

Learnable Gaussian Prototype Guided Ordinal Learning for Fine-Grained WSI Grading

Yingjiao Deng¹, Qing Zhang¹, Hui Tang³, Chunhua Zhou², Lili Gao², Jiansheng Wang¹, Li Sun¹, Yan Wang¹, Duowu Zou², Hongkai Xiong¹, and Qingli Li^{1*}

¹ Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai, 200241, China

² Department of Gastroenterology, Rujin Hospital, School of Medicine, Shanghai JiaoTong University, Shanghai, China

³ The First People Hospital of Yunnan Province, Yunnan, China

Abstract. Advances in digital pathology have enabled large-scale automated grading through deep learning-based analysis of whole slide images (WSIs), which is essential for cancer diagnosis and prognosis. However, existing WSI classification methods typically treat grades as independent categories, ignoring the ordinal and progressive nature of their relationships. Overlooking this intrinsic relationship leads to unstable boundaries and inconsistent predictions. Although ordinal regression and classification models have been explored to model the label continuity, they primarily enforce output orderings instead of learning continuous feature representations, leading to discrete embeddings that ignore inter-grade similarity. To address these problems, we propose GPO-MIL, an ordinal multiple-instance learning framework that constructs a severity-aligned Gaussian prototype space to model disease progression. A global set of Gaussian prototypes is learned and reordered by their correlation with slide grades, forming an interpretable benign-to-severe axis that guides responsibility-based prototype attention. An ordinal-aware optimization further stabilizes this progression structure and enhances grading consistency. Experiments on our in-house pancreatic cytology dataset and two public histopathology grading datasets demonstrate consistent gains in grading accuracy and calibration, along with interpretable prototypes that align with pathological progression.

Keywords: Multiple Instance Learning · Ordinal classification · Pathology · Whole Slide Image

1 Introduction

Pancreatic cancer is one of the most lethal malignancies, and reliable early cytological assessment is essential for treatment planning [6]. Liquid-based cytology (LBC), obtained through endoscopic ultrasound-guided aspiration, offers a minimally invasive diagnostic tool in which pathologists assign grades from benign to

* Corresponding author. Email: qlli@cs.ecnu.edu.cn

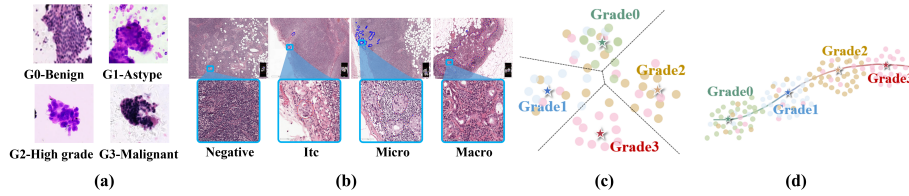


Fig. 1. (a) LBC patch examples across four grades. (b) Camelyon17 visualizations of metastasis severity. (c) Feature space without ordinal structure. (d) Severity-aligned progression feature space.

malignant based on nuclear atypia and architectural disorder [13]. As illustrated in Fig 1 (a), cytological atypia evolves gradually along a biological progression trajectory, and the boundaries between adjacent grades are often subtle, especially in pancreatic LBC, where malignant cells may be sparse. These factors make manual grading labor-intensive and susceptible to variability, underscoring the need for automated WSI-level grading that can capture the underlying progression of cytological changes.

Multiple instance learning (MIL) [20] has become the dominant paradigm for weakly supervised WSI grading. To enhance pathology grading, recent MIL research has incorporated prototype-driven [12], complexity-calibrated morphological modeling [22], and graph-based fusion approaches [19]. These methods strengthen feature aggregation and enhance morphological representation for histopathology grading. However, they treat grading as a conventional multi-class classification problem, implicitly assuming discrete boundaries between disease stages, as illustrated in Fig. 1 (c). Such formulations overlook a key medical prior: morphological progression is gradual rather than abrupt, as also observed in histopathology, as shown in Fig. 1 (b). Ignoring this continuity prevents the model from capturing inter-grade proximity and yields less stable predictions near ambiguous boundaries.

Ordinal classification (OC) methods offer a principled way to model ordered labels and have shown strong performance in age estimation [3, 16, 4], retinal disease grading [23], and inflammatory severity prediction [15, 18]. Current OC methods, such as margin-based [14] and ranking-based [15] variants, enforce ordinal relationships mainly through the loss function and depend on image-level supervision. As a result, they do not explicitly capture a continuous progression structure in the learned representations. Moreover, they assume homogeneous feature sources and cannot directly handle the strong intra-slide heterogeneity inherent in WSI-based analysis. Thus, ordinal relationships have not been effectively incorporated into prototype-guided representation learning for WSI grading, leaving a critical gap in modeling disease progression.

To address this limitation, we introduce **Gaussian Prototype-guided Ordinal Multiple Instance Learning (GPO-MIL)**, which explicitly models lesion progression in the MIL framework. A global Gaussian mixture proto-

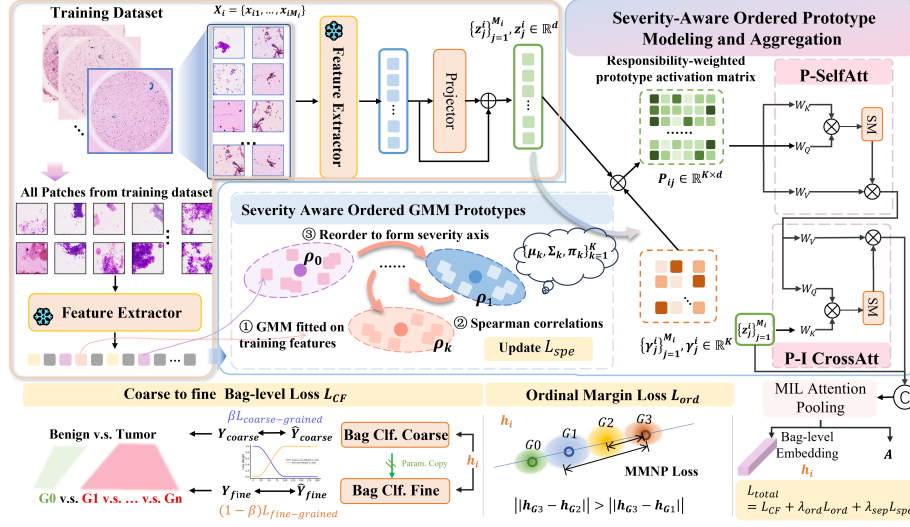


Fig. 2. Overview of the proposed GPO-MIL framework. A global Gaussian prototype space is built and reordered into a severity-aligned axis. Instance responsibilities guide prototype-conditioned attention, and an ordinal-aware optimization enforces progression consistency and discriminative grading.

type space is constructed and reordered by correlation with slide-level grades, forming a severity-aligned axis. Responsibility-guided prototype attention then conditions instance aggregation on progression semantics, highlighting grade-defining regions. Additionally, an ordinal-aware optimization strategy that incorporates prototype separation, monotonicity constraints, and hierarchical supervision helps training stabilize and enhance discriminability. Experiments on in-house pancreatic LBC, CAMELYON17, and PANDA demonstrate consistent gains over prior MIL methods while offering an interpretable progression axis that links latent prototypes to disease severity.

2 Method

2.1 Problem Formulation and Overall Framework

Given the training dataset denoted as $D = \{(X_i, y_i)\}_{i=1}^N$, where X_i denotes the i th whole slide image and $y_i \in \{0, 1, \dots, C-1\}$ is its corresponding grading label, satisfying $0 < 1 \dots < C-1$ in diagnostic severity. Following standard multiple instance learning, each WSI is divided into M_i non-overlapping image patches $X_i = \{x_{ij}\}_{j=1}^{M_i}$. A pretrained feature extractor $f_\theta(\cdot)$ extracts instance embeddings: $z_{ij} = f_\theta(x_{ij}) \in \mathbb{R}^d$, $Z_i = \{z_{ij}\}_{j=1}^{M_i}$. Conventional MIL methods aggregate instance embeddings Z_i into a bag-level representation s_i and perform a categorical softmax classification. However, such a formulation treats grades as independent classes and ignores the progressive nature of pathological grading.

In practice, adjacent grades share overlapping morphological patterns, and higher grades progressively encompass more aggressive lesion subtypes. Thus, modeling an ordered latent severity structure is more appropriate than assuming independent class boundaries. Motivated by this, we introduce **GPO-MIL**, a Grading-oriented Prototype Ordering MIL framework that constructs a severity-aligned prototype space and integrates it into MIL aggregation through ordinal-consistent optimization. The framework is illustrated in Fig 2. The framework enables the model to capture continuous disease progression patterns while maintaining discriminative power between adjacent grades.

2.2 Severity-Aware Ordered Prototype Modeling and Aggregation

Severity-Aware Ordered Prototype Space Construction To capture heterogeneous lesion patterns and introduce a severity-consistent inductive bias, we model all instance embeddings using a K-component Gaussian Mixture Model (GMM), which is defined as:

$$p(z) = \sum_{k=1}^K \pi_k \mathcal{N}(z|\mu_k, \Sigma_k), \quad (1)$$

where π_k are the mixture weights ($\sum_k \pi_k = 1$) and (μ_k, Σ_k) denote prototype means and covariances. Diagonal covariances are used for stability in high-dimensional embedding space.

The GMM is fitted on training instance features before MIL optimization. Posterior responsibilities for each instance are computed as:

$$r_{ij,k} = p(k|z_{ij}) = \frac{\pi_k \mathcal{N}(z_{ij}|\mu_k, \Sigma_k)}{\sum_{t=1}^K \pi_t \mathcal{N}(z_{ij}|\mu_t, \Sigma_t)}. \quad (2)$$

Slide-level prototype composition is obtained by averaging: $w_{i,k} = \frac{1}{M_i} \sum_{j=1}^{M_i} r_{ij,k}$, representing the latent pathological mixture for slide i .

Since mixture components are permutation-invariant, prototype indices initially lack intrinsic ordinal structure. To align prototypes with ordinal disease severity, we compute Spearman correlations $\rho_k = \text{corr}(\{w_{i,k}\}_{i=1}^N, \{y_i\}_{i=1}^N)$, and reorder mixture components in ascending ρ_k to form a benign-to-severe prototype axis. This re-indexing imposes semantic ordering without changing the underlying probabilistic model.

During end-to-end MIL training, responsibilities are computed in a differentiable form:

$$r_{ij,k} = \text{softmax}_k \left(\frac{\log \pi_k + \log \mathcal{N}(x_i|\mu_k, \Sigma_k)}{\tau} \right), \quad (3)$$

where τ controls assignment sharpness. To ensure instability, all GMM parameters remain frozen for the first e_{warm} epochs; afterwards, only prototype means $\{\mu_k\}$ are updated, allowing mild task-specific refinement while preserving the established ordinal alignment.

Responsibility-Guided Instance Aggregation: The ordered prototypes are injected into the MIL aggregator via responsibility-weighted decompositions:

$$P_{ij,k} = r_{ij,k} z_{ij}, \quad (4)$$

producing prototype-aligned activations $P_{ij} \in \mathbb{R}^{K \times d}$.

To propagate prototype semantics and capture interactions among latent pathological subtypes, we introduce a **two-stage prototype-conditioned attention module**. (1) Prototype Self-Attention (P-SelfAtt) is applied along the prototype dimension: $H_{ij} = \text{SelfAttn}(P_{ij})$ enabling the model to encode co-occurrence structures among latent lesion components and capture progressive patterns consistent with the ordinal prototype axis. (2) Prototype-to-Instance Cross-Attention (P-I CrossAtt) conditions the original instance embedding on the learned prototype semantics, enabling prototype-guided refinement of local features: $\hat{z}_{ij} = \text{CrossAttn}(H_{ij}, z_{ij})$, where H_{ij} serves as queries and values. This step enriches instance representation with severity-aware structural context.

The enhanced embedding $\hat{z}_{ij} = \text{Concat}(\hat{z}_{ij}, z_{ij})$ is finally aggregated through attention-based MIL pooling to obtain the WSI representation h_i . By grounding aggregation explicitly on the ordered prototype axis, the model selectively emphasizes grade-defining morphological cues while suppressing background interference.

2.3 Ordinal-Aware Optimization

The final objective combines hierarchical grading, ordinal metric learning, and prototype regularization:

$$L_{\text{total}} = L_{\text{CF}} + \lambda_{\text{ord}} L_{\text{ord}} + \lambda_{\text{sep}} L_{\text{sep}}. \quad (5)$$

Coarse-to-Fine Hierarchical Grading Loss. Since fine-grained boundaries are subtle, directly optimizing fine-grained objectives may lead to unstable convergence. We adopt a smooth coarse-to-fine schedule:

$$L_{\text{CF}} = \beta(e) L_{\text{coarse}} + (1 - \beta(e)) L_{\text{fine}}, \quad (6)$$

where $\beta(e)$ gradually decreases after warm-up, enabling progressive refinement from coarse grouping to fine-grained separation.

Ordinal Metric-Learning via Multi-Margin N-Pair Loss. To enforce a monotonic embedding structure at the bag-level feature consistent with grade ranks, we adopt the Multi-Margin N-Pair (MMNP) loss [14]:

$$L_{\text{ord}} = \sum_{(h_j, y_j) \in \mathcal{H}^+} \sum_{(h_k, y_k) \in \mathcal{H}^-} \max(0, m_{y, y_k} + \cos(h, h_k) - \cos(h, h_j)) \quad (7)$$

where $m_{y, y_k} = m_1 + \dots m_{y_k - y}$ is the cumulative ordinal margin between the two grades. Following prior work, MMNP is optimized during the first 50 epochs, after which the ordinal margins are frozen to preserve the learned grade ordering.

Table 1. Comparison with state-of-the-art MIL methods on PANDAS, Camelyon17, and LBC. Results are reported as mean_{std} over five-fold. Best results are in bold.

Method	PANDAS			CAMELYON17			LBC		
	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1
CLAM-SB [11]	92.64 _{0.33}	71.20 _{0.92}	65.17 _{1.55}	87.37 _{4.35}	76.60 _{8.09}	60.00 _{9.69}	80.80 _{2.34}	62.47 _{2.62}	50.00 _{3.53}
CLAM-MB [11]	92.79 _{0.38}	71.12 _{0.89}	65.12 _{1.28}	90.12 _{2.13}	83.40 _{2.87}	69.61 _{4.69}	80.95 _{2.79}	62.05 _{2.79}	48.68 _{4.99}
TransMIL [17]	91.17 _{0.49}	67.11 _{1.70}	59.57 _{1.64}	89.04 _{1.31}	78.80 _{5.71}	60.37 _{4.55}	75.56 _{2.48}	52.34 _{4.40}	42.62 _{2.13}
ABMIL [7]	92.34 _{0.36}	69.82 _{0.95}	62.71 _{1.78}	88.38 _{2.63}	79.00 _{5.22}	62.66 _{2.99}	78.85 _{2.78}	58.52 _{1.59}	51.02 _{3.49}
S4MIL [5]	91.40 _{0.87}	67.72 _{1.84}	61.70 _{3.00}	92.03 _{1.72}	87.60 _{3.26}	70.88 _{5.29}	80.01 _{2.35}	62.88 _{5.53}	51.89 _{6.95}
MambaMIL [21]	91.51 _{0.63}	68.46 _{1.73}	62.25 _{2.33}	86.20 _{2.81}	77.80 _{3.87}	62.84 _{6.08}	79.36 _{5.01}	60.94 _{6.44}	50.07 _{8.18}
DTFD-AFS [24]	89.48 _{0.50}	63.91 _{1.49}	51.33 _{3.51}	88.30 _{2.35}	83.80 _{3.31}	61.44 _{4.32}	81.72 _{1.59}	62.20 _{3.24}	44.65 _{4.90}
DTFD-MMS [24]	85.04 _{1.94}	55.00 _{2.74}	46.04 _{2.46}	71.61 _{6.99}	72.8 _{8.75}	35.86 _{9.30}	68.15 _{2.72}	40.43 _{2.37}	22.93 _{2.68}
HMIL [8]	92.17 _{0.50}	70.26 _{1.35}	64.06 _{1.99}	86.58 _{6.38}	86.11 _{4.55}	61.30 _{6.69}	78.93 _{1.23}	60.02 _{4.78}	51.81 _{5.35}
DGRMIL [25]	90.91 _{0.75}	67.89 _{1.54}	61.33 _{1.77}	86.11 _{6.72}	80.0 _{7.07}	54.99 _{9.30}	74.32 _{3.58}	60.27 _{3.89}	45.10 _{4.80}
Ours	94.03 _{0.42}	73.20 _{1.43}	66.48 _{1.38}	94.17 _{2.32}	88.31 _{4.20}	71.20 _{7.98}	82.69 _{2.31}	67.20 _{3.87}	54.20 _{4.27}

Prototype Separation Regularization. To avoid prototype collapse and maintain distinct pathological semantics, we impose cosine-based separation loss:

$$L_{\text{sep}} = \frac{1}{K(K-1)} \sum_{k \neq j} \max(0, \cos(\mu_k, \mu_j) - \tau), \quad (8)$$

where τ is the maximum allowable similarity.

3 Experiments

3.1 Dataset and Implementation Details

Dataset. In this work, we evaluated our method on our in-house pancreatic LBC dataset (LBC) and two public pathology dataset CAMELYON17 [1] and PANDA [2]. **LBC** contains H&E-stained liquid-based cytology WSIs collected in-house under approved institutional ethics protocols. Each case was reviewed by two senior cytopathologists, with disagreements resolved by consensus. The dataset includes four progressive cytological grades: G1 Benign (327), G2 Atypia (202), G3 High grade (131), and G4 Malignant (298). **CAMELYON17** comprises 500 lymph-node WSIs with the standard four-grade setting: negative (318), ITC (36), micro-metastasis (59), and macro-metastasis (87). **PANDA** includes 10,616 prostate cancer WSIs with ISUP grades 0–5, distributed as 2892/2666/1343/1242/1249/1224 cases.

Evaluation Metrics. We evaluate grading performance using three standard metrics: accuracy (ACC), the area under the ROC curve (AUC), and macro F1-score (F1).

Implementation Details. For CAMELYON17 and PANDA, WSIs are preprocessed following the CLAM [11] thresholding pipeline, while LBC slides adopt blank-region removal due to cytology-specific sparsity. All slides are tiled into non-overlapping patches of size 256×256 at level 0 for PANDA/LBC and level 1 for CAMELYON17. Patch embeddings are extracted using a pretrained CONCH [10] image encoder, yielding 512-dimensional features. Experiments follow 5-fold

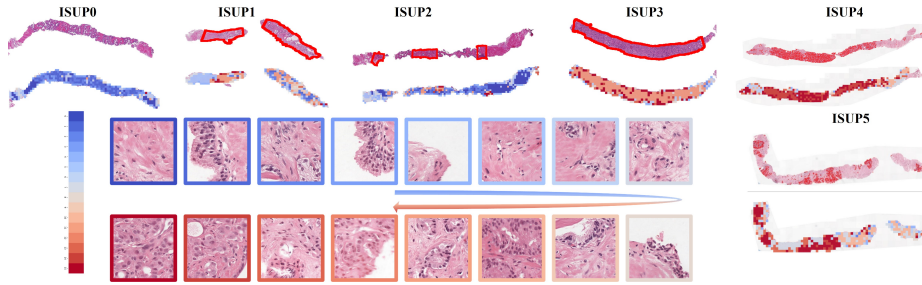


Fig. 3. Visualization on the PANDA dataset, showing the WSI, annotation, zoom-in images, and the corresponding prototype-activation map, where severity-ordered prototypes reveal a smooth transition from low- to high-grade patterns.

cross-validation, with splits of 6:2:2 for Camelyon17 and 7:1.5:1.5 for PANDA and LBC. All experiments are conducted on an NVIDIA GeForce RTX 3090.

To ensure a fair comparison, all MIL methods are trained for 200 epochs with Adam ($lr = 1 \times 10^{-4}$) with early-stop. For our GPO-MIL, we adopt the gated attention pooling module from ABMIL [7] as the MIL pooling operator. The initialization of ordered GMM using diagonal-covariance, implemented by scikit-learn, max iter 200, init with k-means and reg covar is set to $1e-6$. For the loss function, the λ_{sep} and λ_{ord} are set to 0.1, and 0.3. The number of prototypes is set to $K = 16$ for PANDA and $K = 8$ for CAMELYON17 and LBC.

3.2 Main Results

Table 1 reports the performance of GPO-MIL against state-of-the-art MIL models. Across all dataset, our method achieves the highest AUC, ACC, and F1 scores. The improvements are consistent on both tissue-based histopathology and cytology slides, demonstrating that modeling grading as an ordered prototype axis generalizes well across heterogeneous pathology modalities. Moreover, GPO-MIL exhibits comparable or lower performance variance than existing methods, indicating that the integration of ordinal structure and prototype modeling does not introduce instability during optimization. Overall, these results confirm that explicitly aligning latent instance distributions with disease severity yields more reliable and discriminative slide-level predictions compared with purely categorical MIL approaches.

We visualize ordered Gaussian prototypes by mapping each patch to the prototype with highest responsibility $k^* = \arg \max_k r_{ij,k}$ and recoloring the slide according to prototype severity. On PANDA (Fig. 3), low-grade slides are dominated by low-severity prototypes (blue), while high-grade slides show extensive activation of severe prototypes (red). On CAMELYON17 (Fig. 4), subtle lesions such as ITC or Micro-metastasis are highlighted by higher-severity prototypes, whereas benign lymphoid tissue consistently maps to low-severity ones. Zoomed visualizations confirm that severe prototypes correspond to malignant glands or

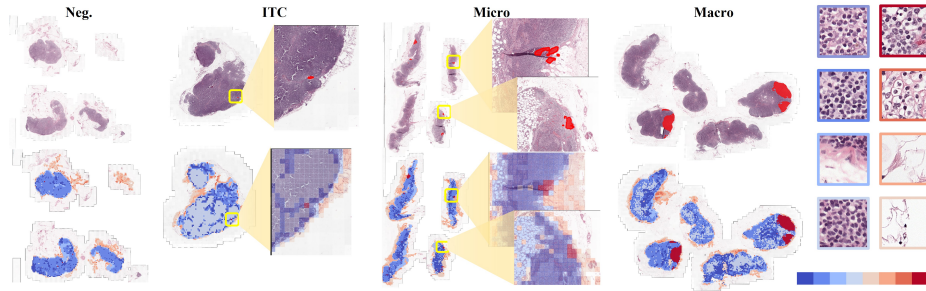


Fig. 4. Visualization on the CAMELYON17 dataset, including the WSI, annotation, zoom-in images, and the associated prototype-activation map, while severity-ordered prototypes illustrate a continuous progression from benign to metastatic patterns.

Table 2. Evaluation of the main components and loss functions.

		PANDAS			Camelyon17			LBC		
		AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1
Main Module	ABMIL-baseline	92.34	69.82	62.71	88.38	79.00	62.66	78.85	58.52	51.02
	w.o. sev	92.89	71.02	63.60	90.28	83.19	67.23	80.38	64.29	52.91
	w.o. GMM-up	93.03	72.27	64.79	92.19	85.49	69.38	79.31	64.35	52.48
	w.o. SA	92.02	69.29	60.76	93.08	88.05	70.19	81.03	66.32	53.17
	w.o. CA	92.18	70.96	64.07	93.21	87.13	70.54	81.92	66.41	53.94
Loss	w.o. L_{ord}	92.87	72.28	64.99	92.48	82.09	70.67	80.92	65.86	53.19
	w.o. L_{sep}	89.21	66.23	56.27	91.39	83.25	69.38	79.25	64.03	52.17
	L_{fine}	93.03	71.69	65.23	93.95	86.76	70.25	81.46	66.03	53.97
	$L_{krank}[15]$	93.76	72.49	64.97	92.49	84.12	70.82	81.36	65.21	53.08
	$L_{POE}[9]$	93.18	71.93	65.89	92.91	83.21	70.94	81.05	65.95	53.26
Ours		94.03	73.2	66.48	94.17	88.31	71.23	82.69	67.2	54.2

small metastatic clusters, demonstrating that the learned prototype axis captures meaningful progression cues.

3.3 Ablation Study

Table 2 summarizes the ablation results. Removing severity ordering (“w.o. sev”) or disabling prototype refinement (“w.o. GMM-up”) consistently reduces performance, showing the necessity of a severity-aligned and adaptable prototype space. Eliminating self-attention (“w.o. SA”) or cross-attention (“w.o. CA”) also degrades results, confirming their roles in modeling latent subtype interactions and injecting prototype semantics into instance features. For loss functions, discarding L_{ord} weakens monotonic grade separation, while removing L_{sep} leads to prototype collapse and the largest drop. Alternative ordinal losses $L_{krank}[15]$ and $L_{POE}[9]$ offer only modest gains and still fall short of the MMNP loss.

4 Conclusion

In this paper, we introduced GPO-MIL, a severity-aligned prototype-guided MIL framework for ordinal histopathological grading. By constructing an ordered Gaussian prototype space and coupling it with responsibility-aware attention and ordinal optimization, the method effectively captures progressive lesion patterns overlooked by prior MIL approaches. Experiments on LBC and public pathology datasets show consistent improvements and enhanced interpretability, validating the effectiveness of severity-aligned prototypes for reliable grading.

Acknowledgments. This work is supported by National Natural Science Foundation of China (Grant No. 62475072, 42527804, 42530110), Fundamental Research Funds for the Central Universities, the Science and Technology Commission of Shanghai Municipality (Grant No. 25xtcx00600, 22DZ2229004), China Postdoctoral Science Foundation (Grant No. 2025M783508).

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

References

1. Bándi, P., Geessink, O., Manson, Q., Van Dijk, M., Balkenhol, M., et al.: From Detection of Individual Metastases to Classification of Lymph Node Status at the Patient Level: The CAMELYON17 Challenge. *IEEE Transactions on Medical Imaging* **38**(2), 550–560 (Feb 2019). <https://doi.org/10.1109/TMI.2018.2867350>
2. Bulten, W., Kartasalo, K., Chen, P.H.C., Ström, P., Pinckaers, H., Nagpal, K., Cai, Y., et al.: Artificial intelligence for diagnosis and Gleason grading of prostate cancer: The PANDA challenge. *Nature Medicine* **28**(1), 154–163 (Jan 2022). <https://doi.org/10.1038/s41591-021-01620-2>
3. Cao, W., Mirjalili, V., Raschka, S.: Rank consistent ordinal regression for neural networks with application to age estimation. *Pattern Recognition Letters* **140**, 325–331 (Dec 2020). <https://doi.org/10.1016/j.patrec.2020.11.008>
4. Corbetta, V., Beets-Tan, R., Cardoso, J.S., Silva, W.: Understanding the Impact of Client Heterogeneity on Ordinal Classification in Federated Medical Image Analysis. In: *Medical Imaging with Deep Learning* (Jan 2025)
5. Fillioux, L., Boyd, J., Vakalopoulou, M., Cournède, P.H., Christodoulidis, S.: Structured State Space Models for Multiple Instance Learning in Digital Pathology (Jun 2023). <https://doi.org/10.48550/arXiv.2306.15789>
6. Hayat, U., Croce, P.S., Saadeh, A., Desai, K., Appiah, J., Khan, S., et al.: Current and emerging treatment options for pancreatic cancer: a comprehensive review. *Journal of Clinical Medicine* **14**(4), 1129 (2025)
7. Ilse, M., Tomczak, J.M., Welling, M.: Attention-based deep multiple instance learning. *arXiv preprint arXiv:1802.04712* (2018)
8. Jin, C., Luo, L., Lin, H., Hou, J., Chen, H.: HMIL: Hierarchical Multi-Instance Learning for Fine-Grained Whole Slide Image Classification. *IEEE Transactions on Medical Imaging* **44**(4), 1796–1808 (Apr 2025). <https://doi.org/10.1109/TMI.2024.3520602>

9. Li, W., Huang, X., Lu, J., Feng, J., Zhou, J.: Learning Probabilistic Ordinal Embeddings for Uncertainty-Aware Regression. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 13891–13900. IEEE (Jun 2021). <https://doi.org/10.1109/CVPR46437.2021.01368>
10. Lu, M.Y., Chen, B., Williamson, D.F.K., Chen, R.J., Liang, I., Ding, T., Jaume, G., Odintsov, I., Le, L.P., Gerber, G., Parwani, A.V., Zhang, A., Mahmood, F.: A visual-language foundation model for computational pathology. *Nature Medicine* **30**(3), 863–874 (Mar 2024). <https://doi.org/10.1038/s41591-024-02856-4>
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. Ma, M., Luo, F., Ma, B., Liu, S., Lv, X., Huang, P.: A Multi-instance Learning Network with Prototype-instance Adversarial Contrastive for Cervix Pathology Grading. *Medical Image Analysis* **109**, 103880 (Mar 2026). <https://doi.org/10.1016/j.media.2025.103880>
13. Pan, Y., Ran, T., Zhang, X., Qin, X., Zhang, Y., Zhou, C., Zou, D.: Adequacy of EUS-guided fine-needle aspiration and fine-needle biopsy for next-generation sequencing in pancreatic malignancies: A systematic review and meta-analysis. *Endoscopic Ultrasound* **13**(6), 366–375 (2024)
14. Pitawela, D., Carneiro, G., Chen, H.T.: CLOC: Contrastive Learning for Ordinal Classification with Multi-Margin N-pair Loss. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15538–15548 (2025)
15. Schwab, E., Cula, G.O., Standish, K., Yip, S.S.F., Stojmirovic, A., Ghanem, L., Chehoud, C.: Automatic Estimation of Ulcerative Colitis Severity from Endoscopy Videos using Ordinal Multi-Instance Learning. <https://doi.org/10.48550/arXiv.2109.14685>
16. Shah, J., Siddiquee, M.M.R., Su, Y., Wu, T., Li, B.: Ordinal classification with distance regularization for robust brain age prediction. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 7882–7891 (January 2024)
17. 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)
18. Shiku, K., Nishimura, K., Suehiro, D., Tanaka, K., Bise, R.: Ordinal Multiple-instance Learning for Ulcerative Colitis Severity Estimation with Selective Aggregated Transformer. 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) pp. 4290–4299 (Feb 2025). <https://doi.org/10.1109/WACV61041.2025.00421>
19. Tu, G., Li, W., Lin, Y., Xu, Z., He, J., ontors: Gammil: A graph attention-guided multi-scale fusion multiple instance learning model for the who grading of meningioma in whole slide images. *Biomedical Signal Processing and Control* **105**, 107652 (2025)
20. Waqas, M., Ahmed, S.U., Tahir, M.A., Wu, J., Qureshi, R.: Exploring multiple instance learning (mil): A brief survey. *Expert Systems with Applications* **250**, 123893 (2024). <https://doi.org/https://doi.org/10.1016/j.eswa.2024.123893>
21. Yang, S., Wang, Y., Chen, H.: MambaMIL: Enhancing Long Sequence Modeling with Sequence Reordering in Computational Pathology (Mar 2024). <https://doi.org/10.48550/arXiv.2403.06800>
22. Yu, J., Wang, X., Fan, W., Guo, Y., Fu, J., Xu, Y.: Exploring Complexity-Calibrated morphological distribution for whole slide image classification

- and difficulty-grading. *Medical Image Analysis* **105**, 103707 (Oct 2025). <https://doi.org/10.1016/j.media.2025.103707>
23. Yu, Q., Xie, J., Nguyen, A., Zhao, H., Zhang, J., Fu, H., et al.: CLIP-DR: Textual Knowledge-Guided Diabetic Retinopathy Grading with Ranking-Aware Prompting. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, vol. 15001, pp. 667–677. Springer Nature Switzerland (2024). https://doi.org/10.1007/978-3-031-72378-0_62
 24. 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: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 18780–18790. IEEE, New Orleans, LA, USA (Jun 2022). <https://doi.org/10.1109/CVPR52688.2022.01824>
 25. Zhu, W., Chen, X., Qiu, P., Sotiras, A., Razi, A., Wang, Y.: DGR-MIL: Exploring Diverse Global Representation in Multiple Instance Learning for Whole Slide Image Classification. In: *Computer Vision – ECCV 2024*. pp. 333–351. Springer Nature Switzerland, Cham (2025). https://doi.org/10.1007/978-3-031-72920-1_19