IIB-MIL: Integrated Instance-Level and Bag-Level Multiple Instances Learning with Label Disambiguation for Pathological Image Analysis

Qin Ren[1,2], Yu Zhao[1(✉)], Bing He[1(✉)], Bingzhe Wu[1], Sijie Mai[3], Fan Xu[1,4], Yueshan Huang[1,5], Yonghong He[2], Junzhou Huang[6], and Jianhua Yao[1(✉)]

[1] AI Lab, Tencent, Shenzhen 518000, China
jianhuayao@tencent.com
[2] Shenzhen International Graduate School, Tsinghua University, Shenzhen 518071, China
[3] School of Electronic and Information Technology, Sun Yat-sen University, Guangzhou 510006, China
[4] ShanghaiTech University, Shanghai 201210, China
[5] Shanghai Jiao Tong University, Shanghai 200240, China
[6] University of Texas at Arlington, Arlington, TX 76019, USA

Abstract. Digital pathology plays a pivotal role in the diagnosis and interpretation of diseases and has drawn increasing attention in modern healthcare. Due to the huge gigapixel-level size and diverse nature of whole-slide images (WSIs), analyzing them through multiple instance learning (MIL) has become a widely-used scheme, which, however, faces the challenges that come with the weakly supervised nature of MIL. Conventional MIL methods mostly either utilized instance-level or bag-level supervision to learn informative representations from WSIs for downstream tasks. In this work, we propose a novel MIL method for pathological image analysis with integrated instance-level and bag-level supervision (termed IIB-MIL). More importantly, to overcome the weakly supervised nature of MIL, we design a label-disambiguation-based instance-level supervision for MIL using Prototypes and Confidence Bank to reduce the impact of noisy labels. Extensive experiments demonstrate that IIB-MIL outperforms state-of-the-art approaches in both benchmarking datasets and addressing the challenging practical clinical task. The code is available at https://github.com/TencentAILabHealthcare/IIB-MIL.

Keywords: computational pathology · multi-instance learning · label disambiguation · prototype · confidence bank

Q. Ren and Y. Zhao—Equally-contributed authors.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43987-2_54.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14225, pp. 560–569, 2023.
https://doi.org/10.1007/978-3-031-43987-2_54

1 Introduction

Pathology is widely recognized as the gold standard for disease diagnosis [15]. As the demand for intelligently pathological image analysis continues to grow, an increasing number of researchers have paid attention to this field [12,14, 25]. However, pathological image analysis remains a challenging task due to the complex and heterogeneous nature [19] of obtained whole slide images (WSIs), as well as their huge gigapixel-level size [20]. To address this issue, multiple instance learning (MIL) [1] is usually applied to formulate pathological image analysis tasks into weakly supervised learning problems. In the MIL setting, the entire WSI is regarded as a bag and tiled patches are instances. The primary challenge of MIL arises from its weakly supervised nature, i.e. only the bag-level label for the entire WSI is provided, while labels for individual patches are usually unavailable. Although MIL-based methods have shown impressive potential in solving a wide range of pathological image analysis tasks including cancer grading and subtype diagnosis [23], prognosis prediction [18], genotype-related tasks such as gene mutation prediction [4], etc., it is still an open question regarding learning an informative and effective representation of the entire WSI for down-streaming task based on MIL architecture.

Current MIL methods can be broadly categorized into two types: bag-level MIL and instance-level MIL. Bag-level MIL [9,17], also known as embedding-based MIL, involves converting patches (instances) into low-dimensional embeddings, which are then aggregated into WSI (bag)-level representations to conduct the analysis tasks [22]. The aggregator can take different architectures such as an attention module [7,13], convolutional neural network (CNN), Transformer [16], or graph neural network [10,28]. Instance-level MIL [2,8,24], on the other hand, focuses its learning process at the instance level, and then obtains the bag-level prediction by simply aggregating instance predictions. Bag-level MIL incorporates instance embeddings to create a bag representation, converting the MIL into a supervised learning problem. Furthermore, it can extract contextual information and correlations between instances. Nonetheless, Bag-level MIL needs to learn informative embeddings of instances and adjust the contributions of these instance embeddings to generate the bag representation simultaneously, which faces the risk of obtaining a suboptimal model given the limited training samples in practice. The instance-level MIL, however, faces the problem of noisy labels, which is caused by the common strategy of assigning the WSI labels to patches and the fact that there are lots of patches irrelevant to the WSI labels [3,6].

Considering these conventional MIL methods usually utilize either bag-level or instance-level supervision, leading to suboptimal performance. In this paper, we format the instance-level MIL as a noisy label learning task and propose to solve it by designing an instance-level supervision based on the label disambiguation [21]. Then we propose to combine bag-level and instance-level supervision to improve the performance of MIL. The bag-level and instance-level supervision can corporately optimize the instance embedding learning process and well-learned instance embeddings can facilitate the aggregation module to generate the bag representation. The co-supervision design also makes the MIL to be a

multi-task learning framework, where the bag-level supervision channel works to globally summarise the WSI for prediction and the instance-level supervision channel can locally identify key relevant patches. The detailed contributions can be summarized as follows:

1) We propose a novel MIL method for pathological image analysis that leverages a specially-designed residual Transformer backbone and organically integrates both Transformer-based bag-level and label-disambiguation-based instance-level supervision for performance enhancement.
2) We develop a label-disambiguation module that leverages prototypes and confidence bank to tackle the weakly supervised nature of instance-level supervision and reduce the impact of assigned noisy labels.
3) The proposed framework outperforms state-of-the-art (SOTA) methods on public datasets and in a practical clinical task, demonstrating its superiors in WSI analysis. Besides, ablation studies illustrate the superiority of our co-supervision design compared to using only one type of supervision.

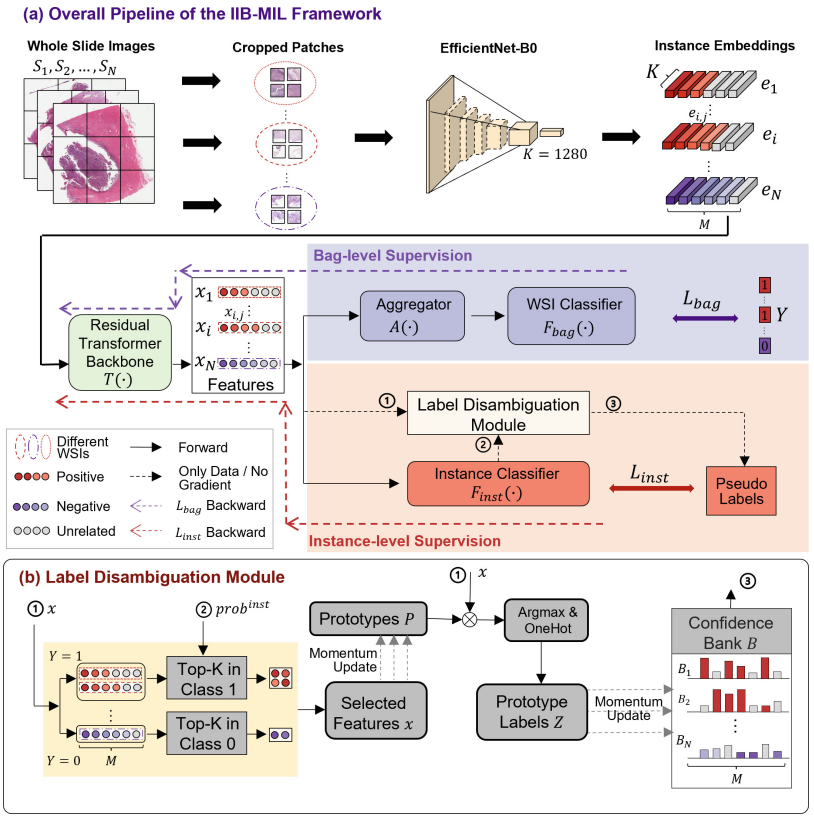

Fig. 1. (a) Overall framework of the IIB-MIL.(b) The detailed diagram of the label-disambiguation-based instance-level supervision. (Details are given in section: 2.1)

2 Method

2.1 Overview

The overall framework of the proposed IIB-MIL is shown in Fig. 1. Similar to previous works [27], IIB-MIL first transforms input huge-size WSI to a set of patch embeddings to simplify the following learning task using a pre-trained encoder, i.e. EfficientNet-B0. Then a specially-designed residual transformer backbone works to calibrate the obtained patch embeddings and encode the context information and correlation of patches. After that, IIB-MIL utilizes both a transformer-based bag-level and a label-disambiguation-based instance-level supervision to cooperatively optimize the model, where the bag-level loss is calculated referring to the WSI labels, while the instance loss is calculated referring to pseudo patch labels calibrated by the Label-Disambiguation module. Since bag-level supervision channel is trained to globally summarise information of all patches for prediction, the bag-level outputs are used as the final predictions during the test stage.

2.2 Problem Formulation

Assume there is a set of N WSIs denoted by $S = \{S_1, S_2, ..., S_N\}$. Each WSI S_i has a WSI-level label $Y_i \in \{1, ..., C\}$, where C represents category number. In each S_i, there exist M_i tiled patches without patch-level labels. To reduce the computational cost, we used a frozen pre-trained encoder to transform patches into K dimensional embeddings $\{e_{i,j}|e_{i,j} \in \mathbb{R}^K, i \in [1, N], j \in [1, M]\}$. Our proposed IIB-MIL comprehensively integrates obtained embeddings $\{e_{i,j}, ...\}$ to generate accurate WSI classification.

2.3 Backbone Network

Before bag-level and instance-level supervision channels, we design a residual transformer backbone $T(\cdot) : \mathbb{R}^K \rightarrow \mathbb{R}^D$ to calibrate the obtained patch embeddings and encode the context information and correlation of patches. $T(\cdot)$ maps patch embeddings $\{e_{i,j}, ...\}$ to a lower-dimensional feature space, denoted as $\{x_{i,j}, ...\}$, where $x_{i,j} = T(e_{i,j})$, $x_{i,j} \in \mathbb{R}^D$ is the calibrated embedding, $T(\cdot)$ is composed of transformer layers and skip connections (Details are given in the supplementary.).

2.4 Instance-Level Supervision

At the core of instance-level supervision is the label disambiguation module, which serves to rectify the imprecise labels that have been assigned to patches. It comprises prototypes and a confidence bank, takes instance features and instance classifier predictions as inputs, and generates soft labels as outputs (Fig. 1 (b)). The prototypes, denoted as $P \in \mathbb{R}^{C \times D}$, are initialized with all-zero vectors and employ momentum-based updates using selected instance features x with

the highest probability $prob^{inst}$ of belonging to their corresponding categories. Prototype labels z are determined based on the proximity of patch features to the prototypes. Confidence $B \in \mathbb{R}^{N \times M \times C}$ is initialized with all WSI labels and uses momentum-based updates with z. Detailed steps are summarized as follows:

Step 1: Obtain the instance classifier output. The instance-level classifier, denoted as $F_{inst}(\cdot)$, takes $x_{i,j} \in \mathbb{R}^D$ as input and outputs the predicted instance probability $prob_{i,j}^{inst} \in \mathbb{R}^C$, as:

$$prob_{i,j}^{inst} = softmax(F_{inst}(x_{i,j})), \tag{1}$$

The probability that $x_{i,j}$ is predicted as class c is denoted as $prob_{i,j,c}^{inst} \in \mathbb{R}^1$.

Step 2: Obtain the prototype labels. At t time, the prototype vector for the category c is $P_{c,t} \in \mathbb{R}^D$. To update $P_{c,t}$, we select a set of instance features $Set_{c,t}$ that have the highest probabilities $prob_{i,j,c}^{inst}$ of belonging to category c. Specifically, we define $Set_{c,t}$ as:

$$Set_{c,t} = \{x_{i,j} | arg\ Top_K(prob_{i,j,c}^{inst}), j \in [1, M], i \in \{i | Y_i = c\}\}, \tag{2}$$

where K is the number of top instance features to select. Then, we use a momentum-based update rule to obtain $P_{c,t+1}$:

$$P_{c,t+1} = \alpha \cdot P_{c,t} + (1 - \alpha) \cdot x_{i,j}, \text{if } x_{i,j} \in Set_{c,t}, \tag{3}$$

where α is the momentum coefficient that automatically decreases from $\alpha = 0.95$ to $\alpha = 0.8$ across epochs. Then, we can obtain prototype labels $z_{i,j} \in \mathbb{R}^C$ using the following equation:

$$z_{i,j} = \text{OneHot}(\arg\max_c(P \cdot x_{i,j}^T)), \tag{4}$$

The resulting prototype label $z_{i,j} \in \mathbb{R}^C$ is a one-hot vector that indicates the category of the j-th instance in the i-th WSI.

Step 3: Obtain Soft Labels from the Confidence Bank Specifically, at time t, the pseudo-target $B_{i,j,t} \in \mathbb{R}^C$ of the instance embedding $e_{i,j}$ is updated by the following:

$$B_{i,j,t} = \beta \cdot B_{i,j,t-1} + (1 - \beta) \cdot z_{i,j}, \tag{5}$$

where β is the momentum update parameter with a default value of $\beta = 0.99$.

Step 4: Calculate Instance-Level Loss. We compute instance-Level Loss using the cross-entropy function:

$$L_{inst} = -\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{C} B_{i,j,c} \cdot log(prob_{i,j,c}^{inst}), \tag{6}$$

Here, $B_{i,j,c}$ and $prob_{i,j,c}^{inst}$ are the c-th component of the pseudo-target $B_{i,j}$ and predicted probability $prob_{i,j}^{inst}$, respectively.

2.5 Bag-Level Supervision

For bag-level supervision, instance features $x_i \in \mathbb{R}^{M\times D}$ go through a transformer-based aggregator $A(\cdot) : \mathbb{R}^{M\times D} \to \mathbb{R}^D$ and a WSI classifier $F_{bag}(\cdot) : \mathbb{R}^D \to \mathbb{R}^C$ in turn (Architecture details are given in the supplementary.). Then we obtain the predicted probability of WSI S_i as:

$$prob_i^{bag} = softmax(F_{bag}(A(x_i))). \tag{7}$$

The bag-level loss function is given by:

$$\mathcal{L}_{bag} = -\sum_{i=1}^{N} prob_i^{bag} \cdot \log(Y_i), \tag{8}$$

where $Y_i \in \mathbb{R}^C$ is the label of WSI S_i.

2.6 Training

In the training phase, We employ a warm-up strategy in which we update only the Prototypes and do not update the Confidence Bank during the first few epochs. Our approach is trained end-to-end, and the total loss function is :

$$\mathcal{L} = \mathcal{L}_{\text{bag}} + \lambda\mathcal{L}_{\text{inst}}, \tag{9}$$

where λ is the hyperparameter that controls the relative importance of the two losses.

3 Experiments

3.1 Dataset

We evaluate our model with three datasets. (1) LUAD-GM Dataset: The objective is to predict the epidermal growth factor receptor (EGFR) gene mutations in patients with lung adenocarcinoma (LUAD) using 723 Whole Slide Image (WSI) slices, where 47% of cases have EGFR mutations. (2) TCGA-NSCLC and TCGA-RCC Datasets: Cancer type classification is performed using The Cancer Genome Atlas (TCGA) dataset. The TCGA-NSCLC dataset comprised two subtypes, lung squamous cell carcinoma (LUSC) and lung adenocarcinoma (LUAD), while the TCGA-RCC dataset included three subtypes: renal chromophobe cell carcinoma (KICH), renal clear cell carcinoma (KIRC), and renal papillary cell carcinoma (KIRP).

3.2 Experiment Settings

The dataset was randomly split into three parts: training, validation, and testing, with 60%, 20%, and 20% of the samples, respectively. WSIs were preprocessed by cropping them into 1120 × 1120 patches, without overlap. The proposed model was implemented in Pytorch, trained on a 32GB TESLA V100 GPU, using AdamW [11] optimizer. The batch size was set to 4, with a learning rate of $1e^{-4}$ and a weight decay of $1e^{-5}$.

Table 1. The performance of IIB-MIL compared with other SOTA methods.

Models	LUAD-GM		TCGA-NSCLC		TCGA-RCC	
	AUC (%)	Accuracy	AUC (%)	Accuracy	AUC (%)	Accuracy
ABMIL [7,12]	52.44	54.55	86.56	77.19	97.02	89.34
CNN-MIL [20]	45.28	44.92	78.64	69.86	69.56	61.17
DSMIL [9]	78.53	71.53	89.25	80.58	98.4	92.94
CLAM-SB [13]	78.49	70.14	86.37	78.47	90.21	76.60
CLAM-MB [13]	82.33	75.70	88.18	81.80	97.23	88.16
ViT-MIL [5]	76.39	70.14	93.77	84.22	97.99	89.66
TransMIL [16]	77.29	74.45	96.03	88.35	98.82	94.66
SETMIL [27]	83.84	76.38	96.01	89.27	99.01	95.20
DTFD [26]	82.41	76.03	97.37	92.34	99.00	96.90
IIB-MIL(Ours)	**85.62 (+1.78)**	78.77	**98.11(+0.74)**	90.91	**99.57 (+0.56)**	95.24

Table 2. Ablation studies and model analysis of IIB-MIL.

Models	LUAD-GM		TCGA-NSCLC		TCGA-RCC	
	AUC (%)	Accuracy	AUC (%)	Accuracy	AUC (%)	Accuracy
w/o Instance	84.01 (−1.61)	76.20	95.89 (−2.22)	89.50	98.93 (−0.64)	93.12
w/o Label Disambiguation	84.23 (−1.39)	65.07	96.40 (−1.71)	89.47	98.97 (−0.6)	94.18
w/o Bag	84.67 (−0.95)	71.23	97.65 (−0.46)	91.39	99.01 (−0.56)	91.53
IIB-MIL	**85.62**	78.77	**98.11**	90.91	**99.57**	95.24
warmup = 1	84.81	75.34	96.14	89.47	99.06	92.59
warmup = 5	85.37	76.71	97.32	91.87	99.25	93.65
warmup = 10	**85.62**	78.77	**98.11**	90.91	**99.57**	95.24
warmup = 50	85.36	77.40	97.51	89.95	99.42	93.12
$\lambda = 0.1$	83.50	71.92	97.59	95.69	99.11	92.06
$\lambda = 1$	83.19	71.23	98.05	91.39	99.05	93.12
$\lambda = 5$	**85.62**	78.77	**98.11**	90.91	**99.57**	95.24
$\lambda = 10$	85.60	76.03	96.51	89.47	99.23	89.95

4 Results and Discussion

4.1 Comparison with State-of-the Art Methods

Table 1 presents a performance comparative analysis of IIB-MIL in relation to other SOTA methods, including ABMIL [7,12], CNN-MIL [20], DSMIL [9], CLAM [13], ViT-MIL [5], TransMIL [16], SETMIL [27], and DTFD [26]. All methods were evaluated in three tasks, namely gene mutation prediction (with or without EGFR mutation), TCGA-NSCLC subtype classification, and TCGA-RCC subtype classification. IIB-MIL achieved AUCs of 85.62%, 98.11%, and 99.57%. We can also find IIB-MIL outperformed other SOTA methods, in the three tasks with at least 1.78%, 0.74%, and 0.56% performance enhancement (AUC), respectively.

4.2 Ablation Studies

We conducted ablation studies to assess the efficacy of each component in IIB-MIL. The results, in Table 2, indicate that all of the designed components,

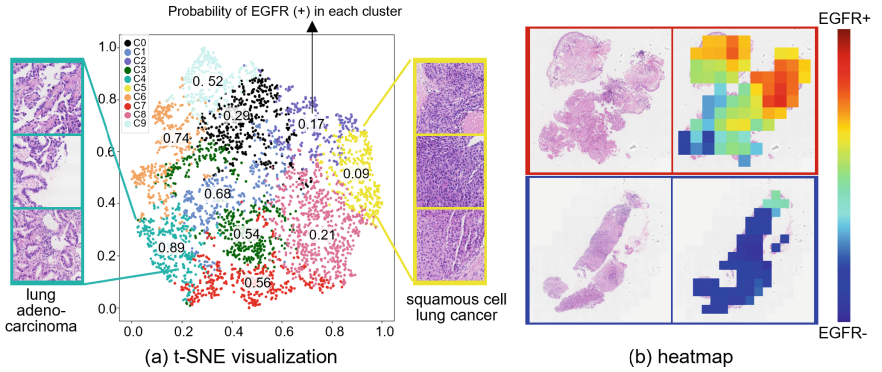

Fig. 2. (a)t-SNE plot of the patch features obtained from the backbone;(b) Example heatmaps of IIB-MIL on WSIs with known EFGR mutation labels.

including the label disambiguation module, instance-level supervision, and bag-level supervision, contribute to the success of IIB-MIL. We also investigated the impact of the warm-up epoch number and found that selecting an appropriate value, such as $warmup = 10$, can lead to better model performance. Furthermore, we examined the impact of the weighting factor λ, and the outcomes indicated that assigning greater importance to instance-level supervision ($\lambda = 5$) helps IIB-MIL enhance its performance, thus demonstrating the effectiveness of the designed label-disambiguation-based instance-level supervision.

4.3 Model Interpretation

Figure 2(a) shows the t-SNE plot of the obtained patch features from the backbone of the IIB-MIL. The patches are unsupervisedly clustered into groups based on their features, indicated by various colors. The numbers displayed within each group represent the average likelihood of the EGFR mutation predicted by the patches. With the help of the label-disambiguation-based instance-level supervision, IIB-MIL can identify highly positive and negative related patches to the WSI-label, i.e., the cyan-blue group and yellow group. Double-checked by pathologists, we find that the cyan-blue group consists of patches from lung adenocarcinoma and the yellow group consists of patches from the squamous cells. This finding aligns with the domain knowledge of pathologists. Figure 2(b) investigates the contribution of each patch in predicting EGFR mutation. The resulting heatmap shows the decision mechanism of IIB-MIL in the accurate distinguishment between EGFR mutation-positive and negative samples.

5 Conclusion

This paper presents IIB-MIL, a novel MIL approach for pathological image analysis. IIB-MIL utilizes a label disambiguation module to establish more precise instance-level supervision. It then combines the instance-level and bag-level

supervision to enhance the performance of the IIB-MIL. Experimental results demonstrate that IIB-MIL surpasses current SOTA techniques on publicly available datasets, and holds significant potential for addressing more complex clinical applications, such as predicting gene mutations. Furthermore, IIB-MIL can identify highly relevant patches, providing pathologists with valuable insights into underlying mechanisms.

References

1. Amores, J.: Multiple instance classification: review, taxonomy and comparative study. Artif. intell. **201**, 81–105 (2013)
2. Campanella, G., et al.: Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat. Med. **25**(8), 1301–1309 (2019)
3. Chikontwe, P., Kim, M., Nam, S.J., Go, H., Park, S.H.: Multiple instance learning with center embeddings for histopathology classification. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12265, pp. 519–528. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59722-1_50
4. Coudray, N., et al.: Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat. Med. **24**(10), 1559–1567 (2018)
5. Dosovitskiy, A., et al.: An image is worth 16×16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
6. 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)
7. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International Conference on Machine Learning, pp. 2127–2136. PMLR (2018)
8. Lerousseau, M., Vakalopoulou, M., Classe, M., Adam, J., Battistella, E., Carré, A., Estienne, T., Henry, T., Deutsch, E., Paragios, N.: Weakly supervised multiple instance learning histopathological tumor segmentation. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12265, pp. 470–479. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59722-1_45
9. 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)
10. Li, R., Yao, J., Zhu, X., Li, Y., Huang, J.: Graph CNN for survival analysis on whole slide pathological images. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 174–182. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_20
11. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
12. Lu, M.Y., et al.: Ai-based pathology predicts origins for cancers of unknown primary. Nature **594**(7861), 106–110 (2021)
13. 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. Nat. Biomed. Eng. **5**(6), 555–570 (2021)

14. Noorbakhsh, J., et al.: Deep learning-based cross-classifications reveal conserved spatial behaviors within tumor histological images. Nat. Commun. **11**(1), 6367 (2020)
15. Rubin, R., Strayer, D.S., Rubin, E., et al.: Rubin's Pathology: Clinicopathologic Foundations of Medicine. Lippincott Williams & Wilkins (2008)
16. 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. Adv. Neural Inf. Process. Syst. **34**, 2136–2147 (2021)
17. 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)
18. Skrede, O.J., et al.: Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. The Lancet **395**(10221), 350–360 (2020)
19. Srinidhi, C.L., Ciga, O., Martel, A.L.: Deep neural network models for computational histopathology: a survey. Med. Image Anal. **67**, 101813 (2021)
20. Tellez, D., Litjens, G., van der Laak, J., Ciompi, F.: Neural image compression for gigapixel histopathology image analysis. IEEE Trans. Pattern Anal. Mach. Intell. **43**(2), 567–578 (2019)
21. Wang, H., et al.: Pico: contrastive label disambiguation for partial label learning. arXiv preprint arXiv:2201.08984 (2022)
22. Wang, X., Yan, Y., Tang, P., Bai, X., Liu, W.: Revisiting multiple instance neural networks. Pattern Recogn. **74**, 15–24 (2018)
23. Wetstein, S.C., et al.: Deep learning-based breast cancer grading and survival analysis on whole-slide histopathology images. Sci. Rep. **12**(1), 15102 (2022)
24. Xu, G., et al.: Camel: a weakly supervised learning framework for histopathology image segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10682–10691 (2019)
25. Zhang, H., Meng, Y., Qian, X., Yang, X., Coupland, S.E., Zheng, Y.: A regularization term for slide correlation reduction in whole slide image analysis with deep learning. In: Medical Imaging with Deep Learning, pp. 842–854. PMLR (2021)
26. Zhang, H., et al.: 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)
27. Zhao, Y., Lin, Z., Sun, K., Zhang, Y., Huang, J., Wang, L., Yao, J.: Setmil: spatial encoding transformer-based multiple instance learning for pathological image analysis. In: Medical Image Computing and Computer Assisted Intervention-MICCAI 2022: 25th International Conference, Singapore, 18–22 September 2022, Proceedings, Part II, pp. 66–76. Springer, Heidelberg (2022). https://doi.org/10.1007/978-3-031-16434-7_7
28. Zhao, Y., et al.: Predicting lymph node metastasis using histopathological images based on multiple instance learning with deep graph convolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4837–4846 (2020)