Few-Shot Lymph Node Metastasis Classification Meets High Performance on Whole Slide Images via the Informative Non-Parametric Classifier

Yi Li[1 [0000−0002−7840−2611]], Qixiang Zhang[1 [0009−0002−3158−9471]], Tianqi Xiang[1 [0000−0002−5550−1721]], Yiqun Lin[1 [0000−0002−7697−0842]], Qingling Zhang[3 [0000−0002−5287−264X]], and Xiaomeng Li[1,2(✉) [0000−0003−1105−8083]]

[1] Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, China
eexmli@ust.hk

[2] HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, Futian, Shenzhen

[3] Department of Pathology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China

Abstract. Lymph node metastasis (LNM) classification is crucial for breast cancer staging. However, the process of identifying tiny metastatic cancer cells within gigapixel whole slide image (WSI) is tedious, time-consuming, and expensive. To address this challenge, computational pathology methods have emerged, particularly multiple instance learning (MIL) based on deep learning. But these methods require massive amounts of data, while existing few-shot methods severely compromise accuracy for data saving. To simultaneously achieve few-shot and high performance LNM classification, we propose the informative non-parametric classifier (INC). It maintains informative local patch features divided by mask label, then innovatively utilizes non-parametric similarity to classify LNM, avoiding overfitting on a few WSI examples. Experimental results demonstrate that the proposed INC outperforms existing SoTA methods across various settings, with less data and labeling cost. For the same setting, we achieve remarkable AUC improvements over 36.76% on CAMELYON16. Additionally, our approach demonstrates excellent generalizability across multiple medical centers and corrupted WSIs, even surpassing many-shot SoTA methods over 7.55% on CAMELYON16-C. Code is available at https://github.com/xmed-lab/INC.

Keywords: Few-Shot, Non-Parametric, WSI, Lymph Node Metastasis

1 Introduction

Pathological Whole Slide Image (WSI) serves as the gold standard for cancer diagnosis. For breast cancer, the lymph node metastasis (LNM) classification is an essential diagnostic staging task, according to the tumor, nodes, metastasis staging system. In clinical practice, it is tedious, time-consuming, and expensive

to search tiny metastatic cancer cells from multiple lymph nodes within gigapixel WSI. Therefore, deep learning based computational pathology schemes [1,17] emerged to this end.

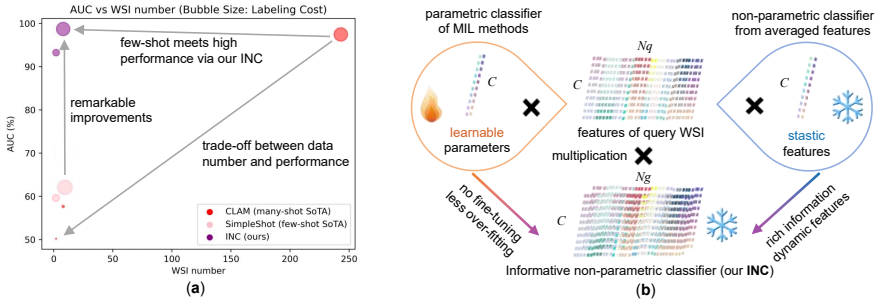

Fig. 1. (a) Few-shot LNM classification meets high performance via the proposed INC method on CAMELYON16 dataset [1]. (b) INC is a non-parametric classifier, maintaining rich information without fine-tuning. The notion of a parametric classifier can be regarded as an FC layer with fixed weights, while a non-parametric classifier (e.g., KNN) uses the globally averaged feature. In comparison, our INC method keeps all local features and matches them dynamically according to individual test examples.

Currently, the predominant WSI classification technique is multiple instance learning (MIL) [12,19,22,11], including techniques like feature fusion [14], self-attention [12,22], pretraining [4,13] etc. [18,15]. Notably, conventional MIL methods require massive data, *e.g.*, Campanella et al. [2] used thousands of training WSIs for LNM classification. However, acquiring a large training set can be laborious, costly, and demands significant training computational resources. To address these issues, few-shot methods have been proposed for WSI classification tasks, which only use several training examples at hundreds of times less data, *e.g.*, text prompt learning [20], data-efficient MIL [19], prototype [5] with general few-shot learning [24]. Besides, some pathological few-shot techniques [7,26,3,9] are specially designed for patch classification, which are not directly applicable to WSI classification without a crucial patch aggregation process.

Despite the achievement in reducing data cost, performances of existing methods are obviously sacrificed using fewer WSIs. For example, the AUC of CLAM [19] dramatically drops to 57.65% from 97.45% as Fig. 1 (a). In other words, it is a trade-off between few data and high performance. While extra information can compensate for the performance loss owing to the data scarcity. For instance, leveraging weak labels [16] significantly enhances the model's capability beyond that of the unsupervised method [27] for gland segmentation. Motivated by this, we aim to use fewer WSIs with mask labels to save data and achieve high performance at the same time. Importantly, many WSI-level labels cost more labeling efforts than a few patch-level masks (when mask labels $\lesssim \frac{1}{30}$ WSI labels), indicating that our scheme saves both data and annotation costs.

Another motivation is that, parametric classifier (*e.g.*, FC layer) in MIL is prone to overfit on a few WSI examples or specific medical centers (see MIL methods in Tab. 3). Thus, we expect a non-parametric classifier without fine-tuning on a few WSIs. For existing non-parametric methods, like KNN evaluation protocol for pretrained models [4], or Prototype [5] and SimpleShot [24] based on mask label, they use the global static feature as Fig. 1 (b). Unfortunately, these methods discard rich information of local patches to obtain the global feature, resulting in suboptimal performance. Besides, some KNN based MIL methods for text [10], CT [25] etc. are not designed for gigabyte WSI.

Based on the above motivations, we proposed the informative non-parametric classifier (INC) to simultaneously achieve few-shot and high performance LNM classification. Specifically, it maintains all deep features of local patches divided by mask label, then utilizes non-parametric similarity between the informative gallery WSI bags and the query bag to classify LNM. Experimentally, our method surpasses existing SoTA MIL methods across various settings. Under the same setting on CAMELYON16 [1] dataset, the AUC improvements are nontrivial over 36.76%. Importantly, our method saves the labeling cost when applying much less data, with better performances. Furthermore, we also achieve outstanding performance when generalizing to multiple medical centers on CAMELYON17 [17], and the AUC on corrupted WSIs of CAMELYON16-C [28] even surpasses many-shot SoTA over 7.55%.

2 Methodology

The primary benefit of the INC lies in its ability to preserve rich local features, thereby avoiding information loss in previous parametric classifier methods such as MIL, or non-parametric methods that rely on mean feature values (*e.g.*, KNN and Prototype [5]). In contrast to static classifiers with fixed parameters or global features, INC functions as a dynamic classifier by incorporating bags of local features. This unique approach guarantees the retention of the most comprehensive information when compared to other classifiers, enabling it to dynamically adapt to different medical centers or scanners without the risk of overfitting. Here, we firstly elaborate how we generate the classification logit based on similarity from bags of local features. Then we describe a further retrieval aggregation step, which considers multiple related logits for slide-level prediction.

2.1 Informative Similarity Logit for Non-parametric Classifier

The core step of the proposed INC is to produce a classification logit from bags of local features. This is accomplished by computing the similarity between informative gallery feature bags and query bags, taking into account both the cancer region and the normal region. In light of the fact that the similarity stems from informative local features, it is referred to as the informative similarity logit.

To be specific, we firstly extract the patch features via a self-supervised pretrained encoder $f(\cdot)$, which returns L2 normalized features. Then, we get

the bag of local gallery features $\boldsymbol{G} \in \mathbb{R}^{N_g \times C}$ (support WSIs) and query features $\boldsymbol{Q} \in \mathbb{R}^{N_q \times C}$ (a test WSI) as:

$$\boldsymbol{G} = f(\boldsymbol{X}_g), \quad \boldsymbol{Q} = f(\boldsymbol{X}_q), \tag{1}$$

where $\boldsymbol{X}_g, \boldsymbol{X}_q$ indicate gallery and query patch instances, at instances quantity of N_g and N_q, respectively. Then the target similarity matrix $\boldsymbol{S} \in \mathbb{R}^{N_g \times N_q}$ between gallery bag and query bag are calculated by matrix multiplication as:

$$\boldsymbol{S} = \boldsymbol{G}\boldsymbol{Q}^\top. \tag{2}$$

The above cosine similarity matrix contains rich information between gallery feature bag and query bag. To get the prediction score, we need logits for each query via reducing $\boldsymbol{S}$ along the gallery dimension (N_g). In the step, we firstly divide $\boldsymbol{S}$ into a set of positive similarity matrix $\boldsymbol{S}^+ \in \mathbb{R}^{N_p \times N_q}$ and negative similarity matrix $\boldsymbol{S}^- \in \mathbb{R}^{N_n \times N_q}$ as:

$$\boldsymbol{S} = \{\boldsymbol{S}^+, \boldsymbol{S}^-\}, \quad s.t. \;\; \boldsymbol{S}^+ = \boldsymbol{S}_{idx(\mathbf{m}=1)}, \;\; \boldsymbol{S}^- = \boldsymbol{S}_{idx(\mathbf{m}=0)}, \tag{3}$$

where the mask label $\mathbf{m} \in \mathbb{R}^{N_g}$ is used to divide positive gallery related similarities $\boldsymbol{S}^+ = \boldsymbol{S}_{idx(\mathbf{m}=1)}$ and negative similarities $\boldsymbol{S}^- = \boldsymbol{S}_{idx(\mathbf{m}=0)}$ via indexing on the gallery dimension. Note, N_p, N_n indicates the number of cancer patches and normal patches, respectively.

To dynamically utilize suitable similarities for each query, we take the top-k gallery similarities of $\boldsymbol{S}^+, \boldsymbol{S}^-$ as $\hat{\boldsymbol{S}}^+ \in \mathbb{R}^{N_k \times N_q}, \hat{\boldsymbol{S}}^- \in \mathbb{R}^{N_k \times N_q}$ at top-k size N_k. Then we generate the informative similarity logit $\mathbf{s} \in \mathbb{R}^{N_q}$ for each query as:

$$\mathbf{s} = \frac{\sum \hat{\boldsymbol{S}}^+ - \sum \hat{\boldsymbol{S}}^-}{N_k}, \tag{4}$$

where top related cancer gallery patches and normal patches are dynamically used as the non-parametric classifier.

2.2 Retrieval Aggregation of Informative Non-Parametric Classifier

To get the slide-level prediction from the patch logit in Eq. 4, we need a further step to aggregate the patch logits. Our solution is to select the highest logit in Eq. 4, then conduct retrieval within the bag of query features to aggregate discriminative and related logits. Let $\boldsymbol{Q}_1$ to be the query feature at the highest similarity logit, namely $idx(max(\mathbf{s}))$ using max and indexing operation, respectively. Then, we can retrieve the index array $\mathbf{r}$ of related patch queries as:

$$\mathbf{r} = idx(\boldsymbol{Q}_1\boldsymbol{Q}^\top > t), \tag{5}$$

where cosine similarities $\boldsymbol{Q}_1\boldsymbol{Q}^\top$ higher than a threshold t are indexed into $\mathbf{r}$.

To aggregate the related logits according to their importance, we need an aggregation weight $\mathbf{w}$ for each logit according to the similarity $\hat{\mathbf{s}}$ from $\mathbf{r}$. Then

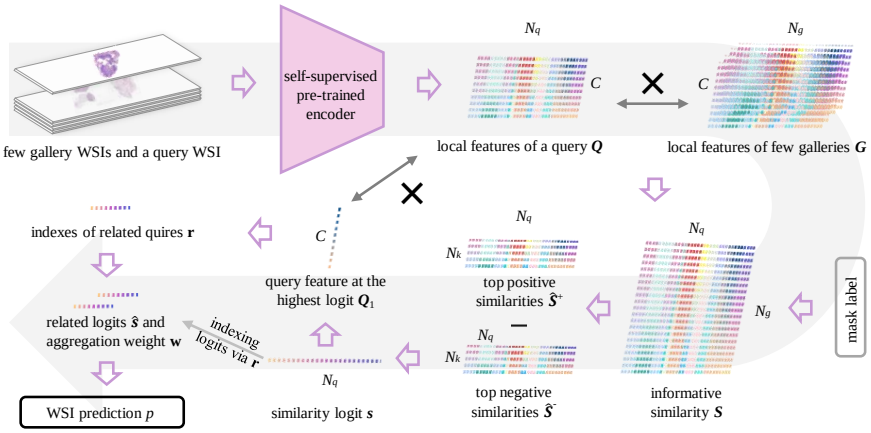

Fig. 2. Data flow of the proposed INC. Note, $\times$ indicates matrix multiplication, and each symbol can be found from Eq. 1 to Eq. 7.

we multiply it with a softmax temperature τ before the softmax operation $\sigma(\cdot)$. As above descriptions, we have the formula of aggregation weights:

$$\mathbf{w} = \sigma(\hat{\mathbf{s}} \cdot \tau) \quad s.t. \quad idx(\hat{\mathbf{s}}) = \mathbf{r}. \tag{6}$$

In the last, we multiply the related logits $\hat{\mathbf{s}}$ with the transposed weight $\mathbf{w}^\top$ to realize the weighted sum. Until here, we get the slide-level prediction p from the aggregation of related query logits as:

$$p = \hat{\mathbf{s}}\mathbf{w}^\top. \tag{7}$$

3 Experiments

Dataset. The main dataset is CAMELYON16 [1], consisting of 399 H&E stained WSIs for breast cancer LNM classification. The train-val set has 270 slides, and there are 129 test slides. To test the generalization, we introduced CAMELYON17 [17] and CAMELYON16-C [28]. Specifically, CAMELYON17 contains 5 medical centers, where 500 labeled WSIs (1 invalid) are used to test. Besides, CAMELYON16-C is an augmented dataset simulating WSI scanning corruption from CAMELYON16. Following Li et al. [15], 243 WSIs (9:1) of CAMELYON16 are used to train many-shot baselines. Besides, CAMELYON17 (499 WSIs) and CAMELYON16-C (129 test WSIs) are used to test only, using models trained from CAMELYON16 to evaluate generalization as Li et al. [15].

Implementation and evaluation. We use a self-supervised pretrained model [13] based on backbone ViT-S/8 [8] as the image encoder. The patch size is 256 at $\times$20 magnification. To fit the input size of ViT-S/8 [8], we apply five-crop at side 224 without other augmentation to extract features and use the mean feature as the local feature. *For a fair comparison, we apply this backbone, weights, and*

Table 1. Few-shot results on CAMELYON16, CAMELYON17 [17] and CAMELYON16-C [28], using 8 WSIs in total. Note, all the methods are implemented in the same settings for fair comparison, except results with † based on CLIP [21]. Besides, * indicates the usage of mask labels as ours, where SimpleShot [24] is a general few-shot method, combined with Prototype [5].

Method	CAMELYON16		CAMELYON17		CAMELYON16-C	
	AUC (%)	Acc. (%)	AUC (%)	Acc. (%)	AUC (%)	Acc. (%)
†TOP [20]	75.41	-	-	-	-	-
KNN-mean	$58.12_{\pm 8.02}$	$41.09_{\pm 2.45}$	$45.80_{\pm 2.75}$	$37.72_{\pm 2.20}$	$57.16_{\pm 7.94}$	$40.62_{\pm 3.42}$
KNN-max	$59.45_{\pm 8.01}$	$40.93_{\pm 1.92}$	$50.91_{\pm 7.23}$	$41.84_{\pm 10.74}$	$59.77_{\pm 7.96}$	$43.26_{\pm 7.09}$
CLAM-MB [19]	$57.65_{\pm 6.79}$	$48.68_{\pm 6.79}$	$52.49_{\pm 6.42}$	$49.26_{\pm 9.87}$	$57.12_{\pm 3.58}$	$48.53_{\pm 7.86}$
*Prototype [5]	$48.83_{\pm 2.88}$	$49.77_{\pm 2.57}$	$57.50_{\pm 1.52}$	$64.13_{\pm 0.71}$	$48.69_{\pm 2.88}$	$59.53_{\pm 2.10}$
*SimpleShot [24]	$61.95_{\pm 3.89}$	$59.22_{\pm 2.38}$	$71.58_{\pm 2.50}$	$69.98_{\pm 1.69}$	$62.40_{\pm 2.75}$	$66.67_{\pm 2.59}$
INC (ours)	$\mathbf{98.71}_{\pm 0.22}$	$\mathbf{95.35}_{\pm 0.49}$	$\mathbf{89.42}_{\pm 0.81}$	$\mathbf{85.37}_{\pm 1.76}$	$\mathbf{95.27}_{\pm 0.76}$	$\mathbf{85.89}_{\pm 2.61}$

crop strategy to all implemented baselines. The few-shot WSI numbers are 1, 2, 4, 8 (default shot). All the hyperparameters are gird searched on the validation set with detailed experiments in Supplementary B. For baselines, we studied the implementations in Supplementary A for fair and strong baselines. For stable results, we randomly split the dataset 5 times (fixed random seed at 1024 for good reproducibility), and report the average Area Under the Curve (AUC) and Accuracy (ACC) with corresponding standard deviation (±).

Table 2. Few-shot results on CAMELYON16 [1] (8 WSIs in Tab. 1). Note, 2 WSIs is also 1-shot (1 WSI per class) for TOP [20], KNN, CLAM-MB [19], while others use 2 positive WSIs. All the baselines use the same self-supervised encoder [13], but † indicates the CLIP [21] encoder, and * indicates the usage of mask labels as ours.

Method	1 WSI		2 WSIs		4 WSIs	
	AUC (%)	Acc. (%)	AUC (%)	Acc. (%)	AUC (%)	Acc. (%)
†TOP [20]	-	-	67.14	-	69.67	-
KNN-mean	-	-	$50.49_{\pm 10.86}$	$39.22_{\pm 3.52}$	$53.71_{\pm 13.17}$	$43.10_{\pm 3.05}$
KNN-max	-	-	$45.03_{\pm 11.75}$	$39.69_{\pm 6.19}$	$55.01_{\pm 15.94}$	$46.82_{\pm 12.20}$
CLAM-MB [19]	-	-	$50.19_{\pm 12.75}$	$43.88_{\pm 7.99}$	$52.89_{\pm 9.33}$	$45.74_{\pm 8.26}$
*Prototype [5]	$52.61_{\pm 7.77}$	$49.30_{\pm 7.94}$	$44.15_{\pm 4.33}$	$42.95_{\pm 6.37}$	$43.43_{\pm 8.80}$	$40.16_{\pm 7.38}$
*SimpleShot [24]	$59.82_{\pm 9.96}$	$51.16_{\pm 10.53}$	$57.23_{\pm 8.56}$	$51.32_{\pm 7.28}$	$55.74_{\pm 13.36}$	$49.77_{\pm 10.12}$
INC	$\mathbf{87.33}_{\pm 8.72}$	$\mathbf{82.48}_{\pm 10.20}$	$\mathbf{93.24}_{\pm 7.01}$	$\mathbf{88.68}_{\pm 4.00}$	$\mathbf{94.15}_{\pm 1.18}$	$\mathbf{88.84}_{\pm 1.81}$

Remarkable improvements in few-shot LNM classification. We compare our results with previous SoTA methods as Tab. 1 in fair implementation, where the encoder, crop strategy, and data split are the same. The results suggest our INC has surpassed SoTA few-shot baselines at large margins. Specifically, on CAMELYON16 [1], our AUC is $98.71_{\pm 0.22}$ which is higher than the baseline in the same setting (SimpleShot [24]) by 36.76%, and the gain of accuracy is 36.13%. For the results of TOP [20] (only single data split), the improvement

Table 3. Our few-shot INC with 8 WSIs meets high performance on CAMELYON16, CAMELYON17 [17] and CAMELYON16-C [28], even beyond SoTA many-shot methods using 243 WSIs at fair implementation. Notably, our overall labeling cost is comparable or less than these many-shot methods. Note, signs ‡ and † indicate referenced results with partial results, and † means single test without repeats.

Method	CAMELYON16		CAMELYON17		CAMELYON16-C	
	AUC (%)	Acc. (%)	AUC (%)	Acc. (%)	AUC (%)	Acc. (%)
many-shot						
†dMIL-Trans [6]	91.69	81.40	-	-	-	-
†UNI [5]	<u>97.5</u>	-	-	-	-	-
‡MHIM-MIL [23]	$96.49_{\pm 0.65}$	$92.48_{\pm 0.35}$	-	-	-	-
‡Trans.+FT [15]	$96.7_{\pm 0.3}$	-	71.7	-	85.7	-
MIL-mean	$59.30_{\pm 3.35}$	$64.81_{\pm 3.27}$	$66.63_{\pm 1.84}$	$66.21_{\pm 2.81}$	$57.87_{\pm 6.17}$	$55.50_{\pm 3.66}$
MIL-max	$89.36_{\pm 0.44}$	$86.51_{\pm 0.46}$	<u>88.07</u>$_{\pm 0.40}$	$77.24_{\pm 1.85}$	$73.31_{\pm 2.37}$	$64.65_{\pm 1.46}$
MIL-top5 [19]	$96.16_{\pm 2.37}$	$91.16_{\pm 1.44}$	$60.04_{\pm 3.52}$	$57.07_{\pm 3.39}$	$85.02_{\pm 3.16}$	$61.50_{\pm 10.04}$
ABMIL [12]	$92.02_{\pm 0.41}$	$90.17_{\pm 0.22}$	$83.34_{\pm 1.24}$	$79.93_{\pm 1.83}$	$84.59_{\pm 2.40}$	<u>82.08</u>$_{\pm 2.20}$
CLAM-SB [19]	$97.25_{\pm 0.16}$	$92.56_{\pm 0.62}$	$87.03_{\pm 2.36}$	$79.48_{\pm 3.99}$	$86.42_{\pm 2.46}$	$74.81_{\pm 1.61}$
CLAM-MB [19]	$97.45_{\pm 1.81}$	<u>94.57</u>$_{\pm 2.21}$	$85.00_{\pm 1.23}$	$76.07_{\pm 2.37}$	<u>87.72</u>$_{\pm 2.02}$	$75.35_{\pm 4.23}$
TransMIL [22]	$95.77_{\pm 2.24}$	$90.39_{\pm 2.11}$	$88.02_{\pm 1.63}$	<u>81.35</u>$_{\pm 2.74}$	$87.31_{\pm 2.83}$	$72.75_{\pm 2.56}$
few-shot						
INC (ours)	**98.71**$_{\pm 0.22}$	**95.35**$_{\pm 0.49}$	**89.42**$_{\pm 0.81}$	**85.37**$_{\pm 1.76}$	**95.27**$_{\pm 0.76}$	**85.89**$_{\pm 2.61}$

is also remarkable at 23.30%. When generalized to CAMELYON17 [17] and CAMELYON16-C [28], the proposed INC still works best, where the AUC gain is 17.84% and 32.87% on these datasets, respectively.

Besides the results of 8 WSIs in Tab. 1, we list more results using fewer WSIs in Tab. 2 on CAMELYON16 [1]. From these tables, we find INC is significantly higher than existing methods. Specifically, it surpasses the second by 27.51%, 26.10%, 24.48%, 23.30% on AUC for 1, 2, 4, 8 WSIs, respectively, and the corresponding accuracy improvements are 31.32%, 37.36%, 39.07%, 19.22%, respectively. These remarkable results can be attributed to our effective utilization of crucial local data (patch) and task-specific designs, which suit WSI much better than existing methods borrowed from the general CV. It suggests that maintaining rich local information is necessary, instead of generating global information like Prototype [5] and SimpleShot [24]. Also, it's more suitable to used non-parametric classifier instead of parametric classifier (CLAM-MB [19]) in few-shot setting.

Few-shot INC meets high performance beyond many-shot methods. Besides great successes in few-shot setting, our INC even shows higher results (97.5% vs. our 98.7%) than many-shot methods (243 WSIs vs. our 8 WSIs) as Tab. 3. Notably, the labeling cost is also reduced when applying 8 or less masks (30 times labeling cost for quantity $\leq \frac{8}{243}$). These results indicate our method takes advantages in both performance and labeling and data cost.

In real practice, the WSI scanners are different across hospitals with variations. Thus, the model train from one hospital often meets obvious perfor-

Table 4. Ablation study on the CAMELYON16 [1] using 8 WSIs showed that the mask label (SimpleShot) only slightly improves the baseline (KNN). However, the core improvements come from informative local features for similarity logit, and the retrieval aggregation further enhances the performance.

Setting	Mask Label	Informative Similarity Logit	Retrieval Aggregation	AUC (%)
KNN-mean	✗	✗	✗	$58.12_{\pm 8.02}$
SimpleShot	✔	✗	✗	$61.95_{\pm 3.89}$
Ours Eq. 1-4	✔	✔	✗	$96.43_{\pm 0.49}$
Ours Eq. 1-7	✔	✔	✔	$\mathbf{98.71}_{0.22}$

mance drops when applied to other medical centers. For example, the parametric MIL method, CLAM-MB [19] produces good results on CAMELYON16 at AUC 97.45%, but it turns to 85% on CAMELYON17 and 87.72% on CAMELYON16-C, where the decreases are 12.45% and 9.73%, respectively. In comparison, our INC shows stronger generalization ability at AUC 89.42% and 95.27%, which greatly exceeds previous methods. Specifically, INC surpasses the best many-shot method by 7.55% on CAMELYON16-C. Notably, non-parametric classifiers (few-shot baselines) present similar results for different datasets, which support our motivation of non-parametric classifiers fine.

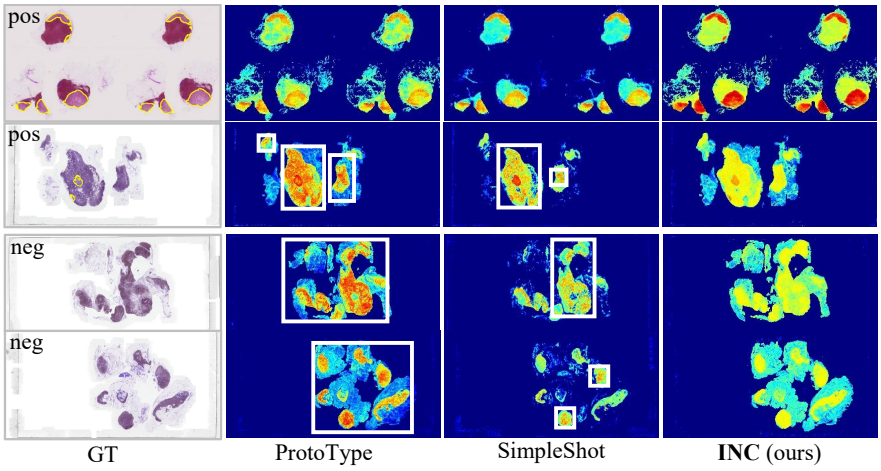

Fig. 3. Visualizations on CAMELYON16 datasets. Ground-truth tumors are circled in yellow, predicted LNM regions are red, and false predictions are noted by white boxes.

Ablation study. In Tab. 4, we find that the usage of mask label in SimpleShot can only improve the KNN baseline by 3.83% in AUC only. While our informative similarity logit improves the baseline by 40.59%. It suggests that keeping rich local information is very crucial to make good use of mask label in few-shot

setting. Besides, the retrieval aggregation can further enhance the results at a high level, which demonstrates the significant effectiveness of the proposed INC. **Visualization results.** We further depict the visualizations in Fig. 3 via the query logits $\hat{\mathbf{s}}$ in Eq. 4. Since KNN methods use global features without local information, we compare our method with Prototype and SimpleShot. We find our INC obviously shows better results than other baselines. Specifically, baselines predict more false negatives. Besides, we draw retrieval processes in Supplementary C to show how INC works.

4 Conclusion

In summary, this study presents a novel, data-efficient, and effective informative non-parametric (INC) classifier for the classification of lymph node metastasis (LNM). In contrast to existing many-shot methods that heavily depend on large datasets, or few-shot methods that seriously compromise performance, INC successes to make few-shot LNM classification meet high performance, even beyond existing SoTA many-shot methods. Furthermore, the INC does not require fine-tuning and avoids overfitting on specific medical centers or scanners, with less labeling cost using only a few masks. Comprehensive experimental results demonstrate that the INC attains the new SoTA in LNM classification, meanwhile offering several notable advantages.

Acknowledgement. This work was supported in part by grants from the National Natural Science Foundation of China under Grant No. 62306254, grants from Hong Kong Innovation and Technology Fund under Projects ITS/030/21, grants from the Foshan HKUST Projects under Grants FSUST21-HKUST10E and FSUST21-HKUST11E and Project of Hetao Shenzhen-Hong Kong Science and Technology Innovation Cooperation Zone (HZQB-KCZYB-2020083).

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

References

1. Bejnordi, B.E., Veta, M., Van Diest, P.J., Van Ginneken, B., Karssemeijer, N., Litjens, G., Van Der Laak, J.A., Hermsen, M., Manson, Q.F., Balkenhol, M., et al.: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. Jama **318**(22), 2199–2210 (2017)
2. 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)
3. Chen, J., Jiao, J., He, S., Han, G., Qin, J.: Few-shot breast cancer metastases classification via unsupervised cell ranking. IEEE/ACM transactions on computational biology and bioinformatics **18**(5), 1914–1923 (2019)

4. Chen, R.J., Chen, C., Li, Y., Chen, T.Y., Trister, A.D., Krishnan, R.G., Mahmood, F.: Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16144–16155 (2022)
5. 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.: A general-purpose self-supervised model for computational pathology. arXiv preprint arXiv:2308.15474 (2023)
6. Chen, Y., Shao, Z., Bian, H., Fang, Z., Wang, Y., Cai, Y., Wang, H., Liu, G., Li, X., Zhang, Y.: dmil-transformer: Multiple instance learning via integrating morphological and spatial information for lymph node metastasis classification. IEEE Journal of Biomedical and Health Informatics (2023)
7. Deuschel, J., Firmbach, D., Geppert, C.I., Eckstein, M., Hartmann, A., Bruns, V., Kuritcyn, P., Dexl, J., Hartmann, D., Perrin, D., et al.: Multi-prototype few-shot learning in histopathology. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 620–628 (2021)
8. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
9. He, Y., Li, X.: Whole-slide-imaging cancer metastases detection and localization with limited tumorous data. arXiv preprint arXiv:2303.10342 (2023)
10. Hellman, S., Murray, W.R., Wiemerslage, A., Rosenstein, M., Foltz, P., Becker, L., Derr, M.: Multiple instance learning for content feedback localization without annotation. In: Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications. pp. 30–40 (2020)
11. Huang, S.C., Chen, C.C., Lan, J., Hsieh, T.Y., Chuang, H.C., Chien, M.Y., Ou, T.S., Chen, K.H., Wu, R.C., Liu, Y.J., et al.: Deep neural network trained on gigapixel images improves lymph node metastasis detection in clinical settings. Nature Communications **13**(1), 3347 (2022)
12. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
13. Kang, M., Song, H., Park, S., Yoo, D., Pereira, S.: Benchmarking self-supervised learning on diverse pathology datasets. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3344–3354 (2023)
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 (CVPR). pp. 14318–14328 (June 2021)
15. Li, H., Zhu, C., Zhang, Y., Sun, Y., Shui, Z., Kuang, W., Zheng, S., Yang, L.: Task-specific fine-tuning via variational information bottleneck for weakly-supervised pathology whole slide image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7454–7463 (2023)
16. Li, Y., Yu, Y., Zou, Y., Xiang, T., Li, X.: Online easy example mining for weakly-supervised gland segmentation from histology images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 578–587. Springer (2022)
17. Litjens, G., Bandi, P., Ehteshami Bejnordi, B., Geessink, O., Balkenhol, M., Bult, P., Halilovic, A., Hermsen, M., van de Loo, R., Vogels, R., et al.: 1399 h&e-stained sentinel lymph node sections of breast cancer patients: the camelyon dataset. GigaScience **7**(6), giy065 (2018)

18. Lu, M.Y., Chen, B., Zhang, A., Williamson, D.F., Chen, R.J., Ding, T., Le, L.P., Chuang, Y.S., Mahmood, F.: Visual language pretrained multiple instance zero-shot transfer for histopathology images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19764–19775 (2023)
19. 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)
20. Qu, L., Fu, K., Wang, M., Song, Z., et al.: The rise of ai language pathologists: Exploring two-level prompt learning for few-shot weakly-supervised whole slide image classification. Advances in Neural Information Processing Systems **36** (2024)
21. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PMLR (2021)
22. 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)
23. Tang, W., Huang, S., Zhang, X., Zhou, F., Zhang, Y., Liu, B.: Multiple instance learning framework with masked hard instance mining for whole slide image classification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4078–4087 (2023)
24. Wang, Y., Chao, W.L., Weinberger, K.Q., Van Der Maaten, L.: Simpleshot: Revisiting nearest-neighbor classification for few-shot learning. arXiv preprint arXiv:1911.04623 (2019)
25. Xu, C., Qi, S., Feng, J., Xia, S., Kang, Y., Yao, Y., Qian, W.: Dct-mil: deep cnn transferred multiple instance learning for copd identification using ct images. Physics in Medicine & Biology **65**(14), 145011 (2020)
26. Yang, J., Chen, H., Yan, J., Chen, X., Yao, J.: Towards better understanding and better generalization of few-shot classification in histology images with contrastive learning. arXiv preprint arXiv:2202.09059 (2022)
27. Zhang, Q., Li, Y., Xue, C., Li, X.: Morphology-inspired unsupervised gland segmentation via selective semantic grouping. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 281–291. Springer (2023)
28. Zhang, Y., Sun, Y., Li, H., Zheng, S., Zhu, C., Yang, L.: Benchmarking the robustness of deep neural networks to common corruptions in digital pathology. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 242–252. Springer (2022)