy. Let $\bar{g}:\mathbb{R}^d \to [0,1]^2$ be WEG-MoE's gating function. The bag-level prediction by WEG-MoE is

$$\hat{\boldsymbol{y}}_n^{(\mathrm{weg})} = \mathrm{softmax}\Big(\sum\nolimits_{m=1}^{M} \bar{w}_{n,m}^{(G)}\, \boldsymbol{p}_{n,m}^{(G)} + \bar{w}_{n,m}^{(X)}\, \boldsymbol{p}_{n,m}^{(X)}\Big), \tag{6}$$

$$\text{where} \quad \big[\bar{w}_{n,m}^{(G)}, \bar{w}_{n,m}^{(X)}\big] = \mathrm{softmax}\big(\bar{g}\big(\boldsymbol{h}_{n,m}^{(G)}\big)\big). \tag{7}$$

During training, after independently pre-training each modality model f_G and f_X, we force $\boldsymbol{p}_{n,m}^{(X)} = \mathbf{0}$ and optimize only $\bar{g}$'s parameters. In this way, WEG-MoE determines whether the graph modality is useful for each patch and uses the image modality as necessary.

3 Experiments

Requirement for the explainable lymphoma subtyping is to simultaneously address (1) accurate classification of multiple subtypes, (2) identification of subtype-specific ROIs within a WSI, and (3) capture the spatial distribution of characteristic cells present in the ROIs. In this section, we show that our approach meets the requirement.

3.1 Experiment Setup

Because public datasets have only a single subtype label [25,8] or ROIs without full WSIs [21], these datasets are unsuitable for evaluation based on the requirement as mentioned above. Then, we constructed a private dataset comprising 1,233 lymphoma WSIs diagnosed at Kurume University, with labels of DLBCL, FL, and Reactive subtypes (411 per subtype). For training the cell classifier, 237,544 cell images were carefully selected to prevent test data leakage. Annotation of cell labels by a pathologist was completed in a few minutes using the annotation method introduced in Sec. 2.1. Each patch was clipped at 512×512 pixels from a non-background area and selected if it had at least 100 cells.

We compared the performance of our WEG-MoE with several baselines. Unimodal baselines included ResNet50 [12], UNI2 [4], GIN [27], and HACT-Net [22]. Multi-modal baselines included concat [18], Pathomic Fusion [5], MCAT [6], mutual attention [3], and MoE [11]. To adapt all cases to MIL, AdditiveMIL [15] and TransMIL [23] served as (f_X, A_X) and (f_G, A_G), respectively. In AdditiveMIL, ψ_X and ψ_G were three fully connected (FC) layers with ReLU activation. In UNI2, an FC layer followed UNI2's feature extractor, and only this FC layer was trained. In GIN, the feature extractor consisted of four GIN layers. In HACT-Net, we used deep features as node features, following [22]. In multi-modal baselines, UNI2 was ϕ_X, and GIN was ϕ_G, and $\boldsymbol{h}_{n,m}^{(X)}$ and $\boldsymbol{h}_{n,m}^{(G)}$ with latent dimension

Table 1. Mean and standard deviation across five cross-validation sets: accuracy, each class AUC, and mean AUC. The table is divided into image modality, graph modality, and multi-modality sections by lines from the top. Best performance is in bold.

Method	Accuracy	DLBCL AUC	FL AUC	Reactive AUC	AUC mean
ResNet50[12]	0.842 ± 0.032	0.956 ± 0.022	0.870 ± 0.059	0.973 ± 0.016	0.933 ± 0.045
UNI2[4]	0.904 ± 0.014	0.980 ± 0.010	0.948 ± 0.015	0.988 ± 0.010	0.972 ± 0.017
GIN[27]	0.829 ± 0.019	0.964 ± 0.012	0.807 ± 0.047	0.916 ± 0.025	0.896 ± 0.065
HACT-Net[22]	0.761 ± 0.077	0.957 ± 0.022	0.856 ± 0.045	0.943 ± 0.023	0.919 ± 0.045
concat[18]	0.907 ± 0.012	0.980 ± 0.011	0.954 ± 0.002	$\mathbf{0.989 \pm 0.008}$	0.975 ± 0.015
Pathomic Fusion[5]	0.902 ± 0.015	0.975 ± 0.017	0.949 ± 0.016	0.986 ± 0.012	0.970 ± 0.015
MCAT g2i[6]	0.901 ± 0.011	0.973 ± 0.010	0.953 ± 0.013	0.988 ± 0.004	0.971 ± 0.014
MCAT i2g[6]	0.881 ± 0.019	0.975 ± 0.013	0.951 ± 0.015	0.986 ± 0.006	0.971 ± 0.015
mutual attention[3]	0.906 ± 0.018	0.972 ± 0.017	0.956 ± 0.008	0.987 ± 0.007	0.972 ± 0.013
MoE[11]	0.907 ± 0.012	0.977 ± 0.013	0.951 ± 0.016	0.988 ± 0.010	0.972 ± 0.015
WEG-MoE (ours)	$\mathbf{0.911 \pm 0.011}$	$\mathbf{0.983 \pm 0.009}$	$\mathbf{0.961 \pm 0.011}$	0.988 ± 0.010	$\mathbf{0.977 \pm 0.012}$

$d = 256$ were fused. In Pathomic Fusion, we reduced the feature dimension to 48 before the direct product. In MCAT, we employed two flows to handle its directional attention flow: graph-to-image (g2i) and image-to-graph (i2g). In MoE and WEG-MoE, g and $\bar{g}$ were three FC layers with ReLU activation.

We used ten NVIDIA A100 80GB GPUs. In WEG-MoE, we processed ten WSIs simultaneously in parallel, taking approximately one hour in total from WSI scanning to test prediction. Our implementation is publicly available[6].

3.2 Results

Table 1 shows classification performance. We can see that WEG-MoE outperforms the others in classification performance. Especially since WEG-MoE outperforms MoE, we can confirm the effectiveness of WEG-MoE's gating strategy.

Figure 3 shows explainability results with a representative case: (A) class-wise attention of each subtype, (B) frequency by cell type, and (C) frequency of cell adjacency. (B) and (C) are computed from high-attention regions (top 25%). Dotted lines show distribution within pathologist-supervised ROIs. Below, we contrast what the pathologist expects with explanations by our framework.

DLBCL. **Expectation:** DLBCL has spread lesions with increased and diffusely distributed LBCs. **Attention:** Figure 3(A-1) shows higher attention for DLBCL overall than other subtypes, reflecting that LBCs are diffusely distributed overall. **Cell frequency:** Figure 3(B-1) indicates increased LBCs compared to other subtypes. **Cell distribution:** LBCs in DLBCL are adjacent to a greater variety of cells than in Reactive (Fig. 3(C-1) vs (C-2)). **Validity:** These results are consistent with the expectation, and the distribution of LBC is close to it in the supervised ROI, so they are aligned with the pathologist's view.

FL. **Expectation:** FL lesions are follicles with densely packed cells and numerous CCs. **Attention:** In specific regions, Fig. 3(A-2) shows higher attention

[6] https://github.com/mdl-lab/Explainable-Malignant-Lymphoma-Classifier

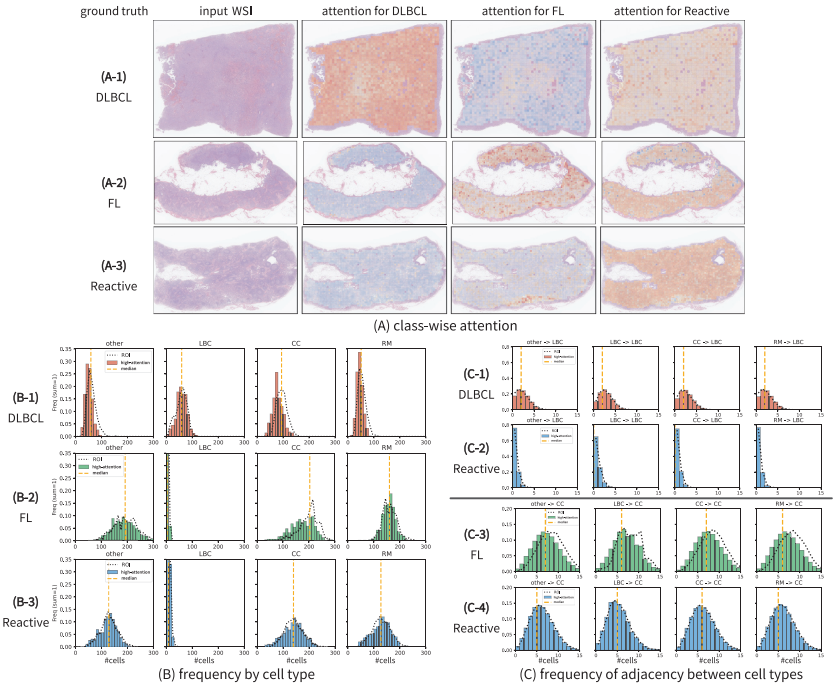


Fig. 3. (A) Class-wise attention, where higher attention is red, and lower is blue. (B) Frequency by cell types. (C) Frequency of adjacency between cell types. In (B) and (C), the black dotted lines present the results computed with the ROI supervised by a pathologist, which can be interpreted as the ground truth, and the orange dotted lines present the median of the distribution. In (C), for example, "LBC→CC" indicates the number of CCs that are connected by edges to LBC. Each data in (B) and (C) is from the top 25% of instance-level class-wise attention scores for the correct class.

for FL, while it is lower for Reactive and DLBCL. **Cell frequency:** Figure 3(B-2) indicates increased CCs compared to other subtypes. **Cell distribution:** The number of CCs adjacent to various cells in FL (Fig. 3(C-3)) exceeds those in Reactive (Fig. 3(C-4)). **Validity:** These results indicate that the follicles show attention and have the characteristics of FL rather than Reactive, so they are aligned with the pathologist's view.

Reactive. **Expectation:** Reactive tissue shows normal follicular structure and inter-follicular tissue throughout, with RM often at follicle boundaries. **Attention:** Figure 3(A-3) shows high attention for Reactive overall, with locally high FL attention in the lower part. **Cell frequency:** Reactive shows fewer LBCs than DLBCL (Fig. 3(B-3) vs (B-1)) and fewer CCs, RMs, and other cells than FL (Fig. 3(B-3) vs (B-2)). **Cell distribution:** Reactive shows fewer cell adjacencies for both LBCs and CCs compared to DLBCL and FL, respectively. **Validity:** Areas with high FL attention appear as FL to pathologists, while other regions display Reactive characteristics rather than malignant patterns, validat-

ing the model's attention allocation. Due to the Reactive ROIs covering areas both within and outside follicles, it can not be observed that the RM frequency is high at follicular borders. When comparing cell distributions, the Reactive case displays non-malignant patterns distinct from DLBCL and FL. These patterns closely match the distribution in pathologist-identified ROIs, confirming the overall appropriateness of the model's interpretations.

4 Conclusion

We proposed an explainable multi-modal MIL framework for the subtyping of malignant lymphoma. This framework can explain not only localized ROIs through class-wise attention but also the frequency and spatial distribution of characteristic cell types based on the labeled cell graph. Our experiments confirmed the appropriateness of these explanations based on the pathologist's assessment and showed state-of-the-art classification performance.

Acknowledgments. This work is partly supported by the Japan Science and Technology Agency (JST), CREST JPMJCR21D3, and the Japan Society for the Promotion of Science (JSPS), Grants-in-Aid for Scientific Research 23H00483 and 24KJ1049.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abbas, S.F., Le Vuong, T.T., Kim, K., Song, B., Kwak, J.T.: Multi-cell type and multi-level graph aggregation network for cancer grading in pathology images. Medical Image Analysis **90**, 102936 (2023)
2. Baranwal, M., Krishnan, S., Oneka, M., Frankel, T., Rao, A.: Cgat: Cell graph attention network for grading of pancreatic disease histology images. Frontiers in Immunology **12**, 727610 (2021)
3. Cai, G., Zhu, Y., Wu, Y., Jiang, X., Ye, J., Yang, D.: A multimodal transformer to fuse images and metadata for skin disease classification. The Visual Computer **39**(7), 2781–2793 (2023)
4. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F., Jaume, G., Chen, B., Zhang, A., Shao, D., Song, A.H., Shaban, M., et al.: Towards a general-purpose foundation model for computational pathology. Nature Medicine (2024)
5. Chen, R.J., Lu, M.Y., Wang, J., Williamson, D.F., Rodig, S.J., Lindeman, N.I., Mahmood, F.: Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. IEEE Transactions on Medical Imaging (2020)
6. Chen, R.J., Lu, M.Y., Weng, W.H., Chen, T.Y., Williamson, D.F., Manz, T., Shady, M., Mahmood, F.: Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4015–4025 (2021)
7. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)

8. Fernandez-Pol, S., Natkunam, Y., Vrabac, D., Rojansky, R., Advani, R., Rajpurkar, P., S, Ng, A.Y.: H&e and immunohistochemical stain images of 209 cases of diffuse large b-cell lymphoma linked with cytogenetic features and clinical outcomes (version 1) [data set] (2022). https://doi.org/10.7937/NVA3-N783
9. Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N.: Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Medical Image Analysis **58**, 101563 (2019)
10. Hanna, M.G., Parwani, A., Sirintrapun, S.J.: Whole slide imaging: technology and applications. Advances in Anatomic Pathology **27**(4), 251–259 (2020)
11. Hashimoto, N., Hanada, H., Miyoshi, H., Nagaishi, M., Sato, K., Hontani, H., Ohshima, K., Takeuchi, I.: Multimodal gated mixture of experts using whole slide image and flow cytometry for multiple instance learning classification of lymphoma. Journal of Pathology Informatics **15**, 100359 (2024)
12. 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)
13. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
14. Jaume, G., Pati, P., Bozorgtabar, B., Foncubierta, A., Anniciello, A.M., Feroce, F., Rau, T., Thiran, J.P., Gabrani, M., Goksel, O.: Quantifying explainers of graph neural networks in computational pathology. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8106–8116 (2021)
15. Javed, S.A., Juyal, D., Padigela, H., Taylor-Weiner, A., Yu, L., Prakash, A.: Additive mil: Intrinsically interpretable multiple instance learning for pathology. Advances in Neural Information Processing Systems **35**, 20689–20702 (2022)
16. Krishna, A., Gupta, R.K., Kurian, N.C., Jeevan, P., Sethi, A.: Heterogeneous graphs model spatial relationship between biological entities for breast cancer diagnosis. In: Ahmadi, S.A., Pereira, S. (eds.) Graphs in Biomedical Image Analysis, and Overlapped Cell on Tissue Dataset for Histopathology. pp. 97–106. Springer Nature Switzerland, Cham (2024)
17. Maron, O., Lozano-Pérez, T.: A framework for multiple-instance learning. Advances in neural information processing systems **10** (1997)
18. Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D.A., Barnholtz-Sloan, J.S., Velázquez Vega, J.E., Brat, D.J., Cooper, L.A.: Predicting cancer outcomes from histology and genomics using convolutional networks. Proceedings of the National Academy of Sciences **115**(13), E2970–E2979 (2018)
19. Nair, A., Arvidsson, H., Gatica V, J.E., Tudzarovski, N., Meinke, K., Sugars, R.V.: A graph neural network framework for mapping histological topology in oral mucosal tissue. BMC bioinformatics **23**(1), 506 (2022)
20. Nishiyama, D., Fukuchi, K., Akimoto, Y., Sakuma, J.: Camri loss: Improving the recall of a specific class without sacrificing accuracy. IEICE TRANSACTIONS on Information and Systems **106**(4), 523–537 (2023)
21. Orlov, N., Chen, W., Eckley, D., Macura, T., Shamir, L., Jaffe, E., Goldberg, I.: Automatic classification of lymphoma images with transform-based global features. IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society **14**, 1003–13 (07 2010). https://doi.org/10.1109/TITB.2010.2050695
22. Pati, P., Jaume, G., Foncubierta-Rodríguez, A., Feroce, F., Anniciello, A.M., Scognamiglio, G., Brancati, N., Fiche, M., Dubruc, E., Riccio, D., Di Bonito, M., De

Pietro, G., Botti, G., Thiran, J.P., Frucci, M., Goksel, O., Gabrani, M.: Hierarchical graph representations in digital pathology. Medical Image Analysis **75**, 102264 (2022)
23. 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)
24. Sims, J., Grabsch, H.I., Magee, D.: Using hierarchically connected nodes and multiple gnn message passing steps to increase the contextual information in cell-graph classification. In: MICCAI Workshop on Imaging Systems for GI Endoscopy. pp. 99–107. Springer (2022)
25. The Cancer Genome Atlas Research Network: The cancer genome atlas lymphoid neoplasm diffuse large b-cell lymphoma (tcga-dlbc). Genomic Data Commons Data Portal (2016), https://portal.gdc.cancer.gov/projects/TCGA-DLBC
26. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2097–2106 (2017)
27. Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks? In: International Conference on Learning Representations (2019)
28. Zhou, Y., Graham, S., Koohbanani, N.A., Shaban, M., Heng, P.A., Rajpoot, N.: Cgc-net: Cell graph convolutional network for grading of colorectal cancer histology images. In: The IEEE International Conference on Computer Vision (ICCV) Workshops (2019)