ORCGT: Ollivier-Ricci Curvature-based Graph Model for Lung STAS Prediction

Min Cen[1], Zheng Wang[2], Zhenfeng Zhuang[2], Hong Zhang[3], Dan Su[4], Zhen Bao[4], Weiwei Wei[5], Baptiste Magnier[6,7], Lequan Yu[8], and Liansheng Wang[2(✉)]

[1] School of Artificial Intelligence and Data Science, University of Science and Technology of China, Hefei, China
cenmin0127@mail.ustc.edu.cn
[2] School of Informatics, Xiamen University, Xiamen, China
{zhengwang, zhuangzhenfeng}@stu.xmu.edu.cn, lswang@xmu.edu.cn
[3] School of Management, University of Science and Technology of China, Hefei, China
zhangh@ustc.edu.cn
[4] Zhejiang Cancer Hospital, Hangzhou, China
{sudan, baozhen}@zjcc.org.cn
[5] Medicine AI Lab, Clinvita-Intelligent Pathology Center, Shanghai, China
wei01@clinvita.vip
[6] EuroMov Digital Health in Motion, Univ Montpellier, IMT Mines Ales, Ales, France
baptiste.magnier@mines-ales.fr
[7] Service de Médecine Nucléaire, Centre Hospitalier Universitaire de Nîmes, Université de Montpellier, Nîmes, France
[8] Department of Statistics and Actuarial Science, The University of Hong Kong, Pok Fu Lam, Hong Kong SAR.China
lqyu@hku.hk

Abstract. Tumor Spread Through Air Spaces (STAS), identified as a mechanism of invasion, has been substantiated by multiple studies to be associated with lower survival rates, underscoring its significant prognostic implications. In clinical practice, pathological diagnosis is regarded as the gold standard for STAS examination. Nonetheless, manual STAS diagnosis is characterized by labor-intensive and time-consuming processes, which are susceptible to misdiagnosis. In this paper, we attempt for the first time to identify the underlying features from histopathological images for the automatic prediction of STAS. Existing deep learning-based methods usually produce undesirable predictive performance with poor interpretability for this task, as they fail to identify small tumor cells spread around the main tumor and their complex correlations. To address these issues, we propose a novel Ollivier-Ricci Curvature-based Graph model for STAS prediction (ORCGT), which utilizes the information from the major tumor margin to improve both the accuracy and interpretability. The model first extracts the major tumor margin by a tumor density map with minimal and coarse annotations, which enhances the visibility of small tumor regions to the model. Then, we develop a

M. Cen and Z. Wang——Contributed equally.

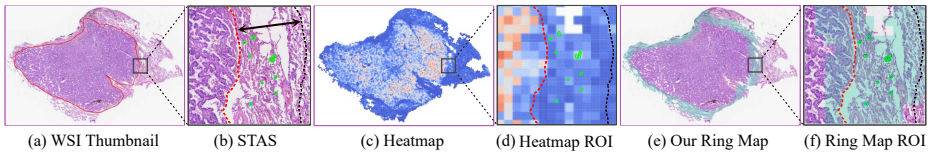

Fig. 1. (a)-(b): The STAS annotation. The red line in WSI is the annotation of major tumor margin and the green circle (very small) is the annotations of STAS; (c)-(d): Attention heatmap from ABMIL [8] cannot highlight the critical tumor margin region; and (e)-(f): The extracted ring map of ORCGT.

Pool-Refined Ollivier-Ricci Curvature-based module to enable complex interactions between patches regardless of long distances and reduce the negative impact of the over-squashing phenomenon among patches linked by negative curvature edges. Extensive experiments conducted on our collected dataset demonstrate the effectiveness and interpretability of the proposed approach for predicting lung STAS. Our code is available at https://github.com/zhengwang9/ORCGT.

Keywords: Histopathology · Tumor Spread Through Air Spaces · Oliver-Ricci curvature · Lung cancer

1 Introduction

Lung cancer is one of the most common malignant tumors worldwide with high mortality and incidence [24], causing increasing emergency in studying the mechanism of lung cancer metastasis. In clinical contexts, Spread Through Air Spaces (STAS) is characterized by micropapillary clusters, solid nests, or individual cells extending within air spaces beyond the boundary of the primary tumor, as depicted in Fig. 1(a) and (b) [7]. STAS as an invasion pattern of pulmonary adenocarcinomas (ADC), is confirmed to be highly correlated to the cancer recurrence rate and the survival of lung ADC patients by statistical analysis [13,3]. Hence, precise identification of STAS plays a crucial role in aiding clinicians in postoperative treatment decisions for patients. Unfortunately, the manual interpretation of WSI is a tedious and time-consuming process with a high risk of misdiagnosis, even with an experienced pathologist [21]. While computer-assisted WSI evaluation could significantly enhance the accuracy, efficiency, and reproducibility of STAS detection, it has only received little attention in the field.

Recently, deep learning techniques have exhibited remarkable performance in WSI analysis [14,10,17]. Several graph-based models have been adapted to model the interrelation among patches and capture global information [2], achieving excellent results in survival analysis tasks. However, we empirically observe that directly applying existing multiple-instance learning-based WSI analysis approaches leads to suboptimal and uninterpretable predictive performance for STAS detection. This could be attributed to the minute size of STAS relative to the main tumor region and its proximity to the main tumor area; see Fig. 1(b).

As shown in Fig. 1(a)-(d), the state-of-the-art method ABMIL [8] pays the most attention to the main tumor region rather than the small tumor regions (i.e. STAS), which neglects the contribution of small tumor regions to the label and hinders modeling the complex interaction among different tumor regions.

To tackle these issues, we propose a novel Oliver-Ricci Curvature-based Graph Model for Lung STAS Prediction (ORCGT). Our approach consists of two innovative modules: Major Tumor Margin Extraction and Ollivier-Ricci Curvature-based Graph Network. In the first module, we restrict the model's input to the ring-like invasive region surrounding the major tumor area (i.e. major tumor margin) (e.g., Fig. 1(e)), which significantly increases the visibility of each small tumor region to the model. In the second module, we use a graph neural network to model complex interactions of patches in the ring-like invasive region and thus enhance the detection of STAS. However, interactions between patches in the graph are hindered by both the long distances among patches and the over-squashing phenomenon during message passing. For the long-distance issue, a hierarchical graph pooling strategy is used to encourage long-distance patch interactions by dynamically shrinking the graph scale. For the over-squashing issue, we introduce Oliver-Ricci Curvature (ORC) to mitigate the distortion of interactions between patches. Concretely, since the patch interactions will be severely distorted between nodes with edges of negative curvature and such edges commonly exist in the constructed graph from each WSI (*e.g.*, Fig 4(d)), the ORC-based aggregation helps restore interactions by adaptively adjusting aggregation weights based on the curvature values.

Our contributions can be summarized as follows: (a) We develop a novel ORCGT framework to predict STAS in WSI that only needs limited coarse annotations. To the best of our knowledge, our work is the first one using deep learning to predict STAS in histological images. (b) We propose to restrict the model's attention only to the ring-like major tumor margin, leading to sound interpretability of model predictions on each WSI. (c) We enhance the interaction among patches by respectively using the ORC to handle distorted interactions and pooling layers to model long-distance interactions, whose effectiveness is carefully validated in our ablation study. (d) Extensive experiments on our collected dataset show the advantage of our proposed framework in WSI-based STAS prediction.

2 Methodology

Fig. 2 illustrates the framework of our proposed ORCGT, composed of two modules: Major Tumor Margin Extraction Module and Ollivier-Ricci Curvature-based Graph Module. Firstly, our proposed Tumor Density Map segmentation module is employed to delineate the major tumor margin. Subsequently, plaques within this margin contribute to graph construction, using their features and coordinates. Additionally, the ORC is integrated to capture the manifold characteristics of the graph. A Pool-Refined Curvature-based graph layer (PRCG) is introduced to improve the prediction of STAS.

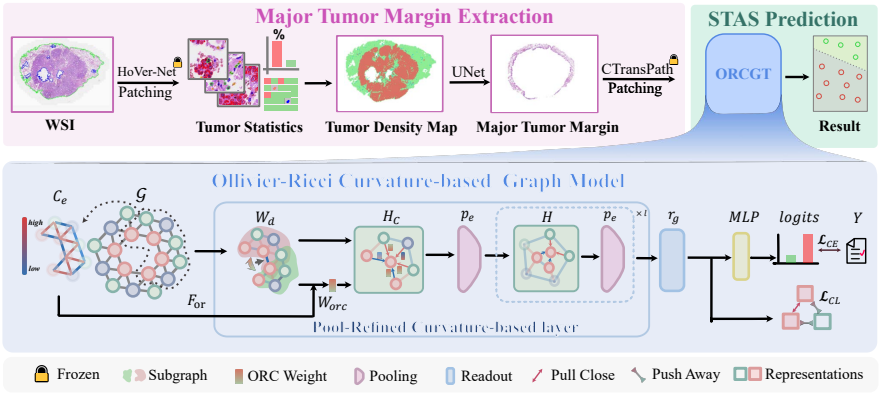

Fig. 2. Overview of our proposed ORCGT, which consists of two modules: Major Tumor Margin Extraction Module and Ollivier-Ricci Curvature-based Graph Module.

2.1 Major Tumor Margin Extraction

We initiate our approach by delineating clinically significant regions along the major tumor boundary within WSI. Initially, we employ a pretrained HoVer-Net[6] to classify tumor patches based on their cell count. To refine the delineation of the major tumor boundary, we identify patches as false positives if more than six of their 8-adjacent patches exhibit negative classification results, subsequently excluding them from the major tumor region. Subsequently, we organize the classification results of all patches within the WSI based on their coordinates, forming a slide-level tumor density map. This density map serves as input for training a tumor region segmentation network to derive the mask for the major tumor region. Finally, adhering to the pathological definition of STAS [7], we identified critical regions: the region of the major tumor margin containing STAS by extending the boundary of the major tumor region inward by n_1 patches and outward by n_2 patches.

2.2 Ollivier-Ricci Curvature-based Graph Construction

Specifically, after the extraction of the major tumor margin, patches within the major tumor margin are treated as nodes V in the ring graph $G{=}(V, E)$. The node features $V{=}(X_1, X_2, \ldots, X_n) \in \mathbb{R}^{n\times d}$ are obtained by pretrained CtransPath model [23]. n is the number of extracted patches and d is the dimension of features. Then the edges are constructed by approximate k nearest neighbor through Hierarchical NSW [12]. All edges $E = \{e_{ij}\}$ are undirected. To model the transmission of dependencies between nodes, we further calculate the ORC for all edges of G. The ORC of edge e_{ij} between node i and node j can be formulated as:$R_{ij} = 1 - \frac{W_D(P_i,P_j)}{dist(i,j)}$, where $W_D(P_i, P_j)$ means the Wasserstein distance [15] between two probability distributions P_i and P_j of node i and node j. $dist(\cdot,\cdot)$ is the Euclidean distance. The probability distribution of node i is defined as

$$P_x(x_i) = \begin{cases} \alpha & \text{if } x_i = x \\ (1-\alpha)/d(x) & \text{if } x_i \in \mathcal{N}(x), \text{ where } \mathcal{N}(x) \text{ the set of adjacent nodes} \\ 0 & \text{otherwise} \end{cases}$$

of node x and $\mathrm{d}(\cdot)$ indicates the degree of nodes, α is a hyperparameter of the probability distribution. Then we can get the Curvature $C = \{R_{ij}\}$ of Graph G. Finally, the curvature graph $G^c = (V, E, C)$ is constructed for each WSI.

2.3 Ollivier-Ricci Curvature-based Graph Network

We propose a novel Ollivier-Ricci Curvature-based Graph Network for predicting STAS status, with a specific focus on the Ring Map. This network addresses the presence of bottleneck regions, which often lead to over-squashing and distorted interactions among patches.

Pool-Refined Curvature-based Graph Layer. Due to the fact that negative ORC typically indicates over-squashing phenomena [22]. We construct an edge weight function that is inversely related to negative ORC. Therefore, we form the curvature vector following [19], formally:

$$F_{\mathrm{or}}(ij) = \left(\frac{1+\mathbf{e}^{-R_{ij}}}{2}, \frac{1+\mathbf{e}^{-2\times R_{ij}}}{2}, \ldots, \frac{1+\mathbf{e}^{-k\times R_{ij}}}{2} \right)^T, \tag{1}$$

where R_{ij} represents the ORC of the edge between node i and node j. Then the edge's curve-based weight can be obtained through $W_{orc}(ij) = W^T F_{\mathrm{or}}(ij) + b$. This curvature weight $W_{orc}(ij)$ is incorporated into the original GCN to form the Curvature-based Graph layer H_c, in which the node features of the layer l can be computed by the following formula:

$$X_x^l = \sigma\left(\sum_{y \in \mathcal{N}(x)\cup\{x\}} \frac{1}{\sqrt{\mathrm{d}(x)}\sqrt{\mathrm{d}(y)}} W_{orc} W_{\mathrm{GCN}} X_y^{l-1} \right), \tag{2}$$

X_x^l denotes the feature of node x after l-th layer, σ the activation function. Then an Edge Pooling Layer [4] was adopted to downsize G^c, while maintaining its pivotal structural and feature information, which also assists local GCN models in learning long-distance dependencies more quickly and facilitates the recognition of the tiny spread tumor regions. Subsequently, we apply a standard GCN layer H to further refine the node features, followed by an additional pooling operation for enhanced feature aggregation.

Finally, the proposed whole Pool-Refined Curvature-based Graph (PRCG) convolution layer denoted $\mathcal{G}_{PRCG}$ is formulated as:

$$\begin{aligned} G^l(V^l, E^l, C^l) &= \mathcal{G}_{PRCG}(G^{l-1}(V^{l-1}, E^{l-1}, C^{l-1})) \\ &= (Pe(H(Pe(H_c(G^{l-1}(V^{l-1}, E^{l-1}, C^{l-1}))))), \end{aligned} \tag{3}$$

where $G^l(V^l, E^l, C^l)$ represents the Graph after the l-th layer. Our Ollivier-Ricci Curvature-based Graph Network is made up of multiple PRCG layers.

After obtaining the aggregated graph yield by our PRCG layers, we learn a global attention-based pooling layer r_g to adaptively compute a weighted sum of all node features $V^l = (X_1^l, X_2^l, \ldots, X_n^l)$ obtained from $G^l(V^l, E^l, C^l)$ to be aggregated into the slide-level feature. Finally, an MLP classifier is used to get the predictive logits of STAS.

Training Strategy. Due to the class imbalance present in the dataset, we employed two loss functions to supervise the learning process of our model: weighted cross-entropy loss $\mathcal{L}_{\tilde{\text{CE}}}$ and supervised contrastive loss $\mathcal{L}_{\text{CL}}$:

$$\begin{cases} \mathcal{L}_{\tilde{\text{CE}}} = -\dfrac{1}{M}\displaystyle\sum_{i=1}^{M} w_i(y_i \ln(\hat{y}_i) + (1-y_i)\ln(1-\hat{y}_i)), \\ \mathcal{L}_{\text{CL}} = -\displaystyle\sum_{i=1}^{M} \dfrac{1}{M_{y_i}-1}\sum_{j=1}^{M} l_{i\neq j} l_{y_i=y_j} \ln\left[\dfrac{\mathbf{e}^{(s_{i,j}/t)}}{\mathbf{e}^{(s_{i,j}/t)} + \sum_{k=1}^{M} l_{y_i\neq y_k}\mathbf{e}^{(s_{i,k}/t)}}\right], \end{cases} \tag{4}$$

where M denotes the mini-batch size, y_i and y_j represent the labels of the anchor sample i and sample j, respectively. M_{y_i} denotes the count of samples with label y_i in a mini-batch. $\hat{y}_i$ is the predicted probability, w_i is the weight assigned to the i^{th} sample. $l_{i\neq j}$ is a similar indicator function. $s_{i,j}$ is the cosine similarity between the sample i and j. Considering $\lambda_{\tilde{CE}}$ and λ_{CL} as loss weights, the overall loss function is defined as $\mathcal{L} = \lambda_{\tilde{\text{CE}}} \cdot \mathcal{L}_{\tilde{\text{CE}}} + \lambda_{\text{CL}} \cdot \mathcal{L}_{\text{CL}}$.

3 Experiments

Datasets Our dataset is collected containing 284 pathologically confirmed lung adenocarcinoma H&E-stained WSIs, of which 208 are classified as negative (without STAS), and 76 as positive (with STAS). There are 23 slides annotated with rough major tumor boundary only for the training of major tumor region segmentation network. We use five-fold cross-validation for evaluation and further split the training set into two groups randomly: training set (75%) and validation set (25%). We adopt 5 evaluation metrics, including the area under the receiver operating characteristic curve (AUROC), the area under the precision-recall curve (AUPRC), balanced accuracy (Acc) [1], as well as the F1 score (F1_{STAS}) and Recall (Recall_{STAS}) for the STAS category. The best model for the validation set is used to evaluate the testing set.

Table 1. Performance comparison of STAS prediction using five-fold cross validation on our collected dataset.

Method	AUROC	AUPRC	Acc	F1_{STAS}	Recall_{STAS}
CLAM-SB [11]	0.701	0.501	0.730	0.326	0.264
ABMIL [8]	0.712	0.508	**0.732**	0.340	0.276
TransMIL [18]	0.601	0.401	0.692	0.200	0.224
Exphormer [20]	0.615	0.392	0.608	0.388	0.488
PatchGCN [2]	0.708	0.501	0.700	0.422	0.500
ORCGT (Ours)	**0.753**	**0.556**	0.724	**0.502**	**0.512**

Table 2. Ablation analysis on our collected dataset.

Method	AUROC	AUPRC	Acc	$F1_{STAS}$	$Recall_{STAS}$
ORCGT (Ours)	**0.753**	**0.556**	0.724	**0.502**	0.512
w/o Ring Map	0.730	0.474	0.720	**0.502**	0.528
w/o Curvature	0.738	0.538	**0.740**	0.494	0.514
w/o Pooling	0.692	0.521	0.684	0.484	**0.584**
w/o contrasive loss	0.714	0.500	0.728	0.438	0.420

Experimental Setup The proposed framework was implemented using PyTorch [16] and PyTorch Geometric [5] on a workstation equipped with 8 NVIDIA GeForce RTX 3090 GPUs. All slides were partitioned into non-overlapping patches of size 256px ×256px at a magnification of 20×. The feature dimension for each patch extracted by CtransPath [23] is 1×768. Due to data imbalance, the weights for the cross-entropy loss were adjusted to 0.35 for samples without STAS and 0.65 for those with STAS. The initial learning rate was set to 0.0002 with Adam optimizer [9], and the batch size was fixed at 4 for 120 epochs. n_1 and n_2 are set to 5 and 10 experimentally (see Section 2.1). The loss weights $\lambda_1 = \lambda_2 = 1$.

Comparison with Other Methods We first evaluated the five-fold cross validation experimental results of our proposed method with two groups of state-of-the-art WSI analysis methods: (1) attention-based algorithms including: CLAM-SB [11], ABMIL [8], TransMIL [18], and (2) graph-based algorithms including: PatchGCN [2], Exphormer [20]. Table 1 shows the comparison results on our dataset. In general, our method achieves an outstanding result in all five evaluation metrics of predicting STAS. It can be found that our proposed method outperforms all the compared algorithms with the best AUROC of 0.753, AUPRC of 0.556, and $Recall_{STAS}$ of 0.512. Compared to other algorithms, ORCGT improves at least 4.8% in classification AUPRC, which is a particularly significant metric in data imbalanced situations similar to ours.

In addition, we achieved better results in $F1_{STAS}$ and $Recall_{STAS}$ which is regarded with importance.

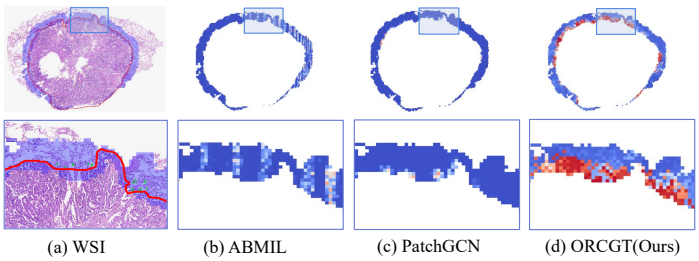

Fig. 3. Attention heatmaps comparsion for variable methods: ABMIL, PatchGCN and ORCGT. The red line in (a) WSI is the annotation of major tumor margin and the green circle (very small) is the annotations of STAS.

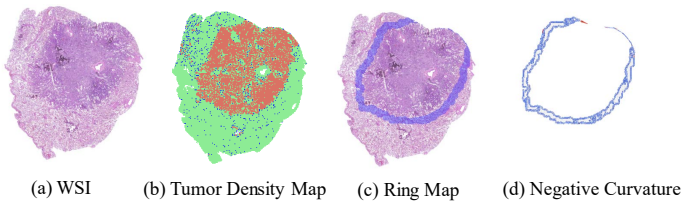

Fig. 4. Major Margin Extraction and Curvature Visualization. In (b), red points denote the predicted major tumor region, blue points represent potential STAS patches, which may be STAS or isolated cancer within the major tumor area patches, while green points denote normal patches. In (c), the blue ring is the Ring Map.

Ablation Study To investigate the effects of four ORCGT key components, we further conduct an ablation study. Table 2 reports the results. In first row, we used all cropped patches in WSIs instead of the patches in Ring Maps; in the second row, we replaced the PRCG with a normal graph convolution layer. The third row removes all pool layers from our model. And the last row, we removed the contrasive loss from our training strategy and only use the cross-entropy loss. It's worth noting that AUPRC has decreased by 8.2% with all patches, demonstrating the importance of major tumor margin extraction. The results of the ablation analysis demonstrate that all the modules employed in our study effectively enhanced the predictive performance of the model to some degree.

Attention Map Visualization Fig. 3 depicts an H&E-stained WSI with STAS annotations and its Ring Map, along with attention heatmaps generated by three different methods: Fig. 3(b) ABMIL; Fig. 3(c) PatchGCN; Fig. 3(d) ORCGT (Ours) with patches in the Ring Map. In addition to our proposed ORCGT, these two models exhibit good experimental evaluation metrics. Therefore, we choose to visualize the attention heatmaps of these two models to compare their interpretability with our approach. The sample depicted in Fig. 3 exhibits the presence of STAS and is accurately predicted by ORCGT. In the attention map, it is evident that both ABMIL and PatchGCN fail to capture the relationship between the patches of the main tumor area and STAS, whereas our ORCGT successfully does so. In the attention heatmap generated by ORCGT, regions with higher attention encompass the STAS areas.

Major Tumor Margin Extraction and Curvature-based Ring Map Visualization Fig. 4(b) illustrates reconstructed tumor density map color-coded based on tumor types, the red areas selected through local tumor density appear to align more closely with the contours of the main tumor region in the H&E slide compared to the cancerous regions obtained directly from HoVer-Net [6]. Fig. 4 (c) presents the Ring Map prediction results, demonstrating relatively accurate ring predictions that encompass the STAS region. Fig. 4 (d) shows that the edges

of graphs with negative curvature, suggesting that emphasizing the inner region of the circle helps to capture the long-distance relationship between the major tumor margin and STAS and reduce distorted interactions across patches.

4 Conclusion

We propose a novel framework called the Oliver-Ricci Curvature-based Graph Model for Lung STAS Prediction (ORCGT) with histological WSIs. Unlike existing methods that fail to capture STAS, our approach restricts the model's input to the ring-like invasive region surrounding the major tumor area and designed a PRCG layer to simultaneously improve the detection of STAS and model complex interactions among major tumor margin and STAS. Experimental results demonstrate that our method achieves superior classification performance and interpretability, with each module proving effective. This has significant implications for improving the efficiency of STAS diagnosis and reducing costs.

Acknowledgments. This work was supported by National Natural Science Foundation of China (Grant No. 62371409, 12171451), the Research Grants Council of Hong Kong (T45-401/22-N and 27206123), HongKong Innovation and Technology Fund (ITS/274/22), and Anhui Center for Applied Mathematics.

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

References

1. Brodersen, K.H., Ong, C.S., Stephan, K.E., Buhmann, J.M.: The balanced accuracy and its posterior distribution. In: 2010 20th international conference on pattern recognition. pp. 3121–3124. IEEE (2010)
2. Chen, R.J., Lu, M.Y., Shaban, M., Chen, C., Chen, T.Y., Williamson, D.F., Mahmood, F.: Whole slide images are 2d point clouds: Context-aware survival prediction using patch-based graph convolutional networks. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24. pp. 339–349. Springer (2021)
3. Dai, C., Xie, H., Su, H., She, Y., Zhu, E., Fan, Z., Zhou, F., Ren, Y., Xie, D., Zheng, H., et al.: Tumor spread through air spaces affects the recurrence and overall survival in patients with lung adenocarcinoma $>$ 2 to 3 cm. Journal of thoracic oncology **12**(7), 1052–1060 (2017)
4. Diehl, F.: Edge contraction pooling for graph neural networks. arXiv preprint arXiv:1905.10990 (2019)
5. Fey, M., Lenssen, J.E.: Fast graph representation learning with pytorch geometric. arXiv preprint arXiv:1903.02428 (2019)
6. 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)

7. Han, Y.B., Kim, H., Mino-Kenudson, M., Cho, S., Kwon, H.J., Lee, K.R., Kwon, S., Lee, J., Kim, K., Jheon, S., et al.: Tumor spread through air spaces (stas): prognostic significance of grading in non-small cell lung cancer. Modern Pathology **34**(3), 549–561 (2021)
8. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
9. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
10. Li, F., Yang, Y., Wei, Y., He, P., Chen, J., Zheng, Z., Bu, H.: Deep learning-based predictive biomarker of pathological complete response to neoadjuvant chemotherapy from histological images in breast cancer. Journal of translational medicine **19**, 1–13 (2021)
11. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature biomedical engineering **5**(6), 555–570 (2021)
12. Malkov, Y.A., Yashunin, D.A.: Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. IEEE transactions on pattern analysis and machine intelligence **42**(4), 824–836 (2018)
13. Mino-Kenudson, M.: Significance of tumor spread through air spaces (stas) in lung cancer from the pathologist perspective. Translational Lung Cancer Research **9**(3), 847 (2020)
14. Niehues, J.M., Quirke, P., West, N.P., Grabsch, H.I., van Treeck, M., Schirris, Y., Veldhuizen, G.P., Hutchins, G.G., Richman, S.D., Foersch, S., et al.: Generalizable biomarker prediction from cancer pathology slides with self-supervised deep learning: A retrospective multi-centric study. Cell Reports Medicine **4**(4) (2023)
15. Panaretos, V.M., Zemel, Y.: Statistical aspects of wasserstein distances. Annual review of statistics and its application **6**, 405–431 (2019)
16. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems **32** (2019)
17. Sha, L., Osinski, B.L., Ho, I.Y., Tan, T.L., Willis, C., Weiss, H., Beaubier, N., Mahon, B.M., Taxter, T.J., Yip, S.S.: Multi-field-of-view deep learning model predicts nonsmall cell lung cancer programmed death-ligand 1 status from whole-slide hematoxylin and eosin images. Journal of pathology informatics **10**(1), 24 (2019)
18. 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)
19. Shen, C., Ding, P., Wee, J., Bi, J., Luo, J., Xia, K.: Curvature-enhanced graph convolutional network for biomolecular interaction prediction. Computational and Structural Biotechnology Journal (2024)
20. Shirzad, H., Velingker, A., Venkatachalam, B., Sutherland, D.J., Sinop, A.K.: Exphormer: Sparse transformers for graphs. arXiv preprint arXiv:2303.06147 (2023)
21. Theissig, F., Kunze, K., Haroske, G., Meyer, W.: Histological grading of breast cancer: interobserver, reproducibility and prognostic significance. Pathology-Research and Practice **186**(6), 732–736 (1990)
22. Topping, J., Di Giovanni, F., Chamberlain, B.P., Dong, X., Bronstein, M.M.: Understanding over-squashing and bottlenecks on graphs via curvature. arXiv preprint arXiv:2111.14522 (2021)

23. Wang, X., Yang, S., Zhang, J., Wang, M., Zhang, J., Yang, W., Huang, J., Han, X.: Transformer-based unsupervised contrastive learning for histopathological image classification. Medical image analysis **81**, 102559 (2022)
24. Wong, M.C., Lao, X.Q., Ho, K.F., Goggins, W.B., Tse, S.L.: Incidence and mortality of lung cancer: global trends and association with socioeconomic status. Scientific reports **7**(1), 14300 (2017)