PAMIL: Prototype Attention-based Multiple Instance Learning for Whole Slide Image Classification

Jiashuai Liu[1], Anyu Mao[1], Yi Niu[1], Xianli Zhang[4], Tieliang Gong[1], Chen Li[1] (✉), and Zeyu Gao[2,3] (✉)

[1] School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
[2] Department of Oncology, University of Cambridge, UK
[3] CRUK Cambridge Centre, University of Cambridge, UK
[4] Interactive Entertainment Group Global, Tencent, Singapore
ljs1007599414@stu.xjtu.edu.cn

Abstract. Digital pathology images are not only crucial for diagnosing cancer but also play a significant role in treatment planning, and research into disease mechanisms. The multiple instance learning (MIL) technique provides an effective weakly-supervised methodology for analyzing gigapixel Whole Slide Image (WSI). Recent advancements in MIL approaches have predominantly focused on predicting a singular diagnostic label for each WSI, simultaneously enhancing interpretability via attention mechanisms. However, given the heterogeneity of tumors, each WSI may contain multiple histotypes. Also, the generated attention maps often fail to offer a comprehensible explanation of the underlying reasoning process. These constraints limit the potential applicability of MIL-based methods in clinical settings. In this paper, we propose a Prototype Attention-based Multiple Instance Learning (PAMIL) method, designed to improve the model's reasoning interpretability without compromising its classification performance at the WSI level. PAMIL merges prototype learning with attention mechanisms, enabling the model to quantify the similarity between prototypes and instances, thereby providing the interpretability at instance level. Specifically, two branches are equipped in PAMIL, providing prototype and instance-level attention scores, which are aggregated to derive bag-level predictions. Extensive experiments are conducted on four datasets with two diverse WSI classification tasks, demonstrating the effectiveness and interpretability of our PAMIL. The code is available at https://github.com/Jiashuai-Liu/PAMIL

Keywords: Multiple instance learning · Prototype learning · Whole slide image classification.

1 Introduction

The rise of digital pathology has driven substantial advancements in applying artificial intelligence to analyze Whole Slide Images (WSI). Nevertheless, the necessity for expert pathologists to annotate gigapixel images presents hurdles for fully supervised methods in thoroughly processing WSIs [9]. As a result, Multiple Instance Learning (MIL) has gained prominence [4], employing adeptly trained feature extractors and aggregators to merge instance-level (refers to image patch) information, thereby facilitating predictions for the entire bag (refers to WSI).

Among MIL-based WSI classification methods, incorporating attention mechanisms into the MIL framework has demonstrated remarkable classification capabilities [10,16,20]. However, the application of these methods in real clinical scenarios is hampered by two main limitations. On the one hand, they lack the interpretability that is recognized by pathologists [19,26,25]. While attention-based MIL provides a form of interpretability by quantifying the importance of each instance through the attention network, it fails to disclose the humanly comprehensible reasons behind the high attention scores of certain instances. On the other hand, they fall short of supporting complex yet common pathological diagnostic tasks. Existing MIL methods are mostly suitable for tasks involving one label per WSI, while a single tumor slide often displays multiple histopathological phenotypes, due to the heterogeneity of tumors.

Fortunately, incorporating prototype learning with MIL seems to be a possible solution. ProtoMIL [19] attempted to enhance the interpretability of the inference process by prototype learning, building upon case-based reasoning akin to human thinking processes. However, the initialization of ProtoMIL's prototypes is random within the model, and the design of loss function does not ensure alignment between the prototypes and instances distributions. In addition, the prototypes in ProtoMIL are predefined with specific categories, limiting their ability to adapt and learn in multi-label scenarios.

To address these limitations, we propose a Prototype Attention-based Multiple Instance Learning (PAMIL), which embeds prototypes into the attention mechanism to quantify the similarity between prototypes and instances. The prototype, serving as a globally shared parameter, offers a global case-based interpretation of the model inference. Subsequently, we design two branches, prototype representation and instance representation, to perform feature aggregation and collaboratively derive bag-level predictions. The two branches assign weights to prototypes and instances respectively, providing prototype and instance-level interpretation. Different from ProtoMIL, our prototypes don't have predefined categories, which enables its adaptability to both multi-label and multi-class classification tasks. This flexibility allows prototypes to incorporate representations of different subtypes during training. Lastly, we devised an optimization strategy and incorporated regularization terms to ensure the stability of the model training process. The efficacy of PAMIL was assessed across four datasets, covering both multi-label and multi-class classification tasks, demonstrating that PAMIL achieved state-of-the-art WSI classification performance across all datasets, simultaneously offering an extensively comprehensible reasoning process.

2 Related Work

2.1 Multiple Instance Learning

Classification of WSI via MIL can be categorized as instance-level and embedding-level approaches. The instance-level approaches predict individual instances and aggregate these predictions to yield a bag-level prediction [27,1,7]. However, these approaches inadequately consider the inter-instance relationships within the bag, resulting in imprecise final predictions. The embedding-level approaches commence by extracting instance-level features, which are subsequently aggregated into bag-level features using various strategies [24,13,8,6]. To improve bag-level performance and interpretability in embedding-based MIL, Ilse *et al.* [10] introduced an attention mechanism. This approach utilizes attention scores to assign significant importance weights to instances, thereby aiding in the aggregation of bag-level features through a selective summarization process. Notably, involving attention mechanism has demonstrated significant efficacy across diverse WSI analysis tasks [16,20,14,29,11]. Furthermore, some studies have integrated the benefits of both embedding-level and instance-level approaches, leveraging their respective strengths to enhance model performance [21,18].

2.2 Prototype Learning

Prototype learning finds extensive application in natural language processing and computer vision, aiming to optimize the feature space by preserving a global prototype. Chen *et al.* [2] proposed a pioneering approach that employs prototype learning to enhance the interpretability of the model inference process. On this basis, the majority of prototype-based interpretability methods [3,5,12,17] model the feature space using several global prototypes, then the similarity between prototypes and local regions is utilized to elucidate the reasoning process behind image recognition. Recently, the combination of prototype learning and MIL has been extensively investigated, which aims to guide the distribution of the instance space [19,28,15,23]. For instance, Rymarczyk *et al.* [19] integrated the concept of ProtoPNet into the ABMIL framework, effectively modeling the human reasoning process in MIL. Concurrently, PMIL [28] employs a dual clustering approach to identify prototypes and incorporates metric learning to refine the feature space of instances, further improving the model's performance. However, previous methods often struggle to balance prototype interpretability with bag-level performance. Some rely solely on prototype-instance similarity for prediction, which is insufficient for complex tasks like multi-label classification. Others focus on guiding the feature space through prototypes but overlook the importance of interpretability. In this paper, we integrate prototypes with the attention mechanism, developing a prototype attention approach. This approach improves interpretability through prototype reasoning while maintaining strong bag-level classification performance.

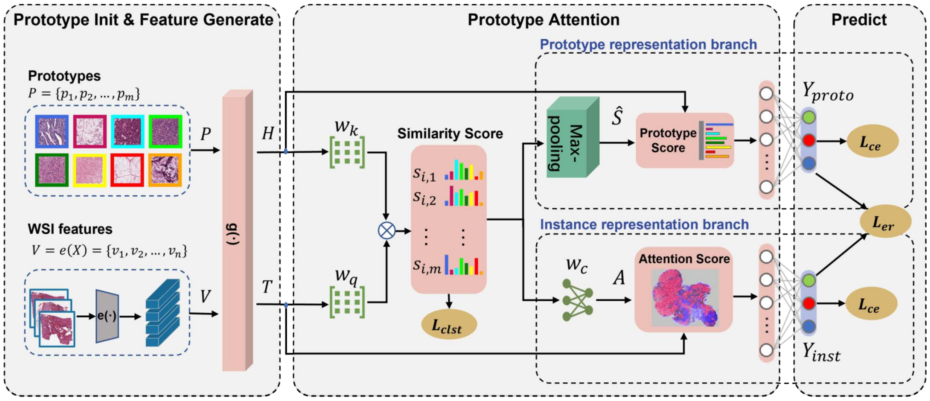

Fig. 1. Overview of PAMIL.

3 Method

3.1 Preformulation

For a WSI dataset, each slide is considered as a bag, and the patches serve as instances. Each bag is denoted as $X = \{x_1, x_2, \cdots, x_n\}$, where x_i represents the i-th instance of bag X, with n instances in total. For multiple instance learning, only the bag-level label is given, represented as $Y = \{Y_1, Y_2, \cdots, Y_k\}$, where $Y_i \in \{0, 1\}$. If at least one instance in the bag belongs to class Y_i, then $Y_i = 1$.

3.2 PAMIL for WSI Classification

The PAMIL framework is illustrated in Fig. 1. Initially, all instances go through a pre-trained encoder $e(\cdot)$ to generate feature embeddings $V = \{v_1, v_2, \ldots, v_n\}$. Next, we incorporate prototypes into the attention mechanism to assess the similarity between prototypes and instances. For feature aggregation, we use two interactive branches for instance and prototype features, respectively. This approach constrains prototypes to be selected from all available patches, thus enhancing the interpretability of the inference process in PAMIL.

Instead of using randomly initialized prototypes [19], we adopt K-means clustering twice on the instance embeddings within all bags, obtaining the initial global prototype embeddings $P = \{p_1, p_2, \cdots, p_m\}$. Both P and V are then processed by a dimensionality reduction layer $g(\cdot)$ to get compressed embeddings H for prototypes and T for instances. Next, we apply cross-attention to derive the similarity matrix $S = \{s_{i,j} | i = 1, \cdots, n, j = 1, \cdots, m\}$ between embeddings v_i and p_j:

$$S = \tanh\left(w_q T\right)^{\top} \times \operatorname{sigmoid}\left(w_k H\right) \qquad (1)$$

where w_q and w_k represent the learnable weight matrices for T and H, respectively. Then, we design two interactive branches that utilize the prototype-level and instance-level representations for generating bag-level predictions, respectively, supporting the corresponding reasoning processes for interpretation.

In the instance representation branch, a learnable weight matrix w_c is used to map the relationship between instances and prototypes S to the relationship between instances and categories A, *i.e.*, $A = w_c \cdot S$. A also can be regarded as the attention matrix for all instances. Additionally, w_c implicitly models the correlation between prototypes and categories, thereby providing category-related guidance for updating prototypes during back-propagation. The bag-level prediction Y_{inst} is generated by aggregating embeddings T with A:

$$Y_{inst} = f_{cls1}\left(\hat{A} \cdot T\right), \tag{2}$$

where $\hat{A} = \mathrm{softmax}(A) \in \mathbb{R}^{n \times c}$ represent the softmax normalized attention matrix, $f_{cls1}(\cdot)$ is a classifier with one hidden layer.

In the prototype representation branch, the similarity matrix S is also used to aggregate compressed prototype embeddings H into another bag-level feature which is then fed into a classifier f_{cls2} to obtain another prediction for this bag:

$$Y_{proto} = f_{cls2}\left(\hat{S} \cdot H\right), \tag{3}$$

where $\hat{S} \in \mathbb{R}^{m \times c}$ represent the similarity matrix S after the max-pooling operation, serving as a metric for prototype-to-bag similarity. It critically functions as the interpretability for each prototype during the inference in this branch. The final prediction of the model is the average of the two probability values.

Finally, three different losses are used for model optimization:

$$\begin{aligned} & L = \lambda_{ce} \cdot L_{ce} + \lambda_{er} \cdot L_{er} + \lambda_{clst} \cdot L_{clst} \\ & \begin{cases} L_{ce} = BCE\left(Y_{inst}, Y\right) + BCE\left(Y_{proto}, Y\right) \\ L_{er} = KL\left(Y_{proto} \| Y_{inst}\right) \\ L_{clst} = \frac{1}{n}\sum_i \max_j\left(\mathrm{softmax}\left(s_{i,j}\right)\right) + \frac{1}{m}\sum_j \max_i\left(\mathrm{softmax}\left(s_{i,j}\right)\right) \end{cases} \end{aligned} \tag{4}$$

where BCE stands for the binary cross-entropy loss function, and KL represents the Kullback-Leibler divergence. L_{er} is the equivalent regularization loss, which ensures consistency between Y_{proto} and Y_{inst}. L_{clst} is the clustering loss, aimed at refining the shared embedding space of instances and prototypes.

Prototype Optimization and Interpretation. We optimized the prototypes in a multi-stage manner similar to [19]. After initialization, the prototypes are refined according to loss functions and are associated with categories via the weight matrix w_c. Subsequently, a prototype projection operation is executed, during which prototypes are substituted with the most similar instances within all bags. This adjustment enables us to interpret the model reasoning process by the similarities between prototypes and instances.

The two branches in PAMIL provide interpretations at both the instance level and the prototype level, both of which can be traced back to the similarity matrix

S. In the instance representation branch, the attention matrix A determines the contribution of each instance to different categories. It can be traced back to the similarity matrix S between instances and each prototype. In the prototype representation branch, $\hat{S}$ signifies the importance of the prototypes to the bag. This importance can be linked to the most significant instance in the bag, that is, the instance closest to the prototype in the embedding space.

4 Experiments

Datasets. The proposed PAMIL was evaluated on four public datasets for multi-class or multi-label WSI classification. For the multi-class classification task, a bag has only one positive category, meaning that at least one of the instances in the bag is of the same category as the bag, and the rest are negative instances. For this task, we derive two datasets from TCGA (The Cancer Genome Atlas): non-small cell lung cancer (NSCLC) and renal cell carcinoma (RCC). For the multi-label classification task, a bag could have multiple labels as positive. Each dimension of the label indicates whether there are corresponding instances in the bag. For this task, we select the Stomach Cancer (STAD) dataset from TCGA and the publicly available SICAPv2 dataset[22], which specifically focuses on prostate Gleason grading. The details are as follows: (1) TCGA-NSCLC consists of 937 slides, specifically 447 slides of Lung Adenocarcinoma (LUAD) and 490 slides of Lung squamous cell carcinoma (LUSC). (2) TCGA-RCC consists of 660 slides, specifically 299 slides of Clear Cell RCC (KIRC), 258 slides of Papillary RCC (KIRP) and 103 slides of chromophobe RCC (KICH). (3) TCGA-STAD consists of 339 slides, specifically 218 slides of Highly Differentiated (HD) , 265 slides of Poorly Differentiated (PD) and 50 slides of Mucinous (Muc). (4) SICAPv2 comprises 18,426 cropped patches obtained from 153 slides labeled with G3, G4, G5, and normal.

Baseline and Evaluation Metrics. The baseline consists of five attention-based MIL methods: CLAM[16], DSMIL[14], DTFDMIL[29], Additive MIL[11], TransMIL[20] and a prototype-based MIL method: ProtoMIL[19]. Most of these methods reached state-of-the-art at the time. To assess the classification performance, we employ accuracy and area under the curve (AUC) scores as evaluation metrics. The accuracy is calculated using a threshold of 0.5 in all experiments. We perform five-fold cross-validation to evaluate our model on all datasets.

4.1 Experiments and Results

Except for SICAPv2, we generate non-overlapping patches from WSIs at 40× magnification with 2048 × 2048 pixels and at 20× magnification with 1024 × 1024 pixels. These patches are then extracted by a pre-trained ResNet-50 to get feature vectors. When initializing the prototypes, we first cluster patch feature vectors of each slide to 10 cluster centers, and then cluster all cluster centers to obtain 8 initialization prototypes. During the training process, we set λ_{ce} to

1, λ_{er} to 0.4 and λ_{clst} to -0.2. The training process utilize the Adam optimizer with a learning rate set to 0.0001. The results of the comparative experiments are presented in Table 1.

Table 1. Comparison Results (presented in %).

Method	TCGA-STAD		SICAPv2		TCGA-NSCLC		TCGA-RCC	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
CLAM	81.26±4.04	80.80±4.32	85.52±3.95	**82.20±4.97**	**90.29±2.23**	81.82±3.62	97.59±1.56	90.02±3.92
Additive MIL	80.30±4.30	74.43±1.42	82.57±4.60	71.90±4.20	86.71±4.10	**82.65±2.80**	95.85±2.88	87.76±4.40
DSMIL	80.33±3.89	75.18±1.63	85.71±6.09	71.87±2.85	88.71±2.80	81.10±1.73	**98.21±0.85**	89.40±1.86
DTFDMIL	76.82±6.08	73.22±2.73	82.03±6.15	68.43±2.62	86.47±2.95	80.18±2.31	95.56±1.66	85.47±3.52
TransMIL	80.49±5.32	75.91±3.04	85.52±6.59	71.97±2.41	88.83±3.39	78.44±4.25	97.42±0.91	90.01±2.68
ProtoMIL	-	-	-	-	70.41±1.28	63.34±8.10	76.26±8.71	59.70±9.07
PAMIL	**84.75±3.55**	**81.66±3.67**	**87.57±3.93**	81.50±2.98	89.85±2.10	81.92±2.56	97.58±1.21	**90.17±4.02**

From Table 1, we can see that compared with all the state-of-the-art methods, PAMIL achieves competitive WSI classification performance. Moreover, compared with traditional prototype-based MIL (ProtoMIL), PAMIL has achieved huge improvements, 20-30% in AUC. Furthermore, for multi-label classification datasets, PAMIL outperforms all other methods in AUC, due to the well-designed prototype optimization strategy. Beyond the competitive bag-level performance, PAMIL can offer a comprehensive explanation of its reasoning process.

4.2 Visualization of Interpretability

To illustrate the interpretability of PAMIL, we visualize the entire inference process. Two selected samples from the TCGA-STAD dataset are shown in Fig. 2. We use similar colors to mark prototypes belonging to the same category, *i.e.*, blue for "HD", red for "PD", green for "Muc", and yellow for "Normal".

We take the first slide as an example, which is predicted to have "HD" and "PD" subtypes. The predictions of both branches could be traced back to the similarity score S, as shown in Fig. 2(c), which is represented as the similarity between patches of this WSI and patches obtained by prototype projection. In the instance representation branch, the attention matrix A represents the importance of each patch to the category, as depicted in Fig. 2(d), and it is derived from the weighted sum of similarity score S with w_c as the weight. Compared with Fig. 2(c), a category correspondence is found between the attention score and the prototype similarity. In the prototype representation branch, the prototype score expresses the importance of the prototypes to this WSI. We choose the prototypes with top-4 prototype scores and show the patches that are most similar in the WSI, as depicted in Fig. 2(b).

4.3 Ablation Study

To further verify the design of each module in the model, we conducted an ablation study. The results are presented in Table 2. We proceeded to ablate

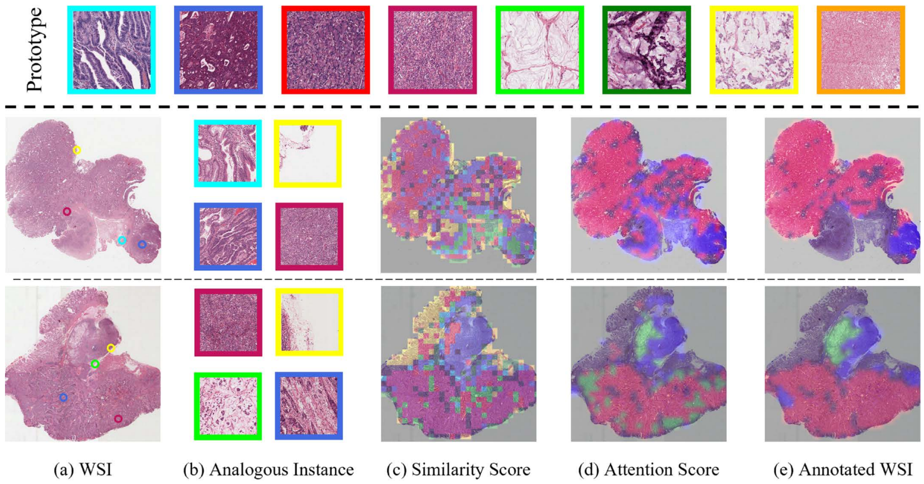

Fig. 2. Visualization of PAMIL reasoning process. Prototypes predicted to be of the same category according to matrix w_c are labeled with similar colors. (a) The original WSI. (b) Patches with the highest prototype similarity, marked in (a) accordingly. (c) The similarity score map, which labels each patch as the category of the most similar prototype. (d) The patch prediction masks from the attention matrix. (e) The annotation masks from the pathologist.

the cluster initialization of the prototype, the instance representation branch (IRB), the prototype representation branch (PRB), as well as the equivalent regularization loss (L_{er}) and clustering loss (L_{clst}).

Table 2. Ablation study for PAMIL (presented in %).

Method	TCGA-STAD		SICAPv2		TCGA-NSCLC		TCGA-RCC	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
w/o Proto Init	81.46±5.27	76.31±6.44	82.79±3.70	73.61±2.01	89.12±2.12	82.33±3.08	97.07±1.14	89.10±4.36
w/o IRB	81.72±4.72	76.46±3.72	80.84±6.72	73.67±7.43	87.83±3.77	81.12±4.27	96.74±1.52	89.72±4.74
w/o PRB	82.95±5.14	74.56±11.85	87.10±5.16	81.45±3.79	89.02±1.43	82.13±2.86	97.29±0.83	89.11±2.94
w/o L_{er}	82.31±4.53	80.34±4.72	84.97±2.94	79.30±3.43	89.29±2.60	81.62±2.69	97.20±1.62	89.13±4.95
w/o L_{clst}	81.16±5.22	76.95±4.53	85.27±5.28	77.92±6.66	89.16±2.71	**82.64±4.52**	96.16±3.18	88.07±6.24
ours	**84.75±3.55**	**81.66±3.27**	**87.57±3.93**	**81.50±2.98**	**89.85±2.10**	81.92±2.56	**97.58±1.21**	**90.32±4.20**

The results show that each module and the loss function designed in PAMIL effectively improve the bag-level prediction performance. The L_{er} loss plays a crucial role in stabilizing the two-branch prediction. Interestingly, without the L_{er} loss, the model's performance deteriorates even compared to the single-branch prediction. Additionally, the results of the ablation experiments demonstrate that the two branches of the model can achieve promising results when making independent predictions. This suggests that both branches can be effectively utilized as separate backbone models for further research endeavors.

5 Conclusion

In this paper, we propose a novel method for MIL that leverages prototype attention across two inference branches. Our method incorporates prototype learning to facilitate case-based interpretability and tackle the complexities of multi-label classification. A key feature of our approach is the ability of our prototypes to encapsulate category information independently of pre-defined categories throughout the training phase. Through experimental analysis of four datasets, we assessed the efficacy of the proposed PAMIL.

Acknowledgements. This work has been supported by the Key Research and Development Program of Ningxia Hui Nationality Autonomous Region under Grant 2022BEG-02025 and 2023BEG02023, the National Nature Science Foundation of China (62106191), the Cancer Research UK Cambridge Centre [CTRQQR-2021-100012], and the Mark Foundation for Cancer Research [RG95043], GE HealthCare.

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

References

1. Campanella, G., Hanna, M.G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K.J., Brogi, E., Reuter, V.E., Klimstra, D.S., Fuchs, T.J.: Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature medicine **25**(8), 1301–1309 (2019)
2. Chen, C., Li, O., Tao, D., Barnett, A., Rudin, C., Su, J.K.: This looks like that: Deep learning for interpretable image recognition. Advances in neural information processing systems **32** (2019)
3. Chen, Z., Bei, Y., Rudin, C.: Concept whitening for interpretable image recognition. Nature Machine Intelligence **2**(12), 772–782 (2020)
4. Couture, H.D., Marron, J.S., Perou, C.M., Troester, M.A., Niethammer, M.: Multiple instance learning for heterogeneous images: Training a cnn for histopathology. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018. pp. 254–262. Springer (2018)
5. Donnelly, J., Barnett, A.J., Chen, C.: Deformable protopnet: An interpretable image classifier using deformable prototypes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10265–10275 (2022)
6. Gao, Z., Hong, B., Li, Y., Zhang, X., Wu, J., Wang, C., Zhang, X., Gong, T., Zheng, Y., Meng, D., et al.: A semi-supervised multi-task learning framework for cancer classification with weak annotation in whole-slide images. Medical Image Analysis **83**, 102652 (2023)
7. Gao, Z., Hong, B., Zhang, X., Li, Y., Jia, C., Wu, J., Wang, C., Meng, D., Li, C.: Instance-based vision transformer for subtyping of papillary renal cell carcinoma in histopathological image. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021. pp. 299–308. Springer (2021)

8. Gao, Z., Mao, A., Wu, K., Li, Y., Zhao, L., Zhang, X., Wu, J., Yu, L., Xing, C., Gong, T., et al.: Childhood leukemia classification via information bottleneck enhanced hierarchical multi-instance learning. IEEE Transactions on Medical Imaging (2023)
9. Hou, L., Samaras, D., Kurc, T.M., Gao, Y., Davis, J.E., Saltz, J.H.: Patch-based convolutional neural network for whole slide tissue image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2424–2433 (2016)
10. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International Conference on Machine Learning. pp. 2127–2136. PMLR (2018)
11. 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)
12. Kim, E., Kim, S., Seo, M., Yoon, S.: Xprotonet: Diagnosis in chest radiography with global and local explanations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15719–15728 (2021)
13. Li, B., Li, Y., Eliceiri, K.W.: 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. pp. 14318–14328 (2021)
14. Li, B., Li, Y., Eliceiri, K.W.: 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. pp. 14318–14328 (2021)
15. Li, X., Yang, B., Chen, T., Gao, Z., Huang, M.: Promil: A weakly supervised multiple instance learning for whole slide image classification based on class proxy. Expert Systems with Applications **238**, 121800 (2024)
16. 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)
17. Nauta, M., Schlötterer, J., van Keulen, M., Seifert, C.: Pip-net: Patch-based intuitive prototypes for interpretable image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2744–2753 (2023)
18. Qu, L., Luo, X., Liu, S., Wang, M., Song, Z.: Dgmil: Distribution guided multiple instance learning for whole slide image classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 24–34. Springer (2022)
19. Rymarczyk, D., Pardyl, A., Kraus, J., Kaczyńska, A., Skomorowski, M., Zieliński, B.: Protomil: Multiple instance learning with prototypical parts for whole-slide image classification. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 421–436. Springer (2022)
20. 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)
21. Sharma, Y., Shrivastava, A., Ehsan, L., Moskaluk, C.A., Syed, S., Brown, D.: Cluster-to-conquer: A framework for end-to-end multi-instance learning for whole slide image classification. In: Medical Imaging with Deep Learning. pp. 682–698. PMLR (2021)

22. Silva-Rodríguez, J., Colomer, A., Sales, M.A., Molina, R., Naranjo, V.: Going deeper through the gleason scoring scale: An automatic end-to-end system for histology prostate grading and cribriform pattern detection. Computer methods and programs in biomedicine **195**, 105637 (2020)
23. Vu, Q.D., Rajpoot, K., Raza, S.E.A., Rajpoot, N.: Handcrafted histological transformer (h2t): Unsupervised representation of whole slide images. Medical Image Analysis **85**, 102743 (2023)
24. Wang, X., Chen, H., Gan, C., Lin, H., Dou, Q., Tsougenis, E., Huang, Q., Cai, M., Heng, P.A.: Weakly supervised deep learning for whole slide lung cancer image analysis. IEEE transactions on cybernetics **50**(9), 3950–3962 (2019)
25. Wu, J., He, K., Mao, R., Li, C., Cambria, E.: Megacare: Knowledge-guided multiview hypergraph predictive framework for healthcare. Information Fusion **100**, 101939 (2023)
26. Wu, J., Zhang, R., Gong, T., Liu, Y., Wang, C., Li, C.: Bioie: Biomedical information extraction with multi-head attention enhanced graph convolutional network. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 2080–2087. IEEE (2021)
27. Xu, Y., Zhu, J.Y., Chang, E., Tu, Z.: Multiple clustered instance learning for histopathology cancer image classification, segmentation and clustering. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 964–971. IEEE (2012)
28. Yu, J.G., Wu, Z., Ming, Y., Deng, S., Li, Y., Ou, C., He, C., Wang, B., Zhang, P., Wang, Y.: Prototypical multiple instance learning for predicting lymph node metastasis of breast cancer from whole-slide pathological images. Medical Image Analysis **85**, 102748 (2023)
29. Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S.E., Zheng, Y.: Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18802–18812 (2022)